مبادئ الإحصاء

وقل رب زدني علما
جامعة دمشق
Damascus University

منشورات جامعة دمشق
كلية الآداب والعلوم الإنسانية
قسم المكتبات والمعلومات

مبادئ الإحصاء

تأليف

الدكتور مطانيوس مخول
الأستاذ المساعد في قسم
الإحصاء التطبيقي
كلية الاقتصاد – جامعة دمشق

الدكتور عدنان غانم
الأستاذ المساعد في قسم
الإحصاء التطبيقي
كلية الاقتصاد – جامعة دمشق

1426- 1427 هـ

جامعة دمشق

2005 - 2006 م

وقل رب زدني علما
جامعة دمشق
Damascus University

فهرس المحتويات

المحتويات	الصفحة
مقدمة	**13**
الفصل الأول: ماهية علم الإحصاء	**15**
1-1 - لمحة تاريخية عن تطور علم الإحصاء	17
1-2- مفاهيم علم الإحصاء	18
1-3- شروط تطبيق الطرائق العملية على البيانات العددية	20
1-4- استعمالات الإحصاء	21
1-5- مراحل البحث الإحصائي (الدراسة الإحصائية)	26
1-6- أسئلة	28
الفصل الثاني: جمع البيانات الإحصائية	**29**
2-1- تحديد المشكلة وماهية الدراسة التحضيرية	31
2-2- تحديد إطار الدراسة	32
2-2-1- تحديد الوحدة الإحصائية	32
2-2-1-1- أنواع الوحدات الإحصائية	33
2-2-1-1-1- الوحدات الإحصائية البسيطة	33
2-2-1-1-2- الوحدات الإحصائية المركبة	34
2-2-2- تحديد المجتمع الإحصائي (تعريفه ، الاعتبارات ، أنواعه)	34
2-3- أهمية مرحلة جمع البيانات الإحصائية	34
2-4- مصادر جمع البيانات الإحصائية (تاريخية- داخلية- ميدانية)	35
2-4-1- الحصر الشامل	35
2-4-2- العينات	35
2-4-2-1- أنواع العينات (العينات العشوائية وغير العشوائية)	37
2-4-2-2- أخطاء العينات وكيفية معالجتها	47

2-5- تصنيف البيانات حسب مصادرها 48
2-5-1- بيانات ثانوية 48
2-5-2- بيانات أولية 49
2-6- طرائق الحصول على البيانات الإحصائية 49
2-6-1- المقابلة الشخصية 49
2-6-2- العدادون 49
2-6-3- الاستمارة أو الكشف البريدي والشروط الواجب توافرها في الاستمارة 49
2-6-4- الهاتف والانترنت 51
2-7- أسئلة 52
الفصل الثالث: عرض البيانات الإحصائية وتبويبها **55**
3-1- العرض الجدولي 57
3-1-1- التبويب 57
3-1-2- ماهية الجداول الإحصائية وشروطها 58
3-1-3- أنواع الجداول الإحصائية 60
3-1-4- كيفية ترتيب البيانات الإحصائية وأساليب ترتيبها 61
3-1-5- الجداول التكرارية 62
3-1-5-1 ماهية الجداول التكرارية 62
3-1-5-2- كيفية إعداد الجداول التكرارية (مزاياها ومساؤها) 63
3-1-5-3- خصائص وصفات التوزيعات التكرارية 69
3-2- العرض البياني 71
3-2-1- ماهية العرض البياني وغايته 71
3-2-2- أنواع العرض البياني 72
3-2-2-1- المصورات الإحصائية والخرائط الجغرافية 72

3-2-2-2- الأعمدة البيانية (البسيطة ، المقسّمة) 72
3-2-2-3- رسوم المساحات والأحجام البيانية " الدوائر الإحصائية " 75
3-2-2-4- رسوم المنحنيات البيانية 76
3-2-2-4-1- منحنيات التكرار وأنواعها 77
3-2-2-4-2- منحنيات السلاسل الزمنية 79
3-3- أسئلة وتمارين 80
الفصل الرابع: مقاييس التحليل الإحصائي (لظاهرة واحدة) ـــــ 85
4-1 - النسب والمعدلات 87
4-1-1- مفهوم النسب والمعدلات 87
4-1-2- استعمالات النسب والمعدلات 89
4-2 -مقاييس النزعة المركزية 91
4-2-1- الوسط الحسابي 92
4-2-1-1- مفهوم الوسط الحسابي 92
4-2-1-2- كيفية حساب الوسط الحسابي[بيانات غير مبوبة وبيانات مبوبة] 92
4-2-1-3- خواص الوسط الحسابي 96
4-2-1-4- مزايا الوسط الحسابي ومساوئه 98
4-2-2- الوسيط 99
4-2-2-1- تعريف الوسيط 99
4-2-2-2- كيفية إيجاد الوسيط لبيانات غير مبوبة وبيانات مبوبة 99
4-2-2-3- مزايا الوسيط ومساوئه 103
4-2-3- المنوال 104
4-2-3-1- مفهوم المنوال وكيفية حسابه 104
4-2-3-2- مزايا المنوال ومساوئه 107
4-3 - العلاقة بين المتوسطات الثلاث " علاقة بيرسون " 108
4-4 - مقاييس التشتت 110
4-4-1- المدى 112

4-4-1-1- مفهومه وكيفية حسابه 112
4-4-1-2- استعمالاته 113
4-4-2- الانحراف المتوسط 113
4-4-2-1- تعريفه وكيفية حسابه 113
4-4-2-2- مزايا الانحراف المتوسط ومساوئه 115
4-4-3- الانحراف الربيعي 116
4-4-3-1- مفهومه وكيفية إيجاده 116
4-4-3-2- المزايا والمساوئ 120
4-4-4- الانحراف المعياري 120
4-4-4-1- ماهية الانحراف المعياري وكيفية حسابه 120
4-4-4-2- خواص الانحراف المعياري واستخداماته 125
4-4-4-3- مزايا الانحراف المعياري وعيوبه 131
4-5- أسئلة وتمارين 131
الفصل الخامس: مقاييس التحليل الإحصائي (لظاهرتين) ـــــ الارتباط والانحدار **135**
5-1- مقدمة في تحليل الارتباط 137
5-2- تحليل الانحدار 138
5-2-1- تحديد طبيعة المتغيرات 138
5-2-2- شكل الانتشار 139
5-2-3- مقاييس الارتباط وأنواعه 140
5-3- الانحدار الخطي 141
5-3-1- معادلة الانحدار الخطي (خواصها ، استعمالاتها) 141
5-3-2 الخطأ المعياري للتقدير 148
5-3-3- تحليل التباينات 150
5-3-4- اختبار معنوية معامل الارتباط 154
5-4- تحليل الارتباط للبيانات النوعية 156

5-4-1- معامل الاقتران 157
5-4-2- معامل ارتباط الرتب " سبيرمان " 159
5-5- مسائل وتمارين 161
الفصل السادس : العينات وتوزيع المعاينة ـــــــــــــــــ 165
6-1- أهمية العينات ومجالات تطبيقها 167
6-2- أنواع العينات 171
6-2-1- العينات العشوائية (شروطها وأنواعها) 171
6-2-2- العينات غير العشوائية وأنواعها 172
6-3- توزيعات المعاينة وخصائصها 173
6-3-1- ماهية توزيع المعاينة 173
6-3-2- خصائص توزيع العينة 175
- توزيع معاينة الأوساط الحسابية 175
- توزيع معاينة الانحرافات المعيارية 178
- توزيع معاينة النسب المئوية 179
- توزيعات معاينة الفروق 181
6-4- تحديد حجم العينة 182
6-4-1- تحديد حجم العينة للمتغيرات الكمّية 182
6-4-2- تحديد حجم العينة للمتغيرات النوعية 184
6-5- مسائل وتمارين غير محلولة 187
الفصل السابع: الاستدلال الإحصائي للعينات كبيرة الحجم ـــــ 189
- مقدمة 191
7-1- التقديرات الإحصائية وأنواعها 191
7-1-1- التقدير النقطي 191
7-1-2- التقدير المجالي 192
7-1-2-1- تقدير مجال الثقة للأوساط الحسابية 193
7-1-2-2- تقدير مجال الثقة للانحرافات المعيارية 195
7-1-2-3- تقدير مجال الثقة للنسب المئوية 196
7-1-2-4- تقدير مجال الثقة للفروق 197

7-2- الاختبارات الإحصائية و أنواعها 201
7-2-1- اختبارات تتعلق بمجتمع واحد 207
7-2-2- اختبارات تتعلق بمجتمعين 214
7-3- مسائل وتمارين غير محلولة 219
الفصل الثامن: الاستدلال الإحصائي للعينات صغيرة الحجم ـــــ 223
8-1- توزيع $t-$ ستودنت 225
8-2- اختبار الفرضيات 226
8-2-1- اختبار الفرضيات للمتوسط (μ) [حالة مجتمع واحد] 226
8-2-2- اختبار الفرضيات للمتوسط (μ) [حالة مجتمعين] 229
8-2-3- تقدير متوسط المجتمع (μ) 232
8-2-4-اختبار الفرق بين متوسطين (حالة العينات غير المستقلة) 233
8-3- توزيع كاي مربع χ^2 235
8-3-1- اختبار الفرضيات للتباين σ^2 238
8-3-2- تقدير تباين المجتمع σ^2 239
8-4 توزيع فيشر (F) واختبار نسبة تباينين مجتمعين طبيعيين 242
8-5- أسئلة وتمارين 248
الفصل التاسع : الإحصاء اللامعلمي ـــــــــــــــــ 251
9-1- اختبار حسن المطابقة (كاي مربع χ^2) 254
9-2- ملاحظات حول اختبار حسن المطابقة 261
9-3- اختبار استقلال الجداول (2×2) و (2×3) و (3×3) 265
9-4- تصحيح ياتس 274
9-5- أسئلة وتمارين 278
الفصل العاشر: مبادئ في الاحتمالات ـــــــــــــــــ 285
10-1- مفاهيم أولية (الاحتمال ، التجربة ، الحادثة) 287
10-2- خصائص الاحتمال 288

10-2-1- الاحتمال النظري 289
10-2-2- الاحتمال التجريبي 290
10-3- أنوع الحوادث 291
10-3-1- الحادث الأكيد ، الحادث المستحيل ، الحادث الممكن 291
10-3-2- الحوادث المتنافية وغير المتنافية 292
10-3-3- الحوادث المستقلة 292
10-3-4- الحوادث الشرطية 292
10-4- جمع الاحتمالات وضربها 293
10-4-1- جمع الاحتمالات 293
10-4-2- ضرب الاحتمالات 295
10-4-3- الاحتمال الشرطي 296
10-4-4 نظرية بايز 298
10-4-5- شجرة الاحتمال " الشجرة البيانية " 300
10-5 مسائل وتمارين 304
- بعض الجداول الإحصائية ـــــــــــــــــــــــــــــــ 313
- المصطلحات الإحصائية ـــــــــــــــــــــــــــــــ 231
- قائمة المراجع ـــــــــــــــــــــــــــــــ 333

وقل رب زدني علما
جامعة دمشق
Damascus University

المقدمة

تساعد عملية البحث العلمي على فهم كيفية تفاعل الأحداث التي تربطها علاقات متبادلة ، فهي تمدنا في أغلب الأحيان بالوسائل التي تجعلنا قادرين على التكيف مع المحيط بنا ، لأن عملية البحث العلمي تشتمل على كافة طرائق جمع البيانات الرقمية للوصول إلى استنتاجات تتعلق بخصائص هذه البيانات واختيار ما يلائمها من أساليب إحصائية تتوافق معها ، لأن هذه الأساليب تنظم عملية البحث للبيانات الرقمية وكيفية تطبيقها للوصول إلى نتائج دقيقة .

إذ تُعد معرفة كيفية إجراء البحوث بالنسبة للمكتبيين وخبراء الأرشيف ، وأخصائي المعلومات على درجة كبيرة من الأهمية ، وذلك لاعتبارات عدة، منها على سبيل المثال لا الحصر : التطورات الكثيرة التي حدثت في العالم، والتعقيدات التي نشأت عن تداخل المفاهيم ، وتشابك الأنظمة ، مما أدى إلى ظهور أنماط من المؤسسات ذات طبيعة خاصة ، بالإضافة إلى تغير دور المكتبي من دور الوسيط ما بين المعلومات المسجلة ومستخدميها من خلال إيجاد مكان المعلومات وتحديدها، إلى تحليل تلك المعلومات وتقييمها وتفسيرها لصالح المستفيد .

يفتقد علم المكتبات والمعلومات الإطار النظري الرياضي المنظور الذي يساعد على تفسير أغراض وأنشطة هذا المجال التخصصي وشرحها ، لذا يعمد الكثير من المكتبين على تطوير النظريات والقوانين المتعلقة بسلوكيات هذا المجال التخصصي ، سواءً من حيث الكم فيها أم من حيث الكيف ، والاستفادة من تطور مناهج البحث العلمي في التخصصات الأخرى وذات الصلة بها ، لأن فهم نتائج البحوث ذات الصلة وتطبيقها في أي مجال معرفي، يؤدي إلى تطوير الأفكار القابلة للبحث ، وتكفي لتغطية الجوانب البحثية للتخصص كلها.

لقد بات معروفاً ، بأن تدريس الإحصاء لم يعد مقتصراً على تخصص بعينه، بل أصبح يشمل كافة التخصصات ، اجتماعية كانت ، أم اقتصادية ، أم تطبيقية ، أم إنسانية ، ... الخ ، لأن الإحصاء أصبح بالنسبة لهذه التخصصات الطابع المميزّ للمنهج العلمي المتبع فيها ، وبخاصة تخصص المكتبات والمعلومات الذي أصبح بأمس الحاجة إلى الإحصاء لاستخدامه في مجال البحوث التي تؤدي إلى معارف جديدة ، وفي تحليل البيانات التي تفسر الظواهر المعلوماتية ، لأن المكتبي هو المستخدم أنواع البيانات جميعها والتي يستفيد من دراستها في تطوير كفاءته المهنية ورفعها لتقديم خدمات أفضل سواءً للمؤسسة التي يعمل بها ، أم للمجتمع .

نأمل في فصول هذا الكتاب ، أن نكون قدمنا مفهوماً واضحاً لعلم الإحصاء والتحليل الإحصائي ، ومدى فائدتهما لتخصص المكتبات والمعلومات ، إذ تجلى هدفنا في تزويد طلاب قسم المكتبات والمعلومات ، والمهتمين في هذا التخصص بالمفاهيم الإحصائية الأكثر شيوعاً واستخداماً، والتي تفيدهم في مجال تخصصهم .

يتكون هذا الكتاب من عشرة فصول ، توزعت على أهم مراحل البحث الإحصائي ، إذ تناولنا في الفصل الأول ماهية علم الإحصاء ، وفي الفصل الثاني جمع البيانات الإحصائية ، وفي الفصل الثالث عرض البيانات الإحصائية وتبويبها، وفي الفصلين الرابع والخامس لمقاييس التحليل الإحصائي سواءً لظاهرة واحدة (النسب والمعدلات ، ومقاييس النزعة المركزية ومقاييس التشتت) أم ظاهرتين (الارتباط والانحدار) ، وفي الفصلين السادس والسابع الاستدلال الإحصائي: للعينات كبيرة الحجم والعينات صغيرة الحجم ، وفي الفصل التاسع الإحصاء اللامعلمي ، وفي الفصل العاشر والأخير لمبادئ في الاحتمالات .

وفي نهاية الكتاب ، قمنا بوضع الجداول الإحصائية التي تساعد الباحث في فهم أصول التحليل الإحصائي ، والتي رأيناها ضرورية وتتعلق بفصول هذا الكتاب ، بالإضافة إلى قائمة بالمصطلحات الإحصائية والمراجع العربية والأجنبية .

إن إعداد هذا الكتاب ، هو إعداد جماعي ومسؤولية مشتركة ، إلاَّ أن الأمانة العلمية تقتضي تخصيص دور كل مؤلف على حدة ، فقد قام الدكتور مطانيوس مخول بكتابة الفصول : الأول والثاني والثالث والرابع والسادس والسابع ، والدكتور عدنان غانم بكتابة الفصول : الخامس والثامن والتاسع والعاشر .

لقد بذلنا جهدنا في تقديم الأفكار التي تثير اهتمام طلاب قسم المكتبات والمعلومات والمختصين منهم في هذا المجال في أهمية البحث والتحليل الإحصائيين ، وإننا إذا نضع هذا الكتاب بين أيديهم ، نأمل منهم جميعاً أن يتفضلوا بتقديم أي نقد يساهم في إغناء موضوعاته وإخراجه مستقبلاً بشكل نجعل منه أكثر نفعاً وفائدة ، وبما يسمح لنا في متابعة مسيرة البحث العلمي الملتزم والجاد ، وأخيراً نقدم شكرنا وتقديرنا لكل من ساهم في إخراج هذا الكتاب ، والله ولي التوفيق .

دمشق في 7 كانون الأول 2005

المؤلفان

الدكتور عدنان غانم & الدكتور مطانيوس مخول

الفصل الأول
ماهية علم الإحصاء

1-1 - لمحة تاريخية عن تطور علم الإحصاء
1-2- مفاهيم علم الإحصاء
1-3- شروط تطبيق الطرائق العملية على البيانات العددية
1-4- استعمالات الإحصاء
1-5- مراحل البحث الإحصائي (الدراسة الإحصائية)
1-6 - أسئلة

وقل رب زدني علما
جامعة دمشق
Damascus University

1-1- لمحة تاريخية عن تطور علم الإحصاء

يحتل علم الإحصاء مكاناً هاماً في مضمار العلم والمعرفة ، إذ يعد العلم المساعد لأغلب العلوم الاقتصادية والاجتماعية والطبيِّة والديمغرافية والفلكية وغيره ا، فهو علم عُرف منذ أقدم العصور ويُعد أيضاً من العلوم الحديثة التي دخلت جميع مجالات الحياة، إذ ظهرت أهميته المتزايدة من كونه يقدم الأدوات الضرورية لتحليل الظواهر الاقتصادية والاجتماعية والطبيِّة والديمغ رافية وغيرها، حيث إن المشاهدة أو الملاحظة ا لإحصائية هي عبارة عن " معرفة الشروط الضرورية للتأثير الفعال لماهية تطور الظواهر المدروسة وذلك تحت تأثير القوانين الناظمة لها، كما أنها الأساس في الإدراك العلمي وصولاً إلى الهدف المحدد" .

ارتبط مفهوم علم الإحصاء منذ القدم بفكرة الأرقام والأعداد ، حيث إن حاجة الإنسان للعدد بدأت حينما احتاج إلى العد ، وحاجته إلى العد بدأت حينما بدأ بمقارنة الأشياء فاستخدم أصابعه في العد . ولما كانت أصابعه لا تكفي ، ففكر عدّها بالحصى ، فنشأ عن ذلك مصطلح لغوي هو"الإحصاء" وهو يعني باللاتينية Statistics ، وهو مشتق من كلمتين Status الوضع أو الحالة ، و State تعني الدولة ، ومن هاتين الكلمتين اشتقت كلمة Statistics في القرن الثامن عشر، وتعني " إشراف الدولة " أو" مجموعة المعلومات المستقاة عن حالة الدولة " ، أي لكل دولة بياناتها الإحصائية ، وبناءً عليه ، يمكننا أن نقسم مراحل تطور علم الإحصاء إلى ثلاث مراحل:

1- **العصر الإحصائي الأول (مرحلة الإحصاء ا لأول)**: الذي يعد اللبن ة الأولى لانطلاقة هذا العلم " نقطة البدء " ، إذ استخدم في القرن العشرين ما قبل الميلاد لدى الصينيين حيث يُروى أن إمبراطور الصين كان يملك إحصاءات دقيقة عن السكان والأراضي المزروعة ، ثم عُرف الإحصاء لدى الرومان ، فالملك أوغسست " وضع ميزاناً لثروة إمبراطوريته" ، و كذلك عُرف في عهد المصريين القدامى والبطالمة والعباسيين غير أن غايات الإحصاء في ذلك العصر تجلت في :

حصر أعداد المواطنين وذلك لمعرفة المقدار العددي للقوة البشرية العاملة للرجال من جهة ، ولمعرفة مقدرة الدولة على الدفاع من جهة ثانية ؛

حصر الأراضي المزروعة ومن يعمل بها.

2- **العصر الإحصائي الثاني**: وفيه رُسخ الإحصاء كعلم يتصل بفكرة الأرقام والأعداد ويسعى إلى جمع البيانات والمعلومات عن الظواهر المختلفة والتعبير عنها بصورة رقمية ، وقد عُرف آنذاك باسم " الحساب العددي السياسي " ومن أنصاره العالم الاقتصادي وليم بيتي وإدمون هالي في القرنين السادس والسابع عشر للميلاد .

وعليه يمكننا أن نعرف " الرقم الإحصائي " بأنه عبارة عن مؤشر يعكس حالة الظاهرة المدروسة (حجم المساحات المزروعة أو غير المزروعة في بلد ما وفي لحظة زمنية معينة، عدد سكان مجتمع ما في زمن ما ، عدد الكتب المستعارة من مكتبة ما خلال اليوم الواحد أو خلال الشهر ،...الخ) .

3- **العصر الإحصائي الثالث** : وفيه ظهرت أهمية الإحصاء كمنهج علمي يشمل جميع المبادئ والأسس والطرائق العلمية التي تُتبع في جمع البيانات وعرضها وتحليلها وتفسيرها . فعمل الباحث الإحصائي إذن ، يبدأ بجمع البيانات ولا ينتهي إلاّ بعد أن يرتبها ويحللها ويفسرها، ومن ثم التنبؤ بما ستؤول إليه الظاهرة المدروسة مستقبلاً، وفي هذا العصر تطور علم الإحصاء بشكل ملحوظ وخاصة في القرنين السابع عشر والثامن عشر للميلاد ، نتيجة إدخال النظريات الرياضية عليه ، لكنه حقق تقدماً ملحوظاً في القرن التاسع عشر للميلاد ، لاسيما عندما ظهرت أبحاث العالم الإحصائي الإنكليزي كارل بيرسون (Carl Pearson)، الذي لقب بأبي الإحصاء ، وذلك عندما أدخل نظريات الاحتمال الرياضية في مجال علم الوراثة ثم تبعه ابن بلده رولاند فيشر(Roland Fisher) وغيره م من الإحصائيين أمثال ستودنت (Student) وماركوف (Marcove) ، وولد ، وكندل ، وكرامر ، وجميعهم استخدموا النظريات الرياضية في مجالا ت الإحصاء ، كما أنهم أسسوا جداول إحصائية تفسر العلاقة بين الظواهر موضوع الدراسة تفسيراً علمياً محدداً بدقة ، ومنها : جداول $t-$ ستودنت ، جداول كاي مربع χ^2 ، جداول فيشر F ، سينديكور، ... وغيرها .

هذا وقد حظي الإحصاء باهتمام جميع الدول ، لأنه يُعد الأساس الضروري الذي تعتمد عليه أية خطة مستقبلية تحددها السياسة العليا للدولة لتحقيق أهداف معنية ، بحيث كلما كانت البيانات دقيقة ومفصلة ، كلما كانت نتائج الخطة أقرب إلى الواقع .

1-2- مفاهيم علم الإحصاء:

قبل أن نُعرِّف الإحصاء ، فإن لكلمة الإحصاء مدلولين:

1. أن الإحصاء قائم على أنه مجموعة من البيانات أو الأرقام أو الأعداد ، كأن يقال مثلاً :

- إحصاءات طوارئ العمل في صناعة ما عام 2003، ويفضل هنا استخدام عدد حوادث طوارئ العمل بدلاً من إحصاءات ؛

- عدد حوادث السير في مدينة دمشق عام 2004 .

2. إن الإحصاء قائم على أنه مجموعة من الأساليب أو الطرائق العلمية التي تُطبق في معالجة البيانات العددية .

وفي إطار هذا المدلول الأخير نجد تعريفات عدة للإحصاء ، إلاّ أنها تصب في إعطاء فكرة واضحة عن الإحصاء ، وإن اختلفت في الشكل فإن مضمونها واحدٌ ، ومنها الآتي :

-الإحصاء هو : " مجموعة من الطرائق العلمية مهمتها جمع البيانات العددية وعرضها وتحليليها "؛

-الإحصاء " يشمل جميع الطرائق العلمية والوسائل و الأدوات الحديثة التي تساعد المحلل على الوصول إلى قرار سليم تجاه مشكلات تتحكم فيها العشوائية من خلال قوانين رياضية معروفة" ، أي أن الإحصاء أحد فروع علم الرياضيات ؛

- الإحصاء هو "علم يهتم باستخدام الطرائق العلمية والعملية وتطبيقها لجمع البيانات العددية وتنظيمها وتلخيصها وعرضها وتحليلها بغية الوصول إلى نتائج مقبولة وقرارات سليمة في ضوء التحليل ".

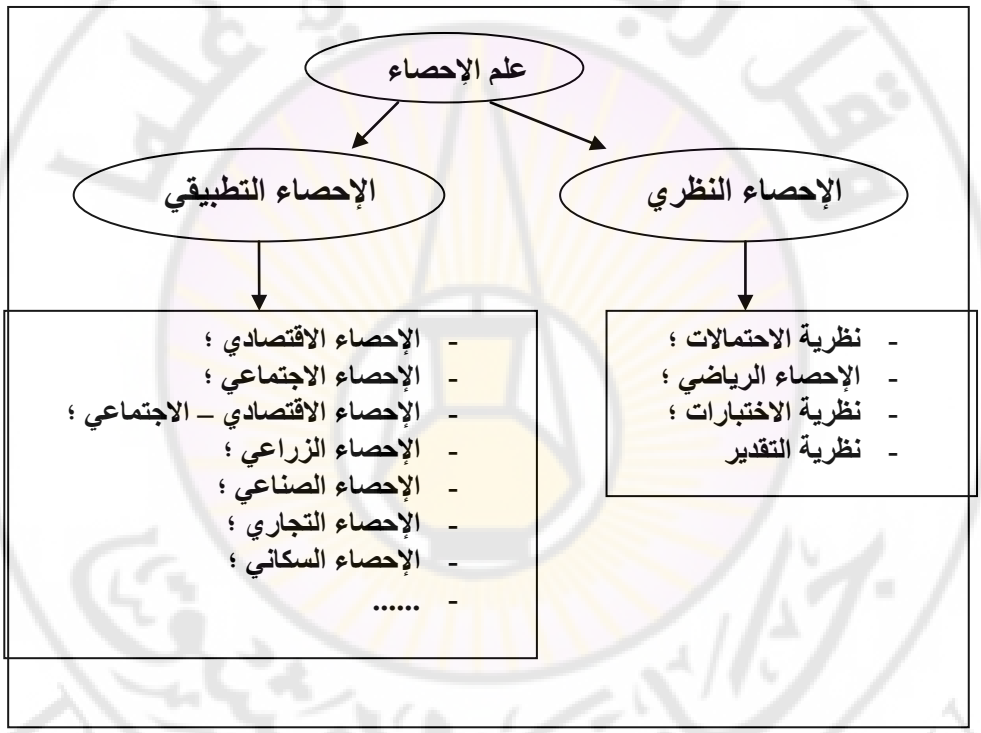
واستناداً إلى ما سبق ، يمكننا أن نميزّ بين :

- **<u>الإحصاء الوصفي أو التطبيقي</u> :** هو أحد فرعي علم الإحصاء ويهدف إلى وصف ظاهرة معينة وتحليلها ، وذلك بغية الوصول إلى نتائج ، وتقتصر غاية هذا الإحصاء فقط على جمع بيانات عن الظاهرة المدروسة وتبويب تلك البيانات وعرضها بيانياً أو جدولياً ثم وصف تغيرها، فهذا الإحصاء يستخدم في تلخيص معلومات عن مجموعة حالات بحثية مُقيسة على ضوء متغير أو أكثر ، ولا ينتج عن بياناتها أي نوع من أنواع الاستنتاج أو الاستنباط ، ويستخدم في تحديد العلاقات القائمة بين متغيرات البحث .
- **<u>الإحصاء الإستقرائي (ا لنظري)</u>:** وهو فرع من علم الإحصاء يهتم بالشروط الواجب توافرها ليكون الاستدلال سليماً ، أي ذلك الإحصاء الذي يطبق فيه منطق الاستنباط ، بغرض تقييم النتائج عن المجتمع المدروس ككل ، اعتماداً على معلومات جُمعت عن طريق العينة

العشوائية للحالات البحثية (محاولة تطبيق نتائج الجزء على الكل)، كما يطلق عليه بالإحصاء الاستنتاجي أيضاً .

وبناءً عليه ، يمكننا أن نبّين من خلال الشكل رقم (1-1) أهم الفروع العلمية لعلم الإحصاء وما يتفرع عنها من فروع علمية أخرى ، كالآتي :

الشكل رقم (1-1) : الفروع العلمية لعلم الإحصاء وما يتفرع عنها



1-3- شروط تطبيق الطرائق العملية على البيانات العددية

تجدربنا الإشارة هنا إلى أنه يشترط لتطبيق الطرائق العلمية في معالجة البيانات العددية أن تكون هذه البيانات العددية مكونة من وحدات متجانسة في الصفة أو في صفات نعتبرها رئيسية تبعاً لغرض الدراسة ، فعند انتفاء هذا الشرط ألاَّ وهو " **التجانس** " لا يمكن تطبيقها.

مثال: دراسة أوزان طلاب قسم المكتبات في كلية الآداب بجامعة دمشق وأطوالهم عام 2005 . أخذت عيرة مؤلفة من 500 طالب وطالبةلقياس أوزانهم أو أطوالهم . والسؤال هنا، هل الصفة في هذه العينة المأخوذة متجانسة أم لا ؟ طبعاً لا ، فللصفة في هذه العينة غير متجانسة نظراً لوجود طلاب ذكور وطالبات إناث ، لذا يجب أخذ عيرة من الطلاب الذكور وأخرى من الإناث، كل على حدة ، لكي تكون الصفة في أي من العينتين متجانسة ، لكن يمكن تطبيقها فيما لو كانت الصفة المدروسة في عينة طلاب قسم المكتبات " صفة الذكاء "،لأن الذكاء لدى جميع الطلبة واحداً، غير أن صفة التجانس غير محققة فيما إذا كانت طُبقت في قياس مستوى الذكاء ، لأن هذا متوقف على عدة عوامل تؤثر فيه مثل : عدد ساعات الدراسة والجهد المبذول في التحصيل والمواظبة على الدوام في الدروس والوسط الاجتماعي والعائلي المحيط ... الخ .

وخلاصة القول: إن " علم الإحصاء هو طريقة من طرائق البحث العلمي التي يمكن استخدامها في مختلف العلوم الكونية ... "

هذا ويمد الإحصاء الباحث الإحصائي بأدوات وطرائق علمية ، لكن يجب على الباحث مراعاة الآتي :

1 -**تجنب المغالاة في جمع البيانات العددية** ، لأن مهمته الاختيار بين الفرضيات الإحصائية واختبارها وتعميم نتائج العين ة على المجتمع الإحصائي (وضع الفرضية – اختبارها- تعميم نتائجها) ؛

2 -**إلمام الباحث ولحدٍ كبير بالموضوع المرغوب في دراسة أحد جوانبه دراسة إحصائية** ، أي ملاءمة الأساليب الإحصائية مع موضوع الدراسة وهدفها ، كي لا يسيء استخدام الطرائق الإحصائية ، أي يجب أن يكون لدى الباحث إلمام بطبيعة الظواهر المدروسة وبالأساليب التي تتلاءم معها .

1- 4- استعمالات الإحصاء :

الإحصاء علم قائم بنفسه لأنه يتميز بمجموعة من الخصائص منها:

- أنه يدرس الناحية الكمية بارتباط وثيق مع الكيف وفي ظروف محددة في الزمان والمكان ، أي ربط قيم الظاهرة المدروسة وأرقامها في المكان والزمان، مثال: عدد حوادث السير في مدينة دمشق في عام ما ؛

- أنه يدرس الناحية الكمية للظواهر الاقتصادية والاجتماعية كثيرة العدد (مجتمع عدد وحداته كبيرة جداً) ، أ ي قياس تلك الظواهر ، مثل عدد السكان في عام ما ؛
- أنه يدرس الناحية الكمية للظواهر المقسمة إلى فئات حسب مؤشرات لها عدة احتمالات ، كالسكان حسب العمر والنوع (ذكور- إناث) أو حسب المستوى التعليمي (يقرأ ولا يكتب ، يكتب ويقرأ ، أمي ، ...)أو حسب الحالة الزواجية (أعزب ، متزوج ، مطلق ،أرمل) .... الخ ؛
- أنه يدرس تطور الظواهر خلال الزمن ، مثل تغير أسعار الكتب الثقافية أو الترفيهية خلال فترات زمنية مختلفة ، وهذا يتطلب معرفة الأسباب الكامنة وراء ارتفاع هذه الأسعار(أسعار تنافسية أم نتيجة التضخم أم ...) خلال الفترات الزمنية المختلفة ؛
- أنه يدرس التأثير المتبادل بين الظواهر الاقتصادية والاجتماعية، مثل تساقط المطر يؤدي إلى محصول زراعي جيد وهذا بدوره يؤدي إلى زيادة الدخل وزيادة عدد عقود الزواج ، أي ظاهرة بيئية (أحوال جوية) ارتبطت بظاهرة اقتصادية (الدخل) و ارتبطت بظاهرة اجتماعية أخرى (الزواج) .

وبناءً على ذلك ، يستعمل الإحصاء في مختلف مجالات الحياة الكونية سواءً أكانت جغرافية أم اقتصادية أم زراعية أم طبيِّة أم فلكية أم... الخ ، و مع تزايد المسؤوليات الملقاة على عاتق الدول والمؤسسات نتيجة التقدم العلمي والتق اني وتطور أساليب الإنتاج ، تزايدت بالتالي استعمالات الإحصاء ، إ ذ غدت الأساليب الإحصائية ركناً أساسياً في اتخاذ القرارات ورسم السياسات على أسس علمية صحيحة ، ومن هذه الاستعمالات :

1 - **في الاقتصاد والتحليل الاقتصادي :** يحتل الإحصاء في الوقت الحاضر أهمية متزايدة في تطوير الاقتصاد ونموه ، لأن البيانات الإحصائية أصبحت تتخذ أساساً في السياسات الاقتصادية وبالوقت ذاته يمكن اعتبارها دعامة للنظريات الحديثة ، وإذا ما تتبع المرء الاتجاهات الثلاث (الاتجاه الكمي ، الاتجاه الاستدلالي ، الاتجاه التاريخي) التي تطور بها علم الاقتصاد وجعلت من علماء الاقتصاد، يعتمدون على الإحصاء في دراساتهم وأبحاثهم وتطوير نظرياتهم اعتماداً متزايداً، وبناءً عليه تتضح الصلة الوثيقة بين الاقتصاد والإحصاء . هذا ويستخدم الإحصاء في اغلب الدراسات الاقتصادية التي تهدف عادة إلى التنبؤ والتخطيط سواءً على مستوى مشروع صناعي ، أم على مستوى قطاع ، أم على مستوى

الاقتصاد الوطني ككل ، وهذا يقتضي استخدام بعض المؤشرات والمقاييس الإحصائية اللازمة كأدوات في التحليل الاقتصادي (كأرقام القياسية وغيرها) في دراسة الأسعار والأجور والاستثمار والادخار والاستهلاك والتصدير والاستيراد.

2 - **في إدارة الأعمال :** يلعب علم الإحصاء دوراً مميزاً في مختلف الدراسات الإدارية لما يقدمه من طرائق وأساليب إحصائية تُعد منهجاً علمياً وأداة للبحث في مجال إدارة الأعمال (بحوث العمليات، نظرية القرارات الإدارية ، ...) ، فالإحصاء يساهم في :

- دراسة اتجاه المبيعات وبيان الآثار الموسمية والدورية لها بهدف إعداد الخطط المستقبلية للمؤسسات والمنشآت المدروسة ؛
- تتبع حجم المخزون من المواد الخام ومستلزمات الإنتاج وغيرها ؛
- دراسة احتياجات المستهلكين ورغباتهم وأذواقهم (دراسة تسويقية لربط الإنتاج بمنافذ البيع) ؛
- توقع المبيعات في ضوء أساليب الإنتاج والخطط الإنتاجية (معرفة الاتجاه العام لحركة الأسواق) ؛
- تحديد درجة الجودة والرقابة عليها.

3 - **في الدراسات المحاسبية :** يستخدم الإحصاء بشكل كبير جداً في مجال الأعمال والدراسات المحاسبية ، مثل : تحليل الموازنات المالية ، والبحث عن مؤشرات وضوابط محاسبية ، والتنبؤ بالأرباح المحققة وتوزيعها ، وتحديد القيمة المضافة على مختلف مستويات الاقتصاد الوطني .

4 - **في الدراسات السكانية (الديمغرافية) :** يطبق علم الإحصاء في مجال الدراسات السكانية بغية التعرّف على أهم المؤشرات الديمغرافية المتعلقة بحالة السكان(حجم السكان في لحظة زمنية معينة ، تركيب السكان وفق أهم الخصائص الديمغرافية كالعمر والنوع و ...) وبحركة السكان الطبيعية (ولادات ووفيات) والمكانية الديناميكية (الهجرة من وإلى المجتمع) ، لأن قياسها الإحصائي يساعد في فهم الظواهر السكانية وتحليلها للكشف عن العلاقات السببِّية وتحديد العوامل التي تؤثر فيها وتتأثر بها ، ولهذا نرى الباحث الديمغرافي يسعى لاستخدام الطرائق الإحصائية في دراسة جميع العمليات الديمغرافية الناجمة عن حركتي السكان الطبيعية والمكانية .

5 - **في مجالات المال والتأمين :** ثّمة اعتبارات عدة تجعل من الإحصاء ذا أهمية كبيرة في مجال العلوم المالية والمصرفية بسبب أن ظواهر هذه العلوم تمتاز:

- بالتغير المستمر الذي يجعل التنبؤ بها يقتضي استخدام أساليب إحصائية متطورة؛
- بالسلوك البشري المعقد الذي يشكل أحد العوامل الضرورية في أثناء دراسة الأسواق المالية ؛
- التغيرات التي تتعرض لها القروض المصرفية الممنوحة من المصارف في دولة ما وفي فترة زمنية معينة تبعاً لآجال سدادها وللأغراض التي منحت من أجلها.

ويمكننا أن نذكر بعضاً من الظواهر الممكن دراستها إحصائياً : النقد المتداول ، المبيعات ، المخزون ، الودائع ، القروض ، أسعار الفائدة ، أسعر الصرف ، رؤوس الأموال ، الدخل القومي ، الناتج المحلي ، حركة الأوراق المالية ، أسعار البورصة ، أسعار الأسهم والسندات ... الخ ، حيث تحتاج كل ظاهرة من الظواهر المذكورة بيانات خاصة بها تتفق والأساليب الإحصائية المطبقة عليها (كالرياضيات الإكتوارية في مجال التأمين) .

6 - **في المكتبات والمعلومات :** بات معلوماً بأنَّ عملية البحث العلمي هي رؤية فكرية فعالة ومتميزة ، لأنها تساعد على فهم كيفية تفاعل الأحداث التي تجمعها علاقات مشتركة وإدراكها ، حيث تمدنا في أغلب الأحيان بالسبل التي تساعدنا على التكيف مع البيئة المحيط بنا، هذا وتشتمل عملية البحث العلمي على كل من مناهج جمع البيانات العددية وطرائق الوصول إلى استنتاجات تتعلق بخصائص هذه البيانات التي يتحدد نمط دراستها بالتطورات السابقة ذات الصلة بالموضوع المراد دراسته ، لذا تُعد معرفة كيفية إجراء البحوث على درجة كبيرة من الأهمية للمكتبيين وخبراء الأرشيف وأخصائي المعلومات أكثر من غيرهم من المختصين، وذلك لاعتبارات كثيرة ، منها الآتي :

- إن التطورات الأخيرة التي حدثت في العالم والتعقيدات التي نشأت عن تداخل المفاهيم وتشابك الأنظمة ، انعكس على آلية عمل المكتبيين وأصبح دورهم كوسطاء ما بين المعلومات المسجلة ومستخدمها ، إلاَّ أن هذا الدور لم يقتصر فقط على إيجاد مكان المعلومات وتحديده ، بل تعدى ذلك إلى تحليل تلك المعلومات وتقييمها وتفسيرها لصالح المستفيد ، لأن

معظم المعلومات ترد في أشكال بحوث ومقالات وتقارير تحتم على المكتبيين التعرف عليها واختبارها ليتمكنوا من بثّها ، حيث باستطاعة المكتبي القيام بهذا الدور بكفاءة عالية ، إذا ما تمكن من إجادة اللغات ، والتعرف على مناهج البحث العلمي وفهمها التي تساعده كثيراً علي تقديم خدمات معلومات متوازنة أكثر فائدة وأكثر جدية محققاً بذلك أحد الأهداف الرئيسة للتخصص .

- يوجد بعدان آخران لدور المكتبي :

1- يعتبر المكتبي مستخدماً لشتى أنواع البيانات التي يستفيد من دراستها في تطوير كفاءته المهنية ورفعها ، وذلك لتقديم خدمات معلومات أفضل للمؤسسة التي يعمل بها ؛

2- يعتبر عدد المكتبيين المشاركين حالياً في مشاريع الأبحاث العلمية متواضعاً ، إلاّ أن هذا العدد يتوقع له الزيادة قريباً ، فيما إذا كانت المشاريع البحثية متعلقة بالأنشطة الرئيسة لمجال المكتبات والمعلومات . ومع ذلك يجد المكتبيون أنفسهم مشاركين في أبحاث ودراسات تتعلق بمجالات أخرى مثل أمناء المكتبات المدرسية العامة والخاصة ، وأيضاً في دراسات تتعلق بالتعليم وغيرها ؛

وبناءً على ذلك ، تفيد التحليلات الإحصائية لأبحاث الإنتاج الفكري في مجال علوم المكتبات والحالات الأخرى ذات العلاقة ، في بلورة ازدياد أهمية البحوث التي تجري في مجال الإدارة ، الأمر الذي من شأنه أن يساعد المكتبي على القيام بدوره على أكمل وجه ، وذلك من خلال قيامه بعمليات تحليل للبيانات في مجال المكتبات أو شبكات المعلومات ومراكز المعلومات ليتوصل إلى معارف جديدة تكمن غايتها في تزويد المكتبيين والمستفيدين من المكتبات بالمعلومات الدقيقة والموثوق بها، لتساعدهم على اتخاذ القرار الأفضل ، وتحدد جودة البيانات المستخرجة من الأبحاث وفقاً لأهميتها ، ومدى مشاركتها في زيادة فعالية الخدمات المكتبية ورفع كفاءة المستفيدين من المكتبات.

هذا وتتوقف زيادة قدرة المكتبات في القيام بدورها في إجراء البحوث وتقييمها على عاملين أساسيين هما :

- ما يتعلق بالمشكلات الناتجة بسبب التطورات التي حدثت في مجال تقنيات الاتصالات والتكنولوجيا ؛

- ما يتعلق بالتوقعات البحثية المرتبطة بالمسؤولية العامة والمهنية المتخصصة ، التي تأخذ شكل بيانات إحصائية .

وخلاصة القول : إن المكتبي – كمستخدم للمعلومات – يرى بأن البيانات الإحصائية قد أصبحت جزءاً هاماً من الإنتاج الفكري للمجالات التي يتعامل معها ، فالكتب والدوريات والتقارير وغيرها من المصادر المعرفية، أصبحت أقرب ما تكون إلى الدراسات الكمِّية التي تقتضي على المكتبي معرفة كيفية التعامل مع هذه البيانات الرقمية كباحث ، سواءً أكانت صفته مصدراً للمعلومات أم كمدقق لها.

7 - **في العلوم الأخرى :** الطِّب ، التربية ، الزراعة ، الفلك ، علم الاجتماع ، علم النفس ، علم الحياة ، علم الأجناس ،.... الخ .

1 5 - مراحل البحث الإحصائي (الدراسة الإحصائية)

كان البحث العلمي فيما مضى أكثر ما يهتم بالعلاقات السببية ، أما الآن فإن اهتمام الباحث العلمي ينصب على نواح أخرى غير السببية ، إذ أصبح الهدف الرئيس له الحصول على علاقات تربط مختلف الظواهر بعضها ببعض ، ولا يمكن أن يُسمى البحث بحثاً علمياً إلاّ إذا سار تبعاً لمراحل لا تختلف في شيء عن المراحل التي يتبعها الشخص العادي في تفكيره عندما تصادفه مشكلة من مشاكل الحياة اليومية، وعليه يمكننا تصنيف مراحل البحث الإحصائي إلى ثلاث مراحل أساسية هي :

1- جمع البيانات الإحصائية:

- الشعور بوجود مشكلة أو مجال يحتاج إلى دراسة أو بحث (اختيار مجال البحث وتحديد المشكلة وفروضها والقيام بالدراسة التحضيرية)؛
- تحديد إطار الدراسة أو البحث (تحديد الوحدة الإحصائية وتحديد المجتمع الإحصائي) ؛
- جمع البيانات الإحصائية(وعلى هذه المرحلة تتوقف المراحل التالية ، وهذا ما سنوضحه بالتفصيل في الفصل الثاني).

2- تحليل البيانات الإحصائية :

1- تفريغ هذه البيانات وتبويبها وتصنيفها وفق مجموعات متجانسة ؛

2- عرض البيانات (جدولياً أو بيانياً) ؛

3- تحليل البيانات الإحصائية (إجراء بعض الحسابات اللازمة لبعض القيم المعبرة عن الظاهرة المدروسة).

3 – عرض نتائج التحليل واستنباط الأمور المتعلقة بها ، استخلاص النتائج والتنبؤ بالمستقبل بعد اختبار جودة النتائج المستخلصة ودقتها .

وبالاستناد إلى ما سبق ، يمكننا أن نلخص مراحل البحث الإحصائي أو الدراسة الإحصائية في شكل توضيحي يبيِّن لنا كل ما تتضمنه كل مرحلة من مراحله وكالآتي :

الشكل رقم (2): شكل توضيحي لأهم مراحل البحث الإحصائي

هذا وسنفرد مبحثاً خاصاً بكل مرحلة من مراحل البحث الإحصائي لتوضيح كل ما يتعلق بها بالتفصيل ، ولكي يفهم القارئ كيفية قيامه بدراسة إحصائية يتوخى الهدف المرجو منها .

1-6- أسئلة

1. عرِّف المفاهيم الآتية : المشاهدة الإحصائية ؛ الرقم الإحصائي ؛ علم الإحصاء الوصفي ، علم الإحصاء الاستقرائي ؛ علم الإحصاء .
2. ما هو علم الإحصاء ، وما هي فروعه العلمية ، وما الفرق فيما بينها؟
3. ما هي شروط تطبيق الإحصاء في معالجة البيانات العددية ؟
4. ما هي الاعتبارات الواجب مراعاتها من قبل الباحث في تطبيق الطرائق الإحصائية ؟
5. ما هي الخصائص التي يتميز بها علم الإحصاء ؟
6. اذكر أهمية استخدام علم الإحصاء في الاقتصاد والتحليل الاقتصادي؟
7. اذكر ما يساهم به علم الإحصاء في إدارة الأعمال ؟
8. ما هي الاعتبارات التي تجعل من الإحصاء ذا أهمية كبيرة في مجال المال والتأمين ؟
9. بيّن أهمية تطبيق الإحصاء في مجال المكتبات والمعلومات ؟
10. لماذا تُعد معرفة كيفية إجراء البحوث ذا أهمية كبيرة للمكتبيين ؟
11. على ماذا تتوقف زيادة قدرة المكتبات على القيام بدورها في إجراء البحوث ؟
12. اذكر أهم مراحل العمل الإحصائي (الدراسة الإحصائية) ؟

الفصل الثاني
جمع البيانات الإحصائية

2-1- تحديد المشكلة وماهية الدراسة التحضيرية
2-2- تحديد إطار الدراسة
2-2-1- تحديد الوحدة الإحصائية
2-2-1-1- أنواع الوحدات الإحصائية
2-2-1-1-1- الوحدات الإحصائية البسيطة
2-2-1-1-2- الوحدات الإحصائية المركبة
2-2-2- تحديد المجتمع الإحصائي (تعريفه ، الاعتبارات ، أنواعه)
2-3- أهمية مرحلة جمع البيانات الإحصائية
2-4- مصادر جمع البيانات الإحصائية (- تاريخية - داخلية - ميدانية)
2-4-1- الحصر الشامل
2-4-2- العينات
2-4-2-1- ماهية العينات
2-4-2-2- أنواع العينات (العينات العشوائية وغير العشوائية)
2-4-2-3- أخطاء العينات وكيفية معالجتها
2-5- تصنيف البيانات حسب مصادرها
2-5-1- بيانات ثانوية
2-5-2- بيانات أولية
2-6- طرائق الحصول على البيانات الإحصائية
2-6-1- المقابلة الشخصية
2-6-2- العدادوين
2-6-3- الاستمارة أو الكشف البريدي والشروط الواجب توافرها في الاستمارة
2-6-4- الهاتف والانترنت

وقل رب زدني علما
جامعة دمشق
Damascus University

تعتمد عملية اتخاذ القرار في أي مجال من مجالات الحياة اقتصادية كانت أم إدارية أم اجتماعية أم سياسية أم ... الخ ، على تحليل البيانات العددية المتوافرة عنها ، وهذا يتوقف على دقة عملية جمع البيانات الإحصائية وموثوقتها ، لأن عملية جمع البيانات الإحصائية تُعد من أهم مراحل العمل الإحصائي وعليها تتوقف نتائج التحليل الإحصائي، لأنه ليس لنتائج التحليل الإحصائي أية قيمة ما لم تكن البيانات الإحصائية التي قام التحليل على أساسها قد جمعت بشكل أصولي . هذا وتُعد عملية جمع البيانات الإحصائية الخطوة الأكثر كلفة من أية مرحلة أخرى من المراحل اللازمة للقيام بالدراسة الإحصائية ، لذا لا بد من القيام بدراسة كافية وواضحة لكل حيثيات المشكلة المراد دراستها وقبل البدء بعملية البيانات الإحصائية عنها ، إذا ما أُريد الحصول على بيانات موثوقة ذات قيمة علمية ـ حيث لا يوجد أسلوب معين يمكن اتباعه للتغلب على كافة مشكلات الجمع الإحصائي وللوصول إلى بيانات دقيقة بأقل التكاليف . هذا وتشتمل عملية جمع البيانات الإحصائية ثلاث خطوا ت أساسية ، هي :

1- تحديد المشكلة
2- تحديد إطار الدراسة أو البحث
3- جمع البيانات الإحصائية عن الظاهرة المدروسة .

2-1- تحديد المشكلة وماهية الدراسة التحضيرية

يدفعنا الشعور بوجود مشكلة ما إلى تحديد المشكلة المراد دراستها بشكل واضح ودقيق، أي يجب على الباحث أن يحدد المشكلة موضوع الدراسة بدقة متناهية ، قبل الشروع بالدراسة الإحصائية تخوفاً من ضياع الوقت ، ولكي لا تتراكم لديه معلومات أو بيانات إحصائية كثيرة لا تمت بصلة إلى موضوع الدراسة ، كما يجب على الباحث الأخذ بالحسبان بعض الاعتبارات في اختيار موضوع دراسته مثل : الاهتمام الشخصي ، وأهمية البحث ، والتركيز على المشكلة ، وتمويل البحث ، وتوافر مصادر البيانات، هذا ويجب على الباحث أيضاً صياغة فرضيات بحثه بشكل دقيق وواضح وأن تكون هذه الفرضيات مترابطة فيما بينها وقابلة للاختبار ، وبعد ذلك يضع خطة عمل لبحثه تتضمن أهم ما يجب القيام به[جانب نظري يوضح فيه أهمية المشكلة المراد دراستها و أبعادها ، وجانب تطبيقي يحدد فيه منهجية البحث (وصفي، استدلالي ، تجريبي ، ميداني) ومصادر الحصول على البيانات] ، بالإضافة إلى ذلك يتوجب على الباحث الاطلاع على جميع الأبحاث والدراسات ذات الصلة بموضوع دراسته ، لاسيما

قبل الصياغة النهائية للمشكلة التي يرغب في دراستها ، وهذا ما يمكن أن نطلق عليه " بالدراسة التحضيرية " ، وذلك للأسباب التالية:

1- إلمام الباحث بمعلومات كافية عن المصطلحات التاريخية والشروط الراهنة لموضوع الدراسة الإحصائية ؛

2- اكتشاف الباحث نتيجة االاطلاع والتحضير بأن المشكلة المراد دراستها قد سبقه غيره بدراستها وقام بحلها، وبالتالي أصبحت دراسته غير ضرورية وغير مجدية؛

3- اكتشاف الباحث نتيجة الدراسة التحضيرية أن ثَّمة دراسات ناجحة وجدت بهذا الشأن وما عليه غير التجديد أو الاستعانة بها أو أنه يكتشف أن هذه الدراسات مستحيلة فيجب عليه تبيان الأسباب أو الصعوبات التي واجهت هذه الدراسة.

2-2- تحديد إطار البحث

2-2-1- تحديد الوحدة الإحصائية

يُعد تحديد الوحدة الإحصائية من أهم المشكلات التي تواجه الباحث وعليها تتوقف الدراسة الإحصائية وعملية الجمع الإحصائي للبيانات عن الظاهرة أو الظواهر موضوع الدراسة ، لذا لا بد لنا من تعريف الوحدة الإحصائية بدقة تامة ، وعليه نُعرّف الوحدة الإحصائية : " بأنها مجموعة من الأشياء أو الظواهر المتجانسة فيما بينها فيما يتعلق بصفة أو بصفات معينة يجب توافرها في كل مفردة حتى يمكن اعتبارها وحدة إحصائية " ، ومن خلال هذا التعريف يمكننا توضيح الشروط الواجب تحققها في الوحدة الإحصائية، وهي كالآتي :

1- أن تكون ملائمة لغرض الدراسة ومعرفة تعريفاً دقيقاً واضحاً ، بحيث يمكن استبعاد الوحدات غير المتجانسة معها في الصفة أو في الصفات الرئيسية ؛

2- أن تكون ثابتة ومستقرة ويستثنى منها الوحدات النقدية لأنها مرنة وفي تغير مستمر بين فترة زمنية وأخرى .

وبعبارة أخرى يمكننا تعريف الوحدة الإحصائية : " بأنها أصغر جزء مستقل تجري عليه الدراسة الإحصائية وهي ثابتة ومستقرة وتشترك بصفة أو أكثر ".

2-2-1-1- أنواع الوحدات الإحصائية

يمكننا تصنيف الوحدات الإحصائية في نوعين هما :

1- وحدات إحصائية بسيطة.

2- وحدات إحصائية مركبة.

2-2-1-1-1- الوحدات الإحصائية البسيطة

الوحدات الإحصائية البسيطة ه ي تلك الوحدات التي تشمل على خاصة واحدة للموضوع الذي تصفه ، و التي يمكن أن تستعمل للجمع والتحليل والعرض الإحصائي ، وهي تقسم بدورها إلى وحدات عد ووحدات قياس ووحدات نقد .

1 - **وحدات العّد :** وهي الوحدات التي تستخدم في عمليات العّد ، و تشمل تعداد جميع الأشياء المادية والنوعية أو صفاتها مثل: (طبيب ، مدرسة ، جامعة ، مكتبة ، طالب ، مكتبي ، مجلد ، كتاب ، قصة ، رواية ، ....) وإن هذه الوحدات تطبق على الأشياء التي ليس لها واحدة قياس فيزيائية أو رياضية ؛

2 - **وحدات القياس:** وهي الوحدات التي تستعمل في قياس الأشياء المادية التي لا يمكن عدها ، وهذا يعني وجود أنظمة كاملة للأوزان والمقاييس (كالطن ، والغالون، والمتر ، والميل ، و... الخ) في معظم بلدان العالم . هذا وتُعد وحدات القياس مرضية تماماً ، لأنها تتمتع بجميع صفات الوحدة الإحصائية المثلى ، إلاَّ أنها تشكو من بعض الصعوبات المصطنعة (نتيجة اختلاف أنظمة القياس، فمثلاً البلدان العربية والأوروبية تستخدم مقياس المتر ، والكيلوغرام والليتر ،... في حين تستخدم البلدان الأخرى كالولايات المتحدة الأمريكية وبريطانيا اليرد ، والرطل ، و... الخ) ، والطبيعية وهي أكثر خطورة ، لأن هذا النوع من الوحدات لا يقيس إلاّ الأشياء المادية ولا يأخذ بالحسبان الجودة ؛

3 -**وحدات النقد :** وهي الوحدات التي يمكن اعتبارها نوعاً من أنواع وحدات القياس كون النقد مقياساً للقيمة ،إلاَّ أنّ ل ها مشكلات خاصة غير شائعة بالنسبة لوحدات القياس الأخرى ، فهي مثلاً غير ثابتة، إذ تخضع بين الحين والآخر إلى تقلبات تطرأ على مستوى الأسعار والقوة الشرائية لها ، لذا تعتبر أية مقارنة بالوحدات النقدية في أوقات مختلفة مقارنة مضَّللة بسبب مرونة قيمتها.

2-2-1-1-2- الوحدات الإحصائية المركبة :

الوحدات الإحصائية المركبة هي الوحدات التي تتألف على الأقل من وحدتين إحصائيتين بسيطتين ،

مثل : الوحدة الحرورية وهي عبارة عن الكمية اللازمة لرفع حرارة 1 سم^3 من الماء سانتغراداً واحداً على مستوى سطح البحر ؛ سرعة السيارة (كم/ سا) ؛ الكثافة السكانية (شخص / كم^2).

2-2-2- تحديد المجتمع الإحصائي

يجب على الباحث عند قيامه بالدراسة الإحصائية أن يحدد بدقة تامة إطارها الزماني والمكاني ، لمّا لهذا التحديد من أهمية كبرى عند الشروع في توسيع ذلك الإطار أو تضييقه ، وهذا يخضع إلى عدة اعتبارات كالهدف من الدراسة ، والإمكانات المادية والكوادر البشرية المتاحة، وعليه نعّرف " المجتمع الإحصائي " : بأنه مجموع الوحدات الإحصائية ذات الصفة المدروسة ، وهو مفهوم مرن يمكن أن يتلاءم مع أي دراسة إحصائية ، أي بمعنى أنه يمكن توسيع إطاره المكاني والزماني أو تضييقه، في ضوء الاعتبارات الآتية ، منها:

- الهدف من الدراسة ؛
- الوقت اللازم لإنهاء الدراسة ؛
- الإمكانات المادية والبشرية المتاحة.

هذا وقد يكون المجتمع الإحصائي محدوداً ، أي يمكن حصر مفرداته مثل (عدد المكتبات الخاصة والعامة في سورية لعام ما ، عدد طلاب جامعة دمشق في عام معين) ، إنتاج معمل المسامير في يوم ما ، ... ، وهو ذو طبيعة متقطعة ، أو يكون غير محدود ، أي لا نستطيع حصر مفرداته ، مثل : عدد أسماك البحر ، عدد النجوم في السماء ، رمي قطعة النقد مرات لا متناهية وظهور (كتابة أو شعار)، ... الخ وهو ذو طبيعة مستمرة.

2-3- أهمية مرحلة جمع البيانات الإحصائية

وبناءً على ما سبق تأتي خطوة جمع البيانات الإحصائية ، والتي تعتبر من أهم خطوات أو مراحل البحث الإحصائي لأن عليها تتوقف نتيجة التحليل ، فكلما جمعت البيانات بدقة ، كلما كانت النتائج الناجمة عن التحليل أكثر صحة ودقة والعكس من ذلك يؤدي إلى نتائج مضلّلة ومشوهة ، أي ينتج عن عدم جمع البيانات بشكل أصولي عدم دقة نتائج التحلي ل الإحصائي حتى ولو استخدم ت

أفضل الأساليب الإحصائية ، لأن الاستخدام الأخير لا يعوض عن عدم صحة البيانات التي اتخذت أساساً للدراسة الإحصائية أو تحيزّها .

2- 4- مصادر جمع البيانات الإحصائية

تختلف أساليب جمع البيانات الإحصائية باختلاف مصدر الحصول على هذه البيانات ، لذا نميزّ بين :

1 - **المصادر التاريخية (الرسمية)** : يمكن جمع ا لبيانات من سجلات رسمية نتيجة التسجيل تبعاً للقوانين والأنظمة المرعية في المجتمع (سجلات الأحوال المدنية) ، ومن التقارير السنوية التي تصدرها الهيئات الرسمية المرتبطة بالدولة (المكتب المركزي للإحصاء ، وزارة الصحة ، الهيئات والمنظمات الدولية ، ...) . يتميز هذا الأسلوب بالسرعة والسهولة وقلة التكاليف ودورية البيانات ووفرتها ، إلاَّ أن البيانات التي يوفرها محدودة بمجالات خاصة ومحددة؛

2 - **المصادر الداخلية** : يقصد بالمصادر الداخلية مجموعة البيانات التي تجمعها الهيئات والمؤسسات بصفة مستمرة ودورية لأغراض التنظيم والإدارة داخل هذه المؤسسات ، وتتميز هذه بقلة التكاليف وسرعة الحصول عليها بالإضافة إلى كونها أكثر تخصصاً ؛

3 - **المصادر الميدانية**: يقصد بالمصادر الميدانية استقاء البيانات من مصادرها الأصلية، حيث يقوم الباحث الإحصائي بجمعها وفق طريقتين : الحصر الشامل والعينة الإحصائية .

2-4-1- الحصر الشامل

يقوم الحصر الشامل على التعداد الكامل لجميع الوحدات الإحصائية التي يتألف منها المجتمع الإحصائي المدروس ، وهذا يجري كل عشر سنوات أو خمس سنوات حسب الإمكانات المادية والبشرية لكل دولة ، تتميز هذه الطريقة بالدقة ، غير أنه يصعب تنفيذها نتيجة ارتفاع تكاليفها وحاجتها إلى كادر إحصائي مدرب ومؤهل .

2-4-2- العينات

تقوم هذه الطريقة على جمع المعلومات أو البيانات عن طريق دراسة جزء من المجتمع الإحصائي ،وذلك بأخذ عينة تمثله تمثيلاً صادقاً ، ثم تعميم نتائجها بعد الدراسة والتحليل على المجتمع الذي أخذت منه .

وتجدر الإشارة هنا ، إلى بعض الاعتبارات التي يأخذها الباحث عند اختيارطريقة جمع البيانات الإحصائية ، طريقة الحصر الشامل أم طريقة العيرة ، ومنها:

1- الوقت والمال المخصصّان لإنهاء الدراسة ؛

2- طبيعة المجتمع الإحصائي ونوعه ، عما إذا كان المجتمع محدوداً أم غير محدود ، فمثلاً عدد سكان المجتمع الصيني كبير جداً يفضل أخذ عينة بدلاً من الحصر الشامل؛

3- نوع الدراسة ، فإذا كانت الدراسة سترؤدي إلى فقدان بعض الوحدات موضوع الدراسة وتلفها، فإننا نلجأ إلى العيرة ومن ثم نعمم نتائجها ، مثل تحليل دم الإنسان ، ومعرفة مدى صلاحية البيض المنتج للأكل في مزرعة ما؛

4- تعذر الوصول إلى جميع عناصر المجتمع الإحصائي وإجراء الدراسة عليها ، كدراسة كميات الأمطار الهاطلة منذ عام 1940 وحتى الآن وأثرها في كميات إنتاج القمح في سورية خلال تلك الفترة الزمنية ؛

5- الحاجة الملحة إلى نتائج سريعة لاتخاذ قرار معين ، وخير مثال شركات تصنيع الأدوية .

وعليه يفضل الباحث طريقة العيرة بدلاً من الحصر الشامل للأسباب الآتية:

1- العيرة غير مكلفةمن حيث المال والوقت ؛

2- عدم إمكانية توفر الكوادر البشرية المؤهلة والتجهيزات الضرورية للقيام بالحصر الشامل ؛

3- تعرضها للأخطاء الاحتمالية والتحيّز ، غير أنه يمكن التقليل والحّد منها، بسبب أنها جزء من المجتمع الإحصائي المدروس ؛

4- الاستفادة منها في عملية الاستدلال ، أي الاستدلال على صفات المجتمع المأخوذة منه.

وعندما يقرر الباحث الأخذ بطريقة العينة ، فثّمة مشكلة تواجهه ألاَّ وهي تحديد حجم العينة المراد سحبها ونرمز لها بـ (n) ، وهذا يتوقف على عدة عوامل ، أهمها :

- مقدار الأموال المخصصِّة لجمع البيانات المطلوبة ؛
- مقدار الثقِّة التي يرغب الباحث في توافرها في البيانات التي يقوم بجمعها؛
- الكوادر البشرية المؤهلة والمتاحة للقيام بجمع المعلومات المرغوب فيها.

وعليه يمكننا تعريف " العينة " : بأنها مجموعة من مفردات المجتمع الإحصائي يتم جمعها بحيث تكون ممثلة تمثيلاً صحيحاً له ، بقصد دراسة ظاهرة معينة على

هذه المجموعة من المفردات للوصول إلى نتائج يمكن تعميمها على ذلك المجتمع الذي أُخذت منه عند احتمال ما ، وإن أخذ مثل هذه العينة الممثلة لمجتمعها الإحصائي يتطلب أمرين ، هما :

- تحديد هدف الدراسة الإحصائية ، عاماً كان أم خاصاً ؛
- تحديد حجم المجتمع الإحصائي ونرمز له بـ (N) .

1-2-4-2- أنواع العينات

تصنف العينات حسب أسلوب سحبها من المجتمع الإحصائي إلى عينات عشوائية (احتمالية أو علمية) وأخرى غير عشوائية (شخصية) ، إذ تحتل العينات العشوائية أهمية كبيرة نظراً لاعتمادها على نظرية الاحتمالات وإمكانية استخدامها في اختبار الفرضيات والتعميم من العينة على مجتمعها الإحصائي وحساب مقدار الأخطاء الاحتمالية التي تقل مع ازدياد حجم العينة وذلك بعكس أخطاء التحيّز فهي ثابتة بالرغم من ازدياد حجم العينة ، وسوف نوجز ما يتعلق بكل من العينات العشوائية وغير العشوائية .

2- 4-2-1-1- العينات العشوائية

تعد العينات العشوائية بمختلف أنواعها الأسلوب الوحيد الذي يعطي مقاييس كميّة لقياس الدقة أو الثقة ويسمح بتقويم مقدار الأخطاء الاحتمالية أو أخطاء الحظ والصدف. وتعّرف العينة العشوائية بأنها : " العينة التي اختيرت باستخدام أساليب السحب العلمية في الاختيار العشوائي كالسحب بالقرعة أو بأسلوب الدواليب أو كرات اليانصيب أو استعمال جداول الأرقام العشوائية ، أي بمعنى إعطاء كل مفردات المجتمع الإحصائي الفرصة أو الاحتمال نفسيهما ليتم اختيارها " .

ولكي تتصف العينات بالعشوائية يجب مراعاة الشروط الآتية :

1 - أن تكون مفردات المجتمع الإحصائي المراد أخذ عينة منه متجانسة من حيث طبيعتها ونوعية الدراسة المرغوب إجراءها على هذه المفردات ؛

2- أن تكون مفردات المجتمع الإحصائي مستقلة بعضها عن بعض ، أي أن انتقاء أي مفردة من مفردات المجتمع لا يرتبط بسحب أو عدم سحب أية مفردة أخرى ؛

3- أن يكون احتمال انتقاء أية مفردة من مفردات المجتمع الإحصائي معروفاً أو يمكن حسابه ؛

4- أن يتم انتقاء مفردات العينة من المجتمع الإحصائي المعن ي بدون تحيزّ ، أي أن يتصف الانتقاء بالعشوائية .

ونشير هنا إلى أن عملية سحب مفردات العينة العشوائية يتم بأحد شكلين، هما:

- **السحب مع الإعادة** : وذلك بأن نسمح بإمكانية اختيار مفردة اختيرت سابقاً ، أي أن المفردة التي اختيرت تعاد ثانية إلى المجتمع حتى لا تسقط من عملية الاختيار ، وإنما تتاح لها فرصة الاختيار مرة أخرى، أي احتمال اختيارها ثابتاً ويساوي لـ ($\frac{1}{N}$) ، حيث إن (N) هو عدد مفردات المجتمع الإحصائي المراد أخذ عينة عشوائية منه ؛
- **السحب بدون إعادة** : وهو رفض إعادة أية مفردة تم اختيارها سابقاً ، أي أن المفردة التي تم اختيارها لن تتاح لها فرصة أخرى للاختيار ، فمثلاً : احتمال اختيار إحدى مفردات المجتمع الإحصائي للمرة الأولى هو ($\frac{1}{N}$) لا يساوي احتمال اختيار مفردة أخرى منه في المرة الثانية ($\frac{1}{N-1}$) وهكذا .

وأياً كان شكل السحب لمفردات العينة ، فإنه يتم سحب مفرداتها تبعاً للأساليب الآتية :

1. **أسلوب القرعة** : يقتضي هذا الأسلوب ترقيم كافة مفردات المجتمع الإحصائي بأرقام متسلسلة بحيث تسجل هذه الأرقام على كرات أو قصاصات ورقية متجانسة فيما بينها وتوضع في كيس أو وعاء كبير وتخلط جيداً ، وبعدها نقوم بسحب إحدى الكرات أو القصاصات مع إغماض العينين ، ونكرر هذه العملية عدداً من المرات بمقدار عدد مفردات العينة المراد تشكيلها ، لكننا نشير إلى أنه بعد أية عملية سحب يجب أن تخلط المفردات بشكل جيد . وما يصيب هذا الأسلوب أنه يشكل عبئاً مزعجاً إذا كان حجم المجتمع الإحصائي الذي ستؤخذ منه العينة كبيراً جداً ، بالإضافة إلى أن أدوات الخلط قد لا تكون مرضية .
2. **أسلوب الدواليب** : يتطلب هذا الأسلوب ترقيم جميع مفردات المجتمع الإحصائي بأرقام متتالية ثم نجهز عدداً من الدواليب المتوازنة والمتماثلة بمقدار عدد مرتبات العدد (N) ، ونتأكد من أن هذه الدواليب ذات توازن تام وتعمل بصورة جيدة ، ثم نقسّم محيط كل دولاب إلى عشرة أقسام متساوية وبعدها تلصق على كل منها الأرقام 0 ، 1 ، 2 ، 3 ، 4 ، 5 ، 6 ، 7 ، 8 ، 9 . وبعدها نثبت سهماً خارج محيط هذه الدواليب للدلالة على الأرقام في عملية السحب بعد تدوير تلك الدواليب وبقوة شديدة وبآن واحد ، كما هو الحال في سحب يانصيب معرض دمشق الدولي، وما يميز هذا الأسلوب تخلصنا من عملية تشكيل الإطار.

3. **أسلوب جداول الأرقام العشوائية** : وهي عبارة عن جداول من أرقام عشرية من 0 إلى 9 مسجلة في جداول خاصة ترتب فيها الأرقام بشكل عشوائي، بحيث يكون احتمال سحب أيٍّ منها معلوماً أو مساوياً لسحب رقم آخر منها ، وهي تتكون من عشرين صفحة وكل صفحة منها تتألف من عشرة أعمدة وخمسين سطراً ، بحيث يكون كل عمود يتكون من أربعة أرقام ، كما في المقطع الآتي المأخوذ من جدول الأرقام العشوائية (انظر ملحق الجداول الإحصائية في نهاية الكتاب) :

1534	7106	2836	7873	5574	7545	7590	5574	1202	7712
6128	8993	4102	2551	0330	2358	6427	7067	9325	2454
6047	8566	8644	9343	9297	6751	3500	8754	2913	1258
0806	5201	5705	7355	1448	9562	7514	9205	0402	2427
9915	8274	4525	5695	5752	9630	7172	6988	0227	4264
.	.	.	.	.	.	.	.	.	.
3522	5601	6197	6051	3470	8283	5702	0103	8726	5282

ويتم أولاً اختيار إحدى صفحات الجداول العشوائية بشكل عشوائي وأيضاً اختيار كل من رقم العمود ورقم السطر في تلك الصفحة بشكل عشوائي ، ومن ثُم نختار مفردات العينة ، فمثلاً : إذا فرضنا وجود مجتمع مكون من 250 مفردة ، وأردنا اختيار عينة عشوائية مكونة من 25 مفردة (10% من حجم المجتمع) ، فيتم اختيار مفردات العينة كالآتي :

بعد الاختيار العشوائي لرقم صفحة من صفحات جداول الأرقام العشوائية ولعمود ما وسطر ما من تلك الصفحة ، فإننا نقاطع السطر مع العمود في تلك الصفحة مكوناً من ثلاثة أرقام (لأن حجم المجتمع مكوناً من 250 مفردة)، وتكون أول مفردة ، بحيث نأخذ كل ثلاثة أرقام متتالية في ذلك العمود على أساس التتابع الرأسي حتى نحصل على مجموعة من المفردات تساوي حجم العينة ، مع استبعاد المفردة التي رقمها يفوق حجم المجتمع ، وكذلك استبعاد المفردة التي يتكرر رقمها في حالة السحب بدون إعادة ، وعليه يمكننا توضيح مثالنا أعلاه ، من أننا حصلنا على :

867	632	629	050	086	929	769	796	202	176
624	668	013	854	383	796	285	043	265	142
314	878	156	368	935	575	209	955	348	685
819	028	652	413	629	421	709	012	551	345
287	342	012	517	077	211	242	287	369	306
255	285	326	597	521	716	204	004	538	325
996	255	719	907	241	740	227	121	842	314
204	019								

وبالتالي تكون مفردات العينة هي الأرقام الآتية :

142	050	086	202	176	012	156	209	013	043
012	077	211	242	028	241	227	121	204	004
131	026	038	204	019					

حيث استبعدت كافة الأرقام التي تزيد عن 250 ، لأن المجتمع يقتصر فقط على 250 رقماً وليس غير .

وعليه نميزّ عدة أنواع للعينات العشوائية ، منها :

❖ العينة العشوائية البسيطة

تُعد العين ة العشوائية البسيط ة من أبسط العينات العشوائية و أكثرها أصالة بالعشوائية ، إذ يراعى فيها تكافؤ الفرص ويعطى لكل مفردة من مفردات المجتمع الإحصائي موضوع الدراسة الاحتمال نفسه بأن تنتقى ضمن العينة ، ويتم اختيار مفرداتها بأحد أساليب السحب العشوائي التي سبق شرحها .

وما يؤخذ على هذا النوع من العينات الآتي :

1- عدم تمثيلها للطبقات المكونة للمجتمع الإحصائي ، خاصة إذا كان المجتمع الإحصائي مكوناً من طبقات غير متجانسة من حيث الظاهر ، كطلبة إحدى الجامعات ، فهو مؤلف من طلاب ذكور وطلاب إناث ، وتوزع كل من طلاب الذكور متفاوت عنه لدى الطالبات الإناث ، لاسيما إذا كانت البيانات المرغوب في جمعها على أساس النوع ؛

2- ارتفاع التكاليف المادية والوقت اللازمين لإنهاء الدراسة ، لاسيما إذا كان حجم المجتمع كبيراً جداً ، أو إذا كانت مفردات المجتمع الإحصائي منتشرة

بصورة واسعة في مناطق شاسعة وبعيدة عن مكان الإشراف على العمل الميداني بجمع البيانات عن تلك المناطق بالإضافة إلى صعوبة الوصول إليها أحياناً.

❖ العينة العشوائية الطبقية

هي إحدى البدائل للعينة العشوائية البسيطة والتي يمكن استعمالها لزيادة دقة النتائج ، فيما إذا كان المجتمع الإحصائي مؤلفاً من طبقات عدة بينها اختلاف واضح ، إذ يتم تقسيم المجتمع إلى طبقات متجانسة من حيث الظاهرة موضوع الدراسة ، ثم يسحب عدد معين من مفردات كل طبقة بشكل عشوائي وفق إحدى الطرائق الآتية :

1. **طريقة التخصيص المتساوي** : توزع مفردات العينة بالتساوي على طبقات المجتمع ؛
2. **طريقة التخصيص النسبي** : يؤخذ من كل طبقة عدد من المفردات يتناسب مع حجم الطبقة إلى حجم المجتمع ، حيث يؤخذ من كل طبقة نفس العدد لتكون في مجموعها حجم العينة ؛
3. **طريقة التخصيص الأمثل** : يؤخذ بالحسبان عند اختيار المفردات من كل طبقة الأمور التالية : حجم الطبقة ، وتوزيع الظاهرة في كل طبقة ، وتكلفة جمع البيانات من كل طبقة ، غير أن الطريقة الثانية – طريقة التخصيص النسبي – هي الطريقة الأكثر شيوعاً واستخداماً ، وبموجبها يتم تقسيم المجتمع الإحصائي موضوع الدراسة إلى طبقات متجانسة وغير متداخلة ، فإذا كان حجم مجتمع الدراسة (N) واستطعنا تقسيم هذا المجتمع إلى طبقات عددها (m) وكانت أحجام هذه الطبقات :

$$N_1, N_2, N_3, ..., N_m$$

فإن :

$$N = N_1 + N_2 + N_3 + ... + N_m = \sum_{i=1}^{m} N_i$$

وأردنا اختيار عينة حجمها (n) من مجتمع الدراسة ، فإننا نختار من كل طبقة عدداً من المفردات يتناسب طردياً مع حجم كل طبقة ، وهنا نميز بين حالتين :

1. **عندما يكون تباين الظاهرة متساو في الطبقات المختلفة** : فإذا رمزنا للمفردات المختارة من هذه الطبقات بـ $n_1, n_2, n_3, ...n_m$ فإن

$$n = n_1 + n_2 + n_3 + ... + n_m = \sum_{i=1}^{m} n_i \quad \text{بحيث يكون :}$$

$$n_1 = \frac{N_1}{N} * n \quad ، \quad n_2 = \frac{N_2}{N} * n \quad ، \quad n_3 = \frac{N_3}{N} * n \quad ، \ldots \quad ، \quad n_m = \frac{N_m}{N} * n$$

فمثلاً إذا أردنا أن نختار عينة عشوائية طبقية حجمها $n=100$ من مجتمع مكون من $N=1000$ شخصاً ، ينقسمون إلى ذكور وإناث (400 ، 600)، وعلى التوالي ، فإننا سنختار :

- من الذكور $n_1=\frac{N_1}{N}*n=\frac{400}{1000}*100=40$

- من الإناث $n_2=\frac{N_2}{N}*n=\frac{600}{1000}*100=60$

وعليه تكون العينة العشوائية **الطبقية** $n=n_1+n_2=40+60=100$.

مثال : لقد كان من المرغوب فيه تشكيل عينة من (100) مكتبة من المكتبات (1000) المتوافرة في بلد ما ، وكنا نعلم بأن المكتبات المدرسية تمثل (50%) والمكتبات العامة (30%) والمكتبات الأكاديمية (15%) والمكتبات المتخصِّصة (5%) ، ففي هذه الحالة تُشكل العينة الطبقية بنسبة تساوي نسبتها في مجتمع الدراسة ، أي يكون:

-50 مكتبة مدرسية ؛

-30 مكتبة عامة ؛

-15 مكتبة أكاديمية ؛

-5 مكتبات متخصِّصة

مثال آخر : عند دراسة نمط المعيشة وتكاليفها في بلد ما ، وكان المرغوب فيه الحصول على نتائج الدراسة على مستوى الريف والمدينة ، وهما الطبقتان الأساسيتان اللتان يتألف منهما مجتمع البلد المدروس ، في مثل هذه الحالات تؤخذ العينة من كل من الريف والمدينة بطريقة تتيح دراسة كل طبقة من هاتين الطبقتين .

2. **عندما يكون تباين الظاهرة غير متساو في الطبقات المختلفة** :

فإننا نختار من كل طبقة عدداً من المفردات يتناسب طرداً مع حجم هذه الطبقة وتباين الظاهرة داخل الطبقة ، أي :

$$n_1=\frac{N_1*\sigma_1}{\sum_{i=1}^{m}N_i*\sigma_i}*n \quad ، \ldots . ، \quad n_m=\frac{N_m*\sigma_m}{\sum_{i=1}^{m}N_i*\sigma_i}*n$$

وبشكل عام يكون : $n_i=\frac{N_i*\sigma_i}{\sum_{i=1}^{m}N_i*\sigma_i}*n$

حيث إن σ_i هو الانحراف المعياري للظاهرة داخل الطبقة وهو الجذر التربيعي للتباين ، ويستخدم هذا النوع من العينات عندما تكون :

1) الحاجة ماسة إلى جمع بيانات عن كل طبقة من طبقات المجتمع ، فإنه يفضل معاملة كل طبقة وكأنها مجتمع مستقل ؛
2) الظروف الإدارية تستدعي عملية التقسيم إلى طبقات ، فقد نجد الجهة القائمة على إجراءات البحث عدة هيئات أو لجان ميدانية وكل منها لها القدرة على جمع البيانات عن جزء معين من مجتمع الدراسة ؛
3) مشكلات المعاينة مختلفة بين طبقة وأخرى ، كما هي الحال في المجتمعات البشرية كنزلاء الفنادق أو المشافي يختلفون عن أولئك الذين يعيشون في بيوتهم، لأن الوصول إلى وحدات العينة يختلف بين حالة وأخرى ، كما هي الحال في المعاينة الإدارية لمعرفة ظروف المكتبات لبلد ما ، فإننا نقسم المكتبات الموجودة فيه إلى مكتبات كبيرة وأخرى صغيرة بحسب عدد العمال المستخدمين في تلك المكتبات ، بحيث يوزع حجم العينة بين هاتين الطبقتين ؛
4) الغاية هي الوصول إلى تقدير أفضل لثوابت المجتمع .

❖ العينة العشوائية المنتظمة

يمتاز هذا النوع من العينات بسهولة التطبيق ، عندما يكون عدد مفردات العينة كبيراً ($n \succ 50$) ، بالإضافة إلى سهولة استخدامها من السجلات ، كما يمكن استخدامها عندما نرغب في ترتيب مفردات المجتمع الإحصائي بانتظام.

وما يميز هذا النوع من العينات هو تساوي بعد المفردات بعضها عن بعض انطلاقاً من المفردة الأولى المجددة بشكل عشوائي ، بحيث يحدد هذا البعد بتقسيم عدد مفردات المجتمع (N) على عدد مفردات العينة (n)، وهذا ما يدعى " بطول فترة السحب "، فمثلاً : كان لدينا مجتمعٌ حجمه (N) وأردنا اختيار عينة حجمها (n) ، فإننا نرقم مفردات المجتمع من 1 إلى N ، ثم نختار مفردة ما بشكل عشوائي ولتكن (k) فإن المفردة الثانية هي ($k+\frac{N}{n}$) والمفردة الثالثة ($k+2*\frac{N}{n}$) وهكذا بحيث تكون المفردة الأخيرة ($k+\frac{(n-1)*N}{n}$) ، بفرض لدينا مجتمع حجمه 25 وأردنا اختيار عينة حجمها 5 ، فإن طول فترة السحب $= \frac{N}{n} = \frac{25}{5} = 5$ ، فالعينات الممكن سحبها من الحجم نفسه متوقف على تحديد المفردة الأولى بشكل عشوائي ، فمثلاً حُددت تلك المفردة ، وكانت مساوية لـ 3 فتكون مفردات العينة كالآتي :

3 ، 8 ، 13 ، 18 ، 23

مثال : إذا أردنا أخذ عينة حجمها (40) بطاقة من فهرس بطاقات مكتبة ما يتألف من (1000) بطاقة ، ما علينا في هذه الحالة سوى تحديد إحدى البطاقات بشكل عشوائي ولتكن حاملة للرقم (5) ، ثم نحدد طول فترة السحب $= \frac{N}{n} = \frac{1000}{40} = 25$ ، فتكون البطاقة الثانية حاملة للرقم (30) وهكذا ، أي : 5 ، 30 ، 55 ، 80 ، 105 ، 130 ، 155 ، 180 ، 205 ، 230 ، 255، 280 ، 305 ، 330 ، 355 ، 380 ، 405 ، 430 ، 455 ، 480 ، 505 ، 530 ، 555 ، 580 ، 605 ، 630 ، 655 ،680 ، 705 ، 730 ، 755 ، 780 ، 805 ، 830 ؛ 855 ، 880 ، 905 ، 930 ، 955 ، 980 .

وما يعاب على هذا النوع من العينات الآتي :

- عدم تمثيلها للمجتمع الإحصائي الذي أخذت منه ، فيما إذا كان هناك تكرار دوري في الطريقة المرتبة معها مفردات المجتمع الإحصائي ، ومثال على ذلك إذا أراد أحد المكتبيين إجراء مسح ميداني يستطلع فيه رأي مواطني المنطقة عن الخدمات التي تقدمها مكتبته ، وبغرض تحديد مجتمع الدراسة قرر استخدام قائمة أرقام المنازل السكنية المحيطة بالمكتبة ليستطلع رأي قاطنيها ،لأنهم يمثلون فئة المستفيدين من خدمات المكتبة ، وبتطبيق أسلوب العينة العشوائية المنتظمة وقع اختياره العشوائي على منزل يقع على الزاوية في إحدى المنازل السكنية ، فإنه سيتم اختيار جميع المنازل الواقعة على الزاوية أيضاً في المنازل السكنية الأخرى في المنطقة المشمولة بالدراسة ، غير أن ميزات المنازل الواقعة على الزوايا مختلفة عن غيرها من البيوت من حيث الكلفة والسعر ومن حيث وجهة نظرهم تجاه الخدمات التي تقدمها المكتبة ؛
- تحيزّها لاسيما عند تقدير متوسط الطول أو الوزن لمجتمع إحصائي مؤلف من ألف جندي قد رتبوا في صفوف متساوية بحسب الطول .

❖ العينة العشوائية العنقودية

من الملاحظ بأن استعمال كل من العينات العشوائية البسيطة والطبقية يتطلب توفر قائمة عن مفردات المجتمع الإحصائي المدروس ، لكن وفي حالات كثيرة لا تتوافر لدينا مثل هذه القوائم ، لذا يصبح من المتعذر اختيار عينة من ذلك المجتمع ، وهذا يتطلب منا القيام بمسح شامل ، وليس باستطاعتنا إجراؤه ، فعندها نلجأ إلى تقسيم مفردات المجتمع الإحصائي إلى مجموعات نسمي كلاً منها عنقوداً ، ثم نختار عينة عشوائية بسيطة من بين تلك العناقيد .

يستخدم هذا النوع من العينات بالدراسات المتعلقة بالقطاع الزراعي وبالدراسات السكانية، إذ يقسم المجتمع الإحصائي إلى مناطق جغرافية مميزة، حيث تسحب عينة من هذه المناطق ثم يسحب بصورة عشوائية عدد من المفردات من ضمن كل منطقة من المناطق التي تتضمنها العينة ، فمثلاً إذا أردنا دراسة حجم الأسرة في مدينة دمشق ، ولنفترض أن هذه المدينة قد قُسمت إلى عدد من مناطق العد ، حيث يقوم الباحث بسحب عدد من تلك المناطق بشكل عشوائي ومن ثم يسحب من كل منطقة عينة عشوائية بسيطة من أسر تلك المنطقة ، فكل منطقة هي عنقود ، وبعد أخذ البيانات المطلوبة عنهم ، فإنه بالإمكان تحديد حجم الأسرة في مدينة دمشق ، وبما أنه تم اختيار المناطق عشوائياً وكذلك المفردات المأخوذة من كل منطقة من تلك المناطق عشوائياً ، فمن الممكن حساب أخطائها الاحتمالية.

❖ العينة العشوائية المتعددة المراحل

يستخدم هذا النوع من العينات في حال كان المجتمع الإحصائي كبيراً ويصعب فيه تحديد إعداد أطر تشمل كل مفرداته ، فنعمد إلى تقسيم المجتمع الإحصائي إلى وحدات أولية ثم نختار عينة من هذه الوحدات كمرحلة ، ثم نقسم كل وحدة من الوحدات الأولية المختارة إلى وحدات ثانوية ثم نأخذ منها عينة كمرحلة ثانية ، وبعدها نقسم كل وحدة من الوحدات الثانوية المختارة إلى وحدات أصغر نختار منها عينة كمرحلة ثالثة وهكذا ...

وعليه تعرّف العينة العشوائية المتعددة المراحل بأنها العينة التي تختار مفرداتها، بحيث تمر بعدة مراحل في كل مرحلة الوحدات الأصغر من المجتمع ثم تليها الوحدات الأصغر منها وهكذا حتى نصل إلى المفردة التي نحصل منها على البيانات ، أي تلك العينة التي يتطلب اختيار مفرداتها أكثر من مرحلة ، فإذا تم هذا الاختيار على مرحلتين ، فإن العينة تسمى بثنائية المراحل . فإذا كان لدينا عدد من صفوف الثالث الثانوي في إحدى المدارس الثانوية في مدينة حلب ، فإنه يجب اختيار عينة من الصفوف ثم عينة من الطلاب ، وهذا يعني بأن جمع البيانات قد تم على مرحلتين :

- مرحلة أولى : اختيار عينة من الصفوف (الوحدات الأولية) ؛
- مرحلة ثانية : اختيار عينة الطلاب من الصفوف المختارة (الوحدات الثانوية) .

2- 4-2-1-2- العينات غيرالعشوائية

ويطلق على هذا النوع من العينات أحياناً بالعينات الشخصية وهي عبارة عن تلك العينات التي يتم الحصول على بياناتها بشكل غير عشوائي ، أي التي لا يمكن تطبيق نظرية الاحتمال وأصول الاستدلال الإحصائي عليها ، وإن تطبيقها ضمن تحفظات ما يضمن كي تكون نتائج التحليل صحيحة وغير مضّللة ، وإلاَّ كان العكس من ذلك ،

وهذا يعطيها أهمية وخاصة في كونها تقدم مقاييس إحصائية وصفية ذا ت دورهام في توفير خلاصات مفيدة وقيّمة للباحث بالإضافة لاستعمالها كأدوات لا غنى عنها عند إجراء محاكمة عقلانية ، ومع ذلك وجب الحذر عند استعمالها وتعميم نتائجها ، ومن أهم أنواعها نذكر :

- **عينات الحصص (أو الكوتا)**

وهي تشبه العينات العشوائية الطبقية ، إذ يؤخذ عدد من المفردات من كل طبقة من الطبقات المكونة للمجتمع الإحصائي بما يتناسب مع عدد المفردات التي تتألف منها كل طبقة ، إلاّ أنها تختلف عن العينات العشوائية الطبقية فقط بأسلوب السحب المطبق في سحب مفرداتها ، بالإضافة إلى عدم التمكن من قياس أخطائها الاحتمالية ، فهي تتعرض لأخطاء التحيزّ ، وتستعمل هذه العينات في مجالات كثيرة منها استطلاع الرأي وفي المسح التسويقي .

- **العينات المنتقاة**

وهي ذلك النوع من العينات غير العشوائية التي يقوم الباحث اعتماداً على حكمته ومهارته في انتقاء مفردات المجتمع التي يجب أن تتضمنه العينة ، حيث لا يوجد أي مقياس يمكن معه قياس الأخطاء الاحتمالية في هذه العينات، ويطلق على هذا النوع من العينات غير العشوائية " بالعينات الغرضية " أو " عينات الخبرة " .

وتعطي العينات المنتقاة نتائج جيدة عند استعمالها في مجالات معينة ، وخاصة عند اختيار السلع الداخلة في تركيب الأرقام القياسية لأسعار الجملة أو المفرق، كما أنها تعطي نتائج أكثر دقة عندما يتعلق الأمر بتقدير متوسط الظاهرة المدروسة ، في حين تكون النتائج أقل دقة عندما يتعلق الأمر بتشتت الظاهرة وانتشارها ، لأن مقاييس التشتت المحسوبة من هذه العينات غالباً ما تكون متحيزة نحو الأدنى ، وكمثال عليها :

لنفرض أننا نريد إجراء بحث حول العلاقة بين المكتبات العامة والسلطات الإدارية والتنفيذية في منطقة ما، فما علينا سوى استطلاع رأي مديري المكتبات المشهود لهم بنشاطهم وفعاليتهم والذين يفهمون دورهم الاجتماعي والسياسي في المنطقة ، وذلك أفضل من استقاء المعلومات عن طريق العاملين في مكتبات المنطقة .

هذا ونشير هنا إلى إمكانية اختبار صلاحية تمثيل هذا النوع من العينات لمجتمع الدراسة المعني ، من خلال اتباع الآتي :

- استخدام النتائج التي حصلنا عليها من العينة المنتقاة لإجراء استبيان أو مقابلات نطبق عليها أسلوب العينة العشوائية ؛
- القيام بالبحث عن أدلة تؤيد النتائج التي توصلنا إليها ، عن طريق قراءة التقارير السنوية وسجلات المكتبة ، والإنتاج الفكري في مجال المكتبات ومراجعتها؛
- اتباع الملاحظة على سلوكيات مديري المكتبات ومطابقتها مع نتائجنا .

- **العينات كبير ة الحجم**

وهي العينات التي تقوم على أساس أخذ عينات من حجم كبير جداً من مفردات المجتمع الإحصائي المرغوب في دراسته ، استناداً إلى الاعتقاد الشائع : " بأنه كلما ازداد حجم العينة فإن المعلومات التي تعطيها تكون أكثر دقة " . فهذا يكون صحيحاً عندما يكون حجم المجتمع الإحصائي محدوداً وكان حجم العينة المأخوذة منه مقترباً جداً من حجم المجتمع الإحصائي نفسه ، لأن أخطاء العينة تقل في مثل هذه الحالات ، أما إذا كان حجم المجتمع الإحصائي المأخوذة منه العينة غير محدود ، فإن زيادة حجم العينة سيؤدي إلى تقليل الأخطاء الاحتمالية فيما إذا سحبت بشكل عشوائي ، لأن العلاقة عكسية بين الأخطاء الاحتمالية وحجم العينة العشوائية، غير أن هذه العلاقة غير صحيحة فيما إذا كانت العينة غير عشوائية ، ويستعمل هذا النوع من العينات في التنبؤ بانتخابات الرئاسة في بعض البلدان .

2-4-2-3- أخطاء العينة

تُعد العينات العشوائية المأخوذة بأحسن الأساليب وفي أحسن الحالات قريبة من العينات المثالية، إلاَّ أنها تختلف بمقاييسها الإحصائية الو صفية بشكل أو بآخر عن المقاييس الإحصائية الو صفي ة المقابلة لها في المجتمعات الإحصائية التي أخذت منها ، وهذه الاختلافات تدعى بأخطاء العينة التي يمكن تقسيمها إلى أخطاء ترافق عملية اختيار مفردات العينة وأخرى تنشأ عن حدوث تقصير أو إهمال في خطوات الدراسة بالعينة، أي تقسم أخطاء العينة إلى :

1 -**الأخطاء الاحتمالية أو أخطاء الحظ والصدف** : وهي عبارة عن الأخطاء التي تصاحب عملية اختيار مفردات العينة، وتنشأ عن أسباب متعددة، صغيرة ومعقدة ، لكنها من حجم متساوٍ وتأثيرها في الاتجاه السلبي أو الإيجابي متكافئ الاحتمال، مما يجعلها تميل في المدى البعيد لإلغاء بعضها بعضاً ، وبالتالي سيكون تأثيرها في النتائج قليلاً عندما تؤخذ هذه النتائج في مجموعها ، وهذه الأخطاء يمكن تحديدها بطريقة أو بأخرى ضمن شروط معينة، ويمكن التخلص من تأثيرها في نتائج التحليل الإحصائي أو إنقاصه لدرجة كبيرة بأخذ عينة من حجم أكبر ، غير أن هذا مكلفٌ من حيث الوقت والنفقات ، وقد يُعرض بعض المفردات للتلف من أجل فحصها في بعض الحالات كما ه ي الحال في مراقبة الجودة إحصائياً ، مما يصبح أمراً غير عملي عند أخذ عينات كبيرة الحجم، لذا يتوجب على الباحث الاقتناع بقياس أخطاء الحظ والصدف وأخذها بالحسبان بشكل كامل عند قيامه بتحليل العينة واستخلاص أية معلومات عنها .

2- أخطاء التحيّز أو الأخطاء المنتظمة : وتنشأ لأي سبب أو لمجموعة من الأسباب التي تميل لإحداث أخطاء في نفس الاتجاه، سواءً كان هذا الاتجاه موجباً أو سالباً ، وتبقى أخطاء التحيّز ثابتة بالرغم من زيادة حجم العينة، ومن هذه الأسباب ما يلي :

- عدم مراعاة الأساليب العلمية في سحب العينات ؛
- الاختيار القصدي للعينة ؛
- عدم تنفيذ العينة بشكل كامل أو الاستعاضة عن بعض مفرداتها بمفردات أخرى ؛
- استعمال إطار معيب في اختيار العينة ؛

ويمكننا معالجة أخطاء التحيّز بإحدى طريقتين هما :

1- إدخال تعديل على النتائج عندما تُعرف طبيعة التحيّز وحجمه ؛

2- استخدام أساليب تؤدي إلى التقليل من فرصة ارتكاب أخطاء التحيّز إلى الحد الأدنى، لاسيما عندما يكون حجم التحيّز واتجاهه غير معلومين وهذا يوافق أغلب الحالات .

2-5- تصنيف البيانات حسب مصادرها

تصنف البيانات الإحصائية التي يجمعها الباحث تبعاً لمصادرها إلى بيانات ثانوية وأخرى بيانات أولية :

2-5-1- البيانات الثانوية : وهي تشمل جميع البيانات الإحصائية المتوفرة منشورة كانت أم غير منشورة ، ومن الأمثلة عليها البيانات التي تصدرها الجهات الرسمية والحكومية كالمجموعات الإحصائية لأي بلد ، أو البيانات الإحصائية لإحدى الوزارات في دولة ما، أو سجلات إحدى المكتبات العامة لبلد معين ، وتتميز بأنها :

1- توفر في الوقت والتكلفة ؛

2- تسمح للباحث في توسيع مدى دراسته ضمن الوقت والمال المخصصّين لذلك.

ومن مساوئها ، الآتي :

- تجمع لأغراض مختلفة ، قد لا تلائم غرض الدراسة الإحصائية ؛
- ليس للباحث أية رقابة على مدى صحتها ، أي يعتريها النقص أحياناً؛
- قد تكون الوحدة الإحصائية المستخدمة غير مرضية ؛
- قد تكون الثقة بها مختلفة باختلاف مصدر الحصول عليها والهدف منها .

2-5-2- البيانات أولية : غالباً ما يجد الباحث بأن البيانات الثانوية الموثوق بها غير ملائمة لغرض الدراسة ، لذا لابد من القيام بجمع البيانات الإحصائية بنفسه ، وهنا يطرح السؤال : لماذا يفضل الباحث البيانات الأولية على البيانات الثانوية ؟

يفضل الباحث البيانات الأولية على البيانات الثانوية لأن البيانات الثانوية الموثوق بها غير ملائمة لغرض الدراسة ، لذلك يلجأ الباحث إلى استخدام البيانات الأولية لما له من إمكانية التحكم في فرز هذه البيانات بما يلائم هدف دراسته ولكي تكون الوحدة الإحصائية المستخدمة أكثر ملائمة مع غرض دراسته .

2- 6 – طرائق الحصول على البيانات الإحصائية

2-6-1- <u>طريقة المقابلة الشخصية</u> : يقوم الباحث وفق هذه الطريقة بنفسه أو بواسطة أشخاص مدربين على البحث بشكل جيد بمقابلة المبحوث، وتوجيه الأسئلة بشكل مباشر كما هي واردة في الاستمارة وتسجيل الإجابات في المكان المخصص لها ، حيث تُعد هذه الطريقة من أكثر الطرائق المطبقة في الحصول على بيانات إحصائية أكثر دقة وصحة ، وهي تخدم الباحث باعتباره مطلعّاً على أهداف الدراسة وأغراضها وله المقدرة على فرز البيانات الهامة من تلك غير الهامة ، إلاَّ أن محاذيرها: مكلفة من حيث الوقت والمال ، إذا لم يحدد مدى الدراسة الإحصائية .

2-6-2- <u>طريقة العدادون</u> : وهي تعتمد على استخدام أشخاص لجمع البيانات يسمون عدادون ،و العداد : هو الشخص المكلف من قبل الجهة أو الهيئة التي تخَّوله بجمع البيانات الإحصائية عن بعض الظواهر . هذا وبموجب هذه الطريقة يمكن الحصول على بيانات إحصائية موثوقة بتفويض عدد من العدادين للقيام بجمعها، وهي أكثر كلفة من الطريقة الأولى (المقابلة الشخصية) بسبب استخدام عدد كبير من ذوي الخبرة .

2-6-3- <u>طريقة الاستمارة البريدية (الاستبيان البريدي)</u> : وهو أسلوب ذو كلفة أقل وبالوقت ذاته أقل كفاية ، لأن هذا الأسلوب يتطلب إثارة الحافز المناسب لدى المستجوب لتأمين إجابته عن طريق إرسال كتاب مرافق بقائمة أسئلة مُعدة بعناية ودقة ، للتعويض ما أمكن عن الاتصال المباشر بين الباحث والمستجوب ، ومن هذه الحوافز: عدم الإكراه ، الاهتمام ، الولاء ، الربح ....

وعليه يعّرف الاستبيان بأنه : " كشف يتضمن الأسئلة المراد طرحها من قبل الباحث بهدف جمع البيانات والمعلومات عن الظاهرة المرغوب في دراستها " ، هذا وعند تصميم الاستبيان يجب مراعاة الأسس الآتية :

- اسم الجهة المشرفة على البحث والأهداف المرجوة منه ؛
- بيانات شخصية عن المستجوب إن كان هو وحدة البحث مثل : الاسم، العنوان ، الحالة الاجتماعية ، الحالة التعليمية ، ... الخ ؛
- الأسئلة والاستفسارات المطلوب الإجابة عليها ؛
- الإشارة إلى سرية البيانات وعدم استخدامها لأغراض أخرى ؛
- ملاحظات جامع البيانات وتعليمات معينة من خلال إعطاء كل استبيان رقم خاص به ليسهل عملية تفريغ البيانات فيما بعد ؛

كما يجب الأخذ بالحسبان الشروط الواجب توفرها عند صياغة أسئلة الاستمارة الإحصائية ، وهي كالآتي :

1. أن تحتوي أقل عدد ممكن من الأسئلة ؛
2. أن تكون الأسئلة واضحة وغير غامضة (لا يوجد فيها أي إبهام أو غموض)؛
3. لا تحتوي على أمر يقلق بال المستجوب أو يضر بمصالحه ؛
4. أن تدور الأسئلة حول معلومات حديثة لا قديمة ؛
5. تجنب قدر الإمكان الإجابات الإنشائية ؛
6. عدم إفساح المجال في صيغة الأسئلة للإجابة بتحيزّ ؛
7. من المرغوب فيه أن توضع أسئلة معاكسة في الاستمارة ، شريطة عدم إطالة قائمة الأسئلة إطالة غير مقبولة ؛
8. يستحسن أن تكون الأسئلة مصاغة بشكل تكون الإجابة " بنعم " أو "لا ".

ولهذا الأسلوب ثمّة مزايا ومساوئ ، نذكر منها الآتي :

مزايا الاستبيان	مساوئ الاستبيان
• سهولة الوصول إلى المبحوثين وقلة التكاليف ؛ • تصلح في الحالات التي تحتوي أسئلة محرجة ؛ • تسمح للمبحوث بالتفكير بالإجابة الدقيقة ؛ • التخلص من تأثير البحث في المبحوث عن طريق الإيحاء وبالتالي التخلص من أثر تحيّز الباحث .	• لا تصلح في حال كثر عدد الأميين بين أفراد البحث ؛ • قد يؤدي هذا الأسلوب إلى الحصول على أجوبة متحيزّة ، لاسيما عندما يرغب بعض أفراد البحث من توجيه البحث باتجاه معين وغياب بعض الإجابات من الأفراد المهملين ؛

2-6-4- <u>الهاتف والانترنت</u> : بهذه الحالة يقوم الباحث بتوجيه أسئلة استبيانه إلى المبحوث عن طريق الاتصال المباشر بهم من خلال استخدام الهاتف أو عن طريق نشر هذا الاستبيان على شبكة الانترنت ، إلاَّ أن ما يميزّ طريقة استخدام الهاتف الآتي :

1. السرعة في الحصول على البيانات ؛
2. إفساح المجال في توضيح الأسئلة للمبحوث .

وما يعيبها الآتي :

- لا تصلح إلاَّ للأشخاص الذين يملكون هاتفاً ؛
- تحتاج إلى تكاليف مرتفعة ؛
- لا تصلح في الحالات التي تتطلب ملاحظة مباشرة .

غير أن مزايا استخدام طريقة نشر الاستبيان عبر شبكة الانترنت يتجلى في السرعة في الحصول على البيانات وقلة التكاليف ، أما مساوئها في عدم امتلاك جميع الأفراد في مجتمع البحث للحاسب الشخصي أو الهاتف أو كليهما ، بالإضافة إلى عدم إلمام جميع الأفراد بكيفية استخدام الحاسب أو استخدام شبكة الانترنت إن توافرت لديهم .

2-7- أسئلة

1 - ما هي الدراسة التحضيرية ؟
2 - ما هي أسباب الاهتمام بالدراسة التحضيرية ؟
3 - ما هي الوحدة الإحصائية ، وما هي أنواعها ؟
4 - ما هي صفات الوحدة الإحصائية ، أو ما الشروط الواجب تحققها في الوحدة الإحصائية ؟
5 - ما هي أنواع الوحدات الإحصائية ، اذكرها واشرح كل منها ؟
6 - ما هي أنواع الوحدات الإحصائية البسيطة ؟
7 - هل تعتبر الوحدة النقدية مقياساً ثابتاً للقيمة في أوقات مختلفة ؟
8 - لماذا تعتبر المقارنة بالوحدات النقدية في أوقات مختلفة مقارنة مضّللة ؟
9 - لماذا يعتبر مفهوم المجتمع الإحصائي مفهوماً مرناً ؟
10 -ما هي اعتبارات تحديد المجتمع الإحصائي ؟
11 - عرّف المجتمع الإحصائي ؟
12 - ما الفرق بين المجتمع الإحصائي المحدود وغير المحدود ؟
13 - ماذا ينتج عن عدم جمع البيانات العددية أو الإحصائية بشكل غير أصولي ؟
14 - لماذا تعتبر خطوة جمع البيانات الإحصائية من أهم خطوات الدراسة الإحصائية ؟
15 - ما هي الاعتبارات في اختيار أسلوب جمع البيانات ؟
16 - لماذا يفضل غالباً الباحث الأخذ بأسلوب العينة بدلاً من أسلوب الحصر الشامل ؟
17 - ما هي مصادر الحصول على البيانات الإحصائية ؟
18 - ما الفرق بين أسلوب الحصر الشامل وأسلوب العينة ؟
19 - ما هي الأساليب العلمية المستخدمة في السحب العشوائي ؟
20 - ما هي العينة العشوائية وما شروطها ؟
21 - ما هي أنواع العينات؟
22 - اذكر أنواع كل من العينات العشوائية والعينات غير العشوائية ؟
23 - ما الفرق بين العينات العشوائية والعينات غير العشوائية ؟

24 - ما الفرق بين العينة العشوائية الطبقية وعينة الحصص ؟
25 - ما هي أخطاء العينة ، وكيف تعالج ؟
26 - ما هو التحيّز وما هي أسبابه وكيف تعالج ؟
27 - صنف البيانات الإحصائية تبعاً لمصادرها ؟
28 - ما الفرق بين البيانات الثانوية والبيانات الأولية ؟
29 - ما هي فوائد البيانات الثانوية ؟
30 - لماذا يفضل الباحث البيانات الأولية على البيانات الثانوية ؟
31 - ما هي طرائق جمع البيانات ؟ ما هي الشروط الواجب توافرها في الاستمارة الإحصائية ؟
32 - اذكر مزايا الاستمارة الإحصائية ومساوئها؟
33 - اذكر مزايا ومساوئ كل من : طريقة الهاتف وطريقة الانترنت .

وقل رب زدني علما
جامعة دمشق
Damascus University

الفصل الثالث
عرض البيانات الإحصائية

3-1- العرض الجدولي
3-1-1- التبويب
3-1-2- ماهية الجداول الإحصائية وشروطها
3-1-3- أنواع الجداول الإحصائية
3-1-4- كيفية ترتيب البيانات الإحصائية وأساليب ترتيبها
3-1-5- الجداول التكرارية
3-1-5-1 ماهية الجداول التكرارية
3-1-5-2- كيفية إعداد الجداول التكرارية (مزاياها ومساوئها)
3-1-5-3- خصائص التوزيعات التكرارية وصفاتها
3-2- العرض البياني
3-2-1- ماهية العرض البياني وغايته
3-2-2- أنواع العرض البياني
3-2-2-1- المصورات الإحصائية والخرائط الجغرافية
3-2-2-2- الأعمدة البيانية (البسيطة ، المقسّمة)
3-2-2-3- رسوم المساحات والأحجام البيانية " الدوائر الإحصائية "
3-2-2-4- رسوم المنحنيات البيانية
3-2-2-4-1- منحنيات التكرار وأنواعها
3-2-2-4-2- منحنيات السلاسل الزمنية

وقل رب زدني علما
جامعة دمشق
Damascus University

تستخدم أساليب عرض البيانات بشكل كبير في المجالات والعلوم كلها ، نظراً لأهميتها في العرض الموجز لمكونات الظاهرة المدروسة أو لكيفية التعبير عن محتوى الظواهر المدروسة بإيجاز، لذا لابد من تصنيف البيانات الإحصائية وتلخيصها بعد جمعها ، لكي تغدو صالحة لعملية التحليل ، نتيجة لوجود أرقام متباينة وكثيرة وغير منسقة ، أي لا يمكن أن يتم تحليل البيانات التي جُمعت عن ظاهرة ما ، ما لم نسعَ إلى تفريغها وتبويبها . وبهذا الصدد سوف نتطرق لكل من : العرض الجدولي والعرض البياني .

3-1- العرض الجدولي

3-1-1- التبويب

كي يتم عرض البيانات الإحصائية بشكل جدولي ، لابد من تبويبها ، إذ تعتمد عملية التبويب على تجميع المعلومات المتجانسة أو المتشابهة وتصنيفها في فئات معينة ، والغاية من ذلك هو : " عرض أكثر ما يمكن من البيانات بأكبر ما يمكن من الوضوح والإيجاز " ، و عليه يمكننا أن نلخص أهم خطوات التبويب كالآتي :

- تحديد أصناف الدراسة المرغوب فيها ، أي تصنيف الإجابات في أصناف محددة بوضوح ومعرفة تعريفاً جيداً ؛

– تبويب الإجابات في الأصناف المحددة لها ، أي جمع هذه الإجابات المجموعة في كل صنف من الأصناف أو تلخيصها؛

– إعداد الجدول الإحصائي الملائم ، أي تسجيل نتائج هذا الجمع في جداول إحصائية ملائمة .

وهذه العملية تدعى بكاملها بالتبويب ، وعليه يعرّف "التبويب" بأنه : "جمع مفردات البيانات في مجموعات أو أصناف بحسب كمياتها أو ميزاتها المشتركة " . فمثلاً لو قمنا بدراسة الكتب التي يختارها المستفيدون من إحدى المكتبات العامة تبعاً لماهية الكتب المتوافرة في هذه المكتبة لكانت الإجابات متباينة وكثيرة منسقة وبحاجة إلى تبويب إلى بنود أو أصناف متجانسة ، حيث يتضمن كل بند أو صنف المفردات التي تشترك في صفة معينة أو صفات عدة يرتبط بعضها ببعض مما يسوغ الجمع بينها . لنأخذ مثلاً ظاهرة اختيار المستفيدين من ماهية الكتب المتوافرة والمتاحة في مكتبة ما ، فنجد في كل بطاقة من بطاقات الكتب المصنفة في تلك المكتبة تصنيفاً محدداً : قصص وروايات ، كتب علمية ، كتب ثقافية ، كتب

اقتصادية ، كتب اجتماعية ، كتب سياسية ، كتب عامة ، صحف ، مجلات ، .... الخ ، لذا لا بد لنا من إعداد جدول تفريغ عددي يضم عمودين : الأول يبين تصنيف الكتب المتوافرة في المكتبة وفق ما صنفناه سابقاً ، والثاني عدد المستفيدين المقابل لكل صنف من الأصناف المشار إليها سابقاً ، ثم نبدأ بفرز الإجابات بوضع الإشارة في العمود المناسب والسطر المناسب من جدول التفريغ كالآتي :

الجدول رقم (1) : عدد المستفيدين من الكتب المتوافرة في المكتبة وفق أصناف تبويبها

أصناف الكتب	عدد المستفيدين	المجموع (التكرار)
قصص وروايات	//// ~~////~~	9
كتب علمية	////~~////~~ ~~////~~	18
كتب ثقافية	///—~~////~~ ~~////~~ ~~////~~	23
كتب اقتصادية	—~~////~~ ~~////~~ —~~////~~	15
كتب اجتماعية	///—~~////~~ —~~////~~	13
كتب سياسية	// —~~////~~ —~~////~~	12
كتب عامة	/ —~~////~~ —~~////~~	11
صحف	~~////~~ ~~////~~	10
مجلات	/// ~~////~~	8
المجموع		119

تُعد هذه الطريقة في التفريغ والتبويب يدوية وبدائية ، وتستخدم في حال كون العينة محدودة العدد ، أما في عمليات التبويب الكبيرة ، فإننا نلجأ إلى التبويب الآلي ، وإلى الاعتماد على الحاسبات الالكترونية القادرة على حل الكثير من مشكلات التفريغ والتبويب بأسرع وقت وأقل النفقات .

3-1-2- ماهية الجداول الإحصائية وشروطها

بعد الانتهاء من عملية جمع البيانات الإحصائية وتبويبها ، يقوم الباحث بعرضها ضمن جدول ، حيث ترتب بشكل متسلسل ومنطقي بغية السماح للقارئ بإجراء المقارنة ما بين الصفات المدروسة بسهولة ، وبأخذ صورة عددية عنها ، لنأخذ مثلاً البيانات المتعلقة بعناوين عدد من الكتب المتوافرة في إحدى المكتبات ، وكما هي معروضة في الجدول رقم (2) الآتي:

الجدول رقم (2) : يبين عدد الكتب المتوافرة في إحدى المكتبات تبعاً للعنوان

عنوان الكتاب	قصص أطفال	روايات	كتب دينية	كتب علمية	كتب بلغات أجنبية	مجلات	صحف
عدد الكتب	12	18	6	14	25	10	8

المصدر : فرضي

وعليه يُعرِّف الجدول الإحصائي **:** " بأنه عبارة عن ترتيب البيانات العددية بخطوط أفقيه وعمودية بشكل يسهل قراءتها في أي اتجاه من هذين الاتجاهين " .

وبشكل عام تكمن غاية الجدول الإحصائي : في اختصار البيانات وملاحظة صفاتها العامة . هذا ولكي نحصل على الغاية المرجو ة من الجدول الإحصائي ، لابد أن تتوافر فيه بعض الشروط الواجب مراعاتها عند إعداد ه ، ومنها :

1. كتابة عنوان الجدول بشكل واضح يتضمن محتواه ؛
2. كتابة مصدر الحصول على المعلومات في أسفل الجدول ، على اعتبار أن جميع المصادر ليست على درجة واحدة من الثِّقة ؛
3. إيراد بعض الملاحظات في أسفل الجدول لشرح بعض المعلومات الواردة فيه عند الضرورة ؛
4. وضع رقم للجدول، وإن ورد أكثر من جدول في بحث ما يعطى له رقم متسلسل ؛
5. عرض مجاميع أعمدة الجدول وسطوره ، لما في ذلك من فائدة لاسيما عند المقارنة ؛
6. ذكر الوحدة الإحصائية المستخدمة ؛
7. ترقيم أعمدة الجدول للإشارة إليها دون ذكر عنوان الجدول ؛
8. التناسق في إعداد الجدول الإحصائي مع البيانات المعطاة ؛
9. عند عرض البيانات الإحصائية يمكن غالباً تدوير الأرقام الواردة فيها ، لأجل الاختصار وسهولة القراءة ؛
10. إذا لم تتوافر بيانات في سطر أو خانة توضع ثلاث نقاط (...) أما إذا كانت المعلومة مساوية للصفر يوضع خط مستقيم صغير(-) .

3-1-3- أنواع الجداول الإحصائية

يمكن تصنيف الجداول الإحصائية حسب الهدف من إعدادها إلى جداول عامة وجداول خاصة.

- **الجداول العامة:** هي تلك الجداول التي تنشأ بدون أي هدف سوى تسجيل نتائج البيانات الإحصائية المجموعة بشكل يسهل استعمالها وبدون أن تكون خاضعة لأي أصول تحليلية ماعدا العد والاختصار، إلاّ أن غايتها تكمن في إمكانية استخدامها كأساس لدراسات لاحقة، وخير مثال على ذلك الجداول الإحصائية المنشورة في المجموعات الإحصائية ،والتقارير التي تنشرها بعض الهيئات أو الوزارات... الخ .
- **الجداول الخاصة:** وهي جداول مشتقة من البيانات المسجلة في الجداول العامة ، أي تلك التي تعرض ملخصاً عن الجداول العامة ، إما بشكل مجاميع أو أرقام قياسية ، وتكون غايتها التأكيد على بعض الحقائق التي لانكشف عنها بوضوح الجداول العامة ، أو عرض خلاصة المعلومات التي تم الحصول عليها نتيجة القيام بإجراء تحليل إحصائي معقد وشاق .

كما تقسم الجداول الإحصائية حسب طريقة العرض وعدد الصفات إلى :

1 - **الجدول البسيط**: وهو جدول تعرض فيه البيانات حسب صفة واحدة ، أي جدول يتكوّن من عمودين أحدهما يبين الحالات الممكنة ، والآخر يبين عدد المفردات المقابلة لكل حالة من الحالات ، وكمثال عليه، الجدول رقم (2) .

2 - **الجدول المزدوج**: وهو الجدول الذي تعرض فيه البيانات بدلالة صفتين بدلاً من صفة ، وكمثال على ذلك الآتي :

الحالة الزواجية / العمر بالسنوات	أعزب	متزوج	أرمل	مطلق
15				
16				
17				
.				
.				

3 - **الجدول المركب**: وهو جدول تعرض فيه البيانات لأكثر من صفتين .

الحالة الاجتماعية	أعزب		متزوج		أرمل		مطلق	
النوع / الحالة التعليمية	ذكور	إناث	ذكور	إناث	ذكور	إناث	ذكور	إناث
يقرأ ولا يكتب								
يكتب ولا يقرأ								
أمي								
ابتدائي								
.								
.								

3-1- 4- كيفية ترتيب البيانات الإحصائية وأساليب ترتيبها

إن أسلوب ترتيب البيانات الإحصائية داخل الجداول الإحصائية يتحدد بحسب طبيعة البيانات من جهة (جغرافية أو كمية أو زمنية ...) ، وبحسب الهدف الذي سنستخدم من أجله هذه البيانات من جهة أخرى ، فيما إذا ستعرض في جداول عامة أم في جداول خاصة ، هذا ومن أهم أساليب ترتيب البيانات الإحصائية:

- أسلوب الترتيب التاريخي : ويتم بإيراد البيانات في الجدول حسب تسلسلها التاريخي ، (من الأقدم إلى الأحدث أو من الأحدث إلى الأقدم) ، كأن ترتب الكتب في المكتبات حسب تاريخ إصدارها أو نشرها؛
- أسلوب الترتيب الأبجدي : وتشتمل ترتيب المعلومات ب حسب حروفها الأبجدية ، كما هي الحال في تبويب فهارس الكتب والمراجع في المكتبات ؛
- أسلوب الترتيب الكمي: يؤخذ بهذا الأسلوب عندما يكون عامل الكم أو الحجم هو الأساس المرغوب في تأكيده ، لذا ترتب المعلومات

وفق هذا الأسلوب إما بشكل تصاعدي (من الأصغر إلى الأكبر) أو بشكل تنازلي (من الأكبر إلى الأصغر) ، إذ يؤخذ هنا عدد الصفحات بالحسبان ، فيما إذا كنا أمام ترتيب الكتب المتوافرة في مكتبة ما وفقاً لحجمها ؛

- أسلوب الترتيب الجغرافي : كأن ترتب الكتب المتوافرة في مكتبة ما وفقاً للتقسيم الجغرافي لبلدان القارات الخمس ؛
- أسلوب الترتيب حسب عدد المؤشرات المراد تبويبها (مؤشر ، مؤشرين أو أكثر) .

3-1- 5 - الجداول التكرارية

3-1- 5-1- ماهية الجداول التكرارية

تصنف البيانات الإحصائية حسب طبيعتها إلى بيانات كمية وأخرى نوعية ، فالبيانات النوعية التي يطلق عليها أيضاً " النوعيات " وهي عبارة عن تلك الميزات التي قد يتصف بها الشيء أولا يتصف والتي لا يمكن قياسها (كاللغيوم ، والألوان ، والحالة الاجتماعية للفرد ، وأسماء الكتب ، وأسماء المؤلفين ، وأسماء المستفيدين من خدمات مكتبة ما ، . .. الخ) ، وهذا ما يطلق عليه في علم المكتبات والمعلومات " بالقياس الاسمي " الذي لا يتضمن أي دلالة إحصائية، فهو مجرد تخصيص أرقام للمفردات، لكي نتمكن من وضعها في قوائم تشير إلى مكان وجود المفردات على القائمة ، كأن نُعطي للنوع البشري : 1 = ذكر ، 2 = أنثى ، فلو حصل الشخص على الرقم (1) في متغير النوع ، فهذا يعني أنه ذكر، بمعنى أن هذا الرقم لا يشير إلى أي ترتيب نظامي فيما يتعلق بالنوع ، وعليه يمكننا قياس جمهور المكتبة اسمياً ، وذلك من خلال تصنيفه إلى : مستفيدين وغير مستفيدين ، وفئة المستفيدين تصنف بدورها إلى ذكور وإناث ، أو موظفين وعمال ، أما العاملين بالمكتبة فيتم تصنيفهم إلى ثلاث فئات : الكوادر المؤهلة ، والكوادر شبه المؤهلة ، والكوادر الكتابية .

تشكل مثل هذه التصنيفات الخطوات الأولية لأي مستوى من مستويات القياس ، واعتماداً على هذه التصنيفات ، يمكننا التوصل إلى إجابات عن بعض الأسئلة :

- أي الفئتين (المستفيدين ، غير المستفيدين) تشتمل على نسبة أكبر من الذكور أم من الإناث ؟
- أي الفئتين تحتوي على نسبة أكبر من خريجي الجامعات ؟
- ....

أما البيانات الكمية والتي نطلق عليها " المتغيرات " وهي عبارة عن القيم العددية لتلك الصفات المشتركة التي تتصف بها مجموعة من الأشياء بدرجات مختلفة ومقاسة.

مثال : عمال منشأة ما حسب العمر، وأسعار الكتب أو الصحف أو المجلات ، والأجور الشهرية للمكتبيين العاملين في مكتبة ما ، .... الخ ، وهنا يمكننا التمييز بين :

- المتغيرات المستمرة : وهي عبارة عن المتغيرات التي يمكن قياسها مهما صغرت قيمتها ، مثل: العمر ، والطول ، والوزن ،... الخ ؛
- المتغيرات المتقطعة : وهي عبارة عن متغيرات لا يمكن قياسها إلاَّ بوحدات إحصائية كاملة ، الطلبة، والسكان ، والكتب ، ... الخ .

هذا وإن الإحصاء يهتم بتحليل المتغيرات ، لذا وجب القيام بتبويبها ، الأمر الذي يتطلب ترتيب المفردات حسب الحجم وهذا ما يسمى " بالترتيب" ، ويدعى بعلم المكتبات والمعلومات " بالقياس الترتيبي " – وهو نظام ترتيبي بالأفضليات التي يتبعها المستفيد - ، الذي يُعد شكلاً مفيداً من أشكال العرض الجدولي سواءً بشكل تصاعدي أم بشكل تنازلي ، فهو يحافظ على هوية المفردات ،إذ بالإمكان التأكد من تكرار القيم ومدى التغير الحاصل فيها ،إلاَّ أنه يُشكّل عبئاً ثقيلاً عندما يكون عدد المفردات كبيراً جداً ، ويقتصر استعماله فقط على الملاحق الإحصائية التي تربط في نهاية الدراسة الإحصائية ، بهدف تزويد القارئ ببيانات خام تمكنه من التحقق بنفسه من نتائج الدراسة ، إذا ما أراد ذلك ، وللتغلب على محاذير الترتيب نقوم باختصاره إلى جداول تكرارية ، وعليه يُعرّف " الجدول التكراري " بأنه: " عبارة عن تنظيم المفردات في فئات متسلسلة بحيث يظهر أمام كل فئة عدد مرات تكرارها في البيان الأصلي الذي أُخذت منه " ، أو هو " الجدول الذي يبين أو يعطي فكرة واضحة عن التكرارات المقابل ة لكل فئة من فئات الظاهرة المدروسة " .

3-1- 5-2- كيفية إعداد الجداول التكرارية (مزاياها ومساوئها)

نتبع لإعداد جدول تكراري ثلاثة خطوات ، هي :

1. تحديد طول الفئة ؛
2. كيفية كتابة حدود الفئات ؛
3. إيجاد أو حساب التكرارات " التفريغ ".

ونظراً لأهمية كل خطوة من هذه الخطوات ، فإننا سنتناول كل منها كالآتي:

1 - **تحديد طول الفئة** : تعني فترة الفئة أو طولها المدى العددي الواقع بين حدي الفئة الأدنى والأعلى ، كأن نقول (0 وأقل من 5) ، فالحد الأدنى في هذه الفئة المعطاة هو (0) في حين حدها الأعلى هو (5) ، أي جميع القيم الواقعة ما بين الصفر والتي تقل عن الخمسة ، وينطبق الأمر نفسه على جميع فئات الجدول التكراري ، وبالتالي القيمتين الواقعتين على أقصى كل من طرفي الفئة تدعيان بحدي الفئة .

ولتحديد طول الفئة نعتمد على علاقة ستورجز "H. A. STURGES " كالآتي :

$$C = \frac{I}{1+3.322*(\log n)} = \frac{(M-m)}{1+3.322*(\log n)}$$

حيث تمثل :

C طول الفئة ؛

I المدى العددي [أي الفرق بين (M) أكبر مفردة في البيان الأصلي و (m) أصغر مفردة فيه)؛

$\log n$ لوغاريتم حجم العينة أو عدد المفردات الواردة في البيان الأصلي .

ويجب التنويه هنا إلى ضرورة تدوير الناتج بالنسبة لطول الفئة C إذا كان كسراً وتقريبه إلى أقرب عدد من مضاعفات العشرة بصورة عامة .

كما تجدر الإشارة بنا ، إلى أن علاقة ستورجز "STORGES" لإيجاد طول الفئة تخضع إلى اختبار الوسط الحسابي ما قبل التبويب وما بعده . أي يجب أن لا يختلف الوسط الحسابي المحسوب بدلالة الجدول التكراري عنه المحسوب بدلالة البيانات الواردة في البيان الأصلي بشكل جوهري ، فإذا كان الاختلاف جوهرياً وجب تعديل طول الفئة المستخرج من علاقة ستورجز من خلال تقسيم المدى العددي على عدد الفئات المراد أخذها في الجدول التكراري المنوي تشكيله ، أي :

$$C = \frac{(M-m)}{K} = \frac{I}{K}$$

حيث إن :

K هو عدد الفئات المراد أخذها بالحسبان .

كما ننوه هنا ، أن المقصود بالبيانات غير المبوبة هي البيانات ما قبل التبويب أو البيانات الواردة في البيان الأصلي ، في حين نقصد بالبيانات المبوبة تلك البيانات الواردة في الجدول التكراري أو البيانات ما بعد التبويب .

ومما سبق يمكننا القول : بأن طول الفئة يكون سوياً ، فيما إذا كان الوسط الحسابي ما قبل التبويب لا يختلف كثيراً عنه ما بعد التبويب ، وأن الجدول التكراري يكون جدولاً سوياً ، فيما إذا حقق الشرطين التاليين:

- شرط تعريف الجدول التكراري ، وه و عبارة عن الجدول الذي ينظم المفردات في فئات متسلسلة بحيث يظهر أمام كل فئة عدد مرات تكرارها في البيان الأصلي؛
- أن يكون الوسط الحسابي المحسوب ما قبل التبويب هو نفسه ما بعد التبويب.

ونشير هنا أنه ، كلما كان طول فئات الجدول التكراري صغيراً يكون الاحتمال قليلاً بوجود اختلاف بين الوسط الحسابي المحسوب من البيانات الإحصائية نفسها قبل التبويب عنه بعد التبويب ، وعلى العكس من ذلك ، كلما كان طول فئات الجدول التكراري أكبر ازداد الاحتمال بوجود اختلاف بين الوسط الحسابي المحسوب ما قبل التبويب وما بعده .

ويكون الجدول التكراري :

- **جدولاً منتظماً**: عندما تكون أطوال فئاته متساوية وملاحظة تمركز مفرداته حول قيمة متوسطها ومدى انتشارها أو ابتعادها حول هذه القيمة ؛
- **جدولاً غير منتظم** : عندما تكون أطوال فئاته غير م تساوية وصعوبة ملاحظة تمركز مفرداته حول متوسطها ومدى ابتعادها عن هذه القيمة .

2 - **تحديد كيفية كتابة حدود الفئات** : يتم اختيار حدِّي الفئة بحيث يكون وسطها يمثل المفردات الواقعة فيها ، واعتماداً على علاقة ستورجز في إيجاد طول الفئة يمكننا كتابة حدود فئات الجدول التكراري بوضوح تام منعاً لأي التباس أو تداخل فيما بينها ، نحدد نوعية المجال العددي ، وكالآتي :

- المجال المغلق $[a,b]$ وهو المجال الذي تنتمي إليه كافة القيم الواقعة بين a,b ، بما فيها القيم a,b ، أي : $a \le x \le b$ ؛
- المجال المفتوح $]a,b[$ وهو المجال الذي تنتمي إليه كافة القيم الواقعة بين a,b ، ما عدا a,b ، أي : $a \prec x \prec b$ ؛
- المجال نصف المفتوح من اليسار $[a,b[$ وهو المجال الذي تنتمي إليه كافة القيم الواقعة بين a,b ، بما فيها القيمة b ما عدا القيمة a ، أي : $a \prec x \le b$ ؛

- المجال نصف المفتوح من اليمين $[a,b[$ وهو المجال الذي تنتمي إليه كافة القيم الواقعة بين b,a ، بما فيها القيمة a ما عدا القيمة b ، أي : $a \le x \prec b$ ؛

وبناءً على ذلك نميّز بين :

- الجدول التكراري المغلق : يسمى الجدول التكراري مغلقاً ، إذا كان محدد المبدأ البداية والنهاية ؛
- الجدول التكراري المفتوح : يسمى الجدول التكراري مفتوحاً ، إذا كانت فئاته الحدية مفتوحة (... وأقل من 5 ، وأكثر من 60) .

وتجدر بنا الإشارة هنا ، إلى أنه عندما تتعلق البيانات بظاهرة ذات صفتين مشتركتين ، فإننا نلجأ إلى الجداول التكرارية المزدوجة ، كالآتي :

X / Y	$[3-5[$	$[5-7[$	$[7-9[$	$\sum$
$[3-5[$	3	2	-	5
$[5-7[$	2	4	6	12
$[7-9[$	-	1	3	4
$\sum$	5	7	9	21

3 - **إيجاد أو حساب التكرارات " التفريغ "** : بعد تحديد طول فئات الجدول التكراري وحدوده ، نقوم بتوزيع مفردات الترتيب الواردة في البيان الأصلي عن الظاهرة المدروسة وذلك من خلال قراءة المفردات مفردة تلو الأخرى ووضع إشارة (/) لكل مفردة في الفئة التي تقع ضمنها ، بحيث إذا ما حصلنا على أربعة خطوط مائلة وآخر أفقي ، شكلت حزمة مؤلفة من خمسة مفردات يسهل عدها وكتابة ما يقابلها رقمياً ، أي ترجمة هذه الإشارات على شكل أرقام تقابل فئات تضمنيها ، وهذا ما يطلق عليه " بالتفريغ " ، ويطلق على الأرقام المقابلة لكل فئة من فئات الجدول التكراري المنشئ " بالتكرارات " : التي هي عبارة عن عدد مرات ورود المفردات في البيان الأصلي الذي أُخذت منه ، [انظر ما سبق في الجدول رقم (1)] .

لنأخذ الآن مثالاً توضيحياً عن كيفية إعداد جدول تكراري ، ولتكن لدينا البيانات العددية عن أطوال 50 طالباً من طلاب قسم المكتبات الإناث ، كالآتي :

165	165	165	164	163	161	160	159	158	157	155	150
171	171	171	171	171	171	170	169	167	166	165	165
177	177	177	176	175	173	172	171	171	171	171	171
187	185	185	183	183	183	182	181	179	178	177	177
										190	190

1 - نحسب طول الفئة من علاقة ستورجز :

$$C = \frac{I = (M - m)}{1 + 3.322 * (\log n)} = \frac{190 - 150}{1 + 3.322 * (\log 50)} = \frac{40}{1 + 3.322 * (1.689)} = 6.02$$

وبتدوير الناتج يكون طول الفئة مساوياً لـ 6 ؛

2 - نقوم بكتابة حدود الفئات حيث نختار أصغر مفردة في الترتيب ، إذا لم تحدد بصورة مسبقة ، كحد أدنى للفئة الأولى ، إذ يضاف إليها طول الفئة 6 فنحصل على الحد الأعلى لهذه الفئة وهكذا على التوالي نحصل على حدود الفئات الأخرى، وكما هو موضح في الجدول التالي ؛

3 - إيجاد التكرارات " التفريغ " ، كالآتي :

فئات الطول (بالسنتيمتر)	التفريغ	التكرار (عدد الطلاب)
150 – 156	//	2
156 – 162	////	5
162 – 168	//// ////	9
168 – 174	//// //// ////	15
174 - 180	//// ////	9
180 – 186	//// //	7
186 - 192	///	3
$\sum$		50

وانطلاقاً من الجداول التكرارية البسيطة (التوزيعات التكرارية البسيطة) يمكننا إنشاء جداول تكرارية صاعدة و أخرى هابطة ، كما يمكننا عرض هذه التكرارات بشكل نسبة مئوية من مجموع التكرارات ، وفي هذه الحالة يسمى

الجدول التكراري أو التوزيع التكراري توزيعاً تكرارياً نسبياً ، كما هو مبين في الجدول رقم (3) الآتي:

الجدول رقم (3) : يبين التكرارات المطلقة والنسبية لأجور الموظفين في بلد ما لعام 2004

فئات أجور الموظفين **(بألوف الليرات السورية)**	التكرارات المطلقة			التكرارات النسبية		
	عدد العمال التكرار f_i	ت. ت. ص (تكرار تجميع صاعد) $f_i \uparrow$	ت. ت .هـ (تكرار تجميع هابط) $f_i \downarrow$	التكرار النسبي $f_i\%$	تكرار تجميع صاعد نسبي (ت. ت.ص.ن) $f_i \uparrow \%$	تكرار تجميع هابط نسبي (ت. ت.هـ.ن) $f_i \downarrow \%$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
5 -8	4	4	28	14.29	14.29	100.00
8 -11	7	11	24	25.00	39.29	85.71
11-14	9	20	17	32.14	71.43	60.71
14-17	6	26	8	21.43	92.86	28.57
17-20	2	28	2	7.14	100.00	7.14
المجموع	28			100.00		

واعتماداً على الجدول رقم (3) نستنتج المفاهيم الآتية :

- **التكرار النسبي**: هو عبارة عن حاصل قسمة تكرار كل فئة إلى مجموع التكرارات مضروبة في (100) ، أي أنه يعبر عن نسبة المفردات التي قيمها تنتمي إلى كل فئة ، كما في العمود (5) ؛

- **التكرار التجميع الصاعد** (ت . ت. ص) : هو عبارة عن عدد المفردات في جدول تكراري تقل عن الحد الأعلى ولكل فئة ، كما في العمود (3) ؛

- **التكرار التجميع الهابط** (ت . ت. هـ) : هو عبارة عن المفردات في جدول تكراري تزيد عن الحد الأدنى ولكل فئة ، كما في العمود (4) ؛

- **التكرار التجميعي الصاعد النسبي** (ت. ت. ص. ن) : يعبر عن نسبة المفردات التي تقل فيها عن الحد الأعلى ولكل فئة ، وذلك من خلال تقسيم التكرار التجميعي الصاعد على مجموع التكرارات ومن ثم ضربها في (100) ، كما في العمود (6) ؛

- **التكرار الهابط النسبي** (ت. ت. هـ . ن) : فهو يعطي نسبة المفردات التي تزيد قيمتها عن الحد الأدنى ولكل فئة ، وذلك من خلال تقسيم التكرار التجميعي الهابط على مجموع التكرارات مضروباً في (100) ، كما في العمود (7) .

هنا لا بد لنا ، من ذكر كل ما تتمتع به الجداول التكرارية من مزايا ومساوئ ، كما يلي :

- **مزايا الجداول التكرارية** :
  1. إن البيانات ملخصة بشكل سهل للعرض الموجز ؛
  2. يمكن ملاحظة مقدار التشتت بنظرة سريعة ؛
  3. ملاحظة أي ميل لتمركز التكرارات حول قيمة معينة بسهولة.
- **مساوئ الجداول التكرارية :**

1- تفقد المفردات معناها نظرا لدمج عدة أرقام أو مفردات معاً في فئة واحدة ، ولهذا السبب يعتبر وسط الفئة ممثلاً صحيحاً لقيمة مفردات هذه الفئة ، أي أنه لو كان متوسط مفردات كل فئة واردة في الترتيب توافق تماماً نقطة الوسط لكل فئة ، أي تعامل كل مفردة من مفردات الفئة معاملة وسط هذه الفئة ، لذا يعتبر الوسط الحسابي المحسوب قبل التبويب أكثر دقة منه بعد التبويب ؛

$$\text{وسطي الفئة } x_i = \text{الحد الأدنى للفئة} + \frac{\text{طول الفئة}}{2}$$

أو :

$$\text{وسطي الفئة } x_i = \frac{(\text{الحد الأعلى للفئة} + \text{الحد الأدنى للفئة})}{2}$$

$$\text{طول الفئة } C_i = \text{الحد الأعلى للفئة} - \text{الحد الأدنى للفئة}$$

2- صعوبة حساب احد مقاييس النزعة المركزية ، لاسيما الوسط الحسابي ، والوسط الهندسي ، خاصة عندما تكون إحدى فئاته مفتوحة .

3-1-5-3- خصائص وصفات التوزيعات التكرارية

تتمتع التوزيعات التكرارية بالخصائص الآتية :

1 - تظهر مفردات هذه التوزيعات ميلاً واضحاً للتمركز حول قيمة معينة ، وتتناقص تدريجياً عن كل من طرفي هذه القيمة ، وتدعى هذه القيمة " بالنزعة المركزية "، وتقاس بمقاييس يطلق عليها مقاييس النزعة المركزية ، ومنها : الوسط الحسابي ، الوسيط ،المنوال ، الوسط الهندسي ، الوسط التوافقي .

2 - بالرغم من ميل المفردات الواضح إلى التمركز حول قيمة معينة ، فإنها تبتعد وتختلف عنها بالوقت ذاته ، وهذا الاختلاف عن قيمة النزعة المركزية يدعى " بالتشتت " ، ويقاس بمقاييس تدعى بمقاييس التشتت ومنها: المدى ؛ الانحراف المتوسط ؛ الانحراف المعياري ؛ الانحراف الربيعي .

3 - تختلف التوزيعات التكرارية بعضها عن بعض اختلافاً كبيراً من حيث التماثل ، فالمنحنى المتماثل هو ذلك الذي تنطبق حوافيه على حوافي التوزيع الطبيعي، ويشبه شكله شكل الجرس المقلوب ، وإذا لم يكن متماثلاً قلنا عنه ، بأنه ملتوٍ ، أي يعاني من التواء – فالالتواء يقيس درجة التماثل أو البعد عنها - ، ويقاس الالتواء ، بمعامل يطلق عليه معامل الالتواء ويعطى كالآتي :

$$\text{معامل الالتواء} = \frac{\text{الوسط الحسابي} - \text{المنوال}}{\text{الانحراف المعياري}} = \frac{3(\text{الوسط الحسابي} - \text{الوسيط})}{\text{الانحراف المعياري}}$$

إذا كانت قيمة معامل الالتواء تساوي الصفر يكون التوزيع التكراري المدروس متماثلاً، أما إذا كانت قيمته أكبر من الصفر (موجباً) ، كان التوزيع التكراري ملتوياً نحو اليمين ، وإذا كانت قيمته أصغر من الصفر (سالباً) كان التوزيع التكراري ملتوياً نحو اليسار ، هذا ويمكننا التعرف على طبيعة التوزيع التكراري من خلال المقارنة بين مقاييس النزعة المركزية (الوسط الحسابي، والوسيط ، والمنوال) ، فنجد:

- إذا كان الوسط الحسابي أكبر من الوسيط وأكبر من المنوال ، كان التوزيع التكراري ملتوياً موجباً (التواء نحو اليمين وهو عبارة عن أي توزيع تكراري تتصاعد فيه القيم بشدة لتصل إلى الذروة ثُم تنحدر ببطء شديد ، فتكون فيه القمة نحو اليسار بعكس جهة الالتواء)، أي

$$\overline{X} \succ W \succ M_o$$

حيث تمثل : $\overline{X}$ الوسط الحسابي ؛ و W الوسيط ؛ و M_o المنوال ؛

- إذا كان الوسط الحسابي مساوياً للوسيط وللمنوال ، كان التوزيع التكراري متماثلاً ، أي التوزيع التكراري لا يعاني من أي التواء، فهو إذن أي توزيع تكراري تتصاعد فيه القيم ببطء شديد لتصل إلى الذروة ثُم تنحدر ببطء شديد لتلامس المحور الأفقي ، أي : $\overline{X} = W = M_o$ ؛

- إذا كان الوسط الحسابي أصغر من الوسيط وأصغر من المنوال ، كان التوزيع التكراري ملتوياً سالباً (التواء نحو اليسار، وهو عبارة عن أي توزيع تكراري تتصاعد فيه القيم ببطء شديد لتصل إلى الذروة ثُم تنحدر بشدة لتلامس المحور الأفقي ، فتكون فيه القمة نحو اليمين بعكس جهة الالتواء) ، أي $\overline{X} \prec W \prec M_o$.

4- تختلف التوزيعات التكرارية في العلاقات الموجودة بين التكرارات العظمى ومدى التشتت ، إذ تتمركز التكرارات في بعض التوزيعات بشكل كبير في فئات قليلة – مدى تشتتها ضيقاً جداً - ، وبعضها يُظهر تمركزاً أقل وتشتتاً أكبر ، فهذه العلاقة بين ارتفاع المنحنى التكراري والتشتت يطلق عليها (بالتفرطح أو لتفلطح) ويقاس بمقياس " معامل التفرطح " .

والتفرطح يقيس درجة تدبب قمة منحنى التوزيع التكراري بالمقارنة مع منحنى التوزيع الطبيعي، ومعامل التفرطح يعطى كالآتي :

$$\beta_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^4}{n\sigma^4} = 3$$

ومن هذه العلاقة يلاحظ الآتي :

1 - إذا كانت قيمة معامل التفرطح مساوية لـ 3 ، كان التوزيع التكراري متماثلاً (توزيع طبيعي) ؛

2 - إذا كانت قيمة معامل التفرطح أكبر من 3 ، كان التوزيع التكراري مدبباً؛

3 - إذا كان معامل التفرطح أصغر من 3 ، كان التوزيع التكراري مفرطحاً .

3-2- العرض البياني

3-2-1- ماهية العرض البياني وغايته

يتم عرض المعلومات التي تم جمعها وتبويبها بطرق مختلفة ، كما وأشرنا من خلال العرض الجدولي عند تبويبنا للأرقام المعبرة عن الظاهرة المدروسة ، الأمر

الذي يعطينا صورة عددية كاملة ومرتبة ومتسلسلة بشكل منطقي ومعقول يسهل علينا مجال المقارنة بين الصفات والخواص المدروسة ، لكننا نسعى في أحيان أخرى إلى عرض المعلومات بشكل يلفت النظر للقارئ بصورة سريعة وخاطفة تثبت في ذهنه وذاكرته ، فنعمد إلى استعمال طريقة العرض البياني .

فالعرض البياني هو عبارة عن التمثيل التصويري للبيانات الإحصائية ، الذي يأخذ عادة شكل مصورات ورسوم بيانية وأشكال هندسية ، يكون ا لهدف منه ، إبراز خصائص وميزات البيانات الإحصائية التي تكون أقل وضوحاً ، فيما لو عرضت بأشكال أخرى .

ويُعد ا لعرض البياني مكملاً للعرض الجدولي ، حيث أن العرض الجدولي للبيانات يشمل كل التفاصيل الصغيرة التي يسمح بها التبويب ، في حين أن العرض البياني يحرص على إبراز الاختلاف الأكثرأهمية.

وقد تطورت أساليب العرض البياني وتشعبت ومن الصعب حصرها جميعاً ، لذا سنكتفي بذكر كل منها وشرحها بإيجاز مقتضب : المصورات والخرائط الإحصائية ، والأعمدة البيانية ، ورسوم المساحات والأحجام البيانية " الدوائر الإحصائية "، بالإضافة إلى منحنيات التكرار ومنحنيات السلاسل الزمنية .

3-2-2 – أنواع العرض البياني

3-2-2-1- المصورات والخرائط الإحصائية

ويقوم هذا النوع من العرض البياني على تنقيط المصورات الجغرافية أو تظليلها ، وتستعمل في عرض أية بيانات إحصائية موزعة توزيعاً جغرافياً في فترة زمنية ما، مثل توزع عدد المكتبيين في مكتبات مدينة دمشق عام 2005 ،إلاّ أنه يتعذر في هذا النوع إمكانية مقارنة ظاهرة في فترتين زمنيتين على مصور واحد ، وإنما يعطى مصور خاص بكل فترة زمنية ، وتكون طرق التمثيل على المصورات بواسطة دوائر ومربعات أو بشكل نقط ، تمثل كل نقطة قيمة محددة .

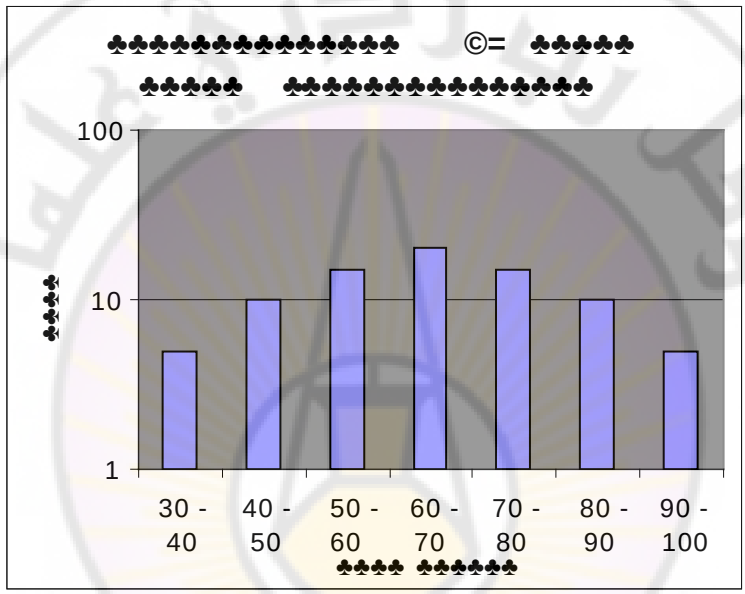
3-2-2-2- الأعمدة البيانية (البسيطة والمقسّمة)

يطلق على طريقة العرض البياني بواسطة الأعمدة البيانية ب المستطيلات ، لأنها تعتمد على المستطيلات بشكل أساسي ، إذ تكون قواعد ها على محور السينات وأطوالها تتناسب مع قيمة الظاهرة المدروسة .

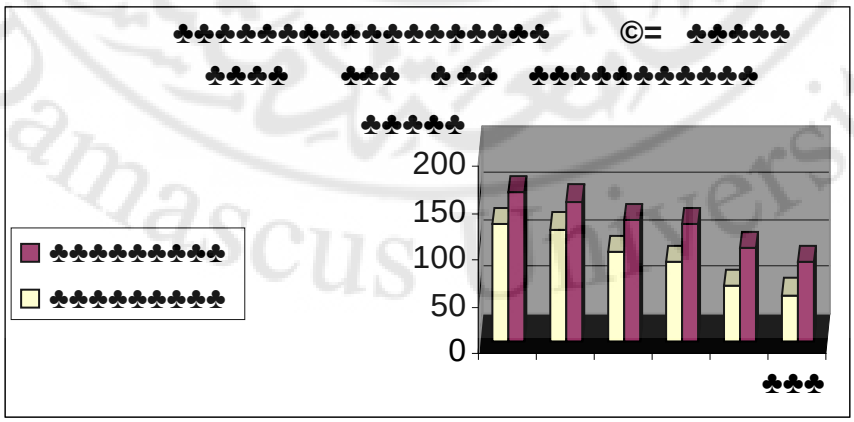
وعليه تعرّف الأعمدة البيانية : بأنها أبسط الأشكال الهندسية التي تستعمل في عرض البيانات الإحصائية ومقارنتها ، وهي عبارة عن مستطيلات ترسم بعرض

واحد ، ولكن بأطوال مختلفة ، بحيث يدل الطول على تكرار(عدد) المعلومة الإحصائية . وتستعمل الأعمدة البيانية :

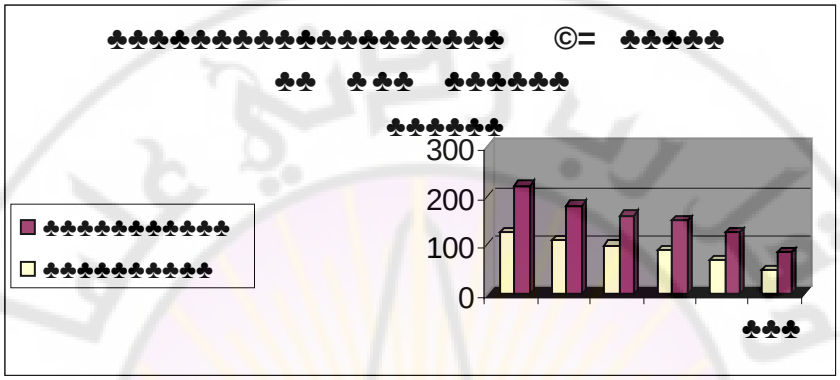
1 - للمقارنة بين قيم ظاهرة واحدة ولعدة فترات زمنية ، مثل تطور عدد المكتبات في مدينة دمشق خلال فترة زمنية ممتدة من عام 1990 – 2002 ، كما في الشكل رقم (3) الآتي :



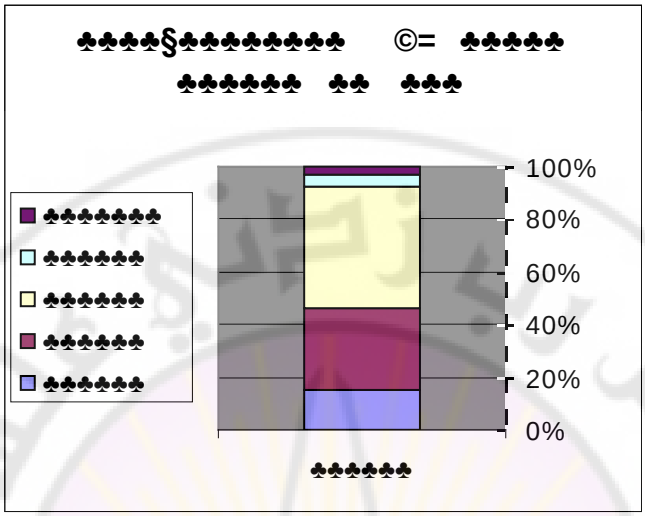
2 - للمقارنة بين صفتين من صفات ظاهرة ما أو عرض ظاهرتين أو أكثر ولنفس الفترة الزمنية ، كأن نقول توزع طلاب قسم المكتبات في جامعة دمشق في عام ما حسب النوع (ذكور وإناث) ، كما في الشكل رقم (4) الآتي :



3 -لمقارنة مجموعة من الظواهر مع مثيلاتها في فترات زمنية متجانسة ، كما في الشكل رقم (5) الآتي :



ونميزّ هنا بين: أعمدة بيانية بسيط ة وأخرى مقسِّمة ، فالأعمدة البيانية البسيطة تستعمل لتمثيل بيانات إحصائية لفترة زمنية واحدة أو لعرض سلسلة أو سلاسل زمنية بيانياً ، كما هي موضحة في الشكل رقم (3) ، هذا وعند عرض سلسلة زمنية بيانياً ، فإن نظام ترتيب الأعمدة يجب أن يكون بحسب التسلسل الزمني ، كما أن المسافات الفاصلة بين الأعمدة على محور السينات يجب أن تتناسب مع أطوال الفترات الزمنية، كما أنه يجب تمييز الأعمدة بألوان مختلفة بعضها عن بعض إذا كانت لظواهر مختلفة [كما في الشكلين رقم (4) و(5)]، هذا ونذكر هنا بالمحافظة على قاعدة الصفر التي هي نقطة التعامد الإحداثي ، وإغفالها يؤدي إلى إعطاء صورة مشوهة ومضلّلة ، أما الأعمدة البيانية المقسِّمة فهي تقوم على تجزئة مكونات القيم الإجمالية إلى أجزائها ، وبالتالي يقسم العمود الواحد إلى الأجزاء المكونة له عوضاً عن استخدام الأعمدة كوحدات كاملة إجمالية ، كمعرفة عدد المكتبات حسب تصنيفها في إحدى المدن في عام ما ، وهذا ما نعرضه في الشكل رقم (6) الآتي :



3-2-2-3- رسوم المساحات والأحجام البيانية " الدوائر الإحصائية "

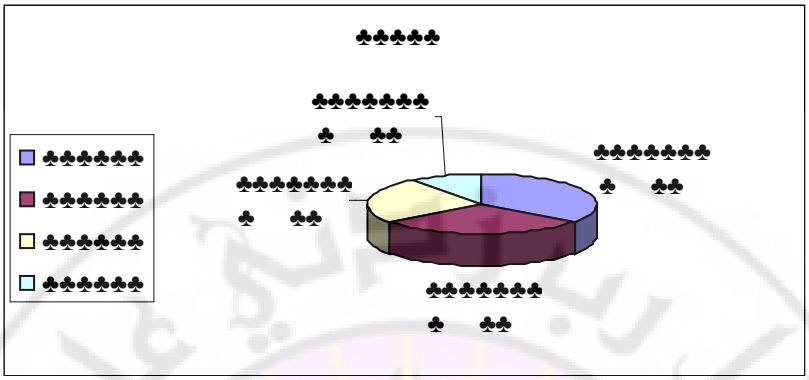
وذلك باستخدام رسوم بيانية للمساحة في شكل دوائر أو مستطيلات أو مربعات أو ... الخ، وتستعمل في :

1 - إظهار التركيب أو نسب التوزيع الهيكلي لمقومات ظاهرة من الظواهر المدروسة ، مثل توزع عدد طلاب قسم المكتبات في جامعة دمشق خلال السنوات الأربع في عام2004 ، ولمعرفة كيفية توزعهم بالتمثيل الدائري نتبع الخطوات الآتية :

- نرسم دائرة من حجم يتناسب ومكان العرض ؛
- نحسب :

طول القوس = (حجم الظاهرة / مجموع حجم الظاهرة المدروسة)×360

- وباستخدام الفرجار والمنقلة نحدد الزاوية الموافق لطول القوس على الدائرة المحددة . وعليه يكون لمثالنا المُعطى ، كما في الشكل رقم (7) الآتي :



2 - مقارنة عدة ظواهر مع بعضها: إذ تمثل كل ظاهرة بدائرة تتناسب مساحتها مع حجم الظاهرة ، ويتم معرفة نصف قطر كل دائرة ، بأخذ أحد نصفي الأقطار إلى الآخر الذي يجب أن يساوي جذر مساحة إحدى الدوائر إلى الأخرى على الترتيب ، لأن مساحة الدائرة تتناسب مع مربع نصف قطرها .

ليكن لدينا على سبيل المثال عدد المكتبات في مدينة دمشق عام 1996 يساوي 760 مكتبة ، وعددها عام 2000 يساوي 1500 مكتبة ، فيكون :

$\frac{R_2}{R_0} = \sqrt{\frac{1500}{760}} = 1.4$ ، وإذا اخترنا نصف قطر إحدى الدوائر بـ

(2 سم) ، فيكون نصف قطر الدائرة الثانية ($2 * 1.4 = 2.8$ سم) ، وما يؤخذ على هذا النوع صعوبة مقارنة مساحة الدوائر فيما بينها.

3-2-2-4- رسوم المنحنيات البيانية

تعرّف الرسوم البيانية للمنحنيات تبعاً لطبيعة البيانات الإحصائية المراد عرضها:" بأنها عبارة عن منحنيات الجداول التكرارية ومنحنيات السلاسل الزمنية " ، وهي تشمل كل الرسوم التي تُمثل فيها البيانات الإحصائية ، إما بشكل خط منحنٍ أو بشكل سلسلة من الخطوط المستقيمة المتصلة .

والمنحني هو عبارة عن خط منحنٍ ليس فيه أجزاء مستقيمة ، ويطلق على كل الخطوط البيانية التي تنشأ عن طريق وصل خطوط مستقيمة قصيرة بعضها ببعض ، هذا وتدعى منحنيات الجداول التكرارية بمنحنيات التكرار ، وهي تشتمل على معلومات لا تختلف عن تلك التي تعرضها الجداول التكرارية ، وتكمن غاية منحنيات التكرار في:

1- معرفة الاتجاه العام للظاهرة ؛

2- توضيح طبيعة التوزيع التكراري من حيث مقاييس (النزعة المركزية ، ومدى التشتت ، و التماثل) لوضع النموذج الرياضي الذي يخضع له التوزيع .

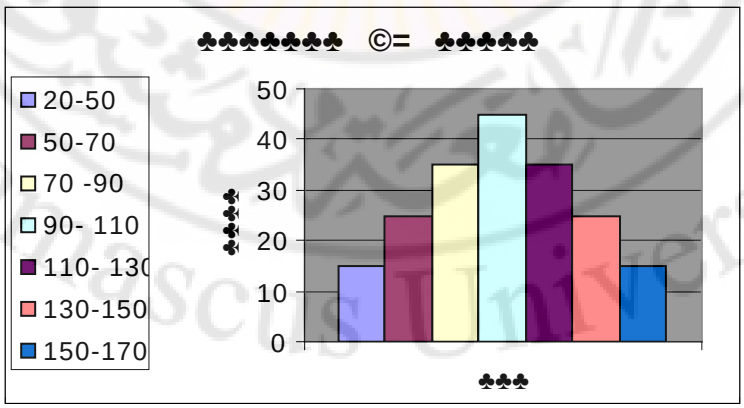
3-2-2-4-1- منحنيات التكرار وأنواعها

يهدف تمثيل التوزيعات التكرارية بيانياً إلى إعطاء فكرة سريعة وواضحة عن العلاقة بين الفئات وبين التكرارات ، إذ تُمثل الفئات على المحور الأفقي ، وتُمثل التكرارات على المحور العمودي ، ونشير هنا ، في حال كانت أطوال فئات الجداول التكرارية غير متساوية أي أن التوزيعات التكرارية غير منتظمة ، فيجب علينا القيام بتعديل التكرارات لكي تتناسب أطوال أضلاع المستطيلات وقواعدها في التمثيل البياني ، فنحسب ما يسمى بالتكرار المعدل ، والذي يعطى كالآتي :

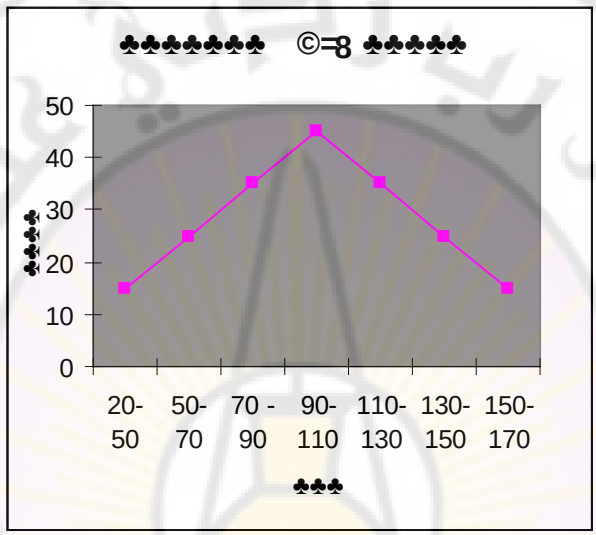
التكرار المعدل = (التكرار الأصلي ÷ طول الفئة)

وبناءً على ما سبق ، يتم عرض التوزيعات التكرارية بيانياً في ثلاثة أشكال ، هي : المدرج التكراري ، المضلع التكراري ، المنحنى التكراري.

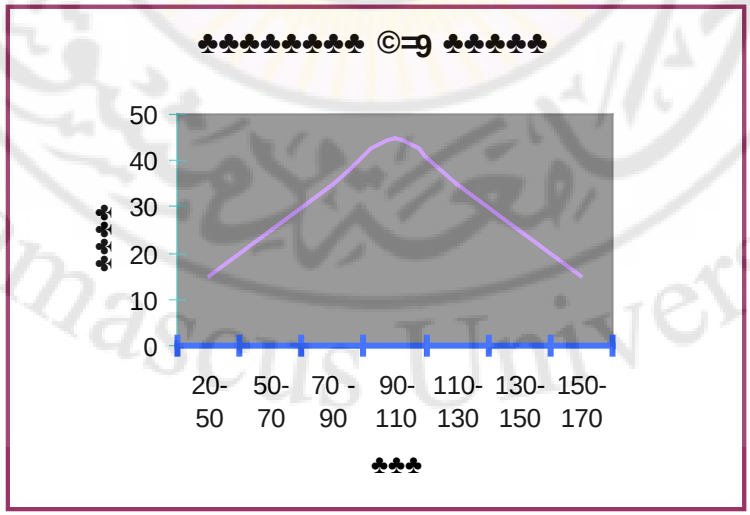
1- المدرج التكراري : وهو عبارة عن أعمدة بيانية متلاصقة بعضها ببعض ، حيث تُمثل الفئات (قيم المتغيرات) على المحور الأفقي والتكرارات على المحور العمودي بحيث تكون ارتفاعات الأعمدة متناسبة وتكرارات كل فئة ، ويستخدم المدرج التكراري في قياس متغيرات في مستوى القياس الاسمي للقياس ، كما في الشكل رقم (7) الآتي :



2- المضلع التكراري : نحدد على المحور الأفقي مركز كل فئة أو وسطها ، ثم نرفع عمودا بطول تكرار الفئة ، وبعدها نوصل نهايات هذه الأعمدة بخطوط مستقيمة فنحصل على ما يسمى بالمضلع التكراري ، الذي هو عبارة عن المدرج التكراري ، بعد أخذ مراكز أوساط الفئات وتوصل بخطوط مستقيمة من الأعلى ، كما في الشكل رقم (8) الآتي :

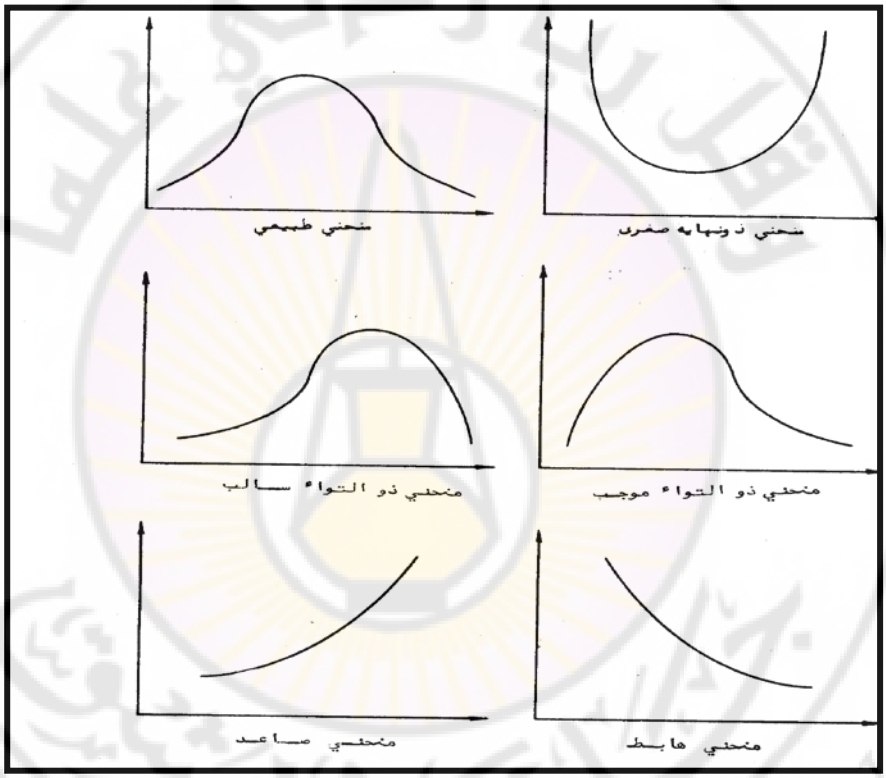


3- المنحنى التكراري : ما هو إلاَّ المضلع التكراري بعد أن أزيلت زواياه ، كما في الشكل رقم (9) الآتي :



وهناك أنواع عدة ومختلفة للمنحنيات التكرارية ، نذكر منها التالي :
المنحنى الطبيعي ، والمنحنى ذا الالتواء المنحني السالب أو الموجب ، والمنحنى الهابط ، والمنحنى الصاعد ، ومنحنى القطع المكافئ (ذا النهاية الصغرى) ، كما في الشكل رقم (10) الآتي :

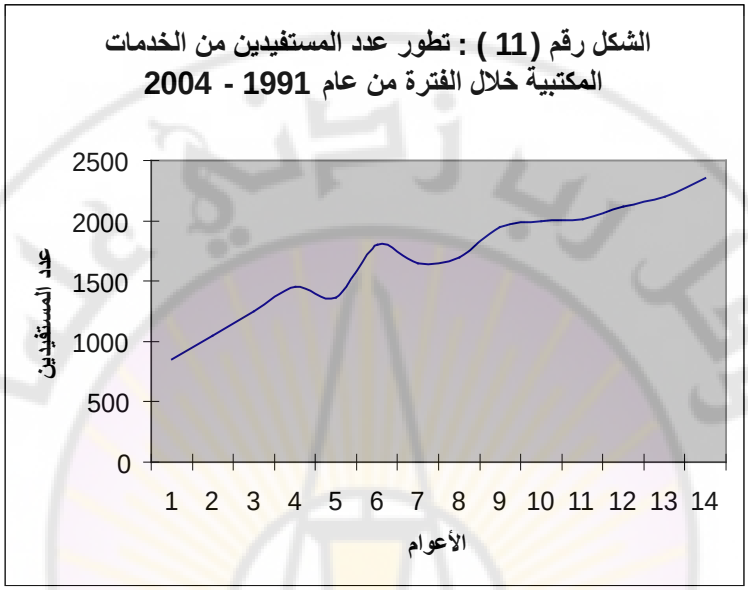
الشكل رقم (10) : أشكال المنحنى التكراري



3-2-2-4-2- منحنيات السلاسل الزمنية

تستند طريقة الخطوط البيانية في أساسها على إمكانية تمثيل أية علاقة بين متغيرين بصورة بيانية ، حيث نفترض أن أحد المتغيرين مستقل والآخر تابع له ، وتستعمل هذه الطريقة في السلاسل الزمنية ، حيث المحور الأفقي يمثل السنوات كمتغير مستقل ، والمحور العمودي يمثل تطور ظاهرة ما بمرور الزمن أي المتغير التابع ،

كما في الشكل رقم (11) الآتي ، الذي يوضح لنا تطور عدد المستفيدين من الخدمات المكتبية في جامعة ما خلال الفترة من 1991 – 2004 :



أما إذا كان لدينا سلسلتان زمنيتان أو أكثر ، ونريد تمثيلهما على محور إحداثي واحد بغرض المقارنة ، فإنه يمكننا ذلك ، باتباع إحدى الطريقتين لتوحيد وحدات القياس :

- بمساواة مقاييس الرسم على أساس الأوساط الحسابية ، أي أن نعدل قيم إحدى السلسلتين فقط ، وذلك من خلال الآتي :
  نقسم الوسط الحسابي للسلسلة الأولى على الوسط الحسابي للسلسلة الثانية ، ثم نضرب الناتج بكل قيمة من قيم السلسلة الثانية المراد مقارنة قيمها مع قيم السلسلة الأولى ؛
- بتمثيل بيانات كل من قيم السلسلتين كنسب مئوية .

3-3- أسئلة وتمارين

1. ما هي غاية التبويب ؟
2. ما هو الجدول الإحصائي ، وما هي غايته ؟
3. ما هي الشروط الواجب توافرها في الجدول الإحصائي السليم ؟
4. عرّف التبويب ؟

5. ما الفرق بين التبويب والترتيب ؟
6. ما هي خطوات التبويب ؟
7. عرّف الجداول الإحصائية حسب غايتها ؟
8. ما الفرق بين الجداول الإحصائية الخاصة والعامة من حيث الهدف ؟
9. كيف ترتب البيانات داخل الجداول الإحصائية ؟
10. ما هي أساليب الترتيب المستخدمة في الجداول الإحصائية ؟
11. عرّف الجداول الإحصائية حسب طريقة عرضها وعدد صفاتها ؟
12. عرّف البيانات الإحصائية بعد تصنيفها وفقاً لطبيعتها ؟
13. ما الفرق بين البيانات الكمية والبيانات النوعية ؟
14. ما هي المتغيرات وأنواعها ؟
15. عرف الجدول التكراري وبيّن خطوات إنشائه ؟
16. ما هو الاختبار الذي تخضع له علاقة ستورجز ؟
17. ما هو التكرار ؟ وماذا يقصد بالتفريغ ؟
18. متى يكون الجدول التكراري سوياً ؟
19. ما الفرق بين الجدول التكراري المنتظم وغير المنتظم ؟
20. كيف يتم اختيار حدي الفئة ؟
21. ما هي مزايا الجدول التكراري ومساوئه؟
22. متى يعتبر وسط الفئة ممثلاً صحيحاً لقيمة مفردات هذه الفئة ؟
23. لماذا تعتبر المقاييس الإحصائية المحسوبة قبل التبويب أكثر دقة منها بعد التبويب ؟
24. ما هي خصائص التوزيعات التكرارية ؟
25. عرّف المفاهيم الآتية : الالتواء ، التفرطح ، التكرار التجميعي الصاعد ، التكرار التجميعي الهابط ، التكرار التجميعي الصاعد النسبي ، التكرار التجميعي الهابط النسبي ، الرسوم البيانية تبعاً لأشكالها ، العرض البياني ، المدرج التكراري ، المضلع التكراري ، المنحنى التكراري ، القياس الاسمي، القياس الترتيبي .
26. ما الفرق بين الأعمدة البيانية البسيطة والأعمدة البيانية المقسّمة ؟
27. ما هو العرض البياني بواسطة الأعمدة البيانية ؟

28.ما هي قاعدة الصفر ، وما هي نتائج إغفالها ؟

29.ما فائدة منحنيات التكرار ؟

30. ما الفرق بين المضلع التكراري والمنحنى التكراري ؟

31.أخذت عينة عشوائية من إحدى المكتبات حجمها (40) ، وكانت كالآتي : رواية ، قصة أطفال ، رواية ، كتاب ديني ، قصة أطفال ، كتاب علمي ، مجلة ، صحيفة ، مجلة ، قصة أطفال ، رواية ، كتاب أجنبي ، كتاب أجنبي، مجلة ، صحيفة ، كتاب ديني ، كتاب علمي ، مجلة ، رواية ، مجلة ، رواية، قصة أطفال ، رواية ، كتاب ديني ، قصة أطفال ، كتاب علمي ، مجلة ، صحيفة ، مجلة ، قصة أطفال ، رواية ، كتاب أجنبي ، كتاب أجنبي، مجلة ، صحيفة ، كتاب ديني ، كتاب علمي ، مجلة ، رواية ، مجلة .

32.والمطلوب : قدم جدولاً لهذه البيانات ؟

33. لدينا الجدول الآتي المتضمن عدد المكتبات العامة في إحدى البلدان :

الأعوام	1995	1996	1997	1998	1999	2000
عدد المكتبات العامة	23	25	33	27	30	20

والمطلوب : اعرض بيانات الجدول أعلاه باستخدام الأعمدة البيانية البسيطة ؟

34.ليكن لدينا جدول يحتوي بيانات عن عدد المكتبات التخصصية والمكتبات الدينية في مدينة حلب خلال الفترة من عام 1995 - 2000 ، كالآتي :

الأعوام	1995	1996	1997	1998	1999	2000
عدد المكتبات التخصصية	25	28	33	35	30	27
عدد المكتبات الدينية	5	7	8	10	12	15

والمطلوب : اعرض بيانات الجدول باستخدام الأعمدة البيانية المقسّمة ؟ ومن ثم بطريقة الخطوط البيانية ؟

35. ليكن لدينا جدول يحتوي بيانات عن عدد المراجع في المجالات المختلفة في إحدى المكتبات الجامعية عام 2005 :

مجالات المراجع	علوم إنسانية	علوم اقتصادية	علوم طبيّة	علوم اجتماعية	علوم هندسية	علوم نفسية
عدد المراجع	100	25	65	45	35	25

والمطلوب : اعرض بيانات الجدول باستخدام طريقة الدوائر الإحصائية ؟

36. أخذت عينة عشوائية مؤلفة من (30) مكتبة من مكتبات إحدى البلدان ، فوجد أن متوسط الكتب المعارة أسبوعياً ، كالآتي :

152	151	150	148	146	145	140	137	135	132
140	168	167	166	165	162	162	160	158	155
170	168	169	169	166	168	171	155	145	155

<u>والمطلوب</u> :

- أنشىء الجدول التكراري لهذه المعلومات ؟
- أوجد الجدول التكراري التجميعي الصاعد والهابط ؟
- أوجد الجدول التكراري النسبي العادي والتجميعي الصاعد والهابط ؟
- ارسم كلا من : المدرج التكراري ، المضلع التكراري ، المنحنى التكراري.

37. ليكن لدينا التوزيع التكراري الآتي :

فئات الأسعار	]5 – 10[	]10 – 15[	]15 – 20[	]20 – 25[	]25 – 30[	]30 – 35[
عدد الكتب	15	25	70	90	80	80

<u>والمطلوب</u> :

- ارسم المدرج والمضلع والمنحنى لهذا التوزيع ؟
- أوجد عدد الكتب التي أسعارها تتراوح ما بين (15 – 25) وحدة نقدية ؟
- ما عدد الكتب التي يزيد سعرها عن (15) وحدة نقدية ؟
- ما عدد الكتب التي يقل سعرها عن (20) وحدة نقدية ؟
- ما نسبة الكتب التي تتراوح أسعارها ما بين (5 – 20) وحدة نقدية ؟
- ما نسبة الكتب التي يزيد سعرها عن (25) وحدة نقدية ؟
- ما نسبة الكتب التي يقل سعرها عن (30) وحدة نقدية .

وقل رب زدني علما
جامعة دمشق
Damascus University

الفصل الرابع
مقاييس التحليل الإحصائي
(لظاهرة واحدة)

4 1 - النسب والمعدلات
4-1-1- مفهوم النسب والمعدلات
4-1-2- استعمالات النسب والمعدلات
4-2- مقاييس النزعة المركزية
4-2-1- الوسط الحسابي
4-2-2- الوسيط
4-2-3- المنوال
4-3- العلاقة بين المتوسطات الثلاث " علاقة بيرسون "
4 4 مقاييس التشتت
4-4-1- المدى
4-4-2- الانحراف المتوسط
4-4-3- الانحراف الربيعي
4-4-4- الانحراف المعياري

Damascus University

وقل رب زدني علما
جامعة دمشق
Damascus University

تكلمنا في الفصول السابقة عن جمع المعلومات الإحصائية عن الظاهرة المدروسة وتبويبها في جداول إحصائية وتمثيلها بيانياً ، إلاَّ أن المعلومات التي نحصل عليها غير قابلة للتحليل العلمي الأولي ، وذلك لكونها تشكل حجماً كبيراً من القيم المطلقة، ولعدم إمكانية المقارنة بين عناصر الظاهرة المدروسة في المقارنات الآنية أو الزمنية لعدم وجود أساس مشترك ، لذا تمثل الخطوات السابقة من الوجهة الإحصائية المرحلة التمهيدية للتحليل الإحصائي ، الذي يهدف إلى الحصول على نتائج دقيقة وواضحة للغاية ، ولهذا يلجأ الباحث الإحصائي إلى تطبيق طرائق حسابية وإحصائية تساعد ه بدقة في تحقيق الهدف واتخاذ القرار السليم بشأن ما تؤول إليه الظاهرة المدروسة ، وذلك من خلال حساب النسب والمعدلات ، وحساب مقاييس النزعة المركزية ، وحساب التشتت، وحساب الالتواء والتفرطح .

4-1- النسب والمعدلات

4-1-1- مفهوم النسب والمعدلات

إن التمييز بين خصائص النسب والمعدلات من الأمور التي تصعب على الباحث العادي ، فكثيراً ما يخلط بين النسب والمعدلات ، لذلك يجب الإشارة إلى الفرق بينهما ، فالنسبة بين عددين هي حاصل قسمة أحدهما على الآخر ، أما المعدل فهو كثافة حدوث حدث معين خلال وحدة قياسية ، فمثلاً نقيس نسبة الإناث إلى مجموع الطلبة في صف ما ، أو نسبة عدد الإناث إلى عدد الذكور في منطقة ما ، أو نسبة عدد الكتب العلمية إلى مجموع الكتب المتوافرة في مكتبة ما ، أو نسبة عدد الكتب القصصية إلى عدد الكتب غير القصصية ، أما إذا قسمنا عدد الكيلومترات التي قطعتها سيارة على عدد الساعات التي استغرقتها الرحلة فنحصل عندئذٍ على معدل السرعة في الساعة ، أو إذا حسبنا كمية الأمطار في وحدة المساحة حصلنا على معدل هطول الأمطار ، كما أنه إذا قسمنا عدد المستفيدين من استعارة الكتب في مكتبة ما في اليوم أو في الأسبوع أو في ... إلى عدد الكتب المتوافرة في هذه المكتبة حصلنا على معدل الاستفادة من الكتب المستعارة اليومي أو الأسبوعي أو ... الخ .

إن استعمالات النسب والمعدلات في مجال الإحصاء الحيوي متعددة والتفريق بينها ضروري جداً ، فيطلق ، مثلاً ، لفظ المعدلات على ناتج قسمة عدد الحوادث الحيوية (مواليد ، وفيات ، زواج ، طلاق ، ... الخ) خلال فترة زمنية معينة – هي عادة السنة - على متوسط عدد السكان في تلك السنة ، لأنها تقيس سرعة الإنجاب أو الانقراض أكثر منها قياس العلاقة الموجودة بين عددين ، بينما يطلق لفظ " نسبة " على ناتج قسمة عدد الذكور إلى عدد السكان الكلي . غير أن

استعمالات النسب والمعدلات في مجال علم المكتبات والمعلومات يأخذ طابعاً آخر، لأن النسبة تأخذ معياراً لقياس قضية مسَّلم بها ، عندما تكون وحدات القياس ثابتة ومعيارية ، وتسمح لنا بإجراء تحليلات دقيقة ، فمثلاً ، إذا كانت المكتبة (A) تحتوي على (500000) كتاب ، والمكتبة (B) تحتوي على (400000) كتاب، والمكتبة (C) تحتوي على
(300000) كتاب ، فإنه يمكننا القول بأن :

- مقتنيات المكتبة (A) من الكتب أكبر من مقتنيات المكتبة (B) من الكتب بـ (100000) كتاب ، أي بنسبة (25%) ؛
- مقتنيات المكتبة (A) من الكتب أكبر من مقتنيات المكتبة (C) من الكتب بـ (200000) كتاب ، أي بنسبة (66.7%) ؛
- مقتنيات المكتبة (B) من الكتب أكبر من مقتنيات المكتبة (C) من الكتب بـ (100000) كتاب ، أي بنسبة (33.3%) ؛

ونشير هنا ، إلى أن قسمة عدد مقتنيات المكتبة (A) من الكتب على عدد مقتنيات المكتبة (B) من الكتب يدعى " بالتناسب " ، وحاصل القسمة إذا ما ضُرب في مئة أو في ألف سُمي بالنسبة المئوية أو بالنسبة الألفية .

ونشير هنا ، إلى أن النسبة المئوية تكون غير ملائمة في الحسابات القومية وفي بعض الدراسات التحليلية ، لسببن هما :

1 - النسب المئوية لا يضاف بعضها إلى بعض ؛

2 - النسب المئوية غير قابلة لعملية الانعكاس في الزمن .

لهذا ، إذا ما أردنا معرفة التغير الذي يطرأ على ظاهرة ما بين فترتين زمنيتين ، تطرقنا إلى ما يسمى " بنسبة التغير " ، الذي هو شكل من أشكال النسب ، وهو يعبر عن مدى الزيادة أو النقصان في الظاهرة المدروسة بين الفترتين ، ويمكن إيجاده من خلال قسمة قيمة الظاهرة في نهاية الفترة على قيمة الظاهرة نفسها في بداية الفترة ، أي :

$$\text{نسبة التغير} = \frac{\text{قيمة الظاهرة في نهاية الفترة}}{\text{قيمة الظاهرة نفسها في بداية الفترة}}$$

$$\text{معدل النمو خلال الفترة} = \frac{\text{قيمة الظاهرة في نهاية الفترة} - \text{قيمة الظاهرة في بداية الفترة}}{\text{قيمة الظاهرة في بداية الفترة}}$$

معدل النمو = نسبة التغير – 1
أو :
نسبة التغير = معدل النمو + 1

مثال (4-1) : تم ترقية أحد موظفي المكتبات إلى منصب إداري أعلى ، فأصبح يتقاضى مرتباً قدره (16000) وحدة نقدية ، بدلاً من (12000) وحدة نقدية ، فما هو معدل التغير الذي طرأ على مرتب هذا الموظف بعد ترقيته ؟
إن معدل النمو الذي طرأ على مرتب الموظف بعد ترقيته هو $\left(\frac{16000-12000}{12000}=33.3\%\right)$ ، وهذا يعني بأن مرتبه قد ازداد بعد ترقيته بنسبة (33.3%) عما كان عليه قبل الترقية ، أو أن مرتبه كان قبل ترقيته مساوياً لـ (75%) مما أصبح عليه بعد الترقية ، أو بعبارة أخرى ، كان مرتبه قبل الترقية أقل بـ (25%) ما بعد الترقية .

4-1-2- استعمالات النسب والمعدلات

تستعمل النسب والمعدلات في تحليل الظواهر بأشكال متعددة ، تبعاً لطبيعة الظاهرة موضوع الدراسة التي يراد القيام بها ، ومن أهم استعمالات النسب في :

- إعطاء فكرة عن التركيب الهيكلي لظاهرة ما : أي مقارنة الأجزاء بالكل بهدف إعطاء صورة واضحة عن التركيب الهيكلي ، حيث ننسب كل جزء من الظاهرة المدروسة إلى مجموع ما تشكله هذه الظاهرة المدروسة ، ثُم نضربها في مئة فنحصل على النسبة المئوية .
لنأخذ مثلاً على ذلك عدد الكتب المتوافرة في مكتبة كلية الآداب والعلوم الإنسانية لعام 2004 تبعاً لعنوان تصنيفها، كما هو موضح في الجدول رقم (4) الآتي :

الجدول رقم (4) : يبين توزع الكتب المصنفة تبعاً للعنوان في مكتبة كلية الآداب والعلوم الإنسانية عام 2004

عنوان الكتب المصنفة	عدد الكتب	% النسبة
قصص	1050	6.84
روايات	2500	16.29
كتب أجنبية	3200	20.85
كتب دينية	800	5.21
كتب اقتصادية	750	4.49
كتب اجتماعية	2550	16.60
كتب عامة	4500	29.32
المجموع	15350	100.00

المصدر : فرضي

- مقارنة التطور الحاصل في ظاهرة ما خلال فترة معينة من الزمن : حيث نفترض أن قيمة الظاهرة في إحدى السنوات هي الأساس ، ومن ثُم نقارن بقية قيم الظاهرة المدروسة بالنسبة لها . لنأخذ تطور الإنفاق على شراء الكتب لمكتبة ما خلال الفترة الممتدة من عام 1995 وحتى 2000 ،علماً بأن عام 1995 هو عام الأساس ، كما هو مبين في الجدول رقم (5) الآتي :

الجدول رقم (5) : يبين تطور الأنفاق على شراء الكتب خلال الفترة من 1995 – 2000 بالنسبة لعام 1995

الأعوام	الإنفاق على شراء الكتب (بألوف الوحدات النقدية)	التطور بالنسبة لعام 1995 (%)	الزيادة أو النقص (%)
1995	500	100.00	-
1996	650	130.00	+ 30
1997	850	170.00	+ 70
1998	1050	210.00	+ 110
1999	1250	250.00	+ 150
2000	1500	300.00	+ 200

المصدر : فرضي

- مقارنة التطور الحاصل في ظاهرة ما بين فترة وأخرى : وتتم المقارنة في هذه الحالة بنسبة قيمة الظاهرة في سنة ما إلى قيمتها في السنة التي تسبقها مباشرة ، بدلاً من نسبتها إلى أساس ثابت ، لنأخذ بيانات الجدول السابق رقم (5) ، فيكون التغير ، كما هو موضح في الجدول رقم (6) الآتي :

الجدول رقم (6) : تطور الإنفاق على شراء الكتب من عام لآخر

الأعوام	الإنفاق على شراء الكتب (بألوف الوحدات النقدية)	التطور من عام لآخر (%)	الزيادة أو النقص (%)
1995	500	-	-
1996	650	130.00	+ 30.00
1997	850	130.77	+ 30.77
1998	1050	123.53	+ 23.53
1999	1250	119.00	+ 19.00
2000	1500	120.00	+ 20.00

المصدر : فرضي

- جميع المقارنات زمنياً ومكانياً : إن الأرقام المطلقة غير مجدية في المقارنات المكانية والزمنية ، لأن الأساس في أغلب الأحيان غير موحد ، لذا نلجأ إلى ما يدعى بعلم الإحصاء " بالمعايرة " ؛
- النسب والمعدلات الحيوية : تلك التي تتعلق بالسكان وتركيبهم الهيكلي والحيوي ، كمعدلات الوفاة والخصوبة والزواج والطلاق والترمل وغيرها ، أي تلك المتوافرة في الدراسات السكانية والاجتماعية.

4- 2- مقاييس النزعة المركزية

عند تحليل التوزيعات التكرارية سواء كانت منفصلة أو مقارنة توزيعيين تكراريين أو أكثر ، وجب بالضرورة تلخيص كل توزيع من هذه التوزيعات التكرارية بمقاييس إحصائية وصفية لأجل الوصف الدقيق لأي توزيع تكراري ، ولهذا لابد لنا من استخدام مقاييس إحصائية ، لكي يكون الوصف كاملاً.

لاحظنا فيما سبق ، بأن إحدى خصائص التوزيعات التكرارية ، و هي ميل المفردات الواضح إلى التمركز حول قيمة معينة والتناقص التدريجي عن كل من جانبي هذه القيمة، التي أطلقنا عليها بما يسمى " بالنزعة المركزية " ، التي تقاس بمقاييس تدعى بمقاييس النزعة المركزية ، ومنها : - الوسط الحسابي ؛ - الوسيط؛ - المنوال ؛ الوسط الهندسي ؛ - الوسط التوافقي .

هذا وإن كل متوسط من المتوسطات المذكورة يقيس نقطة النزعة المركزية بطريقة مختلفة عن غيرها ، فيما إذا كان التوزيع التكراري غير متماثلاً ، وفي هذه الحالة، فإن كل متوسط له قيمة مختلفة عن غيره ، لذا وجب على الإحصائي اختيار المتوسط الملائم ، الذي يصف شكل النزعة المركزية وذلك تبعاً لطبيعة البيانات الإحصائية ونوعها، وهذا يتطلب منا التعرّف على ماهية كل مقياس من المقاييس المشار إليها، وسنقتصر في منهجنا هذا ، فقط على المتوسطات الثلاث الأولى (الوسط الحسابي ، الوسيط ، المنوال) التي سوف نشرحها بالتفصيل الدقيق ، كما سنرى لاحقاً .

إلاَّ أننا نشير هنا إلى أهمية مقاييس النزعة المركزية ، لقدرتها على إعطائنا فكرة عامة وسهلة وواضحة عن قيم الظاهرة المدروسة ، وسوف نقسم هذه المقاييس إلى الآتي :

- مقاييس كميِّة وتشتمل : الوسط الحسابي ، الوسط التربيعي ، الوسط الهندسي ، الوسط التوافقي ؛
- مقاييس موضعية أو مكانية وتشتمل : الوسيط ، الربيع الأول والربيع الثالث ؛
- مقاييس تكرارية وتشمل : المنوال .

4-2-1- الوسط الحسابي

4-2-1-1- مفهوم الوسط الحسابي

يعرف الوسط الحسابي $\overline{X}$ لمجموعة من القيم $X_1, X_2, X_3, ..., X_n$ بأنه حاصل قسمة مجموع هذه القيم على عددها ، أي :

$$\overline{X} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ويعني الوسط الحسابي ، بأنه يساوي قيمة كل مفردة من مفردات المجموعة ، فيما إذا كانت جميعها متماثلة وبقي عددها ومجموع قيمها ، أو بعبارة أخرى ، إنه تلك القيمة لو أعطيت لكل مفردة من مفردات المجموعة ، لكان مجموع هذه القيم الجديدة يساوي مجموع القيم الأصلية نفسها .

4-2-1-2- كيفية حساب الوسط الحسابي

ثَّمة أسلوبان لإيجاد قيمة الوسط الحسابي لبيانات غير مبوبة وثلاثة أساليب لإيجاده لبيانات مبوبة ، وكل من الأساليب المطبقة يؤدي إلى النتيجة نفسها التي يمكن الحصول عليها من إحداها، وسوف نلخص هذه الأساليب ، وحسب ما أشرنا إليه ، كالآتي :

الأسلوب	بيانات غير مبوبة	بيانات مبوبة
المباشر	$\overline{X}=\frac{x_1+x_2+x_3+...+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}$ حيث أن : x_i قيمة المفردة	$\overline{X}=\frac{f_1.x_1+f_2.x_2+f_3.x_3+...+f_nx_n}{\sum_{i=1}^{n}f_i}=\frac{\sum_{i=1}^{n}f_ix_i}{\sum_{i=1}^{n}f_i}$ حيث أن: x_i وسط الفئة ، و f_i تكرار الفئة
الانحرافات	$\overline{X}=\overline{x}_0+\frac{\sum_{i=1}^{n}(x_i-\overline{x}_0)}{n}=\overline{x}_0+\frac{\sum_{i=1}^{n}d_i}{n}$ حيث أن : $\overline{x}_0$ وسط حسابي فرضي ، و $d_i=(x_i-\overline{x}_0)$	$\overline{X}=\overline{x}_0+\frac{\sum_{i=1}^{n}f_i.(x_i-\overline{x}_0)}{\sum_{i=1}^{n}f_i}=\overline{x}_0+\frac{\sum_{i=1}^{n}f_i.d_i}{\sum_{i=1}^{n}f_i}$ حيث أن : $\overline{x}_0$ وسط حسابي فرضي ، و $d_i=(x_i-\overline{x}_0)$
الانحرافات المختصرة		$\overline{X}=\overline{x}_0+\frac{\sum_{i=1}^{n}f_i(\frac{x_i-\overline{x}_0}{c_i})}{\sum_{i=1}^{n}f_i}=\overline{x}_0+\frac{\sum_{i=1}^{n}f_i.d_i^{`}}{\sum_{i=1}^{n}f_i}.c_i$ حيث أن : $\overline{x}_0$ وسط حسابي فرضي ، و $d_i^{`}=(x_i-\overline{x}_0)/c_i$ ، و c_i طول الفئة

مثال(4-2): ليكن لدينا عدد المكتبات المتنقلة تبعاً لعدد مرات التوقف ، كما يلي : 5 ؛ 6 ؛ 9 ؛ 10 ؛ 14 ؛ 16 ، والمطلوب : إيجاد قيمة الوسط الحسابي لهذه البيانات بأسلوبي المباشر والانحرافات ؟

1- وفق الأسلوب المباشر يكون :

$$\overline{X}=\frac{x_1+x_2+x_3+...+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{5+6+9+10+14+16}{6}=\frac{60}{6}=10$$

وهذا يعني : أن المكتبة المتنقلة تتوقف بالمتوسط (10) مرات .

2- وفق أسلوب الانحرافات نعمد أولاً لتحديد قيمة الوسط الحسابي الفرضي $\overline{x}_0$ وليكن مساوياً لـ (9) ، ومن ثّم نوجد انحرافات كل قيمة عن الوسط الفرضي المُعطى ، أي $d_i=(x_i-\overline{x}_0)$ ، وهي على التوالي: (- 4 ، -3 ، 0 ، 1 ، 5 ، 7) فيكون المجموع الجبري لها مساوياً لـ (6) ، وبعدها نطبق العلاقة التي يمكننا وفقها إيجاد قيمة الوسط الحسابي ، فيكون :

$$\overline{X}=\overline{x}_0+\frac{\sum_{i=1}^{n}(x_i-\overline{x}_0)}{n}=\overline{x}_0+\frac{\sum_{i=1}^{n}d_i}{n}=9+\frac{6}{6}=9+1=10$$

وهي نفس النتيجة التي توصلنا إليها وفق الأسلوب المباشر .

مثال (4-3) : لتكن البيانات التي تبين لنا مقدار ما ينفقه الفرد الواحد أسبوعياً (بمئات الوحدات النقدية) على شرائه للمجلات، فأعطت الجدول التكراري الآتي :

فئات الإنفاق	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100	المجموع
عدد المجلات	8	10	20	10	8	56

والمطلوب : إيجاد قيمة الوسط الحسابي بالأساليب الثلاثة ؟

الحل : لكي نوجد الوسط الحسابي وفق الأساليب الثلاثة المشار إليها أعلاه ، وجب علينا أن نعتمد على بيانات الجدول المساعد الآتي ، علماً بأن الوسط الحسابي الفرضي المتخذ هو ($\bar{x}_0 = 75$) :

فئات الإنفاق	عدد المجلات f_i	وسط الفئة x_i	$f_i * x_i$	$d_i = (x_i - \bar{x}_0)$	$f_i d_i = f_i(x_i - \bar{x}_0)$	$d_i' = (x_i - \bar{x}_0)/c_i$	$f_i.d_i' = f_i.(x_i - \bar{x}_0)/c_i$
50 – 60	8	55	440	-20	-160	-2	-16
70-60	10	65	650	-10	-100	-1	-10
80-70	20	75	1500	0	0	0	0
90-80	10	85	850	10	100	1	10
100-90	8	95	760	20	160	2	16
$\sum$	**56**		**4200**	**0**	**0**	**0**	**0**

- نطبق الأسلوب المباشر:

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i * x_i}{\sum_{i=1}^{n} f_i} = \frac{4200}{56} = 75$$

أي لو كان الإنفاق على شراء المجلات متساوياً ، لكان متوسط ما ينفقه الفرد أسبوعياً على شرائه للمجلات هو 7500 وحدة نقدية .

- لو طبقنا أسلوب الانحرافات :

$$\bar{X} = \bar{x}_0 + \frac{\sum_{i=1}^{n} f_i.(x_i - \bar{x}_0)}{\sum_{i=1}^{n} f_i} = \bar{x}_0 + \frac{\sum_{i=1}^{n} f_i.d_i}{\sum_{i=1}^{n} f_i} = 75 + \frac{0}{56} = 75$$

- أما إذا طبقنا أسلوب الانحرافات المختصرة ، فيكون :

$$\overline{X} = \overline{x}_0 + \frac{\sum_{i=1}^{n} f_i(\frac{x_i - \overline{x}_0}{c_i})}{\sum_{i=1}^{n} f_i} = \overline{x}_0 + \frac{\sum_{i=1}^{n} f_i.d_i`}{\sum_{i=1}^{n} f_i}.c_i = 75 + \frac{0}{56} = 75$$

ونلاحظ من مقارنة النتائج التي حصلنا عليها من تطبيق الأساليب الثلاثة في حساب قيمة الوسط الحسابي ولنفس البيانات ، أننا قد حصلنا على نتيجة واحدة . وهنا يجب علينا الإشارة إلى أن أسلوب الانحرافات المختصرة لا يمكن تطبيقه ، إلاَّ إذا كانت أطوال فئات الجدول التكراري متساوية ، وفي مثل هذه الحالة ، أي عندما تكون أطوال فئات الجدول التكراري غير متساوية يمكننا تطبيق الأسلوبين الآخرين ، الأسلوب المباشر وأسلوب الانحرافات .

هذا ويجدر بنا أيضاً ، أن نشير إلى بعض ملاحظات :

1- عندما تكون القيم غير متساوية من حيث أهميتها ، وجب ترجيح أو تثّقيل تلك القيم بما يتناسب وأهمية كل منها بوزن ما ، وهذا ما يسمى " بالوسط الحسابي المثّقل أو المرجح ، ويعطى كالآتي :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i * w_i}{\sum_{i=1}^{n} w_i}$$

حيث تمثل w_i الوزن الذي ثُقلت بها جميع القيم .

2- عندما تتغير القيم عن بعضها البعض بعدد ثابت ، فإنها تؤلف سلسلة حسابية ، وعندها يحسب الوسط الحسابي بسهولة من خلال :

$$\overline{X} = \frac{X_{\max} + X_{\min}}{2}$$

حيث إن : $X_{\max}$ أكبر مفردة من قيم السلسلة و $X_{\min}$ أصغر مفردة من قيم نفس السلسلة .

مثال (4-4): لو أخذت ظاهرة المكتبات المتنقلة حسب مرات توقفها ، القيم الآتية: 1 ؛ 3 ؛ 5 ؛ 7 ؛ 9 ؛ 11 ؛ 13 ؛ 15 . لكانت قيمة الوسط الحسابي لها مباشرة تساوي لـثمانية مرات توقف ، أي : $\overline{X} = \frac{X_{\max} + X_{\min}}{2} = \frac{15+1}{2} = 8$ ؛

3- عندما يكون لدينا عينتان أو أكثر ، فإننا نميز حالتين لكل من البيانات المبوبة وغير المبوبة ، والتي نعرضها ، كما يلي :

الحالة	بيانات غير مبوبة	بيانات مبوبة
تساوي العينات في أحجامها	$\overline{X}_w = \frac{\overline{X}_1 + \overline{X}_2}{2}$	$\overline{X}_w = \frac{\overline{X}_1 + \overline{X}_2}{2}$
عدم تساوي العينات في أحجامها	$\overline{X}_w = \frac{n_1 * \overline{X}_1 + n_2 * \overline{X}_2}{n_1 + n_2}$	$\overline{X}_w = \frac{\overline{X}_1 * \sum f_1 + \overline{X}_2 * \sum f_2}{\sum f_1 + \sum f_2}$

4 - عندما تكون المعطيات الإحصائية المتوافرة بشكل تكرارات نسبية ، فعندها نطبق الآتي :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i . \frac{f_i}{\sum_{i=1}^{n} f_i} .k)}{\sum_{i=1}^{n} \frac{f_i}{\sum_{i=1}^{n} f_i} .k} = \frac{\sum_{i=1}^{n} x_i . f_r}{\sum f_r}$$

حيث إن : $f_r = \frac{f_i}{\sum_{i=1}^{n} f_i} * k$.

4 -2 -1 -3 خواص الوسط الحسابي

يتمتع الوسط الحسابي بعدد من الخواص ، نذكر منها ، الآتي :

خاصة 1- المجموع الجبري لانحرافات قيم مجموعة عن الوسط الحسابي يساوي الصفر ، أي :

- لبيانات غير مبوبة $\sum_{i=1}^{n} (x_i - \overline{X}) = 0$

- لبيانات مبوبة $\sum_{i=1}^{n} f_i (x_i - \overline{X}) = 0$

خاصة 2- لا تتغير قيمة الوسط الحسابي إذا ضربنا جميع التكرارات بقيمة ثابتة صحيحة أو كسرية ، أي :

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i (M * f_i)}{\sum_{i=1}^{n} M * f_i} = \frac{M \sum_{i=1}^{n} x_i * f_i}{M \sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i * \frac{f_i}{M}}{\sum_{i=1}^{n} \frac{f_i}{M}} = \frac{\frac{1}{M} * \sum_{i=1}^{n} x_i * f_i}{\frac{1}{M} * \sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i}$$

خاصة 3- إذا ضربنا أو قسمنا كل قيمة من قيم التوزيع التكراري بقيمة ثابتة ولتكن (r)، فإن قيمة الوسط الحسابي تزيد (تنقص) بمقدار (r) مرة ، أي :

$$r * \overline{X} = \frac{\sum_{i=1}^{n} x_i * r * f_i}{\sum_{i=1}^{n} f_i} = \frac{r \sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i} = r * \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i}$$

$$\frac{1}{r} * \overline{X} = \frac{\sum_{i=1}^{n} x_i * \frac{f_i}{r}}{\sum_{i=1}^{n} f_i} = \frac{\frac{1}{r} * \sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i} = \frac{1}{r} * \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i}$$

خاصة 4- إذا أضفنا إلى (طرحنا من) كل قيمة من قيم التوزيع التكراري قيمة معينة ، فإن الوسط الحسابي يزيد (ينقص) بالقيمة نفسها ، أي :

$$\overline{X} + \overline{X}` = \frac{\sum_{i=1}^{n} (x_i + \overline{X}`) * f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i} + \frac{\overline{X}` \sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i}$$

$$\overline{X} - \overline{X}` = \frac{\sum_{i=1}^{n} (x_i - \overline{X}`) f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i * f_i}{\sum_{i=1}^{n} f_i} - \frac{\overline{X}` \sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i}$$

خاصة 5- مجموع مربع انحرافات قيم مجموعة عن الوسط الحسابي هو أصغر من مجموع انحرافات قيم المجموعة عن أية قيمة أخرى ، أي :

- لبيانات غير مبوبة : $\sum_{i=1}^{n}(x_i - \overline{X})^2 = mim$

- لبيانات مبوبة : $\sum_{i=1}^{n} f_i(x_i - \overline{X})^2 = mim$

4-1-2-4- مزايا الوسط الحسابي ومساوئه

يتصف الوسط الحسابي بعدد من المزايا ، نذكر منها :

1 - وضوح معناه وتعريفه ، ومن أكثر المتوسطات استعمالاً ؛

2 - هو مفهوم محدد رياضياً بدقة ، ويتصف بميزات رياضية لا يتصف بها سواه من المتوسطات ، باستثناء الوسط الهندسي والوسط التوافقي ؛

3 - سهولة حسابه ، إلاَّ أنه يتطلب بعض الوقت ؛

4 - عندما يكون له قيمة تمثيلية ، يستعمل كضابط أو معيار تقارن به المفردات الأخرى في التوزيع، أو كمعيار للحكم على كافة مفردات التوزيع . ويكون للوسط الحسابي قيمة تمثيلية ، عندما يكون التوزيع التكراري لبيانات الظاهرة المدروسة متماثلاً ، لا يعاني من أي التواء ؛

5 - إذا سحبت عدة عينات من مجتمع ما، فإن الوسط الحسابي أقل تذبذب من بقية المتوسطات (كالوسيط والمنوال) ، لذا يعتبر أفضل تقدير لمتوسط المجتمع.

وبالرغم من المزايا التي يتمتع بها الوسط الحسابي ، غير أنه يتمتع أيضاً ببعض المساوئ ، ومنها :

1 - يتأثر بالقيم المتطرفة الموجودة على أحد طرفي التوزيع ، أي أن قيمة الوسط الحسابي تنجذب نحو القيم المتطرفة ، ويكون للوسط الحسابي قيمة تمثيلية ، فيما إذا كانت التوزيعات التكرارية متماثلة أو قريبة من التماثل ، أما في التوزيعات التكرارية الملتوية بشكل كبير ، يفقد الوسط الحسابي قيمته التمثيلية ؛

2 - عدم إمكانية حسابه ، عندما تكون إحدى فئات الجدول التكراري مفتوحة ، لأن حسابه يتطلب معرفة أوساط جميع فئات الجدول التكراري ، لذا نكون عند حالتين ، هما :

- إما أن يفترض بشكل تحكمي قيمة أوساط الفئات المفتوحة ؛
- أو أن تهمل الفئات المفتوحة بالكامل من الحساب .

وكلتا الحالتين غير مرغوب فيهما ، لذا يفضل افتراض قيمة معقولة لنقاط وسط الفئة المفتوحة ؛

3 - عدم إمكانية حسابه لبيانات نوعيّ ؛

4 - لا يمكن إيجاده بيانياً .

4- 2-2- الوسيط

4-2-2-1- تعريفه

إذا رتبنا قيم متغير بصورة تصاعدية أو تنازلية ، فقد يكون مهماً أن نعين وضعية قيمة معينة ، كأن نقول مثلاً : إن قيم ثلثي الظاهرة المدروسة هي أعلى من قيمة ما ، وثلثه أقل من هذه القيمة ، فمقاييس كهذه هي مقاييس وضعية ، والوسيط هو من أحد أهم هذه المقاييس ، حيث يُعرّف " الوسيط " لمجموعة من القيم : بأنه القيمة التي يتساوى على طرفيها عدد المفردات بشكل تصاعدي أو تنازلي ، ومن خلال هذا التعريف لا بد لنا من أن نقوم بترتيب البيانات إما بشكل تصاعدي أو بشكل تنازلي ، ومن ثّم نوجد ترتيب موقع الوسيط ، والذي يُعطى بالعلاقة :

$$T_w = \frac{n+1}{2}$$

حيث تمثل :

T_w ترتيب موقع الوسيط ؛ n حجم العينة أو عدد مفردات الظاهرة المدروسة .

4-2-2-2- كيفية حساب الوسيط لبيانات غير مبوبة ولبيانات مبوبة

لإيجاد قيمة الوسيط للبيانات غير المبوبة يجب علينا أولاً ترتيب تلك البيانات تصاعدياً أو تنازلياً ، ومن ثّم نوجد ترتيب موقع الوسيط ، والذي يُعطى بالعلاقة :

$$T_w = \frac{n+1}{2}$$

بفرض لدينا مجموعة من القيم $\{x_1, x_2, x_3, ..., x_n\}$ مرتبة بشكل تصاعدي ، أي $x_1 \le x_2 \le x_3 \le ... \le x_n$ ، وكان عدد القيم n عدداً فردياً ، فإننا نسمي وسيطاً العدد x_i الذي ترتيبه $T_w = \frac{n+1}{2}$ ، أما إذا كان عدد القيم n عدداً زوجياً ، فالوسيط هو بالتعريف الوسط الحسابي للقيمتين اللتين تقعان في الوسط ، أي الوسط الحسابي للعددين اللذين ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$.

مثال(4-5) : إذا كانت لدينا القيم (10 ، 9 ، 8 ، 7 ، 5 ، 4 ، 3 ، 2 ، 1) مرتبة بشكل تصاعدي ، فإن وسيط هذه المجموعة من القيم هو القيمة (5) ، لأن ترتيب

الوسيط $T_w = \frac{n+1}{2} = \frac{9+1}{2} = 5$ ، أما إذا أضفنا عدداً آخر إلى قيم المجموعة السابقة ، وليكن العدد (12) ، لأصبح عدد القيم المجموعة زوجياً ، ففي هذه الحالة يكون الوسيط لها مساوياً لقيمة الوسط الحسابي للعددين اللذين ترتيبهما $\frac{n}{2} = \frac{10}{2} = 5$ و $\frac{n}{2}+1 = \frac{10}{2}+1 = 6$ ، أي : $\frac{1}{2}(5+7) = 6$ ، وبناءً عليه ، فالبيانات التي عددها زوجي ، ليس لها وسيط ، وإنما وسيطها هو متوسط القيمتين اللتين يقع بينهما الوسيط .

ويمكننا إيجاد قيمة الوسيط لبيانات مبوبة (خاضعة لتوزيع أو جدول تكراري) ، بإحدى الطريقتين الآتيتين :

1 - بيانياً : تقوم هذه الطريقة على رسم كل من المنحنى التكراري التجميعي الصاعد والمنحنى التكراري الهابط ، فتكون نقطة تقاطع المنحنين بعد إسقاطها على المحور الأفقي ، هي قيمة الوسيط ، (انظر الشكل رقم (7)).

2 - رياضياً : تأخذ هذه الطريقة بالحسبان الفئة الوسيطة ، التي هي الفئة التي تحوي الوسيط والتي يكون تكرارها التجميعي الصاعد أكبر أو يساوي $\frac{\sum_{i=1}^{n} f_i}{2} = \frac{n}{2}$ ، (أي : $\sum_{i=1}^{n} f_i = n$) ، وذلك من خلال تطبيق العلاقة الآتية :

$$W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w$$

حيث تمثل :

W قيمة الوسيط ؛

L_w الحد الأدنى للفئة الوسيطية ؛

$\frac{\sum_{i=1}^{n} f_i}{2}$ ترتيب الوسيط ؛

$\sum_{i=1}^{w-1} f_i$ مجموع التكرارات التي تسبق الفئة الوسيطية ؛

f_w تكرار الفئة الوسيطية ؛

C_w طول الفئة الوسيطية .

ومن هذه العلاقة ، يمكننا أن نشرح خطوات حساب الوسيط لبيانات مبوبة ، كالآتي :

1 - نوجد التكرار التجميعي الصاعد ؛

2 - نوجد ترتيب الوسيط $\frac{\sum_{i=1}^{n} f_i}{2}$ ؛

3 - نحدد الفئة الوسيطية اعتماداً على التكرار التجميعي الصاعد ومن خلال تعريفها، فهي تمكننا من تحديد كل من : الحد الأدنى للفئة الوسيطية وطولها ومجموع التكرارات التي تسبقها ، بالإضافة لتكرار الفئة الوسيطية ؛

4 - نطبق العلاقة أعلاه .

مثال (4-6): أخذت عينة عشوائية من (100) مكتبي ، لمعرفة مقدار إنفاقه الشهري (بمئات الوحدات النقدية) على شرائه للمجلات ، فأعطت الجدول التكراري الآتي :

فئات الإنفاق	30 -40	50- 60	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100	المجموع
عدد المجلات	8	10	20	24	20	10	8	100

والمطلوب : إيجاد قيمة الوسيط لبيانات هذه العينة العشوائية بيانياً ورياضياً؟

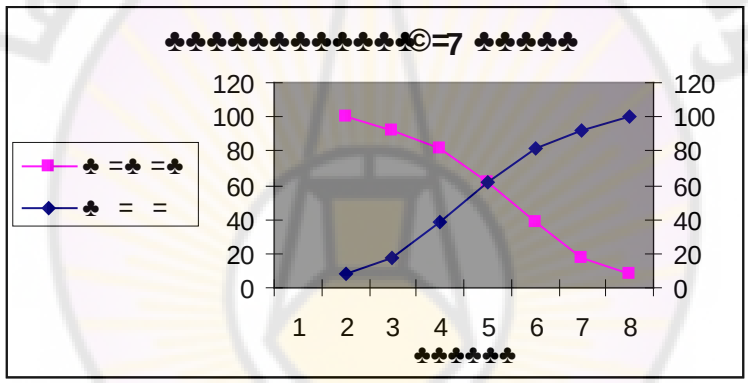
إن حساب الوسيط من جدول تكراري يعتمد بشكل أساسي على أن القيم موزعة توزيعاً منتظماً ضمن كل فئة ، ولحساب قيمته لبيانات المثال أعلاه ، نعتمد على الجدول الآتي:

فئات الأنفاق	عدد المجلات f_i	ت. ت. ص ↑ f_i	ت . ت . هـ ↓ f_i
30 – 40	8	8	100
40 – 50	10	18	92
50 – 60	20	38	82
60 – 70	24	62	62
70 – 80	20	82	38
80 – 90	10	92	18
90- 100	8	100	8
$\sum$	100		

بناءً على تعريف الفئة الوسيطية والتكرار التجميعي الصاعد نحدد الفئة الوسيطية ، وهي (70 – 60) ، وإذا ما طبقنا علاقة إيجاد الوسيط السابقة ، تكون قيمة الوسيط :

$$W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w = 60 + \frac{\frac{100}{2} - 38}{24} * 10 = 65$$

وتعني قيمته هنا ، أن نصف ما ينفقه المكتبي على شرائه للمجلات يكون أقل من 6500 وحدة نقدية والنصف الآخر أكبر من هذه القيمة . أما قيمة الوسيط البيانية ، فتكون كما في الشكل رقم (7) الآتي :



مثال(4-7): أخذت عينة عشوائية من المكتبات المدرسية في مدينة دمشق ، لمعرفة مشترياتها(بألوف الليرات) من الكتب المدرسية للعام الدراسي الجديد ، فأعطت الجدول التكراري الآتي :

فئات المشتريات	- 25	25 – 40	40 – 55	55 – 70	70 – 85	85 – 100
عدد المكتبات	20	22	35	35	28	18

والمطلوب : أوجد قيمة وسيط المشتريات لتلك المكتبات المدرسية ؟

الحل: ننشئ الجدول المساعد ، كالآتي :

فئات المشتريات	عدد المكتبات المدرسية f_i	ت. ت . ص ↑ f_i
- 25	20	20
25 – 40	22	42
40 – 55	35	77
55 – 70	35	112
70 – 85	28	140
85 – 100	18	158
$\sum$	158	

وبتطبيق علاقة إيجاد الوسيط ، علماً بأن الفئة الوسيطية هي : (55 – 70) ، فنجد :

$$W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w = 55 + \frac{\frac{158}{2} - 77}{35} * 15 = 55.86$$

وبالطريقة نفسها التي عرفنا بها الوسيط ، يمكننا أن نعرف الربيع الأول الذي هو القيمة التي ترتيبها $\frac{n}{4}$ ، أي القيمة التي تقع بحيث يكون ربع المفردات أقل منه وثلاثة أرباع المفردات أكبر ، ونرمز له بـ Q_1 ، والربيع الثالث هو القيمة التي ترتيبها $\frac{3n}{4}$ ، أي القيمة التي تقع بحيث يكون ثلاثة أرباع المفردات أصغر منه والربع الآخر أكبر منه ، ونرمز له بـ Q_3 ، لأن الوسيط في الواقع ما هو إلاَّ الربيع الثاني الذي ترتيبه $\frac{2n}{4} = \frac{n}{2}$ ، هذا وسوف نتطرق لكيفية إيجادهما ثانية ، عندما نوجد أحد مقاييس التشتت ألاَّ وهو الانحراف الربيعي الذي يعتمد عليهما ، وبالطريقة نفسها يمكننا تعريف العشر الأول بترتيب $\frac{n}{10}$ وهكذا دواليك .... ، لأن جميع هذه القيم هي قيم موضعية .

4-2-2-3- مزايا الوسيط ومساوئه

يتمتع الوسيط بعدة مزايا ، منها :

1- إن مجموع القيم عن وسيطها بالقيمة المطلقة هي أقل من مجموع انحرافات تلك القيم عن أيّ قيمة أخرى بالقيمة المطلقة :

$$\sum|X_i - W| \prec \sum|X_i - K|$$

2- يتحرر من عيوب الوسط الحسابي ، أي عدم تأثره النسبي بوجود قيم متطرفة على أحد جانبي التوزيع ؛

3- إمكانية حسابه في جدول تكراري ، فيما إذا كانت إحدى فئاته مفتوحة ، شريطة أن لا تكون الفئة الوسيطية هي الفئة المفتوحة ؛

4- سهل الفهم والحساب ، غير أنه يتطلب جزءاً من الوقت ؛

5- يمكن تحديد الوسيط لمجموعة من السلاسل ، ولو لم تكن قابلة للقياس الكمي .

كما أنه يتصف ببعض المساوئ ، نذكر منها :

1 -طرق حسابه تقريبيّ لاتصافه بميزات جبرية قليلة مقارنة مع تلك التي يتصف بها الوسط الحسابي ،لأن جداء الوسيط في مجموع التكرارات يُعطي قيمة مختلفة عن مجموع قيم الترتيب ، ويكون ذلك صحيحاً فقط عندما يكون التوزيع التكراري متماثلاً ؛

2 -قابل للتعرض أكثر من بقية المتوسطات لخطأ الصدفة ، خاصة إذا كان عدد المفردات صغيراً ، فهو إذن لا يصلح كمقياس للنزعة المركزية ، عندما يكون عدد المفردات صغيراً ، لأن قيمته غير مستقرة ، فهي تختلف اختلافاً كبيراً كلما أضيفت مفردات جديدة إلى العينة ؛

3 - وما يمكن ملاحظته ، أنه إذا كان مقدار الالتواء شديداً فإن قيمة الوسيط تنجذب باتجاه القيم المتطرفة ، ويكون أقرب إلى نقطة التمركز العظمى ودوماً بالمقارنة مع الوسط الحسابي .

4-2-3- المنوال

4-2-3-1- مفهوم المنوال وكيفية حسابه

يُعرّف المنوال بأنه : " عبارة عن القيمة الأكثر تكراراً واستعمالاً وشيوعاً بين مفردات المجموعة " ، فمثلاً لدينا هذه المجموعة من القيم :

- 8 ؛ 2 ؛ 3 ؛ 4 ؛ 6 ، ليس لها أي منوال ؛
- 8 ؛ 8 ؛ 5 ؛ 4 ؛ 8 ، لهذه المجموعة من القيم لها منوال واحد هو القيمة 8 وهو القيمة الأكثر تكراراً بين مفردات المجموعة ؛
- وقد يكون لبعض البيانات أكثر من منوال واحد ، كأن يكون لدينا على سبيل المثال القيم الآتية :

5 ؛4 ؛5 ؛ 2؛ 3؛ 3؛ 4، فلهذه القيم لها (3) منوال ، وهي (3 ؛ 4 ؛ 5)، وفي هذه الحالة عندما يكون لمجموعة من بيانات القيم أكثر من منوال

واحد، فإنه يدل على أن هذه البيانات غير متجانسة فيما بينها أو أنها قد سحبت من مجتمع إحصائي آخر غير المجتمع الإحصائي المدروس .

إن قيمة المنوال بيانياً هي القيمة على المحور الأفقي للجملة الإحداثية التي تقابل أعلى نقطة على المنحنى التكراري ، وهو بعكس الوسط الحسابي والوسيط ، لأن قيمته لا تتأثر أبداً وبأية حال من الأحوال بقيم المفردات الأخرى باستثناء القيم المجاورة له مباشرة ، لذا يُعد المنوال من أكثر المتوسطات صحة من حيث إعطاؤه القيمة الممثلة للنزعة المركزية ، ويستعمل للبيانات الكمية والنوعية معاً وأكثر للبيانات النوعية .

لإيجاد المنوال لبيانات غير مبوبة يتم من خلال تعريفه ، فمثلاً لدينا القيم الآتية ، التي تمثل عدد المكتبات المتنقلة وعدد مرات التوقف في إحدى البلدان:

عدد المكتبات المتنقلة	1	2	3	4	5	6	7	8	9
عدد مرات التوقف	10	12	18	25	40	35	20	15	9

من ملاحظة القيم الواردة في الجدول ، نجد أن المكتبة المتنقلة (5) هي المكتبة الأكثر توقفاً ، وهي تكون قيمة المنوال ، وبخاصة عندما يكون السؤال ما هي المكتبة المتنقلة الأكثر توقفاً ؟

غير أننا نجد عدة طرائق لإيجاد المنوال لبيانات مبوبة ، ومنها: طريقة الرافعة – طريقة الفروق " طريقة بيرسون " – الطريقة البيانية .

1 - طريقة الرافعة : تعتمد هذه الطريقة على تكرارات كل من الفئة المنوالية والتي تسبقها والتي تليها ، بالإضافة إلى طول الفئة المنوالية ، أي :

$$M_0 = L_{M_0} + \frac{f_{m2}}{f_{m2} + f_{m1}} * c_{M_0}$$

حيث إن :

M_0 قيمة المنوال ؛

L_{M_0} الحد الأدنى للفئة المنوالية ، والفئة المنوالية هي الفئة المقابلة لأكبر تكرار في الجدول التكراري ؛

f_{m1} التكرار السابق للفئة المنوالية ؛

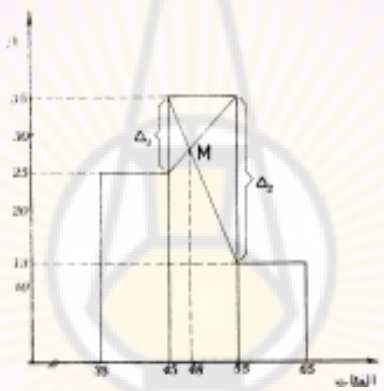
f_{m2} التكرار اللاحق للفئة المنوالية ؛

C_{M_0} طول الفئة المنوالية .

وتعتبر هذه الطريقة قليلة التطبيق ، لأنها تهمل تكرار الفئة المنوالية من الحسبان .

2 - طريقة الفروق " طريقة بيرسون " : تعتبر هذه الطريقة الأكثر استعمالاً من الطرائق الأخرى التي تم ذكرها ، وتتلخص بتحديد موقع المنوال في الفئة المنوالية ، ويتم هذا بتقسيم هذه الفئة إلى قسمين متناسبين مع الفرق بين تكرار الفئة المنوالية والفئة ما قبل المنوالية أولاً ، والفرق بين تكرار الفئة المنوالية والفئة ما بعد المنوالية ثانياً ، وتتفق هذه الطريقة مع الطريقة البيانية ، التي تقوم على رسم المدرج التكراري لبيانات جدول تكراري المراد إنشاؤه ، حيث نصل بخط مستقيم بين نهاية الفئة ما قبل المنوالية ونهاية الفئة المنوالية، كما نصل بخط مستقيم آخر بين بداية الفئة المنوالية وبداية الفئة ما بعد المنوالية ، فنقطة تقاطع المستقيمين مع بعضهما ، ولتكن النقطة M ، تُسقط على المحور الأفقي ، فتكون هي قيمة المنوال ، كما في الشكل رقم (12) الآتي :

f_i



الشكل رقم (12) : تحديد قيمة المنوال بيانياً

ويمكن تحديد قيمة المنوال رياضياً ، وكما أوجدها " بيرسون " بالاعتماد على الطريقة البيانية الأنفة الذكر ، والتي أعطاها بالعلاقة الآتية :

$$M_0 = L_{M_0} + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c_{M_0}$$

حيث تمثل :

M_0 قيمة المنوال ؛

L_{M_0} الحد الأدنى للفئة المنوالية ؛

Δ_1 الفرق بين تكرار الفئة المنوالية وتكرار الفئة ما قبل المنوالية ؛

Δ_2 الفرق بين تكرار الفئة المنوالية وتكرار الفئة ما بعد المنوالية ؛

C_{M_0} طول الفئة المنوالية .

مثال (4-8) : في دراسة إحصائية عن أعمار المستفيدين من الكتب المستعارة من مكتبة مركز البحوث والدراسات الاستراتيجية في بلد ما، أخذت عينة عشوائية مؤلفة من (200) من المستفيدين ، فأعطت بياناتها الجدول التكراري الآتي :

فئات الأعمار	20-30	30-40	40-50	50-60	60-70	70-80
عدد الكتب المستعارة	20	30	50	40	35	25

والمطلوب : أوجد قيمة المنوال لبيانات الجدول التكراري أعلاه ؟

الحل : لإيجاد قيمة المنوال نحدد الفئة المنوالية وهي المقابلة لأكبر تكرار من تكرارات الجدول التكراري المعطى ، وهي (50 – 40) ، ومن ثّم نطبق العلاقة الآتية :

$$M_0 = L_{M_0} + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c_{M_0} = 40 + \frac{(50-30)}{(50-30)+(50-40)} * 10 = 46.67 \text{ سنة}$$

وهذا يعني أن أغلب الكتب المستعارة من قبل المستفيدين أعمارهم 46.67 سنة .

4-2-3-2- مزايا المنوال ومساوئه

يتصف المنوال ببعض المزايا ، نذكر منها :

1 - أنه سهل الفهم ولا يتطلب أكثر من ذكر تعريفه لجعل معناه واضحاً ؛
2 - إن قيمته لا تتأثر بالقيم المتطرفة ؛
3 - إمكانية حسابه من جدول تكراري مفتوح ، شرط أن لا تكون الفئة المعنية في إيجاده هي الفئة المفتوحة؛
4 - له فائدة للبيانات النوعية ، مثل : اللون ، والحالة الاجتماعية ، ... الخ .

وبالرغم من المزايا التي يتصف بها ، فهو يتصف ببعض المساوئ ، ومنها :

1 - يتم حسابه بأساليب كلها تقريبية ، يمكن أن تعطي نتائج مختلفة اختلافاً جوهرياً ، عندما تطبق على بيانات الجدول التكراري الواحد ، وهذا يعني بأن قيمته النهائية تتوقف على كيفية إعداد الجدول التكراري من حيث طول الفئة وتحديد حدودها ، فهو إذن مقياساً غير مستقر للنزعة المركزية ؛
2 - ليس له أي معنى ، إذا لم يكن هناك تمركز قوي في توزيع الظاهرة المدروسة ، فهو إذن لا يتمتع بميزات جبرية كالوسط الحسابي ؛
3 - لا يمكن إيجاد قيمته في التوزيعات التكرارية ذات المنحنى النوني.

4- 3- العلاقة بين المتوسطات الثلاث " علاقة بيرسون "

عندما تكون التوزيعات التكرارية متماثلة أو قريبة من التماثل ، تكون قيمة المتوسطات الثلاثة متطابقة ، وتكون قيمة الوسيط متوسطة ما بين الوسط الحسابي والمنوال ، وهذا ما يزيد من أهمية الوسيط .

وبناءً على ذلك ، أوجد " كارل بيرسون " علاقة رياضية تربط بين هذه المتوسطات الثلاث ، والتي تُعطى كالآتي :

$$(\overline{X} - M_0) = 3(\overline{X} - W)$$

واعتماداً على هذه العلاقة ، يمكن حساب أحد المتوسطات بدلالة المتوسطين الآخرين ، لكن فقط في التوزيعات التكرارية المتماثلة ، أي :

$$\overline{X} = W = M_0$$

$$\overline{X} = \frac{3W - M_0}{2}$$

$$W = \frac{2\overline{X} + M_0}{3}$$

وتجدر الإشارة هنا ، أنه لمعرفة طبيعة التوزيع التكراري ، عما إذا كان متماثلاً أو بعيداً عن التماثل ، يكفي أن نحسب على الأقل مقياسين من مقاييس النزعة المركزية ، وهذا يتوقف على نوع الجدول التكراري ، عما إذا كان مفتوحاً(نأخذ الوسيط والمنوال) أم إذا كان مغلقاً (الوسط الحسابي وأي مقياس آخر سواءً الوسيط أم المنوال) ، واستناداً لذلك تتم المقارنة (انظر فيما سبق صفات وخصائص التوزيعات التكرارية).

مثال (4-9) : في دراسة إحصائية عن الأجور الشهرية (بمئات الوحدات النقدية) لعمال إحدى المطابع في مدينة دمشق ، أخذت عينة عشوائية من (140) عامل من عمالها ، فأعطت بياناتها الجدول التكراري الآتي :

الأجر الشهري	80 -110	110 -140	140 -170	170 -200	200 -230	$\sum$
عدد العمال	20	30	40	30	20	140

والمطلوب : ما هي طبيعة التوزيع الذي تخضع له بيانات الجدول التكراري المعطى أعلاه ؟

الحل :

لمعرفة طبيعة توزيع بيانات الجدول التكراري المعطى ، نوجد مقياسين على الأقل من مقاييس النزعة المركزية بالاعتماد على بيانات هذا الجدول المساعد :

الأجر الشهري	عدد العمال f_i	وسط الفئة x_i	$f_i * x_i$	$f_i \uparrow$	$f_{i\downarrow}$
80 – 110	20	95	1900	20	140
110 – 140	30	125	3750	50	120
140 – 170	40	155	6200	90	90
170 – 200	30	185	5550	12	50
200 - 230	20	215	4300	140	20
$\sum$	140		21700		

- الوسط الحسابي : مئة وحدة نقدية $\overline{X} = \frac{\sum_{i=1}^{n} f_i * x_i}{\sum_{i=1}^{n} f_i} = \frac{21700}{140} = 155$ ؛

- الوسيط :

مئة وحدة نقدية $W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w = 140 + \frac{\frac{140}{2} - 50}{40} * 30 = 155$ ؛

- المنوال :

$$M_0 = L_{M_0} + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c_{M_0} = 140 + \frac{(40-30)}{(40-30)+(40-30)} * 30 = 155$$

مئة وحدة نقدية .

بمقارنة المقاييس الثلاث نجد تطابقاً فيما بينهما ، وبالتالي طبيعة توزيع بيانات الجدول التكراري المعطى في المثال (4- 9) متماثلة ، فهو لا يعاني من أي التواء .

لو أضفنا إلى بيانات الجدول الواردة في المثال (4-9) الفئتين (260 - 230) و (390 – 260) وكان عدد العمال المقابل لهما على التوالي : (18) و (2) ، فما طبيعة التوزيع للبيانات الجديدة ؟ وعليه تصبح بيانات الجدول المساعد كالآتي:

الأجر الشهري	عدد العمال f_i	وسط الفئة x_i	$f_i * x_i$	$f_i \uparrow$	$f_{i\downarrow}$
80 – 110	20	95	1900	20	160
110 – 140	30	125	3750	50	140
140 – 170	40	155	6200	90	110
170 – 200	30	185	5550	12	70
200 - 230	20	215	4300	140	40
230 – 260	18	245	4410	158	20
260 - 290	2	275	550	160	2
$\sum$	160		26660		

- الوسط الحسابي : مئة وحدة نقدية $\overline{X} = \frac{\sum_{i=1}^{n} f_i * x_i}{\sum_{i=1}^{n} f_i} = \frac{26660}{160} = 166.625$ ؛

- الوسيط :

$$W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w = 140 + \frac{\frac{160}{2} - 50}{40} * 30 = 162.5$$ مئة وحدة نقدية

- المنوال :

$$M_0 = L_{M_0} + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c_{M_0} = 140 + \frac{(40-30)}{(40-30)+(40-30)} * 30 = 155$$

مئة وحدة نقدية .

بالمقارنة نلاحظ أن : $\overline{X} = 166.625 \succ W = 162.5 \succ M_0 = 155$ ، أي توزيع بيانات الجدول الجديد غير متماثل ، فهو يعاني من التواء موجب ونحو اليمين .

4-4- مقاييس التشتت

سبق أن تطرقنا إلى مقاييس النزعة المركزية ، غير أن هذه المقاييس لا تعطي فكرة كاملة عن حقيقة المجموعة التي يمثلها التوزيع ، فعلى الرغم من تجمع المفردات حول قيمة معينة ، نجد أنها بالوقت ذاته تقترب وتبتعد عن هذه القيمة بمقادير مختلفة وهذا الاختلاف يدعى " بالتشتت " ، الذي يعرّف بأنه : " تناثر أو انتشار قيم الظاهرة المدروسة حول أحد قيم المتوسطات (أحد مقاييس النزعة

المركزية) " ، والمقاييس التي تقيس ظاهرة التشتت يطلق عليها " مقاييس التشتت " ، التي هي دوماً مقاييس متممة ومكملة لمقاييس النزعة المركزية ، لأنها تُبرز بعض الميزات التي لا يمكن لمقاييس النزعة المركزية أن تبرزها.

مثال (4-10) : ليكن لدينا مجموعتان من مرتبات خمس من المكتبيين العاملين (بالدولار) في إحدى المكتبات ، كالآتي :

المجموعة (1)	800	800	800	800	800
المجموعة (2)	400	500	600	1100	1400

ولو حسبنا الوسط الحسابي لمرتبات هؤلاء المكتبيين في كل مجموعة لوجدنا بأنه متساوٍ للآخر ، وهو : $\overline{X}_1 = \overline{X}_2 = 800$ ، إلاَّ أن تباعد مفردات كل مجموعة عن وسطها الحسابي مختلف ، فهو في المجموعة (1) مساوياً للصفر، في حين أنه متفاوت وأكبر من الصفر في المجموعة (2) ، وهذا يدل على أن الوسط الحسابي للمجموعة (1) يعتبر مقياساً صادقاً لمفردات هذه المجموعة ، في حين أنه في المجموعة (2) أقل تمثيلاً ، وذلك بسبب عدم تجانس مفردات هذه المجموعة وانتشارها، أي أن مرتبات المكتبيين في المجموعة (1) أكثر تجانساً ومصداقية من مرتبات المجموعة (2) ، وعليه فالتشتت هو : عبارة عن درجة تباعد مفردات المجموعة كل عن الآخر ، ودراسة التشتت تنحصر لأي مجموعة في معرفة التجانس بين مفرداتها .

وبشكل عام ، فبقدر ما يكون تشتت مجموعة من القيم أكبر، بقدر ما يكون متوسطها أقل تمثيلاً أو متجانساً والعكس بالعكس ، أي : " كلما قل التشتت زاد التجانس بين المفردات " ، لذا ومن أبرز استعمالات مقاييس التشتت ، الآتي :

1 - تحديد المقدرة التمثيلية لأحد المتوسطات ؛

2 - دورها في التحليل الإحصائي : في دراسة الاستدلال الإحصائي ، وفي الارتباط والانحدار، وفي السلاسل الزمنية ، وفي الدراسات الإحصائية التحليلية للأسعار والدورات الاقتصادية وتوزيع الدخل ، وفي معرفة التغيرات في أعمار السكان ،..... الخ .

هذا وتصنف مقاييس التشتت في صنفين ،هما :

1. مجموعة المقاييس التي توضح مدى وكيفية اختلاف القيم فيما بينها ، أي أنها قائمة على أساس المفهوم القائل : " بأن تباعد المفردات أو انتشارها يمكن وصفه بحجم المدى أو الاختلاف بين قيمتين في مجموعة القيم سواءً أكانت أكثر تطرفاً كأسلوب " المدى " أو قيماً أخرى منتقاة لأسباب ما ، كأسلوب" نصف المدى الربيعي " أو ما يطلق عليه " بالانحراف الربيعي " ؛

2. مجموعة المقاييس التي تقيس انحرافات أو اختلافات المفردات حول أحد المتوسطات ، كالانحراف المتوسط أو كالانحراف المعياري ، فهي توضح مدى تناثر القيم حول أحد المتوسطات.

وكلا الصنفين يحسبان بشكل مطلق وبشكل نسبي ، وهذا ما يقودنا إلى التعرّف على كل من مقاييس التشتت المطلقة والنسبية ، وما هو الاختلاف بينهما :

- مقاييس ا لتشتت المطلقة : وهي تقيس مقدار التشتت بالوحدات التي تقيس قيم الظاهرة نفسها ، أي أن وحداتها هي وحدات القيم الأصلية نفسها. وتستخدم هذه المقاييس من أجل المقارنة بين عدة مجموعات من القيم ، فيما إذا كانت المجموعات ا لمراد المقارنة فيما بينها من وحدات إحصائية واحدة أو كان متوسط كل مجموعة من تلك المجموعات لا يختلف كثيراً عن غيره في المجموعات الأخرى ؛
- مقاييس التشتت النسبية : وهي تقيس التشتت على شكل نسب مئوية ، ولها مدلول أكبر ، خاصةً إذا كانت المقارنة المرغوب فيها بين بيانات من وحدات إحصائية مختلفة ، أو يختلف متوسط كل منها اختلافاً كبيراً بالرغم من استخدام وحدات إحصائية واحدة.

4-4-1- المدى

4-4-1-1- مفهومه وكيفية حسابه

يُعد المدى من أبسط مقاييس التشتت ويعرّف بأنه : " الفرق بين أكبر القيم وأصغرها في مجموعة من القيم " ، ويعبر عنه ، كالآتي :

$$R = X_{\max} - X_{\min}$$

فمثلاً لو كان لدينا مرتبات مجموعة من الإداريين في مكتبة ما (باليورو)، وعلى التوالي :

1200 ، 1250 ، 1300 ، 1400 ، 1450 ، 1600 ، فيكون المدى لهذه المرتبات مساوياً

لـ (400) يورو ، أي :

$$R = X_{\max} - X_{\min} = 600 - 1200 = 400$$

هذا وبالرغم من بساطته وسهولة حسابه ، فهو لا يستخدم للأسباب الآتية :

1 - يتأثر بالقيم الحدية فقط ، فهو لا يأخذ بالحسبان جميع قيم المجموعة ، كما أنه لا يكشف عن انتشار البيانات الواقعة بين هذين الحدين ؛

2 - في حالة مقارنة بين مجموعتين مختلفتين من حيث الحجم ، يكون المدى مقياساً مضاداً للتشتت ؛

3 - أما إذا كانت البيانات مبوبة فيحسب المدى بأخذ الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى من الجدول التكراري المراد دراسته ، أو بأخذ الفرق بين مركز الفئة الأخيرة ومركز الفئة الأولى من الجدول التكراري ، شرط أن لا تكون إحدى الفئتين أو كلاهما مفتوحة ، لأنه وفي مثل هذه الحالة لا يمكن حسابه .

أما إذا أردنا المقارنة بين مجموعتين من البيانات المختلفة في وحدات قياسها أو في مقاييسها الإحصائية ، وجب علينا أن نحسب مقياسها النسبي ، وعليه نحسب المدى النسبي ، كالآتي :

- حال كان الوسط الحسابي معروفاً : $R\% = \frac{X_{\max} - X_{\min}}{\overline{X}} * 100$

- في حال كان الوسط الحسابي غير معروف : $R\% = \frac{X_{\max} - X_{\min}}{X_{\max} + X_{\min}} * 100$

4-4-1-2- استعمالاته

يستعمل المدى في الحالات التالية :

- في مجال الرقابة على الجودة – السيطرة النوعية للإنتاج ، لضمان كون الوحدات المنتجة بين حدين أعلى وأدنى ؛
- في قياس تشتت الأسعار اليومية للسندات المالية المستحقة ، أسعار الصرف عند الافتتاح في بداية اليوم والأقفال في نهايته ؛
- في قياس التغير في درجات الحرارة ؛
- في قياس التداول اليومي لمقتنيات المكتبة بين المستفيدين منها ؛
- في الدراسات المخبرية لقياس سرعة تثقّل الدم في الساعة الأولى والثانية .

4-4-2- الانحراف المتوسط

4-4-2-1- تعريفه وكيفية حسابه

يعرّف الانحراف المتوسط الذي نرمز له بـ (D_m) ، بأنه : " الوسط الحسابي لمجموع الانحرافات عن أحد مقاييس النزعة المركزية ، إما الوسيط أو الوسط الحسابي ، وذلك بالقيمة المطلقة " ، وبتعبير آخر : " فهو متوسط انحرافات قيم المفردات بشكل مطلق عن أحد مقاييس النزعة المركزية " .

وبناءً على هذا التعريف يمكننا حساب الانحراف المتوسط لبيانات غير مبوبة ، كالآتي :

-الانحراف عن الوسط الحسابي : $D_m = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n}$

-الانحراف عن الوسيط : $D_m = \frac{\sum_{i=1}^{n} |X_i - W|}{n}$

مثال (4-11) : إذا كان التداول خلال أيام الأسبوع لمقتنيات إحدى المكتبات كما يلي :

1100 ، 1200 ، 1300 ، 1500 ، 1600 ، 1700 ، فالمطلوب : إيجاد قيمة الانحراف المتوسط لهذه المقتنيات مرة بدلالة الوسط الحسابي ، ومرة أخرى بدلالة الوسيط ؟

الحل :

- نوجد أولاً قيمة كل من : الوسط الحسابي

$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{8400}{6} = 1400$ ، والوسيط $W = 1500$ ؛

- ثّم نوجد انحرافات قيم هذه المقتنيات المتداولة مرة عن وسطها الحسابي ومرة أخرى عن وسيطها وذلك بالقيم المطلقة ، أي كالآتي :

X_i	1100	1200	1300	1500	1600	1700	$\sum$
$\lvert X_i - \overline{X} \rvert$	300	200	100	100	200	300	1200
$\lvert X_i - W \rvert$	400	300	200	0	100	200	1200

- نطبق علاقات إيجاد الانحراف المتوسط وفق كلتا الطريقتين السابقتين ،

فيكون : $D_m = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n} = \frac{1200}{6} = 200$

و $D_m = \frac{\sum_{i=1}^{n} |X_i - W|}{n} = \frac{1200}{6} = 200$

غير أنه يحسب لبيانات مبوبة وذلك بأخذ الانحرافات المطلقة لأوساط فئات الجدول التكراري عن وسطها الحسابي أو وسيطها ، بعدها يتم جداؤها بالتكرارات المقابلة لها ، ومن ثّم نقوم بتقسيم حاصل الجداء على مجموع التكرارات ، كالآتي :

-الانحراف عن الوسط الحسابي : $$D_m = \frac{\sum_{i=1}^{n} f_i \left| x_i - \overline{X} \right|}{\sum_{i=1}^{n} f_i}$$

-الانحراف عن الوسيط : $$D_m = \frac{\sum_{i=1}^{n} f_i \left| x_i - W \right|}{\sum_{i=1}^{n} f_i}$$

وتجدر الإشارة هنا ، إلى أنه عندما يكون التوزيع التكراري بعيداً عن التماثل يفضل حساب الانحرافات عن الوسيط ، وذلك لأن مجموع الانحرافات المطلقة عن الوسيط هو دوماً أقل قيمة معنية ، وإن الانحراف المتوسط عن الوسط الحسابي هو دوماً مقياس أكبر للتشتت منه عن الوسيط ، إلاَّ في حال توافق قيمة الوسط الحسابي مع الوسيط (حالة التوزيعات المتماثلة) .

أما إذا أردنا المقارنة بين مجموعتين من البيانات المختلفة في وحدات قياسها أو في مقاييسها الإحصائية ، وجب علينا أن نحسب مقياسها النسبي ، فالمجموعة ذات الانحراف المتوسط النسبي الأقل هي المجموعة ذات التجانس الأكبر في مقياس نزعتها المركزية ، وعليه نحسب الانحراف المتوسط النسبي ، كالآتي :

-بدلالة الوسط الحسابي : $$D_m\% = \frac{D_m}{\overline{X}} * 100$$

-بدلالة الوسيط : $$D_m\% = \frac{D_m}{W} * 100$$

4-4-2-2- مزايا الانحراف المتوسط ومساوئه

هذا وسنلخص كل من مزايا الانحراف الربيعي ومساوئه ، كالآتي :

مزايا الانحراف المتوسط	مساوئ الانحراف المتوسط
• سهولة حسابه وفهمه ؛ • يأخذ بالحسبان جميع المفردات، ويعتبر مقياساً مرضياً للتشتت مقارنة مع المدى والانحراف الربيعي ، وأسهل حساباً من الانحراف المعياري .	• عدم ملاءمته للحسابات الجبرية بسبب وجود القيم المطلقة (إهمال الإشارات الجبرية) ؛ • تقتصر قدرته فقط على قياس التشتت ؛ • يُعد أقل تجاوباً مع العمليات الجبرية مقارنة مع الانحراف المعياري ؛ • عدم إمكانية حسابه في جداول تكرارية مفتوحة.

4-4-3- الانحراف الربيعي

4-4-3-1- مفهومه وكيفية إيجاده

الانحراف الربيعي هو مقياس موضعي للتشتت يقوم على الترتيب ويساوي نصف الفرق الربيعين الثالث والأول، ويطلق عليه أيضاً بنصف المدى الربيعي . وبناءً على هذا التعريف يمكننا أن نكتب :

$$Q = \frac{Q_3 - Q_1}{2}$$

حيث إن :

Q الانحراف الربيعي أو نصف المدى الربيعي ؛

Q_1 الربيع الأول ؛

Q_3 الربيع الثالث .

لحساب الانحراف الربيعي لبيانات غير مبوبة ، نتبع الآتي :

1 - نرتب البيانات بشكل تصاعدي أو تنازلي ثّم نحدد ترتيب موقع كل من الربيع الأول والثالث لإيجاد قيمتهما ، وذلك كما يلي :إذا كانت البيانات مرتبة بشكل تصاعدي $x_1 \le x_2 \le x_3 \le ... \le x_n$ ، فإننا نعمد لإيجاد ترتيب موقع كل من :

الربيع الأول $T_{Q_1} = \frac{n+1}{4}$ والربيع الثالث $T_{Q3} = \frac{3n+1}{4}$ ، وبناءً على ترتيبهما نحدد قيمة كل منهما ، ومن ثّم نحسب قيمة الانحراف الربيعي وفق العلاقة المحددة أعلاه .

مثال (4-12) : إذا كان التداول بالمواد المكتبية خلال أيام أسبوع ، كما يلي : 1050 ، 1100 ، 1120 ، 1260 ، 1280 ، 1310 ، 1340 ، فما هي قيمة الانحراف الربيعي لهذه المجموعة من القيم ؟

نلاحظ بأن القيم قد رتبت بشكل تصاعدي ، فما علينا إلاَّ أن نوجد قيمتي كل من الربيعين الأول والثالث من خلال ترتيب موقعهما ، أي :

$T_{Q_1} = \frac{n+1}{4} = \frac{7+1}{2} = 4$ ، أي أن المفردة التي ترتيبها الرابع هي قيمة الربيع الأول ، وهي في مثالنا 1260 ، وبما أن $T_{Q3} = \frac{3(n+1)}{4} = \frac{3*(7+1)}{2} = \frac{24}{4} = 6$ فالمفردة التي ترتيبها (6) هي قيمة الربيع الثالث ، وهي في بيانات مثالنا 1310 ، وعليه تكون قيمة الانحراف الربيعي لبيانات مثالنا :

$$Q = \frac{Q_3 - Q_1}{2} = \frac{1310 - 1260}{2} = 25$$

أما لو وجد ترتيب الربيع الأول أو الثالث بين قيمتين تؤخذ النسبة المئوية للمدى العددي الواقع فيما بينهما الربيع الأول أو الثالث وتضاف إلى المفردة التي ترتيبها (...) فتكون عندئذٍ هي قيمة الربيع الأول أو الثالث.

2 - أما حساب الانحراف الربيعي لبيانات مبوبة ، نطبق الآتي :

$$Q_j = L_{Q_j} + \frac{\frac{(j)*\sum_{i=1}^{n} f_i}{4} - \sum_{i=1}^{Q_j} f_i}{f_{Q_j}} * C_{Q_j}$$

حيث تمثل :

$j = 1,2,3,4$ ترتيب الربيع ؛

Q_j قيمة الربيع من الترتيب j ؛

f_{Q_j} تكرار الفئة الربيعية j ؛

$\sum_{i=1}^{Q_j} f_i$ مجموع التكرارات التي تسبق الفئة الربيعية من الترتيب j ؛

C_{Q_j} طول الفئة الربيعية ذي الترتيب j .

واستناداً لما سبق ، نتبع الخطوات الآتية :

-إيجاد التكرارات التجميعية الصاعدة ؛

-نحدد الفئة الربيعية المراد إيجادها ، فمثلاً الفئة الربيعية الأولى وتُعرّف بأنها : الفئة التي تحوي الربيع الأول والتي يكون تكرارها التجميعي الصاعد مساوياً أو أكبر من $\frac{\sum_{i=1}^{n} f_i}{4}$ ، أما الفئة الربيعية الثالثة فهي الفئة التي تحوي الربيع الثالث والتي يكون تكرارها التجميعي الصاعد مساوياً أو أكبر من $\frac{3*\sum_{i=1}^{n} f_i}{4}$ ؛

-واستناداً إلى الفئات المشار إليها يمكننا عندها تطبيق العلاقات الرياضية سواءً لحساب قيمة كل من الربيع الأول والثالث أم قيمة الانحراف الربيعي .

أما إذا أردنا المقارنة بين مجموعتين من البيانات المختلفة في وحدات قياسها أو في مقاييسها الإحصائية ، وجب علينا أن نحسب مقياسها النسبي ، فالمجموعة ذات الانحراف الربيعي الأقل هي المجموعة ذات التجانس الأكبر في مقياس نزعتها المركزية ، وعليه نحسب الانحراف الربيعي النسبي ، كالآتي :

$$Q\% = \frac{Q_3 - Q_1}{2 * W} * 100$$

وفي الحالة التي لا يتوافر لدينا فيها قيمة للوسيط ، يمكننا إيجاد الانحراف الربيعي النسبي من خلال العلاقة الآتية:

$$Q\% = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$$

ونشير هنا ، إلى أن الوسيط كمقياس للنزعة المركزية يقابله كمقياس للتشتت دوماً الانحراف الربيعي .

مثال (4-13) : قامت وزارة الثقافة في بلد ما ، بأخذ عينة عشوائية من (160) مكتبة متنقلة ، بغية معرفة المسافات (بالكم) التي تقطعها المكتبات المتنقلة أثناء تقديم خدماتها، فأعطت بياناتها الجدول التكراري الآتي :

المسافة المقطوعة	100- 80	110- 100	130- 110	160- 130	180- 160	190- 180	210- 190	$\sum$
عدد المكتبات المتنقلة	9	24	28	32	30	25	12	160

ولقد أخذت عينة عشوائية من بلد أخر ومن الحجم نفسه، فوجد أن وسيطها (145) كم ، ونصف مداها النسبي (25) كم ، فأي العينتين تعرض تشتتاً أكبر حول وسيطها ، عينة الجدول التكراري أم عينة البلد الآخر؟

الحل :

- ننشئ الجدول المساعد في إيجاد التكرارات التجميعية الصاعدة :

المسافة المقطوعة	100- 80	110- 100	130- 110	160- 130	180- 160	190- 180	210 - 190	$\sum$
عدد المكتبات المتنقلة	9	24	28	32	30	25	12	160
$f_{i\uparrow}$	9	33	61	93	123	148	160	

- نوجد قيم كلٍ من :

- الوسيط :

$$W = L_w + \frac{\frac{\sum_{i=1}^{n} f_i}{2} - \sum_{i=1}^{w-1} f_i}{f_w} * C_w = 130 + \frac{\frac{160}{2} - 61}{32} * 30 = 147.8$$

- الربيع الأول :

$$Q_j = L_{Q_j} + \frac{\frac{(j) * \sum_{i=1}^{n} f_i}{4} - \sum_{i=1}^{Q_j} f_i}{f_{Q_j}} * C_{Q_j} = 110 + \frac{\frac{1*160}{4} - 33}{28} * 20 = 115$$

- الربيع الثالث :

$$Q_j = L_{Q_j} + \frac{\frac{(j) * \sum_{i=1}^{n} f_i}{4} - \sum_{i=1}^{Q_j} f_i}{f_{Q_j}} * C_{Q_j} = 160 + \frac{\frac{3*160}{4} - 93}{30} * 20 = 178$$

- الانحراف الربيعي : $Q = \frac{Q_3 - Q_1}{2} = \frac{178 - 115}{2} = 31.5$
- الانحراف الربيعي النسبي لبيانات الجدول التكراري :

$$Q\% = \frac{Q_3 - Q_1}{2 * W} * 100 = \frac{178 - 115}{2 * 147.8} * 100 = 21.31\%$$

- الانحراف الربيعي النسبي لبيانات عينة البلد الآخر :

$$Q\% = \frac{Q_3 - Q_1}{2 * W} * 100 = \frac{25}{145} * 100 = 17.24\%$$

- المقارنة والقرار : من مقارنة الانحراف الربيعي النسبي لبيانات الجدول التكراري معه لبيانات عينة البلد الآخر ، نلاحظ بأن الانحراف الربيعي النسبي لعينة البلد الآخر هو أقل منه لدى بيانات الجدول التكراري ، وهذا يدل على عينة البلد الآخر تعرض تشتتاً أقل وتجانساً أكبر بوسيطها من بيانات الجدول التكراري .

4-4-3-2- المزايا والمساوئ

هذا وسنلخص كل من مزايا الانحراف الربيعي ومساوئه ، كالآتي :

مزايا الانحراف الربيعي	مساوئ الانحراف الربيعي
• سهل الحساب والوضوح في معناه ؛ • إمكانية حسابه عند وجود إحدى الفئات مفتوحة في الجدول التكراري ؛ • أقل تأثراً بالقيم الشاذ ة والمتطرفة على جانبي التوزيع وذالك مقارنة مع مقياس المدى.	• عدم قابليته للعمليات الجبرية ؛ • عدم أخذه لجميع المفردات بالحسبان، فهو يأخذ 50% منها ويهمل 50% الأكثر، فـقيمته ليست دقيق ة كمقياس للتشتت نتيجة إهماله لنصف المفردات وعدم انتشاره .

4-4-4- الانحراف المعياري

4-4-4-1- ماهية الانحراف المعياري وكيفية حسابه

يعتبر الانحراف المعياري من أكثر مقاييس التشتت أهمية ، لأنه مفهوم جبري محدد بدقة ، ويعتمد على خصائص رياضية ملائمة ، لذا يطبق في أغلب الدراسات.

ويُعرّف الانحراف المعياري لمجموعة من القيم بأنه : " الجذر التربيعي لمتوسط مجموع مربع انحرافات القيم عن وسطها الحسابي " ، فهو يدل على مدى انتشار أو ابتعاد كل قيمة من القيم التي تأخذها الظاهرة المدروسة عن وسطها الحسابي .

وثّمة أسلوبان لإيجاد قيمة الانحراف المعياري لبيانات غير مبوبة ، هما :

- الأسلوب المباشر :

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$$

أو

$$S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2}$$

- أسلوب الانحرافات :

$$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X}_0)^2}{n}-\left[\frac{\sum_{i=1}^{n}(X_i-\overline{X}_0)}{n}\right]^2}=\sqrt{\frac{\sum_{i=1}^{n}d_i^2}{n}-\left(\frac{\sum_{i=1}^{n}d_i}{n}\right)^2}$$

حيث إن : $\overline{X}_0$ وسط حسابي فرضي ، و $d_i=(x_i-\overline{X}_0)$ ، وكل من هذه الأساليب المطبقة تؤدي إلى النتيجة نفسها التي يمكن الحصول عليها من إحداها .

مثال (4-14) : ليكن لدينا عدد الساعات التدريبية التي يقضيها مجموعة من طلاب قسم المكتبات في المكتبة أسبوعياً ، كالآتي : 15 ، 9 ، 14 ، 6 ، 23 ، 19 ، 33 ، فما هو الانحراف المعياري لهذه البيانات ؟

الحل :

نحسب أولاً الوسط الحسابي لهذه البيانات ، فيكون : $\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}=\frac{119}{7}=17$ ، ثُم نحسب انحرافات هذه القيم عن وسطها الحسابي ، كالآتي :

i	1	2	3	4	5	6	7	$\sum$
X_i	15	9	14	6	23	19	33	119
X_i^2	225	81	196	36	529	361	1089	2517
$X_i-\overline{X}$	-2	-8	-3	-11	6	2	16	
$(X_i-\overline{X})^2$	4	64	9	121	36	4	256	494
$X_i-\overline{X}_0=d_i$	1	-6	-1	-9	8	4	18	15
d_i^2	1	36	1	81	64	16	324	523

بالتطبيق على ما سبق ، نجد :

-الأسلوب المباشر : $S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}}=\sqrt{\frac{494}{7}}=8.4$

- أو $S = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2} = \sqrt{\frac{2517}{7} - \left(\frac{119}{7}\right)^2} = 8.4$

-أسلوب الانحرافات ، حيث أن قيمة $\overline{X}_0 = 15$:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X}_0)^2}{n} - \left[\frac{\sum_{i=1}^{n}(X_i - \overline{X}_0)}{n}\right]^2} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n} - \left(\frac{\sum_{i=1}^{n} d_i}{n}\right)^2} = \sqrt{\frac{523}{7} - \left(\frac{15}{7}\right)^2} = 8.4$$

أما أساليب إيجاده لبيانات مبوبة ، فهي الآتي :

- الأسلوب المباشر : $S = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \overline{X})^2}{\sum_{i=1}^{n} f_i}}$

حيث إن: x_i وسط الفئة ، و f_i تكرار الفئة .

- أسلوب الانحرافات :

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i (X_i - \overline{X}_0)^2}{\sum_{i=1}^{n} f_i} - \left[\frac{\sum_{i=1}^{n} f_i (X_i - \overline{X}_0}{\sum_{i=1}^{n} f_i}\right]^2} = \sqrt{\frac{\sum_{i=1}^{n} d_i}{n} - \left(\frac{\sum_{i=1}^{n} d_i}{n}\right)^2}$$

حيث إن : $\overline{X}_0$ وسط حسابي فرضي ، و $d_i = (x_i - \overline{X}_0)$.

-أسلوب الانحرافات المختصرة :

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i \left(\frac{x_i - \overline{X}_0}{C_i}\right)^2}{\sum_{i=1}^{n} f_i} - \left[\frac{\sum_{i=1}^{n} f_i \left(\frac{x_i - \overline{X}_0}{C_i}\right)}{\sum_{i=1}^{n} f_i}\right]^2} = \sqrt{\frac{\sum_{i=1}^{n} f_i (d_i`)^2}{\sum_{i=1}^{n} f_i} - \left(\frac{\sum_{i=1}^{n} f_i (d_i`)}{\sum_{i=1}^{n} f_i}\right)^2}$$

حيث إن : $d_i^{`} = \left(\frac{x_i - \overline{X}_0}{C_i} \right)$

- أسلوب انحرافات X_i عن نفسها : $S = \sqrt{\frac{\sum_{i=1}^{n} f_i X_i^{\,2}}{\sum_{i=1}^{n} f_i} - \left[\frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i} \right]^2}$

وبما أن الانحراف المعياري يُعد مقياساً معيارياً للتشتت ، فهو يستخدم للمقارنة بين مجموعتين أو أكثر من البيانات المختلفة في وحدات قياسها أو في مقاييسها الإحصائية، فيما إذا كانت قابلة للمقارنة ، وعليه نحسب المقياس النسبي لكل مجموعة ، ثِّم نقارن فيما بين مقاييسها النسبية ، فالمجموعة ذات الانحراف المعياري النسبي الأقل هي المجموعة ذات التشتت الأقل والتجانس الأكبر في مقياس نزعتها المركزية ، ألاَّ وهو الوسط الحسابي ، وعليه نحسب الانحراف المعياري النسبي أو ما يطلق عليه بمعامل الاختلاف ، الذي يُعطى :

$$V = \frac{S_x}{\overline{X}} * 100$$

مثال (4-15) : في دراسة إحصائية قامت بها وزارة الثقافة ، بغية معرفة عدد سنوات الخبرة لموظفي الوزارة العاملين في مجال المكتبات والمراكز الثقافية ، فقد أخذت عينة عشوائية مؤلفة من (100) موظف من موظفي الوزارة في مجال المكتبات والمراكز الثقافية التابعة لها ، فأعطت بياناتها الجدول التكراري الآتي :

فئات سنوات الخدمة	19- 17	21 - 19	23 - 21	25 - 23	27 - 25	30 - 27	$\sum$
عدد الموظفين	2	12	18	36	18	14	100

كما أخذت الوزارة عينة عشوائية من حجم مماثل من موظفي الوزارة في الإدارة العامة للوزارة ، فكان متوسط عدد سنوات الخدمة فيها للموظف مساوياً لـ 25 سنة ، وبمعامل اختلاف قدره (12%) ، فالمطلوب :

1- ما هو أفضل مقياس من مقاييس النزعة المركزية وما يقابلها من مقاييس التشتت لبيانات الجدول التكراري المعروض أعلاه ، وما هو السبب في اختيارك لهذا المقياس ؟

2- أي العينتين تبدي تشتتاً أكبر في مقياس نزعتها المركزية ؟

الحل :

1- إن أفضل مقياس من مقاييس النزعة المركزية لبيانات الجدول التكراري في المثال (4-15) هو الوسط الحسابي ، ويقابله من مقاييس التشتت الانحراف المعياري ، والسبب وراء هذا الاختيار هو أنهما يتمتعان بميزات جبرية لا تتمتع بها المقاييس الأخرى ، وذلك لأن فئات الجدول التكراري المذكور جميعها مغلقة .

2- ننشئ الجدول المساعد في إيجاد كلٍ من الوسط الحسابي والانحراف المعياري لبيانات الجدول التكراري المعطى ، كالآتي :

فئات سنوات الخدمة	عدد الموظفين f_i	وسط الفئة x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
17 – 19	2	18	36	324	648
19 – 21	12	20	240	400	4800
21 – 23	18	22	396	484	8712
23 – 25	36	24	864	578	20808
25 – 27	18	26	468	676	12168
27 - 30	14	28.5	399	812.25	11371.5
$\sum$	100		2403		58507.5

- فتكون قيمة الوسط الحسابي لهذه البيانات :

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{2403}{100} = 24.03 \text{ سنة خدمة} \quad ؛$$

- وقيمة الانحراف المعياري :

$$S = \sqrt{\frac{\sum_{i=1}^{n} f_i X_i{}^2}{\sum_{i=1}^{n} f_i} - \left[\frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}\right]^2} = \sqrt{\frac{117607.5}{100} - \left(\frac{2403}{100}\right)^2} = 2.76 \text{ سنة خدمة} \quad ؛$$

- أما قيمة معامل الاختلاف : $V = \frac{S_x}{\overline{X}} * 100 = \frac{2.76}{24.03} * 100 = 11.49\%$ ؛

- المقارنة والقرار : بمقارنة معامل الاختلاف الذي حسبناه مع معامل الاختلاف المعطى للعينة الأخرى ، نلاحظ بأن معامل الاختلاف الذي حسبناه هو أقل من

الآخر المعطى ، وبالتالي تكون بيانات الجدول التكراري أقل تشتتاً وأكبر تجانساً بوسطها الحسابي من بيانات العينة الأخرى ، وعليه نثق ببيانات الجدول التكراري.

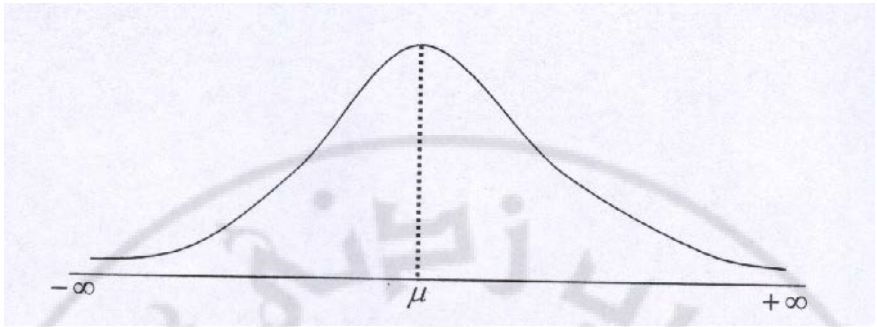
4-4-4-2- خواص الانحراف المعياري واستعمالاته

يعتبر التباين الذي هو مربع الانحراف المعياري مهماً للغاية في لتحليل الإحصائي، وهو يستعمل في أغلب التطبيقات وبصورة خاصة في مجال الاستدلال الإحصائي، وقد اكتسبت أهميته بشكل خاص من تمتعه بالميزات الجبرية ، فلو كانت قيمته صغيرة دلّ ذلك على تمركز المعطيات حول الوسط الحسابي ، وعلى العكس تماماً عندما تكون قيمته كبيرة فيدلّ على عدم التجانس وعلى التشتت الأكبر عن قيمة الوسط الحسابي لقيم الظاهرة المدروسة . ولعل من أهم خواص الانحراف المعياري ، تلك المتعلقة بنظرية تشيبيشيف وتطبيقاتها على التوزيع الطبيعي المعياري .

يُعد التوزيع الطبيعي من أهم التوزيعات العشوائية المستمرة ، نظراً لاستخداماته الواسعة والمتعددة في مختلف المجالات التطبيقية ، إذ يتم تمثيل التوزيع الطبيعي بواسطة منحنى يأخذ شكل الجرس ومعرّف بمؤشرين هما : الوسط الحسابي $\overline{X}$ والانحراف المعياري S_x (للعينة) ، أ و الوسط الحسابي μ والانحراف المعياري σ_x للمجتمع ، وبالإضافة لذلك نذكر بعضاً من الخصائص الأخرى التي يتصف بها هذا التوزيع ، ومنها :

- أنه متماثل ومتناظر ، فالمساحات الواقعة على طرفي الوسط الحسابي متساوية ، أي لو أطبقنا الطرف الأيمن منه على الطرف أيسر لانطبقت على بعضها الآخر ؛
- النهايات الطرفية تمتد إلى ما لانهاية ، أي أنها تكاد تلامس المحور الأفقي ، فهي متقاربة وموازية له ؛
- يمثل منحنى التوزيع الطبيعي عدداً غير محدود أو عدداً كبيراً من الحالات .

إن منحنى كثافة الاحتمال متناظر حول الوسط الحسابي ويمتد من $[-\infty,+\infty]$ ، الذي يرمز له غالباً بـ $N(\mu,\sigma_x)$ أو $N(\overline{X},S_x)$ ، فالشكل رقم (14) يبين لنا ذلك ، كالآتي :



الشكل رقم (14) : منحنى كثافة الاحتمال للتوزيع الطبيعي

إن تابع كثافة الاحتمال يأخذ الشكل الآتي :

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}*e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma_x}\right)^2}$$

إن المساحة المحصورة بين المنحنى ومحور السينيات تساوي الواحد الصحيح ، أي :

$$\int_{-\infty}^{+\infty} f(x).d_x=\int_{-\infty}^{+\infty}\left[\ \frac{1}{\sigma_x\sqrt{2\pi}}*e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma_x}\right)^2}\ \right]*d_x=1$$

وعندما نعبر عن X_i بدلالة الوحدات المعيارية ، عندها نجري تحويلاً للمتحول العشوائي X_i على الشكل الآتي:

- بدلالة المجتمع $$Z=\left|\frac{X_i-\mu}{\sigma_x}\right|$$

- بدلالة العينة $$Z=\left|\frac{X_i-\overline{X}}{S_x}\right|$$

وبتعويض قيمة Z في كثافة احتمال التوزيع الطبيعي أعلاه ، نجد :

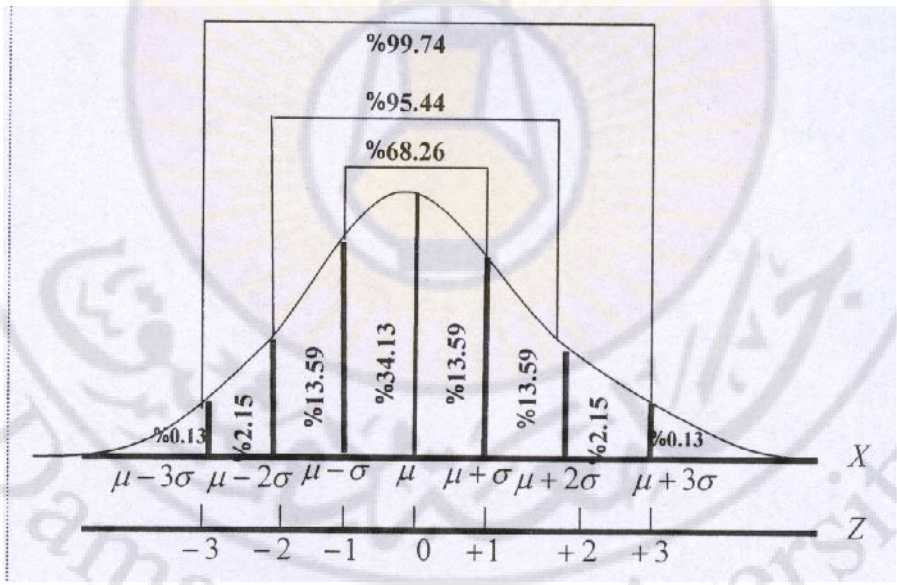
$$\int_{-\infty}^{-\infty}\frac{1}{\sigma_{x\sqrt{2\pi}}}*e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma_x}\right)^2}*d_x=\int_{-\infty}^{-\infty}\frac{1}{\sqrt{2\pi}}*e^{-\frac{Z^2}{2}}.d_z=1$$

وبذلك نحصل على متحول عشوائي جديد يدعى بالمتحول العشوائي المعياري ، وسطائه : الوسط الحسابي المساوي للصفر ، وتباينه يساوي الواحد الصحيح ، وكثافة الاحتمال تعطى ، كالآتي :

$$\int_{-\infty}^{+\infty} f(z).d_z = \int_{-\infty}^{-\infty} \frac{1}{\sqrt{2\pi}} * e^{-\frac{Z^2}{2}}.d_z = 1$$

حيث تدعى Z بالدرجة المعيارية ، وهي عبارة عن إمكانية تحويل أي قيمة خاضعة للتوزيع الطبيعي إلى قيمة طبيعية معيارية (خاضعة لتوزيع طبيعي معياري) ، والتي تستخدم في مقارنة المشاهدات من مجموعات مختلفة ، وبعبارة أخرى ، هي مدى درجة ابتعاد قيمة X_i عن وسطها الحسابي $\overline{X}$ بالانحرافات المعيارية .

واعتماداً على تابع كثافة الاحتمال لتوزيع الطبيعي المعياري ، يمكننا معرفة نسبة أو احتمال أو عدد القيم التي توجد في مجال محدد حول الوسط الحسابي ، في حال كان التوزيع التكراري خاضعاً للتوزيع الطبيعي وسطائه ($N(\mu,\sigma_x)$ أو $N(\overline{X},S_x)$) ، أي : ($\overline{X}=0,\sigma_x=1$) ، فالتمثيل البياني لهذا التوزيع موضح في الشكل رقم (15) ، كالآتي :



الشكل رقم (15) : شكل المنحنى الطبيعي مع تحديد المساحات

وبالنظر إلى الشكل رقم (15) وإلى جدول المساحات ما تحت المنحنى الطبيعي في الجداول الإحصائية الملحقة ، نذكر بعض المجالات العددية الشهيرة ، ومنها :

- $\Pr\{\overline{X}-1S_x \le X \le \overline{X}+1S_x\} = 68.27\%$

$$\Pr\{\overline{X} - 1.96S_x \le X \le \overline{X} + 1.96S_x\} = 95\% \quad -$$

$$\Pr\{\overline{X} - 2S_x \le X \le \overline{X} + 2S_x\} = 95.45\% \quad -$$

$$\Pr\{\overline{X} - 2.58S_x \le X \le \overline{X} + 2.58S_x\} = 99\% \quad -$$

$$\Pr\{\overline{X} - 3S_x \le X \le \overline{X} + 3S_x\} = 99.73\% \quad -$$

مثال (4-16) : أخذت عيرة عشوائية من 1000 مكتبي ، فتبين أن متوسط مرتبهم الأسبوعي 3000 وحدة نقدية ، وبانحراف معياري قدره 300 وحدة نقدية ، فإذا علمت بأن مرتبات المكتبيين في المجتمع الذي أخذت منه هذه العينة تتوزع توزعاً طبيعياً ، والمطلوب:

1- أوجد عدد المكتبيين الذين يتراوح أجرهم الشهري بين 2700 وحدة نقدية و 3600 وحدة نقدية ؟

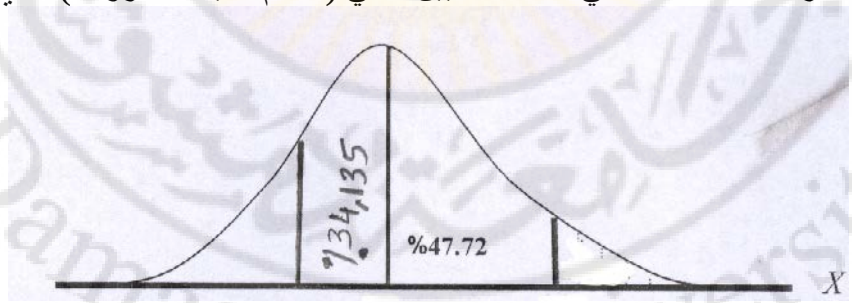
2- أوجد احتمال أن يكون مكتبياً مرتبه الأسبوعي يقل عن 2400 وحدة نقدية ؟

3- ما هي نسبة المكتبيين الذين يزيد مرتبهم الأسبوعي عن 2400 وحدة نقدية ؟

4- ما هو احتمال أن يكون المرتب الأسبوعي لمكتبي ما أكبر من 3300 وحدة نقدية .

الحل :

1 - لإيجاد عدد المكتبيين الذين يتراوح مرتبهم الأسبوعي ما بين 2700 و 3600 وحدة نقدية ، وجب علينا أن نوجد المساحة المحصورة تحت المنحنى الطبيعي وبين هاتين القيمتين ، بعد تحويلهما إلى قيمتين طبيعيتين معياريتين ، ثُم نقوم بضرب هذه المساحة في عدد المكتبيين الكلي (حجم العينة المدروسة) ، أي :



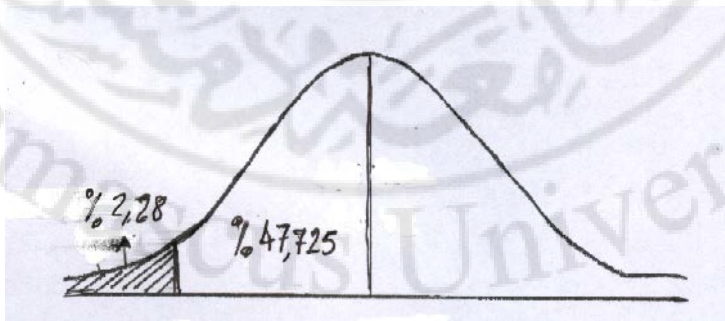
$-1\sigma_x \quad \overline{X} = 3000 \quad +2\sigma_x$

$$\Pr\left\{\frac{X_i-\overline{X}}{S_x}\le X\le\frac{X_i-\overline{X}}{S_x}\right\}=\Pr\left\{\frac{2700-3000}{300}\le X\le\frac{3600-3000}{300}\right\}=$$

$$\Pr\{-1\le X\le 2\}=\Pr(0\le X\le -1)+\Pr(0\le X\le 2)=\frac{0.6827}{2}+\frac{0.9545}{2}=0.34135+0.47725=0.8186$$

عدد المكتبيين = المساحة التي أوجدناها مضروبة في حجم العينة ، أي :
0.8186 × 1000 = 818.6 مكتبياً ، تقريباً 819 مكتبياً ؛
<u>وبطريقة أخرى</u> ، نوجد مدى ابتعاد 3300 عن 3000 بالانحرافات المعيارية من خلال تحويل القيمة 3300 إلى قيمة طبيعية معيارية ، أي :

$Z=\left|\frac{X_i-\overline{X}}{S_x}\right|=\left|\frac{3300-3000}{300}\right|=|-1|$ فهي تحصر ما مساحته تحت المنحنى الطبيعي ما مقداره $\frac{\%68.27}{2}=\%34.135=0.34135$ ، ثُم نوجد مدى ابتعاد القيمة الثانية وبنفس الطريقة ، أي : $Z=\left|\frac{X_i-\overline{X}}{S_x}\right|=\left|\frac{3600-3000}{2}\right|=|2|$ ، فهي تحصر ما مساحته تحت المنحنى الطبيعي ما قيمته $\frac{\%95.45}{2}=\%47.725=0.47725$ ، فالمساحة المطلوبة هي مجموع المساحتين ما بين $[-1\sigma_x,+2\sigma_x]$ ، أي :
0.8186 = 0.34135 + 0.47725 ، وعليه يكون :
عدد المكتبيين = 0.8186 * 1000 = 818.6 ≅ 819 مكتبياً .
3 - إن احتمال أن يكون مكتبياً مرتبه الأسبوعي يقل عن 2400 وحدة نقدية ، هو :

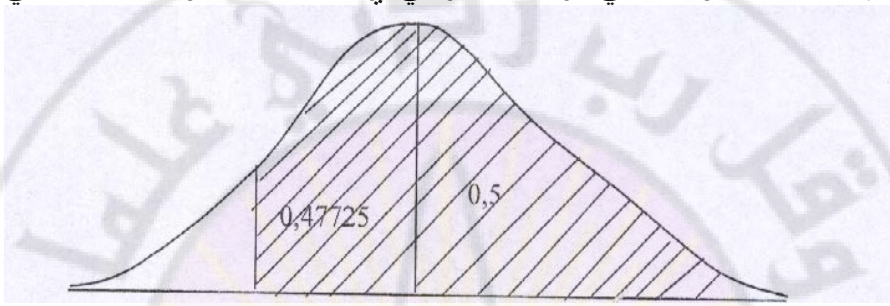


$-2\sigma_x$ $\overline{X}=3000$

$$\Pr\left\{X \le \frac{X_i - \overline{X}}{S_x}\right\} = \Pr\left\{X \le \frac{2400-3000}{300}\right\} = \Pr(-\infty \le X \le -2)$$

$$= \Pr(0 \le X \le -\infty) - \Pr(0 \le X \le -2) = 0.5 - \frac{0.9545}{2} = 0.5 - 0.47725 = 0.02275 = 2.275\%$$

4 - إن نسبة أن يكون مكتبي مرتبه الأسبوعي يزيد عن 2400 وحدة نقدية ، هي :

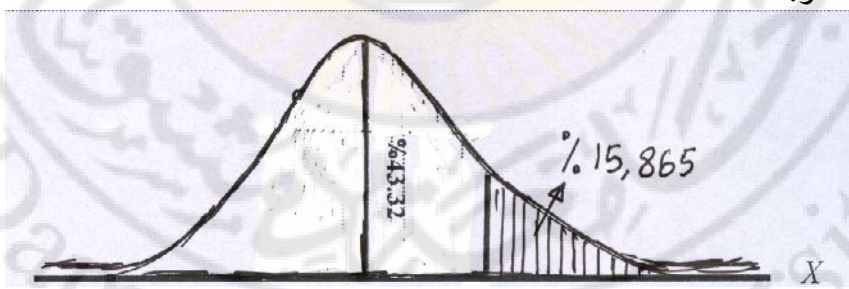


$-2\sigma_x$ $\overline{X} = 3000$

$$\Pr\left\{X \ge \frac{X_i - \overline{X}}{S_x}\right\} = \Pr\left\{X \ge \frac{2400-3000}{300}\right\} = \Pr(-2 \le X \le +\infty)$$

$$= \Pr(-2 \le X \le 0) + \Pr(0 \le X \le +\infty) = 0.5 + \frac{0.9545}{2} = 0.5 + 0.47725 = 0.977725 = 97.77\%$$

5 - إن احتمال أن يكون المرتب الأسبوعي لمكتبي ما أكبر من 3300 وحدة نقدية ، هو:



$\overline{X} = 3000 \ +1\sigma_x$ $+\infty$

$$\Pr\left\{X \ge \frac{X_i - \overline{X}}{S_x}\right\} = \Pr\left\{X \ge \frac{3300-3000}{300}\right\} = \Pr(0 \le X \le +\infty) - \Pr(1 \le X \le +\infty)$$

$$= 0.5 - \frac{0.6827}{2} = 0.5 - 0.34135 = 0.15865 = 15.865\%$$

4-4-4-3- مزايا الانحراف المعياري وعيوبه

يتمتع الانحراف المعياري بعدة مزايا ، من أهمها الآتي :

1 - يقوم على أساس انحراف المفردات الوسطي عن أحد مقاييس النزعة المركزية، وبخاصة الوسط الحسابي ؛

2 - يأخذ بالحسبان جميع المفردات ، فهو إذن أكثر شمولاً من أسلوبي المدى والانحراف الربيعي ؛

3 - تعتبر عملية التربيع للانحرافات بغية التخلص من الإشارات السالبة عملية رياضية صحيحة ، ليس كما في الانحراف المتوسط الذي يهمل الأخذ بها؛

4 - هو مفهوم مجرد وأكثر صعوبة من حيث الفهم من بقية المتوسطات ؛

5 - تسمح قيمته في معرفة المقدرة التمثيلية للوسط الحسابي ؛

6 - يمكننا بواسطته معرفة النسب أو المساحات أو الاحتمالات أو القيم ضمن مجالات محددة بدلالة التوزيع الطبيعي ؛

7 - يتمتع بميزات جبرية كثيرة كالوسط الحسابي .

وبالرغم من مزاياه ، إلاَّ أنه يتصف ببعض المساوئ ، منها :

1 - يتأثر بالقيم المتطرفة أكثر من الانحراف المتوسط ، وبأخذ الجذر التربيعي لمتوسط مجموع مربع انحرافات القيم عن وسطها الحسابي ، لا يمكن التخلص من تأثره بتلك القيم المتطرفة، لذا تكون قيمته عددياً أكبر من أي مقياس آخر للتشتت ماعدا المدى ؛

2 - ليس له أي مدلول إلاَّ عند مقارنته بانحرافات معيارية لمجموعات أخرى من القيم ؛

3 - قيمة الانحراف المعياري لعينة عشوائية من مجتمع إحصائي ما ، هي أقل من قيمة الانحراف المعياري لهذا المجتمع الإحصائي التي أخذت منه؛

4 - لا يمكن حسابه في توزيعات تكرارية أحد فئاته مفتوحة.

4-5- أسئلة وتمارين

1 - ما الفرق بين النسبة والمعدل؟

2 - ما هي العلاقة بين معدل النمو ومعدل التغير ؟

3 - ما هي استخدامات النسب والمعدلات ؟

4 - لتكن لدينا بيانات عما هو موجود في إحدى مكتبات أحد الأحياء لمدينة حمص ، كما هو مبين في الجدول الآتي :

نوع الكتب	قصصية	روايات	كتب تخصصية	كتب هندسية
عدد الكتب	1200	1500	80	50

والمطلوب :

أ - أوجد نسبة عدد الكتب القصصية إلى الكتب الروائية ؟

ب- ما هي نسبة عدد الكتب الهندسية والتخصصية إلى عدد الكتب القصصية ؟

ت- أوجد نسبة عدد الكتب الهندسية إلى عدد الكتب التخصصية ؟

5 - لقد كان المرتب الشهري لأحد مدراء المكتبات في عام ما مساوياً لـ (18000) وحدة نقدية ، وأصبح في العام التالي مساوياً لـ (22000) وحدة نقدية ، فما هو معدل التغير الذي طرأ على مرتب هذا المدير ؟

6 - ما هي نقطة التمركز ؟

7 - متى يجب تطبيق طريقة الانحرافات المختصرة في إيجاد الوسط الحسابي؟

8 - متى يكون للوسط الحسابي قيمة تمثيلية ؟

9 - متى يفقد الوسط الحسابي قيمته التمثيلية ؟

10 ما هي استعمالات الوسط الحسابي عندما يكون له قيمة تمثيلية ؟

11 متى تنجذب قيمة الوسيط نحو القيم المتطرفة ؟

12 هل يعتبر الوسيط ذا قيمة تمثيلية لعينة صغيرة الحجم ، ولماذا ؟

13 ما هي الفئة الوسيطية ، ولماذا نقوم بتحديد الفئة الوسيطية ؟

14 على ماذا تتوقف قيمة المنوال النهائية ؟

15 ما مبرر وجود أكثر من نقطة تمركز في جدول تكراري ؟

16 متى يمكن استخدام العلاقة بين المتوسطات " علاقة بيرسون " ؟

17 ما المقصود بطبيعة توزيع تكراري ما ، وكيف يمكن معرفة تلك الطبيعة ؟

18 عرّف كلا من : الالتواء ، التفرطح ، التشتت ، الربيع الأول ، الربيع الثالث ، الفئة المنوالية ، الفئة الربيعية الأولى ، الفئة الربيعية الثالثة ، الانحراف المتوسط ، الانحراف المعياري ، نصف المدى الربيعي ، الدرجة المعيارية ؟

19 لماذا تعتبر مقاييس التشتت مقاييس مكملة لمقاييس النزعة المركزية ؟

20 ما هي استعمالات مقاييس التشتت ؟

21 -ما الفرق بين مقاييس التشتت المطلقة والنسبية، ومتى يمكننا استخدام كل منها؟

22 -ما هو مقياس النزعة المركزية الذي يقابل نصف المدى الربيعي كمقياس للتشتت؟

23 -متى يفضل حساب الانحراف المتوسط بدلالة الوسيط ؟

24 -ليكن لدينا مقتنيات سبع مكتبات عامة في إحدى الدول ، وكالآتي :

المكتبة	A	B	C	D	E	F	G
عدد الكتب	6200	4300	3800	7400	5100	6500	5900

والمطلوب :

1 - ما هو العدد الوسطي من مقتنيات هذه المكتبة من الكتب ؟

2 - ما هي المكتبة ذات العدد الأكبر من الكتب ؟

3 - ما هي المكتبة التي يتساوى عن طرفيها عدد المكتبات ؟

25 -سُحبت عينتان عشوائيتان من عمال الطباعة في مدينة حلب ، بغية معرفة أجورهم الأسبوعية (مقدرة بمئات الليرات) ، فأعطت بياناتها الجدول التكراري الآتي :

الأجور الأسبوعية	22 - 18	28 - 22	36 - 28	42 - 36	46 - 42	المجموع
عدد عمال في المطبعة(آ)	14	28	36	28	14	120
عدد عمال في المطبعة (ب)	14	26	25	20	15	100

والمطلوب :

1 - أوجد كلا من الوسط الحسابي والوسيط والمنوال لكل مطبعة على حدة ؟

2 - ما قيمة متوسط الأجر الأسبوعي لعمال المطبعتين معاً ؟

3 - هل تختلف طبيعة توزيع أجور عمال المطبعة (آ) عنه لعمال المطبعة (ب) ؟

4 - أيَّ العينتين تعرض تشتتاً أكبر حول وسطها الحسابي،عينة عمال المطبعة (آ) أم عمال المطبعة
(ب) ، دعّم إجابتك إحصائياً ؟

5 - أيَّ العينتين تعرض تشتتاً أكبر حول وسيطها ، عينة عمال المطبعة (آ) أم عمال المطبعة (ب) ، دعّم إجابتك إحصائياً ؟

6 - اعتماداً على (3) أوجد الآتي :

a) ما احتمال أن يكون عامل ما أجره الأسبوعي أقل من 2200 وحدة نقدية ؟

b) ما نسبة العمال الذين يتراوح أجرهم الأسبوعي ما بين 2200 و3600 وحدة نقدية ؟

c) ما عدد العمال الذين يتراوح أجرهم الأسبوعي 3600 و3800 وحدة نقدية ؟

d) ما احتمال أن يكون أجر العامل أكبر من 2200 وحدة نقدية ؟

e) ما هي نسبة العمال الذين أجورهم الأسبوعية أقل من 3600 وحدة نقدية ؟

26 - أخذت عينة عشوائية من أعمار المستفيدين من إعارة الكتب من مكتبة مركز البحوث الدراسات الاستراتيجية في إحدى الدول ، فأعطت نتائجها الجدول الآتي :

فئات الأعمار	30- 40	40- 50	50- 55	55- 65	65- 75	75- 85
عدد الكتب	20	60	140	55	40	10

والمطلوب :

1 - ما قيمة كل من الربيع الأول والثالث لأعمار المستفيدين من إعارة الكتب من مكتبة المركز ؟

2 - ما قيمة نصف المدى الربيعي النسبي لهذه الأعمار ؟

3 - ما هو العمر الذي تتمركز حوله الكتب المعارة ؟ أوجده ؟

27 - في دراسة إحصائية عن عدد رواد مكتبة الآداب وعدد ساعات دراستهم الأسبوعية فيها ، تبين من أخذ عينة عشوائية من 500 من روادها ، أن متوسط الساعات الأسبوعية 18 ساعة وبانحراف معياري 6 ساعات ، علماً بأن توزيع العينة العشوائية يخضع إلى التوزيع الطبيعي ، والمطلوب :

a) ما نسبة الرواد الذين تقل ساعات دراستهم عن 24 ساعة أسبوعياً ؟
b) ما احتمال الرواد الذين ساعات دراساتهم يتراوح ما بين 12 و30 ساعة أسبوعياً ؟
c) ما عدد الرواد الذين تزيد ساعاتهم الأسبوعية عن 6 ساعات أسبوعياً .

الفصل الخامس

مقاييس التحليل الإحصائي (لظاهرتين)
الارتباط والانحدار

5-1- مقدمة في تحليل الارتباط

5-2- تحليل الانحدار

5-2-1- تحديد طبيعة المتغيرات

5-2-2- رسم شكل الانتشار

5-2-3- مقاييس الارتباط وأنواعه

5-3- الانحدار الخطي

5-3-1- معادلة الانحدار الخطي (خواصها واستعمالاتها)

5-3-2- الخطأ المعياري للتقدير

5-3-3- تحليل التباينات

5-3-4- اختبار معنوية معامل الارتباط

5-4- تحليل الارتباط للبيانات النوعية

5-4-1- معامل الاقتران

5-4-2- معامل ارتباط الرتب " سبيرمان "

5-5- تمارين غير محلولـه .

وقل رب زدني علما
جامعة دمشق
Damascus University

1-5- مقدمة في تحليل الارتباط

من المعروف أن كل ظاهرة من الظواهر ، لا تنشأ من تلقاء نفسها بدون الغير ، وإنما تقع ضمن نطاق تأثيرات الظواهر الأخرى ، هذا يعني أن هناك علاقة بين الظواهر ، ويتوجب علينا معرفة هذه العلاقة وطبيعتها ، وفق أسلوب إحصائي يدعى بالارتباط .

إن دراسة الارتباط تقوم على بحث العلاقة بين ظاهرتين أو أكثر ، لمعرفة ما إذا كان تغير ، إحدى الظاهرتين مرتبطاً بتغير الأخرى ، ومعرفة طبيعة العلاقة ونوعها ، وشدتها ، بين الظاهرتين المدروستين. فمثلاً هناك علاقة بين الدخل والإنفاق ، وبين الدخل والادخار ، وبين الإنتاج والمبيعات ، وبين معدلات النمو ومعدلات الاستثمار ، وبين العمر وضغط الدم ... الخ .

إن هذه العلاقات بين المتغيرات السابقة تسمى بالارتباط ، وإن تحليل العلاقات الارتباطية بين المتغيرات يدعى بتحليل الانحدار ، فالانحدار يبحث في تفسير تغير أحد المتحولات بدلالة مجموعة أخرى من المتحولات ، بينما يهدف الارتباط إلى قياس شدة العلاقة بين عدة متحولات.

ومن الضروري الإشارة إلى أن وجود علاقة ارتباطية بين ظاهرتين لا يعني بالضرورة وجود علاقة السببية بينهما ، فقد يكون الارتباط نتيجة عوامل خارجة عن نطاق المتغيرين موضوع الدراسة ، بالإضافة إلى ذلك أن بعض الحالات تدل على وجود علاقة ارتباط زائفة بين المتغيرات ، بمعنى أن مقاييس الارتباط تدل على وجود علاقة قوية بين المتغيرين ، إلاَّ أن طبيعة المتغيرين تدل على أن الارتباط ليس له أي مدلول ، مثل العلاقة بين عدد رواد مكتبة ما في دولة ما وأسعار الكتب في دولة أخرى، فالارتباط يجب أن يكون منطقياً من حيث الطبيعة أو الظروف المحيطة بها .

الشروط الواجب توافرها في دراسة العلاقة الارتباطية بين ظاهرتين :

- أن تكون العلاقة بين المتغيرات قائمة على المنطق .
- أن تكون إحدى الظواهر مسببة والأخرى ناتجة .
- أن تكون الظاهرتان قابلتين للقياس بوحدة قياس معينة لكل منها (مستقلة أو مشتركة) .

- أن تكون القياسات المأخوذة للظاهرتين متقابلة من حيث الزمان والمكان أو كليهما معاً .

2-5- تحليل الارتباط

1-2-5- تحديد طبيعة المتغيرات

نقصد بطبيعة المتغيرات تحديد كل من المتغير المستقل والمتغير التابع ، فالمتغير المستقل هو المتغير الذي تكون تغيراته غير مرتبطة بتغيرات الظاهرة الثانية ونرمز له بـ (X) ، بينما المتغير التابع هو المتغير الذي تكون تغيراته مرتبطة أو مرهونة بتغيرات المتغير المستقل ونرمز له بـ (y)،على سبيل المثال العلاقة بين الدخل والإنفاق ،نجد أن الدخل هو المتغير المستقل (x) والإنفاق هو المتغير التابع (y) وكذلك الأمر العلاقة بين العمر وضغط الدم ، حيث العمر هو المتغير المستقل (x) وضغط الدم هو المتغير التابع (y) وأيضاً المبيعات (x) والإنتاج (y) .

فمن أجل تحديد أي المتغيرين هو التابع وأيهما المستقل ، هناك عدة طرق أو قواعد متبعة في التحديد:

أ) طريقة المحاكمة أو المناقشة العقلانية : فمثلاً عند دراسة العلاقة بين الدخل والإنفاق ، فمن المنطق أن الدخل هو المتغير المستقل والإنفاق هو المتغير التابع ، على اعتبار أن أي تغير في الدخل يرافقه تغيراً في الإنفاق والعكس غير صحيح ، وأيضاً العلاقة بين الإنتاج والمبيعات ، فالذي يحدد الإنتاج هو المبيعات والعكس غير صحيح .

ب) طريقة الأسبقية الزمنية : تعتمد هذه الطريقة اعتبار المتغير الأسبق في الحدوث زمنياً هو المتغير المستقل ، والثاني هو المتغير التابع .

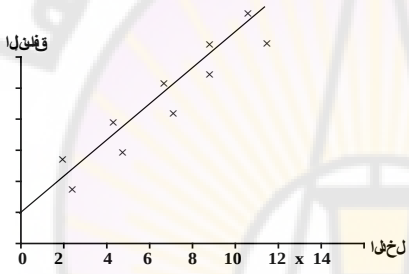
جـ) إذا صادف حدوث المتغيرين بآنٍ واحد فالباحث له الخيار في تحديد أي المتغيرات هو المستقل وأيهما هو التابع .

وخلاصة القول : إن الطريقة الأكثر استخداماً في تحديد المتغير المستقل والتابع هي طريقة المحاكمة العقلانية ثم الأسبقية الزمنية في الحدوث .

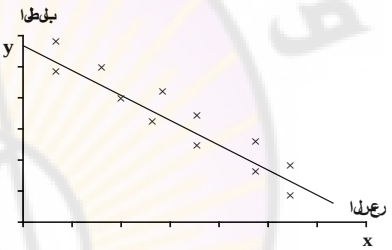
5-2-2- رسم شكل الانتشار

لتحليل سلوك متغيرين أو ظاهرتين ، نقوم بتمثيل قيم الأزواج المتقابلة للظاهرتين على مستوى ديكارتي واحد ، بحيث تمثل قيم (x) على المحور الأفقي وقيم (y) على المحور العمودي .

إن الهدف من رسم شكل الانتشار ، هو تحديد نوع العلاقة بين المتغيرين وطبيعتها ، هل العلاقة بين المتغيرين المدروسين علاقة خطية ، أم من الدرجة الثانية ، أو آسية ، وهل الارتباط طردي (موجب) أم عكسي (سالب) ، أم انعدام ارتباط ، لذلك نميز هنا، بين أشكال الانتشار التالية للعلاقة بين متغيرين :



شكل (6 - 1)



شكل (6 - 2)

نلاحظ من الشكل (6-1) أن نقاط الانتشار متوضعة حول خط مستقيم ، تتزايد فيه قيم (y) مع تزايد قيم (x) ، وهذا يعني أن العلاقة خطية طردية بين المتغيرين (x, y) وإن شكل معادلة التمثيل من الدرجة الأولى

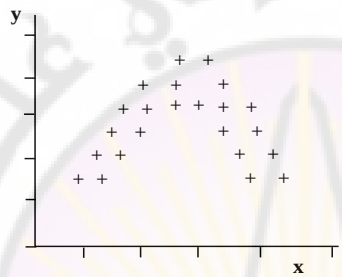
$y = f(x)$.

ومن الشكل رقم (6-2) نجد أن النقاط أيضاً منتشرة حول خط مستقيم ، وتتناقص فيه قيم (y) مع تزايد قيم (x) ، أي أن العلاقة خطية عكسية .. مثال السعر وحجم الطلب على السلع .

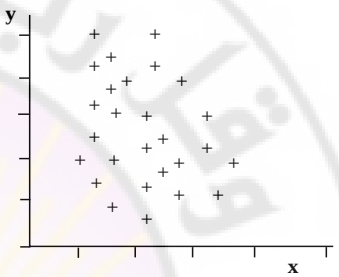
ومن الشكل رقم (6-3) نلاحظ أن نقاط الانتشار مبعثرة بدون ترابط حول اتجاه معين مما يدل على أنه لا توجد علاقة بين المتغيرين (انعدام الارتباط) .

ومن الشكل رقم (6-4) : نجد أن نقاط الانتشار تقع داخل حزمة لها شكل منحنٍ ويستنتج من ذلك أن الارتباط غير خطي ، وأن معادلة التمثيل من الشكل التالي :

$$y = a_0 + a_1 x + a_2 x^2$$



شكل (6 - 4)



شكل (6 - 3)

يتضح مما تقدم ، أن إجراء تحليل الارتباط بين متغيرين أو ظاهرتين يعتمد بالدرجة الأولى على تصور العلاقة القائمة بينهما ، وأن الهدف من رسم شكل الانتشار هو تحديد نوع العلاقة بين المتغيرات وطبيعتها.

5-2-3- مقاييس الارتباط وأنواعه

ثمّة مقاييس للارتباط نذكر ، منها :

a) معادلة التقدير أو الانحدار : وهي العلاقة التي تصف العلاقة بين المتغيرين موضوع الدراسة ، وتكون غايتها : تقدير القيم التي يمكن أن يأخذها المتغير التابع بدلالة المتغير المستقل ؛

b) الخطأ المعياري للتقدير : وهو يبّين مقدار انحراف القيم الفعلية للمتغير التابع عن قيمها المقابلة لها والمحسوبة من معادلة التقدير أو الانحدار ، وغايته تكمن في تحديد المقدرة التمثيلية لمعادلة التقدير ؛

c) معامل الارتباط : وهو يقيس شدة العلاقة أو درجة العلاقة بين المتغيرين ، وأن قيمته تتراوح بين $[-1,+1]$ ويرمز له بـ r ، وتربيع هذا المعامل يُعطي ما يسمى بمعامل التحديد الذي يدل على النسبة المئوية من تباين المتغير التابع التي يفسرها تباين المتغير المستقل ، وقيمته تتراوح بين

$r^2 \in [0,+1]$ ولا يمكن أن تكون سالبة ، وسوف نتناول هذه المقاييس بالتفصيل ، كما سنرى فيما بعد .

ونشير هنا إلى أهم أنواع الارتباط ،ومنها :

- الارتباط البسيط المستقيم ؛
- الارتباط البسيط غير المستقيم ؛
- الارتباط الخطي المتعدد ؛
- الارتباط الذاتي ؛
- الارتباط الجزئي ؛
- ارتباط الرتب ومعامل الاقتران ، وسوف تقتصر دراستنا فقط على كل من الارتباط البسيط المستقيم وارتباط الرتب ومعامل الاقتران .

5-3- الانحدار الخطي

5-3-1- معادلة الانحدار الخطي (خواصها – استعمالاتها)

إذا كان لدينا متغيران (x , y) أحدهما متغير مستقل (x) والثاني متغير تابع (y) لظاهرتين قيد الدراسة ، مثل العلاقة بين الدخل والإنفاق والعمر وضغط الدم ... الخ ، فإنه يوجد أسلوبان لتحديد ثوابت معادلة التقدير والتي تأخذ الصيغة التالية:

$$\tilde{y}_i = a + bx_i$$

حيث :

a تشير إلى ثابت خط الانحدار ، أي قيمة ($\tilde{y}_i$) عندما (x = 0) .

b معامل الانحدار ، (ميل خط الانحدار) ، ويدل على مقدار التغير في y (القيم المقدرة) لكل وحدة تغير واحدة في المتغير المستقل (x) .

ولتحديد هذين الثابتين a و b ، نعتمد إما :

1- أسلوب الرسم الحر :

لتوضيح هذا الأسلوب ، نسوق المثال التالي : أخذت عينة عشوائية مكونة من (7) أسر في مدينة ما ، بغية دراسة العلاقة بين الدخل والإنفاق السنوي ، لهذه الأسر ،

على الثقافة ، فأعطت البيانات التالية : وقد قيس كل من الدخل والإنفاق بآلاف الوحدات النقدية :

الدخل (x)	38	32	42	48	40	44	50
الإنفاق (y)	24	21	27	30	27	33	36

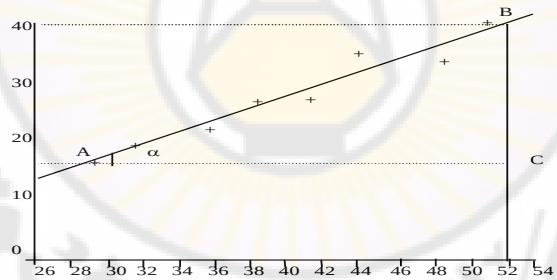
والمطلوب : إيجاد معادلة خط الانحدار ، بأسلوب الرسم الحر ، ثم بين بالاعتماد على شكل الانتشار

الناتج ، نوع العلاقة بين المتغيرين وطبيعتها ؟

الحل :

من أجل تحديد نوع العلاقة بين المتغيرين وطبيعتها ، نرسم شكل الانتشار ، ونلاحظ من الشكل أن العلاقة بين المتغيرين خطية ، على اعتبار أنه يمكن وصف العلاقة المتوسطة بين المتغيرين بمعادلة خط مستقيم من الشكل:

$$\tilde{y}_i = a + bx_i$$



شكل (6 – 5)

وواضح أيضاً من الشكل (6-5) ، أن نقاط الانتشار متوضعة حول خط مستقيم تتزايد فيه قيم (y) بتزايد قيم (x) ونستنتج من ذلك أن العلاقة بين المتغيرين طردية موجبة .

ولتحديد الثابتين وإيجاد المعادلة (معادلة التقدير) بأسلوب الرسم الحر نقوم بتوفيق أو تمرير مستقيم ، يمر من بين نقاط الانتشار ، على اعتبار أنه أفضل مستقيم يمثل العلاقة المتوسطة بين المتغيرين . ثم نحدد على الخط الممهد الذي رسمناه النقطتين

(A , B) ثم نسقط من النقطة (B) خطاً عمودياً على المحور (x) يوازي محور (y) ، ومن النقطة (A) نرسم خطاً أفقياً يوازي المحور (x) ويتقاطع مع المحور العمودي الذي قمنا برسمه في النقطة (c) فيكون ميل المستقيم (b) هو عبارة عن " $\tan(\alpha)$ "

$$\tan(\alpha) = \frac{Bc}{Ac} = b$$

أما الثابت (a) فهو عبارة عن نقطة تقاطع خط الانحدار (الخط الممهد) مع المحور (y). وبالتالي نحصل على ثوابت المعادلة .

ولكن هذا الأسلوب غير دقيق ، لأنه يمكن تمرير عدد غير متناه من المستقيمات من بين نقاط الانتشار ، وسوف نحصل لكل مستقيم على ثوابت (a , b) مختلفة من مستقيم لآخر . لهذا السبب فإن هذه الطريقة قليلة الاستخدام في التوفيق وتعتمد بشكل كبير على خبرة الباحث ومهارته .

2- أسلوب المربعات الصغرى : (المعادلات الطبيعية)

إن أسلوب المربعات الصغرى ، يبني نظريته ، على أساس أن أفضل مستقيم يمثل العلاقة المتوسطة بين (x , y) هو ذلك المستقيم الذي يجعل الفروق بين القيم النظرية والفعلية أقل ما يمكن ، أي الذي يحقق العلاقة :

$$\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 = \min$$

نبدل $\tilde{y}_i$ بقيمتها في المعادلة الأساسية :

$$\tilde{y}_i = a + bx_i$$

فتصبح العلاقة $F = \sum (y_i - a - b\,x_i)^2$

ولحساب (b, a) نحسب المشتقات الجزئية الأولى للتابع F بالنسبة لكل من (a , b) فنجد أن مشتقه بالنسبة لـ a يساوي :

$$\frac{\partial F}{\partial a} = \frac{\partial}{\partial a}\sum (y_i - a - b x_i)^2$$

$$= \sum \frac{\partial}{\partial a}(y_i - a - b x_i)^2$$

$$= \sum 2(y_i - a - b x_i)\cdot(-1)$$

وبالإصلاح نحصل على المعادلة الطبيعية الأولى :

$$\sum y_i = na + b\sum x_i \quad (1)$$

ولإيجاد المعادلة الثانية نشتق F بالنسبة لـ b فنجد أن :

$$\frac{\partial F}{\partial a} = \sum 2 (y_i - a - b x_i)(-x_i)$$

وبالإصلاح نحصل على المعادلة الثانية :

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2 \quad (2)$$

نسمي هاتين المعادلتين بالمعادلتين الطبيعيتين لحساب الثابتين (a , b) ، مثال : وبالعودة إلى التمرين السابق ، المتعلق بالدخل والإنفاق السنوي على الثقافة لـ (7) أسر ، نقوم بإيجاد معادلة خط الانحدار بأسلوب المربعات الصغرى ، ومن أجل ذلك ننشئ جدولاً مساعداً لحساب المجاميع لـ

$$: \sum x , \sum y , \sum x^2 , \sum x y$$

i	x	y	x . y	x^2	y^2	$\tilde{y}$	$(y - \tilde{y})^2$	$(\tilde{y} - \bar{y})^2$
1	38	24	912	1444	576	25.21	1.4641	9.42
2	32	21	672	1024	441	20.59	0.1681	59.13
3	42	27	1134	1764	729	28.29	1.6641	0.001
4	48	30	1440	2304	900	32.91	8.4681	21.43
5	40	27	1080	1600	729	26.75	0.0625	2.34
6	44	33	1452	1936	1089	29.83	10.0489	2.40
7	50	36	1800	2500	1296	34.45	2.4025	38.06
∑	297	198	8490	12572	5760	203.03	24.2783	132.781

وبالتعويض في المعادلات :

$$\sum y = n\,a + b \sum x \qquad (1)$$

$$\sum x \,.\, y = a \sum x + b \sum x^2 \qquad (2)$$

$$198 = 7a + 294\,b \qquad (1)$$

$$8490 = 294\,a + 12572\,b \qquad (2)$$

نضرب المعادلة (1) بقيمة (42) (طريقة الحذف بالتعويض) نجد أن :

$$83490 = 294\,a + 12348 \qquad (3)$$

وبطرح (3) من (2) $\qquad 174 = 0 + 224\,b$

$$\mathbf{b = \frac{174}{224} = 0.77}$$

وبالتعويض في (1) نجد $\qquad 198 - 234\,(0.77) = 7\,a$

$$a = \frac{28.38}{7} = -4.05$$

نعوض في معادلة خط الانحدار :

$$\tilde{y}_i = -4.05 + 0.77X_i$$

التفسير : تمثل (a) القيمة الوسطية للمتغير التابع ($\tilde{y}$) عندما تكون قيمة $x = 0$ ، وهي تسمى ثابت مستقيم الانحدار ، وتمثل (b) قيمة معامل الانحدار ، و بما أن إشارتها موجبة ، فإن ه ذا يعني أن العلاقة طردية بين المتغيرين ، و هي تعني أن ($\tilde{y}$) تزداد بمقدار (b) وسطياً مع ازدياد قيمة (x) وحدة واحدة .

3- أسلوب المربعات الصغرى (الطريقة المباشرة) : يمكن حساب الثوابت بالطريقة المباشرة وفق الصيغ التالية:

$$b = \frac{\sum x \cdot y - n\bar{x}\,\bar{y}}{\sum x^2 - n(\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

وبالتطبيق ينتج لدينا :

$$b = \frac{8490 - 7(42)\ (28.28)}{12572 - 7(42)^2} = 0.77$$

$$a = 28.28 - 0.77(42) = -4.05$$

$$\boxed{\tilde{y} = -4.05 + 0.77\,x}$$

- **خواص معادلة الانحدار**

1- إن خط الانحدار يمر بنقطة إحداثياتها ($\bar{x}$, $\bar{y}$) .

2- مجموع الفروق بين القيم الفعلية والقيم المقدرة للمتغير التابع أصغر ما يمكن .

3- إن مجموع القيم النظرية يساوي مجموع القيم الفعلية ، أي : $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \tilde{y}_i$.

- **استعمالات معادلة الانحدار**

1- استخدامها في التقدير : أي في الحصول على تقدير القيمة الوسطية للمتغير التابع من أجل قيمة معينة من قيم (x) في العينة وذلك بالتعويض في معادلة خط الانحدار .

$$\boxed{\tilde{y} = -4.05 + 0.77\,x}$$

من أجل قيم (x) الواردة في التمرين السابق نجد أن :

$$\boxed{\tilde{y} = -4.05 + 0.77\,x}$$

$$\boxed{\tilde{y} = -4.05 + 0.77\,(38) = 25.21}$$

نلاحظ أن القيمة الوسطية (النظرية) من أجل $x = 38$ هي القيمة $\tilde{y} = 25.21$.

وهكذا بالنسبة لبقية القيم الواردة في الجدول السابق (يترك للطالب التحقق من ذلك).

2- استخدامها في التنبؤ : أي في الحصول على توقع قيمة المتغير التابع (y) بدلالة قيم (x) الواقعة ضمن مجال تغيرات (x) ، وذلك لضمان بقاء العلاقة خطية بين المتغيرات وهذا ما يدعى بالتمديد الداخلي .

تمرين : قدر إنفاق أسرة يبلغ دخلها (33) وحدة نقدية من معادلة الانحدار السابقة ، نعوض في معادلة خط الانحدار :

$$\tilde{y}_i = -4.05 + 0.77x_i = -4.05 + 0.77(33) = 21.36$$

وهي القيمة المتوقعة للإنفاق ، لأسرة دخلها (33) وحدة نقدية ، أما بالنسبة لتقدير قيم ($\tilde{y}$) من أجل قيم (x) الواقعة خارج المجال الذي تحدده بيانات العينة ، ففي هذه الحالة لا نستطيع استخدام معادلة خط الانحدار في إجراء التنبؤ أو التقدير، لأننا لا نستطيع أن ن ضمن استمرارية العلاقة الخطية بين المتغيرين موضوع الدراسة خارج قيم تحول المتغير المستقل والتي على أساسها تم إيجاد معادلة التقدير .

رسم خط الانحدار :

من أجل الرسم البياني لخط الانحدار نوضح ذلك من خلال التمرين السابق ، ونأخذ المعادلة :

$$\tilde{y}_i = -4.05 + 0.77\, x_i$$

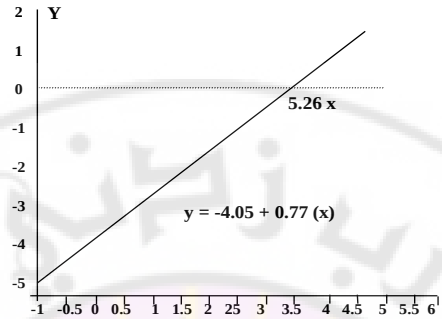
ونضع : $\tilde{y}_i = 0$

فنجد أن : $-4.05 + 0.77\, x = 0$ ، ومنه :

$$4.05 = 0.77\, x \Rightarrow x = \frac{4.05}{0.77} = 5.25$$

ثم نضع : $x=0$ ، فنجد أن :

$$\tilde{y}_i = -4.05$$



شكل (6 – 6)

5-3-2- الخطأ المعياري للتقدير لمعادلة خط الانحدار

هو عبارة عن مقياس للتشتت ، ويقيس انتشار أو تشتت (y) الفعلية حول($\tilde{y}$) أي حول خط الانحدار ، بمعنى أن هذا المقياس يبين مقدار انحراف (y) عن ($\tilde{y}$) المحسوبة من معادلة خط الانحدار ، والخطأ المعياري للتقدير يعرف رياضياً ، بأنه الجذر التربيعي لمجموع مربع انحرافات القيم المقدرة عن القيم الفعلية للمتغير التابع ، ويعطى بالعلاقة التالية :

$$S_{y\tilde{y}} = \sqrt{\frac{\sum (y - \tilde{y})^2}{n}} = \sqrt{\frac{\sum y^2 - \left(a\sum y + b\sum xy\right)}{n}}$$

وذلك بعد تعويض y بقيمتها وإصلاح العلاقة .

إن استخدامات هذا المقياس تشابه تماماً استخدامات الانحراف المعياري بالنسبة لمقاييس النزعة المركزية ، فهو يستخدم كمعيار لمقدار الثقة التي نضعها في القيم المحسوبة ($\tilde{y}_i$) من معادلة خط الانحدار بدلالة المتغير (x) .

وكلما كبرت قيمة هذا المقياس كلما قلت الثقة بالقيم المقدرة وأصبحت المصداقية أقل ، وبالعكس كلما قلت قيمة هذا المقياس كلما ازدادت الثقة بالقيم المقدرة $\tilde{y}$ والمحسوبة من معادلة خط الانحدار ، أي كلما ازداد التقارب بين القيم الفعلية والقيم النظرية ، وإذا كانت قيمة هذا المقياس $S_{y\tilde{y}} = 0$ ، فهذا يعني أن جميع نقاط

الانتشار تقع على الخط المستقيم دون انتشار أو تشتت وهذا ما ندعوه بالارتباط التام ، وفيه يكون :

$$\tilde{y} = bx$$
$$y_i = \tilde{y}_i$$
$$a = 0$$
$$\bar{y} = \bar{x}$$

وأخيراً نشير إلى أ ن مربع هذا الخطأ المعياري للتقدير ($S^2_{y\tilde{y}}$) لمعادلة خط الانحدار يسمى التباين غير المفسر .

مثال : بالعودة إلى التمرين السابق ، والمتعلق بدخول الأسر ونفقاتها على الثقافة.

والمطلوب : حساب الخطأ المعياري للتقدير لمعادلة خط الانحدار ؟

- بتطبيق الصيغة السابقة :

$$S_{y\tilde{y}} = \sqrt{\frac{\sum (y - \tilde{y})^2}{n}} = \sqrt{\frac{24.2783}{7}} = 1.862$$

- أو :

$$S_{y\tilde{y}} = \sqrt{\frac{\sum y^2 - (a\sum y + b\sum yx)}{n}}$$
$$= \sqrt{\frac{5760 - \left(4.05(198) + 0.77(8490)\right)}{7}} = 1.862$$

وبتربيع الناتج نحصل على قيمة التباين غير المفسر أي :

$$S^2_{y\tilde{y}} = (1.862)^2 = 3.468$$

ويستخدم الخطأ المعياري للتقدير لحساب التقدير المجالي لقيم ($\tilde{y}$) من العلاقة :

$$\Pr\left(Z_{\alpha/2} \le Z \le Z_{\alpha/2}\right) = 1 - \alpha$$
$$Z = \frac{\tilde{y} - y}{S_{y\tilde{y}}}$$

وبالتعويض بالعلاقة نحصل على التقدير المجالي لقيمة ($\tilde{y}$) .

$$\tilde{y} - Z_{\alpha/2} \cdot S_{\tilde{y}\, y} \le y \le \tilde{y} + Z_{\alpha/2} \cdot S_{\tilde{y}\, y}$$

من أجل احتمال ($1 - \alpha$) يمكن إيجاد القيمة المجالية لـ (y) باستخدام (y) وبالعودة إلى التمرين السابق ، نقوم بتقدير قيمة الإنفاق لآسرة دخلها (33) وحدة نقدية عند مستوى دلالة $\alpha = 0.05$ ؟

نقوم أولاً بحساب التقدير النقطي لـ y :

$$\tilde{y} = -4.05 + 0.77\,(33) = 21.36$$

ومن ثم التقدير المجالي ، فنجد :

$$\tilde{y} - Z_{\alpha/2} \cdot S_{\tilde{y}\, y} \le y \le \tilde{y} + Z_{\alpha/2} \cdot S_{\tilde{y}\, y}$$

$$21.36 - 1.96 \cdot (1.862) \le y \le 21.36 + 1.96\,(1.862)$$

$$17.17 \le y \le 25.009$$

وإن التفسير الرياضي لذلك هو أنه باحتمال 95 % ، أو بمستوى دلالة ($\alpha = 5$ %) (إن إنفاق الأسرة التي يبلغ دخلها (33) وحدة نقدية ، لن يقل عن 17.17 وحدة نقدية ولن يزيد عن 25.009 وحدة نقدية ومن الضروري الإشارة إلى أن حجم العينة يجب أن يكون $n > 30$ مفردة حتى يكون التقدير صحيحاً ، أما إذا كان حجم العينة أقل من ذلك ، أي : $n \le 30$ ، في هذه الحالة نستخدم توزيع (t) ستودنت ، وتصبح علاقة الخطأ المعياري للتقدير :

$$S_{y\tilde{y}} = \sqrt{\frac{\Sigma\,(y - \tilde{y})^2}{n - 1}}$$

5-3-3- تحليل التباينات

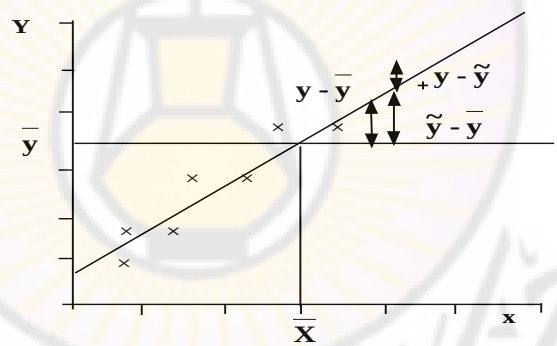
لقد رأينا من خلال دراستنا للتمرين السابق ، (المتعلق بالعلاقة بين الدخل والإنفاق على الثقافة) ، أن استخدام معادلة خط الانحدار قد ساعد على التقليل من أخطاء التقدير ، فوجدنا أن الخطأ المعياري للتقدير أقل بكثير من الانحراف المعياري للوسط الحسابي للمتغير التابع (y) ، ويعرف التباين الكلي (S_y^2) بأنه مجموع مربع الانحرافات للمتغير التابع (y) عن وسطها الحسابي ويعطي بالعلاقة التالية ، وذلك لعينات $n > 30$:

$$S_y^2 = \frac{\sum (y - \tilde{y})^2}{n} \qquad \text{أو} \qquad S_y^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n} \right)^2$$

ومن الممكن إظهار التباين الكلي على شكل الانتشار كما هو موضح بالرسم (انظر 6 - 7) ويقسم التباين الكلي إلى قسمين :

1 – التباين غير المفسر : هو مربع الخطأ المعياري للتقدير $S^2_{y\tilde{y}}$ الذي تحدثنا عنه سابقاً وبعبارة أخرى هو عبارة عن التباين الذي لا تفسره معادلة خط الانحدار .

2 – التباين المفسر : هو عبارة عن التباين الذي تفسره معادلة خط الانحدار ، وكلما زادت قيمته دل ذلك على إمكانية تحديد قيم المتغير (y) بدلالة المتغير (x) بدقة أكبر وتزداد قيمة هذا المعامل كلما أزداد الفرق بين $\tilde{y}$ و y ، وبالتالي يكون الارتباط تاماً ، وفي هذه الحالة يكون التباين الكلي مساوياً للتباين المفسر .



شكل (6 - 7)

ويعطى التباين المفسر بالعلاقة التالية :

$$S^2\,\tilde{y} = \frac{\sum \left(\tilde{y} - \bar{y} \right)^2}{n}$$

أو :

$$S^2\,\tilde{y} = \frac{b \sum \left(x y - \bar{x} \sum y \right)}{n}$$

وبالعودة إلى التمرين السابق ، نقوم بحساب كلٍ من التباين المفسر والتباين غير المفسر:

- التباين المفسر : $S^2\,\tilde{y} = \frac{\sum\left(\tilde{y} - \bar{y}\right)^2}{n} = \frac{132.78}{7} = 18.96$

- أما التباين غير المفسر: هو عبارة عن مربع الخطأ المعياري للتقدير الذي تم حسابه سابقاً : $S^2_{y\tilde{y}} = (1.862)^2 = 3.467$

التباين الكلي = التباين المفسر + التباين غير المفسر .

$$S^2_y = S^2_{\tilde{y}} + S^2_{y\tilde{y}}$$

$$22.436 = 18.969 + 3.467$$

ومنه نستنتج أن هذه القيمة هي نفسها قيمة التباين الكلي (مربع الانحراف المعياري).

$$S^2_y = (4.737)^2 = 22.436$$

معامل الارتباط (بيرسون) : يستخدم هذا المقياس لقياس درجة العلاقة بين متغيرين أو شدتها ، وإن قيمته تتراوح بين [1 + , 1-] أي أن $r \in$ [1 + , 1-] ولكن إشارته دوماً من إشارة (b) معامل الانحدار ، وكلما اقتربت قيمة هذا المعامل من ($\pm$ 1) دل ذلك على قوة أكبر للعلاقة بين المتغيرين وبالعكس .

نحن نعلم :

$$S^2_y = S^2_{\tilde{y}} + S^2_{y\tilde{y}}$$

وبقسمة الطرفين على S^2_y نجد :

$$\frac{S^2\,y}{S^2\,y} = \frac{S^2\,y\,\tilde{y}}{S^2\,y} + \frac{S^2\,\tilde{y}}{S^2\,y}$$

$$1 = \frac{S^2\,y\,\tilde{y}}{S^2\,y} + \frac{S^2\,\tilde{y}}{S^2\,y}$$

ومن العلاقة الأخيرة :

$$\frac{S^2\,\tilde{y}}{S^2\,y} = 1 - \frac{S^2 y\,\tilde{y}}{S^2\,y}$$

نرمز للتباين المفسر النسبي $\frac{S^2\,\tilde{y}}{S^2\,y}$ بـ r^2 وهو ما يعرف بمعامل التحديد ، فنجد أن :

$$r^2 = 1 - \frac{S^2 y\,\tilde{y}}{S^2\,y}$$

وهو أيضاً يساوي :

$$\mathbf{r^2 = \frac{S^2\,\tilde{y}}{S^2\,y}}$$

ويمكن البرهان على أن **معامل التحديد** للارتباط الخطي هو عبارة عن مربع معامل الارتباط ، وعندما يكون الارتباط تاماً ، أي $r = 1$ ، فإنه في هذه الحالة يكون $S_y^2 = S_{\tilde{y}}^2$ وتكون قيمة ($r^2 = 1$) ، وعندما يكون الارتباط معدوماً ($r = 0$) تكون قيمة $S_{y\tilde{y}}^2 = 0$ وبالتالي فإن ($r^2 = 0$) ، ومما سبق يمكننا أن نستخلص أن $r^2 \in [\,0\,,\,1\,]$ ، وهو يدل على النسبة المئوية من تباين المتغير التابع ، التي يفسرها تباين المتغير المستقل ، أو هو النسبة المئوية من التباين الكلي ، التي تفسرها معادلة خط الانحدار . وبالعودة إلى التمرين السابق ، والمطلوب : حساب معاملي الارتباط والتحديد وفسرهما ؟

$$r^2 = 1 - \frac{S^2 y\,\tilde{y}}{S^2\,y} = 1 - \frac{3.467}{22.436} = 0.84$$

$$r^2 = 1 - \frac{S^2\,\tilde{y}}{S^2\,y} = \frac{18.969}{22.436}\,0.84$$

والتفسير هو : أن 84 % من تباين المتغير التابع ، يفسرها تباين المتغير المستقل ، أو أن 84 % من التباين ، تفسره معادلة خط الانحدار ، وإن 16 % من التباين لا تفسره المعادلة . أي 84 % من التباين ، تفسره معادلة خط الانحدار ، وإن 16 % من التباين لا تفسره المعادلة . أي 84 % من تباين الإنفاق يفسرها تباين الدخل و 16 % تعود لمتغيرات أخرى لم تدخل في معادلة خط الانحدار .

ولحساب معامل الارتباط r ، يمكننا في هذه الحالة أخذ الجذر التربيعي لمعامل التحديد ، فنجد أن :

$$\mathbf{r = \sqrt{r^2} = \sqrt{0.84} = 0.91}$$

وإن قيمة معامل الارتباط ، تدل على أن العلاقة بين المتغيرين قوية في مثالنا .

كما يمكن حساب معامل الارتباط بالطرق المباشرة الآتية :

$$r = \sqrt{\frac{a\sum y + b\sum xy - n\bar{y}^2}{\sum y^2 - n\bar{y}^2}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{n.\sigma_x.\sigma_y}$$

أو من العلاقة المحسوبة منها :

$$\mathbf{r = \frac{n\sum xy - \sum x\sum y}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}}$$

وبالتطبيق على التمرين السابق نجد أن :

$$r = \sqrt{\frac{-4.05(198) + 0.77(8490) - 7(28.28)^2}{5760 - 7(28.28)^2}} = 0.92$$

$$r = \frac{7(8490) - (294)(198)}{\sqrt{[7(12572) - (86436)][7(5760) - (39204)]}} 0.92$$

5-3-4- اختبار معنوية معامل الارتباط

إذا سحبنا عدداً كبيراً من العينات العشوائية حجمها (n) وحسبنا معاملات الارتباط ، من هذه العينات ، فإن هذه المعاملات تخضع لمنحنى توزيع طبيعي (التوزيع العيني لمعاملات الارتباط) إذا كان المجتمع الإحصائي (N) يتوزع طبيعياً ، بغض النظر عن حجم العينات المسحوبة .

أما إذا كان المجتمع ، لا يتوزع طبيعياً وكانت حجوم العينات المسحوبة ($n > 30$) فإن توزيع معاملات الارتباط تقترب من شكل التوزيع الطبيعي ، وإذا كان $n \leq 30$ والمجتمع لا يخضع لتوزيع طبيعي فإن شكل التوزيع لمعاملات الارتباط يخضع لتوزيع (t) ستودنت .

هذا ويتم اختبار معامل الارتباط المحسوب من العينة بالعلاقة :

$$t = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

وذلك من أجل عدد من درجات الحرية : $V = (n-2)$

ومستوى دلالة محدد .... 10 % , 5 % = α

ويتم حساب درجات الحرية من عدد المشاهدات المستقلة في العينة مطروحاً منها العدد(2) .

نقارن النتيجة التي نحصل عليها من الاختبار السابق (t) بالقيمة النظرية المستخرجة من الجداول الخاصة بالتوزيع (t) .

وبالعودة إلى التمرين السابق ، والمطلوب : اختبار معنوية معامل الارتباط ، أو بيان الأهمية الإحصائية لمعامل الارتباط عند مستوى دلالة $\alpha = 0.05$

الحل :

لقد وجدنا أن : $r = 0.92$ ، وبما أن : $n = 7$ و $V = 7 - 2 = 5$

وبالتطبيق نجد أن :

$$t = r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$= 0.91 \frac{\sqrt{7-2}}{\sqrt{1-(0.91)^2}} = 4.914$$

من خلال مقارنة هذه القيمة $t = 4.914$ (المسحوبة) مع القيمة الجدولية $t = 2.57$ المقابلة لمستوى دلالة $\alpha = 0.05$ ، نجد أن (النظرية) $t > t$ المحسوبة) وبالتالي نرفض فرضية الاستقلال ، والتي تنص بأنه لا توجد علاقة بين الظاهرتين المدروستين في المجتمع الذي سحبت منه العينة ، ونقرر أن هناك ارتباط حقيقي في العينة ، أو بمعنى آخر، إن الارتباط في العينة لا يساوي الصفر ، والعينة المسحوبة تمثل المجتمع الإحصائي الذي سحبت منه ، ولمعامل الارتباط أهمية إحصائية .

الخطأ المعياري لمعامل الارتباط (S.E.C.r) ، الذي يعطي بالعلاقة التالية:

$$S.E.C.r = \sqrt{\frac{1 - r^2}{n - 2}}$$

وبالعودة إلى التمرين السابق والمتعلق بدخول الأسر ونفقاتهم على الثقافة ، نجد الخطأ المعياري لمعامل الارتباط يساوي :

$$S.E.C.r = \sqrt{\frac{1 - 0.84}{7 - 2}} = 0.17$$

ويستخدم هذا المقياس لبيان ، فيما إذا كان معامل الارتباط ، في المجتمع الإحصائي المسحوب منه العينة مساوياً للصفر أم لا ، وبالتالي الحكم على معنوية معامل الارتباط في العينة ، فإذا كان معامل الارتباط في العينة أكبر أو يساوي ثلاثة أمثال خطئه المعياري أي (S.E. C. r) $r \geq 3$ فالعينة المسحوبة تكون عشوائية وتمثل المجتمع الإحصائي الذي سحبت منه ، أفضل تمثيل ، ولمعامل الارتباط في العينة أهمية إحصائية ، أي بمعنى أن معامل الارتباط للمجتمع الإحصائي الذي أخذت منه هذه العينة العشوائية لا يمكن أن يساوي الصفر ، وإنما أية قيمة أخرى موجبة كانت أم سالبة .

وإذا كان (S.E.C.r) $r < 3$ فالعينة المسحوبة لا تمثل مجتمعها الإحصائي ، الذي سحبت منه ، وليس لمعامل الارتباط أية أهمية إحصائية ، والارتباط الذي تبينه العينة يعود لقوى الحظ والصدف .

وبالنسبة للتمرين ، فإن معامل الارتباط أكبر من ثلاث ة أمثال خطئه المعياري ، وبالتالي فإن لمعامل الارتباط أهمية إحصائية ويعبر عن الارتباط بين هاتين الظاهرتين في المجتمع الإحصائي .

5-4- تحليل الارتباط للبيانات النوعية

فيما سبق تكلمنا عن دراسة العلاقة الارتباطية بين مؤشرين قابلين للقياس بواحدة قياس معينة ، وبالتالي يمكننا أن نعبر عنهما بصورة كميّة حسب الزمان والمكان ، غير أننا نلاحظ في حياتنا اليومية كثيراً من الظواهر غير القابلة للقياس الكمي لأسباب قاهرة ، مثل : تصنيف الكتب المتوافرة في مكتبة ما حسب عنوانها أو حسب المادة التي تتضمنها ، أو مستوى تأهيل المكتبيين (مؤهل جداً ، متوسط

التأهيل ، متدني التأهيل) ، أو درجة استيعاب الطالب لمحتويات مقرر ما ، حيث يمكننا القول بأن استيعابه : ممتاز ، جيد ، مقبول ، ضعيف ، أو الحالة التعليمية للفرد ، أو الحالة الاجتماعية للفرد ، .... الخ .

في حياتنا العملية نجد علاقات ارتباطية بين مؤشرين غير قابلين للقياس(مثل العلاقة بين مستوى التأهيل للمكتبيين ونوعهم) أو على الأقل بين مؤشر قابل للقياس وآخر غير قابل للقياس (كالعلاقة بين العمر للفرد ومستواه التعليمي) .

لدراسة مثل هذه العلاقات الارتباطية بين مؤشرات غير مقاسة ، يمكننا أن نطبق معامل الاقتران أو معامل ارتباط الرتب ، وهذا ما سنتطرق له بالتفصيل بعد قليل .

5-4-1- معامل الاقتران

يستخدم معامل الاقتران عندما يكون لدينا متغيران نوعيان ، ولكل منهما قيمتان أو صفتان فقط ، مثل العلاقة بين نوع المستفيدين من مكتبة ما، مما يتوافر فيها من كتب أو ما شابه وبين نوع القصص التي يقوم المستفيدون بإعارتها من تلك المكتبة (قصص طويلة أم قصيرة) ، أو العلاقة بين تأهيل المكتبيين

(مؤهل وغير مؤهل) ونوعهم (ذكور ، إناث) ، العلاقة بين الإصابة بمرض عضال (مصاب ، غير مصاب) لمجموعة من الأشخاص (مدخنين وغير مدخنين) ،.... الخ . فمعامل الاقتران هو ذاك المعامل الذي يقيس متانة الارتباط بين مؤشرين لكل منهما حالتان فقط ، ولإيجاد قيمة معامل الاقتران ، فإننا نحسب تكرارات المفردات حسب المتغيرات والصفات ، كالآتي :

المتغير آ / المتغير ب	الصفة 1	الصفة 2	$\sum$
الصفة 1	n_{11}	n_{12}	$n_{1.}$
الصفة 2	n_{21}	n_{22}	$n_{2.}$
$\sum$	$n_{.1}$	$n_{.2}$	n

حيث أن :

$$n = n_{11} + n_{12} + n_{21} + n_{22}$$

أو

$$n = n_{.1} + n_{.2} = n_{1.} + n_{2.}$$

ويحسب معامل الاقتران من خلال العلاقة الآتية :

$$r_c = \frac{n_{11} * n_{22} - n_{12} * n_{21}}{n_{11} * n_{22} + n_{12} * n_{21}}$$

وكلما كانت قيمة r_c مقتربة من الواحد الصحيح كانت العلاقة بين المتغيرات قوية جداً ، وإذا كانت قيمة $r_c \succ 05$ كان هناك علاقة بين المتغيرات المدروسة.

مثال : بهدف دراسة العلاقة الكائنة بين نوع المستفيدين (ذكور أم إناث) ونوع القصص (طويلة أم قصيرة) المستعارة من إحدى المكتبات ، قام أحد الباحثين بجمع بيانات عن 150 مستفيداً من القصص المتوافرة في مكتبة ما، فكانت كالآتي:

نوع القصص / نوع المستفيد	قصص قصيرة	قصص طويلة	$\sum$
ذكور	70	10	80
إناث	10	60	70
$\sum$	80	70	150

والمطلوب : هل من علاقة بين نوع المستفيد ونوع القصة المستعارة ؟

نلاحظ هنا ، أننا أمام متغيرين ولكل منهما صفتان ، لذا وجب علينا أن نحسب معامل الاقتران ، كما يلي :

$$r_c = \frac{n_{11} * n_{22} - n_{12} * n_{21}}{n_{11} * n_{22} + n_{12} * n_{21}} = \frac{70 * 60 - 10 * 10}{70 * 60 + 10 * 10}$$

$$= \frac{4200 - 100}{4200 + 100} = \frac{4100}{4300} = 0.9535$$

وهذا يدل على وجود علاقة قوية بين نوع المستفيد ونوع القصة ، ويلاحظ أن الذكور يميلون نحو القصص القصيرة ، بينما تميل الإناث نحو القصص الطويلة.

مثال آخر : أخذت عينة عشوائية من 100 طالب ، فحصلنا على النتائج الآتية:

الوضع النهائي للطالب / النوع	ناجح	راسب
ذكر	55	7
أنثى	35	3

والمطلوب : هل من علاقة ارتباطية بين الوضع النهائي للطالب ونوعه ؟

الحل :

نطبق علاقة إيجاد معامل الاقتران ، فنجد أن :

$$r_c = \frac{n_{11} * n_{22} - n_{12} * n_{21}}{n_{11} * n_{22} + n_{12} * n_{21}} = \frac{55*3 - 7*35}{55*3 + 7*35}$$

$$= \frac{165 - 245}{165 + 245} = \frac{-80}{410} = -0.195$$

وهذا يدل على عدم وجود اقتران بين النوع والوضع النهائي للطالب ، فقيمة معامل الاقتران قليلة جداً ، أي ليس هناك من علاقة بين النوع والنجاح .

5-4-2- معامل ارتباط الرتب ، سبيرمان (Sperman) :

إن معامل ارتباط بيرسون ، يستخدم لقياس شدة أو قوة العلاقة بين المتغيرات الكمية ، وفي حالة البيانات النوعية فنستخدم معامل ارتباط سبيرمان ، وتعرف البيانات النوعية: بأنها عبارة عن الأشياء أو الصفات غير القابلة للقياس، مثل (اللون ، العرق ، الجنسية ، الحالة الزواجية) .

إن معامل ارتباط الرتب لسبيرمان ، يعطي قياساً للارتباط للبيانات الكمية والنوعية المتقابلة التي لها صفة الترتيب ، ويعطى بالعلاقة التالية :

$$\mathbf{r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}}$$

حيث تشير :

r_s : معامل ارتباط الرتب .

n : حجم العينة لأزواج من القيم (x , y) .

d : الفرق بين رتب أزواج القيم (x , y) .

مثال : لدينا عينة عشوائية من (7) أسر لدراسة الدخل والإنفاق على الثقافة ، والمطلوب : حساب معامل ارتباط الرتب لسبيرمان ؟

الدخل X	الإنفاق Y	رتب x	رتب y	الفرق d	مربع الفرق d^2
38	24	2	2	0	0
32	21	1	1	0	0
42	27	4	3.5	0.5	0.25
48	30	6	5	1	1
40	27	3	3.5	-0.5	0.25
44	33	5	6	-1	1
50	36	7	7	0	0
					$\sum d^2 = 2.5$

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)} = 1 - \frac{6\left(2\cdot 5\right)}{7\left(49-1\right)} = 1 - \frac{13}{336} = 0.96$$

ملاحظة : لقد تم إعطاء الرتبة 3.5 للقيمتين المتساويتين 27 للمتحول التابع ، وذلك بحسابها من العلاقة الآتية : $\frac{i(i+1)}{2}$ ، أي : $\frac{3(3+1)}{2} = 3.5$ ، ويلاحظ في الغالب أن قيمة معامل ارتباط بيرسون ، وقيمة معامل ارتباط هي قيم متقاربة .

5-5- تمريـــنات غير محلولـة

تمرين (1) : قامت إحدى الشركات برصد إنفاقها على الإعلان وحجم المبيعات ، وتم التوصل إلى النتائج التالية :

الإنفاق على الإعلان	90	80	80	85	87	92	85	90	93	95
حجم المبيعات	5	2	4	6	4	7	5	6	5	7

وقدرت المبيعات والإنفاق على الإعلان بألوف الوحدات النقدية ، والمطلوب:

1 - ارسم شكل الانتشار ، لبيانات الجدول السابق ثم استنتج نوع العلاقة بين المتغيرين وطبيعتها.

2 - أوجد معادلة خط الانحدار وفسر الثوابت .

3 - احسب الخطأ المعياري للتقدير لمعادلة خط الانحدار وبين أهميته .

4 - احسب معاملي الارتباط والتحديد وفسرهما .

تمرين (2) : يبين الجدول التالي سنوات الخبرة ، وحجم المبيعات ، ب ألوف الوحدات النقدية ، لعدد من مندوبي المبيعات في إحدى شركات بيع مستلزمات المكتبات :

رقم البائع	1	2	3	4	5	6	7	8	9	10
سنوات الخبرة	1	3	4	4	6	8	10	10	11	13
حجم المبيعات	80	97	92	102	103	111	119	123	117	136

والمطلوب :

1 - احسب معامل الانحدار وثابت مستقيم الانحدار وفسرهما .

2 - ارسم معادلة خط الانحدار الناتج وبين على الشكل التباين غير المفسر .

3 - قدر باحتمال % 95.5 حجم المبيعات ، لبائع عدد سنوات خبراته (12) سنة .

4- احسب التباين الكلي .

5- احسب الخطأ المعياري للتقدير لمعامل الارتباط وبين استخداماته .

تمرين (3) : تم التوصل من خلال دراسة أجريت ، عن العلاقة بين دخول (14) أسرة ونفقاتها على الثقافة في مدينة دمشق إلى المعلومات التالية :

$$\sum Y = 483, \sum X = 560, \sum XY = 19960, \sigma_y = 1.7, \sigma_x = 3$$

حيث تمثل (x) الدخل الذي يتراوح بين [59 , 28] وحدة نقدية ، وتمثل (y) الإنفاق الذي يتراوح بين

[47 , 27] وحدة نقدية . وقد قيس كل من الدخل والإنفاق بآلاف الوحدات النقدية ، والمطلوب :

1 - ما هي الشروط الواجب توافرها في دراسة العلاقة الارتباطية بين متغيرين .

2 - أوجد معادلة خط الانحدار ثم بين طبيعة العلاقة بين المتغيرين .

3 - ارسم خط الانحدار ثم اذكر خواص هذا الخط .

4 - قدر النفقات لأسرة دخلها (60) وحدة نقدية ؟ برر إجابتك بما تراه مناسباً.

5 - ميز بين المفاهيم التالية ، الارتباط ، الارتباط التام ، الارتباط المعدوم.

6 - بين الأهمية الإحصائية لمعامل الارتباط في العينة ثم اختبره وفق توزيع (t) .

تمرين (5) : بفرض وجود علاقة خطية بين حجم الطلب وسعر السلعة ، ثم التوصل من خلال دراسة أجريت عن العلاقة بين سعر السلعة ، وحجم الطلب على هذه السلعة إلى المعلومات التالية :

$\sum xy = 1596, n = 14, \sum x^2 = 6888, \bar{x} = 18, \bar{y} = 11, \sum y^2 = 2290$ ، علماً أن حجم الطلب على هذه السلعة عندما توزع مجاناً كان (20) وحدة ، والمطلوب :

1 – أوجد معادلة خط الانحدار وبين الغاية أو الهدف من حسابها .

2 – احسب معاملي التحديد والارتباط وفسرهما .

3 – ارسم مستقيم الانحدار الناتج .

4 – ما الفرق بين الارتباط والانحدار .

تمرين (6) : تم التوصل من خلال دراسة عن العلاقة بين الدخل والإنفاق الشهري على الثقافة لعينة عشوائية من (8) أسر في مدينة حلب إلى البيانات التالية:

$$\sum xy = 1895, \sum x^2 = 5864, \bar{x} = 25, \bar{y} = 8.625, \sum y^2 = 651$$

حيث تمثل (x) الدخل الشهري الذي يتراوح بين [31 , 18] وحدة نقدية وتمثل (y) الإنفاق الذي يتراوح بين [14 , 3] وحدة نقدية ، وقد قيس كل منهما بمئات الوحدات النقدية ، والمطلوب :

1 – احسب التباين غير المفسر واشرح مدلوله .

2 – ارسم خط الانحدار وحدد عليه التباين المفسر .

3 – قدر باحتمال 95.5 % إنفاق أسرة يبلغ دخلها (28) وحدة نقدية ، ثم قدر النفقات الشهرية على الثقافة لأسرة دخلها الشهري (35) وحدة نقدية .

4 – احسب معامل التحديد وفسره ؟ ثم بين الأهمية الإحصائية لمعامل الارتباط .

تمرين (7) : أخذت عينة عشوائية من درجات (10) طلاب من كلية التجارة لمقرري الإحصاء والرياضة فحصلنا على الجدول التالي :

رقم الطالب	1	2	3	4	5	6	7	8	9	10
درجته في الإحصاء	70	50	85	40	90	90	44	70	60	80
درجته في الرياضة	67	50	79	41	85	90	40	60	59	73

والمطلوب :

1 – حدد المتغير المستقل والمتغير التابع ، مبرراً إجابتك بما تراه مناسباً .

2 – أوجد معامل ارتباط الرتب لسبيرمان ، وفسره .

3 – أوجد معامل ارتباط بيرسون .

4 – ارسم شكل الانتشار وبين نوع العلاقة بين المتغيرين وطبيعتها .

تمرين (8) : أخذت عينة عشوائية من (8) أسر حيث جمعت معلومات تتعلق بدخولها وادخاراتها فتبين ما يلي :

$\sum X = 520 \qquad \sum X^2 = 34200 \qquad \sum X Y = 7400$

$\sum Y = 112 \qquad \sum Y^2 = 1689$

حيث (x) تمثل دخل الأسرة وتمثل (y) ادخارها . والمطلوب :

1 – احسب ثوابت المعادلة وفسرها .

2 – ارسم مستقيم الانحدار الناتج واذكر خواصه .

3 – هل من أهمية إحصائية لمعامل الارتباط ، علل بما تراه مناسباً .

4 – لقد وجد أن أسرة دخلها (56) وحدة نقدية وتدخر (18) وحدة نقدية ، فهل تعتقد أن ادخارها يقع ضمن المجال وذلك باحتمال 95 % .

تمرين (9) : أخذت عينة عشوائية من 100 طالب ، فحصلنا على النتائج الآتية:

التأهيل / عدد المكتبيين	مؤهلين	غير مؤهلين
نفذوا المعدل	155	15
لم ينفذوا المعدل	30	70

والمطلوب : هل من علاقة ارتباطية بين التأهيل وعدد المكتبيين ؟

الفصل السادس

العينات وتوزيع المعاينة

6-1- أهمية العينات ومجالات تطبيقها

6-2-أنواع العينات

6-2-1- العينات العشوائية (شروطها وأنواعها)

6-2-2- العينات غير العشوائية وأنواعها

6-3- توزيعات المعاينة وخصائصها

6-3-1-ماهية توزيع المعاينة

6-3-2- خصائص توزيع العينة

- توزيع معاينة الأوساط الحسابية
- توزيع معاينة الانحرافات المعيارية
- توزيع معاينة النسب المئوية
- توزيعات معاينة الفروق

6-4- تحديد حجم العينة

6-4-1- تحديد حجم العينة للمتغيرات الكمية

6-4-2- تحديد حجم العينة للمتغيرات النوعية

6-5- مسائل وتمارين وغير محلولة

وقل رب زدني علما
جامعة دمشق
Damascus University

1-6 - أهمية العينات ومجالات تطبيقها

إننا نعيش في عصرِ يصعب الحكم فيه على صحة الأشياء بدون أدلة موضوعية والبحث الواسع ، عصرٌ مليء بالمشكلات العلمية المختلفة والظواهر المتعددة.
ولدراسة أية مشكلة علمية نحتاج إلى جمع كل ما يتعلق بتلك المشكلة من معلومات ، تسمى مجموعة العناصر المتعلقة بتلك المشكلة بالمجتمع الإحصائي ، وهو مفهوم مرن يمكن أن يتلاءم مع أية دراسة إحصائية ، أي بمعنى توسيع إطاره الزماني والمكاني أو تضيقه تحت اعتبارات منها: الهدف من الدراسة والوقت اللازم لإنهاء تلك الدراسة والإمكانات المادية والبشرية المتاحة.
من هنا تبدأ أية دراسة إحصائية بالبيانات الخام المتوفرة عنها ، والتي يتم جمعها بإحدى الطريقتين الآتيتين :
1- طريقة الحصر الشامل أو ما يسمى بالتعداد العام ، حيث يتم جمع المعلومات عن كل مفردة من مفردات المجتمع المدروس ومن ثم إجراء التحليل المنهجي اللازم : كتعداد السكان العام ، المسح الصناعي ، المسح الزراعي ، وغيرها.
2- طريقة العينة ، حيث يتم جمع المعلومات عن عدد محدود من مفردات المجتمع شريطة تمثيلها له ، ومن ثم يتم تعميم نتائج هذه المفردات على كل المجتمع.
في الواقع العملي وفي مجال البحث العلمي تقوم معارفنا وقراراتنا على أساس العينات ، التي احتل استخدامها مكاناً هاماً سواءً في الماضي البعيد أم القريب ، وخاصة بعد أن ارتبطت مع بداية القرن العشرين للميلاد بتطور الرياضيات ونظرية الاحتمالات وتطور العلوم الأخرى ، الأمر الذي ساعد في صياغة الأبحاث المستندة إليها في أشكال رياضية محددة وذات قيم ة عملية، بحيث أمكن أن يطلق عليها اسم نظرية العينات ، وهكذا تطورت نظرية العينات منذ الحرب الكونية الثانية ، وتزايد استعمالها كوسيلة للحصول على البيانات الإحصائية ، وتراكمت نظريات كافية تستند إليها طرائق المعاينة واتسعت تطبيقاتها في مجالات عدة، لاسيما في ميادين العلوم الاقتصادية والاجتماعية والإدارية وعلم النفس والزراعة وغيرها، والأمثلة على ذلك كثيرة ، بحيث لم يَعد من الضروري التأكيد على أهمية دور العينات كمصدر رئيس للإحصاءات الرسمية.
تعتمد الكثير من الأبحاث على العينة كوسيلة للحصول على بياناتها الإحصائية ، ومن هذه الأبحاث على سبيل المثال لا الحصر نذكر:
- بحث ميزانية الأسرة ، لمعرفة الأرقام القياسية لتكاليف المعيشة وغيرها من الأرقام القياسية .
- جمع البيانات عن أسعار السلع والخدمات.

- مراقبة الإنتاج لمعرفة نسبة التالف في إنتاج آلة معينة لأجل تحسين ظروف الإنتاج الداخلية في المصنع، ولكي تكون الوحدات المنتجة مطابقة للمواصفات الفنية المطلوبة.
- البحوث الحيوية والعلمية.
- في أبحاث التسويق لمعرفة ميل المستهلكين ورغبات هم وأذواقهم نحو نوع المنتج ومواصفاته.

يستعمل الباحث في المسوحات المتخصصة بعض تقنيات القياس الخاصة بالموضوع والتي يتعذر تطبيقها على كافة عناصر المجتمع مثل : مسوحات التغذية أو صحة العائلة وبعض المسوحات الزراعية التي تستخدم تقنيات حديثة لقياس المردود الزراعي في الهكتار، وتجرى هذه المسوحات بالعينة. وكثيراً ما نلجأ إلى طريقة العينة لدراسة بعض المجتمعات مثل تقدير إنتاج المنشآت الصغيرة التي لا يشملها التعداد الصناعي أو تقدير بعض المعطيات المتعلقة بالتعداد السكاني ومن نوع خاص.

وثمة مجال آخر لاستخدام العينات يتجلى في تقييم نوعية نتائج التعداد الشامل والمسوحات الكبيرة ، وتسمى هذه العملية بالتدقيق الآجل للتعداد، وبناءً عليه نجد أنَّ للعينات مجالات تطبيقية عديدة وأنَّ استعمالها أمكننا التعرف على دقائق جملة من الموضوعات وتحليلها.

ونذكر هنا بعض الحالات التي يفضل بها استخدام طريقة العينة على طريقة الحصر الشامل :

أ - تلف عناصر المجتمع وفقدانهم نتيجة أخذ مشاهدات من تلك العناصر(مدى صلاحية إنتاج مزرعة ما من البيض المعد للأكل) ؛

ب - تعذر الوصول إلى جميع عناصر المجتمع وإجراء الدراسة عليها (كمية الأمطار الهاطلة منذ عام 1950 وحتى الآن وأثرها في كميات إنتاج محصول القمح في سوريا خلال نفس الفترة الزمنية) ؛

ت - تقييد أغلب الدراسات الإحصائية بمقدار من التكاليف والوقت والجهد المخصص لإنهائها؛

ث - يتطلب إجراء المسح الشامل عدداً كبيراً من الأشخاص لجمع البيانات، الأمر الذي يؤدي إلى أخطاء متعددة لأسباب منها : المهارة والخبرة الفردية بين العاملين في جمع البيانات الإحصائية ؛

ج - الحاجة الملحة إلى نتائج سريعة لاتخاذ قرار معين (شركة دوائية) ؛

ح - طبيعة المجتمع الإحصائي من حيث كونه :

- مستمراً ، حيث تكون عناصر المجتمع غير قابلة للعد (كمخزون النفط أو الفوسفات في بلد ما) ،

- متقطعاً ، أي أنَّ عناصر المجتمع قابلة للعد ومهما كبرت(عدد أفراد المجتمع الصيني) .

وبناءً على ما سبق نلخص أوجه المقارنة بين مزايا كل من طريقتي الحصول على البيانات الإحصائية ومساوئهما في الجدول الآتي :

الطريقة / أوجه المقارنة	الحصر الشامل	العينة
المزايا	- إذا كان الإطار سليماً، فإننا نحصل على نتائج أكثر دقة؛ - لا مجال لوجود أخطاء عشوائية، نتيج شمول الدراسة لجميع مفردات المجتمع	- تؤدي إلى تخفيض التكاليف والوقت اللازمين لذلك؛ - إمكانية تطبيقها مهما كان نوع وطبيعة المجتمع .
المساوئ	- تحتاج إلى تكاليف مادية وبشرية كبيرة ووقت أطول وجهد أكبر ؛ - لا يمكن تطبيقها ميدانياً، وخاصة في المجتمعات المستمرة والمتقطعة ؛ - تتعرض لأخطاء التحيّز لدرجة كبيرة بسبب كبر حجمها .	- تعطي نتائج جزئية أو تقريبية ؛ - وجود أخطاء عشوائية ؛ - تعترض لأخطاء التحيّز بدرجة قليلة بسبب صغر حجمها .

يتحتم على الباحث أثناء دراسته الإحصائية تحديد عما إذا كان سيختار بين إحدى الطريقتين المذكورتين سابقاً في جمع بياناته، وهنا يجب عليه المفاضلة بين الطريقتين وفقاً لـ :

- طبيعة المجتمع الإحصائي ؛
- طبيعة الظاهرة المدروسة ؛
- طبيعة البيانات ؛
- الإمكانات المتاحة للباحث .

غير أنه وبشكل عام ، يكون أسلوب العينة هو المفضل في أغلب الأحيان لاعتبارات عدة ، منها :

1- الحصول على نتائج دقيقة بتكاليف أقل وبسرعة أكبر ؛
2- عدم توفر مجتمع الدراسة بالكامل غالب الأحيان ؛
3- عدم إمكانية توفر الكوادر البشرية المؤهلة والتجهيزات الضرورية للقيام بالحصر الشامل .

وثّمة مشكلة تواجه الباحث عندما يقرر لأخذ بأسلوب العينة ، ألاَّ وهي تحديد حجم العينة المراد سحبها من المجتمع الإحصائي الممثلة له ونرمز لحجم العينة بـ (n) ، وهذا يتوقف على عوامل عدة أهمها :

- مقدار الأموال المخصصة لجمع البيانات المطلوبة ؛
- مقدار الثّقة التي يرغب الباحث توافرها في البيانات التي يقوم بجمعها ؛
- الكوادر البشرية المؤهلة والمتاحة للقيام بجمع المعلومات المرغوب فيها .

وعليه يمكننا تعريف العينة بأنها : " مجموعة من مفردات المجتمع الإحصائي يتم جمعها بحيث تكون ممثلة تمثيلاً صحيحاً له، بقصد دراسة ظاهرة معينة على هذه المجموعة من المفردات للوصول إلى نتائج يمكن تعميمها على ذلك المجتمع الذي أُخذت منه عند احتمال ما ". وإن أخذ عينة ممثلة لمجتمعها الإحصائي يتطلب توفر أمرين هما :

- تحديد الهدف من الدراسة الإحصائية (عاماً كان أم خاصاً) ؛
- تحديد المجتمع الإحصائي ونرمز له بـ (N) وهو عبارة عن جميع الوحدات الإحصائية المؤلفة له، بحيث أنَّ الوحدة الإحصائية هي أصغر جزء مستقل تجري عليه الدراسة الإحصائية وهي ثابتة ومستقلة وتشترك بصفة أو أكثر.

وتجدر بنا الإشارة إلى " إطار العينة " الذي هو عبارة عن مجموعة الطرائق والأساليب التي يمكن بواسطتها التعرف على جميع وحدات المجتمع الإحصائي ، وقد يكون قائمة أو خريطة أو رسم بياني يوضح جميع مفردات المجتمع موضوع الدراسة بحيث يؤدي إلى :

- شمول الاختيار لكل مفردات المجتمع الإحصائي ؛
- استخدام طريقة تؤدي إلى اختيار عينة ممثلة للمجتمع الإحصائي ؛
- اختيار حجم العينة الذي يؤدي إلى الحصول على تقديرات مقبولة ؛
- المعرفة التامة بدرجة تجانس المجتمع الإحصائي من حيث الظاهرة موضوع الدراسة .

كما لا بد لنا هنا من ذكر أهم الخطوات الأساسية لتصميم العينة ، ومنها :

1 - تحديد المشكلة المراد دراستها والهدف منها ؛
2 - تعريف المجتمع الإحصائي المراد أخذ عينة الدراسة منه وتحديده ، والتأكد من شرط تجانس مفرداته ، بحيث لا يجوز استبدال المجتمع المعروف بآخر مماثل له أو تحويل إجراء الدراسة من فترة زمنية لأخرى؛
3 - تحديد المعلومات المطلوبة بما يتناسب وهدف المشكلة موضوع الدراسة ؛
4 - تحديد جمع المعلومات سواءً عن طريق المقابلة الشخصية أ م الاستمارة البريدية أم التلفون أم الانترنت وغيرها ؛

5 - تحديد الإطار لأن هذا يمكننا من حصر جميع وحدات المجتمع وتغطي تها وبالتالي اختيار العينة المطلوبة ؛
6 - تحديد حجم العينة وتكاليفها اعتماداً على الدقة المطلوبة وعلى تباين المجتمع ؛
7 - إجراء اختبار سابق لتقييم بعض الجوانب الغامضة في أسئلة الاستمارة الإحصائية الموضوعة عن المشكلة المدروسة، لأن هذا يفيد في تذليل بعض الصعوبات التي تواجه الباحث عند إجراء الدراسة وخاصة تحديد حجم العينة ؛
8 - إجراء مراجعة كاملة لكافة المعلومات التي حصل عليها وتصحيح الأخطاء أو حذفها ، بالإضافة إلى تبويب تلك المعلومات للقيام بدراستها وتحليلها ؛
9 - تقدير معالم المجتمع الإحصائي بواسطة العينة ، تلك المعالم المتأثرة بعدة عوامل كتجانس المجتمع أو احتمال سحب وحدات العينة أو حجم العينة المسحوبة .

6-2- أنواع العينات

تصنف العينات حسب أسلوب سحبها من المجتمع الإحصائي إلى عينات عشوائية (احتمالية أو علمية) وأخرى غير عشوائية (شخصية) ، إذ تحتل العينات العشوائية أهمية كبيرة نظراً لاعتمادها على نظرية الاحتمالات وإمكانية استخدامها في اختبار الفرضيات والتعميم من العينة على مجتمعها الإحصائي وحساب مقدار الأخطاء الاحتمالية التي تقل مع ازدياد حجم العينة وذلك بعكس أخطاء التحيزّ فهي ثابتة بالرغم من ازدياد حجم العينة (سنتطرق لذلك لاحقاً) ، وسوف نوجز ما يتعلق بكل من العينات العشوائية وغير العشوائية .

6-2-1- العينات العشوائية (شروطها وأنواعها)

تعد العينات العشوائية بمختلف أنواعها الأسلوب الوحيد الذي يعطي مقاييس كميّة لقياس الدقة أو الثقة ويسمح بتقويم مقدار الأخطاء الاحتمالية أو أخطاء الحظ والصدف. وتعّرف العينة العشوائية بأنها : " العينة التي اختيرت باستخدام أساليب السحب العلمية في الاختيار العشوائي كالسحب بالقرعة أو بأسلوب الدواليب أو كرات اليانصيب أو استعمال جداول الأرقام العشوائية ، أي بمعنى إعطاء كل مفردات المجتمع الإحصائي الفرصة نفسها ليتم اختيارها " .
ولكي تتصف العينات بالعشوائية يجب مراعاة الشروط الآتية :

1 - أن تكون مفردات المجتمع الإحصائي المراد أخذ عينة منه متجانسة من حيث طبيعتها ونوعية الدراسة المرغوب إجراؤها على هذه المفردات ؛
2- أن تكون مفردات المجتمع الإحصائي مستقلة بعضها عن بعض ، أي أن انتقاء أي مفردة من مفردات المجتمع لا يرتبط بسحب أو عدم سحب أية مفردة أخرى ؛
3- أن يكون احتمال انتقاء أية مفردة من مفردات المجتمع الإحصائي معروفاً أو يمكن حسابه ؛
4- أن يتم انتقاء مفردات العينة من المجتمع الإحصائي المعن ي بدون تحيزّ ، أي أن يتصف الانتقاء بالعشوائية .

وعليه نميزّ عدة أنواع للعينات العشوائية ، منها :

- العينة العشوائية البسيطة ؛
- العينة العشوائية الطبقية ؛
- العينة العشوائية المنتظمة ؛
- العينة العشوائية العنقودية ؛
- العينة العشوائية المتعددة المراحل .

6-2-2- العينات غير العشوائية

ويطلق على هذا النوع من العينات أحياناً بالعينات الشخصية وهي عبارة عن تلك العينات التي يتم الحصول على بياناتها بشكل غير عشوائي ، أي التي لا يمكن تطبيق نظرية الاحتمال وأصول الاستدلال الإحصائي عليها ، وإن تطبيقها ضمن تحفظات ما يضمن كي تكون نتائج التحليل صحيحة وغير مضّللة ، وإلاَّ كان العكس من ذلك ، وهذا يعطيها أهمية وخاصة في كونها تقدم مقاييس إحصائية وصفية ذات دور هام في توفير خلاصات مفيدة وقيّمة للباحث بالإضافة لاستعمالها كأدوات لا غنى عنها عند إجراء محاكمة عقلانية ، ومع ذلك وجب الحذر عند استعمالها وتعميم نتائجها ، ومن أهم أنواعها نذكر :

- عينات الحصص أو الكوتا ؛
- العينات المنتقاة ؛
- العينات الكبير ة الحجم . (لقد شرحت أنواع العينات العشوائية وغير العشوائية بالتفصيل في الفصل الثاني من مفردات هذا الكتاب).

6-3- توزيعات المعاينة وخصائصها

6-3-1- ماهية توزيع المعاينة

كما نعلم تصنف البيانات الإحصائية حسب طبيعتها إلى بيانات نوعية (نوعيات) وأخرى بيانات كمّية (متغيرات) ، حيث إنَّ " النوعيات " هي تلك الميزات التي يتصف بها الشيء أو لا يتصف بها، ولكن ليس بدرجات متغيرة ومقاسة ، أي يمكن وصف ميزات المجتمع الإحصائي بشكل نسبة مئوية للوحدات التي تتصف بها إحدى الصفات النوعية (كنسبة الأمية في مجتمع ما عام 2000) ، في أنَّ " المتغيرات " هي تلك القيم العددية لتلك الميزات المشتركة التي تتصف بها مجموعة من الأشياء بدرجات مختلفة ومقاسة (كأجور العمال الشهرية لإحدى المؤسسات في عام ما) .

وأيّ كان نوع البيانات الإحصائية المتوفرة (نوعية أم كميّة) ، فإن لها ثلاثة توزيعات تكرارية ، فمعرفة هذه التوزيعات كل على حدة ، وفهم العلاقات القائمة فيما بينها ضروري لمسائل الاستدلال الإحصائي (اختبار الفرضيات أو التقدير الإحصائي) ، وهذه التوزيعات هي :

1 - **توزيع المجتمع الإحصائي** : يتميز توزيع كل مجتمع إحصائي بمقاييس إحصائية وصفية تدل على خصائصه، مثل : الوسط الحسابي، الوسيط ، الانحراف المعياري ،النسبة المئوية لتكرار الصفة... الخ تدعى هذه المقاييس الإحصائية الوصفية المحسوبة بدلالة المجتمع الإحصائي بالثوابت الإحصائية، أي كل مقياس منها هو ثابت إحصائي ، ويخضع هذا التوزيع لشكل توزيع المنحنى الطبيعي ، فمثلاً لو كان لدينا بيانات إحصائية عن أوزان عمال صناعة ما في سورية عام 2000 أو أطوالهم أو دخولهم الشهرية ... الخ فإن مفردات كل مجتمع من هذه المجتمعات تشكل توزيعاً تكرارياً يطلق عليه بتوزيع المجتمع الإحصائي .

2 - **توزيع العينة** : بفرض أننا قمنا بسحب عينة عشوائية حجمها (n) من مجتمع إحصائي ما ، فإن توزيعها يشبه توزيع مجتمعها الإحصائي الذي أخذت منه، لأنها قد أخذت منه بشكل عشوائي ، أما المقاييس الإحصائية الوصفية المحسوبة بدلالتها كالوسط الحسابي والوسيط والمنوال والانحراف المعياري والنسبة المئوية لتكرار الصفة في العينة و... التي يطلق عليها بالتوابع الإحصائية ، التي تتوافق تقريباً مع الثوابت الإحصائية المقابلة لها ، لأن توزيع العينة يشبه توزيع المجتمع الإحصائي الذي سحبت منه هذه العينة ، لكن قد تختلف التوابع الإحصائية عن الثوابت الإحصائية المقابلة لها أو عن توابع إحصائية أخرى، وهذا الاختلاف قد يكون:

a– اختلافاً ظاهرياً ناشئاً عن الأخطاء الاحتمالية أو أخطاء الحظ والصدف ؛

b– اختلاف حقيقي (جوهري) ينشأ عن عدم سحب العينة بشكل عشوائي أو أنها قد سحبت من مجتمع إحصائي يختلف عن المجتمع الإحصائي المرغوب فيه .

3- توزيع المعاينة : إذا كان لدينا مجتمع إحصائي مؤلف من (N) وحدة ، فإننا نستطيع أن نسحب منه عدداً من العينات العشوائية بحجم (n) يساوي $C_N^n = \frac{N!}{n!(N-n)!}$ ، فإن لكل عينة من هذه العينات وسطها الحسابي ووسيطها وانحرافها المعياري و... وتدعى هذه المقاييس الإحصائية الوصفية أيضاً بالتوابع الإحصائية ، وإذا حسبنا لكل عينة من هذه العينات وسطها الحسابي فإنه سيكون لدينا C_N^n وسط حسابي (عدد من الأوساط الحسابية المطابق لعدد العينات التي لدينا) ، ونظراً لأن هذه الأوساط قد تختلف عن بعضها بعضاً بسبب الأخطاء الاحتمالية ، فإنها تشكل فيما بينها توزيعاً تكرارياً يطلق عليه بتوزيع العينة للأوساط الحسابية أو توزيع معاينة الأوساط الحسابية ، وبشكل مشابه نحصل على توزيعات المعاينة للانحرافات المعيارية، للوسيط ،للنسب المئوية وغيرها .

يستعمل المنحنى الطبيعي كتوزيع معاينة نظري في معالجة مسائل الاستدلال الإحصائي وبغض النظر عن التابع الإحصائي موضوع الدراسة ، سواءً أكان وسطاً حسابياً أم انحرافاً معيارياً أم نسباً مئوية ، شرط أن يكون حجم العينة $(n \geq 30)$ ،لأنه وحسب نظرية الحد المركزي للإحصاء الرياضي: " أنه كلما ازداد حجم العينة وأصبح أكثر من 30 مفردة ، فإن شكل توزيعها يقترب من شكل توزيع المنحنى الطبيعي " ، غير أنَّ توزيع معاينة النسب المئوية يخضع لتوزيع ثنائي الحدين الذي يقترب من توزيع المنحنى الطبيعي عندما يصبح حجم العينة أكبر من 100 مفردة .

وبناءً على ما سبق ، نستطيع أنَّ نُعرّف توزيع المعاينة : " بأنه التوزيع التكراري لأحد المقاييس الإحصائية الوصفية المحسوب من عدد كبير من العينات العشوائية متساوية الحجم من مجتمع إحصائي واحد ، وبشكل عام ، فإن توزيع المعاين يصف الكيفية التي تختلف فيها التوابع الإحصائية عن الثوابت الإحصائية المقابلة لها عندما تحدد قوى الحظ وحدها الوحدات التي سوف تتضمنها العينة من مجموع وحدات المجتمع الإحصائي .

6-3-2- خصائص توزيعات العينات

1 - توزيع معاينة الأوساط الحسابية : إذا كان عدد العينات العشوائية المسحوبة من مجتمع إحصائي حجمه (N) كبيراً جداً (حالة سحب مع الإعادة) ، فإنه يمكننا القول : بأنَّ الوسط الحسابي لتوزيع معاينة أي تابع إحصائي يتفق تماماً مع الثابت الإحصائي المقابل له . وبناءً على ذلك ، فالوسط الحسابي لتوزيع معاينة الأوساط الحسابية يساوي الوسط الحسابي للمجتمع الإحصائي ، والوسط الحسابي لتوزيع معاينة الانحرافات المعيارية يساوي بالضبط الانحراف المعياري للمجتمع الإحصائي ... الخ ، ودوماً عندما يكون حجم العينة $(n \geq 30)$. هذا ويمكننا التعبير عن ذلك رياضياً :

$$\overline{X} = E(\overline{x}_i) = \frac{\overline{x}_1 + \overline{x}_2 + ... + \overline{x}_n}{n} = \frac{\sum_{i=1}^{n} \overline{x}_i}{n}$$

حيث $\overline{X}$ يمثل الوسط الحسابي للمجتمع الإحصائي ، و $\overline{x}_i$ متوسطات العينات الممكنة ، و $E(\overline{x}_n)$ يمثل الأمل الرياضي للوسط الحسابي لأوساط العينات ، الذي يساوي متوسط مجموع الأوساط الحسابية لهذه العينات وهو يساوي الوسط الحسابي للمجتمع .

أما الانحراف المعياري للمجتمع الإحصائي يساوي لمتوسط مجموع الانحرافات المعيارية لهذه العينات العشوائية المسحوبة منه ، أي :

$$\sigma_x = \frac{s_1 + s_2 + ... + s_n}{n} = \frac{\sum_{i=1}^{n} s_i}{n}$$

وتمثل σ_x الانحراف المعياري للمجتمع الإحصائي .

s_i الانحراف المعياري للعينة i حيث $(i = 1,2,3,...,n)$.

ويطلق على الانحراف المعياري للأوساط الحسابية للعينات العشوائية المتساوية الحجم بالخطأ المعياري للوسط الحسابي ، ويساوي : $\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}}$ حيث أنَّ :

σ_x الانحراف المعياري للمجتمع .

$\sigma_{\overline{x}}$ الخطأ المعياري للوسط الحسابي .

n حجم العينة (عدد مفردات العينة) .

وإذا لم يكن الانحراف المعياري للمجتمع معروفاً ، فإنه يمكننا تقدير الخطأ المعياري للوسط الحسابي بالاعتماد على الانحراف المعياري للعينة ، أي :

$$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{n-1}}$$

وتمثل s_x الانحراف المعياري للعينة .

أما إذا كان حجم المجتمع الإحصائي (N) صغيراً (حالة السحب دون إعادة) ، فإنه وجب تعديل الخطأ المعياري بالكمية $\sqrt{\frac{N-n}{N-1}}$ ، فيصبح لدينا :

- $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

- $$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{n-1}} \cdot \sqrt{\frac{N-n}{N-1}}$$

مثال (6- 1) : إذا كان لدينا مجتمع يمثل أعمار خمسة أطفال ، وكانت أعمارهم على التوالي :

2 ، 3 ، 6 ، 8 ، 11 ، وقمنا بسحب جميع العينات العشوائية الممكنة من حجم طفلين من هذا المجتمع ، والمطلوب :

1 - أوجد الوسط الحسابي للمجتمع وانحرافه المعياري ؟

2 - أوجد الوسط الحسابي لتوزيع أوساط العينات والخطأ المعياري لها ؟

الحل :

1 - الوسط الحسابي للمجتمع $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{2+3+6+8+11}{5} = 6$

- الانحراف المعياري للمجتمع $\sigma_x = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}}$

$$= \sqrt{\frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5}} = 3.29$$

2 - نرغب في سحب كافة العينات من حجم $(n = 2)$ ، فنكون أمام حالتين :

أ - عدد العينات المسحوبة مساوياً لعدد المرتبات مع التكرار :

$N^n = 5^2 = 25$ وهي كالآتي :

(2 ، 2) ، (2 ، 3) ، (2 ، 6) ، (2 ، 8) ، (2 ، 11) ، (3 ، 2) ، (3 ، 3) ، (3 ، 6) ،
(3 ، 8) ، (3 ، 11) ، (6 ، 2) ، (6 ، 3) ، (6 ، 6) ، (6 ، 8) ، (6 ، 11) ، (8 ، 2) ،
(8 ، 3) ، (8 ، 6) ، (8 ، 8) ، (8 ، 11) ، (2، 11) ، (11 ، 3) ، (11 ، 6) ، (11 ، 8) ،
(11 ، 11) .

وتكون الأوساط الحسابية لتلك العينات على التوالي :

(2) ، (2,5) ، (4) ، (5) ، (6,5) ، (2,5) ، (3) ، (4,5) ، (5,5) ، (7) ، (4) ،
(4,5) ، (6) ، (7) ، (8,5) ، (5) ، (5,5) ، (7) ، (8) ، (9,5) ، (6,5) ، (7) ،
(8,5) ، (9,5) ، (11) .

تتوزع هذه الأوساط الحسابية للعينات المحسوبة من المجتمع توزيعاً طبيعياً ، وعليه يكون الوسط الحسابي للأوساط الحسابية :

$$\overline{X} = \frac{\sum_{i=1}^{N^n} \overline{x}_i}{N^n} = \frac{150}{25} = 6$$

وهذه النتيجة تؤكد بأنَّ الوسط الحسابي للمجتمع هو الأمل الرياضي للأوساط الحسابية للعينات ، أي بمعنى أنَّ الوسط الحسابي للمجتمع يساوي الوسط الحسابي لتوزيع معاينة الأوساط الحسابية ، أما الخطأ المعياري للأوساط الحسابية فيساوي :

$$\sigma_{\bar{x}} = \sqrt{\frac{(2-6)^2 + (2{,}5-6)^2 + \ldots + (11-6)^2}{25}} = 2{,}32$$

وإن هذه القيمة توافق تماماً تلك التي نحصل عليها من خلال تطبيق العلاقة :

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{3{,}29}{\sqrt{2}} = 2{,}32$$

ب - عدد العينات المسحوبة مساوياً لعدد المتوافقات دون تكرار ، أي :

$$C_N^n = \frac{N!}{n!(N-n)!} = \frac{5!}{2!(5-2)!} = 10$$

فالعينات هي :

(2 ، 3) ، (2 ، 6) ، (2 ، 8) ، (2 ، 11) ، (3 ، 6) ، (3 ، 8) ، (3 ، 11) ، (6 ، 8) ، (6 ، 11) ، (8 ، 11) .

وأوساطها على التوالي :

(2,5) ، (4) ، (5) ، (6,5) ، (4,5) ، (5,5) ، (7) ، (7) ، (8,5) ، (9,5).

ومنه الوسط الحسابي للأوساط الحسابية :

$$\overline{X} = \frac{\sum_{i=1}^{C_N^n} \overline{x}_i}{C_N^n} = \frac{2,5+4+5+...+9,5}{C_5^2} = \frac{60}{10} = 6$$

أما الخطأ المعياري لتوزيع الأوساط الحسابية فيحسب من العلاقة :

$$\sigma_{\overline{x}} = \sqrt{\frac{\sum_{i=1}^{n}(\overline{x}_i - \overline{x})^2}{n}} = \sqrt{\frac{(2,5-6)^2 + (4-6)^2 + ... + (9,5-6)^2}{10}} = 2,01$$

وهذا يوافق تطبيق العلاقة الآتية :

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{3,29}{\sqrt{2}} \cdot \sqrt{\frac{5-2}{5-1}} = 2,01$$

3- توزيع معاينة الانحرافات المعيارية : عندما نسحب عدداً كبيراً من العينات العشوائية المتساوية الحجم من مجتمع إحصائي ما ، نحصل من كل منها على قيمة للانحراف المعياري ، فتشكل هذه القيم من الانحرافات المعيارية لتلك العينات فيما بينها توزيعاً تكرارياً يطلق عليه بتوزيع معاينة الانحرافات المعيارية ، وهذا التوزيع يقترب من التوزيع الطبيعي مع تزايد حجم العينة .

ويطلق على الانحراف المعياري للتوزيع التكراري لعدد كبير جداً من الانحرافات المعيارية لعينات عشوائية من حجم واحد ومن مجتمع إحصائي واحد بالخطأ المعياري للانحراف المعيارية ويرمز له بـ σ_σ ، والذي يحسب من العلاقات الآتية:

- حال كون الانحراف للمجتمع معلوماً : $\sigma_\sigma = \frac{\sigma_x}{\sqrt{2n}}$

ولمثالنا السابق يكون : $\sigma_\sigma = \frac{3,29}{\sqrt{2*2}} = 1,645$

- حال كون الانحراف المعياري للمجتمع غير معلوم ، يستعاض عنه بدلالة الانحراف المعياري للعينة ، لأنه يمكن اعتبار قيمة الانحراف المعياري للعينة العشوائية كتقدير مناسب لقيمة الانحراف المعياري للمجتمع الإحصائي الذي أخذت منه على اعتبار أنَّ حجمها كبي رٌ ، لأنه وحسب قانون الأعداد الكبيرة الذي ينص على أنه : " كلما ازداد حجم العينة ، فإن الاحتمال يقترب من اليقين ، وأنَّ الفرق بين التابع الإحصائي والثابت الإحصائي المقابل له يصبح أصغر من أية قيمة مهما صغرت " ، أي :

$$\sigma_\sigma = \frac{s_x}{\sqrt{2n}}$$

- أما إذا كان حجم العينة صغيراً ، فإننا نضرب قيمة الانحراف المعياري للعينة بالعامل المصحح $\sqrt{\frac{n}{n-1}}$ ، ليكون التقدير للانحراف المعياري للمجتمع الإحصائي بدلالة العينة العشوائية المأخوذة منه تقدير غير متحيزّ .

وتجدر بنا الإشارة هنا إلى بعض الملاحظات حول التقدير المتحيزّ ، ومنها :

- يعتبر متوسط توزيع معاينة الأوساط الحسابية مساوياً لمتوسط المجتمع ، أي أنَّ متوسط العينة هو تقدير غير متحيزّ لمتوسط المجتمع ، عندما تكون تلك العينة قد أخذت منه بشكل عشوائي ؛
- متوسط توزيع معاينة التباينات $S_x^2.\frac{n}{n-1}$ يساوي تباين المجتمع σ_x^2 ، وهو تقدير غير متحيزّ لتباين المجتمع ، عندما يكون حجم العينة العشوائية أقل من ثلاثين مفردة $(n \prec 30)$ ، أما إذا حجم تلك العينة أكبر أو يساوي ثلاثين مفردة $(n \geq 30)$ فإن تباينها S_x^2 هو تقدير غير متحيزّ لتباين المجتمع σ_x^2 ؛
- إذا كان توزيع المعاينة لعينتين لهما الوسط الحسابي نفسه ، فإن العينة ذات التباين الأقل تسمى مقدراً غير متحيزّاً للوسط الحسابي ، في حين أنَّ العينة الأخرى هي مقدر متحيزّ .

3 - **توزيع معاينة النسب المئوية** : شرحنا سابقاً بأنَّ البيانات الإحصائية تصنف حسب طبيعتها إلى بيانات كميّة

(متغيرات) وأخرى نوعية (نوعيات) ، والنوعيات هي عبارة عن تلك الميزات التي قد يتصف بها الشيء أو لا ، بحيث يمكن التعبير عنها بنسب مئوية ، أي بمعنى أنه يمكن وصف ميزات المجتمع الإحصائي بشكل نسب مئوية للوحدات التي تتصف بها إحدى الصفات النوعية ، كأن نقول بأن 20% من سكان بلد ما هم عاطلون عن العمل ، فهذه النسبة تصف المجتمع الإحصائي بهذه الصفة (العطل عن العمل) .

وبشكل عام ، فإننا نرمز للنسبة المئوية لتكرار حدوث الصفة في المجتمع الإحصائي بـ (p) وللنسبة المئوية لتكرار عدم حدوث الصفة فيه بـ $[q = (1 - p)]$ ، وإذا سحبنا عدداً كبيراً من العينات العشوائية متساوية الحجم من هذا المجتمع الإحصائي ، فسيكون لدينا توزيع للمعاينة من قيم النسب المئوية (P') ، وإن هذا التوزيع – كما سبق وأشرنا – يخضع لتوزيع ثنائي الحدين وسطاؤه (الوسط الحسابي = np ، والانحراف المعياري $\sigma_x = \sqrt{npq}$) ، لكن وعندما يكون حجم العينة كبيراً $(n \geq 100)$ فإن شكل التوزيع يقترب من توزيع المنحنى الطبيعي ، وعليه يكون في حال كون المجتمع محدداً والسحب مع الإعادة :

1 - الوسط الحسابي لتوزيع معاينة القيم (p') مساوي للقيمة (p) ؛

2 - الخطأ المعياري لتوزيع معاينة النسب المئوية معطى بالآتي :

- إذا عرفت قيمة (p) : $\sigma_p = \sqrt{\frac{p.q}{n}} = \sqrt{\frac{p(1-P)}{q}}$

حيث تمثل σ_p الخطأ المعياري للنسب المئوية .

-إذا لم تعرف قيمة (p) يمكن اعتبار قيمة (p') المحسوبة من إحدى العينات كتقدير لقيمة (p) استناداً لقانون الأعداد الكبيرة ، أي :

$$\sigma_p = \sqrt{\frac{p'.q'}{n}} = \sqrt{\frac{p'(1-p')}{n}}$$

أما إذا كان المجتمع الإحصائي محدوداً والسحب بدون إعادة ، يصبح :

- قيمة (p) معلومة : $\sigma_p = \sqrt{\frac{p.q}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

- قيمة (p) غير معلومة : $\sigma_p = \sqrt{\frac{p'.q'}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

4 - **توزيع معاينة الفروق** : لنفرض أن لدينا مجتمعاً إحصائياً حجمه N ، وسحبنا منه كافة العينات العشوائية من حجم ما، فإننا نحصل على توزيع لأوساط هذه العينات ولنرمز لها بـ :

$$\overline{X}_{11}, \overline{X}_{12}, \overline{X}_{13}, ..., \overline{X}_{1n}$$

وقمنا ثانية ، بسحب كافة العينات العشوائية الممكنة منه ولكن بحجم آخر ، فإننا نحصل على توزيع آخر لأوساط هذه العينات الجديدة ، ونرمز لها بـ :

$$\overline{X}_{21}, \overline{X}_{22}, \overline{X}_{23}, ..., \overline{X}_{2n}$$

وإذا قمنا الآن بطرح توزيع أوساط العينات الأخيرة من توزيع أوساط العينات الأولى ، حصلنا على توزيع جديد يمثل توزيع معاينة فروق الأوساط الحسابية ، وهو توزيع طبيعي ، يتصف بالآتي :

- $\overline{X}_{1-2} = \overline{X}_1 - \overline{X}_2$

أي أنَّ الوسط الحسابي لتوزيع الفروق يساوي الفرق بين الوسطين الحسابيين لتوزيع العينتين ؛

- أما الخطأ المعياري لتوزيع الفروق بين الوسطين الحسابين $\sigma_{\overline{x}_1-\overline{x}2}$ فيساوي :
  - حال كانت ا لانحرافات المعيارية للمجتمع معلومة :

$$\sigma_{\overline{x}_1-\overline{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

  - حال كانت الانحرافات المعيارية للمجتمع غير معلومة ، يستعاض عنها بالانحرافات المعيارية للعينات :

$$\sigma_{\overline{x}_1-\overline{x}2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- الخطأ المعياري لتوزيع الفروق بين الانحرافين المعياريين فيساوي :
  - الانحرافات المعيارية للمجتمع معلومة :

$$\sigma_{(\sigma_1-\sigma 2)} = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

  - الانحرافات المعيارية للمجتمع غير معلومة :

$$\sigma_{(\sigma_1-\sigma 2)} = \sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}$$

- ويمكن الحصول على نتائج مقابلة لتوزيع معاينة الفروق بين النسب ، فيكون :

- $p_{1-2} = p'_1 - p'_2$
- p معلومة $\sigma_{p_1-p2} = \sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}$
- p غير معلومة $\sigma_{p_1-p2} = \sqrt{\frac{p'_1.q'_1}{n_1} + \frac{p'_2.q'_2}{n_2}}$

ونشير هنا ، إلى أنَّ جميع القيم التي حصلنا عليها ستفيدنا كثيراً في فهم ما يعالجه الاستدلال الإحصائي سواءً ما يتعلق بالتقدير الإحصائي للثوابت الإحصائية في حدود ثقة معينة أو ما يتعلق باختبار الفرضيات ، كما سنرى فيما بعد .

6-4- تحديد حجم العينة

تعتبر مسألة تحديد حجم العينة الواجب سحبها لدراسة ظاهرة ما من المسائل الهامة في نظرية العينات ، إذ يتوقف تحديد حجم العينة عادة وفقاً لنوع العينة المستخدم والهدف من الدراسة وطبيعة المجتمع المدروس أو الظاهرة المدروسة وكذلك التكلفة ودرجة الدقة المطلوبة بالبيانات، إذ يعتمد الباحث في تحديد حجم العينة على درجة الثقة التي يريدها من نتائجه فقد تكون % 90 ، 95% ، 95,5% ، 99% ... ، كما يعتمد أيضاً على حدود الخطأ الذي يمكن أن يفرضه أو يتوقعه أو يحدده عن طريق خصائص الظاهرة في المجتمع المدروس وهذا ما نطلق عليه بالخطأ المسموح بارتكابه ، وسنقتصر هنا فقط على تحديد حجم العينة العشوائية البسيطة تبعاً لطبيعة البيانات نوعية كانت أم كميّة .

6-4-1- تحديد حجم العينة للمتغيرات الكمية

لتحديد حجم العينة العشوائية البسيطة في تقدير المتوسط ، يتطلب منا معرفة كل من تباين الظاهرة المدروسة في المجتمع ودرجة الثقة (احتمال الدقة) المرغوبة في البيانات ، وكذلك مقدار الخطأ المسموح بارتكابه (d) ، لأن الخطأ المعياري للوسط الحسابي يساوي نسبة مقدار الخطأ المسموح بارتكابه إلى مقدار الأخطاء المعيارية المقابلة لاحتمال الدقة (z) ، أي :

$$\text{الخطأ المعياري للوسط الحسابي } \sigma_{\bar{x}} = \frac{\text{الخطأ المسموح بارتكابه } (d)}{\text{الأخطاء المعيارية المقابلة لاحتمال الدقة } (z)}$$

وبما أنَّ $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ فيكون : $\frac{\sigma_x}{\sqrt{n}} = \frac{d}{z}$ وبتربيع الطرفين والإصلاح نجد أن :

$n = \frac{z^2.\sigma_x^2}{d^2}$ وهو عبارة عن حجم العينة العشوائية البسيطة في حالة السحب مع الإعادة ،

أما حجمها في حالة السحب بدون إعادة ، فيكون : $\frac{\sigma_x}{\sqrt{n}}.\sqrt{\frac{N-n}{N-1}} = \frac{d}{z}$

وبتربيع الطرفين والإصلاح نجد أنَّ : $n = \frac{N.z^2.\sigma_x^2}{(N-1)d^2 + z^2.\sigma_x^2}$

مثال (6-2) : ما هو حجم العينة العشوائية الواجب سحبها من مجتمع الكتب لمكتبة ما لتقدير وسطي عدد الصفحات التي يحويها الكتاب ، مع العلم أن $N = 2000$ و $\sigma_x^2 = 600$ و $d = 5$ وفي حالة $z = 3$ ؟

الحل :

- نطبق دستور حجم العينة بدون إعادة :

صفحة

$$n = \frac{N.z^2.\sigma_x^2}{(N-1).d^2 + z^2.\sigma_x^2} = \frac{(2000).(3)^2.(600)}{(2000-1).(5)^2 + (3)^2(600)} = 195$$

- نطبق دستور حجم العينة مع الإعادة :

صفحة $n = \frac{z^2.\sigma_x^2}{d^2} = \frac{(3)^2.(600)}{(5)^2} = 216$

وفي أغلب الأحيان ، لا نعرف قيمة σ_x ، كما أنه ليس هناك عينة مسحوبة يمكن التقدير منها ، وفي هذه الحالة يمكننا تقدير قيمة σ_x من خلال جداء $\frac{1}{6}$ في المدى العددي ، الذي يمكن أن يقع به متوسط الظاهرة المدروسة .

مثال (6- 3) : رغبت وزارة الثقافة في معرفة متوسط عمر العمال في مطبعة ما، وكانت لديها معلومات تفيد بأنَّ العمال يبدأون العمل في الثامنة عشر من العمر ويحالون على التقاعد في الستين منه ، فما هو حجم العينة العشوائية المراد سحبها لتقدير هذا المتوسط ، إذا رغبنا التأكد لكل الحالات العملية على أنَّ لا يزيد الخطأ في التقدير عن 1,5 سنة ؟

الحل :

$$\sigma_x = \frac{1}{6}(60-18) = 7$$

$$n = \frac{z^2.\sigma_x^2}{d^2} = \frac{(3)^2.(7)^2}{(1{,}5)^2} = 196$$

196 عاملاً هو حجم العينة العشوائية المراد سحبها لتقدير متوسط عمر العمال في تلك المطبعة ، وبما أنَّ المجتمع الإحصائي كبير جداً هنا (المطابع التابعة لوزارة الثقافة) والسحب مع الإعادة ، **وبشكل عام** ، عندما لا يعرف لدينا حجم المجتمع الإحصائي الذي يجب أنَّ تسحب منه عينة عشوائية ، فيكون السحب مع الإعادة .

ومما سبق ، نلاحظ الآتي :

- أنَّ حجم العينة العشوائية البسيطة بدون إعادة هو أصغر من حجمها مع الإعادة ؛
- أنَّ كلفة العينة بدون إعادة هي أقل من العينة مع الإعادة من أجل نفس الخطأ المسموح بارتكابه ؛
- أنَّ حجم العينة العشوائية البسيطة يتناسب طرداً مع تباين الظاهرة المدروسة ، فكلما كان التباين σ_x^2 كبيراً ، كان حجم العينة كبيراً أيضاً وبالعكس ؛
- يتناسب حجم العينة العشوائية البسيطة طرداً مع درجة الثقة المرغوب فيها، وعكسياً مع الخطأ المرتكب (المسموح بارتكابه) .

6-4-2- تحديد حجم العينة للمتغيرات النوعية

يمكننا تحديد حجم العينة العشوائية البسيطة لتقدير نسبة الظاهرة في المجتمع (p) ، بصورة مشابهة لما سبق ،

أي أنَّ :

الخطأ المعياري للنسب $\sigma_p = \sqrt{\frac{p.q}{n}}$ ، كما أنَّ :

$$\text{الخطأ المعياري للنسب } \sigma_p = \frac{\text{الخطأ المسموح بارتكابه } (d)}{\text{الأخطاء المعيارية المقابلة لاحتمال الدقة } (z)}$$

وعليه يكون : $\sqrt{\frac{p.q}{n}} = \frac{d}{z}$ وبتربيع الطرفين والإصلاح نجد أنَّ حجم العينة يكون مساوياً

في حالة السحب مع الإعادة لــ : $n = \frac{z^2.p.q}{d^2}$

أما في حالة السحب بدون إعادة يكون حجم العينة مساوياً لــ :

$$n = \frac{N.z^2 p.q}{(N-1).d^2 + z^2.p.q}$$

وتنطبق هنا الملاحظات نفسها التي ذكرناها سابقاً .

مثال (6 – 4) : في أحد الأحياء من مدينة ما تقطن 4000 أسرة ، وكان المطلوب تقدير نسبة الأسر المستفيدة من الخدمات التي تقدمها مكتبة الحي بخطأ مرتكب لا يزيد عن 5% ، وكانت النسبة الحقيقية للمستفيدين مساوية لــ 45% ، فما هو حجم العينة العشوائية الواجب سحبها لتقدير النسبة الحقيقية للمستفيدين عند احتمال دقة 95,5% ؟

الحل :

بما أنَّ z المقابلة لاحتمال الدقة 95,5% تساوي لــ 2 ، وعليه يكون :

- حجم العينة مع الإعادة :

$$n = \frac{z^2.p.q}{d^2} = \frac{(2)^2.(45).(55)}{(5)^2} = 396 \text{ مستفيداً}$$

- حجم العينة بدون الإعادة :

$$n = \frac{N.z^2.p.q}{(N-1).d^2 + z^2.p.q} = \frac{(4000)(2)^2(45)(55)}{(4000-1)(5)^2 + (2)^2(45)(55)} = 360 \text{ مستفيداً}$$

ويجدر بنا هنا ، أنَّ نشير إلى الآتي :

1 - في معظم الأحيان لا يتوفر لدينا معلومات عن قيمة تحقق الظاهرة المدروسة في المجتمع الإحصائي (p) ، وفي هذه الحالة فقد جرت العادة على اعتبارها مساوية لــ 50 % .

مثال (6- 5) : لقد كان من المرغوب به ، دراسة حالة الأنفاق لتحديد النسبة المئوية للأسر التي تنفق قسماً من دخلها على شراء الكتب والمجلات ، فما هو

حجم العينة الواجب سحبها عشوائياً لتعطي باحتمال قدره 95,5% ، النسبة المئوية الحقيقية لإنفاق ا لأسر على شراء الكتب والمجلات ، على أنَّ لا يزيد الخطأ المرتكب في التقدير عن 5% ؟

الحل :

أسرة $n = \frac{z^2.p.q}{d^2} = \frac{(2)^2.(50).(50)}{(5)^2} = 400$ وهو حجم العينة العشوائية الواجب سحبها لتقدير النسبة الحقيقية للأسر التي تنفق قسماً من دخلها على شراء الكتب والمجلات .

2 -وفي بعض الأحيان تتوافر لدينا معلومات عن قيمة تحقق الظاهرة المدروسة في المجتمع (p) ، إلاَّ أنها تتراوح بين قيمتين كبرى وصغرى ، فإننا نأخذ بالقيمة الأقرب إلى % 50 ، وهنا نميز بين حالتين :

- $(5 \le p \le 10)$ فتكون قيمة $p = 10$ ؛
- $(60 \le p \le 80)$ فتكون قيمة $p = 60$.

مثال (6- 6) : لقد كان من المرغوب فيه تحديد نسبة العمال حملة الإجازة الجامعية في المكتبات ، وقد قدرت هذه النسبة بين % 12 و % 22 ، فما هو حجم العينة الواجب سحبها عشوائياً بغية تقدير النسبة الحقيقية لحملة الإجازة الجامعية في هذه المكتبة على أن لا يزيد الخطأ المرتكب في التقدير عن % 5 ، وأن لا يقل احتمال الدقة عن % 95,5 ؟

الحل :

نلاحظ هنا الحالة (1) ، وبالتالي تكون قيمة $(p = 22)$ وعليه نجد :

عاملاً يحمل إجازة جامعية $n = \frac{z^2.p.q}{d^2} = \frac{(2)^2(22)(78)}{(5)^2} = 275$

مثال (6- 7) : رغب أحد الباحثين في تقدير نسبة الأسر التي لا يزيد عدد أفرادها عن ثلاثة أشخاص ، وكان الاعتقاد السائد بأن نسبة هذه الأسر يتراوح بين % 60 و % 70 ، فما هو حجم العينة المراد سحبها عشوائياً لتقدير النسبة الحقيقية للأسر التي لا يزيد عدد أفرادها عن ثلاثة أشخاص ، على أن لا يزيد الخطأ المسموح بارتكابه في التقدير عن % 4 وأن لا يقل احتمال الدقة عن % 95,5 ؟

الحل :

تكون قيمة ($p = 60$) وفقاً للحالة (2) ، فنجد :
أسرة لا يزيد عدد أفرادها عن ثلاثة أشخاص

$$n = \frac{z^2.p.q}{d^2} = \frac{(2)^2(60)(40)}{(4)^2} = 600 \quad .$$

5-6- مسائل وتمارين غير محلولة

1 - لماذا يعد المجتمع الإحصائي مفهوماً مرناً ؟
2 - اذكر بعض الأبحاث التي تعتمد أسلوب العينة في الحصول على بياناتها الإحصائية ؟
3 - ما هي الحالات التي يفضل بها استخدام أسلوب العينة على أسلوب الحصر الشامل ؟
4 - على ماذا يعتمد الباحث أثناء مفاضلته في اختيار الأسلوب المناسب في جمع بياناته الإحصائية ؟
5 - ما هي اعتبارات تفضيل الباحث أسلوب العينة على أسلوب الحصر الشامل ؟
6 - ما هي العوامل التي يتوقف عليها حجم العينة المراد سحبها من المجتمع الإحصائي موضوع الدراسة ؟
7 - ما هي الخطوات الأساسية في تصميم العينة ؟
8 - ما الفرق بين العينات العشوائية وغير العشوائية ؛ ثم اذكر أنواع كل منها ؟
9 - ما هي العينة العشوائية وما هي شروطها ؟
10 - ما هي أخطاء العينة وكيف يتم معالجتها ؟
11 - ما هي الأسباب التي تؤدي إلى التحيّز ؟
12 - عرّف كلا من : توزيع المعاينة - توزيع معاينة الأوساط الحسابية – توزيع معاينة الانحرافات المعيارية - توزيع معاينة الفروق - توزيع معاينة النسب المئوية .
13 - ما المقصود بالخطأ المعياري : للوسط الحسابي – للانحراف المعياري – للفروق – للنسب المئوية ؟
14 - متى يكون التقدير غير متحيّز ؟
15 - لقد كان من المرغوب فيه تحديد نسبة العمال المصابين بأمراض تنفسية في إحدى المطابع، وقد قدرت هذه النسبة بناءً على خبرة سابقة بين 6% و10% ، كما رغب في تحديد نسبة العمال حملة الشهادة الثانوية في تلك المطابع ، فما ه و حجم العينة المراد سحبها عشوائياً لتقدير كل من :
أ - النسبة الحقيقية للمصابين بأمراض تنفسية في هذه المطبعة ؟

بـ- النسبة الحقيقية للعمال من حملة الشهادة الثانوية في هذه المطبعة ، شرط أن لا يزيد الخطأ في التقدير وبكلتا الحالتين عن % 3 ولا يقل احتمال الدقة عن % 95 ؟

16 - لقد كان من المرغوب فيه تقدير متوسط الأنفاق الشهري لأسر إحدى المدن السورية ، حيث كان التباين في متوسط إنفاق الأسر مساوياً لـ 12 ، فما هو حجم العينة العشوائية الواجب سحبها ، شرط أن لا يتجاوز الخطأ المرتكب في التقدير عن % 4 وأن لا يقل احتمال الدقة عن % 99.73 ؟

Damascus University

الفصل السابع
الاستدلال الإحصائي للعينات كبيرة الحجم

مقدمة

7-1- التقدير الإحصائي وأنواعه

7-1-1- التقدير النقطي

7-1-2- التقدير المجالي

7-1-2-1- تقدير مجال الثقة للأوساط الحسابية

7-1-2-2- تقدير مجال الثقة للانحرافات المعيارية

7-1-2-3- تقدير مجال الثقة للنسب المئوية

7-1-2-4- تقدير مجال الثقة للفروق

7-2- الاختبارات الإحصائية

7-2-1- اختبارات تتعلق بمجتمع واحد

7-2-2- اختبارات تتعلق بمجتمعين

7-3- مسائل وتمارين غير محلولة

وقل رب زدني علما
جامعة دمشق
Damascus University

مقدمة

سنتناول في هذا الفصل ما يعالجه الاستدلال الإحصائي بما يخص العينات كبيرة الحجم فقط ، تاركين ما يخص العينات صغيرة الحجم لفصل آخر . وقد سبق وأشرنا بأن الاستدلال الإحصائي يعالج مسألتين أساسيتين ، هما :

أ- تقدير المقاييس الإحصائية الوصفية للمجتمع (الثوابت الإحصائية ، مثل : الوسط الحسابي، الانحراف المعياري ، ...) بدلالة المقاييس الإحصائية الوصفية للعينة (التوابع الإحصائية ، مثل : الوسط الحسابي ، الانحراف المعياري ، ...) ، أي التقدير الإحصائي .

ب- يعتمد اتخاذ القرارات المتعلقة بالمجتمع بناءً على بيانات مستمدة من العينة (قرارات إحصائية) على فروض عن المجتمع موضوع الدراسة ، التي يمكن أن تكون صحيحة أو غير صحيحة (الفروض الإحصائية) ، لذا لا بد من التأكد من صحة أو عدم صحة تلك الفروض من خلال اختبارها ، بما يسمى بالاختبارات الإحصائية .

7-1- التقدير الإحصائي وأنواعه

يعد تقدير ثوابت المجتمع الإحصائي بدلالة إحدى العينات العشوائية المأخوذة منه، من أحد الأسباب الرئيسة لتحليل العينات ، إذ يمكن التمييز بين نوعين من التقدير هما :

7-1-1- التقدير النقطي

التقدير النقطي هو عبارة عن قيمة واحدة فقط يأخذها الثابت الإحصائي المقدر بدلالة التابع الإحصائي المقابل والمحسوب من تلك العين ات العشوائية المسحوبة من المتجمع الإحصائي المراد تقدير أحد ثوابته الإحصائية ، أي تقدير الثابت بقيمة واحدة ، فمثلاً : يكون أفضل تقدير للوسط الحسابي للمجتمع الإحصائي هو الوسط الحسابي المحسوب من عينة عشوائية سحبت من ذلك المجتمع ، على اعتبار أن هناك احتمالاً كبيراً ليكون الوسط الحسابي للعينة قريباً جداً من الوسط الحسابي للمجتمع غير المعلوم ، وما ينطبق على الوسط الحسابي ينطبق على الثوابت الإحصائية الأخرى ، كالانحراف المعياري والنسب المئوية وغيرها ... ولتوضيح ما سبق ، لنفترض أننا سحبنا عينة عشوائية مكونة من (8) عمال من مجتمع إحصائي يتألف من عمال مطبعة ما في مدينة حلب، فكانت أجورهم الأسبوعية بالليرات السورية، كالآتي :

740 ؛ 760 ؛ 800 ؛ 820 ؛ 870 ؛ 920 ؛ 980 ؛ 990 ، فيكون متوسط الأجر الأسبوعي لمجتمع هذه ا المطبعة هو قيمة متوسط الأجر لهذه العينة العشوائية

المسحوبة منه ، أي $\frac{6880}{8} = 860$ ليرة سورية . وهكذا نجد إن مسألة التقدير النقطي تتلخص في الحصول على قيمة ما للتابع الإحصائي المطلوب ، بحيث يمكن أن تعد أفضل تقدير للثابت الإحصائي غير المعروف ، إلاَّ أنَّ هذا التقدير يجب أن يتصف بالآتي :

- **عدم التحيّز** : يقال عن التابع الإحصائي بأنه تقدير غير متحيّز للثابت الإحصائي ، إذا كان الثابت الإحصائي هو الأمل الرياضي للتابع الإحصائي : $\hat{\mu} = E(X) = \overline{X}$ أو $\hat{p} = p$.
- **فعالية التقدير** : فكلما قل التباين زادت فعالية التقدير ، أي التباين عن الثابت الإحصائي أقل ما يمكن ، $Var(\hat{\mu}) \to \min$.
- **التقارب** : ويقال عن تقدير إنه متقارب إذا كانت قيمة التابع الإحصائي تقارب قيمة الثابت الإحصائي كلما زاد حجم العينة .

7-1-2- التقدير المجالي (تقدير المدى)

إن تقدير المدى هو عبارة عن تقدير قيمة الثابت الإحصائي ضمن مجال محدد عند احتمال معين بدلالة التابع الإحصائي المقابل له ، أخذين بالحسبان الخطأ المعياري للتابع الإحصائي المراد التقدير بدلالته ، فنحصل على قيمة للثابت الإحصائي تقع عند ذلك الاحتمال بين حدين أعلى وأدنى ، وذلك بافتراض أنَّ العينة العشوائية تتوزع توزيعاً طبيعياً ، إذ يطلق على هذين الحدين اسم حدّي الثِّقة أو مجال الثِّقة أو فترة الثِّقة ، وعلى احتمال وقوع التقدير بين هذين الحدين باحتمال الثقة (درجة الدقة)، وإن المتمم الحسابي لاحتمال الثقة يسمى بمستوى الدلالة ، فإذا رمزنا بـ (α) لمستوى الدلالة فإن درجة الثِّقة تساوي $(1-\alpha)$ ، التي تعبر وتدل على المساحة المحصورة تحت المنحنى الطبيعي ، وبناءً على جدول المساحات ما تحت المنحنى الطبيعي نورد القيم المختلفة لاحتمالات الثقة المستخدمة في الحياة العملية:

احتمال الثقة	% 68,27	% 80	% 90	% 95	% 95,45	% 98	% 99	% 99,73
مستوى الدلالة (α)	% 31,73	% 20	% 10	% 5	% 4,55	% 2	% 1	% 0,27
قيمة المتحول $z_{\alpha/2}$	1	1,28	1,645	1,96	2	2,33	2,58	3

ويلاحظ أنه كلما زدنا مستوى الدلالة ، نقصت المساحة المعتمدة للقيم ، أي جعلنا المدى أقل تشتتاً حول الثابت الإحصائي ، وهذا يعني ، أن تقديراتنا أصبحت أدق

وأكثر قرباً من الثابت الإحصائي ، أما إذا قللنا من مستوى الدلالة المتخذ اتسع المدى العددي الذي يُحتمل أن يقع به الثابت الإحصائي ، وبالتالي قلت الفائدة العملية من وراء إنشائه والعكس بالعكس .
هذا واعتماداً على ما سبق (توزيع المعاينة) ، سنتطرق هنا إلى التقدير المجالي لكل من : الأوساط الحسابية ، الانحرافات المعيارية ، النسب المئوية ، الفروق .

7-1-2-1- تقدير مجال الثِّقة للأوساط الحسابية

لنفترض أن لدينا مجتمعاً إحصائياً حجمه N وسحبنا منه جميع العينات العشوائية المتساوية الحجم ، فإنها تشكل فيما بينها توزيعاً طبيعياً ، وكان الوسط الحسابي لهذا المجتمع μ غير معروف ، أما انحرافه المعياري σ_x معروفاً ، ولنفرض أننا أخذنا عينة محددة من تلك العينات المسحوبة وحسبنا وسطها الحسابي ، فإنه بالإمكان تقدير الحدود التي يقع ضمنها الوسط الحسابي للمجتمع كالآتي :

حدا الثِّقة = التابع الإحصائي $\mp z_{\frac{\alpha}{2}}$. الخطأ المعياري للتابع الإحصائي

وعليه نكتب : $$\overline{X} \mp z_{\frac{\alpha}{2}} . \frac{\sigma_x}{\sqrt{n}}$$

وبشكل عام ، قد تكون قيمة الانحراف المعياري للمجتمع σ_x غير معروفة ، وفي هذه الحالة يمكننا الحصول على حدّي الثِّقة السابقة باستخدام التقدير $\overline{X}$ والانحراف المعياري المحسوب من العينة ، شرط أن يكون حجمها $(n \geq 30)$ ، وذلك كما يلي :

$$\overline{X} \mp t_{\frac{\alpha}{2}} . \frac{S_x}{\sqrt{n-1}}$$

حيث إن ($z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2}}$ في العينات الكبيرة الحجم) .

مثال (7-1) : لدينا مجتمع انحرافه المعياري معلوم ويساوي 30 ، أخذنا منه عينة عشوائية مكونة من 225 وحدة ، فكان وسطها الحسابي مساوياً لـ 168 ، والمطلوب : تقدير الوسط الحسابي للمجتمع إذا كانت :
أ- درجة الثِّقة % 95,45 ؛ ب- درجة الثِّقة % 99 .

الحل :

أ- نطبق العلاقة الآتية : $\overline{X} \mp z_{\frac{\alpha}{2}} . \frac{\sigma_x}{\sqrt{n}}$ فيكون :

$$168 \mp 2 * \frac{30}{\sqrt{225}} \Rightarrow [164 - 172]$$

وهذا يعني بأن متوسط المجتمع يقع بين 164 و 172 وحدة ، أي أننا باحتمال قدره 95,45 % واثقون بأن متوسط المجتمع لن يقل عن 164 وحدة ولن يزيد عن 172 وحدة ، أي :

$$p_r \{164 \le \mu \le 172\} = 0,9545$$

ب- نطبق العلاقة الآتية :

$$\overline{X} \mp z_{\frac{\alpha}{2}} . \frac{\sigma_x}{\sqrt{n}}$$

فيكون : $168 \mp 2,58 * \frac{30}{\sqrt{225}} \Rightarrow [162,84 - 173,17]$

أي أننا على ثِّقة وباحتمال قدره 99 % بأنَّ قيمة الوسط الحسابي لهذا المجتمع تتراوح بين 162,84 و73,17 وحدة.

مثال (7- 2) : أخذنا عينة عشوائية من مجتمع لا نعرف قيمة انحرافه المعياري وبحجم 122 طالباً ، فوجدنا أن متوسط طول الطالب فيها يساوي 170 سم والانحراف المعياري هو 9 سم ، والمطلوب : تقدير متوسط طول الطلبة الذين سحبت منهم هذه العينة، فيما إذا كانت درجة الثِّقة : أ- 95 % ؛ ب- 99,73 % .

الحل :

أ- نطبق العلاقة الآتية : $\overline{X} \mp t_{\frac{\alpha}{2}} . \frac{S_x}{\sqrt{n-1}}$

بالتعويض نجد : $170 \mp 1,96 * \frac{9}{\sqrt{122-1}} \Rightarrow 170 \mp 1,96 * \frac{9}{11}$

أي أنَّ حدي الثِّقة هما : [171,6 ، 168,4] ، أي أنَّ متوسط طول الطلبة ضمن درجة ثِّقة 95% يتراوح بين 168,4 سم و 171,6 سم .

ب- نطبق نفس العلاقة السابقة : $\overline{X} \mp t_{\frac{\alpha}{2}} . \frac{S_x}{\sqrt{n-1}}$ فنجد :

$$170 \mp 3 * \frac{9}{\sqrt{122-1}} \Rightarrow 170 \mp 3 * \frac{9}{11} \Rightarrow [167{,}55 - 172{,}45]$$

أي أنَّ متوسط طول الطلبة ضمن درجة ثِّقة 99,73 % يتراوح بين 167,55 سم و 172,45 سم .

7-1-2-2- تقدير مجال الثِّقة للانحرافات المعيارية

يمكننا اعتماداً على ما سبق أن نكتب :

حدا الثِّقة = التابع الإحصائي $\mp z_{\frac{\alpha}{2}} *$ الخطأ المعياري للتابع الإحصائي

أي أنّ حدود الثِّقة للانحراف المعياري σ_x لمجتمع يتوزع توزيعاً طبيعياً ، كما هي مقدرة من عينة عشوائية سحبت منه وانحرافها المعياري S_x ، تعطى كالآتي :

$$S_x \mp z_{\frac{\alpha}{2}} . \frac{\sigma_x}{\sqrt{2n}}$$

ولإيجاد حدود الثِّقة نستخدم S_x كتقدير لـ σ_x ، فيصبح : $S_x \mp z_{\frac{\alpha}{2}} . \frac{S_x}{\sqrt{2n}}$

وإذا أردنا معرفة حجم العينة الواجب سحبها عشوائياً من مجتمع إحصائي لا يختلف انحرافه المعياري الحقيقي σ_x عن الانحراف المعياري للعينة المسحوبة منه باحتمال معين لا يزيد عن نسبة الخطأ في تقدير الانحراف المعياري، إذ نكتب:

$$n = \frac{t_{\frac{\alpha}{2}}^2 . S_x^2}{2.d^2}$$

حيث وكما أشرنا سابقاً أنَّ ($z_{\frac{\alpha}{2}}^2 = t_{\frac{\alpha}{2}}^2$) في العينات الكبيرة الحجم .

مثال (7- 3) : إذا كان الانحراف المعياري للعمر الإنتاجي لعينة من 200 آلة طباعة كان 100 ساعة ، فأوجد :

1- حدود الثِّقة للانحراف المعياري لجميع آلات الطباعة من النوع نفسه عند احتمال قدره 95 % مرةً و 99 % مرةً أخرى ؟

2- حجم العينة الواجب سحبها عشوائياً من جميع آلات الطباعة المتماثلة بحيث نكون واثقين باحتمال قدره 99,73 % بأنّ الانحراف المعياري الحقيقي لن يختلف عن الانحراف المعياري للعينة بأكثر من 5 % ؟

الحل :

1- حدود الثِّقة للانحراف المعياري لجميع آلات الطباعة σ_x يعطى :

$$S_x \mp z_{\frac{\alpha}{2}} . \frac{\sigma_x}{\sqrt{2n}}$$

وبما أننا نقوم بتقدير الانحراف المعياري للمجتمع ، فإننا سنستخدم الانحراف المعياري للعينة كتقدير لـ σ_x

فيكون : $S_x \mp t_{\frac{\alpha}{2}} . \frac{S_x}{\sqrt{2n}}$

- الـ 95 % حدود الثِّقة هي :

$$100 \mp 1{,}96 * \frac{100}{\sqrt{2 * 200}} \Rightarrow 100 \mp 1{,}96 * \frac{100}{20}$$

$$\Rightarrow 100 \mp 1{,}96 * 5 \Rightarrow [90{,}2 - 109{,}8]$$

أي أننا واثقون باحتمال قدره 95 % بأنَّ الانحراف المعياري لجميع آلات الطباعة من نفس النوع سيقع بين 90,2 و 109,8 ساعة .

- الـ 99 % حدود الثِّقة هي :

$$100 \mp 2{,}58 * \frac{100}{\sqrt{2 * 200}} \Rightarrow 100 \mp 2{,}58 * \frac{100}{20}$$

$$\Rightarrow 100 \mp 2{,}58 * 5 \Rightarrow [87{,}1 - 112{,}9]$$

أي أننا نكون باحتمال قدره 99 % واثقين من أنَّ الانحراف المعياري لجميع آلات الطباعة من النوع نفسه سيقع بين 87,1 ساعة و 112,9 ساعة .

2 - حجم العينة يكون كما يلي :

$$n = \frac{t_{\frac{\alpha}{2}}^2 . S_x^2}{2.d^2} = \frac{(3)^2 . (100)^2}{2.(5)^2} = 1800 \text{ آلة}$$

7-1-2-3- تقدير مجال الثِّقة للنسب المئوية

اعتماداً على توزيع معاينة النسب المئوية وعلى القاعدة الأساسية في إنشاء حديِّ الثِّقة حول قيمة التابع الإحصائي ، يمكننا أن نقدر أيضاً حدود الثِّقة لنسبة تحقق الصفة في المجتمع الإحصائي الذي سحبت منه جميع العينات العشوائية المتساوية الحجم وحُسب بدلالة كل منها نسبة تحقق نفس الصفة ، فإننا نضع حدود الثِّقة لنسب المجتمع كالآتي :

$$p \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{p.q}{n}}$$

مثال (7 – 4) : في استطلاع للرأي العام بالعينة سحبت عينة عشوائية حجمها 100 من المستفيدين من خدمات مكتبة الحي ، حيث دلت على أنَّ 55 % منهم يستفيدون من هذه الخدمات ، فأوجد 99 % حدود الثِّقة للنسبة بين جميع المستفيدين من خدمات مكتبة الحي ؟

الحل :

نطبق العلاقة الآتية : $p \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{p.q}{n}}$ فنجد :

$$55 \mp 2{,}58 * \sqrt{\frac{55 * 45}{100}} \Rightarrow 55 \mp 2{,}58 * (4{,}975) \Rightarrow [42{,}16 - 67{,}84]$$

أي أننا باحتمال قدره 99 % واثقون بأن النسبة بين جميع المستفيدين من خدمات مكتبة الحي لن تقل عن 42,16 % ولن تزيد عن 67,84 % .

7-1-2-4- تقدير مجال الثِّقة للفروق

إذا أخذنا عينتين عشوائيتين يقترب توزيع معاينتهما من التوزيع الطبيعي للمجتمع الإحصائي الذي أخذتا منه ، فإنه يمكننا تقدير فترة الثِّقة لثوابت المجتمع الإحصائي بدلالة تلك العينات ، ونميزّ الآتي :

1- **تقدير مجال الثِّقة للفرق بين الأوساط الحسابية :**

$$(\overline{X}_1 - \overline{X}_2) \mp z_{\frac{\alpha}{2}} . \sigma_{\overline{x}_1 - \overline{x}2}$$

حيث إنَّ : $\sigma_{\overline{x}_1 - \overline{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

فيكون : $(\overline{X}_1 - \overline{X}_2) \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

وإذا كانت تباينات المجتمعات الإحصائية غير معلومة ، يستعاض عنها بدلالة تباينات العينات

(العينات مستقلة) ، فتصبح :

$$(\overline{X}_1 - \overline{X}_2) \mp t_{\frac{\alpha}{2}} . \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

2 - **تقدير مجال الثِّقة للفرق بين الانحرافات المعيارية :**

$$(S_1 - S_2) \mp z_{\frac{\alpha}{2}} . \sigma_{\sigma_1 - \sigma 2}$$

غير أنَّ :

$$\sigma_{\sigma-\sigma 2} = \sqrt{\frac{\sigma_1^2}{2.n_1} + \frac{\sigma_2^2}{2.n_1}}$$

فيصبح :

$$(S_1 - S_2) \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{\sigma_1^2}{2.n_1} + \frac{\sigma_2^2}{2.n_2}}$$

وإذا كانت تباينات المجتمعات الإحصائية غير معلومة ، يستعاض عنها بدلالة تباينات العينات المستقلة ، فيكون :

$$(S_1 - S_2) \mp t_{\frac{\alpha}{2}} . \sqrt{\frac{S_1^2}{2.n_1} + \frac{S_2^2}{2.n_2}}$$

3 - تقدير مجال الثِّقة لفروق النسب المئوية :

$$(p_1' - p_2') \mp z_{\frac{\alpha}{2}} . \sigma_{p_1 - p2}$$

وبما أنَّ :

$$\sigma_{p-p2} = \sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}$$

فيكون :

$$(p_1' - p_2') \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}$$

وإذا كانت النسب p_1 و p_2 في المجتمعين غير معلومة ، تقدر بدلالة نسبة العينات p_1' و p_2' :

$$(p_1' - p_2') \mp z_{\frac{\alpha}{2}} . \sqrt{\frac{p_1'.q_1'}{n_1} + \frac{p_2'.q_2'}{n_2}}$$

مثال (7 – 5) : أخذت عينة عشوائية من 150 كتاباً من الصنف A فكان متوسط عدد صفحاته 1400 صفحة وانحرافها المعياري 120 صفحة ، كما أخذت عينة عشوائية أخرى من 100 كتاب من الصنف B ، فوجد متوسط عدد صفحاته 1200 صفحة وانحرافها المعياري 80 صفحة ، والمطلوب :

أ - إيجاد 99 % حدود الثِّقة للفرق بين متوسط عدد الصفحات لمجتمعي الصنفين A و B ؟

ب - إيجاد 95 % حدود الثِّقة للفرق بين الانحرافات المعيارية لمجتمعي الصنفين A و B ؟

الحل :

أ- نطبق العلاقة : $(\overline{X}_1 - \overline{X}_2) \mp t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

بالتعويض نجد : $(1400-1200) \mp 2{,}58 * \sqrt{\frac{(120)^2}{150} + \frac{(80)^2}{100}}$

$$200 \mp 2{,}58 * (12{,}65) \Rightarrow 200 \mp 32{,}64 \Rightarrow [167{,}36 - 232{,}64]$$

أي أننا واثقون باحتمال قدره % 99 بأن الفرق بين متوسطات المجتمعات تقع بين 167,36 صفحة
و 232,64 صفحة .

ب- نطبق العلاقة : $(S_1 - S_2) \mp t_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S_1^2}{2.n_1} + \frac{S_2^2}{2.n_2}}$

بالتعويض نجد : $(120-80) \mp 1{,}96 * \sqrt{\frac{(120)^2}{2*150} + \frac{(80)}{2*100}}$

$$40 \mp 1{,}96 * (8{,}94) \Rightarrow 40 \mp 17{,}52 \Rightarrow [22{,}48 - 57{,}52]$$

أي نحن واثقون باحتمال قدره % 95 بأن فروق الانحرافات المعيارية للمجتمعين لن تقل عن 22,48 صفحة ولن تزيد عن 57,52 صفحة .

مثال (7 – 6) : في عينة عشوائية مكونة من 600 من البالغين و 400 من المراهقين الذين شاهدوا برنامجاً تلفزيونياً معيناً ، وجد أن 300 من البالغين و 100 من المراهقين يفضلون هذا البرنامج، فما هي 95,5% لحدود الثِّقة للفرق بين نسبة كل من البالغين ونسبة كل من المراهقين الذين شاهدوا البرنامج ويفضلونه ؟

الحل :

نوجد كل من $p'_1 = \frac{300}{600} = 50\%$ و $p'_2 = \frac{100}{400} = 25\%$.

نطبق العلاقة الآتية : $(p'_1 - p'_2) \mp z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p'_1.q'_1}{n_1} + \frac{p'_2.q'_2}{n_2}}$

بالتعويض نجد : $(50-25) \mp 2 * \sqrt{\frac{(50*50)}{600} + \frac{(25*75)}{400}}$

$$25 \mp 2 * (2,98) \Rightarrow [19,04 - 30,96]$$

أي أننا واثقون باحتمال قدره % 95,5 أنَّ الفرق الحقيقي يقع بين % 19,04 و % 30,96 .

مثال (7 – 7) : في مطبعة (أ) لإنتاج الكتب المدرسية ، كان متوسط عدد صفحات الكتاب 1500 صفحة وبانحراف معياري قدره 250 صفحة ، بينما في المطبعة (ب) كان متوسط عدد صفحات الكتاب 1300 صفحة وبانحراف معياري قدره 125 صفحة ، فإذا سحبنا عينة عشوائية من 125 كتاباً من كل من المطبعتين ، فما احتمال كون متوسط عدد صفحات الكتاب في المطبعة (أ) أكبر منه في المطبعة (ب) ، على الأقل 150 صفحة ؟

الحل :

المعطيات هي : $\sigma_1 = 250$ ؛ $\sigma_2 = 125$ ؛ $\overline{X}_1 = 1500$ ؛ $\overline{X}_2 = 1300$ ؛ $n_1 = n_2 = 125$.

الفرق بين الوسطين الحسابين هو : $(\overline{X}_1 - \overline{X}_2) = 1500 - 1300 = 200$ صفحة .

وإن الخطأ المعياري لفروق الأوساط الحسابية :

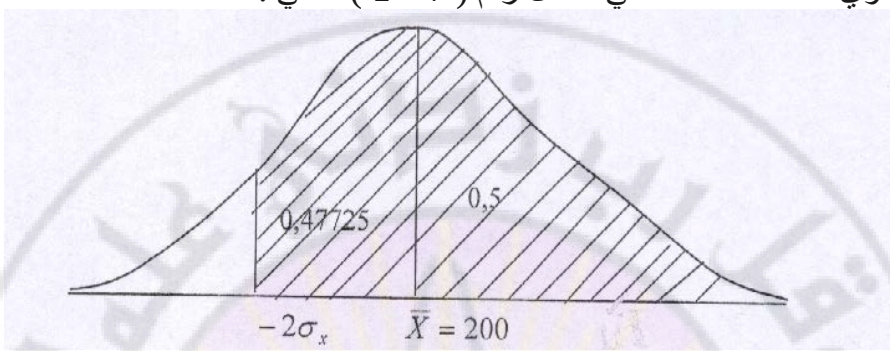
$$\sigma_{\overline{x}_1 - \overline{x}2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(250)^2}{125} + \frac{(125)^2}{125}} = 25 \text{ صفحة} .$$

هذا التوزيع الجديد للفرق بوسطائه : الوسط الحسابي الذي يساوي 200 صفحة وبانحراف معياري قدره 25 صفحة هو توزيع طبيعي بغض النظر عن التوزيعات الأصلية ، حيث تمثل القيمة 200 وسطي الفروق بين أوساط العينتين ، ولإيجاد احتمال كون الفرق على الأقل 150 صفحة نوجد أولاً :

$$z = \frac{X_i - \overline{X}}{\sigma_x} = \frac{150 - 200}{25} = -2$$

وهي تقابل ما مساحته تحت المنحنى $\frac{95,45\%}{2} = 47,725\%$ وهي عبارة عن المساحة المحصورة بين 150 و 200 ومن ثم نضيف إليها % 50 وهي عبارة عن المساحة المقابلة للقيم 200 وأكثر ، فيكون الاحتمال المطلوب : 0,97725 = 0,50 + 0,47725

أي أنَّ احتمال كون متوسط عدد صفحات الكتاب في المطبعة (أ) أكبر من متوسط عدد صفحات الكتاب في المطبعة (ب) يساوي لـ % 97,725 ، وهو يساوي المساحة المظللة في الشكل رقم (7 – 1) التالي :



الشكل رقم (7– 1)

2-2- اختبار الفرضيات

تطرقنا فيما سبق إلى التقدير الإحصائي (النقطي ، المجالي) لمؤشرات المجتمع والمتعلقة بخاصة ما من هذا المجتمع ، والآن سنتطرق إلى اختبار الفرضيات الخاصة بهذه المؤشرات بهدف اتخاذ القرار بقبول أو برفض تلك الفرضية .

فالفرضية الإحصائية هي تحديد قيمة ثابت إحصائي ، أي تعيين خاصة من خواص التوزيع الاحتمالي للمتغير العشوائي ، وكما رأينا ، فإن كل توزيع يتمتع بخواص معينة ، فمثلاً : التوزيع الطبيعي يتميز بالوسط الحسابي والانحراف المعياري ونرمز له بـ $N(\mu, \sigma^2)$ ، وتوزيع ثنائي الحدين يتميز بمعرفة نسبة تحقق p أو نسبة عدم تحقق q الصفة المراد دراستها ، لذا فإن معرفة طبيعة البيانات الإحصائية المتوفرة (نوعية أم كمّية) يساعدنا في اختيار أسلوب الاختبار الإحصائي الملائم لتلك البيانات ، أي معرفة طبيعة التوزيع الذي تخضع له تلك البيانات ، عما إذا كانت خاضعة لتوزيع طبيعي أم لتوزيع ثنائي الحدين .

هذا ويمكن تحديد القيم الحقيقية لهذه الخواص ، بالاستناد إلى المعلومات المتوفرة عن جميع مفردات المجتمع ، وكل قيمة من هذه القيم الحقيقية الثابتة تسمى كما سبق وأشرنا بـ : " الثابت الإحصائي للمجتمع " ، وتتحدد الفرضية الإحصائية التي نريد اختبارها بتحديد قيمة هذا الثابت .

لنفرض مثلاً ، أننا نقوم بدراسة توزيع مجتمع إحصائي معين ، انحرافه المعياري معلوم ويساوي الواحد الصحيح ، أما وسطه الحسابي μ مجهول ، فالفرضية الممكن اعتمادها هي أن يكون التوزيع طبيعياً معيارياً أم لا ؟

لذا لا بد من إخضاع تلك الفرضية لاختبار يسمح لنا بأن نختار بين فرضين H_0 و H_1 بالاعتماد على عينة ذات حجم معين مسحوبة من المجتمع الإحصائي المدروس ، أي يتم الاختبار وفق ما يلي :

1 - أن يكون التوزيع طبيعياً معيارياً ، أي الفرضية المعتمدة صحيحة ، والوسط الحسابي يساوي الصفر ، أي : $H_0 : \overline{X} = \mu = 0$ ويطلق على H_0 بفرضية العدم .

2 - أن يكون التوزيع غير معياري ، أي الفرضية المعتمدة غير صحيحة ، والوسط الحسابي لا يساوي الصفر ، أي : $H_1 : \overline{X} \neq \mu \neq 0$ كما يطلق على H_1 بالفرضية البديلة (فرضية نفي العدم).

وبشكل عام ، تقوم الاختبارات الإحصائية على أساس تحديد احتمال تحقق الفرضيات الإحصائية المدروسة ، ومدى الثِّقة الممكن وضعها في الفرضية الإحصائية المعتمدة .

ونلاحظ هنا ، بأنَّ الفرضية المعتمدة قد تكون صحيحة وقد تكون خاطئة ، وللتأكد من صحتها أو خطئها ، فإننا نلجأ إلى اختبار الفرضية نفسها ، ويتم الاختبار بسحب عينة عشوائية من المجتمع الإحصائي المدروس ، وبالاستناد إلى قيم هذه العينة نقرر صحة الفرضية أو خطأها ، أي أنه تتم المقارنة بين تابع إحصائي وثابت إحصائي ، فإن كان الفرق بينهما صغيراً فمرده للأخطاء الاحتمالية أو أخطاء الحظ والصدف وقبلنا الفرضية ، أي أنَّ نتائج العينة لا تختلف عن نتائج المجتمع المأخوذة منه ، أما إذا كان الفرق بين التابع الإحصائي والثابت الإحصائي المقابل له كبيراً رفضنا الفرضية المعتمدة وقلنا بأن نتائج العينة تختلف عن نتائج المجتمع الإحصائي المأخوذة منه (اختلافاً حقيقياً أو جوهرياً).

وفي الحقيقة ، إن اتخاذ القرار بقبول أو برفض الفرضية يصاحبه نوعان من الأخطاء المحتمل ارتكاب إحداها:

1 - **خطأ من النوع الأول** : نرتكب خطأ من النوع الأول إذا رفضنا فرضية العدم H_0 وكانت في حقيقتها صحيحة، واحتمال ارتكاب هذا الخطأ يدعى بمستوى الدلالة α ويحدد مسبقاً ويعبر عن احتمال الرفض الخاطىء ، أي رفض فرضية صحيحة .

2 - **خطأ من النوع الثاني** : نرتكب خطأ من النوع الثاني ، إذا قبلنا فرضية العدم H_0 وكانت في حقيقتها خاطئة ، واحتمال ارتكاب هذا الخطأ هو β وهو يعبر عن احتمال القبول الخاطىء ، أي قبول فرضية خاطئة ، والجدول الآتي يبين لنا الحالات المختلفة للأخطاء المرتكبة :

نتيجة الاختبار / حقيقة الفرضية	قبول الفرضية	رفض الفرضية
صحيحة	قرار سليم واحتماله $(1-\alpha)$	رفض خاطىء احتماله (α) (خطأ من النوع الأول)
خاطئة	قبول خاطىء احتماله (β) (خطأ من النوع الثاني)	قرار سليم واحتماله $(1-\beta)$ ويسمى بقوة الاختبار

ولكي يكون اختبار الفرضيات جيداً أو تكون قواعد اتخاذ القرار جيدة ، يجب أن تصمم بحيث تؤدي إلى التقليل من أخطاء القرار ، وهذا ليس بالأمر السهل ، عندما يكون حجم العينة ثابتاً ، لأنه عند تقليل احتمال ارتكاب أحد أنواع الخطأ يصاحبه زيادة في احتمال ارتكاب النوع الآخر من الخطأ .
ومن الناحية العملية فإن أحد أنواع الخطأ قد يكون أكثر خطورة من النوع الآخر ، وبهذا يجب الوصول إلى حل وسط لصالح تحديد الخطأ الأكثر خطورة ، هذا وتعتبر الأخطاء من النوع الأول الأكثر خطورة وفداحة في أغلب الأحيان ، وعليه يتم تحديد مستوى الدلالة (α) من خلال المقارنة بين هذين النوعين من الأخطاء ، فلو كنا أمام مستوى دلالة 5 % وكان ارتكاب الخطأ من النوع الأول الأكثر خطورة وفداحة فإننا نأخذ مستوى دلالة أقل 1 % ، أما إذا كنا أمام مستوى دلالة 1 % وكان ارتكاب الخطأ من النوع الثاني هو الأكثر فداحة وخطورة فإننا نأخذ مستوى دلالة أكبر 5 % . ولتجنب الوقوع في هذين النوعين من الأخطاء إلى أقصى حدٍ ممكن يتم من خلال " التحليل التتابعي" و هو نوع من الاختبارات الإحصائية الحديثة المستخدمة في تحديد الحد الأقصى لاحتمال ارتكاب هذين النوعين من الأخطاء بآن معاً ، غير أنَّ الطريقة الوحيدة للتقليل من هذين النوعين من الأخطاء هو زيادة حجم العينة، وهذا قد لا يكون ممكناً في كثير من الأحيان .

مثال : لنفرض أننا نرغب في اختبار (100) فرضية منها (80) فرضية خاطئة و (20) فرضية صحيحة ، وكنا أمام مستوى دلالة 5 % ، فإننا نرفض جميع الفرضيات الخاطئة دون أن نرتكب أي خطأ من النوع الأول و سنرفض فرضية واحدة من الفرضيات الصحيحة ، أي يحدث خطأ واحد في كل (100) قرار ، وهذا يدل على أن الحد الأقصى لاحتمال ارتكابنا أخطاء من النوع الأول هي تلك النسبة التي يشير إليها مستوى الدلالة ، عندما تكون كافة الفرضيات المختبرة صحيحة ، إلاَّ أنه في أغلب الأحيان تكون النسبة لاحتمال ارتكابنا أخطاء من النوع الأول أقل من تلك التي يشير إليها مستوى الدلالة المتخذ.

هذا ولإجراء أي اختبار عن فرضية إحصائية ما (عينات كبيرة الحجم أم صغيرة الحجم) ، يراعى إتباع الخطوات الآتية :

1 - **تحديد الفرضية** :

- فرضية العدم H_0 : وهي الفرضية التي تصاغ على أمل رفضها ، أي بمعنى لا يوجد اختلاف حقيقي أو جوهري بين التابع الإحصائي للعينة والثابت الإحصائي للمجتمع ، فيما إذا كانت المقارنة بين تابع وثابت (اختبار يتعلق بمجتمع واحد) ، أما إذا كانت المقارنة بين تابع إحصائي وتابع إحصائي آخر (اختبار يتعلق بمجتمعين) ، فتنص الفرضية H_0 عندئذٍ على أنه لا يوجد اختلاف جوهري بينهما ، وهذا ما سنوضحه فيما بعد .
- الفرضية البديلة H_1 : وهي على خلاف فرضية العدم H_0 ، أي عند رفض الفرضية H_0 ، فإننا نقبل الفرضية H_1 ، أي يوجد اختلاف جوهري فيما بين المراد الاختبار حولهما .

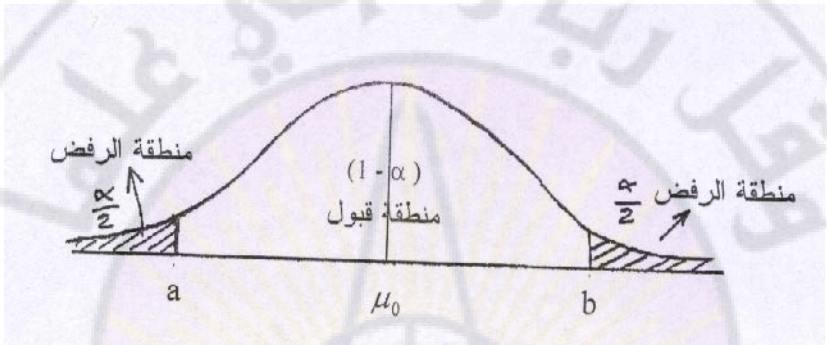
2 - **تحديد مستوى الدلالة** (α) أو درجة الثِّقة أو احتمال الدقة $(1-\alpha)$.

3 - **تحديد قيمة** $z_{\frac{\alpha}{2}}$ ، z_α **النظرية المقابلة لمستوى الدلالة المُعطى ،** التي تؤخذ من جدول المساحات لتوزيع المنحنى الطبيعي ، أي تحديد النقاط الفاصلة بين منطقتي القبول أو الرفض في ضوء ذلك المستوى .

4 - **تحديد قيمة الاختبار** Z والمحسوبة في ضوء البيانات المعلومة (تختلف باختلاف المؤشر المراد اختباره) .

5 - **المقارنة واتخاذ القرار الإحصائي** من خلال مقارنة قيمة $z_{\frac{\alpha}{2}}$ أو z_α مع قيمة الاختبار Z المحسوبة، فالقرار الإحصائي يقسم مجال قيمة الاختبار Z المحسوبة إلى منطقتين ، تسمى إحداهما بمنطقة القبول والأخرى بمنطقة الرفض ، ومنطقة القبول هي المنطقة التي تتكون من قيمة الاختبار Z المحسوبة وكثيرة الحدوث إذا كانت H_0 صحيحة ، أما منطقة الرفض فهي المنطقة التي تتكون من قيمة الاختبار Z قليلة الحدوث إذا كانت H_0 صحيحة ، وبناءً على ذلك يكون القرار برفض H_0 إذا وقعت قيمة الاختبار في منطقة الرفض ، ويكون القرار قبول H_0 إذا وقعت قيمة الاختبار قي منطقة القبول ، هذا ويمكن تحديد منطقتي القبول أو الرفض اعتماداً على الفرضية البديلة H_1 والتي نميز فيها ثلاثة حالات هي :

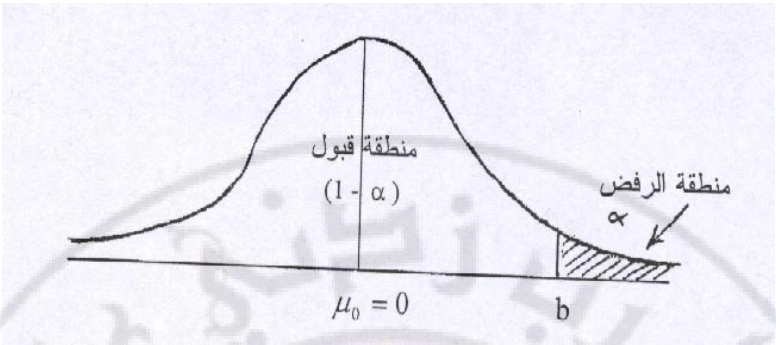
أ- الفرضية $H_0 : \mu = \mu_0 = 0$ ، فالفرضية البديلة $H_1 : \mu \neq \mu_0 \neq 0$ وهنا نكون أمام خيارين: إما أن تكون $\mu \succ \mu_0 \succ 0$ أو أن تكون $\mu \prec \mu_0 \prec 0$ ، أي أننا أمام اختبار من اتجاهين ، وفي هذه الحالة نضع نصف α في كل من طرفي توزيع قيمة الاختبار، كما في الشكل رقم(7 – 2) الآتي :



الشكل رقم (7 –2)

وتدعى المنطقة بين a و b بالمنطقة الحرجة (بمنطقة القبول ، والمنطقة التي على الطرفين بمنطقة الرفض ، والقيمتان a و b اللتان تفصلان هاتين المنطقتين بالقيم الحرجة ، أي $p_r(z \prec a) = p_r(z \succ b) = \frac{\alpha}{2}$ ، أي أننا نقبل H_0 إذا وقعت قيمة Z المحسوبة بين القيمتين a و b ونرفضها إذا وقعت في المنطقة المظللة التي تمثل منطقة الرفض .

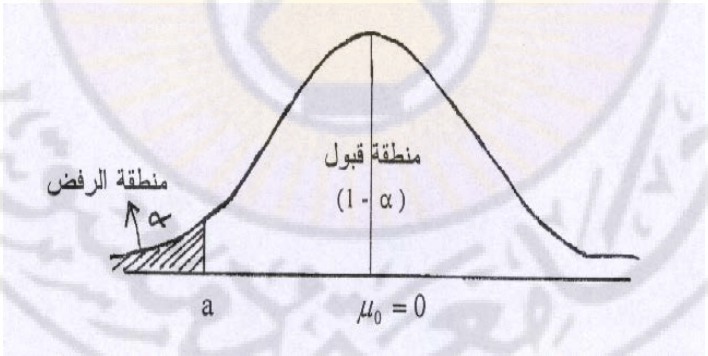
ب- فرضية العدم $H_0 : \mu = \mu_0 = 0$ والفرضية البديلة $H_1 : \mu \succ \mu_0 \succ 0$ وفي هذه الحالة تكون منطقة الرفض في جهة اليمين فقط ، لذلك نضع α بكاملها في الطرف الأيمن من توزيع قيمة الاختبار Z ، وهذا ما يطلق عليه اختبار من طرف واحد أو من اتجاه واحد ، $p_r(z \succ b) = \alpha$ كما في الشكل رقم (7 – 3) الآتي :



الشكل رقم (7 – 3)

ونقبل الفرضية H_0 إذا وقعت قيمة الاختبار Z المحسوبة في منطقة القبول ونرفضها إذا وقعت في المنطقة المظللة .

ج- فرضية العدم $H_0 : \mu = \mu_0 = 0$ في حين أن الفرضية البديلة H_1 : $\mu \prec \mu_0 \prec 0$ وفي هذه الحالة نضع α في الطرف الأيسر من توزيع قيمة الاختبار Z ، وهذا ما يطلق عليه أيضاً باختبار من طرف واحد أو من اتجاه واحد ، $p_r(z \prec a) = \alpha$ كما في الشكل رقم (7 – 4) الآتي :



الشكل رقم (7 – 4)

ونقبل H_0 إذا وقعت قيمة الاختبار Z المحسوبة في منطقة القبول ونرفضها إذا وقعت في المنطقة المظللة .

وبناءً على ما سبق نميزّ إذا كانت :

1 - قيمة التابع الإحصائي أكبر أو أقل من قيمة الثابت الإحصائي ، كان الاختبار من اتجاهين ؛

2 - قيمة التابع الإحصائي أكبر أو أفضل أو أحسن من قيمة الثابت الإحصائي ، كان الاختبار من اتجاه واحد ويقع على الطرف الأيمن من التوزيع ؛

3 - قيمة التابع الإحصائي أقل أو أسوأ من قيمة الثابت الإحصائي ، كان الاختبار من اتجاه واحد ويقع على الطرف الأيسر من التوزيع .

وفيما يلي جدول يعطي بعض القيم الحرجة لـ z لكل من الاختبارات في اتجاه واحد والاختبارات من اتجاهين لمستويات دلالة مختلفة ، والقيم الأخرى التي يمكن الحصول عليها باستخدام جدول المساحة تحت المنحنى الطبيعي:

مستوى الدلالة α	0,10	0,05	0,01	0,005	0,002
قيم حرجة z_α لاختبار من اتجاه واحد	– 1,28 1,28	– 1,645 1,645	– 2,33 2,33	– 2,58 2,58	– 2,88 2,88
قيم حرجة $z_{\frac{\alpha}{2}}$ لاختبار من اتجاهين	– 1,645 1,645	– 1,96 1,96	– 2,58 2,58	– 2,81 2,81	– 3,08 3,08

هذا ويمكن تقسيم الاختبارات الإحصائية إلى ما يلي :

- اختبارات تتعلق بمجتمع واحد ؛
- اختبارات تتعلق بمجتمعين ؛
- اختبارات تتعلق بأكثر من مجتمعين (لن نتطرق لها في مفردات هذا المنهاج) ، إذ نقتصر هنا على دراسة بعض الاختبارات التي تتعلق بمجتمع واحد أو مجتمعين ، بالاعتماد على توزيعات المعاينة التي مرت معنا سابقاً ، ونكتفي بتوزيعات المعاينة للمجتمعات غير المحدودة أو السحب مع الإعادة ، أما إذا كانت المعاينة بدون إعادة من المجتمعات المحدودة فيجب تعديل نتائجها ، كما أشرنا إليه سابقاً .

7-2-1- اختبارات تتعلق بمجتمع واحد : وهي تلك الاختبارات التي تتعلق بمقارنة قيمة التابع الإحصائي مع قيمة الثابت الإحصائي المقابل له ، وبناءً عليه نتطرق إلى :

▪ **اختبار الأوساط الحسابية** : وهو مقارنة قيمة الوسط الحسابي المحسوب بدلالة العينة العشوائية كتابع إحصائي مع قيمة الوسط الحسابي المحسوب من المجتمع الإحصائي الذي أخذت منه العينة كثابت إحصائي ، قد تكون قيمته معلومة أو مفترضة أو محددة بصورة مسبقة حسب المواصفات المرغوب فيها ، حيث يكون توزيع معاينة التوابع الإحصائية في هذه الحالة موزعة توزيعاً طبيعياً حول وسط

حسابي يساوي بالضبط قيمة الثابت الإحصائي المقابل له ، وعليه نطبق خطوات اختبار الفرضيات الإحصائية ونوجد قيمة الاختبار Z كالآتي :

$$Z = \frac{\left|\overline{X} - \mu\right|}{\sigma_{\bar{x}}} = \frac{\left|\overline{X} - \mu\right|}{\frac{\sigma_x}{\sqrt{n}}}$$

حيث إنَّ :

$\overline{X}$ قيمة التابع الإحصائي وهو الوسط الحسابي للعينة ؛

μ قيمة الثابت الإحصائي وهو الوسط الحسابي للمجتمع الإحصائي ؛

$\sigma_{\bar{x}}$ الخطأ المعياري للوسط الحسابي ، $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ وعند الضرورة نستخدم الانحراف المعياري للعينة S_x كتقدير لـ σ_x (انظر ما سبق وما سيأتي في الفصل الثالث الاستدلال الإحصائي للعينات الصغيرة الحجم) .

مثال (7– 8) : بفرض أنَّ متوسط الأجر الشهري لعمال إحدى المكتبات في بلد ما كان مساوياً لـ 2500 ل.س وبانحراف معياري يساوي 200 ل.س ، وكان اهتمام القائمين في هذه المكتبة معرفة متوسط نفقات العمال على السكن لكي يتخذوا قراراً ، عما إذا كان بإمكان هؤلاء العمال من تسديد الأقساط الشهرية في حال تم توفير بيوت سكنية لهم ، وللتأكد سحبت عينة من 100 عامل ، فوجد متوسط أجره الشهري يساوي 2400 ل.س ، والمطلوب : هل يختلف متوسط الأجر الشهري للعامل عن متوسط الأجر الشهري في هذه المكتبة ، وذلك عند مستوى دلالة 5 % ثّم 1 % ؟

الحل :

نطبق خطوات اختبار الفرضيات كالآتي :

1 - تحديد الفرضية : - فرضية العدم H_0 : $\overline{X} = \mu$ وفرضية نفي العدم أو البديلة H_1 : $\overline{X} \neq \mu$.

2 - تحديد مستوى الدلالة 5 % و 1 % .

3 - تحديد قيمة الاختبار النظرية ، والاختبار هنا من اتجاهين :

- عند مستوى دلالة 5 % تكون قيمة الاختبار النظرية

$$z_{\frac{\alpha}{2}} = z_{\frac{0,05}{2}} = z_{0,025} = 1,96$$

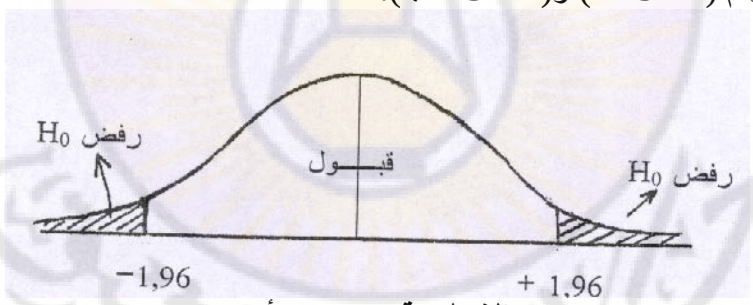
- أما عند مستوى دلالة % 1 فتكون قيمة الاختبار النظرية $z_{\frac{\alpha}{2}} = z_{\frac{0,01}{2}} = z_{0,005} = 2,58$

4 - تحديد قيمة الاختبار Z المحسوبة من العلاقة :

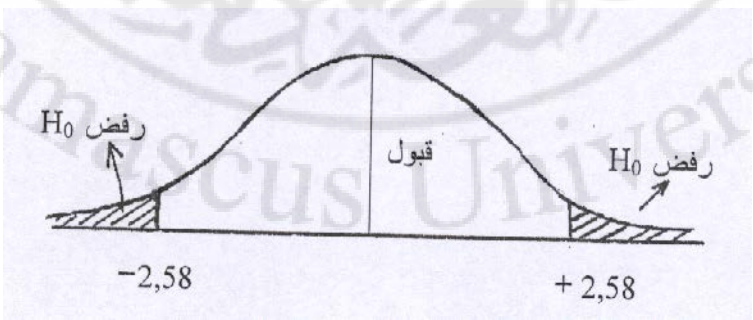
$$Z = \frac{\left|\overline{X} - \mu\right|}{\frac{\sigma_x}{\sqrt{n}}} = \frac{\left|2400 - 2500\right|}{\frac{200}{\sqrt{100}}} = \left|-5\right| \quad .$$

5 - المقارنة واتخاذ القرار : $Z = 5 \succ z_{\frac{0,05}{2}} = 1,96$ وبما أن Z المحسوبة أكبر من قيمة الاختبار النظرية عند مستوى الدلالة % 5 ، لذلك فإننا نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 ، أي يختلف متوسط الأجر الشهري للعامل عن متوسط الأجر الشهري لعمال هذه المكتبة .

نجد عند مستوى الدلالة 1% بأن $Z = 5 \succ z_{\frac{0,01}{2}} = 2,58$ وبالتالي يكون القرار كما هو الحال عند مستوى الدلالة % 5 ، رفض فرضية العدم H_0 ، وهذا موضح في الشكل رقم (7- 5– أ) و(7 – 5 – ب).



الشكل رقم (7–5 – أ)



الشكل رقم (7 – 5 – ب)

ومعنى رفض الفرضية هنا يعود إلى حدوث تحيّز في أسلوب سحب هذه العينة ، فالعينة ليست عشوائية ، مما أدى إلى اختلاف جوهري بين قيمة التابع الإحصائي وقيمة الثابت الإحصائي المقابل له إلى درجة بلغ معها (5) أخطاء معيارية .

مثال (7 – 9) : ترغب إحدى المؤسسات في شراء ورق للطباعة بغية استعمالها في طباعة الكتب ، وقد حددت مواصفات شرط أن يكون متوسط السماكة 2500 ملغ وبانحراف معياري قدره 300 ملغ ، وإن شحنة قد وردت إليها ، فقمنا بسحب عينة عشوائية من 100 ماعون من هذه الشحنة ، فأظهرت أن متوسط السماكة قدره 2450 ملغ ، والمطلوب : هل يجب قبول هذه الشحنة أم لا عند مستوى دلالة 1 % ؟

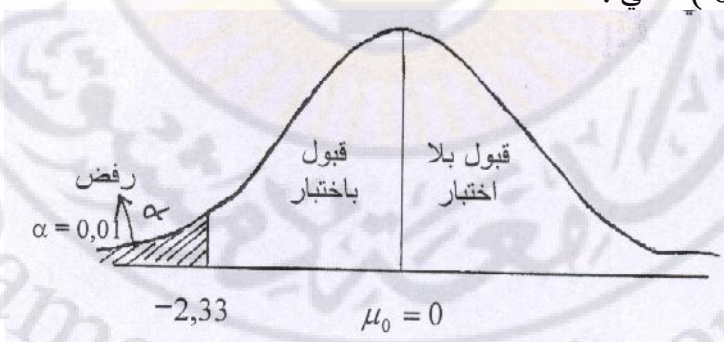
الحل :

1- الفرضية : - $H_0 : \overline{X} = \mu$ و $H_1 : \overline{X} \prec \mu$.

2- تحديد مستوى الدلالة : $\alpha = 0,01$.

3- تحديد منطقتي الرفض والقبول : نلاحظ هنا بأن الاختبار من اتجاه واحد ومن اليسار ، لأن مواصفات الشحنة أقل مما هو محدد من قبل المؤسسة ، إذ تقبل جميع القضبان التي متوسط متانتها أكبر من المحدد دون اختبار ، وعليه يكون :

$z_\alpha = z_{0,01} = 2,33$ ، وهذا موضح في الشكل رقم (7 – 6) الآتي :



الشكل رقم (7 – 6)

1 - نحدد قيمة الاختبار : $Z = \dfrac{\left|\overline{X} - \mu\right|}{\dfrac{\sigma_x}{\sqrt{n}}} = \dfrac{\left|2450 - 2500\right|}{\dfrac{300}{\sqrt{100}}} = \left|-1,67\right|$.

2 - المقارنة واتخاذ القرار : نلاحظ بأن $Z = 1,67 \prec z_{\alpha} = 2,33$ لذلك نقبل فرضية العدم H_0 ، أي لا يوجد اختلاف جوهري بين مواصفات الشحنة الواردة وبين المواصفات المحددة من قبل المؤسسة ، لذا نقرر قبول الشحنة ، والفرق المشاهد يعود للأخطاء الاحتمالية .

- **اختبار النسب المئوية** : وهو مقارنة نسبة تحقق الظاهرة المدروسة p' في العينة كتابع إحصائي مع نسبة تحقق الظاهرة p في المجتمع الإحصائي الذي أخذت منه هذه العينة ، حيث نطبق نفس خطوات الاختبار السابقة ، غير أننا نحسب قيمة الاختبار Z بالعلاقة الآتية :

$$Z = \frac{|p' - p|}{\sigma_p} = \frac{|p' - p|}{\sqrt{\frac{p.q}{n}}}$$

مثال (7- 10) : إذا كان معلوماً أنَّ حوالي 25 % من المستهلكين يفضلون المجلات من النوع (أ) وقد قامت الشركة الموزعة لهذا النوع من المجلات بحملة دعائية ، ولبيان مدى نجاح هذه الحملة ، أخذت عينة من المستهلكين حجمها 230 مستهلكاً ، فوجد من بينهم 69 مستهلكاً يفضلون هذا النوع ، والمطلوب : هل أدت الحملة الدعائية إلى زيادة نسبة المستهلكين لهذا النوع من المجلات أم لا ، عند مستوى دلالة 5 % ؟

الحل :

- نعلم إن النسبة في المجتمع $p = 25\%$ ونحسب النسبة في العينة $p' = \frac{69}{230} = 30\%$.

1- نضع الفرضيتين كما يلي :

$$H_0 \rightarrow p' = p = 25\%$$
$$H_1 \rightarrow p' \neq p \neq 25\%$$

2- مستوى الدلالة $\alpha = 0,05$.

3 - تحديد قيمة $z_{\frac{\alpha}{2}}$ لأن الاختبار من اتجاهين ، وعليه $z_{\frac{0,05}{2}} = 1,96$.

4 - قيمة الاختبار : $Z = \frac{|p' - p|}{\sqrt{\frac{p.q}{n}}} = \frac{|30 - 25|}{\sqrt{\frac{(25).(75)}{230}}} = 1,75$.

5 - المقارنة واتخاذ القرار : نلاحظ بأنَّ $Z = 1,75 \prec z_{\frac{0,05}{2}} = 1,96$ لذا فإننا نقبل الفرضية H_0 والتي تدعو إلى أنَّ الحملة الدعائية لهذه الشركة لم تؤثر في سلوك المستهلكين لهذا النوع من الشاي وذلك عند مستوى دلالة قدره 5 % أو بدرجة ثِّقة 95 % .

مثال (7 – 11) : تعاقدت إحدى المؤسسات التجارية على استيراد نوع من آلات الطباعة ، وقد حددت مواصفات تقضي بأن لا يزيد عدد القطع التي لا تحقق هذه المواصفات في الشحنة عن 3 % ، وقد وردت إليها شحنة تحتوي على مئة ألف قطعة من هذه الآلات، أُخذت منها عينة عشوائية من 125 قطعة ، ولدى إجراء الكشف عليها تبين أنَّ 6 % لم تكن محققة للمواصفات المطلوبة ، والمطلوب : هل يتم قبول الشحنة أم ترفض ؟

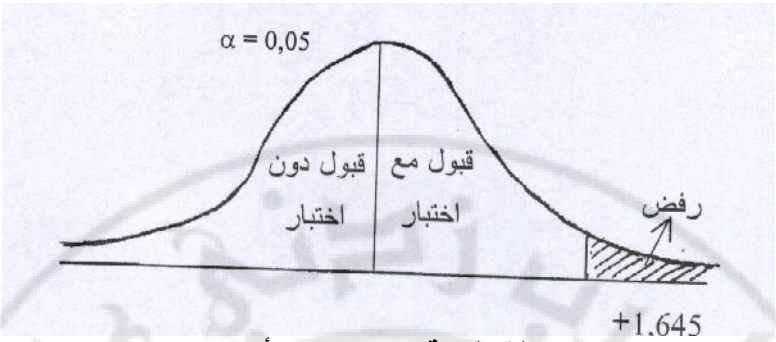
الحل :

1 - الفرضية :

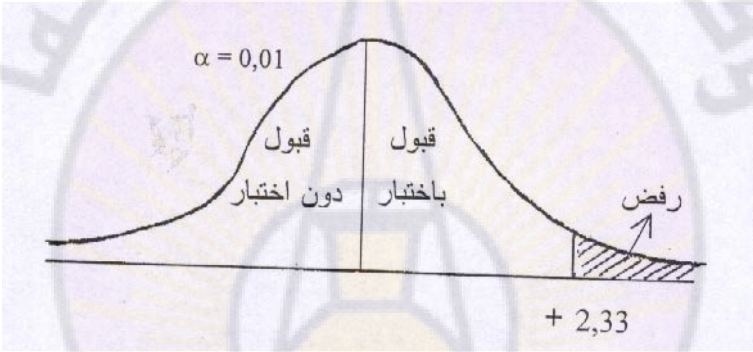
$$H_0 \rightarrow p' = p = 3\%$$
$$H_1 \rightarrow p' \succ p \succ 3\%$$

2 - مستوى الدلالة لم يعطَ هنا ، لذا نأخذ مرة $\alpha = 0,01$ ومرة أخرى $\alpha = 0,05$.

3 - تحديد قيمة z_α النظرية المقابلة لمستويات الدلالة المتخذة 0,01 و 0,05 ، والاختبار من اتجاه واحد ومن اليمين $z_{0,01} = 2,33$ و $z_{0,05} = 1,645$ والشكل رقم (7 – 7 – أ) و (7- 7 – ب) يوضحان مناطق الرفض والقبول كالآتي :



الشكل رقم (7 – 7 – أ)



الشكل رقم (7 – 7 – ب)

4 - قيمة الاختبار : $Z = \dfrac{|p' - p|}{\sqrt{\dfrac{p.q}{n}}} = \dfrac{|3-6|}{\sqrt{\dfrac{(3).(97)}{125}}} = |-1{,}97|$

5 - المقارنة واتخاذ القرار :

- عند مستوى دلالة 5 % : $Z = 1{,}97 \succ z_{0,05} = 1{,}645$ نرفض الفرضية H_0 ونرفض الشحنة .
- عند مستوى دلالة 1 % نجد : $Z = 1{,}97 \prec z_{0,01} = 2{,}33$ نقبل الفرضية H_0 وتقبل الشحنة ، غير أنَّ القرار هنا غير حاسم ، لذا ينصح بسحب عينة عشوائية جديدة وإعادة إجراء الاختبار إلى أن نصل إلى قرار حاسم من حيث القبول أو الرفض .

7-2-2- اختبارات تتعلق بمجتمعين : وهي تلك الاختبارات التي تتعلق بمقارنة تابع إحصائي مع تابع إحصائي آخر مقابل له ، أي أننا نكون أمام عينتين من حجم n_1 و n_2 من المحتمل أنهما سحبتا عشوائياً من مجتمع إحصائي واحد أو مجتمعين إحصائيين متماثلين من حيث ثوابتهما الإحصائية وحسبت من كل منهم التوابع الإحصائية اللازمة ، بحيث نطبق هنا توزيع معاينة الفروق بين التوابع الإحصائية التي تتوزع توزيعاً طبيعياً حول وسط حسابي يساوي الصفر ، وعليه تكون قيمة الاختبار Z وفقاً للتوابع الإحصائية المراد اختبارها كالآتي :

1 - **اختبارات الأوساط الحسابية** :

$$Z = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - 0}{\sigma_{\overline{x}_1 - \overline{x}2}} = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

وبما أنَّ $\mu_1 = \mu_2$ يكون $\mu_1 - \mu_2 = 0$ وحيث إن :

$\overline{X}_1$ الوسط الحسابي للعينة (1) ؛

$\overline{X}_2$ الوسط الحسابي للعينة (2) ؛

$\sigma_{\overline{x}_1 - \overline{x}2}$ الخطأ المعياري لفروق الأوساط الحسابية ، ونستخدم الانحرافات المعيارية للعينات

S_1 و S_2 كتقدير لـ σ_1 و σ_2 في حال كانت غير معلومة .

2 - **اختبارات الانحرافات المعيارية** :

$$Z = \frac{\left|S_1 - S_2\right| - 0}{\sigma_{\sigma_1 - \sigma 2}} = \frac{\left|S_1 - S_2\right| - 0}{\sqrt{\frac{\sigma_1^2}{2.n_1} + \frac{\sigma_2^2}{2.n_1}}}$$

حيث إنَّ :

$\sigma_1 - \sigma_2 = 0$ لأن $\sigma_1 = \sigma_2$ ؛

S_1 الانحراف المعياري للعينة (1) ؛

S_2 الانحراف المعياري للعينة (2) ؛

$\sigma_{\sigma-\sigma 2_1}$ الخطأ المعياري لفروق الانحرافات المعيارية ، ونستخدم الانحرافات المعيارية للعينات

S_1 و S_2 كتقدير لـ σ_1 و σ_2 في حال كانت غير معلومة .

ويستخدم هذا الاختبار للتأكد من تماثل شكل توزيع العينة (1) مع شكل توزيع العينة (2) .

3- **اختبارات النسب المئوية** :

$$Z = \frac{|p'_1 - p'_2| - 0}{\sigma_p} = \frac{|p'_1 - p'_2| - 0}{\sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}}$$

حيث إن :

$p_1 - p_2 = 0$ ؛

p'_1 نسبة تحقق الظاهرة في العينة (1) ؛

p'_2 نسبة تحقق الظاهرة في العينة (2) ؛

σ_p الخطأ المعياري لفروق النسب المئوية ، ويمكن الاستعانة بالعينات في حال كون p_1 و p_2 غير معلومة .

ويلاحظ عند إجراء الاختبار في أي حالة من الحالات الثلاث التي أشرنا إليها قبل قليل ، أنه في حال كان الفرق حقيقياً بين التابع الإحصائي المراد مقارنته مع التابع الإحصائي الآخر ، فإننا نكون أمام أمرين هما :

- إذا كان معلوماً بشكل قاطع أنَّ العينتين قد أُخذتا من مجتمع إحصائي واحد ، فإن الفرق يعود إلى أسلوب السحب الذي لم يكن عشوائياً سواءً بالنسبة لإحدى العينتين أو كليهما .
- أما إذا لم يكن معلوماً بشكل قاطع وأنَّ العينتين قد أُخذتا من مجتمع إحصائي واحد ، فإن الفرق يعود إلى أن العينتين قد سحبتا من مجتمعين إحصائيين مختلفين بثوابتهما الإحصائية .

مثال (7 – 12) : أُعطى اختبار لفصلين يتكون الأول من 40 طالباً والثاني من 50 طالباً، حيث كان متوسط العلامات في الفصل الأول74 وبانحراف معياري 8، وكان متوسط علامات الفصل الثاني 78 وبانحراف معياري 7 ، والمطلوب : هل

يختلف متوسط علامات الفصل الأول عنه في الفصل الثاني عند احتمال قدره 95 % ؟

الحل :

لنفرض بأن الفصلين مسحوبان من مجتمعين متوسطاتهما هي μ_1 و μ_2 وعليه نقرر بين فرضين هما : $H_0 \rightarrow \mu_1 = \mu_2$ والاختلاف يرجع للحظ والصدف .

$H_1 \rightarrow \mu_1 \neq \mu_2$ والاختلاف حقيقي (جوهري) .

- مستوى الدلالة $\alpha = 1 - 095 = 0,05$ والاختبار من اتجاهين فتكون : $z_{\frac{\alpha}{2}} = z_{\frac{0,05}{2}} = 1,96$.

- قيمة الاختبار : $Z = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}} = \frac{|74 - 78| - 0}{\sqrt{\frac{(8)^2}{40} + \frac{(7)^2}{50}}} = |-2,49|$

هنا استخدمنا الانحرافات المعيارية للعينات كتقدير لـ σ_1 و σ_2 .

- المقارنة واتخاذ القرار : نلاحظ بأن $Z = 2,49 \succ z_{\frac{0,05}{2}} = 1,96$ فإننا نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 ، أي أنَّ هناك اختلافاً حقيقياً بين متوسطات علامات الفصل الأول عنه في الفصل الثاني ، وهذا يعود إلى أن العينتين قد سحبتا من مجتمعين إحصائيين مختلفين بثوابتهما الإحصائية .

مثال (7 – 13) : إذا كان متوسط أوزان 60 طالباً من المشاركين في النشاط الرياضي في قسم المكتبات هو 70,5 كغ وبانحراف معياري 2,5 كغ ، بينما كان متوسط وزن 70 طالباً لم يظهروا اهتماماً بالمشاركة في النشاط الرياضي في الكلية نفسها هو 69,8 كغ وبانحراف معياري قدره 2,6 كغ ، والمطلوب : هل تعتقد بأنّ الطلبة الذين يساهمون في النشاط الرياضي أثقل وزناً من غيرهم في قسم المكتبات ؟

الحل :

- الفرضية : $H_0 \rightarrow \overline{X}_1 = \overline{X}_2$ لا يوجد اختلاف حقيقي بين متوسط أوزان الطلبة ، في حين أنَّ $H_1 \rightarrow \overline{X}_1 \succ \overline{X}_2$ يوجد اختلاف حقيقي ، أي أنَّ متوسط أوزان المساهمين بالنشاط الرياضي أكبر منه لدى غير المساهمين بالنشاط الرياضي .

- مستوى الدلالة عند 0,05 و 0,01 والاختبار من اتجاه واحد ومن اليمين ، فتكون : $z_{0,05} = 1,645$ و $z_{0,01} = 2,33$.

- قيمة الاختبار : $$Z = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}} = \frac{\left|70,5 - 69,8\right| - 0}{\sqrt{\frac{(2,5)^2}{60} + \frac{(2,6)^2}{70}}} = \left|1,562\right|$$

- المقارنة واتخاذ القرار :

نلاحظ بأنَّ $Z = 1,562 \prec z_{0,01} = 2,33$ و $Z = 1,562 \prec z_{0,05} = 1,645$

وبالتالي نقبل الفرضية H_0 ولا يوجد فرق حقيقي بين متوسط أوزان الطلبة و إن الفرق المشاهد يعود لأخطاء الحظ والصدف .

مثال (7 – 14) : في دراسة حول متوسط الدخل الشهري لإحدى ا لمطابع ، أخذت عينة عشوائية من 600 عامل من مطبعة تابعة لتلك المطبعة الأم ، فوجد أن متوسط الدخل الشهري للعامل مساوي لـ 5950 ل.س وبانحراف معياري قدره 750 ل.س ، كما أخذ ت عينة عشوائية من 400 عامل من م طبعة أخرى تابعة المطبعة الأم نفسها ، فوجد أن متوسط الدخل الشهري للعامل فيها مساوي لـ 5250 ل.س وبانحراف معياري قدره 700 ل.س ، والمطلوب :

1 - هل تعتقد باحتمال قدره % 99 ، بأن متوسط الدخل الشهري في المطبعة الأولى يختلف عنه في المطبعة الأخرى ؟

2 - هل تعتقد باحتمال قدره % 95 ، بأن هناك تماثلاً في الدخل الشهري للمطبعة الأولى والمطبعة الأخرى ؟

الحل :

- طـ1- الفرضية : $H_0 \rightarrow \overline{X}_1 = \overline{X}_2$ لا يوجد اختلاف حقيقي بين متوسط الدخل الشهري في المطبعة الأولى عنه في المطبعة الأخرى مقابل وجوده $H_1 \rightarrow \overline{X}_1 \neq \overline{X}_2$

- مستوى الدلالة $\alpha = 1 - 0,99 = 0,01$ والاختبار من اتجاهين ، فتكون : $z_{\frac{\alpha}{2}} = z_{\frac{0,01}{2}} = 2,58$.

- قيمة الاختبار : $$Z = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}} = \frac{\left|5950 - 5250\right| - 0}{\sqrt{\frac{(750)^2}{600} + \frac{(700)^2}{400}}} = \left|15,05\right|$$

- المقارنة واتخاذ القرار :

نلاحظ بأنَّ $Z = 15,05 \prec z_{0,005} = 2,58$ وبالتالي نرفض الفرضية H_0 ونقبل الفرضية H_1 ، أي أن هناك اختلافاً حقيقياً بين متوسط الدخل الشهري في المطبعة الأولى عنه في المطبعة الأخرى .

ط2 - الفرضية : $H_0 \to S_1 = S_2$ يوجد تماثل في الدخل الشهري في المطبعة الأولى والمطبعة الأخرى ، مقابل عدم وجود تماثل فيما بينهما $H_1 \to S_1 \neq S_2$.

- مستوى الدلالة $\alpha = 1 - 0,95 = 0,05$ والاختبار من اتجاهين فيكون : $z_{\frac{\alpha}{2}} = z_{\frac{0,05}{2}} = 1,96$.

- قيمة الاختبار : $$Z = \frac{|S_1 - S_2| - 0}{\sqrt{\frac{S_1^2}{2.n_1} + \frac{S_1^2}{2.n_2}}} = \frac{|750 - 700| - 0}{\sqrt{\frac{(750)^2}{2*600} + \frac{(700)^2}{2*400}}} = |1,52|$$

- المقارنة واتخاذ القرار :

نلاحظ بأنَّ $Z = 1,52 \prec z_{0,025} = 1,96$ وبالتالي نقبل الفرضية H_0 ، أي هناك تماثل في توزيع الدخل الشهري للمطبعة الأولى معه في المطبعة الأخرى .

مثال (7 – 15) : عند دعوة الشباب إلى الخدمة الإلزامية فإنهم يتعرضون لفحص طبي لبيان مدى لياقتهم الصحية ، وبناءً على خبرات سابقة تبين أن نسبة الشباب من المدن الذين هم غير لائقين صحياً يساوي تقريباً نسبة الشباب من الريف الذين هم غير لائقين صحياً . افترض أن اللجنة الطبِّية قد قررت إعطاء فترة شهر لمعالجة الشباب الذين تم استدعاؤهم للخدمة الإلزامية على أمل تقليل الحالات المرفوضة ، ولقد تم استدعاء 3200 شاب من المدن و 4500 شاب من الريف ، وبعد مُضي شهر من الاستدعاء والمعالجة لمن هم بحاجة إلى العلاج رُفض لأسباب صحية 1056 شاباً من المدن و 1215 شاباً من الريف ، والمطلوب : هل تعتقد باحتمال قدره 99 % أن نسبة الشباب المرفوضين لأسباب صحية في المدن تختلف بصورة حقيقية عن نسبتهم في الريف ؟

الحل :

نرمز بـ p_1' لنسبة الشباب المرفوضين لأسباب صحية في المدن :

$$p_1' = \frac{1056}{3200} = 33\% \quad ؛$$

وبـ p'_2 لنسبة الشباب المرفوضين لأسباب صحية في الريف :

$p'_2 = \frac{1215}{4500} = 27\%$.

- الفرضية : $H_0 \rightarrow p'_1 = p'_2$ لا يوجد اختلاف حقيقي بين النسبتين مقابل وجوده $H_1 \rightarrow p'_1 \neq p'_2$.
- مستوى الدلالة $\alpha = 1 - 0,99 = 0,01$ والاختبار من اتجاهين فتكون : $z_{\frac{\alpha}{2}} = z_{\frac{0,05}{2}} = 2,58$.
- قيمة الاختبار :

$Z = \frac{\left|p'_1 - p'_2\right| - 0}{\sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}} = \frac{\left|33 - 27\right|}{\sqrt{\frac{(29,5).(70,5)}{3200} + \frac{(29,5).(70,5)}{4500}}} = 5,69$ وذلك

لأن : $p = \frac{1056 + 1215}{3200 + 4500} = 29,5\%$.

- المقارنة واتخاذ القرار : نلاحظ بأن $Z = 5,69 \succ z_{\frac{0,05}{2}} = 2,58$ وبالتالي نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 ، أي هناك اختلاف حقيقي في نسبة الشباب المرفوضين لأسباب صحية في المدن عنها في الريف .

3-7- مسائل وتمارين غير محلولة

1 - ماذا يعالج الاستدلال الإحصائي ؟
2 - ما هو التقدير الإحصائي وما هي أنواعه ؟
3 - ما الفرق بين التقدير النقطي والتقدير المجالي ؟
4 - بماذا يتصف التقدير الجيد ؟
5 - ما المقصود بحديّ الثِّقة ؟
6 - ما هو الاختبار ؟
7 - ما هي الفرضية الإحصائية ؟
8 - ما هي الأخطاء التي تصاحب عملية اتخاذ القرار ؟
9 - ما الفرق بين الخطأ من النوع الأول والخطأ من النوع الثاني ؟
10 - كيف يتم تحديد مستوى الدلالة ، وماذا يعني مستوى الدلالة ؟

11 - ما هي خطوات اختبار فرضية ما ؟
12 - عرّف كلا من : منطقة الرفض ، منطقة القبول ، المنطقة الحرجة ، القيم الحرجة .
13 - ما الفرق بين الاختبار من اتجاهين والاختبار من اتجاه واحد ؟ ثُّم بيّن الحالات التي يكون فيها الاختبار من اتجاهين أو من اتجاه واحد ؟
14 - ما هي أنواع الاختبارات الإحصائية ؟
15 - ما الفرق بين الاختبارات المتعلقة بمجتمع واحد والاختبارات المتعلقة بمجتمعين ؟
16 - لماذا يختلف توزيع معاينة الاختبارات المتعلقة بمجتمع واحد عن توزيع معاينة الاختبارات المتعلقة بمجتمعين ؟
17 - ما هي القيمة المختبر حولها في الاختبارات المتعلقة بمجتمعين ؟
18 - ماذا يعني رفض الفرضية H_0 في الاختبارات التي تتعلق بمجتمع واحد ؟ وماذا يعني رفضها أيضاً في الاختبارات التي تتعلق بمجتمعين ؟
19 - تبين من دراسة شاملة أجريت على كافة عمال إحدى المطابع في سورية ، أن متوسط الأجر الشهري للعامل الواحد في هذه المطبعة يبلغ 3600 ل.س والانحراف المعياري 208 ل.س ، وقد رغبت وزارة الثقافة بعد فترة من إجراء الدراسة معرفة نسبة العمال الأميين في هذه المطبعة ، فقامت بسحب عينتين ، الأولى من عمال دمشق والثانية من عمال حلب فحصلت على المعلومات الآتية :

اسم العينة	حجمها	متوسط الأجر الشهري	الانحراف المعياري	نسبة العمال الأميين
دمشق	169	3576	100	8 %
حلب	169	3504	110	17 %

<u>والمطلوب</u> :
1 - هل تمثل هذه العينات أصدق تمثيل المجتمع الإحصائي الذي سحبت منه ، إذا كان مستوى الدلالة
المعتمد 5 % ؟
2 - أوجد باحتمال 99 % نسبة الأميين في هذه الصناعة اعتماداً على المعلومات التي حصلنا عليها من عينة دمشق أولاً ثُّم من عينة حلب ؟

3 -دع جانباً النتائج التي حصلت عليها من الفقرات السابقة ، فهل تعتقد أن نسبة العمال الأميين في مدينة دمشق مختلفة جوهرياً عن نسبة العمال الأميين في مدينة حلب معتمداً % 10 كمستوى دلالة ؟

4 -ما هي القيمة التي اختبرت حولها الفرق بين التابعين الإحصائيين موضوع الطلب السابق ، وماذا تمثل هذه القيمة ؟

21 – بيّنت الدراسات الشاملة التي أجرتها وزارة الثقافة في المطابع التابعة لها ، أن متوسط الأجر الشهري للعامل الواحد يبلغ 3300 ل.س وبانحراف معياري 700 ل.س ، وللتأكد من ذلك سحبت عينت ان من عمال المطابع التابعة لها ، فأعطت الآتي :

$$n_1 = 120; \overline{X}_1 = 3200; S_1 = 750; n_2 = 100; \overline{X}_2 = 3170; S_1 810$$

والمطلوب :

1 -أيَّ العينتين تمثل المجتمع الإحصائي الذي سحبت منه أصدق تمثيل ، مدّعماً رأيك بالحسابات اللازمة ؟ ثُم بينّ معنى العينة التي حصلت عليها واذكر أنواعها ؟

2 -ما هي أسباب ا لاختلاف بين التابع الإحصائي والثابت الإحصائي ؟

3 -في ضوء نتائج العينة (1) :

أ- قدّر باحتمال % 95,5 متوسط الأجر الشهري للعامل في تلك المطابع ؟

ب- أوجد نسبة العمال الذين يقل أجرهم الشهري عن 1700 ل.س ؟

ج- أوجد عدد العمال الذين يتراوح أجرهم الشهري بين 1470 و 3950 ل.س .

4- ما الفرق بين الخطأ المعياري للوسط الحسابي والخطأ المعياري للانحراف المعياري ، وما المقصود بكل منهما ؟

5- هل تعتقد أن هناك تماثلاً في التوزيعات التكرارية لأجور العمال لهاتين المطبعتين ، ثُم بينّ القيمة المختبر حولها وماذا تمثل ؟

6- لقد كان من المرغوب به تحديد النسبة الحقيقية للعمال الأميين في المطبعة (2) ، وقد قدرت هذه النسبة بناءً على خبرات سابقة بأنها لا تتجاوز % 10 ، فهل كان حجم العينة المحسوبة أعلاه كافٍ لتقدير النسبة الحقيقية للعمال الأميين ، على أن لا يزيد الخطأ في القدير عن % 4 لكل الحالات العملية.

وقل رب زدني علما
جامعة دمشق
Damascus University

الفصل الثامن

الاستدلال الإحصائي للعينات صغيرة الحجم

8-1- توزيع t – ستودنت
8-2- اختبار الفرضيات
8-2-1- اختبار الفرضيات للمتوسط (μ) [حالة مجتمع واحد]
8-2-2- اختبار الفرضيات للمتوسط (μ) [حالة مجتمعين]
8-2-3- تقدير متوسط المجتمع (μ)
8-2-4- اختبار الفرق بين متوسطين (حالة العينات غير المستقلة)
8-3- توزيع كاي مربع χ^2
8-3-1- اختبار الفرضيات للتباين σ^2
8-3-2- تقدير تباين المجتمع σ^2
4-8 توزيع فيشر (F) واختبار نسبة تباينين مجتمعين طبيعيين
8-5- أسئلة وتمارين

وقل رب زدني علما
جامعة دمشق
Damascus University

مقدمة:

نحن نعلم أن توزيع المعاينة لتقدير μ يتبع التوزيع الطبيعي في إحدى حالتين:

. عندما يكون حجم العينة المسحوبة والمدروسة $n \geq 30$ حتى ولو لم يكن المجتمع خاضعاً للتوزيع الطبيعي.

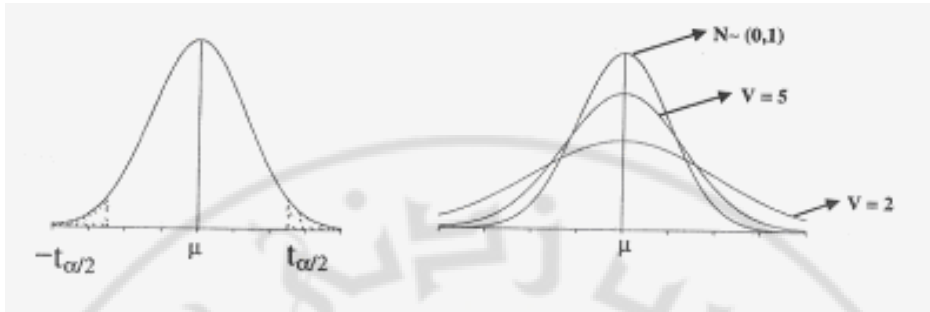
. عندما يكون المجتمع المدروس والتي سحبت منه العينة خاضعاً أساساً للتوزيع الطبيعي، وأن يكون الانحراف المعياري للمجتمع σ معلوماً حتى لو كانت ($n < 30$) مفردة.

وفي الحالة الثانية إذا كان σمجهولاً فإن توزيع المعاين ة يتبع توزيع آخر يدعى توزيع ستودنت وهي الحالة التي يكون فيها للمجتمع شكل التوزيع الطبيعي وحجم العينة صغيراً ($n <30$) والانحراف المعياري للمجتمع σ مجهولاً.

8-1 – توزع (t) ستودنت:

وهو عبارة عن توزيع احتمالي مستمر يستخدم في العينات صغيرة الحجم ($n<30$) ويعتمد عليه بشكل أساسي في الاختبارات والتقدير الإحصائي، والتوزيع (t) يشبه التوزيع الطبيعي من حيث التماثل حول الصفر، لكنه أكثر انبساطاً منه، ويقترب من التوزيع الطبيعي القياسي عندما $n \geq 30$ ، أي عندما يزداد تغيير عدد درجات الحرية والتي هي عبارة عن: ($v = n-1$).

وكلما كانت درجات الحرية قليلة فإن المنحني يكون أكثر تفلطحاً وتشتتاً بالمقارنة مع التوزيع الطبيعي، وهناك جداول خاصة لقيم (t) لاحتمالات مختلفة وفق عدد درجات الحرية في التوزيع، ويمكن اعتبار الانحراف المعياري ($S = \sigma$) تقدير غير متحيز.



وأن أهم تطبيقات توزيع ستودنت (t) هي إمكانية استخدامه كتوزيع معاينة للعينات الصغيرة المسحوبة من مجتمع خاضعٍ للتوزيع الطبيعي وانحرافه المعياري σ مجهولاً ويعطى تقديره بواسطة s ، وإن المتحول العشوائي الخاضع لهذا التوزيع يعطى في هذه الحالة بالعلاقة:

$$t = \frac{\left|\overline{X} - \mu\right|}{s_x / \sqrt{n}} = \frac{\left|\overline{X} - \mu\right|}{\frac{S_x}{\sqrt{n}}}$$

8-2- اختبار الفرضيات

8-2-1- اختبار الفرضيات للمتوسط (μ) (حالة مجتمع واحد)

يقوم هذا الاختبار على مقارنة متوسط المجتمع μ (أو الثابت الإحصائي) مع الإحصاء $\overline{X}$ (أي التابع الإحصائي) المسحوب من عينة عشوائية (N < 30) ، علماً بأن σ للمجتمع غير معلوم ، والهدف من ذلك بيان فيما إذا كان الفرق بين μ و ($\overline{X}$) معنوياً أم لا عند درجات حرية (v) معينة ويتم ذلك وفق خطوات الاختبار:

1 . نقوم صياغة الفرضيتين :

$H_0 : \mu = \overline{X}$

$H_1 : \mu \neq \overline{X}$ والاختبارمن اتجاهين

$H_1: \mu > \overline{X}$ أو $\mu < \overline{X}$ والاختبار من اتجاه واحد

$$t = \frac{|X - \mu|}{s_x / \sqrt{n}} = \frac{|X - \mu|}{\frac{S_x}{\sqrt{n}}}$$

3 . اتخاذ القرار: نقارن قيمة (t) المحسوبة من العلاقة السابقة مع القيمة الجدولية $t_{\frac{\alpha}{2}}$ لاختبار من اتجاهين أو مع القيمة الجدولية t_α لاختبار من اتجاه واحد ، وذلك عند درجات من الحرية v = n -1 ، وبالتالي نحصل على منطقة الرفض ، وعليه نجد من خلال المقارنة بين ($t_{\frac{\alpha}{2}}$) و(t) :

. إما أن تكون $t < t_{(\alpha/2,v=n-1)}$ (المحسوبة) وعندها نقبل فرضية العدم ، وفي هذه الحالة تكون (t) قد وقعت ضمن منطقة القبول والفرق بين (μ) و($\overline{X}$) يكون غير معنوي أي أن $\mu = \overline{X}$.

– وإما أن تكون $t > t_{(\alpha/2, = n-1)}$ وعندها نرفض فرضية العدم H_0 ، وبالتالي نستنتج أن (t) وقعت ضمن منطقة الرفض ، لذلك نرفض فرضية العدم ونقبل الفرضية البديلة (H_1) ونقرر بأن الفرق بين (μ) و($\overline{X}$) معنوي أي أن $\mu \neq \overline{X}$.

تمرين (1): صممت إحدى الآلات لتعطي (80) غ من القهوة عند وضع قطعة النقود الملائمة، ولمعرفة ما إذا كانت الآلة تعمل كما يجب، سحبت عينة عشوائية

مكونة من (16) فنجاناً من القهوة لقياس القهوة بكل منها. فإذا علمت أن الوسط الحسابي للعينة وانحرافها المعياري هما $\overline{X}$ = 75 غرام وانحرافها المعياري S = 8 غرام ، اختبر الفرض القائل أن الآلة تعمل كما يجب (μ = 80) في مقابل الفرض البديل بأن الآلة لا تعمل كما يجب عند مستوى دلالة 0.01 = α مفترضاً أن التوزيع الأصلي للبيانات طبيعي وأن تباينه σ^2 مجهول .

1 . صياغة الفرضية: $H_0 : \overline{X} = \mu = 8$

$$H_1 : \overline{X} \neq \mu \neq 8$$

2 . الاختبار الإحصائي

$$t = \frac{\left|\bar{x} - \mu\right|}{s_x / \sqrt{n}} = \frac{|75 - 80|}{8/\sqrt{16}} = 2.5$$

3 . اتخاذ القرار: من جداول (t) ثنائية الجانب وعند عدد من درجات الحرية (v = n-1) أي v =15 وعند مستوى دلالة (معنوية) α= 0.01 نجد أن:

$t_{(\alpha/2,v)} = t_{(0.01/2,15)} = 2.947$ ومن خلال المقارنة نجد أن:

$t = 2.5 < t = 2.947$ وبالتالي نقبل فرضية العدم H_0 وأن الآلة تعمل كما يجب.

تمرين (2): يعتقد أحد بائعي آلات العرض أن متوسط ربحه في الآلة الواحدة يقل عن (500) وحدة نقدية، لارتفاع نسبة الخصم التي يقدمها لعملائه، ولمعرفة مدى صحة هذا الاعتقاد سحبت عينة عشوائية بحجم (25) فاتورة بيع سابقة فوجد أن متوسط الربح في العينة هو (485) وحدة نقدية بانحراف معياري (S = 45) وحدة نقدية، هل تؤيد هذه البيانات الفرض القائل بأن متوسط الربح يقل عن (500) وحدة نقدية وذلك عند مستوى معنوية (α = 0.05) وذلك بافتراض أن أرباح الآلات تتبع التوزيع الطبيعي.

الحل:

1 . نقوم بصياغة الفرضيتين كما يلي : $H_0 : \mu = 500$

$H_1 : \mu < 500$ وهو اختبار أحادي الجانب

2 . الاختبار الإحصائي:

$$t = \frac{\left|\bar{x} - \mu\right|}{s_x / \sqrt{n}} = \frac{|485 - 500|}{45 / \sqrt{25}} = 1.66$$

3 . اتخاذ القرار: نبحث في جداول ستودنت عن قيمة (t) أحادية الجانب المقابلة لمستوى دلالة $\alpha = 0.05$ ولدرجة حرية $v = n-1 = 24$ ، تساوي :

$t_{(\alpha, v)} = t_{(0.05,24)} = 1.170$

من خلال المقارنة نجد $t_{(\alpha , v)} = 1.710 > t = 1.66$ وبالتالي نقبل H_0 ونقول إن الاعتقاد كان خاطئاً بأن متوسط الربح يقل عن 500 وحدة نقدية.

8-2-2- اختبار الفروض للفرق بين متوسطين (حالة مجتمعين)

يقوم هذا الاختبار على مقارنة الفرق بين إحصائيتين محسوبتين من مجتمعين مستقلين سحبت منهما وبشكل عشوائي ، عينتان مستقلتان بحجمين $(n_1 , n_2) < 30$ ، وضمن شروط تساوي التباينات في هذين المجتمعين وغير المعلومين أي ($\sigma_1^2 = \sigma_2^2$) ، وذلك على النحو التالي:

1 . صياغة الفرضية: $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

2 . الاختبار الإحصائي

$$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - d_0}{\frac{S_p}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

حيث إن : $d_0 = \mu_1 - \mu_2 = 0$

وَ

$$S_p = \sqrt{\frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

3 . اتخاذ القرار: نبحث في جداول ستودنت عن قيمة (t) الجدولية حسب نوع الاختبار أحادي الجانب أو ثنائي الجانب عن: $t_{(\alpha\ ,\ v)}$ أو $t_{(\alpha\ /2,\ v)}$ ، والمقابلة لدرجة حرية $v = n-2$ ، ومن خلال المقارنة بين $t_{(\alpha\ ,\ v)}$ و(t) أي الجدولية والمحسوبة نحدد قبول أو رفض فرضية العدم H_0.

تمرين: يرغب مدير تدريب الأفراد في إحدى المكتبات العامة معرفة إذا كان لطريقتين مختلفتين في التدريب نفس التأثير، لذلك سحبت عينتان مستقلتان من العاملين في المكتبة عشوائياً وتم تدريب كل منهما بطريقة من الطريقتين السابقتين وبعد انتهاء مدة التدريب أجري لهما امتحان موحد لقياس درجة الكفاءة للأفراد فحصلنا على البيانات التالية:

العينة الأولى: $s_1^2 = 36$ ، $\overline{X}_1 = 85$ ، $n_1 = 18$

العينة الثانية: $S_2^2 = 36$ ، $\overline{X}_2 = 80$ ، $n_2 = 12$

وبافتراض أن درجات اختبار جميع العاملين تتبع التوزيع الطبيعي، وأن التباينات متساوية ($\sigma_1^2 = \sigma_2^2$)

المطلوب: اختبر الفرض القائل بأن هاتين الطريقتين بالتدريب متساويتان عند مستوى دلالة $\alpha = 0.05$

الحل:

1 . صياغة الفرضية:

$$H_0 = \mu_1 = \mu_2$$
$$H_1 = \mu_1 \neq \mu_2$$

2 . الاختبار الإحصائي :

$$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right| - d_0}{\frac{S_p}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$Sp = \sqrt{\frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}} = \sqrt{\frac{36(18-1) + 34(12-1)}{18+12-2}} = 5.93$$

$$t = \frac{\left|(85-80) - 0\right|}{(5.93)/\sqrt{1/18 + 1/12}} = 2.27$$

3 . اتخاذ القرار: من جداول (t) ومن أجل اختبار ثنائي الجانب وعند مستوى دلالة $\alpha = 0.05$ وعدد درجات حرية $v = n1 + n2 - 2 = 18 + 12 - 2 = 28$ ، فنجد أن:

$$t_{\left(\frac{\alpha}{2}, v\right)} = t_{\left(\frac{0.05}{2}, 28\right)} = 2.048$$

ومن خلال المقارنة نجد أن $t = 2.27 > t_{\left(\frac{0.05}{2}, 28\right)} = 2.048$ ، وبالتالي نفرض H_0 ونقول إن هناك فرقاً بين طريقتي التدريب.

8-2-3- تقدير متوسط المجتمع (μ):

إذا سحبنا عينة عشوائية (n <30) من مجتمع إحصائي ما فإنه يمكن إجراء تقدير لمتوسط المجتمع (μ) بدلالة ($\overline{X}$) للعينة المسحوبة عشوائياً وذلك على النحو التالي:

$$\Pr\left[\overline{X} - t_{\left(\frac{\alpha}{2},n-1\right)}.\frac{S_x}{\sqrt{n}}\right] \le \mu \le \left[\overline{X} + t_{\left(\frac{\alpha}{2},n-1\right)}.\frac{S_x}{\sqrt{n}}\right] = 1-\alpha$$

وبالتالي يمكن القول بأنه باحتمال معين (α – 1) فإن متوسط المجتمع يقع ضمن المجال السابق.

تمرين: لقياس السرعة القصوى لسيارة جديدة، أجرينا (16) تجربة فحصلنا على متوسط السرعة القصوى $\overline{X}$ = 200 K بالساعة وانحراف معياري S = 10K بالساعة والمطلوب: أوجد مجال الثقة لمتوسط السرعة القصوى لهذا النوع الجديد من السيارات وذلك باحتمال (0.99).

الحل:

$$\Pr\left[\overline{X} - t_{\left(\frac{\alpha}{2},n-1\right)}.\frac{S_x}{\sqrt{n}}\right] \le \mu \le \left[\overline{X} + t_{\left(\frac{\alpha}{2},n-1\right)}.\frac{S_x}{\sqrt{n}}\right] = 1-\alpha$$

لدينا الاحتمال 0.99 = α -1 ومنه α = 0.01 نبحث في جداول (t) الموضوعة لاختبار ثنائي الجانب عند مستوى معنوية 0.01 وعند درجات حرية 15 = 1- n.

$$t_{\left(\frac{\alpha}{2},n-1\right)} = t_{\left(\frac{0.01}{2},15\right)} = 2.946$$

ومنه $$\Pr(200-(2.946)(\frac{10}{\sqrt{16}})) \le \mu \le (200+(2.946)(\frac{10}{\sqrt{16}})) = 0.99$$

$$\Pr(192.63 \le \mu \le 207.3) = 0.99$$

وبالتالي يمكن القول إنه باحتمال 99% فإن متوسط السرعة لهذا النوع من السيارات لن يقل عن 192.63 كم بالساعة ولن يزيد عن 207.3 كم بالساعة.

8-2-4- اختبار الفرق بين متوسطين (حالة العينات غير المستقلة)

في هذه الحالة نجد البيانات في كل زوج من هذه الأزواج تكون غير مستقلة، وأن تباين الفرق بين متوسطي العينتين لا يساوي مجموع تبايني المتوسطين، وإنما تستخدم الفروق بين المشاهدات (D) ، ثم حساب قيمة الاختبار الإحصائي التالي :

$$T = \frac{\overline{D}}{S_{\overline{D}}}$$

حيث إن : $D = x_i - y_i$ ، وإن : $\overline{D} = \frac{\sum D}{n}$ متوسط الفروق

$S_{\overline{D}} = \frac{S_D}{\sqrt{n}}$ الخطأ المعياري المقدر لمتوسط الفروق

$S_D^2 = \frac{\sum D^2 - \frac{\left(\sum D\right)^2}{n}}{n-1}$ تباين الفروق

ولنوضح هذا الاختبار من خلال التمرين التالي:

تمرين 2: يرغب المسؤول عن التعليم بإحدى المناطق التعليمية في معرفة ما إذا كان مستوى الطالب في اللغة الإنكليزية يختلف عن مستواه في اللغة العربية:

سحبت عينة عشوائية مؤلفة من (11) طالباً فحصلنا على البيانات الآتية:

والمطلوب: هل تؤدي هذه البيانات إلى قبول الفرض القائل بأن مستوى الطلبة في اللغة الإنكليزية لا يختلف عن مستواهم في اللغة العربية وذلك عند مستوى معنوية $\alpha = 0.02$. علماً أن درجات كل مادة تتبع التوزيع الطبيعي والعينات غير مستقلة.

اللغة الإنكليزية	اللغة العربية	D	D^2
88	85	3	9
93	96	−3	9
75	65	10	100
68	73	−15	25
98	88	10	100
58	65	−7	49
79	80	−1	1
83	75	8	64
88	93	−5	25
52	48	4	16
67	70	3	9
$\sum$		11	407

الحل: نحسب الفرقات D ثم نحسب مربعاتها D^2 ، فنجد من الجدول أن :

$$\overline{D} = \frac{\sum D}{n} = \frac{11}{11} = 1$$

$$S_D^2 = \frac{\sum D^2 - \frac{\left(\sum D\right)^2}{n}}{n-1} = \frac{407 - \frac{(11)^2}{11}}{11-1} = 39.6$$

$$S_{\overline{D}} = \frac{S_D}{\sqrt{n}} = \sqrt{\frac{39.6}{11}} = 1.89$$

1 . نقوم بصياغة الفرضيتين :

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

2 . الاختبار الإحصائي:

$$t = \frac{\overline{D}}{S_{\overline{D}}} = \frac{1}{1.89} = 0.52$$

3 . اتخاذ القرار: من خلال جدول t والاختبار من اتجاهين وعند مستوى دلالة (معنوية) $\alpha = 0.02$ درجات حرية (V = n-1 = 11 -1 = 10) ، نجد أن :

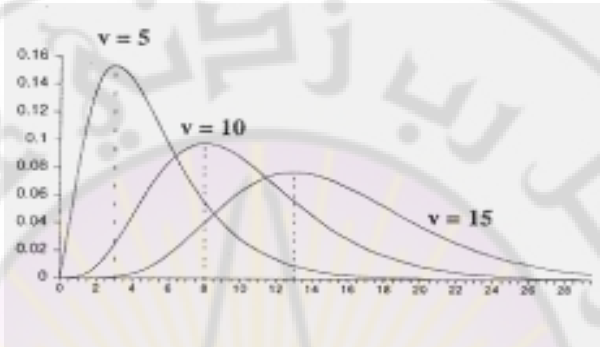
$$t = 0.52 < t_{\left(\frac{\alpha}{2}, v\right)} = 2.764$$

وبالتالي نقبل بفرضية العدم H_0 القائلة بعدم وجود فرق في مستوى الطلاب في المادتين.

8-3- توزيع كاي مربع χ^2

تعريف: إن توزيع كاي مربع χ^2 هو عبارة عن توزيع احتمالي مستمر، يستخدم في الاختبارات للعينات الصغيرة الحجم، ويعتمد بشكل أساسي على درجات الحرية في اختبار الفرق بين مجموعتين من البيانات (واقعية ونظرية) لمعرفة مدى التطابق النسبي بينهما.

ويتألف توزيع χ^2 من مجموعة من المنحنيات التي يتغير شكلها بتغير عدد درجات الحرية (V) فعندما يكون v =2 فإن منحني التوزيع يأخذ شكل الراء المقلوبة ، وبزيادة عدد درجات الحرية يصبح شكل المنحني ملتوياً نحو اليمين.



شكل (8-2)

ويقترب شكل منحني χ^2 من شكل التوزيع الطبيعي عندما يصبح عدد درجات الحرية (v > 30) ونحن نعلم أن الأوساط الحسابية للعينات الصغيرة الحجم تتوزع وفق التوزيع (t) ، أما بالنسبة للتباينات S^2 والمحسوبة من عينات صغيرة فإن شكل توزيعها الاحتمالي ملتوي نحو اليمين وتتوزع وفق كاي مربع (χ^2) ، وإن متحولها العشوائي يعطى بالعلاقة التالية:

$$\chi^2 = (n-1)\frac{S^2}{\sigma^2}$$

ويوجد جداول خاصة للقيم النظرية لكاي χ^2 مقابلة لاحتمالات مختلفة (وبالتالي مستويات معنوية مختلفة) ووفق عدد مختلف من درجات الحرية، وإذا كان عدد درجات الحرية v > 30 ، فإننا نجري التحويل $y = \sqrt{2\chi^2}$ ، فيكون خاضعاً

للتوزيع الطبيعي الذي توقعه $\sqrt{2n-1}$ وتباينه يساوي الواحد ، ويكون المتحول $Z=\left(\sqrt{2\chi^2}-\sqrt{2n-1}\right)$ خاضعاً للتوزيع الطبيعي المعياري $N(0,1)$.

8-3-1- اختبار الفروض للتباين σ^2

يقوم هذا الاختبار على مقارنة الثابت الإحصائي (المعلمة) σ^2 مع التابع الإحصائي S^2 المحسوبة من عينة عشوائية صغيرة $n<30$ والهدف من ذلك بيان فيما إذا كان الفرق بين σ^2 و S^2 معنوياً أم لا وذلك عند درجات حرية $v = n-1$ ومستوى معنوية α وذلك وفق الخطوات التالية:

1 . صياغة الفرضية: $H_0 : \sigma^2 = S^2$

$$H_1 : \sigma^2 \neq S^2$$

2 . الاختبار الإحصائي: $$\chi^2 = (n-1)\frac{S^2}{\sigma^2}$$

3 . اتخاذ القرار: من خلال جداول χ^2 (وهي جداول معطاة لاتجاه واحد فقط).

نوجد من جداول χ^2 قيمة χ^2_α الجدولية المقابلة لمستوى دلالة α ، ودرجات حرية $v = n-1$ ، ثم نقوم بالمقارنة بين χ^2 والمحسوبة والتي حصلنا عليها من الاختبار مع χ^2 الجدولية فنكون أمام حالتين:

. إما أن تكون (الجدولية) $\chi^2 \leq$ (المحسوبة) χ^2 عند مستوى دلالة α ودرجات حرية v ، وهذا يعني χ^2 وقعت ضمن منطقة القبول ، فإننا نقبل فرضية العدم H_0 ، ونقر بأن الفرق بين S^2, σ^2 غير معنوي ، أي أن $\sigma^2 \neq S^2$

. وإما أن تكون (الجدولية) $\chi^2 >$ (المحسوبة) χ^2 وهذا يعني أن χ^2 وقعت ضمن منطقة الرفض ، فإننا نرفض فرضية لعدم H_0 ونقرر بأن الفرق بين S^2, σ^2 معنوي ، أي أن $\sigma^2 \neq S^2$.

تمرين 2: بفرض أن أعمار الطلاب بالجامعات السورية تتبع التوزيع الطبيعي بانحراف معياري قدره(6) سنوات. سحبت منهم عينة عشوائية بحجم (25) طالباً ،

فوجد أن الانحراف المعياري للأعمار في العينة يساوي (5)، هل تدل هذه البيانات على أن تباين أعمار الطلاب بهذه الجامعة يختلف عن تباين أعمار الطلاب بالجامعات السورية عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . نقوم بصياغة الفرضيتين كما يلي : $H_0 : \sigma^2 = S^2$

$H_1 : \sigma^2 \neq S^2$

2 . نحسب الاختبار الإحصائي: $\chi^2 = (n-1)\frac{S^2}{\sigma^2} = (25-1)\frac{(5)^2}{(6)^2} = 16.66$

3 . اتخاذ القرار: من جداول توزيع χ^2 نجد أن :

$$\chi^2_{(\alpha,v)} = \chi^2_{(0.05,24)} = 36.41$$

وبالمقارنة نجد أن : $16.66 \prec 36.41$ ، وبالتالي نقبل فرضية العدم H_0 ونقول إن تباين أعمار الطلاب في تلك العينة لا يختلف عن أعمار الطلاب في الجامعات السورية بشكل عام.

8-3-2- تقدير تباين المجتمع (σ^2)

إذا كان لدينا عينة عشوائية ($n < 30$) سحبت من مجتمع إحصائي فإنه يمكن إجراء تقدير مجال الثقة لتباين المجتمع (σ^2) بدلالة (S^2) المعلومة من العينة المحسوبة وذلك من خلال المجال الاحتمالي الآتي:

$$\Pr\left[(n-1)\frac{S^2}{\chi^2_{\left(\frac{\alpha}{2}\right)}}\right] \leq \sigma^2 \leq \left[(n-1)\frac{S^2}{\chi^2_{\left(1-\frac{\alpha}{2}\right)}}\right] = 1-\alpha$$

وبالتالي يمكن القول بأن σ^2 يقع ضمن المجال السابق باحتمال $1 - \alpha$ أي بمستوى دلالة α.

تمرين:نعود إلى التمرين السابق المتعلق بأعمار الطلاب في الجامعات السورية ، **والمطلوب:** أوجد مجال ثقة (0.95) لتباين أعمار الطلاب في هذه الجامعة.

الحل:

من أجل القيام بهذا التقدير، فإننا نستخدم قيمتين لـ χ^2 الجدولية هما : $X^2_{a/2}$ (من اليمين) و $\chi^2_{\left(1-\frac{\alpha}{2}\right)}$ (من اليسار) :

$$\chi^2_{\frac{\alpha}{2}} = \chi^2_{\frac{0.05}{2}} = \chi^2_{0.025} = 39.4$$

$$\chi^2_{\left(1-\frac{\alpha}{2}\right)} = \chi^2_{\left(1-\frac{0.025}{}\right)} = \chi^2_{0.975} = 12.4$$

$$(24)\frac{(25)}{39.4} \le \sigma^2 \le (24)\frac{(25)}{12.4}$$

$$15.23 \le \sigma^2 \le 48.39$$

وبالتالي نجد أنه : باحتمال 95% فإن تباين المجتمع لن يقل عن (15.23) ولن يزيد عن (48.39).

تمرين: بفرض أن تباين الأطوال لجميع الطلاب في جامعة دمشق هو (30) سم، وللتأكد من ذلك سحبت عينة عشوائية من (51) طالباً من طلاب إحدى الكليات التابعة لها، فكان تباين الأطول ($S^2 = 25$). بفرض أن الأطوال تتبع التوزيع الطبيعي.

1 . اختبر الفرض القائل بتساوي التباينين عند مستوى معنوية $\alpha = 0.05$

2 . أوجد مجال ثقة (0.95) للتباين الحقيقي لأطوال طلاب جامعة دمشق.

الحل:

1 . صياغة الفرضية: $H_0 : \sigma^2 = S^2$

$H_1 : \sigma^2 \neq S^2$

2 . الاختبار الإحصائي: $\chi^2 = (n-1)\frac{S^2}{\sigma^2} = (50)\frac{(25)}{(30)} = 41.67$

3 . اتخاذ القرار: من خلال جداول χ^2 نجد أن:

$\chi^2_{(\alpha,v)} = \chi^2_{(0.05,50)} = 67.6$

من خلال المقارنة نجد أن :

$\chi^2 = 41.67 < \chi^2_{0.05} = 67.6$

لذلك نقبل فرضية العدم H_0 ونقول إن التباينين متساويين في المجتمع والعينة.

حل الطلب الثاني نستخدم قيمتين لـ χ^2 الجدولية هما : $\chi^2_{a/2}$ و $\chi^2_{1-a/2}$

وذلك عند مستوى معنوية $\alpha = 0.05$ وعدد درجات حرية $v = 50$

بتطبيق القانون:

$$(n-1)\frac{S^2}{\chi^2_{a/2}} \leq \sigma^2 \leq (n-1)\frac{S^2}{\chi^2_{1-a/2}}$$

$$(50)\frac{(25)}{71.4}\sigma^2 \leq (50)\frac{(25)}{32.4}$$

$$17.5 \leq \sigma^2 \leq 38.6$$

تمرين: سحبت عينة عشوائية من (25) طالباً بإحدى الجامعات فوجد أن تباين إنفاقهم السنوي هو ($S^2 = 8000$) افترض أن الإنفاق السنوي لجميع طلبة الجامعات يتبع التوزيع الطبيعي تباين قدره (9000)، هل تدل هذه البيانات على أن تباين إنفاق طلبه هذه الجامعة أكبر معنوياً من تباين إنفاق طلبه كل الجامعات وذلك عند مستوى معنوية $\alpha = 5\%$.

الحل:

1 . صياغة الفرضية: $H_0 : \sigma^2 = S^2$

$H_1 : \sigma^2 > S^2$

2 . الاختبار الإحصائي: $\chi^2(n-1)\frac{S^2}{\sigma^2} = \frac{(24)(8000)}{9000} = 21.33$

3 . اتخاذ القرار: من جداول χ^2 نجد أن: $\chi^2_{(0.05,24)} = 36.41$ ومن خلال المقارنة نجد أن: $\chi^2 = 21.33 < \chi^2_a = 36.41$

وبالتالي نقبل فرضية العدم H_0 ونرفض الفرضية H_1 ونقول أن تباين النفقات لطلبة هذه الجامعة ليس أكبر معنوياً من تباين إنفاق طلبه كل الجامعات (الفرق ليس معنوياً).

8-4- توزيع فيشر (F)

نحن نعلم أن توزيع (t) والاختبارات المتعلقة بهذا التوزيع يفترض أن تكون التباينات متجانسة ومتساوية في المجتمعات المسحوبة منها العينات أي ($\sigma_1^2 = \sigma_2^2$)، وعند إجراء اختبار معنوية الفروق للأوساط الحسابية للعينات الصغيرة لابد من إجراء اختبار التجانس وذلك باستخدام لتوزيع (F) لاختبار تساوي تباييني مجتمعين.

يستخدم توزيع (F) أساساً لاختبارات تساوي تباييني مجتمعين، وهو عبارة عن توزيع احتمالي مستمر، وأن إحصائية الاختبار لهذا التوزيع تعطى بالعلاقة التالية :

$$F = \frac{S_1^2}{S_2^2}$$

وكما هو معلوم لدينا أن توزيعي (t) و (χ^2) يعتمدان اعتماداً كاملاً على معلمة واحدة فقط هي درجات الحرية، أما التوزيع (F) فإنه يعتمد على معلمتين:

- درجات الحرية في البسط v_1 للتباين S_1^2 وهي تساوي $v_1 = n_1 - 1$؛
- درجات الحرية في المقام v_2 للتباين S_2^2 وهي تساوي $v_2 = n_2 - 1$.

ويتحدد هنا التوزيع بمعرفة قيمة كل من (V_1 و V_2) وتسحب قيمة S_1^2 باستخدام بيانات العينة العشوائية المكونة من (n_1) مفرده من المجتمع الأول، وبالمثل فإن المعلمة (v_2) و(S_2^2) تحسبان من خلال بيانات العينة العشوائية المكونة من (n_2) مفرد ة من مفردات المجتمع الثاني، وبالتالي فإن التوزيع (F) يختلف باختلاف درجات الحرية في البسط أو المقام أو كل يهما، وإن إحصائية الاختبار (F) تأخذ قيماً موجبة دوماً بين الصفر وما لانهاية، ومن الناحية العملية يستخدم التباين الأكبر S_1^2 في البسط والأصغر S_2^2 في المقام، كما يكون الواحد الصحيح حداً أدنى لهذه النسبة وكلما كانت درجات الحرية (v_1, v_2) صغيرة فإن منحني F يكون ملتوياً جهة اليمين وكلما زادت درجات الحرية فإن شكل المنحني يتجه نحو التماثل (الشكل 3-8) أما قيمة (F) الجدولية فنحصل عليها من خلال نقطة تقاطع السطر مع العمود في الجدول حيث:

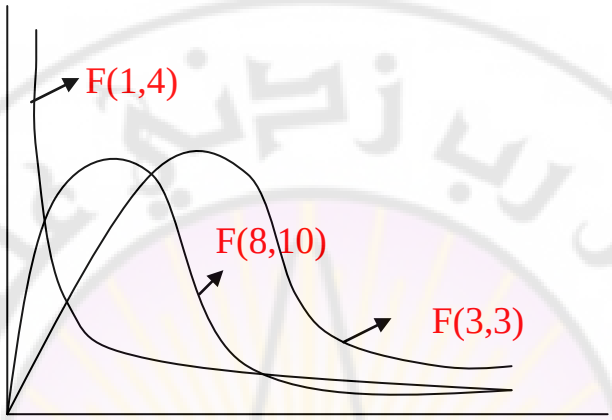
$V_1 = n_1-1$ للتباين الأكبر S_1^2 من السطر.

$V_2 = n_2-1$ للتباين الأصغر S_2^2 من العمود.

وفي بعض الحالات يتم أخذ أقرب قيمة لدرجات الحرية في حالة عدم توافرها في الجدول ونرمز للقيمة الجدولية:

$$F(\alpha, v_1, v_2) = F(\alpha, n_1 - 1, n_2 - 1)$$

وجداول (F) فيشر هي جداول من اتجاه واحد بسبب القيمة الموجبة للتوزيع.



الشكل (8-3)

ويتم إجراء الاختبار الإحصائي على النحو التالي:

1 . صياغة الفرضية:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

2 . الاختبار الإحصائي: $F = \frac{S_1^2}{S_2^2}$

3 . اتخاذ القرار: من خلال جداول التوزيع (F) فوجد قيمة $F_{(\alpha, v1, v2)}$ المقابلة لـ مستوى معنوية $\alpha = 0.01$ أو $\alpha = 0.05$ ودرجات حرية v_1 تمثل سطر الجدول وV_2 تمثل عمود جدول F.

ومن خلال المقارنة بين $F_{(\alpha, v1, v2)}$ الجدولية و F المحسوبة التي حصلنا عليها نكون أمام حالتين:

. إذا كانت (الجدولية) $F_{(\alpha, v1, v2)}$ > (المحسوبة) F ، فإننا نقبل فرضية العدم H_0 ونقول إن الفرق بين التباينين غير معنوي أي أن: $\sigma_1^2=\sigma_2^2$

. أما إذا كانت : (الجدولية) $F_{(\alpha, v1, v2)}$ < F (المحسوبة) ، فإننا نرفض فرضية العدم H_0 ونقول إن الفرق بين التباينين معنوي أي أن:

$$\sigma_1^2\neq\sigma_2^2$$

مع العلم أن الاختبار قد يكون من اتجاه واحد أو من اتجاهين برغم أن الجداول تكون معطاة دوماً من اتجاه واحد، وفي الحالة الثانية يجب تقسيم مستوى المعنوية على (2).

تمرين (1): أرادت إحدى المؤسسات الاجتماعية أن توضح فيما إذا كان أسلوب النقاش الجماعي يؤدي إلى تقارب الآراء حول موضوع معين، فقامت المؤسسة باختيار عينة حجمها (16) شخصاً وقسمتهم عشوائياً إلى مجموعتين، الأولى مكونة من ($n_1 = 6$) أشخاص طرحت عليهم الموضوع وطلبت منهم إعلان رأيهم النهائي بعد النقاش الجماعي وأعطى المختصين علامات تقييمية للآراء المطروحة ووجد أن التباين بين الآراء ($S_1^2 = 77.9$)، أما المجموعة الثانية والتي حجمها ($n_2 = 10$) أشخاص فقد طلب من أفرادها أن يبدوا آراءهم على الموضوع دون إجراء نقاش مسبق ووجد أن التباين في الآراء ($S_2^2 = 194.2$) والمطلوب:

اختبر الفرض القائل بأن النقاش يؤدي إلى تقليل التباين في الآراء:

الحل:

1 . صياغة الفرضية:

$$H_0 : \sigma_1^2=\sigma_2^2$$
$$H_1 : \sigma_1^2>\sigma_2^2$$

2 . الاختبار الإحصائي: $F = \frac{S_2^2}{S_1^2} = \frac{194.2}{77.9} = 2.49$

3 . اتخاذ القرار: من جداول (F) نجد أن: $F_{(0.05,9.5)} = 4.77$ من خلال المقارنة نجد أن:

$F = 2.49 < F_{(\alpha,v_1,v_2)} = 4.77$ وبالتالي نقبل فرضية العدم H_0 ونقول إن النقاش الجماعي حول هذا الموضوع لم يكن كافياً لتقليل التباين في الآراء فيما لم يتم النقاش.

تمرين (2): بفرض أننا نرغب في معرفة ما إذا كان تشتت درجات القراءة لهؤلاء الذين تعلموا القراءة باستخدام طريقة الصوت أكبر من تشتت الذين تعلموا باستخدام طريقة الصورة. سحبت عينة عشوائية مكونة من (25) مفردة من المجتمع الأول فكان تباينها (S_1^2 = 40) وسحبت عينة أخرى مستقلة مكونة من (30) مفردة من المجتمع الثاني فوجد أن تباينها (S_2^2 = 25)، إذا كانت ($\alpha = 0.05$)هل تؤدي هذه البيانات إلى قبول الفرض القائل بأن تباين درجات اختبار القراءة باستخدام طريقة الصوت أكبر من تباين درجات اختبار القراءة باستخدام طريقة الصورة؟

الحل:

1 . صياغة الفرضية:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

2 . الاختبار الإحصائي: $F = \frac{S_2^2}{S_1^2} = \frac{40}{25} = 1.6$

3 . اتخاذ القرار: من جداول (F) فيشير نجد أن:

$$F_{(\alpha,v_1,v_2)} = F_{(\alpha,n_{1-1},n_{2-1})} = F_{(0.05,24.29)} = 1.94$$

ومن خلال المقارنة نجد أن: $F = 1.6 < F_{(\alpha,v_1,v_2)} = 1.94$

وبالتالي نقبل فرضية العدم H_0 ونرفض الفرضية البديلة أي أن تشتت درجات القراءة باستخدام طريقة الصوت ليس أكبر بدرجة كافية من تشتت درجات القراءة بطريقة الصوت.

تمرين (3): بفرض أننا نريد معرفة إذا كان تشتت أطوال الطلاب يساوي تشتت أطوال الطالبات، افترض أن σ_1^2 يمثل تباين أطوال الطلاب (المجتمع الأول)، وأن σ_2^2 يمثل تباين أطوال الطالبات (المجتمع الثاني)، وأن الأطوال تتبع التوزيع الطبيعي، سحبت عينتان مستقلتان من هذين المجتمعين، فإذا علمت أن العينة الأولى تتكون من (16) طالباً وتباين الأطوال فيها (S_1^2 = 200) وأن العينة الثانية تتكون من (25) طالبة وتباين أطوالها (S_2^2 = 600). اختبر الفرض القائل بعدم تساوي تبايني المجتمعين عند مستوى دلالة

$\alpha = 0.05$

الحل:

1 . صياغة الفرضية:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

2 . الاختبار الإحصائي:

$$F = \frac{S_2^2}{S_1^2} = \frac{600}{200} = 3$$

نلاحظ هنا أن الاختبار من اتجاهين لذلك يجب تقسيم مستوى المعنوية على (2):

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

$$V_1 = n_2 - 1 = 24$$

$$V_2 = n_1 - 1 = 15$$

$$F_{(\alpha/2, v_1, v_2)} = F_{(0.025, 24, 15)} = 2.7$$

$$F = 3 > F_{(\alpha/2, v_1, v_2)} = 2.7$$

وبالتالي نرفض فرضية العدم H_0 ونقبل H_1 الفرضية البديلة ونقول إن التباينين في المجتمعين غير متساويين.

8-5- تمارين ومسائل

تمرين (1): تدعي إحدى المستشفيات الحكومية بأن متوسط أوزان الأطفال عند الولادة أقل من (3.5) كغ وبغية التحقق من ذلك سحبت عينة عشوائية من (16) طفلاً فوجد أن متوسطها $\overline{X}$ = 3 كغ وبانحراف معياري قدره (1) كغ، بفرض أن أوزان الأطفال عند الولادة تتبع التوزيع الطبيعي، هل تدل هذه البيانات على أن متوسط الوزن أقل من (3.5) كغ عند مستوى معنوية $\alpha = 0.01$.

تمرين (2): في اختبار أجري على (16) مصباحاً كهربائياً وجدنا أن متوسط عمر المصباح $\overline{X}$ = 3000 ساعة وبانحراف معياري (20) ساعة **والمطلوب:** أوجد مجال الثقة لمتوسط عمر المصابيح في المجتمع باحتمال 95%.

تمرين (3): أجري اختبار نوعين جديدين من السيارات لمعرفة الاختلافات بينهما في التحكم في تلوث الهواء الناتج منهما، اختبر النوع (A) لمدة (5) أيام فوجد أن متوسط مقياس التلوث $\overline{X1}$ = 202.2 وانحراف معياري 2.52، واختبر النوع (B) لمدة (5) أيام فوجد أن متوسط مقياس التلوث $\overline{X2}$ = 206 وانحراف معياري (2.57) **والمطلوب:** هل يمكن اعتبار متوسط التلوث من السيارة (A) يكافىء متوسط التلوث من السيارة (B) بفرض أن التباينين $\sigma_1^2 = \sigma_2^2$ في المجتمعين متساويان، وذلك عند مستوى دلالة $\alpha = 0.05$.

تمرين (4): صمم اختبار لقياس قدرة الطفل على تعلم كلمات جديدة فكانت النتائج كما يلي:

العينة الأولى: $\overline{X}_1 = 95$ ، $S_1^2 = 40$ ، $n_1 = 10$

العينة الثانية: $S_2^2 = 36$ ، $\overline{X}_2 = 97$ ، $n_2 = 8$

والمطلوب: اختبر الفرض القائل بتساوي قدرة أطفال العينتين على تعلم الكلمات الجديدة مع العلم أن درجات الاختبار تتبع التوزيع الطبيعي والتباين ان مجهولان ومتساويان $\sigma_1^2 = \sigma_2^2$ عند مستوى معنوية $\alpha = 0.05$.

تمرين (5): ترغب شركة في اختبار مدى اهتراء نوعين من إطارات السيارات (A, B)، لهذا قامت الشركة بتركيب هذه الإطارات على (5) سيارات بحيث كانت الإطارات على إحدى الجهتين من النوع (A) وعلى الجهة الأخرى من النوع (B) وبعد استعمالها المسافة نفسها وقياس مقدار الاهتراء حصلت الشركة على النتائج الآتية:

رقم السيارة	الاهتراء A	الاهتراء B	D	D^2
1	10.6	10.2	0.4	0.16
2	9.8	9.4	0.4	0.16
3	12.3	11.8	0.5	0.25
4	9.7	9.1	0.6	0.36
5	8.8	8.3	0.5	0.25

والمطلوب: هل تؤدي هذه البيانات إلى قبول الفرض القائل بأن مقدار الاهتراء في النوع B لا يختلف عن مقدار الاهتراء في النوع A وذلك عند مستوى معنوية $\alpha = 0.05$ ، بفرض أن مقدار الاهتراء يتبع التوزيع الطبيعي.

تمرين (6): سحبت عينة عشوائية مكونة من (10) مفردات من مجتمع طبيعي فوجد أن تباينها (S2 = 15)، أوجد فترة ثقة (0.95) لتباين المجتمع.

تمرين (7): يعتقد البعض أن تباين عدد الوحدات التي ينتجها العمال في الساعة سيزيد بعد توقيعهم على أحد العقود عما كان قبل توقيع هذا العقد، علماً أن التباين قبل توقيع العقد هو (σ^2 = 80). سحبت عينة عشوائية من (25) عامل إنتاج وسجل إنتاجهم في الساعة بعد توقيع العقد فوجد أن تباين العينة هو (90)،

افترض أن عدد الوحدات المنتجة في الساعة تتبع التوزيع الطبيعي، ثم اختبر الفرض السابق القائل بزيادة التباين بعد توقيع العقد وذلك عند مستوى معنوية α = 0.05.

تمرين (8): يدعي أحد صانعي المنسوجات القطنية بأن تباين قوة مقاومة القطع للخيوط القطنية التي تنتجها شركة أقل من تباين قوة مقاومة القطع لخيوط النايلون، حيث يُعتقد أن مقاومة القطع بالكغ للنوعين تتبع التوزيع الطبيعي.

سحبت عينة عشوائية مؤلفة من (41) بكره من خيوط القطن فوجد أن تباينها (S_1^2 = 65) وسحبت عينة أخرى مستقلة مؤلفة من (25) بكرة من خيوط النايلون فوجد أن تباينها (S_2^2 = 78) ، هل تؤيد هذه البيانات الادعاء السابق عند مستوى معنوية (α = 0.05).

تمرين (9): لمقارنة طريقتين من طرق تعليم القراءة، سحبت عينتان مستقلتان من الأطفال حيث استخدمت الطريقة (A) لتعليم أطفال إحدى العينتين واستخدمت الطريقة (B) لتعليم أطفال العينة الأخرى (بفرض أن تحديد الطريقة المستخدمة لتعليم كل عينة يتم عشوائياً)، والبيانات التالية تمثل نتيجة اختبار موحد أعطي لأطفال العينتين:

الطريقة A: $n_1 = 25$ $S_1^2 = 108$

الطريقة B: $n_2 = 30$ $S_2^2 = 95$

بفرض أن الدرجات تتبع التوزيع الطبيعي. اختبر الفرض القائل بتساوي تباين مجتمعي درجات اختبار القراءة بالطريقتين وذلك عند مستوى معنوية (0.05).

تمرين (10): بفرض أن المشتريات السنوية لعملاء إحدى الشركات تتبع التوزيع الطبيعي، ترغب هذه الشركة في معرفة ما إذا كان تباين المشتريات السنوية للرجال يختلف عن تباين المشتريات السنوية للسيدات، سحبت عينة عشوائية من (31)

سيدة فوجد أن تباينها يساوي (225)، كما سحبت عينة أخرى مستقلة من (25) رجلاً فوجد أن تباينها يساوي (180). هل تؤدي هذه البيانات إلى رفض فرضية العدم القائلة بتساوي التباينين في مقابل الفرض البديل القائل بعدم تساوي التباينين عند مستوى دلالة
.($\alpha = 0.10$)

الفصل التاسع
الإحصاء اللامعلمي

9-1- اختبار حسن المطابقة (كاي مربع χ^2)
9-2- ملاحظات حول اختبار حسن المطابقة
9-3- أسئلة وتمارين

Damascus University

وقل رب زدني علما
جامعة دمشق
Damascus University

وقل رب زدني علما
جامعة دمشق
Damascus University

مقدمة:

معظم أساليب اختبار الفروض التي ناقشناها سابقاً تعرف بالطرق المعلمية Parametric Methods وذلك لأنها تهتم بمعالم (ثوابت) المجتمع مثل ($\overline{X}, \sigma^2, \overline{P}$)، وعلى الرغم من ذلك يوجد العديد من المواقف التي لا نستطيع فيها تطبيق الطرق المعلمية ، لاسيما عندما ندرس الظواهر النوعية التي تكون فيها المقاييس أو المشاهدات في صور رتب أو حالات ، وعندها يصبح من الأفضل استخدام الطرق اللامعلمية Nonparametric ، أضف إلى ذلك أن الطرق المعلمية تعتمد على افتراض معرفة التوزيع الاحتمالي للمجتمع ، فعند إجراء اختبارات الفروض الإحصائية للمتوسط والتباين أو للفرق بين متوسطين والنسبة بين تباينين افترضنا أن للمجتمعات المدروسة توزيعات طبيعية أو قريبة من الطبيعية ، أما بالنسبة للطرق اللامعلمية فإنها لا تفترض شكلاً معيناً لتوزيع المجتمع الأصلي ، وبذلك تكون طرق حرة التوزيع distribution free، ومع أن العبارتين (لا معلمي) و (حرة التوزيع) ليس لهما المعنى نفسه تماماً، فعبارة (لا معلمي) تستخدم لوصف اختبارات الفروض التي لا تتضمن أية معلمة من معالم المجتمع، و (حرة التوزيع (توزع حر)) تستخدم لوصف اختبارات الفروض التي لا تفترض شكلاً معنياً لتوزيع المجتمع الأصلي.

لقد ظهرت اختبارات إحصائية عديدة لا تتطلب شكلاً خاصاً لتوزيع المجتمع الأصلي لتحل محل بعض الاختبارات للفروض المعلمية، لذا ينصح باستخدام الطرق اللا معلمية عندما لا نستطيع معرفة شكل توزيع المجتمع الأصلي، لكن هذا لا يعني بالطبع أن الطرق اللامعلمية مفضلة دائماً على الطرق المعلمية ، فإذا تحقق شرط تبعية البيانات للتوزيع الطبيعي ينصح باستخدام الطرق المعلمية والتي تؤدي في هذه الحالة إلى نتائج أدق، أما الطرق اللامعلمية تستخدم فقط عندما يتأكد الباحث من عدم إمكانية استخدام الطرق المعلمية مع العلم أن الطرق

اللامعلمية تتطلب عدداً أقل من الفروض وسهلة الشرح والفهم والتطبيق، وتستخدم هذه الطرق خصوصاً، إذا كانت البيانات في صورة رتب أو حالات وليست في صورة مقاييس كمية.

9-1- اختبار حسن المطابقة (كاي مربع χ^2):

تعتمد الدراسات الإحصائية على فرض نموذج إحصائي تخضع له البيانات المتوفرة. فعلى سبيل المثال فإن درجات الطلاب تخضع للتوزيع الطبيعي، وأعداد حوادث السير تخضع لتوزيع بواسون وعدد المصابيح غير الصالحة في مصنع تخضع لتوزيع ثنائي الحدين وهكذا... إن السؤال الذي يطرح نفسه هو كيف يمكن التأكد من أن البيانات المتوفرة تخضع لهذا النموذج الإحصائي المعطى أو ذاك ؟ بمعنى آخر نود اختبار حسن مطابقة النموذج الإحصائي للبيانات المتوفرة، وهناك عدة اختبارات لهذا الغرض منها:

اختبار كاي مربع χ^2 لحسن المطابقة، وتقوم فلسفة هذا الاختبار على حساب مقياس يعبر عن مدى التطابق بين المشاهدات الحقيقية (التكرارات الحقيقية) مع المشاهدات المتوقعة (التكرارات المتوقعة) والبرهان عما إذا كان النموذج الإحصائي صحيحاً، بمعنى آخر ، إن حسن المطابقة عبارة عن اختبار فيما إذا كان توزع المفردات للعينة يوافق أو يطابق ما هو مفروض أو متوقع أن يكون توزعها في المجتمع ، فإذا رمزنا بـ E للتكرارات المتوقعة و Q للتكرارات الحقيقية المعروفة من خلال العينة فإنه يمكن حساب الاختلاف أو مدى التطابق بين هذه التكرارات من خلال العلاقة الآتية:

$$\chi^2 = \sum_{i=1}^{r} \frac{(Q-E)^2}{E}$$

وهو عبارة عن متحول عشوائي يخضع لتوزيع χ^2 ذي v= r – 1 درجة حرية ، ويتم الاختبار من خلال الخطوات التالية:

1 . نقوم بصياغة الفرضيتين كما يلي :

H_0: تنص على أن التكرارات الحقيقية Q تتطابق مع التكرارات المتوقعة E.

H_1: تنص على أن التكرارات الحقيقية Q لا تتطابق مع التكرارات المتوقعة E.

2 . نحسب قيمة الاختبار الإحصائي: $\chi^2 = \sum_{i=1}^{r} \frac{(Q-E)^2}{E}$

3 . اتخاذ القرار: من جداول التوزيع χ^2 نوجد قيمة χ^2 الجدولية $\chi^2(\alpha,v)$ عند مستوى معنوية α وعدد درجات حرية v = r -1 ثم نقارنها مع قيمة χ^2 المحسوبة، وهنا نكون أما حالتين :

– إما أن تكون (الجدولية $\chi^2(\alpha,v)$ > χ^2 (المحسوبة) ، وعندها نقبل فرضية العدم H_0 ، وبالتالي يمكن القول إن التكرارات الحقيقية والمتوقعة متطابقة .

– وإما أن تكون (الجدولية) χ^2 (α, v) > χ^2 (المحسوبة) ، وعندها نرفض فرضية العدم H_0 ، بالتالي يمكن القول إن التكرارات الحقيقية والمتوقعة غير متطابقة .

تمرين (1): رغبت إحدى المؤسسات الاجتماعية معرفة رأي السكان بتنظيم الأسرة، فقامت المؤسسة بأخذ عينة عشوائية من (50) شخصاً بالغاً ، فوجد أن (32) منهم يؤيدون الفكرة و(18) منهم يعارضون، هل يمكن القول إن الآراء تنقسم بشكل متساوٍ حول هذا الموضوع وذلك عند مستوى معنوية α= 0.05.

الحل:

1 . صياغة الفرضية: $H_0 : P_1 = P_2 = 1/2$

$H_1 : P_1 \neq P_2 \neq 1/2$

3 . نحسب قيمة الاختبار الإحصائي:

$$\chi^2 = \sum_{i=1}^{r} \frac{(Q-E)^2}{E}$$

لحساب المجموع السابق نضع الجدول التالي:

التكرارات / الآراء	Q	E	Q - E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
مؤيد	32	25	7	49	1.96
معارض	18	25	−7	49	1.96
$\sum$	50	50	–	–	$\chi^2 = 3.29$

4 . اتخاذ القرار: من جداول التوزيع χ^2 وعند $\alpha = 0.05$ و $v = 2-1 = 1$ نجد

أن: $\chi^2_{(\alpha,v)} = \chi^2_{(0.05,1)} = 3.84$

ومن خلال المقارنة نجد أن:

χ^2 (المحسوبة) $=3.92 > \chi^2_{(\alpha, V)} = 3.84$

وبالتالي نرفض فرضية العدم H_0 ونقول إن الآراء في هذا المجتمع لا تتقسم بالتساوي.

تمرين (2): يهتم الباحثون في إحدى الدول بمعرفة الحالة التعليمية لجميع من لهم الحق في الانتخاب، وقد تم من خلال خبرة سابقة تصنيف جميع من لهم حق الانتخاب حسب حالتهم التعليمية إلى خمسة أقسام: ابتدائي، إعدادي، ثانوي، جامعي، دراسات عليا وأن النسب المقابلة لهذه الأقسام كانت على النحو التالي:

الحالة التعليمية	ابتدائي	إعدادي	ثانوي	جامعي	دراسات عليا
النسبة المئوية	%30	%35	%20	%10	%5

أخذت عينة عشوائية من (1000) ناخب فوجد أن (270) طالباً في المرحلة الابتدائية ، (315) طالباً في المرحلة الإعدادية ، (230) طالباً في المرحلة الثانوية ، (125)طالباً طالباً في الجامعة ، (60) في مرحلة الدراسات العليا ، والمطلوب: اختبار فيما إذا كان التوزيع الحالي للحالة التعليمية لمن لهم حق الانتخاب في هذه الدولة ما زال هو نفس التوزيع السابق الذي تم الحصول عليه من خلال خبرة سابقة وذلك عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة الفرضية: $H_0 : P_i = Q_i$

علماً بأن الحالة التعليمية تتوزع حسب النسب الآتية:

(30%، 35%، 20%، 10%، 5%).

$H_1 : p_i \neq Q_i , (i = 1, 2 , 3 , 4 , 5)$

2 . نحسب قيمة الاختبار الإحصائي:

$$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

لذلك نحسب التكرارات المتوقعة E من العلاقات التناسبية التالية :

$$E_1 = \frac{(30)(1000)}{100} = 300 \quad ، \quad E_2 = \frac{(35)(1000)}{100} = 350 \quad ، \quad E_3 = \frac{(20)(1000)}{100} = 200$$

$$E_4 = \frac{(10)(1000)}{100} 100 \quad ، \quad E_5 = \frac{(5)(1000)}{100} = 50$$

ثم نضع جدول التكرارات الحقيقية والمتوقعة:

$\frac{(Q-E)^2}{E}$	$(Q-E)^2$	Q - E	E	Q	الحالية التعليمية / الخلايا
3.00	900	−30	300	270	ابتدائي
3.50	1225	−35	350	315	إعدادي
4.50	900	30	200	230	ثانوي
6.25	625	25	100	125	جامعي
2.00	100	10	50	60	دراسات عليا
χ^2 =19.25	–	–	1000	1000	$\sum$

3 . اتخاذ القرار : من جداول التوزيع χ^2 نجد أن: $\chi^2_{(0.05.4)} = 9.84$

ومن خلال المقارنة نجد: $\chi^2_{(\alpha ,V)} = 9.48$ **<19.2=** χ^2، لذلك نرفض فرضية العدم H_0 ونقول إن التوزيع للحالة التعليمية قد تغير عن السابق.

تمرين (3): رغب أحد الباحثين معرفة آراء المواطنين بشأن الاتجاه نحو عمل المرأة وقد تم من خلال خبرة سابقة تصنف جميع الآراء والنسب المقابلة لها على النحو التالي:

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	$\sum$
النسب المئوية	%30	%20	%10	%25	%15	%100

ثم قام الباحث باختيار عينة عشوائية تتكون من (200) مواطن لاستطلاع آرائهم بشأن الاتجاه حول عمل المرأة من خلال اختبار إجابة واحدة للسؤال الموجه إليهم، وهو: هل تؤيد عمل المرأة خارج البيت في بعض الأعمال الموجودة في المجتمع؟ فكانت النتائج على النحو التالي:

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	$\sum$
عدد المواطنين	75	55	15	45	10	%100

المطلوب: اختبر الفرض القائل بأن هناك تشابهاً بين نسب استجابات العينة مع النسب في المجتمع، من خلال الخبرة السابقة عند مستوى ($\alpha = 0.05$).

الحل:

1 . صياغة الفرضية: $H_0 : P_i = Q_i$

تتوزع الحالة (الآراء) على النحو (30%، 20%، 10%، 25%، 15%).

$$H_1 : P_i \neq Q_i$$

(أي لا تتوافق النسب في هذه الحالة مع النسب السابقة).

2 . نحسب قيمة الاختبار الإحصائي: $\chi^2 = \sum \frac{(Q-E)^2}{E}$

نحسب التكرارات المتوقعة E:

$$E_1 = \frac{(30)(200)}{100} = 60 \quad ، \quad E_2 = \frac{(20)(200)}{100} = 40 \quad ، \quad E_3 = \frac{(10)(200)}{100} = 20$$

$$E_4 = \frac{(25)(200)}{100} 50 \quad ، \quad E_5 = \frac{(15)(200)}{100} = 30$$

ولنضع النتائج في الجدول التالي :

الآراء	Q	E	Q - E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
مؤيد بشدة	75	60	15	225	3.750
مؤيد	55	40	15	225	5.625
محايد	15	20	−5	25	1.250
معارض	45	50	−5	25	0.5
معارض بشدة	10	30	−20	400	13.333
$\sum$	200	200	−	−	$\chi^2 = 24.548$

3 . اتخاذ القرار: من جداول التوزيع χ^2 نجد أن:

$$\chi^2\ (\alpha, v) = \chi^2\ (0.05, 4) = 9.487$$

ومن خلال المقارنة نجد أنَّ:

$$\chi^2 = 24.548 > \chi^2\ (\alpha, v) = 9.487$$

لذلك نرفض فرضية العدم H_0 أي أن النسب لم تعد كالسابق وليس هناك تشابه بين استجابات العينة والمجتمع .

تمرين (4): أخذت عينة عشوائية من (200) طفل موزعين حسب الجنس كالآتي:(112) ذكور و(88) إناث **والمطلوب:** اختبر الفرضية القائلة بأن نسبة الجنس عند الولادة متساوية عند مستوى معنوي (0.05) الحل:

1 . صياغة الفرضية: $H_0 = P_1 = P_2 = 1/2$

$H_1 = P_1 \neq P_2 \neq 1/2$

ولنضع جدول التكرارات الحقيقية والمتوقعة:

التكرارات / الجنس	Q	E	Q - E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
ذكور	112	100	12	144	1.44
إناث	88	100	−12	144	1.44
$\sum$	200	200	–	–	χ^2 =19.25

3 . اتخاذ القرار: من جداول χ^2 نجد أن: $\chi^2_{(0.05.1)} = 3.841$

ومن خلال المقارنة نجد أن: $\chi^2_{(a, v)} = 3.841 < \chi^2 = 2.88$

بالتالي نقبل بفرضية العدم H_0 ونقول إن نسبة الجنس عند الولادة متساوية.

9-2 – ملاحظات حول اختبار حسن المطابقة

1 . شاهدنا في بعض التطبيقات أننا افترضنا أ ن نسبة الاستجابات في المجتمع الأصلي كانت متساوية لكل الاختبارات (التوزيع بالتساوي) لكن هذه ليست هي الحال في جميع التجارب أو التطبيقات أو البحوث، فقد تكون التكرارات المتوقعة E تتوزع وفقاً لقواعد أو نسب معينة معروفة في ضوء خبرة سابقة.

2 . عند استخدام اختبار χ^2 لجودة المطابقة، ينبغي ألا تقل القيمة المتوقعة (التكرار المتوقع عن (5) أي عن 20% في العدد الكلي، ويمكن في بعض الأحيان اللجوء إلى ضم الخلايا (أي التكرارات المتوقعة) التي أقل من (5) إلى بعضها البعض، أما إذا كانت هذه الخلية هي الخلية الوحيدة التي تقل عن (5) فلا تؤخذ بعين الاعتبار وتبقى كما هي.

3 . هناك إمكانية دمج عدة قيم محسوبة لـ χ^2 في قيمة واحدة، بمعنى إذا كان لدينا قيم كاي مربع لعدة عينات خاصة بالموضوع المدروس نفسه ومسحوبة في المجتمع نفسه ، فإن هذه القيم يمكن إضافتها لاتخاذ قرار أفضل من القرار الذي أتخذ على أساس البيانات الخاصة بعينة واحدة فقط ، ويجب أن تكون درجات الحرية في العينة متساوية في كل التجارب وتكون درجة الحرية لاختبار χ^2 الكلية تساوي مجموع جميع درجات الحرية بالتجارب، وبعبارة أخرى فإن شروط دمج قيم χ^2 هو:

1 . أن تكون الجداول في التجارب في النمط نفسه .

2 . أن تكون التجارب تحمل درجات الحرية نفسها .

ولتوضيح ذلك نأخذ التمرين التالي : بفرض أن البيانات التالية تمثل توزيع عدد الكتب المستعارة في المكتبتين A وB في أسبوع معين ،

المكتبة A:

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة	$\sum$
عدد الكتب المستعارة	24	28	12	10	7	8	11	100

المكتبة B:

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة	$\sum$
عدد الكتب المستعارة	9	10	10	8	5	12	16	70

والمطلوب:

1 . اختبر الفرض القائل بأن نسبة الكتب المستعارة في كل يوم من يومي السبت والأحد تساوي 25% بينما تساوي نسبة الكتب المستعارة في كل يوم من أيام الأسبوع الباقية هي 10% وذلك عند مستوى معنوية $\alpha = 0.05$.

2 . إذا علمت أن قيمة χ^2 لتوزيع الكتب المستعارة في هذا الأسبوع في المكتبة C تساوي (9.8) وفي المكتبة D تساوي (15.516). فما هو القرار الواجب اتخاذه على مستوى الدولة ككل عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة الفرضية:

$H_0 = P_1 = 25\%, P_2 = 10\%$ أي الكتب المستعارة في كلٍ من A, B تتوزع بنسبة 25% للسبت والأحد،و(10%) لباقي للأسبوع

$H_1 = P_1 \neq 25\%, P_2 \neq 10\%$ أي ليس هناك فارق معنوي بين تكرارات الجدول والنسب المذكورة.

2 . الاختبار الإحصائي: $\chi^2 = \sum \frac{(Q-E)^2}{E}$

وجدول التكرارات للمكتبة A هو:

الخلايا	Q	E	Q - E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
1	24	25	1	1	0.04
2	28	25	3	9	0.36
3	12	10	2	4	0.4
4	10	10	0	0	0
5	7	10	3	9	0.9
6	8	10	2	4	0.4
7	11	10	1	1	0.1
$\sum$	100	100	–	–	$\chi^2 = 2.2$

3 . المقارنة واتخاذ القرار: من جداول التوزيع χ^2 نجد أن:

$$\chi^2_{(0.05,6)} = 12.6$$

وبالمقارنة $\chi^2 = 2.2 < \chi^2_{(\alpha, v)} = 12.6$

لذلك نقبل فرضية العدم H_0 ونقول بأن ا لكتب المستعارة في المكتبة A تتوزع حسب النسب المذكورة.

أما جدول التكرارات في المكتبة B فهو:

الخلايا	Q	E	Q - E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
1	9	17.5	−8.5	72.25	4.13
2	20	17.5	−7.5	56.25	3.21
3	10	7	3	9	1.28
4	8	7	1	1	0.14
5	5	7	2	4	0.57
6	12	7	5	25	3.57
7	16	7	9	81	11.57
$\sum$	70	70	–	–	$\chi^2 = 24.48$

وبالمقارنة $\chi^2 = 24.48 > \chi^2_{(\alpha, v)} = 12.6$

لذلك نرفض فرضية العدم H_0 ونقول إن نسبة الكتب المستعارة في المكتبة B لا تتوزع حسب النسب المذكورة .

بعد مناقشة قبول ورفض فرضية العدم لكل عينة على حدة من خلال الجدول التالي ، نقوم بدمج قيم χ^2 في جدول موحد كالآتي :

المدنية	χ^2	v	النتيجة الأولية
A	2.2	6	قبول H_0
B	24.48	6	رفض H_0
C	9.8	6	قبول H_0
D	15.516	6	قبول H_0

$\sum$	–	52	24

نحسب χ^2 الجدولية عند عدد من درجات الحرية مساوي (24) ومستوى معنوية $\alpha = 0.05$، فنجد أن :

$$\chi^2_{(\alpha, v)} = \chi^2_{(0.05, 24)} = 36.4$$

بالمقارنة نجد أن: $\chi^2 = 52 > \chi^2 (\alpha, v) = 36.4$

وبالتالي نرفض فرضية العدم H_0 ونقول إن الكتب المستعارة من المكتبات تتوزع بغير النسب المفروضة .

9-3- اختبار الاستقلال لجداول التوافق (2×2)و(3× 2) و(3× 3)

يعتمد هذا الاختبار على اختبار التطابق النسبي بين التكرارات الحقيقية والتكرارات المتوقعة لظاهرتين أو لمتغيرين ويؤدي هذا الاختبار إلى اتخاذ قرار بشأن استقلال متغير ما إحصائياً عن متغير آخر، لذلك فإن المشكلة هنا هي تحديد ما إذا كان هناك علاقة ما بين مجموعتين من الصفات لمتغيرين في المجتمع، ويتم الاستدلال الإحصائي عليها من خلال بيانات العينة.

وعلى الرغم من ذلك فإن وجود علاقة ما في العينة لا يتطلب بالضرورة وجودها في المجتمع، والسبب في ذلك أن بيانات العينة قد تظهر وجود علاقة ما بين المتغيرين نتيجة لعوامل الصدفة، وذلك على الرغم من استقلال هذين المتغيرين في المجتمع، وبالتالي فإن العلاقة التي تظهرها بيانات العينة يجب أن تكون واضحة بدرجة تكفي لاستنتاج وجودها وفي المجتمع. وطريقة إجراء اختبار χ^2 للاستقلال تشبه بدرجة كبيرة طريقة إجراء اختبار χ^2 لجودة المطابقة، وفي الحقيقة فإن

إحصائيتي الاختبار في الحالتين متشابهتان بدرجة واضحة، ولإجراء الاختبار فإننا نقوم أولاً بتبويب بيانات العينة في جدول مزدوج ويتم الاختبار على النحو التالي:

1 . صياغة فرضية العدم (فرضية الاستقلال) : وتنص على أنه لا توجد علاقة بين الظاهرتين الأولى والثانية، وكل ظاهره مستقلة عن الأخرى ، أما الفرضية البديلة : فهي أنه توجد علاقة بين الظاهرتين المدروستين .

2 . الاختبار الإحصائي: $$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

ويتم إعداد جدول التوافق الذي يبحث عن مدى التطابق بين Q وE من خلال بيانات العينة، على اعتبار كل ظاهرة مقسمة إلى قسمين أو أكثر.

∑	B	A	الظاهرة I / الظاهرة II
مجموع السطر A	Q2	Q1	A
مجموع السطر B	Q4	Q3	B
المجموع الكلي	مجموع العمود B	مجموع العمود A	∑

ويتم حساب التكرارات المتوقعة E على النحو التالي:

$$\text{المتوقع E} = \frac{\text{مجموع السطر} * \text{مجموع العمود}}{\text{المجموع الكلي}}$$

ثم يتم إنشاء الجدول المتعلق بالتكرارات الحقيقية والتكرارات المتوقعة كما يلي:

$\frac{(Q-E)^2}{E}$	$(Q-E)^2$	Q – E	E	Q	الخلايا
.	.	.	E1	Q1	1
.	.	.	E2	Q2	2

3	Q3	E3	.	.	.
4	Q4	E4	.	.	
$\sum$					$\chi^2 =$

3 . اتخاذ القرار: من خلال جداول التوزيع χ^2 نوجد قيمة χ^2 الجدولية عند مستوى معنوية α وعدد درجات حرية $V = (C - 1)(r-1)$ (حيث C, r عدد الأسطر والأعمدة) ثم نقارن بين قيمتي χ^2 الجدولية و χ^2 المحسوبة فنجد أن: إما (الجدولية) $\chi^2_{(\alpha ،r)} > \chi^2$ (المحسوبة) ، وعندها نقبل فرضية الاستقلال ونقول ليس هناك علاقة بين الظاهرتين ، وإما أن تكون $\chi^2 > \chi^2_{(\alpha ،v)}$ (المحسوبة) وبالتالي نرفض فرضية الاستقلال ونقول هناك علاقة بين الظاهرتين ولتوضيح ذلك لدينا التمرين التالي:

تمرين (1):

قام أحد الباحثين باختيار عينة عشوائية من الطلبة الذكور والإناث في المرحلة الثانوية واستطلع آراءهم بالفرع الذي يرغبون اختياره لمواصلة الدراسة (علمي، أدبي) فكانت النتائج على النحو التالي:

التخصص / الجنس	علمي	أدبي	$\sum$
ذكور	110	90	200
إناث	40	60	100
$\sum$	150	150	300

والمطلوب:

اختبار فيما إذا كان هناك علاقة بين التخصص والجنس أم لا، عند مستوى معنوية $\alpha = 0.01$ الحل:

1 . صياغة فرضية العدم : (فرضية الاستقلال) لا توجد علاقة بين الجنس (ذكور، إناث) وبين التخصص، وكل ظاهرة مستقلة عن الأخرى.

2 . الاختبار الإحصائي:

$$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

نحسب التكرارات المتوقعة E:

$$E1 = \frac{(200)(150)}{300} = 100 \qquad E3 = \frac{(100)(150)}{300} = 50$$

$$E2 = \frac{(200)(150)}{300} = 100 \qquad E4 = \frac{(100)(150)}{300} = 50$$

نضع جدول التكرارات الحقيقية والتكرارات المتوقعة بشكل مبسط ، كما يلي :

$\frac{(Q-E)^2}{E}$	$(Q-E)^2$	Q – E	E	Q	الخلايا
1	100	10	100	110	ذكور علمي
1	100	−10	100	90	ذكور أدبي
2	100	−10	50	40	إناث علمي
2	100	10	50	60	إناث أدبي
χ^2 = 6		–	300	300	$\sum$

3 . اتخاذ القرار: من جداول التوزيع χ^2 وعند مستوى دلالة $\alpha = 0.01$ وعدد درجات الحرية

v = (2 . 1) (2 − 1) نجد أن : v = 1 ، ونجد أن χ^2 (0.01.1) = 6.634

وبنتيجة المقارنة نقبل فرضية الاستقلال ونقول ليس هناك علاقة بين الجنس واختيارهم للتخصص.

تمرين (2):

قام أحد الباحثين باختيار عينة عشوائية من (100) عامل لدراسة علامة المؤهل العلمي ودرجة إتقانهم للعمل نحصل على الجدول التالي:

المؤهل / تقييم الآراء	جامعي	غير جامعي	$\sum$
جيد	15	15	30
سيء	25	45	70
$\sum$	40	60	100

والمطلوب: اختبار فيما إذا كان هناك علاقة بين المؤهل العلمي ودرجة الإتقان للعمل عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة فرضية العدم (فرضية الاستقلال) : لا توجد علاقة بين المؤهل العلمي والآراء وكل ظاهرة مستقلة عن الأخرى.

2 . الاختبار الإحصائي:

$$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

ونحسب التكرارات المتوقعة:

$$E1 = \frac{(30)(40)}{100} = 12 \qquad E3 = \frac{(70)(40)}{100} = 28$$

$$E2 = \frac{(30)60)}{100} = 18 \qquad E4 = \frac{(70)(60)}{100} = 42$$

وجدول التكرارات الحقيقية والمتوقعة يصبح كما يلي :

الخلايا	Q	E	Q – E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
جيد / جامعي	15	12	3	9	0.75
جيد / غير جامعي	15	18	–3	9	0.5

سيء / جامعي	25	28	−3	9	0.32
سيء / غير جامعي	45	42	3	9	0.21
$\sum$	100	100			$\chi^2 = 1.78$

3 . اتخاذ القرار: من خلال جداول التوزيع χ^2 عند مستوى $\alpha = 0.05$ ودرجات حرية

$v = (2 . 1)(2 - 1) = 1$ ، نجد أن : $\chi^2_{(0.05.1)} = 3.841$

ومن خلال المقارنة:

$\chi^2 = 1.78 < \chi^2_{(\alpha ،v)} = 3.841$ ، لذلك نقبل فرضية الاستقلال ونقول بأنه لا توجد علاقة بين المؤهل العلمي (جامعي، غير جامعي) وبين إتقان العمل (جيد، سيء) الظاهرتان مستقلتان .

جداول (2 × 3) :

تمرين (3): إن البيانات التالية توضح نتائج دراسة حول تسلم المرأة لمناصب سياسية:

الآراء / النوع	مؤيد	معارض	محايد	$\sum$
ذكور	80	40	40	160
إناث	100	80	60	240
$\sum$	180	120	100	400

والمطلوب: اختبار فيما إذا كان هناك علاقة بين الرأي والجنس عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة فرضية العدم (فرضية الاستقلال) : لا توجد علاقة بين الآراء والجنس وكل ظاهرة مستقلة عن الأخرى.

2 . الاختبار الإحصائي:

$$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

ونحسب التكرارات المتوقعة:

$$E1 = \frac{(180)(160)}{400} = 72 \qquad E_3 = 48 \qquad E_5 = 40$$

$$E2 = \frac{(120)(160)}{400} = 108 \qquad E_4 = 72 \qquad E_6 = 60$$

وجدول تحديد التكرارات الحقيقية والمتوقعة:

الخلايا	Q	E	Q – E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
ذكور مؤيد	80	72	8	64	0.89
إناث مؤيد	100	108	8	64	0.59
ذكور معارض	40	48	8	64	1.33
إناث معارض	80	72	8	64	0.89
ذكور محايد	40	40	0	0	0
إناث محايد	60	60	0	0	0
$\sum$	400	400			χ^2 = 3.7

3 . اتخاذ القرار: من خلال جداول التوزيع χ^2 عند مستوى α = 0.05 ودرجات حرية

v = (2 . 1) * (2 – 1) = 1 ، نجد أن χ^2 (0.05.2) = 5.99 وبالمقارنة نجد أن :

$\chi^2 = 3.7 < \chi^2_{(\alpha,r)} = 5.99$ وبالتالي نقبل فرضية الاستقلال ونقول بأنه لا توجد علاقة بين الآراء والنوع .

جداول (3 × 3)

تمرين (4): قام عالم اجتماع بإجراء تجربة لتحديد ما إذا كان الإنجاز العلمي لدكاترة الجامعات يتأثر بحالتهم الاجتماعية، حيث يتم تصنيف الحالة الزوجية إلى أعزب ومتزوج ومطلق، كما تم تصنيف المستوى العلمي للدكاترة إلى جيد متوسط، سيء، وكانت النتائج كما يلي:

الحالة الزوجية / الإنجاز العلمي	أعزب	متزوج	مطلق	$\sum$
جيد	56	71	12	139
متوسط	47	163	38	248
سيء	14	42	85	141
$\sum$	117	276	135	538

والمطلوب: اختبر الفرض القائل بعدم وجود علاقة بين الإنجاز العلمي والحالة الزوجية عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة فرضية (فرضية الاستقلال) : لا توجد علاقة بين الآراء والجنس وكل ظاهرة مستقلة عن الأخرى.

2 . الاختبار الإحصائي:

$$\chi^2 = \sum \frac{(Q-E)^2}{E}$$

ونحسب التكرارات المتوقعة:

$$E_1 = \frac{(189)(117)}{528} = 30.80 \qquad E_3 = 35.54$$

$$E_2 = 72.66 \qquad E_4 = 54.95$$

وكذلك E_5, E_6, E_7, E_8 .

وجدول التكرارات الحقيقية والمتوقعة يأخذ الشكل التالي:

الخلايا	Q	E	Q – E	$(Q-E)^2$	$\frac{(Q-E)^2}{E}$
جيد أعزب	56	30.80	25.2	635.04	20.62
جيد متزوج	71	72.66	−1.66	2.75	0.04
جيد مطلق	12	35.54	−23.54	554.13	15.59
متوسط أعزب	47	54.95	−7.95	63.20	1.15
متوسط متزوج	163	129.64	33.36	1112.89	8.58
متوسط مطلق	38	63.41	−25.41	645.21	10.18
سيء أعزب	14	31.24	−17.24	297.21	9.51
سيء متزوج	42	73.70	−31.7	1004.89	13.63
سيء مطلق	85	36.05	48.95	2396.10	66.47
$\sum$	528	528			χ^2 = 145.77

3 . اتخاذ القرار: من جداول التوزيع χ^2 عند مستوى $\alpha = 0.05$ ودرجات حرية $v = (3 . 1)\ (3 - 1) = 4$ ، نجد أن : $\chi^2\ (0.05.4) = 9.487$

ومن خلال المقارنة: $\chi^2 = 145.77 > \chi^2_{(\alpha ،v)} = 9.487$ ، لذلك نقبل فرضية الاستقلال ونقول بأنه لا توجد علاقة بين الإنجاز العلمي والحالية الزواجية لدكاترة الجامعات.

9-4- تصحيح ياتس:

يستخدم هذا التصحيح في اختبارات الاستقلال لجداول التوافق (2 × 2) و(3 × 2) إذا كان أحد التكرارات المتوقعة أقل من (10) في إحدى الخلايا، ويقوم على طرح (½) من كل تكرار حقيقي أكبر من التكرار المتوقع E المقابل له، وإضافة (½) إلى كل تكرار حقيقي Q إذا كان أقل من التكرار المتوقع E المقابل له (وذلك لجميع التكرارات الحقيقية). وينتج عن هذا التصحيح تقليل الفرق بين التكرارات الحقيقية والتكرارات المتوقعة باعتبار أن التكرارات الحقيقية تعتمد على قيم صحيحة، مع العلم أن التكرارات الحقيقية تختلف وتقفز قفزات ذات درجات منفصلة ، بينما جدول χ^2 الذي يمثل توزيع درجات χ^2 يعتبر ذا قيم متصلة ، لذلك يستخدم من أجل تخفيف هذا القفزات ، ومع ذلك عندما تكون التكرارات (في العينة) كبيرة الحجم لا نجد ضرورة لاستخدام هذا التصحيح ويصبح دون معنى.

تمرين: أجرى أحد الباحثين الاجتماعيين بحثاً عن دور الاتصال عن طريق معرفة اتجاهات الناس إزاء برامج الإذاعة فسأل عينة من الأفراد عددها (43) شخصاً السؤال التالي:

هل تجد من الأسهل أن تستمع إلى الأخبار بدلاً من قراءتها، فأجاب (بنعم) .

(10) أشخاص من أصل (19) شخصاً من رجال الأعمال ، وأيضاً أجابوا (بنعم)

(20) شخصاً من أصل (24) شخصاً من العمال ، فإذا علمت أن الإجابات تنحصر بنعم أو لا، والسؤال: اختبر فيما إذا كان هناك علاقة بين نوع الإجابة والمهنة عند مستوى معنوية $\alpha = 0.05$.

الحل:

1 . صياغة فرضية العدم (فرضية الاستقلال) : لا توجد علاقة بين نوع الإجابة (نعم، لا) والمهنة، وكل منهما تستقل عن الآخر.

2 . الاختبار الإحصائي: $\chi^2 = \sum \frac{(Q-E)^2}{E}$

المهنة / الإجابة	رجال أعمال	عمال	نضع جدول التكرارات النظرية: ∑
نعم	10	20	30
لا	9	4	13
∑	19	24	43

ثم نحسب التكرارات المتوقعة E:

$$E_1 = \frac{(30)(19)}{43} = 13.26$$

$$E_2 = \frac{(30)(24)}{43} = 16.74$$

$$E_3 = 5.74$$

$$E_4 = 7.26$$

نلاحظ أن بعض التكرارات المتوقعة أقل من 10 ، أي E < 10 والعينة صغيرة لذلك لابد من إجراء تصحيح ياتس :

– فإذا كان E > Q نضيف ½ لـ Q

. وإذا كان E >Q نطرح ½ من Q ، كما هو واضح من خلال الجدول التالي:

الخلايا	Q	E	Q^1	$Q^1 - E$	$(Q^1-E)^2$	$\frac{(Q^1-E)^2}{E}$
1	10	13.26	10.5	−2.76	7.62	0.57
2	20	16.74	19.5	+2.76	7.62	0.45
3	9	5.74	8.5	+2.76	7.62	1.33
4	4	7.26	4.5	−2.76	7.62	1.05

χ^2 = 3.4					43	$\sum$

3 . اتخاذ القرار: من خلال جداول χ^2 عند $\alpha = 0.05$ و $v = 1$ نجد أن :

$\chi^2 = 3.84$ وبالمقارنة : $\chi^2 = 3.4 < \chi^2_{(\alpha ،v)} = 3.84$

وبالتالي نقبل فرضية الاستقلال، ونقول بأنه لا يوجد علاقة بين نوع الإجابة والمهنة.

ملاحظات:

1 . في بعض الأحيان قد تعطى البيانات على شكل نسب مئوية في هذه الحالة لابد من تحويل النسب المئوية إلى أرقام مطلقة ومن ثم إجراء الاختبار.

2 . نحن نعلم أن فكرة جداول الاستقلال (2 × 2) قائمة على فكرة الاستقلال بين الظاهرتين المدروستين أو عدم الاستقلال، لذلك في بعض الأحيان بعد أن نقوم بإجراء الاختبار للجداول (2 × 2) ونرغب بمعرفة النسبة المئوية التي تحدد مدى الارتباط أو الاستقلال بين الظاهرتين عن طريق استخدام معامل الاقتران الذي هو عبارة عن مقياس يقيس شدة العلاقة الارتباطية بين ظاهرتين في الجداول (2 × 2) وتعطى قيمته بالعلاقة:

$$r_c = \frac{AD - BC}{AD + BC}$$

وكلما اقتربت قيمة هذا المعامل من الواحد كانت العلاقة قوية، بمعنى إذا كانت $r_c > \%50$ دل ذلك على وجود علاقة بين المتغيرات وبالعكس ، لذلك يستخدم هذا المعامل أحياناً لدعم القرار المتخذ في اختبار (2 × 2) فيما إذا كان قرارنا صحيحاً أم لا. فعلى سبيل المثال لو كان قرارنا قبول فرضية الاستقلال وقمنا بحساب معامل الاقتران ، وكان $r = 0.29$ فإن هذا يدل على أن $r_c < \%50$

فالعلاقة ضعيفة بين الظاهرتين، وهذا يؤكد قرار الاختبار الإحصائي بقبول فرضية الاستقلال، أي لا علاقة بين الظاهرتين المدروستين.

3 . بالأسلوب السابق نفسه عندما يكون لدينا جداول (3 × 3) نستخدم معامل التوافق

$$r_A = \sqrt{\frac{Gr - 1}{Gr}}$$

$$Gr = \sum \left(\frac{f_{ij}^2}{f_i * f_j} \right)$$

وكلما اقتربت قيمة r_A من الواحد كانت العلاقة قوية ومتوافقة (أي توافق قوي بين البيانات)، وأيضاً يستخدم هذا المعامل لدعم القرار المتخذ في اختبار (3 × 3) فيما إذا كان قرارانا صحيحاً أم لا، فإذا كان $r_A > 0.50$ فالقرار المتعلق بالاختبار هو رفض فرضية الاستقلال، وبالعكس إذا كان $r_A < 0.50$ كان القرار هو قبول فرضية الاستقلال.

9-5- تمارين غير محلولة

تمرين (1): وجه أحد الباحثين سؤالاً لعينة عشوائية من الطلاب حجمها (500) طالب حول إجراءات الامتحانات بطريقة الأتمتة (الخيارات المتعددة) وجاءت التكرارات لبدائل خمسة على النحو:

البدائل (الآراء)	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	$\sum$
التكرارات	200	60	100	120	20	500

فإذا كانت النسبة المتوقعة طبقاً للدراسة هي:

الآراء	مؤيد بشدة	مؤيد	محايد	معارض	معارض بشدة	$\sum$
النسبة المئوية	%50	%12	%18	%17	%3	%100

والمطلوب: اختبر الفرض القائل بأن هناك تشابهاً بين نسب استجابات العينة والنسب المتوقعة طبقاً للدراسات السابقة عند مستوى معنوية $\alpha = 0.05$.

تمرين (2): تتوزع فئات أحد المجتمعات على النحو التالي:

الفئات	A	B	C	D	E	$\sum$
النسبة	%15	%30	%12	%33	%10	100

أُخذت عينة عشوائية من هذا المجتمع فجاء توزيع الفئات على النحو التالي:

الفئات	A	B	C	D	E	$\sum$
النسبة	24	40	20	44	12	140

والمطلوب: هل يمكن القول بصحة عشوائية العينة (أي هل العينة عشوائية فعلاً) ، وذلك عند مستوى معنوية $\alpha = 0.01$.

تمرين (3): البيانات التالية تمثل عدد الطلبة المسجلين في عشرة مجموعات تقدم بتدريس مقرر الإحصاء :

رقم المجموعة	1	2	3	4	5	6	7	8	9	10	$\sum$
المسجلون	29	31	20	15	30	20	23	28	25	29	250

والمطلوب: هل تدل هذه البيانات على عدم تفضيل الطلبة لأي مجموعة على المجموعات الأخرى وذلك عند مستوى معنوية $\alpha= 0.01$.

تمرين(4) : رغب أحد الباحثين معرفة رغبات المتعلمين في نوع الكتب المفضلة لديهم فاختار عينة عشوائية من (56) شخصاً وطلب من كل واحد منهم أن يختار الكتب التي يفضلها أكثر من غيرها من بين أربعة أنواع منها هي (كتب تربوية،اجتماعية ، اقتصادية ، سياسية) وطلب منهم عدم اختيار أكثر من نوع واحد فوجد أن(18) منهم يفضلون الكتب التربوية، (8) منهم يفضلون الكتب الاجتماعية، (19) منهم يفضلون الكتب الاقتصادية و(11) منهم يفضلون الكتب السياسية، والمطلوب: هل رغبات هؤلاء الأشخاص تمثل رغبات جميع السكان المتساوية في الرغبات لكل نوع من أنواع الكتب.

تمرين (5) : يعتقد أن جميع الموظفين في القطاع العام في إحدى الدول العربية بحسب خبرة سابقة يتوزع حسب الفئات التالية:

الفئات	1	2	3	4	5
النسب المئوية	%10	%15	%25	%30	%20

أُخذت عينة عشوائية تمثل هذا المجتمع حجمها (210) موظفاً فوجد أنها تتوزع على هذه الفئات على النحو التالي:

الفئات	1	2	3	4	5
عدد الموظفين	15	35	60	50	50

والمطلوب: اختبر صحة هذا الاعتقاد عند مستوى معنوية $\alpha = 0.05$

تمرين (6): تهتم إدارة إحدى الكليات بمعرفة تقديرات النجاح لجميع المقررات، وقد تم من خلال خبرة سابقة تصنف هذه التقديرات لأحد المقررات على النحو التالي:

التقديرات	ضعيف	مقبول	جيد	جيد جداً	ممتاز
النسب المئوية	%30	%30	%25	%10	%5

أخذت عينة عشوائية من (1000) طالب ، فوجد أن: (250) طالباً تقديرهم ضعيف ، (200) طالب تقديرهم مقبول ، (100) طالب تقديرهم جيد ، (150) طالباً تقديرهم جيد جداً ، (100) طالب تقديرهم ممتاز ، والمطلوب: اختبار فيما إذا كان التوزيع الحالي للتقديرات مازال نفس التوزيع السابق الذي تم الحصول عليه من خبرة سابقة عند مستوى معنوية ($\alpha= 0.05$) .

تمرين (7): في عينة عشوائية من (100) عامل من عامل الطباعة وجد أن من بين (70) عاملاً متعلماً (15) عاملاً مدخناً ومن بين غير المتعلمين وجد (14) عاملاً لا يدخنون والمطلوب:هل يمكن إرجاء صفة التدخين أصلاً إلى صفة التعلم عند مستوى معنوية $\alpha = 0.05$.

تمرين (8): البيانات التالية عبارة عن نتائج دراسة تتعلق بالزواج المدني في إحدى المدن

الرأي / الجنس	مؤيد	معارض	محايد	$\sum$
ذكور	70	15	35	120
إناث	55	10	15	80
$\sum$	125	25	50	200

والمطلوب: اختبر فيما إذا كان هناك علاقة بين الرأي والجنس

تمرين (9): في عينة عشوائية من (100) طالب وطالبة كانت أعداد الناجحين والراسبين في امتحان مادة الإحصاء كما يلي:

النتيجة / الجنس	ناجح	راسب	$\sum$
طالب	40	15	55
طالبة	35	10	45
$\sum$	75	25	100

المطلوب: اختبر فيما إذا كان هناك علاقة بين الجنس عند الطلبة ونتائجهم في الامتحان عند مستوى معنوية $\alpha = 0.05$.

تمرين (10): سحبت عينة عشوائية مكونة من (330) من الناخبين في إحدى الدول لمعرفة رأيهم في عقوبة الإعدام حيث تم تصفيتهم طبقاً لا نتمائاتهم الحزبية، والجدول التالي يمثل النتائج التي حصلنا عليها.

الرأي / الجنس	مؤيد	معارض	محايد	$\sum$
الحزب (A)	62	52	24	138
الحزب (B)	98	68	26	192
$\sum$	160	120	50	330

والمطلوب: اختبر الفرض القائل باستقلال كل من الانتماء الحزبي والرأي في عقوبة الإعدام عند مستوى معنوية $\alpha = 0.05$.

تمرين (11): في دراسة حول عمل المرأة خارج المنزل، أُخذت عينة عشوائية فكانت النتائج كالآتي:

الرأي / الجنس	مؤيد	معارض	محايد	$\sum$
ذكور	40	32	18	90
إناث	60	40	20	120
$\sum$	100	72	38	210

المطلوب: اختبار فيما إذا كان هناك بين الجنس والرأي علاقة عند مستوى معنوية $\alpha = 0.05$.

تمرين (12): بفرض أن البيانات التالية لعينة عشوائية من (1000) ناخب في إحدى الدول التي تم تصنيفهم حسب انتمائهم الحزبي وحسب تأييدهم أو معارضتهم لمبدأ معين:

الانتماء الحزبي / الرأي	الحزب A	الحزب B	$\sum$
معارض للمبدأ	250	200	450
مؤيد للمبدأ	400	150	550
$\sum$	650	350	1000

والمطلوب: اختبر الفرض القائل بعدم وجود علاقة بين الانتماء الحزبي ومعارضة أو تأييد المبدأ عند مستوى معنوية
$\alpha = 0.05$.

تمرين (13): يرغب البعض في معرفة فيما إذا المستوى العلمي في الثانوية العامة له علاقة بالاستمرار في الدراسة الجامعية، سحبت عينة عشوائية مكونة من

(1000) طالب من حملة الثانوية العامة حيث تم تصنيفهم حسب التقدير العام في الثانوية إلى:

ممتاز، جيد جداً، جيد، وحسب الاستمرار في الدراسة الجامعية أو عدم الاستمرار فيها

التقدير العام / استمرارية الدراسة	ممتاز	جيد جداً	جيد	∑
الاستمرارية	90	150	160	400
عدم الاستمرارية	110	250	240	600
∑	200	400	400	1000

والمطلوب: اختبر الفرض القائل بعدم وجود علاقة بين تقدير النجاح في الثانوية العامة وبين الاستمرار في الدراسة الجامعية عند مستوى معنوية $\alpha = 0.05$.

تمرين (14): أجريت دراسة لمعرفة ما إذا كان درجة إتقان العمل مستقلة عن التقدير العام في الدراسة الجامعية، سحبت عينة عشوائية من (1000) من العاملين وتم تصنيف بياناتهم في الجدول التالي:

درجة إتقان العمل / التقدير العام	ممتاز	متوسط	سيء	∑
جيد جداً	100	50	50	200
جيد	200	120	80	400
مقبول	200	130	70	400
∑	500	300	200	1000

والمطلوب: اختبر الفرض القائل بعدم وجود علاقة بين التقدير العام في الجامعة وبين درجة إتقان العمل.

تمرين (15): بفرض أننا نرغب في معرفة ما إذا كان ت الكثافة السكانية ومعدل الجرائم مستقلين عن بعضهما . سحبت عينة عشوائية مكونة من (1000) حي من أحياء إحدى المدن، فحصلنا على التكرارات التالية:

معدل الجريمة / الكثافة السكانية	منخفض	متوسط	مرتفع	∑
مرتفع	100	100	100	300
متوسط	250	100	50	400
منخفض	150	100	50	300
∑	500	300	200	1000

والمطلوب: اختبر الفرض القائل باستقلال المتغيرين، عند مستوى معنوية 0.01 = α.

تمرين (16): عُين مُحاضران لتدريس مقرر الإحصاء ، وكانت البيانات التالية تمثل توزيع التقديرات لكل من هؤلاء المحاضرين :

التقديرات / المحاضر	ممتاز	جيد	مقبول	∑
X	62	52	24	138
Y	98	68	26	192
∑	160	120	50	330

والمطلوب: اختبر فرض العدم القائل بعدم وجود علاقة بين التقديرات وبين نوعيه المحاضر

الفصل العاشر

مبـادئ الاحتمـالات

10-1- مفاهيم أولية (الاحتمال ، التجربـــة ، الحادثـــة)
10-2- خصائص الاحتمال (الاحتمال النظـري، الاحتمال التجريبي)
10-3- أنواع الحوادث
10-3-1- لحادث الأكيد – الحادث المستحيل – الحادث الممكن
10-3-2- الحوادث المتنافية وغير المتنافية
10-3-3- الحوادث المستقلة وغير المستقلة
10-3- 4- الحوادث الشرطية
10-4- جمع الاحتمالات وضربها
10-4-1- جمع الاحتمالات
10-4-2- ضرب الاحتمالات
10-4-3- الاحتمال الشرطـي
10-4-4- نظريــة بايــز
10-4-5- شجرة الاحتمال (الشجرة البيانية) .
10-5- تمارين غير محلولـة

وقل رب زدني علما
جامعة دمشق
Damascus University

1-10- مفاهيم أولية

إن كلمة (احتمال) تستخدم كثيراً في حياتنا اليومية . فمثلاً يقال احتمال فوز نادٍ رياضي معين في إحدى المباريات ، واحتمال سقوط المطر ، واحتمال ظهور رقم من جراء رمي حجر النرد ، واحتمال ظهور كرة حمراء أثناء عملة السحب من صندوق يحتوي كرات حمراء وبيضاء ، واحتمال أن تكون قطعة رديئة من إنتاج معين ... الخ .

إن كلمة احتمال تستخدم للتعبير عن إمكانية تحقق حدث بذاته غير مؤكد الحدوث ، لذلك فإن عدد كبير من الأفراد يستخدمون كلمة احتمال بتعبيرات مختلفة، ودرجة الثقة في حدوث حادثة ما مختلفة من شخص لأخر ، لذلك نشأت الحاجة إلى وضع مقاييس رقمية احتمالية بدلاً من التعبيرات المختلفة لوقوع الحدث أو عدمه ، وبذلك يمكن تعريف الاحتمال بما يلي : هو العدد النسبي الذي يعبر عن إمكانية تحقق حدث معين خلال التجربة .

إن العلم الذي يدرس هذه المقاييس وعلاقتها بعضها ببعض يسمى علم الاحتمالات.

بشكل عام يمكن تقسيم علم الإحصاء إلى قسمين : الإحصاء الوصفي ، والإحصاء الرياضي ، فالإحصاء الوصفي يتعلق بكيفية جمع البيانات وعرضها على شكل جداول أو رسوم بيانية ، أو يعرف بأنه عملية اختصار للبيانات في مقاييس إحصائية وصفية

(مقاييس النزعة المركزية والتشتت ، الانحدار والارتباط) أما الإحصاء الرياضي، فهو مبني على نظرية الاحتمالات ويهتم بالمستقبل غير الأكيد ، أي المستقبل المتضمن عنصر المخاطرة والذي يمكن حسابه من خلال نظرية الاحتمالات التي تساعدنا في اتخاذ القرارات المناسبة عندما يوجد هناك مخاطرة . لقد بدأت نظرية الاحتمالات بالظهور بعد الثورة الصناعية الكبرى في أوربا ودخول مفهوم الإنتاج الآلي في عملية الإنتاج ، وأيضاً ظهور المنافسة بين التجار مما جعل الاقتصاديين يهتمون بالاحتمالات لمعرفة التوقعات المحتملة وتعد نظرية الاحتمالات أو قوانين المصادفة Laws of chances من النظريات الرياضية المهمة التي تتخصص في دراسة الحوادث والمتغيرات والظواهر التي تتميز بعدم تأكد حدوثها ، ولا تبرز أهمية نظرية الاحتمالات لتطبيقاتها المهمة في شتى المجالات فحسب بل أن هناك علوماً ونظريات أخرى أصبحت ترتكز أساساً على هذه النظرية منها نظرية الإحصاء Statistics theory ونظرية القرارات Decision theory ونظرية المعلومات Information theory . لقد دخلت نظرية الاحتمالات بتطبيقاتها

المختلفة في شتى المجالات التقنية والهندسية والإنسانية والاقتصادية والعسكرية ، وبشكل عام تقوم نظرية الاحتمالات على مفهومين أساسيين :

1 - التجربـــة :

هي كل عملية تؤدي إلى ملاحظة أو قياس ظاهرة ماء ، مثال : رمي قطعة النقد في الهواء ، أو إلقاء حجر النرد في الهواء ، هذا النوع من التجارب تسمى التجارب العشوائية لأن نتائج هذا النوع من التجارب يخضع لتأثيرات عشوائية على اعتبار لو أعيدت التجربة لأكثر من مرة واحدة فأنها تعطي في كل مرة نتائج مختلفة ، وهناك أيضاً نوع أخر من التجارب تسمي التجارب المحددة حيث تكون نتائج هذا النوع من التجارب ثابتة ومستقرة لو أعيدت وبنفس الظروف في كل مرة فإنها تعطي دائماً نفس النتيجة . مثال ذلك إذا كان لديك ميزان حساس وطلب منك وزن كتلة معينة ، وأعدت عملية الوزن للكتلة بذلك الجهاز وتحت نفس الظروف لأكثر من مرة واحدة فأنك ستحصل دائماً على النتيجة نفسها .

2 - الحـادثـة :

هي الواقعة التي تحدث نتيجة التجربة أو قد لا تحدث ، مثلاً ظهور الرقم (6) عند رمي حجر النرد، وظهور الصورة عند رمي قطعة النقد في الهواء ، لذلك عند رمي حجر النرد يلاحظ أن لهذا الحجر (6) أوجهاً ولا يمكن أن نحصل إلا على وجه واحد (إما الوجه الأول أو الثاني ... الخ) ، هذه النتائج التي نحصل عليها نتيجة التجربة تدعى مجموعة الحوادث الممكنة

2-10- خصائص الاحتمال

لنرمز لاحتمال تحقق الحادث A بالرمز P(A) فنجد أنه يتمتع بالخصائص التالية :

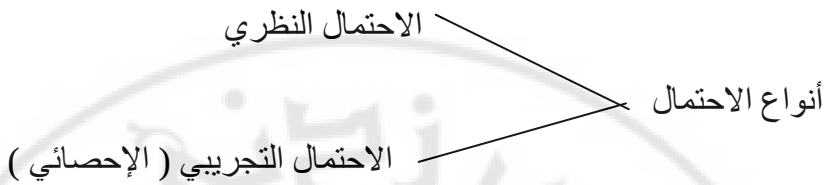
1 - يجب أن يقع الاحتمال بين الصفر والواحد مع إمكانية شموله لهذين العدديين ، أي عندما يكون صفراً فهذا احتمال الحادث المستحيل ، وإما يساوي الواحد وهذا احتمال الحصول على الحادث الأكيد ، أي أن :

$$0 \leq P(A) \leq 1$$

2 - يجب أن يساوي الصفر أو يكون أكبر منه ، وبالتالي لا يكون سالباً.

$$0 \leq P(A)$$

يجب أن يكون مجوع احتمال حصول جميع الحالات الممكنة يساوي الواحد .



10-2-1- الاحتمال النظري :

$$\text{احتمال النجاح} = \frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}}$$

يمكن تعريفه وحسابه ، كالآتي :

مثال : إن احتمال ظهور الوجه (6) عند رمي حجر النرد لمرة واحدة

$$P(A) = 1/6$$

احتمال النجاح + احتمال الإخفاق = 1

أي احتمال وقوع الحادث + احتمال عدم وقوع الحادث يساوي الواحد.

تمرين:إن احتمال ظهور الوجه (3) أو(4) عند رمي حجر النرد،ويحسب كما يلي:

$$\left.\begin{array}{l} P(A) = 1/6 = (3) \\ P(B) = 1/6 = (4) \end{array}\right\} \Rightarrow p(A \cup B) = 1/6 + 1/6 = 1/3$$

الوجه الوجه

وكذلك نجد أن احتمال عدم ظهور الوجه (3) أو (4) عند رمي حجر النرد

$$P(D) = 1 - P(A \cup B)$$
$$= 1 - (2/6) = 2/3$$

من خلال ذلك نجد أنه لحساب الاحتمال لابد من معرفة عدد الحالات الممكنة وعدد الحالات الملائمة ، إضافة لذلك ، أن الاحتمال النظري مبني على نتيجة المنطق الرياضي وليس نتيجة التجربة ، أي عُرِّف على أساس المنطق الرياضي وليس على أساس التجربة ، وهذا يعني أنه يمكن للتجربة أن تثبت عكس ذلك ، فلو رمينا حجر النرد (6) مرات ، هل يمكن ضمان الحصول على الوجه (1) مرة واحدة كل ست مرات ، و (2) مرة واحدة كل خمس مرات ، ..... الخ،

طبعاً لا يمكن ذلك ، بمعنى آخر إن الاحتمال النظري هو احتمال مثالي تقارن به النتائج الأخرى ، ويتحقق عندما يكون عدد مرات إجراء التجربة كبيراً جداً .

10-2-2- الاحتمال التجريبي (الإحصائي) :

هو الاحتمال المبني على التجربة لعدد كثير من المرات وتحديد حالة النجاح وحالة الإخفاق ومن ثم معرفة نسبة حالة النجاح من الحالات الأساسية ويمكن تعريفه على النحو التالي :

$$\text{الاحتمال التجريبي} = \frac{\text{عدد مرات النجاح}}{\text{عدد مرات إجراء التجربة}}$$

من خلال ذلك نلاحظ أن الفرق بين الاحتمال النظري والاحتمال التجريبي هو أن الاحتمال النظري يعطي النتيجة مباشرة بالحساب النظري قبل القيام بالتجربة بينما الاحتمال التجريبي هو عبارة عن التكرار النسبي للحادثة الملائمة بعد القيام بالتجربة .

تمرين : رمينا قطعة من النقد (100) مرة وكان الوجه الملائم هو الصورة وظهرت (51) مرة ، فالاحتمال التجريبي $= \frac{51}{100} = 0.51$ ؛

طبعاً يستخدم هذا الأسلوب إذا كان إمكانات التجربة غير منته ، ومع ذلك يقترب الاحتمال التجريبي من الاحتمال النظري كلما زدنا من عدد مرات إجراء التجربة ، لكن تبقى قيمة الاحتمال التجريبي قيمة تقريبية للأسباب التالية :

أ) أننا غير قادرين على إعادة التجربة من أجل كل حادث عدداً لا متناهياً من المرات .

ب) لا يمكن إحصائياً أن نحدد بشكلٍ دقيق نهاية النسبة: $\frac{\text{الحالات الملائمة}}{\text{الحالات الممكنة}}$

$n \to \infty$

ومما سبق ، يمكننا التعرض إلى ما يسمى بقانون الأعداد الكبيرة ، الذي ينص على أن الاحتمال التجريبي $\neq$ الاحتمال النظري ، إلا عندما يكون عدد مرات إجراء التجربة كبيراً جداً ، وفي هذه الحالة فإن قيمة الاحتمال التجريبي تقترب إلى حد كبير من قيمة الاحتمال النظري ، ويصبح الفرق بين القيمتين أصغر من أية قيمة موجبة مهما صغرت أي أن :

(عدد موجب صغير جداً) $\varepsilon >$ | الاحتمال النظري - الاحتمال التجريبي |

لكن على الرغم من استعمالات قانون الأعداد الكبيرة في نظرية الاحتمالات لا يعطي التأكيد بأن الاحتمال التجريبي مساوياً الاحتمال النظري بشكل مطلق ، بل على العكس هناك دوماً حالة الشك في عدم التساوي فقانون الأعداد الكبيرة ينص أن :

الاحتمال النظري $\rightarrow$ نهاية الاحتمال التجريبي

$$n \rightarrow \infty$$

وهذه حقيقة مطلقة وليست تجريبية .

ويقابل الاحتمال التجريبي في الجداول التكرارية ، التكرار النسبي الذي يساوي:

تكرار فئة ما

التكرار الكلي

والاحتمال التجريبي عبارة عن نسبة محصورة بين الصفر والواحد .

10-3- أنواع الحوادث

10-3-1- الحادث الأكيد – الحادث المستحيل – الحادث الممكن

عند حدوث التجربة يمكن أن نحصل على إحدى الحالات التالية :

1 - الحادث الأكيد : نقول عن حادث إنه أكيد ، إذا كان تحققه مؤكداً ، بمعنى احتمال وقوع الحادث الأكيد الواحد الصحيح .

مثال : ظهور أحد الأرقام الستة عند رمي حجر النرد هو حادث أكيد .

2 - الحادث المستحيل : نقول عن حادث إنه حادث مستحيل إذا كان غير قابل للتحقق أبداً ، بمعنى احتمال وقوع الحادث المستحيل يساوي الصفر.

مثال : ظهور على الرقم (7) أو الحصول عليه عند رمي حجر النرد ، هو حادث مستحيل .

3 - الحادث الممكن : هو الحادث الممكن حدوثه أو عدم حدوثه ، بمعنى احتمال وقوع الحادث تساوي قيمة محصورة بين الصفر والواحد .

10-3-2- الحوادث المتنافية وغير المتنافية

نقول عن حادثين A , B إنهما متنافيان إذا كان وقوع أحدهما ينفي وقوع الأخر ، بمعنى آخر إن الحوادث المتنافية تشكل مجموعة جزئية غير متقاطعة ، مثل أوجه النرد ، ويكون الحادثان A , B غير متنافيين إذا كان وقوع أحدهما لا يمنع من وقوع الحادث الأخر ، أي أن الحوادث غير المتنافية هي حوادث جزئية متقاطعة ، مثل ظهور عدد فردي في حجر النرد .

10-3-3- الحوادث المستقلة وغير المستقلة

نقول عن حادثين A , B إنهما مستقلان إذا كان حدوث أحدهما مستقل عن حدوث الأخر ، أي لا يؤثر احتمال حدوثه . ويكون الحادثان غير مستقلين إذا كان حدوث أحدهما يؤثر في احتمال حدوث الأخر. تجدر الإشارة إلى أن الحوادث المتنافية هي بالضرورة غير مستقلة ولكن العكس ليس بالضرورة صحيحاً، وهذا يعني أن الحوادث المستقلة ليست دائماً متنافية ، و أن الحوادث غير المتنافية قد تكون مستقلة وقد تكون غير مستقلة .

10-3-4- الحوادث الشرطية

نقول عن حادثين A , B إن أحدهما حادث شرطي للآخر ، إذا كان الحادثان A , B غير مستقلين ، بمعنى أن هذا الحادث لا يمكن أن يتحقق ما لم يتحقق الحادث الأخر.

مثال : لدينا في صندوق (10) كرات مؤلفة من 4 كرات سوداء و 6 كرات بيضاء ، ولنفرض أننا سنسحب مع الإعادة كرتين من هذا الصندوق ، ل نرمز إلى حادث السحب الأول لكرة سوداء بـ (A) فيكون احتمال تحققه $p(A)=\frac{4}{10}$ ، ولنرمز إلى حادث السحب الثاني لكرة بيضاء (B) ، ويكون احتمال تحققه $p(B)=\frac{6}{10}$.

- من خلال ذلك نلاحظ أن الحادثين (A) , (B) غير متنافيين على اعتبار أن حدوث (A) لا ينفي حدوث (B) .

- ونلاحظ أيضاً إذا كان السحب مع الإعادة ، أي أن سحب الكرة الثانية يتم بعد إعادة الكرة الأولى إلى الصندوق ، في هذه الحالة الحادثان (A) , (B) مستقلان عن بعضهما بعض. أما إذا كان السحب مع عدم الإعادة ، أي إذا كان سحب الكرة الثانية يتم دون إعادة الكرة الأولى إلى الصندوق ، ففي هذه الحالة يكون الحادثان

(A) , (B) غير مستقلين بعضهما عن بعض ، لأن $p(A)=\frac{4}{10}$ ويكون $p(B/A)=\frac{6}{9}$ ، حيث رمزنا بـ $p(B/A)$ للحادث الشرطي الذي يُشترط لتحقق حدوث (B) حدوث (A) بصورة مسبقة ويقال عن الحادث (B) إنه حادث شرطي .

- الحوادث المتكاملة : هي الحوادث التي تشكل مجموعة متنافية من جميع الحوادث الممكنة .

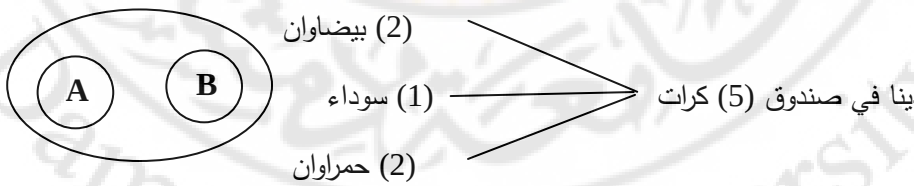
10-4- جمع الاحتمالات وضربها

10-4-1- جمع الاحتمالات

أ - الحوادث متنافية : إذا كان الحادثان A , B متنافيين بالتبادل بمعنى حدوث أحدهما ينفي وقوع الآخر، فإن احتمال حصول واحد منهما أو كليهما يساوي مجموع حدوث كلٍ منهما ، على اعتبار أن احتمال اتحاد أو اجتماع حادثين غير متقاطعين هو مجموع احتمال كل منهما ، وأن قاعدة الجمع للحوادث المتنافية هي:

$$P(A \cup B) = P(A) + P(B)$$

تمرين :



سحبنا كرة واحدة بشكل عشوائي من هذا الصندوق ، والمطلوب : ما احتمال الحصول على كرة بيضاء أو سوداء ؟

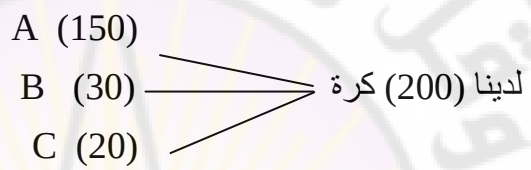
الحل :

من خلال التمرين نلاحظ أن الحوادث متنافية (بيضاء أو سوداء) لذلك في هذه الحالة نطبق فرضية جمع الاحتمالات :

$$\left.\begin{array}{l} P(A \cup B) = P(A) + P(B) \\ = 2/5 + 1/5 = 3/5 \end{array}\right\}$$

تمرين : يوجد في صندوق كبير 200 كرة ، منها (150) كرة طبعنا عليها حرف (A) و (30) كرة طُبع عليها (B) و (20) كرة طُبع عليه ا حرف (C) ، والمطلوب : ما احتمال سحب كرة من النوع الأول (A) أو النوع الثاني (B) . مع العلم أن جميع الكرات من الحجم نفسه .

الحل :



من خلال النص نجد أن الحوادث متنافية لذلك نطبق فرضية جمع الاحتمالات:

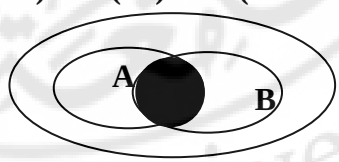
$$\left.\begin{array}{l} P(A) = 150/200 \\ P(B) = 30/200 \end{array}\right\} P(A \cup B) = P(A) + P(B)$$

$$= 150/200 + 30/200$$

$$= 180/200$$

ب - الحوادث غير المتنافية أو المتقاطعة : إذا كان وقوع أحدهما لا يمنع وقوع الآخر، أي أن هناك إمكانية حدوث A , B بآن واحد ، في هذه الحالة احتمال الحدوث يساوي إلى مجموع احتمال وقوع كل منهما ناقصاً احتمال وقوعهما معاً في أن واحد .

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



تمرين : لدينا (20) كرة مرقمة من (1) إلى (20) ، ما احتمال سحب كرة تحمل رقم يقبل القسمة على (4) أو على (5) .

الحل :

لو أمعنا النظر في التمرين لوجدنا أن الحوادث غير متنافية ، على اعتبار وقوع أحدهما لا يمنع وقوع الآخر ، ومن جهة أخرى نلاحظ أن أرقام الكرات التي تقبل القسمة على (4) هي التالية :

(4) ، (8) ، (12) ، (16) ، (20)

5 4 3 2 1

وأرقام الكرات التي تقبل القسمة على (5) هي (5) ، (10) ، (15) ، (20)

1 2 3 4

وهنا نلاحظ بأن الكرة رقم (20) تقبل القسمة على (4) و (5) بآن واحد ، وبذلك يتحقق (التقاطع) وهي إمكانية حدوث A , B بآن واحد .

الحل :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 5/20 + 4/20 - 1/20 = 8/20$$

10-4-2- ضرب الاحتمالات

إذا كانت الحوادث مستقلة ، بمعنى حدوث أحدهما مستقل عن حدوث الآخر ، أي لا يؤثر في احتمال تحققه ، وفي هذه الحالة نجد أن احتمال حدوثها معاً يساوي جداء احتمال حدوث كل واحد من هذه الحوادث ، أي :

$$P(A \cap B) = P(A) \cdot P(B)$$

تمرين : لو رمينا قطعتي نرد مرة واحدة فما هو احتمال الحصول على الوجه (6) في كلا القطعتين .

الحل :

من خلال ذلك نلاحظ أن الحوادث مستقلة وبالتالي نطبق قاعدة ضرب الاحتمالات

$$P(A \cap B) = P(A) \cdot P(B)$$

$$\left.\begin{matrix} P(A)=1/6 \\ P(B)=1/6 \end{matrix}\right\} = 1/6 \cdot 1/6 = 1/36$$

النرد الأول

النرد الثاني

تمرين : ثلاثة دواليب حظ يحمل كل منها (10) أرقام من الصفر إلى الـ (9) أديرت الدواليب معاً ، ما احتمال ظهور الرقم (159) .

الحل :

إن الحوادث في هذا التمرين مستقلة لذا نطبق قاعدة ضرب الاحتمالات .

لدينا :

$$\left.\begin{array}{l} P(A)=1/10 \\ P(B)=1/10 \\ P(C)=1/10 \end{array}\right\} P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$$

$$p(A\cap B\cap C)=1/10\times 1/10\times 1/10=1/1000$$

10-4-3- الاحتمال الشرطي

إذا كان A , B حادثين معرفين في فضاء العينة (S) فإن احتمال حدوث A علماً أن B قد حدث يساوي إلى النسبة بين احتمال حدوث تقاطع A, B معاً مقسوماً على احتمال B .

والتعبير المتبع لهذا الاحتمال هو P (A / B) ويقرأ احتمال حدوث (A) علماً أن الحادث (B) قد وقع ويعرف على النحو التالي $P(A/B)=\frac{P(A\cap B)}{P(B)}$ و

$P(B)>0$

تمرين :

إذا كان احتمال أن ينجح عارف هو (1/3) واحتمال أن ينجح عارف وأمين هو (1/4) . أوجد احتمال نجاح أمين إذا علمت أن عارف قد نجح :

الحل :

نرمز بـ (B) لنجاح عارف واحتماله يساوي P(B)=1/3 ، ونرمز بـ (A) لنجاح أمين ، فيكون $A\cap B$ نجاح عارف وأمين ، وإن احتمال تحققه

$p(A \cap B) = \frac{1}{4}$ ، ويكون B / A نجاح أمين بشرط نجاح عارف :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/3} = 3/4 .$$

تمرين : في إحدى المدن (40 %) من المواطنين لهم شعر أسود اللون ، (25 %) لهم عيون سوداء اللون ، (15 %) لهم شعر أسود وعيون سوداء اختير مواطن من هذه المدينة بطريقة عشوائية ، والمطلوب :

- ما احتمال أن تكون عيناه سوداوين إذا كان شعره أسود .
- ما احتمال أن يكون شعره أسود إذا كانت عيناه سوداوين .

الحل :

نلاحظ أنه لدينا :

(15 %)	(25 %)	(40 %)
شعر وعيون سوداء	عيون سوداء	شعر أسود

نستخدم الرموز التالية :

شعرهم غير أسود $\overline{B}$, (40 %) شعرهم أسود (B)

عيونهم غير سوداء $\overline{A}$, (25 %) عيونهم سوداء (A)

(15 %) شعرهم وعيونهم سوداء ($A \cap B$)

- وبذلك نجد أن احتمال أن تكون عيناه سوداوين إذا كان شعره أسود :

$$\mathbf{P(A/B) = \frac{P(A \cap B)}{P(B)}}$$

$P(A \cap B)=15/100$ احتمال أن يكون للمواطن شعر أسود وعيون سوداء .

$P(B) = 40/100$ احتمال أن يكون المواطن ذا شعر أسود .

$P(A/B) = \frac{15/100}{40/100} = 15/40$ احتمال أن تكون عيناه سوداوين إذا كان شعره أسود .

- وكذلك نجد أن احتمال أن يكون شعره أسود إذا كانت عيناه سوداوان :

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

$$P(B \cap A) = 15/100$$

$$P(A) = 15/100$$

$$P(B/A) = \frac{15/100}{25/100} = 15/25$$

تمرين : تشير السجلات لإحدى الشركات إلى أن احتمال غياب أحد العمال يوم الإثنين هو (0.4) ، يوم الجمعة هو (0. 2) ويوم الإثنين ويوم الجمعة هو = (0.1)،احسب احتمال غياب العامل يوم الجمعة، علماً بأنه كان متغيباً يوم الإثنين ؟

لنرمز لحادث تغيبه يوم الجمعة بـ A ولحادث تغيبه يوم الإثنين بـ B ، فنجد أن :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.2} = 0.5$$

10-4-4- نظرية بايز

تهتم نظرية بايز في حساب احتمال أن يكون هناك سبب ما هو مصدر حدوث حادثة معينة ، وأن هذه الحادثة يرجع حدوثها إلى عدد من الأسباب المعروف احتمال حدوث كل منها، أي أن التجربة الإحصائية متعددة المراحل ويجب علينا تحديد احتمال كل مرحلة من هذه المراحل،وتعطي بالعلاقة التالية :

$$P(A_i/B) = \frac{P(B/A_i)\,P(A)}{\sum_{k=1}^{n} P(B/A_k) \cdot P(A_k)}$$

تمرين : لدينا (3) صناديق،الأول يحتوي (3) كرات بيضاء و (3) كرات سوداء والثاني يحتوي على (1) كرة بيضاء و (3) كرات سوداء ، والثالث يحتوي على (3) بيضاء و(2) سوداء ، أثناء عملية السحب العشوائي لكرة واحدة من الصناديق

حصلنا على كرة بيضاء ، ما احتمال أن تكون هذه الكرة قد سحبت من الصندوق الأول ؟

الحل :

نستخدم الرموز التالية : احتمال السحب من الصندوق الأول $P(A_1)$ ، واحتمال السحب من الصندوق الثاني $P (A_2)$ ، واحتمال السحب من الصندوق الثالث $P (A_3)$ ، وعليه يكون :

$$P (A_1) = 1/3$$
$$P (A_2) = 1/3$$
$$P (A_3) = 1/3$$

(الحدث الثاني) نرمز إلى الكرة البيضاء (B) ، وبـ $P (B / A)$ لاحتمال سحب الكرة البيضاء من أحد الصناديق ، فنجد أن :

احتمال سحب كرة بيضاء من الصندوق الأول $P (B / A_1) = 3/6$.

احتمال سحب كرة بيضاء من الصندوق الثاني $P (B / A_2) = 1/4$.

احتمال سحب كرة بيضاء من الصندوق الثالث $P (B / A_3) = 3/5$.

بتطبيق نظرية بايز :

$$\mathbf{P(A/B)} = \frac{\mathbf{P(A)} \cdot \mathbf{P(B/A)}}{\sum_1^n \mathbf{P(A)} \cdot \mathbf{P(B/A)}}$$

احتمال سحب كرة بيضاء من أحد الصناديق

$$P(A_1/B) = \frac{P(A_1)\cdot P(B/A_1)}{P(A_1)\cdot P(B/A_1)+P(A_2)\cdot P(B/A_2)+P(A_3)\cdot P(B/A_3)} = \frac{\frac{1}{3}*\frac{3}{6}}{\frac{1}{3}*\frac{3}{6}+\frac{1}{3}*\frac{3}{5}+\frac{1}{3}*\frac{3}{4}}$$

وهو المطلوب .

10-4-5- شجرة الاحتمال (الشجرة البيانية)

تستخدم لحساب احتمال حادث معين ما ، وهذا الحادث هو نتيجة لحوادث متعددة، ويمثل أصل الشجرة فضاء العينة و تمثل فروع الشجرة تقسيم لفضاء العينة ، وتستخدم الشجرة البيانية أيضاً في تمثيل نتائج التجربة الإحصائية متعددة المراحل .

طرق السحب العشوائي :

- السحب بإرجاع : أي عند سحب (n) كرة من صندوق به (N) من الكرات فإذا سحبت الكرة الأولى وأعيدت هذه الكرة إلى الصندوق ، قبل إجراء السحب الثاني ، في هذه الحالة يكون لدينا السحب بإرجاع .

- السحب بدون إرجاع : يقال عن السحب بدون إرجاع إذا لم يتم إعادة الكرة إلى الصندوق ويكون عدد طرق سحب الكرة الثانية بدون إرجاع الكرة الأولى (N - 1) طريقة ، وهكذا لباقي السحب .

تمرين : صندوق يحتوي (10) بطاريات جافة متشابهة منها (7) بطاريات صالحة و (3) تالفة، اختيرت بطاريتان عشوائياً وبدون إرجاع . والمطلوب :

– ما احتمال أنهما صالحتان .

– ما احتمال أن تكون واحدة صالحة والأخرى تالفة .

– ما احتمال أن البطاريتين تالفتان .

الحل :

نستعين برسم الشجرة البيانية في حل هذه المسألة

1 - $P(A/A) = (7/10)(6/9)$

2 - $P(B/A) + P(A/B) = (7/10\ (3/9) + (3/10)(7/9)$

3 – $P(B/B) = p(C) = (3/10)(2/9)$

ونحن بشكل عام من المفروض أن نتأكد من البداية أن مجموع الاحتمال ممكنة = 1

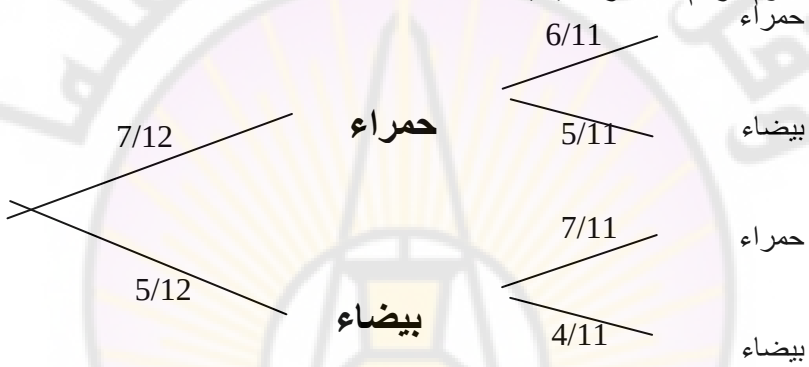
$(7/10\ (6/9) + (7/10)(3/9) + (3/10\ (7/9) + (3/10)(2/9) = 1$

تمرين : يوجد في جعبة (12) كرة متشابهة منها (5) كرات بيضاء و (7) كرات حمراء ، سحبت كرتان من الجعبة عشوائياً بدون إرجاع . والمطلوب :

– ما احتمال أن الكرتين حمراوان .

– ما احتمال أن تكون كرة حمراء وكرة بيضاء .

الحل : نقوم برسم الشجرة البيانية :

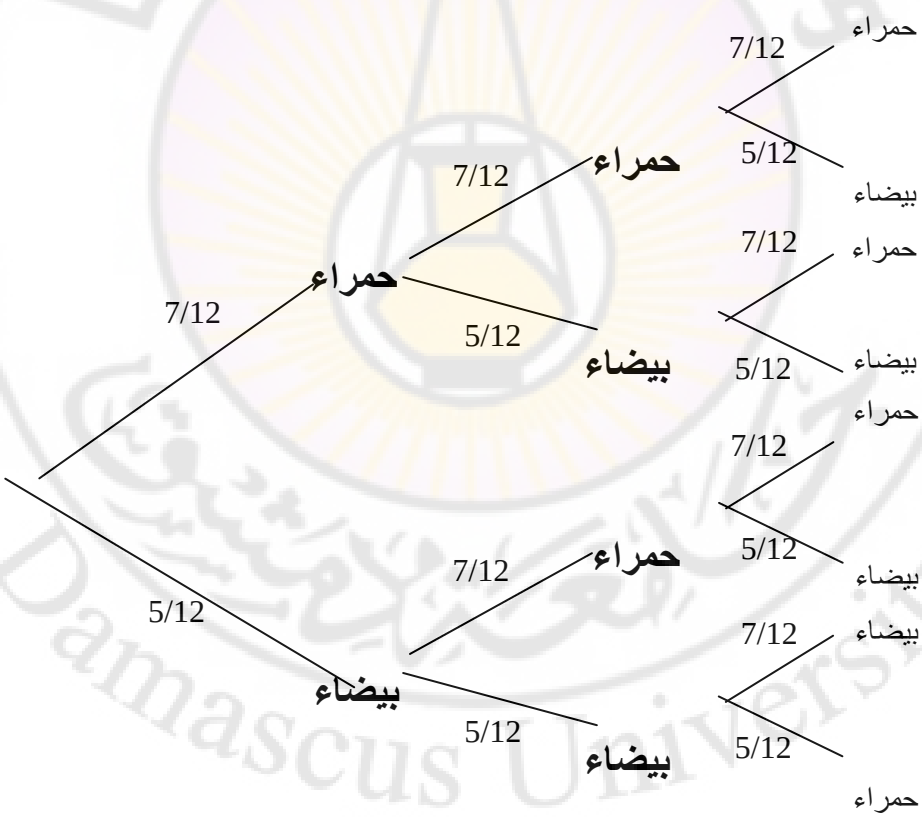


1- $P(A/A) = (7/12)(6/11)$

2 - $P(B/A) = (7/12)(5/11) + (5/12)(7/11)$

فيما لو أعدنا التمرين السابق وسحبت ثلاث كرات بشكل عشوائي ومستقل، والسحب بإرجاع ،

والمطلوب : أوجد الاحتمالات التالية باستخدام الشجرة البيانية .


7/12
حمراء
5/12
بيضاء
حمراء
7/12
حمراء
5/12
بيضاء
7/12
حمراء
5/12
بيضاء
حمراء
7/12
حمراء
5/12
بيضاء
بيضاء
7/12
بيضاء
5/12
حمراء

3 - احتمال أن تكون الكرات الثلاث حمراء :

$$P(B) = (7/12)\ (7/12)\ (7/12) = (7/12)^3$$

4 - ما احتمال أن تكون الكرات الثلاث بيضاء :

$$P(A) = (5/12)\ (5/12)\ (5/12)$$

5 - ما احتمال أن تكون كرتان حمراوين وواحدة بيضاء :

$$P(C) = [\ (7/12)\ (7/12)\ (5/12)\ ] \times 3$$

6 - ما احتمال أن تكون كرة واحدة حمراء على الأقل :

$$P(D)=(7/12)\times 3+[(7/12)(7/12)(5/12)] \times 3+[(7/12)(7/12)(5/12)]\times 3$$

10-5- تمارين غير محلولة

تمرين (1) : في أحد الصفوف (30) طالباً وطالبة . منهم (10) طلاب ذكور والباقي إناث نصف أعين الطلاب والطالبات من اللون الأزرق ، اختبر أحد أعضاء الصف بطريقة عشوائية . ما احتمال أن يكون الشخص المختار طالباً أو عيونه زرقاوين .

تمرين (2) : لدى تاجر ثلاث مجموعات من المصابيح :

أ – تحتوي المجموعة الأولى على (500) مصباح منها (2 %) معطوبة .

ب – تحتوي المجموعة الثانية على (300) مصباح منها (5 %) معطوبة .

جـ- تحتوي المجموعة الثالثة على (200) مصباح منها (1 %) معطوبة .

وقد تم خلط تلك المصابيح جيداً ثم اختير مصباح بصورة عشوائية فتبين أنه معطوب .

والمطلوب : ما احتمال انتماء ذلك المصباح إلى كل مجموعة من المجموعات الثلاث .

تمرين (3) : يحتوي صندوق (4) كرات حمراء مرقمة من (1) إلى (4) ، و (3) كرات بيضاء مرقمة من (1) إلى (3) ، وجميع هذه الكرات من الحجم نفسه ، والمطلوب :

- سحبت بصورة عشوائية كرة من الصندوق ، ما احتمال أن تكون هذه الكرة حمراء أو تحمل رقم زوجي؟

- أعيدت الكرة المسحوبة سابقاً ، ثم سحبت الكرات بالتتالي ، ما احتمال تعاقب الكرات الحمراء والبيضاء في عملية السحب .

تمرين (4) : ثلاث صيادين أطلقوا النار على هدف ، الأول رمى طلقه واحدة والثاني رمى طلقتين والثالث أربع طلقات ، والمطلوب :

- ما احتمال أصابه الهدف .

- ما احتمال عدم أصابه الهدف .

تمرين (5) : أطلق رامي ثلاث طلقات متلاحقة على هدف ، إذا علمت أن احتمال إصابة الهدف (0.6) والمطلوب :

- ما هو احتمال إصابة الهدف بطلقة واحدة فقط .

- احسب عدد الطلقات الواجب رميها على الهدف ليكون احتمال إصابة الهدف (0.9) على الأقل .

تمرين (6) : ثلاثة رجال يطلقون النار على هدف واحد وكل منهم يطلق طلقه واحدة فإذا كان احتمال إصابة الأول للهدف (0.4) واحتمال إصابة الثاني لهدف (0.5) واحتمال إصابة الثالث لهدف (0.1) ، والمطلوب :

- ما احتمال إصابة الهدف بطلقة واحدة .

- ما احتمال إصابة الهدف بطلقتين فقط .

- ما احتمال إصابة الهدف بثلاث طلقات .

- ما احتمال إصابة الهدف بطلقة واحدة على الأقل .

تمرين (7): يوجد في أحد الصفوف الدراسة (30) طالباً وطالبة ، (12) طالباً من بينهم (4) طلاب متفوقين و (18) طالبه من بينهن (8) طالبات متفوقات ، أختبر أحد الطلبة بصورة عشوائية ليكون ممثلاً عن هذا الصف . والمطلوب :

ما احتمال أن يكون الممثل :

- طالباً متفوقاً

- طالبة متفوقة

- متفوقاً (طالباً أو طالبة)

تمرين (8) : في أحد المختبرات (10) حاسبات ، (6) منها تعمل بلغة البيسك فقط و (3) تعمل بلغة الفورتران فقط و (1) تعمل بكلتا اللغتين ، فإذا اختيرت إحدى الحاسبات بصورة عشوائية ، ما احتمال :

- أن تكون الحاسبة تعمل بلغة البيسك ولا تعمل بلغة الفورتران .

- أن تكون الحاسبة تعمل بلغة البيسك والفورتران معاً .

- أن تكون الحاسبة تعمل بلغة البيسك .

- أن تكون الحاسبة تعمل بلغة الفورتران .

- أن تكون الحاسبة تعمل بلغة البيسك أو الفورتران .

تمرين (9) : ثلاثة متبارين a ,b ,c يصوبون معاً إلى هدف واحد واحتمال إصابتهم للهدف على الترتيب 2/5 , 1/2 , 1/4 . والمطلوب :

- ما احتمال إصابة اثنين منهم للهدف عندما يصوب الثلاثة معاً بنفس الوقت.

تمرين (10) : لنفترض أنه لدينا مصنعاً للغزل والنسيج ، حيث إن نسبة القطع المعطوبة في مرحلة الغزل هي (% 20) وفي مرحلة النسيج (% 15) وفي مرحلة الصباغة (% 10) ، سحب ثوب من الإنتاج الكلي عشوائياً فكان معطوباً . والمطلوب : ما احتمال أن يكون هذا العطب قد حصل في مرحلة الغزل .

تمرين (11) : ليكن لدينا مصنع مكون من (3) مجموعات من الآلات M1 , M2 M3 , ولنفترض أن (% 90) من إنتاج المجموعة الأولى سلع جيد ة و (% 80) من إنتاج المجموعة الثانية هو أيضاً سلع جيد ة و (% 25) من إنتاج المجموعة الثالثة هو سلع معطوبة . والمطلوب :

أ - أثناء تجميع الإنتاج الكلي للسلع ، سحبت سلعة من هذا الإنتاج عشوائياً ، ما احتمال أن تكون السلعة المسحوبة معطوبة .

ب - أثناء تجميع الإنتاج الكلي للسلع سحبت سلعة عشوائياً فكانت معطوبة ، ما احتمال أن تكون هذه السلعة قد أنتجت من المجموعة الأولى .

تمرين (12) : يتم تصويب بندقيتين (I , II) على نفس الهدف ، أطلقت (9) طلقات من البندقية (I) في نفس الوقت الذي أطلقت فيه (10) طلقات من البندقية (II) ، فإذا علمت أنه بالمتوسط من كل (10) طلقات من البندقية (I) (8) فقط تصيب الهدف ومن البندقية (II) (7) فقط تصيب الهدف ، وأثناء التصويب أصيب الهدف بطلقة ولم يكن معلوماً من أي بندقية جاءت الطلقة . والمطلوب :

1- ما هو احتمال إصابة الهدف من البندقية II .

2- ما هو احتمال إصابة الهدف من البندقية I .

3- ما هو احتمال إصابة الهدف بطلقة .

تمرين (13) : لدينا (5) صناديق من المرطبات تحتوي الصناديق (1 ,2 , 3) على (10) زجاجات عصير برتقال و (10) ليمون ، ويحتوي الصندوقين (4 , 5) على (5) زجاجات برتقال و (15) ليموناً ، والمطلوب :

1- سحبنا بشكل عشوائي زجاجة من أحد الصناديق ، ما احتمال أن تكون الزجاجة عصير برتقال .

2- سحبنا مع الإعادة (3) زجاجات من الصندوق الثالث ، ما احتمال أن تكون الزجاجة عصير ليمون .

تمرين (14) : لون الوجه ين الأول والثاني من حجر النرد باللون الأحمر ® والوجهان الثالث والرابع باللون الأصفر (J) والوجه ين الخامس والسادس باللون الأزرق (B)، والمطلوب :

1- حدد الفضاء الاحتمالي لتجربة رمي حجر النرد مرتين متتاليتين واحسب احتمال الحصول على كل من نتائج هذه التجربة .

تمرين (15) : صندوقان A , B يحوي الصندوق A على (9) أوراق مرقمة من (1 إلى 9) ويحوي الصندوق B على (17) ورقة مرقمة من (1 إلى 17) . اختير صندوق من الصندوقين بطريقة عشوائية واختيرت منه ورقة ، والمطلوب:

1- ما احتمال أن تكون الورقة المختارة تحمل رقماً زوجياً .

2- إذا كان رقم الورقة المختارة زوجياً ، فما احتمال أن تكون الورقة المختارة من الصندوق A .

تمرين (16) : شركة أحذية لها ثلاثة مصانع إنتاجية ، المصنع الأول ينتج 40) (% من إنتاج الشركة وفيه (% 2) من البضاعة معطوبة ، والمصنع الثاني ينتج (% 35) من إنتاج الشركة وفيه (% 10) من البضاعة معطوبة، أما المصنع الثالث فينتج (% 25) من إنتاج الشركة وفيه (40%) من البضاعة تكون معطوبة، فإذا اشتريت حذاء من إنتاج هذه الشركة ما احتمال أن يكون :

1 – معطوباً ؛ 2 - غير معطوب ؛ 3 - إذا اشتريت حذاء من إنتاج هذه الشركة ووجدته معطوباً ، ما احتمال كونه من المصنع الأول أو الثاني .

تمرين (17) : إذا علمت أن الصندوق (A) فيه (3) كرات بيضاء و (5) كرات سوداء وفي الصندوق (B) (5) كرات سوداء و (3) كرات بيضاء ، سحبنا عشوائياً مع الإعادة كرتين من الصندوق (A) وسحبنا بدون إعادة كرتين من الصندوق (B) . والمطلوب :

1- أحسب احتمال الحصول على (4) كرات بيضاء .

2- أحسب احتمال الحصول على كرة سوداء واحدة على الأقل .

تمرين (18) : يحتفظ مشفى بسيارتي إسعاف للطوارئ واحتمال أن تكون السيارة جاهزة للتحرك عند الحاجة إليها هو (0,9) إذا علمت أن توفر إحدى السيارتين مستقل عن الأخرى . والمطلوب :

1- أوجد احتمال أن لا تتوفر أي منهما .

2- ما هو احتمال تلبية الطلب لسيارة إسعاف في حالة الطوارئ .

تمرين (19) : إذا كان لدينا احتمال هطول مطر في يوم معين (0,1) واحتمال وجود رياح نشطة في ذلك اليوم (0,05) واحتمال وجود مطر ورياح نشطة (0,03) . والمطلوب :

1- أوجد احتمال وجود مطر أو رياح نشطة في ذلك اليوم .

2- أوجد احتمال هطول مطر في ذلك اليوم علماً أن الرياح نشطة .

تمرين (20) : كيس يحتوي على (25) كرة متماثلة منها (10) كرات بيضاء و (5) كرات سوداء والباقي حمراء ، سحبت عينه من ثلاث كرات الواحدة تلو

الأخرى عشوائياً . أوجد احتمال أن تكون الكرات الثلاث من اللون نفسه ، إذا كان : 1- السحب مع الإعادة ؛ 2- السحب بدون إعادة .

تمرين (21) : يتنافس أحمد ، محمد ، حسين على المرتبة الأولى في امتحان مقرر الإحصاء ، فإذا كان احتمال فوز أحمد = (2/3) احتمال فوز محمد ، واحتمال فوز حسين ضعف احتمال فوز محمد ، فما هو احتمال فوز كل منهم .

تمرين (22) : قذفت قطعة نقود مرات عديدة في الهواء فكان احتمال ظهور الصورة ضعف احتمال ظهور الكتابة في كل مرة . والمطلوب : أوجد احتمال ظهور الصورة والكتابة .

تمرين (23) : إذا كانت (A , B) حادثتين،حيث كان P (A ∪ B)=3/4 ، P(B) 2/5 ، P(A) = x . أوجد قيمة x في كل من الحالات التالية :

1- إذا كان A و B متنافيتين .

2- إذا كان A و B مستقلتين .

تمرين (24) : يتضمن صندوق (6) قطع حمراء مرقمة من (1) إلى (6) وكذلك (6) قطع بيضاء مرقمة من (1) إلى (6) سحبنا قطعة بصورة عشوائية . والمطلوب :

1- ما احتمال أن تكون حمراء .

2- ما احتمال أن تكون عليها رقم زوجي .

3- ما احتمال أن تكون حمراء وعليها رقم زوجي .

تمرين (25) : اختير شخصان على التوالي من مجموعة بها (7) رجال و (5) نساء فإذا كان الشخص الثاني امرأة فما احتمال أن يكون الشخص الأول امرأة أيضاً .

تمرين (26) : ثلاثة فصول في كل منها (25) طفلاً ، أطفال الفصل الأول كلهم أولاد والثاني كلهم بنات والثالث مناصفة . اختير فصل عشوائي ومنه اختير طفل ما هو احتمال :

1 – أن يكون ولد

2 – إذا علمنا إنه ولد ، فما احتمال أن يكون في الفصل الأول .

تمرين (27) : صندوقان في كل منهما (3) كرات بيضاء و (2) كرة حمراء نقلت كرة من الصندوق الأول إلى الصندوق الثاني وسحبت كرة من الصندوق الثاني ، فما احتمال أن تكون الكرة المسحوبة بيضاء .

تمرين (28) : نعرف الحوادث التالية على فضاء العينة (S) رمي حجر النرد لمرة واحدة ، الحادثة (A) : حصولنا على عدد زوجي ، والحادثة (B) : حصولنا على عدد يقبل القسمة على (3) بدون باقي . والمطلوب : أوجد الاحتمالات الآتية :

$1 - P(A \cup B)$

$$\mathbf{2 - P\left(\overline{A} \cap B\right)}$$

تمرين (29) : بيَّن وجه الخطأ في كل من العبارات التالية :

1 - الاحتمال التجريبي = $\frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}}$

2 - الاحتمال $0 \le P(A) \le 1$

3 - الاحتمال لا يكون سالباً .

4 - الاحتمال النظري مبني على المنطق الرياضي وليس نتيجة التجربة

5 - احتمال النجاح + احتمال الإخفاق $\neq 1$

6 - الاحتمال يستخدم للتعبير عن حدث مؤكد الحدوث

7 - الإحصاء الرياضي مبني على نظرية الاحتمالات ويهتم بالمستقبل غير المؤكد

8 - الحادثة كل عملية تؤول إلى ملاحظة أو قياس ظاهرة ما

9 - في الحوادث المتنافية بالتبادل :

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

10 - عندما يكون عدد مرات التجربة كبيراً فإن الاحتمال التجريبي يقترب من الاحتمال النظري

11 – التكرار النسبي = $\frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الممكنة}}$

12- احتمال وقوع الحادث الأكيد يساوي الصفر .

13 - ظهور رقم (7) عند رمي حجر النرد هو حادث أكبر .

14 - في الحوادث المستقلة $P (A \cap B) = P (A) P (B)$

15 - المتغير العشوائي هو نتائج التجربة الإحصائية غير المعروفة مسبقاً

16 - المتغير العشوائي المنفصل يأخذ قيماً صحيحة فقط

17 - من شروط تابع التوزيع الاحتمالي $\sum_{x=1}^{n} f (x_i) \neq 1$

18 - الأمل (التوقع) الرياضي هو قيمة المتغير مضروباً باحتمال الحصول عليها .

19 - إن فضاء العينة لرمي حجر النرد لمرة واحدة

$S : \{ 1 , 2 , 3 , 4 , 5 , 6 \}$

20 - إن فضاء العينة لرمي قطعة نقود مرتين متتاليتين هو :

$S : \{ H T , T H , T T \}$

21 - إن فضاء العينة لرمي قطعة نقود سوية ثلاث مرات متتالية هو على النحو التالي:

$S : \{ (HHH) , (HHT) , (HTH) , (HTT) , (TTH) , (THH) , (THT) \}$

22 – من عيوب التعريف الإحصائي للاحتمال (التكرار النسبي) أن يكون فضاء العينة منتهياً ، ولكل عنصر من عناصر فضاء العينة احتمال الحدوث نفسه .

23 – إذا عرفنا على فضاء العينة لرمي حجر النرد مرة واحدة ، الحادث A على النحو التالي : $A = \{7\}$ فإن A حادثاً أكيداً .

24 – إذا عرفنا على فضاء العينة لرمي حجر نرد لمرة واحدة الحادث A على النحو التالي :

$A = \{ 1 , 5 , 7 \}$ فإن $\overline{A} = \{2,4\}$.

25 - احتمال أن ينجح خالد في امتحان الفيزياء هو (0.95 -) .

26 - احتمال أن ينجح علي في الإحصاء هو (0,9) واحتمال أن لا ينجح (0,15) .

27 - احتمال أن يفوز الفريق الوطني لكرة القدم في مباراته القادمة مع منتخب الإمارات هو (0,75) واحتمال أن يتعادل (0,09) واحتمال أن يتعادل أو يفوز(0,95) .

28 - احتمال أن ينجح عارف في الإحصاء هو (0,9) واحتمال أن ينجح في مقرري الإحصاء والرياضيات (0,95) .

29 - احتمال أن تستقبل عيادة طبيب أقل من (5) مرضى في فترة ما قبل الظهر هو (0,62) واحتمال أن تستقبل (5) مرضى أو أكثر هو (0,25) .

30 - $$P\left(\overline{A}\cap\overline{B}\right)=1-P\left(A\cup B\right)$$

31 - $$P\left(A/\overline{B}\right)=\frac{P(A)-P\left(A\cap B\right)}{1-\left(\overline{B}\right)}$$

32 - $$P\left(B/A\right)=\frac{P\left(A\cap B\right)}{P(A)}$$

تمرين (30) : أجب باختصار عن الأسئلة التالية :

ما الفرق بين الحادثة والتجربة ؟

2- ما الفرق بين الاحتمال النظري والاحتمال التجريبي .

ما الفرق بين المتغيرات العشوائية المنفصلة والمتغيرات العشوائية المتصلة ؟

4- ميِّز بين المفاهيم التالية :

الحادث الأكبر ، الحادث المستحيل ، الحادث الممكن .

5- متى يكون التابع ، تابع كثافة احتمالي ؟

ملاحق بالجداول الإحصائية

جدول رقم (I): مقطع من جدول الأرقام العشوائية

94015	46874	32444	48277	59820	96163	64654	25843	41145	42820
74108	88222	88570	74015	25704	91035	01755	14750	48968	38603
62880	87873	95160	59221	22304	90314	72877	17334	39283	04149
11748	12102	80580	41867	17710	59621	06554	07850	73950	79552
17944	05600	60478	03343	25852	58905	57216	39618	49856	99326
66067	42792	95043	52680	46780	56487	09971	59481	37006	22186
54244	91030	45547	70818	59849	96169	61459	21647	87417	17198
30945	57589	31732	57260	47670	07654	46376	25366	94746	49580
69170	37403	86995	90307	94304	71803	26825	05511	12459	91314
08345	88975	35841	85771	08105	59987	87112	21476	14713	71181
27767	43584	85301	88977	29490	69714	73035	41207	74699	09310
13025	14338	54066	15243	47724	66733	47431	43905	31048	56699
80217	36292	98525	24335	24432	24896	43277	58874	11466	16082
10875	62004	90391	61105	57411	06368	53856	30743	08670	84741
54127	57326	26629	19087	24472	88779	30540	27886	61732	75454
60311	42824	37301	42678	45990	43242	17374	52003	70707	70214
49739	71484	92003	98086	76668	73209	59202	11973	02902	33250
78626	51594	16453	94614	39014	97066	83012	09832	25571	77628
66692	13986	99837	00582	81232	44987	09504	96412	90193	79568
44071	28091	07362	97703	76447	42537	98524	97831	65704	09514
41468	85149	49554	17994	14924	39650	95294	00556	70481	06905
94559	37559	49678	53119	70312	05682	66986	34099	74474	20740
41615	70360	64114	58660	90850	64618	80620	51790	11436	38072
50273	93113	41794	86861	24781	89683	55411	85667	77535	99892
41396	80504	90670	08289	40902	05069	95083	06783	28102	57816
25807	24260	71529	78920	72682	07385	90726	57166	98884	08583
06170	97965	88302	98041	21443	41808	68984	83620	89747	99882
60808	54444	74412	81105	01176	28838	36421	16489	18059	51061
80940	44893	10408	36222	80562	71944	92638	40333	67054	16067
19516	90120	46759	71643	13177	55292	21036	82808	77501	97427
49386	54480	23604	23554	21785	41101	91178	10174	29420	90438
06312	88940	15995	69321	47458	64809	98189	81851	29651	84215
60942	00307	11897	92674	40405	68032	96717	54244	10701	41393
92329	98932	78284	46347	71209	92061	39448	93136	25722	08564
77936	63574	31384	51924	85561	29671	58137	17820	22751	36518
38101	77756	11657	13897	95889	57067	47648	13885	70669	93406
39641	69457	91339	22502	92613	89719	11947	56203	19324	20504
84054	40455	99396	63680	67667	60631	69181	96845	38525	11600
47468	03577	57649	63266	24700	71594	14004	23153	69249	05747
43321	31370	28977	23896	76479	68562	62342	07589	08899	05985
64281	61826	18555	64937	13173	33365	78851	16499	87064	13075
66847	70495	32350	02985	86716	38746	26313	77463	55387	72681
72461	33230	21529	53424	92581	02262	78438	66276	18396	73538
21032	91050	13058	16218	12470	56500	15292	76139	59526	52113
95362	67011	06651	16136	01016	00857	55018	56374	35824	71708
49712	97380	10404	55452	34030	60726	75211	10271	36633	68424
58275	61764	97586	54716	50259	46345	87195	46092	26787	60939
89514	11788	68224	23417	73959	76145	30342	40277	11049	72049
15472	50669	48139	36732	46874	37088	73465	09819	58869	35220
12120	86124	51247	44302	60883	52109	21437	36786	49226	77837

المصـــدر :

The Rand Corporation, A. Million Random Digits with 100,000 Normal Deviates (Glencoe, III, Free Press, 1955), P. 4, Reprinted with Permission of the Rand Corporation.

جدول رقم : (II) قيم تابع التوزيع الطبيعي

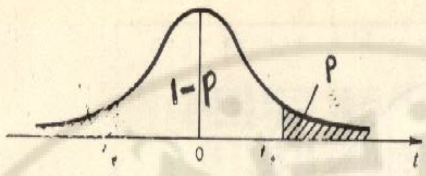
=

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

جدول القيم الكبيرة لـ z *Table pour les grandes valeurs de* z

z	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
Φ(z)	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

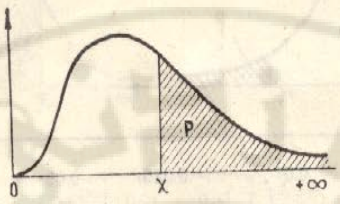
جدول رقم : (III) القيم الحدية (الحرجة) للمتحول t_p الخاضع لتوزيع (ستودنت)



K \ P	0,005	0,01	0,025	0,05	0,1	0,15
1	63,657	31,821	12 706	6,314	3,078	1 963
2	9,925	6 965	4,303	2 920	1 886	1 386
3	5,841	4,541	3,182	2 353	1 638	1 250
4	4,604	3,747	2 776	2,132	1,533	1,190
5	4 032	3 365	2 571	2 015	1 476	1 156
6	3,707	3,143	2,447	1,943	1 440	1,134
7	3 499	2 998	2,365	1 895	1,415	1 119
8	3,365	2,896	2,306	1 860	1 397	1,108
9	3,250	2,821	2 262	1 833	1 383	1 100
10	3 169	2 764	2.228	1 812	1 372	1 093
11	3 106	2,718	2 201	1 796	1,363	1 088
12	3,055	2,681	2,179	1 782	1,356	1 083
13	3,012	2 650	2,160	1 771	1 350	1 079
14	2 977	2 624	2 145	1 761	1 345	1 076
15	2 947	2 602	2 131	1 753	1 341	1 074
16	2,921	2,583	2 120	1 746	1 337	1 071
17	2 898	2 567	2 110	1 740	1 333	1 069
18	2,878	2,552	2 101	1 734	1 330	1 067
19	2 861	2,539	2 093	1 729	1 328	1 066
20	2 845	2,528	2 086	1 725	1 325	1 064
21	2 831	2,518	2 080	1 721	1 323	1 063
22	2 819	2 508	2,074	1 717	1 321	1 061
23	2 807	2,500	2 069	1 714	1 319	1 060
24	2 797	2,492	2 064	1 711	1 318	1 059
25	2 787	2,485	2 060	1 708	1 316	1 058
26	2 779	2,479	2 056	1 706	1 315	1 058
27	2 771	2,473	2 052	1 703	1 314	1 057
28	2 763	2,467	2 048	1 701	1 313	1 056
29	2 756	2,462	2 045	1 699	1 311	1 055
30	2 750	2,457	2,042	1 697	1 310	1 055
∞	2 576	2,326	1 960	1,645	1,282	1,036

SOURCE Reproduit du Tableau IV dans Sir Ronald A Fisher *Statistical Methods for Research Workers* 13e edition Oliver and Boyd Ltd Edimbourg 1963 avec l'autorisation des editeurs et de l'executeur litterraire de Sir Ronald Fisher

جدول رقم (IV): القيم الحدية للمتحول χ_p الخاضع $\chi^2_{\alpha_{(x)}}$



ν	$P = 0{,}90$	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,662	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

جدول رقم (V) :القيم الحدية F_p الخاضع لـ $F_{n_1,n2}(x)$

$f(F; n_1, n_2)$

$F_{0,05}$ $F_{0,01}$ F

n_2 | n_1 degrés de liberté (pour le plus grand carré moyen)

n_2		1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	n_2
1		161 4 052	200 4 999	216 5 403	225 5 625	230 5 764	234 5 859	237 5 928	239 5 981	241 6 022	242 6 056	243 6 082	244 6 106	245 6 142	246 6 169	248 6 208	249 6 234	250 6 258	251 6 286	252 6 302	253 6 323	253 6 334	254 6 352	254 6 361	254 6 366	1
2		18,51 98,49	19,00 99,00	19,16 99,17	19,25 99,25	19,30 99,30	19,33 99,33	19,36 99,34	19,37 99,36	19,38 99,38	19,39 99,40	19,40 99,41	19,41 99,42	19,42 99,43	19,43 99,44	19,44 99,45	19,45 99,46	19,46 99,47	19,47 99,48	19,47 99,48	19,48 99,49	19,49 99,49	19,49 99,49	19,50 99,50	19,50 99,50	2
3		10,13 34,12	9,55 30,82	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,88 27,67	8,84 27,49	8,81 27,34	8,78 27,23	8,76 27,13	8,74 27,05	8,71 26,92	8,69 26,83	8,66 26,69	8,64 26,60	8,62 26,50	8,60 26,41	8,58 26,35	8,57 26,27	8,56 26,23	8,54 26,18	8,54 26,14	8,53 26,12	3
4		7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,04 14,80	6,00 14,66	5,96 14,54	5,93 14,45	5,91 14,37	5,87 14,24	5,84 14,15	5,80 14,02	5,77 13,93	5,74 13,83	5,71 13,74	5,70 13,69	5,68 13,61	5,66 13,57	5,65 13,52	5,64 13,48	5,63 13,46	4
5		6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,45	4,82 10,27	4,78 10,15	4,74 10,05	4,70 9,96	4,68 9,89	4,64 9,77	4,60 9,68	4,56 9,55	4,53 9,47	4,50 9,38	4,46 9,29	4,44 9,24	4,42 9,17	4,40 9,13	4,38 9,07	4,37 9,04	4,36 9,02	5
6		5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	4,03 7,79	4,00 7,72	3,96 7,60	3,92 7,52	3,87 7,39	3,84 7,31	3,81 7,23	3,77 7,14	3,75 7,09	3,72 7,02	3,71 6,99	3,69 6,94	3,68 6,90	3,67 6,88	6
7		5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 7,00	3,73 6,84	3,68 6,71	3,63 6,62	3,60 6,54	3,57 6,47	3,52 6,35	3,49 6,27	3,44 6,15	3,41 6,07	3,38 5,98	3,34 5,90	3,32 5,85	3,29 5,78	3,28 5,75	3,25 5,70	3,24 5,67	3,23 5,65	7
8		5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,19	3,44 6,03	3,39 5,91	3,34 5,82	3,31 5,74	3,28 5,67	3,23 5,56	3,20 5,48	3,15 5,36	3,12 5,28	3,08 5,20	3,05 5,11	3,03 5,06	3,00 5,00	2,98 4,96	2,96 4,91	2,94 4,88	2,93 4,86	8
9		5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,80	3,29 5,62	3,23 5,47	3,18 5,35	3,13 5,26	3,10 5,18	3,07 5,11	3,02 5,00	2,98 4,92	2,93 4,80	2,90 4,73	2,86 4,64	2,82 4,56	2,80 4,51	2,77 4,45	2,76 4,41	2,73 4,36	2,72 4,33	2,71 4,31	9
10		4,96 10,04	4,10 7,56	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,21	3,07 5,06	3,02 4,95	2,97 4,85	2,94 4,78	2,91 4,71	2,86 4,60	2,82 4,52	2,77 4,41	2,74 4,33	2,70 4,25	2,67 4,17	2,64 4,12	2,61 4,05	2,59 4,01	2,56 3,96	2,55 3,93	2,54 3,91	10
11		4,84 9,65	3,98 7,20	3,59 6,22	3,36 5,67	3,20 5,32	3,09 5,07	3,01 4,88	2,95 4,74	2,90 4,63	2,86 4,54	2,82 4,46	2,79 4,40	2,74 4,29	2,70 4,21	2,65 4,10	2,61 4,02	2,57 3,94	2,53 3,86	2,50 3,80	2,47 3,74	2,45 3,70	2,42 3,66	2,41 3,62	2,40 3,60	11
12		4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,92 4,65	2,85 4,50	2,80 4,39	2,76 4,30	2,72 4,22	2,69 4,16	2,64 4,05	2,60 3,98	2,54 3,86	2,50 3,78	2,46 3,70	2,42 3,61	2,40 3,56	2,36 3,49	2,35 3,46	2,32 3,41	2,31 3,38	2,30 3,36	12
13		4,67 9,07	3,80 6,70	3,41 5,74	3,18 5,20	3,02 4,86	2,92 4,62	2,84 4,44	2,77 4,30	2,72 4,19	2,67 4,10	2,63 4,02	2,60 3,96	2,55 3,85	2,51 3,78	2,46 3,67	2,42 3,59	2,38 3,51	2,34 3,42	2,32 3,37	2,28 3,30	2,26 3,27	2,24 3,21	2,22 3,18	2,21 3,16	13

تتمة الجدول رقم(V)

n_2	n_1 degrés de liberté (pour le plus grand carré moyen)																							n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4,60 **8,86**	3,74 **6,51**	3,34 **5,56**	3,11 **5,03**	2,96 **4,69**	2,85 **4,46**	2,77 **4,28**	2,70 **4,14**	2,65 **4,03**	2,60 **3,94**	2,56 **3,86**	2,53 **3,80**	2,48 **3,70**	2,44 **3,62**	2,39 **3,51**	2,35 **3,43**	2,31 **3,34**	2,27 **3,26**	2,24 **3,21**	2,21 **3,14**	2,19 **3,11**	2,16 **3,06**	2,14 **3,02**	2,13 **3,00**	14
15	4,54 **8,68**	3,68 **6,36**	3,29 **5,42**	3,06 **4,89**	2,90 **4,56**	2,79 **4,32**	2,70 **4,14**	2,64 **4,00**	2,59 **3,89**	2,55 **3,80**	2,51 **3,73**	2,48 **3,67**	2,43 **3,56**	2,39 **3,48**	2,33 **3,36**	2,29 **3,29**	2,25 **3,20**	2,21 **3,12**	2,18 **3,07**	2,15 **3,00**	2,12 **2,97**	2,10 **2,92**	2,08 **2,89**	2,07 **2,87**	15
16	4,49 **8,53**	3,63 **6,23**	3,24 **5,29**	3,01 **4,77**	2,85 **4,44**	2,74 **4,20**	2,66 **4,03**	2,59 **3,89**	2,54 **3,78**	2,49 **3,69**	2,45 **3,61**	2,42 **3,55**	2,37 **3,45**	2,33 **3,37**	2,28 **3,25**	2,24 **3,18**	2,20 **3,10**	2,16 **3,01**	2,13 **2,96**	2,09 **2,89**	2,07 **2,86**	2,04 **2,80**	2,02 **2,77**	2,01 **2,75**	16
17	4,45 **8,40**	3,59 **6,11**	3,20 **5,18**	2,96 **4,67**	2,81 **4,34**	2,70 **4,10**	2,62 **3,93**	2,55 **3,79**	2,50 **3,68**	2,45 **3,59**	2,41 **3,52**	2,38 **3,45**	2,33 **3,35**	2,29 **3,27**	2,23 **3,16**	2,19 **3,08**	2,15 **3,00**	2,11 **2,92**	2,08 **2,86**	2,04 **2,79**	2,02 **2,76**	1,99 **2,70**	1,97 **2,67**	1,96 **2,65**	17
18	4,41 **8,28**	3,55 **6,01**	3,16 **5,09**	2,93 **4,58**	2,77 **4,25**	2,66 **4,01**	2,58 **3,85**	2,51 **3,71**	2,46 **3,60**	2,41 **3,51**	2,37 **3,44**	2,34 **3,37**	2,29 **3,27**	2,25 **3,19**	2,19 **3,07**	2,15 **3,00**	2,11 **2,91**	2,07 **2,83**	2,04 **2,78**	2,00 **2,71**	1,98 **2,68**	1,95 **2,62**	1,93 **2,59**	1,92 **2,57**	18
19	4,38 **8,18**	3,52 **5,93**	3,13 **5,01**	2,90 **4,50**	2,74 **4,17**	2,63 **3,94**	2,55 **3,77**	2,48 **3,63**	2,43 **3,52**	2,38 **3,43**	2,34 **3,36**	2,31 **3,30**	2,26 **3,19**	2,21 **3,12**	2,15 **3,00**	2,11 **2,92**	2,07 **2,84**	2,02 **2,76**	2,00 **2,70**	1,96 **2,63**	1,94 **2,60**	1,91 **2,54**	1,90 **2,51**	1,88 **2,49**	19
20	4,35 **8,10**	3,49 **5,85**	3,10 **4,94**	2,87 **4,43**	2,71 **4,10**	2,60 **3,87**	2,52 **3,71**	2,45 **3,56**	2,40 **3,45**	2,35 **3,37**	2,31 **3,30**	2,28 **3,23**	2,23 **3,13**	2,18 **3,05**	2,12 **2,94**	2,08 **2,86**	2,04 **2,77**	1,99 **2,69**	1,96 **2,63**	1,92 **2,56**	1,90 **2,53**	1,87 **2,47**	1,85 **2,44**	1,84 **2,42**	20
21	4,32 **8,02**	3,47 **5,78**	3,07 **4,87**	2,84 **4,37**	2,68 **4,04**	2,57 **3,81**	2,49 **3,65**	2,42 **3,51**	2,37 **3,40**	2,32 **3,31**	2,28 **3,24**	2,25 **3,17**	2,20 **3,07**	2,15 **2,99**	2,09 **2,88**	2,05 **2,80**	2,00 **2,72**	1,96 **2,63**	1,93 **2,58**	1,89 **2,51**	1,87 **2,47**	1,84 **2,42**	1,82 **2,38**	1,81 **2,36**	21
22	4,30 **7,94**	3,44 **5,72**	3,05 **4,82**	2,82 **4,31**	2,66 **3,99**	2,55 **3,76**	2,47 **3,59**	2,40 **3,45**	2,35 **3,35**	2,30 **3,26**	2,26 **3,18**	2,23 **3,12**	2,18 **3,02**	2,13 **2,94**	2,07 **2,83**	2,03 **2,75**	1,98 **2,67**	1,93 **2,58**	1,91 **2,53**	1,87 **2,46**	1,84 **2,42**	1,81 **2,37**	1,80 **2,33**	1,78 **2,31**	22
23	4,28 **7,88**	3,42 **5,66**	3,03 **4,76**	2,80 **4,26**	2,64 **3,94**	2,53 **3,71**	2,45 **3,54**	2,38 **3,41**	2,32 **3,30**	2,28 **3,21**	2,24 **3,14**	2,20 **3,07**	2,14 **2,97**	2,10 **2,89**	2,04 **2,78**	2,00 **2,70**	1,96 **2,62**	1,91 **2,53**	1,88 **2,48**	1,84 **2,41**	1,82 **2,37**	1,79 **2,32**	1,77 **2,28**	1,76 **2,26**	23
24	4,26 **7,82**	3,40 **5,61**	3,01 **4,72**	2,78 **4,22**	2,62 **3,90**	2,51 **3,67**	2,43 **3,50**	2,36 **3,36**	2,30 **3,25**	2,26 **3,17**	2,22 **3,09**	2,18 **3,03**	2,13 **2,93**	2,09 **2,85**	2,02 **2,74**	1,98 **2,66**	1,94 **2,58**	1,89 **2,49**	1,86 **2,44**	1,82 **2,36**	1,80 **2,33**	1,76 **2,27**	1,74 **2,23**	1,73 **2,21**	24
25	4,24 **7,77**	3,38 **5,57**	2,99 **4,68**	2,76 **4,18**	2,60 **3,86**	2,49 **3,63**	2,41 **3,46**	2,34 **3,32**	2,28 **3,21**	2,24 **3,13**	2,20 **3,05**	2,16 **2,99**	2,11 **2,89**	2,06 **2,81**	2,00 **2,70**	1,96 **2,62**	1,92 **2,54**	1,87 **2,45**	1,84 **2,40**	1,80 **2,32**	1,77 **2,29**	1,74 **2,23**	1,72 **2,19**	1,71 **2,17**	25
26	4,22 **7,72**	3,37 **5,53**	2,98 **4,64**	2,74 **4,14**	2,59 **3,82**	2,47 **3,59**	2,39 **3,42**	2,32 **3,29**	2,27 **3,17**	2,22 **3,09**	2,18 **3,02**	2,15 **2,96**	2,10 **2,86**	2,05 **2,77**	1,99 **2,66**	1,95 **2,58**	1,90 **2,50**	1,85 **2,41**	1,82 **2,36**	1,78 **2,28**	1,76 **2,25**	1,72 **2,19**	1,70 **2,15**	1,69 **2,13**	26

تتمة الجدول رقم(٧)

n_2	n_1 degrés de liberté (pour le plus grand carré moyen)																							n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4,21 **7,68**	3,35 **5,49**	2,96 **4,60**	2,73 **4,11**	2,57 **3,79**	2,46 **3,56**	2,37 **3,39**	2,30 **3,26**	2,25 **3,14**	2,20 **3,06**	2,16 **2,98**	2,13 **2,93**	2,08 **2,83**	2,03 **2,74**	1,97 **2,63**	1,93 **2,55**	1,88 **2,47**	1,84 **2,38**	1,80 **2,33**	1,76 **2,25**	1,74 **2,21**	1,71 **2,16**	1,68 **2,12**	1,67 **2,10**	27
28	4,20 **7,64**	3,34 **5,45**	2,95 **4,57**	2,71 **4,07**	2,56 **3,76**	2,44 **3,53**	2,36 **3,36**	2,29 **3,23**	2,24 **3,11**	2,19 **3,03**	2,15 **2,95**	2,12 **2,90**	2,06 **2,80**	2,02 **2,71**	1,96 **2,60**	1,91 **2,52**	1,87 **2,44**	1,81 **2,35**	1,78 **2,30**	1,75 **2,22**	1,72 **2,18**	1,69 **2,13**	1,67 **2,09**	1,65 **2,06**	28
29	4,18 **7,60**	3,33 **5,42**	2,93 **4,54**	2,70 **4,04**	2,54 **3,73**	2,43 **3,50**	2,35 **3,33**	2,28 **3,20**	2,22 **3,08**	2,18 **3,00**	2,14 **2,92**	2,10 **2,87**	2,05 **2,77**	2,00 **2,68**	1,94 **2,57**	1,90 **2,49**	1,85 **2,41**	1,80 **2,32**	1,77 **2,27**	1,73 **2,19**	1,71 **2,15**	1,68 **2,10**	1,65 **2,06**	1,64 **2,03**	29
30	4,17 **7,56**	3,32 **5,39**	2,92 **4,51**	2,69 **4,02**	2,53 **3,70**	2,42 **3,47**	2,34 **3,30**	2,27 **3,17**	2,21 **3,06**	2,16 **2,98**	2,12 **2,90**	2,09 **2,84**	2,04 **2,74**	1,99 **2,66**	1,93 **2,55**	1,89 **2,47**	1,84 **2,38**	1,79 **2,29**	1,76 **2,24**	1,72 **2,16**	1,69 **2,13**	1,66 **2,07**	1,64 **2,03**	1,62 **2,01**	30
32	4,15 **7,50**	3,30 **5,34**	2,90 **4,46**	2,67 **3,97**	2,51 **3,66**	2,40 **3,42**	2,32 **3,25**	2,25 **3,12**	2,19 **3,01**	2,14 **2,94**	2,10 **2,86**	2,07 **2,80**	2,02 **2,70**	1,97 **2,62**	1,91 **2,51**	1,86 **2,42**	1,82 **2,34**	1,76 **2,25**	1,74 **2,20**	1,69 **2,12**	1,67 **2,08**	1,64 **2,02**	1,61 **1,98**	1,59 **1,96**	32
34	4,13 **7,44**	3,28 **5,29**	2,88 **4,42**	2,65 **3,93**	2,49 **3,61**	2,38 **3,38**	2,30 **3,21**	2,23 **3,08**	2,17 **2,97**	2,12 **2,89**	2,08 **2,82**	2,05 **2,76**	2,00 **2,66**	1,95 **2,58**	1,89 **2,47**	1,84 **2,38**	1,80 **2,30**	1,74 **2,21**	1,71 **2,15**	1,67 **2,08**	1,64 **2,04**	1,61 **1,98**	1,59 **1,94**	1,57 **1,91**	34
36	4,11 **7,39**	3,26 **5,25**	2,86 **4,38**	2,63 **3,89**	2,48 **3,58**	2,36 **3,35**	2,28 **3,18**	2,21 **3,04**	2,15 **2,94**	2,10 **2,86**	2,06 **2,78**	2,03 **2,72**	1,98 **2,62**	1,93 **2,54**	1,87 **2,43**	1,82 **2,35**	1,78 **2,26**	1,72 **2,17**	1,69 **2,12**	1,65 **2,04**	1,62 **2,00**	1,59 **1,94**	1,56 **1,90**	1,55 **1,87**	36
38	4,10 **7,35**	3,25 **5,21**	2,85 **4,34**	2,62 **3,86**	2,46 **3,54**	2,35 **3,32**	2,26 **3,15**	2,19 **3,02**	2,14 **2,91**	2,09 **2,82**	2,05 **2,75**	2,02 **2,69**	1,96 **2,59**	1,92 **2,51**	1,85 **2,40**	1,80 **2,32**	1,76 **2,22**	1,71 **2,14**	1,67 **2,08**	1,63 **2,00**	1,60 **1,97**	1,57 **1,90**	1,54 **1,86**	1,53 **1,84**	38
40	4,08 **7,31**	3,23 **5,18**	2,84 **4,31**	2,61 **3,83**	2,45 **3,51**	2,34 **3,29**	2,25 **3,12**	2,18 **2,99**	2,12 **2,88**	2,07 **2,80**	2,04 **2,73**	2,00 **2,66**	1,95 **2,56**	1,90 **2,49**	1,84 **2,37**	1,79 **2,29**	1,74 **2,20**	1,69 **2,11**	1,66 **2,05**	1,61 **1,97**	1,59 **1,94**	1,55 **1,88**	1,53 **1,84**	1,51 **1,81**	40
42	4,07 **7,27**	3,22 **5,15**	2,83 **4,29**	2,59 **3,80**	2,44 **3,49**	2,32 **3,26**	2,24 **3,10**	2,17 **2,96**	2,11 **2,86**	2,06 **2,77**	2,02 **2,70**	1,99 **2,64**	1,94 **2,54**	1,89 **2,46**	1,82 **2,35**	1,78 **2,26**	1,73 **2,17**	1,68 **2,08**	1,64 **2,02**	1,60 **1,94**	1,57 **1,91**	1,54 **1,85**	1,51 **1,80**	1,49 **1,78**	42
44	4,06 **7,24**	3,21 **5,12**	2,82 **4,26**	2,58 **3,78**	2,43 **3,46**	2,31 **3,24**	2,23 **3,07**	2,16 **2,94**	2,10 **2,84**	2,05 **2,75**	2,01 **2,68**	1,98 **2,62**	1,92 **2,52**	1,88 **2,44**	1,81 **2,32**	1,76 **2,24**	1,72 **2,15**	1,66 **2,06**	1,63 **2,00**	1,58 **1,92**	1,56 **1,88**	1,52 **1,82**	1,50 **1,78**	1,48 **1,75**	44
46	4,05 **7,21**	3,20 **5,10**	2,81 **4,24**	2,57 **3,76**	2,42 **3,44**	2,30 **3,22**	2,22 **3,05**	2,14 **2,92**	2,09 **2,82**	2,04 **2,73**	2,00 **2,66**	1,97 **2,60**	1,91 **2,50**	1,87 **2,42**	1,80 **2,30**	1,75 **2,22**	1,71 **2,13**	1,65 **2,04**	1,62 **1,98**	1,57 **1,90**	1,54 **1,86**	1,51 **1,80**	1,48 **1,76**	1,46 **1,72**	46
48	4,04 **7,19**	3,19 **5,08**	2,80 **4,22**	2,56 **3,74**	2,41 **3,42**	2,30 **3,20**	2,21 **3,04**	2,14 **2,90**	2,08 **2,80**	2,03 **2,71**	1,99 **2,64**	1,96 **2,58**	1,90 **2,48**	1,86 **2,40**	1,79 **2,28**	1,74 **2,20**	1,70 **2,11**	1,64 **2,02**	1,61 **1,96**	1,56 **1,88**	1,53 **1,84**	1,50 **1,78**	1,47 **1,73**	1,45 **1,70**	48

Statistical Terminology

المصطلحات الإحصائية

إنكليزي - عربي

A	
Absolute error	- الخطأ المطلق
Absolute value	- القيمة المطلقة
Accuracy	- دقة
Adjustment data	- تعديل البيانات
Aggregate value	- القيمة التجميعية
Allowance value	- حدود السماح
Analysis of covariance	- تحليل تمام التباين
Analysis of variance	- تحليل التباين
Applied statistics	- الإحصاء التطبيقي
Arbitrary constant	- ثابت اختياري
Area sampling	- المعاينة المساحية
Arithmetic mean	- الوسط الحسابي
Array	- ترتيب
Association	- اقتران
Assumption	- فرض
Asymmetry	- عدم التماثل
Asymmetrical	- غير متماثل
Asymptotic	- تقريبي
Attribute	- صفة
Automatic	- تلقائي ، آلي
Average	- متوسط
Axis	- محور
Axis of abscissas	- المحور السيني أو الأفقي
Axis of ordinates	- المحور العمودي
B	
Base	- أساس أو قاعدة
Base period	- فترة الأساس
Basic statistics	- إحصاءات أساسية

Battery of test	- مجموعة الاختبارات
Bell-shaped curve	- منحنى جرسي
Between groups	- بين المجموعات
Bias	- تحيزّ
Binominal distribution	- توزيع ثنائي الحدين
Bivariate	- ذو متغيرين
Bivariate table	- جدول ذو متغيرين
Breakdown	- تقسيم
Business cycles	- دورات اقتصادية
Business indicators	- مؤشرات اقتصادية

C

Calculation	- حساب
Calculus of probability	- حساب الاحتمالات
Census	- تعداد
Census enumerator	- عداد
Census schedule	- استمارة التعداد
Causation	- سببية
Central tendency	- النزعة المركزية
Chain indexes	- الأرقام القياسية المتسلسلة
Chart	- خريطة
Chi-Squared test	- اختبار كاي مربع
Classification	- تصنيف
Cluster sampling	- المعاينة العنقودية
Code	- رمز
Coefficient	- معامل
Coefficient of correlation	- معامل الارتباط
Coefficient of rank correlation	- معامل الارتباط الجزئي
Column	- عمود
Compute	- يحسب
Confidence intervals	- فترات الثقة
Confidence limits	- حدود الثقة
Consistency	- اتساق

Constant	- ثابت
Contingency	- توافق
Contingency table	- جدول التوافق
Continuous	- متصل ، مستمر
Control	- مراقبة
Coordinate	- إحداثي
Coordinate axis	- محور الإحداثيات
Corrected rate	- معدل مصحح
Correction factor	- عامل التصحيح
Correlation	- ارتباط
Correlation table	- جدول ارتباطي
Correlation ratio	- نسبة الارتباط
Cost of living	- نفقة المعيشة
Covariable	- متغاير
Covariance	- تمام التباين
Critical value	- قيمة حرجة
Curves	- منحنيات
Cycle	- دورة
Cyclical variations	- تغيرات دورية
G	
Gaussian curve	- منحنى غوس
Geometric mean	- وسط هندسي
Graduation	- تدريج
Graph	- رسم بياني
Graphic analysis	- تحيلي بياني
Graphic method	- طريقة بيانية
Graphic presentation	- عرض بياني
Gross national product	- الإنتاج القومي الإجمالي
Grouping	- تجميع / تصنيف
Grouping error	- خطأ التجميع
Growth curve	- منحنى النمو
D	
Data	- بيانات

Deduction	- استنباط
Degrees of freedom	- درجات الحرية
Dependent variable	- متغير تابع
Derivative function	- دالة مشتقة
Design of experiments	- تصميم التجارب
Determinant	- محدّدة
Deviation	- انحراف
Diagram (figure)	- شكل
Differential	- تفاضلي
Discrete	- متقطع ، منفصل
Dispersion	- تشتت
Distribution	- توزيع
Distribution function	- دالة التوزيع
Double sampling	- معاينة مزدوجة
Downward bias	- تحيزّ نحو الأدنى
Downward trend	- اتجاه هابط

E

Econometrics	- اقتصاد قياسي
Effect	- تأثير
Efficiency of estimates	- فعالية التقدير
Elimination	- حذف
Enumerator	- عداد
Equation	- معادلة
Error	- خطأ
Error of estimate	- خطأ التقدير
Error of variance	- تباين الأخطاء
Estimate /Estimation	- تقدير
Estimation equation	- معادلة التقدير
Event	- واقعة / حادثة
Expected value	- القيمة المتوقعة
Experimental error	- الخطأ التجريبي
Experimental units	- وحدات تجريبية
Exponential curve	- منحنى آسي

Exponential equation	- معادلة آسية
Extrapolation	- استكمال خارجي
F	
"F" table	- جدول " ف"
Factor	- عامل
Factor analysis	- تحليل عوامل
Factorial analysis	- تحيلي عاملي
Family budget	- ميزانية الأسرة
Fluctuations	- تقلبات
Forecasting	- تنبؤ
Forms	- استمارات
Frame	- إطار
Frequency	- تكرار
Frequency curve	- منحنى التكرار
Frequency density	- كثافة التكرار
Frequency distribution	- توزيع التكرار
Frequency polygon	- مضلع التكرار
Frequency table	- جدول التكرار
Function	- دالة
Function relationship	- علاقة دالة
H	
- دالة توافقية Harmonic equation	
- وسط توافقي Harmonic mean	
- رب الأسرة Head of household	
- مدرج تكرار Histogram	
- متجانس Homogeneous	
- أسرة Household	
- تعداد المساكن Housing census	
- افتراض / فرضية Hypothesis	
I	
Ideal number	- الرقم الأمثل
Identity	- متطابقة
Independent variable	- متغير مستقل

Index	- دليل أو مقياس
Index number	- الرقم القياسي
Infinite	- لا نهائي
Integration	- تكامل
Interclass correlation	- ارتباط بين الفترات
Interpolation	- استكمال داخلي
Interval	- فترة ، مجال
Interview	- مقابلة / استجواب
Inverse correlation	- ارتباط عكسي
Inverse function	- دالة عكسية
J	
Joint distribution	- توزيع مشترك
Join regression	- انحدار مشترك
J- curve	- منحنى رائي
K	
Kurtosis	- تفرطح
L	
Labels	- علامات مميزّة
Labor force	- قوة العمل
Large sample	- عينات كبيرة
Law of large numbers	- قانون الأعداد الكبيرة
Least squares	- المربعات الصغرى
Least squares method	- طريقة المربعات الصغرى
Level of living	- مستوى المعيشة
Level of prices	- مستوى الأسعار
Likelihood	- إمكان
Limit	- نهاية أو حد
Linear	- خطي / مستقيم
Linear correlation	- ارتباط خطي
Linear equation	- معادلة خطية
Linear interpolation	- استكمال خطي
Linear regression	- انحدار خطي

Linear trend	- اتجاه مستقيم
List	- قائمة
Living conditions	- ظروف معيشية
Logarithmic	- لوغاريتم
Logistic curve	- منحنى لوجستي
Long cycles	- دورات طويلة
Lorenz curve	- منحنى لورنز
M	
Master sample	- عينة رئيسة
Matrix	- مصفوفة
Maximum variation	- أقصى اختلاف
Mean	- وسط
Mean absolute error	- وسط الخطأ المطلق
Median	- الوسيط
Model	- نموذج
Moment	- عزم
Moving average	- متوسط متحرك
Multi-phase sampling	- معاينة متعددة الأوجه
Multiple correlation	- ارتباط متعدد
Multiple factor analysis	- تحليل عاملي متعدد
Multi-Stage sampling	- معاينة متعددة المراحل
Multivariate analysis	- تحليل المتغيرات المتعددة
N	
National accounts	- حسابات قومية
National income	- الدخل القومي
Negative correlation	- ارتباط سالب
Non-Linear correlation	- ارتباط غير خطي
Non-Statistics	- إحصاء لا معلمي
Normal curve	- منحنى طبيعي
Normal distribution	- توزيع طبيعي
Normal equation	- معادلة طبيعية
Null hypothesis	- فرضية العدم

O
- مشاهدة Observation - متعامد Orthogonal
P
- قطع مكافئ Parabola - معلمة ، ثابت Parameter - ارتباط جزئي Partial correlation - انحدار جزئي Partial regression - نسبة مئوية Percentage - دورية Periodicity - تقلبات دورية Periodic movements - مجتمع (إحصائي) Population(statistical) - ارتباط موجب Positive correlation - تنبؤ Prediction - تغيرات الأسعار Price change - الرقم القياسي للأسعار Price index - احتمال Probability - منحنى الاحتمال Probability curve - كثافة الاحتمال Probability density - التوزيع التناسبي للعينة Proportional allocation of the sample - عينة عمدية Purposive sample - اختيار عمدي Purposive selection
Q
- تحليل وصفي Qualitative analysis - مراقبة الجودة Quality control - تحليل كمي Quantitative analysis - الرقم القياسي للكميات Quantity index - استبيان / استمارة Questionnaire - عينة الحصص Quota sampling
R
- توزيع عشوائي Random distribution - خطأ عشوائي Random error - أرقام عشوائية Random numbers

- عينة عشوائية Random sample
- اختيار عشوائي Random selection
- متغير عشوائي Random variable
- مدى Range
- ارتباط الرتب Rank correlation
- معامل ارتباط الرتب Rank correlation coefficient
- معدل Rate
- نسبة Ratio
- تقدير نسبي Ratio estimate
- اختبار النسب Ratio test
- انحدار Regression
- معامل الانحدار Regression coefficient
- معادلة الانحدار Regression equation
- تقدير الانحدار Regression estimate
- خط الانحدار Regression line
- عينة ممثلة Representative sample
- جذر متوسط مربع الانحراف Root-mean square deviation
- خطأ التقريب Rounding error
- صف Row

S

- عينة Sample
- مسح بالعينة Sample census
- تصميم العينة Sample design
- معاينة Sampling
- خطأ المعاينة Sampling error
- طريقة المعاينة Sampling method
- وحدة المعاينة Sampling unit
- مقياس الرسم Scale
- انتشار Scatter
- شكل الانتشار Scatter diagram
- الرقم القياسي الموسمي Seasonal index
- تغيرات موسمية Seasonal variation

Secondary trend	- اتجاه ثانوي
Second degree curve	- منحنى من الدرجة الثانية
Secular trend	- اتجاه عام
Sensitivity	- حساسية
Sequential analysis	- تحليل تتابعي
Significance	- معنوية
Simple correlation	- ارتباط بسيط
Size of sample	- حجم العينة
Small samples	- عينات صغيرة
Source of error	- مصدر الخطأ
Spurious correlation	- ارتباط وهمي
Square	- مربع
Standard normal distribution	- توزيع طبيعي معياري
Standard deviation	- انحراف معياري
Standard error	- خطأ معياري
Standard error of estimate	- خطأ معياري للتقدير
Standard error of the difference between means	- خطأ معياري للفرق بين الوسطين
Statistical induction	- استنتاج إحصائي
Statistical inference	- استدلال إحصائي
Statistical method	- الطريقة الإحصائية
Statistical probability	- احتمال إحصائي
Statistical processing	- التجهيز الإحصائي
Statistics	- إحصاء
Stratified sample	- عينة طبقية
Stratum	- طبقة
Survey	- مسح (استقصاء)
Symmetrical	- متماثل
System	- نظام
Systematic	- نظامي
Systematic sampling	- معاينة منتظمة
T	
"t" –table	- جدول " ت" ستودنت

Table	- جدول
Tabulation	- جدول أو تبويب
Test	- اختبار
Test of goodness of fit	- اختبار جودة التوفيق
Test of homogeneity	- اختبار التجانس
Test of significance	- اختبار المعنوية
Third degree curve	- منحنى من الدرجة الثالثة
Time series	- سلسلة زمنية
Time series curve	- منحنيات السلاسل الزمنية
Trend analysis	- تحليل الاتجاه
Trend line	- خط الاتجاه
Trial census	- تعداد تجريبي
True mean	- وسط حقيقي
Type	- نوع ، نمط
Typical value	- قيمة نمطية
U	
U-curve	- منحنى ذو شعبتين
Unbiased	- غير متحيزّ
Unit of measurement	- وحدة القياس
Universe (statistical)	- مجتمع (إحصائي)
Upward trend	- اتجاه صاعد
V	
Variable	- متغير
Variance	- تباين
Variance analysis	- تحليل التباين
Variance ratio	- نسبة التباين
Variation	- اختلاف
Variation coefficient	- معامل اختلاف
Vital statistics	- إحصاء حيوي
W	
Weights	- تثقيلات ، أوزان
Weight average	- متوسط مثقل أو مرجح

X	
- محور السينات	**X-axis**
Y	
- محور العينات	**Y-axis**
Z	
- اختبار (ز)	**Z-test**

قائمة المراجع بالعربية

-أبو صالح ، محمد صبحي ؛ عوض ، عدنان محمد ، 1990 – مقدمة في الإحصاء – مركز الكتب الأردني .

-أبو عمة ، عبد الرحمن ، راضي ، الحسني عبد البر ، الهندي ، محمود إبراهيم ،1995 ، الإحصاء والاحتمالات ، الرياض ، جامعة الملك سعود .

-البلداوي ، عبد الحميد ، الإحصاء للعلوم الإدارية والتطبيقية ، 1997 ، عمان ، جامعة الإسراء .

-البشير ، زين العابدين عبد الرحيم ، 1997 ، الاستدلال الإحصائي ، الرياض ، جامعة الملك سعود .

-الجاعوني ، فريد ؛ لإسماعيل ، حبيب ؛ غانم ، عدنان ، 1999 ، مبادئ الإحصاء ، جامعة صنعاء .

-الجراد ، خلف مطر ، مجموعى محاضرات في الرياضيات الاقتصادية والإدارية ، 1996 ، جامعة دمشق .

-الحلاق ، عمر ؛ عبيدو ، أميرة ، 1991 ، المدخل إلى الإحصاء الحيوي ، منشورات جامعة حلب .

-العتوم ، شفيق ، 1982 – مقدمة في الأساليب الإحصائية ، الجامعة الأردنية – كلية الاقتصاد والعلوم الإدارية .

-العلي ، إبراهيم محمد، 1979 ، مدخل في نظرية العينات – منشورات جامعة حلب .

-العلي ، إبراهيم محمد ، 1980 ،الإحصاء الرياضي ، جامعة حلب .

-العلي ، إبراهيم محمد ، 1985 ، نظرية الاحتمالات ، جامعة حلب .

-الأفندي ، عبد القادر ، 1975 ، نظرية الإحصاء ، جامعة حلب .

-أوسيليفان ، جورج ، بانكروفت ، جوردن ، 1983 ، الرياضيات والإحصاء لدراسات المحاسبة والأعمال ، دار ماكجرو هيل للنشر ، ترجمة مقدسي ، حمال سامي ، ومراجعة السيد محمد الغزي .

-حميدان ، عدنان ؛ مخول ، مطانيوس ؛ الجاعوني ، فريد ؛ ناصر آغا ، عمار ، 2005 ، الإحصاء التطبيقي ، منشورات جامعة دمشق .

-حميدان ، عدنان ؛ الجاعوني ، فريد ؛ ناصر آغا ، عمار؛ العواد ، منذر ، 2003 ، مبادئ الإحصاء ، منشورات جامعة دمشق .

-حميدان ، عدنان ؛ النعيمي ، قاسم ؛ ناصر آغا ، عمار ، 2004 ، مبادئ الإحصاء ، منشورات جامعة دمشق، مركز التعليم المفتوح .

-حيدر ، ناظم ، 1982 ، مبادىء الإحصاء – منشورات جامعة دمشق .

-عبد العزيز ، عمر عبد الجواد ؛ بلعربي ، عبد الحفيظ ، 1999 ، مقدمة في الطرق الإحصائية مع تطبيقات تجارية – دار زهرات للنشر والتوزيع – الأردن .
-عبد المجيد ، سمير ، 1994 ، علم الإحصاء ، الرياض ، دار مرامر للطباعة والنشر .
-عربش ، شفيق ؛ مخول ، مطانيوس وآخرون ، محاضرات في مبادئ الإحصاء خلال السنوات الماضية في كلية الاقتصاد بجامعة دمشق .
-علي ، إبراهيم ، 1998 ، مبادئ عامة في العينات وبعض أنواع العينات – ورشة العمل حول مسح الهجرة الداخلية في الجمهورية العربية السورية .
-عودة ، أحمد ، 1991 ، مقدمة في النظرية الإحصائية ، جامعة الملك سعود ، الرياض .
-شجبيل ، موراى ر . ، 1972 ، سلسلة ملخصات سشوم – نظريات ومسائل في الإحصاء – دار ماكجرو هيل للنشر .
-رمضان ، محمد ، 1984 ، الإحصاء الاجتماعي والعينات ، جامعة دمشق .
-طيوب ، محمود ؛ غانم ، عدنان ، 2002 ، مبادئ الإحصاء ، جامعة صنعاء .
-صبح ، محمد ، مبادئ الإحصاء والاحتمال ، جامعة دمشق .
-كابوس ، أمل ، 1983 ، الإحصاء الحيوي – منشورات جامعة حلب .
-كاربنتر ، رأي ل . ؛ فاسو ، إلين ستوري ؛ ترجمة : حسب الله سيد ؛ غندور ، محمد جلال سيد محمد ، 1998 ، الإحصاء للمكتبيين ، دار المريخ للنشر ، الرياض .
-كنجو ، أنيس ، 1995 ، تقنية المعاينة الإحصائية ، جامعة الملك سعود ، الرياض .

المراجع بالأجنبية

- Günter , C.; Ebner, H., 1983, Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen, Volk und Wissen , Volkseigener Verlag , Berlin.
-Klitzsch,W.; Hellmund, U.; Schumann, K., 1992, Grundlagen der Statistik , Verlag moderne industrie , Landsberg , Germany.
- Leiner B., 1994 , Stichprobentheorie, 3.Auflage, R.Oldenbourg Verlag München,Wien.
- Lohse, H.; Ludwig, R., 1982 , Prüfstatistik ,VEB Fachbuchverlag Leipzig .
- Lohse, H.; Ludwig, R.; Röhr, M., 1982 , Statisitsche Verfahren für Psychologen,Pädagogen und Soziologen, Volk und Wissen , Volkseigener Verlag ,Berlin.
- Neubauer,W., 1994 , Statistische Methoden ,Verlag franz Vahlen München .
- Schaich, E.; Köhle, D.; Scheitzer, W.; Wegner, F.; 1993 , Statistik I + II , Verlag franz Vahlen München .

وقل رب زدني علما
جامعة دمشق
Damascus University

اللجنة العلمية :

الأستاذ الدكتور إبراهيم العلي
الأستاذ المساعد الدكتور فريد الجاعوني
الأستاذ المساعد الدكتور هاني عمران

المدقق اللغوي :

الدكتور غسان السيد

- حقوق الطبع و الترجمة و النشر محفوظة لمديرية الكتب و المطبوعات -