مركز الإحصاء
STATISTICS CENTRE

دليل مبادئ التحليل الإحصائي

أدلة المنهجية والجودة- دليل رقم (10)

فهرس المحتويات

مقدمة عامة 3

الفصل الاول

1| عرض البيانات الإحصائية 4

1. مقدمة 4
2. تلخيص وعرض البيانات 4

الفصل الثاني

2| المقاييس الإحصائية 11

1. مقاييس النزعة المركزية 11
2. مقاييس التشتت 16
3. مقاييس الالتواء 23
4. مقاييس التفرطح 23

الفصل الثالث

3| التقدير الإحصائي واختبار الفرضيات 24

1. التقدير 24
2. اختبار الفرضيات 28

الفصل الرابع

4| الارتباط والانحدار الخطي البسيط 33

1. الارتباط 33
2. الانحدار 37

المراجع 42

مقدمة عامة

ضمن إطار عمل المركز على توفير أدلة إحصائية تساهم في أطلاع المستخدمين على منهجيات العمل الاحصائي بكافة انواعه سواء كان تنفيذ مسوح او استطلاعات، أو كان إجراءات معالجة وتدقيق بيانات أو قياس ومتابعة جودة العمل. يصدر هذا الدليل والذي يتضمن مبادئ عمليات التحليل الاحصائي الوصفي والاستدلالي وذلك لغايات تعريف المستخدمين بأهم الطرق المتبعة لتحليل وعرض البيانات، وبناء المقاييس الإحصائية، إضافة الى إجراءات التقدير الإحصائي واختبار الفرضيات، ومعرفة علاقات الارتباط ونماذج الانحدار الخطي بين متغيرين.

إن طرق التحليل الاحصائي وفق النظرية الإحصائية متعددة ومتشابكة تبعا لأنواع وإعداد المتغيرات وأنواع العلاقات بينها، يستعرض هذا الدليل المبادئ الأساسية لطرق التحليل الوصفي للبيانات. وإذا احتاج المحلل الى استخدام طرق اخرى، يمكنه الرجوع الى المراجع والنظريات الإحصائية المتعمقة في هذا المجال.

ولتغطية المبادئ الأساسية في عمليات التحليل الاحصائي، تم إعداد هذا الدليل والذي يتضمن أربعة فصول، يستعرض الفصل الأول طرق عرض البيانات الإحصائية بأنواعها سواء كانت بيانات إفرادية او بيانات مجمعة ضمن فترات ، إضافة الى طرق وأساليب بناء الجداول التكرارية النسبية والتجميعية، أما الفصل الثاني فيتناول المقاييس الإحصائية المختلفة وألية إعدادها، سواء كانت مقاييس النزعة المركزية المتمثلة بالمتوسط الحسابي والوسيط والمنوال وغيره ، أو مقاييس تشتت وانتشار البيانات كالتباين والانحراف المعيار والانحراف المتوسط وغيرها. أما الفصل الثالث فيتناول جانب الإحصاء الاستدلالي المتمثل بالتقدير الاحصائي سواء كانت تقديرات بنقطة واحدة او تقديرات ضمن فترات ثقة بمستوى ثقة محدد مسبقا. إضافة الى بناء واختبار الفرضيات الإحصائية البسيطة الخاصة بمؤشر النسبة او المتوسط الحسابي للمجتمع. أخيرا يتضمن الفصل الرابع أنواع معاملات الارتباط المتمثلة بمعامل ارتباط بيرسون الخاص بالمتغيرات المتصلة او معامل ارتباط سبيرمان الخاص بالبيانات المنفصلة، أضف الى ذلك معاملات الارتباط الجزئي بين عدد من المتغيرات بافتراض ثبات أحد هذه المتغيرات. كما ويتناول الفصل الرابع عملية بناء نموذج الانحدار الخطي الثنائي البسيط من خلال حساب تقديرات معالم الانحدار وبناء النموذج واختبار مدى دقته وكفاءته، واليات اجراء التنبؤ لقيم المتغير المعتمد بافتراض وجود قيم المتغير المستقل.

.

الفصل الاول

1 | عرض البيانات الإحصائية

1.1 مقدمة

علم الإحصاء هو أحد العلوم التي تختص بطرق جمع وتنظيم وتلخيص وعرض وتحليل البيانات، وذلك للوصول إلى نتائج مقبولة وقرارات سليمة على ضوء هذا التحليل.وكما هو معلوم أيضا هناك جانب من علم الإحصاء وصفيا، بحيث يعرف الإحصاء الوصفي على أنه عبارة عن الطرق الخاصة بتنظيم وتلخيص المعلومات، وذلك بهدف فهم هذه المعلومات. وبناء عليه، تأتي أهمية تناول موضوع عرض البيانات في هذا الفصل، على أنه أحد الطرق العلمية التي تتصف بتنظيم البيانات وتقديمها بطريقة سهلة للمستخدم، و بصورة يسهل فهمها والمقارنة بين مفرداتها واستنتاج بعض النتائج الأولية منها، فالطريقة التي يتمّ من خلالها عرض البيانات في العديد من المجالات ذات أهمية كبرى، فهناك طرق تشجع المتلقي على التفاعل بشكل أكبر من الطرق الأخرى، وقد تستعمل البيانات في المناهج المدرسية، وفي قطاع الأعمال، والاقتصاد، والأبحاث، وفي الاحصاءات، وفي العديد من المجالات الأخرى، لذلك يوجد العديد من الطرق المتبعة والتي يمكن من خلالها عرض البيانات بالطريقة المثلى، وبشكل يؤدّي الغرض كاملاً.

2.1 تلخيص وعرض البيانات

بعد ان يتم جمع البيانات يكون من الصعب دراستها وفهمها دون تنظيمها وجدولتها، وكثيرا ما يكون هدف الباحث من عرض البيانات هو جذب انتباه القارئ نحو العلاقة بين المتغيرات التي يدرسها او المقارنة بين المجاميع من البيانات، لذا يعمد الباحث الى تبسيطها وذلك بعرضها بأشكال معبرة وهادفة، وعلى أساس ذلك لابد من عرض هذه البيانات بشكل يتسم بالدقة والوضوح. ومن طرق عرض البيانات الآتي:

1.2.1 العرض بالجداول (Tabulation):

وتضم مجموعة من أساليب التبويب أهمها:

1. الجدول التكراري:

إن أول مرحلة لعرض البيانات الاحصائية تتكون من تصميم جدول التوزيع التكراري (Frequency Distribution)، وهو عبارة عن جدول ينظم ويلخص البيانات الإحصائية سواء كانت وصفية أو كمية، فيوزعها على عمودين ومجموعة من الصفوف، يمثل العمود الأول الصفة للبيانات الوصفية أو الفئة للبيانات الكمية، والثاني يمثل تكرار الفئة أو الصفة، وكما يظهر عدد المشاهدات من البيانات التي تقع في كل صف.

جدول 1: التوزيع التكراري لجنس الحاصلين على درجة البكالوريوس في إحدى المؤسسات (بيانات وصفية)

الجنس	التكرار
ذكر	23
أنثى	26

جدول 2: التوزيع التكراري لمتوسط أجور مجموعة من الموظفين في إحدى المؤسسات. (بيانات كمية)

الأجر (بالدرهم)	التكرار
1500	1
5000	3
6500	5
10500	1
11200	3
14000	3
16500	2
18000	1
21500	1
المجموع	20

تكوين الفئات في الجدول التكراري للمتغيرات الكمية

إن الغرض من عمل فئات أو فترات منتظمة (متساوية الطول) للبيانات هو تخفيض حجم البيانات المعروضة، من خلال تجميع القيم المتقاربة في مجموعات. ولا يوجد قواعد ثابتة لتحديد أطوال الفئات وعددها، إلا أنه يفضل ألا يكون عدد الفئات صغيرا فتضيع معالم التوزيع وتفقد كثيرا من التفاصيل. كما لا يكون كبير جدا فتضيع الحكمة من التجميع في فئات. ويعتمد تحديد عدد الفئات وطول كل فئة إلى حد كبير على الخبرة، وعلى مدى البيانات (وهو الفرق بين أكبر قيمة وأصغر قيمة للبيانات)، ولتوضيح كيفية عمل الفئات المنتظمة نأخذ بيانات جدول (2) وتكون الخطوات كالتالي:

1. نحسب طول المدى R: $R = 21500 - 1500 = 20000$
2. نختار مثلا عدد الفئات = 5 فئات
3. نحسب طول الفئة (L) بأن نقسم المدى على عدد الفئات: $L = 20000 / 5 = 4000$.
4. نختار أصغر قيمة في البيانات لتكون بداية الفئة الأولى المقربة، ويضاف إليها طول الفئة فنحصل بذلك على بداية الفئة الثانية، وهكذا لباقي الفئات. فمثلا بداية الفئة الأولى في جدول (2) هي $1500 + 4000 = 5500$
5. لإيجاد نهاية أي فئة نضيف إلى بدايتها طول الفئة مطروحا منه واحد، ونهاية الفئة الأولى مثلا هي 5499

ويتكون جدول التوزيع التكراري من خانتين. الأولى يكتب بها حدود فئات الأجر المقربة والثانية يكتب بها التكرار، وبناء عليه تم تكوين الفئات للبيانات وكما يبين الجدول أدناه.

جدول 3: حدود الفئات وتكرارها لبيانات جدول التوزيع التكراري في جدول 2.

حدود الفئات	التكرار
5499 - 1500	4
9499 - 5500	5
13499 - 9500	4
17499 - 13500	5
21499 - 17500	2
المجموع	20

2. الجدول التكراري النسبي والتكراري المئوي

ويمكن تكوين جدولين آخرين من جدول التوزيع التكراري وهما: الجدول التكراري النسبي والجدول التكراري المئوي ويتكون كل منهما من خانتين مثل الجدول التكراري العادي ولكن خانة التكرار في النسبي يكتب بها التكرار النسبي، وهو عبارة عن التكرار لأي فئة مقسوما على مجموع التكرارات. ويكون مجموع التكرار النسبي لجميع الفئات مساويا للواحد الصحيح. وأما خانة التكرار المئوي يكتب بها التكرار المئوي ويمكن الحصول عليها بضرب التكرار النسبي في 100. ويلاحظ أن مجموع التكرارات المئوية يساوى 100، كما هو موضح بالجدول (4) وباستخدام البيانات في جدول (2).

جدول 4: الجدول التكراري النسبي والمئوي

حدود الفئات	التكرار النسبي	التكرار المئوي
5499 - 1500	0.20	20%
9499 - 5500	0.25	25%
13499 - 9500	0.20	20%
17499 - 13500	0.25	25%
21499 - 17500	0.10	10%
المجموع	1.00	100%

3. الحدود الحقيقية (الفعلية) للفئات ـ Exact Interval Limits

نلاحظ مما سبق، بأن الفئات غير متصلة ببعضها، أي أن هناك بعض القيم بين فئة وأخرى لم يتم تغطيتها، لذا يفضل تحديد الحد الأعلى والأدنى الفعليين للفئة، وتحديد مراكز الفئات، وذلك لاستخدامها لاحقا في طرق عرض البيانات، والتي يتم حسابها من خلال المعادلات التالية:

1. الحد الاعلى الفعلي للفئة المطلوبة= الحد الاعلى للفئة المطلوبة + نصف وحدة طول
2. الحد الادنى الفعلي للفئة المطلوبة = الحد الادنى الفئة المطلوبة – نصف وحدة طول
3. مراكز الفئة (هو القيمة التي تتوسط الفئة) = (الحد الأعلى للفئة + الحد الأدنى للفئة) ÷ 2

ويبين الجدول أدناه الحدود الفعلية للفئات ومراكزها، المحسوبة من بيانات الجدول 2.

جدول 4: الحدود الفعلية للفئات ومراكز الفئات المحسوبة من بيانات الجدول 2:

حدود الفئات الحقيقية	مركز الفئة	التكرار
5499.5 - 1499.5	3499.5	4
9499.5 - 5499.5	7499.5	5
13499.5 - 9499.5	11499.5	4
17499.5 - 13499.5	15499.5	5
21499.5 - 17499.5	19499.5	2

4. الجدول التكراري المتجمع (Cumulative Frequency Table)

في كثير من الأحيان يكون اهتمامنا منصبا على عدد القراءات التي تكون أصغر من أو تساوى مقدارا معينا، لذلك يتم تكوين التكرار المتجمع الصاعد بجمع قيمة التكرار في الفئة الى قيمة جميع التكرارات بالفئات السابقة، ويسمى هذا التكرار بالتكرار المتجمع. ويمكن كتابة الجدول التكراري المتجمع الصاعد المكون من خانتين، الأولى يكتب فيها أقل من الحد الأدنى الحقيقي للفئة الأولى (بدلا من حدود الفئة الأولى وكذلك لباقي الفئات حتى نصل إلى الفئة الأخيرة فيكتب لها سطرين الأول منهما: أقل من الحد الأدنى الحقيقي، والثاني منها أقل من الحد الأعلى الحقيقي)، ويتضح ذلك في جدول رقم (5).

جدول 5: التوزيع التكراري المتجمع لأجور الموظفين للبيانات في جدول (2)

حدود الفئات المقربة	الحدود الدنيا للفئات الحقيقية	التكرار المتجمع
ــــــ	1499.5 >	0
5499 - 1500	5499.5 >	4
9499 - 5500	9499.5 >	9
13499 - 9500	13499.5 >	13
17499 - 13500	17499.5 >	18
21499 - 17500	21499.5 >	20

2.2.1 العرض البياني

استخدم الاحصائيون تقنيات الرسم البياني لوصف البيانات بطريقة أفضل. إذ أن العرض البياني يساعد في إعطاء صورة سريعة لوصف ظاهرة طبيعية معينة، فهو يعطي صورة عن الوضع من خلال النظر وبدون الدخول في تفصيلات، ثم يتيح إمكانية المقارنات، واستخلاص بعض المؤشرات والتفسيرات. وفيما يلي نستعرض أهم أنواع العرض البياني:

1. التمثيل بالنقاط Dot Plot

يعرف على أنه طريقة لعرض وتلخيص وتمثيل البيانات من خلال نقاط، حيث تعبر كل نقطة على المحور العمودي عن مقدار تكرار الفئة او الطبقة. ويستخدم هذا التمثيل في عمليات التحليل للتعرف على خصائص التوزيع الاحصائي للبيانات. كما ويستخدم التمثيل بالنقاط في ملاحظة القيم المتطرفة او الفجوات في مجموعة البيانات.

مثال: الرسم التالي يبين التمثيل بالنقاط لعلامات 30 طالب في مادة العلوم، حيث يبين أن أكثر العلامات تكرارا هي العلامة 70، كما ويبين الرسم توزيع البيانات اذ يبدأ بالعلامة 65 ومن ثم يتزايد تكرار العلامات عند 70، وتنخفض بعد العلامة 70، الى أن تصل للعلامة 90 وهي الأقل تكرارا.

شكل (1): التمثيل بنقاط لعلامات مجموع من الطلاب

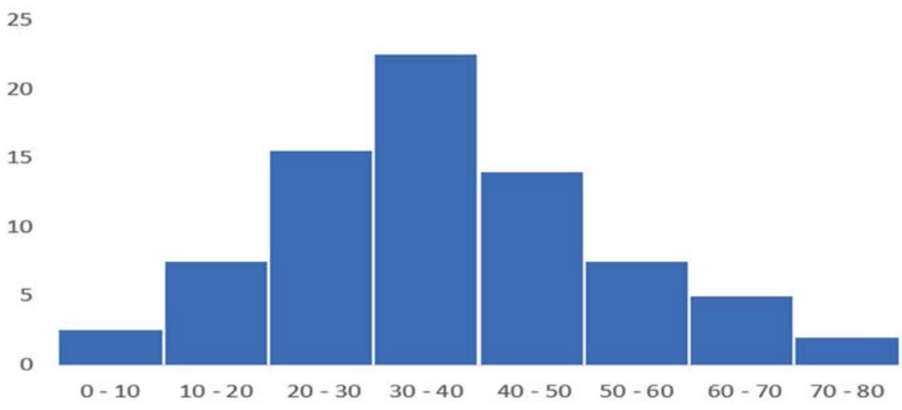
		X				
		X				
		X	X			
		X	X	X		
	X	X	X	X	X	
	X	X	X	X	X	X
	X	X	X	X	X	X
	X	X	X	X	X	X
العلامة	**65**	**70**	**75**	**80**	**85**	**90**

2. المدرج التكراري Histogram

هو عبارة عن طريقة او اسلوب لعرض وتلخيص البيانات الفئوية يستخدم لمعرفة نوع وخصائص التوزيع الاحتمالي للبيانات، ويقوم على تقسيم مدى البيانات الى مجموعات، حيث يتم تمثيل هذه المجموعات بأعمدة لكل منها، يمثل عرض العمود طول الفئة الفعلي، ويمثل ارتفاع العمود تكرار قيم البيانات في تلك الفئة.

مثال: يبين الرسم التالي المدرج التكراري لفئات الأعمار في مجتمع ما، بحيث يمثّل ارتفاع العمود عدد الأفراد بالألف، بينما عرض العمود يمثل امتداد الفئة العمرية، فمثلاً الأفراد في الفئة العمرية (30 – 40) يبلغ عددهم في المجتمع حوالي 22 ألفاً.

شكل (2): المدرج التكراري لفئات الأعمار



3. مخطط الساق والاوراق Steam and Leaf

يعتبر تمثيل الساق والأوراق واحد من طرق الرسم البياني المستخدمة لوصف البيانات الكمية والتي تستخدم بشكل واسع لتحليل البيانات، عندما تكون البيانات المتوفرة قليلة العدد نسبيا. ويتشابه التمثيل بهذه الطريقة إلى حد ما مع المدرج التكراري، وفي العادة يرفق رسم الساق والاوراق بالجدول التكراري للبيانات.

مثال: المخطط ادناه يوضح مخطط الساق والأوراق لعلامات 25 طالب (العلامة من 100). حيث أن الساق يمثل خانة العشرات، والأوراق تمثل خانة الآحاد. فمثلا العلامات 55،55،56،59 (هي الأقل في مجموعة العلامات) تم تمثيلها ضمن الساق الأول وهو العدد (5)، والأوراق تضمنت الأرقام 5، 5، 6، 9. بينما مثلا الأرقام 100، 100 هما العلامتين الاعلى في المجموعة. والمكونة من ثلاث خانات، نخصص الخانتين الأوليتين 10 كساق، ونمثل الخانة الأخيرة بالورقة (0).

شكل (3): التمثيل بمخطط الساق والأوراق لعلامات مجموعة من الطلاب

الساق	الأوراق					
5	5	5	6	9		
6	2	3	5	6		
7	2	5	6	8	9	9
8	1	5	7	7	9	
9	2	3	5	6		
10	0	0				

4. التمثيل بالصندوق والنقاط (Box and Dot plot)

هو عبارة عن تمثيل بياني يبين توزيع وانتشار البيانات، ويمكن من خلاله تحديد ما إذا كان هناك بيانات متطرفة (شاذة) او غير منسجمة مع مجموعة البيانات الرئيسية. اما طريقة رسم الصندوق فتكون من خلال تحديد الربيع الأول للبيانات أي القيمة التي تحصر أقل منها 25% من البيانات، والربيع الثالث وهي القيمة التي تحصر أقل منها 75% من البيانات. وتحديد الوسيط بينهما. إضافة الى القيمة الصغرى للبيانات التي تحدد الطرف الأول للرسم، والقيمة العظمى التي تحدد الطرف الاخر.

يتم حساب نصف المدى الربيعي من خلال طرح مقدار الربيع الثالث من مقدار الربيع الاول وقسمة الناتج على 2 .

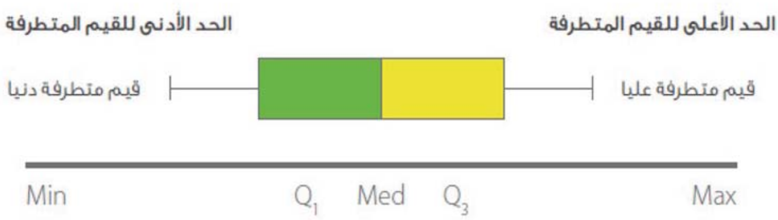
بناء على ذلك يتم حساب كل من :

الحد الادنى للقيم المتطرفة= الربيع الثالث + (نصف المدى الربيعي مضروبا في 1.5)

الحد الاعلى للقيم المتطرفة = الربيع الاول + (نصف المدى الربيعي مضروبا في 1.5)

البيانات التي تقع خارج الحدود الدنيا والعليا للقيم المتطرفة بأنها غير منسجمة او أحيانا تدعى قيمة متطرفة او شاذة.

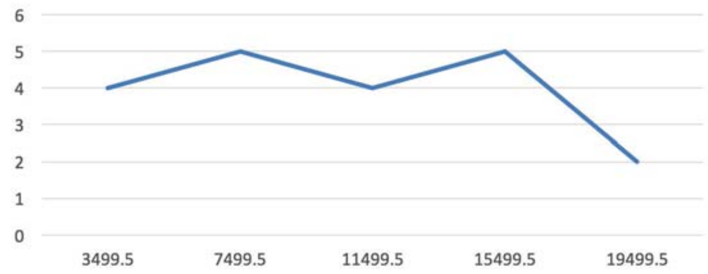
شكل (4) التمثيل بالصندوق والنقاط



5. المضلع التكراري polygon

وهو عبارة عن خطوط مستقيمة مضلعة تعبر عن حجم الظاهرة داخل محورين عامودي وأفقي، ويعبر عن البيانات برسم أكثر وضوحا من المدرج التكراري، ويصبح المضلع ممهدا كلما زاد عدد القيم في الصفوف وعدد المشاهدات. ويمكن رسم المضلع التكراري من خلال برنامج اكسل (Excel) بحيث يمثل المحور الأفقي مركز الفئات الخاصة بالظاهرة المدروسة، ويمثل المحور العمودي التكرار.

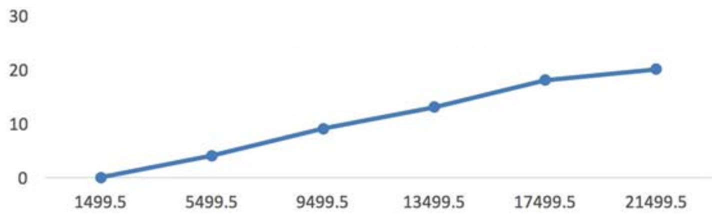
شكل (5) المضلع التكراري



6. المنحنى التكراري المتجمع الصاعد (Curve Frequency)

وهو منحنى يصف حالة أو ظاهرة معينة، حيث يمكن رسم المنحنى التكراري المتجمع الصاعد من خلال برنامج اكسل (Excel) على محورين متعامدين الأفقي يمثل مركز الفئات الفعلية، والعمودي يمثل التكرارات المتجمعة الصاعدة وتوضع النقاط في الرسم أعلى الحدود الدنيا الحقيقية للفئات بحيث يكون الارتفاع ممثلا للتكرار المتجمع الصاعد. ويتبين ذلك في الرسم البياني التالي الذي يمثل بيانات جدول 6.

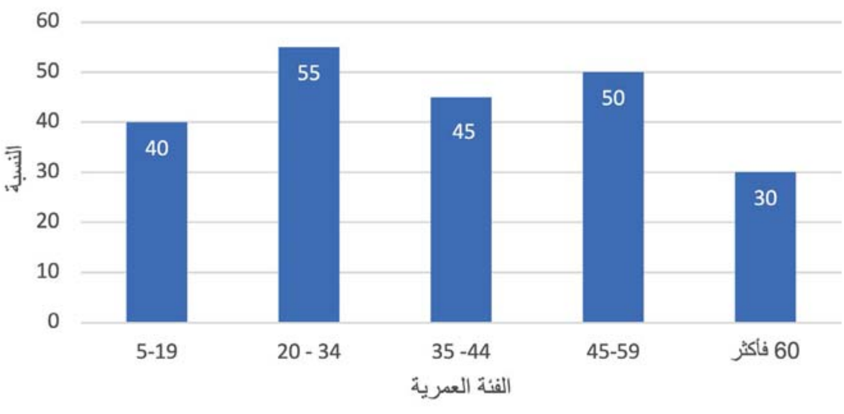
شكل (6) المنحنى التكراري المتجمع



7. الأعمدة Column Charts

تسهّل هذه الطريقة على المتلقي قراءة البيانات وعقد المقارنات بين القيم المختلفة بطريقة سهلة جداً، ممّا يسهل أيضاً عملية اتخاذ القرارات المختلفة اعتماداً على ما يراه أمامه، في هذه الطريقة يتم تمثيل الأرقام بأعمدة طولها متناسب مع القيمة التي تعبر عنها، بحيث يكون العمود الأطول للرقم ذو القيمة الأعلى والعكس صحيح، أمّا عدد الأعمدة فهو يتناسب مع عدد القيم التي يعبر عنها. من الممكن أن يتضمن المتغير المطلوب تمثيله بالأعمدة على تصنيف واحد، بحيث يتم تمثيله بعمود واحد، ويسمى هذا الرسم بالأعمدة. والرسم البياني التالي مثال على هذا الشكل من الرسوم، حيث يبين الاهتمام بالأرقام الإحصائية في مجتمع ما لكل فئة عمرية.

شكل (7) توزيع الافراد حسب الفئة العمرية



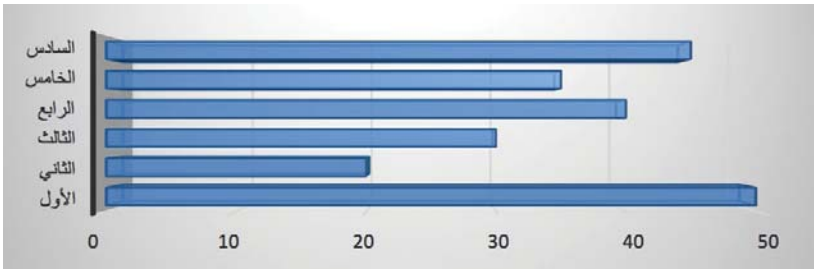
ويمكن أن يتضمن المتغير المطلوب تمثيله بالأعمدة أكثر من تصنيف، بحيث يتم تمثيل كل تصنيف بعمود منفصل ضمن الفترة الواحدة، مثل أعداد الطلبة (ذكور، إناث، مجموع) خلال سنة معينة، ويسمى هذا الرسم بالأعمدة المجمعة. والرسم البياني التالي مثال على هذا الشكل من الرسوم، حيث يبين أعداد الطلبة المقبولين في جامعة ما حسب السنوات.

شكل (8) أعداد الطلبة المقبولين في جامعة ما، حسب الجنس خلال السنوات 2001 - 2005

8. الخطوط الافقية Bar Charts

هي خطوط افقية، تأتي وفق ترتيب معين سهلة في القراءة وأقل تشويشا من الأعمدة. والرسم البياني التالي مثال على هذا الشكل من الرسوم، حيث يبين أعداد الطلبة في مدرسة ما لسنة معينة، حسب الصفوف من الأول الى السادس.

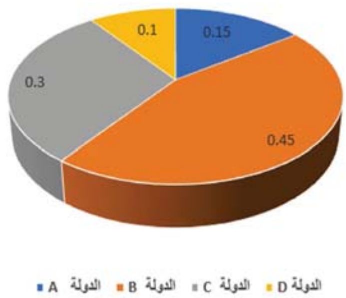
شكل (9) أعداد الطلبة حسب الصفوف



9. الدائرة البيانية (Pie Chart)

هي دائرة مقسمة الى شرائح او قطاعات، تستخدم لعرض الأهمية النسبية للمجتمع، ووضعها ضمن مجموعات مختلفة للمتغير النوعي، وتعتبر ذات الاستعمال الواسع جداً في تمثيل وعرض البيانات، بسبب ما توفره من سهولة كبيرة جداً في قراءتها. والرسم البياني التالي مثال على هذا الشكل من الرسوم، حيث يبين صادرات بلد ما، حسب دول المقصد.

شكل (10) التوزيع النسبي للصادرات حسب الدول



3.2.1 الصور

تعتبر هذه الطريقة من أكثر طرق عرض البيانات التي تثير في نفس المتلقّي الاستمتاع بشكل كبير أثناء تفاعله مع ما يتمّ عرضه أمامه من بيانات، كما وتمتاز هذه الطريقة بقدرتها العالية على جعل المتلقي قادراً على حفظ البيانات المعروضة لأطول فترة ممكنة، فالإنسان يفضّل ويحب هذه الطريقة في تلقي البيانات. تعتمد هذه الطريقة أساساً على تمثيل الأرقام صورياً بشكل يحفظ لها دلالتها.

2 | المقاييس الاحصائية

لقد تناولنا في الفصل السابق عرض البيانات الإحصائية وتلخيصها في جداول تكرارية أو رسوم بيانية، بهدف الحصول على بعض الخصائص لمجتمع الدراسة، ولكن تلك الطرق غير كافية لوصف البيانات، لذلك لا بد من وجود مقاييس عددية تصف هذه البيانات. وسنتعرض في هذا الفصل إلى اثنين من المقاييس الإحصائية وهي؛ مقاييس النزعة المركزية، ومقاييس التشتت. وسنتناول في هذا الفصل، هذه المقاييس بشيء من التفصيل، حيث أن لكل مقياس ميزاته ومحدداته، التي تعتمد على طبيعة البيانات والهدف من استخدامها.

1.2 مقاييس النزعة المركزية (Measures of central tendency)

تعرف مقاييس النزعة المركزية أو مقاييس الموقع أو المتوسطات، على أنها مقاييس عددية تحدد موقع التوزيع للبيانات، ويمكن تعريف المتوسطات بأنها القيمة النموذجية الممثلة لمجموعة من البيانات، والتي تميل إلى الوقوع في المركز، لذلك تسمى المتوسطات بمقاييس النزعة المركزية. وهي مهمة في حالة المقارنة بين التوزيعات المختلفة للبيانات. وتكون فائدتها أكثر في حالة التوزيعات المتشابهة في طبيعتها وأشكالها ولكنها مختلفة في مواقعها. فمثلا: عند دراسة الإنفاق لعينة من الأسر في الريف، وأخرى في والحضر، فإنه يمكننا المقارنة بينهما من خلال هذه المقاييس. وسوف نستعرض أهم مقاييس النزعة المركزية أدناه، حيث أن لكل منها ميزاته ومحدداته.

1.1.2 الوسط (المتوسط) الحسابي Mean

هو قيمة تتجمع حولها مجموعة من القيم، ويعتبر من أهم مقاييس النزعة المركزية والأكثر استخداما في الإحصاء والحياة العملية، ويستخدم عادة في الكثير من المقارنات بين الظواهر المختلفة.

ويحسب الوسط الحسابي رياضيا، بجمع قيم عناصر المجموعة المراد إيجاد وسطها، ويقسم المجموع على عدد العناصر، ويرمز له بالرمز:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

مثال 1: إذا كانت أجور 5 موظفين في إحدى الشركات (بالدولار) هي: 250، 280، 320، 450، 370، فإن الوسط الحسابي يحسب لها كما يلي:

$$\bar{x} = \frac{250 + 280 + 320 + 450 + 370}{5} = 334$$

1.1.1.2 الوسط الحسابي للبيانات المبوبة (الجداول التكرارية)

إذا كان لدينا عدد k من الفئات ذات المراكز $(x_1, x_2, \dots, x_{kn})$ ولها تكرارات (f1, f2, …fk) على الترتيب فإن الوسط الحسابي يعطى بالعلاقة الآتية:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k} = 1/n \sum_{i=1}^{k} f_i x_i$$

مثال 2: احسب متوسط أعمار الطلاب في الجدول التالي:

فئات العمر	6 - 5	8 - 7	10 - 9	12 - 11
عدد الطلاب	4	6	5	9

الحل: لتبسيط إجراءات الحل ننشئ الجدول التالي:

فئات العمر	مركز الفئات (x)	التكرار (f)	$f_i x_i$
6 - 5	5.5	4	22
8 - 7	7.5	6	45
10 - 9	9.5	5	47.5
12 - 11	11.5	9	103.5
المجموع		24	218

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{n} = \frac{\sum_{i=1}^{n} x_i f_i}{n}$$

2.1.1.2 أهم مميزات الوسط الحسابي:

- مقياس سهل حسابه ويخضع للعمليات الجبرية بسهولة، ويعتبر أكثر المقاييس استخداما في الإحصاء.
- يأخذ في الاعتبار جميع القيم محل الدراسة.
- يكون المتوسط الحسابي محصورًا دائمًا بين أكبر وأصغر قيمة في العيّنة
- مجموع انحرافات القيم عن الوسط الحسابي للعينة يساوي صفرا.

3.1.1.2 بعض محددات الوسط الحسابي:

- يتأثر بالقيم الشاذة (المتطرفة) وهي القيم الكبيرة جدا أو الصغيرة جدا مقارنة بباقي القيم.
- يصعب حسابه في حالة الجداول التكرارية المفتوحة، حيث يتطلب ذلك معرفة مركز كل فئة.
- لا يمكن حسابه في حالة البيانات الوصفية.

2.1.2 الوسيط Median

يُعَرِّف علماء الإحصاء الوسيط (Median): بأنه المقياس الذي يستخدم لقياس القيمة المتوسطة التي تكون القيم الأكثر منها تساوي القيم الأقل منها، أو بعبارة أخرى: هو المقياس الذي يقوم بعملية فصل متساوٍ للنصف الأعلى من البيانات عن النصف الأدنى، بحيث يأخذ بالاعتبار ترتيب البيانات، ويختلف حساب الوسيط في حالة البيانات غير المبوبة عنها في المبوبة.

1.2.1.2 الوسيط للبيانات غير المبوبة

لحساب الوسيط نتبع الخطوات التالية:

- ترتيب البيانات (المشاهدات) ترتيبا تصاعديا أو تنازليا.
- تحديد رتبة الوسيط والتي تساوي 2/(n+1). حيث n هو عدد قيم البيانات، فإذا كان عدد البيانات فرديا، يكون الوسيط المشاهدة التي تقع في المنتصف، وإذا كان عددها زوجيا فإن الوسيط هو متوسط المشاهدتين اللتين تقعان في المنتصف.

مثال 3: الوسيط للبيانات التالية: 52، 15، 102، 68، 44

الحل: يتم ترتيب البيانات تصاعديا وتحديد رتبة كل بيان كالتالي:

قيمة الوسيط	**15**	**44**	**52**	**68**	**102**
رتبة الوسيط	**1**	**2**	**3**	**4**	**5**

من خلال البيانات أعلاه نحدد رتبة الوسيط = 3 وعليه فإن قيمة الوسيط هي 52.

مثال 4: الوسيط للبيانات التالية: 52، 15، 102، 68، 44، 72

الحل: يتم ترتيب البيانات تصاعديا وتحديد رتبة كل بيان كالتالي:

قيمة الوسيط	15	44	52	60	68	72	102
رتبة الوسيط	1	2	3	3.5	4	5	6

من خلال البيانات نحدد رتبة الوسيط = 3.5 وفي هذه الحالة تكون رتبة الوسيط بين الرتب (3، 4)، ويتم حسابها كما يلي:

الوسيط = $\frac{68 + 52}{2}$ = 60، وفي هذه الحالة يقع الوسيط بين المشاهدتين (52، 68).

2.2.1.2 الوسيط للبيانات المبوبة

إذا كان لدينا عدد k من الفئات ذات المراكز ($x_1, x_2, \dots, x_{kn}$) ولها تكرارات (f1, f2, …fk) على الترتيب.

نتبع الخطوات التالية لحساب الوسيط حسابيا:

- نكون الجدول المتجمع الصاعد باستخدام الحدود الحقيقية.
- نجد رتبة الوسيط.
- نحدد الفئة الوسيطية وهي الفئة التي يكون تكرارها المتجمع اكبر من او يساوي تكرار رتبة الوسيط.

نحدد تكرار الفئة السابقة للفئة الوسيطية (f1) . ونحدد طول الفئة الوسيطية (L) . ومن ثم نحدد الحد الادنى الفعلي للفئة الوسيطية (A) . ويكون مقدار الوسيط حسب المعادلة التالية::

$$Med = A + \frac{(n/2 - f_1)}{f_2 - f_1} L$$

مثال (5): الوسيط لأعمار الطلاب في المثال (2) السابق:

الحل: نكون جدول التكرار المتجمع الصاعد (كما تم شرحه في الفصل السابق) بحيث يصبح كالآتي:

التكرار المتجمع (f)	**فئات العمر المتجمع**
0	4.5 >
4	6.5 >
10	8.5 >
15	10.5 >
24	12.5 >

نحسب (n/2)، فيكون حاصل القسمة 12 وهو محصور على عمود التكرار المتجمع بين 15 و 10، وبتطبيق المعادلة يكون:

$$A = 8.5, \quad f_1 = 10, \quad f_2 = 15, \quad L = 15 - 10 = 3$$

وبتطبيق قانون الوسيط نحصل على:

$$Med = 8.5 + \frac{(12 - 10)}{15 - 10} \times 3 = 9.7 \text{ سنة}$$

3.2.1.2 مميزات الوسيط:

- لا يتأثر بالقيم المتطرفة. يمكن إيجاده في حالة البيانات الوصفية التي يمكن ترتيبها.
- مجموع الانحرافات المطلقة عن الوسيط أقل ما يمكن مقارنة بأي قيمة حقيقية.

3.2.1.2 محددات الوسيط:

- لا يأخذ جميع القيم في الاعتبار عند حسابه.
- لا يسهل التعامل معه في التحاليل الإحصائية والرياضية.

3.1.2 المنوال Mode

يعرف المنوال على أنه القيمة الأكثر تكرارا في مجموعة البيانات. ويكثر استخدامه في حالة البيانات الوصفية، لمعرفة النمط (المستوى) الشائع. وقد يكون لمجموعة البيانات منوال واحد ولذلك يطلق عليها وحيدة المنوال، أو يكون لها أكثر من منوال وتسمى متعددة المنوال. وقد لا يكون لمجموعة البيانات أي منوال وبذلك تسمى عديمة المنوال.

مثال (6): المنوال من البيانات التالية: 8، 6، 4، 2، 8، 15، 8

الحل: يوجد لهذه البيانات منوال واحد وهو القيمة 8.

مثال (7): المنوال من البيانات التالية: 8، 6، 4، 2، 8، 15، 8، 12، 4

الحل: يوجد لهذه البيانات منوال واحد وهو القيمة 8.

وفي حالة البيانات المبوبة أو الجداول التكرارية لا يمكن القول بأن قيمة معينة يكون لها أكبر تكرار لأن القيم تذوب داخل الفئات المختلفة، ولذلك يمكن القول بأن هناك فئات منواليه وهي الفئات التي يقابلها أعلى تكرار.

1.3.1.2 مميزات المنوال:

- مقياس سهل حسابه ولا يتأثر بالقيم الشاذة.
- يمكن إيجاده للقيم الوصفية والتوزيعات التكرارية المفتوحة.

2.3.1.2 محددات المنوال:

- عند حساب المنوال لا تؤخذ جميع قيم البيانات في الاعتبار. قد يكون لبعض البيانات أكثر من منوال وبذلك لا يمكن تحديد قيمة وحيدة للمنوال.

4.1.2 الوسط الهندسي Geometric Mean

الوسط الهندسي (GM) هو نوع من المتوسطات أو المعدّلات التي تقيس النزعة المركزية أو القيمة النموذجية لمجموعة معطيات، يتم حسابه عن طريق حساب الجذر من الدرجة الـ (n) لحاصل ضرب حدود المجموعة، حيث (n) هو عدد الحدود. أي المتوسط الهندسي للقيم $(x_1, x_2, \dots, x_n)$ هو:

$$GM = \sqrt[n]{x_1 . x_2, \dots, . x_n}$$

$$GM = \sqrt[2]{2 \times 8} = 4$$

$$GM = \sqrt[3]{1 \times 2 \times 4} = 2$$

مثال (8): الوسط الهندسي للقيم 2، 8

الحل: الجذر التربيعي لحاصل ضربهما، أي:

$$GM = \sqrt[2]{2 \times 8} = 4$$

مثال (9): الوسط الهندسي للقيم 1، 2، 4

الحل: الجذر التكعيبي لحاصل ضربهما، أي:

$$GM = \sqrt[3]{1 \times 2 \times 4} = 2$$

1.4.1.2 مميزات ومحددات الوسط الهندسي

- من مميزاته أنه لا يتأثر بالقيم المتطرفة
- بينما لايمكن استخدامه مع البيانات التي تضم قيما سالبة او صفر.

5.1.2 الوسط التوافقي Harmonic Mean

يكون الوسط التوافقي مفيد، إذا كانت المتغيرات على شكل نسب، فهو يستخدم عندما يكون مقلوب المتغير له دلاله كأن يعين نسبة بين متغيرين مرتبطين مثل السرعة بالنسبة للزمن. والوسط التوافقي H لمجموعة من القيم هو مقلوب الوسط الحسابي لهذه القيم، أي أن:

$$H = \frac{1}{\bar{x}}$$

مثال (10): أحسب الوسط التوافقي للقيم 10، 7، 8، 6، 14، 9

الحل: نحسب الوسط الحسابي للقيم، فيكون

$$\bar{x}_{\text{غير المبوبة}} = \frac{10+7+8+6+14+9}{6} = 9$$

$$H = \frac{1}{\bar{x}} = \frac{1}{9}$$

هذا المثال للبيانات غير المبوبة، وبالنسبة للبيانات المبوبة يتم تطبيق نفس المعادلة اعلاه، وذلك بعد حساب الوسط الحسابي الذي يتم للبيانات المبوبة.

2.2 مقاييس التشتت (Measures of Dispersion)

مقاييس التشتت هي مقاييس عددية تستخدم لقياس درجة تجانس (تقارب) أو تشتت (تباعد) مفردات البيانات عن بعضها البعض. ومقاييس التشتت تستخدم لوصف مجموعة البيانات، وكذلك لمقارنة مجموعات البيانات المختلفة، إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

1.2.2 المدى Range

يعتبر المدى من أسهل مقاييس التشتت تعريفا وحسابا، حيث أنه يعطينا فكرة سريعة عن مدى تفرق البيانات ويرمز له بالرمز (R). ويعرف المدى لمجموعة من البيانات بالصيغ التالية:

1.1.2.2 في حالة البيانات غير المبوبة:

- المدى (R) = أكبر قيمة – أصغر قيمة

مثال 11: المدى للبيانات التالية: 54، 89، 65، 70، 95، 47

الحل: المدى = 95 – 47 = 84

2.1.2.2 أما في حالة البيانات المبوبة فإن المدى يعرف بأكثر من طريقة، نذكر منها الطريقتين الآتيتين:

- المدى (R) = مركز الفئة العليا – مركز الفئة الدنيا.
- المدى (R) = الحد الأعلى للفئة العليا – الحد الأدنى للفئة الدنيا.

مثال 12: المدى للفئات العمرية في الجدول التالي:

الحل: نقوم بتحديد حدود ومراكز الفئات كما مر معنا في الفصل السابق والمبين جدول البيانات.

فئات العمر	15 - 6	25 - 16	35 - 26	45 - 36	55 - 46	65 - 56
حدود الفئات الفعلية	15.5 - 5.5	25.5 - 15.5	35.5 - 25.5	45.5 - 35.5	55.5 - 45.5	65.5 - 55.5
مركز الفئات	10.5	20.5	30.5	40.5	50.5	60.5
التكرار (f)	10	16	14	6	9	5

حسب الطريقة الأولى: المدى

(R) = 60.5 – 10.5 = 50

حسب الطريقة الثانية: المدى

(R) = 65.5 – 5.5 = 60

ونلاحظ اختلاف كل من الطريقتين في حساب قيمة المدى. ولكن غالبا ما تستخدم الطريقة الأولى في إيجاد المدى.

3.1.2.2 مميزات المدى

- سهل التعريف والحساب.
- يعطى فكرة سريعة عن طبيعة البيانات، ويستخدم كثيرا في ظواهر الحياة المختلفة مثل مراقبة جودة الإنتاج وكذلك في وصف طبيعة الأحوال الجوية.

4.1.2.2 محددات المدى

- يعتمد في حسابه على قيمتين من البيانات، ولا يأخذ بالاعتبار باقي القيم.
- يتأثر بالقيم الشاذة، وبالتالي فهو لا يعطى صورة صادقة عن طبيعة البيانات. لذلك فهو مقياس تقريبي.

5.1.2.2 نصف المدى الربيعي Mid – Range Quartile

نصف المدى الربيعي هو نصف المدى بين الربيع الأول (Q_1)، والربيع الثالث (Q_3) ويرمز له بالرمز Q، ويعرف بالصيغة التالية:

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

حيث Q_1 يمثل القيمة التي يسبقها ربع البيانات بعد ترتيبها تصاعديا، و Q_3 يمثل القيمة التي يسبقها ثلاثة أرباع البيانات بعد ترتيبها تصاعديا. وبناء عليه فإن الفرق بين Q_1، Q_3 يسمى المدى الربيعي وهو النصف الأوسط للبيانات.

6.1.2.2 حساب نصف المدى الربيعي للبيانات غير المبوبة

- ترتب البيانات تصاعديا.
- نجد قيمة الربيع الأول (Q_1).
- نجد قيمة الربيع الأول (Q_3).
- تطبيق لعلاقة الرياضية السابقة.

مثال 13: أحسب نصف المدى الربيعي للبيانات التالية: 53، 89، 65، 70، 95، 47، 74، 86

الحل: ترتيب البيانات تصاعديا فتصبح: 47، 53، 65، 70، 74، 86، 89، 95

الربيع الأول في البيانات أعلاه هو القيمة التي ترتيبها الثاني من بين البيانات، حيث ان رتبة الربيع هي حاصل ضرب عدد القيم n في نسبة الربيع كأن تكون 25% أو 50%، أو 75%. وعليه:

$$Q_1 = X_{(2)} = 53$$

$$Q_3 = X_{(6)} = 86$$

$$Q = \frac{86 - 53}{2} = 69.5$$

7.1.2.2 نصف المدى الربيعي للبيانات المبوبة

يحسب نصف المدى الربيعي لهذه البيانات بطريقة الفروق، ويتم حساب الربيع الأول والثالث حسب العلاقات المبينة أدناه والتي تم شرحها سابقا لحساب الوسيط، مع اختلاف بسيط بوضع مقدار n/2 عند حساب الربيع الثالث (Q_3):

$$Q_1 = A_1 + \frac{(n/4 - f_1)}{f_2 - f_1} L$$

$$Q_2 = A_2 + \frac{(3n/4 - f_3)}{f_4 - f_3} L$$

حيث؛ A_1 الحد الحقيقي للتكرار السابق لفئة الربيع الأول، وA_2 الحد الحقيقي للتكرار السابق لفئة الربيع الثالث، وتمثل L طول الفئة ويساوى الحد الأدنى للفئة التالية مطروحا منه الحد الأدنى للفئة السابقة، ويمثل f_1: التكرار المتجمع السابق لتكرار (Q_1) و f_3 التكرار المتجمع السابق لتكرار (Q_3). ويمثل f_2 التكرار المتجمع اللاحق لتكرار (Q_1)، و f_4 التكرار المتجمع اللاحق لتكرار (Q_3)

مثال 14: جد نصف المدى الربيعي (Q) حسابيا للفئات العمرية في مثال 12.

الحل: ننشئ جدول التكرار المتجمع الصاعد كما هو مبين أدناه، ومن ثم نحسب Q من العلاقات المبينة أعلاه.

التكرار المتجمع الصاعد	حدود الفئات
0	5.5
10	15.5
26	25.5
40	35.5
46	45.5
55	55.5
60	65.5

$$n = 60, \quad n/4 = 15, \quad 3n/4 = 45, \quad L = 10$$

$$Q_1 = 15.5 + \frac{(15 - 10)}{26 - 10} \quad 10 = 18.6$$

$$Q_3 = 35.5 + \frac{(45 - 24)}{39 - 24} \quad 10 = 43.8$$

(نصف المدى الربيعي) $$Q = \frac{18.6 - 43.8}{2} = 31.2$$

8.1.2.2 مميزات نصف المدى الربيعي:

- لا يتأثر بالقيم الشاذة أو المتطرفة.
- يمكن حسابه من التوزيعات التكرارية المفتوحة من الطرفين.

9.1.2.2 محددات نصف المدى الربيعي:

- لا يأخذ جميع القيم في الاعتبار.
- لا يسهل التعامل معه في التحليل الإحصائي.

2.2.2 الانحراف المتوسط Mediterranean Deviation

يعرف الانحراف المتوسط بأنه متوسط الانحرافات المطلقة للبيانات عن وسطها الحسابي، ويرمز له بالرمز MD.

1.2.2.2 الانحراف المتوسط للبيانات غير المبوبة

يتم حساب الانحراف المتوسط للبيانات غير المبوبة من خلال العلاقة التالية:

$$MD = 1/n\sum_{i=1}^{n} |x_i - \bar{x}|$$

حيث X_i تمثل المشاهدات، n عدد المشاهدات.

مثال 13: الانحراف المتوسط للبيانات التالية: 5، 9، 7، 14، 11، 8، 12، 6

الحل: بعد حساب الوسط الحسابي نكون الجدول التالي:

$$\bar{x} = 1/n\sum x = 72/8 = 9$$

x	$x - \bar{x}$	$\|x - \bar{x}\|$
5	-4	4
9	0	0
7	-2	2
14	5	5
11	2	2
8	-1	1
12	3	3
6	-3	3
72	0	20

من بيانات الجدول أعلاه، وباستخدام علاقة الانحراف المتوسط للبيانات غير المبوبة نحسب قيمته:

$$MD = 9/8 = 1.13$$

2.2.2.2 الانحراف المتوسط للبيانات المبوبة

يتم حساب الانحراف المتوسط للبيانات المبوبة من خلال العلاقة التالية:

$$MD = 1/n\sum_{i=1}^{n} f_i |x_i - \bar{x}|$$

حيث X_i تمثل مركز الفئة، و f_i تكرار المشاهدات، n مجموع التكرارات.

مثال 14: جد الانحراف المتوسط للفئات العمرية في مثال 10.

الحل: نكون الجدول أدناه بعد حساب الوسط الحسابي:

$$\bar{x} = 1/n\sum_{i=1}^{n} f_x = 1860/60 = 31$$

مركز الفئات (X)	التكرار (f)	f_x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$	$f \lvert x - \bar{x} \rvert$
10.5	10	105	-20.5	20.5	205
20.5	16	328	-10.5	10.5	168
30.5	14	427	-0.5	0.5	7
40.5	6	243	9.5	9.5	57
50.5	9	454.5	19.5	19.5	175.5
60.5	5	302.5	29.5	29.5	147.5
المجموع	60	1860			760

من بيانات الجدول أعلاه، وباستخدام علاقة الانحراف المتوسط للبيانات المبوبة نحسب قيمته:

$$MD = 760/60 = 12.7$$

3.2.2 الإنحراف المعياري Standard Deviation

يعتبر الانحراف المعياري من أهم وأفضل مقاييس التشتت وأكثرها شيوعا واستخداما في التحليل الإحصائي. ونظرا لكون الانحراف المعياري هو الجذر التربيعي لتباين البيانات، لا بد من تعريف التباين Variance الذي هو متوسط مربع انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز، وفكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها، فالتباين يكون كبيرا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس. ويمكن حساب التباين من العلاقة التالية:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_{i-\bar{X}})^2}{N}}$$

حيث $x - \bar{x}$ انحراف البيانات عن متوسطاتها الحسابية، N عدد مفردات البيانات.

ويعرف الانحراف المعياري على أنه الجذر التربيعي للتباين ويرمز له بالرمز σ. وكما هو الحال في التباين، فالزيادة في قيمته تدل على درجة كبيرة في تشتت أو تذبذب وتباعد البيانات، والعكس اذا انخفضت قيمته. ومن خلال التباين يمكن حساب الانحراف المعياري لمجتمع إحصائي بالعلاقة التالية:

$$s^2 = \frac{\sum_{i=1}^{n}(x_{i-\bar{X}})^2}{n-1}$$

1.3.2.2 الانحراف المعياري للبيانات غير المبوبة

في حالة العينة التي حجمها n المأخوذة من المجتمع، فإن التباين يرمز له بـ S^2، ويعرف بالعلاقة:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_{i-\bar{X}})^2}{n-1}}$$

ومن العلاقة السابقة للتباين، يرمز للانحراف المعياري بـ S، وتحسب قيمته بالعلاقة التالية:

مثال 15: الانحراف المعياري لأعمار ستة طلاب (x) في المرحلة الابتدائية: 5، 8، 6، 9، 7، 10

الحل: نكون الجدول أدناه بعد حساب الوسط الحسابي:

$$\bar{x} = {}^{1}/_{n} \sum x = {}^{45}/_{6} = 7.5$$

$(x - \bar{x})^2$	$(x - \bar{x})$	(x)
6.25	-2.5	5
0.25	0.5	8
2.25	-1.5	6
0.25	-0.5	7
2.25	1.5	9
6.25	2.5	10
17.5	0	المجموع

من بيانات الجدول أعلاه، وباستخدام علاقة التباين للبيانات غير المبوبة نحسب قيمته:

$$s^2 = \frac{17.5}{6-1} = 3.5$$

ومن التباين نحسب الانحراف المعياري بالعلاقة التالية:

$$s = \sqrt{3.5} = 1.871$$

2.3.2.2 الانحراف المعياري للبيانات المبوبة

إذا كان لدينا عدد k من الفئات ذات المراكز $(x_1, x_2 \ldots x_k)$ ولها تكرارات $(f_1, f_2 \ldots f_k)$، على الترتيب فإن حساب التباين للبيانات المبوبة يتم من خلال العلاقة التالية:

$$s^2 = \frac{\sum_{i=1}^{n} f_i (x_{i-\bar{x}})^2}{n-1}$$

ومن التباين يحسب الانحراف المعياري بالعلاقة التالية:

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_{i-\bar{x}})^2}{n-1}}$$

مثال 16: احسب الانحراف المعياري للفئات العمرية في المثال 10.

الحل: نكون الجدول أدناه بعد حساب الوسط الحسابي:

$$\bar{x} = 1/n \sum_{i=1}^{n} f_x = 1/60 \times (1860) = 31$$

مركز الفئات (X)	التكرار (f)	f_x	$x - \bar{x}$	$\lvert x - \bar{x} \rvert$	$f \lvert x - \bar{x} \rvert$
10.5	10	105	-20.5	420.25	4202.5
20.5	16	328	-10.5	110.25	1764
30.5	14	427	-0.5	0.25	3.5
40.5	6	243	9.5	90.25	541.5
50.5	9	454.5	19.5	380.25	3422.25
60.5	5	302.5	29.5	870.25	4351.25
المجموع	60	1860	-	-	14285

من بيانات الجدول أعلاه، وباستخدام علاقة التباين للبيانات المبوبة نحسب قيمته:

$$s^2 = 242.12$$

ومن التباين يحسب الانحراف المعياري بالعلاقة التالية:

$$s = 15.56$$

4.2.2 القيمة (الدرجة) المعيارية Standard Value

القيمة المعيارية هي مقياس يقيس الانحرافات عن الوسط الحسابي بوحدات من الانحراف المعياري ويرمز لها بالرمز Z. فإذا كان لدينا المتغير X وله القيم $(x_1, x_2 \ldots x_n)$ والتي لها المتوسط $(\bar{x})$، والانحراف المعياري S، فإن Z تحسب من العلاقة التالية:

$$Z_i = \frac{x_i - \bar{x}}{S}$$

مثال 16: إذا كانت علامة أحد الطلاب 86 في مادة المحاسبة، وكان متوسط العلامات هو 77، وذلك بانحراف معيارى 11 درجة، وإذا كانت علامته في مادة الاقتصاد 96، و كان متوسط العلامات هو 84، وذلك بانحراف معياري 17 درجة، ففي أي المقررين كان أداء الطالب أفضل؟

الحل: باستخدام العلاقة أعلاه تحسب الدرجة المعيارية للمادتين.

- الدرجة المعيارية للمحاسبة:

$$Z_i = \frac{87 - 77}{11} = 0.82$$

- الدرجة المعيارية للاقتصاد:

$$Z_i = \frac{96 - 84}{11} = 0.7$$

وهذا يدل على أن أداء الطالب في المحاسبة أفضل من أدائه في الاقتصاد بالرغم من أن درجته في المحاسبة أقل.

3.2 مقاييس الالتواء Skewness

الالتواء (Sk) هو درجة عدم التماثل أو الانحراف عن التماثل، فإذا كان منحنى توزيع الشكل العام للبيانات له طرف على يمين مركز التوزيع أطول من الطرف الأيسر، فان التوزيع يسمى ملتوي لليمين أو أن له التواء موجب، وإذا حدث العكس يقال إن التوزيع ملتوي لليسار أو أنه سالب الالتواء.

هناك الكثير من الطرق لقياس الالتواء في التوزيع التكراري أو مجموعة من البيانات , وسنذكر منها كما في العلاقات التالية:

$$Sk = \frac{3(\bar{x}-M)}{S}$$

$$Sk = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{S^3(N-1)}$$

حيث $\bar{x}$ الوسط الحسابي، وM الوسيط، وS الانحراف المعياري، X_i قيم المتغير.

ويعطي هذا المقياس النسبي إشارة سالبة للالتواء جهة اليسار، وإشارة موجبة للالتواء جهة اليمين، ويمتد الالتواء من (3-) في الالتواء السالب الى (3+) في الالتواء الموجب، ويتلاشى الالتواء عندما يصبح الفرق بين الوسيط والوسط الحسابي صفراً وذلك عندما يكون التوزيع اعتدالياً أو ما يسمى بالتوزيع الطبيعي.

4.2 مقاييس التفرطح Kurtosis

هو مقياس يقيس درجة علو أو انخفاض أي منحنى توزيع تكراري بالنسبة للمنحنى الطبيعي للبيانات، وهو منحنى متماثل حول الرأس يمر بالمتوسط، فإذا كان للتوزيع قمة مرتفعة (أكبر من التوزيع الاعتدالي) يقال أنه مدبب Leptokurtic. وإذا كان التوزيع ذو قمة مسطحة يقال أنه مفلطح Platykurtic، وإذا كانت قمة التوزيع متوسطة (ليست مدببة وليست مفلطحة) يسمى متوسط التفلطح Mesokurtic. وصفة التفلطح ليس لها علاقة بالمتوسط الحسابي للتوزيع فقد يكون هناك أكثر من توزيع لهم نفس المتوسط الحسابي ولكن يختلف شكل المنحنى من مدبب أو مسطح

وحيث أن ارتفاع قمة التوزيع الاعتدالي تساوي 3 تقريباً، فان التوزيع يكون مفلطحاً عندما يكون معامل التفلطح أقل من 3، ويكون التوزيع مدبباً عندما يكون معامل التفلطح أكبر من 3. يحسب معامل التفلطح من الصيغة الرياضية التالية:

$$Sk = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{S^4(N-1)}$$

3 | التقدير الإحصائي واختبار الفرضيات

تعتمد دراسة خصائص أي مجتمع إحصائي على طبيعة وأسلوب التعامل مع مفرداته. فعندما تؤخذ كل المفردات بأسلوب العد الشامل او التعداد فأن دراسة خصائصه تتم عبر التعرف على المؤشرات الإحصائية لتوزيع المجتمع في ضوء قيم معلماته (Parameters) وخصائصها، ومن هذه المعلمات المتوسط، الوسيط، المعدل، الانحراف المعياري، الخ. وعند دراسة المجتمع ليس من خلال عملية العد الشامل انما من خلال اخذ عينة من مفرداته، فتتم دراسة خصائص المجتمع بإجراء عملية تقدير احصائي للمعلمات من معطيات العينة المختارة. وتسمى كل من هذه المقدرات بالاحصاءة Statistic. الا ان قرار قبول المقدرات واعتمادها في دراسة خصائص المجتمع الذي سحبت منه العينة، مرتبط بعملية تقييم لتلك المقدرات، لأن تقدير المؤشر الاحصائي المسحوب من مفردات العينة قد لا يساوي معلمة المجتمع. وان وجود فروق بين هذين المؤشرين قد يكون بالإمكان تجاهلها إذا كانت بسيطة. وليس لها خطورة على قيمة المقدر، ولكن قد يكون الفرق واضحا وجوهريا.

ان عملية التقدير الاحصائي تهدف الى إيجاد أفضل مقدر او أكثر لمعلمات المجتمع. اما اختبار الفرضيات الإحصائية فينطوي على بناء أساليب تعتمد على البيانات قيد الدراسة لاتخاذ قرار بشأن فرضية تصاغ قبل التعامل مع بيانات العينة. الا ان التمييز بين عمليتي التقدير والاختبار لا يعكسه فصل هاتين العمليتين فهما مترابطتان، ما يستدعي الحاجة الى عرض عن كل منهما.

1.3 التقدير

ترتبط عملية التقدير بمجموعة من مشكلات إحصائية، يجري التعامل معها بصيغة الاستدلال الذي يقود الى تصورات دقيقة، قدر الإمكان، للبحث في قيمة او أكثر من قيم معلمات المجتمع. وتتم عملية التقدير اما بالسعي للحصول على قيمة مقدرة محددة (Point Estimate) مشتقة من بيانات عينة من المجتمع، بحيث نحاول جعلها أقرب ما يمكن الى قيمة المعلمة الحقيقية، او بحساب حدود يتوقع ان تقع القيمة الحقيقية للمعلمة ضمنها باحتمال معين، وكلما كان ذلك الاحتمال عاليا كلما كانت هناك موثوقية أكبر في حصولنا على القيمة الحقيقة للمعلمة ضمن مدى الثقة.

1.1.3 التقدير بنقطة:

ان تقدير النقطة هو اجراء يتقرر بموجبه اعتماد المقدر $\overline{\Theta}$ لمعلمة المجتمع Θ، ولا يعتمد الفرق الذي يعبر عنه بالمقدار

$(\overline{\Theta} - \Theta)$ للحكم على دقة عملية التقدير، ويستعاض عنه أحيانا بمربع الفرق $(\overline{\Theta} - \Theta)^2$ وذلك للتخلص من اثر الإشارة في الفروق، او أحيانا يستعاض عنه بالقيمة المطلقة، حيث يتم اختيار المقدر الذي يجعل القيمة المتوقعة لمربعات الفروق بين قيمة المعلمة وقيمة المقدر أقل ما يمكن، ويسمى المقدر حينها بالمقدر ذو أقل متوسط مربعات خطأ.

مثال:

إذا كان لدينا مجتمع أسر امارة ابوظبي، وكان المطلوب هو التعرف على مؤشر متوسط إنفاق الفرد في الامارة. فان ذلك يتطلب بيانات شاملة لكافة افراد سكان امارة ابوظبي ويتم سؤالهم عن مقدار متوسط كل فرد، مما يعني اجراء عملية مسح شامل لكافة الافراد الامر الذي سيؤدي الى كلفة مالية عالية والى وقت زمني طويل ونتيجة المؤشر وهو متوسط إنفاق الفرد لن تكون بالمستوى المطلوب من الدقة وذلك لكبر حجم عملية جمع البيانات وكثرة الفرق الميدانية العاملة واحتمالات الأخطاء المختلفة التي ستنعكس بالنهاية على قيمة معلمة المجتمع.

ان البديل المناسب لعملية المسح الشامل هو اجراء مسح بالعينة من خلال اختيار عينة من افراد الامارة واستيفاء بيانات الانفاق لكل فرد من افراد العينة وحسب متوسط انفاق الفرد من بيانات العينة أي مقدار $\overline{\Theta}$ واعتباره تقديرا لمعلمة متوسط انفاق مجتمع افراد الامارة أي مقدار Θ.

السؤال المطروح هنا، كيف نقيس مقدار دقة القيمة المقدرة للمؤشر. ان المقدر يتضمن نسبة خطأ محددة ولكنها غير معروفة وهذا الخطأ من مصدرين رئيسيين الأول من العينة ويدعى بالأخطاء العينية، والثاني من إجراءات وعمليات جمع البيانات ويدعى بالأخطاء غير العينية والتي لا يمكن قياسها وانما يمكن الحد منها لتكون في أقل مستوى من خلال ضبط إجراءات وعمليات جمع ومعالجة البيانات.

اما خطأ العينة فيمكن قياسه بالاعتماد على مقدار الانحراف المعياري S لبيانات افراد العينة العشوائية البسيطة التي تشتمل على n من الوحدات، من خلال حساب ما يعرف بخطأ المعاينة وهو: $\frac{S}{\sqrt{N}} \times \left(\frac{N - n}{N}\right)$، حيث N هو حجم المجتمع الكلي.

مثال:

اذا كان المطلوب هو تقدير متوسط حجم الاسرة في دولة الامارات فأن معلمة المجتمع المبحوث هنا هو متوسط حجم الاسرة، اخذت عينة حجمها 5000 أسرة، وتم استيفاء بيان عدد الافراد لكل اسرة، بعد ذلك تم حساب مقدر لمتوسط حجم الاسرة في الدولة اعتمادا على بيانات العينة وكان المقدر هو 6.4.

وعند حساب مقدار الانحراف المعياري لبيانات العينة وجد ان S=2.121، من هنا ان مقدار خطأ المعاينة يحسب من خلال قسمة الانحراف المعياري على الجذر التربيعي لحجم العينة، ويتم تجاهل المعامل $(\frac{N-n}{N})$ باعتبار انه سيكون مقدار صغير جدا عندما تكون حجم العينة كبير.

بناء على ما سبق ان خطأ المعاينة هو: $\frac{2.121}{\sqrt{5000}} = 3.0\%$

2.1.3 التقدير بفترة (فترة الثقة)

ان أسلوب التقدير بفترة وكما أسلفنا يتم من خلال تحديد بالاعتماد على بيانات العينة حدود لفترة تدعى فترة الثقة، يتوقع ان تكون معلمة المجتمع متواجدة فيها بمستوى ثقة محدد مسبقا كأن يكون 95% مثلا او 90% أو غير ذلك.

ان تحديد كل من الحد الأدنى والحد الأعلى لفترة الثقة، يتطلب افتراض ان البيانات تتوزع توزيع احتمالي طبيعي Normal distribution، و/أو يكون حجم او عدد مفردات العينة كبير نسبيا.

بناء على ما سبق فان فترة الثقة تتحدد بالاعتماد على التقدير النقطي للمعلمة مضافا اليه ومطروح من مقدار w، ويعبر عن ذلك بالمقدار التالي:

$$W = Z\left(1-{}^{a}/_{2}\right)\frac{S}{\sqrt{n}}$$

حيث أن: $Z\left(1-{}^{a}/_{2}\right)$ هو مقدار التوزيع الطبيعي المعياري عند مستوى الثقة المحدد مسبقا، فمثلا عند مستوى ثقة 95% يكون مقدار $a=5\%$، ومقدار الثابت $Z\left(1-{}^{a}/_{2}\right)$ من جدول التوزيع الطبيعي 1.96 اما عند مستوى ثقة 90% فأن هذا المقدار 1.654، وعند مستوى ثقة 85% فهو 1.28. هذا ويدعى المقدار w بهامش الخطأ او حد الخطأ في فترة الثقة.

بناء على ما سبق تبنى فترة الثقة على النحو التالي:

$$\left[\bar{\theta} - Z_{(1-{}^{a}/_{2})}\frac{S}{\sqrt{n}}\ ,\ \ \bar{\theta} + Z_{(1-{}^{a}/_{2})}\frac{S}{\sqrt{n}}\right]$$

يمكن القول بأن مقدار معلمة المجتمع يتوقع ان تكون ضمن هذه الفترة باحتمال او بمستوى ثقة مقداره $\left(1-{}^{a}/_{2}\right)\%$

مثال (1):

بالاعتماد على المثال السابق، اذا كان المطلوب هو تقدير متوسط حجم الاسرة في الدولة بفترة عند مستوى ثقة مقداره 95%، في هذه الحالة نحسب حد الخطأ على النحو التالي:

$$W = \frac{1.96 \times 2.121}{\sqrt{5000}} = 0.0588$$

اما فترة الثقة فهي عبارة عن التالي:

$$\left[6.4 - 0.0588,\ 6.4 + 0.0588\right]$$

وهي تساوي

$$\left[6.34,\ 6.46\right]$$

مثال (2)

تنتج احدى الشركات الغذائية نوع من العصير زنة العبوة 125 جرام، قام مدير مراقبة الإنتاج بسحب عينة عشوائية حجمها 36 عبوة وقياس كمية الكربوهيدرات بالجرام، ووجد ان متوسط كمية الكربوهيدرات 12 جرام، والانحراف المعياري 2.4 جرام، فاذا اراد قسم مراقبة الإنتاج تقدير فترة ثقة 95% لمتوسط كمية الكربوهيدرات في العبوات اذا كان وزن الكربوهيدرات يتبع التوزيع الطبيعي، فان حد الخطأ هو

$$W = \frac{Z(1-{}^{a}/_{2})S}{\sqrt{n}} = \frac{1.96 \times 2.4}{\sqrt{36}} = 0.784$$

وبالتالي فأن فترة الثقة هي: $[11.22,\ 12.78]$

إذا اننا نثق بنسبة 95% بأن كمية الكربوهيدرات يتراوح وزنها بين (11.22 الى 12.78).

ملاحظة، اذا كانت بيانات المجتمع لا تخضع للتوزيع الطبيعي او كان عددها قليل (أقل من 30 مشاهدة)، في هذه الحالة ولحساب فترة الثقة نفترض خضوع البيانات للتوزيع الاحصائي (t) وبدرجات حرية مقدارها n-1. بالاعتماد على القيم الجدولية للتوزيع (t) عند مستوى ثقة معينة يمكن الحصول على قيمة (t) من الجدول الاحصائي الخاص به. ونضعها في المعادلة أعلاه عوضا عن قيمة توزيع Z قيمة (t) الجدولية عند مستوى الثقة المحدد مسبقا ($1-a$).

$$\left[\bar{\theta} - t_{(1-{}^{a}/_{2})}\frac{S}{\sqrt{n}}\ ,\ \ \bar{\theta} + t_{(1-{}^{a}/_{2})}\frac{S}{\sqrt{n}}\right]$$

3.1.3 التقدير الإحصائي في البيانات غير الموزونة

في هذا الجزء يقصد بالبيانات غير الموزونة أي البيانات التي ليس لجميعها الأهمية النسبية او الوزن ذاته في المجتمع الذي تنتمي اليه، ويكون الاختلاف ناتج عن اختلاف الظاهرة التي تمثلها البيانات داخل أجزاء المجتمع نفسه.

فعندما تكون البيانات تختلف في أهميتها داخل المجتمع الواحد فأن حساب قيمة المؤشر من تلك البيانات لا تتم بالطريقة المباشرة، أي بالتعامل مع البيانات جميعها على انه لها نفس الوزن.

فمثلا اذا كان جزء (i) من المجتمع المبحوث يتضمن W_iمن البيانات وكانت قيمة المشاهدة لتلك البيانات هي X_i، وكانت أجزاء المجتمع متعددة (i<1) فأن قيمة المؤشر Θ سواء كان متوسط او نسبة تساوي:

$$\Theta = \frac{\sum W_i \bar{x}}{\sum W_i}$$

فمثلا عند دراسة الاسر في عدد من المدن للتعرف على متوسط إنفاق الاسرة بشكل كامل، ليس جميع الاسر لديها نفس الوزن او الأهمية وبالتالي للحصول على متوسط إنفاق الاسرة الواحدة لا يجوز ان تجمع بيانات إنفاق جميع الاسر في جميع المدن وتقسم على عددها من اجل الحصول على متوسط إنفاق الاسرة. انما هناك متغير حجم او وزن المدينة التي توجد الاسرة بها يلعب دورا مؤثرا في حجم انفاقها. إذا في هذه الحالة يحسب المتوسط بالشكل التالي:

$$\bar{X} = \frac{\sum W_i \bar{x}}{\sum W_i}$$

حيث ان:

$\bar{x}_i$: متوسط المشاهدات في فئة المجتمع (i)

W_i: هو الوزن او الأهمية النسبية لفئة المجتمع (i)

مثال:

إذا كان عدد الاسر ومتوسط حجم الاسرة في عدد من المدن هو كما في الجدول ادناه، فكيف يمكن حساب متوسط حجم الاسرة الكلي.

المدينة	متوسط حجم الاسرة (x)	عدد الاسر (w)	X. W
أ	5.5	1800	9900
ب	6.3	2500	15750
ج	4.7	3000	14100
د	5.8	800	4640
هـ	7.0	1200	8400
المجموع		9300	52790

في هذا المثال ان متوسط حجم الاسرة على المستوى الكلي هو 5.7

$$\overline{X} = \frac{\sum W_i \overline{x}}{\sum W_i} = \frac{52790}{9300} = 5.7$$

ان ما سبق ينطبق أيضا على مؤشر النسبة او المعدل، فاذا كان لدينا عدد من المجتمعات الجزئية وكانت نسبة المتغير في كل جزء من الأجزاء مختلفة فتكون الاوزان او الاهميات النسبية للبيانات مختلفة، وبالتالي فأن النسبة المقدرة كمؤشر احصائي على مستوى المجتمع ككل هي على النحو التالي:

$$p = \frac{\sum W_i p_i}{\sum W_i}$$

حيث ان p هو النسبة او المعدل على مستوى المجتمع الكلي، بينما p_i هي النسبة او المعدل لقيم المتغير في الجزء (i) من المجتمع.

مثال:

اذا كان نسبة الإصابة بمرض معين لدى فئات السكان تختلف بحسب النوع الاجتماعي، وكانت هذه النسب كما في الجدول التالي، فأن النسبة الكلية للإصابة بالمرض على مستوى المجتمع ككل تحسب كما في الجدول التالي:

النوع الاجتماعي	نسبة الاصابة (p)	عدد الأفراد (w)	X. W
ذكور	20.4	2800	57120
اناث	8.7	5200	45240
المجموع		8000	102360

وفي هذا المثال ان نسبة الإصابة بالمرض في المجتمع ككل هي:

$$p = \frac{\sum W_i p_i}{\sum W_i} = \frac{102360}{8000} = 12.8$$

اما أسلوب تقدير مؤشر المجموع τ بدلا من المتوسط او النسبة، فيتم اعتمادا على مؤشرات المجموع في أجزاء المجتمع المختلفة (i). أي ان

$$\tau = \sum \tau_i = \sum N_i \bar{x}_i$$

حيث ان:

τ_i هو مجموع قيم المؤشر في الجزء (i) من المجتمع.

N: هو عدد وحدات المجتمع الكلية.

مثال:

اخذت عينة من المنشآت الاقتصادية من كل مدينة بهدف تقدير متوسط الإيرادات على مستوى المجتمع، علما بأن الأهمية النسبية الممثلة بأعداد المنشآت في المدن غير متساوية أي لابد هناك من اوزان نسبية على مستوى كل مدينة من المدن.

المدينة	حجم العينة المختارة من المنشآت (n)	مجموع الإيرادات (x)	عدد المنشآت الكلي N	متوسط الإيرادات في المنشأة الواحدة	المجموع الكلي للإيرادات
a	30	15000	150	500	75000
b	20	10000	90	500	45000
c	10	8500	60	850	51000
المجموع	60	33500	300		171000

وفق الجدول أعلاه وباستخدام المعادلة السابقة تقدير المجموع الكلي للإيرادات هو عبارة عن مجموع العمود الأخير في الجدول:

$$\tau = \sum \tau_i = \sum N_i \bar{x}_i = 150 x 500 + 90 x 500 + 60 x 850 = 171000$$

2.3 اختبار الفرضيات

1.2.3 مفهوم اختبار الفرضيات

في البداية تم التطرق إلى وسائل دراسة معالم المجتمع المجهولة وذلك عن طريق فترات الثقة لهذه المعالم واستخدامها كمعلومة مساندة في عملية اتخاذ القرارات، حيث يتم استخدام بيانات عينة عشوائية مسحوبة من المجتمع المراد تقدير معالمه لحساب فترة الثقة المطلوبة عند مستوى ثقة $(1-a)$. وعليه فان النتيجة المحصلة من خلال فترات الثقة يمكن غالباً صياغتها نصيا بالشكل التالي:

باحتمال $(1-a)$ نحن متأكدون بان فترة الثقة سوف تحتوي على القيمة الحقيقية المجهولة لمعلمة المجتمع.

ومن الملاحظ أن فترة الثقة يتم حسابها بالاعتماد على بيانات عينة عشوائية، ليتم استخدام تلك الفترة في عمليات لاستدلال الإحصائي حول القيمة الحقيقية لمعلمة المجتمع. ولكن في الواقع العملي غالبا ما يكون هنالك ادعاء مسبق حول قيمة المعلمة المجهولة. وليس بالضرورة أن يكون الادعاء مرتبط بقيمة محددة حيث يمكن أن يكون الادعاء ذا صيغة رياضية، كان ينص مثلا على أن قيمة المعلمة لا تزيد عن قيمة محددة أو أن تكون أكبر من قيمة محددة. في هذه الحالة يكون الهدف من الاستدلال الإحصائي أكثر تحديداً منه في عملية حساب فترة ثقة، حيث يكون منصباً حول البحث في مصداقية الادعاء المطروح وبالتالي الوصول إلى قرار بقبول أو رفض الادعاء.

يطلق على عملية التعامل مع الافتراضات والحكم على مصداقيتها بعملية اختبار الفرضيات. وتوجد علاقة بين كل من حساب فترة ثقة واختبار الفرضيات، حيث يمكن القول بان اختبار الفرضيات تعطي معلومة أكثر استخداماً في اتخاذ القرارات من المعلومة المحصلة من حساب فترات الثقة. بيد انه يمكن الاعتماد على فترات الثقة في بعض الحالات للوصول إلى نتائج حول صحة فرضية من عدمها.

في عمليات اختبار الفرضيات يكون هنالك ادعاء أو افتراض يراد اختباره، ويتم في البداية افتراض عدم صحة الادعاء ومن ثم استخدام بيانات الدراسة لإثبات العكس، أي إثبات صحة الادعاء. وتلك الآلية تعطي اختبار الفرضيات قوة نابعة من تلافي التحيز وعدم الدقة، حيث أن الضعف في أداء الدراسة وجمع البيانات يصب في مصلحة عكس الادعاء ومن ثم لا يمكن قبول ادعاء إلا إذا كان هنالك مؤشر إحصائي قوي على ذلك.

2.2.3 أنواع الفرضيات الإحصائية

تنقسم الفرضيات الإحصائية إلى قسمين:

1. الفرضية البديلة: Alternative Hypotheses H_a

وهو ما يود الباحث أن يثبت صحته، ويوصي به في كثير من الأحول.

1. الفرضية العدم: Null Hypotheses H_o

وهو ما يود الباحث أن يثبت ضده.

هناك ثلاثة اتجاهات لصياغة الفرضيات البديلة:

الفرضية البديلة H_a	الفرضية العدم H_o
$H_a : \mu \neq \mu o$	$H_o : \mu = \mu o$
$H_a : \mu > \mu o$	$H_o : \mu \leq \mu o$
$H_a : \mu < \mu o$	$H_o : \mu \geq \mu o$

حيث أن μo هو قيمة متوسط المجتمع تحت صحة الفرضية العدم

ويلاحظ أن الفرضية العدم مصاحب دائماً بعلامة =، لذلك يمكن كتابة الفرضية العدم على الصورة:

$H_o : \mu = \mu o$

القرارات التي يمكن اتخاذها حول الفرضية العدم H_o، هناك أربعة احتمالات ممكنة حول الفرضية العدم هي:

القرار	الفرضية العدم صحيح	الفرضية العدم غير صحيح
رفض الفرضية العدم H_o	α	(1 - β)
قبول الفرضية العدم H_o	(1 - α)	β

اذا الاحتمالات هي:

1. احتمال رفضية الفرضية العدم اذا كان صحيح (احتمال وقوع خطأ من النوع الأول)

P(reject Ho when Ho is correct) =α

2. احتمال قبول الفرضية العدم اذا كان صحيح

P(accept Ho when Ho is correct) = (1−α)

3. احتمال قبول الفرضية العدم اذا كان غير صحيح (احتمال وقوع خطأ من النوع الثاني)

P(accept Ho when Ho is uncorrected) =β

4. احتمال رفض الفرضية العدم اذا كان غير صحيح

P(reject Ho when Ho is uncorrected) = (1−β)

والخطأ من النوع الأول يمكن التحكم فيه، والذي يحدده الباحث قبل الاختبار، ويطلق على احتمال وقوعه اسم مستوى المعنوية (α)، وأغلب القيم المستخدمة هي 0.05، 0.01

وكمحصلة يمكن القول بان لكي يتم وضع فرضية عدم وفرضية بديلة رياضيا لابد من تحقيق عدة شروط هي:

- يجب في البداية تحديد المعلمة المجهولة القيمة والمطلوب إجراء الاختبار عليها، حيث يمكن أن تكون متوسط مجتمع أو الفرق بين متوسطين أو نسبة حدوث حدث في مجتمع أو الفرق بين نسبتين أو تباين مجتمع أو نسبة تباينين.
- يتم تحديد القيمة المقابلة للمعلمة المجهولة والمتعلقة بالادعاء المطلوب اختباره
- يتم تحديد اتجاه العلاقة بين المعلمة والقيمة المقابلة، والتي يلزم أن تكون في إحدى ثلاث صيغ هي < أو > أو، والتي ستكون الفرضية البديلة الممثلة للادعاء.
- يتم في هذه الخطوة صياغة فرضية العدم، حيث تضم مكونات الفرضية البديلة مع تبديل العلاقة الرياضية بين المعلمة والقيمة المقابلة مع تغيير إشارة المتباينة لتعكس الحالة المقابلة للفرضية البديلة، وبالتالي لتمثل عكس الادعاء.

مثال 1

بين فيما يلي معطيات الادعاء (المعلمة، القيمة المقابلة والعلاقة الرياضية) وقم بصياغة كل من فرضية العدم والفرضية البديلة.

1. ادعاء مدير أحد الدوائر الاقتصادية بان الوقت المستغرق في المتوسط لصيانة أي آلة اقل من 12 ساعة.
2. يدعي أحد المصانع الوطنية للبطاريات الكهربائية بان متوسط عمر البطارية المنتجة بواسطة المصنع أكثر من 1.5 سنة.
3. يدعي إحدى الباحثين بان نسبة الطلاب الحاصلين على إنذارات أكاديمية في جامعة الشيخ خليفة اقل من 0.30 من إجمالي عدد الطلاب في الجامعة.
4. يدعي إحدى المستثمرين بان نسبة الربح في المتوسط من الاستثمار في اسهم سوق ابوظبي المالي لا تساوي (تختلف عن) 0.10

الحل (1)

المعلمة: متوسط الوقت المستغرق لصيانة آلة (μ)

القيمة المقابلة: 12 يوم

العلاقة الرياضية: اقل من (>)

الفرضيات:

$Ho: \mu \geq 12$

$Ha: \mu < 12$

الحل (2)

المعلمة: متوسط عمر البطارية (μ)

القيمة المقابلة: 1.5 سنة

العلاقة الرياضية: أكثر من (<)

الفرضيات:

$Ho: \mu \leq 1.5$

$Ha: \mu > 1.5$

الحل(3)

المعلمة: نسبة الطلاب الحاصلين على إنذارات (P)

القيمة المقابلة: (0.30)

العلاقة الرياضية: اقل من (>)

الفرضيات:

$Ho: P \geq 0.3$

$Ha: P < 0.3$

الحل (4)

المعلمة: نسبة الربح من الاستثمار في الأسهم (P)

القيمة المقابلة: 10%

العلاقة الرياضية: لا تساوي (≠)

الفرضيات:

$Ho: P = 0.1$

$Ha: P \neq 0.1$

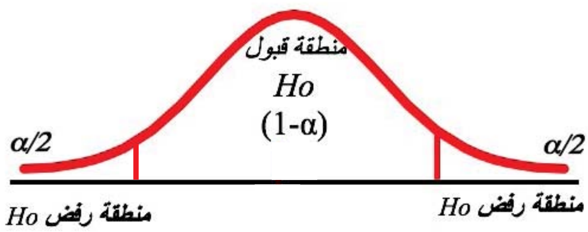
ويمكن تلخيص ما سبق فيما يلي:

- الخطأ من النوع الأول هو رفض فرضية عدم صحيحة ويرمز لاحتمال وقوعه بالرمز α ويطلق علية مصطلح مستوى المعنوية.
- الخطأ من النوع الثاني هو قبول فرضية عدم خاطئة ويرمز لاحتمال وقوعه بالرمز β.
- مستوى الثقة (1-α) هو احتمال قبول فرضية عدم صحيحة.
- قوة الاختبار (1-β) هو احتمال رفض فرضية عدم خاطئة.

3.2.3 أنواع اختبارات الفروض

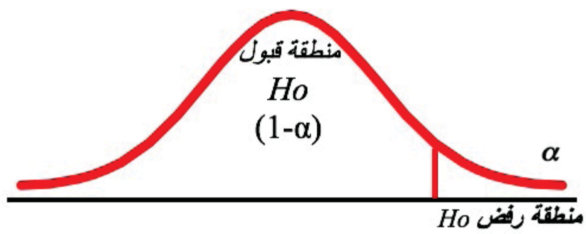
هناك نوعان لاتجاهات الفروض، يتحدد نوع الاتجاه المستخدم بناء على نوع الفرض البديل كما يلي:

1. الاختبار في اتجاهين (اذا كان الفرض البديل $(H_a: \mu \neq \mu_0$ في هذه الحالة تقع منطقة الرفض في طرفي المنحنى.

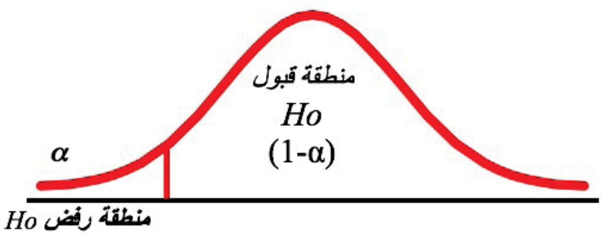


2. الاختبار في اتجاه واحد بمعنى أن منطقة الرفض α تقع جميعها في طرف المنحنى الأيمن، أو في الطرف الأيسر كما يلي:

- اذا كان الفرض البديل Ha :μ>μo وقعت منطقة الرفض في الطرف الأيمن من المنحنى كما هو مبين في الشكل ادناه:



- اذا كان الفرض البديل Ha :μ< μo وقعت منطقة الرفض في الطرف الأيسر من المنحنى كما هو مبين في الشكل ادناه:



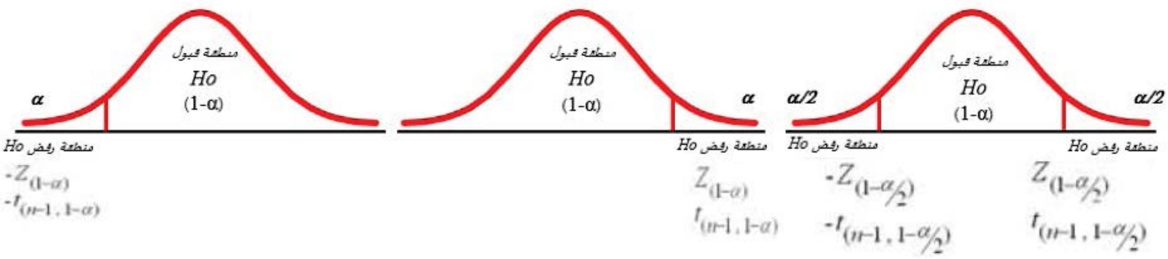
خطوات إجراء اختبار ات الفروض

1. صياغة الفروض

الفرضية البديلة *Ha*	الفرضية العدم *Ho*
Ha :μ ≠μo	*Ho* :μ=μo
Ha :μ > μo	*Ho* :μ≤μo
Ha :μ<μo	*Ho* :μ≥μo

2. تحديد مستوى المعنوية α وتوزيع المعاينة، وتحديد مناطق القبول والرفض.

توزيع المعاينة، أما توزيع طبيعي قياسي أو توزيع t بدرجات حرية (n-1)، ويتم استخراج القيم الحرجة من الجداول والتي تحدد مناطق القبول أو الرفض كما هي مبينة في الشكل التالي:



3. حساب إحصائية الاختبار

باستخدام بيانات العينة، ومتوسط المجتمع تحت صحة الفرض العدم Ho: μ= μo، يمكن حساب قيمة تسمى «إحصائية الاختبار» أو القيمة المحسوبة، وتتحدد حسب معلومية تباين المجتمع أو عدم معلوميته كما هو في الجدول أدناه:

حجم العينة	إحصائية الاختبار (القيمة المحسوبة)	تباين المجتمع σ^2
لا يشترط حجم معين للعينة	$= Z^* = \dfrac{(\bar{x} - \mu_o)}{\dfrac{\sigma}{\sqrt{n}}}$	معلوم
n ≤ 30	$= t^* = \dfrac{(\bar{x} - \mu_o)}{\dfrac{S}{\sqrt{n}}}$	غير معلوم
n > 30	$= t^* \sim Z^* = \dfrac{(\bar{x} - \mu_o)}{\dfrac{S}{\sqrt{n}}}$	غير معلوم

4. اتخاذ القرار بخصوص الفرضية العدم

اذا وقعت القيمة المحسوبة (خطوة رقم 3) في منطقة الرفض (خطوة رقم 2)، فترفض فرضية العدم أي بمعنى اخر تقبل الفرضية البديلة.

4 | الارتباط والانحدار الخطي البسيط

1.4 الارتباط

يوفر تحليل الارتباط وسيلة للاستدلال على قوة العلاقة بين متغيرين أو اكثر، أي أن الارتباط هو مقياس للدرجة التي تتغير فيها قيم المتغير بأسلوب منتظم. وهو يعتبر مؤشر كمي لتحديد درجة الاعتماد على متغير أو اكثر في التنبؤ بقيم متغير آخر. من المهم معرفة ما يمكن أن يوفره التحليل الارتباط وبنفس الأهمية يتوجب معرفة مالا يمكن أن يوفره هذا النوع من التحليل. فتحليل الارتباط لا يقدم أية معلومات للتنبؤ بقيم متغير ما، كما انه لا يوفر أي مؤشر فيما لو كانت العلاقة بين المتغيرات سببية، انما يستطيع التحليل تحديد فقط فيما لو كان درجة التباين المشترك ذات دلالة. ولذا تعرف العلاقة بين الظاهرتين او متغيرين بالارتباط. وقد يكون الارتباط طرديا بمعنى ان تتغير الظاهرتين في نفس الاتجاه بحيث اذا زادت احدى الظاهرتين تميل الثانية الى الزيادة وبالعكس. وقد يكون الارتباط عكسيا بمعنى ان تتغير الظاهرتان في اتجاهين متضادين بحيث اذا زادت احدى الظاهرتين تميل الثانية الى النقصان وبالعكس.

يلاحظ ان قيمة معامل الارتباط هي قيمة عددية نسبية تنحصر بين 1+ و1- ولا تكون هذه القيمة 1+ و1- الا اذا كان الارتباط تاما.

1.1.4 أنواع الارتباط:

يقسم الارتباط الى عدة أنواع، وذلك بحسب نوع المتغير المراد قياسه، اذ هناك متغيرات كمية مقاسة، وهناك متغيرات أخرى نوعية، قياسها لا يعتمد على كميات عددية.

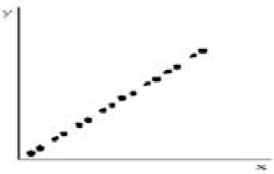
1.1.1.4 معامل الارتباط للظواهر للمتغيرات الكمية:

ويشمل دراسة العلاقة فيما بين الظواهر المقاسة، وهي الظواهر القابلة للقياس الكمي او العددي. وهذا يشمل جميع الظواهر التي يمكن التعبير عنها بصورة رقمية كالطول والدخل وكمية الإنتاج وغير ذلك من الظواهر التي يمكن التعبير عنها رقميا. ويقسم الى عدة انواع

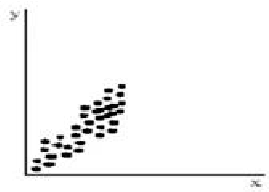
معامل الارتباط البسيط (معامل بيرسون)

وهو معامل ارتباط يحدد مقدار او حجم العلاقة واتجاهها بين متغيرين اثنين، وذلك في الحالات او الظواهر التي تقتصر فيها الدراسة على متغيرين. مثال قد يكون من المطلوب التعرف على حجم العلاقة واتجاهها بين اطوال مجموعة من الأشخاص واوزانهم. او قد يكون الهدف مثلا التعرف على حجم واتجاه العلاقة بين مقدار الدخل الشهري وحجم الانفاق الشهري للأسر في مجتمع ما.

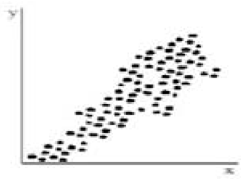
للارتباط عدة أنواع يمكن التعرف عليها من خلال كل من مقدار معامل الارتباط ومن خلال اتجاه العلاقة بين المتغيرين بالاعتماد على لوحة انتشار البيانات.



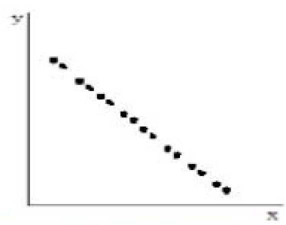
ارتباط طردي تام



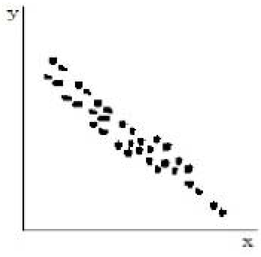
ارتباط طردي قوي



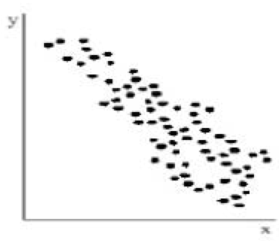
ارتباط طردي



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي

والجدول التالي يلخص أنواع الارتباط واتجاه العلاقة بين المتغيرين

قيمة معالم الارتباط	نوع علاقة الارتباط
+1	ارتباط طردي تام
من 0.70 الى 0.99	ارتباط طردي قوي
من 0.50 الى 0.69	ارتباط طردي متوسط
من 0.01 الى 0.49	ارتباط طردي ضعيف
0	لا يوجد ارتباط خطي

وكذلك الحال وبنفس المستوى تكون علاقة الارتباط عكسية في حالة كانت إشارة معامل الارتباط سالبة.

اما المعادلة العامة لحساب معامل الارتباط بين قيم متغيرين X وY فهي على الشكل التالي

$$r^2_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}$$

حيث ان:

$\sum xy$: مجموع حاصل ضرب قيم x في قيم y

$\sum x$: مجموع قيم المتغير x

$\sum y$: مجموع قيم المتغير y

$\sum x^2$: مجموع مربعات قيم المتغير x

$\sum y^2$: مجموع مربعات قيم المتغير y

مثال:

سجلت قراءات تقريبية لحجم الإنتاج (x) وحجم الصادرات (y) خلال عدة سنوات البيانات التالية:

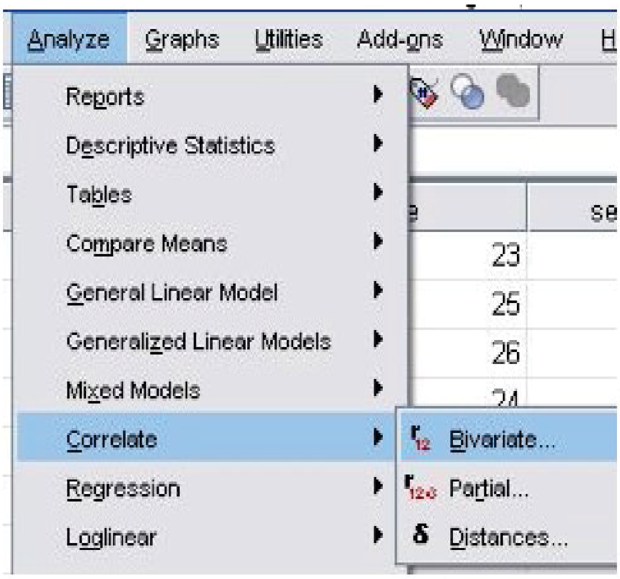
y^2	x^2	xy	الصادرات (y) بالمليون	الإنتاج (x) بالمليون
4	9	6	2	3
4	16	8	2	4
4	4	4	2	2
1	4	2	1	2
1	4	2	1	2
1	4	2	1	2
$\sum y^2 = 15$	$\sum x^2 = 41$	$\sum xy = 24$	$\sum y = 9$	$\sum x = 15$

بناء على ما سبق، ان معامل ارتباط بيرسون يحسب كالتالي:

$$r^2_{xy} = \frac{6(24) - (15)(9)}{\sqrt{\left((6x41) - 15^2\right)\left((6x15) - 9^2\right)}} = 0.65$$

وبما ان معامل الارتباط يساوي 0.65 فيمكن ان تصنف العلاقة بين كل من حجم الإنتاج والصادرات بانها علاقة ارتباط طردي متوسط.

هـذا ومـن الممكـن الاعتمـاد علـى البرمجيـات الإحصائيـة التـي تمكـن مـن حسـاب معامـل الارتبـاط بسـهولة جـدا، فمثـلا يمكـن اسـتخدام برنامـج SPSS مـن خـلال الصفحـة التاليـة لحسـاب معامـلا الارتبـاط بيـن متغيريـن.



معامل الارتباط المتعدد

هـو معامـل ارتبـاط يوضـح العلاقـة بيـن متغيـر تابـع واحـد وعـدد مـن المتغيـرات المسـتقلة الأخـرى. علـى سـبيل المثـال يسـتخدم هـذا المعامـل للتعـرف علـى نـوع علاقـة الارتبـاط بيـن حجـم الإنتـاج للدونـم الواحـد مـن القمـح و كل مـن كميـة الامطـار، كميـة السـماد، درجـة الحـرارة. فـي هـذه الحالـة يقيـس هـذا المعامـل الارتبـاط بيـن حجـم الإنتـاج كمتغيـر تابـع ومنظومـة مـن المتغيـرات المسـتقلة الأخـرى التـي يتبـع لهـا هـذا المتغيـر.

اما معادلة حساب معامل الارتباط المتعدد فتكون على النحو التالي:

$$R^2_{1.23} = \sqrt{\frac{R^2_{12} + R^2_{13} - 2R_{12}R_{13}R_{23}}{1 - R^2_{23}}}$$

حيث إن

R^2_{12}: مربع معامل الارتباط البسيط بين المتغيرين الأول والثاني.

R^2_{13}: مربع معامل الارتباط البسيط بين المتغيرين الأول والثالث

R_{12}: معامل الارتباط البسيط بين المتغيرين الأول والثاني

R_{23}: معامل الارتباط البسيط بين المتغيرين الثاني والثالث

مثال:

أراد مـدرب السـباحة معرفـة العلاقـة بيـن زمـن سـباحة (100) متـر حـرة (المتغيـر التابـع) وكل مـن الانعكاسـات العضليـة (المتغيـر المسـتقل الأول) وانعـكاس الجهـاز الـدوري التنفسـي (المتغيـر المسـتقل الثانـي) وكان معامـل الارتبـاط البسـيط بيـن المتغيـرات كمـا يلـي:

* زمن الأداء والانعكاسات العضلية (0.82) R_{12}
* زمن الأداء وانعكاس الجهاز الدوري التنفسي (0.86) R^2_{13}
* الانعكاسات العضلية وانعكاس الجهاز الدوري التنفسي (0.80) R_{23}

وبالتطبيق بالمعادلة أعلاه نحصل على معامل الارتباط المتعدد $R^2_{1.23} = 0.88$

معامل الارتباط الجزئي

هـو معامـل ارتبـاط يقيـس العلاقـة بيـن متغيريـن اثنيـن بافتـراض ثبـات تأثيـر المتغيـر الثالـث علـى كل المتغيريـن. ويرمـز لـه بالرمـز: $\rho_{12.3}$. فيسـتخدم مثـلا لأيجاد قـوة او حجـم العلاقة بيـن متغير ضغـط الـدم وقيـاس السـكر في الـدم، بافتـراض مثـلا ثبات مسـتوى الكوليسـترول علـى العلاقة. اما قانـون حسـاب معامـل الارتبـاط الجزئي بيـن متغيـر Y فيكـون مـن خـلال الصيغـة التاليـة:

$$\rho_{y2.1} = \frac{r^2{}_{y2} - r^2{}_{y1}\, r^2{}_{12}}{\sqrt{(1-(R^2{}_{y1})^2)(1-(R^2{}_{12})^2)}}$$

حيث ان:

$r^2{}_{y2}$: هو معامل الارتباط البسيط بين المتغير y والمتغير الثاني

$r^2{}_{12}$: معامل الارتباط البسيط بين المتغيرين الأول والثاني

$r^2{}_{y1}$: هو معامل الارتباط البسيط بين المتغير y والمتغير الأول

مثال:

أرادت مؤسسـة للدعايـة والإعـلان معرفـة العلاقـة بيـن عـدد المسـتجيبين لإعلاناتهـا y وحجـم الإعـلان المنشـور فـي الصحيفـة x_1 وعـدد الصحـف الموزعـة x_2 التـي تنشـر الإعـلان وحصلـت المؤسسـة علـى البيانـات التاليـة:

عدد المستجيبين بالمئات (y_i)، حجم الإعلان بالإنش (x_1)، عدد الصحف الموزعة بالآلف (x_2)

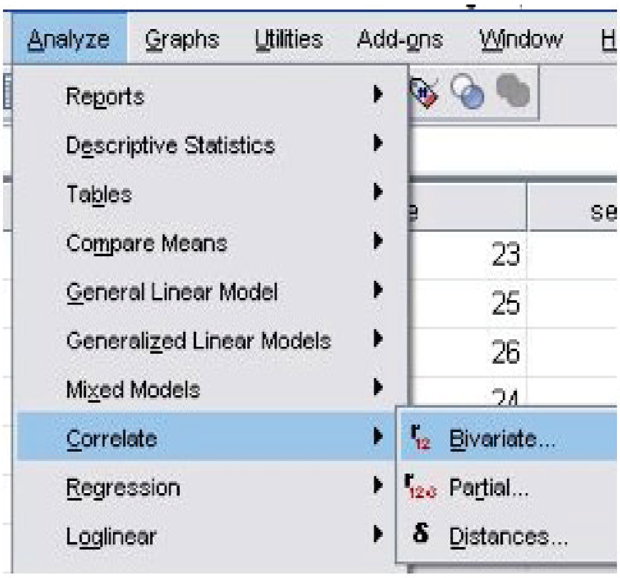
هذا وتم الحصول على النتائج التالية:

$$r^2{}_{12} = 0.741, \qquad r^2{}_{y2} = 0.931, \qquad r^2{}_{y1} = 0.909$$

بتطبيق الصيغة الرياضية أعلاه لحساب معامل الارتباط الجزئي $\rho_{y2.1}$ وعلى النحو التالي:

$$\rho_{y2.1} = \frac{0.931 - 0.909 \times 0.741}{\sqrt{(1 - 0.826)(1 - 0.549)}} = 0.92$$

كذلـك هنالـك العديـد مـن الحـزم البرمجيـة الإحصائيـة الجاهـزة التـي تمكـن مـن حسـاب معامـل الارتبـاط الجزئـي. ففـي حزمـة SPSS يتـم بعـد ادخـال البيانـات اللازمـة حسـاب معامـل الارتبـاط الجزئـي بيـن المتغيـرات مـن خـلال النافـذة التاليـة:



2.1.1.4 معامل الارتباط للظواهر غير المقاسة (معامل سبيرمان للرتب)

هناك بعض الظواهر لا يمكن قياسها عدديا، وقد تكون على شكل صفات او على رتب ومن هذه الظواهر، الحالة الصحية للأفراد والتدخين فلا يوجد مقياس كمي لقياس الحالة الصحية او عادة التدخين وكل ما نستطيع ان تقوم بة هو تصنيف الافراد من حيث الحالة الصحية الى أصناف متدرجة ابتداء من السيئة وانتهاء بالحالة الجيدة او الممتازة وهكذا ينطبق على بقية الظواهر المماثلة مثل الرتب أي تحويل القيم الرقمية الى رتب، ويقسم إلى عدة أنواع وأكثرها شيوعا هو معامل ارتباط الرتب لسبيرمان. فاذا كان لدينا متغيرين X وY، وكانت قيم كل منهما عددها n، فان معامل ارتباط سبيرمان بينهما هو

$$R = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

حيث d_i هو مقدار الفرق بين رتبة قيمة X ورتبة قيمة Y المناظرة لها.

n: هو عدد ازواج القيم لكل من X وY.

مثال:

الجدول التالي يوضح درجات مجموعة من الطلاب فى اختبار تم إجراؤه على نفس الطلاب مرتين متتاليتين والمطلوب حساب قيمة معامل ارتباط الرتب لسبيرمان بين درجات الاختبارين؟

درجة الاختبار الأول (X)	2	5	9	8	2
درجة الاختبار الثاني (Y)	4	6	7	4	3

المتغير Y يوجد به رقمان متساويان هما (4،4) وترتيبهما (2،3) إذا يأخذ كل منهم متوسط الترتيب (2+3)/2 = 5/2 = 2.5.

X	Y	رتب X	رتب Y	d_i	d_i^2
3	4	2	2.5	-0.5	0.25
5	6	3	4	-1	1
9	7	5	5	0	0
8	4	4	2.5	1.5	2.25
2	3	1	1	0	0
المجموع					3.5

وبالاعتماد على المعادلة السابقة فأم معامل ارتباط سبيرمان للرتب هو:

$$R = 1 - \frac{6(3.5)}{5(24)} = 0.825$$

2.4 الانحدار

تحليل الانحدار Regression Analysis هو تحليل يمكننا من إيجاد معادلة رياضية تربط بين متغير تابع ومتغير أو متغيرات مستقلة. فمثلا يمكننا باستخدام تحليل الانحدار دراسة العوامل التي تؤثر في زيادة الطلب على المنتج وتحديد نموذجا (معادلة) رياضيا لهذه العلاقة. هذا النموذج يجعلنا قادرين ليس فقط على فهم طبيعة العلاقة وتحديد العوامل المؤثرة فعلا بل إنه يجعلنا قادرين على توقع تأثير تغير أي متغير من هذه المتغيرات المستقلة على المتغير التابع.

ان الحاجة لاستخدام هذا الانحدار كثيرة ومتنوعة. فالمهندس يحتاج لدراسة العوامل التي تؤثر في ارتفاع درجة حرارة الغازات المستخدمة في عملية ما وقد يكون لديه العديد من العوامل التي يريد أن يعرف تأثيرها الحقيقي. باستخدام الانحدار فإن هذا المهندس يستطيع تحديد العوامل المؤثرة وإهمال تلك غير المؤثرة ويمكنه توقع التغير الذي يحدث في درجة حرارة الغازات نتيجة لتغير محدد في أي من تلك المتغيرات المؤثرة. ومدير الموارد البشرية يريد تحديد العوامل التي تؤثر على أداء العاملين الجدد من بين عدة عوامل مثل السن وتقدير التخرج وجامعة الدراسة وغيرها. فيمكنه باستخدام تحليل الانحدار معرفة ما هي العوامل التي لا تؤثر ولا ترتبط بأداء العاملين الجدد وتلك المؤثرة ويمكنه الحصول على نموذجا رياضيا يمكنه من توقع وفهم حجم تأثير تلك العوامل على الأداء.

كما ذكرنا سابقا ان معامل الارتباط هو قياس مدى العلاقة بين الظواهر. ولكن كثيرا ما نحتاج في دراسة هذه الظواهر التعرف على طبيعة العلاقة بينها، فقد تكون على صورة خط مستقيم او على صورة منحني. ويعرف خط الانحدار بانة الخط الذي يمثل العلاقة بين متغيرين او هو طريقة بيانية تمثل العلاقة بين الظواهر ويستخدم الانحدار في تقدير قيمة أحد المتغيرين إذا عرف المتغير الاخر.

التمثيل البياني للمتغيرات

أن الخطوة الأولى في دراسة العلاقة بين متغيرين هو أجراء تحليل بياني تصويري، حيث يساعد الفحص البصري للبيانات في توفير المعلومات التالية:

- التعرف على درجة التباين المشترك وهو مؤشر لدرجة الارتباط بين المتغيرين.
- التعرف على مدى وتوزيع نقاط عينة البيانات.
- التعرف فيما اذا كان هناك ظهور نقاط متطرفة.
- تعيين شكل العلاقة بين المتغيرين.
- تعيين نوع العلاقة.

يعتمد تحليل الانحدار على العلاقة بين متغيرين أو اكثر. والتحليل هنا يقوم على أساس وجود متغير تابع وآخر مستقل (متبوع). فبمجرد تحديد العلاقة الرياضية بين المتغيرين يسهل تحديد المتغير التابع بمعرفة بيانات المتغير المستقل (المتبوع).

فاذا كانت الواردات مثلا تتأثر بالدخل القومي فبتحديد هذه العلاقة كميا يمكن التنبؤ بالواردات بمجرد معرفة الدخل القومي المتوقع. ومن الناحية الرياضية إذا اعتمد المتغير التابع Y في قيمته على مقدار التغير في قيمة المتغير المستقبل X فانه يعبر عن Y بانها دالة في X وهو ما يسمى بالانحدار. أما معامل الانحدار فهو المؤشر الذي يبين لنا مدى التغير الذي يحصل في متغير ما (المتغير التابع) نتيجة تغير وحدة واحدة من المتغير الاخر(المتغير المستقل).

1.2.4 أنواع تحليل الانحدار

هناك نوعان من تحليل الانحدار أولهما هو الانحدار الخطي وهو الأكثر انتشارا. الانحدار الخطي يعني أننا ندرس العلاقة الخطية. أما النوع الثاني فهو الانحدار غير الخطي والذي نحتاجه عند دراسة علاقات على شكل منحنى وليس خطا مستقيما. الانحدار الخطي هو الأكثر شيوعا. والانحدار الخطي له نوعان بسيط ومتعدد فالبسيط يحاول التنبؤ بالعلاقة بين متغير ما وعامل واحد يؤثر فيه والمتعدد يحاول التنبؤ بالعلاقة بين متغير ما وعدة عوامل تؤثر فيه. نستعرض في هذا الدليل النوع الأول وهو الانحدار الخطي البسيط، ولإجراء عمليات التحليل باستخدام الانحدار غير الخطي او الانحدار الخطي المتعدد يمكن الرجوع الى كتب الإحصاء المتخصصة بذلك.

الانحدار الخطي البسيط

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد على إنتاجية الدونم.
- دراسة أثر الكلفة على الانتاج.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- أثر الدخل على الإنفاق الاستهلاكي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

نموذج الانحدار الخطي

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبئ منه، على المتغير الثاني ويسمى بالمتغير الحقيقي التابع أو المتنبئ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$y_i = \beta_0 + \beta_1 x_i + e_i , \quad i = 1, 2, 3, \dots , n$$

حيث ان:

x_i : هي قيم المتغير المستقل x

β_1 : هي مقدار معامل الانحدار، وهي القيمة التقديرية لمعامل الانحدار.

β_0 : هي مقدار تقاطع خط الانحدار مع المحور العمودي.

e : هو مقدار الخطأ العشوائي وهو الفرق ما بين قيم y الحقيقية والقيمة التقديرية لها.

تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار (β_1 , β_0) في النموذج باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right)}{\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x، هو الوسط الحسابي لقيم $\bar{y}$، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ، ويطلق على هذا التقدير « تقدير معادلة انحدار y على x.

معاملا التحديد R^2

ان المعيار الحقيقي لقوة علاقة الانحدار ومدى تمثيلها للنموذج التحليلي هو معامل التحديد Coefficient of Determination، هو مربع معامل الارتباط (ويرمز له بالرمز (r^2) وهى قيمة موجبة دائماً وتشير لقوة العلاقة الارتباطية بين أي ظاهرتين فعلى سبيل المثال إذا كان معامل الارتباط ($r = 0.80$) فهذا يدل على وجود علاقة انحدار موجبة وتكون قوة العلاقة ($r^2 = 0.64$) ويمكننا الحصول على النسبة المئوية للارتباط من خلال ضرب ($r^2 \times 100$) لنحصل على النسبة المئوية لقوة العلاقة الارتباطية

مثال

فيما يلي بيانات عن كمية البروتين اليومي بالجرام ومقدار الزيادة في الوزن بالكجم، وذلك لعينة من حجمها 10.

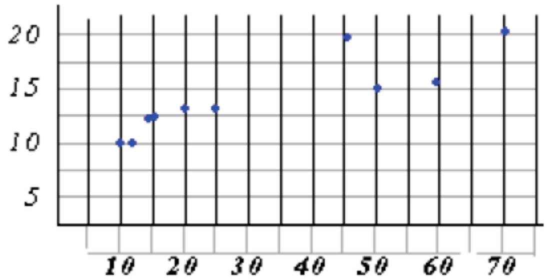
كمية البروتين	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن	10	10	12	12	13	13	19	15	16	20

المطلوب:

- رسم نقط الانتشار، وما هو توقعاتك لشكل العلاقة؟
- تقدر معادلة انحدار الوزن على كمية البروتين.
- تفسر معادلة الانحدار.
- مقدار الزيادة في الوزن عند إعطاء الشخص 50 جرام من البروتين؟ وما هو مقدار الخطأ العشوائي؟
- رسم معادلة الانحدار على نقط الانتشار في المطلوب (1).

الحـل:

رسم نقط الانتشار: مقدار الزيادة y



من المتوقع أن يكون لكمية البروتين أثر طردي (إيجابي) على مقدار الزيادة في الوزن.

تقدير معادلة الانحدار.

بفرض أن x هي كمية البروتين، y هي مقدار الزيادة في الوزن، يمكن تطبيق المعادلتين اعلاه، ومن ثم يتم حساب المجاميع التالية:

كمية البروتين x	الزيادة في الوزن y	xy	x^2
10	10	100	100
11	10	110	121
14	12	168	196
15	12	180	225
20	13	260	400
25	13	325	625
46	19	874	2116
50	15	750	2500
59	16	944	3481
70	20	1400	4900
320	140	5111	14664

المجاميع المطلوبة

$$\sum x = 320$$

$$\sum y = 140$$

$$\sum xy = 511$$

$$\sum x^2 = 14664$$

إذا الوسط الحسابي:

$$\bar{x} = \frac{\sum x}{n} = \frac{320}{10} = 32$$

$$\bar{y} = \frac{\sum y}{n} = \frac{140}{10} = 14$$

- بتطبيق المعادلة الأولى أعلاه يمكن حساب $\hat{\beta}_1$ كما يلي:

$$\hat{\beta}_1 = \frac{6310}{44240} = 32$$

- بتطبيق المعادلة الثانية يمكن حساب $\hat{\beta}_0$ كما يلي:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 14 - (0.1426)(32) = 9.4368$$

- إذا معادلة الانحدار المقدرة، هي:

$$\hat{y} = 9.44 + 0.143\,x$$

- تفسير المعادلة:
- الثابت $\hat{\beta}_0 = 9.44$: يدل على أنه في حالة عدم استخدام البروتين قي التغذية، فإن الوزن يزيد 9.44 كجم.
- معامل الانحدار $\hat{\beta}_1 = 0.143$: يدل على أنه كلما زادت كمية البروتين جرام واحد، حدث زيادة في الوزن بمقدار 0.143 كجم، أي زيادة مقدارها 143 جرام.
- مقدار الزيادة في الوزن عند استخدام 50 غرام بروتين هو:

$$\hat{y} = 9.44 + 0.143\,(50) = 16.59$$

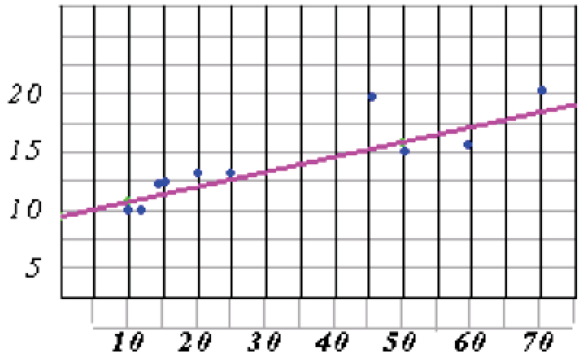
وأما مقدار الخطأ العشوائي هو:

$$\hat{e}_{x=50} = y_{x=50} - \hat{y}_{x=50} = 15 - 16.59 = -1.59$$

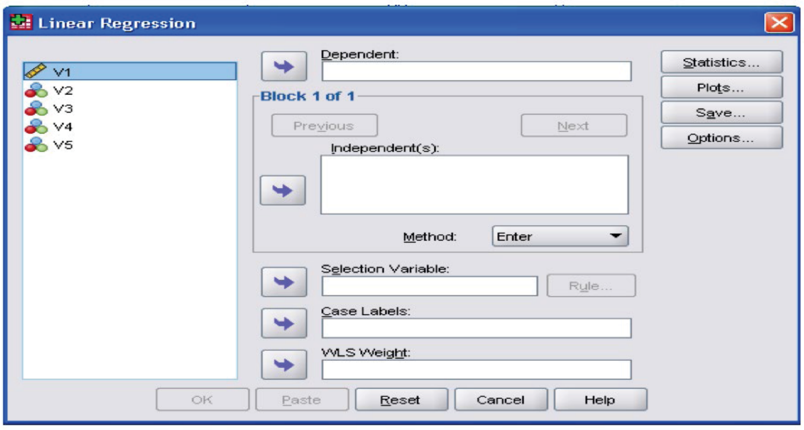
- رسم معادلة الانحدار على نقط الانتشار.

يمكن رسم معادلة خط مستقيم إذا علم نقطتين على الخط المستقيم.

إذا خط معادلة الانحدار هو:



- أما معامل التحديد فيكون في المثال السابق (100×0.83 = 83%)، ويمكن تفسير نسبة الارتباط بأن 83% من الظاهرتين يتأثر كل منهما بالأخر أو أن 83% من التغير في أحد الظاهرتين ينتج من التغير في الظاهرة الأخرى
- من جانب اخر، يمكن تطبيق احد الحزم الإحصائية مثل SPSS، لأجراء تحليل الانحدار الخطي وبكل سهولة وذلك من خلال النافذة التالية:



المراجع

- التحليل الاحصائي للبيانات (www.pitt.edu/~super1/ResearchMethods/Arabic/statistical-ar.pdf)
- مبادئ الإحصاء (https://ar.wikibooks.org/wiki/إحصاء/مبادئ_الإحصاء)
- طرائق عرض وتحليل البيانات إحصائياً (https://sites.google.com/site/drmohama/statand-control)
- دليل تصميم وتنفيذ المسوح الإحصائية، أدلة المنهجية والجودة – دليل رقم (8)
- سلامة المهدي وآخرون، دليل كتابة تقارير إحصائية سهلة الاستخدام، الجهاز المركزي للإحصاء، العراق.
- مقاييس التشتت (www.faculty.ksu.edu.sa)
- اعداد وكتابة التقارير، مركز الخبراء لعمل الأبحاث والرسائل العلمية
- أمجد الغامدي، التقارير الإدارية الأهمية والأهداف والمفهوم، جامعة الملك فيصل

مـركـز الإحـصـاء
STATISTICS CENTRE

الرؤية: ببياناتنا نمضي نحو غدٍ أفضل
Vision: Driven by data for a better tomorrow

www.scad.gov.ae
adstatistics