

جامعة النيل الأبيض

محاضرات الإحصاء

إعداد الأستاذ / فياض محمد عظيم أحمد

المحاضرة الأولى

مقدمة

علم الإحصاء: هو العلم الذي يبحث في:

- طرق جمع البيانات الصحيحة والدقيقة عن ظاهرة ما، ثم تلخيص هذه البيانات (بيانياً أو جدولياً) في صورة مبسطة يسهل معها قراءتها واستخدامها.
- وصف هذه البيانات ثم تحليلها واستخراج النتائج منها، واتخاذ القرارات المناسبة.
- دراسة علاقة الظاهرة بباقي الظواهر وتقدير قيمة الظاهرة في المستقبل.

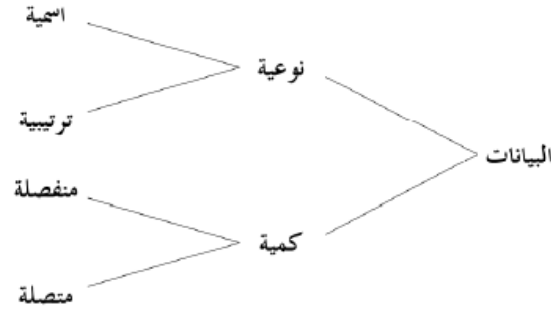
أقسام علم الإحصاء

- الإحصاء الوصفي: يختص بجمع البيانات وعرضها ووصفها وتلخيصها.
- الإحصاء الاستدلالي: يختص باستخلاص وتفسير النتائج واتخاذ القرارات.

مصطلحات إحصائية:

- المجتمع: هو مجموعة من الأفراد أو المشاهدات أو الأشخاص والتي نرغب في دراسة وتحليل خصائصه، وهناك نوعين: مجتمع محدود أو نهائي و مجتمع غير محدود أو لا نهائي.
- العينة: هي أي مجموعة جزئية من المجتمع، وتعبّر عنه أصدق تمثيل.
- المعلمة: عبارة عن قيمة تعبر عن بيانات المجتمع.
- إحصاء العينة (المقياس): عبارة عن قيمة تعبر عن بيانات العينة.
- البيانات: عبارة عن مجموعة القيم أو القياسات للمتغير الذي يرافق الأفراد أو عناصر المجتمع. وقد تكون في شكل أرقام أو صفات أو رموز، ويمكن تصنيف البيانات على النحو

الآتي:



البيانات الاسمية: عبارة عن اسم أو وصف لأي عنصر أو مفردة في المجتمع.

البيانات الترتيبية: عبارة عن اسم أو وصف يعبر عن التفضيل أو الترتيب لأي عنصر في المجتمع.

البيانات المنفصلة: عبارة عن قيم تدل على صفة يمكن عدّها، وتأخذ قيم صحيحة فقط.

البيانات المتصلة: عبارة عن قيم تدل على صفة يمكن قياسها، وتأخذ جميع القيم الصحيحة والكسرية.

مثال: حدد نوع البيانات التالية:

البيان	بيانات اسمية	بيانات ترتيبية	بيانات منفصلة	بيانات متصلة
عدد أسهم شركة مساهمة المخصصة للفرد				
عدد الإداريين في إحدى الأقسام بالجامعة				
الدرجات العلمية لعضو هيئة التدريس				
رضا المستهلك عن منتج معين من إحدى الشركات				
كمية الخبر الموجودة في أحد أنواع الاقلام				
أنواع الألوان المستخدمة في طباعة كتاب معين				
عدد الغائبين من الموظفين في أحد الأيام				
معدلات الطلاب في إحدى الشعب بإحدى الجامعات				

التوزيعات التكرارية

1.1. التوزيعات التكرارية : هي عبارة عن جداول تلخص فيها بيانات العينة إلى فئات، ولكل فئة تكرار تتحدد قيمته حسب البيانات. وتسمى البيانات بعد تلخيصها في جداول تكرارية *بـالبيانات الهبوية*. وعادة ما نلجأ لهذا الأسلوب إذا كانت لدينا عدد كبير من البيانات.

مثال (1.1) (مثال على البيانات الوصفية) البيانات الآتية تمثل المؤهلات العلمية لعينة من (35) موظف بإحدى الشركات، والمطلوب تلخيص هذه البيانات في جدول تكراري.

ثانوي	ثانوي	دكتوراه	ثانوي	جامعي	جامعي	جامعي
ابتدائي	ثانوي	جامعي	ثانوي	متوسط	جامعي	جامعي
ثانوي	ثانوي	متوسط	ثانوي	ثانوي	جامعي	دكتوراه
جامعي	جامعي	ثانوي	ثانوي	جامعي	ثانوي	جامعي
ثانوي	ثانوي	جامعي	ثانوي	ثانوي	جامعي	متوسط

الحل:

جدول تكراري → جدول تفريري

التكرار f	العلامات	الفئات c
2	//	دكتوراه
13	###	جامعي
16	### /	ثانوي
3	///	متوسط
1	/	ابتدائي
35	(Σ) المجموع	

عدد الموظفين (f)	المؤهل العلمي (c)
2	دكتوراه
13	جامعي
16	ثانوي
3	متوسط
1	ابتدائي
35	Σ

مثال (1.2): (مثال على البيانات المنفصلة) فيما يلي عدد الغيابات لعينة من (30) موظف بإحدى الشركات:

0	3	0	0	3	0
2	2	0	1	2	1
0	0	1	2	4	0
4	2	1	0	1	0
0	2	0	1	3	2

المطلوب: تلخيص بيانات العينة في جدول تكراري.

جدول تكراري ← جدول تفرغي

الفئات c	العلامات	التكرار f
0	### ### //	12
1	### /	6
2	### //	7
3	///	3
4	//	2
Σ		30

عدد الغيابات (c)	عدد الموظفين (f)
0	12
1	6
2	7
3	3
4	2
Σ	30

مثال (1.3): (مثال على البيانات المتصلة) البيانات التالية تمثل الأجور اليومية بالريال لعينة مكونة من (50) عامل بإحدى المصانع:

47	36	40	55	75	53	46	43	21	10
66	56	46	35	47	32	52	48	41	30
27	25	57	15	37	22	63	21	61	62
54	42	35	49	39	32	45	31	72	50
65	18	79	23	48	44	32	51	44	42

المطلوب: تلخيص بيانات العينة في جدول تكراري.

الحل: نتبع الخطوات التالية:

1. نحسب المدى (R) والذي يُعرف بأنه الفرق بين أكبر قيمة وأقل قيمة

$$R = \max - \min = 79 - 10 = \boxed{69}$$

2. نوجد عدد الفئات (k)

$$\begin{aligned} k &= 1 + 3.3 \times \log n = 1 + 3.3 \times \log(50) \\ &= 1 + (3.3 \times 1.69897) = 1 + 5.61 = 6.61 \approx \boxed{7} \end{aligned}$$

3. نحدد طول الفئة (h)

$$h = \frac{R}{k} = \frac{69}{7} = 9.86 \approx \boxed{10}$$

ملحوظة: تم استخدام قاعدة (ستيرجس) لإيجاد طول الفئة

$$h = \frac{R}{K} = \frac{R}{1 + 3.3 \log n}$$

4. نكون الجدول التفرغي، بأن نضع الفئات في العمود الأول من الجدول، مع ملاحظة أن الفئة الأولى لا بد أن تبدأ أو تشمل أصغر قيمة، والفئة الأخيرة لا بد أن تشمل أكبر قيمة، ثم نكمل الحل كالمعتاد.

جدول تكراري			جدول تفرغي		
عدد العمال (f)	الأجور (c)		التكرار f	العلامات	الفئات c
3	10 -		3	///	10 -
6	20 -		6	### /	20 -
10	30 -		10	### ###	30 -
15	40 -		15	### ### ###	40 -
8	50 -		8	### ///	50 -
5	60 -		5	###	60 -
3	70 - 80		3	///	70 - 80
50	Σ		50		Σ

ملاحظات هامة:

1. يرمز للفئات بالرمز c وللتكرار بالرمز f وحجم العينة بالرمز n ، ويجب أن يكون مجموع التكرارات مساوي لحجم العينة، أي $\Sigma f = n$
2. في الفئة الأولى "10 -" المقصود بها من 10 إلى أقل من 20، كما أن القيمة 10 تمثل الحد الأدنى للفئة الأولى، وينطبق القول على بقية القيم.
3. طول الفئة هو الفرق بين الحد الأعلى للفئة والحد الأدنى للفئة.

H.W(1)

فيما يلي بيانات درجات ٧٠ طالب في الامتحان النهائي لمقرر مادة الإحصاء

56	65	70	65	55	60	66	70	75	56
60	70	61	67	61	71	67	62	71	66
68	72	57	68	72	59	57	71	69	75
72	62	67	73	58	63	66	73	63	65
58	73	74	76	74	80	81	60	74	58
76	82	77	83	77	85	91	78	94	72
79	64	57	79	55	87	64	88	78	62

المطلوب : تكوين جدول تكراري ذي فئات

كوّن جدول توزيع تكراري لعلامات (30) طالب في امتحان ما كانت كما

46	49	48	58	54	50
40	62	37	48	54	75
54	48	59	45	34	58
47	61	49	44	68	39
63	56	43	57	40	45

المحاضرة الثانية

التمثيل البياني

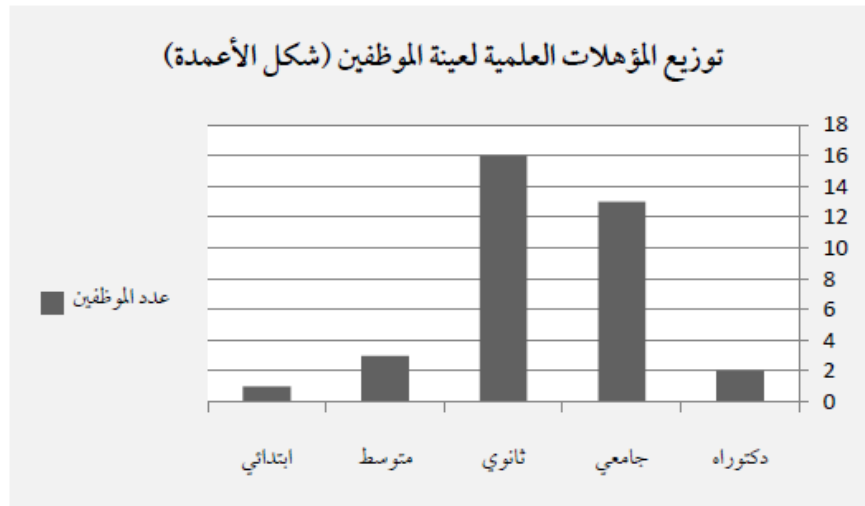
يمكن وصف البيانات النوعية والبيانات الكمية المنفصلة

باستخدام شكل الأعمدة، أما البيانات الكمية يمكن تمثيلها بالمدراج التكراري أو بالمضلع التكراري أو بالمنحنى التكراري، وخطوات تكوين هذه الأشكال على النحو التالي:

1. رسم محورين؛ الأفقي يمثل الفئات C والرأسي يمثل التكرار f .
2. رسم أعمدة طولها يعتمد على عدد التكرارات، وتكون غير متجاورة إذا كانت البيانات نوعية أو كمية منفصلة، ومتجاورة إذا كانت البيانات كمية ، ويسمى الشكل في هذه الحالة بالمدراج التكراري.
3. لتكوين المضلع التكراري نضع نقاط في منتصف أعمدة المدرج التكراري ثم نقوم بوصلها بخطوط مستقيمة، أما المنحنى التكراري فنقوم بوصل النقاط في منتصف الأعمدة يدوياً.

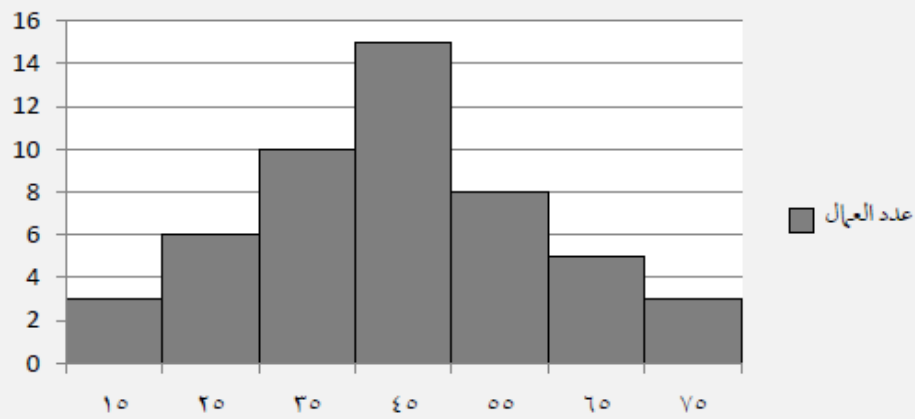
مثال (1.4): مثل التوزيعات التكرارية في الأمثلة السابقة بما يناسبها من أشكال بيانية.

الحل:



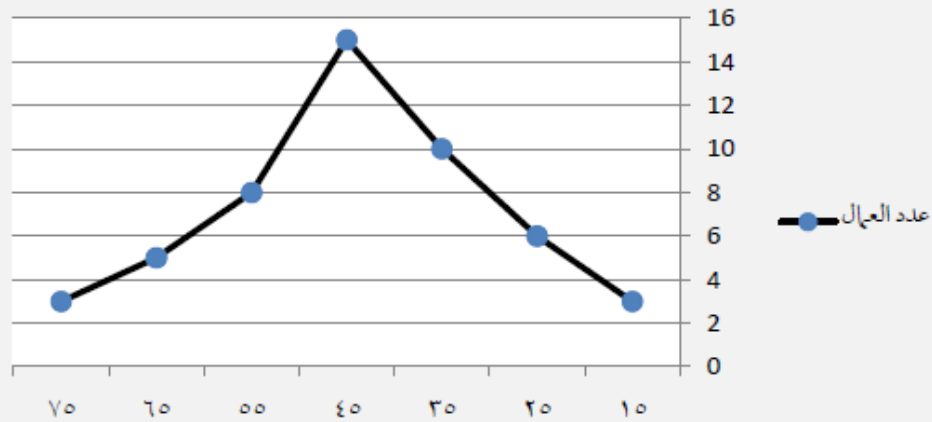
رسم توضيحي 1: توزيع أعداد الموظفين حسب مؤهلاتهم العلمية (مثال 1.1)

توزيع أعداد العمال حسب فئات الأجور (المدرج التكراري)



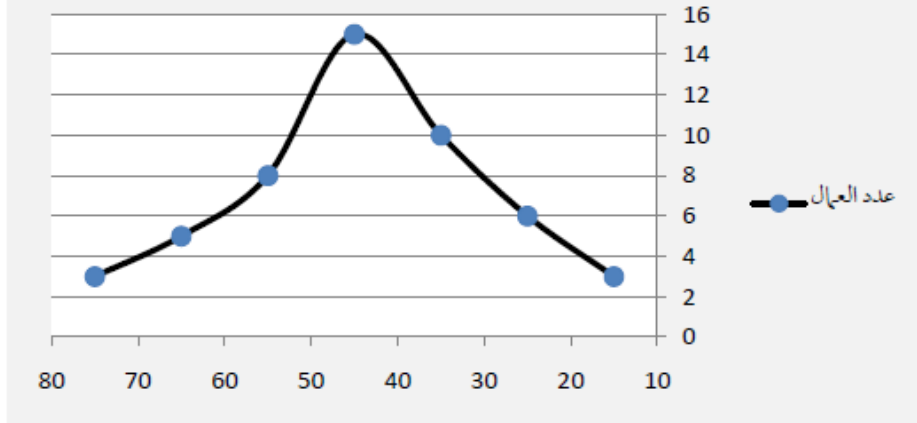
رسم توضيحي 3: توزيع أعداد العمال حسب فئات الأجور (مثال 1.3)

توزيع أعداد العمال حسب فئات الأجور (المضلع التكراري)



رسم توضيحي 4: توزيع أعداد العمال حسب فئات الأجور (مثال 1.3)

توزيع أعداد العمال حسب فئات الأجور (المنحنى التكراري)



رسم توضيحي 5: توزيع أعداد العمال حسب فئات الأجور (مثال 1.3)

1.3. الجداول التكرارية المتجمعة : الجدول التكراري البسيط يعطينا فقط عدد المفردات في كل فئة، ولكننا نحتاج إلى معرفة بيانات تجميعية لا نحصل عليها من الجداول البسيطة مباشرة، ولكن نحصل عليها من الجداول المتجمعة. وهي نوعان:

- **جدول متجمع صاعد :** وهو جدول ذو عمودين، عمود للفئات ويكتب "أقل من الحد الأعلى للفئة"، وعمود للتكرارات ويكتب "تكرار متجمع صاعد" (ت. م. ص.) ونحصل عليه بتجميع التكرارات في الجدول البسيط من بداية الجدول إلى نهايته.
- **(للإطلاع فقط) جدول متجمع نازل :** وهو جدول ذو عمودين، عمود للفئات ويكتب "الحد الأدنى للفئة فأكثر"، وعمود للتكرارات ويكتب "تكرار متجمع نازل" (ت. م. ن.) ونحصل عليه بتجميع التكرارات في الجدول البسيط من نهاية الجدول إلى بدايته.

مثال (1.5): كون الجدول المتجمع الصاعد والنازل لمثال أجور العمال (1.3)

الحل:

جدول متجمع نازل		جدول متجمع صاعد		جدول تكراري	
ت. م. ن.	الحد الأدنى للفئة فأكثر	ت. م. ص.	أقل من الحد الأعلى للفئة	عدد العمال	الأجور
50	10 فأكثر	3	أقل من 20	3	10 -
47	20 فأكثر	9	أقل من 30	6	20 -
41	30 فأكثر	19	أقل من 40	10	30 -
31	40 فأكثر	34	أقل من 50	15	40 -
16	50 فأكثر	42	أقل من 60	8	50 -
8	60 فأكثر	47	أقل من 70	5	60 -
3	70 فأكثر	50	أقل من 80	3	70 - 80
				50	Σ

1.4. التمثيل البياني للتوزيع التكراري المتجمع: يتم تمثيل الجدول المتجمع الصاعد بيانياً بالمنحنى

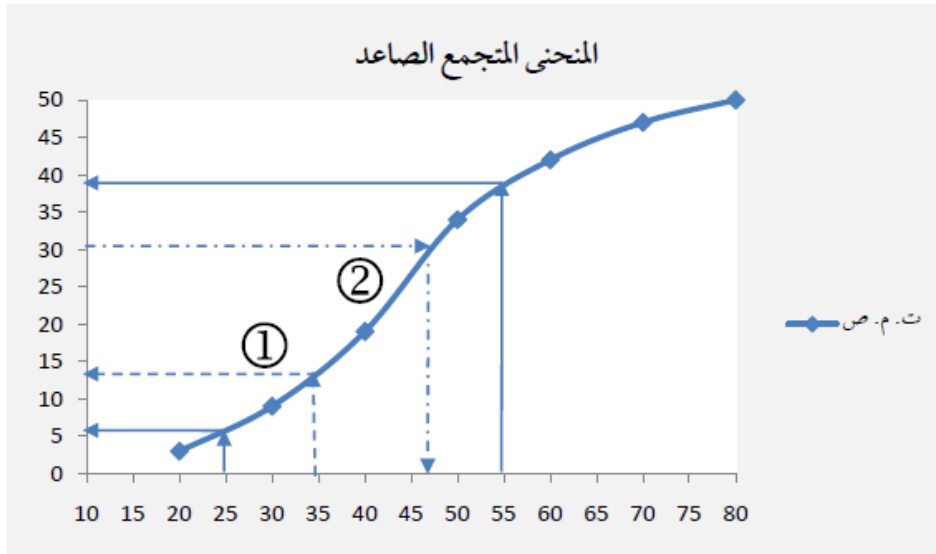
المتجمع الصاعد، كذلك يتم تمثيل الجدول المتجمع النازل بيانياً بالمنحنى المتجمع النازل. وهذه المنحنيات تعطينا البيانات التي لا توفرها الجداول المتجمعة مباشرة، ولإنشاء المنحنى المتجمع، تتبع الخطوات التالية:

1. رسم محورين؛ الأفقي يمثل الحد الأعلى للفئة والرأسي يمثل ت. م. ص. في حالة التكرار المتجمع الصاعد، وفي حالة التكرار المتجمع النازل يمثل المحور الأفقي الحد الأدنى للفئة والرأسي يمثل ت. م. ن.
2. نحدد نقاط مناسبة حسب التكرار المتجمع، ثم نصلها يدوياً.

مثال (1.6): باستخدام الجدول المتجمع الصاعد في المثال (1.5)، ارسم المنحنى المتجمع الصاعد. ومن الرسم أوجد:

1. عدد العمال اللذين يقل أجورهم عن 35 ريال.
2. الحد الأعلى للأجور الذي بلغه 30 عامل.
3. نسبة عدد العمال اللذين تتراوح أجورهم اليومية بين 25 و 55 ريال.

الحل:

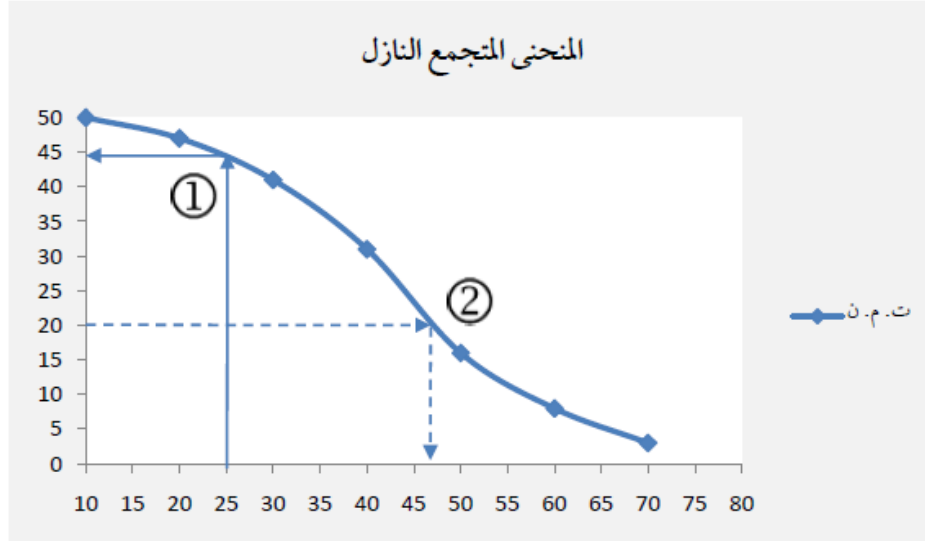


1. عدد العمال اللذين يقل أجورهم عن 35 ريال = 13 عامل
2. الحد الأعلى للأجور الذي بلغه 30 عامل = 46 ريال
3. نسبة عدد العمال اللذين تتراوح أجورهم اليومية بين 25 و 55 ريال =
$$\{ (39 - 6) \div 50 \} \times 100 = 66\%$$

مثال (1.7) (للإطلاع فقط) باستخدام الجدول المتجمع النازل في المثال (1.5)، ارسم المنحنى المتجمع النازل. ومن الرسم أوجد:

1. عدد العمال الذين تصل أجورهم 25 ريال فأكثر.
2. الحد الأدنى للأجور الذي بلغه 20 عامل.

الحل:



1. عدد العمال الذين تصل أجورهم 25 ريال فأكثر = 44 عامل
2. الحد الأدنى للأجور الذي بلغه 20 عامل = 47 ريال

H.W(2)

الجدول التالي يمثل علامات ٣٠ طالب مبوبة في جدول تكراري كما يلي

الفئات	34-39	40-45	46-51	52-57	58-63	64-69	70-75
التكرارات	3	6	8	5	6	1	1

مثل هذا الجدول بكل من الطرق الآتية

١/ المدرج التكراري ، المنحني التكراري

2/ المضلع التكراري ، المنحني التكراري المتجمع الصاعد

المحاضرة الثالثة

المقاييس الوصفية في الإحصاء

مقاييس النزعة المركزية

غالباً ما نلاحظ أن البيانات تميل إلى التركز حول قيمة معينة وفي

هذه الحالة يمكن استخدام هذه "القيمة المركزية" لتمثيل هذه المجموعة من البيانات والمقاييس المستخدمة في التعرف على هذه القيمة المركزية تسمى "مقاييس النزعة المركزية". ويجب أن يتوفر في هذه المقاييس الصفات الآتية لكي يكون المقياس جيداً:

- أن يعتمد المقياس في حسابه على كل المشاهدات.
- أن يكون المقياس سهل الحساب والفهم.
- أن يتوفر فيه القابلية للتعامل الجبري.
- ألا يتأثر بالقيم الشاذة أو المتطرفة.

من بين هذه المقاييس: الوسط الحسابي (أو المتوسط) والوسيط والمنوال ، ولا تتوفر كل الصفات السابق ذكرها في مقياس واحد، ولكن كل مقياس من هذه المقاييس يفضل استخدامه في حالات معينة ولا يفضل استخدامه في حالات أخرى، وفيما يلي مقارنة بين خصائص المتوسطات المذكورة:

الخاصية	المقاييس		
	المتوسط	الوسيط	المنوال
يعتمد المقياس في حسابه على كل المشاهدات	✓	x	x
المقياس سهل الحساب والفهم	✓	✓	✓
يتوفر فيه القابلية للتعامل الجبري	✓	✓	✓
لا يتأثر بالقيم الشاذة أو المتطرفة	x	✓	✓

الوسط الحسابي (Arithmetic Mean)

يعرف الوسط الحسابي بأنه مجموع القيم مقسوماً على عددها n فإذا رمزنا للمتغير الملاحظ

بالرمز x فإننا نرمز للوسط الحسابي بالرمز $\bar{x}$ حيث

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

مثال (2.1): احسب الوسط الحسابي للأجور اليومية بالدولار للعينة التالية المكونة من خمس عمال بإحدى القطاعات:

50 70 80 90 60

الحل:

$$\bar{x} = \frac{\sum x}{n} = \frac{(50+70+80+90+60)}{5} = \frac{350}{5} = \boxed{70\$}$$

حساب الوسط الحسابي من الجداول التكرارية:

أولاً إذا كانت البيانات في شكل جدول تكراري غير مبوب (بسيط) حيث القيم المختلفة

$x_1, x_2, x_3, \dots, x_k$ وتكراراتها كانت $n_1, n_2, n_3, \dots, n_k$

علي التوالي فإن المتوسط يعطي بالعلاقة :

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i}$$

Example

الجدول التالي يعطي أوزان 300 رجل بالكجم اوجد الوسط الحسابي لهذه الأوزان.

مجموع	99	92	85	78	71	64	57	50	43	الوزن
300	9	26	38	51	71	50	34	16	5	التكرار

Solution

نكون صف (عمود) ثالث يتم فيه ضرب الأوزان في التكرارات المقابلة لها

مجموع	99	92	85	78	71	64	57	50	43	الوزن
300	9	26	38	51	71	50	34	16	5	التكرار
21685	891	2392	3230	3978	5041	3500	1938	800	215	$n_i x_i$

ثم نطبق في القاعدة:

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{21685}{300} = 72.283$$

علية فان الوسط الحسابي يساوي 72.28

(ثانياً) حالة البيانات المبوبة: تتبع الخطوات التالية:

1. نضيف للجدول التكراري عمود يمثل مركز الفئة ونرمز له بالرمز X ؛ ولحساب مراكز

الفئات تتبع ما يلي:

• نحسب مركز الفئة الأولى = (الحد الأدنى للفئة الأولى + الحد الأدنى للفئة الثانية) ÷ 2

• نحسب مركز الفئة الثانية = مركز الفئة الأولى + طول الفئة (h)

• نحسب مركز الفئة الثالثة = مركز الفئة الثانية + h ، وهكذا حتى ننتهي من جميع

الفئات.

2. نكون عمود آخر يحتوي على حاصل ضرب عمود التكرارات في عمود مركز الفئة.

3. نستخدم القانون التالي:

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

حيث: $\sum f \cdot x$: حاصل ضرب عمود التكرارات في عمود مركز الفئة

$\sum f$: مجموع التكرارات

مثال (2.2): استخدم الجدول التكراري الخاص بأجور العمال اليومية (مثال 1.3) لإيجاد ما يلي:

- مراكز الفئات المختلفة
- حساب الوسط الحسابي (أو المتوسط)

الحل: حيث أن طول الفئة h يساوي 10، فإن:

(1)	(2)	(3)	(4) = (2) × (3)
فئات الأجور (c)	عدد العمال (f)	مراكز الفئات (x)	$f \cdot x$
10 -	3	$(10+20) \div 2 = 15$	45
20 -	6	$15+h = 15+10 = 25$	150
30 -	10	35	350
40 -	15	45	675
50 -	8	55	440
60 -	5	65	325
70 - 80	3	75	225
Σ	50 Σf		2210 $\Sigma f \cdot x$

$$\bar{x} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{2210}{50} = \boxed{44.2} \text{ ريال}$$

المحاضرة الرابعة

الوسيط (Median)

تعريفه: هو المفردة التي تقسم مفردات العينة بعد ترتيبها تصاعدياً (أو تنازلياً) إلى قسمين متساويين، بحيث يكون عدد المفردات الأصغر منها في القيمة مساوياً لعدد المفردات الأكبر منها في القيمة، ويرمز له بالرمز (m) .
طرق حسابه:

(أولاً) حالة البيانات الغير مبوبة: نتبع الخطوات التالية:

1. نرتب مفردات العينة حسب قيمها إما تصاعدياً أو تنازلياً.
2. إذا كان حجم العينة فردي نستخدم تعريف الوسيط مباشرة، إما إذا كان حجم العينة زوجي فنقوم بإيجاد متوسط قيمة المفردتين التي تتوسط بقية المفردات.

مثال (2.3): احسب وسيط الأجور اليومية بالدولار لعيتتي العاملين بإحدى القطاعات التاليتين:

60	90	80	70	50	: (1) العينة
100	60	90	80	70	: (2) العينة

الحل:

العينة (1): لحساب قيمة الوسيط، نرتب القيم تصاعدياً فتصبح

50 60 70 80 90

من تعريف الوسيط، نجد أن قيمة الوسيط $m = 70S$

العينة (2): لحساب قيمة الوسيط، نرتب المفردات حسب قيمها تصاعدياً فتصبح

50 60 70 80 90 100

نجد أن قيمة الوسيط

$$m = (70 + 80) \div 2 = 75S$$

(ثانياً) حالة البيانات المبوبة: نتبع الخطوات التالية:

1. نحدد ترتيب الوسيط والذي يحسب من العلاقة: $c_1 = \frac{\sum f}{2}$
2. نكون الجدول التكراري المتجمع الصاعد.
3. نحدد فئة الوسيط؛ بالبحث عن الفئتين التي تتراوح قيمة c_1 بين ت. م. ص. الخاص بهما، ثم اختيار الفئة ذات أكبر ت. م. ص. بينهما.
4. نستخدم العلاقة التالية:

$$m = L + \frac{c_1 - c_2}{c_3} \times h$$

حيث:

- L : الحد الأدنى لفئة الوسيط
- c_1 : ترتيب الوسيط
- c_2 : ت. م. ص. السابق لفئة الوسيط
- c_3 : التكرار الأصلي لفئة الوسيط
- h : طول الفئة

مثال (2.4): احسب وسيط أجور العمال اليومية (مثال 1.3).

الحل:

ت. م. ص.	أقل من الحد الأعلى للفئة	عدد العمال	الأجور
3	أقل من 20	3	10 -
9	أقل من 30	6	20 -
$c_2 = 19$	أقل من 40	10	30 -
34	أقل من 50	$c_3 = 15$	$L = 40 -$
42	أقل من 60	8	50 -
47	أقل من 70	5	60 -
50	أقل من 80	3	70 - 80

$c_1 = 25$

$$m = L + \frac{c_1 - c_2}{c_3} \times h = 40 + \frac{25 - 19}{15} \times 10 = \boxed{44} \text{ ريال}$$

المنوال (Mode)

تعريفه: هو قيمة المفردة التي تتكرر أكثر من غيرها، أو هو المفردة ذات القيمة الأكثر شيوعاً أو تكراراً، ويرمز له بالرمز (D) .

ملاحظة: المنوال يمكن استخدامه كذلك في حالة البيانات الوصفية (الغير مبوبة فقط).

طرق حسابه:

(أولاً) حالة البيانات الغير مبوبة: باستخدام التعريف مباشرة.

مثال (2.6): احسب الأجر الشائع لعينتي العاملين بإحدى القطاعات التاليتين:

60	90	50	70	50	: (1) العينة
70	60	50	70	70	: (2) العينة

الحل:

العينة (1): من تعريف المنوال، نجد أن $D = 50$

العينة (2): من تعريف المنوال، نجد أن $D = 70$

(ثانياً) حالة البيانات المبوبة: نتبع الخطوات التالية:

1. نحدد فئة المنوال؛ وهي الفئة ذات أكبر تكرار.

2. نستخدم العلاقة التالية:

$$D = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

L : الحد الأدنى لفئة المنوال

Δ_1 : الفرق بين أكبر تكرار والسابق له

Δ_2 : الفرق بين أكبر تكرار واللاحق له

h : طول الفئة

مثال (2.7): احسب أجر العمال الشائع (مثال 1.3).

الحل

الاجور C	10-	20-	30-	40-	50-	60-	70-80
العمال F	3	6	10	15	8	5	3

$$D = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 40 + \frac{(15-10)}{(15-10)+(15-8)} \times 10 = 40 + \frac{5}{5+7} \times 10 = \boxed{44.17} \text{ ريال}$$

ملاحظة:

1. درسنا الوسط الحسابي والوسيط والمنوال . نجد أن الوسط الحسابي أدق المتوسطات الثلاثة لأنه يأخذ في الاعتبار جميع فئات التوزيع ، يليه الوسيط ، وأخيرا المنوال لأن طرق حسابه تقريبية ، وعلى ذلك إذا ذكر لفظ المتوسط دون تحديد فيقصد به الوسط الحسابي .
2. ذكر لفظ "الشائع" ، مثل الطول الشائع ، العمر الشائع الخ ، يقصد به المنوال .
3. عادة ما تكون قيم المتوسطات قريبة من بعضها البعض .

H.W(3)

1/ أوجد الوسط الحسابي – والوسيط – والمنوال لكل من مجموعات الأعداد الآتية ؟

1/ 2، 6، 8، 10، 100

2/ 4، 5، 3، 3، 2، 2، 2

3/ 8، 12، 4، 4، 4، 14، 14، 14

2/ أوجد الوسط الحسابي – والوسيط – والمنوال للتكراري التالي ؟

فئات	4-7	8-11	12-15	16-19	20-23	24-27	مجموع
تكرارات	3	8	14	18	11	6	60

3/ أوجد الوسط الحسابي والمنوال للتكراري ؟

مفردة	4	5	7	9	12	15	22	مجموع
تكرار	5	16	4	10	15	18	2	70

المحاضرة الخامسة

مقاييس التشتت

مقاييس التشتت Measure of Variation

درسنا في الفقرة السابقة مقاييس النزعة المركزية ، ولكن غالباً ما تكون هذه المقاييس غير كافية لتمثل الواقع بشكل كامل أو لمقارنة مجموعتين من المشاهدات أو أكثر . إذا اعتبرنا المجموعتين التاليتين من القيم :

المجموعة الأولى : 5 , 6 , 8 , 10 , 12 , 14 , 15

المجموعة الثانية : 1 , 2 , 5 , 10 , 15 , 18 , 19

لوجدنا أن الوسط الحسابي لكل مجموعة هو 10 كما أن الوسيط هو نفسه للمجموعتين ويساوي 10 أيضاً ومع ذلك فهناك فرق بين المجموعتين حيث تختلف مفردات المجموعة الأولى عن مفردات المجموعة الثانية ، كما أن قيم المجموعة الثانية موزعة على مدى أوسع من المجموعة الأولى ويمكن أن نقول إن تشتت المجموعة الثانية أكبر منه في المجموعة الأولى . ويمكن قياس درجة التشتت بعدة مقاييس منها : المدى ، الانحراف عن المتوسط، التباين، والانحراف المعياري .

المدى : Range

المدى هو الفرق بين أكبر قيمة وأصغر قيمة في مجموعة ما، فمدى المجموعة الأولى هو $15-5=10$.

بينما مدى المجموعة الثانية $19-1=18$ لاحظ أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى . أما المدى لقيم معطاة في جدول توزيع تكراري فيحسب من الفرق بين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا . ففي الجدول (5-3) المدى هو $70 = 75 - 5$.

الانحراف المتوسط: Mean deviation

يعتبر الانحراف عن المتوسط أو (الانحراف المتوسط) أحد مقاييس التشتت . ويعرف الانحراف المتوسط بأنه متوسط الفروق للبيانات عن وسطها الحسابي بقيمها المطلقة فإذا كانت لدينا مجموعة البيانات

$$x_1, x_2, \dots, x_n$$

وسطها الحسابي $\bar{x}$ فإنه يمكن حساب الانحراف المتوسط بالعلاقة التالية :

$$M_D = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

أما من أجل قيم معطاة في جدول توزيع تكراري فإن الانحراف المتوسط يعطى بالعلاقة التالية:

$$M_D = \frac{1}{n} \sum_{i=1}^k |x_i - \bar{x}| f_i$$

حيث :

$$n = \sum_{i=1}^k f_i \text{ و } f_i \text{ تكرار الفئة ،}$$

لنوجد على سبيل المثال الانحراف المتوسط لمجموعة البيانات التالية :

$$5, 6, 8, 10, 12, 14, 15$$

إن المتوسط لهذه القيم $\bar{x} = 10$ فيكون الانحراف المتوسط

$$M_D = \frac{1}{7} [|5-10| + |6-10| + |8-10| + |10-10| + |12-10| + |14-10| + |15-10|] = 22/7$$

وبنفس الطريقة نجد أن الانحراف المتوسط للمجموعة الأخرى

1,2,5,10,15,18,19 هو 44/7

لاحظ من هذين المثالين أنه كلما كان الانحراف المتوسط كبيراً كلما كان التباعد بين القيم كبيراً وكلما كان صغيراً كانت القيم متقاربة .

الانحراف المعياري: Standard Deviation

تعريفه: هو أدق مقاييس التشتت وأكثرها استخداماً، ويعرف بأنه الجذر التربيعي لمتوسط مربعات

انحرافات (فروق) القيم عن وسطها الحسابي، ويرمز له بالرمز (s_x) .

نجد أن متوسط مربعات انحرافات (فروق) القيم عن وسطها الحسابي يسمى التباين ، أي أن الانحراف المعياري هو جزر التباين .

وقيمة التباين لابد أن تكون موجبة أو تساوي الصفر

وكلما إقتربت قيمة التباين من الصفر ، أي كلما إقتربت قيمة الانحراف المعياري من الصفر كلما أصبحت البيانات قريبة من التجانس .

طرق حسابه:

(أولاً) حالة البيانات الغير مبوبة: نستخدم القانون التالي:

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

حيث: $\sum x^2$: مجموع مربعات قيم مفردات العينة n : حجم العينة

مثال (2.9): احسب الانحراف المعياري للأجور اليومية بالدولار للعينة التالية المكونة من خمس عمال

بإحدى القطاعات: 60 90 80 70 50

الحل:

$$\bar{x} = \frac{350}{5} = \boxed{70\$}$$

$$\frac{\sum x^2}{n} = \frac{60^2 + 90^2 + 80^2 + 70^2 + 50^2}{5} = 5100$$

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{5100 - 4900} = \sqrt{200} = \boxed{14.14\$}$$

(ثانياً) حالة البيانات المبوبة: نتبع الخطوات التالية:

1. نتبع خطوات حساب المتوسط الحسابي.
2. نضيف عمود جديد هو حاصل ضرب عمود مركز الفئة (عمود 3) في عمود حاصل ضرب عمود التكرارات في عمود مركز الفئة (عمود 4).
3. نستخدم القانون التالي:

$$s_x = \sqrt{\frac{\sum f \cdot x^2}{\sum f} - \bar{x}^2}$$

مثال (2.10): اوجد الانحراف المعياري لأجور العمال (مثال 1.3).

الحل: من مثال (2.2) نعلم أن:

$$\bar{x} = \frac{\sum f \cdot x}{\sum f} = \frac{2210}{50} = \boxed{44.2}$$

(1)	(2)	(3)	(4)	(5) = (3) × (4)
الأجور (c)	عدد العمال (f)	x	f · x	f · x ²
10 -	3	15	45	675
20 -	6	25	150	3750
30 -	10	35	350	12250
40 -	15	45	675	30375
50 -	8	55	440	24200
60 -	5	65	325	21125
70 - 80	3	75	225	16875
Σ	50 = Σ f		2210 = Σ f · x	109250 = Σ f · x²

$$s_x = \sqrt{\frac{\sum f \cdot x^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{109250}{50} - 44.2^2} = \sqrt{2185 - 1953.64} = \sqrt{231.36} = \boxed{15.21} \text{ ريال}$$

المحاضرة السادسة

2.2.2. معامل الاختلاف

لمقارنة تشتت مجموعتين أو أكثر من البيانات ، فإننا لا نستخدم مقياس التشتت السابق ذكرها مباشرة ، وذلك بسبب:

- قد تكون البيانات ذات وحدات قياس مختلفة.
- قد تكون الأوساط الحسابية للظاهرتين مختلفة، لذلك نستخدم ما يسمى بمقياس التشتت النسبي، وأهم هذه المقاييس هو ما يعرف بمعامل الاختلاف.

تعريف معامل الاختلاف: هو معامل نسبي يستخدم للمقارنة بين تشتت ظاهرتين أو أكثر مختلفتين أو حتى متشابهتين في وحدة القياس. والظاهرة التي معامل اختلافها أكبر تكون أكثر تشتتاً من الأخرى. ويستخدم في حسابه العلاقة الآتية:

$$cv.(x) = \frac{s_x}{\bar{x}} \times 100$$

مثال (2.11): اوجد معامل الاختلاف لأجور العمال (مثال 1.3).

الحل: من مثال (2.2) و (2.10) نعلم أن:

$$\bar{x} = 44.2$$

$$s_x = 15.21$$

$$cv.(x) = \frac{s_x}{\bar{x}} \times 100 = \frac{15.21}{44.2} \times 100 = 34.41\%$$

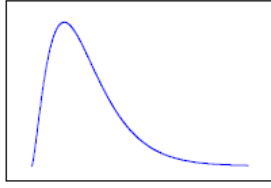
2.3. الالتواء

الالتواء: هو بعد المنحنى التكراري عن التماثل. ويقصد بالتماثل أنه إذا أسقطنا عموداً من قمة المنحنى التكراري وقسمه إلى قسمين منطبقين يكون التوزيع متماثلاً. والعكس فيكون التوزيع غير متماثل أي ملتو إما إلى جهة اليمين أو إلى جهة اليسار.

ويقاس الالتواء بعدة مقاييس، منها:

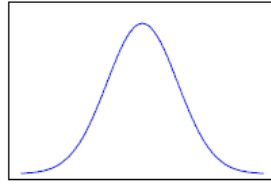
$$s.k.(I) = \frac{\bar{x} - D}{s_x}$$

$$s.k.(II) = \frac{3(\bar{x} - m)}{s_x}$$



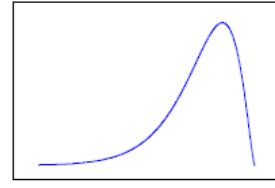
التوزيع غير متماثل
وملتو من جهة اليمين

$$\bar{x} \neq m \neq D$$



التوزيع متماثل

$$\bar{x} = m = D$$



التوزيع غير متماثل
وملتو من جهة اليسار

$$\bar{x} \neq m \neq D$$

ودلالة قيمة أي من المعاملين على النحو التالي:

$$s.k. \begin{cases} < 0 \quad (-) & \text{التوزيع ملتو لليسار} \\ = 0 & \text{التوزيع متماثل} \\ > 0 \quad (+) & \text{التوزيع ملتو لليمين} \end{cases}$$

مثال (2.12): ادرس التواء التوزيع التكراري لأجور العمال (مثال 1.3)، باستخدام:

الحل:

$$s.k.(I) = \frac{\bar{x} - D}{s_x} = \frac{44.2 - 44.17}{15.21} = 0.002$$

$$s.k.(II) = \frac{3(\bar{x} - m)}{s_x} = \frac{3(44.2 - 44)}{15.21} = 0.04$$

ملاحظة:

عادة ما نحصل علي نتائج مختلفة لمعامل التواء، وهذا يناقض بعضه إذا أن كل معامل يقيس الالتواء علي أساس يخالف المعاملات الاخرى.

عند مقارنة إلتواء توزيعات مختلفة يجب استخدام نفس المعامل عند المقارنة.

امثلة :

2.4.1. الجدول التالي يمثل التوزيع التكراري لفئات الدخل الشهري لعينة من الأسر (مئات الريالات)

بإحدى المدن:

فئات الدخل	62 -	66 -	70 -	74 -	78 -	82 -	86 - 90
عدد الأسر	3	8	20	21	14	10	4

المطلوب:

1. حساب الوسط الحسابي، والانحراف المعياري.

2. حساب الوسيط والمتوال.

3. دراسة إلتواء التوزيع التكراري.

الحل:

فئات الدخل	عدد الأسر	x	$f \cdot x$	$f \cdot x^2$
62 -	3	64	192	12288
66 -	8	68	544	36992
70 -	20	72	1440	103680
74 -	21	76	1596	121296
78 -	14	80	1120	89600
82 -	10	84	840	70560
86 - 90	4	88	352	30976
Σ	80		6084	465392

1.

$$\bar{x} = \frac{\sum f \cdot x}{\sum f} = \frac{6084}{80} = 76.05 \quad \text{مئات الريالات}$$

$$s_x = \sqrt{\frac{\sum f \cdot x^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{465392}{80} - 76.05^2} = \sqrt{33.8} = 5.8 \quad \text{مئات الريالات}$$

2.

ت. م. ص.	أقل من الحد الأعلى للفئة	عدد الأسر	فئات الدخل
3	أقل من 66	3	62 -
11	أقل من 70	8	66 -
31	أقل من 74	20	70 -
52	أقل من 78	21	74 -
66	أقل من 82	14	78 -
76	أقل من 86	10	82 -
80	أقل من 90	4	86 - 90
		80	Σ

$$m = L + \frac{c_1 - c_2}{c_3} \times h = 74 + \frac{40 - 31}{21} \times 4 = \boxed{75.71} \quad \text{مئات الريالات}$$

$$D = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = 74 + \frac{1}{1 + 7} \times 4 = 74 + \frac{1}{8} \times 4 = \boxed{74.5} \quad \text{مئات الريالات}$$

٣- نستخدم المنوال ل الإلتواء

$$s.k.(I) = \frac{\bar{x} - D}{s_x} = \frac{76.05 - 74.5}{5.8} = 0.27$$

التوزيع غير متماثل وملتو الليمين.

2.4.2. إذا كانت y تمثل الاستهلاك الأسبوعي الشخصي بالريال، x تمثل الاستهلاك الأسبوعي لوقود

السيارات (باللتر)، استخدم المعلومات التالية لتحديد أي الظاهرتين أكثر تشتتاً.

$$\bar{y} = 75 \text{ ريال} \quad , \quad s_y = 15 \text{ ريال}$$

• التوزيع التكراري للاستهلاك الأسبوعي لوقود السيارات:

فئات الاستهلاك	18 -	28 -	38 -	48 -	58 -
عدد المستهلكين	26	59	105	79	31

الحل: نقوم بحساب معامل الاختلاف للمتغير (أو الظاهرة) x ومن ثم نحسب معامل الاختلاف

للمتغير (أو الظاهرة) y ثم نقارن بينهما.

فئات الاستهلاك	عدد السيارات	x	fx	fx^2
18 -	26	23	598	13754
28 -	59	33	1947	64251
38 -	105	43	4515	194145
48 -	79	53	4187	221911
58 - 68	31	63	1953	123039
Σ	300		13200	617100

$$\bar{x} = \frac{\sum f \cdot x}{\sum f} = \frac{13200}{300} = \boxed{44} \text{ لتر}$$

$$s_x = \sqrt{\frac{\sum f \cdot x^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{617100}{300} - 44^2} = \sqrt{121} = \boxed{11} \text{ لتر}$$

$$\therefore cv.(x) = \frac{s_x}{\bar{x}} \times 100 = \frac{11}{44} \times 100 = 25\%$$

كما أن:

$$\therefore cv.(y) = \frac{s_y}{\bar{y}} \times 100 = \frac{15}{75} \times 100 = 20\%$$

لذا يمكن القول أن ظاهرة الاستهلاك الأسبوعي لوقود السيارات أكثر تشتتاً من ظاهرة الاستهلاك

الاسبوعي الشخصي

2.5. تمارين

2.5.1. الجدول الآتي يوضح توزيع عينة من (100) موظف حسب فئات الزيادة التي حصلوا عليها في الراتب (بعشرات الريالات):

فئات الزيادة	30 -	40 -	50 -	60 -	70 -	80 -	90 -
عدد الموظفين	4	11	20	36	17	8	4

1. ارسم المدرج التكراري للتوزيع، ثم حدد منه قيمة المتوال.
2. ارسم المنحنى المتجمع الصاعد ومن الرسم أوجد:
 - أ. عدد الموظفين الذين حصلوا على الزيادة أقل من 750 ريال.
 - ب. الحد الأعلى للزيادة التي حصل عليها 30 موظف.
 - ج. وسيط الزيادة في الراتب.

2.5.2. الجدول الآتي يوضح توزيع عينة من (200) موظف بإحدى الوزارات حسب أعمارهم بالسنة.

فئات العمر	20 -	25 -	30 -	35 -	40 -	45 -	50 -	55 -
عدد الموظفين	10	17	24	43	34	30	23	19

1. احسب الوسط الحسابي والانحراف المعياري.
2. ارسم المنحنى المتجمع الصاعد ومن الرسم أوجد:
 - أ. نسبة عدد المشتغلين الذين يقل عمرهم عن 42 سنة.
 - ب. الحد الأعلى للعمر الذي بلغه 120 موظف.
 - ج. وسيط العمر.
3. ارسم المنحنى التكراري، هل المنحنى متماثل؟ دلل على إجابتك بحساب معامل الالتواء.

2.5.3. الجدول الآتي يوضح توزيع عينة من الموظفين بإحدى شركات القطاع الخاص حسب فئات الراتب (آلاف الريالات).

فئات الراتب	62 -	66 -	70 -	74 -	78 -	82 -	86 -
عدد الموظفين	3	8	20	21	14	10	4

1. احسب كل من الوسط الحسابي والانحراف المعياري.

2. الوسيط (بالحساب و بالرسم).

3. المتوال (بالحساب و بالرسم).

4. معامل الاختلاف.

5. معاملي الالتواء.

2.5.4. الجدول الآتي يوضح توزيع الدخل السنوي (بآلاف الريالات) لعينة من الأسر بإحدى المدن:

فئات الدخل	40-	48-	56-	64-	72-	80-	88-
عدد الأسر	8	16	24	36	30	18	8

1. الانحراف المعياري للتوزيع.

2. الوسيط.

3. الدخل الشائع لهذه الأسر.

4. معامل الاختلاف.

5. هل هذا التوزيع متماثل؟ علك إجابتك.

2.5.5. الجدول الآتي يوضح توزيع عينة من (500) موظف حسب الساعات الإضافية التي حققها

أسبوعياً بإحدى الشركات:

الساعات الإضافية	2 -	3 -	4 -	5 -	6 -	7 -	8 -	9 -
عدد الموظفين	98	118	101	95	50	20	10	8

احسب المتوسط والتباين لعدد الساعات الإضافية، ثم احسب معامل الاختلاف.

2.5.6. البيانات التالية تمثل أسعار منتجات معينة (بالريال) مختارة من ثلاثة مراكز تجارية مختلفة:

عينة (1): 5 6 3 1 2 7

عينة (2): 6 3 1 2 4 8

عينة (3): 3 1 5 2 4 1

احسب معامل الاختلاف لكل عينة، أي العينات أكثر تشتتاً؟

2.5.7. الجدول الآتي يوضح توزيع عينة من المرضى بمرض معين بإحدى المستشفيات حسب الساعات التي قضوها حتى تماثلوا للشفاء:

الساعات	15 -	19 -	23 -	27 -	31 -
عدد المرضى	6	14	42	10	8

1. ارسم المدرج و المضلع التكراري للتوزيع.
2. احسب الوسط الحسابي والانحراف المعياري للتوزيع.
3. أوجد منوال التوزيع.
4. احسب معامل الاختلاف.
5. هل هذا التوزيع متماثل؟ علل إجابتك.

2.5.8. الجدول الآتي يوضح توزيع أسعار عينة من العقارات بمدينة جدة حسب سعرها (بمئات الآلاف ريال)

فئات الأسعار	15 -	17 -	19 -	21 -	23 -
عدد العقارات	3	7	21	5	4

1. احسب الانحراف المعياري للتوزيع.
 2. احسب وسيط التوزيع، واستخدمه في دراسة تماثل التوزيع.
- إذا كان الوسط الحسابي والانحراف المعياري لمساحة نفس العقارات هو 600، 150 (متر مربع) على التوالي. فأيهما أكثر تشتتاً؟ السعر أم المساحة.

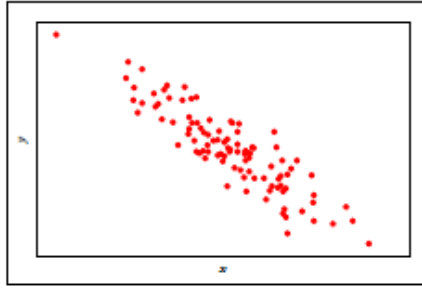
المحاضرة السابعة

مبادئ الارتباط والانحدار الخطي

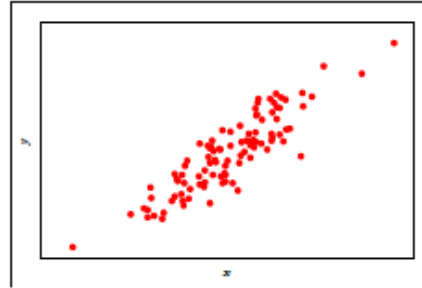
3.1. الارتباط

الارتباط: هو علاقة بين متغيرين (x, y) بمعنى أنه إذا تغير أحد المتغيرين فإن الآخر قد يتبعه. إما في نفس الاتجاه فيكون الارتباط طردي، كعلاقة الصادرات بالميزان التجاري (رسم توضيحي 6). أو في الاتجاه المضاد فيكون الارتباط عكسي، كعلاقة الواردات بالميزان التجاري (رسم توضيحي 7). ويقال أن المتغيرين مستقلين عندما ينعدم الارتباط، كعلاقة دخل الفرد بوزنه (رسم توضيحي 8).

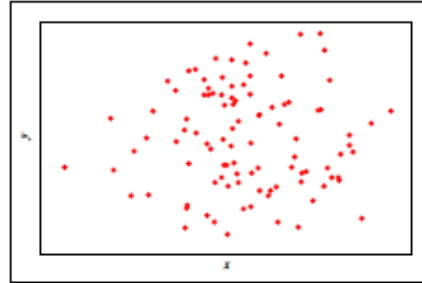
ملاحظة: الرسوم التوضيحية 6، 7، 8 تسمى أشكال الانتشار.



رسم توضيحي 7: الارتباط العكسي



رسم توضيحي 6: الارتباط الطردي



رسم توضيحي 8: لا يوجد ارتباط (استقلال)

معامل الارتباط: والذي يرمز له بالرمز r ، عبارة عن مقياس كمي نسبي يقيس قوة الارتباط بين متغيرين، حيث تتراوح قيمته بين 1 و -1، أي أن $-1 \leq r \leq 1$. يقال أن الارتباط طردي تام إذا كان معامل الارتباط $r = +1$ ، ويقال أن الارتباط عكسي تام إذا كان معامل الارتباط $r = -1$. يقال عن المتغيرين أنها مستقلان إذا كان $r = 0$.

ملاحظة: كلما اقتربت قيمة معامل الارتباط إلى 1 كلما كان الارتباط الطردي قوياً بين الظاهرتين (أو المتغيرين) وكلما اقتربت قيمته إلى الصفر كلما كان الارتباط الطردي ضعيفاً. ونفس القول ينطبق على الارتباط العكسي.

3.1.1. معامل الارتباط الخطي

يقيس العلاقة بين متغيرين كميين، ويسمى أيضاً بمعامل ارتباط بيرسون، ويحسب من العلاقة التالية:

$$r_p = \frac{\frac{\sum xy}{n} - \bar{x} \bar{y}}{s_x s_y}$$

حيث:

$\sum xy$	:	مجموع حاصل ضرب x في y
$\bar{x}$	:	متوسط المتغير (أو الظاهرة) x
$\bar{y}$	:	متوسط المتغير (أو الظاهرة) y
s_x	:	الانحراف المعياري للمتغير (أو الظاهرة) x
s_y	:	الانحراف المعياري للمتغير (أو الظاهرة) y

ملاحظة: قيمة معامل الارتباط الخطي التي تساوي الصفر تدل على عدم وجود علاقة ارتباط خطية فقط بين المتغيرين محل الدراسة.

مثال (3.1): لدراسة علاقة الصادرات بالميزان التجاري خلال عدة سنوات، أخذنا عشر قراءات تقريبية لقيمة صادرات المملكة العربية السعودية (x) وقيمة الميزان التجاري (y) بعشرات المليارات ريال كما يلي:

y	1	3	8	7	6	5	7	8	12	12
x	9	11	17	18	19	16	16	19	23	23

مصدر البيانات الأصلية: موقع مصلحة الإحصاءات العامة على شبكة الإنترنت

هل توجد علاقة ارتباط خطية؟ ما نوعها وما مدى قوتها؟

الحل:

x	y	xy	x ²	y ²
9	1	9	81	1
11	3	33	121	9
17	8	136	289	64
18	7	126	324	49
19	6	114	361	36
16	5	80	256	25
16	7	112	256	49
19	8	152	361	64
23	12	276	529	144
23	12	276	529	144
Σ	171	69	1314	585
	= Σx	= Σy	= Σxy	= Σx ² = Σy ²

$$\bar{x} = \frac{\sum x}{n} = \frac{171}{10} = 17.1, \quad \bar{y} = \frac{\sum y}{n} = \frac{69}{10} = 6.9$$

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{3107}{10} - (17.1)^2} = \sqrt{310.7 - 292.41} = \sqrt{18.29}$$

$$s_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{585}{10} - (6.9)^2} = \sqrt{58.5 - 47.61} = \sqrt{10.89}$$

$$r_p = \frac{\frac{\sum xy}{n} - \bar{x} \bar{y}}{s_x s_y} = \frac{\frac{1314}{10} - (17.1)(6.9)}{\sqrt{(18.29)(10.89)}} = \frac{131.4 - 117.99}{14.11} = \frac{13.41}{14.11} \approx 0.95$$

من الملاحظ أن علاقة الارتباط الخطي بين قيمة صادرات المملكة العربية السعودية وقيمة الميزان التجاري موجودة وهي علاقة ارتباط طردية قوية.

مثال (3.2): سُجلت ست قراءات تقريبية لحجم الإنتاج وحجم صادرات النفط الخام بالمملكة العربية السعودية (بالمليار برميل) خلال عدة سنوات كما يلي:

حجم الصادرات (y)	2	2	2	1	1	1
حجم الإنتاج (x)	3	4	2	2	2	2

مصدر البيانات الأصلية: موقع وزارة البترول والثروة المعدنية على شبكة الإنترنت

ادرس وجود علاقة ارتباط خطية بين حجم الإنتاج وحجم صادرات النفط الخام.

الحل:

	x	y	xy	x^2	y^2
	3	2	6	9	4
	4	2	8	16	4
	2	2	4	4	4
	2	1	2	4	1
	2	1	2	4	1
	2	1	2	4	1
Σ	15	9	24	41	15
	$= \Sigma x$	$= \Sigma y$	$= \Sigma xy$	$= \Sigma x^2$	$= \Sigma y^2$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{6} = 2.5, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{9}{6} = 1.5$$

$$s_x = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{41}{6} - (2.5)^2} = \sqrt{6.83 - 6.25} = \sqrt{0.58}$$

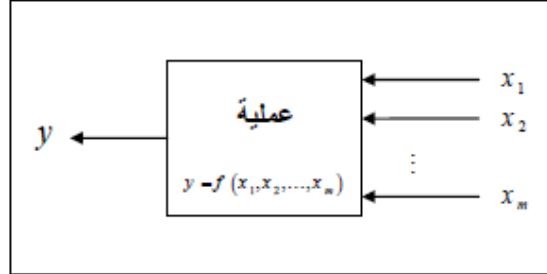
$$s_y = \sqrt{\frac{\Sigma y^2}{n} - \bar{y}^2} = \sqrt{\frac{15}{6} - (1.5)^2} = \sqrt{2.5 - 2.25} = \sqrt{0.25}$$

$$r_p = \frac{\frac{\Sigma xy}{n} - \bar{x} \bar{y}}{s_x s_y} = \frac{\frac{24}{6} - (2.5)(1.5)}{\sqrt{(0.58)(0.25)}} = \frac{4 - 3.75}{\sqrt{0.145}} = \frac{0.25}{0.38} \approx 0.66$$

من الملاحظ أن علاقة الارتباط الخطي بين حجم الإنتاج وحجم صادرات النفط الخام علاقة طردية متوسطة.

3.2. تحليل الانحدار الخطي البسيط

تحليل الانحدار: عبارة عن أسلوب إحصائي يقوم بصياغة دالة رياضية لعملية ذات عوامل مؤثرة عدة $x_1, x_2, \dots, x_m$ لوصف متغير ناتج y من هذه العملية والتحكم به وتوقع قيم غير معروفة له. ويكون شكل الدالة الرياضية تكون على الصورة $y = f(x_1, x_2, \dots, x_m)$



تسمى المتغيرات $x_1, x_2, \dots, x_m$ بالمتغيرات المستقلة، والمتغير الناتج y بالمتغير التابع. تُعرف الدالة السابقة بدالة الانحدار، وأبسط حالة لهذه الدالة عندما يكون للعملية متغير مستقل واحد فقط يرتبط مع المتغير التابع بعلاقة خط مستقيم، أي أن: $\hat{y} = b_0 + b_1x$ (*)

حيث:

b_0 : الجزء المقطوع من محور y ميل الخط المستقيم أو معامل انحدار x على y (أو y/x)

وتحسب القيمتين b_0 و b_1 من العلاقتين التاليتين:

$$b_1 = \frac{\sum xy - \bar{x} \bar{y}}{s_x^2} \quad b_0 = \bar{y} - b_1 \bar{x}$$

ولإيجاد أي قيمة مقدرة جديدة $\hat{y}_h$ ، نعوض بقيمة معلومة للمتغير المستقل ولتكن x_h في المعادلة (*).

مثال (3.5): لدراسة علاقة الاستهلاك المحلي (y) بالإنتاج (x) لمادة الإسفلت (بالمليون برميل) خلال

عدة سنوات، أخذنا عشر قراءات تقريبية كما يلي:

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

أوجد معادلة الانحدار الخطي البسيط، وتوقع قيمة الاستهلاك المحلي عندما يصل إنتاج 16000000 برميل.

الحل:

x	y	xy	x^2
10	6	60	100
13	8	104	169
15	9	135	225
14	8	112	196
9	7	63	81
7	6	42	49
6	5	30	36
6	6	36	36
5	5	25	25
5	5	25	25
Σ	90	65	632
	$= \Sigma x$	$= \Sigma y$	$= \Sigma xy$
			$= \Sigma x^2$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{90}{10} = 9, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{65}{10} = 6.5$$

$$s_x^2 = \frac{\Sigma x^2}{n} - \bar{x}^2 = \frac{942}{10} - (9)^2 = 94.2 - 81 = 13.2$$

$$b_1 = \frac{\frac{\Sigma xy}{n} - \bar{x} \bar{y}}{s_x^2} = \frac{\frac{632}{10} - (9)(6.5)}{13.2} = \frac{63.2 - 58.5}{13.2} \approx 0.36$$

$$b_0 = \bar{y} - b_1 \bar{x} = 6.5 - (0.36)9 = 6.5 - 3.24 \approx 3.26$$

∴ معادلة خط الانحدار البسيط في هذه الحالة: $\hat{y} = 3.26 + 0.36x$ ، ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج 16000000 برميل، نحول وحدة هذه القيمة من برميل إلى مليون برميل بالقسمة على مليون أي أن القيمة المستخدمة في توقع الاستهلاك هي $x_h = 16$ ، وبالتعويض في المعادلة السابقة نجد أن:

$$\hat{y}_h = b_0 + b_1 x_h = \hat{y} = 3.26 + 0.36(16) = 9.02$$

أي أن الاستهلاك المحلي قد يصل إلى 9.02 مليون برميل، أي ما يعادل 9020000 برميل خلال السنة.

H.W

3.4.1. لدراسة العلاقة بين الدخل (x) والاستهلاك (y) بمئات الريالات في مدينته ما، أخذت عينة من الأسر فأعطت النتائج الآتية:

x	5	4	5	6	9	10	9	12	11	9
y	5	4	5	5	8	6	8	11	10	8

1. احسب معامل ارتباط بيرسون بين الظاهرتين.

3.4.2. البيانات الآتية توضح المبالغ المنصرفة على الدعاية (x) (بالآلاف الريالات) لإحدى المؤسسات في عدة مناطق وحجم المبيعات (y) (بالآلاف الريالات) في تلك المناطق:

x	5	3	3	7	6	3	1
y	20	15	10	25	12	18	5

1. احسب معامل ارتباط بيرسون بين الظاهرتين علماً بأن:

$$\sum y^2 = 1843, \sum xy = 481$$

2. قدر حجم المبيعات عندما يصل المنصرف على الدعاية 4500 ريال.

3.4.3. البيانات الآتية تمثل التكلفة الحدية (x) لإنتاج السلعة بالريال وإجمالي المنتج السنوي (y) بالمليون طن لثمانية مؤسسات:

x	5	3	3	7	6	3	1
y	20	15	10	25	12	18	5

1. احسب معامل ارتباط بيرسون بين الظاهرتين.

2. قدر حجم المبيعات عندما يصل المنصرف على الدعاية 4500 ريالاً.