المحاضرة الأولى

أولاً : مفهوم علم الإحصاء ، تبويب وعرض البيانات ، أقسام علم الإحصاء ، التوزيع التكراري ، تمثيل البيانات (المدرج التكراري ، المضلع التكراري ، المنحنى التكراري).

ثانياً : مقاييس النزعة المركزية (الوسط الحسابي ، الوسط الترجيحي ، الوسط الهندسي ، الوسط التوافقي ، الوسيط ، المنوال ، العلاقة بين الأوساط الإحصائية).

ثالثاً : مقاييس التشتت :(المدى ، الانحراف المتوسط ، الانحراف المعياري، الخطأ المعياري ، معامل الاختلاف).

رابعاً : نظرية الاحتمالات : التوزيعات الاحتمالية المتصلة ، التوزيع الطبيعي ، توزيع t الاحتمالي ، توزيع x^2 ، توزيع f.

خامساً : اختبار الفرضيات الإحصائية (الاختبار الإحصائي ، اختبار متوسط مجتمع واحد ، اختبارات تتعلق بمتوسطين).

سادساً : الارتباط (معامل ارتباط بيرسون ، معامل ارتباط سبيرمان).

سابعاً : تحليل الانحدار (البسيط والمتعدد) وتحليل التباين.

ثامناً : السلاسل الزمنية.

الإحصاء :

هو علم يدرس كل الظواهر الطبيعية ويتعامل مع الأعداد ، ويعرف علم الإحصاء بأنه العلم الذي يهتم بتوفير الحقائق الرقمية للظواهر المختلفة ومن ثم ترتيبها وعرضها ثم تحليلها للوصول إلى نتائج محددة بدقة بهدف فهم الظاهرة من جهة ووضع المقترحات المختلفة لمتابعة سيرها المستقبلي من جهة أخرى.

ولعل أبسط التعريفات هو الذي يرى علم الإحصاء بأنه علم جمع وتصنيف وتبويب البيانات وتحيلها وتفسيرها.

مراحل العملية الإحصائية :

1- جمع البيانات : وهي المعلومات الأولية العددية التي يتم الحصول عليها من المصادر الحكومية أو الخاصة.

2- تنظيم البيانات :أن البيانات التي يتم الحصول عليها تنظيم عادة بجداول إحصائية أو برسوم بيانية لغرض معالجتها رياضياً ولسهولة الاطلاع عليها ومعرفة بعض الدلائل الأولية.

3- المعالجة الرياضية : تتم معالجة البيانات رياضياً وذلك لاستخراج نتائج عددية لها دلالة إحصائية مثل المتوسطات أو مقاييس التشتت أو معاملات الارتباط وغيرها.

4- التفسير أو الاستنتاج : وتعد من أهم مراحل العملية الإحصائية وبدونها تبقى النتائج مجرد أرقام صماء لا معنى لها ويتطلب التفسير قدراً كافية من الأمانة وعدم التحيز والإلمام التام بالموضوع المبحوث.

أقسام علم الإحصاء :

يمكن تصنيف علم الإحصاء إلى نوعين رئيسيين :

1- الإحصاء الوصفي : ويستخدم لغرض وصف الحقائق وتحويلها إلى أرقام وعرضها بشكل مناسب ، والوسائل المستخدمة فيه هي :

أ- العرض البياني : ويستخدم للتعبير عن البيانات الإحصائية باستخدام جداول أو خرائط أو رسوم بيانية والهدف هو توضح بعض الاتجاهات العامة عن الظاهرة محل الدراسة.

ب- الدراسة الرياضية : وتتم عن طريق حساب بعض المقاييس الإحصائية كمقاييس النزعة المركزية ومقاييس التشتت والارتباط والانحدار.

2- الإحصاء التحليلي : ويهتم بالتوصل إلى بعض المعالم عن المجتمع محل الدراسة ويستخدم في ذلك :

أ- التقدير : بالاستفادة من معالم عينة أو عينتين مثل الوسط الحسابي أو الانحراف المعياري في تقدير معالم المجتمع الذي سحبت منه العينة.

ب- اختبار الفروض : ويعني استخدام البيانات التي تجمعت في دراسة ما للوصول إلى قرار بشأن الفروق التي وضعت في بداية الدراسة ، والقرار يكون بقبول أو رفض الفرضية ، ويسمى هذا النوع ن الإحصاء بالإحصاء الاستدلالي أو الاستنتاجي.

تبويب وعرض البيانات :

تبقى البيانات التي تم جمعها بأسلوب العينة أو غيرها بدون فائدة ما لم يتم تنظيمها وجدولتها ، لذا يعتبر تبويب وعرض البيانات من الخطوات الأساسية للحصول على نتائج ذات قيمة.

وتقسم البيانات إلى قسمين هنا :

1- بيانات نوعية : وهي مجموع الحقائق والمعلومات التي تصف الظاهرة والتي ترتبط بصفة معينة من صفاتها ، ولا يمكن قياسها مباشرة بالأرقام العددية مثل الحالة الاجتماعية (غني ، متوسط الحال ، فقير) الجنس (ذكر ، أنثى) ، لذا يجب تحويلها إلى بيانات رقمية لكي تطبق عليها المقاييس الإحصائية.

2- بيانات كمية : وهي مجموعة الحقائق والمعلومات التي يعبر عنها بشكل أرقام مثل (كمية الإنتاج ، أعداد العاملين ، الطول ، الوزن) وتطبق عليها المقاييس الإحصائية مباشرة.

المجتمع الإحصائي Statistical Population :

المجتمع هنا عبارة عن جميع القيم أو المفردات التي يمكن أن يأخذها المتغير والتي ترغب بالحصول على استنتاجات حولها ، فهي تهم الباحث أو متخذ القرار ، فمثلاً إذا كانت دراستنا متعلقة بأطوال موظفي المؤسسة فأن المجتمع في هذه الحالة هو موظفو تلك المؤسسة ، ويقسم المجتمع الإحصائي إلى قسمين :

1- المجتمع المحدود : وهو المجتمع الذي يمكن حصر عدد مفرداته على سبيل المثال طول طلبة جامعة تكريت.

2- المجتمع غير المحدود : وهو المجتمع الذي من الصعب أو المستحيل حصر عدد مفرداته مثل الأسماك في النهر أو النجوم في الفضاء الخارجي.

العينة Simple :

العينة هي جزء من المجتمع وتمثل مجموعة من المشاهدات اختيرت بطريقة ما منه ، فعندما يتعذر دراسة المجتمع لأسباب فنية أو اقتصادية أو علمية تتم الاستعاضة عن دراسته بدراسة العينة ومنها نستطيع أن نستنتج خواص المجتمع الأصلي الذي أخذت منه العينة.

تبويب البيانات الإحصائية :

التوزيع التكراري Frequency Distribution :

ويعتبر من أهم الوسائل المستخدمة في تبويب وعرض البيانات الإحصائية المتعلقة بمختلف الظواهر والتي يتم الحصول عليها من مصادر مختلفة ، حيث تكون في هيئتها الأولى غير مرتبة ، ولكي تكون ذات معنى يجري ترتيبه (تصاعدياً أو تنازلياً) وتصنيعها لكل صنف صفة مميزة ويسمى كل صنف.

المدى :

عبارة عن فترات متسلسلة تضم مجموعة بيانات محصورة بين حدين هما الحد الأعلى ويمثل القيمة التي تقع في نهاية الفئة ، والحد الأدنى والذي يمثل القيمة التي تقع في بداية الفئة.

لذا فالجدول التكراري يضم عمودين أساسيين هما (الفئات والتكرارات) ويمكن استخراج منهما أعمدة إضافية تستخدم لغرض وصف البيانات الإحصائية وتحليلها.

ولغرض إعداد جدول تكراري نتبع الخطوات الآتية :

1- إيجاد مدى التغيير Range.

2- اختيار وتحديد الفئات.

3- إيجاد طول الفئة.

4- تحديد حدود الفئات (الحد الأدنى والحد الأعلى).

5- إيجاد عدد التكرارات لكل فئة.

مدى التغيير : هو الفرق بين الحد الأعلى والأدنى للبيانات (Range).

طول الفئة : الحد الأعلى – الحد الأدنى + 1.

مثال :

تم دراسة أعداد العاملين بأجر في 20 معمل للمواد الغذائية وكانت الأعداد كما يلي :

3 , 3 , 2 , 2 , 4 , 3 , 2 , 4 , 1 , 4 , 3 , 3 , 5 , 4 , 2 , 3 , 1 , 5 , 2 , 1

عدد العاملين	التكرارات f	التكرار النسبي	التكرار النسبي المئوي
1	3	0,15	0,15 x 100 = 15
2	5	0,25	25

3	6	0,3	30
4	4	0,2	20
5	2	0,1	10
	$\sum = 20$	$\sum = 1$	$\sum = 100$

التكرار النسبي :

هو نسبة تكرار كل فئة إلى مجموع التكرارات (التكرار الكلي).

مثال :

البيانات التالية تمثل الأجر الاسبوعي بالدولار مكونة من 50 عامل غير مؤهل في إحدى المؤسسات.

30	38	42	51	34	42	54	34	42	39	40	50	52
38	47	35	53	28	34	41	35	31	41	36	25	36
32	37	44	45	37	45	46	29	46	38	26	84	40
33	44	45	44	40	31	47	43	27	53	41		

أجور العاملين	التكرارات f	مركز الفئة	التكرار النسبي	التكرار النسبي المئوي
25 – 29	5	27	0,1	10
30 – 34	8	32	0,16	16
35 – 39	10	37	0,2	20
40 – 44	13	42	0,26	26
45 – 49	8	47	0,16	16
50 – 54	6	52	0,12	12
	$\sum = 50$		$\sum = 1$	$\sum = 100$

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى+الحد الأعلى}}{2}$$

طول الفئة = الحد الأعلى للفئة – الحد الادنى + 1

التكرار المتجمع الصاعد :

وفيه يتم البدء بأصغر الفئات ووضع التكرارات أمام الفئات بحيث يتضمن التكرار المقابل لكل فئة مجموع التكرارات للفئات الأقل منها.

التكرار المتجمع النازل :

وفيه يتم البدء بأصغر الفئات ووضع التكرارات أمام الفئات بحيث يتضمن التكرار المقابل لكل فئة مجموع التكرارات الأكبر منها.

مثال :

للبيانات الآتية استخرج التكرار المتجمع الصاعد والتكرار المتجمع النازل :

الفئات	التكرارات	التكرار المتجمع الصاعد	التكرار المتجمع النازل
50 – 59	3	3	30
60 – 69	5	8	27
70 – 79	10	18	22
80 – 89	8	26	12
90 – 99	4	30	4
	$\sum = 30$		

تحويل البيانات الغير مبوبة إلى البيانات المبوبة :

مثال :

حول البيانات الآتية التي تمثل درجات 25 طالب في مادة الإحصاء إلى بيانات مبوبة :

30	84	60	36	70	55	62	47	39	82	25	70	71
35	89	30	57	48	33	88	75	46	77	35	62	

الحل :

لغرض تحويل البيانات نتبع الخطوات الآتية :

1- نجد المدى : قانون المدى = أكبر قيمة – أقل قيمة + 1

$R = 89 - 25 + 1 = 65$

2- نجد عدد الفئات وحسب الصيغة الآتية : (صيغة Yale)

$N = 2.5 \times \sqrt[4]{n}$

$N = 2.5 \times \sqrt[4]{25} = 5.59 = 6$

3- نستخرج طول الفئة :

$C = \frac{R}{N} \rightarrow C = \frac{65}{6} = 10.833 = 11$

ملاحظة : دائماً طول الفئة يقرب نحو الأعلى بغض النظر عن الكسر.

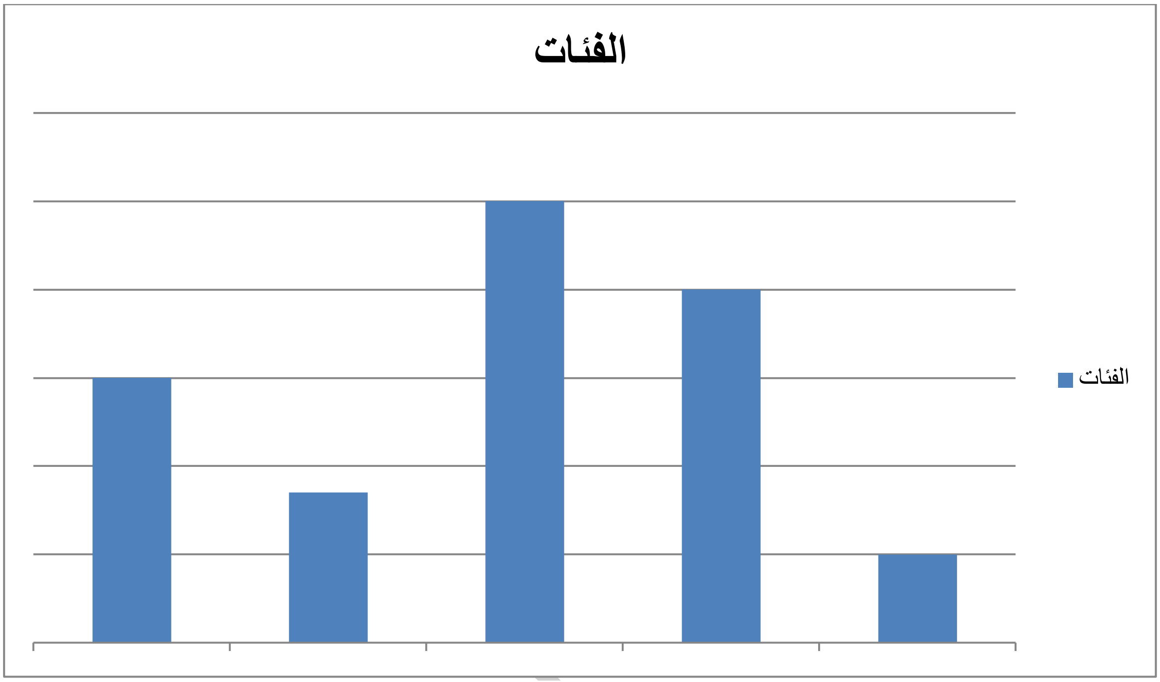
تمثيل البيانات :

ويتم بواسطة أشكال ورسوم بيانية وتدخل ضمن أدوات الإحصاء الوصفية ، وهي تكون أما بديلة عن الجدول التكراري أو استكمالاً له ، وتمتاز بالبساطة والفعالية وإعطاء فكرة سريعة عن البيانات ومنها :

1- المدرج التكراري Histogram :

وهو عبارة عن مجموعة من المستطيلات المتلاصقة قاعدة كل منها تساوي طول الفئة وارتفاعها يساوي تكرارها.

التكرارا



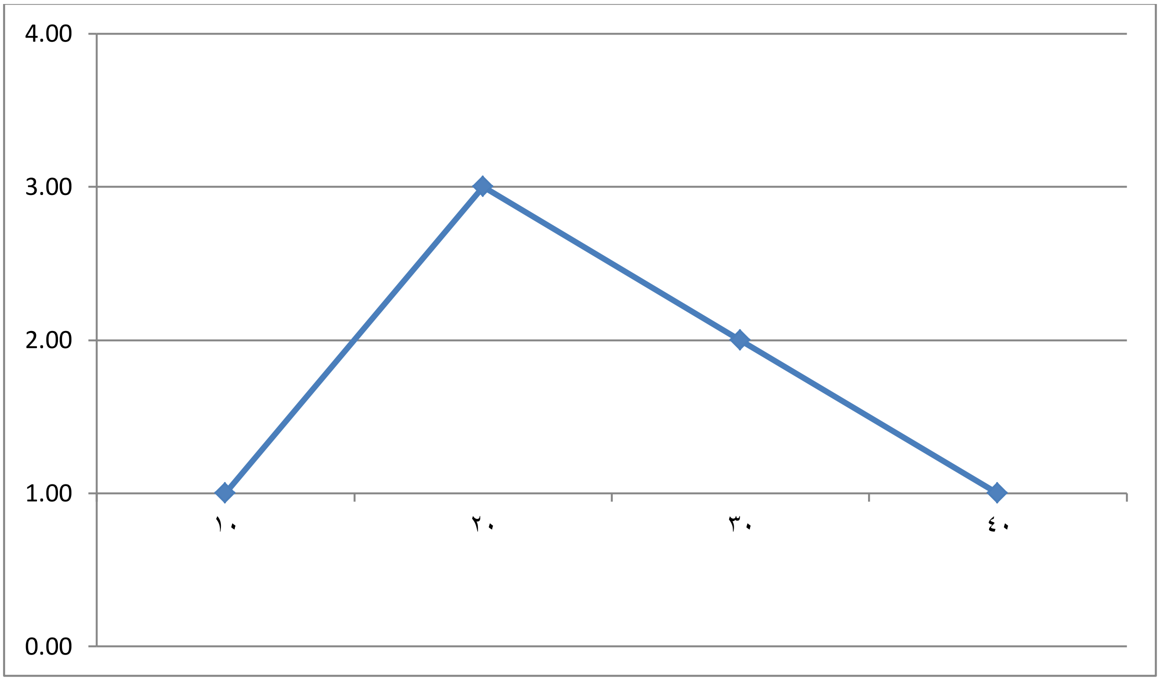
طول الفئة

الفئات

2- المضلع التكراري :

ويمثل سلسلة من المستقيمات تصل النقاط الممثلة (مراكز الفئات) وتكرارها.

التكرارا



الفئات

3- المنحى التكراري :

يرسم نفس طريقة الرسم المضلع التكراري ولكن بدلاً من أن نصل بين النقاط بمستقيمات سوف يتم إيصالها بمنحنيات متصلة ليصبح منحني واحد ويمر بجميع هذه النقاط.

4- الدائرة البيانية Pie Chart :

وتستخدم لتوضيح بيانات معينة وبشكل خاص المتغيرات النوعية مثل (أنواع القطاعات الاقتصادية ، الأقسام الإدارية ، التوزيع السكاني ... الخ) حيث تقسم إلى قطاعات بحيث تمثل مساحات القطاع البيانات الخاصة التي تعود إلى حقل من حقول البيانات.

ولما كانت مساحة القطاع تتناسب مع زاويته كانت النسبة بين مساحات القطاعات المختلفة في الدائرة الواحدة كالنسبة بين الزوايا المركزية للقطاعات ، ولذلك تقسم الزاوية المركزية الكلية للدائرة وهي (°360) إلى زوايا مركزية تتناسب مع أعداد البيانات.

مثال :

أتضح من أحد البحوث الاجتماعية أن هناك أسرة تنفق دخلها البالغ 2750 دولار سنوياً على النحو الآتي:

الغذاء 1200 ، الملابس 400 ، المسكن 850 ، متفرقة 300

المطلوب : رسم دائرة بيانية لتمثيل هذه البيانات.

الحل :

أ- باستخدام صيغة النسبة والتناسب = $\frac{\text{المشاهدة} \times 360}{\text{المجموع الكلي للمشاهدات}}$

الزاوية المركزية لقطاع الغذاء = $\frac{1200}{2750} \times 360 = 157.1$

الزاوية المركزية لقطاع الملابس = $\frac{400}{2750} \times 360 = 52.36$

الزاوية المركزية لقطاع المسكن = $\frac{850}{2750} \times 360 = 111.27$

الزاوية المركزية لقطاع المتفرقة = $\frac{300}{2750} \times 360 = 39.27$

ب- باستخدام صيغة النسبة المئوية لكل مشاهدة :

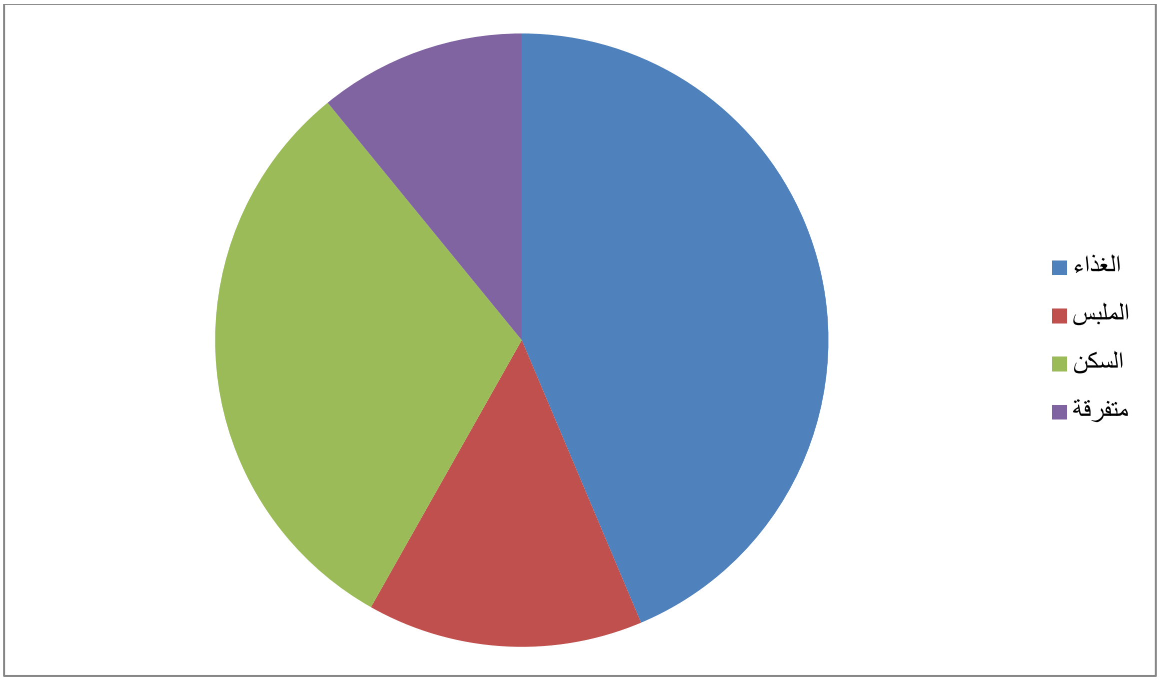
النسبة المئوية = $\frac{\text{المشاهدة}}{\text{مجموع المشاهدات}} \times 100$

النسبة المئوية لقطاع الغذاء = $\frac{1200}{2750} \times 100 = 43.54\%$

النسبة المئوية لقطاع الملابس = $\frac{400}{2750} \times 100 = 14.54\%$

النسبة المئوية لقطاع المسكن = $\frac{850}{2750} \times 100 = 30.91\%$

النسبة المئوية لقطاع المتفرقة = $\frac{300}{2750} \times 100 = 10.91\%$



مقاييس النزعة المركزية :

وجدنا في الفصل السابق أهمية تنظيم البيانات الإحصائية وتلخيصها في جداول تكرارية وتمثيلها بيانياً وذلك في توضيح شكل التوزيعات التكرارية للبيانات الإحصائية وطبيعتها بشكل عام ، ولكن الحاجة تدعو إلى إيجاد مؤشرات تلخص البيانات بأقل قدر من التفصيل أو نموذج يمثل المجموعة الإحصائية ومفرداتها أو معيار تقاس بالنسبة إلى هذه المفردات وتقارن بواسطته المجموعة كلها بالنسبة إلى المجموعات الإحصائية الأخرى.

لقد وجد أن معظم القيم بمختلف الظواهر الطبيعية تتركز عامة في الوسط أو قريباً منه ، إذ يحدث في كثير من التوزيعات أن يتراكم عدد كبير من القيم حول قيمة معينة ، ويقل هذا التراكم تدريجياً كلما أبتعد المتغير عن هذه القيمة ، وهذا التجمع أو التراكم حول قيمة ما يسمى بالنزعة المركزية للتوزيع ، ونسبة القيمة التي يحدث حولها التراكم بمقياس النزعة المركزية ومن أهم هذه المقاييس هي (الوسط الحسابي ، الوسط الهندسي ، الوسط التوافقي ، الوسيط ، المنوال) بالإضافة إلى أوساط ومقاييس أخرى.

أولاً : الوسط الحسابي ($\bar{x}$) The Arithmetic Mean :

وهو من أبسط مقاييس النزعة المركزية وأوسعها انتشاراً من ناحية الاستخدام ، ومن مميزاته :

أ- يمكن استخدامه في أغلب البيانات الإحصائية ولمختلف الظواهر.

ب- كونه باقٍ بنظر الاعتبار جميع القيم الإحصائية أو البيانية.

ج- لا يحتاج لإيجاده إلى تنظيم البيانات.

د- دائماً يكون مجموع انحرافات القيم عن وسطها الحسابي يساوي (صفر).

إلا أن من عيوبه أنه يتأثر بالقيم الشاذة لأنه يأخذ جميع البيانات.

ويمكن إيجاد الوسط الحسابي بالصيغ التالية :

1- بالنسبة للبيانات غير المبوبة :

$$x = \frac{\sum xi}{n}$$

حيث أن :

xi = القيم المتغيرة.

n = عدد القيم.

مثال :

من البيانات الآتية استخرج الوسط الحسابي :

10 , 15 , 13 , 12 , 20 , 17 , 18 , 10

الحل :

$$x = \frac{\sum xi}{n} = \frac{120}{8} = 15$$

مثال :

استخرج قيمة Z التي تجعل للقيم التالية وسطاً حسابياً مقداره (25) :

30 , 17 , 20 , 37 , Z

الحل :

n x = 25

$$x = \frac{\sum xi}{n} \rightarrow 25 = \frac{\sum xi}{n} \rightarrow \sum xi = 125$$

$$\therefore Z = 125 - 104 = 21$$

مثال :

جد الوسط الحسابي للبيانات الآتية :

2 , 3 , 5 , 2 , 4 , 2 , 52

الحل :

$$x = \frac{\sum xi}{n} = \frac{70}{7} = 10$$

من المثال أعلاه نلاحظ :

أ- أن قيمة الوسط الحسابي تساوي (10) مع أن معظم القيم كانت أقل من (5) وهذا ما يفسر لنا تأثير الوسط الحسابي في القيم الشاذة ، أي أنه تأثر بالقيمة الأخيرة (52).

ب- يمكن التأكد من صحة الحل عن طريق طرح القيم من الوسط الحسابي ، فإذا كان مجموعها يساوي صفر فهذا يعني أن الحل صحيح.

2- الوسط الحسابي بالنسبة للبيانات المبوبة :

ويمكن إيجاده عن طريق الصيغة الآتية :

$$x = \frac{\sum fi\, xi}{\sum fi}$$

حيث أن :

xi = مراكز الفئات

fi = التكرارات

مثال :

للجدول التكراري الآتي استخرج الوسط الحسابي :

الفئات	التكرارات fi	مركز الفئة xi	fi xi
0 – 4	6	2	12
5 – 9	10	7	70
10 – 14	15	12	180
15 – 19	30	17	510
20 – 24	20	22	440
25 – 29	8	27	216
30 – 34	2	32	64
	91		$\sum 1492$

$$X = \frac{\sum fi\, xi}{\sum fi} = \frac{1492}{91} = 16.395$$

ثانياً : الوسط الترجيحي (X_w) Weighted Average :

لاحظنا عند حساب الوسط الحسابي أننا تعاملنا مع كل قيم المتغير معاملة واحدة ولكن في بعض الأحيان تختلف قيم المتغير وفقاً لنسبتها في العينة أو المجتمع ، ولهذا يتم استخدام الوسط الترجيحي كوسط معبر عن هذه البيانات ، ويحسب من خلال الصيغة الآتية :

$$X_w = \frac{\sum ni\, xi}{\sum ni}$$

حيث أن :

ni = وزن المتغير

xi = قيم المتغير

مثال :

جد المعدل الطالب (س) :

ni xi	xi الدرجة	ni عدد الوحدات	المادة
210	70	3	رياضيات
250	50	5	كيمياء
240	60	4	إحصاء
160	80	2	فيزياء
90	90	1	ديمقراطية
$\sum 950$		$\sum 15$	

$X_w = \frac{\sum ni\ xi}{\sum ni} = \frac{950}{15}$ 63.33

ثالثاً : الوسط الهندسي Geometric Mean G :

1- للبيانات غير المبوبة :

ويمكن إيجاده بالصيغة الآتية :

$$\text{Log G} = \frac{\log x1 + \log x2 + \log x3 \ldots \log xn}{n}$$

مثال :

جد الوسط الهندسي للبيانات الآتية :

12 – 45 – 50

الحل :

$$\text{Log G} = \frac{\log 12 + \log 45 + \log 50}{3}$$

$$\text{Log G} = \frac{4.43}{3} \rightarrow \text{Log G} = 1.477$$

G = 30

2- الوسط الهندسي للبيانات المبوبة :

ويحسب من الصيغة الآتية :

$$\text{Log G} = \frac{\sum fi\ Log\ xi}{\sum fi}$$

مثال :

استخدم التوزيع التكراري الذي يتضمن توزيع عدد من الطلبة في جامعة ما حسب أوزانهم بالكيلو غرام لإيجاد الوسط الهندسي :

الحل :

fi Log xi	Log xi	مركز الفئة xi	أعداد الطلبة fi	الفئات
8.891	1.778	61	5	60 – 62
32.511	1.806	64	18	63 – 65
76.696	1.826	67	42	66 – 68
49.817	1.845	70	27	69 – 71
14.906	1.863	73	8	72 – 74
$\sum = 182.822$			$\sum = 100$	

$$\text{Log G} = \text{Log G} = \frac{\sum fi\ Log\ xi}{\sum fi}$$

$$\text{Log G} = \frac{182.822}{100} = 1.8282$$

$$\text{G} = 67.3$$

ملاحظة :

1- يتضح مما سبق أن الوسط الهندسي لقيم مختلفة موجبة دائماً أصغر من الوسط الحسابي.

2- يكثر استخدام الوسط الهندسي في الأرقام القياسية للأسعار أو إيجاد متوسط لعدد من النسب أو في معدل المتغيرات في المبيعات أو السكان ... الخ.

3- لا يمكن إيجاد الوسط الهندسي إلا إذا كانت مجموعة القيم موجبة ، وأن تأثره بالقيم المتطرفة يكون أقل منه في الوسط الحسابي.

رابعاً : الوسط التوافقي Harmonic Mean H :

1- للبيانات غير المبوبة :

الوسط التوافقي هو مقلوب الوسط الحسابي :

$$\text{X} = \frac{\sum xi}{n} \qquad \text{H} = \frac{n}{\sum \frac{1}{xi}}$$

ملاحظة : مقلوب القيم يعني واحد تقسيم القيمة = $\frac{1}{xi}$

مثال :

استخرج الوسط التوافقي للقيم التالية :

5 , 17 , 26 , 14 , 9 , 18 , 22

الحل :

xi	$\frac{1}{xi}$
5	$\frac{1}{5} = 0.2$
17	0.058
26	0.038
14	0.071
9	0.111
18	0.055
22	0.075
	$\sum = 0.578$

$\sum = 0.578$

$H = \frac{n}{\sum \frac{1}{xi}} \rightarrow H = \frac{7}{0.578} = 12.11$

2- الوسط التوافقي للبيانات المبوبة :

ويأخذ الصيغة التالية :

$H = \frac{\sum fi}{\sum fi\left(\frac{1}{xi}\right)} \rightarrow H = \frac{\sum fi}{\sum \frac{fi}{xi}}$

المثال السابق (أوزان الطلبة) :

fi / xi	مركز الفئة xi	أعداد الطلبة fi	الفئات
0.082	61	5	60 – 62
0.281	64	18	63 – 65
0.626	67	112	66 – 68
0.385	70	27	69 – 71
0.109	73	8	72 - 74
$\sum = 1.483$		100	

$H = \frac{100}{1.483} = 67.43$

يستخدم الوسط التوافقي في إيجاد المتوسطات للمعدلات الزمنية مثل إيجاد متوسط القراءة لمجموعة من الأفراد بدلالة عدد الكلمات في الدقيقة.

الوسيط Median Me :

الوسيط لمجموعة من القيم هو القيمة التي يقع ترتيبها وسط المجموعة عند ترتيب هذه القيم تصاعدياً أو تنازلياً ، أي بعبارة أخرى هي القيمة التي تقسم المجموعة (مجموعة البيانات) إلى قسمين متساويين إذا كان عدد القيم فردي ، أما إذا كان عدد القيم زوجي فسيكون هناك قيمتان تتوسطان القيم ويكون الوسيط هنا هو متوسط تلك القيم بعد ترتيب القيم تصاعدياً أو تنازلياً ويمكن حسابه كما يلي :

أ- للبيانات غير المبوبة :

مثال :

استخرج الوسيط للبيانات الآتية :

30 , 57 , 26 , 40 , 72 , 85 , 37

الحل :

- نرتب القيم تصاعدياً أو تنازلياً.

- نلاحظ عدد القيم (n) فإذا كان عدد القيم فردي فأن الوسيط يقابل القيمة التي نحصل عليها من القانون :

$$\left(\frac{n+1}{2}\right) \rightarrow \frac{7+1}{2} = 4$$

26
30
37
40
57
82
85

عدد القيم هنا فردي = 7

= Me

∴ الوسيط هو القيم ذات الترتيب الرابع = 40

أما إذا كان عدد القيم زوجياً فأن الوسيط يقابل القيمة $\frac{n}{2}$ والتي تليها ، حـ .جميعها وتقسيمها على 2.

مثال :

جد الوسيط للبيانات الآتية :

30 , 57 , 26 , 40 , 72 , 85 , 75

الحل :

- نرتب القيم تصاعدياً أو تنازلياً :

26
30
37
40
57

$$Me = \frac{40+57}{2} = 48.5$$

= Me = 48.5

عدد القيم هنا زوجي = 8

82
85

∴ الوسيط هو القيم ذات الترتيب الرابع = 40

ب- الوسيط للبيانات المبوبة :

يتم حساب الوسيط في البيانات المبوبة عن طريق الخطوات الآتية :

$$Me = Li + \left[\frac{(\sum fi)/2 - fi}{fi}\right] x\ C$$

حيث أن :

Li : الحد الأدنى للفئة الوسيطة.

$\sum fi$: مجموع التكرارات.

Fi : التكرار المتجمع الصاعد للفئة قبل الوسيطية.

Fi : التكرار الحقيقي للفئة الوسيطية.

C : طول الفئة.

لإيجاد الوسيط يتم اتباع الخطوات الآتية :

1- عمل جدول تكرار متجمع صاعد.

2- إيجاد التكرار الوسيط.

3- تحديد الفئة الوسيطية وهي الفئة المقابلة للتكرار المتجمع الصاعد الذي يلي ترتيب التكرار الوسيط مباشرة.

مثال :

للجدول التكراري الآتي استخرج الوسيط :

الفئات	التكرار	Fi الصاعد
0 – 4	6	6
5 – 9	17	23
10 – 14	23	46
15 – 19	30 التكرار الحقيقي	76
20 – 24	21	97
25 – 29	14	111
30 – 34	10	121
35 – 39	3	124
	$\sum = 124$	

الحل :

1- نعمل تكرار متجمع صاعد.

2- نستخرج التكرار الوسيط وذلك عن طريق :

$$\frac{\sum fi}{2} = \frac{124}{2} = 62$$

3- نلاحظ وقوع التكرار الوسيط على المتجمع الصاعد ، فإذا كان يقع بين تكرارين نأخذ التكرار الذي يليه مباشرة ، أما إذا كان التكرار الوسيط موجود فأنها نأخذه نفسه.

4- نحدد الفئة الوسيطية : وهي الفئة المقابلة للـ 76 على المتجمع الصاعد.

∴ :

Li = الحد الأدنى للفئة الوسيطة = 15

Fi = القيمة على المتجمع الصاعد التي تسبق التكرار الوسيط = 6

fi = التكرار الحقيقي للفئة الوسيطية = 30

C = طول الفئة الوسيطية = 5

5- نطبق القانون $Me = Li + \left[\frac{(\sum fi)/2 - fi}{fi}\right] \text{x } C$

$$Me = 15 + + \left[\frac{62-46}{30}\right] \text{x } 5$$

$$Me = 17.666$$

لغرض التأكد من دقة الناتج يجب أن تكون قيمة الوسيط محصورة أو قريبة بين الحد الأدنى والحد الأعلى للفئة الوسيطية.

المنوال Mode Mo :

وهي القيمة التي تتكرر أكثر من غيرها من المفردات ويسمى أيضاً (الصفة الأكثر شيوعاً) لذا فأنه يفضل حينما يكون المطلوب معرفة القيمة الأكثر شيوعاً كما في حالات التسوق والتسويق ، حيث يكون من المطلوب معرفة بيانات عن السلع والمواصفات الخاصة للشائع منها ويمكن حسابه كما يلي :

أ- للبيانات غير المبوبة :

هي القيمة التي تتكرر أكثر من غيرها.

مثال :

أحسب المنوال للمفردات الآتية :

6 , 3 , 4 , 9 , 7 , 8 , 4

∴ قيمة المنوال هي القيمة رقم 4

ب- بالنسبة للبيانات المبوبة هناك طريقتان :

1- طريقة الفروق :

لإيجاد المنوال نستعين بالمثال التالي :

مثال :

للجدول التكراري التالي استخرج المنوال :

التكرار	الفئات
8	5 – 9
7	10 – 14
24	15 – 19
32	20 – 24
20	25 – 29
12	30 – 34
8	25 – 39
2	40 – 44

الحل :

1- نستخرج الفئة المنوالية : وهي الفئة التي تقابل أكبر تكرار.

حيث أن :

$\Delta 1$: تكرار الفئة المنوالية ناقص التكرار السابق :

$\Delta 1 = 32 - 24 = 8$

2- نستخرج $\Delta 1$ و $\Delta 2$:

حيث أن :

$\Delta 1$: تكرار الفئة المنوالية ناقص التكرار السابق :

$\Delta 2 = 32 - 24 = 8$

$\Delta 2$: تكرار الفئة المنوالية ناقص التكرار اللاحق :

$\Delta 2 = 32 - 20 = 12$

3- نطبق قانون المنوال :

$$Mo = Li + \left(\frac{\Delta 1}{\Delta 1 + \Delta 2}\right) x\ C$$

حيث أن :

Li : الحد الأدنى للفئة المنوالية.

C : طول الفئة المنوالية.

$$Mo = 20 + \left(\frac{8}{8+12}\right) x\ 5$$

Mo = 22

2- طريقة العزوم :

$$Mo = Li + \left(\frac{\text{التكرار اللاحق}}{\text{التكرار السابق+اللاحق}}\right) x\ C$$

Mo =

العلاقة بين الأوساط الإحصائية (الوسط الحسابي والوسيط والمنوال) :

يمكن معرفة حساب أي من الأوساط الإحصائية في حالة معرفة الأثنين الآخرين عن طريق العلاقة الآتية :

X – Mo = 3 (X – Me)

مثال :

جد المنوال إذا علمت أن الوسيط = 150 والوسط الحسابي يساوي 120 :

الحل :

120 – Mo = 3 (120 – 150)

120 – Mo = -90

Mo = 120 + 90 = 210

وللتأكد من صحة الحل نعود في المعادلة التي تمثل العلاقة بينهم :

120 – 210 = 3 (120 – 150

-90 = -90

مثال :

جد الوسط الحسابي إذا علمت أن المنوال = 280 والوسيط = 300 :

X - 280 = 3 (X – 300)

X – 280 = 3X – 900

3X – X = 900 – 280

$2X = 620 \rightarrow X = \frac{620}{2} = 310$

نتحقق من صحة الحل :

$310 - 280 = 3\ (310 - 300)$

$30 = 30$

ملاحظة : تستخدم هذه العلاقة في إيجاد الوسط الحسابي للبيانات المفتوحة :

5 – 9
10 – 14
15 – 19
20 – 24
25 - □

بيانات مفتوحة

مقاييس التشتت

لا تعتبر مقاييس التمركز كافة لوصف مجموعة من البيانات وصفاً كاملاً ، فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلاف توزيع بياناتها حول مركزها (درجة تجانس البيانات) ، مثل المجموعتين الآتيتين من البيانات :

(95 , 97 , 100 , 103 , 105) ، (50 , 75 , 100 , 125 , 150)

حيث يلاحظ أن لهما نفس الوسط الحسابي والوسيط وهي (100) ، في حين المجموعتان مختلفتان من ناحية تشتتهما حول المركز أو الوسط ، وأن الدرجة التي تتجه بها البيانات الرقمية للانتشار حول قيمة وسطها تسمى تشتت أو توزيع البيانات ، ومن أهم مقاييس التشتت المدى ، الانحراف المتوسط ، التباين ، والانحراف المعياري.

أولاً : المدى Range :

المدى : هو الفرق بين أكبر قيمة وأصغر قيمة.

مثال :

أحسب المدى لكل من المجموعتين الآتيتين من البيانات :

Y = 12 , 6 , 7 , 3 , 15 , 10 , 18 , 5

A = 9 , 12 , 12 , 12 , 22 , 22 , 22 , 24

$R_Y = 18 - 3 = 15$

$R_A = 24 - 9 = 15$

ويلاحظ أن مداهما متساويان مما يوحي أن المدى أحياناً يكون مظللاً كونه يعتمد على القيمتين الطرفيتين واللتين كثير ما تكون شاذتين ، لذلك ظهرت الحاجة إلى مقاييس أكثر وضوحاً للتشتت.

ثانياً : الانحراف المتوسط The Mean Deviation MD :

وهو متوسط الانحرافات المطلقة عن الوسط الحسابي ويتم حسابه وفق الآتي :

1- البيانات غير المبوبة :

$$MD = \frac{\sum|xi - x|}{n}$$

مثال :

جد الانحراف المتوسط للقيم الآتية :

Xi = 9 , 8 , 6 , 5 , 7

xi	xi - x	$\|xi-x\|$
9	2	2
8	1	1
6	-1	1
5	-2	2
7	0	0
$\sum = 35$	0	6

MD = $\frac{6}{5}$ = 1.2

2- البيانات المبوبة :

MD = $\frac{\sum fi\,|xi-x|}{\sum fi}$

مثال :

جد الانحراف المتوسط لجدول التوزيع التكراري الآتي :

$\|xi-x\|$	xi - x	fi xi	xi مركز الفئة	fi التكرار	الفئات
6,45	-6,45	305	61	5	60 – 62
3,45	-3.45	1152	64	18	63 – 65
0,45	-0.45	2814	67	42	66 – 68
2,55	2,55	1880	70	27	69 – 71
5,55	5.55	580	73	8	72 – 74
		6745		100	

X = $\frac{\sum xi\, fi}{\sum fi}$ = 67,45

MD = $\frac{\sum fi\,|xi-x|}{\sum fi}$

MD = $\frac{226.5}{100}$ = 2,265

ثالثاً : التباين Variance S^2 :

نلاحظ أن مجموع انحراف عناصر (مفردات العينة) عن وسطها الحسابي $\sum(xi-x)$ يساوي صفر ، ذلك بسبب كون قسم من الانحرافات يكون موجباً بينما الآخر يكون سالباً ، وللتغلب على هذه المشكلة فقد تم معالجتها بأخذ القيم المطلقة للانحرافات في الانحراف المتوسط ، ويمكن معالجتها بطريقة أخرى إلا وهي تربيع قيم الانحرافات للحصول على مجموع مربعات الانحرافات ، ويمكن حساب التباين كما يلي :

1- للبيانات غير المبوبة :

$$S^2 = \frac{\sum(xi-x)\dot{2}}{n}$$

مثال :

جد التباين للقيم الآتية :

Xi = 9 , 8 , 6 , 5 , 7

الحل :

xi	xi - x	$(xi – x)^2$
9	2	4
8	1	1
6	-1	1
5	-2	4
7	0	0
$\sum = 35$	0	10

$$S^2 = \frac{10}{5} = 2$$

2- للبيانات المبوبة :

$$S^2 = \frac{\sum fi(xi-x)}{\sum fi}$$

مثال :

جد التباين لجدول التوزيع التكراري الآتي :

$(xi – x)^2$	xi - x	fi xi	مركز الفئة xi	التكرار fi	الفئات
41,6	-6,45	305	61	5	60 – 62
11,9	-3,45	1152	64	18	63 – 65
0,2	-0,45	2814	67	42	66 – 68
6,5	2,55	1880	70	27	69 – 71
30,8	5,55	580	73	8	72 – 74
		6745		100	

الحل :

$fi \, \lvert xi - x \rvert^2$
208
214,2
8,5
175,5
246,4
$\sum$ 852,6

$$X = \frac{6745}{100} = 67,45$$

$$\therefore S^2 = \frac{852,6}{100}$$

$$S^2 = 8,526$$

رابعاً : الانحراف المعياري Standard Deviation S :

عند حساب التباين قمنا بتربيع الانحرافات ، حيث تكون قيمة التباين مقاسة بمربع الوحدات المستخدمة في قياس المشاهدات ، ولا ضير في ذلك إلا أن المشكلة تظهر عندما يكون مربع الوحدات غير ذي معنى أو غير مقبول ، فعند استخدام وحدات قياس مثل عدد العمال أو الكيلو غرام أو الدينار فأن التباين يكون (عامل مربع أو كيلو غرام أو دينار مربع) وهذه كلها غير ذات معنى ، وحلاً لذلك يتم إرجاع وحدات القياس إلى أصلها وذلك بأخذ الجذر التربيعي للتباين وهو ما يسمى بالانحراف المعياري أو (القياسي) إذ يكون مقاساً بالوحدات الأصلية.

1- الانحراف المعياري للبيانات غير المبوبة :

$$S = \sqrt{\frac{\sum(xi-x)2}{n}}$$

مثال :

جد الانحراف المعياري للقيم الآتية :

xi	xi - x	$(xi - x)^2$
9	2	4
8	1	1
6	-1	1
5	-2	4
7	0	0
$\sum 35$	0	10

X = 7

$$S = \sqrt{\frac{\sum(xi-x)2}{n}} = \sqrt{\frac{10}{5}} = 1{,}414$$

$$\therefore = \sqrt{S2} = 35$$

2- الانحراف المعياري للبيانات المبوبة :

ويأخذ الصيغة الآتية :

$$S = \sqrt{\frac{\sum fi(xi-x)2}{\sum fi}}$$

مثال :

جد الانحراف المعياري لجدول التوزيع التكراري الآتي :

جد التباين لجدول التوزيع التكراري الآتي :

الفئات	التكرار fi	تكرار الفئة xi	fi xi	xi – x	$(xi – x)^2$
5 – 9	2	7	14	-13,66	186,6
10 – 14	4	12	48	-8,66	75
15 – 19	8	17	136	-3,66	13,4
20 – 24	6	22	132	1,34	1,79
25 – 29	6	27	162	6,34	40,19
30 – 34	4	32	128	11,34	128,59
	30		620		

$X = \frac{620}{30} = 20,66$

$\therefore S = \sqrt{\frac{1546,64}{30}}$

$S = \sqrt{51,55} = 7,17$

$fi\ \lvert xi - x \rvert^2$
373,2
300
107,2
10,74
241,14
514,36
1546,64

خامساً : الخطأ المعياري Standard Error SE :

وهو عادة ما يصاحب الوسط الحسابي لإعطاء صورة واضحة ودقيقة عن البيانات وبحسب الطريقة الآتية :

$SE = \frac{S}{\sqrt{n}}$

مثال :

للبيانات الآتية استخرج الخطأ المعياري :

xi	xi - x	$(xi – x)^2$
1	-2	4
3	0	0

5	2	4
4	1	1
2	-1	1
15	0	10

$X = \frac{15}{3} = 3$

$S^2 = \frac{(xi - x)\ 2)}{2} = \frac{10}{5} = 2$

$S = \sqrt{S2} \rightarrow \sqrt{2} = 1.414$

$SE = \frac{S}{\sqrt{n}} = \frac{1.414}{\sqrt{5}} = \frac{1.414}{2.236} = 0.63$

ويجب أن يكتب الوسط الحسابي بالشكل الآتي : $Mean = 3 \mp 0,63$

أي أن الوسط الحسابي الحقيقي هو $Mean = 2,37 \sim 3.63$

ملاحظة : العينات ذات الخطأ المعياري الأقل أفضل من العينات ذات الخطأ المعياري الأعلى.

سادساً : معامل الاختلاف C.V Coefficient of Variation :

ويستخدم هذا المؤشر والذي هو عبارة عن نسبة الانحراف المعياري على الوسط الحسابي للمقارنة بين عينتين إذا كانت وحدة القياس بينهما مختلفة (كالوزن والطول) أما إذا كانت وحدات القياس مشتركة فالخطأ المعياري يكفي للمقارنة ، ويحسب وفق الصيغة الآتية :

$C.V. = \frac{S}{x}$

مثال :

من المثال السابق جد معامل الاختلاف C.V :

$C.V = \frac{S}{x} = \frac{1.414}{3} = 0.47$

اختبار الفرضيات :

عندما يقوم الباحث بإجراء بحثه فأنه يختار عينة محدودة العدد طبقاً لإمكانياته ، لأنه لا يستطيع أن يطبق البحث على المجتمع الأصلي بأكمله ، لكنه عندما يستخرج نتيجته فأنه يكون في دالة شك هل النتيجة راجعة إلى مجرد صدفة أم إلى ظاهرة حقيقية في المجتمع الأصلي، ويقتضي هذا تكرار البحث عدة مرات واختيار عينات مختلفة من المجتمع الأصلي للتأكد من أن النتائج التي حصل عليها لا تختلف باختلاف العينات التي يجري عليها البحث ، أن تكرار التجربة يحتاج إلى وقت وجهد ونفقات لذا فأن مقاييس الدلالة الإحصائية ومن خلال اختبار الفرضيات على الباحث هذا التكرار فهي تبين إلى أي حد يستطيع أن يتأكد من ثبات نتائجه وإلى أي حد يستطيع إرجاعها إلى عامل الصدفة.

صياغة الفرضيات الإحصائية :

الفرضية الإحصائية عبارة عن أدعاء (قد يكون صائباً أو خطأ) حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

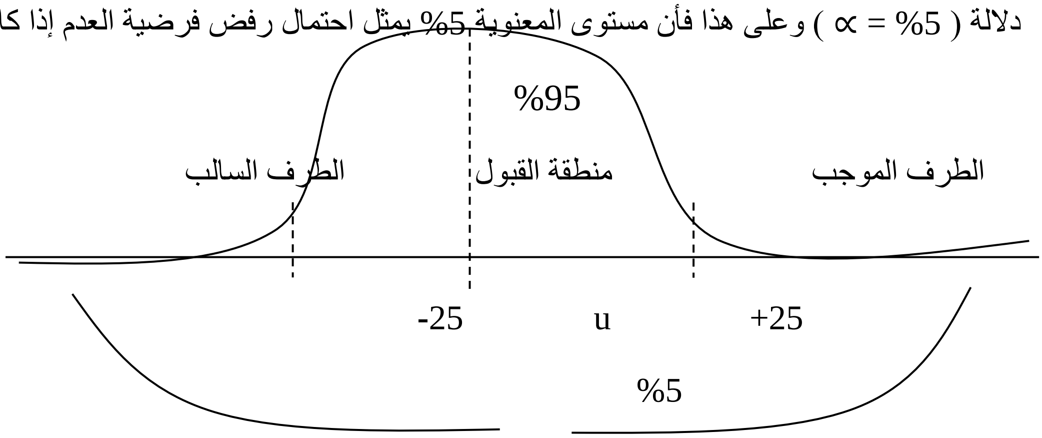
وعامة يصاغ الفرض الإحصائي بشكل عدم وجود اختلاف أو عدم وجود علاقة وتسمى فرضية العدم ويرمز لها بالرمز H_0 وهذه الفرضية تنطلق منها وتتمسك بها ولا نرفضها إلا إذا توفرت دلائل قوية من العينة تقود إلى رفضها ، وإلى جانب فرضية العدم توجد الفرضية البديلة ويرمز لها بالرمز H_1 وهذه الفرضية يجب أن تكون صحيحة في حالة عدم صحة H_0.

يجري الاختبار وتكون نتيجته أما رفض H_0 أو قبولها ، فإذا كان القرار قبول H_0 يعني ذلك أن الاختلاف ناتج عن الصدفة وليس حقيقياً.

تحديد مستوى المعنوية :

كلمة معنوية (Significance) أو مؤكد تعني بأن الفروق بين القيم النظرية للمجتمع والقيم الناتجة من العينة الحقيقية وكبيرة بحيث لا يمكن أن نعزى إلى الصدفة ، وعادة يختار الباحث مستوى المعنوية عند وضع تصميم التجربة منذ البداية ، ويتوقف اختيار مستوى المعنوية على طبيعة البحث نفسه وهناك مستوين معنويان شائعان في أغلب العلوم التطبيقية ، حيث نختار $\propto$ مساوية 5% أو 1% على الأكثر.

ويقصد بمستوى معنوية 5% أن المساحة تحت المنحنى = 1 وأن 95% من القيم يقع داخل مجال قدره m + 2S وهذا يعني أن المفردة التي يكون أغراضها عن المتوسط أكثر من انحرافين معياريين هي من مفردات المجتمع باحتمال 5% وأنها ليست من مفردات المجتمع باحتمال قدره 95% ، فإذا رفضنا اعتبار هذه المفردة ضمن مفردات المنطقة تكون خطأ باحتمال قدره 5% وعلى صواب باحتمال قدره 95% وتسمى منطقة الرفض بمستوى دلالة ($\propto$ = %5) وعلى هذا فأن مستوى المعنوية 5% يمثل احتمال رفض فرضية العدم إذا كانت صحيحة.



أولاً : اختيار متوسط مجتمع واحد :

يتم هنا مقارنة متوسط عينة بمتوسط المجتمع لبيان هل أن العينة تنتمي لهذا المجتمع أم لا ؟ فإذا كانت نتيجة الاختبار ايجابية يعني أن متوسط العينة المحسوب لا يختلف جوهرياً عن متوسط المجتمع.

1- الاختبار في حالة معلومية تباين المجتمع S^2 :

يستند الاختبار إلى فرضية تقول أن الوسط الحسابي للمجتمع يساوي قيمة معينة $H_0 : m = m_0$

حيث أن :

m = الوسط الحسابي للمجتمع.

m_0 = هو قيمة معينة معلومة.

S^2 = تباين لمجتمع معلوم.

ويمكن تحديد منطقة الرفض والقبول اعتماداً على الفرضية البديلة والتي تأخذ الصيغ الآتية :

$H_1 = m \neq m_0$ أي أن الوسط الحسابي للمجتمع لا يساوي القيمة المعينة m_0 وفي هذه الحالة تكون منطقة الرفض على جانبي المنحنى.

$H_1 = m > m_0$ أي أن منطقة الرفض ستقع على يمين منحنى توزيع x.

$H_1 = m < m_0$ أي أن منطقة الرفض ستقع على يسار منحنى توزيع x.

فرضيات التحليل :

- العينة ينبغي أن تكون عشوائية.

- توزيع العينة ينبغي أن يكون طبيعياً أو قريباً منه.

- تباين المجتمع معلوم S^2.

وتمثل الإحصائية Z المنطقة الحرجة من خلال تحويل القيم العشوائية xi إلى قيم طبيعية معيارية zi باستخدام الصيغة الآتية :

$$Z = \frac{x-m}{s/\sqrt{n}} = r\ z = \frac{x-m}{6/\sqrt{n}}$$

حيث أن 6^2 تمثل تباين المجتمع.

مثال :

أظهرت نتائج الإحصاء السكاني لبلد ما إن متوسط حجم الأسرة في ذلك البلد يبلغ 6,5 فرد ، وأن الانحراف المعياري لحجم الأسرة فيه يبغ 3,5 أخذت عينة من 100 أسرة في العاصمة فوجد أن متوسط حجم الأسرة يبلغ 7,1 فرد.

المطلوب : اختبر الادعاء بأن متوسط حجم الأسرة في العاصمة هو أعلى من المتوسط العام لذلك البلد عند مستوى معنوية 5% إذا علمت أن قيمة z الجدولية = 1,645.

الحل :

1- صياغة الفرضية الإحصائية :

أن فرضية العدم هنا :

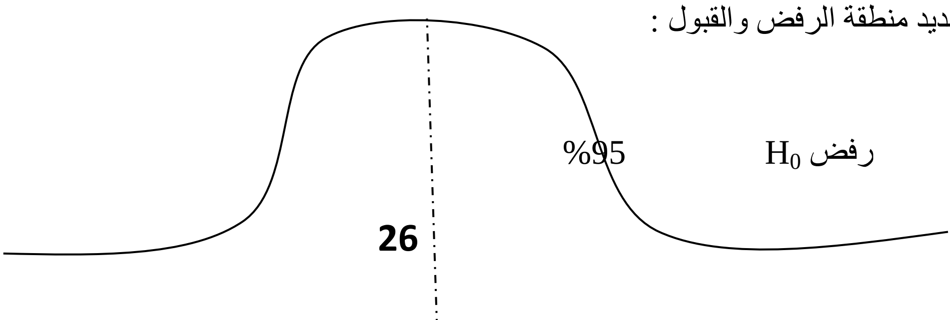
$H_0 = m = 6,5$

$H_1 = m > 6,5$

2- تطبيق الإحصائية :

$$Z = \frac{x-m}{6/\sqrt{n}} \qquad z = \frac{7,1-6,5}{3,5/\sqrt{100}} = \frac{0,6}{0,35} = 1,71$$

3- تحديد منطقة الرفض والقبول :



قبول H_0

الجدولية $Z = 1,645$

4- اتخاذ القرار الإحصائي :

وحيث أن قيمة z المحتسبة والبالغة 1,71 تقع في منطقة رفض H_0 وهي أكبر من قيمة z الجدولية والبالغة 1,645 ، إذاً نرفض فرضية العدم (فرضية تساوي المتوسطين) ونقبل الفرضية البديلة ، أي نستنتج وبثقة 95% أن متوسط حجم الأسرة في العاصمة هو أكبر من متوسط البلد المذكور.

مثال :

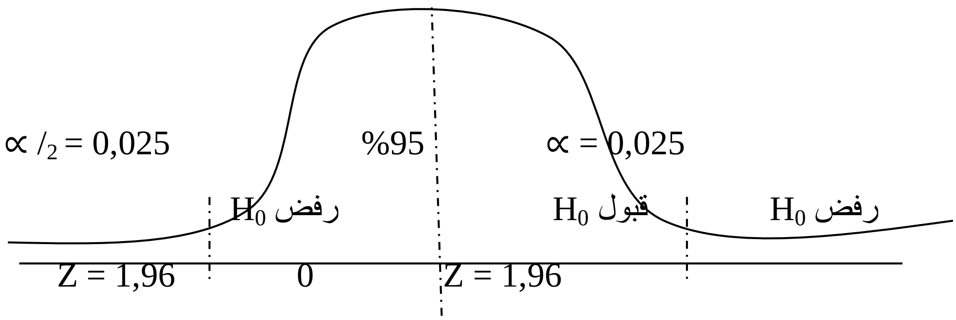
أخذت عينة مكونة من 16 شاباً بالغاً وكان متوسط الطول لهم 156 سم وكان الانحراف المعياري للمجتمع معلوماً 5 سم اختير الفرض القائل أن متوسط المجتمع الذي أخذت منه هذه العينة هو 160 سم ، وذاك عند مستوى معنوية $\propto = 0,05$.

الحل :

$H_0 = m = 160$ cm $\qquad\qquad$ $H = M \neq 160$ cm

$$Z = \frac{x-m}{6\,/\sqrt{n}} = \frac{156-160}{5\,/\sqrt{16}} = 1,6$$

وتحدد منطقة قبول ورفض H_0 بالفترة (1,96 . 1,96-) وباستخدام جدول Z للقيم الحرجة من طرفين عند $\propto = 0,05$ وحيث أن قيمة Z المحتسبة تساوي 1,6- تقع داخل منطقة قبول H_0 لذا تقبل فرضية العدم ، حيث أن متوسط طول الأفراد البالغين في هذا المجتمع هو 160 سم وبدرجة ثقة 95%.



2- الاختبار في حالة مجهولية التباين :

إذا كان تباين المجتمع مجهولاً 6^2 حيث يكون أحياناً من الصعب معرفة ذلك التباين فأنه يكون لدينا حالتين هما :

أ- أن الإحصائية التي تستخدم لاختبار فرضية العدم هي $t_0 = \frac{x-m}{s\,/\sqrt{n}}$ والتي تتبع توزيع t بدرجة حرية 7 وسيتم توضيح ذلك لاحقاً.

ب- استخدام تباين العينة S^2 ليحل محل تباين المجتمع 6^2 على أن يكون حجم العينة المحسوبة كبيراً($n \geq 30$) إذ يكن استخدام العينة الكبيرة S^2 عوضاً عن تباين المجتمع غير المعلوم لأن S^2 هو مقدار جيد 6^2 ولأنه لا يتغير كثيراً من عينة لأخرى مادام حجم العينة كبيراً ($n \geq 30$) وعندئذ فأن الإحصائية $Z = \frac{x-m}{s\,/\sqrt{n}}$ لازالت تقريباً تتوزع توزيعاً طبيعياً قياسياً (Z).

مثال :

في عينة شملت (150) مراجعاً لإحدى المستشفيات ، تم الاستفسار منهم عدد المراجعات للمستشفى خلال السنوات الثلاثة الماضية أتضح بأن المتوسط $x = 4{,}8$ وبانحراف معياري $S = 3{,}1$ ، فهل هذه النتيجة تتطابق مع أدعاء إدارة المستشفى بأن مراجعة كل شخص لا تقل عن 5 مرات ($m \geq 5$) خلال نفس الفترة وعند مستوى معنوية مقداره $\propto = 0{,}01$.

الحل :

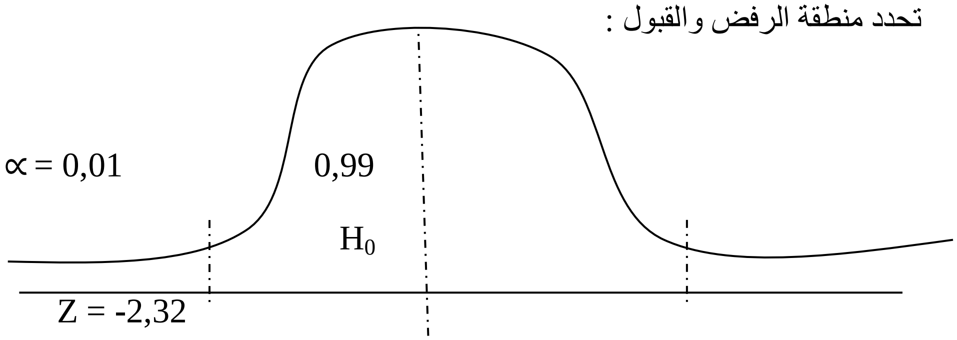
$H_0 = m \geq 5$

$H_1 = m < 5$

وحيث أن الفرضية البديلة تشير إلى قبولها فقط في حالة $m < 5$ فأن الاختبار هو من طرف واحد وأن قيمة Z الجدولية عند مستوى معنوية 0,01 هو 2,32.

$$Z = \frac{x-m}{s/\sqrt{n}} = \frac{4{,}8-5}{3{,}1/\sqrt{150}} = -\,0{,}79$$

تحدد منطقة الرفض والقبول :



وحيث أن قيمة Z المحتسبة تقع في منطقة قبول H_0 ، لذا تقبل فرضية العدم ونستنتج أن أدعاء إدارة المستشفى كان صحيحاً.

ثانياً : اختبارات تتعلق بمتوسطين :

عندما يراد المقارنة بين متوسطي مجموعتين فأن العمل سينصب على اختبار إذا كان الفرق بين المتوسطين يعزى إلى الصدفة أم أن هذا الفرق جوهرياً.

فإذا كان هناك مجتمعين وسطهما الحسابي m_1 و m_2 وتباينهما $6_1^{\,2}$ و $6_2^{\,2}$ يتم اختيار الفرضية القائلة أن الفرق بين المتوسطين يساوي قيمة معينة ، أي :

$H_0 = m_1 - m_2 = d_0$

حيث أن d_0 هي القيمة المعينة المعلومة ، أما الفرضية البديلة H_1 فستكون واحدة من الفرضيات البديلة الآتية :

$H_1 : m_1 - m_2 \neq d_0$

$H_1 : m_1 - m_2 > d_0$

$H_1 : m_1 - m_2 < d_0$

وعندما تكون $d = 0$ فان معنى فرضية العدم H_0 أن المتوسطين متساويين ويعتمد هذا الاختبار على نظرية توزيع المعاينة للمتغير العشوائي (الإحصائية) $(x_1 - x_2)$ فأن كان $(x_1 - x_2)$ هو الفرق بين متوسط العينتين المستقلتين ذات الحجم n_1 و n_2 كبيرين وإذا علم تباين المجتمع $6_1{}^2$ و $6_2{}^2$ فأن :

$$Z = \frac{(x1-x2)}{\sqrt{\frac{S2}{n1}+\frac{s2}{n2}}}$$

مثال :

أجريت دراسة للمقارنة بين متوسطي أعمار سكان المدن والقرى وكان أعمار سكان المدن يتبع توزيع طبيعية وبانحراف معياري $6_1 = 7$ سنوات وأعمار سكان القرى يتبع توزيع طبيعي بانحراف معياري $6_2 = 9$ وقد أخذت عينة من سكان المدن $n_1 = 20$ فقد كان متوسط أعمارهم 63 سنة وكذلك عينة من سكان القرى $n_2 = 25$ شخصاً كان متوسط أعمارهم 60 سنة.

اختير الفرض القائل بعدم وجود اختلاف معنوي بين متوسطين أعمار المجتمعين عند مستوى معنوية 0,01 ، علماً أن قيمة Z الجدولية = 2,58.

الحل :

1- صياغة الفرض الإحصائي :

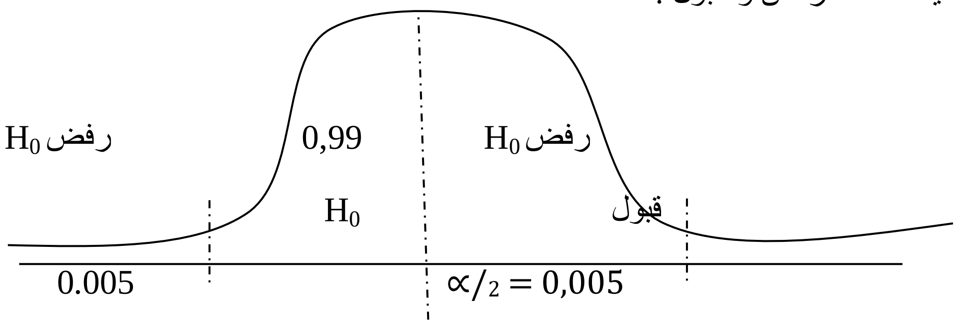
$H_0 : m_1 - m_2 \neq d_0$

$H_1 : m_1 - m_2 \neq d_0$

2- اختيار الإحصائية ثم حسابها :

$$Z = \frac{(x1-x2)}{\sqrt{\frac{6\,2}{n1}+\frac{6\,2}{n2}}} = = \frac{(63-60)}{\sqrt{\frac{(7)2}{20}+\frac{(9)2}{25}}} = 1,2$$

3- تحديد منطقة الرفض والقبول :



4- أتخاذ القرار الإحصائي :

بما أن قيمة $Z = 1,2$ المحتسبة واقعة في منطقة القبول ، لذلك لا يمكن رفض فرضية العدم H_0 القائلة بأن أعمار سكان المدن لا يختلف عن متوسط أعمار سكان القرى وبمستوى ثقة 0,99.

تحليل الارتباط Correlation Analysis :

يهتم تحليل الارتباط بقوة العلاقة بين متغيرين أو أكثر ، مثل العلاقة بين مهارة العاملين والإنتاجية أو بين سعر السلعة والكمية المطلوبة ، وتقاس العلاقة بمؤشر إحصائي يدعى معامل الارتباط ويرمز له بالرمز (r) وهناك عدة أنواع من معاملات الارتباط منها البسيط والجزئي ومعامل ارتباط الرتب ومعامل ارتباط فاي ... الخ.

خصائص معامل الارتباط :

1- تتراوح قيمة معاملة الارتباط بين سالب واحد والواحد الصحيح :

$$-1 \leq r \leq 1$$

إذ يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (∓ 1).

2- تكون قيمته تساوي صفراً عندما يكون المتغيران مستقلان عن بعضهما تماماً ، ويكون مساوٍ للواحد الصحيح عندما يكون الارتباط تاماً.

3- تكون قيمته موجبة عندما يكون الارتباط بين المتغيرين طردياً ويكون قوياً عندما يكون معامل الارتباط قريباً من الواحد الصحيح وضعيفاً عندما يكون قريباً من الصفر.

4- تكون قيمته سالبة عندما يكون الارتباط بين المتغيرين عكسياً ويكون قوياً عندما يكون المقدار السالب قريباً من (1-) ، وضعيفاً عندما يكون المقدار السالب قريباً من الصفر.

أولاً : معامل ارتباط بيرسون :

ويرمز له بالرمز rp وهو معامل ارتباط خطي بسيط يقيس قوة واتجاه العلاقة بين متغيرين فقط ، وهذان المتغيران هما متغيران كميان أي يعبر عنهما بالأرقام ، ويحسب المعامل وفق القانون الآتي :

$$rp = \frac{n\sum xY - \sum x \sum Y}{\sqrt{[n\sum x2] - [n\sum Y - (\sum Y)2]}}$$

مثال :

جد قوة واتجاه العلاقة بين سعر السلعة x والكمية المطلوبة Y باستخدام مقابل ارتباط بيرسون :

X	Y	XY	X^2	Y^2
1	6	6	1	36
2	6	12	4	36
3	4	12	9	16
4	2	8	16	4
5	1	5	25	1

$\sum = 15$	19	43	55	93

$$rp = \frac{5\,(\,43\,) - 15\,(\,19\,)}{\sqrt{[5\,(55) - (15)2][5\,(93) - (19\,)2]}}$$

$$rp = \frac{215 - 285}{\sqrt{[275 - 225][465 - 361]}} = \frac{-70}{\sqrt{5200}}$$

$$rp = \frac{-70}{72} = -\ 0{,}97$$

من قيمة معامل الارتباط يتضح ما يلي :

1- هناك ارتباط قوي بين سعر السلعة والكمية المطلوبة كون أن قيمة معامل الارتباط يساوي 0,97- وهي قريبة جداً من 1-.

2- العلاقة بين السعر والكمية المطلوبة هي علاقة عكسية بدليل الإشارة السالبة لمعامل الارتباط.

3- هذه العلاقة تتفق مع المنطق (كلما ارتفع سعر السلعة انخفض الطلب عليها).

مثال :

إذا كانت ساعات الدراسة في مادة الإحصاء لخمسة طلاب متمثلة ب x وكانت درجات هؤلاء الطلبة متمثلة بـ x ، فهل هناك علاقة بين ساعات الدراسة والدرجات التي حصل عليها الطلبة من خلال معامل ارتباط بيرسون.

X	Y	XY	X^2	Y^2
1	1	1	1	1
2	1	2	4	1
3	2	6	9	4
4	6	24	16	36
5	10	50	25	100
$\sum = 15$	20	83	55	142

$$rp = \frac{5\,(\,83\,) - 15\,(\,20\,)}{\sqrt{[5\,(55) - (15)2][5\,(42) - (20\,)2]}}$$

$$rp = \frac{415 - 300}{\sqrt{[275 - 225][710 - 400]}}$$

$$rp = \frac{115}{124} = -\ 0{,}93$$

من قيمة معامل الارتباط يتضح ما يلي :

1- هناك ارتباط قوي بين عدد ساعات الدراسة ودرجات الطلبة.

2- العلاقة طردية بدلالة الإشارة الموجبة (كلما زادت ساعات الدراسة زادت الدرجات).

3- تتفق مع المنطق.

ملاحظات :

1- قوة معامل الارتباط تعتبر مؤشر مهم على وجود علاقة بين المتغيرين إلا أنه لا يمكن الاعتماد عليها على قوة معامل الارتباط فقط ، وإنما يجب أن يكون معنوياً وذلك من خلال مقارنة r المحتسبة مع r الجدولية في جدول القيم الحرية ، فإذا كانت r المحتسبة أكبر من r الجدولية كان الارتباط معنوي والعكس صحيح.

2- تختلف درجة قوة معامل الارتباط باختلاف العلوم ، فعلى سبيل المثال في العلوم الطبية لا يمكن اعتماد معامل ارتباط أقل من 90% ، أما في العلوم الزراعية فيمكن اعتماد 60% إلا أنه بصورة عامة إذا زادت قيمة r عن 70% يعتبر هناك ارتباط قوي بين المتغيرين.

ثانياً : معامل ارتباط الرتب (سبيرمان) :

وهو معامل ارتباط ثنائي يصلح في المتغيرات الكمية والنوعية وهو أقل من معامل ارتباط بيرسون ، ويمكن إيجاد معامل ارتباط سبيرمان حسب الخطوات الآتية :

1- إعطاء رموز رقمية للبيانات النوعية (الرتب) لكل من X و Y.

2- نستخرج الفرق بين رتب X ورتب Y بعمود جديد هو d.

3- نقوم بتربيع الفرق بعمود آخر هو d^2.

4- تطبيق قانون سبيرمان لإيجاد معامل الارتباط.

$$rs = 1 - \frac{6\sum d2}{n(n2-1)}$$

مثال :

كانت تقديرات مجموعة من الطلبة في مادتي التحليلات X والإحصاء Y كما يأتي :

∴ المطلوب : إيجاد معامل الارتباط بينهما والتعليق عليه.

تحليلات X	Y	رتب X	رتب Y	d	d^2
امتياز	جيد جداً	1	2	-1	1
جيد جداً	امتياز	2	1	1	1
جيد	متوسط	3,5	4,5	-1	1
جيد	جيد	3,5	3	0,5	0,25
متوسط	متوسط	5	4	0,5	0,25
					3,5

$$rs = 1 - \frac{6(3,5)}{5(25-1)} \rightarrow 1 - \frac{21}{5(24)}$$

$$rs = 1 - \frac{21}{120} \rightarrow 1 - 0,175 \rightarrow rs = 0,82$$

التعليق :

1- ان قيمة rs هي 0,82 وهذا يعني أن هناك علاقة ارتباط قوية بين تحصيل الطالب في التحليلات وتحصيله في الإحصاء.

2- أن هذه العلاقة هي علاقة طردية أي أن الطالب المتفوق في التحليلات ويكون متفوق في الإحصاء.

3- تتفق هذه العلاقة من النظريات التربوية.

مثال :

تهتم إحدى المؤسسات بوضع طريقة لاختبار المتقدمين لمراكز وظيفية بدلاً من طريقة المقابلة ، وللمساعدة في البحث عن قرار للاختيار فأن ثمانية من المتقدمين سوف يتعرضون إلى الطريقتين وكانت النتائج كما يلي :

d^2	d	رتب المقابلة	رتب الامتحان	التقدير في المقابلة	التقدير في الامتحان	المتقدم
36	-6	7	1	G	A	1
1	-1	3	2	C	B	2
4	-2	5	3	E	C	3
12,25	3,5	1	4,5	A	D	4
6,25	-2,5	2	4,5	B	D	5
4	-2	8	6	H	F	6
9	3	4	7	D	G	7
4	2	6	8	F	H	8

$$rs = 1 - \frac{6(76,5)}{8(64-1)} \rightarrow 1 - \frac{465}{504} = 76,5$$

$$rs = 1 - 0,9 \rightarrow rs = 0,1$$

وهذا يعني أن العلاقة بين درجة المقابلة ودرجة الامتحان ضعيفة وأن العلاقة طردية بدليل الإشارة الموجبة ، إلا أنها لا يمكن الاعتماد عليها كون أن قيمة معامل الارتباط أقل من 70% ، وهذا يعني أن المتقدم المتفوق في الامتحان ليس بالضرورة يتجاوز اختبار المقابلة.

تحليل الانحدار Regression Analysis :

عرفنا أن الارتباط هو علاقة بين متغيرين يمكن قياسها رقمياً بمعامل عددي يحدد قوة واتجاه العلاقة ، أما الانحدار هو النموذج الرياضي الذي يصف العلاقة بين هذين المتغيرين ، وهو أيضاً الطريقة البيانية التي تصور العلاقة بين المتغيرات ويستخدم الانحدار في تفسير أحد المتغيرين أذا عرف الآخر.

إذن فالانحدار هو محاولة تقدير الأثر الكمي للمتغير النسبي على المتغير النتيجة ، وهذا التقدير يوضح مقدار التغير في المتغير التابع إذا تغير المستقل بمقدار وحدة واحدة أو بنسبة 100%.

والانحدار أما أن يكون بسيط أي بين متغير مستقل والآخر تابع أو أن يكون متعدداً أي بين عدة متغيرات مستقلة ومتغير تابع ، هذا من ناحية ومن ناحية أخرى أما أن يكون خطياً أي هناك علاقة خطية مستقلة بين المتغير المستقل والتابع أو أن يكون لا خطياً مثل الدالة اللوغارتمية ، الأسية ، اللوجستية ... الخ).

وأن أغلب العلاقات بين المتغيرات في جميع العلوم ومنها الهندسة هي علاقات لا خطية ومن العبث الاكتفاء بتقدير العلاقة الخطية واعتبارها ممثلاً عن واقع العلاقة بين المتغيرات ، ومع ذلك يكون موضوع الانحدار الخطي هو الموضوع الأساس لفهم بقية أنواع الانحدار.

الانحدار الخطي البسيط Simple Linear regression :

ويعني انحدار Y على X أي تقدير الأثر الكمي لـ X على Y ويمكن استخدام عدة طرق في تقدير هذا الأثر غير أن أفضل هذه الطرق هي طريقة الانحرافات عن الوسط الحسابي أو ما يسمى (طريقة المربعات الصغرى الاعتيادية) ، وتأخذ معادلة الانحدار الصيغة التالية :

$Y = \propto + BX$

مثال :

من البيانات التالية قدر معادلة انحدار Y على X :

المستقل X	التابع Y	x = (x – x)	y= (Y – Y)	xy	x^2
1	2	-2	-5	10	4
2	6	-1	-1	1	1
3	8	0	1	0	0
4	10	1	3	3	1
5	9	2	2	4	4
15	35	0	0	18	10

$X = \frac{15}{5} = 3$ $Y = \frac{35}{5} = 7$

$Y = \propto + BX$

$B = \frac{\sum xy}{\sum x2} = \frac{18}{10} = 1{,}8$

$\propto = Y – BX \rightarrow \propto = 7 – 1{,}8 (3) = 1{,}6$

$Y = 1{,}6 + 1{,}8 X$

ماذا تعني هذه المعادلة ؟

يبدو من المعادلة المقدمة أنها تحتوي على حدين الاول هو ∝ والتي تساوي 1,6 وتدعى معلمة المقطع وقد تكون أحياناً قيمة رياضية لا معنى لها ، غير أنها في الغالب عبارة عن قيمة Y إذا تم تصفيرها X.

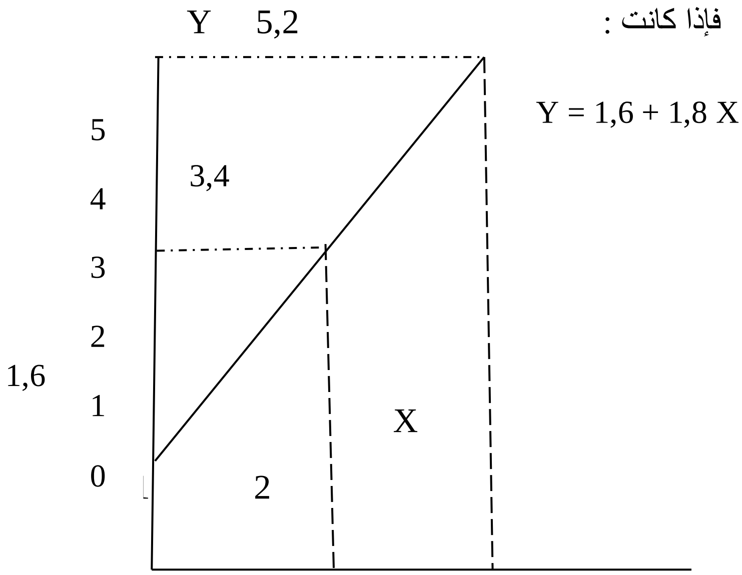
أما B والتي تساوي 1,8 فهي مقدار تأثير X على Y أي أنه إذا زاد X بمقدار وحدة واحدة فأن Y سوف يزداد بمقدار 1,8 وبيانياً فأنها ميل خط الانحدار.

فإذا كانت :

$Y = 1{,}6 + 1{,}8 (0) = 1{,}6$

$Y = 1{,}6 + 1{,}8 (1) = 3{,}4$

$Y = 1{,}6 + 1{,}8 (2) = 5{,}2$



مثال :

جمعت بيانات عن الأسعار والكميات المطلوبة من سلعة ما لمدة خمسة أسابيع ، المطلوب تقدير معادلة الطلب على تلك السلعة مع شرح المعادلة المقدرة ومن ثم رسمها بيانياً.

السعر x	الكمية y	x	y	Xy	X^2	ei	ei^2
5	1	2	-4	-4	4		
4	3	1	-2	-2	1		
3	6	0	1	1	0		
2	7	-1	2	2	1		
1	8	-2	3	3	4		
15	25	0	0	-18	10		

X = 3 Y = 5

$B = \frac{\sum xy}{\sum x2} = \frac{-18}{10} = -1,8$

$\propto = Y\ BX \rightarrow \propto = 5 - (-1,8)(3)$

$\propto = 5 + 5,4 = 10,4$

$Y = 10,4 - 1,8\ X$

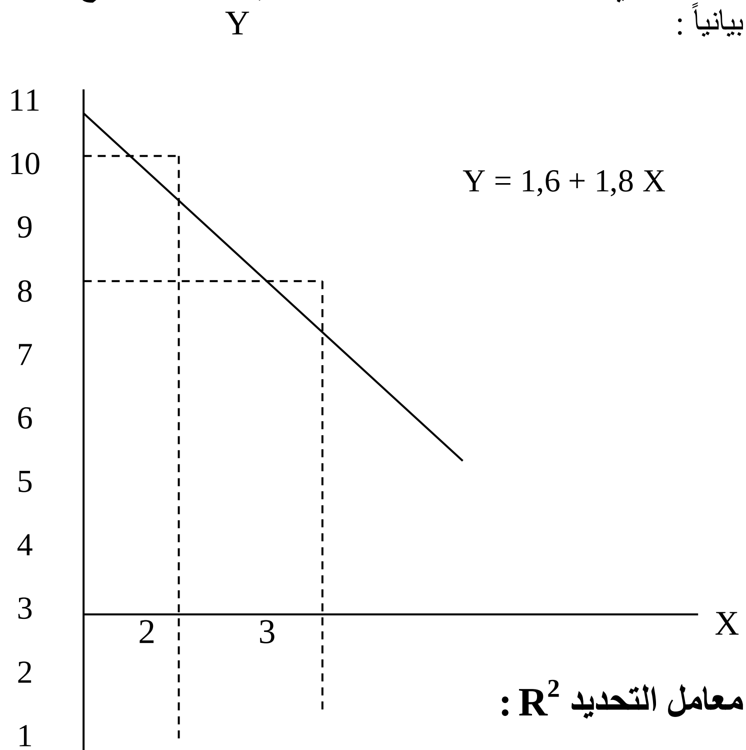
ويبدو من المعادلة المقدرة أن قيمة ∝ تساوي (10,4) وهي تمثل الطلب المستقل عن السعر ، أي أنه لو كانت السلعة مجانية (صفر) لكان أقصى طلب عليها هو (10,4).

أما قيمة B والتي تساوي (1,8-) فهي تمثل العلاقة العكسية بين السعر والكمية المطلوبة بدليل الإشارة السالبة ، أي أنه كلما ازداد السعر بمقدار 1,8 وهذا يتفق مع المنطق الاقتصادي ، ويمكن تمثيل معادلة الانحدار بيانياً :

$Y = 10,4 - 1,8 (0) = 10,4$

$Y = 10,4 - 1,8 (1) = 8,6$

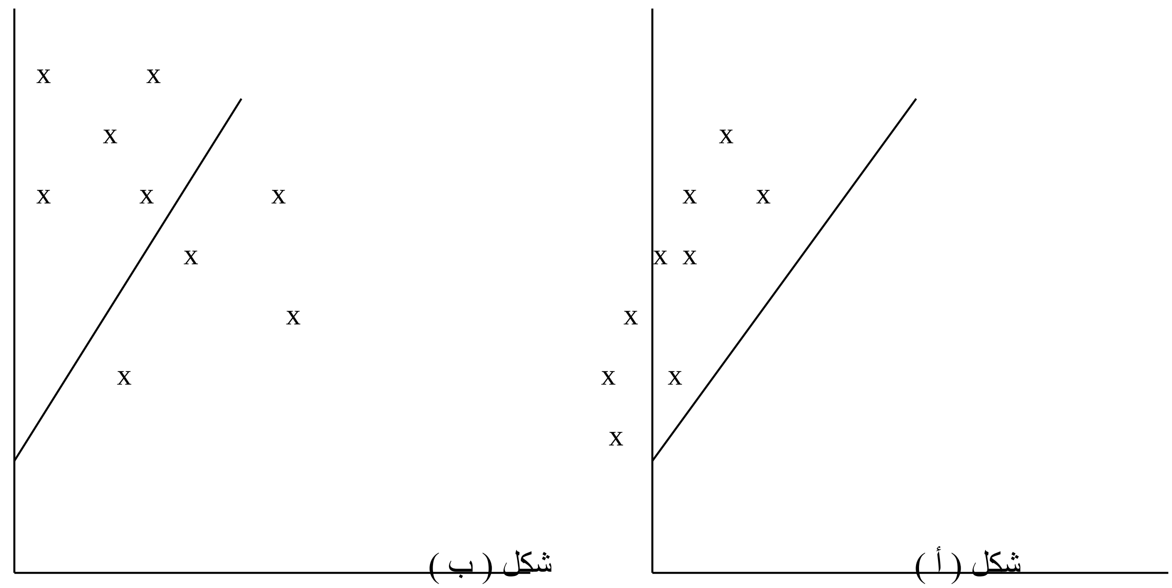
$Y = 10,4 - 1,8 (2) = 6,8$



معامل التحديد R^2 :

بعد تقدير المعلمات وتحديد خط الانحدار بطريقة المر صغرى الاعتيادية ورسم خط الانحدار لابد من قياس جودة هذا الخط من خلال التعرف على جودة التوفيق ـلـ لمشاهدات العينة ولكل من X و Y وبشكل

عام كلما اقتربت المشاهدات الحقيقية من هذا الخط أو وقعت عليه فأن كان التوفيق أفضل وكما مبين فش الشكلين أدناه :



يلاحظ بأن جودة التوفيق للشكل (أ) أفضل منه في الشكل (ب) بمعنى آخر تفسير الانحرافات في المتغير التابع y والناجمة عن تأثير المتغير المستقل x سيكون أفضل في الشكل (أ) منه في الشكل (ب) ، إذ أن الجزء الأكبر من الانحرافات التي تحدث في المتغير التابع y يعود سببها إلى وجود تأثير المتغير المستقل x والموضحة من خلال معادلة الانحدار التقديرية ، أما الجزء المتبقي فيعود إلى تأثير عوامل عشوائية أخرى غير داخلة في المعادلة.

لذلك يطلق على معامل التحديد R^2 بالقوة التفسيرية للمعادلة (النموذج) أي نسبة ما تستطيع المعادلة المقدرة أن تفسره من تغيرات في المتغير التابع الناجمة عن تغيرات في المتغير المستقل.

وتحسب قيمة R^2 وفق القيمة الآتية :

$$R^2 = \frac{B \sum xy}{\sum y^2}$$

وتكون قيمة R^2 بين الصفر والواحد الصحيح ، فكلما اقتربت قيمة R^2 من (1) كانت المعادلة أفضل وأنها تستطيع أن تفسر النسبة الأكبر من التغيرات في المتغير التابع y بإعادتها إلى تغيرات في المتغير المستقل x.

مثال :

البيانات التالية تمثل ساعات الدراسة اليومية لمجموعة من الطلاب ودرجاتهم في مادة الإحصاء ، المطلوب إيجاد القوة التفسيرية للمعادلة المقدرة والتعليق عليه :

ساعات الدراسة X	درجات الطلبة Y	x	y	xy	x^2	y^2
0	1	-2	-4	8	4	16
1	3	-1	-2	2	1	4
1	3	-1	-2	2	1	4
2	5	0	0	0	0	0
3	7	1	2	2	1	4
3	6	1	1	1	1	1
4	10	2	5	10	4	25
14	35	0	0	25	12	54

X = 2 Y = 5

$B = \frac{B\sum xy}{\sum x^2} = \frac{25}{12} = 2{,}08$

$\propto = Y - BX = 5 - 2{,}08\ (2) = 0{,}84$

$Y = \propto + BX \rightarrow Y = 0{,}84 + 2{,}08\ X$

$R^2 = \frac{B\sum xy}{\sum y^2} = \frac{2{,}08\ (25)}{54} = \frac{52}{54} = 0{,}96$

التعليق :

وهذا يعني أن المعادلة المقدرة تستطيع أن تفسر 96% من التغيرات التي تحدث في درجات الطلبة بإعادتها إلى التغيرات في عدد ساعات الدراسة ، أما 4% المتبقية فهي تعود إلى عوامل عشوائية أخرى غير داخلة في المعادلة (كأن تكون الحالة النفسية أو ذكاء الطلبة أو أي متغيرات أخرى.

مثال :

أخذت عينة من 5 نباتات متشابهة وتم قياس قطر ساق النبتة وكمية السماد الكيمياوي المستخدم لكل نبتة وكانت النتائج كما يلي :

كمية السماد X	قطر الساق Y	x	y	xy	x^2	y^2	ei	ei^2
1	3	-2	-3	6	4	9	-0,6	0,36
2	5	-1	-1	1	1	1	0,2	0,04
3	7	0	1	0	0	1	1	1
4	7	1	1	1	1	1	-0,2	0,04
5	8	2	2	4	4	4	-0,4	0,16
15	30	0	0	12	10	16	0	1,6

S.O.V	S.S	d.f	M.S	F
S.S.T	16	4	————————	27
SSR	14,4	1	14,4	
SSE	1,6	3	0,533	

F0,05,4 = 5,5 الجدولية

وبما أن قيمة F المحتسبة والتي تساوي 27 هي أكبر من قيمة F الجدولية والتي تساوي 5,05 لذا نرفض فرضية العدم ونقبل الفرضية البديلة وهذا يعني أن المعنوية الاجمالية للنموذج المقدر جيدة وتعبر عن واقع العلاقة بين كمية السماد الكيمياوي وقطر ساق النبات.

ملاحظة : في الانحدار الخطي البسيط (فقط) قيمة t تساوي F ويتم اختبار عينة B بواسطة اختبار t أيضاً عن طريق مقارنة قيمة t المحتسبة مع قيمة t الجدولية عند مستوى معنوية 0,05 ودرجة حرية n – 1 ، وفي مثالنا تكون قيمة t الجدولية هي : t 0,05 ,4 = 2,015

$T = \sqrt{27} = 5,19$

وهنا أيضاً كانت B معنوية لكون t المحتسبة أكبر من الجدولية وهذا يعني أن هناك تأثير معنوي للسماد الكيمياوي على قطر ساق النبات.

الانحدار الخطي المتعدد Multiple Linear Regression :

تناولنا في الانحدار الخطي البسيط العلاقة بين المتغير التابع (Y) ومتغير مستقل واحد هو (X) ، ولكن في الواقع أن نجد متغير تابع يعتمد على متغير مستقل واحد وأغلب المتغيرات في جميع العلوم ومنها الهندسية تعتمد على عدة متغيرات أخرى ، فالطلب مثلاً يعتمد على سعر السلعة ودخل المستهلك وأسعار السلع البديلة والمكملة ... الخ ، وان إنتاجية الموظف لا يعتمد على متغير التخصص فقط بل تعتمد على مستوى المهارة أو الخبرة أو التدريب ... الخ ، فلابد من معرفة أثر كل متغير مستقل في المتغير التابع وهذه المهمة يقوم بها الانحدار المتعدد بافتراض أن العلاقة خطية بين المتغيرات أو هي بيسا كذلك.

سنقوم بدراسة موضوع الانحدار الخطي المتعدد وباستخدام متغير تابع ومتغيرين مستقلين ، أما إذا زاد عدد المتغيرات المستقلة عن أثنين فيكون من الصعب إنجاز الانحدار بطريقة يدوية وإنما لابد من الاعتماد على البرامج الإحصائية في ذلك (SPSS) مثلاً.

ونأخذ معادلة الانحدار الخطي المتعدد الصيغة الآتية :

$$Y = \propto + B_1 X_1 + B_2 X_2 + B_3 X_3 \dots B_n X_n$$

حيث تشير :

Y : إلى القيمة المحسوبة للمتغير التابع Y.

$\propto$: الحد المستقل أو هي قيمة Y عندما تكون كل من X_1 و X_2 تساوي صفر.

B_1 : وهي معامل X_1 وتمثل المتغير الناجم في Y نتيجة تغير القيمة بمقدار وحدة واحدة بافتراض ثبات قيمة X_2.

B_2 : وهي معامل X_2 وتمثل المتغير الناجم في Y نتيجة تغير X_2 بوحدة واحدة بافتراض ثبات قيمة X_1.

مثال :

توفرت لديك البيانات التالية عن سعر السلعة X_1 ودخل المستهلك X_2 والكمية المطلوبة Y ، المطلوب تقدير معادلة الانحدار الخطي المتعدد ومن ثم شرح المعادلة المقدرة :

X_1	X_2	Y	x_1	x_2	y	x_1^2	x_2^2	y^2	x_1x_2	x_1y	x_2y
5	2	3	2	-4	-5	4	16	25	-8	-10	20
4	3	4	1	-3	-4	1	9	16	-3	4	12
3	5	6	0	-1	-2	0	1	4	0	0	2
2	8	12	-1	2	4	1	4	16	-2	-4	8
1	12	15	-2	6	7	4	36	49	-12	-14	42
15	30	40	0	0	0	10	66	110	-25	-32	84

$X_1 = (X X) , (X Y)$

$X X = \begin{pmatrix} \sum x_1^2 & \sum x_{1x_2} \\ \sum x_{1x_2} & \sum x_2^2 \end{pmatrix}$

$X X = \begin{pmatrix} 10 & -25 \\ -25 & 66 \end{pmatrix}$

منكوس المصفوفة $= \frac{1}{D} (C)$

$D = 10 (66) – (-25) (-25) = 660 – 625 = 35$

$C = \begin{pmatrix} 66 & 25 \\ 25 & 10 \end{pmatrix}$

$B_1 = \frac{1}{35} \begin{matrix} B_1 \\ B_2 \end{matrix} \begin{pmatrix} 66 & 25 \\ 25 & 10 \end{pmatrix} \begin{pmatrix} -32 \\ 84 \end{pmatrix} \begin{matrix} \sum x_1 y \\ \sum x_2 y \end{matrix}$

$B_1 = \frac{66 (-32)+25 (84)}{35} = \frac{-12}{35} = -0,34$

$B_2 = \frac{25 (-32)+10 (84)}{35} = \frac{40}{35} = 1,14$

$\propto = Y – B_1 X_1 – B_2 X_2 = 8 – (-0.34) (3) – 1,14 (6)$

$\propto = 8 + 1,02 – 6,84 = 2,18$

$Y = 2,18 – 0,34 X_1 + 1,14 X_2$

شرح المعادلة :

1- قيمة $\propto = 2,18$ الحد الثابت وتعني أن الكمية المطلوبة تساوي 2,18 عندما يكون كل من سعر السلعة ودخل المستهلك يساوي صفر.

2- قيمة $B_1 = -0,34$ حيث تدل الإشارة السالبة على العلاقة العكسية بين السعر والكمية المطلوبة وهذا يعني إذا ازداد السعر بمقدار وحدة واحدة فأن الكمية المطلوبة سوف تنخفض بمقدار 0,34 (مع استبعاد أثر X_2).

3- قيمة $B_2 = 1,14$ حيث تدل الإشارة الموجبة على العلاقة الطردية بين دخل المستهلك والكمية المطلوبة ، وهذا يعني إذا ازداد دخل المستهلك بمقدار وحدة واحدة (مع استبعاد أثر X_1) فأن الكمية المطلوبة سوف تزداد بمقدار 1,14.

من المثال السابق جد القوة التفسيرية للمعادلة المقدرة مع التعليق :

$$R^2 = \frac{B_1 \sum x_1 y + B_2 \sum x_2 y}{\sum y^2}$$

$$R^2 = \frac{-0,34\,(-32) + 1,14\,(84)}{110}$$

$$R^2 = \frac{10,88 + 95,76}{110} = 0.97$$

وهذا يعني أن المعادلة المقدرة تستطيع أن تفسر 97% من التغيرات التي تحدث في الكمية المطلوبة بإعادتها إلى تغيرات في كل من السعر ودخل المستهلك و 3% تعود إلى عوامل عشوائية أخرى غير داخلة في المعادلة.

معامل التحديد المصحح $\bar{R}^2$ Adjusted R Square :

يأخذ $\bar{R}^2$ بنظر الاعتبار عدد المتغيرات المستقلة الداخلة في النموذجين ، أما في نموذج الانحدار المتعدد فتكون :

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\,(1 - R^2)$$

حيث أن k هو عدد المتغيرات المستقلة.

مثال :

افترض بأن لدينا معادلة انحدار خطي بسيط لقيمة من 20 مشاهد لثلاث متغيرات (متغيرين مستقلين ومتغير تابع) وأن معامل التحديد R^2 لها كان 0,85 أوجد قيمة $\bar{R}^2$:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\,(1 - R^2)$$

$$\bar{R}^2 = 1 - \frac{20-1}{20-1}\,(1 - 0,85) \qquad \bar{R}^2 = 0,85$$

مثال :

نفترض أن لدينا معادلة انحدار خطي متعدد لعينة من 20 مشاهد لثلاث متغيرات (متغيرين مستقلين ومتغير تابع) وأن معامل التحديد R^2 لها كان 0,85 أوجد قيمة $\bar{R}^2$.

$$\bar{R}^2 = 1 - \frac{20-1}{20-1}\,(1 - 0,85)$$

$$\bar{R}^2 = 1 - \frac{19}{18}\,(0,15) = 0,841$$

وهذا يدل على أن معامل التحديد المصحح أقل من معامل التحديد في حالة الانحدار المتعدد.

اختبار t لمعنوية المعالم المقدرة t test :

ويعتبر من أهم وأكثر الاختبارات الشائعة في تحليل الانحدار والذي يطلق عليه اختبار المعنوية الأساسي والذي عن طريقه يمكن الإجابة عن التساؤل التالي :

هل هناك تأثير معنوي للمتغير المستقل X على المتغير التابع Y.

ويمكن الإجابة عن هذا التساؤل بعد استخراج قيمة t المحتسبة وإخضاعها إلى اختبار الفرضيات وإجراء عملية المقارنة بين قيمة t المحتسبة وقيمة t الجدولية والتي يتم استخراجها من جدول القيم الحرجة عند مستوى معنوية معين ودرجة حرية $n - 2$ وهنا نحصل على نتيجتين يتم اختيار منها :

1- حالة رفض H_0 : إذا كانت قيمة t المحتسبة أكبر من قيمة t الجدولية فأنه سيتم رفض فرضية العدم H_0 وقبول الفرضية البديلة وهذا يعني معنوية المعلمة المقدرة وبالتالي هناك تأثير معنوي للمتغير المستقل X على المتغير التابع Y.

2- حالة قبول H_0 : إذا كانت قيمة t المحتسبة أكبر من قيمة t الجدولية فأنه سيتم قبول فرضية العدم ورفض الفرضية البديلة وهذا يعني عدم معنوية المعلمة المقدرة وبالتالي لا يوجد هناك تأثير معنوي للمتغير X على المتغير Y.

كيفية إيجاد قيمة t في الانحدار الخطي المتعدد :

$$tBi = \frac{Bi}{SBi}$$

يجري هذا الاختبار حسب الخطوات التالية :

1- استخراج الخطأ : $\sum e^2 = \sum y^2 - B_1\sum x_1 y + B_{2\sum x_2 y}$

2- استخراج تباين الخطأ : $Se^2 = \frac{\sum ei^2}{n-k-1}$

3- استخراج تباين $\bar{B}_i$: $SB^2_i = \frac{Se^2}{\sum x_i^2}$

4- استخراج الخطأ المعياري : $SB_i = \sqrt{SB\overset{2}{i}}$

5- احتساب t : $tB_i = \frac{Bi}{SB_i}$

وبالعودة إلى مثالنا السابق (أثر السعر والدخل على الكمية المطلوبة) يمكن اختبار معنوية المعالم المقدرة :

1- معنوية B_1 :

$$\sum e^2 = 110 - 106{,}64 = 3{,}36$$

$$Se^2 = \frac{3{,}36}{5-2-1} = 1{,}68$$

$$SB_1^2 = \frac{1{,}68}{10} = 0{,}168$$

$SB_1 = \sqrt{0,168} = 0,409$

$tB_1 = \frac{-0,34}{0,409} = -0,831$

الإشارة السالبة تهمل.

2- معنوية B_2 :

$\sum e^2 = 110 - 106,64 = 3,36$

$Se^2 = \frac{3,36}{5-2-1} = 1,68$

$SB_2^2 = \frac{1,68}{66} = 0,025$

$SB_2 = \sqrt{0,025} = 0,159$

$tB_2 = \frac{1,14}{0,159} = 7,16$