وزارة التعليـــــــم العـــــــالي والبحث العلمي

جامعــة 8 مـــاي 1945 قالمـــة

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قســـم العلـــــــوم التجارية

مطبوعة خاصة بمقياس

الإحصاء I

إعداد الدكتور: بشيشي وليد

السنة الجامعية: 2016/2015

تمهيد:

كلمة الإحصاء تعني من الناحية اللغوية عملية الحصر أو العد للأشياء. و يعتبر الإحصاء فرعا مهما من فروع العلم و المعرفة لأنه يدرس بشكل أساسي الناحية الكمية للظواهر الاقتصادية و الاجتماعية بارتباط وثيق مع الكيف، و ذلك باستخدام الطرق و المبادئ الإحصائية المناسبة، فهو يدرس الظاهرة حسب المكان و علاقتها بالظواهر الأخرى، كما يدرس تطور هذه الظاهرة حسب الزمان و التنبؤ بحجمها في المستقبل آخذا بعين الاعتبار العوامل التي تؤثر على هذه الظاهرة في الماضي و تغير هذه العوامل أو تغير تأثيرها في المستقبل الذي لا غنى عنه لمعرفة حقيقة الظاهرة و التخطيط لها. و لكي يقوم الإحصاء بمهمته لابد من جمع المعلومات الإحصائية عن الظاهرة المدروسة حيث يقوم بعد ذلك بتفريغها و تبويبها، و من ثم تحليلها و دراستها للوصول إلى نتائج و استنتاجات

سؤال: ما مدى الحاجة للإحصاء ؟

ربما يقبل البعض حقيقة الإلمام ببعض مفاهيم الإحصاء بأنها ضرورية في الوقت الحديث لمدى الحاجة الماسة لقدر معين من الإحصاء الذي يساعد بقدر كبير على:

1- وصف و وفهم العلاقات بين الظواهر

2- اتخاذ أفضل القرارات

3- التعامل بنجاح مع التغيرات

تعريف علم الإحصاء

يمكن تعريف علم الإحصاء بأنه العلم الذي يختص بالطرق العلمية لجمع البيانات عن الظواهر، تصنيفها، تلخيصها و عرضها في جداول منظمة و تمثيلها بيانيا على شكل رسوم بيانية، كذلك تحليل البيانات و استخلاص النتائج منها لاستخدامها في اتخاذ القرارات المناسبة و مقارنة الظواهر بعضها ببعض و محاولة استنتاج علاقات بينها.

أقسام علم الإحصاء

ينقسم علم الإحصاء إلى القسمين الرئيسيين الآتيين:

- **القسم الأول: الإحصاء الوصفي**

و يتناول الطرق الخاصة بجمع البيانات و تصنيفها و عرضها جدوليا أو بيانيا، كذلك حساب بعض المقاييس الإحصائية كالمتوسطات و مقاييس التشتت، كذلك دراسة العلاقة التي قد توجد بين ظاهرتين أو أكثر باستخدام مقاييس إحصائية خاصة.

- **القسم الثاني: الإحصاء الاستدلالي (الاستقرائي)**

و يتناول طرق و أساليب التوصل إلى استدلالات من عينات محدودة يتم تعميمها على المجتمع الذي يحتوي على الظاهرة أو الظواهر موضع البحث بالاعتماد على المنطق و نظرية الاحتمالات.

بعض المصطلحات الإحصائية

- **المجتمع:** المجتمع الإحصائي هو عبارة عن جميع الوحدات موضع الدراسة، سواء كانت هذه الوحدات أفرادا أو أشياء أو قياسات ...الخ، فهو مجموعة من المفردات التي تشترك في صفة واحدة أو أكثر. و مجتمع البحث أو الدراسة إما أن يكون عدد أفراده محدودا كعدد سكان دولة ما أو عدد طلاب جامعة ما، أو يكون عدد أفراده غير محدود أو لا نهائي كعدد الأسماك في البحار و المحيطات.
- **العينة:** جزء صغير من المجتمع يلجأ الباحث عادة إلى دراسته، حيث إن العينة تسحب من المجتمع الإحصائي لغرض دراسة صفاته و خصائصه، لذلك يراعى أن تكون هذه العينة تتوفر على أقصى قدر ممكن من دقة التمثيل لهذا المجتمع. و يطلق على الطريقة التي يتم بها اختيار مفردات أو وحدات العينة، أسلوب المعاينة.
- **المتغير:** الصفة التي تتغير من شخص إلى آخر أو من مفردة إلى أخرى تسمى ظاهرة أو المتغير و يرمز له (X) و لكل مفردة أو مشاهدة منها يرمز لها بالرمز (x_i) $i=1,2,\dots,n$ فمثلا عند دراسة أطوال مجموعة من طلبة جامعة ما فإننا نرمز لصفة الطول بالرمز X و طول أي طالب(ة) بالرمز x_i و هذه تسمى بالمشاهدة أو المفردة. و تنقسم المتغيرات إلى نوعين:

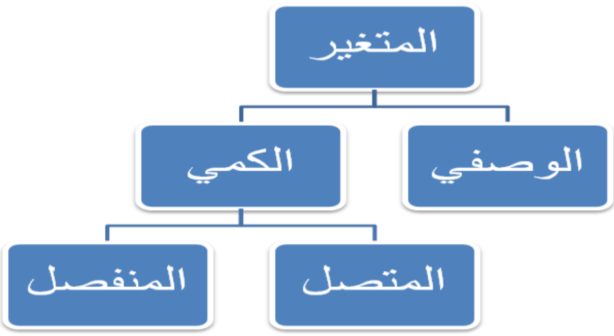
☜ **المتغيرات الوصفية أو النوعية:** و هي تلك الظواهر أو الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون العين، الحالة الاجتماعية، الجنس، إلى غير ذلك من الصفات.

☜ **المتغيرات الكمية:** و هي تلك الظواهر أو الصفات التي يمكن قياسها مباشرة بالأرقام العددية مثل صفة الطول و الوزن... الخ. و يمكن تقسيم المتغيرات الكمية بدورها إلى قسمين:

❈ **المتغيرات المنفصلة أو المتقطعة أو الوثابة:** و هي التي تأخذ المشاهدة أو المفردة فيها قيما متميزة عن بعضها و من أمثلتها عدد أفراد الأسرة أو عدد الوحدات الإنتاجية في مصنع ما. و بصورة عامة، يمكن القول أن البيانات التي نحصل عليها من العد تعتبر بيانات لمتغير منفصل.

❈ **المتغيرات المتصلة أو المستمرة:** و هي تلك المتغيرات التي تأخذ مدى أو مجال معين من القيم و من أمثلتها أطوال الأشخاص، أوزانهم، أعمارهم...الخ.

يتبين من أعلاه أن العلاقة بين المتغيرات تكون كالآتي:



الفصل الأول

1

عرض البيانات

البيانات الأولية (الخام) الخاصة بالدراسة لا يمكن تفسيرها بشكل ملائم و الاستفادة منها وهي بهذه الصورة، لذلك يلجأ الباحث إلى وضع تلك البيانات بشكل جداول مبسطة أو التعبير عنها بالرسوم البيانية لكي يسهل دراستها و تحليلها. و من طرق عرض البيانات نذكر:

1- العرض الجدولي و التوزيع التكراري: و هو عبارة عن عرض البيانات التي تم جمعها ووصفها في جداول منتظمة. و هنا يمكن التمييز بين:

أ-الجداول البسيطة: و هي تلك الجداول التي تتوزع فيها البيانات حسب صفة واحدة و يتألف عادة من عمودين، الأول يمثل الظاهرة و الثاني يمثل عدد المفردات التابعة لكل مشاهدة و هو ما يسمى بالتكرار المطلق و يرمز له بالرمز Fi

مثال: الجدول التالي يبين توزيع عينة من 40 أسرة حسب الحجم

حجم الأسرة	1	2	3	4	5	6	7	8	9	المجموع
التكرار (Fi)	4	8	9	4	3	6	3	2	1	**40**

ب- الجداول المركبة (المزدوجة أو الثنائية): و هي تلك الجداول التي تتوزع فيها البيانات حسب صفتين أو ظاهرتين في نفس الوقت. و بذلك فإن الجدول الثنائي يتكون من الصفوف و التي تمثل فئات أو مجاميع إحدى الظاهرتين، و الأعمدة التي تمثل فئات أو مجاميع الصفة الأخرى. أما المربعات الناتجة من الصفوف و الأعمدة فتحتوي على عدد المفردات أو التكرارات المشتركة بين الظاهرتين. و المثال التالي يبين وصف الجداول الثنائية

مثال: البيانات التالية تمثل توزيع الحالة الاجتماعية حسب الجنس لـــ 50 شخص

الحالة الاجتماعية / الجنس	أعزب	متزوج	مطلق	أرمل	المجموع
ذكور	5	12	7	4	28
إناث	7	8	1	6	22
المجموع	**12**	**20**	**8**	**10**	**50**

أنواع التوزيعات التكرارية:

- **التوزيع التكراري النسبي (Fr) و النسبي المئوي ((%)Fr):** و هو جدول يبين الأهمية النسبية لكل قيمة أو وجه من أوجه الظاهرة محل الدراسة. و يحسب التكرار النسبي و النسبي المئوي بالصورة التالية:

$$Fr = \frac{\text{التكرار المطلق}}{\text{مجموع التكرارات}} = \frac{Fi}{\sum Fi}$$

$$Fr(\%) = \frac{\text{التكرار المطلق}}{\text{مجموع التكرارات}} \times 100 = \frac{Fi}{\sum Fi} \times 100$$

مثال: البيانات التالية تمثل التقديرات التي حصل عليها عينة من 20 طالب في إحدى المسابقات: ممتاز، مقبول، جيد جدا، مقبول، جيد، مقبول، ضعيف، مقبول، مقبول، جيد جدا، جيد، مقبول، جيد، جيد، مقبول، جيد جدا، ضعيف، ممتاز، ضعيف، جيد

أوجه الظاهرة	Fi	Fr	Fr(%)
ضعيف	3	0.15	15
مقبول	7	0.35	35
جيد	5	0.25	25
جيد جدا	3	0.15	15
ممتاز	2	0.10	10
المجموع	**20**	**1**	**100**

ملاحظة هامة: يجب أن يكون دائما:

$$\sum Fr = 1 \quad \text{و بالتالي} \quad \sum Fr(\%) = 100$$

- **التوزيع التكراري المتجمع:** قد نحتاج في بعض الأحيان إلى معرفة عدد القيم أو المفردات (المشاهدات) التي تقل أو تزيد عن قيمة معينة. و هنا نصبح بصدد الحديث عن التوزيعات التكرارية المتجمعة و هي على نوعين:

✵**التوزيع التكراري المتجمع الصاعد (Fcc):** و هو يبحث في عدد القيم التي تقل عن مستوى معين (أقل من). و يحسب وفقا للشكل التالي:

$$Fcc = \begin{cases} 0 \\ F1 \\ F1 + F2 \\ F1 + F2 + F3 \\ . \\ . \\ . \\ \sum Fi \end{cases}$$

مثال: فيما يلي بيان بعدد الأفراد في عينة من 40 أسرة

6	3	5	4	5	6	5	4	3	2	5	3	3	2	6	8	7	7	4	5
5	6	5	7	4	5	4	6	5	5	5	6	4	7	8	2	3	5	4	4

Fcc	Fi	Xi
0		**2**
	3	
3		**3**
	5	
8		**4**
	8	
16		**5**
	12	
28		**6**
	6	
34		**7**
	4	
38		**8**
	2	
40		
-	40	$\sum$

Fcc	الحدود العليا
0	أقل من 2
3	أقل من 3
8	أقل من 4
16	أقل من 5
28	أقل من 6
34	أقل من 7
38	أقل من 8
40	أقل من 9

❋_**التوزيع التكراري المتجمع النازل (Fcd):** و هو يبحث في عدد القيم التي تزيد عن مستوى معين (أكبر من). و يحسب وفقا للشكل التالي:

$$Fcd = \begin{cases} \sum Fi \\ \sum Fi - F1 \\ \sum Fi - F1 - F2 \\ . \\ . \\ . \\ 0 \end{cases}$$

مثال: البيانات التالية تمثل أعمار عينة من 20 عامل في إحدى الشركات:

47	48	41	41	43	30	32	35	53	38
39	40	50	43	46	52	42	47	52	32

بما أن المعطيات السالفة الذكر تخص أعمار الأشخاص، فإننا نصبح بصدد التعامل مع متغيرة متصلة. و للحصول على مختلف الفئات الممثلة للظاهرة المدروسة، يجب المرور على المراحل التالية:

- ✓ حساب المدى العام (طول السلسلة الإحصائية): يرمز له بالرمز ET و يعطى بالعلاقة:

المدى العام= اكبر قيمة في السلسلة – أصغر قيمة في السلسلة

و يعبر عنه رياضيا بــ:

$$ET = Max(xi) - Min(xi) = 53 - 30 = 23$$

- ✓ افتراض العدد المناسب للفئات: بحيث لا يقل عن 5 و لا يزيد عن 15. و ليكن في مثالنا هذا يساوي 6
- ✓ حساب أطوال الفئات (Δ): و يعطى وفقا للعلاقة الرياضية الآتية:

$$\Delta = \frac{\text{المدى العام}}{\text{عدد الفئات}} = \frac{ET}{\text{عدد الفئات}} = \frac{23}{6} = 3.8 \cong 4$$

و منه يصبح لدينا

الفئات	Fi	Fcd
		20
−30	3	
		17
−34	1	
		16
−38	5	
		11
−42	3	
		8
−46	4	
		4
54−50	4	
		0
∑	20	

الحدود الدنيا للفئات	Fcd
30 فأكثر	20
34 فأكثر	17
38 فأكثر	16
42 فأكثر	11
46 فأكثر	8
50 فأكثر	4
54 فأكثر	0

أصناف الجداول التكرارية:

- **الجداول المقفلة و المفتوحة : مثـــال:**

أقل من 3	6 – 3	أقل من 3	6 – 3
6 – 3	10 – 6	6 – 3	10 – 6
10 – 6	14 – 10	10 – 6	14 – 10
14 – 10	14 فأكثر	14 – 10	
		14 فأكثر	
مفتوح من طرفه الأدنى	**مفتوح من طرفه الأعلى**	**مفتوح من كلا طرفيه**	**جدول مقفل**

ملاحظة: كلما كان الجدول مقفلا، كلما كانت العمليات الحسابية أسهل

- **الجداول المنتظمة و غير المنتظمة**

تعريف: الجدول المنتظم هو الجدول الذي تكون فيه أطوال الفئات متساوية، أما الجدول غير المنتظم فهو الجدول الذي تكون فيه أطوال الفئات غير متساوية. و في هذه الحالة، يلجأ الباحث لحساب التكرار المعدل و الذي يمكن الحصول عليه من العلاقة التالية:

$$\text{التكرار المعدل} = \frac{\text{التكرار الأصلي}}{\text{طول الفئة}}$$

مع ملاحظة أنه إذا كان ناتج القسمة يحتوي على كسور فيمكن التخلص منها بضرب جميع التكرارات المعدلة في مقدار ثابت حتى يمكن تسهيل عمليات الرسم و الحساب

مثال: ليكن لدينا التوزيع التكراري التالي

الفئات	**10–**	**15–**	**20–**	**25–**	**30–**	**40–**	**60–50**
Fi	10	70	106	89	133	79	45

يمكن ملاحظة من هذا الجدول أن الفئات غير متساوية الطول، لذا نقوم بحساب التكرار المعدل على النحو التالي:

الفئات	Fi	طول الفئة	Fi المعدل (المقدار الثابت = 10)
10−	10	5	$20 = 10 \times \frac{10}{5}$
15−	70	5	$140 = 10 \times \frac{70}{5}$
20−	106	5	$212 = 10 \times \frac{106}{5}$
25−	89	5	$178 = 10 \times \frac{89}{5}$
30−	133	10	$133 = 10 \times \frac{133}{10}$
40−	79	10	$79 = 10 \times \frac{79}{10}$
60−50	45	10	$45 = 10 \times \frac{45}{10}$

2-العرض البياني (الهندسي) للبيانات

يعتبر العرض الجدولي للبيانات أفضل طريقة لعرض البيانات الإحصائية و أكثرها استخداما، غير أن طبيعة الأرقام قد لا تشجع بعض الناس على قراءتها أو تفهم مدلولها بسرعة و سهولة. لهذا يلجأ الباحث في بعض الأحيان لعرض البيانات بأسلوب آخر زيادة في الإيضاح و هو أسلوب الأشكال و الرسوم البيانية.

و العرض البياني شائع الاستخدام و ذلك لسهولة تفهمه حتى من جانب غير المتخصصين و لوضوحه و سرعة إظهاره للتغير في مختلف الظواهر، كما انه يستخدم لإجراء المقارنات بالرسم.

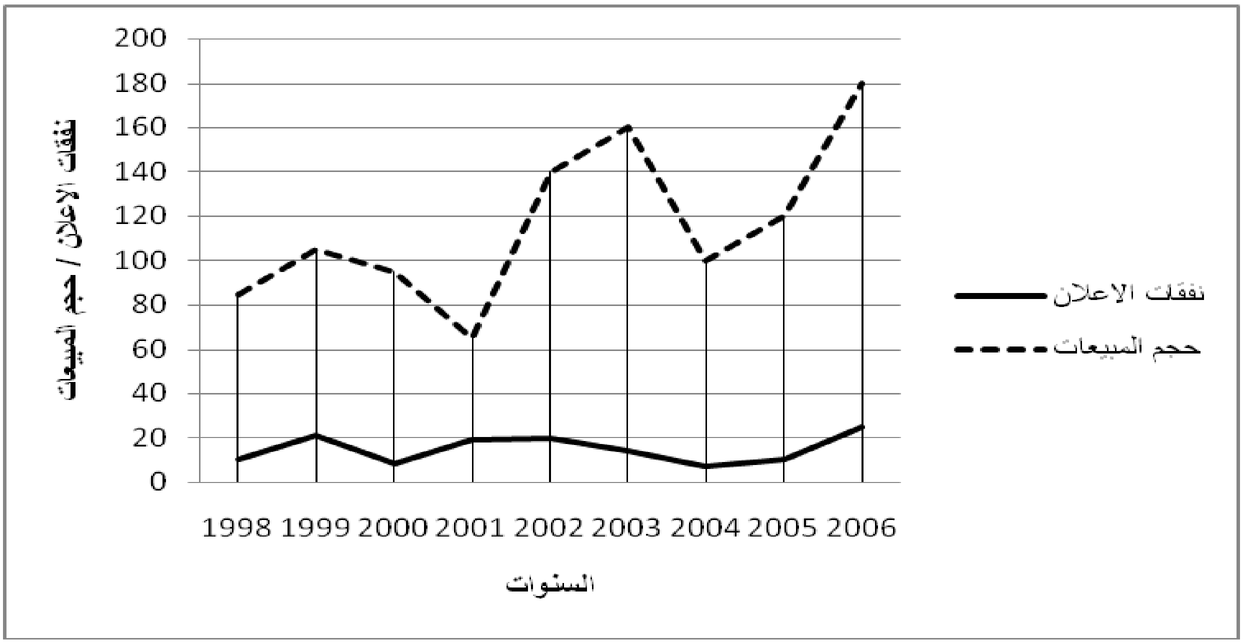
تختلف الرسوم البيانية التي يمكن استخدامها في العرض الهندسي باختلاف طبيعة و نوع البيانات المراد عرضها. و يمكن تقسيم طرق العرض البياني إلى قسمين رئيسيين: عرض البيانات الوصفية (الكيفية) و عرض البيانات الكمية (المنفصلة و المتصلة)

1- العرض الهندسي للبيانات الوصفية: توجد لذلك عدة طرق منها الآتي:

أ- **الخط البياني:** تستخدم هذه الطريقة لتوضيح تطور ظاهرة (أو أكثر) خلال فترة زمنية محدودة (أي لعرض بيانات سلسلة زمنية)، حيث يمثل المحور الأفقي الزمن و المحور الرأسي قيم الظاهرة (أو الظواهر) موضع الدراسة.

مثال 01: الجدول التالي يبين حجم المنفق على الإعلان و حجم المبيعات لشركة ما مختصة في الأجهزة الكهربائية خلال الفترة من 1998 إلى 2006

السنة	**98**	**99**	**00**	**01**	**02**	**03**	**04**	**05**	**2006**
المنفق على الإعلان بآلاف الدولارات	10	21	8	19	20	14	7	10	25
حجم المبيعات بآلاف الدولارات	85	105	95	65	140	160	100	120	180



ب- **الأعمدة (أو المستطيلات):** و هناك أكثر من نوع من الأعمدة:

- **الأعمدة البسيطة:** و فيها يتم تمثيل ظاهرة واحدة بحيث يتم تخصيص عمود رأسي لكل وجه من أوجه الظاهرة

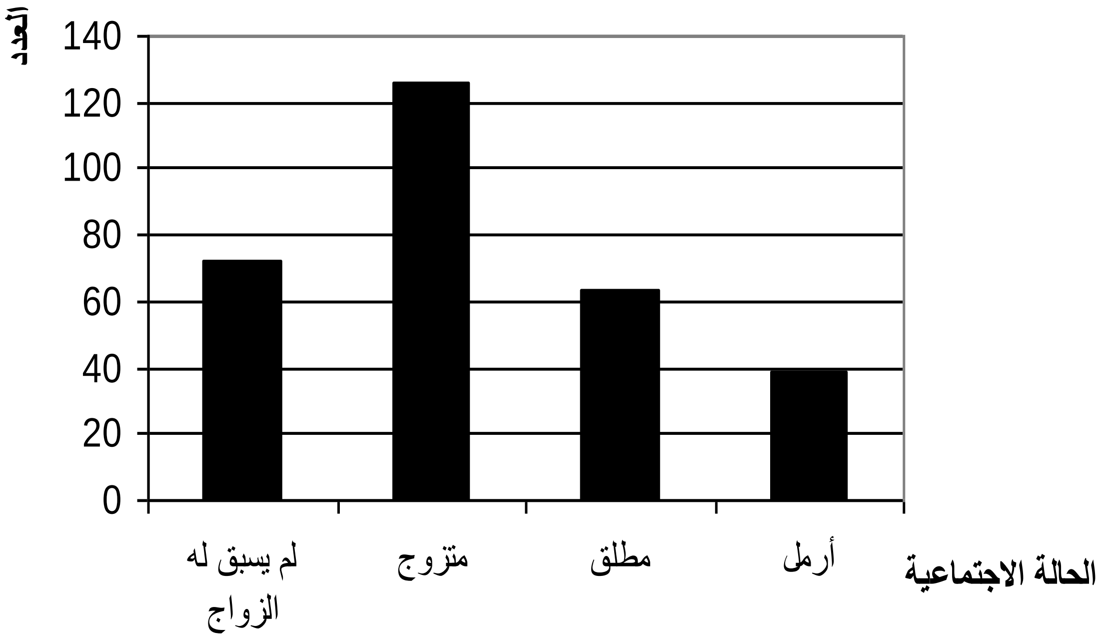
• **الأعمدة المجزأة:** يتم تخصيص عمود لكل وجه من أوجه الظاهرة مع إبراز مكوناتها بتجزئة العمود الممثل للظاهرة، بحيث يكون كل جزء من العمود متناسبا مع قيم الظاهرة المناظرة له.

• **الأعمدة المتلاصقة:** يتم تخصيص عمود أو أكثر تكون متلاصقة للتعبير عن أوجه الظاهرة، بحيث تكون الأعمدة الممثلة للظاهرة متجاورة و يكون طول كل من العمودين المتلاصقين متناسبا مع العدد الذي يمثله.

مثال 02: البيانات التالية تمثل الحالة الاجتماعية لعينة مكونة من 300 شخص بإحدى البلديات:

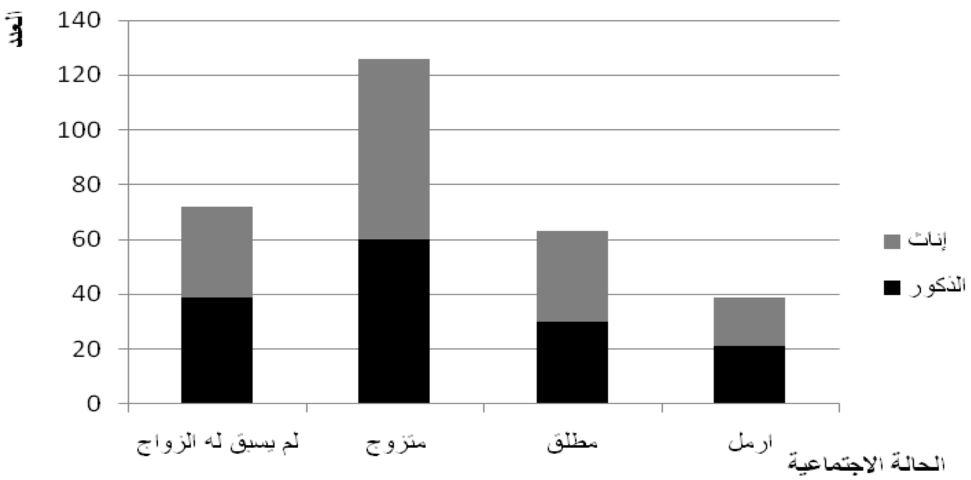
الحالة الاجتماعية / الجنس	لم يسبق له الزواج	متزوج	مطلق	أرمل	المجموع
ذكور	39	60	30	21	150
إناث	33	66	33	18	150
المجموع	**72**	**126**	**63**	**39**	**300**

1- الأعمدة البسيطة

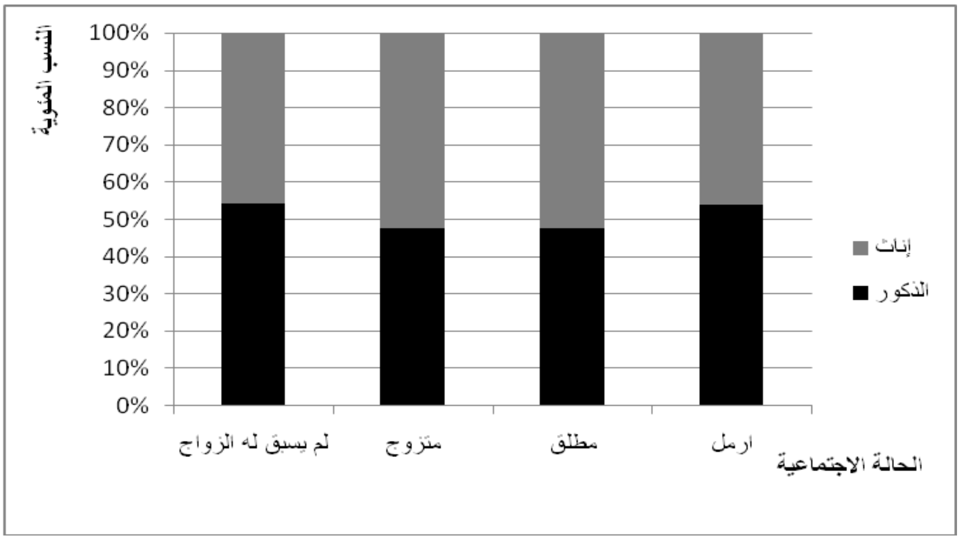


2- الأعمدة المجزأة

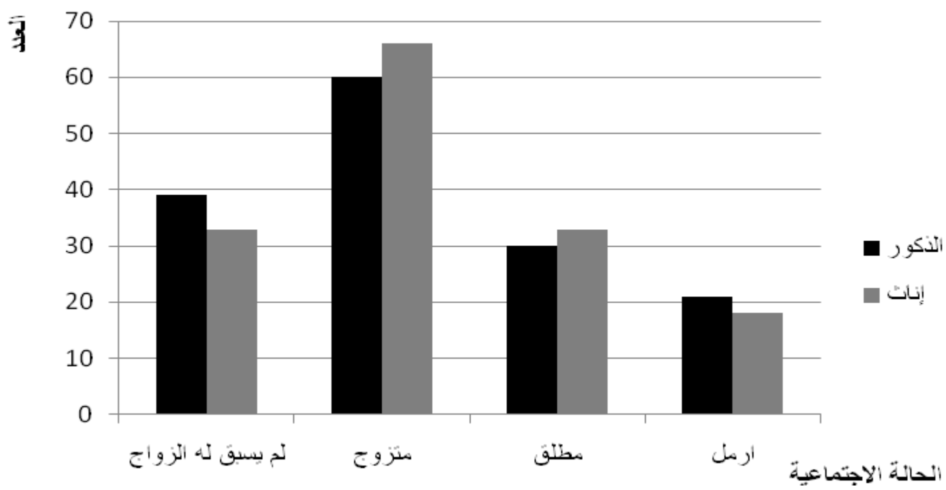
- **باستخدام القيم المطلقة:**



- **باستخدام النسب المئوية:**

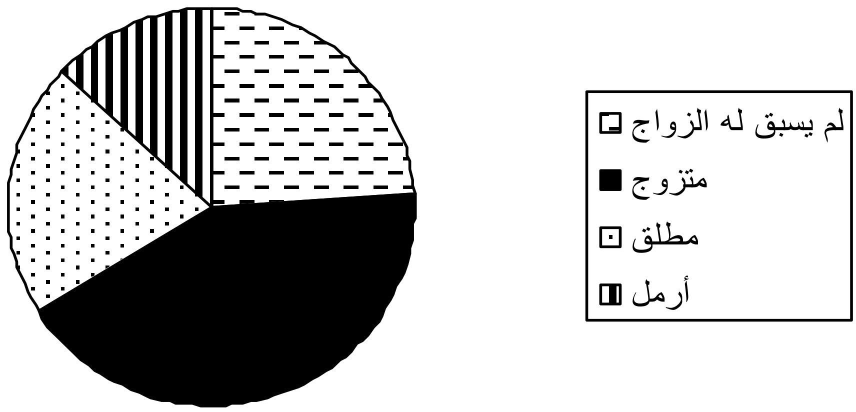


3- الأعمدة المتلاصقة



ج- عرض البيانات بشكل دائري: يتم عرض البيانات من خلال رسم مجموعة من القطاعات الدائرية، بحيث يكتمل شكل الدائرة على أن يكون كل قطاع في الدائرة متناسبا مع عدد المفردات في الوجه الذي يمثله علما أن:

$$\text{زاوية القطاع} = \frac{\text{عدد مفردات الوجه}}{\text{إجمالي المفردات}} \times 360$$



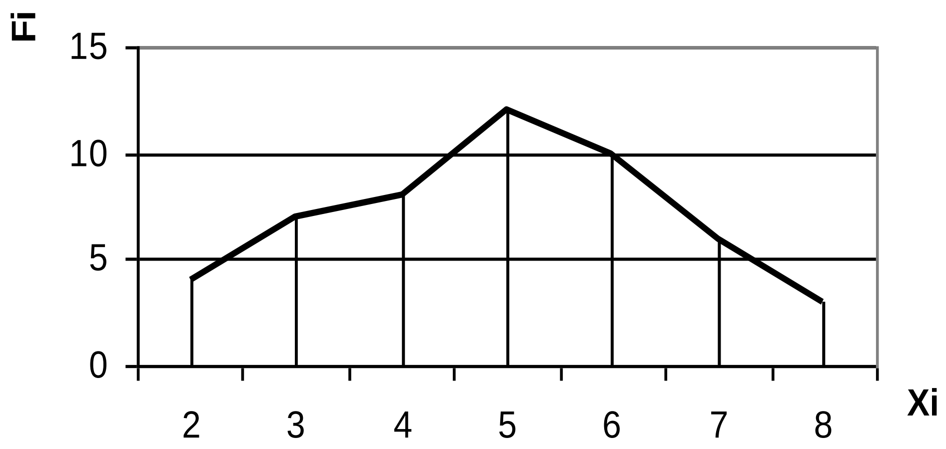
2- العرض الهندسي للبيانات الكمية: و هنا يمكن التمييز بين

أ-حالة المتغير المنفصل: يمكن في هذه الحالة تمثيلا بيانيا مايلي:

- **التكرار المطلق: (البيان بالأعمدة و المضلع التكراري)**

مثال: الجدول التالي عدد العاملين في 50 محلا تجاريا

عدد العاملين	2	3	4	5	6	7	8	المجموع
التكرارات Fi	4	7	8	12	10	6	3	50



- **التكرار المتجمع الصاعد: (المنحنى الدرجي)**

مثال: بناءا على معطيات المثال السابق يمكن حساب قيم Fcc كمايلي:

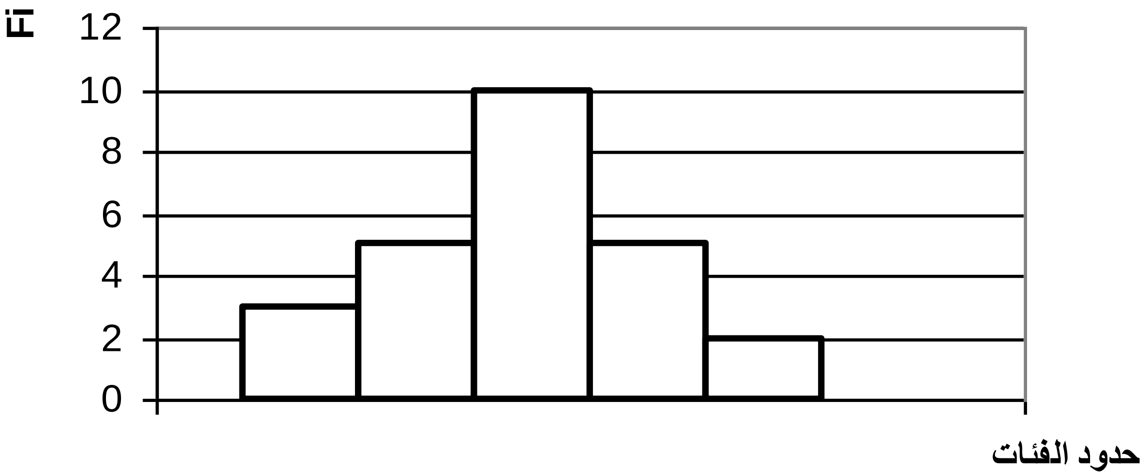
Fcc	Fi	Xi
0		
	4	2
4		
	7	3
11		
	8	4
19		
	12	5
31		
	10	6
41		
	6	7
47		
	3	8
50		
-	50	∑

ب- حالة المتغير المتصل: و هنا كذلك يمكن تمثيل مايلي:

- **التكرار المطلق: (المدرج التكراري)**

مثال: ليكن لدينا التوزيع التكراري التالي

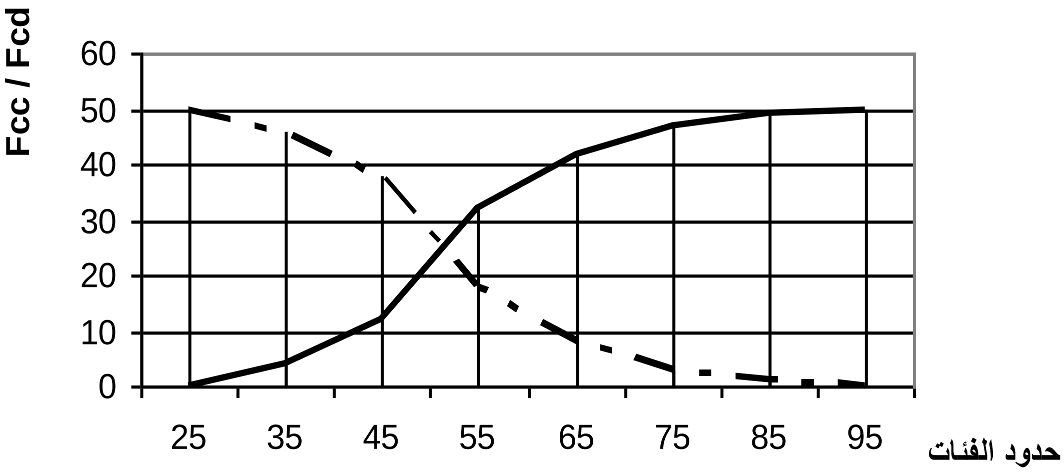
المجموع	**14-12**	**12-10**	**10-8**	**8-6**	**6-4**	**الفئات**
25	2	5	10	5	3	Fi



- **التكرار المتجمع الصاعد (المنحنى التكاملي) و المتجمع النازل (المنحنى التفاضلي)**

مثال:

Fcd	Fcc	Fi	الفئات
50	0		
		4	25−
46	4		
		8	35−
38	12		
		20	45−
18	32		
		10	55−
8	42		
		5	65−
3	47		
		2	75−
1	49		
		1	95−85
0	50		


Fcc / Fcd
60
50
40
30
20
10
0
25
35
45
55
65
75
85
95
حدود الفئات

الفصل الثاني

2

مقاييس النزعة المركزية

مقدمة:

إن معظم القيم لمختلف الظواهر الطبيعية تتركز عادة في الوسط أو قريبة منه، و مقاييس التمركز أو التوسط لأي مجموعة من البيانات التابعة للظاهرة محل الدراسة هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية البيانات بحيث تمثلها أفضل تمثيل و هناك ثلاث مقاييس إحصائية مهمة و شائعة الاستخدام، و يمكن أن تستخدم لتمثيل البيانات الإحصائية و حسب نوعية البيانات. فيمكن أن يكون الوسط الحسابي الذي يعد من أهم المقاييس الإحصائية لكونه يستخدم جميع البيانات الإحصائية، أما المقياس الثاني فهو الوسيط و يعتبر من المقاييس المهمة و المستخدم بشكل واسع جدا و خاصة عندما يكون قسم من البيانات كبيرة جدا أو صغيرة جدا، أو ما تسمى في الإحصاء بالقيم الشاذة، أما المقياس الثالث فهو المنوال و الذي يعد اقل مقاييس النزعة المركزية استعمالا.كما و سيتم في هذا الفصل التعرف على أنواع أخرى من المتوسطات أهمها: الوسط الهندسي، الوسط التوافقي و الوسط التربيعي.

هذا و سنتناول كل مقياس على حدى بنوع من التفصيل من حيث الخصائص و طرق إيجاده.

الوسط الحسابي(La moyenne arithmétique)

تعريف: الوسط الحسابي لمجموعة مشاهدات هو مجموع هذه المشاهدات مقسوما على عددها و يمكن كتابة هذه العلاقة الرياضية بالشكل:

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

كيفية إيجاد الوسط الحسابي:

أ- **إيجاد الوسط الحسابي للبيانات غير المبوبة:** و هنا يمكن التمييز بين

- **البيانات مفردة " غير متكررة " (الوسط الحسابي البسيط)**

تعريف: إذا كان لدينا قيم المشاهدات $X_1, X_2, X_3, \dots, X_n$ ، فإن الوسط الحسابي لهذه المشاهدات $\bar{X}_S$ هو:

$$\bar{X}_S = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

مثال: إذا كان لدينا قيم المشاهدات التالية: 3، 5، 7، 11، 13، 21. و المطلوب إيجاد الوسط الحسابي لهذه البيانات.

الحل: باستخدام العلاقة أعلاه فإن: $\bar{X}_S = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3+5+7+11+13+21}{6} = 10$

☜ **البيانات متكررة (الوسط الحسابي الموزون أو المرجح)**

تعريف: إذا كان لدينا قيم المشاهدات $X_1, X_2, X_3, \ldots, X_n$ و تكراراتها المقابلة على التوالي $F_1, F_2, F_3, \ldots, F_n$ فإن الوسط الحسابي يكون:

$$\bar{X}_P = \frac{F_1X_1 + F_2X_2 + \cdots + F_nX_n}{F_1 + F_2 + \cdots + F_n} = \frac{\sum_{i=1}^{n} F_iX_i}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_iX_i$$

مثال: إذا كان لدينا توزيع مئة طالب حسب العلامات المتحصل عليها في امتحان الإحصاء على النحو التالي.أوجد قسمة الوسط الحسابي لهذه المشاهدات

4	5	6	7	8	9	10	العلامة (X_i)
2	8	13	35	21	16	5	عدد الطلاب (F_i)

الحل:

$\sum FrXi$	$\sum FiXi$	4	5	6	7	8	9	10	X_i
-	733	2	8	13	35	21	16	5	F_i
7.33	-	0.02	0.08	0.13	0.35	0.21	0.16	0.05	Fr

على هذا الأساس، يصبح لدينا

$$\bar{X}_P = \frac{\sum_{i=1}^{7} F_iX_i}{\sum_{i=1}^{7} F_i} = \frac{733}{100} = 7.33 = \sum_{i=1}^{7} Fr_iX_i$$

ب- إيجاد الوسط الحسابي للبيانات المبوبة:

هناك عدة طرق لإيجاد الوسط الحسابي و سوف نستعرض في هذا الصدد أهم الطرق المستخدمة.

طريقة استخدام التكرارات أو طريقة القانون العام:

في هذه الطريقة تكون صيغة الوسط الحسابي بالشكل التالي.

$$\overline{X}_P = \frac{F_1 C_1 + F_2 C_2 + \cdots + F_n C_n}{F_1 + F_2 + \cdots + F_n} = \frac{\sum_{i=1}^{n} F_i C_i}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_i C_i$$

مثال: اوجد الوسط الحسابي لقيم المشاهدات المبوبة بالجدول بالطريقة المباشرة

الفئات	]25-20]	]30-25]	]35-30]	]40-35]	[45-40]	المجموع
التكرار	7	13	21	6	3	50

الحل: نشكل الجدول التالي و الذي يحتوي على جميع الحسابات المطلوبة لهذه الطريقة

الفئات	التكرار F_i	التكرار النسبي Fr	مراكز الفئات C_i	F_iC_i	FrC_i
25-20	7	0.14	22.5	157.5	3.15
30-25	13	0.26	27.5	357.5	7.15
35-30	21	0.42	32.5	682.5	13.65
40-35	6	0.12	37.5	225	4.5
45-40	3	0.06	42.5	127.5	2.55
المجموع	50	1	-	1550	31

$$\overline{X}_P = \frac{\sum_{i=1}^{5} F_i C_i}{\sum_{i=1}^{5} F_i} = \frac{1550}{50} = 31 = \sum_{i=1}^{5} Fr_i C_i$$

- **طريقة استخدام الوسط الفرضي:** في هذه الطريقة تكون صيغة الوسط الحسابي بالشكل التالي

$$\overline{X}_P = \propto + \frac{\sum_{i=1}^{n} F_i W_i}{\sum_{i=1}^{n} F_i}$$

حيث: α هو الوسط الفرضي، و غالبا ما يمثل مركز الفئة المقابلة للأكثر تكرارا و

W_i يمثل انحراف مراكز الفئات عن الوسط الفرضي: $W_i = C_i - \alpha$

مثال: إذا كان لدينا البيانات التالية و المبوبة بالجدول التالي:

الفئات	-30	-40	-50	-60	-70	المجموع
التكرار	2	9	21	11	7	50

المطلوب: إيجاد الوسط الحسابي بطريقة الوسط الفرضي

الحل: نكون الجدول التالي و المتضمن خطوات الحساب الأساسية

الفئات	F_i	C_i	$W_i=C_i-\alpha$	F_iW_i
-30	2	35	-20=55-35	-40
-40	9	45	-10=55-45	-90
-50	**21**	**55**	0=55-55	0
-60	11	65	10=55-65	110
-70	7	75	20=55-75	140
المجموع	50	-	-	120

$$\bar{X}_P = \alpha + \frac{\sum_{i=1}^{n} F_i W_i}{\sum_{i=1}^{n} F_i} = 55 + \frac{120}{50} = 57.4$$

طريقة استخدام الانحرافات المختصرة : في هذه الطريقة تكون صيغة الوسط الحسابي بالشكل التالي

$$\bar{X}_P = \alpha + \frac{\sum_{i=1}^{n} F_i \acute{W}_i}{\sum_{i=1}^{n} F_i} \times K$$

حيث: α هو الوسط الفرضي، W_i يمثل انحراف مراكز الفئات عن الوسط الفرضي، K هو طول الفئة و $\acute{W}_i = \frac{W_i}{L_i}$

مثال: البيانات التالية تمثل أوزان 50 طالبا موزعين في الجدول التالي

الفئات	55-50	60-55	65-60	70-65	75-70	المجموع
الطلاب	7	13	25	3	2	50

المطلوب: إيجاد الوسط الحسابي بطريقة الانحرافات المختصرة

الحل: نكون الجدول الشامل لكافة مراحل الحساب

الفئات	F_i	C_i	W_i	$\acute{W}_i$	$F_i\acute{W}_i$
55-50	7	52.5	-10=62.5-52.5	$\frac{-10}{5}=-2$	-14
60-55	13	57.5	-5=62.5-57.5	$\frac{-5}{5}=-1$	-13
65-60	<u>**25**</u>	<u>**62.5**</u>	0=62.5-62.5	$\frac{0}{5}=0$	0
70-65	3	67.5	5=62.5-67.5	$\frac{5}{5}=1$	3
75-70	2	72.5	10=62.5-72.5	$\frac{10}{5}=2$	4
المجموع	50	-	-	-	-20

$$\bar{X}_P = \propto + \frac{\sum_{i=1}^{n} F_i\acute{W}_i}{\sum_{i=1}^{n} F_i} \times K = 62.5 - \frac{20}{50} \times 5 = 60.5$$

الوسيط (La médiane)

تعريف: الوسيط هو عبارة عن القيمة الأوسطية لمجموعة من القيم <u>رُتبت تصاعديا أو تنازليا</u> في حالة إذا كان عدد القيم فرديا و متوسط القيمتين الأوسطيتين إذا كان عدد القيم زوجيا

التمثيل إذا كان عدد القيم فرديا و الترتيب تصاعديا:

$\underbrace{X\ X\ X}$ المجموعة السابقة و الأدنى — $\underbrace{X}$ الوسيط — $\underbrace{X\ X\ X}$ المجموعة اللاحقة و الأعلى

التمثيل إذا كان عدد القيم زوجيا و الترتيب تصاعديا:

$\underbrace{X\ X\ X}$ المجموعة السابقة و الأدنى — $\underbrace{X\ X}$ القيمتان الأوسطتان — $\underbrace{X\ X\ X}$ المجموعة اللاحقة و الأعلى

الوسيط=متوسط القيمتان

كيفية إيجاد الوسيط:

أ- **إيجاد الوسيط للبيانات غير المبوبة:** يوجد حالتان لحساب الوسيط من هذه البيانات

إذا كان عدد القيم غير المبوبة فرديا

إذا كان لدينا قيم المشاهدات $X_1, X_2, X_3, \ldots, X_n$ ، و كان **n** فرديا. لحساب الوسيط نتبع الخطوات التالية:

- ترتب البيانات ترتيبا تصاعديا أو تنازليا و لكن سنتناول هنا فقط حالة الترتيب التصاعدي
- نجد ترتيب الوسيط من العلاقة:

$$\text{ترتيب الوسيط} = \frac{n+1}{2} \quad \text{حيث } n \text{ عدد القيم}.$$

- نجد قيمة الوسيط و هي القيمة المناظرة لترتيب الوسيط

مثال: إذا كان لدينا قيم المشاهدات التالية 3، 21، 5، 9، 11، 7، 14. اوجد الوسيط لهذه القيم.

الحل: نتبع الخطوات أعلاه

القيمة	3	5	7	9	11	14	21
الترتيب	1	2	3	4	5	6	7

$$\text{ترتيب الوسيط} = \frac{7+1}{2} = 4 \text{ أي الترتيب الرابع}$$

نجد قيمة الوسيط (Me) و هي القيمة التي تناظر الترتيب الرابع و المشار لها بالسهم، فتكون قيمة الوسيط عندئذ **Me=9**

إذا كان عدد القيم غير المبوبة زوجيا

لإيجاد الوسيط لهذه القيم نتبع الخطوات التالية:

- نرتب قيم المشاهدات ترتيبا تصاعديا
- نجد ترتيب الوسيطين من العلاقة التالية:

$$\text{ترتيب } Me_1 (\text{الوسيط الأول}) = \frac{n}{2}$$

$$\text{ترتيب } Me_2 (\text{الوسيط الثاني}) = \frac{n}{2} + 1 \text{ أو } \frac{2+n}{2}$$

- نجد قيم Me_1 و Me_2 المناظرة لترتيبهما

- نجد قيمة الوسيط Me من العلاقة $Me = \frac{Me_{1+Me_2}}{2}$

مثال: اوجد الوسيط لقيم المشاهدات 3، 7، 25، 15، 29، 11، 18، 20

الحل: نتبع الخطوات التالية

القيمة	3	7	11	15	18	20	25	29
الترتيب	1	2	3	4	5	6	7	8

ترتيب الوسيط الأول $= \frac{8}{2} = 4$ **أي الترتيب الرابع**

ترتيب الوسيط الثاني $= \frac{8}{2} + 1 = 5$ **أي الترتيب الخامس**

نجد القيم المناظرة لترتيبهما كما هو مشار بالأسهم فتكون قيمة **$Me_1=15$** و قيمة **$Me_2=18$** و منه قيمة الوسيط هي $Me = \frac{Me_{1+Me_2}}{2} = 16.5$

ب- إيجاد الوسيط للبيانات المبوبة:

لإيجاد الوسيط للبيانات المبوبة نتبع الخطوات التالية:

- نضيف للجدول المعطى فئة سابقة تكرارها صفرا
- نجد عمود للفئات الفعلية العلوية
- نجد عمود التكرار المتجمع الصاعد
- نجد ترتيب الوسيط من العلاقة:

$$\text{ترتيب الوسيط} = \frac{\text{مجموع التكرارات}}{2} = \frac{\sum F_i}{2}$$

- نحدد موقع ترتيب الوسيط بين التكرارات المتجمعة الصاعدة و نشير له بسهم
- نجد الفئة الوسيطية بحديها الفعليين الأدنى و الأعلى و هي الفئة التي تقع تحت السهم الذي يشير لترتيب الوسيط
- نحدد الحد الأدنى للفئة الوسيطية
- نحدد تكرارا المتجمع السابق و اللاحق لترتيب الوسيط
- نحدد طول الفئة الوسيطية

- نجد الوسيط من العلاقة:

$$Me = L_0 + \frac{\frac{\sum F_i}{2} - F_1}{F_2 - F_1} \times K$$

مثال: البيانات التالية تمثل الأجور الأسبوعية لمائة عامل موزعين على النحو التالي

فئات الأجور	70-60	80-70	90-80	100-90	110-100	المجموع
عدد العمال	6	12	47	25	10	100

المطلوب: أوجد ما يلي: عدد العمال الذين رواتبهم اقل من 60 دولار، عدد العمال الذين رواتبهم بين 60 و اقل من 100 دولار، عدد العمال الذين رواتبهم 80 دولار فأكثر و قيمة الوسيط لهذه الأجور

الحل:

الفئات	F_i	Fcc
أقل من 60	0	0
70-60	6	6
80-70	12	
90-80	47	
100-90	25	90
110-100	10	100
المجموع	100	-

التكرار السابق لترتيب الوسيط

التكرار اللاحق لترتيب الوسيط

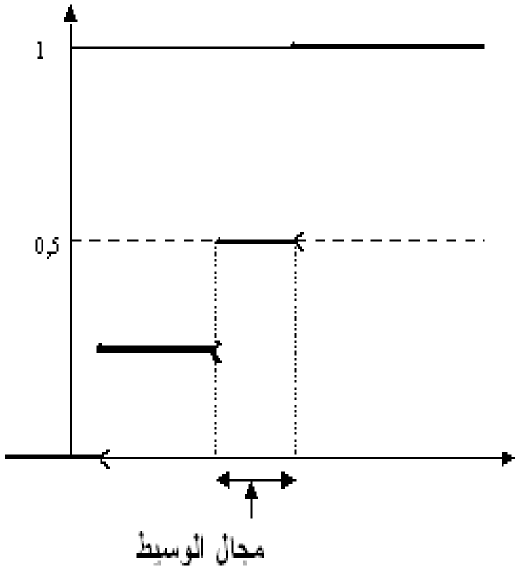
- عدد العمال الذين رواتبهم اقل من 60 دولار= صفر
- عدد العمال الذين رواتبهم بين 60 و اقل من 100 دولار=6+12+47+25=90 عاملا
- عدد العمال الذين رواتبهم 80+ دولار فأكثر=47+25+10=82 عاملا
- قيمة الوسيط لهذه الأجور

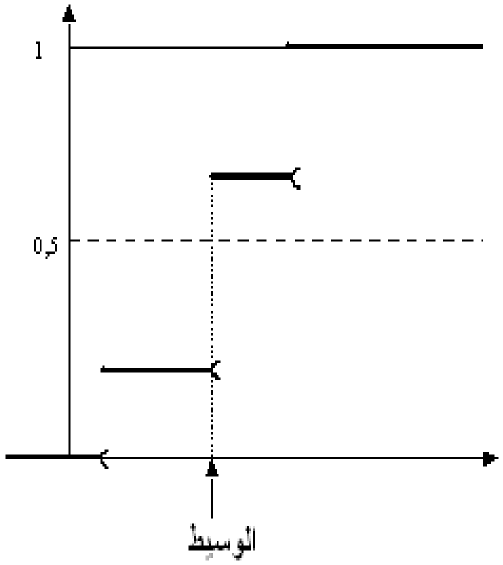
$$Me = L_0 + \frac{\frac{\sum F_i}{2} - F_1}{F_2 - F_1} \times K = 80 + \frac{50 - 18}{65 - 18} \times 10 = 86.8$$

نلاحظ أن قيمة الوسيط وقعت ضمن الفئة الوسيطية [80-90] الواقعة تحت السهم

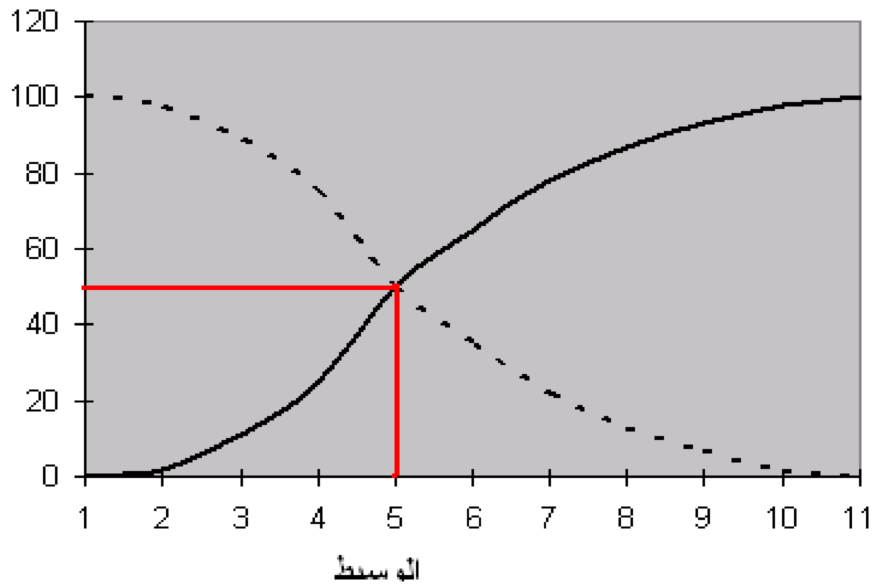
ملاحظة: يمكن تقدير قيمة الوسيط بيانيا وفقا لإحدى الحالتين التاليتين:

1- حالة المتغير المنفصل: و هنا يمكن تحديد قيمة الوسيط بيانيا من خلال الحالتين:





2- حالة المتغير المتصل: في حالة المتغير المتصل، نرسم المنحنى التكاملي و المنحنى التفاضلي و يتم بعدها إيجاد القيمة الفردية ذات الترتيب $X_0=\frac{\sum F_i}{2}=50\%$ من محور التراتيب و إقامة عمود موازي لمحور الفواصل بحيث يلاقي المنحنيين في نقطة تقاطعهما ومنها يسقط عمود على محور الفواصل فيقطعه على بعد يساوي تماما قيمة الوسيط كما هو مبين في الشكل التالي:



المنوال (Le mode)

تعريف: المنوال هو القيمة الأكثر تكرارا أو شيوعا بين قيم المشاهدات.

كيفية إيجاد المنوال:

أ- إيجاد المنوال للبيانات غير المبوبة

- **إذا لم تتكرر أيا من القيم فلا وجود للمنوال**

مثال: لدينا قيم المشاهدات 7، 9، 11، 12، 15.اوجد منوال هذه القيم

الحل: لا يوجد منوال لهذه القيم حيث أ ن أيا منها لم تتكرر

- **إذا تكرر احدها فيكون هناك منوالا واحدا**

مثال: اوجد المنوال لقيم المشاهدات التالية 7، 5،11، 7، 11، 7، 9

الحل: القيمة الأكثر تكرارا هي القيمة 7

- **إذا كان لقيمتين نفس العدد من التكرار، فيكون للقيم منوالان و هكذا تزداد المنوالات بزيادة الأعداد المتساوية للتكرارات**

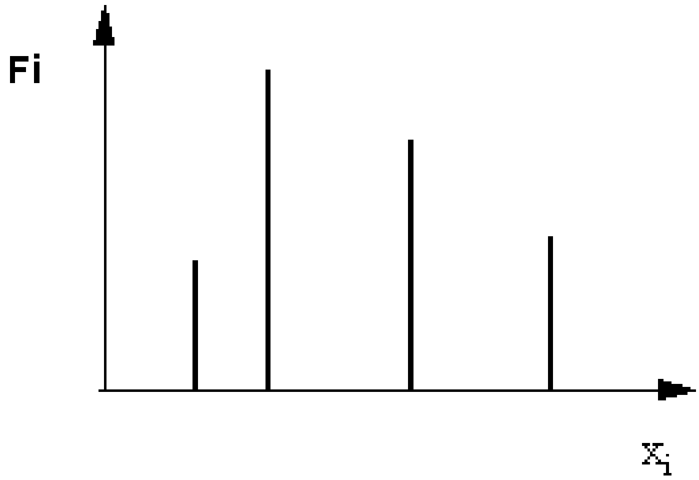
مثال: اوجد المنوال أو المنوالان لقيم المشاهدات التالية 4، 9، 17، 9، 4، 11

الحل: يوجد منوالان هما 4 و 9 لأن لهما نفس التكرار

ملاحظة:

1- إذا كانت قيم السلسلة جد متقاربة ببعضها البعض، ففعالية المنوال و جديته تصبح ضعيفة

2- نستطيع تحديد قيمة المنوال على الشكل البياني: فهي القيمة التي تناسب اكبر عمود.



على عكس الصفات المتقطعة أين يتم تحديد قيمة المنوال بدقة، فانه عندما تتجمع المعطيات في فئات و تفقد هويتها فالتعريف السابق لا ينطبق و نصبح بصدد التحدث عن الفئات المنوالية أو الفئات المسيطرة أي التي تقابل اكبر تكرار . و يمكن في هذه الحالة تقدير قيمة المنوال باستخدام

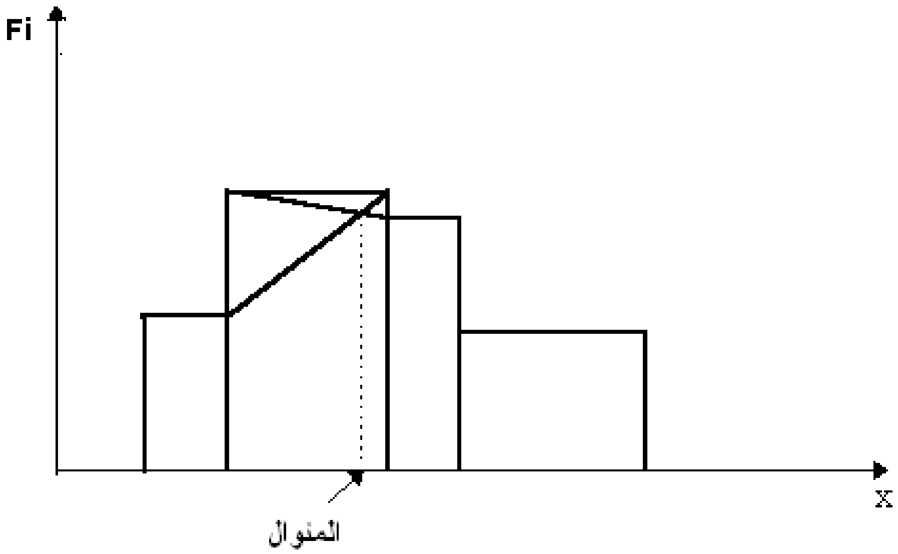
الطريقة التالية و المتمثلة في طريقة الفروق المعتمدة من طرف كارل بيرسون K.Pearson و المعبر عنها رياضيا بالصيغة الاتية:

$$Mo = L_0 + \frac{F_0 - F_1}{(F_0 - F_1) + (F_0 - F_2)} \times K$$

حيث: L_0 هو الحد الأدنى للفئة المنوالية K : طول الفئة المنوالية F_0 : تكرار الفئة المنوالية

F_1: تكرار الفئة السابقة للفئة المنوالية F_2: تكرار الفئة اللاحقة للفئة المنوالية

ملاحظة: هناك طريقة اخرى لتحديد قيمة المنوال و هي الطريقة البيانية حيث بعد رسم المدرج التكراري (على ورقة مليمترية) الممثل للسلسلة المدروسة أين تكون الفئة المنوالية و الفئتين السابقة و اللاحقة لها. نرسم خطين مستقيمين كما هو مبين في الشكل المرافق. و نقطة تقاطع هذين المستقيمين، فاصلتها هي قيمة المنوال بيانيا:



مثال: لتكن السلسلة الإحصائية فيما يخص أطوال 20 طالب بالمتر فكانت النتائج المدونة في الجدول الآتي:

الفئات	1.45 –	1.50 –	1.55 –	1.60 –	1.65 –	1.70 –	1.75 – 1.80
Fi	1	4	7	3	2	1	2

نلاحظ أن الفئة المنوالية هي [1.55 – 1.60[لان عدد الطلبة هو 7 (التكرار المطلق)

حساب المنوال بطريقة الفروق

$$Mo = L_0 + \frac{F_0 - F_1}{(F_0 - F_1) + (F_0 - F_2)} \times K = 1.55 + \frac{(7-4)}{(7-4) + (7-3)} \times 0.05 = 1.57$$

ملاحظة هامة: في حالة التوزيع غير المنتظم تعدل التكرارات قبل تقدير المنوال و قد لا تكون الفئة المنوالية نفسها قبل و بعد تعديل التكرارات.

مثال: لتكن السلسلة الإحصائية المدونة في الجدول التكراري كمايلي:

الفئات	] 10-0]	] 20-10]	] 40-20]	] 70-40]	] 80-70]	المجموع
Fi	50	40	120	150	40	**400**

بما أن الفئات غير متساوية، فيجب تصحيح التكرارات لمعرفة قيمة المنوال. إذ يعتبر خطأ إذا قلنا أن الفئة المنوالية هي الفئة] 70-40] و لكن الفئة المنوالية هي] 40-20] و المتحصل عليها بعد تصحيح التكرارات.

الفئات	-0	-10	-20	-40	80-70	المجموع
Fi	50	40	120	150	40	400
طول الفئة	10	10	20	30	10	-
Fi المعدل	5	4	6	5	4	-

و منه قيمة المنوال (باستعمال طريقة الفروق) تساوي:

$$Mo = 20 + \frac{(6-4)}{(6-4)+(6-5)} \times 10 = 26.7$$

المتوسط الهندسي: La moyenne géométrique

يشيع استخدام هذا النوع من المتوسطات في حساب الأرقام القياسية للأسعار، معدلات التغيير عند تقدير السكان بين سنتي التعداد و حساب متوسط النسب أو المعدلات.

أولا: المتوسط الهندسي البسيط La moyenne géométrique simple

يعرف على انه الجذر النوني لجداء القيم المأخوذة من طرف المتغيرة الإحصائية X و يعبر عنه كالتالي:

$$G_S = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} = \left[\prod_{i=1}^{n} x_i \right]^{\frac{1}{n}}$$

و لتبسيط العمليات الحسابية، يرفع الجذر و يؤخذ اللوغاريتم العشري للطرفين:

$$LogG_S = \frac{1}{n}\log(x_1 \times x_2 \times ... \times x_n) = \frac{1}{n}\sum_{i=1}^{n}\log x_i$$

مثال: حساب المتوسط الهندسي البسيط للسلسلة التالية:2،5،8،17،13،22،33

$$G_S = \sqrt[7]{2\times5\times8\times17\times13\times22\times33} = 10.36$$

و باستعمال طريقة اللوغاريتمات:

$$LogG_S = \frac{1}{7}(\log2+\log5+...+\log33) = 1.0155 \Rightarrow G_S = 10.36$$

ثانيا: المتوسط الهندسي المرجح La moyenne géométrique pondérée

أ/حالة المتغيرة المنفصلة: إذا كانت قيم المتغيرة الإحصائية تتكرر فالمتوسط الهندسي يعبر عنه بالعلاقة التالية:

$$G_P = \sqrt[\sum F_i]{x_1^{F_1} \times x_2^{F_2} \times ... \times x_n^{F_n}}$$

إذا استعملنا اللوغاريتم العشري، فالمتوسط الهندسي يصبح كالتالي:

$$LogG_P = \frac{\sum_{i=1}^{n} F_i.\log x_i}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} \mathrm{Fr}.\log x_i$$

مثال: حساب المتوسط الهندسي للقيم المسجلة في الجدول التالي:

القيم	5	6	7	9	11	13	المجموع
F_i	2	3	5	1	4	5	20
Fr	0.10	0.15	0.25	0.05	0.20	0.25	1

$$G_P = \sqrt{5^2 \times 6^3 \times ... \times 13^5} = 8.56$$

و باستخدام طريقة اللوغاريتمات:

$$LogG_P = \sum_{i=1}^{n} Fr \log x_i = 0.1(\log 5) + 0.15(\log 6) + ... + 0.25(\log 13) = 0.9324 \Rightarrow G_P = 8.56$$

ب/حالة المتغيرة المتصلة: نفس المعادلة بالنسبة لـ G_P المذكور أعلاه مع تعويض X_i بـ C_i و هي مراكز الفئات.

مثال: لو أخذنا المثال المذكور آنفا، و أردنا حساب المتوسط الهندسي المرجح فسوف نجد مايلي:

الفئات	4.5 –	5.5 –	6.5 –	7.5 –	8.5 –	9.5 – 10.5	المجموع
F_i	3	4	2	3	5	3	20
Fr	0.15	0.2	0.1	0.15	0.25	0.15	1
Ci	5	6	7	8	9	10	-

$$G_P = \left[\prod_{i=1}^{n} C_i\right]^{Fr} = 7.39$$

و باستعمال طريقة اللوغاريتمات:

$$LogG_P = \sum_{i=1}^{n} f_i \log C_i = 0.8690 \Rightarrow G_P = 7.39$$

5/ المتوسط التوافقي: La moyenne harmonique

يفضل استخدامه عند حساب معدل التغير، السرعة و متوسط الأسعار متى أعطيت على أساس عدد الوحدات.

أولا: المتوسط التوافقي البسيط La moyenne harmonique simple

يعرف على انه مقلوب المتوسط الحسابي لمقلوبات قيم السلسلة و يعرف رياضيا بالقانون:

$$H_S = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

هذه العلاقة تبين بان مقلوب المتوسط التوافقي لمجموعة من القيم هو المتوسط الحسابي لمقلوبات هذه القيم.

مثال: احسب المتوسط التوافقي للقيم التالية: 2،4،5،8،9

$$H_S = \frac{5}{\frac{1}{2}+\frac{1}{4}+\frac{1}{5}+\frac{1}{8}+\frac{1}{9}} = 4.21$$

ثانيا: المتوسط التوافقي المرجح La moyenne harmonique pondérée

أ/ حالة المتغيرة المنفصلة: يعبر رياضيا عن هذا المتوسط بالعلاقة التالية:

$$H_P = \frac{\sum_{i=1}^{n} F_i}{\sum_{i=1}^{n}\left(\frac{F_i}{x_i}\right)}$$

ب/ حالة المتغيرة المتصلة: نفس المعادلة بالنسبة لـ H_P المذكور أعلاه مع تعويض X_i بـ C_i و هي مراكز الفئات.

$$H_P = \frac{\sum_{i=1}^{n} F_i}{\sum_{i=1}^{n}\left(\frac{F_i}{C_i}\right)}$$

مثال:

الفئات	**4.5 –**	**5.5 –**	**6.5 –**	**7.5 –**	**8.5 –**	**9.5 – 10.5**	**المجموع**
F_i	3	4	2	3	5	3	20
Ci	5	6	7	8	9	10	-
F_i/C_i	0.6	0.67	0.29	0.38	0.56	0.3	2.8

$$H_P = \frac{20}{2.8} = 7.14$$

المتوسط التربيعي: La moyenne quadratique

يستخدم المتوسط التربيعي في الحالات التي لا يمكن فيها استخدام المتوسط الهندسي (أي التي تتضمن قيم سالبة) و المتوسط الحسابي (أي تلغى فيها القيم السالبة و الموجبة بعضها البعض) و بذلك، يكون نادر الاستخدام.

أولا: المتوسط التربيعي البسيط La moyenne quadratique simple

يعرف على انه الجذر التربيعي لمتوسط مربعات مفردات سلسلة إحصائية ما. و يعطى بالعبارة التالية:

$$MQ_S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

مثال: حساب المتوسط التربيعي للقيم التالية:33،22،17،13،8،5،2

$$MQ_S = \sqrt{\frac{1}{7}\left(2^2+5^2+8^2+13^2+17^2+22^2+33^2\right)} = 17.42$$

ثانيا: المتوسط التربيعي المرجح La moyenne quadratique pondérée

أ/ **حالة المتغيرة المنفصلة**: يعبر رياضيا عن هذا المتوسط بالعلاقة التالية:

$$MQ_P = \sqrt{\frac{\sum_{i=1}^{n} F_i x_i^2}{\sum_{i=1}^{n} F_i}} = \sqrt{\sum_{i=1}^{n} Fr.x_i^2}$$

ب/ **حالة المتغيرة المتصلة**: نفس التعبير الرياضي بالنسبة للمعادلة المذكورة أعلاه مع تعويض X_i بـ C_i:

$$MQ_P = \sqrt{\frac{\sum_{i=1}^{n} F_i C_i^2}{\sum_{i=1}^{n} F_i}} = \sqrt{\sum_{i=1}^{n} Fr._i C_i^2}$$

مثال: من خلال المعطيات التالية:

الفئات	4.5 –	5.5 –	6.5 –	7.5 –	8.5 –	9.5 – 10.5	المجموع
F_i	3	4	2	3	5	3	20
Fr	0.15	0.2	0.1	0.15	0.25	0.15	1
Ci	5	6	7	8	9	10	-
C^2_i	25	36	49	64	81	100	-
$Fr. C^2_i$	3.75	7.2	490	9.6	20.25	15	545.8

يمكن إيجاد قيمة المتوسط التربيعي:

$$MQ_P = \sqrt{\sum_{i=1}^{n} f_i C_i^2} = \sqrt{545.8} = 23.36$$

ملاحظات هامة:

1- العلاقة بين $\bar{X}, G, H, MQ$:

* $H \leq G \leq \bar{X} \leq MQ$ و ذلك مهما اختلف عدد المفردات (بسيطة كانت أو مرجحة)

* $G = \sqrt{\bar{X} \times H}$

2- العلاقة بين $Mo, Me, \bar{X}$:

لقد وجد ك.بيرسون بالتجربة أن في التوزيعات القريبة من التماثل، تكون العلاقة بين المتوسطات الموضعية الثلاث كالتالي: $\bar{X} - Mo = 3(\bar{X} - Me)$

و منه: $\bar{X} \approx \frac{3Me - Mo}{2}$ و بذلك يكون $Me = \frac{2\bar{X} + Mo}{3}$

الفصل الثالث

3

مقاييس التشتت

مقدمـــــــة:

شرحنا المتوسطات و ذكرنا أن أيا منها يهدف إلى قياس القيمة المتوسطة للبيانات. إلا أن القيمة المتوسطة وحدها لا تكفي لإعطاء صورة كاملة عن توزيع ظاهرة ما. والوسط وحده لا يعطي فكرة دقيقة عن مجموعة من القيم فلا يبين طبيعتها و لا كيفية توزيعها. كما أن استخدام الوسط فقط لمقارنة عدة مجموعات لا يكفي لإظهار حقيقة المقارنة. فقد يتساوى متوسطا مجموعتين بينما تختلف المجموعتان عن بعضهما كل الاختلاف، فقد تكون مفردات إحدى المجموعتين متقاربة بعضها من بعض و مفردات المجموعة الثانية مبعثرة (متباعدة بعضها عن بعض).

مثال: لنأخذ السلسلتين التاليتين:

السلسلة: 20,20,19,19,18,18,18,17,16,15

$$\overline{X}=Me=Mo=18$$

السلسلة: 44,30,29,19,18,18,9,6,4,3

$$\overline{X}=Me=Mo=18$$

إذا اكتفينا بمقارنة الوسطين الحسابيين للمجموعتين فقد نستنتج أنهما متساويتين أو متشابهتين و هذا غير صحيح لأنهما مختلفتان. فالقيم في السلسلة الأولى متقاربة أو متجانسة بينما نجد أن القيم في السلسلة الثانية أقل تقاربا مع بعضها و أكثر تباعدا أو تشتتا. و التشتت أو الاختلاف لأي مجموعة من القيم يقصد به التباعد أو الاختلاف بين مفرداتها و يقاس التشتت بمقاييس إحصائية تدعى مقاييس التشتت.

تعريـــــف: تظهر مقاييس التشتت الدرجة التي تنتشر بها البيانات الرقمية حول قيمتها المتوسطة أو بعبارة أخرى تظهر مدى اختلاف قيم الفينة المدروسة فيما بينها أو مدى قربها أو بعدها من أحد مقاييس النزعة المركزية و يكون عادة الوسط

الحسابي. و في هذا الصدد يمكن التمييز بين مقاييس التشتت المطلق و مقاييس التشتت النسبي.

1- مقاييس التشتت المطلق : Les paramètres de dispersion absolue

1-**الانحراف المتوسط L'écart moyen :** إن الانحراف المتوسط بالنسبة للقيمة المركزية هو الوسط الحسابي للفوارق بالنسبة لهذه القيمة. و يمكن أن تكون هذه القيمة إما الوسط الحسابي أو الوسيط.

أ- **الانحراف المتوسط الحسابي :**

أولا: تقدير $e_{\overline{X}}$ من بيانات مفردة

يعبر عنه كما يلي: $e_{\overline{X}} = \frac{\sum_{i=1}^{n} |x_i - \overline{X}_S|}{n}$.

مثال: لتكن لدينا السلسلة الإحصائية التالية: 7,6,5,4,3

لإيجاد قيمة $e_{\overline{X}}$ ، نحسب: $\overline{X}_S = \frac{3+4+5+6+7}{5} = 5$

المجموع	7	6	5	4	3	x_i
0	2	1	0	-1	-2	$x_i - \overline{X}_S$
6	2	1	0	1	2	$\lvert x_i - \overline{X}_S \rvert$

$$e_{\overline{X}} = \frac{\sum_{i=1}^{n} |x_i - \overline{X}_S|}{n} = \frac{6}{5} = 1.2$$

ثانيا: تقدير $e_{\overline{X}}$ من توزيع تكراري

1- حالة المتغير الإحصائي المتقطع

يقدر الانحراف المتوسط الحسابي المرجح في هذه الحالة بالصورة التالية:

$$e_{\overline{X}} = \frac{\sum_{i=1}^{n} F_{i.}\left|x_i - \overline{X}_P\right|}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_i.\left|x_i - \overline{X}_P\right|$$

2- حالة المتغير الإحصائي المستمر

تعوض قيم المتغير x_i بمراكز الفئات فيصبح لدينا:

$$e_{\overline{X}} = \frac{\sum_{i=1}^{n} F_{i.}\left|C_i - \overline{X}_P\right|}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_i.\left|C_i - \overline{X}_P\right|$$

ب- **الانحراف المتوسط بالنسبة للوسيط:**

أولا: تقدير e_{Me} من بيانات مفردة

يعبر عنه كما يلي: $$e_{Me} = \frac{\sum_{i=1}^{n} \left|x_i - Me\right|}{n}$$

مثال: لدينا سلسلة إحصائية تخص عدد النقاط المسجلة خلال 09 مباريات في كرة

السلة: 120,110,103,90,86,75,71,60,55.

أحسب قيمة e_{Me}

لدينا: $$n=9 \Rightarrow \frac{n+1}{2}=5 \Rightarrow Me=X_5=86$$

و بالتالي فقيمة الانحراف تساوي:

$$e_{Me} = \frac{|55-86| + \ldots + |120-86|}{9} = \frac{162}{9} = 18$$

ثانيا: تقدير e_{Me} من توزيع تكراري

1- حالة المتغير الإحصائي المتقطع

يقدر الانحراف المرجح في هذه الحالة بالصورة التالية:

$$e_{Me} = \frac{\sum_{i=1}^{n} F_{i.}|x_i - Me|}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_i .|x_i - Me|$$

2- حالة المتغير الإحصائي المستمر

تأخذ قيمة الانحراف المرجح في حالة البيانات المبوبة الصورة التالية:

$$e_{Me} = \frac{\sum_{i=1}^{n} F_{i.}|C_i - Me|}{\sum_{i=1}^{n} F_i} = \sum_{i=1}^{n} Fr_i .|C_i - Me|$$

الانحراف الربيعي L'écart inter-quartile :

تعريف الربيعيات: إذا كان الوسيط يقسم السلسلة الإحصائية إلى قسمين متساويين، فان الربيعيات تقسمها إلى أربعة أجزاء متساوية. يرمز لهذه الربيعيات بالرمز Q_1 و يسمى الربيع الأول أو الأدنى، Q_2 و هو الربيع الثاني و هو يوافق القيمة الوسيط، Q_3 و يدعى الربيع الثالث أو الأعلى.

Q_1 هو العدد الذي يفصل 25% من اصغر القيم في السلسلة عن 75% من القيم الأكثر منها.

Q_3 هو العدد الذي يفصل 25% من اكبر القيم في السلسلة عن 75% من القيم الأصغر منها.

كيفية حساب الربيعيات:

الحالة الأولى : عندما يكون عدد المعطيات n زوجي. يمكن التمييز بين:

1– إذا كان $\frac{n}{2}$ زوجي فان: $Q_1=\frac{X_{\frac{n}{4}}+X_{\frac{n}{4}+1}}{2}$; $Q_2=Me=\frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$; $Q_3=\frac{X_{\frac{3n}{4}}+X_{\frac{3n}{4}+1}}{2}$

2– إذا كان $\frac{n}{2}$ فردي فان: $Q_1=X_{\frac{n+2}{4}}$; $Q_2=Me=\frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$; $Q_3=X_{\frac{3n+2}{4}}$

الحالة الثانية : عندما يكون عدد المعطيات n فردي. يمكن التمييز هنا أيضا بين:

1– إذا كان $\frac{n-1}{2}$ زوجي فان: $Q_1=\frac{X_{\frac{n-1}{4}}+X_{\frac{n-1}{4}+1}}{2}$; $Q_2=Me=X_{\frac{n+1}{2}}$; $Q_3=\frac{X_{\frac{3n+1}{4}}+X_{\frac{3n+1}{4}+1}}{2}$

2– إذا كان $\frac{n-1}{2}$ فردي فان: $Q_1=X_{\frac{n+1}{4}}$; $Q_2=Me=X_{\frac{n+1}{2}}$; $Q_3=X_{\frac{3(n+3)}{4}}$

تعريف: يعرف الانحراف الربيعي على انه نصف المدى بين الربيع الأول Q_1 و الربيع الثالث Q_3 و بذلك يكون:

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

مثال: لتكن لدينا السلسلة الإحصائية التالية و التي تخص أوزان 30 شخص:

95,90,88,87,85,81,79,78,74,72,69,68,68,66,65,65,65,63,61,60,59,
57,56,55,52,50,49,47,42,41

$n=30\Rightarrow\frac{n}{2}=15$ و بالتالي فقيم الربيعيات تكون على الترتيب :

$$Q_1=X_{\frac{n+2}{4}}=X_8=56\ ;\ Q_2=Me=\frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}=\frac{X_{15}+X_{16}}{2}=\frac{65+65}{2}=65\ ;\ Q_3=X_{\frac{3n+2}{4}}=X_{23}=78$$

$$\Leftarrow e_Q=\frac{Q_3-Q_1}{2}=\frac{78-56}{2}=11$$

ملاحظة هامة: في حالة ما إذا كانت المعطيات مبوبة فان قيمة الربيعيات تحدد على من العلاقة التالية:

$$Q_i=L_0+\frac{\frac{in}{4}-F_1}{F_2-F_1}\times K$$

حيث للربيع i يكون: L_0 هو الحد الأدنى لفئة الربيع، $\frac{in}{4}$ هو ترتيب الربيع، F_1 هو التكرار المتجمع الصاعد قبل ترتيب الربيع، F_2 هو التكرار المتجمع الصاعد بعد ترتيب الربيع، K هو طول فئة الربيع

يعتبر الانحراف الربيعي كمقياس للتشتت افضل من المدى إلا انه أيضا يعيبه الاعتماد على قيمتين فقط في حسابه و لكن من مزاياه انه قابل للحساب من الجداول التكرارية المفتوحة.

مثال: لدينا الجدول التالي و الذي يخص توزيع المساكن حسب مدة بنائها.

مدة البناء	**F_i**	**F.c.c**
]4–2]	5	0
]6–4]	20	5
]8–6]	20	25
]10–8]	35	45
]12–10]	15	80
]14–12]	20	95
]16–14]	30	115
]18–16]	45	145
]20–18]	10	190
		200
المجموع	**200**	–

من خلال الجدول يمكن حساب قيم كل الربيعيات و هي على التوالي:

$$Q_1=L_0+\frac{\frac{200}{4}-F_1}{F_2-F_1}\times K \Rightarrow Q_1=8+\frac{50-45}{80-45}\times 2=8.29$$

$$Q_2=Me=L_0+\frac{\frac{200}{2}-F_1}{F_2-F_1}\times K \Rightarrow Q_2=12+\frac{100-95}{115-95}\times 2=12.5$$

$$Q_3=L_0+\frac{\frac{3\times(200)}{4}-F_1}{F_2-F_1}\times K \Rightarrow Q_3=16+\frac{150-145}{190-145}\times 2=16.22$$

و منه فان الانحراف الربيعي يساوي:

$$e_Q=\frac{Q_3-Q_1}{2}=\frac{16.22-8.29}{2}=3.97$$

الانحراف المعياري L'écart type :

أولا: تقدير S_X من بيانات مفردة

لاحظنا عند حساب الانحراف المتوسط أننا قد لجأنا لاستخدام القيمة المطلقة . إلا أن هذه الطريقة غير فعالة في كثير من الأحيان (لا يستخدم لعدم الاستفادة منه كثيرا في مجال الاستدلال الإحصائي) و لتفادي أي إشكال عند حساب الفرق بين كل قيمة و $\bar{X}$ (التغلب على الانحرافات السالبة) نقوم بتربيع هذا الفرق و بالتالي فالعبارة الجديدة تصبح: $\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}_S\right)^2}{n}$. نسمي القيمة التي توافق هذا القانون بالتباين La variance.و يرمز لها بالرمز S_X^2 . $S_X^2=\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}_S\right)^2}{n}$.إذا قمنا بحساب الجذر التربيعي للتباين فإننا نحصل على قيمة تدعى الانحراف المعياري و يرمز لها بالرمز S_X حيث: $S_X=\sqrt{S_X^2}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}_S\right)^2}{n}}$

ملاحظات:

1- في المقام يفضل القسمة على $n-1$ بدلا من n للحصول على التباين لان S_X^2 يعتبر تباين العينة و هو عبارة عن أداة لتقدير σ^2 (وهو تباين المجتمع و الذي في كثير من التطبيقات الإحصائية غير معلوم). و تعطي قيمة S_X^2 افضل تقدير لــ σ^2 لذا يلجأ إليه الإحصائيين عمليا. و عليه فالعبارة الجديدة للتباين و الانحراف المعياري بالتبعية تصبح في الصورة التالية: $S_X=\sqrt{S_X^2}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X}_S)^2}{n-1}}$

2- تتفق قيمة S_X مع درجة التجانس حيث تعني القيمة $S_X=0$ تجانس تام. و كلما ابتعدت قيمته عن الصفر قلت درجة التجانس (أي زاد تشتت المفردات)

3- يمكن استعمال طريقة بديلة لتسهيل حساب التباين. تستبدل العلاقة السابقة بالقانون التالي:

$$S_X^2=\frac{\sum x_i^2-n\bar{X}_S^2}{n-1}=\frac{1}{n-1}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]$$

مثال: حساب الانحراف المعياري للقياسات التالية(باستخدام الطريقة البديلة):33,27,24,18,9,12,21

$$S_X^2=\frac{\sum x_i^2-n\bar{X}_S^2}{n-1}=\frac{1}{n-1}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]=\frac{1}{6}\left[3384-\frac{(144)^2}{7}\right]=70.29\Rightarrow S_X=8.38$$

ثانيا: تقدير S_X من توزيع تكراري

1- حالة المتغير الإحصائي المتقطع

يحسب التباين بتطبيق المعادلة:

$$S_X^2=\frac{\sum_{i=1}^{n}F_i\left(x_i-\overline{X}_P\right)^2}{\sum_{i=1}^{n}F_i-1}=\frac{\sum F_i.x_i^2-n\overline{X}_P^2}{\sum F_i-1}$$

و باستخدام صيغة مكافئة سهلة، تصبح العبارة:

$$S_X^2=\frac{1}{\sum F_i-1}\left[\sum F_i.x_i^2-\frac{\left(\sum F_i.x_i\right)^2}{\sum F_i}\right]\Rightarrow S_X=\sqrt{S_X^2}$$

2- حالة المتغير الإحصائي المستمر : *في هذه الحالة تعوض القيم x_i بمراكز الفئات C_i و بالتالي يصبح التباين يحسب بتطبيق القانون التالي:*

$$S_X^2=\frac{1}{\sum F_i-1}\left[\sum F_i.C_i^2-\frac{\left(\sum F_i.C_i\right)^2}{\sum F_i}\right]\Rightarrow S_X=\sqrt{S_X^2}$$

مثال: لتكن لدينا المعطيات التالية و التي تخص أوزان 30 طالب

الأوزان	]50–40]	]60–50]	]70–60]	]80–70]	]90–80]	]100–90]	المجموع
F_i	4	6	10	4	4	2	30
Ci	45	55	65	75	85	95	–
C_i^2	2025	3025	4225	5625	7225	9025	–
$F_i.C_i$	180	330	650	300	340	190	1990
$F_i.C_i^2$	8100	18150	42250	22500	28900	18050	137950

$$S_X^2=\frac{1}{\sum F_i-1}\left[\sum F_i.C_i^2-\frac{\left(\sum F_i.C_i\right)^2}{\sum F_i}\right]=\frac{1}{29}\left[137950-\frac{(1990)^2}{30}\right]=205.06\Rightarrow S_X=\sqrt{S_X^2}=\sqrt{205.06}=14.32$$

ملاحظات:

1- تستخدم مراكز الفئات لتقدير كل من $\bar{X}_P$ و S_X^2 مما يجعل قيمة $\bar{X}_P \rangle \bar{X}$ لنفس القيم و يرجع هذا لخطأ التبويب (و يصغر هذا الفرق كلما صغر طول الفئة K).

و كذلك بالنسبة لــ S_X^2 فانه بسبب تربيع الفروق فهي تميل إلى التجمع و بذلك يميل S_X^2 و S_X إلى الكبر عن نظيرتهما من البيانات المفردة. و لذلك اقترح شيبارد **SHEPPARD** تصحيحا للتباين من التوزيع التكراري المستمر بتطبيق القانون: $S_X^{*2} = S_X^2 - \frac{K^2}{12}$

و عليه يصبح التباين المصحح في المثال السابق:

$$S_X^{*2} = S_X^2 - \frac{K^2}{12} = 205.06 - \frac{10^2}{12} = 196.73$$

و يكون $S_X^* = 14.03$ بدلا من 14.32

2- من الممكن تسهيل الحسابات أكثر لإيجاد قيمة الانحراف المعياري و ذلك بالاستفادة من إحدى القاعدتين التاليتين:

القاعدة الأولى: إذا كانت لدينا مجموعة من القياسات، و قمنا بطرح من كل القيم مقدارا ثابتا (يسمى متوسطا فرضيا) فان الانحراف المعياري للمجموعة الجديدة من القياسات هو نفسه بالنسبة للمجموعة الأصلية.

مثال: لدينا السلسلة التالية:33,27,24,18,9,12,21 لو طرحنا عددا ثابتا من كل قيمة و لنفرضه 20 فالقيم الجديدة تصبح من الصورة التالية:13,7,4,-2,-11,-8,1

و قيمة الانحراف المعياري تصبح:

$$S_X^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right] = \frac{1}{6}\left[424 - \frac{(4)^2}{7}\right] = 70.29 \Rightarrow S_X = 8.38$$

و هي نفي القيمة الموجودة سابقا باستخدام القيم الأصلية.

القاعدة الثانية: : إذا كانت لدينا مجموعة من القياسات، و قمنا بقسمة كل قياس على مقدار ثابت فان الانحراف المعياري للمجموعة الأصلية هو جداء الانحراف المعياري للمجموعة الجديدة من القياسات بمربع هذا المقدار الثابت.

مثال: تصبح قيم المثال السابق بعد قسمتها على عدد ثابت و لنفرضه 3 كالآتي:11,9,8,6,3,4,7

و قيمة الانحراف المعياري تصبح:

$$S'^2_X=\frac{1}{n-1}\left[\sum x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]=\frac{1}{6}\left[376-\frac{(48)^2}{7}\right]=7.81\Rightarrow S^2_X=9.S'^2_X=70.29\Rightarrow S_X=8.38$$

3-هناك علاقة تجريبية تربط S_X، $e_{\bar{X}}$ و e_Q في حالة التوزيع المتماثل أو محدود الالتواء و هي: $e_{\bar{X}}=\frac{4}{5}S_X$ و $e_Q=\frac{2}{3}S_X$

2- مقاييس التشتت النسبي : Les paramètres de dispersion relative

يقاس كل مقياس من مقاييس التشتت التي استعرضناها بنفس وحدات قياس الظاهرة موضع الدراسة. و عند مقارنة التشتت لمفردات ظاهرتين مختلفتين أو لمفردات ظاهرة واحدة على مستويات مختلفة، يجب أن نحول مقياس التشتت المطلق للمفردات إلى مقياس نسبي حتى يمكن التخلص من تأثير اختلاف وحدات القياس و كذلك التخلص من تأثير اختلاف المتوسط لكل منهما. و لإجراء ذلك يستخدم مقياس يسمى معامل الاختلافCoefficient de variation و يرمز له بالرمز CV و له صورتان:

الصورة الأولى: تعتمد في حسابها على الانحراف المعياري و المتوسط الحسابي و يرمز له بالرمز CV_1 حيث:

$$CV_1=\frac{S_X}{\bar{\bar{X}}}\times 100$$

لاحظ أن معامل الاختلاف لا يعتمد على الوحدات المستخدمة، لهذا السبب يكون مناسبا لمقارنة توزيعات مختلفة للوحدات.

من عيوبه هو صعوبة حسابه عندما يكون $\bar{X}$ قريب من الصفر و لا يستخدم إلا في حالة التوزيعات المقفلة.

الصورة الثانية: تعتمد في حسابها على الانحراف الربيعي و الوسيط و يرمز لها بالرمز CV_2 حيث:

$$CV_2 = \frac{Q_3 - Q_1}{Me} \times 100$$

و يستخدم هذا المقياس في حالة التوزيعات المفتوحة.

مثال1 : قيست أطوال مجموعة من الطلبة فوجد أن متوسط أطوالهم =160سم بانحراف معياري=16سم. و قيست أوزانهم فوجد أن متوسطها=60كغ بانحراف معياري=12كغ. أيهما اكثر تشتتا الأطوال أم الأوزان؟

$$CV_1(L) = \frac{S_X}{\bar{X}} \times 100 = \frac{16}{160} \times 100 = 10\%$$

$$CV_1(P) = \frac{S_X}{\bar{X}} \times 100 = \frac{12}{60} \times 100 = 20\%$$

نستلزم أن الأوزان أكثر تشتتا من الأطوال.

مثال2 : بعد تصحيح أوراق الامتحانات لمجموعتين من الطلبة و جد أن متوسط علامات المجموعة الأولى=16 بانحراف معياري=2 بينما متوسط علامات المجموعة الثانية= 13بانحراف معياري=3. أي المجموعتين اكثر تشتتا؟

$$CV_1(G_1) = \frac{S_X}{\bar{X}} \times 100 = \frac{2}{16} \times 100 = 12.5\%$$

$$CV_1(G_2) = \frac{S_X}{\bar{X}} \times 100 = \frac{3}{13} \times 100 = 23.08\%$$

نستلزم أن المجموعة الثانية اكثر تشتتا من المجموعة الأولى.

الفصل الرابع

4

مقاييس الشكل

مقدمة:

إذا أردنا معرفة شكل التوزيع التكراري فنلجأ عادة إلى استعمال أدوات إحصائية و مقاييس مختصة، حيث تسمح لنا بقياس و بالتالي معرفة مدى التواء شكل التوزيع أو مدى تفرطحه.

إن الحاجة لمعرفة هذا المدى تدفعنا لحساب معاملات تعطينا التقديرات الكمية إما للالتواء (معامل الالتواء) أو للتفرطح (معامل التفرطح). و هذا بدوره يتطلب معرفة كيفية حساب ما يعرف بالعزوم الإحصائية.

تعرف العزوم عادة حول نقطة. و هناك نوعين من النقط اللتين تعرف حوليهما العزوم: قد تعرف حول الصفر و في هذه الحالة تسمى العزوم اللامركزية، و قد تعرف حول المتوسط الحسابي و تدعى عندئذ بالعزوم المركزية – باعتبار أن المتوسط الحسابي هو مركز التوزيع–

و العبارة العامة للعزوم هي: $m_r = \sum_{i=1}^{n} f_i (x_i - b)^r$

العزوم اللامركزية $(b=0)$ Les moments non centrés

يعرف العزم اللامركزي ذو الرتبة **r** و نرمز له بالرمز m_r كالآتي:

أولا: تقدير m_r من بيانات مفردة

$$m_r = \frac{\sum x_i^r}{n} \qquad r = 1,2,3,...$$

و بذلك، تكون قيم العزوم اللامركزية الأربعة الأولى على الصورة:

$$m_1 = \frac{\sum x_i}{n} = \bar{X} \qquad m_2 = \frac{\sum x_i^2}{n} \qquad m_3 = \frac{\sum x_i^3}{n} \qquad m_4 = \frac{\sum x_i^4}{n}$$

ثانيا: تقدير m_r من توزيع تكراري

1- حالة المتغيرة الإحصائية المتقطعة

$$m_r = \frac{\sum F_i . x_i^r}{\sum F_i} = \sum f_i . x_i^r \qquad r = 1,2,3,...$$

و بالتالي نحصل على:

$$m_1 = \frac{\sum F_i . x_i}{\sum F_i} = \sum f_i . x_i = \overline{X} \qquad m_2 = \frac{\sum F_i . x_i^2}{\sum F_i} = \sum f_i . x_i^2 \qquad m_3 = \frac{\sum F_i . x_i^3}{\sum F_i} = \sum f_i . x_i^3$$

$$m_4 = \frac{\sum F_i . x_i^4}{\sum F_i} = \sum f_i . x_i^4$$

حالة المتغيرة الإحصائية المتصلة

$$m_r = \frac{\sum F_i . C_i^r}{\sum F_i} = \sum f_i . C_i^r \qquad r = 1,2,3,...$$

العزوم المركزية $\left(b = \overline{X}\right)$ Les moments centrés

يعرف العزم المركزي ذو الرتبة **r** و نرمز له بالرمز μ_r كالآتي:

أولا: تقدير μ_r من بيانات مفردة

$$\mu_r = \frac{\sum \left(x_i - \overline{X}\right)^r}{n} \qquad r = 1,2,3,...$$

و بالتالي، فقيم العزوم المركزية الأربعة الأولى تكون:

$$\mu_4 = \frac{\sum \left(x_i - \overline{X}\right)^4}{n} \qquad \mu_3 = \frac{\sum \left(x_i - \overline{X}\right)^3}{n} \qquad \mu_2 = \frac{\sum \left(x_i - \overline{X}\right)^2}{n} = \sigma^2$$

$$\mu_1 = \frac{\sum \left(x_i - \overline{X}\right)}{n} = 0$$

ثانيا: تقدير μ_r من توزيع تكراري

1- حالة المتغيرة الإحصائية المتقطعة

$$\mu_r = \frac{\sum F_i (x_i - \bar{X})^r}{\sum F_i} = \sum f_i (x_i - \bar{X})^r \qquad r = 1,2,3,...$$

2 - حالة المتغيرة الإحصائية المتصلة

$$\mu_r = \frac{\sum F_i (C_i - \bar{X})^r}{\sum F_i} = \sum f_i (C_i - \bar{X})^r \qquad r = 1,2,3,...$$

العلاقة بين العزوم المركزية و العزوم اللامركزية

من الواضح انه إذا حسبنا العزوم المركزية بالطريقة المباشرة، فإننا قد نضطر للتقريب في حالة احتواء المتوسط الحسابي على كسور. و لتفادي ذلك، فانه يمكن حساب العزوم المركزية باستخدام العلاقات التالية:

$$\mu_2 = \sigma_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4$$

<u>**مثال:**</u> إيجاد العزوم المركزية للبيانات التالية:2,3,5,6,8,9

$(x_i-\bar{X})^4$	$(x_i-\bar{X})^3$	$(x_i-\bar{X})^2$	$x_i-\bar{X}$	X^4	X^3	X^2	X
150.06	−42.88	12.25	−3.5	16	8	4	**2**
39.06	−15.63	6.25	−2.5	81	27	9	**3**
0.06	−0.13	0.25	−0.5	625	125	25	**5**
0.06	0.13	0.25	0.5	1296	216	36	**6**
39.06	15.63	6.25	2.5	4096	512	64	**8**
150.06	42.88	12.25	3.5	6561	729	81	**9**
378.36	**0**	**37.5**	**0**	**12675**	**1617**	**219**	**33**

حساب العزوم اللامركزية:

$$m_1=\frac{\sum x_i}{n}=\overline{X}=\frac{33}{6}=5.5 \qquad m_2=\frac{\sum x_i^2}{n}=\frac{219}{6}=36.5$$

$$m_3=\frac{\sum x_i^3}{n}=\frac{1617}{6}=269.5 \qquad m_4=\frac{\sum x_i^4}{n}=\frac{12675}{6}=2112.5$$

حساب العزوم المركزية – بالطريقة المباشرة –

$$\mu_1=\frac{\sum\left(x_i-\overline{X}\right)}{n}=0 \qquad \mu_2=\frac{\sum\left(x_i-\overline{X}\right)^2}{n}=\sigma_2=\frac{37.5}{6}=6.25$$

$$\mu_3=\frac{\sum\left(x_i-\overline{X}\right)^3}{n}=0 \qquad \mu_4=\frac{\sum\left(x_i-\overline{X}\right)^4}{n}=\frac{378.36}{6}=63.06$$

حساب العزوم المركزية – بالاعتماد على العزوم اللامركزية –

$$\mu_2=\sigma_2=m_2-m_1^2=36.5-(5.5)^2=6.25$$

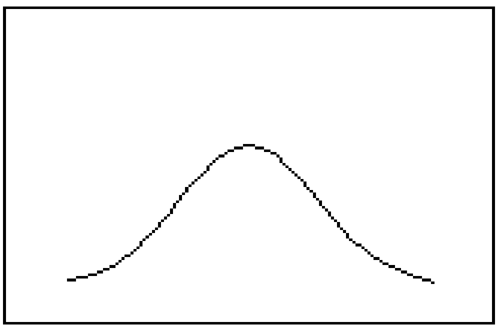
$$\mu_3=m_3-3m_1m_2+2m_1^3=269.5-3[(5.5)(36.5)]+2(5.5)^3=0$$

$$\mu_4=m_4-4m_1m_3+6m_1^2m_2-3m_1^4=63.06$$

الالتواء L'Asymétrie

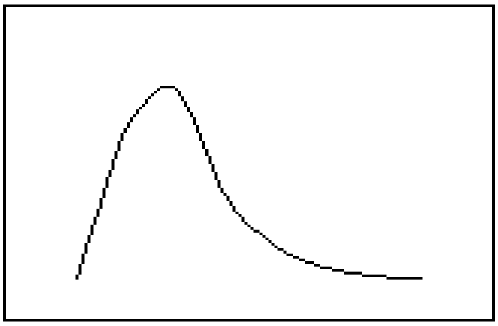
بعد معرفة درجة انتشار التكرارات حول قيمة وسطى، يتم تحديد شكل المنحنى الممثل للتوزيع التكراري للعينة المدروسة. فقد يكون التوزيع التكراري:

*متماثلا إذا كان $\overline{X}=Me=Mo$ أي تكون 50% من القيم على يمين و على يسار هذه المقاييس. و يعتبر هذا الشكل شكلا معياريا أي انه تقاس بالنسبة له كل الأشكال المتبقية و التي سنتطرق إليها.

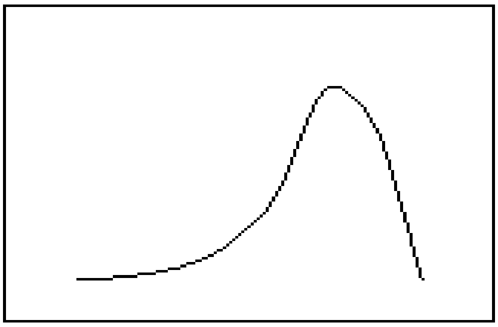


* موجب الالتواء إذا كان ممتدا أكثر نحو اليمين و يكون في هذه الحالة

$$\bar{X} > Me > Mo$$



* سالب الالتواء إذا كان ممتدا أكثر نحو اليسار و يكون عندئذ $\bar{X} < Me < Mo$



من الأشكال الثلاثة، يمكن تعريف الالتواء على انه بعد المنحنى عن التماثل و هو إما التواء موجب أو التواء سالب. و هو يسمح بتلخيص و مقارنة التوزيعات إلى جانب مقاييس النزعة المركزية و مقاييس التشتت التي لا تكفي لوحدها بالقيام بهذه العملية (أي عملية المقارنة). فقد يتساوى توزيعان من حيث المتوسط و الانحراف المعياري و لكنهما تختلفان من حيث الالتواء. و قد يكون التوائهما في اتجاه واحد و لكن بقدرين مختلفين. أو قد تتساوى درجة التوائهما و لكنهما يختلفان في الإشارة. كما يمكن معرفة نوع الالتواء (موجب أو سالب) و درجته (بسيط أو حاد) من شكل المنحنى نفسه، إلا أن هذا لا يعطينا قياسا رقميا للالتواء.

لهذا الغرض، أصبح من المهم التعرف على أداة إحصائية تستخدم لقياس الالتواء تسمى **معامل الالتواء** **Le coefficient d'asymétrie** و يمكن حسابه بإحدى الطريقتين:

أ- **الطريقة الدقيقة:** في التوزيعات المتماثلة، يكون العزم المركزي ذو الرتبة 3 (μ_3) مساو للصفر. أما في التوزيعات غير المتماثلة فقد تكون قيمة μ_3 موجبة أو

سالبة. فكلما كان μ_3 قريبا من الصفر، كلما كان المنحنى قريبا من التماثل. و بالعكس إذا كان μ_3 (موجبا أو سالبا)، كان الالتواء شديدا. و من هذا المنطلق، يمكن استخدام μ_3 كأساس لمقياس الالتواء و يطلق عليه اسم **معامل الالتواء العزمي** أو **معامل فيشر للالتواء** و يعبر عنه رياضيا بالعلاقة:

$$\gamma_F = \frac{\mu_3}{\sigma_3} \quad ,\sigma = \sqrt{\sigma_2} = \sqrt{\mu_2}$$

و جدير بالملاحظة أن معامل التواء التوزيع المتماثل =0. و عندما يكون المنحنى سالب الالتواء، فان إشارة γ_F تكون سالبة. كذلك في حالة الالتواء الموجب تكون إشارة γ_F موجبة.

مثال: إيجاد معامل فيشر γ_F للتوزيع التكراري التالي:

الفئات	]40–30]	]50–40]	]60–50]	]70–60]	]80–70]	]90–80]	]100–90]	المجموع
F_i	4	6	11	13	22	16	8	**80**
C_i	35	45	55	65	75	85	95	−
$C_i-\overline{X}_P$	−35.4	−25.4	−15.4	−5.4	4.6	14.6	24.6	−
$F_i(C_i-\overline{X}_P)^2$	5012.64	3870.96	2608.76	379.08	465.52	3410.56	4841.28	**20588.8**
$F_i(C_i-\overline{X}_P)^3$	**177447.46 −**	**−98322.38**	**−40174.9**	**−2047.03**	**2141.39**	**49794.18**	**119095.49**	**146960.71 −**

$$\mu_2 = \frac{\sum F_i(C_i-\overline{X})^2}{\sum F_i} = \frac{20588.8}{80} = 257.36$$

$$\mu_3 = \frac{\sum F_i(C_i-\overline{X})^3}{\sum F_i} = \frac{-146960.71}{80} = -1837.01$$

و بالتالي تصبح قيمة γ_F:
$$\gamma_F=\frac{\mu_3}{\sigma_3}=\frac{-1837.01}{\left(\sqrt{257.36}\right)^3}=-0.44$$

$\Leftarrow$ المنحنى سالب الالتواء لان إشارة γ_F سالبة.

ب- الطريقة التقريبية: لقد سبق و ذكرنا أن بيرسون قد وجد بالتجربة أن التوزيع لا يعتبر متماثلا و لا ملتويا و إنما قريب من الالتواء إذا تحققت العلاقة: $\bar{X}-Mo=3\left(\bar{X}-Me\right)$

. من هذه العلاقة، يمكن معرفة شكل التوزيع حيث، مثلا، عندما يكون: $\bar{X}-Mo=0 \Leftarrow$ التوزيع متماثل ، $\bar{X}-Me<0 \Leftarrow$ التوزيع سالب الالتواء

و هذا ما ساعد في إيجاد معاملي التواء تقريبيين، سنرمز للأول منهما بالرمز γ_1 و للثاني بالرمز γ_2 حيث:

$\gamma_1=\frac{\bar{X}-Mo}{\sigma}$ و يطلق عليه اسم: **معامل بيرسون الأول للالتواء**

$\gamma_2=\frac{3\left(\bar{X}-Me\right)}{\sigma}$ و يطلق عليه اسم: **معامل بيرسون الثاني للالتواء**

واضح أن المعاملين γ_1 و γ_2 لا يمكن حسابهما في حالة التوزيعات المفتوحة. و لتفادي هذا، يمكن حساب الالتواء من معامل ثالث يرمز له بالرمز γ_3 و هو يعتمد في حسابه على كل من Q_1 ، $Q_2=Me$ و Q_3 . و تعتمد فكرته على أساس انه في حالة التوزيعات المتماثلة، يقع Q_1 و Q_3 على بعدين متساويين من الوسيط $Q_2=Me$. أما في حالة التوزيعات الملتوية، فيختلف بعدا هاذين الربيعيين عن الوسيط و بذلك يكون الفرق بين بعديهما مقياسا للالتواء كالتالي:

$$\gamma_3=\frac{\left(Q_3-Q_2\right)-\left(Q_2-Q_1\right)}{Q_3-Q_1}=\frac{Q_1+Q_3-2Q_2}{Q_3-Q_1}$$

و يطلق عليه اسم: **معامل الالتواء الربيعي**

مثال: إيجاد معاملات الالتواء الثلاثة التقريبية للتوزيع التكراري السابق

الفئات	F_i	F.c.c
]40–30]	4	0
]50–40]	6	4
]60–50]	11	10
]70–60]	13	21
]80–70]	22	34
]90–80]	16	56
]100–90]	8	72
		80
المجموع	80	–

من الجدول نجد أن الفئة المنوالية هي الفئة]80–70] و بالتالي فقيمة المنوال تكون:

$$Mo = L_0 + \frac{F_0 - F_1}{(F_0 - F_1) + (F_0 - F_2)} \times K = 70 + \frac{22-13}{(22-13)+(22-16)} \times 10 = 76$$

كما يمكن حساب قيمة كل من Q_1 ، $Q_2 = Me$ و Q_3 من العلاقات التالية علما أن:

Q_1 ينتمي للفئة]60–50] و $Q_2 = Me$ ينتمي للفئة الوسيطية]80–70] و Q_3 ينتمي للفئة]90–80]

و بالتالي:

$$Q_1 = 50 + \frac{20-10}{21-10} \times 10 = 59.1$$

$$Me = L_0 + \frac{X_0 - F_1}{F_2 - F_1} \times K = 70 + \frac{40-34}{56-34} \times 10 = 72.7$$

$$Q_3 = 80 + \frac{60-56}{72-56} \times 10 = 82.5$$

و عليه فان قيم المعاملات هي:

$$\gamma_1 = \frac{\bar{X} - Mo}{\sigma} = \frac{70.4-76}{\sqrt{257.36}} = -0.35$$

$$\gamma_2 = \frac{3\left(\bar{X} - Me\right)}{\sigma} = \frac{3(70.4-72.7)}{\sqrt{257.36}} = -0.43$$

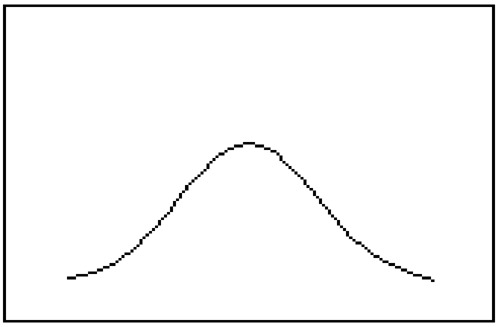
$$\gamma_3 = \frac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \frac{59.1 + 82.5 - 2(72.7)}{82.5 - 59.1} = -0.16$$

يلاحظ أننا حصلنا على نتائج مختلفة و هذا لا يناقض بعضه. إذ أن المعاملات الثلاثة تقريبية و كل منها يقيس الالتواء على أساس يخالف المعاملين الآخرين. و لذلك، فانه عند مقارنة التواء توزيعات مختلفة، يجب استخدام نفس المعامل.

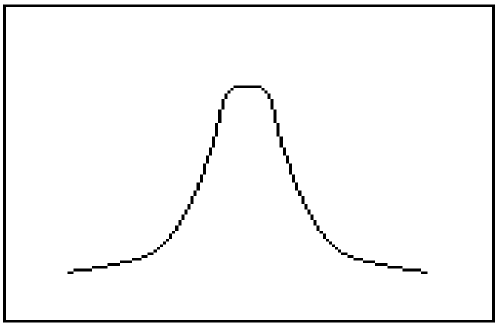
التفرطح L'Aplatissement

يعبر التفرطح عن درجة تدبب قمة منحنى التوزيع (أي مدى اتساع أو ضيق قمة المنحنى). و حسب درجة التسطح ترتب المنحنيات. فعندما:

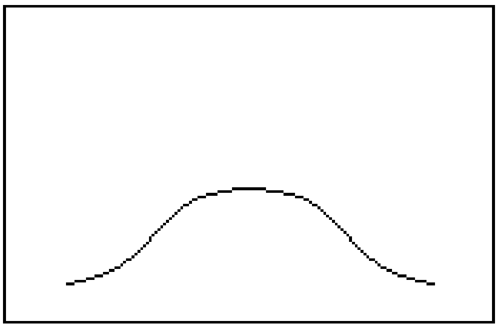
* يكون منحنى التوزيع معتدلا يدعى بالمنحنى متوسط التفرطح



* و عندما تتركز مجموعة البيانات بشدة حول متوسطها الحسابي يكون المنحنى مدببا (أي يكون ضيقا في الوسط و ترتفع قمته)



* عندما تتوزع مجموعة البيانات التي يمثلها المنحنى حول متوسطها الحسابي في مجال غير ضيق يكون المنحنى مفرطحا (أي يكون أكثر اتساعا من وسطه و تنخفض قمته عن قمة المنحنى المعتدل).



يقاس التفرطح بأداة إحصائية تسمى **معامل التفرطح Le coefficient d'aplatissement** و تحسب قيمته كمايلي:

$$K=\frac{\mu_4}{\mu_2^2}=\frac{\mu_4}{\sigma^4}$$

و يدعى في هذه الحالة بـــ: **معامل التفرطح العزمي**

تجدر الإشارة في هذا الصدد أن المنحنى المعتدل (متوسط التفرطح) يعتبر معيارا لتحديد كذلك التفرطح. و وجد عمليا أن **K** في التوزيع المعتدل =3. و منه، يكون الرقم 3 أساسا للتفرقة بين المنحنيات من حيث التدبب و التفرطح.

و عليه فان:

$K=3$ تعني أن التوزيع متماثل القمة أو معتدل

$K>3$ تعني أن التوزيع مدبب القمة

$K<3$ تعني أن التوزيع مفرطح القمة

<u>مثال:</u> إيجاد معامل التفرطح العزمي لمعطيات نفس المثال السابق

$$\mu_2=257.36 \Rightarrow \mu_2^2=66234.17$$

$$\mu_4=\frac{\sum F_i\left(C_i-\overline{X}\right)^4}{\sum F_i}=16344213$$

$$\Rightarrow K=\frac{\mu_4}{\mu_2^2}=\frac{\mu_4}{\sigma^4}=\frac{16344213}{66234.17}=2.47 \quad <3$$

$\Leftarrow$ المنحنى مفرطح

قائمة المراجع

- تيلولت سامية، نماذج الامتحانات في الإحصاء مع الحلول، كلية العلوم الاقتصادية وعلوم التسيير – جامعة الجزائر–، 2009.
- جلاطو جيلالي، الإحصاء التطبيقي مع تمارين ومسائل محلولة، دار الخلدونية، الجزائر، 2009.
- عبد الرزاق عزوز، الكامل في الإحصاء – دروس مفصلة مع تمارين ومسائل محلولة–، ديوان المطبوعات الجامعية، الجزائر، دون ذكر سنة النشر.
- محمد حسين محمد رشيد، الإحصاء الوصفي والتطبيقي والحيوي، دار الصفاء للنشر والتوزيع، عمان، 2008.
- موساوي عبد النور، الإحصاء دار العلوم للنشر والتوزيع، الجزائر، 2009.
- مقدمة في الإحصاء، جامعة الملك عبد العزيز، 1429/1428هـ.