الفصل السابع
اسس احصائية عامة في قياس الفروق الفردية

- تعريف علم الاحصاء
- الاحصاء الوصفي والاستدلالي
- الطرق الاحصائية في البحث العلمي
- اساليب جمع البيانات
- المجتمع الاحصائي والعينة الاحصائية
- التوزيع التكراري وطرق عرض البيانات
- مقاييس النزعة المركزية ومقاييس التشتت
- معاملات الارتباط

تعريف علم الإحصاء:

هو العلم الذي يختص بجمع وتصنيف وتبويب البيانات وعرض وتحليل البيانات واستخلاص النتائج منها.

يقسم الإحصاء بصورة عامة إلى قسمين رئيسين هما

١-الإحصاء الوصفي Descriptive Statistic

- يتضمن طرق جمع البيانات وتصنيفها وعرضها في جداول ورسومات بيانية وحساب بعض المؤشرات الإحصائية.
- عبارة من مجموعة الأساليب الإحصائية التي تعنى بجمع البيانات وتنظيمها وتصنيفها وتلخيصها وعرضها بطريقة واضحة في صورة جداول أو أشكال بيانية وحساب المقاييس الإحصائية المختلفة لوصف متغير ما (أو أكثر من متغير) في مجتمع ما أو عينه منه.

٢-الإحصاء الاستدلالي Inference Statistic

- عبارة عن مجموعة من الأساليب الإحصائية التي تستخدم بغرض تحليل بيانات ظاهرة (أو أكثر) في مجتمع ما على أساس بيانات عينة احتمالية تسحب منه وتفسيرها للتوصل إلى التنبؤ واتخاذ القرارات المناسبة.
- يتضمن طرق اختبار الفرضيات وتقدير المعالم

أهمية علم الإحصاء

يعتبر علم الإحصاء احد الوسائل المهمة في البحث العلمي من خلال استخدام قواعده وقوانينه وطرقه في جمع البيانات و المعلومات اللازمة في البحث العلمي وتحليل هذه المعلومات لغرض الوصول إلى النتائج التي يهدف لها البحث كما وان للإحصاء دورا بارزا في وضع الخطط المستقبلية عن طريق التنبؤ بالنتائج لكافة القطاعات حيث يمكن تطبيق علم الإحصاء في مجالات العلوم الصرفة أو العلوم الإنسانية (مثل الزراعة ،الصناعة ،الطب، الهندسة، الإدارة والاقتصاد...الخ)

الطريقة الإحصائية في البحث العلمي

١- تحديد المشكلة ووضع الفرضيات.
٢- جمع البيانات والمعلومات ذات العلاقة بالبحث.
٣- تصنيف وتبويب وعرض البيانات.
٤- حساب المؤشرات الإحصائية.
٥- تحليل وتفسير النتائج على ضوء فرضية البحث .
٦- تفسير النتائج واتخاذ القرار.

مصادر جمع البيانات

١- المصادر التاريخية: وتشمل الوثائق والمطبوعات والنشرات الإحصائية التي تصدرها الهيئات الحكومية في البلدان المختلفة وتسمى البيانات المجموعة بهذه الطريقة بالبيانات الثانوية مثل تعداد السكان والصادرات والواردات.

٢- المصادر الميدانية: وتشمل البيانات والمعلومات التي يمكن الحصول عليها من مصادرها الأساسية عن طريق المواجهة أو المراسلة أو بأي طريقة أخرى وتسمى البيانات المجموعة بهذه الطريقة بالبيانات الأولية مثل تسجيل حوادث لأي شهر.

أساليب جمع البيانات

١- أسلوب التسجيل الشامل: يتم جمع البيانات هنا عن طريق تسجيل مفردات المجتمع (مفردة ،مفردة)لذلك فان هذا الأسلوب يتطلب وقت وجهد ويفترض عند استخدام هذا الأسلوب إن يكون المجتمع محدود مثل تعداد السكان او حصر عدد العاملين في أجهزة الدولة ويمتاز هذا الأسلوب بكونه يتعامل مع جميع بيانات مجتمع البحث مما يؤدي للحصول على نتائج دقيقة .

٢- أسلوب العينات : وهي عملية جمع البيانات والمعلومات عن مجموعة معينة من مفردات المجتمع مثل استفتاء بعض الأشخاص عن موضوع معين وهذا الأسلوب يحتاج إلى وقت وجهد وموارد اقل مما يحتاجه الأسلوب الأول بالإضافة إلى انه مفيد في حالة دراسة المجتمعات الغير محدودة ولكن دقة النتائج هنا تكون اقل كونه لا يتعامل مع جميع المفردات.

الأخطاء الشائعة في جمع البيانات

١- خطا التحيز: يحدث هذا الخطأ نتيجة جمع البيانات من مصادرها غير الرئيسية أو التحيز لمفردة معينة في المجتمع قيد الدراسة دون الأخرى لسبب أو لأخر .

٢- خطا الصدفة: يحدث هذا الخطأ نتيجة اعتماد الباحث على معلوماته الشخصية أو جمع البيانات ناقصة .

البيانات

هي عملية تقسيم البيانات الى مجاميع صغيرة او اصناف على اساس قاعدة معينة كاشتراكها في بعض الصفات او الخصائص كالمهنة او الحالة الاجتماعية وحسب متطلبات البحث.

تبويب البيانات

وهي عملية ادخال البيانات المصنفة في جداول خاصة بهدف ابراز البيانات وتوضيحها في اضيق نطاق ممكن لتكوين فكرة عنها ويختلف اسلوب تبويب البيانات تبعا لطبيعتها كما يأتي:

١- التبويب الزمني: ومعنى ذلك فرز البيانات الى مجموعات على اساس وحدة زمنية معينة كالسنة او الشهر.

٢- التبويب الجغرافي: ومعنى ذلك فرز البيانات الى مجموعات على اساس وحدة جغرافية معينة مثل محافظة معينة او بلد معين.

٣- التبويب الكمي: ومعنى ذلك فرز البيانات الى مجموعات على اساس وحدات كمية اي الارقام كالطول والوزن....الخ.

٤- التبويب الوصفي: ومعنى ذلك فرز البيانات الى مجموعات على اساس صفة معينة مشتركة مثل الحالة الاجتماعية او لون العين ....الخ.

المجتمع الاحصائي والعينة الاحصائية

١- المجتمع الاحصائي: يعرف المجتمع الاحصائي بانه مجموعة من العناصر او المفردات التي تخص ظاهرة معينة وبهذا يمكن ان تكون عناصر المجتمع الاحصائي افراد او عائلات او موظفين...الخ ويتمثل المجتمع الاحصائي بعدد العناصر او المفردات التي يتضمنها ويرمز له بالرمز N.

٢- العينة الاحصائية: تعرف العينة بانها جزء من مفردات المجتمع الاحصائي يتم اختيارها بطريقة علمية صحيحة وتحمل نفس خصائص المجتمع الاحصائي وتستخدم لأغراض الدراسة في حال تعذر اجراء الدراسة على المجتمع عندما يكون حجمه كبيرا .

المتغيرات العشوائية

هي دالة ذات قيمة حقيقية معرفة في فضاء العينة لأي تجربة عشوائية وعادة ما يرمز لها بالأحرف الانكليزية الكبيرة مثل X,Y,Z وتقسم المتغيرات العشوائية الى قسمين هما:

١- المتغيرات الوصفية: وهي المتغيرات التي تكون مفرداتها غير قابلة للقياس بالأرقام بل تكون على شكل صفات مثل لون العين او تقديرات الطلاب .

٢- المتغيرات الكمية: وهي المتغيرات التي تكون مفرداتها قابلة للقياس بالأرقام مثل اعداد طلبة الجامعات العراقية ،الطول ،العمر ...الخ وتقسم الى قسمين هما:

ا) المتغيرات المتقطعة : وهي المفردات التي تأخذ قيما متميزة عن بعضها وتكون قابلة للعد والحساب ولا تتقبل الكسور مثل عدد الطلاب في كلية معينة او عدد العوائل في منطقة معينة.

ب) المتغيرات المستمرة: وهي المفردات التي تأخذ مدى معين او مجال معين وتكون غير قابلة للعد والحساب و تتقبل الكسور مثل اعمار الافراد او اطوالهم او درجاتهم.

التوزيع التكراري

التوزيع التكراري: و هو عبارة عن ترتيب بيانات المتغير العشوائي الكمي بنوعيها (المستمرة والمتقطعة) حسب اشتراكها بصفة معينة في جداول خاصة تسمى بجداول التوزيع التكراري (Frequency Distribution Tables) مقسمة الى فئات (Classes) وتكرارات (Frequencies) وتكون الفئات على نوعين فئات مغلقة او مفتوحة حسب طبيعة البيانات .

خطوات وضع البيانات في جدول توزيع تكراري

١- تحديد اكبر قيمة في البيانات ويرمز لها ب (V_L) واصغر قيمة فيها ويرمز لها ب (V_S).

٢- ايجاد المدى الكلي للبيانات (Total Range) ويرمز له ب (T.R) وفق الصيغة التالية:

$$T.R=V_L-V_S+1$$

٣- ايجاد عدد الفئات وفق احدى الصيغ التالية

$$K=1+3.3*Log(n)$$

$$K=2.5*$$

٤-ايجاد طول الفئة وهو عبارة عن مقدار سعة الفئة من الارقام وفق الصيغة التالية:

$$L= T.R/K$$

٥ – تحديد الحد الادنى (Lower Limit) لكل فئة والحد الاعلى (Upper Limit) لها كما يأتي:

- في حالة المتغير المتقطع:

$$L.L=V_S$$

$$U.L=L.L+L-1$$

- في حالة المتغير المستمر

$$L.L=VS$$

$$U.L=L.L+L$$

٦- تحديد تكرار الفئة (Class Frequency) وهو عبارة عن جميع المفردات التي تقع ضمن حدود الفئة من حيث القيمة العددية ويرمز له ب (fi) .

٧- ايجاد مركز الفئة (Class Midpoint) وهو القيمة التي تقع في منتصف الفئة ويرمز له ب (Mi) وفق الصيغة التالية: $Mi=(L.L+U.L)/2$

فيكون شكل جدول التوزيع التكراري كالتالي:

١-في حالة المتغير المتقطع يسمى بالتوزيع المغلق أي ان جميع الفئات لها حدود دنيا وعليا كما يأتي:

٢- في حالة المتغير المستمر ويسمى بالتوزيع المفتوح حيث ان جميع الفئات لديها حدود دنيا فقط ماعدا الفئة الاخيرة التي تكون مغلقة أي ان لها حد ادنى واعلى كما ياتي:

مثال (١) : حالة المتغير المتقطع

البيانات التالية تمثل عدد اشجار النخيل التي تمتلكها (٢٠) عائلة فلاحية في مدينة البصرة

Xi:65,56,88,83,75,60,70,65,55,67,45,65,74,72,62,48,69,65,49,49

المطلوب تمثيل هذه البيانات في جدول توزيع تكراري

الحل:

(1) VS=45,VL=89

(2) T.R=VL−VS+1=89−45+1=45

(3) K=1+3.3*Log(n)=1+3.3*Log(20)=5.29=5

(4) L= T.R/K= 45/5=9

والان نضع البيانات في جدول تكراري كالتالي:

عدد الاشجار(الفئات)	التكرار بالإشارات	التكرار	مراكز الفئات
45 −53	////	4	M1=(45+53)/2=49
54−62	~~////~~	5	M2=(54+62)/2=58
63−71	~~////~~ /	6	M3=(63+71)/2=67
72−80	///	3	M4=(72+80)/2=76
81−89	//	2	M5=(81+89)/2=85
المجموع	n=20	20	

يتضح من الجدول أعلاه أن كل فئة تمثل عدد الأشجار وان التكرارات تمثل عدد العوائل التي تمتلكها.

مثال (٢) : حالة المتغير المستمر

البيانات التالية تمثل درجات (١٦) طالبا في احدى المواد

Xi:50,82.5,75,60.5,76.6,80,74,66.5,78.5,72.5,84,71.5,70.5,70,72,58

المطلوب تمثيل هذه البيانات في جدول توزيع تكراري.

(١) VS=٥٠ ,VL=84

(٢) T.R=VL−VS+1=84−50+1=35

(٣) K=2.5*(n)1/4=2.5*(26)1/4=5

(٤) L= T.R/K= 35/5=7

والان نضع البيانات في جدول تكراري كالتالي :

الـــــــــــدرجات (الفئات)	التكرار بالإشارات	التكرار	مراكز الفئات
50 −	/	1	M1=(50+57)/2=53.5
57 −	//	2	M2=(57+64)/2=60.5
64 −	///	3	M3=(64+71)/2=67.5
71−	///// /	6	M4=(71+78)/2=74.5
78 −85	////	4	M5=(78+85)/2=81.5
المجموع	n=16	16	

يتضح من الجدول اعلاه ان كل فئة تمثل الدرجات وان التكرارات تمثل عدد الطلاب الذين حصلوا عليها.

التوزيع التكراري التراكمي او المتجمع

يهتم هذا النوع من التوزيعات بتحديد القيم التي تقل او تزيد عن قيمة معينة مقابل كل فئة من فئات التوزيع وتكون التوزيعات التكرارية المتجمعة على نوعين هما:

ا) التكرار المتجمع الصاعد

يمكن الحصول على التكرار المتجمع الصاعد من خلال تجميع او تراكم تكرارات الجدول الاصلي بدءا بتكرار الفئة الاولى وانتهاء بتكرار الفئة الاخيرة منه الى ان نحصل على مجموع التكرارات كتكرار متجمع صاعد للفئة الاخيرة والشكل التالي يوضح التصميم العام لهذا النوع من التوزيع ويرمز له ب (Fi)

مثال (٣) بالعودة لبيانات مثال رقم (١) جد التكرار المتجمع الصاعد

عدد الاشجار(الفئات)	التكرار	التكرار المتجمع الصاعد
45 -53	4	4
54-62	4	8
63-71	7	15
72-80	3	18
81-89	2	20
المجموع	20	

يتضح من الجدول أعلاه إن عدد العوائل التي تمتلك اقل من ٧١ نخلة هو ١٥ وهكذا.

ب) التكرار المتجمع النازل

يمكن الحصول على التكرار المتجمع النازل من خلال طرح تكرارات الجدول الاصلي من مجموع التكرارات على التوالي بدءا بتكرار الفئة الاولى وانتهاء بتكرار الفئة الاخيرة منه الى ان نحصل على التكرار الاخير كتكرار متجمع نازل للفئة الاخيرة والشكل التالي يوضح التصميم العام لهذا النوع من التوزيع ويرمز له ب (Fi)

مثال (٤) بالعودة لبيانات مثال رقم (١) جد التكرار المتجمع النازل

عدد الأشجار(الفئات)	التكرار	التكرار المتجمع النازل
45 -53	4	20
54-62	4	16
63-71	7	12
72-80	3	5
81-89	2	2
المجموع	20	

يتضح من الجدول اعلاه ان عدد العوائل التي تمتلك اكثر من ٦٣ نخلة هو ١٢ وهكذا.

التوزيع التكراري النسبي المئوي

وهو توزيع اعتيادي تكون التكرارات فيه على شكل تكرارات نسبية مئوية ويرمز له ب (F*i) ويمكن الحصول على التكرار النسبي المئوي وفق الصيغة التالية:

$$F^*i=(fi/n)*100\%$$

مثال (٥) بالعودة لبيانات مثال رقم (٢) جد التكرار النسبي المئوي

الدرجات (الفئات)	التكرار	التكرار النسبي المئوي
50 –	1	6.25%
57 –	2	12.5%
64 –	3	18.75%
71–	6	37.5%
78 –85	4	25%
المجموع	16	

يتضح من الجدول أعلاه إن نسبة ٢٥% من الطلاب تتراوح درجاتهم بين ٧٨ و ٨٥ .

العرض الهندسي للبيانات

وهو عبارة عن تمثيل ووصف البيانات التي يتم جمعها عن ظاهرة معينة بواسطة أشكال بيانية أو رسوم هندسية بهدف إعطاء فكرة واضحة وسهلة وسريعة عن بيانات الظاهرة المدروسة ويمكن استخدام الرسوم البيانية للبيانات الاعتيادية والبيانات المبوبة على حد سواء.

العرض الهندسي للبيانات الاعتيادية

إن البيانات الاعتيادية تعني البيانات التي لا تكون معروضة بشكل جدول توزيع تكراري ويتم تمثيلها بيانيا بالأشكال التالية:

١- المستطيل البياني Rectangle Chart

تتلخص فكرة هذا الشكل باختيار مستطيل ذو قاعدة مناسبة يتم افتراضها ثم يقسم المستطيل الى مستطيلات جزئية تمثل بيانات الصفة المدروسة وفق الصيغة التالية:

طول قاعدة المستطيل الجزئي =(البيانات الجزئية/ البيانات الكلية) * طول قاعدة المستطيل الكلي .

مثال (6):بلغت التكاليف الإنتاجية لإنتاج سلعة معينة (300) دولار كما موضحة بالجدول التالي:

مستلزمات الإنتاج	التكاليف/دولار
أجور	120
مواد خام	60
مصاريف مباشرة	90
مصاريف غير مباشرة	30

الحل: نفرض أن طول قاعدة المستطيل الكلي = 10

طول قاعدة المستطيل الجزئي =(البيانات الجزئية/ البيانات الكلية) * طول قاعدة المستطيل الكلي

طول قاعدة المستطيل الأول (الأجور) = (300/120) *10= 4

طول قاعدة المستطيل الثاني(المواد الخام) =(300/60) *10= 2

طول قاعدة المستطيل الثالث (المصاريف المباشرة)= (90 /300) *10= 3

طول قاعدة المستطيل الرابع (المصاريف غير المباشرة)= (300/30) *10= 1

٢-الاشرطة البيانية Bar- Charts

وهي عبارة عن مجموعة من المستطيلات الافقية او العمودية قواعدها متساوية وتمثل الصفة التي تم على اساسها التبويب (سنة ،محافظة ، وهكذا) وارتفاعاتها تمثل البيانات المقابلة لتلك الصفة والاشرطة البيانية على نوعين هما:

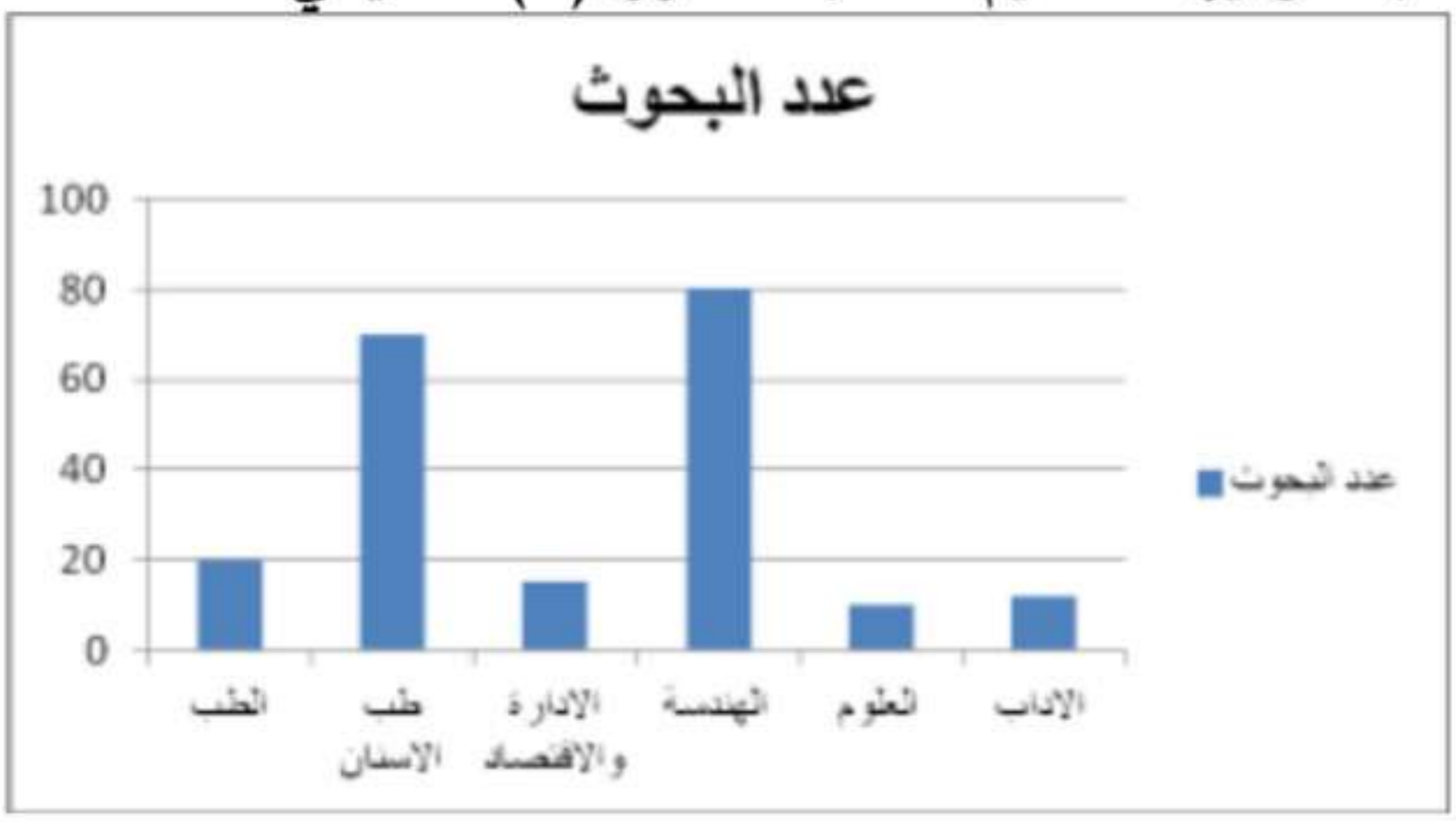
ا) الاشرطة البيانية المفردة

وهي اشرطة بيانية تخص صنف واحد للبيانات مثل عدد الطلبة المقبولين او تطور عدد سكان العراق حسب التعداد السكاني

مثال(٧): البيانات الواردة في الجدول التالي تمثل عدد البحوث العلمية المنجزة من قبل اعضاء هيئة التدريس في جامعة معينة موزعين حسب كلياتهم لعام ٢٠٠٨

الكليات	عدد البحوث
الطب	20
طب الاسنان	70
الادارة والاقتصاد	15
الهندسة	80
العلوم	10
الآداب	12

الحل: يتم رسم الاشرطة البيانية بعد وضع المحور الافقي(السيني) X والذي يمثل الكليات في هذا المثال والمحور العمودي (الصادي)Y والذي يمثل عدد البحوث بعد ان يؤخذ تقسيم مناسب للمحور (Y) كما يأتي :



ب) الاشرطة البيانية المركبة

وهي اشرطة بيانية تخص صنفين او اكثر للبيانات مثل عدد الكتب الموجودة في احدى المكتبات مصنفة حسب انواعها الى كتب علمية وادبية وتاريخية وغيرها او عدد الموظفين في احدى الدوائر مصنفين حسب درجاتهم الوظيفية.

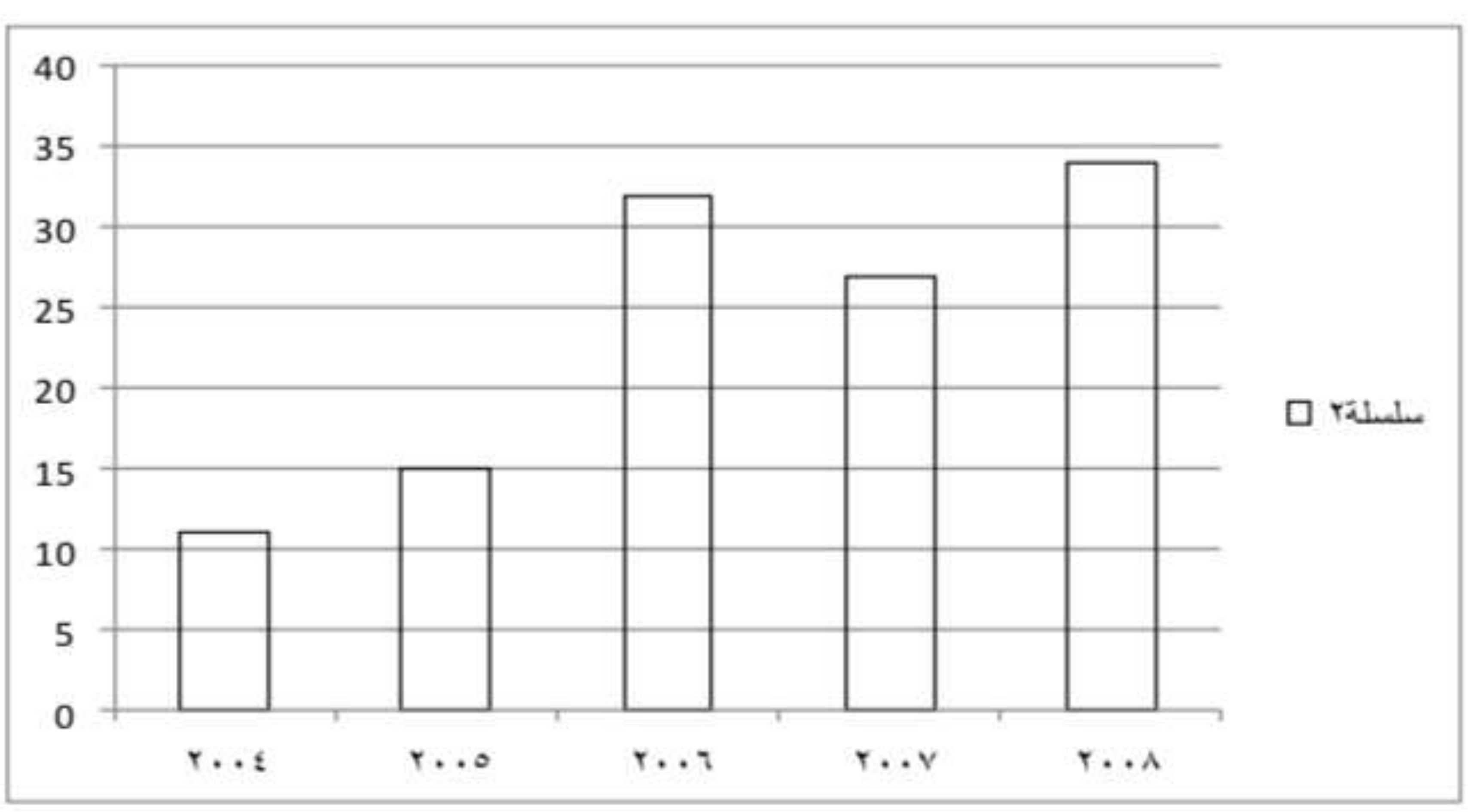
مثـال (٨) : البيانـات التاليـة تمثـل عـدد النـدوات والمـؤتمرات العلميـة التـي عقدتها كليات جامعة معينة:

السنوات	2	2	2	2	2
عدد الندوات	9	1	2	1	1
عدد المؤتمرات	2	1	9	12	16

الحل : نعمل الجدول التالي

السنوات	2	2	2	2	2
عـــــدد	9	1	2	1	1
عـــــدد	2	1	9	1	1
المجموع	1١	1	3	2	3

يتم رسم الاشرطة البيانية بعد وضع المحور الافقي(السيني) X والذي يمثل السنوات في هذا المثال والمحور العمودي (الصادي)Y والذي يمثل المجموع بعد ان يؤخذ تقسيم مناسب للمحور (Y) ومن ثم يرسم شريط ضخم قاعدته السنة وارتفاعه المجموع وبداخله شريطين صغيرين احدهما لعدد الندوات والاخر لعدد المؤتمرات كما يأتي:



٣-الدائرة البيانية Pie-chart

وهي عبارة عن شكل هندسي مثل المستطيل البياني ولكن يتم هنا تمثيل البيانات بقطاعات داخل دائرة بحيث ان مجموع هذه القطاعات تمثل مساحة الدائرة الكلية ويتم تحديد زاوية القطاع وفق الصيغة التالية:

زاوية القطاع = (البيانات الجزئية/ البيانات الكلية) * ٣٦٠°

مثال (٩): بالعودة الى بيانات مثال رقم ٦ ارسم الدائرة البيانية

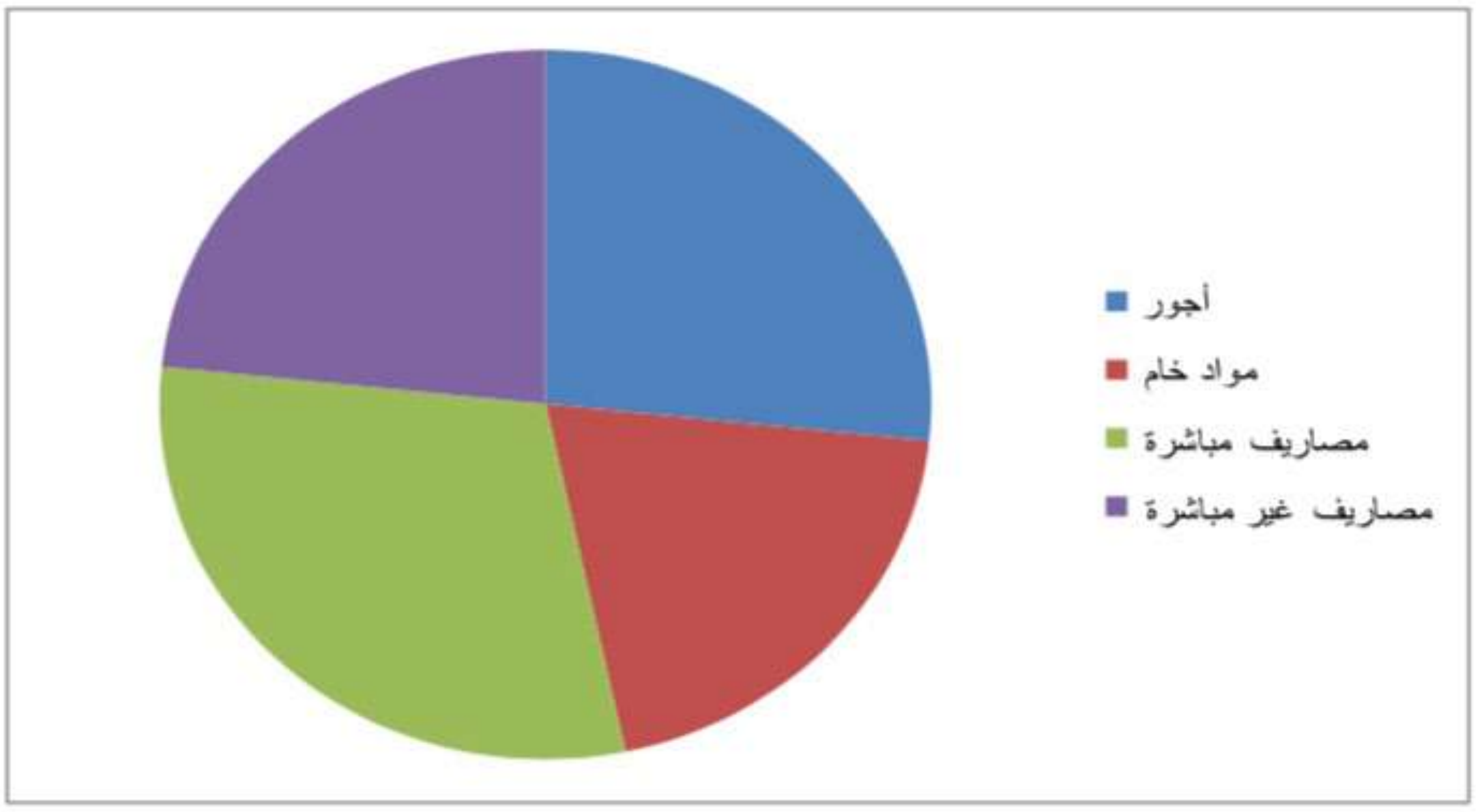
الحل:

زاوية قطاع الاجور = (٨٠/٣٠٠)* ٣٦٠°= ٩٦°

زاوية قطاع المواد الخام = (٦٠/٣٠٠)* ٣٦٠° = ٧٢°

زاوية قطاع المصاريف المباشرة = (٩٠/٣٠٠)* ٣٦٠° =° ١٠٨

زاوية قطاع المصاريف غير المباشرة = (٧٠/٣٠٠)* ٣٦٠° = ٨٤°

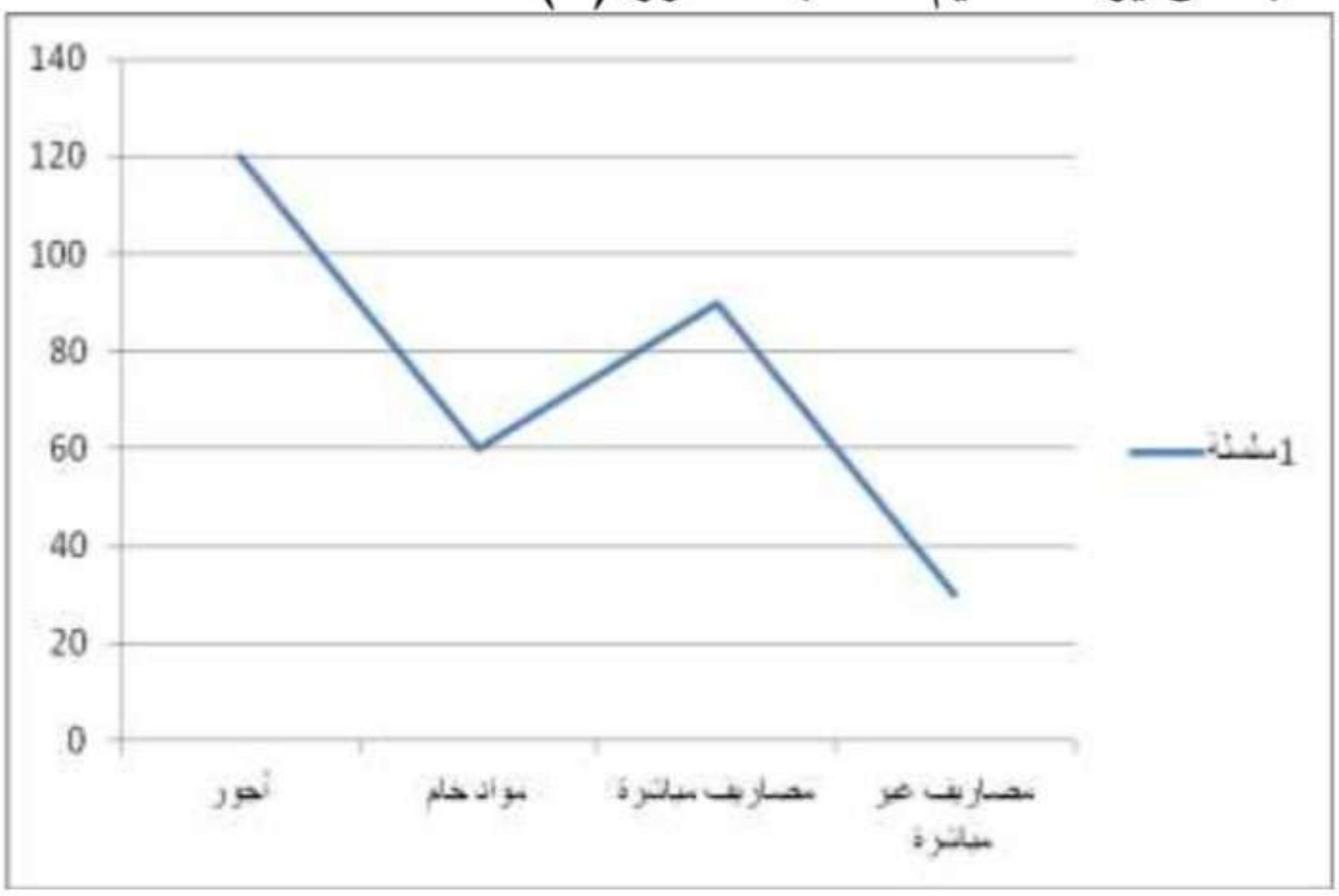


٤- الخط البياني Line-chart

عبارة عن شكل بياني يوضح التغيرات الحاصلة في ظاهرة معينة عبرة فترة محددة من الزمن وهو شكل نافع في اجراء مقارنة بين ظاهرتين او اكثر مقاسة بنفس وحدات القياس على سبيل المثال مقارنة التغيرات الحاصلة بين كميات النفط المنتجة والمصدرة او مقارنة تكاليف انتاج سلعة معينة والارباح المتحققة من مبيعاتها خلال مدة زمنية معينة وغيرها من الامثلة الاخرى.

مثال (١٠): بالعودة الى بيانات مثال رقم ٦ ارسم الخط البياني

الحل: يتم رسم الخط البياني بعد وضع المحور الافقي(السيني) X والذي يمثل مستلزمات الإنتاج في هذا المثال والمحور العمودي (الصادي)Y والذي يمثل التكاليف بعد ان يؤخذ تقسيم مناسب للمحور (Y)

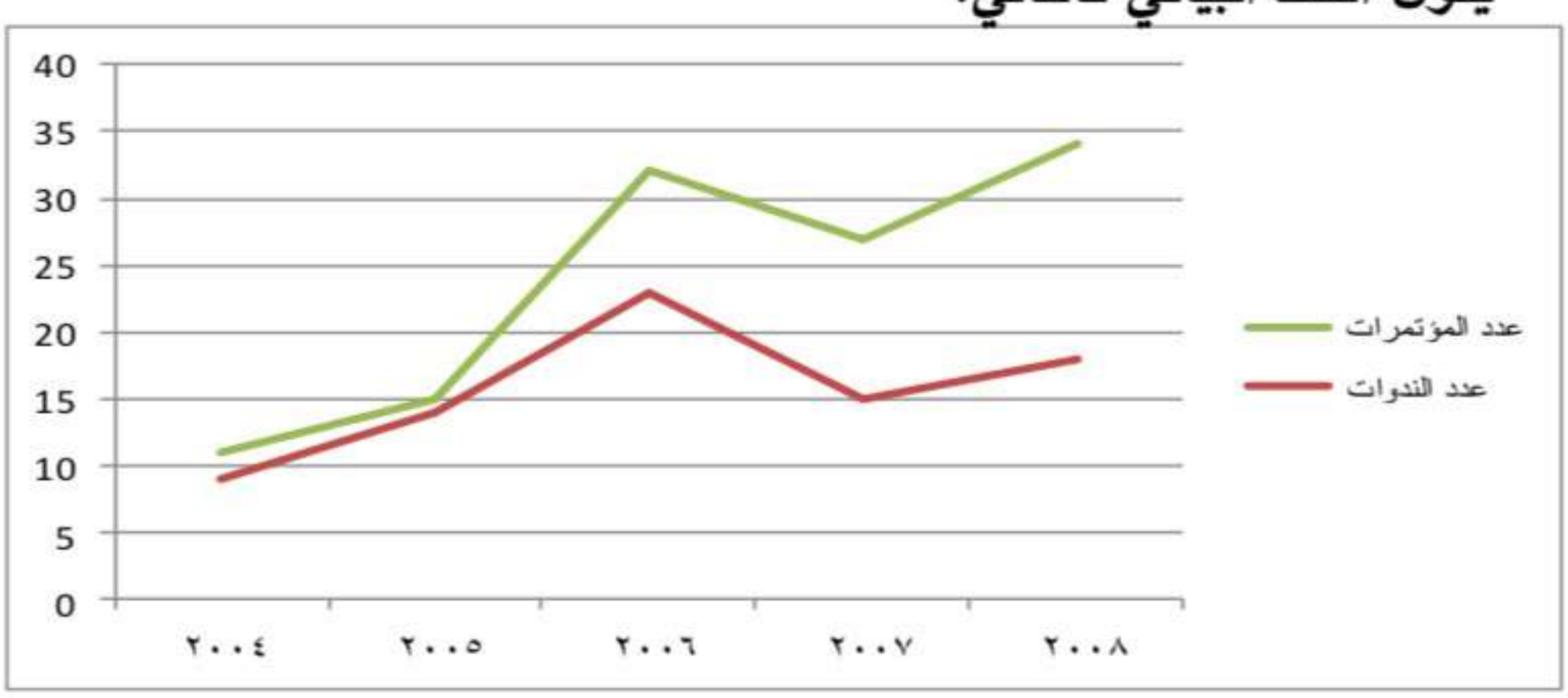


ملاحظة يمكن رسم الخط البياني لصنفين او اكثر من البيانات كما يأتي:

بالعودة لبيانات مثال (٨)

السنوات	2	2	2	2	2
عـــدد	9	1	2	1	1
عـــدد	2	1	9	1	1

يكون الخط البياني كالتالي:



ثانيا: العرض الهندسي للبيانات المبوبة

إن البيانات المبوبة تعني البيانات التي تكون معروضة بشكل جدول توزيع تكراري ويتم تمثيلها بيانيا بالأشكال التالية:

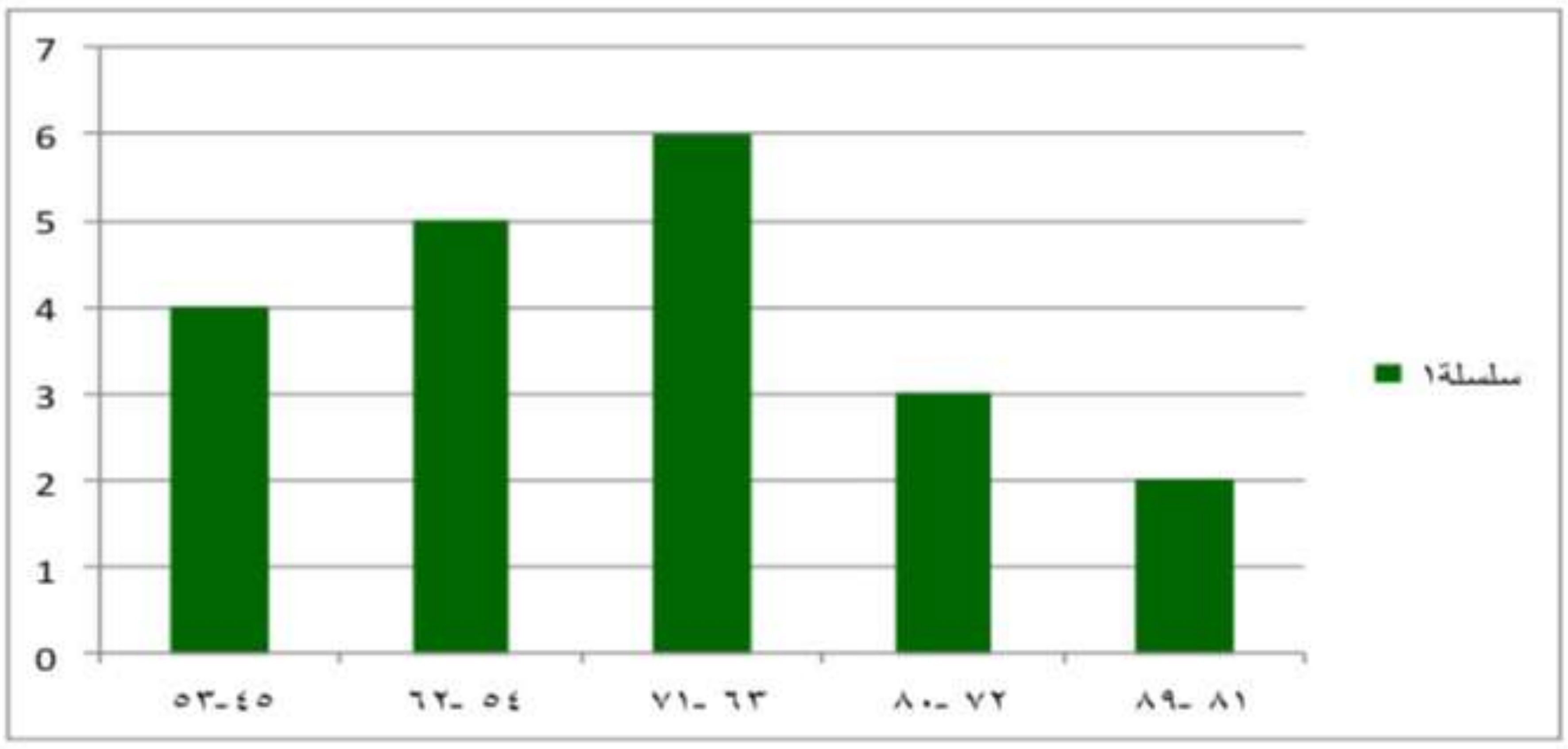
١- المدرج التكراري Histogram

وهو عبارة عن مجموعة من المستطيلات قاعدة كل منها تمثل طول الفئة في التوزيع التكراري وارتفاعها يمثل التكرار المقابل لتلك الفئة أي ان المحور السيني (X) تستقر فيه الفئات والمحور الصادي (Y) تستقر فيه التكرارات هذه المستطيلات تكون منفصلة عن بعضها في حالة المتغير المتقطع ومتصلة مع بعضها في حالة المتغير المستمر وحسب تسلسل فئات التوزيع.

مثال (١١) للتوزيع التكراري التالي ارسم المدرج التكراري

التكرار	عدد الأشجار
4	45 -53
5	54 -62
6	63 -71
3	72 -80
2	81 -89

الحل: المحور السيني (X) تستقر فيه الفئات (عدد اشجار النخيل) والمحور الصادي (Y) تستقر فيه التكرارات (عدد العوائل التي تملكها).



٢- المضلع التكراري Polygon

وهو عبارة عن عدد من المستقيمات التي تتصل ببعضها بواسطة نقاط هذه النقاط تمثل مراكز الفئات أي ان المحور السيني (X) تستقر فيه مراكز الفئات والمحور الصادي (Y) تستقر فيه التكرارات مع مراعاة غلق المضلع بمركزي فئة وهميين (مركز الفئة الاولى – طول الفئة ومركز الفئة الاخيرة + طول الفئة) وبتكرارين مساويين للصفر.

مثال (١٢) للتوزيع التكراري في مثال (١١) ارسم المضلع التكراري

الحل: المحور السيني (X) تستقر فيه مراكز الفئات والمحور الصادي (Y) تستقر فيه التكرارات

مراكز الفئات	التكرار	عدد الأشجار
49	4	45 -53
58	5	54 -62
67	6	63 -71
76	3	72 -80
85	2	81 -89

بعد ذلك نحدد مركزي فئة وهميين من اجل غلق المضلع وذلك بطرح طول الفئة من مركز الفئة الاولى واضافة طول الفئة لمركز الفئة الاخيرة:

Mo=49-9=40

M6=85+9=94

فيكون المضلع بالشكل التالي:

٣-المنحنى التكراري Curve

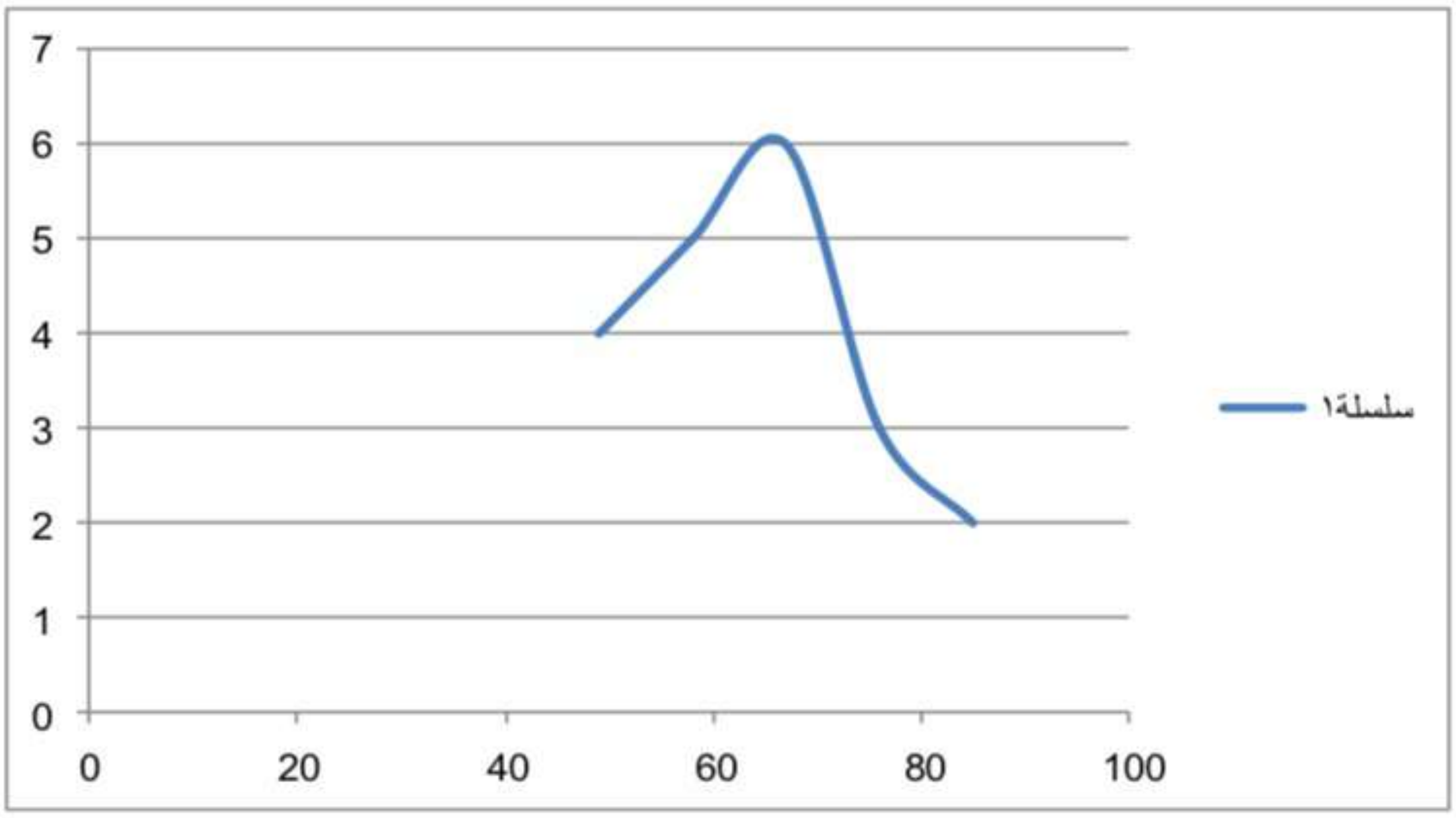
ان فكرة رسم المنحنى التكراري لا تختلف عن رسم المضلع التكراري الا انه يتم ايصال النقاط بخط منحنى بدلا من الخطوط المستقيمة و لا يوجد داعي لغلقه.

مثال (١٣) للتوزيع التكراري في مثال (١١) ارسم المضلع التكراري

الحل: المحور السيني (X) تستقر فيه مراكز الفئات والمحور الصادي (Y) تستقر فيه التكرارات

مراكز الفئات	التكرار	عدد الأشجار
49	4	45 -53
58	5	54 -62
67	6	63 -71
76	3	72 -80
85	2	81 -89

فيكون المنحنى بالشكل التالي:



مقاييس النزعة المركزية

مقدمة:

مقاييس النزعة المركزية هي مقاييس مهمة من المقاييس الإحصائية، والهدف منها تعيين موقع التوزيع وهي مهمة في حالة المقارنة بين التوزيعات المختلفة بشكل عام..

ويمكن تعريفها بأنها القيمة النموذجية الممثلة لمجموعة من البيانات. وحيث أن القيمة النموذجية تميل إلى الوقوع في المركز لذلك فإنها تسمى بمقاييس النزعة المركزية.

تعريف رمز التجميع:

إذا كان لدينا مجموعة من المشاهدات $x_1, x_2, \ldots, x_n$ فإن حاصل جمع هذه المشاهدات يمكن التعبير عنه كما يلي:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

مقاييس النزعة المركزية (Averages and Measures of Central Tendency):

أ. الوسط الحسابي (المتوسط) (Mean or Arithmetic Mean)

أولاً- للبيانات غير المبوبة (non- grouped data-):

إذا كان لدينا مجموعة من المشاهدات للمتغير X وهي $x_1, x_2, \ldots, x_n$ فإن الوسط الحسابي $\bar{x}$ - ويُقرأ (x bar) - يساوي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

ثانياً- للبيانات المبوبة (grouped data):

إذا كان لدينا عدد k من الفئات ذات المراكز $x_1, x_2, \ldots, x_k$ و لها تكرارات $f_1, f_2, \ldots, f_k$ على الترتيب فإن الوسط الحسابي يعطى بالعلاقة التالية:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n}$$

حيث n. مجموع التكرارات.

مميزات الوسط الحسابي:

١. سهل الحساب.
٢. يأخذ جميع القيم بالاعتبار
٣. أكثر المقاييس فهما في الإحصاء

عيوب الوسط الحسابي:

١. يتأثر بالقيم المتطرفة.

٢. لا يمكن حسابه في حالة البيانات الوصفية.

ب. الوسيط (Median)

أولاً- للبيانات غير المبوبة (non- grouped data):

هي القيمة التي تقع في منتصف البيانات بعد ترتيبها تصاعدياً أو تنازلياً ويرمز له بالرمز Med.. وهنا نميز بين حالتين:

١. عندما يكون عدد البيانات فرديا فإن الوسيط هو القيمة الواقعة في منتصف البيانات.

٢. عندما يكون عدد البيانات زوجيا فإن الوسيط هو متوسط القيمتين الواقعتين في منتصف البيانات.

ثانياً- للبيانات المبوبة (grouped data):

الفئة الوسيطية: هي الفئة التي يقع فيها الوسيط.

خطوات إيجاد الوسيط:

- نكون جدول التكرار المتجمع الصاعد.
- نوجد رتبة الوسيط والتي تساوي $\frac{n}{2}$ حيث n عدد البيانات الكلي.
- نحدد مكان الوسيط من الفقرة السابقة ثم نحدد الآتي:

١. بداية الفئة الوسيطية A

٢. التكرار السابق لرتبة الوسيط f_1

٣. التكرار اللاحق لرتبة الوسيط f_2

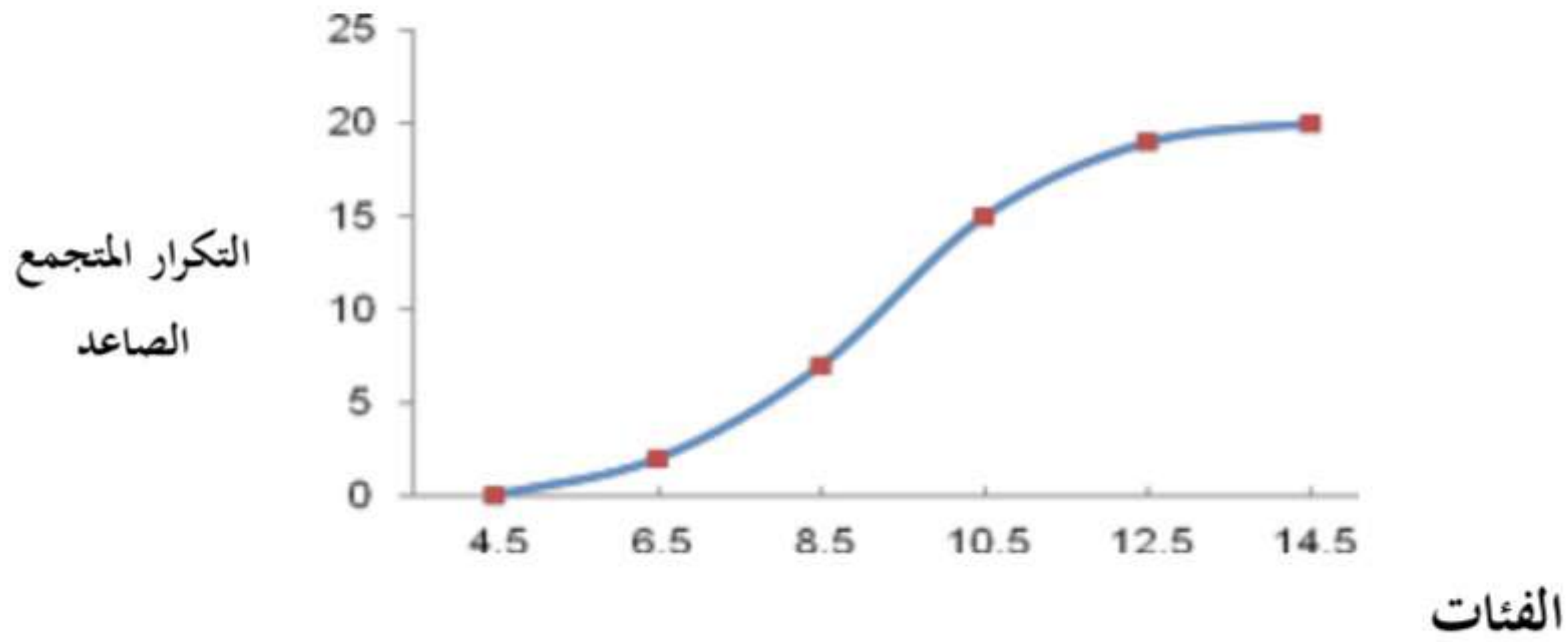
٤. طول الفئة L

ثم نطبق القانون:

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

إيجاد الوسيط بيانياً:

نوجد الوسيط بيانياً وذلك من خلال منحنى التكرار المتجمع الصاعد كما في الرسم التالي والذي يوضح طريقة إيجاد الوسيط للبيانات الواردة في مثال (4).



مميزات الوسيط:

١. لا يتأثر بالقيم المتطرفة.

٢. يمكن إيجاده في حالة البيانات الوصفية ذات العدد الفردي.

عيوب الوسيط:

١. لا يأخذ جميع القيم في الاعتبار.

٢. يصعب التعامل معه في التحاليل الإحصائية والرياضية.

ج. المنوال (Mode)

أولاً- للبيانات غير المبوبة (non- grouped data):

هو القيمة الأعلى تكراراً في مجموعة البيانات ويرمز له بالرمز Mod.. وهنا نميز بين ثلاث حالات:

١. عندما يكون لمجموعة البيانات منوال واحد تسمى وحيدة المنوال.

٢. عندما يكون لمجموعة البيانات أكثر من منوال تسمى متعددة المنوال.

٣. عندما لا يكون لمجموعة البيانات منوال واحد تسمى عديمة المنوال.

ثانياً- للبيانات المبوبة (grouped data):

الفئة المنوالية: هي الفئة المقابلة لأعلى تكرار.

خطوات إيجاد المنوال:

- نوجد أكبر تكرار f وعليه نحدد:

١. بداية الفئة المنوالية A

٢. التكرار السابق للفئة المنوالية f_1

٣. التكرار اللاحق للفئة المنوالية f_2

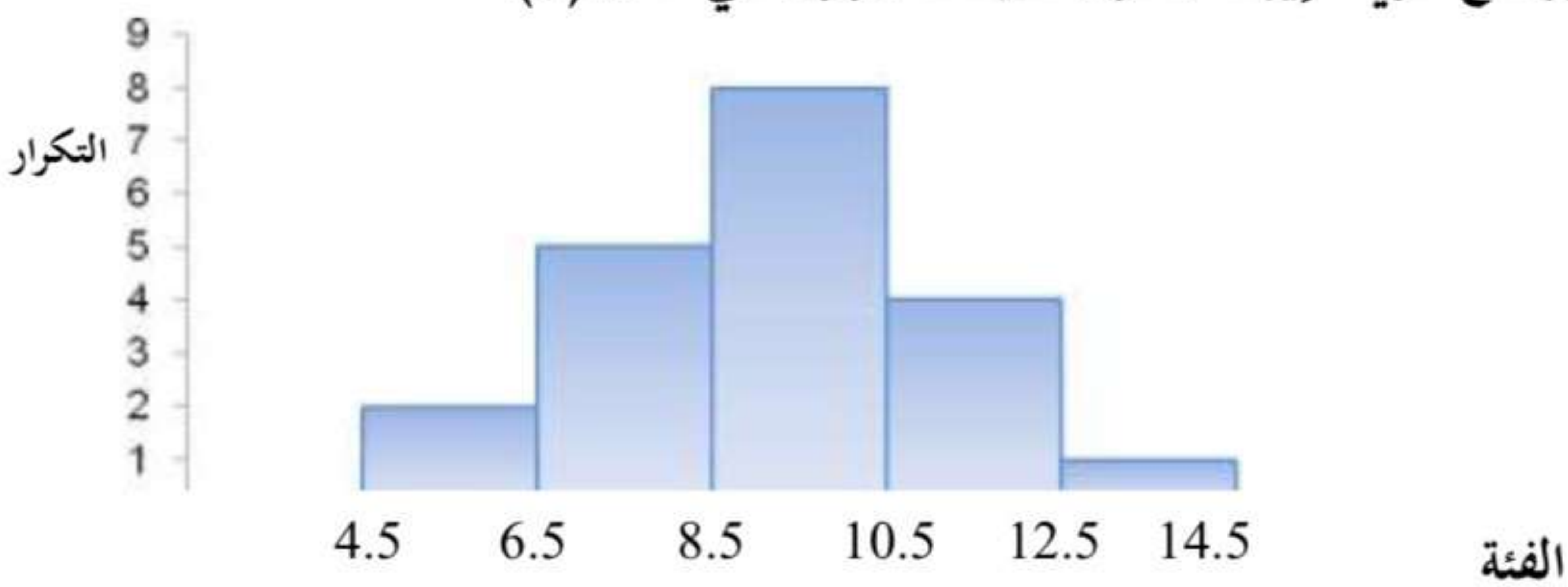
٤. طول الفئة L

ثم نطبق القانون:

$$Mod = A + \frac{f - f_1}{2f - f_1 - f_2} L$$

إيجاد المنوال بيانياً:

نوجد المنوال بيانياً وذلك من خلال المدرج التكراري كما في الرسم التالي والذي يوضح طريقة إيجاد المنوال للبيانات الواردة في مثال (6).



مميزات المنوال:

١. سهل الحساب ولا يتأثر بالقيم المتطرفة.

٢. يمكن إيجاده في حالة البيانات الوصفية ذات العدد الفردي.

عيوب الوسيط:

١. لا يأخذ جميع القيم في الاعتبار.

٢. قد يكون لبعض البيانات أكثر من منوال، وبالعكس قد لا يوجد منوال.

العلاقة بين الوسط الحسابي والوسيط والمنوال:

$$\bar{x} - Med = \frac{\bar{x} - Mod}{3}$$

مقاييس التشتت MEASURES OF DISPERSION

عند مقارنة مجموعتين من البيانات ، يمكن استخدام شكل التوزيع التكراري، أو المنحنى التكراري ، وكذلك بعض مقاييس الترعة المركزية ، مثل الوسط الحسابي والوسيط ، والمنوال ، والإحصاءات الترتيبية ، ولكن استخدام هذه الطرق وحدها لا يكفي عند المقارنة ، فقد يكون مقياس الترعة المركزية للمجموعتين متساوي ، وربما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض ، أو مدى تباعد أو تقارب القيم عن مقياس الترعة المركزية.

ومثال على ذلك ، إذا كان لدينا مجموعتين من الطلاب ، وكان درجات المجموعتين كالتالي:

المجموعة الأولى	63 70 78 81 85 67 88
المجموعة الثانية	73 78 77 78 75 74 77

لو قمنا بحساب الوسط الحسابي لكل مجموعة ، نجد أن الوسط الحسابي لكل منهما يساوي76

درجة ، ومع ذلك درجات المجموعة الثانية أكثر تجانسا من درجات المجموعة الأولى

مثال أوجد الوسط الحسابي والوسيط لدرجات مجموعة من الطلاب في مادتي الرياضيات واللغة الإنجليزية :

درجات الرياضيات : ٤٠ ، ٥٨ ، ٧٢ ، ٨٠ ، ١٠٠

درجات اللغة الإنجليزية : ٦٤ ، ٦٧ ، ٧٢ ، ٧١ ، ٧٦

نجد أن : الوسط الحسابي لكل من الظاهرتين هو ٧٠ درجة وأن الوسيط لكل منهما هو ٧١ درجة

فإذا اكتفينا بمقارنة الوسط والوسيط نجد أن مستوى الطلاب هو نفسه في المادتين وهذا يخالف الواقع حيث إن درجات اللغة الإنجليزية متقاربة من بعضها وتتركز حول وسطها (مثلا) بينما درجات الرياضيات متباعدة ومبعثرة في مدى كبير، وعلى ذلك لا يمكننا اقتصار المقارنة بين الظواهر على متوسطاتها فقط ، بل يجب البحث عن مقياس آخر يبين مدى تقارب أو تباعد مفردات الظواهر بعضها عن بعض ..

من أجل ذلك لجأ الإحصائيين إلى استخدام مقاييس أخرى لقياس مدى تجانس البيانات، أو مدى انتشار البيانات حول مقياس النزعة المركزية، ويمكن استخدامها في المقارنة بين مجموعتين أو أكثر من البيانات، ومن هذه المقاييس: مقاييس التشتت ، والالتواء ، والتفرطح ، وسوف نستعرض في هذا البند بعض هذه المقاييس.

مقاييس التشتت MEASURES OF DISPERSION

- الانحراف المتوسط (Mean Deviation Absolute)
- الانحـراف المعيـاري والتبـاين (Standard Deviation and Variance)
- المدى (Range)
- نصف المدى الربيعي أو الانحراف الربيعي (Interquartile Range)
- التشتت المطلق أو النسبي ومعامل الاختلاف (Coefficient of Variation) و C.V.

١-الانحراف المتوسط (Absolute Mean Deviation)

الانحـراف المتوسـط أو متوسـط الانحرافـات لمجموعـة n مـن الأرقـام $X_1, X_2, \ldots, X_n$ يعرف بما يلي

$$M.D = \frac{\sum_{i=1}^{n} \left| X_i - \overline{X} \right|}{n} = \frac{\sum \left| X - \overline{X} \right|}{n}$$

حيـث $\overline{X}$ هـو الوسـط الحسـابي للأرقـام و $\left| X_i - \overline{X} \right|$ هـو القـيم المطلقـة لانحرافات القيم X_i عن $\overline{X}$.

القيمـة المطلقـة Absolute Deviation لـرقم هـو الـرقم بـدون الإشـارة المرافقة له ويعبر عن ذلك بخطين رأسيين يوضعان حول الرقم، وعلى ذلك فإن :

$$|-4| = 4, |+3| = 3, |6| = 6, |-0.84| = 0.84$$

مثـال : أوجد متوسط الانحرافات لمجموعة الأرقام ٢,٣,٦,٨,١١

$$\overline{X} = \frac{2+3+6+8+11}{5} = 6$$

$$M.D. = \frac{|2-6| + |3-6| + |6-6| + |8-6| + |11-6|}{5}$$

$$= \frac{|-4| + |-3| + |0| + |2| + |5|}{5}$$

$$= \frac{4+3+0+2+5}{}$$

$$= \frac{14}{5} = 2.8$$

إذا كانت $X_1, X_2, \dots, X_k$ تحدث بتكرارات $f_1, f_2, \dots, f_k$ على الترتيب فإن الانحراف المتوسط يمكن كتابته على صورة

$$M.D. = \frac{\sum_{i=1}^{n} f|X_i - \overline{X}|}{n} = \frac{\sum f|X - \overline{X}|}{n}$$

حيث أن $n = \sum_{i=1}^{k} f_i = \Sigma f$ حيث ان $X\,{}'s$ مراكز الفئات $f\,{}'s$

تمثل التكرارات المقابلة لها وهذه الصيغة مفيدة للبيانات المبوبة. في بعض الأحيان يعرف الانحراف المتوسط بدلالة القيمة المطلقة للانحرافات عن الوسيط أو غيره من المتوسطات بدلاً من الوسط الحسابي .

ملاحظــــــــــة : المجموع $\sum_{i=1}^{n} |X_i - a|$ يكون أقل ما يمكن عندما تكون a هي الوسيط

، بمعنى أن متوسط انحرافات القيم عن الوسيط يكون أقل ما يمكن .

لاحظ أنه قد يكون من الأنسب استخدام التعبير متوسط القيم المطلقة للانحرافات MAD للتعبير عن الانحراف المتوسط .

٢-الانحـراف المعيـاري والتبـاين (Standard Deviation and Variance)

- الانحراف المعياري (Deviation Standard)
- التباين (Variance)

الانحراف المعياري

الانحراف المعياري لمجموعة من n رقم $X_1, X_2, ..., X_n$ يعبر عنها بالرمز S

وعلى هذا فان S هي الجذر ألتربيعي لمتوسط مربعات انحرافات القيم عن وسطها الحسابي ويسمى أحياناً جذر متوسط مربع الانحراف .

بالنسبة للبيانات غير المبوبة،

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} \quad \text{and} \quad s = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}}$$

و بالنسبة للبيانات المبوبة،

$$\sigma = \sqrt{\frac{\sum f(X-\mu)^2}{N}} \quad \text{and} \quad s = \sqrt{\frac{\sum f(X-\bar{X})^2}{n-1}}$$

ويعتبر الانحراف المعياري أكثر مقاييس التشتت المطلق شيوعاً في الاستخدام

التباين (Variance)

تباين مجموعة من البيانات يعرف بأنه مربع الانحراف المعياري وبهذا يعرف بـ S^2 ويجب التمييز بين الانحراف المعياري للمجتمع والانحراف المعياري لعينة مسحوبة من هذا المجتمع ، لذا فإننا نستخدم دائماً الرمز S للأخير والرمز σ للأول . وبهذا فإن σ^2, S^2 يمثلان تباين العينة وتباين المجتمع على الترتيب.

ملاحظة : الانحراف المعياري للمجتمع σ للعينة S هما الجذر التربيعي الموجب للتباين المناظر لكل منهما.

تباين المجتمع $\sigma 2$ (الحرف اليوناني سيجما تربيع) وتباين العينة S2 للبيانات غير المبوبة تحسب كآلاتي:

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} \quad \text{and} \quad s^2 = \frac{\sum (X-\bar{X})^2}{n-1}$$

وللبيانات المبوبة ،

$$s^2 = \frac{\sum f(X-\bar{X})^2}{n-1} \quad \text{and} \quad \sigma^2 = \frac{\sum f(X-\mu)^2}{N}$$

بصيغة اخرى

$$S^2 = \frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n-1} \quad \text{and} \quad S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{\sum X^2 - n\bar{X}^2}{n-1}$$

مثال التباين والانحراف المعياري لدرجات ٥٠ طالب

التكرار	الدرجات
3	49.5-59.5
5	59.5-69.5
18	69.5-79.5
16	79.5-89.5
8	89.5-99.5
50	المجموع

$$S^2 = \frac{\sum fX_i^2 - \frac{(\sum fX_i)^2}{n}}{n-1}$$

$$S^2 = \frac{315302.35 - \frac{(3935)^2}{50}}{50-1} = \frac{315302.35 - 309684.5}{49} = 114.65$$

$$S = \sqrt{(114.65)} = 10.71$$

٣- المـــــــدى

هو أبسط مقاييس التشتت ، ويحسب المدى في حالة البيانات غير المبوبة بتطبيق المعادلة التالية.

المدى = اكبر قراءة –اقل قراءة

وأما المدى في حالة البيانات المبوبة له أكثر من صيغة، ومنها المعادلة التالية:

المدى = مركز الفئة الاكبر – مركز الفئة الاقل

مثـــال

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الأسمدة الفسفورية ،

وفيما يلي بيانات كمية الإنتاج من القمح بالطن /هكتار.

4.63 5.08 5.18 5.29 5.18 5.4 6.21 4.8 5.03

والمطلوب حساب المدى.

الحل

المدى = أكبر قراءة – أقل قراءة

أكبر قراءة =6.21 أقل قراءة = 4.63

إذا المدى هو:

Rang=Max−Min

=6.21−4.63

=1.58

المدى يساوي 1.58 طن / هكتار

ومن عيوبه

١- أنه يعتمد على قيمتين فقط ، ولا يأخذ جميع القيم في الحسبان .

يتأثر بالقيم الشاذة

معامل الارتباط Correlation:

معامل الارتباط هو رقم يتراوح بين -١ و١ وهو يبين وجود علاقة خطية بين متغيرين واتجاه تلك العلاقة كما يلي:

+١ تعنى علاقة طردية بمعنى أنه كلما زاد أ زاد ب وكلما قل أ فإن ب يقل

-١ تعني علاقة عكسية بمعنى انه كلما زاد أ فإن ب يقل وكلما قل أ فإن ب يزيد

صفر يعني عدم وجود أي علاقة بين المتغيرين

عندما يقترب معامل الارتباط من إحدى هذه القيم فإنه يدل على ما تدل عليه هذه القيم ولكن بدرجة أقل. فمثلا +٠.٩ تدل على وجود علاقة طردية قوية بين المتغيرين ولكنها ليست مطلقة مثل تلك التي تتوقعها عندما يكون معامل الارتباط يساوي+١.

يسمى معامل الارتباط بمعامل الارتباط لبيرسون Pearson Correlation Coefficient ويشيع تسميته بمعامل الارتباط. ولمعامل الارتباط تطبيقات عديدة فمثلا في مجال التسويق قد تحب أن تدرس إن كان هناك علاقة بين زيادة مبيعات منتجك وزيادة مبيعات سلعة أخرى أو تحسن درجة الحرارة أو تخفيض السعر. وقد تكون مهندسا يريد أن يعرف ما الذي يؤثر على جودة الغاز المنتج هل هو تغير الضغط أم الحرارة أم جودة أي غاز من الغازات الداخلة في العملية الإنتاجية.

طريقة الحساب:

معامل الارتباط يتم حسابه بسهولة عن طريق الحاسوب ولذلك فلسنا بحاجة للدخول في حسابات مملة ولكن من الضروري أن نلقي نظرة على طريقة الحساب لنفهم معنى معامل الارتباط. يتم حساب معامل الارتباط كالتالي

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)S_x S_y}$$

والبسط في هذه المعادلة هو مجموع حاصل ضرب الفارق بين كل قيمة للمتغير الأول ومتوسطه الحسابي في الفارق بين كل قيمة للمتغير الثاني ومتوسطه الحسابي. والمقام هو حاصل ضرب الانحراف المعياري لكل من المتغيرين في عدد البيانات منقوصا منها واحد. هذا في حال أن لدينا عينة من البيانات كأن نأخذ عينة عشوائية من مجموعة كبيرة (المجتمع) وندرس ظاهرة

معينة على هذه العينة. اما عند دراسة المجتمع كله فإن طريقة الحساب تختلف اختلافا طفيفا وتكون كالتالي :

$$\rho = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n)\sigma_x\sigma_y}$$

في هذه الحالة فإن المقام يكون حاصل ضرب الانحراف المعياري للمجتمع لكل من المتغيرين مضروبا في عدد البيانات.

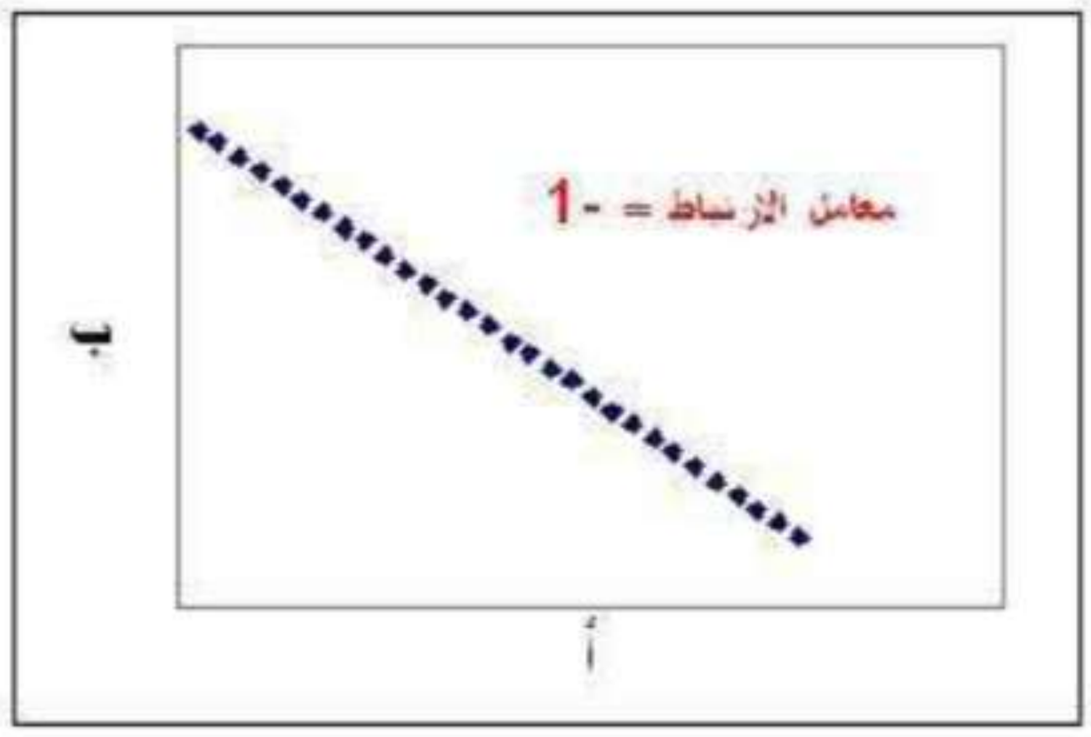
ماذا نفهم من هذه المعادلة المعقدة؟

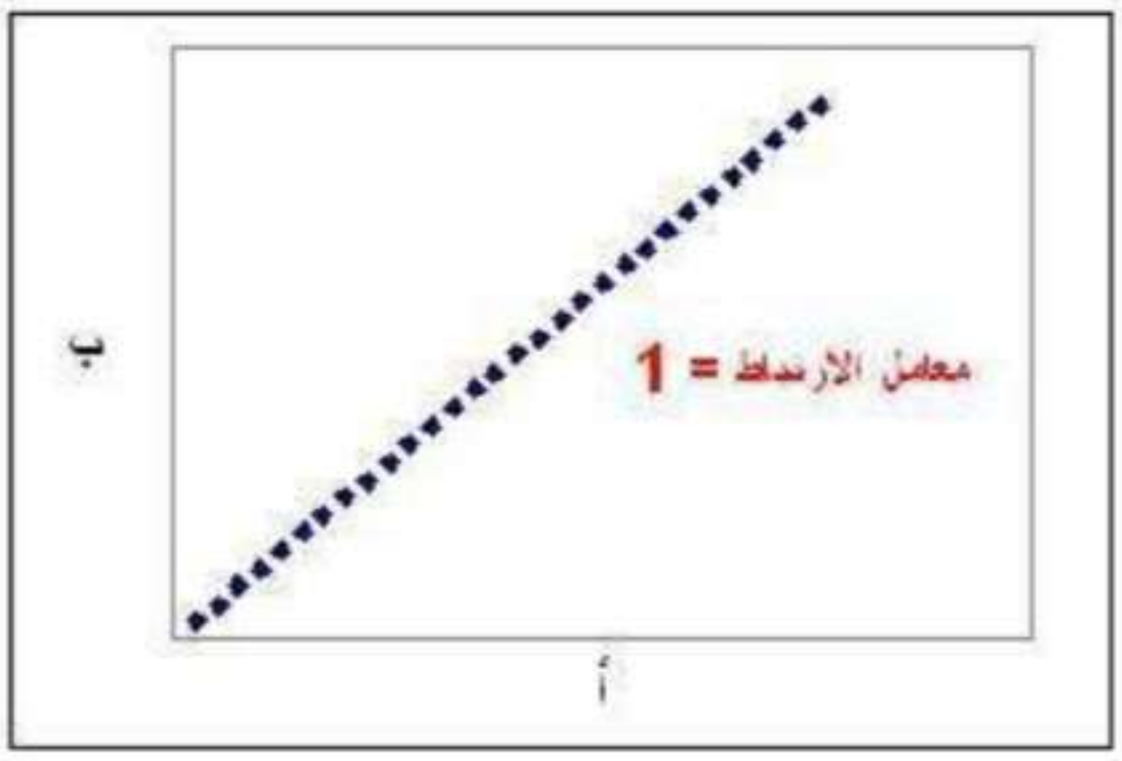
أولا المقام هو حاصل ضرب أرقام موجبة (أكبر من الصفر) فالانحراف المعياري هو دائما رقما موجبا وكذلك عدد البيانات. فمتى يكون معامل الارتباط موجبا ومتى يكون سالبا؟ الأمر يتوقف على البسط. فإذا كان الفارق بين قيمة ما للمتغير الأول ومتوسطه الحسابي موجبا وكان الفارق بين القيمة المقابلة والمتوسط الحسابي للمتغير الثاني موجبا كانت النتيجة موجبة لأن حاصل ضرب قيمة موجبة في قيمة موجبة يساوي قيمة موجبة. وإذا كان كل منهما سالبا فإن الناتج يكون موجيا لأن حاصل ضرب قيمة سالبة في قيمة سالبة يساوي قيمة موجبة. ومعنى ذلك (في الحالة الأولى) أنه عند زيادة المتغير الأول عن متوسطه الحسابي فإن المتغير الثاني يزيد عن متوسطه الحسابي هو الآخر وكذلك (في الحالة الثانية) عند نقصان المتغير الأول عن متوسطه الحسابي فإن نفس الأمر يحدث للمتغير الثاني.

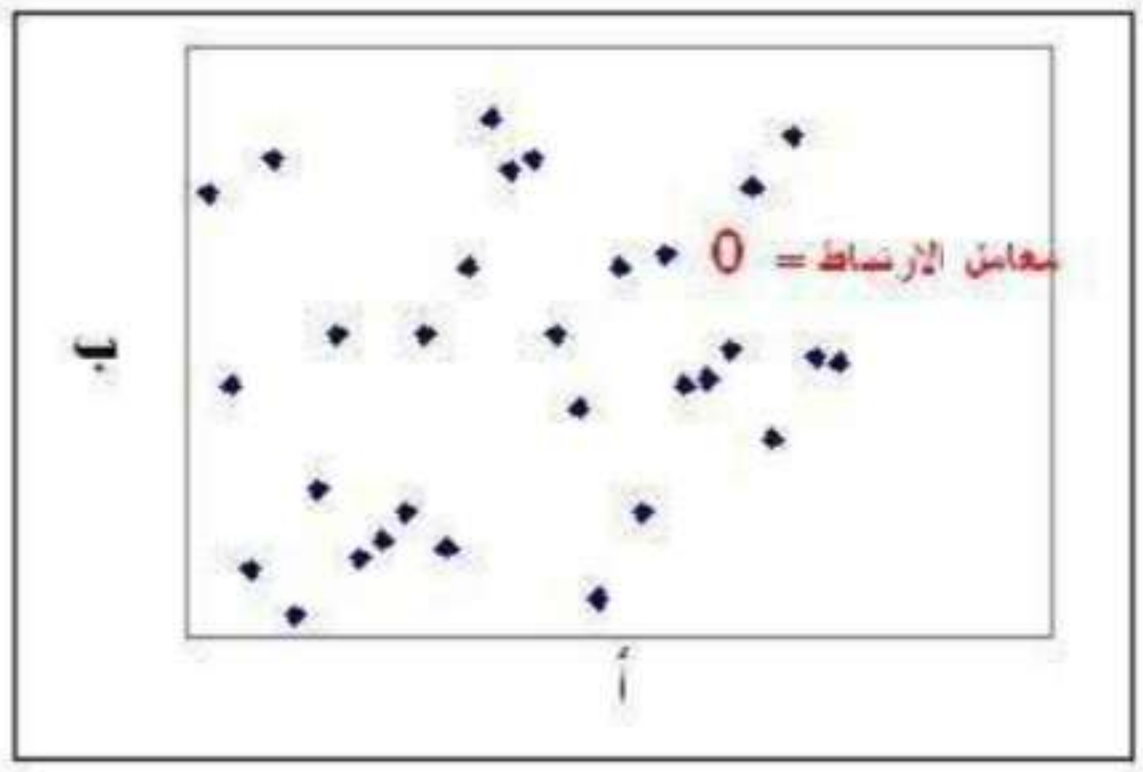
وبالتالي فإنه عندما تكون العلاقة عكسية فإن الناتج يكون سالبا لأن أحد الفارقين سيكون موجبا والآخر سالبا. وهذا يجعلنا نفهم القاعدة بأن معامل الارتباط كلما كان أقرب للواحد الصحيح فإن ذلك يعني وجود علاقة طردية قوية وكلما اقترب من −١ فإن ذلك يعني وجود علاقة عكسية قوية. وكلما اقترب من الصفر فإن ذلك يعني عدم وجود علاقة خطية.

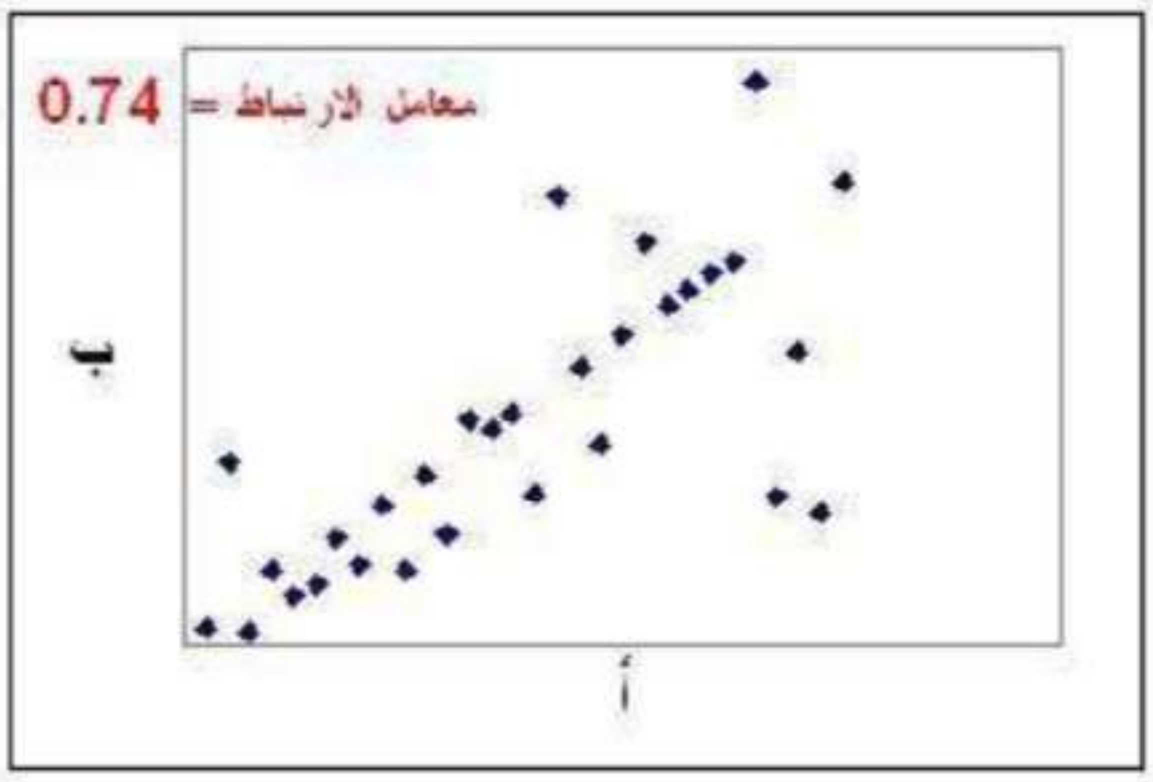
شكل العلاقة:

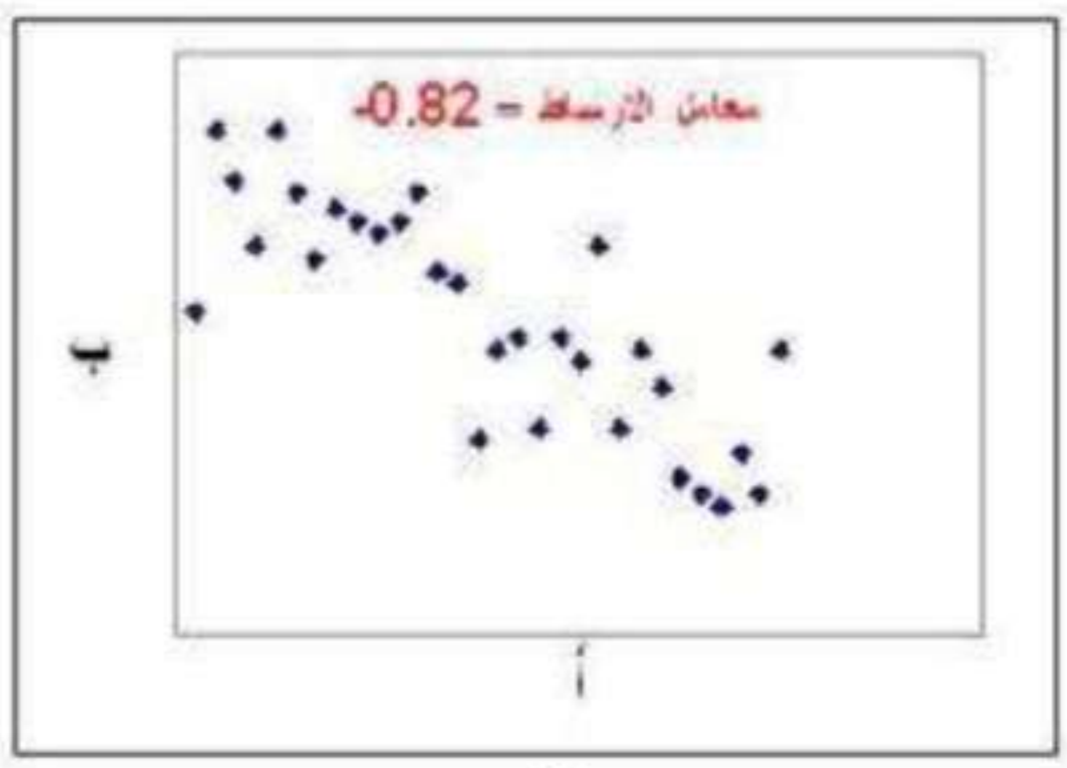
لننظر إلى بعض الرسومات البيانية المرادفة لقيم مختلفة لمعامل الارتباط لنتفهم ما يعنيه هذا الرقم.

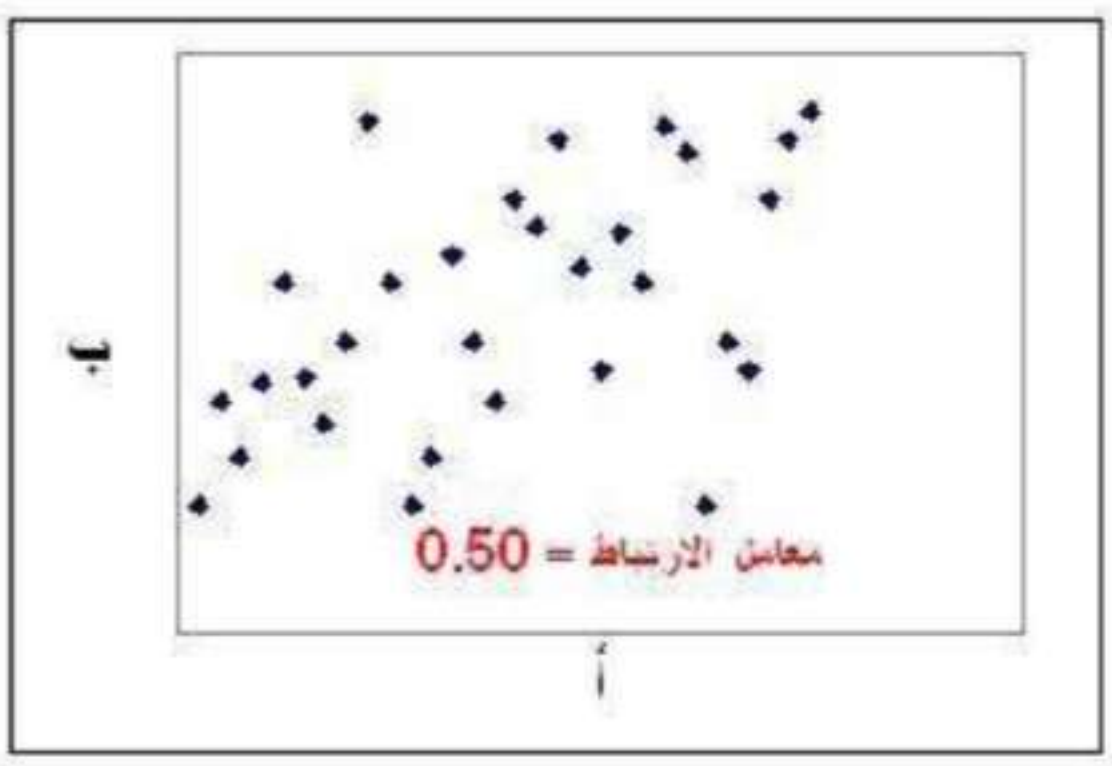






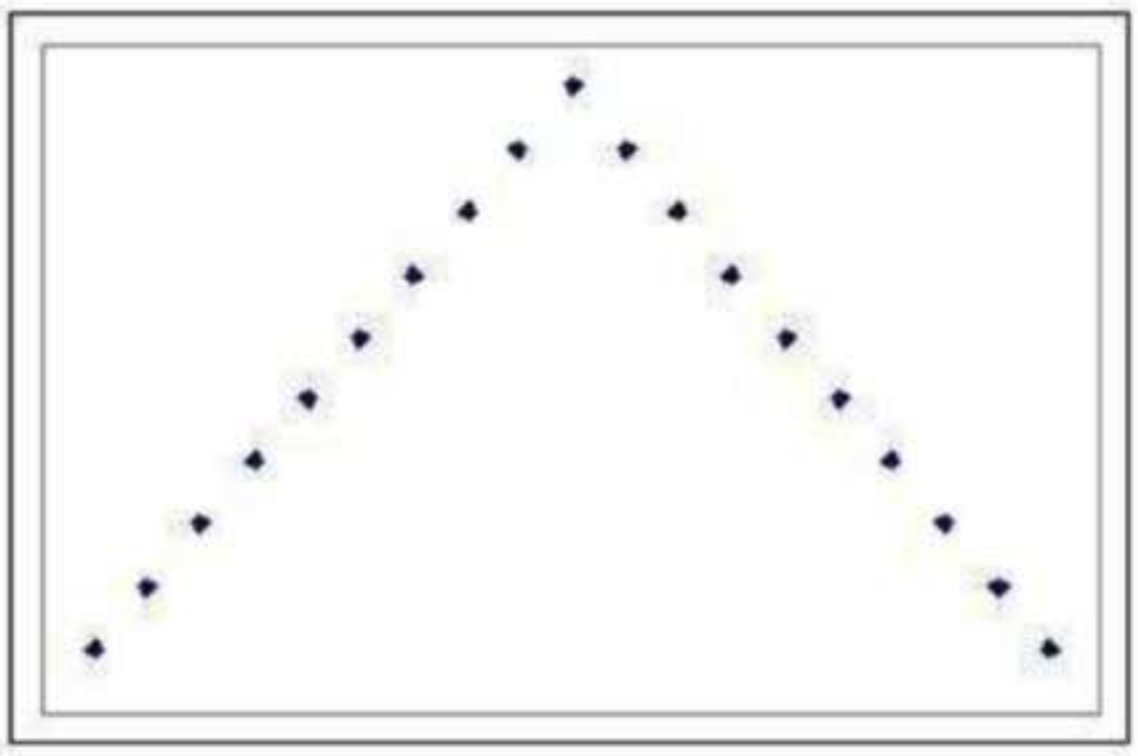


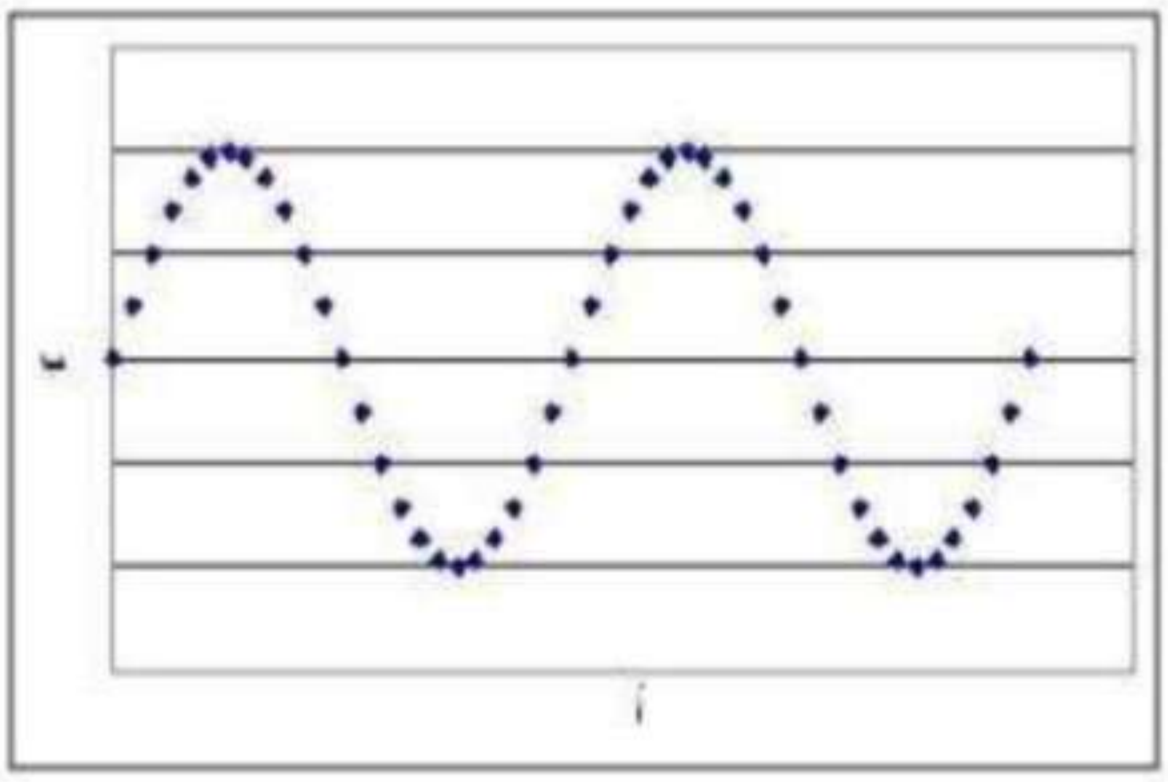




هل توجد علاقة؟

ليس معنى أن يكون معامل الارتباط صفرا أو قريبا من الصفر أنه لا توجد أي علاقة بين المتغيرين. فمعامل الارتباط يبين قوة العلاقة الخطية. والعلاقة الخطية هي علاقة في شكل خط مستقيم فهي علاقة ليس بها منحنيات أو طلوع ونزول. فالعلاقة الخطية تكون طردية أو عكسية فقط. وبالتالي فقد يكون معامل الارتباط يساوي صفرا ولكن توجد علاقة قوية بين المتغيرين ولكنها غير خطية أي أنها ليست على شكل خط مستقيم كما في الامثلة التالية:





ففي هذين الشكلين نرى علاقة واضحة بين المتغيرين ولكنها ليست مجرد علاقة طردية أو عكسية ولا يمكن تمثيلها بخط مستقيم. ففي الحالة الأولى نلاحظ تغير المتغير الثاني بشكل دوري مع المتغير الأول. وفي الحالة الثانية نجد علاقة طردية حتى نقطة ما ثم تتحول العلاقة إلى علاقة عكسية. هذه العلاقات هي علاقات غير خطية ولا يمكن التنبؤ بها بمعامل الارتباط.

بهذا نكون قد استطعنا دراسة شكل العلاقة عن طريق منحنى الانتشار (المنحنى التنقيطي) ومعرفة قوة العلاقة الخطية عن طريق معامل الارتباط. في المقالة التالية إن شاء الله نناقش كيفية الوصول لعلاقة رياضية بين متغير وكل المتغيرات التي تؤثر فيه.

معامل الارتباط الخطي Linear Correlation / بيرسون

بملاحظة المتغير العشوائي ذي البعدين (x , y) بوجود ارتباط أو علاقة بين y , x فإن الهدف من دراسة الارتباط هو قياس قوة الارتباط الخطي بين المتغيرين في حين معامل الارتباط الخطي (linear Coefficient) مقياس لقوة العلاقة الخطية بين x ، y ويقيس مدى تغير y حال زيادة قيمة x فهل y تزداد بزيادة x (ارتباط موجب) أو تنقص بزيادتها (ارتباط سالب) أو لا تتأثر بزيادة x (لا يوجد ارتباط).

معامل الارتباط لمجموعة n من الأزواج المرتبة

$(x_1 , y_1) , (x_2 , y_2) , \ldots , (x_n , y_n)$ هو:

$$r = \frac{\sum (X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum (X-\overline{X})^2}\sqrt{\sum (Y-\overline{Y})^2}}$$

$$r = \frac{\sum_{i=1}^{n} (X_i-\overline{X})(Y_i-\overline{Y})}{(n-1)S_x S_y} \quad \text{Or}$$

$$r = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{\sqrt{n(\sum X_i^2) - (\sum X_i)^2}\sqrt{n(\sum Y_i^2) - (\sum Y_i)^2}} \quad \text{Or}$$

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{\sum X_i^2 - n\overline{X}^2}\sqrt{\sum Y_i^2 - n\overline{Y}^2}} \quad \text{Or}$$

مع ملاحظة: ١) `X الوسط الحسابي للبيانات ١x ، ٢x ، ... ، xn
و `Y الوسط الحسابي للبيانات ١y ، ٢y ، ... ، yn

٢) Sx الانحراف المعياري للبيانات 1x ، 2x ، ... ، xn و Sy الانحراف المعياري للبيانات ١y ، ٢y ، ... ، yn

٣) r معامل الارتباط يعرف بمعامل ارتباط بيرسون (التتابعي أو العزومي) للارتباط (نسبة للعالم كارل بيرسون).

٤) يشترط عند حساب معامل الارتباط لبيرسون أن يكون التوزيع لكلا المتغيرين اعتدالي وأن تكون العينة عشوائية وقيم الفرد لا تعتمد على قيم فرد آخر (استقلالية أفراد العينة).

وفي حالة عدم اعتدالي المتغيرين نستخدم معامل ارتبط آخر سيذكر في حينه (معامل ارتبط سبيرمان أو كندال"تاو"

يعتبر معامل ارتباط بيرسون من أشهر الطرق لقياس معامل الارتباط بين متغيرين نسبيين أو فئويين فيما بينهم وتوجد عدة طرق لحساب معامل ارتباط بيرسون وهي:

- باستخدام الدرجات المعيارية $r = (\sum ZxZy)/n$
- طريقة الانحرافات
- طريقة التغاير (Sxy)
- من البيانات الأصلية
- باستخدام الحاسب الآلي وبرنامج SPSS.
- باستخدام الحاسب الآلي وبرنامج Minitab.

حيث Sxy يسمى معامل التغاير لحساب قوة العلاقة الخطية بين متغيرين ويمكن استخدام القانون $r = Sxy / SxSy$ لحسابه ويمكن استخدام القانون الآتي لطريقة الانحرافات. (راجع المثال للطرق الأربع) وإن مقدار التباين المشترك بين المتغيرين ينتج من تربيع قيمة معامل الارتباط والتباين المشترك Common Variance بين متغيرين يساوي مربع معامل الارتباط بينهم ويعرف بمعامل التحديد Coefficient of determination وهو مقدار التباين في أحد المتغيرين الممكن تحديده بمعرفة التباين في المتغير الآخر فإذا كان $r = 0.8$ فإن التباين يساوي ٠.٦٤ وهو ممكن تفسيره في حين الباقي ١– ٠.٦٤ = ٠.٣٦ جزء لا يمكن تفسيره ويعرف بالتباين العشوائي وهو يبين وجود متغيرات أخرى لم تحتسب أو لم يهتم بها ويسمى معامل عدم التحديد.

عوامل التحكم في معامل ارتباط بيرسون:

- أن تقع نقاط الأزواج (x , y) على خط مستقيم أو تكون قريبة جداً منه حتى تحقق صفة أن العلاقة خطية ($y = ax + b$) ويمكن ملاحظة ذلك من شكل الانتشار. إن لم تكن العلاقة خطية فستخدم معامل آخر.
- مقدار التباين فالعلاقة طردية بين الزيادة في التباين ومعامل الارتباط.
- دقة معامل الارتباط تتأثر بحجم العينة.
- شكل التوزيع وتماثله للمتغيرين يزيد من قيمة معامل الارتباط فإن كان شكلا التوزيع متماثلين فيكون $r = \pm 1$ وإن كانا الالتواء في نفس الاتجاه كان $r = 1$ وإن كان الالتواء في اتجاهين متضادين (احدهم التواءه موجب والآخر سالب) كان $r = -1$

من خصائص معامل الارتباط عدم اعتماده على القيم نفسها بل على تباعدها عن بعضها، لا تتغير قيمة معامل الارتباط بالعمليات الحسابية الأربع الجمع والطرح والقسمة والضرب مع عدد ثابت بالنسبة لقيم x , y .

ثانيا ـ الاحصاء الاستدلالي :

اختبار الفرضيات Test of Hypotheses

يعتبر موضوع اختبار الفرضيات الإحصائية من أهم الموضوعات في مجال اتخاذ القرارات وسنبدأ بذكر بعض المصطلحات الهامة في هذا المجال.

١- الفرضية الإحصائية

هي عبارة عن ادعاء قد يكون صحيحاً أو خطأ حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

تقبل الفرضية في حالة أن بيانات العينة تساند النظرية، وترفض عندما تكون بيانات العينة على النقيض منها، وفي حالة عدم رفضنا للفرضية الإحصائية فإن هذا ناتج عن عدم وجود أدلة كافية لرفضها من بيانات العينة ولذلك فإن عدم رفضنا لهذه الفرضية لا يعنى بالضرورة أنها صحيحة، أما إذا رفضنا الفرضية بناء على المعلومات الموجودة في بيانات العينة فهذا يعنى أن الفرضية خاطئة، ولذلك فإن الباحث يحاول أن يضع الفرضية بشكل يأمل أن يرفضها،.إن الفرضية التي يأمل الباحث أن يرفضها تسمى بفرضية العدم (الفرضية المبدئية) ويرمز لها بالرمز ، ورفضنا لهذه الفرضية يؤدى إلى قبول فرضية بديلة عنها تسمى الفرضية البديلة ويرمز لها بالرمز .

٢- مستوى المعنوية أو مستوى الاحتمال

وهي درجة الاحتمال الذي نرفض به فرضية العدم عندما تكون صحيحة أو هو احتمال الوقوع في الخطأ من النوع الأول ويرمز له بالرمز، وهي يحددها الباحث لنفسه منذ البداية وفي معظم العلوم التطبيقية نختار مساوية 1% أو 5 % على الأكثر.

٣- دالة الاختبار الإحصائية

عبارة عن متغير عشوائي له توزيع احتمالي معلوم وتصف الدالة الإحصائية العلاقة بين القيم النظرية للمجتمع والقيم المحسوبة من العينة.

٤. **القيمة الاحتمالية or . . : (Sig P-value)**

احتمال الحصول على قيمة أكبر من أو تساوي (أقل من أو تساوي) إحصائية الاختبار المحسوبة من بيانات العينة أخذاً في الاعتبار توزيع إحصائية الاختبار بافتراض صحة فرض العدم وطبيعة الفرض البديل . ويتم استخدام القيمة الاحتمالية لاتخاذ قرار حيال فرض العدم.

خطوات اختبار الفرضيات

(١) تحديد نوع توزيع المجتمع

يجب تحديد ما إذا كان المتغير العشوائي الذي يتم دراسته يتبع التوزيع الطبيعي أم توزيع بواسون أم توزيع ذو الحدين أم غيره من التوزيعات الاحتمالية المتصلة أو المنفصلة، معظم التوزيعات الاحتمالية يكون توزيعها مشابهاً للتوزيع الطبيعي خاصة إذا كان حجم العينة كبيراً.

هناك نوعان من الطرق الإحصائية التي تستخدم في اختبار الفرضيات:

(أ) الاختبارات المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها يتبع التوزيع الطبيعي.

(ب) الاختبارات غير المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها لا يتبع التوزيع الطبيعي طبيعي، وكذلك في حالتي البيانات الترتيبية والوصفية.

٢- صياغة فرضيتا العدم والبديلة

مثلاً: عند اختبار أن متوسط المجتمع μ يساوى قيمة معينة μ_0 مقابل الفرضية القائلة بأن μ لا يساوى μ_0، فإن فرضية العدم H_0 والفرضية البديلة H_1 تكون على النحو التالي:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

٣- اختيار مستوى المعنوية α

٤- اختيار دالة الاختبار الإحصائية المناسبة

٥- جمع البيانات من العينة وحساب قيمة دالة الاختبار الإحصائية

٦- اتخاذ القرارات

نرفض H_0 ونقبل H_1 إذا كانت قيمة الاحتمال (Sig. or P-value) أقل من أو تساوي مستوى المعنوية (α)، أما إذا كانت قيمة الاحتمال أكبر من α فلا يمكن رفض H_0.

الاختبار التائي - t -

مقدمة :

يعد اختبار " t " من أكثر اختبارات الدلالة شيوعاً في الأبحاث النفسية والاجتماعية والتربوية ، وترجع نشأته الأولى إلى أبحاث العالم "ستودنت" ولهذا سمى الاختبار بأكثر الحروف تكراراً في اسمه وهو حرف التاء .

ومن أهم المجالات التي يستخدم فيها هذا الاختبار الكشف عن الفروق بين تحصيل الذكور والإناث فى مادة دراسية ما وذلك عن طريق حساب دلالة فرق متوسط تحصيل الذكور عن متوسط تحصيل الإناث .

ويمكن القول أن اختبار " t " يستخدم لقياس دلالة فروق المتوسطات غير المرتبطة والمرتبطة للعينات المتساوية والغير متساوية .

شروط استخدام اختبار ت لدلالة فروق المتوسطات

لا يحق للباحث أن يستخدم اختبار " t " قبل أن يدرس خصائص متغيرات البحث من النواحي التالية :-

١- حجم كل عينة .

٢- الفرق بين حجم عينتي البحث .

٣- مدى تجانس العينة .

٤- مدى اعتدالية التوزيع التكرارى لكل من عينتى البحث .

١- حجم كل عينة

يجب أن يزيد حجم كل من العينتين عن "5" ويفضل أن يزيد عن "30" أما إذا قل حجم أى من العينتين عن "5" فلا يمكن استخدام اختبار "ت" .

٢- الفرق بين حجم عينتى البحث : شرط التقارب

يجب أن يكون حجم عينتى البحث متقارباً فلا يكون مثلاً حجم أحد العينتين "٥٠٠" وحجم الأخرى "٣٠" لأن للحجم أثره على مستوى دلالة " t " .

٣- مدى تجانس العينتين

يقصد بتجانس العينات مدى انتسابها إلى أصل واحد أو أصول متعددة . فإذا انتسبت العينات إلى أصل واحد فهى متجانسة وإذا لم تنتسب العينات إلى أصل واحد فهى غير متجانسة .

وبالطبع يصعب بالنسبة للباحث تحديد أصول العينات لتحديد تجانسها لذا يمكنه استخدام النسبة الفائية لتحديد التجانس .

يحدد تجانس العينتين من خلال حساب قيمة النسبة الفائية حيث تحسب من العلاقة :

$$F = \frac{S_1^2}{S_2^2}$$

حيث أن التباين الأكبر هو التباين الأكبر فى القيمة دون التحيز لأحد العينتين ، والتباين الأصغر هو الأصغر فى القيمة دون التحيز لأحد العينتين .

بالطبع نحصل من القانون السابق على قيمة لـ "ف" تسمى بقيمة ف المحسوبة ولتحديد التجانس نحسب قيمة أخرى تسمى " F " الجدولية ونحصل عليها من جداول "F" الإحصائية عند درجة حرية التباين الأكبر ودرجة حرية التباين الأصغر ومستوى الدلالة الذى قيمته إما "0.05" أو "0.01" حيث نحسب درجات الحرية من القانون التالى :

درجة حرية التباين الأصغر = n -1

حيث "ن" هى عدد أفراد العينة التى تبيانها هو الأكبر .

درجة حرية التباين الأصغر = n-1

حيث "ن" هى عدد أفراد العينة التى تبيانها هو الأصغر .

تحديد التجانس

إذا كانت قيمة "F" المحسوبة > قيمة "F" الجدولية فلا يوجد هناك تجانس .

أما إذا كانت قيمة "F" المحسوبة < قيمة "F" الجدولية فيوجد هناك تجانس .

٤- مدى اعتدالية التوزيع التكرارى لكل من العينتين

يكون التوزيع التكرارى معتدلاً عندما تكون قيمة الالتواء الخاص به محصورة بين القيمتين [3 - , 3+] أى واقعة فى الفترة المغلقة 3 - ، 3+

ويحسب الالتواء من القانون التالى :-

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

حيث:"

$\bar{X}$ هو المتوسط الحسابى ويحسب من العلاقة

$$\bar{X} = \frac{\sum X}{n}$$

حيث : "$\sum X$" هي مجموع القيم ، X هي القيم ، n هي عدد القيم .

"M" هو الوسيط ، ويحسب عن طريق ترتيب القيم تصاعدياً أو تنازلياً ثم اختيار قيمة الوسيط فى حالة أن يكون عدد الأفراد فردياً تكون قيمة الوسيط التى ترتيبها

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

أما إذا كان عدد الأفراد زوجياً فتكون قيمة الوسيط هى متوسط القيمتين اللتان ترتيبهما.

$$\frac{n}{2} \quad , \quad \frac{n}{2} + 1$$

"S" هو التباين ، ويحسب من العلاقة :

$$S^2 = \frac{\sum y^2}{n}$$

من الواضح أن القانون السابق يحسب قيمة التباين فنأخذ للقيمة الناتجة الجذر التربيعى لنحصل على الانحراف المعيارى كالتالى .

$$S = \sqrt{\frac{\sum y^2}{n}}$$

حيث :

S = الانحراف المعيارى

$$X - \bar{M} = y$$

n = عدد القيم

تحديد مدى دلالة "t" من عدمه

سنحصل فى جميع حالات "t" على قيمة لـ " t " نسميها "t المحسوبة" ثم نقارنها بقيمة لـ "t" نحصل عليها من الجداول تسمى " t الجدولية"

- إذا كانت قيمة "" t المحسوبة" > قيمة " t الجدولية" تكون قيمة " t " دالة إحصائية.
- أما إذا كانت قيمة "t المحسوبة" < قيمة "t الجدولية" تكون قيمة "t" ليست دالة إحصائية .

الحالات المختلفة لحساب ت-

١ – **الحالة الأولى** : حساب "t" لدلالة فرق عينتين متجانستين غير متساويتين فى أعداد أفرادهما .

فى هذه الحالة تكون n1 لا تساوى n2حيث n1 ، n2 هما عدد أفراد العينة الأولى والثانية على الترتيب .

تحسب دلالة "t" لفرق عينتين متجانستين ومختلفين فى عدد الأفراد بالمعادلة التالية

$$t = \frac{x_1 - x_2}{\sqrt{\frac{S_1^2(n_1) + S_2^2(n_2)}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_1}\right)}}$$

حيث :

X_1 = المتوسط الحسابى للمجموعة الأولى .

X_2 = المتوسط الحسابى للمجموعة الثانية .

S_1^2= تباين المجموعة الأولى .

S_2^2= تباين المجموعة الثانية .

n_1 = عدد أفراد المجموعة الأولى .

n_2 = عدد أفراد المجموعة الثانية .

مثال :

العينة الأولى	7	4	5	3	8	6	2
العينة الثانية	3	5	15	2	10	13	-

الجدول السابق يوضح درجات مجموعة من الذكور والإناث في اختبار للذكاء والمطلوب حساب قيمة "t "دالة أم لا ؟ عند مستوى دلالة إحصائية 0.05

الحل :

قبل أن نبدأ الحل نلاحظ أن

نعتبر أن العينة الأولى هى "x1" والعينة الثانية هى "x2" ونقوم ببناء الجدول التالى .

X1	x-x1	(x-x)	X2	x-x2	(x-x2)
7	2	4	3	-5	25
4	-1	1	5	-3	9
5	0	0	15	7	49
3	-2	4	2	-6	36
8	3	9	10	4	16
6	1	1	13	5	25
2	-3	9	-	-	-
35	-	28	48	-	148

$$t = \frac{8-5}{\sqrt{\frac{7\times 4+6\times 24.66}{7+6-2}\left(\frac{1}{7}+\frac{1}{6}\right)}}$$

t المحسوبة = - 1.36

تهمل الإشارة السالبة لقيمة "t" دائماً فتصبح :

قيمة "t" المحسوبة = 1.36 .

أن قيمة "t" الجدولية = 3.11.

تحديد دلالة "t"

بمقارنة قيمة "t" المحسوبة بقيمة "t" الجدولية :

نجد أن "t" المحسوبة = ١.٣٦ > "t" الجدولية = 3.11

وبالتالي فإن "t" ليست دالة إحصائية .

٢- **الحالة الثانية** : حساب "t" لدلالة فرق عينتين غير متجانستين وغير متساويتين فى أعداد أفرادهما

تحسب دلالة"t" لعينتين غير متجانستين ومختلفين فى عدد الأفراد بالمعادلة التالية:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1 +} + \frac{S_2^2}{n_2}}}$$

حيث :

$\bar{X}_1$ = المتوسط الحسابى للمجموعة الأولى .

$\bar{X}_2$ = المتوسط الحسابى للمجموعة الثانية .

S_1^2 = تباين المجموعة الأولى .

S_2^2 = تباين المجموعة الثانية .

n_1 = عدد أفراد المجموعة الأولى .

n_2 = عدد أفراد المجموعة الثانية .

مثال :

العينة الأولى	35	17	22	32	19	48	13	19	20
العينة الثانية	11	3	9	10	14	2	7	-	-

الجدول السابق يوضح درجات مجموعة من الذكور والإناث في اختبار للذكاء والمطلوب حساب قيمة "t" هل دالة إحصائية أم لا ؟ عند مستوى دلالة إحصائية ٠.٠٥ ؟

الحل :

قبل أن نبدأ الحل نلاحظ أن

نعتبر أن العينة الأولى هى "x1" والعينة الثانية هى "x2" ونقوم ببناء الجدول التالى .

(x-x2)	x-x2	X2	(x-x1)	x-x1	X1
9	3	11	100	10	35
25	-5	3	64	-8	17
1	1	9	9	-3	22
36	6	14	36	-6	19
36	-6	2	569	23	48
1	-1	7	144	-12	13
-	-	-	36	-6	19
-	-	-	25	-5	20
112	-	56	992	-	225

$$t = \frac{25 - 8}{\frac{110.2}{9} + \frac{16}{7}}$$

بالتعويض فى المعادلة السابقة :

t المحسوبة = 4.46

t الجدولية = 2.33

تحديد دلالة "t

بمقارنة قيمة "t" المحسوبة بقيمة "t" الجدولية

نجد أن "t" المحسوبة = 4.46 > "t" الجدولية = 2.33

وبالتالي فان "t" دالة إحصائية .

٣- **الحالة الثالثة** : حساب "t" لدلالة فرق عينتين غير مرتبطتين ومتساويتين فى أعداد أفرادهما

فى هذه الحالة لا نتحقق من شروط اختبار t" .

تحسب دلالة "t" لفرق عينتين متساويتين فى عدد الأفراد بالمعادلة التالية :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{n-1}}}$$

حيث :

$\bar{X}_1$ = المتوسط الحسابى للمجموعة الأولى .

$\bar{X}_2$ = المتوسط الحسابى للمجموعة الثانية .

S_1^2 = تباين المجموعة الأولى .

S_2^2 = تباين المجموعة الثانية .

n = عدد أفراد العينة الأولى أو الثانية حيث أنهما متساويتان .

مثال :

العينة الأولى	7	4	5	3	8	6	2
العينة الثانية	3	5	15	2	10	13	1

الجدول السابق يوضح درجات مجموعة من الذكور والإناث في اختبار للذكاء والمطلوب حساب قيمة "t" هل دالة إحصائية أم لا ؟ عند مستوى دلالة إحصائية 0.05 ؟

الحل:نعتبر أن العينة الأولى هى "x_1" والعينة الثانية هى "x_2" ونقوم ببناء الجدول التالى.

X1	x-x1	(x-x1)	X2	x-x2	(x-x2)
7	2	4	3	4-	16
4	1-	1	5	2-	4
5	0	0	15	8	64
3	2-	4	2	5-	25
8	3	9	10	3	9
6	1	1	13	6	36
2	3-	9	1	6-	36
35	-	28	49	-	190

حساب قيمة "t" المحسوبة :

$$t = \frac{7-5}{\frac{4+27.14}{7-1}}$$

بالتعويض فى المعادلة السابقة :

t المحسوبة = ‎-0.88

تهمل الإشارة السالبة لقيمة "ت" دائماً فتصبح :

قيمة "ت" المحسوبة =0.88.

نجد أن قيمة "t" الجدولية = 2.18 .

تحديد دلالة "t"

بمقارنة قيمة "t" المحسوبة بقيمة "t" الجدولية

نجد أن "t" المحسوبة = 0.88 > t" الجدولية = 2.18

وبالتالي فان "t" ليست دالة إحصائية .

٤ – **الحالة الرابعة** : حساب "t" لدلالة فرق عينتين مرتبطتين ومتساويتين فى أعداد أفرادهما

يرتبط المتوسطان عندما نجرى اختباراً على مجموعة من الأفراد ثم نعيد نفس الاختبار على نفس المجموعة فى وقت آخر أى أن العينة التى يجرى عليها الاختبار الأول هى نفسها العينة التى يجرى عليها الاختبار الثانى وفى هذه الحالة لا تكون n1 =n2 بل تصبح هى نفسها .

فى هذه الحالة أيضاً لا نتحقق من شروط اختبار "t" .

تحسب دلالة "t" لفرق عينتين متساويتين فى عدد الأفراد بالمعادلة التالية :

$$t = \frac{\sum MF}{\frac{\sum S^2 F}{n(n-1)}}$$

حيث :

MF = متوسط الفروق ويحسب من العلاقة :

F = الفروق = x2 – x1

X_1 هى درجات الاختبار الأول

X_2 هى درجات الاختبار الثانى

n عدد الأفراد فى أى من الاختبارين .

مثال :

درجات الاختبار الأول	26	18	20	24	22	14	23	16	22	11
درجات الاختبار الثانى	23	16	19	21	18	12	24	11	23	9

الجدول السابق يوضح درجات مجموعة من الأطفال في اختبار للذكاء حيث تم إجراء الاختبار مرة ثم بعد إجراء برنامج تدريبى لهم تم إجراء الاختبار مرة أخرى والمطلوب حساب قيمة "t" للفرق بين درجات الاختبارين ومن ثم تحديد هل "t" دالة إحصائية أم لا ؟ عند مستوى دلالة إحصائية 0.05 ؟

الحل :قبل أن نبدأ الحل نلاحظ أن

n_1 هى n_2 نعتبر أن درجات الاختبار الأول هى "x_1" ودرجات الاختبار الثانى هى "x_2" ثم نقوم ببناء الجدول التالى :

X1	X2	f	sf	Sf
26	23	3	1	1
18	16	2	0	0
20	19	1	1-	1
24	21	3	1	1
22	18	4	2	4
14	12	2	0	0
23	24	1-	3-	9
16	11	5	3	9
22	23	1-	3-	9
11	9	2	0	0
-	-	20	-	34

حساب قيمة "t" المحسوبة :

بالتعويض فى المعادلة السابقة :

$$\sqrt{\frac{\frac{2}{34}}{10(1-10)}}$$

t المحسوبة = 3.25

<u>حساب قيمة "ت" الجدولية :</u>

أن قيمة "t" الجدولية = 1.83 .

<u>تحديد دلالة "t"</u>

بمقارنة قيمة t" المحسوبة بقيمة "t" الجدولية

نجد أن "t" المحسوبة = 3.25 > "t" الجدولية = 1.83

وبالتالي فان "t" دالة إحصائية .

<u>اختبار مربع كاي x^2</u>

مقدمة :

ترجع النشأة الأولى لاختبار كا٢ إلى البحث الذي نشره كارل بيرسون فى أوائل القرن العشرين وهى تعد من أهم اختبارات الدلالة الإحصائية وأكثرها شيوعاً لأنها لا تعتمد على شكل التوزيع ولذا فهى تعد من المقاييس اللابارامترية أى مقاييس التوزيعات الحرة ولأنها تحسب لكل خلية من خلايا أى جدول تكرارى ثم تجميع القيم الجزئية للحصول على القيمة لـ x^2 .

وتستخدم كا٢ لحساب دلالة فروق التكرار أو البيانات العددية التي يمكن تحويلها إلى تكرار مثل النسب والاحتمال .

الطريقة العامة لحساب x^2

$$x^2 = \frac{\sum(T_0 - T_e)^2}{T_e}$$

حيث :

T_0: هو التكرار الواقعى الذي يحدث بالفعل والموجود بالجدول .

T_e: هو التكرار المتوقع حدوثه ويختلف حسابه باختلاف نوع الجدول المطلوب حساب x^2 منه .

تحديد مدى دلالة$x2$ من عدمه

فــى جميــع الحــالات نخــرج مــن الحســابات بقيمــة $x2$ المحســوبة نقارنهــا بقيمــة $x2$الجدولية كالتالى :

- إذا كانت$x2$ المحسوبة> x^2الجدولية فان x^2 تكون دالة إحصائية .
- إذا كانتx^2 المحسوبة < x^2 الجدولية فان x^2 ليست دالة إحصائية .

حالات حسابx^2 من الجداول المختلفة :

١ – الحالة الأولى : الطريقة العامة لحساب$x2$ من الجدول التكراري ١×٢ :

يتكـون الجـدول ١×٢ مـن صـف واحـد وعمـودين دون خلايـا المجمـوع إن وجـدت بالجدول .

ولحساب قيمة $x2$ فى هذا الجدول تحسب من القانون العام :

<u>مثال :</u>

<u>الجدول التالي يوضح آراء ٨٠ شخص فى استبيان دار حول رفض أو قبول قضية الزواج العرفي .</u>

الرأي	موافق	غير موافق	مج
التكرار	60	20	80

والمطلوب حساب قيمةx^2 مع بيان مدى دلالتها إحصائيا عند مستوى دلالة 0.05؟

<u>الحل :</u>

$$T_e = \frac{60+20}{20} = 40$$

$$x^2 = \frac{(60-40)^2}{40} + \frac{(20-40)^2}{40}$$

حساب التكرار المتوقع T_e :

كا٢ المحسوبة = 20 .

قيمة كا^٢ الجدولية =3.841.

تحديد مدى دلالة كا^٢ :

نقارن قيمة x^2 المحسوبة بقيمة x^2 الجدولية نجد أن :

قيمةx^2 المحسوبة = 20 > قيمةx^2 الجدولية = 3.841

لذا فانx^2 دالة إحصائية عند مستوى دلالة 0.05 .

٢- الحالة الثانية : الطريقة العامة لحساب x^2 من الجدول التكراري$1 \times n$:

مثال :الجدول التالي يوضح آراء 30 شخص فى استبيان دار حول قضية الزواج العرفي .

الرأي	موافق	لا أدرى	معارض	مج
التكرار	12	2	16	30

والمطلوب حساب قيمة كا٢مع بيان مدى دلالتها إحصائيا عند مستوى دلالة0.05؟

الحل :

حساب كا٢ المحسوبة :

$$x^2 = \frac{(12-10)^2 + (2-10)^2 + (16-10)^2}{10}$$

x2 المحسوبة = 10.4 .

قيمةx2 الجدولية = 5.991 .

تحديد مدى دلالة كا٢ :

نقارن قيمة x^2المحسوبة بقيمة x^2 الجدولية نجد أن

قيمةx^2 المحسوبة = 10.4 > قيمةx^2 الجدولية = 5.991

لذا فان x^2 دالة إحصائية عند مستوى دلالة 0.05 .