المحاضرة الاولى

المتغيرات Variables

يقصد بالمتغير "أي خاصية يمكن قياسها وتتباين قيمها من فرد إلى أخر أو من مجموعة إلى أخرى "، والبيانات الإحصائية التي يتعامل معها الباحث النفسي أو يقوم بجمعها ما هي إلا درجات أو مؤشرات لمقدار الشيء أو الصفة أو الخاصية موضوع القياس لدى الفرد.

أمثلة: متغير الجنس (ذكر، أنثى)، متغير الذكاء، متغير القلق

المتغيرات المستقلة والمتغيرات التابعة

المتغير المستقل هو المتغير الذي يخضع للتحكم والسيطرة وبتغير قيمه أو درجاته تتغير تبعاً لذلك قيم **المتغير التابع.**

فإذا كان هناك متغيرين بينهما علاقة معينة فيمكن التنبؤ بقيمة أحدهما ويعرف في هذه الحالة بالمتغير **التابع** إذا علمت قيمة الأخر وهو المتغير **المستقل**.

أمثلة:

- تأثير الذكاء على التحصيل الدراسي
- أثر التدريس باستخدام الفصول الافتراضية على تحصيل الطلاب في مقرر الاحصاء الاجتماعي

متغيرات مستقلة ومتغيرات مترابطة

عندما يكون لدينا مجموعة من القياسات التي ترتبط أو تؤثر في بعضها البعض يقال للمتغيرات في هذه الحالة **متغيرات مرتبطة** أما إذا كانت القياسات غير مترابطة ولا تؤثر في بعضها البعض فإن المتغيرات في هذه الحالة تكون **متغيرات مستقلة.**

أمثلة:

- إذا أردنا معرفة تأثير الذكاء على التحصيل فيمكن اعتبار الدرجات التي يحصل عليها الأفراد مستقلة ما دامت درجة الفرد لا ترتبط بدرجة غيره من الأفراد
- إذا أردنا معرفة الاختلاف بين تقدير الأم وتقدير الأب للعدوانية عن أطفالهم، فهنا يكون لكل طفل درجتين في العدوانية إحداهما تقدير الأب والأخرى تقدير الأم وهنا يقال أن الدرجات مترابطة

طبيعة البيانات

البيانات الكيفية (النوعية):

هي المشاهدات أو الصفات التي لا يمكن قياسها مباشرة بوسائل القياس المألوفة كالعد والقياس(تكون في صورة غير عددية).

أمثلة: لون العين (أسود، أخضر، عسلي، أزرق)، الجنس (ذكر، أنثى)، تقديرات الطلاب (ممتاز، جيد جداً، جيد، مقبول)، الجنسية (مصري، سعودي، ألماني).

البيانات الكمية (العددية):

هي المشاهدات أو الصفات التي يمكن قياسها مباشرة بوسائل القياس المألوفة (تكون في صورة عددية).

أمثلة: عدد طلاب التعليم الالكتروني، الطول، الوزن، عدد أفراد الأسرة.

أنواع البيانات الكمية

البيانات المنفصلة:

هي المشاهدات أو الصفات التي تأخذ قيماً متمايزة عن بعضها، مما يعني عدم اتصال البيانات، ولا تتضمن كسوراً.

أمثلة: عدد الطلاب الموزعين في كل تخصص أو شعبة أو فصل من فصول مدرسة.

البيانات المتصلة:

هي المشاهدات أو الصفات التي تأخذ مدى معين أو مجال معين من القيم ويمكن توزيعها على خط متصل بدون فواصل بينها لأنها تتضمن كسوراً. **أمثلة:** الطول، والوزن.

أساليب إجراء البحث

أسلوب الحصر الشامل:

يتم فيه جمع البيانات عن الظاهرة موضوع الدراسة من جميع مفردات المجتمع الإحصائي المراد بحثه سواء أكان نطاقه أو مجاله واسعاً أو محدوداً.

أسلوب العينات:

يتم فيه جمع البيانات عن جزء فقط من مفردات المجتمع الإحصائي، ويتم سحب العينة بطريقة ما يساعد في تعميم نتائجها على مجتمع البحث.

أسلوب الحصر الشامل:

مزايا أسلوب الحصر الشامل:

- خال من أخطاء الصدفة (الأخطاء العشوائية أو أخطاء المعاينة)
- يعطي صورة مفصله عن مفردات الظاهرة موضوع الدراسة.

عيوب أسلوب الحصر الشامل:

- الزيادة الكبيرة في التكاليف المادية والبشرية والزمنية.
- طول الوقت اللازم لجمع البيانات يفقد نتائج البحث حداثتها وبالتالي قيمتها.
- وجود مجتمعات بطبيعتها غير محدودة وبالتالي يتعذر تحديد إطار مفرداتها.

أسلوب العينات:

مزايا أسلوب العينات:

- يوفر التكاليف المادية والبشرية والزمنية لإجراء الدراسة.
- زيادة الرقابة والضبط والتحكم في معظم الأسباب المؤدية إلى الأخطاء.
- يصلح للمجتمعات غير المحدودة.

عيوب أسلوب العينات:

- يتعرض أسلوب المعاينة إلى نوع آخر من الأخطاء ينفرد به هذا الأسلوب ويطلق عليه خطأ المعاينة أو خطأ الصدفة وخطأ التحيز.

المجتمع والعينة

المجتمع:

يعرف المجتمع بأنه مجموعة من العناصر، أو المفردات التي تخص ظاهرة معينة محل الدراسة. وهو مصطلح علمي يراد به كل من يمكن أن تعمم عليه نتائج البحث.

العينة:

تعرف العينة بأنها جزء من مفردات المجتمع الإحصائي يتم اختياره بطريقة علمية، ثم دراسة خصائص هذا الجزء لغرض التعرف على خصائص المجتمع الذي اختير منه ذلك الجزء.
ولكي تكون العينة مقبولة من الناحية الإحصائية ينبغي أن تكون عينة ممثلة للمجتمع، أي أنها تحتوي على جميع الخصائص بنفس تواجدها في المجتمع الإحصائي الذي اختيرت منه

البارامترات (المعلمات) والإحصاءات

للمجتمع خصائص متعددة مثل المتوسط والوسيط والانحراف المعياري وكذلك لكل عينة تسحب من هذا المجتمع خصائصها أيضاً وما يتعلق بخصائص المجتمع يسمى مَعلَم أو بارامتر Parameter بينما كل ما يتعلق بخصائص العينات يسمى إحصاءه Statistic ويمكن الاستفادة من إحصاءات العينة تقدير معلمات المجتمع.

الإحصاء الوصفي والإحصاء الاستدلالي

الإحصاء الوصفي يقتصر على الوصف الكمي للظواهر وتصنيفها وتحليلها وعلاقتها بغيرها من الظواهر.

الإحصاء الاستدلالي يتعدى ذلك مستفيداً من نتائج الإحصاء الوصفي في الاستدلال على خصائص المجتمع العام للظاهرة فهو يهدف إلى تقدير خصائص المجتمع استنادا إلى نتائج دراسة عينة منتقاة من هذا المجتمع.

الإحصاء البارامترى والإحصاء اللابارامترى

الأساليب البارامترية (المعلمية): هي الأساليب التي تتطلب استيفاء افتراضات معينة حول المجتمع الذي تسحب منه عينة البحث ومن هذه الافتراضات أن يكون التوزيع طبيعياً وأن يكون هناك تجانس في التباين. والأساليب البارامترية تصلح للبيانات في المستوى الفتري والمستوى النسبي.

الأساليب اللابارامترية (اللامعلمية): هي الأساليب التي تستخدم في الحالات التي لا يكون فيها نوع التوزيع الاحتمالي للأصل الذي سحبت منه العينة معروفاً أو في حالة عدم استيفاء شرط التوزيع الاعتدالي للمجتمع. والأساليب الإحصائية اللابارامترية تصلح في حالة البيانات الرتبية والاسمية.

طرق عرض البيانات

١ ـ**العرض الجدولى للبيانات**
٢ ـ**العرض البيانى للبيانات**

وسوف نتناول فى هذه المحاضرة العرض الجدولى للبيانات بينما نتعرض للعرض البيانى للبيانات فى المحاضرة التالية إن شاء الله تعالى.

أولا: العرض الجدولى

ويقصد بالعرض الجدولى للبيانات أن يتم تلخيص البيانات محل الدراسة وتصنيفها فى صورة جداول تعبر عن القيم التي أخذها المتغير من خلال البيانات التي جمعها و كرار كل قيمة من تلك القيم.

أهمية الجداول الاحصائية:

- تعبر عن الحقائق الكمية المعروضة بعدد كبير من الارقام في جداول بطريقة منظمة
- تلخيص المعلومات الرقمية الكثيرة العدد، المتغيرة القيم، مما يسهل التعرف عليها.
- الاستيعاب وبسهولة عدد كبير من الموضوعات
- اظهار البيانات بأكبر وضوح ممكن وأصغر حيز مستطاع

تكوين الجداول

تتكون اجزاء الجدول مما يلي:

١ ـ **رقم الجدول:** يجب ان يرقم كل جدول حتى تسهل الاشارة اليه.
٢ ـ **العنوان:** يجب أن يعطي كل جدول عنوانا كاملا لتسهيل مهمة استخراج المعلومات منه، ويجب أن يكون هذا العنوان واضحا قصيرا بقدر الامكان، ويستخدم في بعض الاحيان عنوان توضيحي لبعض الجداول وذلك من أجل إعطاء معلومات إضافية عن بيانات الجدول.
٣ ـ **الهيكل الرئيسي:** ويتكون الجدول من أعمدة وصفوف، ويعتبر ترتيب المعلومات في الاعمدة والصفوف أهم خطوة في تكوين الجدول.
٤ ـ **العمود:** كل جدول يتكون من عمود أو اكثر ويوجد لكل عمود عنوان يوضح محتوياته.

٥ - **الحواشي:** قد يحتوي الجدول على مفردات بيانات لا ينطبق عليها عنوان الجدول أو عنوان العمود، ففي هذه الحالة تستعمل الحواشي لتوضيح ذلك وذلك اما بترقيم الملاحظات او باستعمال علامة (*) .. الخ.

٦ - **المصدر:** قد تؤخذ بيانات الجدول من مصادر جاهزة لذلك يجب إظهار المصدر في أسفل الجدول حتى يمكن الرجوع اليه عند الحاجة

أنواع الجداول الاحصائية:

تقسم الجداول تبعا لدرجة تعقيدها الى:

جداول بسيطة: وفيها يتكون كل من موضوع الجدول ومادته من بضع أسطر وخانات تتعلق بالتقسيمات الزمانية (أي الأمور التي يتناولها الجدول أمور تتسلسل حسب السنوات) أو المكانية (أي توزيع الظاهرة حسب المكان) أو مؤشرات وصفية بسيطة وبأرقام بسيطة أيضا.

جداول التوزيع التكراري: وفيها تكون المعطيات مجمعة في فئات بمؤشر أو متغير واحد، ولكل فئة تكراراتها الخاصة عند ذلك المؤشر

جدول التوزيع التكراري المتجمع: وفيه تجمع التكرارات على التوالي من أحد طرفي الجدول الى طرفة الآخر فنحصل على التكرار الكلي (مجموعة التكرارات)، (فاذا بدأ من أعلى الى أسفل الجدول) سمي جدول تكراري متجمع صاعد، (واذا بدأ من أسفل الى أعلى الجدول) سمي جدول تكرار متجمع نازل أو هابط.

الجداول المزدوجة أو المركبة: وهي الجداول التي تتكون من متغيرين أو أكثر، وهذه المتغيرات قد توزع على أعمدة وحقول الجدول بصورة نظامية، تعبر عن الافكار العلمية التي يريد الباحث توضيحها توضيحا عدديا.

وهناك عدة ملاحظات يجب الإنتباه إليها عند عمل جدول التوزيع التكرارى لبيانات المتغير الكمى المتصل:

1- إن تحديد عدد الفئات يتوقف على أمور عدة منها:

أ - عدد المفردات محل الدراسة
ب - انتظام وتوزيع تلك البيانات
ت - طبيعة بيانات المشكلة محل الدراسة

2- **طول الفئة** لا بد أيضاً من تحديده بعناية حيث يمثل الوجه الآخر للعملة مع عدد الفئات، فمن الأفضل أن يكون تحديده بطريقة تجعل مركز الفئة قريباً من تركز البيانات بتلك الفئة بقدر الإمكان حيث يعبر مركز الفئة عن قيمة كل مفردة من المفردات التي تنتمي لتلك الفئة.

3- **أن تكون حدود الفئات واضحة** بحيث لا يكون هناك أي تداخل فيما بينها.

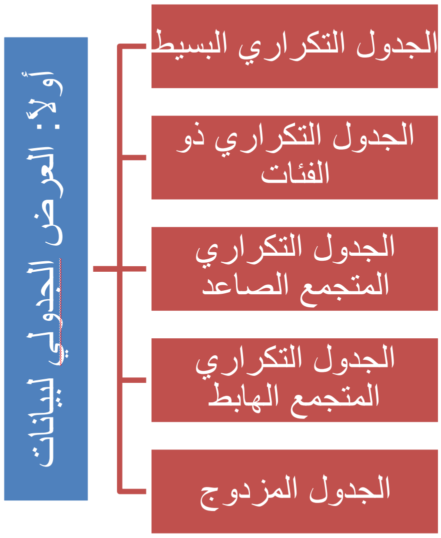
ومن هنا يمكن إعداد جداول التوزيعات التكرارية للمتغيرات المتصلة بثلاث صور هى:

١ - الجداول التكرارية المنتظمة
٢ - الجداول التكرارية غير المنتظمة
٣ - الجداول التكرارية المفتوحة

المحاضرة الثانية

تبويب وعرض البيانات الاحصائية

أولاً: العرض الجدولي للبيانات



تبويب البيانات في جدول تكراري بسيط:

مثـال :

البيانات التالية هى درجات حصل عليها عشرون طالباً فــى مــادة الإحصاء الاجتماعي بالفرقة الأولى قسم الاجتمــاع فــى امتحــان نهاية العام :

10	12	13	15	10	12	14	15	11	12
15	12	10	12	13	15	12	13	10	14

والمطلوب تبويب هذه البيانات فى جدول توزيع تكرارى بسيط ؟

الدرجة	العلامات	التكرار
10	////	4
11	/	1
12	///// /	6
13	///	3
14	//	2
15	////	4
المجموع		20

التقدير	التكرار
مقبول	5
جيد	9
جيد جداً	3
ممتاز	3
المجموع	20

تبويب البيانات في جدول تكراري ذو فئات:

- **المقصود بالفئات**

67	64	68	73	73	54	61	74	60	78
80	74	65	63	60	69	72	66	77	65
74	50	76	69	68	66	78	63	70	55
67	67	64	76	61	72	72	57	65	77
59	71	79	78	58	63	74	66	73	67
61	71	69	68	73	81	64	61	84	55

- **طريقة كتابة الفئات**

١

ف	ك
20-10	5
30-20	20
40-30	50
50-40	25

٢

ف	ك
19-10	5
29-20	20
39-30	50
49-40	25

٣

ف	ك
-10	5
-20	20
-30	50
-40	25

٤

ف	ك
-20	5
-30	20
-40	50
-50	25

مثال :

قام باحث بجمع بيانات تمثل درجات اختبار مادة الحاسب الآلي لخمسين طالباً من طلاب المرحلة الثانية من الثانوية العامة في الجدول التالي :

57	42	51	55	70
53	63	47	60	45
55	82	39	65	33
42	65	61	58	64
55	45	53	52	50
39	63	59	36	25
64	54	49	45	65
78	52	41	42	75
26	48	25	35	30
88	46	55	40	20

والمطلوب هو إعداد جدول توزيع تكراري ذو فئات للجدول السابق؟

حساب المدى = أكبر قيمة – أصغر قيمة = 88 – 20= 68

عدد الفئات = 1 + 3,3 لو (ن)

= 1 + 3,3 × 1,699 = 6,61 7

طول الفئة = المدى / عدد الفئات

= 68/ 7 = 9,71 10

بداية الفئة الأولى هو الحد الأدنى للدرجات (20)

الفئات	العلامات	التكرار
-20	~~////~~	4
-30	~~////~~ /	6
-40	~~////~~ ~~////~~ //	12
-50	~~////~~ ~~////~~ ////	14
-60	~~////~~ ////	9
-70	///	3
90-80	//	2
المجموع		50

تبويب البيانات في الجدول التكراري المتجمع الصاعد:

يقصد بالتكرار المتجمع الصاعد هو تجميع تكرار كل فئة على جميع التكرارات السابقة لها بحيث يكون مجموع التكرار التصاعدي للفئة الأخيرة مساوي لمجموع التكرارات

الفئات	العلامات	التكرار
-20	~~////~~	4
-30	/ ~~////~~	6
-40	// ~~////~~ ~~////~~	12
-50	//// ~~////~~ ~~////~~	14
-60	//// ~~////~~	9
-70	///	3
90-80	//	2
المجموع		50

حدود الفئات	التكرار المتجمع الصاعد
أقل من 20	صفر
أقل من 30	4
أقل من 40	10
أقل من 50	22
أقل من 60	36
أقل من 70	45
أقل من 80	48
أقل من 90	50

تبويب البيانات في الجدول التكراري المتجمع الهابط:

يقصد بالتكرار المتجمع الهابط هو تجميع تكرار كل فئة على جميع التكرارات التالية لها بحيث يكون مجموع التكرار التنازلي للفئة الأولى مساوي لمجموع التكرارات

الفئات	العلامات	التكرار
-20	~~////~~	4
-30	/ ~~////~~	6
-40	// ~~////~~ ~~////~~	12
-50	//// ~~////~~ ~~////~~	14
-60	//// ~~////~~	9
-70	///	3
90-80	//	2
المجموع		50

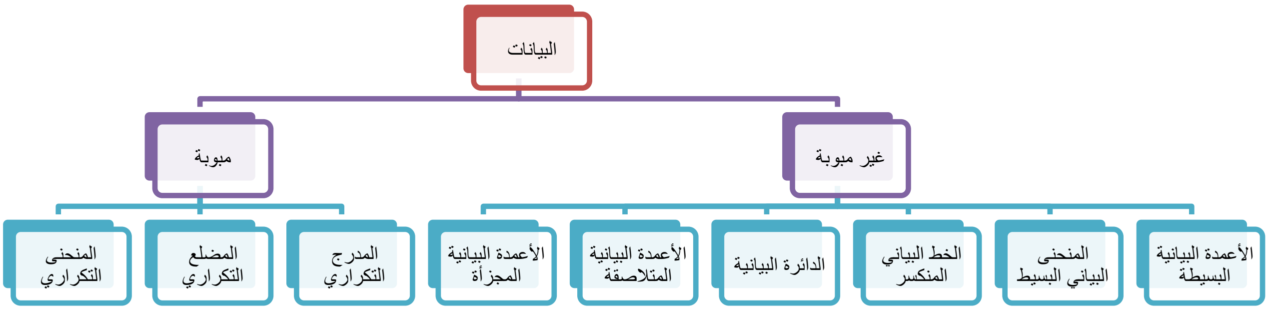
حدود الفئات	التكرار المتجمع الهابط
20 فأكثر	50
30 فأكثر	46
40 فأكثر	40
50 فأكثر	28
60 فأكثر	14
70 فأكثر	5
80 فأكثر	2
90 فأكثر	صفر

تبويب البيانات في الجدول المزدوج:

التوزيع المشترك بين النوع وحضور المحاضرات

	النوع		المجموع
	إناث	ذكور	
عدم الحضور	54	117	**171**
الحضور	348	950	**1298**
المجموع	**402**	**1067**	**1469**

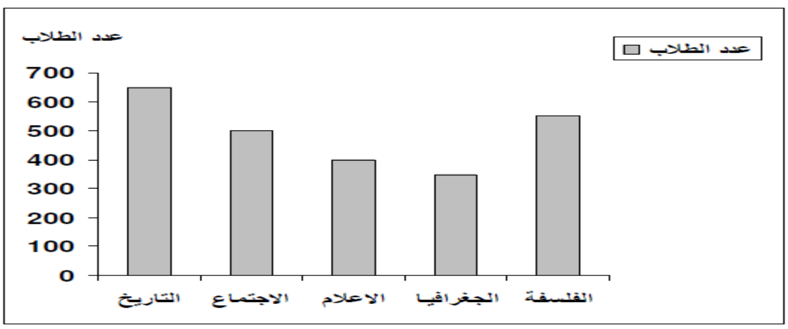
ثانيا: العرض البياني للبيانات



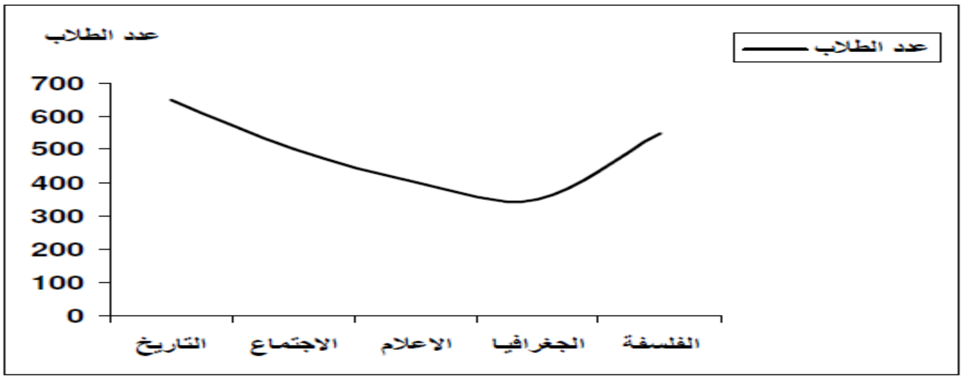
الأعمدة البيانية البسيطة:

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية البسيطة ؟

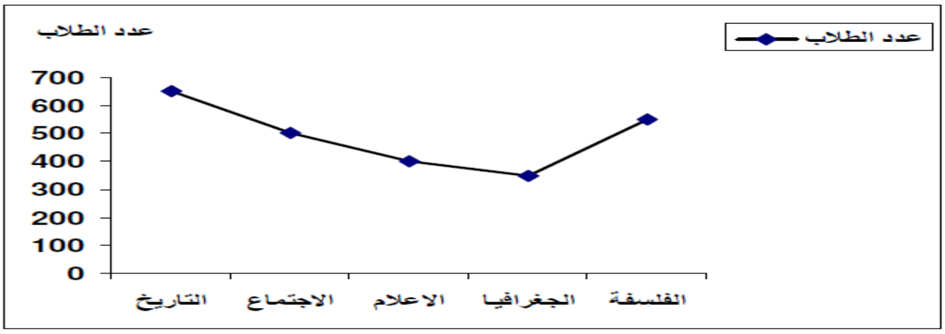
القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



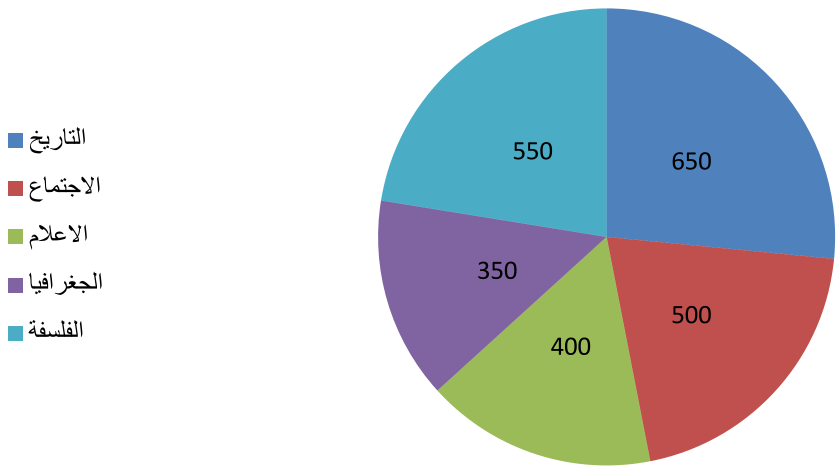
المنحنى البياني البسيط:



الخط البياني المنكسر:



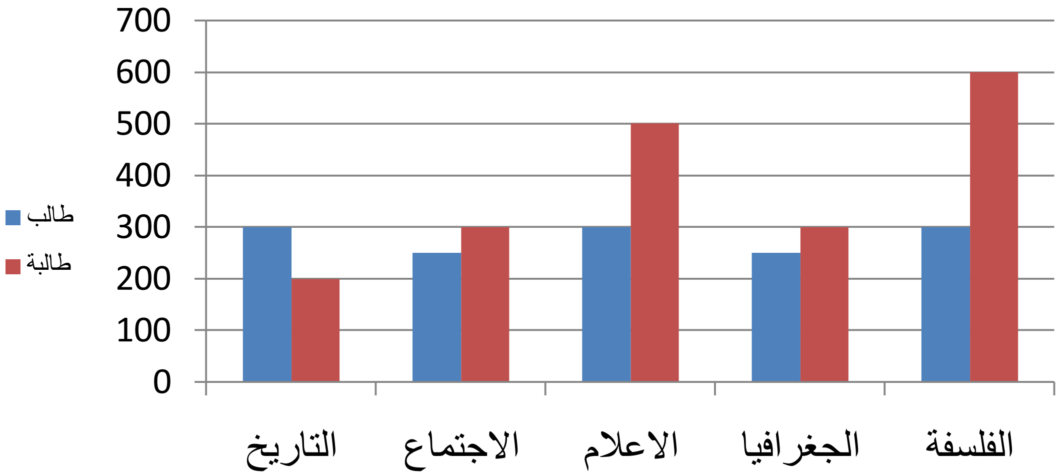
الدائرة البيانية:



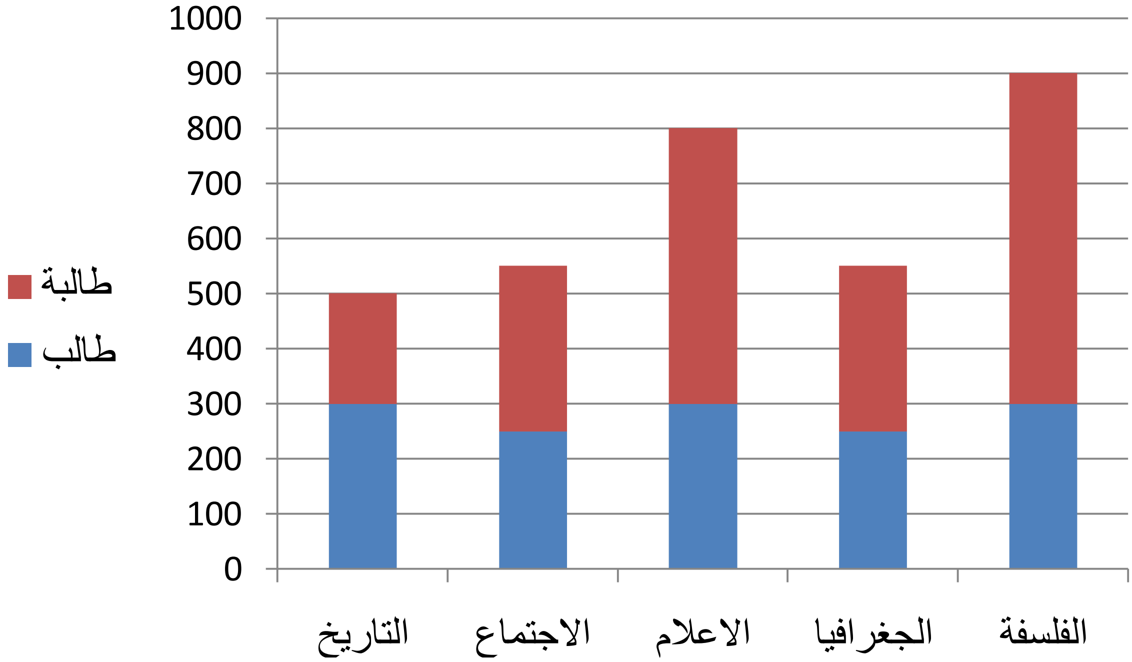
الأعمدة البيانية المتلاصقة:

الجدول التالي يوضح أعداد الطلاب ببعض أقسام كلية الآداب بجامعة الملك فيصل والمطلوب عرض هذه البيانات. باستخدام طريقة الأعمدة المتلاصقة

القسم	التاريخ	الاجتماع	الاعلام	الجغرافيا	الفلسفة
طالب	300	250	300	250	300
طالبة	200	300	500	300	600

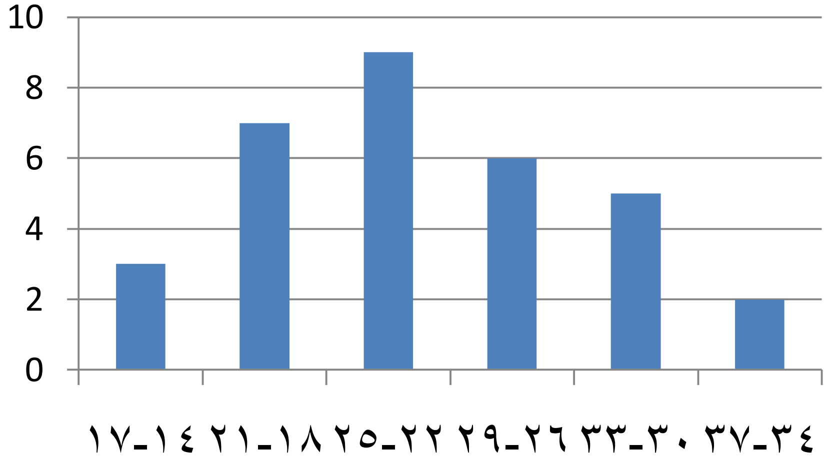


الأعمدة البيانية المجزأة:

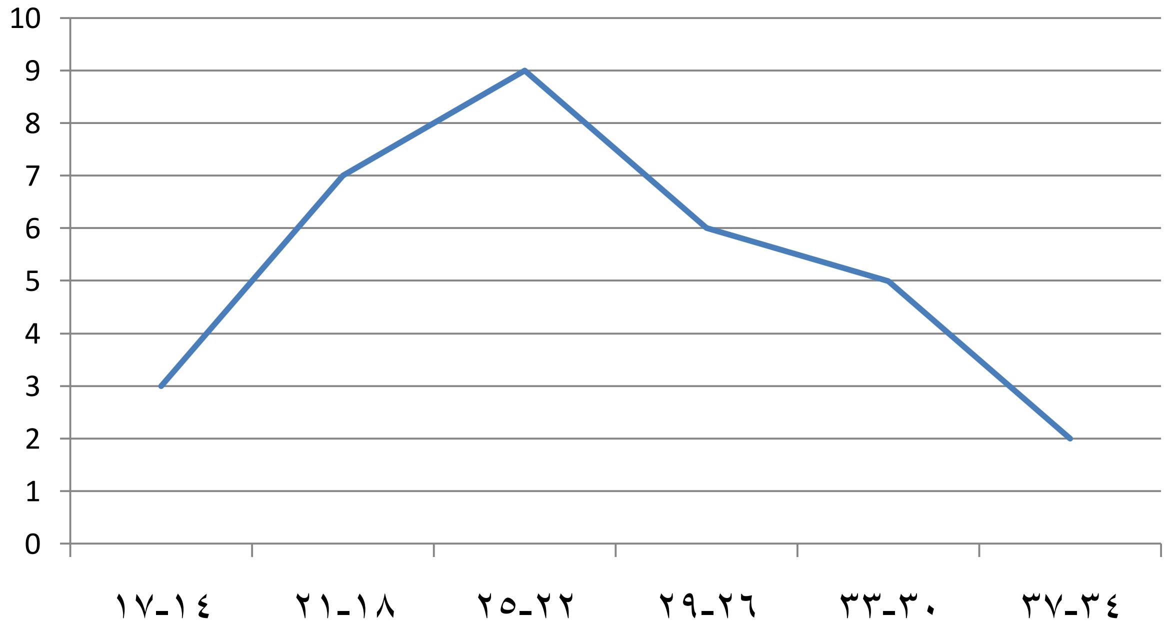


المدرج التكراري:

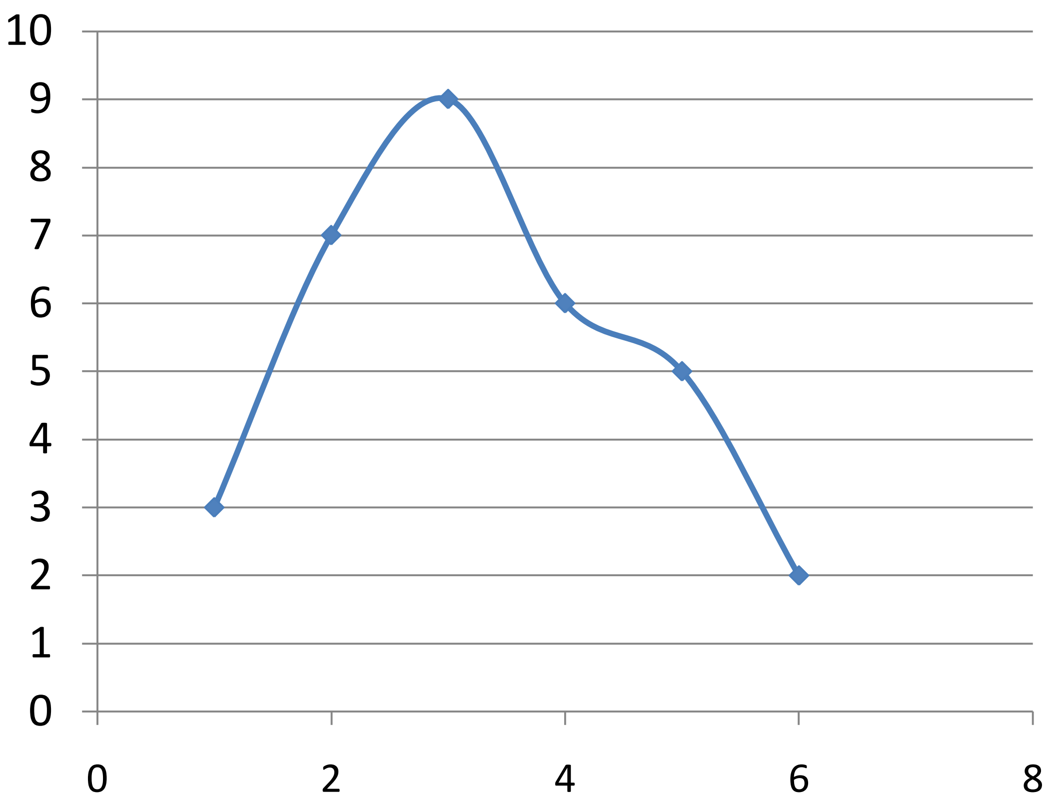
الفئات	التكرارت
17-14	3
21-18	7
25-22	9
29-26	6
33-30	5
37-34	2
المجموع	32



المضلع التكراري



المنحنى التكراري:



<u>تمارين:</u>

1- حصل عدد من الطلاب فى مادة الإحصاء على الدرجات التالية :

5	4	4	5	3	4	2	3	1	2
3	7	4	1	6	3	2	5	3	4
7	3	2	6	5	3	4	2	4	1

المطلوب : تكوين جدول تكرارى بسيط لهذه الدرجات.

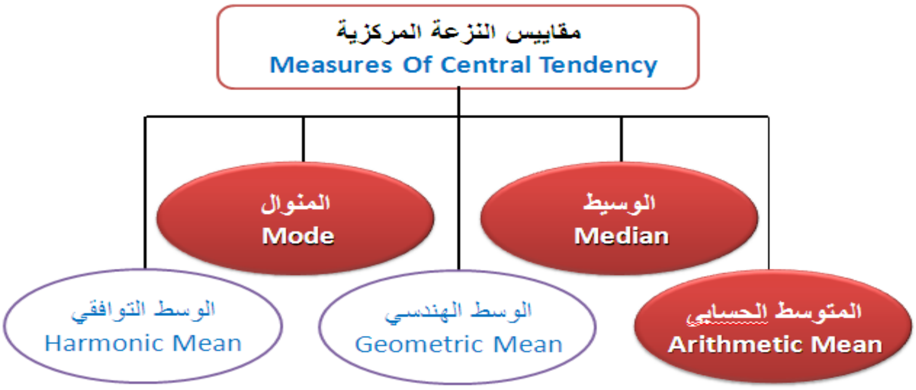
2- تمثل البيانات التالية تقديرات عشرون طالبا فى مادة علم النفس والمطلوب وضعها فى جدول تكرارى بسيط لتلك التقديرات .

ممتاز	مقبول	جيد جدا	مقبول	جيد
جيد جدا	جيد	ضعيف	جيد	مقبول
جيد	ممتاز	مقبول	ضعيف	جيد
جيد جدا	جيد	مقبول	جيد	مقبول

المحاضرة الثالثة

مقاييس النزعه المركزية
المتوسط الحسابي – الوسيط – المنوال

- كل ظاهرة في الحياة العامة لها ميل للتجمع حول نقطة معينة ؛ ومن ثم إذا استطعنا تحديد هذه النقطة فإننا سنصل إلى قيمة متوسطة تتجمع حولها القيم.
- الميل إلى التجمع حول هذه يسمى القيمة بالنزعة المركزية
- وتسمى المقاييس المستخدمة مقاييس النزعة المركزية



يسمى الوسط الحسابي، والوسيط، والمنوال، مقاييس النزعة المركزية لأن كلا منها يحاول أن يصف نقطة تجمع مشاهدات التوزيع

أهمية مقاييس النزعة المركزية

عند معرفتنا بتلك المتوسطات (مقاييس النزعة المركزية) يصبح أمامنا فرصة كبيرة لأن :

- ننظر لمتوسط مجموعة من البيانات لنعرف الكثير عن خصائص تلك المجموعة.
- نعقد مقارنة بين عدة مجموعات من البيانات في وقت واحد وذلك من خلال مقارنة متوسطات تلك المجموعات بعضها ببعض .

الوسط الحسابى **Arithmetic Mean**

يعد من أكثر المقاييس المستخدمة فى الاحصاء حيث انه بسيط وسهل الفهم و يصلح للمقارنة بين المجموعات .

إذا كانت قيم المتغير (x) هـــى $x_1 , x_2 , ... , x_n$ حيث (n) يمثل حجم المجموعة ؛ فإن الوسط الحسابى يمكن التعبير عنه على النحو التالى :-

$$\boxed{\overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x}{n}}$$

الوسط الحسابي = $\frac{\text{مجموع قيم البيانات}}{\text{عددها}}$ أي $\bar{x}=\frac{\sum x}{n}$

س ١ : درجات خمسة طلاب في مقرر ما [الدرجة العظمى 20] هي : 10 , 12 , 7 , 2 , 9 .
أوجد الوسط الحسابي لدرجاتهم .

ج ١ : $\bar{x}=\frac{\sum x}{n}=\frac{9+2+7+12+10}{5}=\frac{40}{5}=8$

من هذا المثال **البسيط** يمكن ملاحظة **الخصائص العامة** التالية للوسط الحسابي :

- يمكن تحديد قيمة الوسط الحسابي بالضبط، كما أن طريقة تحديده سهلة .
- يأخذ في الاعتبار جميع البيانات .
- لا يتأثر بترتيب البيانات .
- لا يُشترط أن يكون الوسط الحسابي عدداً صحيحاً ولا يُشترط أن يكون إحدى قيم البيانات ولكنه قيمة تقع بين أقل قيمة في البيانات وأكبر قيمة فيها .
- يتأثر بالقيم المتطرفة في البيانات [كما يتضح من السؤالين التاليين] .

س ٢ : احسب الوسط الحسابي للقيم:
40, 50, 45, 55, 35

ج ٢ :
$\frac{40+50+45+55+35}{5}=\frac{225}{5}=45$

س ٣ : احسب الوسط الحسابي للقيم:
10 , 15 , 12 , 13 , 900

ج ٣ :
$\frac{40+50+45+55+995}{5}=\frac{285}{5}=57$

- حاصل ضرب قيمة الوسط الحسابي في عدد البيانات = مجموع قيم البيانات

وهذا واضح من تعريف الوسط الحسابي :

$\bar{x}=\frac{\sum x}{n} \xrightarrow{\text{تعني أن}} n\times\bar{x}=\sum x$

فمثلاً

9	2	7	12	10

خمس درجات وسطها الحسابي 8

$\underbrace{5\times 8}_{=40}=\underbrace{9+2+7+12+10}_{=40}$

- إذا أضفنا عدد ثابت c لكل قيمة من قيم البيانات، فإن :

الوسط الحسابي الجديد = الوسط الحسابي القديم + العدد الثابت c

بيانات قديمة

9	2	7	12	10

خمس درجات وسطها الحسابي 8

إضافة 1 لكل قيمة

بيانات جديدة

10	3	8	13	11

خمس درجات وسطها الحسابي 9 [8 + 1 =]

- إذا ضربنا كل قيمة من قيم البيانات في عدد ثابت c ، فإن :

الوسط الحسابي الجديد = الوسط الحسابي القديم x العدد الثابت c

بيانات قديمة: 9 , 2 , 7 , 12 , 10
خمس درجات وسطها الحسابي 8

ضرب كل قيمة في 1.1 ←

بيانات جديدة: 9.9 , 2.2 , 7.7 , 13.2 , 11
خمس درجات وسطها الحسابي 8.8 [8 x 1.1 =]

اعتبر نفسك مدرساً للطلاب الخمسة المذكورين في س ١ [كانت درجاتهم (من 20) كالتالي : 9 , 2 , 7 , 12 , 10]، وأردت أن تحسن من الوسط الحسابي لدرجاتهم ، أيهما أفضل : أن نزيد درجة كل طالب 5 درجات أم نزيد درجة كل طالب 50% من قيمتها ؟ علل إجابتك .

حساب الوسط الحسابي لبيانات غير مبوبة

س : أوجد الوسط الحسابي للأرقام :

5 , 5 , 5 , 5 , 5 , 5 , 3 , 3 , 6 , 6 , 4 , 4 , 4 , 4 , 4 , 2 , 2 , 8 , 8 , 8

ج : بتطبيق مباشر للتعريف :

$$\bar{x}=\frac{(5+5+5+5+5+5)+(3+3)+(6+6)+(4+4+4+4+4)+(2+2)+(8+8+8)}{20}=\frac{96}{20}=4.8$$

لاحظ أن الرقم 5 متكرر 6 مرات ، الرقم 3 مرتان ، والرقم 6 مرتان ، والرقم 4 متكرر 5 مرات ، والرقم 2 مرتان ، والرقم 8 ثلاث مرات ، وبالتالي يمكن عمل العملية الحسابية السابقة كالآتي :

$$\bar{x}=\frac{(6 \times 5)+(2 \times 3)+(2 \times 6)+(5 \times 4)+(2 \times 2)+(3 \times 8)}{6+2+2+5+2+3}$$

$$=\frac{30+6+12+20+4+24}{20}=\frac{96}{20}=\underline{4.8}$$

وهذا يمكن إنجازه بيسر من خلال الجدول التكراري للبيانات كالتالي :

المتغير x	التكرار f	fx
5	6	30
3	2	6
6	2	12
4	5	20
2	2	4
8	3	24
	20	96
	$\sum f=20$	$\sum fx=96$

$$\bar{x}=\frac{\sum fx}{\sum f}=\frac{96}{20}=4.8$$

أي أنه في حالة البيانات الكمية المتقطعة ذات التكرارات يمكن حساب الوسط الحسابي من العلاقة :

$$\bar{x}=\frac{\sum fx}{\sum f}$$

حيث $\sum f$ هو مجموع التكرارات

$\sum fx$ هو مجموع حاصل ضرب كل قيمة في تكرارها

س: من مائة رقم يتكرر الرقم 4 عشرون مرة، والرقم 5 أربعون مرة، والرقم 6 ثلاثون مرة، والباقي كانوا الرقم 7 احسب الوسط الحسابي للمائة رقم .

ج : بتكوين الجدول التكراري للأرقام المذكورة ، ثم بضرب كل قيمة في تكرارها والتجميع [عمود fx] يكون الوسط الحسابي للأرقام المذكورة هو :

الجدول التكراري

المتغير x	التكرار f	fx
4	20	80
5	40	200
6	30	180
7	10	70
	100	530
	$\sum f = 100$	$\sum fx = 530$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{530}{100} = \underline{5.3}$$

حساب الوسط الحسابي لبيانات مبوبة

عندما نتعامل مع بيانات متصلة تُعطى فيها قيم المتغير على صورة فترات، فيمكن اعتبار أن جميع القيم داخل الفترة مطابقة لمركز الفئة ، وبالتالي يمكن استخدام الصيغة السابقة لحساب الوسط الحسابي :

$$\bar{x} = \frac{\sum f\, x_0}{\sum f}$$

حيث $\sum f$ هو مجموع التكرارات ، $\sum f\, x_0$ هو مجموع حاصل ضرب مركز كل فئة في تكرار الفئة

ففي المثال التالي والذي يوضح اطوال سيقان الزهار بالسنتيمتر، يكون الوسط الحسابي لأطوال سيقان الأزهار هو:

الجدول التكراري

الفئة	المتغير x (الطول)	التكرار f	مركز الفئة x_0	fx_0
الأولى	$0 \le x < 20$	4	10	40
الثانية	$20 \le x < 30$	16	25	400
الثالثة	$30 \le x < 35$	12	32.5	390
الرابعة	$35 \le x < 40$	10	37.5	375
الخامسة	$40 \le x < 50$	6	45	270
السادسة	$50 \le x < 60$	2	55	110
		$\sum f = 50$		$\sum f\, x_0 = 1585$

$$\bar{x} = \frac{\sum f\, x_0}{\sum f} = \frac{1585}{50} = \underline{31.7}$$

مزايا وعيوب الوسط الحسابي

من كل ما سبق يمكن استعراض مزايا وعيوب الوسط الحسابي كالتالي :

- يمكن تحديد قيمة الوسط الحسابي بالضبط، كما أن طريقة تحديده سهلة [ميزة].
- يأخذ في الاعتبار جميع البيانات [ميزة] .
- لا يتأثر بترتيب البيانات [ميزة] .
- يتأثر بالقيم المتطرفة في البيانات [عيب] .
- لا يمكن حسابه بالرسم ، أي بيانياً [عيب] .

الوسيط **Median**

استخدام الوسيط في حالة التعامل مع

- **البيانات التي تكثر بها القيم الشاذة.**
- **الجداول التكرارية المفتوحة من أحد طرفيها أو من كليهما.**
- **التوزيعات التكرارية غير المتساوية في طول الفئات.**

تعريف الوسيط :

(ببساطة) يُعرف الوسيط [وسنرمز له بالرمز M] لمجموعة من القيم (المرتبة تصاعدياً أو تنازلياً حسب قيمها) على أنه القيمة التي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد ، أو بتعبير آخر هي القيمة التي في المنتصف

فمثلاً لمجموعة القيم : 13 , 14 , 19 , 16 , 20 , 15 , 18 , 14 , 19 [عددها ٩ قيم أي رقم فردي] ،
إذا قمنا بترتيبها تصاعدياً أو تنازلياً

ترتيب تصاعدي (من الأصغر للأكبر) : 13 14 14 15 **16** 18 19 19 20

مجموعتان متساويتان في العدد

ترتيب تنازلي (من الأكبر للأصغر) : 20 19 19 18 **16** 15 14 14 13

مجموعتان متساويتان في العدد

لاحظ هنا أن عدد القيم n [هنا = ٩] فردي وبالتالي هناك قيمة واحدة في منتصف المجموعة

يكون الوسيط هو العدد الخامس [رتبة الوسيط أي ترتيبه بين القيم] وقيمته **16**

هام جدا : فرق بين رتبة الوسيط وقيمته

أما لمجموعة القيم : 12 , 13 , 14 , 19 , 16 , 20 , 15 , 18 , 14 , 19 [عددها ١٠ قيم (أي رقم زوجي) حيث أضفنا القيمة 12 للمجموعة السابقة]، إذا قمنا بترتيبها تصاعدياً أو تنازلياً

ترتيب تصاعدي (من الأصغر للأكبر)

12 13 14 14 15 16 18 19 19 20

الوسيط هو الوسط الحسابي للقيمتين بالمنتصف

مجموعتان متساويتان في العدد

ترتيب تنازلي (من الأكبر للأصغر)

20 19 19 18 16 15 14 14 13 12

الوسيط هو الوسط الحسابي للقيمتين بالمنتصف

مجموعتان متساويتان في العدد

في هذه الحالة توجد قيمتان بالمنتصف وهما القيمة الخامسة والقيمة السادسة [وهما العددان 16 , 15] ، عندئذ يكون الوسيط هو الوسط الحسابي لهاتين القيمتين ، أي :

$$\frac{15+16}{2} = 15.5$$

إذن من السابق يمكن استنتاج طريقة حساب الوسيط لمجموعة من القيم كالآتي :

- قم أولاً بترتيب البيانات تصاعدياً أو تنازلياً .
- حدد ما إذا كانت هناك قيمة واحدة بالمنتصف أم قيمتين، وهذا يتوقف على قيمة n

فإذا كانت n فردية	وإذا كانت n زوجية
كانت هناك قيمة واحدة في المنتصف رتبتها $\frac{n+1}{2}$ وتكون هذه القيمة هي الوسيط	كانت هناك قيمتان في المنتصف رتبتهما $\frac{n}{2}$ ، $\frac{n}{2}+1$ ويكون الوسط الحسابي لهاتين القيمتين هو الوسيط

فمثلاً

تذكر : الوسط الحسابي لهذه القيم هو :

$$\frac{9+2+7+12+10}{5}=8$$

$n = 5$

9 2 7 12 10

ترتيب تصاعدي

2 7 9 10 12

هناك قيمة واحدة في المنتصف رتبتها :

$$\frac{n+1}{2}=\frac{5+1}{2}=3$$

أي القيمة الثالثة . وتكون تلك القيمة هي الوسيط . أي أن :

الوسيط = 9

$n = 6$

9 2 7 12 10 500

ترتيب تصاعدي

2 7 9 10 12 500

هناك قيمتان في المنتصف رتبتهما :

$$\frac{n}{2}=\frac{6}{2}=3 \quad , \quad 3+1=4$$

أي القيمتان الثالثة والرابعة ، وتكون قيمة الوسيط هي الوسط الحسابي لهاتين القيمتين ، أي :

$$\frac{9+10}{2}=9.5$$

الوسط الحسابي لهذه القيم هو

$$\frac{9+2+7+12+10+500}{6}=90$$

وواضح تأثره كثيراً بالقيمة المتطرفة 500

هل لاحظت أن الوسيط لم يتأثر بالقيمة المتطرفة 500

لاحظ من الأمثلة السابقة أن كلاً من المتوسطين الوسط الحسابي و الوسيط من السهل حسابهما ومن الممكن أن يمثل كل منهما مقياساً للنزعة المركزية للبيانات ، لكن الأفضل (نسبياً هنا) أن نستخدم الوسط الحسابي كمقياس للنزعة المركزية للبيانات حيث أنه يأخذ في الاعتبار جميع قيم البيانات، بينما يهتم الوسيط بقيم البيانات في المنتصف (وذلك بعد ترتيبها) .

مثال آخر : الأجر (بالريال) في الساعة لخمسة عاملين في مكتب هو : 37 , 92 , 32 , 39 , 25 . احسب الوسط الحسابي للأجور ووسيط هذه الأجور . أيهما تفضل كمقياس لمتوسط أجر الساعة ؟ ولماذا ؟

الوسط الحسابي للأجور هو : $\bar{x} = \frac{\sum x}{n} = \frac{25+39+32+92+37}{5} = \frac{225}{5} = 45$

أما لتحديد الوسيط ، فلابد أولاً من ترتيب القيم (تصاعدياً مثلاً) : 25 , 32 , 37 , 39 , 92

وحيث أن عدد القيم فردي ، إذن هناك قيمة واحدة في المنتصف [هي 37] وهي الوسيط

لاحظ في هذا السؤال أن الوسط الحسابي (بالرغم من عدم احتياجه لترتيب القيم وفي نفس الوقت يأخذ في الاعتبار جميع قيم البيانات) إلا أنه تأثر جداً بالقيمة المتطرفة 92 ، في حين لم يتأثر بها الوسيط لأنه يعتمد على البيانات في المنتصف . لذا يُفضل هنا استخدام الوسيط كمقياس للنزعة المركزية حيث يعطي دلالة أفضل لمتوسط الأجور من الوسط الحسابي .

الوسيط لبيانات كمية متصلة:

يمكن حساب الوسيط للبيانات الكمية المتصلة من خلال الرسم

وكذلك من خلال المعادلات الاحصائية بسهولة

الوسيط من خلال الرسم البياني يتم كالتالى:

طريقة تحديد الوسيط من :

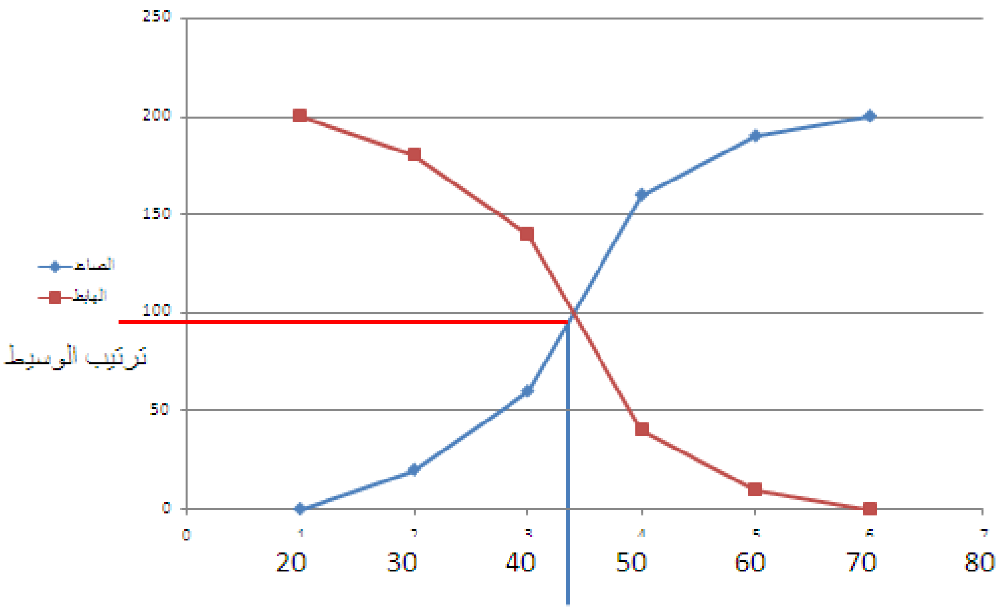
* من المنحنى المتجمع الصاعد فقط
* من المنحنى المتجمع الهابط فقط
* من المنحنيين معا

الجدول التالى يوضح العلاقة بين فئات الدخل بأحد المصانع وعدد العمال والمطلوب من واقع بيانات الجدول حساب الوسيط بالرسم من جدول التكرار المتجمع الصاعد والهابط معاً .

فئات الدخل	20-	30-	40-	50-	70-60
عدد العمال	20	40	100	30	10

الحدود العليا للفئات	ك. م. ص
أقل من ٢٠	صفر
أقل من ٣٠	٢٠
أقل من ٤٠	٦٠
أقل من ٥٠	١٦٠
أقل من ٦٠	١٩٠
أقل من ٧٠	٢٠٠

الحدود العليا للفئات	ك. م. هـ
٢٠ فأكثر	٢٠٠
٢٠ فأكثر	١٨٠
٢٠ فأكثر	١٤٠
٢٠ فأكثر	٤٠
٢٠ فأكثر	١٠
٢٠ فأكثر	صفر



قيمة الوسيط = ٤٤

الوسيط من خلال المعادلات الاحصائية:

مثال : في دراسة جغرافية لعدد من مساحات مجموعة من الأراضي لمنطقة سكنية بالرياض تبين أن التوزيع التكراري لها كما هو مبين .

المطلوب حساب الوسط الحسابي والوسيط لمساحة الأراضي .

المساحة (بالكيلومتر)	عدد قطع الأراضي
1 -	14
3 -	29
5 -	18
7 - 10	9

المتغير x هنا هو مساحة الأرض (بالكيلومتر) ، في حين يمثل عدد قطع الأراضي التكرار f .

أولاً : الوسط الحسابي : نستكمل الجدول التكراري كما هو مبين :

الجدول التكراري

الفئة	المتغير (المساحة) x	التكرار f	المركز x_0	$f\,x_0$
الأولى	$1 \le x < 3$	14	2	28
الثانية	$3 \le x < 5$	29	4	116
الثالثة	$5 \le x < 7$	18	6	108
الرابعة	$7 \le x < 10$	9	8.5	76.5
		$\sum f = 70$		$\sum f\,x_0 = 328.5$

$$\therefore \quad \bar{x} = \frac{\sum f\,x_0}{\sum f} = \frac{328.5}{70} = 4.692857143 \cong \underline{\underline{4.7}}$$

وهنا يتبادر إلى الذهن سؤالان هامان :

أي أن **الفئة الوسيطة** هي تلك الفئة التي يقع داخلها الوسيط

السؤال الأول : هل من الممكن تحديد الفئة الوسيطة من الجدول التكراري مباشرةً أم لازم نعمل الجدول التكراري المتجمع الصاعد ونرسم المضلع التكراري المتجمع الصاعد ؟ .

السؤال الثاني : هل من الممكن [بعد تحديد الفئة الوسيطة] تحديد الوسيط من الجدول التكراري مباشرة دون الحاجة للجدول التكراري المتجمع الصاعد أو المضلع التكراري المتجمع الصاعد ؟ .

والإجابة على السؤالين : نعم يمكن تحديد الفئة الوسيطة من الجدول التكراري مباشرةً ، ثم بعد ذلك يمكن أيضاً من هذا الجدول التكراري تحديد قيمة الوسيط دون أن نحتاج لعمل جدول تكراري متجمع صاعد ورسم المضلع التكراري المتجمع الصاعد ، **وذلك كالتالي:**

بالنسبة للسؤال الأول [تحديد الفئة الوسيطة]

- (١) احسب أولاً نصف مجموع التكرارات .

(٢) ابدأ بالرقم صفر في ذهنك وزود تكرارات الفئات على التوالي وكل مرة قارن بنصف مجموع التكرارات السابق أول ما يزيد الناتج عن نصف المجموع السابق أو يساويه تكون آخر فئة زودنا تكرارها تكون هي الفئة الوسيطة .

ويتم ذلك كالتالي

الجدول التكراري		
التكرار f	المتغير (المساحة) x	الفئة
14	$1 \le x < 3$	الأولى
29	$3 \le x < 5$	الثانية
18	$5 \le x < 7$	الثالثة
9	$7 \le x < 10$	الرابعة
$\sum f = 70$		

- احسب $\frac{1}{2}\sum f$ ⟵ $\frac{1}{2}\sum f = \frac{70}{2} = 35$
- نبدأ بالصفر [في ذهننا]
- نزود على الصفر السابق تكرار الفئة الأولى [14] ينتج

14

14 أقل من 35 ، يبقى الفئة الأولى ليست الفئة الوسيطة

- نزود على الـــ 14 الأخيرة تكرار الفئة الثانية [29] ينتج

43

43 أكبر من 35 ، يبقى الفئة الثانية هي الفئة الوسيطة

وبالنسبة للسؤال الثاني [تحديد الوسيط (بعد ما حددنا الفئة الوسيطية)]

(١) حدد الحد الأدنى للفئة الوسيطية وأيضاً طولها
(٢) احسب ما يُسمى بــ "التكرار المتجمع السابق" = مجموع تكرار الفئات السابقة للفئة الوسيطية
(٣) احسب الوسيط من العلاقة :

الوسيط M = الحد الأدنى للفئة الوسيطية + [(نصف مجموع التكرارات – التكرار المتجمع السابق) / تكرار الفئة الوسيطية × طول الفئة الوسيطية]

الجدول التكراري		
الفئة	المتغير (المساحة) x	التكرار f
الأولى	$1 \le x < 3$	14
الثانية	$3 \le x < 5$	29
الثالثة	$5 \le x < 7$	18
الرابعة	$7 \le x < 10$	9
		$\sum f = 70$

الفئة الوسيطية ← (الثانية)

- الفئة الوسيطية هي الفئة الثانية :

حدها الأدنى 3 وطولها 2 [= 3 – 5] وتكرارها 29

- التكرار المتجمع السابق :

يساوي مجموع تكرارات الفئات السابقة للفئة الوسيطية [أي تكرار الفئة الأولى فقط] = 14

- بالتعويض في القانون السابق :

$$M = 3 + \left[\frac{35-14}{29} \times 2\right] = 3 + \left[\frac{21}{29} \times 2\right] = 3 + 1.44827 = 4.44827 \cong 4.4$$

تُسمى الطريقة الحسابية السابقة (لحساب الوسيط) بــ "طريقة الاستكمال"

المنوال

تعريف المنوال [الشائع]

يُعرف المنوال لمجموعة من القيم على أنه القيمة التي تتكرر أكثر من غيرها أو القيمة الأكثر شيوعاً [لذا يُسمى في بعض الأحيان بالــ "الشائع"] . وأحياناً يُرمز للمنوال بالرمز

فمثلاً :

مجموعة القيم	18	12	11	10	10	9	9	9	7	5	2	2	لها منوال 9
ومجموعة القيم	16	15	12	10	8	5	3	9	ليس لها منوال			[أو عديمة المنوال]
ومجموعة القيم	9	7	7	7	5	5	4	4	4	3	2	لها منوالان 7 , 4

أي أن مجموعة القيم قد تكون وحيدة المنوال [لها منوال واحد] ، وقد تكون عديدة المنوال [منوالان أو أكثر] وقد تكون عديمة المنوال [لا يوجد لها منوال]

4 4 5 5 6 6 7 7

فقد تتسرع وتقول أنها رباعية المنوال ومناولها : 4 , 5 , 6 , 7

لكن [حيث أن جميع القيم لها نفس التكرار] هذه المجموعة الأخيرة عديمة المنوال

والمنوال [مقارنةً بالوسط الحسابي والوسيط] به العديد من العيوب منها :

- أنه لا يأخذ في الاعتبار جميع البيانات ولكنه يهتم فقط بالقيم الأكثر تكراراً .
- أنه قد لا يتواجد أو قد يكون هناك أكثر من منوال للبيانات .

إلا أنه أيضاً يتميز ببعض المزايا منها :

- أنه أسرع في تحديده من الوسط والوسيط
- من الممكن تحديده للتوزيعات التكرارية للبيانات المنفصلة سواء كانت تلك البيانات كمية متقطعة أو نوعية [والبيانات الأخيرة (النوعية) لا يمكن حساب الوسط الحسابي لها أو الوسيط]

بيانات منفصلة

درجات طلاب في مقرر الإحصاء

عدد الطلاب	درجة الطالب
28	12
24	14
39	16
9	18

بيانات كمية متقطعة

لها منوال وحيد وهو "الدرجة 16"

درجات طلاب في مقرر الإنجليزي

عدد الطلاب	درجة الطالب
23	12
30	14
30	16
17	18

بيانات كمية متقطعة

لها منوالان وهما "14 , 16"

درجات طلاب في مقرر الفقه

عدد الطلاب	درجة الطالب
25	12
25	14
25	16
25	18

بيانات كمية متقطعة

ليس لها منوال

سيارات في أحد المواقف

عدد	لون السيارة
10	أحمر R
23	أزرق B
12	أبيض W
5	أصفر Y

بيانات نوعية

لها منوال وهو "اللون الأزرق"

وماذا عن التوزيعات التكرارية للبيانات الكمية المتصلة

$$\text{المنوال} = أ + \frac{ف_1}{ف_1 + ف_2} \times ل$$

أ = الحد ألدنة للفئة المنوالية والمقصود بدايتها .

$ف_1 = ك - ك1$

$ف_2 = ك - ك2$

ك = تكرار الفئة المنوالية

ك1 = تكرار الفئة التى تسبق الفئة المنوالية

ك2 = تكرار الفئة التى تلى الفئة المنوالية

ل = طول الفئة

أوجد المنوال بطريقة بيرسون من الجدول التالي :

فئات الدخل	10-	20-	30-	40-	50-	60-	70-80
عدد العمال	5	12	22	38	22	12	5

	ف	ك	
	10-	5	
	20-	12	
	30-	22	ك1
أ	40-	38	ك
	50-	22	ك2
	60-	12	
	70-80	5	

نحسب $ف_1$ = ك – ك1 = 38 – 22 = 16

نحسب $ف_2$ = ك – ك2 = 38 – 22 = 16

نحسب ل = 10

ثم نعوض في القانون :

$$\text{المنوال} = 40 + \frac{16}{16 + 16} \times 10$$

المنوال = 40 + 5 = 45

مقارنة بين المتوسطات الثلاثة : الوسط ، الوسيط ، المنوال

الوسط الحسابي

مزاياه :

- سهولة حسابه
- يأخذ في الاعتبار جميع البيانات
- لا يحتاج إلى ترتيب معين للبيانات

عيوبه :

- يتأثر بشدة بالقيم المتطرفة
- لا يمكن إيجاده بالرسم [بيانياً]
- لا يمكن حسابه في حالات التوزيعات التكرارية المفتوحة

الوسيط

مزاياه :

- سهولة حسابه حسابياً أو بيانياً
- لا يتأثر بالقيم المتطرفة
- يمكن حسابه في حالة التوزيعات التكرارية المفتوحة

عيوبه :

- يحتاج إلى ترتيب للبيانات أولاً
- لا يأخذ في الاعتبار جميع البيانات

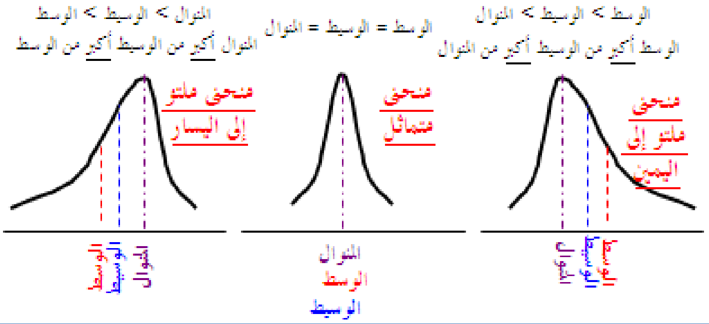
المنوال

مزاياه :

- سهولة حسابه
- لا يتأثر كثيراً بالقيم المتطرفة
- لا يحتاج لترتيب البيانات

عيوبه :

- قد لا يتواجد وقد يكون له أكثر من قيمة



في جميع المنحنيات المبينة ، لاحظ الآتي

المنوال هو القيمة المناظرة لأعلى نقطة في المنحنى

الوسيط يقع دائماً بين الوسط و المنوال

الوسط − المنوال = ٣ × (الوسط − الوسيط)

وهذه العلاقة يمكن وضعها على أي صورة من الصور التالية

$$\text{الوسط} = \frac{(\text{٣} \times \text{الوسيط}) - \text{المنوال}}{\text{٢}}$$

وهذه الصورة مفيدة عندما يكون الوسيط و المنوال معلومان ونريد معرفة الوسط الحسابي

$$\text{المنوال} = (\text{٣} \times \text{الوسيط}) - (\text{٢} \times \text{الوسط})$$

وهذه الصورة مفيدة عندما يكون الوسط الحسابي و الوسيط معلومان ونريد معرفة المنوال

$$\text{الوسيط} = \frac{(\text{٢} \times \text{الوسط}) + \text{المنوال}}{\text{٣}}$$

وهذه الصورة مفيدة عندما يكون الوسط الحسابي و المنوال معلومان ونريد معرفة الوسيط

- فمثلاً إذا كان المنوال لمجموعة من القيم = 95 ، والوسيط لها = 85 ، فإن :

$$\text{الوسط} = \frac{(\text{٣} \times \text{الوسيط}) - \text{المنوال}}{\text{٢}} = \frac{(\text{٣} \times \text{٨٥}) - \text{٩٥}}{\text{٢}} = \frac{\text{٢٢٥} - \text{٩٥}}{\text{٢}} = \frac{\text{١٦٠}}{\text{٢}} = \text{٨٠}$$

- وإذا كان الوسط الحسابي لمجموعة من القيم = 80 ، والوسيط لها = 85 ، فإن :

$$\text{المنوال} = (\text{٣} \times \text{الوسيط}) - (\text{٢} \times \text{الوسط}) = (\text{٣} \times \text{٨٥}) - (\text{٢} \times \text{٨٠}) = \text{٢٥٥} - \text{١٦٠} = \text{٩٥}$$

- وإذا كان الوسط الحسابي لمجموعة من القيم = 80 ، والمنوال لها = 95 ، فإن :

$$\text{الوسيط} = \frac{(\text{٢} \times \text{الوسط}) + \text{المنوال}}{\text{٣}} = \frac{(\text{٢} \times \text{٨٠}) + \text{٩٥}}{\text{٣}} = \frac{\text{١٦٠} + \text{٩٥}}{\text{٣}} = \frac{\text{٢٥٥}}{\text{٣}} = \text{٨٥}$$

سؤال: المنحنى التكراري للبيانات المذكورة في أي من الأمثلة السابقة :

◯ ملتو لليمين   ◯ ملتو لليسار   ◯ متماثل

تمرين: من واقع بيانات الجدول التالي:

الفئات	التكرار
٥ -	٢
١٠ -	٤
١٥ -	٦
٢٠ -	٨
٢٥ -	١٠
٣٠ -	١٦
٣٥ -	٤٠
٤٠ -	٢٤
٤٥ -	١٤
٥٠ -	١١
٥٥ - ٦٠	٥

- احسب الوسط الحسابي بطريقة مراكز الفئات
- احسب الوسيط بطريقتين مختلفتين
- احسب المنوال

المحاضرة 4

مقاييس التشتت

(المدى، الإنحراف المتوسط، التباين، الإنحراف المعياي)

تعريف التشتت

درجة التباعد أو التقارب التي تتجه بها البيانات الكمية للانتشار حول قيمة متوسطة (أحد مقاييس النزعة المركزية) تُسمى تشتت أو تغير البيانات. وتستخدم مقاييس التشتت في المقارنة بين مجموعات البيانات من حيث تشتتها.

كلما قل تشتت البيانات ← كلما اقتربت من متوسطها ← كلما كانت أقرب للتجانس

هل يمكن الاكتفاء بالوسط الحسابي في وصف البيانات؟

إذا كان لدينا 3 مجموعات من الطلاب، كل مجموعة مكونة من خمسة طلاب، وكانت درجاتهم في أحد المقررات كالتالي:

المجموعة الأولى	المجموعة الثانية	المجموعة الثالثة
5,5,5,5,5	3,4,5,6,7	1,2,5,8,9
وسطها الحسابي 5	وسطها الحسابي 5	وسطها الحسابي 5

المجموعات الثلاثة لها وسط حسابي 5 ، لكن في المجموعة الأولى: جميع القيم متساوية وتساوي الوسط 5 ، في حين تنتشر البيانات في المجموعة الثانية حول هذا الوسط بقدرٍ ما ، وفي المجموعة الثالثة تنتشر البيانات حول الوسط بقدر آخر .

أي أن الوسط الحسابي وحده ليس كافياً وحده لوصف البيانات، ولكن لابد من وجود نوع آخر من المقاييس لرصد مدى تشتت البيانات عن تلك القيمة المتوسطة الممثلة للبيانات. هذا النوع من المقاييس هو ما نسميه بـ مقاييس التشتت

أولاً : المدى R :

البيانات الغير مبوبة ← **الفرق بين أكبر قيمة و أصغر قيمة في البيانات و يرمز له بالرمز R**

15 3 7 6 12 18 5 3 13 15

يكون المدى : $R = 18 - 3 = 15$

البيانات المبوبة ← **الفرق بين الحد الأعلى للفئة الأخيرة و الحد الأدنى للفئة الأولى**

الفئة	العمر x
الأولى	$2 \le x < 6$
الثانية	$6 \le x < 12$
الثالثة	$12 \le x < 15$
الرابعة	$15 \le x < 18$

$R = 18 - 2 = 16$

الحد الأعلى للفئة الأخيرة ← 18

الحد الأدنى للفئة الأولى ← 2

وبالرغم من بساطة تحديده إلا أن له بعض العيوب:

- تأثره بالقيم المتطرفة

فمثلاً لمجموعة القيم :	15	3	7	6	12	18	5	3	13	15	يكون المدى	$R=18-3=15$
ولمجموعة القيم :	16	3	14	15	17	18	17	13	14	16	يكون المدى	$R=18-3=15$
	16	15	14	15	17	18	17	13	14	16	يكون المدى	$R=18-15=3$

أي أن المدى واحد للمجموعتين في حين يبدو للعين المجردة أن هناك تشتت للبيانات أكبر في المجموعة الأولى عنه في المجموعة الثانية ، مما يعني أن المدى هنا لا يظهر هذا الفارق.

لذا يُعد المدى مقياساً للتشتت لكنه غير جيد في كثير من الأحيان

- لا يمكن تحديده في حالة التوزيعات التكرارية المفتوحة

الفئة	العمر x
الأولى	$x<6$
الثانية	$6\le x<12$
الثالثة	$12\le x<15$
الرابعة	$15\le x<18$

مفتوح من أسفل

الفئة	العمر x
الأولى	$6\le x<12$
الثانية	$12\le x<15$
الثالثة	$15\le x<18$
الرابعة	$x\ge 18$

مفتوح من أعلى

الفئة	العمر x
الأولى	$x<6$
الثانية	$6\le x<12$
الثالثة	$12\le x<15$
الرابعة	$x\ge 15$

مفتوح من الطرفين

لا يمكن تحديد مدى البيانات

- لا يدخل في حسابه جميع البيانات

مثال :

البيانات التالية لدرجات ذكاء مجموعتين من الأطفال أوجد المدى وقارن بين المجموعتين:

القيمة الكبرى للذكاء	القيمة الصغرى للذكاء	متوسط الذكاء	
112	90	105	المجموعة A
140	75	120	المجموعة B

A المدى $= 140-75=65$ درجة

B المدى $= 112-90=22$ درجة

مثال

البيانات التالية لدرجات الطلاب في مقرر الإحصاء الاجتماعي. احسب المدي لدرجات الطلاب؟

فئات الدرجات	50 —	58 —	66 —	74 —	82 —	90 — 98
عدد الطلاب	3	10	24	40	15	8

المدى $= 98 - 50 = 48$

ثانياً : الانحراف المتوسط [أو متوسط الانحرافات] M.D

يُعرف **الانحراف المتوسط** (أو **متوسط الانحرافات**) [وسنرمز له بالرمز $M.D$] على أنه متوسط القيم المطلقة للانحرافات عن قيمة متوسطة للبيانات [عادةً تكون الوسط الحسابي أو الوسيط] .

فإذا اعتبرنا أن القيمة المتوسطة للبيانات هي الوسط الحسابي ، فإن الانحراف المتوسط لمجموعة من البيانات عددها n يُعطى بـ :

$d = x - \bar{x}$

الانحراف عن الوسط

$$M.D = \frac{\sum |x - \bar{x}|}{n} \quad \text{أو} \quad M.D = \frac{\sum |d|}{n}$$

عدد القيم

ملحوظة هامة : القيمة المطلقة لأي عدد y هي القيمة العددية له دون إشارة ، ونرمز له بنفس الرمز y لكن بين خطين رأسيين | | ، أي نكتب القيمة المطلقة لـ y على الصورة $|y|$. فمثلاً :

$|3| = 3$, $|-3| = 3$, $|2.5| = 2.5$, $|-3.25| = 3.25$

وهكذا .

حيث $d = x - \bar{x}$ هي انحراف القيمة x عن الوسط الحسابي ، $|d|$ هي القيمة المطلقة للانحراف d .

احسب الوسط الحسابي ← ثم ← احسب الانحرافات عن الوسط ← ثم ← احسب القيم المطلقة للانحرافات ← ثم ← احسب متوسط القيم المطلقة

$$\bar{x} = \frac{15+13+3+5+18+12+6+7+3+15}{10} = 9.7$$

15	13	3	5	18	12	6	7	3	15
-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7
5.3	3.3	-6.7	-4.7	8.3	2.3	-3.7	-2.7	-6.7	5.3
5.3	3.3	6.7	4.7	8.3	2.3	3.7	2.7	6.7	5.3

لاحظ أن مجموع الانحرافات = صفر

إذن الانحراف المتوسط هو متوسط القيم المطلقة للانحرافات :

$$M.D = \frac{5.3+3.3+6.7+4.7+8.3+2.3+3.7+2.7+6.7+5.3}{10} = \underline{\underline{4.9}}$$

وسطها الحسابي :

$$\frac{16+14+13+17+18+17+15+14+3+16}{10}=14.3$$

16	14	13	17	18	17	15	14	3	16
-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3
1.7	-0.3	-1.3	2.7	3.7	2.7	0.7	-0.3	-11.3	1.7
1.7	0.3	1.3	2.7	3.7	2.7	0.7	0.3	11.3	1.7

لاحظ أن مجموع الانحرافات = صفر

$$M.D=\frac{1.7+0.3+1.3+2.7+3.7+2.7+0.7+0.3+11.3+1.7}{10}=\underline{\underline{2.64}}$$

أي أن الانحراف المتوسط للمجموعة الثانية من القيم أقل من الانحراف المتوسط للمجموعة الأولى من القيم مما يعني أن المجموعة الثانية أقل تشتتاً من المجموعة الأولى وهذا الأمر لا يمكن ملاحظته عند استخدام المدى كمقياس للتشتت

ويمكن أن يتم حل السؤال السابق وذلك بتنظيم خطواتنا من خلال جداول كالتالي :

المجموعة الأولى $[n=10]$

x	$\bar{x}$	$d=x-\bar{x}$	$\|d\|$
15	9.7	15 – 9.7 = 5.3	5.3
13	9.7	13 – 9.7 = 3.3	3.3
3	9.7	3 – 9.7 = -6.7	6.7
5	9.7	5 – 9.7 = -4.7	4.7
18	9.7	18 – 9.7 = 8.3	8.3
12	9.7	12 – 9.7 = 2.3	2.3
6	9.7	6 – 9.7 = -3.7	3.7
7	9.7	7 – 9.7 = -2.7	2.7
3	9.7	3 – 9.7 = -6.7	6.7
15	9.7	15 – 9.7 = 5.3	5.3
97	97	0	49
$\sum x$		$\sum d$	$\sum\|d\|$

المجموعة الثانية $[n=10]$

x	$\bar{x}$	$d=x-\bar{x}$	$\|d\|$
16	14.3	16 – 14.3 = 1.7	1.7
14	14.3	14 – 14.3 = -0.3	0.3
13	14.3	13 – 14.3 = -1.3	1.3
17	14.3	17 – 14.3 = 2.7	2.7
18	14.3	18 – 14.3 = 3.7	3.7
17	14.3	17 – 14.3 = 2.7	2.7
15	14.3	15 – 14.3 = 0.7	0.7
14	14.3	14 – 14.3 = -0.3	0.3
3	14.3	3 – 14.3 = -11.3	11.3
16	14.3	16 – 14.3 = 1.7	1.7
143	143	0	26.4
$\sum x$		$\sum d$	$\sum\|d\|$

$$\bar{x}=\frac{\sum x}{n}=\frac{97}{10}=9.7 \qquad M.D=\frac{\sum|d|}{n}=\frac{49}{10}=4.9$$

$$\bar{x}=\frac{\sum x}{n}=\frac{143}{10}=14.3 \qquad M.D=\frac{\sum|d|}{n}=\frac{26.4}{10}=2.64$$

- ## وفي حالة البيانات الكمية المتقطعة ذات التكرارات:

- يمكن تحديد الانحراف المتوسط $M.D$ من العلاقة :

$$\boxed{M.D=\frac{\sum f\,|d|}{\sum f}}$$

أي نضرب القيمة المطلقة لانحراف كل قيمة [عن الوسط] في تكرارها ، ثم نقسم الناتج على مجموع التكرارات

فمثلاً : إذا كان المطلوب حساب الانحراف المتوسط للبيانات المبينة بالجدول التكراري :

التكرار f	المتغير x
20	4
40	5
30	6
10	7

الجدول التكراري

المتغير x	التكرار f	fx	$d = x - \bar{x}$	$\|d\|$	$f\|d\|$
4	20	80	$4 - 5.3 = -1.3$	1.3	$20 \times 1.3 = 26$
5	40	200	$5 - 5.3 = -0.3$	0.3	$40 \times 0.3 = 12$
6	30	180	$6 - 5.3 = 0.7$	0.7	$30 \times 0.7 = 21$
7	10	70	$7 - 5.3 = 1.7$	1.7	$10 \times 1.7 = 17$
	100	530			76
	$\sum f = 100$	$\sum fx = 530$			$\sum f\|d\|$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{530}{100} = 5.3$$

$$M.D = \frac{\sum f|d|}{\sum f} = \frac{76}{100} = 0.76$$

انتبه :

مجموع الانحرافات هنا [والذي يجب أن يساوي صفراً] هو $\sum fd$ وليس $\sum d$

- **وفي حالة البيانات الكمية المتصلة :**

- نستخدم نفس العلاقة السابقة لتحديد الانحراف المتوسط $M.D$ ، أي يكون

$$\boxed{M.D = \frac{\sum f|d|}{\sum f}} \quad \text{حيث} \quad \boxed{d = x_0 - \bar{x}}$$

، x_0 تمثل مراكز الفئات

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	$\|d\|$	$f\|d\|$
الأولى	$50 \le x < 60$	6	55	330	$55 - 83.75 = -28.75$	28.75	172.5
الثانية	$60 \le x < 70$	9	65	585	$65 - 83.75 = -18.75$	18.75	168.75
الثالثة	$70 \le x < 80$	15	75	1125	$75 - 83.75 = -8.75$	8.75	131.25
الرابعة	$80 \le x < 90$	12	85	1020	$85 - 83.75 = 1.25$	1.25	15
الخامسة	$90 \le x < 100$	9	95	855	$95 - 83.75 = 11.25$	11.25	101.25
السادسة	$100 \le x < 120$	6	110	660	$110 - 83.75 = 26.25$	26.25	157.5
السابعة	$120 \le x < 180$	3	150	450	$150 - 83.75 = 66.25$	66.25	198.75
		60		5025			945
		$\sum f$		$\sum fx_0$			$\sum f\|d\|$

$$\bar{x} = \frac{\sum f x_0}{\sum f} = \frac{5025}{60} = \underline{\underline{83.75}}$$

$$M.D = \frac{\sum f|d|}{\sum f} = \frac{945}{60} = 15.75$$

ثالثاً : التباين s^2 والانحراف المعياري s

يُعرف متوسط مربعات الانحرافات عن الوسط الحسابي على أنه تباين مجموعة البيانات [ويُرمز له بالرمز s^2] ، ويُعرف الجذر التربيعي للتباين على أنه الانحراف المعياري للبيانات [ويُرمز له بالرمز s] ، أي أن :

التباين $= s^2 = \dfrac{\sum d^2}{n}$ ومنه يكون ← الانحراف المعياري $= s = \sqrt{s^2} = \sqrt{\dfrac{\sum d^2}{n}}$

المجموعة الأولى [$n = 10$]

x	$d = x - \bar{x}$	d^2
15	15 – 9.7 = 5.3	28.09
13	13 – 9.7 = 3.3	10.89
3	3 – 9.7 = -6.7	44.89
5	5 – 9.7 = -4.7	22.09
18	18 – 9.7 = 8.3	68.89
12	12 – 9.7 = 2.3	5.29
6	6 – 9.7 = -3.7	13.69
7	7 – 9.7 = -2.7	7.29
3	3 – 9.7 = -6.7	44.89
15	15 – 9.7 = 5.3	28.09

$$\bar{x} = \frac{\sum x}{n} = 9.7$$

$$s^2 = \frac{\sum d^2}{n} = \frac{274.1}{10} = 27.41$$

$$s = \sqrt{s^2} = \sqrt{27.41} \cong 5.24$$

المجموعة الثانية [$n = 10$]

x	$d = x - \bar{x}$	d^2
16	16 – 14.3 = 1.7	2.89
14	14 – 14.3 = -0.3	0.09
13	13 – 14.3 = -1.3	1.69
17	17 – 14.3 = 2.7	7.29
18	18 – 14.3 = 3.7	13.69
17	17 – 14.3 = 2.7	7.29
15	15 – 14.3 = 0.7	0.49
14	14 – 14.3 = -0.3	0.09
3	3 – 14.3 = -11.3	127.69
16	16 – 14.3 = 1.7	2.89

$$\bar{x} = \frac{\sum x}{n} = 14.3$$

$$s^2 = \frac{\sum d^2}{n} = \frac{164.1}{10} = 16.41$$

$$s = \sqrt{s^2} = \sqrt{16.41} \cong 4.05$$

- وفي حالة البيانات الكمية المتقطعة ذات التكرارات :
- يمكن تحديد التباين s^2 والانحراف المعياري s من :

التباين $= s^2 = \dfrac{\sum fd^2}{\sum f}$ ومنه يكون ← الانحراف المعياري $= s = \sqrt{s^2} = \sqrt{\dfrac{\sum fd^2}{\sum f}}$

فمثلاً : إذا كان المطلوب حساب الانحراف المعياري للبيانات المبينة بالجدول التكراري

الجدول التكراري

المتغير x	التكرار f
4	20
5	40
6	30
7	10

الجدول التكراري

المتغير x	التكرار f	fx	$d = x - \bar{x}$	d^2	fd^2
4	20	80	4 – 5.3 = -1.3	1.69	20 × 1.69 = 33.8
5	40	200	5 – 5.3 = -0.3	0.09	40 × 0.09 = 3.6
6	30	180	6 – 5.3 = 0.7	0.49	30 × 0.49 = 14.7
7	10	70	7 – 5.3 = 1.7	2.89	10 × 2.89 = 28.9
	100	530			81
	$\sum f = 100$	$\sum fx = 530$			$\sum fd^2$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{530}{100} = 5.3$$

$$s^2 = \frac{\sum fd^2}{\sum f} = \frac{81}{100} = 0.81$$

$$s = \sqrt{s^2} = \sqrt{0.81} = 0.9$$

- **وفي حالة البيانات الكمية المتصلة :**

- تُستخدم نفس العلاقة السابقة لتحديد الانحراف المعياري s:

$$\text{التباين} = s^2 = \frac{\sum fd^2}{\sum f} \quad \xrightarrow{\text{ومنه يكون}} \quad \text{الانحراف المعياري} = s = \sqrt{s^2} = \sqrt{\frac{\sum fd^2}{\sum f}} \quad \text{حيث} \quad d = x_0 - \bar{x}$$

أي أنه عند حساب الانحرافات نعتبر أن مركز أي فئة يمثل جميع القيم الموجودة في تلك الفئة:

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	d^2	fd^2
الأولى	$0 \le x < 20$	4	10	40	$10 - 31.7 = -21.7$	470.89	$4 \times 470.89 = 1883.56$
الثانية	$20 \le x < 30$	16	25	400	$25 - 31.7 = -6.7$	44.89	$16 \times 44.89 = 718.24$
الثالثة	$30 \le x < 35$	12	32.5	390	$32.5 - 31.7 = 0.8$	0.64	$12 \times 0.64 = 7.68$
الرابعة	$35 \le x < 40$	10	37.5	375	$37.5 - 31.7 = 5.8$	33.64	$10 \times 33.64 = 336.4$
الخامسة	$40 \le x < 50$	6	45	270	$45 - 31.7 = 13.3$	176.89	$6 \times 176.89 = 1061.34$
السادسة	$50 \le x < 60$	2	55	110	$55 - 31.7 = 23.3$	542.89	$2 \times 542.89 = 1085.78$
		50		1585			5093
		$\sum f$		$\sum fx_0$			$\sum fd^2$

$$\therefore \quad \bar{x} = \frac{\sum f x_0}{\sum f} = \frac{1585}{50} = 31.7 \qquad s^2 = \frac{\sum fd^2}{\sum f} = \frac{5093}{50} = 101.86 \qquad s = \sqrt{s^2} = \sqrt{101.86} \cong 10.09$$

وبنفس الأسلوب يمكن التعامل مع المثال السابق لحساب التباين والانحراف المعياري

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	d^2	$f \times d^2$
الأولى	$50 \le x < 60$	6	55	330	$55 - 83.75 = -28.75$	826.56	4959.38
الثانية	$60 \le x < 70$	9	65	585	$65 - 83.75 = -18.75$	351.56	3164.04
الثالثة	$70 \le x < 80$	15	75	1125	$75 - 83.75 = -8.75$	76.56	1148.4
الرابعة	$80 \le x < 90$	12	85	1020	$85 - 83.75 = 1.25$	1.56	18.72
الخامسة	$90 \le x < 100$	9	95	855	$95 - 83.75 = 11.25$	126.56	1139.04
السادسة	$100 \le x < 120$	6	110	660	$110 - 83.75 = 26.25$	689.06	4134.36
السابعة	$120 \le x < 180$	3	150	450	$150 - 83.75 = 66.25$	4389.06	13167.18
		60		5025			27731.12
		$\sum f$		$\sum fx_0$			$\sum fd^2$

$$\bar{x} = \frac{\sum f x_0}{\sum f} = \frac{5025}{60} = 83.75 \qquad s^2 = \frac{\sum fd^2}{\sum f} = \frac{27731.12}{60} \cong 462.19 \qquad s = \sqrt{s^2} = \sqrt{462.19} \cong 21.5$$

من السابق يتضح أن كلاً من الانحراف المتوسط والانحراف المعياري يعتمدا تماماً في حساباتهما على الوسط الحسابي ، وبالتالي فلهما نفس مزايا وعيوب الوسط الحسابي . أي :

المزايا :

- من السهل حسابهما
- يأخذ في الاعتبار جميع البيانات
- لا يحتاجا لترتيب معين للبيانات

العيوب :

- يتأثرا بشدة بالقيم المتطرفة
- لا يمكن إيجادهما بالرسم (بيانياً)
- لا يمكن حسابهما للتوزيعات التكرارية المفتوحة
- ويمكن تلخيص كل ما يخص الوسط الحسابي والانحراف المتوسط والانحراف المعياري في الآتي :
- للقيم المفردة :

قيم عددها n	الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات
x	$d = x - \bar{x}$	$\|d\|$	d^2
...	...	...	...
...	...	...	...
$\sum x$		$\sum \|d\|$	$\sum d^2$

الوسط الحسابي = $\bar{x} = \frac{\sum x}{n}$

الانحراف المتوسط = $M.D = \frac{\sum |d|}{n}$

$s^2 = \frac{\sum d^2}{n} \longrightarrow s = \sqrt{s^2}$

- ولتوزيع تكراري :

القيم	التكرار		الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات		
x	f	fx	$d = x - \bar{x}$	$\|d\|$	d^2	$f\|d\|$	fd^2
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
	$\sum f$	$\sum fx$				$\sum f\|d\|$	$\sum fd^2$

$\bar{x} = \frac{\sum fx}{\sum f}$

$M.D = \frac{\sum f|d|}{\sum f}$

$\longrightarrow s = \sqrt{s^2}$

- وللبيانات المتصلة :

الفئات	التكرار			الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات		
x	f	مراكز الفئات	fx_0	$d = x_0 - \bar{x}$	$\|d\|$	d^2	$f\|d\|$	fd^2
...	...	x_0	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
	$\sum f$	...	$\sum fx$				$\sum f\|d\|$	$\sum fd^2$

خاصيتان هامتان للانحراف المتوسط والانحراف المعياري :

الخاصية الأولى : إضافة عدد ثابت c لكل قيمة من قيم البيانات لا يؤثر على قيمة الانحرافين المتوسط والمعياري .

الانحراف المتوسط (أو المعياري) الجديد = الانحراف المتوسط (أو المعياري) القديم

الخاصية الثانية : ضرب كل قيمة من قيم البيانات في عدد ثابت c يجعل :

الانحراف المتوسط (أو المعياري) الجديد = الانحراف المتوسط (أو المعياري) القديم × القيمة المطلقة للثابت c

فمثلاً، لو كانت لدينا البيانات التالية والتي توضح درجات مجموعة من الطلاب كالتالي :

بعد إضافة 5 لكل درجة			
x	d	$\|d\|$	d^2
14	1	1	1
7	-6	6	36
12	-1	1	1
17	4	4	16
15	2	2	4
65		14	58

$$\bar{x} = \frac{\sum x}{n} = \frac{65}{5} = 13$$

$$M.D = \frac{\sum |d|}{n} = \frac{14}{5} = 2.8$$

$$s^2 = \frac{\sum d^2}{n} = \frac{58}{5} = 11.6$$

$$s = \sqrt{11.6} \cong 3.4$$

الدرجات الأصلية			
x	d	$\|d\|$	d^2
9	1	1	1
2	-6	6	36
7	-1	1	1
12	4	4	16
10	2	2	4
40		14	58

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{5} = 8$$

$$M.D = \frac{\sum |d|}{n} = \frac{14}{5} = 2.8$$

$$s^2 = \frac{\sum d^2}{n} = \frac{58}{5} = 11.6$$

$$s = \sqrt{11.6} \cong 3.4$$

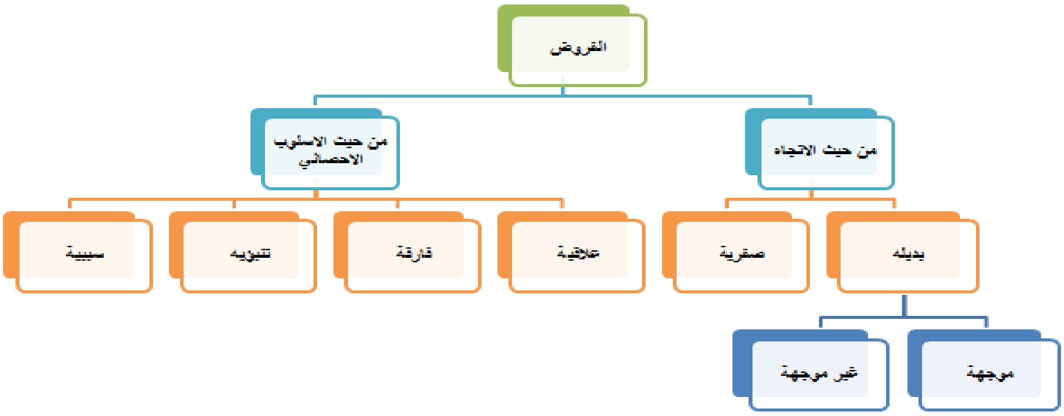
الفروض الإحصائية

يعرف الفرض بأنه إجابة متوقعة لسؤال من الأسئلة التي تراود ذهن الباحث أو المهتم، وهذه الإجابة لا تكون نهائية وإنما خاضعة للدراسة والتحقق من مدى صحتها فإما أن تكون الإجابة صحيحة وإما أن تكون الإجابة خاطئة.

وتوقع الإجابة من جانب الباحث لا يتم من فراغ وإنما بناءً على خلفية نظرية متعلقة بهذا السؤال ونتائج دراسات سابقة حوله.

فمثلاً: يراود ذهن الباحث سؤال مضمونه: ما طبيعة العلاقة بين حب الاستطلاع والقدرة الابتكارية لدى طلاب قسم علم الاجتماع؟ وبناء على الخلفية النظرية ونتائج الدراسات السابقة المرتبطة بطبيعة العلاقة بين المتغيرين يصيغ الباحث إجابة متوقعة لهذا السؤال وهي تمثل إحدى فروض بحثه وتكون صياغة الفرض كالتالي:

- توجد علاقة بين حب الاستطلاع والابتكارية
- لا توجد علاقة بين حب الاستطلاع والابتكارية
- الفرض هو اقتراح لقضية معينة وبالتالي فإن قرار قبولنا هذا الاقتراح كاقتراح صحيح أو رفضنا إياه كاقتراح خاطئ لا بد أن يؤجل حتى نجمع دليل يؤكد قبوله أو رفضه.



- الفرضية الصفرية (فرضية العدم) H_0 (Null Hypothesis) :

هي الفرضية حول معلمة المجتمع التي نجري اختبار عليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصائي من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما. وهي الفرضية التي ننطلق منها ونرفضها عندما تتوفر دلائل على عدم صحتها، وخلاف ذلك نقبلها وتعني كلمة Nul انه لا يوجد فرق بين معلمة المجتمع والقيمة المدعاة (إحصائية العينة).

- الفرضية البديلة (H_a) Alternative Hypothesis :
- هي الفرضية التي يضعها الباحث كبديل عن فرضية العدم و نقبلها عندما نرفض فرضية العدم باعتبارها ليست صحيحة بناء على المعلومات المستقاة من العينة.
- وفي اختبار الفروض يمكن أن نرتكب نوعين من الخطأ:
- الخطأ من النوع الأول Type I error: الخطأ من النوع الأول هو "رفض الفرض العدمي بينما هو صحيح ". أي أنه على الرغم من أن الفرض العدمي في الواقع صحيح وكان من الواجب قبوله فقد تم أخذ قرار خاطئ برفضه. وباختصار شديد فإن الخطأ من النوع الأول هو : " رفض فرض صحيح". ويرمز له بالرمز α .

- الخطأ من النوع الثاني Type II error: وفي المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرض العدمي بينما هو خاطئ " أي أنه على الرغم من أن الفرض العدمي خاطئ وكان من الواجب رفضه فقد تم أخذ قرار خاطئ بقبوله وباختصار شديد فإن الخطأ من النوع الثاني هو " قبول فرض خاطئ ". ويرمز له بالرمز β .

- ويمكن أن نمثل ذلك في الجدول التالي:

القرار / الفرضية	صحيحة (H_0)	خاطئة (H_a)
قبول (H_0)	صواب	خطأ ٢ بيتا (B)
رفض (H_0)	خطأ ١ ألفا (α)	صواب

١. **فرضية صحيحة نتائج العينة تؤيد صحتها. (قبول صواب)**

١. **فرضية صحيحة نتائج العينة غير مؤيدة لصحتها. (رفض صواب)**
وهذا يعطينا خطأ من النوع الأول ألفا (α)

١. **فرضية خاطئة نتائج تؤيد صحتها (قبول خطأ) وهذا يعطينا خطأ من النوع الثاني بيتا (B) ويمكن أن يقلل بزيادة حجم العينة**

١. **فرضية خاطئة نتائج غيرمؤيدة صحتها (رفض خطأ)**

الفروض البحثية:

هي الفروض التي يصيغها الباحث في بحثه بناءً على خلفيته النظرية ونتائج الدراسات السابقة.

1. الفروض العلاقية:

أ. الفرض البديل العلاقي غير الموجه:

توجد علاقة دالة إحصائياً بين الاتجاه نحو الدراسة والبيئة الدراسية

ب. الفرض البديل العلاقي الموجه:

توجد علاقة ايجابية دالة إحصائياً بين الاتجاه نحو الدراسة والبيئة الدراسية

ج. الفرض الصفري العلاقي:

لا توجد علاقة دالة إحصائياً بين نحو الدراسة والبيئة الدراسية

2. الفروض الفارقة:

أ. الفرض البديل الفارق غير الموجه:

توجد فروق ذات دلالة إحصائية بين متوسطي درجات الذكور والإناث في الذكاء الوجداني

ب. الفرض البديل الفارق الموجه:

توجد فروق ذات دلالة إحصائية بين متوسطي درجات الذكور والإناث في الذكاء الوجداني لصالح الذكور

ج. الفرض الصفري الفارق:

لا توجد فروق ذات دلالة إحصائية بين متوسطي درجات الذكور والإناث في الذكاء الوجداني

3. الفروض التنبؤية:

أ. الفرض البديل التنبؤي غير الموجه:

يمكن التنبؤ من المتغيرات المستقلة (الدافعية، وحب الاستطلاع، والقلق) بالمتغير التابع (التحصيل الدراسي) لدى طلاب جامعة الملك فيصل

ب. الفرض البديل التنبؤي الموجه:

يمكن التنبؤ من المتغيرات المستقلة (الدافعية كمنبئ موجب، وحب الاستطلاع كمنبئ موجب، والقلق كمنبئ سالب) بالمتغير التابع (التحصيل الدراسي) لدى طلاب جامعة الملك فيصل

ج. الفرض الصفري التنبؤي :

لا يمكن التنبؤ من المتغيرات المستقلة (الدافعية، وحب الاستطلاع، والقلق) بالمتغير التابع (التحصيل الدراسي) لدى طلاب جامعة الملك فيصل

4. الفروض السببية:

أ. الفرض البديل السببي غير الموجه:

يمكن التوصل إلى نموذج سببي يفسر العلاقة بين المتغيرات المستقلة (المعاملة الوالدية، والذكاء، والضغوط النفسية، والاتجاه نحو الدراسة) والمتغير التابع (مستوى الطموح) لدى طلاب جامعة الملك فيصل

ب. الفرض البديل السببي الموجه:

يمكن التوصل إلى نموذج سببي يفسر العلاقة بين المتغيرات المستقلة (المعاملة الوالدية «تأثير موجب»، والذكاء «تأثير موجب»، والضغوط النفسية تأثير سالب»، والاتجاه نحو الدراسة «تأثير موجب») والمتغير التابع (مستوى الطموح) لدى طلاب جامعة الملك فيصل

ج. الفرض الصفري السببي:

لا يمكن التوصل إلى نموذج سببي يفسر العلاقة بين المتغيرات المستقلة (المعاملة الوالدية، والذكاء، والضغوط النفسية، والاتجاه نحو الدراسة) والمتغير التابع (مستوى الطموح) لدى طلاب جامعة الملك فيصل

الفروض الإحصائية:

ما الفرق بين الفروض البحثية والفروض الاحصائية؟

الفروض البحثية

هي الفروض التي يصيغها الباحث بنفسه في ضوء اطلاعه على الخلفية النظرية ونتائج الدراسات السابقة، وبناء على اطلاعه يحدد اتجاه الفرض هل هو فرض بديل موجه أم فرض بديل غير موجه أم فرض صفري.

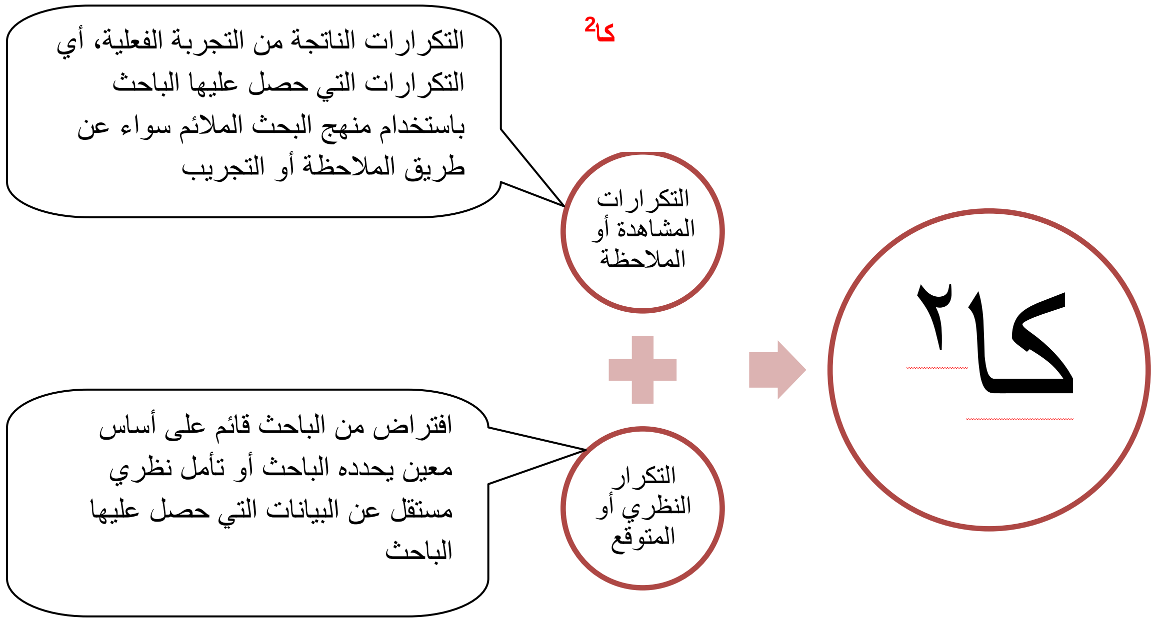
أما الفروض الإحصائية

فتهدف إلى تفسير نتيجة معالجة الأسلوب الإحصائي للفرض البحثي، والذي بناء عليه نتقبل الفرض البحثي أو نرفضه، وبالتالي فالذي يجعلنا نقبل الفرض البحثي ليس الأسلوب الإحصائي فقط ولكن الفرض الإحصائي المرتبط به.

المحاضرة السادسة

مربع كاي

كا^2



هل تحب الإحصاء؟

من المتوقع أن يجيب 50 منهم بـ (نعم) ويجيب 50 الآخرين بـ (لا) وهذا ما يسمى بالتكرار المتوقع Expected Frequency حيث إن:

$$\text{التكرار المتوقع} = \frac{\text{عدد أفراد العينة}}{\text{عدد الاستجابات}}$$

لكن ما حدث أن أجاب 20منهم بـ (نعم) ، وأجاب 80 بـ (لا) وهذا ما يسمى بالتكرار المشاهد أو الملاحظ Observed Frequency

يتعامل اختبار مربع كاي مع تكرارات البيانات الاسمية

هل يتعامل اختبار مربع كاي مع تكرارات البيانات الفترية أو الرتبية؟

السؤال : هل توجد فروق بين من قالوا نعم وبين من قالوا لا

اختبار **كا^2** Chi-squared Test هو أحد اختبارات الدلالة الإحصائية اللابارامترية

يتعامل مع **تكرارات الدرجات** وليس الدرجات نفسها، ويستخدم في دراسة الفروق بين تكرارات استجابات أفراد عينة ما على سؤال أو عدة أسئلة.

ويتم حساب اختبار (كا^2) من المعادلة التالية :

$$x^2 = \sum \frac{(O - E)^2}{E}$$

حيث:

O : التكرار المشاهد Observed Frequency

E : التكرار المتوقع Expected Frequency

$$كا^2 = مجـ \frac{(ت_و - ت_م)^2}{ت_و}$$

حيث :

ت$_و$: هو التكرار الواقعي الذي يحدث بالفعل والموجود بالجدول .

ت$_م$: هو التكرار المتوقع حدوثه ويختلف حسابه بـــاختلاف نـــوع الجدول المطلوب حساب كا^2 منه .

مثال:

الرأي	موافق	لا أدري	معارض	مج
التكرار	12	2	16	30

حساب التكرار المتوقع (ت$_م$) :

$$ت_م = \frac{16 + 2 + 12}{3} = 10$$

ت$_و$	ت$_م$	ت$_و$ - ت$_م$	(ت$_و$ - ت$_م$)^2	$\frac{(ت_و - ت_م)^2}{ت_م}$
12	10	2	4	0.4
2	10	8-	64	6.4
16	10	6	36	3.6
-	-	-	مجموع	10.4

درجة الحرية = عدد الأعمدة – 1 = 3 – 1 = 2

بالبحث فى جداول كا^2 عند درجة حرية = 2 ومستوى دلالة 0.05

نجد قيمة كا^2 الجدولية = 5.991 .

تحديد مدى دلالة كا^2 :

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

قيمة كا^2 المحسوبة = 10.4 > قيمة كا^2 الجدولية = 5.991

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

القرار:

نقارن كا^2 المحسوبة بالجدولية، فعندما تكون قيمة كا^2 المحسوبة **أكبر** من قيمة كا^2 المجدولة فإننا **نرفض** الفرضية الصفرية أو فرض العدم والتي تنص على أنه لا توجد أي علاقة بين المتغيرين ونقبل الفرض البديل والتي تثبت وجود علاقة بين المتغيرين تحت الدراسة.

أما إذا كانت قيمة كا^2 المحسوبة **أقل** من قيمة كا^2 المجدولة فإننا **نقبل** الفرضية الصفرية أو فرض العدم

طريقة أخرى:

الرأي	موافق	لا أدري	معارض	مج
التكرار	12	2	16	30

$$x^2 = \sum \frac{(O - E)^2}{E}$$

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

$$x^2 = \frac{4}{10} + \frac{64}{10} + \frac{36}{10}$$

$$x^2 = 10.4$$

تمرين:

قام باحث بتطبيق استبيان على مجموعة من الأفراد لأخذ أراءهم في قضية الدروس الخصوصية وذلك بتوجيه سؤال واحد إليهم: هل توافق على الدروس الخصوصية (نعم – لا ولكن بشرط – لا)، فحصل على التكرارات التالية:

الاستجابة	نعم	لا ولكن بشروط	لا
التكرار	21	54	14

المطلوب اختبار الفرض البحثي: لا يختلف التكرار التجريبي الذي حصلنا عليه من استجابات الأفراد على قضية الدروس الخصوصية عن التكرار النظري.

مثال:

أراد معلم معرفة علاقة نجاح تلاميذه في المقرر الذي يقوم بتدريسه بأماكنهم في الفصل، فحسب عدد الناجحين في الامتحان وعدد الراسبين وحدد منهم عدد الجالسين في المقاعد الأمامية وعدد الجالسين في المقاعد الخلفية فتوصل إلى الجدول التالي:

	مقاعد أمامية	مقاعد خلفية	المجموع
ناجح	27	9	36
راسب	4	20	24
المجموع	31	29	60

المطلوب اختبار الفرض البحثي:

توجد علاقة بين نجاح التلاميذ في الامتحان وبين أماكنهم في الفصل.

	مقاعد أمامية	مقاعد خلفية	المجموع
ناجح	٢٧	٩	٣٦
راسب	٤	٢٠	٢٤
المجموع	٣١	٢٩	٦٠

	ك و	ك م	(ك و - ك م)	(ك و - ك م)2	(ك و - ك م)2 / ك م
ناجح – مقاعد أمامية	27	18,6	8,4	70,56	3,79
ناجح – مقاعد خلفية	9	17,4	8,4 -	70,56	4,06
راسب – مقاعد أمامية	4	12,4	8,4 -	70,56	5,69
راسب في مقاعد خلفية	20	11,6	8,4	70,56	6,08

المجموع	60	60	صفر		$كا^2 = 19,62$

ك م (لأي خلية) =	**حاصل ضرب مجموعي تكرارات الصف والعمود المنتميان إليهما الخلية**
	المجموع الكلي للتكرارات

الطريقة المختصرة لحساب مربع كاي من الجدول التكراري 2 × 2

	مقاعد أمامية	مقاعد خلفية	المجموع
ناجح	27 أ	9 ب	36 ح
راسب	4 ج	20 د	24 ز
المجموع	31 هـ	29 و	60 ن

$$كا^2 = فاى^2 \times ن$$

حيث :

فاى : هو معامل ارتباط فاى والذى يحسب من العلاقة :

$$فاى = \frac{أ \times د - ب \times جـ}{\sqrt{هـ \times و \times ز \times ح}}$$

$$فاي = \frac{(4 \times 9) - (20 \times 27)}{\sqrt{24 \times 36 \times 29 \times 31}}$$

فاي = 0,57 مربع فاي = 0,33

$كا^٢ = ٠,٣٣ \times ٦٠$

$= ١٩,٦٢$

المحاضرة السابعة

معامل الارتباط

وعندما نقول مقاييس العلاقة نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر مثلا، كأن يكون الهدف معرفة هل هناك علاقة بين مستوى الإنتاجية وجودة المنتج في مصنع ما؟، أي هل كلما زادت الإنتاجية تقل جودة المنتج أو العكس .

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+1) وبين الارتباط السالب التام (-1)

العلاقة الطردية بين المتغيرات: هو تعبير يشير إلى تزايد المتغيرين المستقل والتابع معا، فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجـة تمثله هي (+1) .

العلاقة العكسية بين المتغيرات: هو تعبير يشير إلى تزايد في متغير يقابله تناقص في المتغير الآخر، فإذا كانت الإنتاجية منخفضة ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط سالب، وأعلى درجة تمثله هي (-1) .

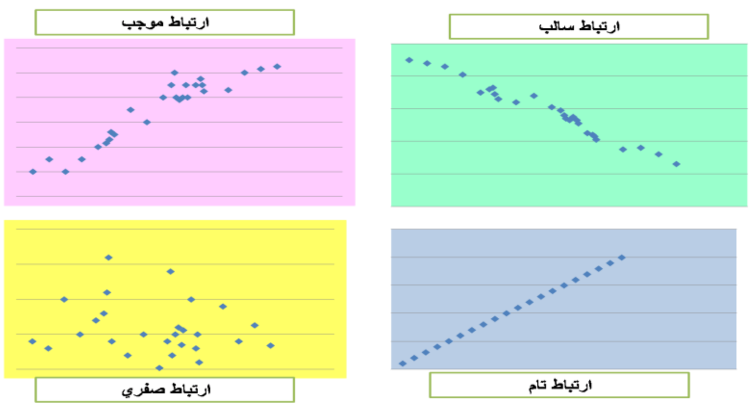
إن معامل الارتباط التام الموجب (+1) يعنى التغير في اتجاه واحد في كلا الظاهرتين مع بقاء الأوضاع النسبية لوحدات الظاهرة ثابتة، سواء كان هذا التغير في اتجاه الزيادة (أي زيادة قيم الظاهرة الأولى تتبعها زيادة في قيم الظاهرة الأخرى)، أو في اتجاه النقص (أي نقص قيم الظاهرة الأولى يتبعها نقص في قيم الظاهرة الأخرى) .

طرق التعرف على العلاقة بين متغيرين وحسابها

أولا: طريقة شكل الانتشار Scatter Diagram :

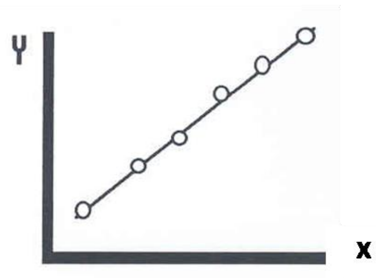
هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذه المحاضرة.

والمقصود **بشكل الانتشار** هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة. والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

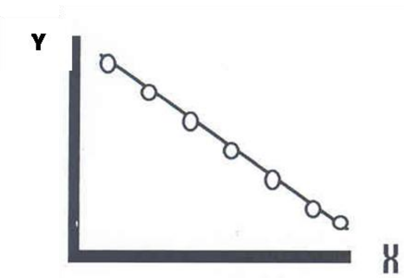


الشكل الأول :

إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة. وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ). ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية. أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد فإن "الارتباط عكسي تام" كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.



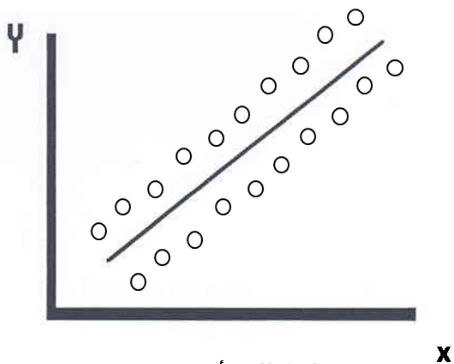
الشكل الأول (أ) ارتباط طردي تام (موجب)



الشكل الأول (ب) ارتباط عكسي تام (سالب)

الشكل الثاني :

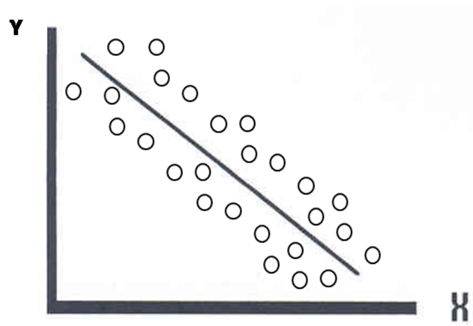
أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ، ب.



الشكل الثاني (أ)

ارتباط موجب قوي

(ارتباط خطي طردي)



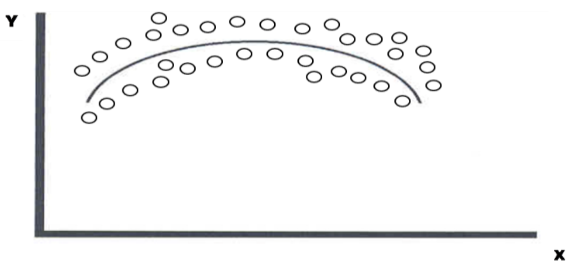
الشكل الثاني (ب)

ارتباط سالب قوي

(ارتباط خطي عكسي)

الشكل الثالث :

وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :

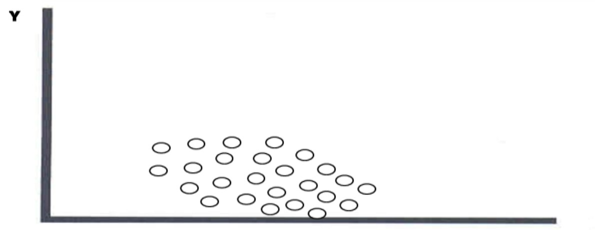


الشكل الثالث

(ارتباط غير خطي)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



الشكل الرابع
(لا توجد علاقة)

ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغيرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + 1، - 1.

* فإذا كانت قيمة معامل الارتباط تساوي + 1 فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.

* وإذا كانت قيمة معامل الارتباط تساوي – 1 فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.

* وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.

* وكلما اقتربت قيمة معامل الارتباط من + 1 أو – 1 كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + 1 أو – 1 فإذا وصلت قيمة المعامل إلى + 1 أو – 1 كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + 1 ولا أصغر من – 1. ويمكن تمثيل قوة العلاقة بالشكل التالي:

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
من ٠.٧٠ إلى ٠.٩٩	ارتباط طردي قوي
من ٠.٥٠ إلى ٠.٦٩	ارتباط طردي متوسط
من ٠.٠١ إلى ٠.٤٩	ارتباط طردي ضعيف
0	لا يوجد ارتباط

يمكن تفسير الارتباط العكسي بنفس الطريقة مع المعاملات السالبة

معامل بيرسون للارتباط الخطى البسيط

Simple Correlation

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام (مثل الارتباط بين العمر والدخل) حيث يقاس العمر بالسنوات ويقاس الدخل بالعملة، بالريال أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي **أربعين** عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون **خمسة آلاف** ريال مثلاً.

وبالتالي فإن أفضل مقياس للارتباط بين مثل هذين المتغيرين – حسب رأي بيرسون – هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير. ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

ويتم حساب معامل ارتباط بيرسون من خلال العلاقــة التالية:

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

حيث :

تعني مجموع حاصل ضرب كل قيمة من X في Y.

تعني مجموع قيم المتغير X .

تعني مجموع قيم المتغير Y .

تعني مجموع مربع قيم المتغير X .

تعني مربع مجموع قيم المتغير X .

تعني مجموع مربع قيم المتغير Y .

تعني مربع مجموع قيم المتغير Y .

عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

مثال :

رغبة إحدى الشركات معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية لهم ، فقاموا بجمع معلومات عن هذا الموضوع وحصلوا على النتائج التالية :

الموظفين	أ	ب	ج	د	هـ	و	ز	ح	ط	ي
ساعات العمل X	٨	٢	٨	٥	١٥	١١	١٣	٦	٤	٦
مستوى الإنتاجية Y	٣	١	٦	٣	١٤	١٢	٩	٤	٤	٥

المطلوب: حساب معامل ارتباط بيرسون للبيانات السابقة .

الموظفين	ساعات العمل X	مستوى الإنتاجية Y	X^2	Y^2	XY
أ	٨	٣	٦٤	٩	٢٤
ب	٢	١	٤	١	٢
ج	٨	٦	٦٤	٣٦	٤٨
د	٥	٣	٢٥	٩	١٥
هـ	١٥	١٤	٢٢٥	١٩٦	٢١٠
و	١١	١٢	١٢١	١٤٤	١٣٢
ز	١٣	٩	١٦٩	٨١	١١٧
ح	٦	٤	٣٦	١٦	٢٤
ط	٤	٤	١٦	١٦	١٦
ي	٦	٥	٣٦	٢٥	٣٠
المجموع	٧٨	٦١	٧٦٠	٥٣٣	٦١٨

نطبق معادلة معامل ارتباط بيرسون كالتالي :

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

$$r = \frac{618 - \frac{(78)(61)}{10}}{\sqrt{\left(760 - \frac{(78)^2}{10}\right)\left((533) - \frac{(61)^2}{10}\right)}} = \frac{618 - 475.8}{\sqrt{(760 - 608.4)(533 - 372.1)}}$$

$$= \frac{1422}{\sqrt{(151.6)(160.9)}} = \frac{1422}{\sqrt{24392.44}} = \frac{1422}{156.2} = 0.91$$

وهذه النتيجة توضح أن درجة الارتباط = 0.91 وهذه النتيجة تعتبر مؤشر على علاقة إيجابية قوية بين ساعات العمل ومستوى الإنتاجية.

ملاحظة مهمة :

من خواص معامل بيرسون للارتباط الخطي أنه لا يتأثر بالعمليات الحسابية التي تجري على المتغيرين y , x. بمعنى أنه لا يتأثر بالطرح (أو الجمع)، ولا بالقسمة (أو الضرب). أي إذا طرحنا (أو جمعنا) **قيمة معينة** من كل قيم x وقيمة أخرى من كل قيم y، أو قسمنا (أو ضربنا) قيم x على قيمة معينة وكل قيم y على قيمة أخرى فإن قيمة معامل الارتباط لا تتغير أي نحصل على القيمة نفسها.

- معامل ارتباط الرتب

Rank Correlation

قد يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، فقد يكون المتغيران **وصفيين ترتيبين** Ordinal **أو أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً ترتيبياً**، أو أن يكون **المتغيران كميين**، ويكون اهتمام الباحث منصباً على الرتب أكثر من القيم. ففي انتخابات مجلس الشيوخ أو النواب الأمريكي مثلاً، يعتبر المرشح الأول هو من حصل على أعلى الأصوات بغض النظر عن عددها، والذي يحصل على عدد أصوات أقل منه مباشرة هو الثاني.. وهكذا.

فإذا كانت رتب المتغيرين تسير في الاتجاه نفسه: بمعنى أن الرتب الأعلى للمتغير الأول تناظرها رتب أعلى للمتغير الثاني كانت العلاقة طردية بينهما. وإذا كانت الرتب الأعلى للمتغير الأول تناظرها رتب أدنى للمتغير الثاني كانت العلاقة بينهما عكسية. ففي مثالنا السابق عن العلاقة بين دخل الناخب وعمره، كان الناخب الأكبر عمراً (بصفة عامة) هو الأعلى دخلاً، فمن الواضح أن العلاقة بينهما طردية، أما إذا كان الناخب الأكبر عمراً (بصفة عامة) هو الأقل مشاركة في العمل السياسي فإننا في هذه الحالة نكون أمام علاقة عكسية.

ولحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

حساب معامل سبيرمان لارتباط الرتب

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأعلى منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأقل منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز **d**) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2 - 1)}$$

حيث : $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين

n هي عدد أزواج القيم

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي :

1 – مجموع الفروق بين الرتب يساوي صفر.

2 – أن قيمة معامل ارتباط الرتب تنحصر بين – 1، + 1 فإذا كانت الرتبة رقم 1 للمتغير الأول تناظرها الرتبة 1 للمتغير الثاني، والرتبة 2 للمتغير الأول تناظرها الرتبة رقم 2 للمتغير الثاني، وهكذا.. فإن معامل ارتباط الرتب يساوي + 1 (ارتباط طردي تام بين الرتب).

وإذا كانت الرتبة رقم 1 (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – 1 (ارتباط عكسي تام بين الرتب).

3 – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي

مثال:

البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي، ومدى ملاءمتها لحاجات الناس.

السؤال الأول	جيدة	مقبولة	ممتازة	جيدة	جيدة جداً	مقبولة	جيدة
السؤال الثاني	جيدة جداً	مقبولة	جيدة جداً	جيدة	جيدة	جيدة	ممتازة

والمطلوب: حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

السؤال الأول X	السؤال الثاني Y	رتب X	رتب Y	d الفرق بين الرتب	d^2 مربعات الفرق
جيدة	جيدة جداً	4	2.5	1.5	2.25
مقبولة	مقبولة	6.5	7	- 0.5	0.25
ممتازة	جيدة جداً	1	2.5	- 1.5	2.25
جيدة	جيدة	4	5	- 1.0	1.00
جيدة جداً	جيدة	2	5	- 3.0	9.00
مقبولة	جيدة	6.5	5	1.5	2.25
جيدة	ممتازة	4	1	3.0	9.00
المجموع				**Zero**	**26.0**

$$r_s = 1 - \frac{6\left(\sum d^2\right)}{n\left(n^2 - 1\right)} = 1 - \frac{6\,(26)}{7\,(49 - 1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

وهذا يعني أن الارتباط بين إجابات المبحوثين بالنسبة للسؤالين هو ارتباط طردي متوسط. وبالتالي فليس بالضرورة أن يكون رأي المجيبين في برامج الضمان الاجتماعي تعني ملاءمتها لحاجات الناس.

لدراسة العلاقة بين تقدير الطالبة في الإحصاء و تقديرها في الرياضيات ، اخترنا خمس طالبات و كانت تقديراتهم كالتالي:

تقدير الإحصاء X	A	C	D	F	A
تقدير الرياضيات y	B	C	B	D	A

رتب y	رتب x	y	x	d	d^2
A	B	5	4	1	1
C	C	3	3	0	0
D	B	2	4	-2	4
F	D	1	2	-1	1
A	A	5	5	0	0
∑					6

$$r_S = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$r_S = 1 - \frac{6 \times 6}{5(5^2 - 1)} = 1 - \frac{36}{5 \times 24} = 1 - \frac{36}{120}$$
$$= 1 - 0.3 = 0.7$$

- **معامل بوينت بايسيريال Point Biserial للارتباط**

يستخدم لقياس الارتباط بين متغير كمي X و متغير اسمي Y مستويين (نعم –لا) أو (ذكر – أنثى) و غيرها.
اشارة معامل الارتباط ليس لها معنى في حالة المتغيرات الوصفية فتقاس قوة العلاقة و ليس اتجاهها.

أوجد قيمة معامل الارتباط بين مشاركة الطالبة في المحاضرة و درجتها في الاختبار للبيانات التالية :

y المشاركة	نعم	نعم	نعم	لا	لا
X درجة الاختبار	15	19	20	15	11

$n_1 = 3$

$$\overline{x_1} = \frac{15+19+20}{3} = \frac{54}{3} = 18$$

$n_2 = 2$

$$\overline{x_2} = \frac{15+11}{2} = \frac{26}{2} = 13$$

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{s_x}\sqrt{\frac{n_1 n_2}{n(n-1)}}$$

$$r_{pb} = \frac{18-13}{13}\sqrt{\frac{3\times 2}{5\times 4}} = \frac{5}{13}\sqrt{\frac{6}{20}} \approx 0.21$$

أوجد قيمة معامل الارتباط بين الإجابة على السؤال الإجباري y و الدرجة الإجمالية y لستة من الطلاب، حيث ١ تعني الإجابة على السؤال و ٠ تعني عدم الإجابة.

Y	1	1	1	0	0	0
X	14	16	19	11	7	8

3 $8 = 8.67$	3 $19 = 16.3$	$n_2 =$ $\bar{x}_2 = \frac{11+7+}{3}$	$S_x = 4.68$	$n_1 = 3$ $\bar{x}_1 = \frac{14+16+}{3}$

$$r_{pb} = \frac{\bar{x}_1 - \bar{x}_2}{s_x}\sqrt{\frac{n_1 n_2}{n(n-1)}}$$

$$r_{pb} = \frac{16.3-8.67}{4.68}\sqrt{\frac{3\times3}{6\times5}} = 1.63 \times \sqrt{\frac{9}{30}} \approx 0.893$$

- معامل الاقتران (معامل فاي) Phi

- يستخدم للعلاقة بين متغيرين اسميين كل منهما ثنائي التقسيم.
- اشارة معامل فاي ليس لها معنى فهو يقيس قوة العلاقة دون اتجاهها.

	X_1	X_2	Sum
Y_1	a	b	a+b
Y_2	c	d	C+d
Sum	a+c	b+d	a+b+c+d

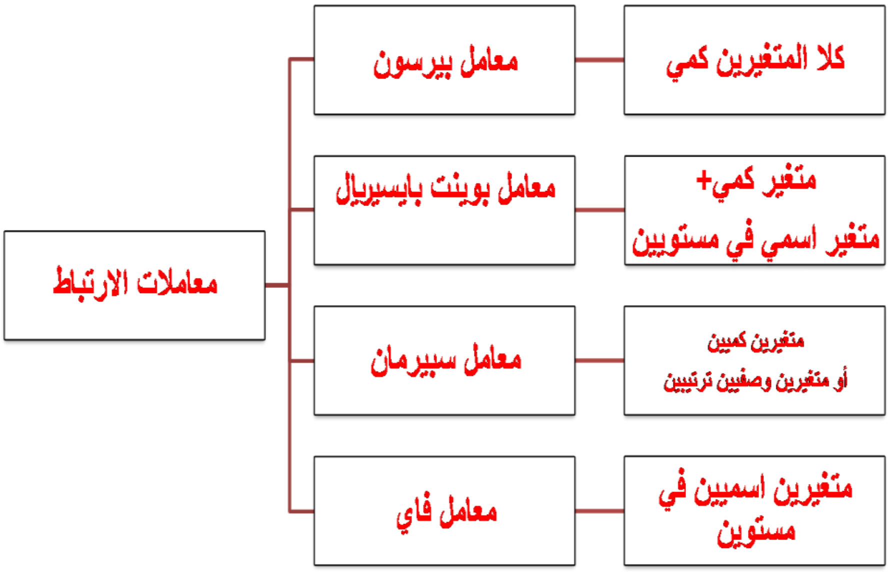
$$r_{\emptyset} = \frac{a\times d\ - b\times C}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

أوجدي قيمة معامل الاقتران بين النوع (ذكر/ أنثى) و بين الاصابة بمرض الاكتئاب (مصاب/ غير مصاب) للبيانات التالية:

	lwhf	غير مصاب	المجموع
ذكر	12	7	19
أنثى	10	5	15
المجموع	22	12	36

$$r_{\emptyset} = \frac{12 \times 5 \ - 7 \times 10}{\sqrt{22 \times 12 \times 19 \times 15}} =$$

$$= \frac{60 - 70}{\sqrt{75240}} = \frac{-10}{274.299} = -0.037$$



المحاضرة 8

اختبار «ت» t. test

يعد اختبار (ت) من أكثر اختبارات الدلالة شيوعاً في الأبحاث النفسية والتربوية. ويستخدم لقياس دلالة فروق المتوسطات غير المرتبطة والمرتبطة، للعينات المتساوية وغير المتساوية.

شروط استخدام اختبار (ت) لدلالة فوق المتوسطات:

١ - حجم كل عينة:

الأصل في اختبار (ت) أنه من مقاييس دلالة العينات الصغيرة ولكن هذا لا يحول دون استخدام (ت) للعينات الكبيرة.

- العينة الصغيرة هي التي يقل حجمها عن 30
- العينة الكبيرة هي التي يزيد حجمها عن 30
- في حالة العينات الصغيرة جداً يتم استخدام البدائل اللابارامترية للدلالة التي تصلح للتوزيعات الحرة غير المقيدة باعتدالية التوزيع.

٢ - الفرق بين حجم العينتين:

من الأفضل أن يكون حجم العينتين متقارباً فلا يكون حجم أحد العينتين 400 وحجم الآخر 50 لأن للحجم أثره على مستوى دلالة (ت).

٣ - مدى تجانس العينتين:

يقاس مدى التجانس بالفرق بين تباين العينتين ولا يقاس هذا الفرق بطرح التباين الأصغر من التباين الأكبر، وإنما يقاس بقسمة التباين الأكبر على التباين الأصغر.

$$ف = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}}$$

$$ف = \frac{ع_{١}^{٢}}{ع_{٢}^{٢}}$$

يتحقق التجانس بين العينتين عندما تصبح ف مساوية للواحد الصحيح، أي عندما يصبح التباين الكبير مساوياً للتباين الصغير.

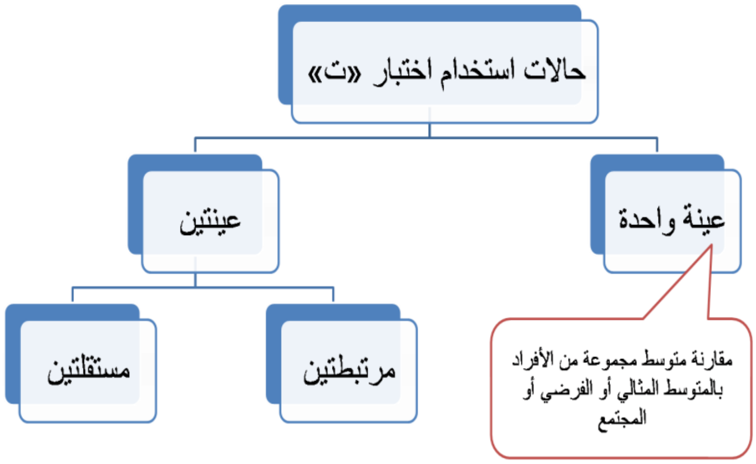
٤ - مدى اعتدالية التوزيع التكراري للعينتين:

نعني بمدى الاعتدالية تحرر التوزيع التكراري من الالتواء، والالتواء اما أن يكون سالباً أو موجباً.

التوزيع الاعتدالي لا التواء له، ويمتد من –3 إلى + 3 مقياس الالتواء التالي:

الالتواء = $\dfrac{\text{٣ (المتوسط – الوسيط)}}{\text{الانحراف المعياري}}$

كلما اقترب الالتواء من الصفر كان التوزيع اعتدالياً، لأن المتوسط في التوزيع الاعتدالي يساوي الوسيط.



- **استخدام اختبار ت للتعرف على دلالة الفرق بين متوسط عينة ما ومحك ثابت:**

ت = $\dfrac{\text{م - س}}{\text{خ}_\text{م}}$

حيث أن ت تمثل النسبة التائية، م متوسط العينة، س متوسط المجتمع أو المحك، خ م الخطأ المعياري للمتوسط.

درجات الحرية = ن – 1

مثال : طبق باحث اختبار في اللغة الانجليزية على مجموعة من المفحوصين عددهم (20) مفحوصاً، فحصل على البيانات التالية:

٦٢	٤٨	٣٠	٣٨	٣٩	٤٠	٤٦	٢٢	٤٠	٣٨
٥٠	١٩	٤١	٤٢	٧٢	١٧	٦٦	٢٤	٣٥	٤٥

المطلوب اختبار الفرض البحثي: يختلف متوسط درجات المجموعة في اللغة الانجليزية عن الدرجة39.

ت = $\dfrac{\text{م - س}}{\text{خ}_\text{م}}$

م = $\dfrac{\text{٨١٤}}{\text{٢٠}}$ **= ٤٠,٧**

الخطأ المعياري للمتوسط = $\dfrac{\text{الانحراف المعياري}}{\text{الجذر التربيعي لحجم العينة}}$

$$\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}}$$

x	$d = x - \bar{x}$	d^2
62	62 – 40.7= 21.3	453.69
48	48 – 40.7= 7.3	53.29
30	30 – 40.7= -10.7	114.49
38	38 – 40.7= -2.7	7.29
39	39 – 40.7= -1.7	2.89
40	40 – 40.7= -0.7	0.49
46	46 – 40.7= 5.3	28.09
22	22 – 40.7= -18.7	349.69
40	40 – 40.7= -0.7	0.49
38	38 – 40.7= -2.7	7.29
50	50 – 40.7= 9.3	86.49
19	19 – 40.7= -21.7	470.89
41	41 – 40.7= 0.3	0.09
42	42 – 40.7= 1.3	1.69
72	72 – 40.7= 31.3	979.69
17	17 – 40.7 = -23.7	561.69
66	66 – 40.7= 25.3	640.09
24	24 – 40.7= -16.7	278.89
35	35 – 40.7= -5.7	32.49
45	45 – 40.7= 4.3	18.49
		4088.2

$$s^2 = \frac{\sum d^2}{n} = \frac{4088.2}{20} = 204.41$$

$$s = \sqrt{s^2} = \sqrt{204.41} \cong \underline{\underline{14.30}}$$

الخطأ المعياري للمتوسط = $\frac{١٤,٣٠}{\sqrt{20}}$ = ٣,٢٠

ت = $\frac{\text{م - س}}{\text{خ م}}$ = $\frac{٣٩ - ٤٠,٧}{٣,٢٠}$ = ٠,٥٣

- يعتمد تطبيق اختبار ت لحساب دلالة الفروق بين متوسطات درجات العينات على

الأولى : تسمى القيمة المحسوبة لاختبار ت يتم حسابها من خلال معادلة خاصة.

الثانية : تسمى القيمة الجدولية لاختبار ت، ويتم حسابها من جدول يسمى جدول ت.

ويعتمد الكشف فى هذه الجداول على ما يسمى بـ "درجات الحرية" .

درجات الحرية = عدد الأفراد – عدد المجموعات

= ن – 1

يتم مقارنة قيمة ت المحسوبة بقيمة ت الجدولية فإذا كانت:

إذا كانت قيمة ت المحسوبة أكبر من قيمة ت الجدولية فذلك يعنى أن (ت) دالة إحصائيا وذلك يعنى أن الفروق بين المتوسطات فروقحقيقية وجوهرية ولها معنى وليست فروقا ظاهرية .

أما إذا كانت قيمة ت المحسوبة أقل من الجدولية فذلك يعنى أن (ت) غير دالة إحصائيا وذلك يعنى أن الفروق بين المتوسطات غير جوهرية بل فروق ظاهرية ليس لها أى تأثير .

الحالات التي يستخدم فيها اختبار (ت) لدى عينة واحدة:

يمكن استخدام اختبار (ت) لدى عينة واحدة في حالات كثيرة منها الحالات التالية:

- دراسة الفرق بين متوسط مجموعة من الأفراد في متغير ما والمتوسط المثالي لهذا المتغير.
- دراسة الفرق بين متوسط التحصيل الدراسي لطلاب فصل دراسي معين في مقرر دراسي أو مقررات دراسية معينة والمتوسط العام للتحصيل الدراسي لطلاب المدرسة أو الإدارة التعليمية أو المحافظة في نفس المقرر أو المقررات الدراسية.
- دراسة الفرق بين متوسط ذكاء مجموعة من الطلاب بمدرسة معينة ومتوسط الذكاء العام لدى طلاب المنطقة أو المحافظة التي تقع بها المدرسة.
- المقارنة بين متوسط أداء مجموعة من الأفراد في شيء ما، ومستوى معين لأداء هذا الشيء.

البيانات المطلوب توافرها لاستخدام اختبار (ت) لدى عينة واحدة:

يحتاج استخدام اختبار (ت) لدى عينة واحدة إلى توافر البيانات التالية:

- البيانات الخام (أو الدرجات الخام) لدى عينة الأفراد موضع الدراسة، أو (متوسط العينة + الخطأ المعياري لمتوسط العينة)، أو (متوسط العينة + الانحراف المعياري لدرجات العينة + عدد أفراد العينة).
- المتوسط المثالي أو الفرضي لدى المجتمع الذي سنقارن به متوسط العينة.

صياغة الفروض عند استخدام اختبار (ت) لدى عينة واحدة:

عند استخدام اختبار (ت) لدى عينة واحدة يمكن صياغة الفروض التالية:

H_0 : لا يوجد فرق دال إحصائيًا بين متوسط عينة البحث والمتوسط العام (أو المثالى أو الفرضى) لدى مجتمع البحث فى المتغير (فرض صفرى).

H1 : يوجد فرق دال إحصائيًا بين متوسط عينة البحث والمتوسط العام (أو المثالي أو الفرضي) لدى مجتمع البحث في المتغير(فرض بديل غير موجه).

يوجد فرق دال إحصائيًا بين متوسط عينة البحث والمتوسط العام (أو المثالي أو الفرضي) لدى مجتمع البحث في المتغير (س)، لصالح متوسط عينة البحث أو لصالح مجتمع البحث. (فرض بديل موجه).

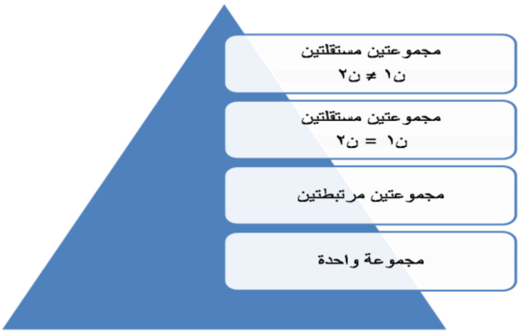
انتهت المحاضرة

المحاضرة 9

اختبار «ت» t. test

مجموعتين

حالات استخدام اختبار ت:



عينتان مرتبطتان

عبارة عن مجموعتين من الدرجات لكنهما ناتجتان عن مجموعة واحدة من الأفراد لكل فرد درجتين على الأقل مثل:

- إجراء قياس قبلي وقياس بعدي لمتغير ما لدى عينة واحدة من الأفراد.
- أو تطبيق اختبارين على مجموعة واحدة أو تطبيق اختبار واحد مرتين على العينة.

اختبار ت لمجموعتين مرتبطتين:

$$ت = \frac{م_ف}{\sqrt{\frac{مج\ ح_ف^2}{ن\,(ن-1)}}}$$

حيث:

م ف = متوسط الفروق ويحسب من العلاقة:

$$ح\ ف = ف - م\ ف$$

$$م\ ف = \frac{مج\ ف}{ن}$$

ف = س1 – س2

س1 درجات الاختبار الأول

س 2 درجات الاختبار الثاني

ن = عدد الأفراد في أي من الاختبارين

مثال1

الاحصاء الاجتماعي	٢٦	١٨	٢٠	٢٤	٢٢	١٤	٢٣	١٦	٢٢	١١
مشروع التخرج	٢٣	١٦	١٩	٢١	١٨	١٢	٢٤	١١	٢٣	٩

س١	س٢	ف	ح ف	(ح ف)٢
٢٦	٢٣	٣	١	١
١٨	١٦	٢	٠	٠
٢٠	١٩	١	-١	١
٢٤	٢١	٣	١	١
٢٢	١٨	٤	٢	٤
١٤	١٢	٢	٠	٠
٢٣	٢٤	-١	-٣	٩
١٦	١١	٥	٣	٩
٢٢	٢٣	-١	-٣	٩
١١	٩	٢	٠	٠
		٢٠		٣٤

$$ت = \frac{م\ ف}{\sqrt{\frac{مجـ\ ح\ ف^2}{ن\ (ن-1)}}}$$

$$م\ ف = \frac{مجـ\ ف}{ن}$$

ح ف = ف – م ف

$$ت = \frac{٢}{\sqrt{\frac{٣٤}{١٠\ (١٠-١)}}}$$

ت = 3,25

مثال2

الاحصاء الاجتماعي	١٠	٥	٦	٧	١٠	٦	٧	٨	٦	٥
مشروع التخرج	٧	٣	٧	٥	٨	٤	٥	٢	٣	٦

س١	س٢	ف	ح ف	(ح ف)٢
١٠	٧	٣	١	١
٥	٣	٢	٠	٠
٦	٧	-١	-٣	٩
٧	٥	٢	٠	٠
١٠	٨	٢	٠	٠
٦	٤	٢	٠	٠
٧	٥	٢	٠	٠
٨	٢	٦	٤	١٦
٦	٣	٣	١	١
٥	٦	-١	-٣	٩
		٢٠		٣٦

$$ت = \frac{م\ ف}{\sqrt{\frac{مجـ\ ح\ ف^2}{ن\ (ن-1)}}}$$

$$ت = \frac{٢}{\sqrt{\frac{٣٦}{١٠\ (١٠-١)}}}$$

ت = 3,16

عينتين غير مرتبطتين (مستقلتين): حيث ن١ = ن٢

$$ت = \frac{م_1 - م_2}{\sqrt{\frac{ع_1^2 + ع_2^2}{ن - 1}}}$$

حيث :

$م_1$ = المتوسط الحسابي للمجموعة الأولى .

$م_2$ = المتوسط الحسابي للمجموعة الثانية .

$ع_1^2$ = تباين المجموعة الأولى .

$ع_2^2$ = تباين المجموعة الثانية .

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

عينتين غير مرتبطتين (مستقلتين)

عبارة عن مجموعتين من الدرجات ناتجة عن مجموعتين مستقلتين من الأفراد مثل (المجموعة التجريبية والمجموعة الضابطة؛ أو الذكور والإناث؛ أو القسم العلمي والقسم الأدبي

ذكور	٧	٤	٥	٣	٨	٦	٢
إناث	٣	٥	١٥	٢	١٠	١٣	١

المجموعة الأولى [$n = 7$]

x	$d = x - \bar{x}$	d^2
7	2	4
4	-1	1
5	0	0
3	-2	4
8	3	9
6	1	1
2	-3	9
35		28

$$\bar{x} = \frac{\sum x}{n} = 5$$

$$s^2 = \frac{\sum d^2}{n} = \frac{28}{7} = 4$$

المجموعة الثانية [$n = 7$]

x	$d = x - \bar{x}$	d^2
3	-4	16
5	-2	4
15	8	64
2	-5	25
10	3	9
13	6	36
1	-6	36
49		

$$\bar{x} = \frac{\sum x}{n} = 7$$

$$s^2 = \frac{\sum d^2}{n} = \frac{190}{7} = 27.14$$

$$\text{ت} = \frac{\text{م}_1 - \text{م}_2}{\sqrt{\frac{\text{ع}_1^2 + \text{ع}_2^2}{\text{ن} - 1}}}$$

$$\text{ت} = \frac{5 - 7}{\sqrt{\frac{4 + 27.14}{7 - 1}}}$$

ت المحسوبة = −0.88

عينتين غير مرتبطتين (مستقلتين): حيث ن١≠ن٢

$$\text{ت} = \frac{\text{م}_1 - \text{م}_2}{\sqrt{\frac{\text{ع}_1^2}{\text{ن}_1} + \frac{\text{ع}_2^2}{\text{ن}_2}}}$$

حيث :

م1 = المتوسط الحسابي للمجموعة الأولى .

م2 = المتوسط الحسابي للمجموعة الثانية .

ع1 2 = تباين المجموعة الأولى .

ع2 2 = تباين المجموعة الثانية .

ن1 = عدد أفراد المجموعة الأولى .

ن2 = عدد أفراد المجموعة الثانية .

العينة الأولى	٣٥	١٧	٢٢	٣٢	١٩	٤٨	١٣	١٩	٢٠
العينة الثانية	١١	٣	٩	١٠	١٤	٢	٧		

$\bar{x} = \frac{\sum x}{n} = 25$

المجموعة الأولى [$n = 9$]		
x	$d = x - \bar{x}$	d^2
35	10	100
17	-8	64
22	-3	9
32	7	49
19	-6	36
48	23	529
13	-12	144
19	-6	36
20	-5	25
225		992

المجموعة الثانية [$n = 7$]		
x	$d = x - \bar{x}$	d^2
11	3	9
3	-5	25
9	1	1
10	2	4
14	6	36
2	-6	36
7	-1	1
56		112

$\bar{x} = \frac{\sum x}{n} = 8$

$s^2 = \frac{\sum d^2}{n} = \frac{112}{7} = 16$

$$ت = \frac{م_1 - م_2}{\sqrt{\frac{ع_1^2}{ن_1} + \frac{ع_2^2}{ن_2}}}$$

$$ت = \frac{٢٥ - ٨}{\sqrt{\frac{١١٠,٢}{٩} + \frac{١٦}{٧}}}$$

$$ت = ٤,٤٦$$

- **يعتمد تطبيق اختبار ت لحساب دلالة الفروق بين متوسطات درجات العينات على حساب درجتين لاختبار ت :**

الأولى : تسمى القيمة المحسوبة لاختبار ت يتم حسابها من خلال معادلة خاصة.

الثانية : تسمى القيمة الجدولية لاختبار ت، ويتم حسابها من جدول يسمى جدول ت.

ويعتمد الكشف فى هذه الجداول على ما يسمى بـ "درجات الحرية" .

درجات الحرية = عدد الأفراد – عدد المجموعات

= ن – 1

يتم مقارنة قيمة ت المحسوبة بقيمة ت الجدولية فإذا كانت:

إذا كانت قيمة ت المحسوبة أكبر من قيمة ت الجدولية فذلك يعنى أن (ت) دالة إحصائيا وذلك يعنى أن الفروق بين المتوسطات فروقحقيقية وجوهرية ولها معنى وليست فروقا ظاهرية .

أما إذا كانت قيمة ت المحسوبة أقل من الجدولية فذلك يعنى أن (ت) غير دالة إحصائيا وذلك يعنى أن الفروق بين المتوسطات غير جوهرية بل فروق ظاهرية ليس لها أى تأثير .

صياغة الفروض عند استخدام اختبار (ت) لمجموعتين:

مجموعتين مرتبطتين:

H_0 : لا توجد فروق دالة إحصائيًا بين متوسطي درجات طلاب قسم الاجتماع في مادتي الإحصاء الاجتماعي ومناهج البحث(فرض صفرى).

H1 : توجد فروق دالة إحصائيًا بين متوسطي درجات طلاب قسم الاجتماع في مادتي الإحصاء الاجتماعي ومناهج البحث(فرض بديل غير موجه).

مجموعتين مستقلتين:

H_0 : لا توجد فروق دالة إحصائيًا بين متوسطي درجات الذكور والإناث في مقرر الإحصاء الاجتماعي(فرض صفرى).

H1 : توجد فروق دالة إحصائيًا بين متوسطي درجات الذكور والإناث في مقرر الإحصاء الاجتماعي(فرض بديل غير موجه).