الفصل الثامن

العمليات الإحصائية في البحث التربوي

ويشتمل على النقاط التالية:

- مفهوم علم الإحصاء
- أقسام علم الإحصاء
- أولاً: مقاييس النزعة المركزية
- ثانياً: مقاييس التشتت (الاختلاف)
- ثالثاً: مقاييس العلاقة

الفصل الثامن

العمليات الإحصائية في البحث التربوي

مقدمة

تعد جدولة البيانات واحدة من الخطوات الأولى في وصف البيانات، فالجدول يحدد لنا وضع كل فرد في العينة بالنسبة للأفراد الآخرين، ويحدد سمات مجموعات الأفراد، إلا أنه لكي نعقد مقارنـات بـين العينات المختلفة من الأفراد فيما يتعلق بسمة معينة، ولكي نحدد كيفية ارتبـاط الـسمات المختلفـة لـنفس الأفراد ببعضها البعض، فإن الأمر يحتاج منا إلى أن نأخذ عملية وصف البيانات إلى مدى أبعد، وهذا المدى لا نصل إليه إلا إذا كنا على دراية بمبادئ الإحصاء الوصفي.

مفهوم علم الإحصاء **Concept Statistic**

إذا عرّفنا "الإحصاء" بأنها القيمة أو الدرجة التي تعبر عـن النتيجـة النهائيـة للعمليـات الرياضية التي تمثل العينة أو المجتمـع الأصـلي فلابـد أن نـشير إلى وجـود ثلاثـة تطـورات في تاريخ الإحصاء تستحق الذكر، الأول نظرية أخطاء القيـاس لجـالتون Galton F وآخـرين عـن تطبيـق المفـاهيم الإحـصائية في العلـوم البيولوجيـة، والثـاني مـا قدمـه فيـشر Fisher مـن

صياغات وابتكارات نظرية وأخيراً الكمبيوتر الذي أدى إلى تسهيل إجراء العمليات المعقدة.

والأصل في كلمة الإحصاء أنها مشتقة من اللفظ اللاتيني "ستاتوس" أو "ستاتو" والذي يستعمل بمعنى الدولة كما يستعمل أيضاً ليشير للمعلومات المتصلة بنظام الدولة ومؤسساتهما وأجهزتها المختلفة وأحوالها، ولذلك أطلق على الإحصاء اسم "ستاتستيك" Statistic ليدل على مجموعة المعلومات الخاصة بالدولة في وقت من الأوقات ثم انتهى به الأمر ليدل حتى الآن على معاني عدة منها:

1- جمع المعلومات التي تبين الأحوال والظروف في البلاد مثل:

- عدد المواليد والوفيات.
- عدد الأذكياء وعدد الأغبياء كما تكشف عنهم اختبارات الذكاء.
- المحاصيل الزراعية والفواكه.
- عدد المتفوقين وعدد المتأخرين دراسياً.
- التجارة الداخلية والخارجية.
- عدد المرضى النفسيين وعدد الأسوياء في مجتمع ما.
- عدد المتعلمين وغير المتعلمين (الأميين).
- عدد المقبولين بناءً على الاختيار المهني.

2- ويعني بالإحصاء إلى جانب ما سبق أنه فرع من فروع العلم له أسلوبه وطريقته وموضوعات البحث الخاصة به.

وعلم الإحصاء ليس مجرد علم يهتم فقط بالبيانات العددية المبوبة وغير المبوبة، وإنما يتضمن النظريات والطرائق الرياضية التي تفيد في جمع وتحليل وتفسير وتمثيل البيان، وتساعد الإحصاء القائم بعملية القياس والتقويم على إعطاء أوصاف على جانب كبير من الدقة العملية، وكلما توصل العلم إلى زيادة في دقة الوصف كلما كان هذا دليلاً على التقدم العلمي.

ويساعد علم الإحصاء القائم بعملية القياس والتقويم على تلخيص النتائج في شكل ملائم مفهوم, فمجرد ذكر الدرجات لا يكفي للمقارنة بين الجنسين, بل أن حساب متوسطي الدرجات قد سهل مهمة المقارنة كثيراً، فالبيانات التي يجمعها الباحث لا تعطي صورة واضحة إلا إذا تم تلخيصها في معامل أو رقم أو شكل توضيحي كالرسوم البيانية.

وقد كان للإحصاء الدور الأكبر في تقدم العلوم التربوية, فهي تستخدم الإحصاء لتفسير نتائج الأبحاث والدراسات بعد تحليل هذه البيانات بالطرائق الإحصائية المناسبة.

فالإحصاء هو طريقة منظمة تسير في خطوات متسلسلة بدءاً من جمع البيانات عن الظاهرة ثم وصف هذه الظاهرة أو تحليل البيانات المتجمعة وفق قواعد وقوانين إحصائية خاصة واتخاذ القرارات المناسبة وغالباً ما تكون هذه القرارات على شكل تعميمات أو تقديرات وذلك من أجل التنبؤ أو لرفض الفرضيات الإحصائية أو عدم رفضها.

ويحصل الباحث الذي يهتم بدراسة ظاهرة ما في أغلب الأحيان على مجموعة من القيم العددية المتعلقة بهذه الظاهرة (البيانات الإحصائية) وتنقسم هذه البيانات إلى نوعين: كمية وكيفية أو نوعية.

1- **البيانات الكمية:** وهي البيانات التي يكون التغير فيها تغيراً من حيث المقدار أي يمكن ترتيب هذه البيانات بحسب مقاديرها وقد يكون المتغير في هذه البيانات متصلاً أو غير متصل.

2- **البيانات النوعية:** وهي البيانات التي يكون التغير فيها تغيراً من حيث النوع ولا يمكن تقسيمها بحسب الأصغر والأكبر تحت تقسيم واحد. ومن أمثلتها عدد الأفراد الذين ينتمون إلى مهنة معينة أو جنس معين أو صف دراسي معين أو طبقاً للون البشرة، وتختلف الطرائق الإحصائية التي تعالج هذين النوعين من البيانات, إلا أنهما يتلقيان عند أكثر من نقطة.

أقسام علم الإحصاء:

يُقسّم الإحصائيون علم الإحصاء إلى قسمين:

1- **الإحصاء الاستدلالي Inferential Statistic:** يركز هذا النوع من الإحصاء على الوصول إلى استنتاجات حول خصائص المجتمع من خلال استخدام المعلومات المتوفرة عن العينة المسحوبة من هذا المجتمع، أي أنه يهدف إلى التعميم من العينة إلى المجتمع، ولهذا فإن الإحصاء الاستدلالي يركز على اختبار الفرضيات المتعلقة بالفروق بين المتوسطات أو النسب المئوية المتعلقة بعينة واحدة أو عينتين أو أكثر أو الفروق بين معاملات الارتباط.

2- **الإحصاء الوصفي Descriptive Statistic:** وهو يهدف إلى وصف مجموعة من البيانات عندما تتوفر, ويركز هذا النوع من الإحصاء على وصف الظاهرة وربما تصنيفها وذلك من خلال استخدام الرسوم والأشكال البيانية والتوزيعات التكرارية أو من خلال استخدام مقاييس النزعة المركزية والتشتت أو من خلال استخدام معاملات الارتباط لدراسة العلاقة بين المتغيرات.

وسنعرض في الصفحات التالية بعض القياسات التي يحتاج إليها الباحث في وصف بياناته، وهي **"مقاييس النزعة المركزية، ومقاييس التشتت، ومقاييس العلاقة"**.

أولاً: مقاييس النزعة المركزية **Central Tendency Measurements**

مقدمة:

تعد مقاييس النزعة المركزية (أو المتوسطات) من أهم المقاييس الإحصائية التي يفكر الباحث في حسابها، بل هي أول مقاييس إحصائية يفكر فيها الباحث السياسي عموماً؛ فمقياس النزعة المركزية لظاهرة ما تعني التعرف على القيمة التي تقع عادة عند مركز التوزيع العددي للقيم المبحوثة.

إن متوسط أي ظاهرة يعبر عن المستوى العام لهذه الظاهرة، فمتوسط مجموعة من القيم هو القيمة التي تعبر عن جميع القيم، أو هو القيمة التي تدور (أو تتركز) حولها باقي

القيم، فمتوسط الدخل لأي بلد يعبر عن المستوى العام للدخل في هذا البلد، كما أن متوسط دخل عمال أحد المصانع هو الدخل الذي تتركز حوله دخول العمال بهذا المصنع، ولذا تسمى المتوسطات بمقاييس النزعة المركزية وذلك لأن قيم أي ظاهرة – عادة – تميل أو تنزع للتركز حول قيمة معينة هي متوسط هذه الظاهرة أو مقياس نزعتها المركزية. فأطوال البالغين تتركز حول رقم معين هو متوسط الطول، وكذلك أوزانهم ومعدلات ذكائهم، وأي ظاهرة أخرى، فالمتوسط - بصفة عامة - هو الذي يعبر عن المستوى العام للظاهرة أي هو الذي يعبر عن جميع قيمها، بمعنى أنه القيمة التي تتركز حولها باقي القيم.

فالنزعة المركزية تشير إلى موقع التوزيع، فهي عبارة عن القيم التي يتمركز حولها القيم لتوزيع ما؛ وذكر المغربي (2007) أن الغرض من دراسة مقاييس النزعة المركزية هي:

- تقوم بوصف مجموعة ما برقم واحد يمثلها ويلخصها ويعبر عنها.
- تفيد في المقارنات بين مجموعات عدة أو مجتمعات.
- تفيد في المقارنات التاريخية بما يمكن من وصف التغير أو التطور في الظاهرة عبر الزمن.

وتشمل مقاييس النزعة المركزية ثلاثة مقاييس هي: الوسط الحسابي والوسيط والمنوال.

وفيما يلي عرض لهذه المتوسطات نبين فيه تعريف كل مقياس وكيفية حسابه ومزاياه وعيوبه.

1-الوسط الحسابي **Arithmetic Mean**:

يعتبر الوسط الحسابي $\bar{x}$ أكثر المتوسطات شهرة وأكثرها استخداماً، بل لعله من أهم المقاييس الإحصائية على الإطلاق، وذلك لما يتمتع به من مزايا وخواص، ولدخوله في حساب الكثير من المقاييس الإحصائية الأخرى كما سيتضح فيما بعد؛ ويمكن حساب الوسط الحسابي للبيانات المبوبة وغير المبوبة، كما يأتي:

أ- الوسط الحسابي للبيانات غير المبوبة:

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوماً على عددها، فإذا كان لدينا n من القيم ، ويرمز لها بالرمز : $x_1, x_2, \ldots, x_n$.

فإن الوسط الحسابي لهذه القيم، ونرمز له بالرمز $\overline{x}$ يحسب بالمعادلة الآتية:

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

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

حيث يدل الرمز Σ على المجموع .

مثال 1:

فيما يأتي درجات 8 طلاب في مادة مناهج البحث التربوي:

34 32 42 37 35 40 36 40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحل: لإيجاد الوسط الحسابي للدرجات نطبق المعادلة الآتية:

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

$$\overline{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$= \frac{34 + 32 + 42 + 37 + 35 + 40 + 36 + 40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مادة مناهج البحث التربوي يساوي 37 درجة.

مثال2:

أوجد المتوسط الحسابي للمشاهدات الآتية والتي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص:5025, 55, 35, 45, 40, 30,

الحل:

$x_1=25,\ x_2=30,\ x_3=40,\ x_4=45,\ x_5=35,\ x_6=55,\ x_7=50$
$n = 7$

المتوسط هو :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_7}{7}$$

$$= \frac{25+30+40+45+35+55+50}{7} = \frac{280}{7} = 40 \quad \text{(كيلوجرامًا)}$$

ب- الوسط الحسابي للبيانات المبوبة:

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هـذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخـل حـدود الفئـة بمركز هذه الفئة ، ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمـة التقديريـة لكـل مفـردة تقـع في هذه الفئة.

فـإذا كانــت k هــي عــدد الفئــات، وكانــت $x_1,\ x_2,\ ......,x_k$ هــي مراكــز هــذه الفئــات، $f_1, f_2, ...,f_k$ هي التكرارات، فإن الوسط الحسابي يحسب بالمعادلة الآتية:

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

مثال: الجدول الآتي يعرض توزيع 40 طالباً حسب أوزانهم:

فئات الوزن	323-4	-3436	363-8	384-0	404-2	424-4
عدد الطلاب	4	7	13	10	5	1

أوجد الوسط الحسابي.

الحــل:

لحساب الوسط الحسابي باستخدام المعادلة يتم إتباع الخطوات الآتية:

1- إيجاد مجموع التكرارات $\sum f$.

2- حساب مراكز الفئات x .

3- ضرب مركز الفئة في التكرار المناظر له (xf) ، وحساب المجموع $\sum xf$

4- حساب الوسط الحسابي بتطبيق المعادلة .

فئات الوزن (C)	التكرارات f	مراكز الفئات x	$x\,f$
323-4	4	2=33 ÷ (32+34)	33=132 × 4
-3436	7	35	35=245 × 7
363-8	13	37	37=481 × 13
384-0	10	39	39=390 × 10
404-2	5	14	41=205 × 5
424-4	1	34	43=43 × 1
المجموع	40		1496

إذا الوسط الحسابي لوزن الطالب هو:

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{1496}{40} = 37.4 \ \ k.g$$

أي أن متوسط وزن الطالب يساوي (k.g 37.4).

مثال: الجدول التالي يمثل توزيع مجموعة من الطلاب حسب فئات الدرجات كما يأتي:

فئات الدرجات Classes	أعداد الطلاب f
2 – 4	3
4 – 6	9
6 – 8	10
8 – 10	5
المجموع	$\sum f = 27$

جد الوسط الحسابي لدرجات الطلاب.

الحل :

الجـدول يقـول أن 3 طـلاب حـصل كـل مـنهم عـلى درجـة تـتراوح بـين (2 أقـل مـن 4) (لكن لا نعلم ما درجة كل منهم بالتحديد)، 9 طلاب حصل كل منهم على درجة تتراوح بين (4 وأقل من 6) (لكن لا نعلم ما درجة كل منهم بالتحديد)، وهكـذا بالنـسبة لبـاقي الفئـات، وفي هـذه الحالـة نحسب مراكز الفئات كأحسن قيم تمثل هذه الفئات، ومركز الفئة هـو القيمـة التـي تقـع في منتـصف الفئة، أي أن :

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى للفئة + الحد الأعلى لها}}{٢}$$

أي أننا نستعيض عن الفئات بمراكزها وهي التي تمثل القيم، وسوف نرمز لها بالرمز x، ثم نكمل الحل كما في الآتي:

فئات الدرجات Classes	أعداد الطلاب f	مراكز الفئات x	حاصل الضرب x. f (مجموع الدرجات)
2 – 4	3	$\frac{2+4}{2}=3$	$3 \times 3 = 9$
4 – 6	9	$\frac{4+6}{2}=5$	$5 \times 9 = 45$
6 – 8	10	$\frac{6+8}{2}=7$	$7 \times 10 = 70$
8 – 10	5	$\frac{8+10}{2}=9$	$9 \times 5 = 45$
المجموع	$\sum f = 27$		$\sum xf = 169$

وبالتعويض في قانون الوسط الحسابي نحصل على :

$$\overline{X} = \frac{\sum xf}{\sum f}$$
$$= \frac{169}{27} = 6.26$$

أي أن الوسط الحسابي يساوي 6.26 درجة.

ملاحظة مهمة :

عند حساب الوسط الحسابي في حالة الفئات نحسب أولاً مراكز الفئات كأحسن قيم تمثل الفئات – كما ذكرنا – ولذلك يقال أن قيمة الوسط الحسابي في حالة الفئات قيمة تقريبية (وليست دقيقة exact) وذلك لأننا نفترض – على سبيل التقريب – أن مركز الفئة هو أحسن قيمة تمثل الفئة لأنه ليست لدينا الدرجات الدقيقة التفصيلية لكل طالب، ومن ذلك نستنتج أنه إذا كانت هناك فئة مفتوحة (بمعنى عدم معرفة أحد حديها) فإنه لا يمكن حساب مركز هذه الفئة، ومن ثم لا يمكن حساب الوسط الحسابي في هذه الحالة.

الوسط الحسابي المرجح أو الموزون **Weighted Arithmetic Mean:**

يواجه الباحث في بعض الأحيان مجموعة من المفردات تتفاوت في درجة أهميتها، لذلك فإنه يحتم على الباحث أن يعامل هذه المفردات بمعاملات ترجيح مختلفة توضح أهمية كل مفردة عن الأخرى، وهذه الطريقة تسمى الوسط الحسابي المرجح ونحصل عليه من مجموع حاصل ضرب كل مفردة في أهميتها النسبية (أوزانها) مقسوماً على مجموع الأوزان المختلفة.

فمثلاً لو أخذنا خمسة طلاب، وسجلنا درجات هؤلاء الطلاب في مادة مناهج البحث التربوي، وعدد الساعات في الأسبوع .

ت	1	2	3	4	5	sum
x (الدرجة)	23	40	36	28	46	173
w (عدد الساعات)	1	3	3	2	4	

نجد أن الوسط الحسابي غير المرجح للدرجة الحاصل عليها الطالب هي:

$$\overline{x} = \frac{\sum x}{n} = \frac{23+40+36+28+46}{5} = \frac{173}{5} = 34.6$$

وإذا أردنا أن نحسب الوسط الحسابي للدرجات x المرجحة بعدد الساعات w، يتم تطبيق المعادلة الآتية:

$$(\overline{w})=\frac{\sum xw}{\sum w}=\frac{23\times1+40\times3+36\times3+28\times2+46\times4}{1+3+3+2+4}$$

$$=\frac{23+120+108+56+184}{13}=\frac{491}{13}=37.769$$

وهذا الوسط المرجح أكثر دقة من الوسط الحسابي غير المرجح .

إذا الوسط الحسابي المرجح $(\overline{w})$ يحسب بتطبيق المعادلة التالية :

$$\boxed{(\overline{w})=\frac{\sum xw}{\sum w}}$$

خصائص الوسط الحسابي:

يتصف الوسط الحسابي بعدد من الخصائص، ومن هذه الخصائص ما يأتي:

1- الوسط الحسابي للمقدار الثابت يساوى الثابت نفسه، أي أنه إذا كانت قيم x هي: $x\ :\ a\ ,\ a\ ,...,\ \ a$ ، فإن الوسط الحسابي هو:

$$\boxed{\overline{x}=\frac{a+a+...+a}{n}=\frac{na}{n}=a}$$

ومثال على ذلك، لو اخترنا مجموعة من 5 طلاب، ووجدنا أن كل طالب وزنه 63)) كيلوجرام، فإن متوسط وزن الطالب في هذه المجموعة هو :

$$\overline{x}=\frac{63+63+63+63+63}{5}=\frac{315}{5}=63\quad k.g$$

2- مجموع انحرافات القيم عن وسطها الحسابي يساوى صفراً، ويعبر عن هذه الخاصية بالمعادلة الآتية:

$$\boxed{\sum(x-\overline{x})=0}$$

ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال سابق تم ذكره، وهو فيما يأتي درجات 8 طلاب في مادة مناهج البحث التربوي: 34, 32, 42, 37, 35, 40, 36, 40 ، والمتوسط الحسابي للدرجة هو $\overline{x}=37$ ، إذا:

x	34	32	42	37	35	40	36	40	296
$(x-\bar{x})$	343-7	323-7	423-7	373-7	353-7	403-7	363-7	43-7	0
$(x-37)$	-3	-5	5	0	-2	3	-1	3	

أي أن : $\sum(x-37)=0$

3- إذا أضيف مقدار ثابت إلى كل قيمة من القيم، فإن الوسط الحسابي للقيم المعدلة (بعد الإضافة) يساوى الوسط الحسابي للقيم الأصلية (قبل الإضافة) مضافاً إليها هذا المقدار الثابت؛ فإذا كانت القيم هي : $x_1, x_2, \dots, x_n$ ، وتم إضافة مقدار ثابت (a) إلى كل قيمة من القيم ، ونرمز للقيم الجديدة بالرمز y ، أي أن $y = x + a$ ، فإن : الوسط الحسابي لقيم y (القيم بعد الإضافة) هو:

$$\boxed{\bar{y} = \bar{x} + a}$$

حيث أن $\bar{y}$ هو الوسط الحسابي للقيم الجديدة، ويمكن التحقق من هذه الخاصية باستخدام بيانات مثال سابق تم ذكره، وهو فيما يأتي درجات 8 طلاب في مادة مناهج البحث التربوي: 34, 32, 42, 37, 35, 40, 36, 40 ، والمتوسط الحسابي للدرجة هو $\bar{x}=37$، وإذا قرر المصحح إضافة5 درجات لكل طالب، فإن الوسط الحسابي للدرجات المعدلة يصبح قيمته {(5+37)=42} ، والجدول الآتي يوضح ذلك:

x	34	32	42	37	35	40	36	40	296
$y=(x+5)$	34+5	32+5	42+5	37+5	35+5	40+5	36+5	40+5	336
	39	37	47	42	40	45	41	45	

نجد أن مجموع القيم الجديدة هو: $\sum y = 336$ ، ومن ثم يكون الوسط الحسابي للقيم الجديدة هو :

$$\bar{y} = \frac{\sum y}{n} = \frac{336}{8} = 42 \quad \rightarrow \quad (\bar{x} + 5 = 37 + 5 = 42)$$

4- إذا ضرب مقدار ثابت (a) في كل قيمة من القيم، فإن الوسط الحسابي للقيم المعدلة (القيم الناتجة بعد الضرب) يساوى الوسط الحسابي للقيم الأصلية (القيم بعد التعديل) مضروباً في هذا المقدار الثابت، أي أنه إذا كان: $y=a\,x$، ويكون الوسط الحسابي للقيم الجديدة y هو:

$$\boxed{\bar{y} = a\ \bar{x}}$$

ويمكن للطالب أن يتحقق من هذه الخاصية باستخدام نفس بيانات المثال السابق، فإذا كان تصحيح الدرجة من 50، وقرر المصحح أن يجعل التصحيح من 100 درجة ، بمعنى أنه سوف يضرب كل درجة في قيمة ثابتة (a=2)، ويصبح الوسط الحسابي الجديد هو: $\bar{y} = a\bar{x} = 2(37) = 74$

5- مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن، أي أن:

$$\boxed{\sum (x-\bar{x})^2 \langle \sum (x-a)^2 \quad if \quad a \neq \bar{x}}$$

وفي المثال السابق فإن : $\sum(x-37)^2 < \sum(x-a)^2$ لجميع قيم $a \neq 37$

مميزات الوسط الحسابي وعيوبه:

يتميز الوسط الحسابي بالمميزات الآتية:

1- أنه سهل الحساب .

2- يأخذ في الاعتبار كل القيم .

3- أنه أكثر المقاييس استخداما وفهما .

ومن عيوب الوسط الحسابي:

1- أنه يتأثر بالقيم الشاذة والمتطرفة .

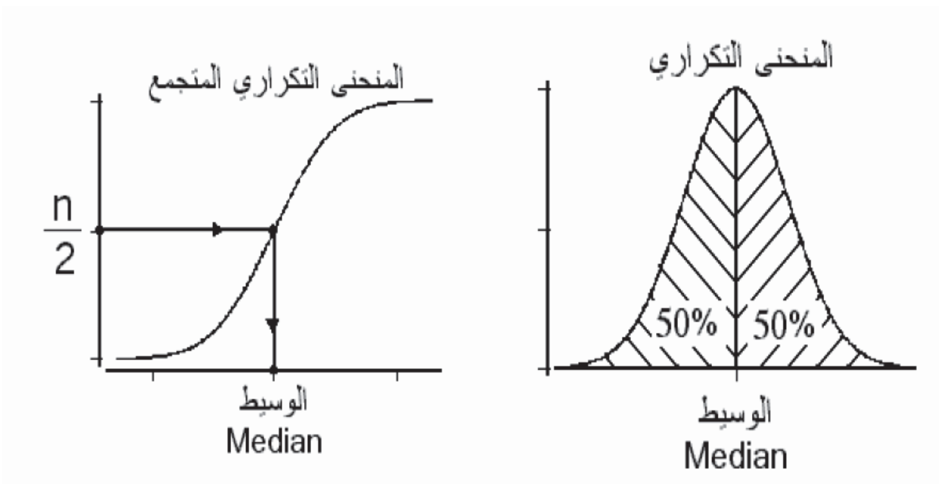
2- يصعب حسابه في حالة البيانات الوصفية .

3- يصعب حسابه في حالة الجداول التكرارية المفتوحة .

2- الوسيط **The Median**:

الوسيط هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم، وهو القيمة التي تقع في منتصف القيم بعد ترتيبها (تصاعدياً أو تنازلياً)، فالوسيط هو القيمة التي تتوسط القيم بعد ترتيبها، فإذا كان عدد القيم فردياً فإنه توجد قيمة واحدة في المنتصف

(بعد الترتيب) تكون هي الوسيط. أما إذا كان عدد القيم زوجياً فإنه توجد قيمتان في المنتصف نجمعهما ونقسم على 2 فنحصل على قيمة الوسيط، وبديهي أننا سنحصل على النتيجة نفسها لو كان الترتيب تصاعدياً أو تنازلياً، أي أن (50%) من البيانات تساوي أو تقل عن الوسيط، و(50%) من البيانات تساوي أو تزيد عن الوسيط، ويرمز للوسيط بالرمز (Med).



أ: الوسيط للبيانات غير المبوبة:

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

- ترتب القيم تصاعدياً .
- تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$
- إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\boxed{\text{الوسيط = القيمة رقم } \left(\frac{n+1}{2}\right)}$$

- إذا كـان عـدد القـيم (n) زوجـي، فـإن الوسـيط يقـع بـين القيمـة رقـم $(n/2)$، والقيمـة رقـم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة الآتية:

$$\text{الوسيط} = \frac{\text{القيمة رقم}\left(\frac{n}{2}\right) + \text{القيمة رقم}\left(\frac{n}{2}+1\right)}{2}$$

مثال: البيانات التالية تمثل أعمار مجموعة من الناخبين :

32 24 20 35 29 فما هو وسيط العمر ؟

الحل :

أولاً: نرتب هذه الأعمار تصاعدياً كما يلي :

35 32 **29** 24 20

ثانياً: نلاحظ أن عدد القيم فردي (يساوي 5) وأنه توجد قيمة واحدة في المنتصف هـي 29 ومـن ثـم فإن قيمة الوسيط تساوي 29 سنة.

مثال: البيانات التالية تمثل دخول بعض الأفراد اليومية بالدولار الأمريكي في إحدى الدول.

11 19 14 18 12 15 أحسب وسيط هذه الدخول ؟

الحل :

أولاً: نرتب هذه الدخول تصاعدياً كما يأتي:

19 18 **15** **14** 12 11

ثانياً: نلاحظ أن عدد القيم زوجي (يـساوي 6) وأنـه توجـد قيمتـان في المنتـصف هـما 15، 14 لـذلك نجمعهما ونقسم على 2، أي أن الوسيط يساوي :

$$\frac{14+15}{2}=14.5 \quad \text{دولاراً}$$

مثال: تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة، وتم زراعتها بمحصول القمح، وتم استخدام نوعين من التسميد هما: النوع (**a**) وجرب على 7 وحدات تجريبية، والنـوع (**b**) وجـرب عـلى 10 وحدات تجريبية، وبعد انتهاء الموسم الزراعي، تم تسجيل إنتاجية الوحدة بالطن / هكتـار، وكانـت على النحو الآتي:

النوع (a)	**1.2**	**2.75**	**3.25**	**2**	**3**	**2.3**	**1.5**			
النوع (b)	**4.5**	**1.8**	**3.5**	**3.75**	**2**	**2.5**	**1.5**	**4**	**2.5**	**3**

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحـل:

أولاً: حساب وسيط الإنتاج للنوع الأول(**a**):

- ترتيب القيم تصاعدياً:

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

- عدد القيم فردي $(n = 7)$
- إذا رتبة الوسيط هي: $\left((n+1)/2 = (7+1)/2 = 4\right)$.
- ويكون الوسيط هو القيمة رقم 4 ، أي أن وسيط الإنتاج للنوع **a** هو:

$Med_a = 2.3$ طن / هكتار

ثانياً: حساب وسيط الإنتاج للنوع الثاني (**b**):

- ترتيب القيم تصاعدياً.

قيمة الوسيط = $\frac{2.5 + 3}{2}$

2.75

الإنتاج	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

- عدد القيم زوجي $(n = 10)$.
- رتبة الوسيط هي: $((n+1)/2 = (10+1)/2 = 5.5)$.
- الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم 5، 6) .

$$Med_b = \frac{2.5 + 3}{2} = 2.75$$ طن / هكتار

وبمقارنة النوعين من السماد، نجد أن وسيط إنتاجية النوع (**a**) أقل من وسيط إنتاجية النوع (**b**)، أي أن: $Med_b > Med_a$.

ب- الوسيط للبيانات المبوبة:

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري، يتم إتباع الخطوات الآتية:

- تكوين الجدول التكراري المتجمع الصاعد .
- تحديد رتبة الوسيط: $\left(\frac{\sum f}{2}\right) = \left(\frac{n}{2}\right)$
- تحديد فئة الوسيط كما في الشكل الآتي:

تكرار متجمع صاعد سابق f_1	**الحد الأدنى لفئة الوسيط** (A)
رتبة الوسيط $(n/2)$ ←	**الوسيط** Med
تكرار متجمع صاعد لاحق f_2	**الحد الأعلى لفئة الوسيط**

- يحسب الوسيط، بتطبيق المعادلة:

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

حيث أن :

L هي طول فئة الوسيط، وتحسب بالمعادلة الآتية:

طول الفئة= الحد الأعلى– الحد الأدنى

$$L = Upper - Lower$$

مثال:

فيما يلي توزيع 50 عجل متوسط الحجم، حسب احتياجاته اليومية من الغذاء الجاف بالكيلوجرام:

فئات الاحتياجات اليومية	1.5 -	4.5 -	7.5 -	10.5 -	13.5 – 16.5
عدد العجول f	4	12	19	10	5

والمطلوب حساب الوسيط: 1-حسابياً 2-بيانياً

الحل:

أولاً : حساب الوسيط حسابياً:

- رتبة الوسيط : $\frac{n}{2} = \frac{\sum f}{2} = \frac{50}{2} = 25$

- الجدول التكراري المتجمع الصاعد:

	تكرار متجمع صاعد	أقل من
	0	1.5
	4	4.5
	f_1 16	A 7.5
رتبة الوسيط	25	Med (الوسيط)
	f_2 35	10.5
	45	13.5
	50	16.5

- **تحديد فئة الوسيط:** وهي الفئة التي تشمل قيمة الوسيط، وهي قيمة أقل منها $(n/2)$ من القيم، ويمكن معرفتها بتحديد التكرارين المتجمعين الصاعدين الذين يقع بينهما رتبة الوسيط $(n/2)$، وفي الجدول أعلاه نجد أن رتبة الوسيط (25) تقع بين

التكرارين المتجمعين(35,16)، ويكون الحد الأدنى لفئة الوسيط هو المناظر للتكرار المتجمع الـصاعد السابق 7.5، والحد الأعلى لفئة الوسيط هو المناظر للتكرار المتجمع الصاعد اللاحق 10.5؛ أي أن فئة الوسيط هي : (7.5-10.5) .

- بتطبيق معادلة الوسيط على هذا المثال نجد أن :

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

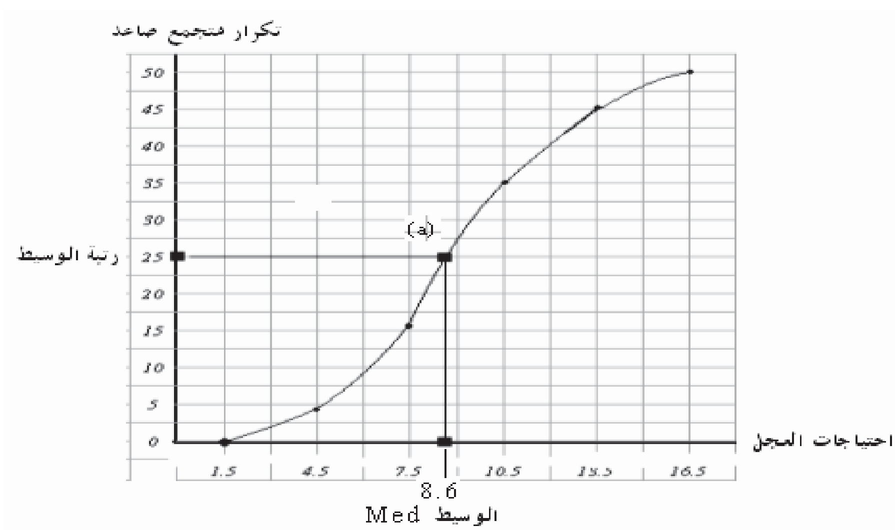
$$A = 7.5 \ , \ f_1 = 16 \ , \ f_2 = 35 \ , \ L = 10.5 - 7.5 = 3$$

إذا الوسيط قيمته هي :

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 7.5 + \frac{25 - 16}{35 - 16} \times 3$$
$$= 7.5 + \frac{9}{19} \times 3 = 7.5 + \frac{27}{19} = 7.5 + 1.421 = 8.921 \ k.g$$

ثانياً: حساب الوسيط بيانياً:

- تمثيل جدول التوزيع التكراري المتجمع الصاعد بيانياً.



- تحديد رتبة الوسيط (25) على المنحنى التكراري المتجمع الصاعد، ثم رسم خط مستقيم أفقي حتى يلقى المنحنى في النقطة (a) .
- إسقاط عمود رأسي من النقطة (a) على المحور الأفقي .
- نقطة تقاطع الخط الرأسي مع المحور الأفقي تعطى قيمة الوسيط .
- الوسيط كما هو مبين في الشكل Med = 8.6 .

مميزات الوسيط وعيوبه:

من مميزات الوسيط:

1- لا يتأثر بالقيم الشاذة أو المتطرفة .

2- سهل في الحساب .

3- مجموع قيم الانحرافات المطلقة عن الوسيط أقل من مجموع الانحرافات المطلقة عن أي قيم أخرى؛ أي أن : $a \neq Med$ ، $\sum |x - Med| \leq \sum |x - a|$.

ومن عيوب الوسيط:

1- أنه لا يأخذ عند حسابه كل القيم في الاعتبار، فهو يعتمد على قيمة أو قيمتين فقط .

2- يصعب حسابه في حالة البيانات الوصفية المقاسة بمعيار أسمي nominal.

3- المنوال:**Mode**

يعرف المنوال بأنه القيمة الأكثر شيوعاً أو تكراراً، ويكثر استخدامه في حالة البيانات الوصفية، لمعرفة النمط (المستوى) الشائع، ويمكن حسابه للبيانات المبوبة وغير المبوبة كما يأتي:

1- حساب المنوال في حالة البيانات غير المبوبة:

المنوال (Mod) = القيمة (المستوى) الأكثر تكرارا

2- حساب المنوال في حالة البيانات المبوبة (طريقة الفروق):

$$\boxed{Mod = A + \frac{d_1}{d_1 + d_2} \times L}$$

حيث أن :

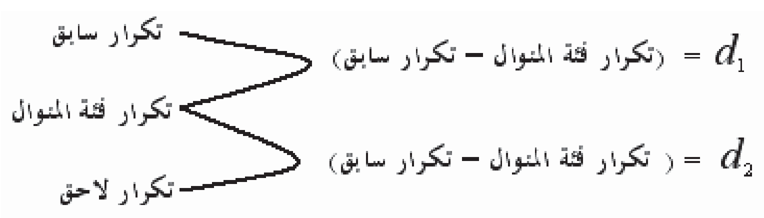
A **:** الحد الأدنى لفئة المنوال (الفئة المناظرة لأكبر تكرار) .

d_1 **:** الفرق الأول = (تكرار فئة المنوال – تكرار سابق)**.**

d_2 : الفرق الثاني = (تكرار فئة المنوال – تكرار لاحق).

L **:** طول فئة المنوال .

فئة المنوال = الفئة المناظرة لأكبر تكرار



مثـال: اختيرت عينات عشوائية من طلاب بعض أقسام كلية التربية، وتم رصد درجات هؤلاء الطلبـة في مادة القياس والتقويم، وكانت النتائج كالآتي:

قسم علوم الحياة	67	58	70	65	77	77	77	75	77	80
قسم الكيمياء	90	95	85	77	65	93	75	60	68	88
قسم الرياضيات	80	86	65	76	88	65	80	69	65	80
قسم الفيزياء	85	72	73	69	69	73	85	69	73	85

والمطلوب حساب منوال الدرجات لكل قسم من الأقسام:

الحـل:

هذه البيانات غير مبوبة، لذا فإن: المنوال = القيمة الأكثر تكراراً

والجدول الآتي يبين منوال الدرجة لكل قسم من الأقسام:

القيمة المنوالية	القيمة الأكثر تكرار	القسم
المنوال = 77 درجة	الدرجة 77 تكررت 4 مرات	قسم علوم الحياة
لا يوجد منوال	جميع القيم ليس لها تكرار	قسم الكيمياء
يوجد منوالان هما: المنوال الأول = 65 المنوال الثاني = 80	الدرجة 65 تكررت 3 مرات الدرجة 80 تكررت 3 مرات	قسم الرياضيات
يوجد ثلاث منوال هي: المنوال الأول = 69 المنوال الثاني = 73 المنوال الثالث = 85	الدرجة 69 تكررت 3 مرات الدرجة 73 تكررت 3 مرات الدرجة 85 تكررت 3 مرات	قسم الفيزياء

مثال: فيما يلي توزيع 30 أسرة حسب الإنفاق الاستهلاكي الشهري لها بالألف دينار .

فئات الإنفاق	2 -	5 -	8 -	11 -	14 - 17
عدد الأسر f	4	7	10	5	4

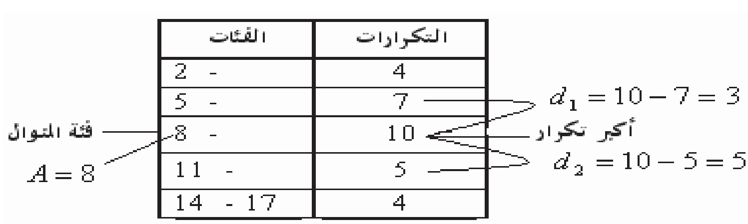
والمطلوب حساب منوال الإنفاق الشهري للأسرة، باستخدام طريقة الفروق .

الحل:

لحساب المنوال لهذه البيانات يتم استخدام معادلة المنوال (mod)= القيمة (المستوى) الأكثر تكراراً ، ويتم إتباع الآتي :

- **تحديد** الفئة **المنوالية:**

الفئة المنوالية: هي الفئة المناظرة لأكبر تكرار: (8-11)



- **حساب الفروق d، حيث أن :**

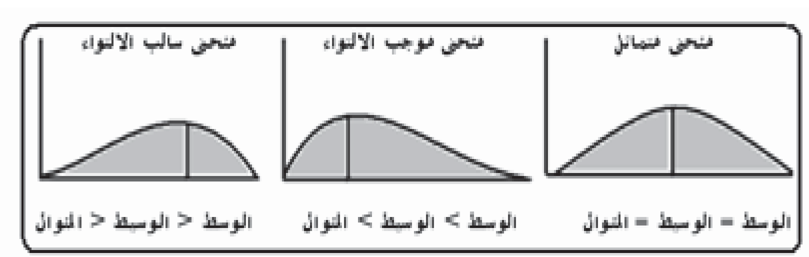
$$d_1 = (10-7) = 3 \quad d_2 = (10-5) = 5$$

- تحديد الحد الأدنى للفئة المنوالية $(A = 8)$ ، وكذلك طول الفئة $(L = 3)$
- وبتطبيق المعادلة الخاصة بحساب المنوال في حالة البيانات المبوبة؛ نجد أن:

$$\begin{aligned} Mod &= A + \frac{d_1}{d_1 + d_2} \times L \\ &= 8 + \frac{3}{3+5} \times 3 = 8 + 1.125 = 9.125 \end{aligned}$$

استخدام مقاييس النزعة المركزية في تحديد شكل توزيع البيانات :

يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري، والذي يعبر عن شكل توزيع البيانات ، كما يأتي:



- يكون المنحنى متماثل إذا كان : الوسط = الوسيط = المنوال .
- يكون المنحنى موجب الالتواء (ملتوي جهة اليمين) إذا كان: الوسط > الوسيط > المنوال.
- يكون المنحنى سالب الالتواء (ملتوي جهة اليسار) إذا كان: الوسط < الوسيط < المنوال.

مثال: قام مدير مراقبة الإنتاج بسحب عينة من 10عبوات من المياه المعبأة للشرب، ذات الحجم5 لتر، والمنتجة بواسطة إحدى شركات تعبئة المياه لفحص كميـة الأمـلاح الذائبـة، وكانـت كـالآتي:**115 123**
119 123 124 119 123 121 123 121

جد حساب الوسط الحسابي، والوسيط، والمنوال، ثم حدد شكل الالتواء لهذه البيانات.

الحل:

- **الوسط الحسابي :**

$$\overline{x} = \frac{\sum x}{n} = \frac{1211}{10} = 121.1$$

- **حساب الوسيط :**

رتبة الوسيط: $(n+1)/2 = (10+1)/2 = 5.5$

ترتيب القيم تصاعدياً:

قيمة الوسيط

122

الطاقة	115	119	119	121	121	123	123	123	123	124
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

عدد القيم = 10 ، وهو عدد زوجي. الوسيط = الوسط الحسابي للقيمتين رقم (5 ,6)

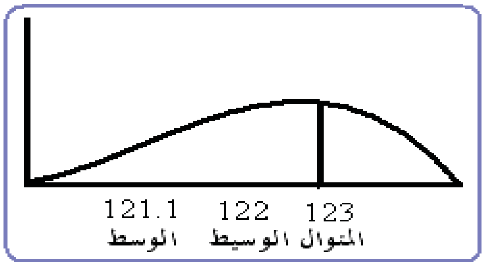
$$Med = \frac{121+123}{2} = \frac{244}{2} = 122$$

- **حساب المنوال :**

المنوال يساوى القيمة الأكثر تكرارا: القيمة 123 تكررت أكثر من غيرها، إذا

$$Mod = 123$$

وبمقارنة الوسط والوسيط والمنوال نجد أن :



نجد أن: الوسط < الوسيط < المنوال، إذا توزيع بيانات كمية الأملاح سالبة الالتواء.

مثال: الجدول التكراري التالي يعرض توزيع 100 عامل في مزرعة حسب الأجر اليومي بالدينار.

الأجر	50 -	70 -	90 -	110 -	130 -	150 -	170 - 190
عدد العمال	8	15	28	20	15	8	6

والمطلوب :

- حساب الوسط والوسيط والمنوال .
- بيان شكل توزيع الأجور في هذه المزرعة .

الحل:

- حساب الوسط والوسيط والمنوال .

1- الوسط الحسابي $\bar{x}$:

فئات الأجر	التكرارات (f)	مراكز الفئات (x)	f x
50 – 70	8	60	480
70 – 90	15	80	1200
90 – 110	28	100	2800
110 - 130	20	120	2400
130 - 150	15	140	2100
150 – 170	8	160	1280
170 - 190	6	180	1080
المجموع	100		11340

$$\overline{x} = \frac{\sum fx}{\sum f} = \frac{11340}{100} = 113.4 \ \ R.S$$

2-الوسيط Med :

رتبة الوسيط : (n/2 =100/2 =50)

تكوين التوزيع التكراري المتجمع الصاعد .

أقل من	تكرار متجمع صاعد
أقل من 50	**0**
أقل من 70	**8**
أقل من 90	$23 \leftarrow f_1$
أقل من 110	$51 \leftarrow f_1$
أقل من 130	**71**
أقل من 150	**86**
أقل من 170	**94**
أقل من 190	**100**

رتبة الوسيط (50)

من الجدول أعلاه نجد أن :

$$\frac{n}{2} = 50 \quad , f_1 = 23 \quad , f_2 = 51 \quad , A = 90 \quad , L = 110 - 90 = 20$$

إذا الوسيط قيمته هي:

$$Med = A + \frac{\frac{n}{2} - f_1}{f_2 - f_1} \times L = 90 + \frac{50 - 23}{51 - 23} \times 20$$

$$= 90 + \frac{27}{28} \times 20 = 90 + \frac{540}{28} = 90 + 19.286 = 109.3 \ \ R.S$$

3- المنوال Mod:

الفئة المنوالية، هي الفئة المناظرة لأكبر تكرار.

أكبر تكرار =28 ، وهو يناظر الفئة التقريبية (90 - 110) .

حساب الفروق: $d_1 = 28-15 = 13$, $d_2 = 28-20 = 8$

الحد الأدنى للفئة : $A = 90$ **طول الفئة :** $L = 110 - 90 = 20$

إذاً المنوال يحسب بتطبيق المعادلة التالية :

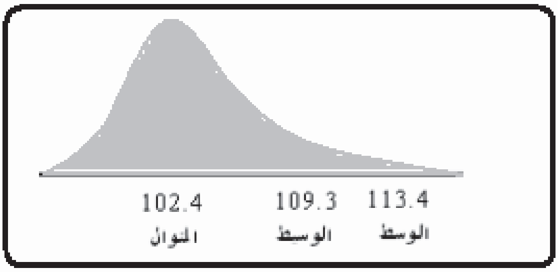
$$Mod = A + \frac{d_1}{d_1+d_2} \times L = 90 + \frac{13}{13+8} \times 20 = 90 + \frac{260}{21} = 1024 \ \ R.S$$

- **بيان شكل التوزيع .**

من النتائج السابقة نجد أن:

الوسط الحسابي : $\overline{x} = 113.4$ **الوسيط :** $Med = 109.3$ **المنوال :** $Mod = 1024$

أي أن : الوسط > **الوسيط** > **المنوال** إذا توزيع بيانات الأجور موجب الالتواء؛ كما هو مبين في الشكل التالي:



4- الرباعيات **Quartiles**:

هو مقياس من مقاييس الموضع Position(*) وهو عبارة عن مجموعة من القيم تقسم التكرار الكلي بنسب معينة. (المغربي، 2011، ص26).

عند تقسيم القيم إلى أربع أجزاء متساوية، يوجد ثلاث إحصاءات ترتيبي تسمى بالرباعيات، وهي:

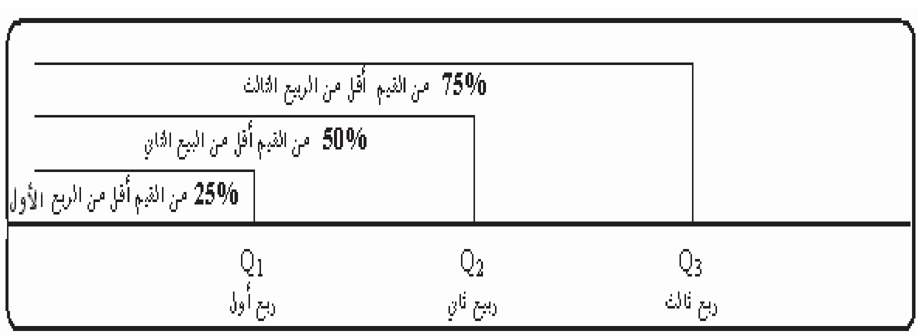
1- **الربيع الأول:** وهو القيمة التي يقل عنها ربع عدد القيم، أي يقل عنها 25% من القيم، ويرمز له بالرمز Q_1.

2- **الربيع الثاني:** وهو القيمة التي يقل عنها نصف عدد القيم، أي يقل عنها 50% من القيم، ويرمز له بالرمز Q_2، ومن ثم يعبر هذا الربيع عن الوسيط.

3- **الربيع الثالث:** وهو القيمة التي يقل عنها ثلاث أرباع عدد القيم، أي يقل عنها 75% من القيم، ويرمز له بالرمز Q_3.

والشكل الآتي يبين أماكن الرباعيات الثلاث.

الرباعيات



(*) مقاييس الموضع (Measures of Position): يقسّم الوسيط القيم إلى مجموعتين من حيث العدد، والربيعيات، والعشيرات، والمئينات مثل الوسيط تفيد نفس الغرض، وتسمى "مقاييس الموضع"؛ وكذلك تسمى "الربيعيات، والعشيرات، والمئينات" "شبيهات الوسيط". (المغربي، 2007، ص171)، (المغربي،2011، ص26).

ولحساب أي من الرباعيات الثلاث، يتم إتباع الآتي:

- بفرض أن عدد القيم عددها n، وأنها مرتبة كالآتي:

القيم مرتبة	$X_{(1)}$	<	$X_{(2)}$	<	$X_{(3)}$	<	$X_{(n)}$
الرتبة	1		2		3		n

- تحديد رتبة الرباعي رقم i، (Q_i) : $R=(n+1)\times\left(\frac{i}{4}\right)$
- إذا كانت R عددا صحيحا فإن قيمة الربيع هو: $Q_i = X_{(R)}$.
- إذا كانت R عدد كسري، فإن الرباعي (Q_i) يقع في المـدى: $X_{(l)} < Q_i < X_{(u)}$ ، ومـن ثـم يحـسب (Q_i) بالمعادلة الآتية:

$$\boxed{Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)})}$$

مثال: فيما يلي كمية الإنتاج اليومي من الحليب باللتر للبقرة الواحدة لعينة حجمهـا 10 أبقـار اختـيرت من مزرعة معينة: **25 23 29 32 34 29 20 18 27 30** ؛ أحسب الرباعيـات الثلاث لكمية الإنتاج، وما هو تعليقك؟

الحل:

لحساب الرباعيات الثلاث، يتم إتباع الآتي:

- **ترتيب القيم تصاعدياً:**

قمة الربيع		22.25			28			30.5		
القيم	18	20	23	25	27	29	29	30	32	34
الرتبة	1	2	3	4	5	6	7	8	9	10
رتبة الربيع		2.75			5.5			8.25		

- **حساب الربيع الأول** (Q_1)**:**

رتبة الربيع الأول هي: $R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{1}{4}\right) = 2.75$

يقع الربيع الأول بين القيمتين: $(20 < Q_1 < 23)$.

- **تطبيق المعادلة:**

$$\boxed{Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)})}$$

نجد أن:

$$l = 2,\ \ R = 2.75\ \ , x_{(l)} = 20\ \ . x_{(u)} = 23$$

إذا :

$$Q_1 = x_{(l)} + (R-l) \times (x_{(u)} - x_{(l)}) = 20 + 0.75(23-20) = 22.25$$

- **حساب الربيع الثاني (الوسيط)** Q_2

رتبة الربيع الثاني هي: $R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{2}{4}\right) = 5.5$

يقع الربيع الثاني بين القيمتين: $(27 < Q_2 < 29)$.

- **تطبيق المعادلة:**

$$\boxed{Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)})}$$

نجد أن:

$$l = 5,\ \ R = 5.5\ \ , x_{(l)} = 27\ \ . x_{(u)} = 29$$

إذا :

$$Q_2 = x_{(l)} + (R-l) \times (x_{(u)} - x_{(l)}) = 27 + 0.5(29-27) = 28$$

- **حساب الربيع الثالث** Q_3

رتبة الربيع الثالث هي: $R = (n+1) \times \left(\frac{i}{4}\right) = (10+1) \times \left(\frac{3}{4}\right) = 8.25$

يقع الربيع الثالث بين القيمتين: $(30 < Q_3 < 32)$.

- **تطبيق المعادلة:**

$$\boxed{Q_i = x_{(l)} + (R-l)(x_{(u)} - x_{(l)})}$$

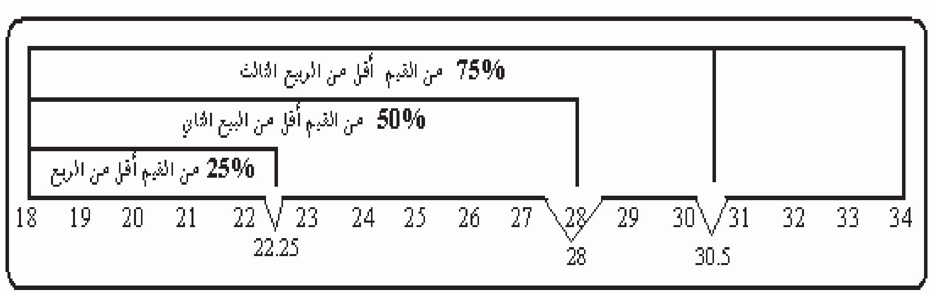
نجد أن:

$$l = 8, \quad R = 8.25 \quad ,x_{(l)} = 30 \quad .x_{(u)} = 32$$

إذا :

$$Q_3 = x_{(l)} + (R-l) \times (x_{(u)} - x_{(l)}) = 30 + 0.25(32-30) = 30.5$$

من النتائج السابقة نجد أن:



- من الأبقار يقل إنتاجه عن 22.25 لتر يومياً.
- 50% من الأبقار يقل إنتاجه عن 28 لتر يومياً.
- 75% من الأبقار يقل إنتاجه عن 30.5 لتر يومياً.

ثانياً: مقاييس التشتت (الاختلاف) **Measures of Dispersion**:

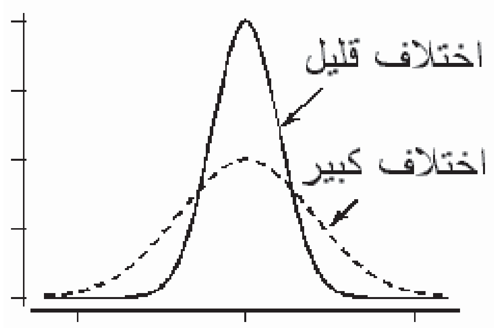
مقدمة:

لقد ذكرنا في المواضيع السابقة بعض مقاييس النزعة المركزية والتي تمثل مقـاييس عدديـة لموضـع أو مكان تركز البيانات لظاهرة ما، وتستخدم مقاييس النزعة المركزية لمقارنـة مجموعـات البيانـات المختلفـة، وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعة البيانات المختلفة، فقد تكون هناك مجموعات من البيانات لها نفـس مقـاييس النزعـة المركزيـة (لهـا نفـس الموضـع) ولكنهـا تختلف في بعض الصفات الأخرى، فمثلاً المثال الآتي يبـين مجمـوعتين مـن البيانـات لهـما نفـس المتوسـط ولكنهما مختلفان في طبيعة تشتتهما.

مثال:

المجموعة	البيانات	المتوسط	
الأولى	59, 61, 62, 58, 60	60	-----------------●--●--●--●--●-----------
الثانية	50, 60, 66, 54, 70	60	-----------●-----●-----●-----●------●--- 60

بالرغم من أن المتوسط يساوي **60** للمجموعتين إلا أن التشتت (أو الاختلاف) بـين القـيم في كـل مجموعة غير متساو؛ فمن الواضح أن بيانات المجموعة الأولى أكثر تقارباً فيما بينها (أقل تشتتاً وتباعداً فيما بينها) من بيانات المجموعة الثانية؛ لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعـة تـشتت (أو تفرق أو اختلاف أو تباعد) البيانات فيما بينها، وهذه المقاييس تسمى مقاييس التشتت أو الاختلاف.



المضلعان التكراريان لتوزيعين لهما نفس مقاييس

النزعة المركزية ولكنهما مختلفين في التشتت

فالتشتت Dispersion هو مدى الاقتراب أو الابتعاد للمفردات حول وسطها الحسابي، فإذا كانت البيانات مركزه حول الوسط الحسابي فإن التشتت يكون صغيراً، أما إذا كانت البيانات مبعثرة بعيداً عن الوسط يكون التشتت كبيراً، إذن التشتت بمقاييسه يبين مدى تجانس لمجموعات.(المغربي،2007،ص321).

فمقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات؛ والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها، فتشتت البيانات يكون صغيراً إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس، وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها، ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة، ومن أشهر مقاييس التشتت هي:

1- المدى Range.
2- نصف المدى الربيعي Semi- Inter-Quartile Range.
3- التباين Variance.
4- الانحراف المعياري Standard Deviation.

5- معامل الاختلاف (أو التغيير) Coefficient of Variation.

6- الدرجات المعيارية Standard Units.

7- الخطأ المعياري Standard Error.

وفيما يأتي عرض ذلك:

أولاً: المدى **Range**:

يعد المدى من أسهل مقاييس التشتت تعريفاً وحساباً ويعطينا فكرة سريعة عن مدى تفرق البيانات، ويعرف المدى لمجموعة من البيانات بالصيغة الآتية:

$$Range = X_{max} - X_{min}$$

حيث أن:

$X_{max} =$ **أكبر قيمة (للبيانات المفردة) = مركز الفترة العليا (للبيانات المبوبة)**

$X_{min} =$ **أصغر قيمة (للبيانات المفردة) = مركز الفترة الدنيا (للبيانات المبوبة)**

مثال: أوجد المدى للمشاهدات الآتية التي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 25, 30, 40, 45, 35, 55, 50

الحل:

$X_{max} = 55$

$X_{min} = 25$

$Range = X_{max} - X_{min} = 55 - 25 = 30$ **كيلو جراماً**

مثال: أوجد المدى R لدرجات 80 طالباً وطالبة في مادة الإحصاء بالمرحلة الثانية في قسم العلوم التربوية والنفسية بكلية التربية للعلوم الإنسانية/ جامعة الأنبار للعام الدراسي 2012-2013م.

فئات درجات الطلبة	عدد الطلبة F
0-10	4
10-20	12
203-0	20
304-0	18
405-0	15
506-0	8
60-70	2
70-80	1
المجموع	80

الحل:

المدى = مركز الفئة العليا - مركز الفئة الدنيا

1- نقوم بحساب مركز الفئة العليا = (70+80) /2 = 75

2- نقوم بحساب مركز الفئة الدنيا = (0+10)/2 = 5

إذن: **المدى** =**75 - 5** =70 R

مميزات المدى (**R**) وعيوبه:

ذكر المغربي (2007) أن للمدى (R) مميزات وعيوب، فمن مميزاته:

1- يفيد المدى (R) في معرفة التشتت للمجتمعات لصغيرة المتجانسة كما في رقابة وخطوط الإنتاج.

2- المدى (R) سهل الحساب وهو من أبسط مقاييس التشتت.

ومن عيوب المدى (**R**):

1- المدى (R) أقل مقاييس التشتت دقة.

2- المدى (R) قيمته مضلله لأنه يعتمد في حسابه على قيمتين فقط (القيمة الصغرى، القيمة الكبرى).

3- لا يمكن حساب المدى (R) في الجداول المفتوحة وذلك لعدم معرفة الحد الأعلى للفئة الأخيرة أو الحد الأدنى للفئة الأولى.

4- لا يمكن استخدام المدى (R) للمقارنة بين توزيعات تختلف وحدات القيم فيها كالمقارنة بين توزيعين أحدهما خاص بالوزن بالكيلوجرام والآخر خاص بالطول بالسنتيمتر.

5- يتأثر بالقيم الشاذة أو المتطرفة، لذلك فهو مقياس تقريبي لا يمكن الاعتماد عليه. (المغربي، 2007،ص322).

ثانياً: نصف المدى الربيعي **Semi- Inter-Quartile Range**:

رأينا أن المدى يتأثر كثيراً بالقيم الشاذة أو المتطرفة، لذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة، وأحد هذه المقاييس هو نصف المدى الربيعي، حيث أن القيم المتطرفة هي القيم الصغيرة جداً أو الكبيرة جداً فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%).

¼ n =25%   ¼ n= 25%

-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-

ربع البيانات الصغيرة Q_1 Q_2 Q_3 ربع البيانات الكبيرة

ويرمز لنصف المدى الربيعي بالرمز (Q) ويعرف بالصيغة الآتية:

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

حيث أن (Q_1) هو الربيع الأول، و(Q_3) هو الربيع الثالث، وقد بينا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال:

أوجد نصف المدى الربيعي لأوزان مجموعة الطلاب الآتية:

67, 65, 69, 58, 55, 71, 72, 70

الحل:

نرتب البيانات ترتيباً تصاعدياً لنحصل على:

55, 58, 65, 67, 69, 70, 71, 72

$$Q_1 = \frac{58+65}{2} = 61.5, \qquad Q_3 = \frac{70+71}{2} = 70.5$$

$$Q = \frac{Q_3 - Q_1}{2} = 4.5$$

مثال:

أوجد نصف المدى الربيعي لأوزان مجموعة الطلاب الآتية:

59, 67, 65, 69, 58, 55, 70, 72, 74

الحل:

نرتب البيانات ترتيباً تصاعدياً لنحصل على:

55, 58, 59, 65, 67, 69, 70, 72, 74

Q1= 59

Q3 = 70

$$Q = \frac{Q_3 - Q_1}{2} = 5.5$$

مثال: الجدول التكراري التالي يبين توزيع 60 مزرعة حسب المساحة المنزرعة بالذرة بالألف دونم .

المساحة	15-20	20-25	253-0	-3035	354-0	-4045
عدد المزارع	3	9	15	18	12	3

والمطلوب حساب نصف المدى الربيعي للمساحة المزروعة بالذرة .

الحـل:

عند حساب الربيع الأول أو الثالث يتبع نفس الأسلوب المستخدم في حساب الوسيط.

- تكوين الجدول التكراري المتجمع الصاعد.
- حساب الرباعي الأول (Q_1).

رتبة الربيعي الأول : n(1/4)=60(0.25)=15

$$f=15 \ , \ f_1=12 \ , \ f_2=27 \ , \ A=25 \ , \ L=5$$

إذاً

$$Q_1 = A + \frac{f - f_1}{f_2 - f_1} L$$

$$= 25 + \frac{15 - 12}{27 - 12}(5) = 25 + \frac{3}{15}(5) = 26$$

حدود المساحة	عدد المزارع
15-	3
20-	9
25-	15
30-	18
35-	12
40-45	3
sum	60

أقل من	تكرار متجمع
15	0
20	3
A 25	f_1 12
	15
30	f_2 27
A 35	45 45
40	57
45	60

- **حساب** الرباعي **الثالث** (Q_3):

موقع الرباعي الثالث: n(3/4)=60(0.75)=45

$$f=45 \ , \ f_1=45 \ , \ f_2=57 \ , \ A=35 \ , \ L=5$$

إذا

$$Q_3 = A + \frac{f - f_1}{f_2 - f_1} L$$

$$= 35 + \frac{45 - 45}{57 - 45}(5) = 35 + \frac{(0)}{15}(5) = 35$$

- **نصف المدى الربيعي:**

$$Q = \frac{Q_3 - Q_1}{2} = \frac{35-26}{2} = 4.5$$

وهي نصف المدى الربيعي للمساحة المزروعة بالذرة

مثال: أوجد نصف المدى الربيعي حسابياً لدرجات الطلاب في الجدول أدناه:

الفئات	404-9	505-9	606-9	70-79	80-89	90-99
عدد الطلاب	2	9	15	11	1	2

الحل: نكون جدول تكرار المتجمع الصاعد كما يأتي:

حدود الفئات	التكرار المتجمع الصاعد
< 39.5	0
< 49.5	2
< 59.5	11
< 69.5	26
< 79.5	37
< 89.5	39
< 99.5	40

$n = 40, \qquad \frac{n}{4} = 10, \qquad \frac{3n}{4} = 30, \qquad L = 10$

$$Q_1 = 49.5 + \left(\frac{10-2}{11-2}\right) 10$$

$$Q_1 = 49.5 + 8.89 = 58.39$$

$$Q_3 = 69.5 + \left(\frac{30-26}{37-26}\right) 10$$

$$Q_3 = 69.5 + 3.64 = 73.14$$

$$Q = \frac{Q_3 - Q_1}{2} = 7.38$$

مميزات نصف المدى الربيعي وعيوبه:

يتميز نصف المدى الربيعي في أنه:

1- لا يتأثر بالقيم الشاذة أو المتطرفة.

2- يمكن حسابه من التوزيعات التكرارية المفتوحة من الطرفين.

ومن عيوب نصف المدى الربيعي أنه:

1- لا يأخذ في الاعتبار جميع البيانات.

2- لا يسهل التعامل معه في التحليل الإحصائي.

ثالثاً: الانحراف المتوسط (**Mean Deviation (MD**):

هو أحد مقاييس التشتت، ويعبر عنه بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي، فإذا كانت $x_1, x_2, \ldots, x_n$ هي القراءات التي تم أخذها عن ظاهرة معينة، وكان ($\overline{x} = \sum x/n$) عبارة عن الوسط الحسابي لهذه القراءات، فإن الانحراف المتوسط (MD) يحسب بتطبيق المعادلة الآتية:

$$\text{M.D.} = \frac{1}{n} \sum_{i=1}^{n} \left| x_i - \overline{x} \right|$$

والسبب في أخذ القيم المطلقة للانحرافات هو أن مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر، وهذه الصيغة تستخدم في حالة البيانات غير المبوبة.

أما في حالة البيانات المبوبة يعطي الانحراف المتوسط من خلال المعادلة الآتية:

$$\text{M.D.} = \frac{\sum_{i=1}^{n} f_i \left| x_i - \overline{x} \right|}{n}$$

مثال: أوجد الانحراف المتوسط لأعمار مجموعة الطلاب 6,5,7,7,8,9,0,5:

الحل: نكون جدول الحل الآتي:

x	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$
6	-1	1
5	-2	2
7	0	0
7	0	0
8	1	1
9	2	2
9	2	2
5	-2	2
56	0	10

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

$$\bar{x} = \frac{56}{8} = 7$$

$$M.D. = \frac{10}{8} = 1.25$$

M.D= 1.25 سنة

مثال: أوجد الانحراف المتوسط لدرجات الطلاب كما في الجدول الآتي:

الفئات	404-9	505-9	606-9	70-79	80-89	90-99
عدد الطلاب	2	9	15	11	1	2

الحل: نكون جدول الحل الآتي:

Classes	x	f	fx	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$
40-49	44.5	2	89	-21.25	21.25	42.5
50-59	54.5	9	490.5	-11.25	11.25	101.25
60-69	64.5	15	967.5	-1.25	1.25	18.75
70-79	74.5	11	819.5	8.75	8.75	96.75
80-89	84.5	2	169	18.75	18.75	37.5
90-99	94.5	1	94.5	28.75	28.75	28.75
		40	2630			325

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

$$\bar{x} = \frac{2630}{40} = 65.75$$

$$\text{M.D.} = \frac{325}{40} = 8.125$$

M.D = 8.125 درجة

رابعاً: الانحراف المعياري **Standard Deviation**:

من الصعب التعامل رياضياً (تحليلياً) مع الانحراف المتوسط، لـذلك دعـت الحاجـة إلى استخدام مقياس للتشتت بنفس قوة الانحراف المتوسط، ولكي يكون من الـسهل التعامـل معه تحليلياً، وبما أن الفكرة هي التخلص من الإشارات للانحرافـات فـإن تربيـع الانحـراف يخلصنا من الإشارة، ولهذا فإن الانحراف المعياري يعرف عن طريـق التبـاين الـذي يعـرف على أنه متوسط مربع انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز (σ^2) للمجتمع، وبالرمز (S^2) للعينات؛ والانحراف المعياري هو الجذر التربيعي لمتوسط مجموع مربعـات انحرافات القيم عن وسطها الحسابي، ونرمز لـه بـالرمز (σ) للمجتمـع، وبـالرمز (S) للعينـة،

أو هو الجذر التربيعي للتباين سواء للمجتمع أو العينة.(المغربي،2007،ص384).

وسوف نتناول طريقة حسابه في حالة البيانات المباشرة والبيانات المبوبة كما يأتي:

الانحراف المعياري في حالة البيانات المباشرة:

إذا كان لدينا مجموعة من البيانات من مجتمع إحصائي عدد مفرداته N على صورة $X_1,X_2,X_3......X_N$ ومتوسط هذه البيانات $\overline{X}$ فإن مربعات انحرافات هذه القيم عن وسطها الحسابي يكون كما في الآتي:

$$(X_1-\overline{X})^2,(X_2-\overline{X})^2,...,(X_N-\overline{X})^2$$

ويعرف التباين σ^2 كالآتي:

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^2$$

والانحراف المعياري σ هو

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

ويفضل عند حساب الانحراف المعياري أن يحسب التباين من المعادلة ويأخذ الجذر التربيعي للنتيجة النهائية لنحصل على معادلة الانحراف المعياري.

أما في حالة العينة التي حجمها (n) المأخوذة من المجتمع فإن الانحراف المعياري في هذه الحالة يرمز له بالرمز (S) والتباين (S^2) ويعرف بقسمة مجموع مربعات الانحرافات على (n-1) ويكتب كما يأتي:

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

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

وذلك في حالة البيانات المباشرة، و(S) يعطى تقديراً أفضل للانحراف المعياري للمجتمع الذي أخذت منه العينة، وإذا كان عدد المفردات كبير (أكثر من 30) فإن قيمة

(S^2) و (σ^2) تكون متساوية تقريباً وذلك من الناحية التطبيقية.

مثال: أحسب الانحراف المعياري لأعمار مجموعة من الطلاب في المرحلة الابتدائية 8,9,7,6,5

الحل:

x	$x-\bar{x}$	$(x-\bar{x})^2$
8	1	1
9	2	4
7	0	0
6	-1	1
5	-2	4
35	0	10

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

$$\bar{x} = \frac{35}{5} = 7$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$S^2 = \frac{1}{5-1}(10) = 2.5$$

$$S = \sqrt{2.5} = 1.581$$

S= 1.581 سنة

خصائص الانحراف المعياري:

من خصائص الانحراف المعياري ما يأتي:

الخاصية الأولى: إذا أضفنا أو طرحنا مقداراً ثابتاً c من جميع القراءات لمجموعة البيانات فإن الانحراف المعياري للقيم الجديدة هو نفسه الانحراف المعياري للقيم الأصلية ويمكن إثبات ذلك بالآتي:

نفرض أن القيم الأصلية هي $x_1,x_2 \ldots x_n$ ونفرض أن القيم الجديدة هي $d_1, d_2 .. dn$

حيث: $d_n = x_n+c, \ldots , d_2=x_2+c, d_1= x_1+c$

$$\because S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}[(d_i \pm c)-(\bar{d}\pm c)]^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2\right.$$

ويمكن أن تستخدم هذه الخاصية في تبسيط البيانات، لاسيما عندما تكون قيمتها كبيرة كما يوضح ذلك المثال الآتي:

مثال: استخدم الخاصية الأولى في حساب الانحراف المعياري لأعمار مجموعة من الطلاب في المرحلة الابتدائية 8,9,7,6,5 باختيار الثابت c يساوي 5:

الحل: نطرح المقدار الثابت من كل القراءات كما هو موضح بالجدول الآتي:

x	$d=x-5$	d^2
8	3	9
9	4	16
7	2	4
6	1	1
5	0	0
	10	30

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

$$S^2 = \frac{1}{4}\left(30-\frac{(10)^2}{5}\right) = 2.5$$

$$S = \sqrt{2.5} = 1.581,$$ سنة

الخاصية الثانية: إذا ضربنا جميع القيم في مقدار ثابت أو قسمناها على مقدار ثابت فإن الانحراف المعياري يتأثر بذلك وسوف نثبت ذلك في حالة الضرب هذا، ويمكن إتباع نفس الخطوات في حالة القسمة كما يأتي:

نفرض أن القيم الأصلية هي $x_1, x_2 \ldots x_n$ فإذا ضربنا هذه القيم في مقدار ثابت c فتكون القيم الجديدة على الصورة $d_1, d_2 .. dn$ حيث:

$$d_1 = cx_1 \,,\; d_2 = cx_2 \; \ldots \; dn = cx_n$$

وعليه فإن:

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{d_i}{c} - \frac{\bar{d}}{c}\right)^2 = \frac{1}{n-1} \frac{1}{c^2} \sum_{i=1}^{n} (d_i - \bar{d})^2$$
$$S_x^2 = \frac{1}{c^2} S_d^2 \Rightarrow S_x = \frac{1}{c} S_d$$

أي أن الانحراف المعياري والقيم الأصلية في حالة الضرب يساوي الانحراف المعياري للقيم الجديدة مقسوماً على المقدار الثابت، أما في حالة القسمة فإنه يمكن إثبات أن:

$$S_x = c\, S_d$$

أي أن الانحراف المعياري للقيم الأصيلة يساوي الانحراف المعياري للقيم الجديدة مضروباً في المقدار الثابت.

الخاصية الثالثة: مجموع مربعات انحرافات القيم عن وسطها الحسابي $\bar{X}$ تكون أصغر من مجموع مربعات انحرافات القيم عن أي وسط فرضي آخر a حيث $a \neq \bar{X}$.

الإثبات:

$$\sum (x-a)^2 = \sum (x + \bar{x} - \bar{x} - a)^2$$
$$= \sum [(x - \bar{x}) + (\bar{x} - a)]^2$$
$$= \sum (x - \bar{x})^2 + n(\bar{x} - a)^2 + 2(\bar{x} - a)\sum (x - \bar{x})$$
$$= \sum (x - \bar{x})^2 + n(\bar{x} - a)^2$$

ونلاحظ أن المقدار $n(\overline{x}-a)^2$ مقدار موجب دائماً، ونستنتج من ذلك أن:

$\sum(x-a) < \sum(x-\overline{x})$ وهو المطلوب إثباته

الخاصية الرابعة: إذا كانت هناك عينتـان مجمـوع تكـرارهما هـو n1 وn2 وتبـاينهما S_1^2 و S_2^2 على الترتيب ولهما نفس المتوسط $\overline{x}$ فإن التباين المشترك هو:

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-1}$$

الإثبات:

نفرض أن المجوعتين هما:

$$x_1, x_2, \dots, x_{n_1}, \quad y_1, y_2, \dots, y_{n_2}$$

$$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\overline{x})^2$$

$$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i-\overline{x})^2$$

$$(n_1-1)\ S_1^2 = \sum_{i=1}^{n_1}(x_i-\overline{x})^2$$

$$(n_2-1)\ S_2^2 = \sum_{i=1}^{n_2}(y_i-\overline{y})^2$$

$$(n_1-1)\ S_1^2 + (n_2-1)\ S_2^2 = \sum_{i=1}^{n_1}(x_i-\overline{x})^2 + \sum_{i=1}^{n_2}(y_i-\overline{y})^2 = \sum_{i=1}^{n_1+n_2}(z_i-\overline{x})^2$$

$$\therefore S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-1}$$

الخاصية الخامسة: الانحراف المعياري لمجموعة من البيانات أكبر من الانحراف المتوسط لها.

الانحراف المعياري للبيانات المبوبة:

إذا كان لدينا عدد k من الفئات ذات المراكز $x_1, x_2 \ldots x_n$ ولها تكرارات $f1, \ f2\ldots. \ fk$ على الترتيب فإن المعادلات (1)، (2)، (3):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \cdots (1)$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \cdots (2)$$

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \right) \qquad \cdots (3)$$

تصبح كالآتي:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2 , \ S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} f_i (x_i - \bar{x})^2} , \qquad \cdots .(4)$$

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} f_i x_i^2 - \frac{(\sum_{i=1}^{n} f_i x_i)^2}{n} \right), \qquad \cdots .(5)$$

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} f_i d_i^2 - \frac{(\sum_{i=1}^{n} f_i d_i)^2}{n} \right), \qquad \cdots .(6)$$

وسوف نبين طريقة حساب الانحراف المعياري باستخدام العلاقات السابقة بالأمثلة الآتية:

مثال: أوجد الانحراف المعياري لدرجات الطلاب كما في الجدول الآتي باستخدام العلاقات (4)، (5)، (6):

الفئات	404-9	505-9	606-9	70-79	80-89	90-99
عدد الطلاب	2	9	15	11	1	2

الحل:

1- باستخدام العلاقة (4) **نكون جدول الحل الآتي:**

Classes	x	f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2\ f$
40-49	44.5	2	89	-21.25	451.56	903.13
50-59	55.5	9	490.5	-11.25	126.56	1139.06
60-69	65.5	15	967.5	-1.25	1.56	23.44
70-79	75.5	11	819.5	8.75	76.56	842.19
80-89	85.5	2	169	18.75	351.56	703.13
90-99	95.5	1	94.5	28.75	826.56	826.56
Total		40	2630			4437.5

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{1}{40}(2630) = 65.75$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \frac{1}{40-1}(4437.5) = 113.78$$

S= 10.67 درجة

2- باستخدام العلاقة (5) **نكون جدول الحل الآتي:**

Classes	x	f	xf	$x^2\ f$
40-49	44.5	2	89	3960.5
50-59	55.5	9	490.5	26732.25
60-69	65.5	15	967.5	62403.75
70-79	75.5	11	819.5	61052.75
80-89	85.5	2	169	12280.5
90-99	95.5	1	94.5	8930.25
Total		40	2630	177360

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} f_i x_i^2 - \frac{(\sum_{i=1}^{n} f_i x_i)^2}{n} \right) = \frac{1}{40-1}(177360 - 172922.5) = 113.78$$

S=10.67 درجة (وهي نفس الدرجة السابقة)

3- باستخدام العلاقة (6) نأخذ المقدار الثابت (الوسط الفرضي) c=64.5 وهو مركز الفئة التي يقابلها أكبر تكرار (وذلك لتبسيط الحسابات) كما هو موضح بجدول الحل الآتي:

Classes	x	f	$d=x-64.5$	df	$d^2 f$
40-49	44.5	2	-20	-40	800
50-59	55.5	9	-10	-90	900
60-69	65.5	15	0	0	0
70-79	75.5	11	10	110	1100
80-89	85.5	2	20	40	800
90-99	95.5	1	30	30	900
Total		40		50	4500

$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{n} f_i d_i^2-\frac{(\sum_{i=1}^{n} f_i d_i)^2}{n}\right)=\frac{1}{40-1}(4500-62.5)=113.78$$

S=10.67 درجة

ملاحظة:

باستخدام الخاصية الثانية نقوم بحل المثال السابق، وقد لاحظنا أنه باستخدام الوسط الفرضي وهو مركز الفئة التي يقابلها أعلى تكرار فد بسطت الحسابات كثيراً، هذا ويمكن تبسيط الحسابات أكثر وذلك بقسمة انحرافات القيم عن الوسط الفرضي على طول الفئة (تستخدم هذه الطريقة في حالة الفئات المنتظمة) وبذلك يكون الحل على النحو الآتي:

Classes	x	f	$d=x-64.5$	$d/10=d'$	$d'f$	$d'^2 f$
40-49	44.5	2	-20	-2	-4	8
50-59	55.5	9	-10	-1	-9	9
60-69	65.5	15	0	0	0	0
70-79	75.5	11	10	1	11	11
80-89	85.5	2	20	2	4	8
90-99	95.5	1	30	3	3	9
Total		**40**			**5**	**45**

$$S_{d'}^2=\frac{1}{n-1}\left(\sum_{i=1}^{n} f_i d_i'^2-\frac{(\sum_{i=1}^{n} f_i d_i')^2}{n}\right)=\frac{1}{40-1}(45-0.625)=1.1378$$

$S_{d'}=1.067$, $S_x=10\,S_{d'}=10.67$ درجة (وهي نفس النتيجة السابقة)

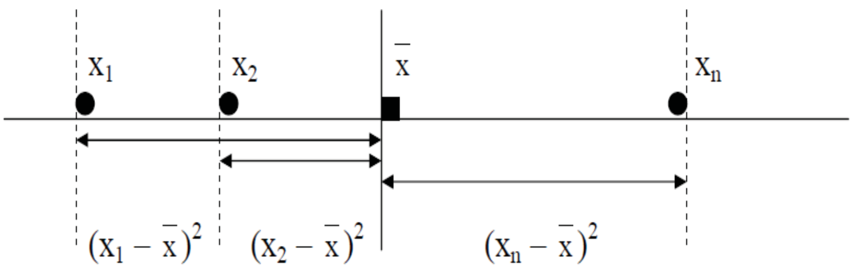
أن مميزات وعيوب الانحراف المعياري هي نفس مميزات وعيوب المتوسط الحسابي الذي مرّ ذكره سابقاً.

خامساً: التباين **Variance**:

التباين الذي يعرف على أنه متوسط مربع انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز (σ^2) للمجتمع، وبالرمز (S^2) للعينات.

وتعتمد فكرة التباين على تشتت أو تباعد البيانات عن متوسطها، فالتباين يكون كبيراً إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

فالتباين هو متوسط مربع انحرافات القيم عن وسطها الحسابي ويرمز له بالرمز (S^2).



القيم (البيانات)	x_n	$\cdots$	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	$\cdots$	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	$\cdots$	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

إذا كانت $x_1, x_2 \ldots x_n$ عينة حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يُعرف كما يأتي:

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

ملاحظات:

1- $S^2 \geq 0$ "دائماً" وكذلك $S \geq 0$ "دائماً".

2- $S^2 = 0 \Longleftrightarrow S = 0 \Longleftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).

3- وحدة S^2 هي وحدة البيانات الأصلية المربعة.

4- يمكن حساب التباين بالصيغة الحسابية الآتية:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

ولحساب تباين العينة باستخدام الصيغة الحسابية السابقة فإننا نحتاج إلى معرفة الكميات الآتية فقط دون الحاجة لمعرفة البيانات الأصلية:

1- حجم العينة = n .

2- مجموع البيانات= $\sum_{i=1}^{n} x_i$

3- مجموع مربعات البيانات= $\sum_{i=1}^{n} x_i^2$

وتستخدم الصيغة الحسابية السابقة لحساب تباين العينة لسببين:

- لأنها أكثر سهولة.
- لأنها أكثر دقة في الحساب عندما يكون هناك تقريب في حساب متوسط العينة.

مثال: أوجد تباين العينة لمجموعة الأوزان (بالكيلوجرام) الآتية:

8.3 ,5.4 ,2.5 ,2.5, 7.1

الحل:

نلخص الحل في الجدول الآتي:

x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	−2.66	7.0756	6.25
2.5	−2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n} (x_i - \overline{x})^2 = 27.832$	$\sum_{i=1}^{n} x_i^2 = 160.96$

$n = 5$

$\sum_{i=1}^{n} x_i = 25.8$

$\sum_{i=1}^{n} (x_i - \overline{x})^2 = 27.832$

$\sum_{i=1}^{n} x_i^2 = 160.96$

متوسط العينة هو:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16 \text{ (كيلوجرامًا)}$$

حساب تباين العينة:

1- باستخدام التعريف:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958 \text{ (كيلوجرامًا مربعًا)}$$

2- باستخدام الصيغة الحسابية:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{160.96 - \frac{(25.8)^2}{5}}{5-1} = \frac{160.96 - 133.128}{4} = 6.958$$

سادساً: الدرجات المعيارية **Standard Units:**

تعد الدرجات المعيارية من أهم مقاييس التشتت النسبي، والدرجة المعيارية تعبر عن بعد الدرجة الخام عن المتوسط الحسابي للمجموعة بدلالة وحدات الانحراف المعياري.

وفي كثير من الأحيان نحتاج إلى مقارنة مفردتين من مجموعتين مختلفتين، وفي هذه الحالة يجب تحويل وحدات كل مفردة إلى وحدات قياسية حتى تكون المقارنة ذات معنى وذلك باستخدام الوسط الحسابي والانحراف المعياري لكل مجموعة. (المغربي، 2011، ص132).

فعند مقارنة درجة طالب في مادة الإحصاء التربوي ومادة مناهج البحث التربوي ليس من المنطق أن نقارن بين مستوى الطالب في كل من المادتين على أساس الدرجة التي حصل عليها، لأن المقارنة ستكون ظالمة، ولكن لابد من مقارنة ذلك بالنسبة لتوزيع درجات الطلاب في كل مادة، ومن ثم نقارن بين موضع كل من هاتين القيمتين على التوزيع الخاص بهما، أي معرفة بعد الدرجة عن الوسط الحسابي بدلالة وحدات الانحراف المعياري ويتم ذلك بطرح الوسط الحسابي لكل توزيع من القيمة الخاصة به وقسمة الناتج على الانحراف المعياري فنحصل على الدرجة المعيارية. (المغربي، 2007، ص480).

لتكن $x_1, x_2,...x_n$ عينة من البيانات حجمها n ومتوسطها $\overline{x}$ وانحرافها المعياري S، ونعرف الدرجة المعيارية للمادة بالصيغة الآتية:

$$z_i=\frac{x_i-\overline{x}}{S} \quad i=1, 2, \dots, n$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2,...x_n$ هي:

$$z_1=\frac{x_1-\overline{x}}{S}, \ z_2=\frac{x_2-\overline{x}}{S}, \ \dots, \ z_n=\frac{x_n-\overline{x}}{S}$$

ملاحظات:

1- $z_i=\frac{x_i-\overline{x}}{S}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.

2- المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i=\overline{x}+Sz_i$

3- الدرجات المعيارية هي عديمة الوحدة لذلك فإنها تستخدم للمقارنة بين المشاهدات المختلفة في المجموعات المختلفة للبيانات.

4- متوسط الدرجات المعيارية = 0

5- الانحراف المعياري للدرجات المعيارية يساوي= 1.

مثال: إذا كان لدينا مجموعة من البيانات متوسطها7 = $\overline{X}$ وانحرافها المعياري S=5 فأوجد:

1- الدرجة المعيارية للقيمة X=9.

2- القيمة الأصلية للدرجة المعيارية Z=0.1

الحل:

1- الدرجة المعيارية للقيمة X=9 **هي:**

$$z=\frac{x-\overline{x}}{S}=\frac{9-7}{5}=\frac{2}{5}=0.4$$

2- القيمة الأصلية للدرجة المعيارية Z=0.1 **هي:**

$$x=\overline{x}+S\,z=7+5\times 0.1=7+0.5=7.5$$

مثال: إذا كانت درجة أحد الطلاب في مادة الإحصاء تساوي (82)، ودرجته في مادة الرياضيات تساوي (89)، وإذا كان متوسط درجات الطلاب في مادة الإحصاء تساوي (75) بانحراف معياري يساوي (10) ومتوسط درجات الطلاب في مادة الرياضيات يساوي (81) بانحراف معياري يساوي (16)، ففي أي المادتين كان أداء الطالب أفضل ؟.

الحل:

المادة	المتوسط $\overline{X}$	الانحراف المعياري S	الدرجة X	الدرجة المعيارية $z=\frac{x-\overline{x}}{S}$
الإحصاء	**75**	**10**	**82**	$z=\frac{82-75}{10}=0.7$
الرياضيات	**81**	**16**	**89**	$z=\frac{89-81}{16}=0.5$

بما أن الدرجة المعيارية لمادة الإحصاء (0.7) أكبر من الدرجة المعيارية لمادة الرياضيات (0.5) فإن أداء الطالب في مادة الإحصاء أفضل من أدائه لمادة الرياضيات بالرغم من أن درجته في مادة الإحصاء أقل من درجته في مادة الرياضيات.

ثالثاً: مقاييس العلاقة **Measures of Correlation**:

معامل الارتباط **Correlation Coefficient**:

هو معامل رقمي يوضح الارتباط بين ظاهرتين أو عدة متغيرات، وكثيراً ما نجد أنفسنا في حاجة لدراسة العلاقة بين متغيرين أو أكثر وتحديد مقدار العلاقة بين المتغيرات، فكلما كانت القيمة العددية لمعاملات الارتباط عالية كلما كانت العلاقات قوية ويكون الناتج قريباً من الواحد الصحيح، ليس ذلك فحسب بل إن معامل الارتباط يبرهن عن اتجاه تلك العلاقة التي قد تكون موجبة فكلما زاد الذكاء زاد التحصيل، وكلما ارتفعت الحرارة تمدد الحديد، وإما أن تكون العلاقة سالبة حيث يرافق زيادة أحد المتغيرين نقصان في المتغير الآخر فكلما زادت الحرارة ذاب الجليد، وكلما زاد الوعي البيئي قل تلوث البيئة وعندما يكون معامل الارتباط بين متغيرين صفراً فإن ذلك دليل على عدم وجود علاقة أو ارتباط بين المتغيرين كالعلاقة بين مهارات القراءة والطول في النمو الإنساني.

أهداف دراسة الارتباط:

1- توضيح العلاقة بين المتغيرين كالعلاقة بين ساعات الاستذكار والمجموع الكلي للدرجات.

2- إمكان تقدير احد المتغيرين إذا ما عرف المتغير الثاني.

3- التحكم أو توقع احتمال الظاهرة أو المتغير.

4- معرفة أسباب وجود العلاقة الارتباطية فقد يكون تغير الظاهرة راجع لتغير ظاهرة أخرى.

شكل الانتشار **Scatter Diagram** :

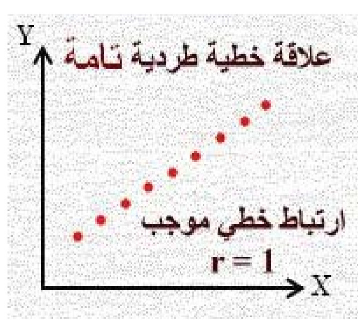
هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة، هذه الوسيلة هي "شكل الانتشار" والتي تصلح إذا كان المتغيران كميين؛ وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا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

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 الآتية تظهر بعض أشكال الانتشار المعروفة:

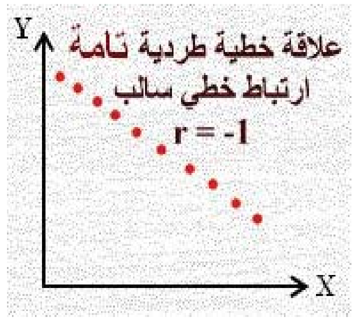
الشكل الأول :

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 الآتي:



ارتباط طردي تام (موجب)

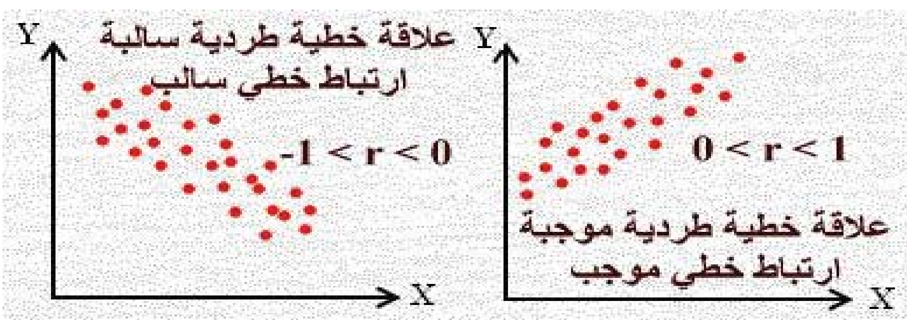
أمـا إذا كانـت العلاقـة عكـسية (وجميـع النقـاط تقـع عـلى خـط مـستقيم واحـد) فـإن " الارتباط عكسي تام " كما في الشكل الآتي:



ارتباط عكسي تام (سالب)

الشكل الثاني :

أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقـة خطية (موجبة أو سالبة) كما في الشكل الآتي:

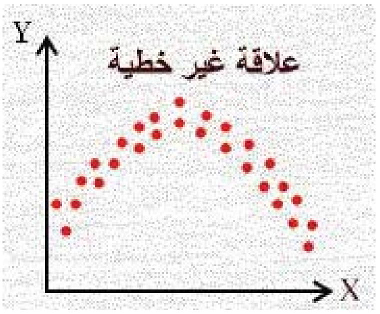


ارتباط موجب قوي (ارتباط خطي طردي)

ارتباط سالب قوي (ارتباط خطي عكسي)

الشكل الثالث :

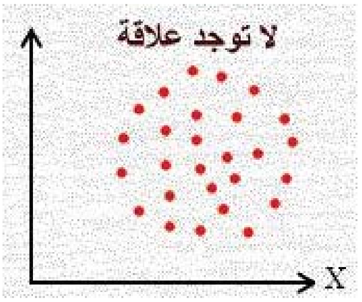
إذا كانت العلاقة تأخذ شكل منحنى فإن الارتبـاط لا يكـون خطيـاً "ارتبـاط غـير خطـي " Non Linear Correlation كما في الشكل الآتي:



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 الآتي:



لا توجد علاقة

تفسير معامل الارتباط:

تتراوح قيمة معامل الارتباط بين القيمة الدنيا (-1) والقيمة العليا (+1)، وتدل إشارة معامل الارتباط على اتجاه العلاقة في حين تدل القيمة على قوة العلاقة.

ويبين الجدول الآتي معايير نسبية يمكن أن تستخدم في تفسير قيم معاملات الارتباط. (عباس وآخرون، 2012، ص307):

قيم معاملات الارتباط وتفسيرها

قيمة معامل الارتباط	تفسير العلاقة
+1	**علاقة طردية تامة**
من (0.60 +) **إلى** (+0.99)	**ارتباط طردي قوي**
من (0.40 +) **إلى** (0.59 +)	**ارتباط طردي متوسط**
من (0.01 +) **إلى** (0.39 +)	**ارتباط طردي ضعيف**
صفر	**عدم وجود ارتباط**
من (0.39 -) **إلى** (0.01 –)	**ضعيفة عكسية**
من (0.59 -) **إلى** (0.40 -)	**متوسطة عكسية**
من (0.99 -) **إلى** (0.60 -)	**قوية عكسية**
-1	**ارتباط عكسي تام**

ويجدر الإشارة إلى أنه في العلوم الإنسانية يصعب إيجاد علاقة تامة بين المتغيرات وذلك لأن معظم المتغيرات لها علاقة بخصائص إنسانية يصعب ضبطها أو عزلها، وهذه الخصائص فيها تداخل وارتباط كبير مع تلك المتغيرات.

أنواع الارتباط:

هناك أربعة أنواع للارتباط هي:

1- الارتباط البسيط (r) **Simple Correlation:**

هو الارتباط الذي يبحث في العلاقة بين متغيرين أثنين فقط كالبحث في العلاقة بين التفكير الرياضي والتحصيل مثلاً.

2- الارتباط الجزئي (.r) **Partial Correlation :**

هو الارتباط الذي يبحث في العلاقة بين متغيرين أثنين بعد ضبط أثر متغير آخر أو متغيرات أخرى على أي من المتغيرين المقصودين أو كليهما، فإذا أراد باحث مثلاً تعرّف

العلاقة بين التفكير الرياضي والتحصيل، وقام بضبط أثر الدافعية في التحصيل فإن معامل الارتباط هنا يُسمى بمعامل الارتباط شبه الجزئي Semi-Partial Correlation.

أما إذا قام الباحث بضبط أثر الدافعية في كل من المتغيرين المقصودين (التفكير الرياضي والتحصيل)، فإن معامل الارتباط الناتج هما يُسمى بمعامل الارتباط الجزئي Partial Correlation.

3- الارتباط المتعدد(**Multiple Correlation (R**:

هو الارتباط الذي يبحث في العلاقة بين متغير معين من جهة وعدد من المتغيرات من جهة ثانية، كالبحث في العلاقة بين التحصيل من جهة وكل من التفكير الرياضي ودافعية الانجاز وموقع الضبط من جهة ثانية.

4- الارتباط الكاتوني **(Canonical Correlation (Rc**:

هو الارتباط الذي يبحث في العلاقة بين مجموعة من المتغيرات من جهة ومجموعة أخرى من المتغيرات من جهة ثانية، كالبحث في العلاقة بين كل من الذاكرة البصرية والذاكرة السمعية من جهة والتحصيل في الرياضيات والتحصيل في اللغة العربية من جهة ثانية.

أساليب حساب معامل الارتباط:

هناك عدد من الأساليب التي تُستخدم لحساب معامل الارتباط، والتي تختلف عن بعضها بعضاً تبعاً لطبيعة المتغيرات المعنية، ويُعد معامل ارتباط بيرسون Pearson أكثر معاملات الارتباط استخداماً، في حين تعد المعاملات الأخرى حالات خاصة من معامل ارتباط بيرسون، وسنتناول بعض معاملات الارتباط:

أولاً: معامل ارتباط بيرسون

Pearson Product Moment Correlation Coefficient (r)

يستخدم معامل ارتباط بيرسون لحساب قيمة معامل الارتباط عندما يكون المتغيرات المراد قياس الارتباط بينهما متغيرات مستقلة أو مستمرة، ويشترط تساوي عدد حالات كل من المتغيرين.

ولحساب قيمة معامل ارتباط بيرسون r نستخدم القانون الآتي:

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

ويكون لمعامل الارتباط (r) الخصائص الآتية:

1- قيمته تساوي صفراً عندما تكون الظاهرتان مستقلتان تماماً.

2- قيمته مقدار موجب عندما يكون الارتباط بين المتغيرين طردياً، ويكون قوياً عندما يكون المقدار الموجب قريباً من الواحد الصحيح، وضعيفاً عندما يكون المقدار الموجب قريباً من الصفر.

3- قيمته مقدار سالب عندما يكون الارتباط بين المتغيرين عكسياً، ويكون قوياً عندما يكون المقدار السالب قريباً من (-1) وضعيفاً عندما يكون المقدار السالب قريباً من الصفر.

مثال: الجدول الآتي يوضح درجات مجموعة مكونة من ثمانية طلاب في كل من مادتي الإحصاء والرياضيات في إحدى الامتحانات، هل هناك علاقة بين تحصيل الطالب في المادتين ؟.

11	16	8	11	15	19	9	13	الإحصاء X
10	14	9	10	15	17	7	15	الرياضيات Y

الحل:

تبسيط البيانات بالجدول، نطرح مقدار ثابت=10 من كل قيم X وقيم Y ونكوّن الجدول الآتي:

x	y	$- 10x = x$	$-10y = y$	$y\ x$	x^2	y^2
13	15	3	5	15	9	25
9	7	-1	-3	3	1	9
19	17	9	7	63	81	49
15	15	5	5	25	25	25
11	10	1	0	0	1	0
8	9	-2	-1	2	4	1
16	14	6	4	24	36	16
11	10	1	0	0	1	0
		22	17	132	158	125

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

$$r = \frac{132 - \frac{(22 \times 17)}{8}}{\sqrt{(158 - \frac{(22)^2}{8})(125 - \frac{(17)^2}{8})}}$$

$$r = 0.93$$

أي يوجد ارتباط قوي جداً بين درجات الطالب في المادتين.

ثانياً: معامل ارتباط الرتب (سبيرمان) **Spearman rank-order (r_s):**

معامل الارتباط الخطي لبيرسون الذي سبق الحديث عنه يقيس مقدار قوة الارتباط بين متغيرين وذلك في حالة البيانات الكمية، لكن في بعض الأحيان يكون مطلوب إيجاد قوة الارتباط بين متغيرين على صورة بيانات وصفية يمكن وضعها في صورة ترتيبية، مثال على هذا تقديرات الطلاب في مادتين مختلفتين، فيكون من الصعب حساب معامل ارتباط بيرسون، لذلك نشأت الحاجة إلى إيجاد مقياس يعطي قوة الارتباط للبيانات الوصفية؛ وهذا المقياس هو ما يسمى بمعامل ارتباط الرتب لسبيرمان، وهو يعطي مقياساً للارتباط في كل من البيانات الكمية والوصفية التي لها صفة الترتيب مثل تقديرات الطلاب، فإنه يمكن إعطاء رتب لها من حيث كبر التقدير وصغره وكذلك البيانات الكمية.

ونلاحظ أن رتب المتغيرين (X,Y) تزيد وتنقص حسب زيادة ونقص كل من قيم المتغيرين (X,Y)، لذلك فإن حساب معامل الارتباط للرتب يقترب كثيراً من معامل ارتباط بيرسون، ولكن يمتاز عنه في السهولة والدقة لاسيما عندما تكون أزواج القيم أقل من 15.

ويعطى معامل ارتباط الرتب بالعلاقة الآتية:

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

حيث r_s معامل ارتباط الرتب لسبيرمان، n تمثل عدد أزواج القيم ((X,Y، d هي الفرق بين رتب أزواج القيم (X,Y).

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يأتي:

1- مجموع الفروق بين الرتب يساوي صفر.

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

3- نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي:

$$\frac{n\,(n+1)}{2}$$

مثال: أوجد معامل ارتباط الرتيب لتقديرات الطلاب في كل من مادتي الإحصاء والرياضيات كما موضح بالجدول الآتي:

الرياضيات X	A	C	C	C	B	D
الإحصاء Y	B	B	D	C	A	E

الحل: نقوم بتلخيص الحل في الجدول الآتي:

الرياضيات x	الإحصاء y	رتبة $a = x$	رتبة $b = y$	$d = a - b$	d^2
A	B	6	4.5	1.5	2.25
C	B	3	4.5	-1.5	2.25
C	D	3	2	1	1
C	C	3	3	0	0
B	A	5	6	-1	1
D	E	1	1	0	0
					6.5

$$r_s = 1- \frac{6\sum d^2}{n(n^2-1)}$$
$$r_s = 1- \frac{6\times 6.5}{6(36-1)}$$
$$r_s = 1- 0.186$$
$$r_s = 0.814$$

أي يوجد ارتباط طردي قوي بين تقديرات مادتي الرياضيات والإحصاء.

مثال: البيانات الآتية تمثل إجابات عينة من سبعة أشخاص حول برنامج 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

المطلوب حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

الحل: ننظم الحل في الجدول الآتي مع ملاحظة ما يأتي:

1- **بالنسبة للسؤال الأول:** فإن التقدير الأعلى سيحصل على الرتبة (1)، والأقل منه مباشرة سيحصل على الرتبة (2) وهكذا.. أي أن الترتيب تنازلي، ونكرر العمل نفسه مع السؤال الثاني.

2- عند حصول إجابتين أو أكثر على التقدير نفسه نعطي لكل إجابة مبدئياً رتبة (كما لو كانوا مختلفين)، ثم نحسب متوسط هذه الرتب، وهذا المتوسط هو الذي يعطى لكل إجابة.

3- ثم نحسب الفروق بين رتب السؤالين ونرمز لها بالرمز (d) ثم نربع هذه الفروق فنحصل على (d^2) ونعوض في القانون، مع ملاحظة أن n=7.

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	- 0.05	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

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

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

$$r_s = 0.54$$

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

مثال: البيانات الآتية تمثل أعداد الساعات التي قرأ فيها عشرة طلاب والدرجات التي حصلوا عليها في امتحان أحد المواد الدراسية:

عدد الساعات X	10	6	12	14	11	6	19	16	3	9
الدرجات y	60	48	83	76	74	58	98	89	37	69

احسب معامل سبيرمان لارتباط الرتب:

الحل: كما في المثال السابق ننظم الحل في الجدول الآتي مع ملاحظة أن n=10.

عدد الساعات X	الدرجات Y	رتب X	رتب Y	الفروق d	d^2
10	60	6	7	- 1	1.00
6	48	8.5	9	- 0.5	0.05
12	83	4	3	1	1.00
14	76	3	4	- 1	1.00
11	74	5	5	0	0
6	58	8.5	8	0.5	0.25
19	98	1	1	0	0
16	89	2	2	0	0
3	37	10	10	0	0
9	69	7	6	1	1.00
المجموع				Zero	4.50

وبالتعويض في القانون حيث $\sum d^2 = 4.5$ ، n = 10 نحصل على:

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(4.5)}{10(1-99)} = 1 - \frac{27}{990} = 1 - 0.027$$

$$r_s = 0.973$$

مما يعني أننا أمام علاقة طردية قوية بين المتغيرين، فكلما زادت عدد الساعات التي يدرسها الطالب في هذا المثال، كلما زادت درجاته في الامتحانات قوة وذلك بما نسبته (97%).

أهم الخواص الإحصائية لمعاملات الارتباط:

1- عدم تأثر معامل الارتباط بزيادة أو نقصان درجات الاختبارات أو قيم المفردات بكمية ثابتة.

2- عندما تزداد قيم التوزيع التكراري العددية لمعاملات الارتباط يلتوي المنحنى، وكلما اقتربت القيم من الواحد الصحيح يميل التوزيع التكراري لمعاملات الارتباط للالتواء الشديد.

3- يصل الارتباط إلى نهايته العظمى عندنا يقترن المتغير الأول اقتراناً تاماً بتغير المتغير الثاني، ويكون معامل الارتباط مساوياً (1) في حالة الارتباط الطردي الكامل، ويساوي (-1) في حالة الارتباط العكسي التام، في حيت يساوي صفراً حال عدم الارتباط.

4- تزداد شدة الارتباط بزيادة القيمة العددية لمعامل الارتباط والعكس صحيح حيث تزداد شدة الارتباط كلما اقترب معامل الارتباط من (1) وتقل كلما اقترب من الصفر.