الفصل الأول

التوزيع الوصفي لجملة من القياسات

(١ ـ ١) اختزال بيان إحصائي وجدول التوزيع التكراري

إن البيانات التي نحصل عليها عند القيام بتنفيذ تجربة أو جمع معلومات إحصائية هي قياسات عددية (كمية) أو وصفية. ومهما أوتينا من الدقة وحسن التتبع فلن يقدم لنا استعراض وتأمل هذه القياسات بطريقة مباشرة وبسيطة، وفي بيانات كبيرة الحجم، إلا القليل جدا عن مدلول هذه القياسات وتفسيرها، وتغيرها بالنسبة لبعضها البعض، ومدى هذا التغيّر. وفي الغالب تبرز فوائد جمة من تصنيف القياسات فيما سنسميه توزيعات تكرارية. وهي تسمح لنا بفهم خصائص وصفية وكمية للبيان الإحصائي، وتفسيره، والحكم عليه بطريقة أكثر موضوعية وأسهل تناولا.

مثال (١ ـ ١)

سألنا عشرة من طلاب الأول ثانوي: «هل ستختار الفرع العلمي أو الفرع الأدبي في العام القادم؟»

وكانت الأجوبة كما يلي:

علمي، علمي، أدبي، علمي، أدبي، أدبي، علمي، علمي، أدبي، علمي

ونلاحظ أن الاختيار «علمي» يظهر ست مرات، أي أن تواتر أو تكرار ظهوره هو 6، بينما يتكرر ظهور الاختيار «أدبي» 4 مرات. ويمكننا ترتيب هذه المعلومات في جدول على الشكل التالي:

جدول (١ ـ ١)

الاختيار	علمي	أدبي
التكرار f	6	4

ونرمز للتكرار بالحرف f.

ويسمى هذا الترتيب للمعلومات التي جمعناها توزيعا تكراريا. فهو يوضح كيف تتوزع الأجوبة العشرة بين الاختيارين المطروحين: علمي، أدبي.

مثال (١ ـ ٢)

يتضمن الجدول (١ ـ ٢) قياس مستوى الهيموغلوبين في الدم لتسعين عاملا ممن يعيشون في مناطق ترتفع ارتفاعا شاهقا عن سطح البحر، والقياسات كما وردت في الجدول تمثل بيانا إحصائيا انتظمت فيه القياسات وفقا لترتيب الحصول عليها أثناء إجراء البحث الإحصائي. فالقياس الأول 18.5 هو مستوى الهيموغلوبين عند أول عامل تناولته التجربة، والقياس الثاني 23.3 هو مستوى الهيموغلوبين عند العامل الثاني الذي تناولته التجربة، وهكذا. ولنفرض أن مما نهتم به في تجربة كهذه، معرفة نسبة العمال الذين يقل مستوى الهيموغلوبين لديهم عن 17. فسيكون الحصول على هذه النسبة من البيان الإحصائي الخام كما ورد في الجدول (١ ـ ٢)، أمرا يستهلك الكثير من الوقت والجهد. وأول ما يخطر بالبال هو تنظيم عرض هذه القياسات بحيث يسهل ترتيبها من الأصغر إلى الأكبر. ولهذا الغرض يمكن إقامة جدول كالجدول (١ ـ ٣)، حيث وضعنا في العمود الأول أعدادا متسلسلة تمثل الرقمين الأولين لقياس (مبتدئين من اليسار) وفي العمود الثاني وضعنا الرقم الثالث (وهو الرقم الأخير) لكل قياس حذاء العدد المناسب، وبحيث تمتد، كما يوضح الجدول، في سطر أفقي، وذلك حسب ترتيب ورودها في البيان. وفي العمود الثالث وضعنا عدد القياسات التي انتظمت أو اصطفت في سطر واحد. وسنطلق على هذه العملية عملية تصفيف البيان الإحصائي الوارد في الجدول (١ ـ ٢).

جدول (١ ـ ٢) قياسات مستوى الهيموغلوبين في الدم لتسعين عاملا

16.2	17.8	22.0	20.6	19.5	19.4	23.2	16.3	18.5
23.0	21.1	19.7	20.8	21.4	24.2	21.6	19.7	23.3
12.2	24.8	16.9	16.1	20.5	20.9	22.7	18.4	21.7
19.7	18.4	17.1	17.8	18.3	17.3	19.3	17.8	17.4
21.0	18.0	20.9	19.1	26.2	15.5	19.2	19.0	17.8
15.9	18.4	20.0	19.5	19.8	17.2	19.2	18.3	20.2
13.1	18.2	20.3	19.4	23.0	17.8	18.4	16.4	19.9
19.3	19.7	23.5	19.9	17.8	14.3	24.1	18.5	20.3
23.5	19.5	19.1	19.7	18.1	17.6	20.8	18.3	20.6
19.0	18.5	15.5	15.6	16.2	18.8	22.4	20.0	18.5

جدول (١ ـ ٣) تصفيف القياسات الواردة في الجدول (١ ـ ٢)

التعداد	الرقم الثالث	الرقمان الأول والثاني
1	2	12
1	1	13
1	3	14
4	5 6 5 9	15
6	8 4 2 1 9 2	16
11	4 8 8 3 2 8 6 8 8 1 8	17
15	5 5 4 3 5 3 4 8 3 1 4 0 4 2 5	18
21	9 7 0 3 2 2 4 5 8 1 5 4 9 7 7 1 7 5 7 3 0	19
12	2 3 6 0 8 9 5 6 8 9 0 3	20
5	7 6 4 1 0	21
3	7 4 0	22
6	3 2 0 5 0 5	23
3	1 2 8	24
0		25
1	2	26

ونلاحظ أن عملية التصفيف هذه هي، في الواقع، عملية فرز وتوزيع القياسات إلى فئات طول كل منها يساوي عشرة أمثال الواحد في المنزلة العشرية الأخيرة من قياسات البيان. أي أن طولها يساوي الواحد الصحيح إذا كانت القياسات معطاة لرقم عشري واحد وطولها واحد في العشرة، إذا كانت القياسات معطاة لرقمين عشريين،

وطولها عشر وحدات إذا كانت القياسات أعدادا صحيحة، وهكذا*.

وقد أصبح الجواب على التساؤل الذي طرحناه سهلا وميسورا، فنظرة إلى الجدول (١ ـ ٣) تبين أن ثلاثة عشر عاملا من بين التسعين عاملا، يقل مستوى الهيموغلوبين عندهم عن 17. وتكون النسبة المطلوبة $\frac{13}{90}$.

والجدير بالملاحظة أن كل ما خسرناه من المعلومات الواردة في البيان الأصلي (الخام) الوارد في الجدول (١ ــ ٢)، كنتيجة للتصفيف، هو الترتيب الزمني للحصول على القياسات. وقد لا يهمنا هذا في شيء أي أننا، عمليا، لم نخسر شيئا. ولكن وقفة تأمل هنا توضح لنا أن عملية التصفيف في بيانات تتضمن قياساتها أكثر من ثلاثة أرقام معنوية ستحتاج إلى جهود كبيرة، وكذلك ستكون الجهود كبيرة في حالة بيانات تتضمن عددا كبيرا من القياسات، مما يجعل التصفيف عملية غير رابحة في مثل تلك البيانات. فالجهود التي نبذلها في التصفيف قد لا تقل، بل قد تفوق، الجهود التي نحتاجها للإجابة على التساؤلات المطروحة مستخدمين البيان الأصلي مباشرة. وتبقى عملية التصفيف مقبولة فقط في بيانات من الحجم المتوسط، كالبيان المعطى في الجدول (١ ــ ٢) أو أصغر حجما، وفي مثل هذه البيانات، ونظرا لكبر عدد الفئات، تبقى إمكانية ظهور فئة خالية لا تتضمن أي قياس إمكانية قائمة، وهو أمر غير مستحسن.

وربما كان المثال السابق كافيا لتوضيح الفكرة التي نريد تقديمها، وهي أننا نحاول اختزال البيان الاحصائي الخام بطريقة تسمح لنا الإجابة عن تساؤلات، أو فهم نواح معينة مهمة من البيان الإحصائي، بسرعة وسهولة. وذلك لقاء فدية نقدمها، إذ نضحي ببعض المعلومات التي كان البيان الأصلي يوفرها لنا، ولكن البيان المختزل لم يعد قادرا على توفيرها. وسنقدم الآن اتجاها عاما ومفيدا لاختزال بيان إحصائي فيما يسمى بجداول التوزيع التكرارية.

* تسمى هذه الطريقة في التصفيف طريقة «الجذع والورقة»

مثال (١ ـ ٣)

قـدمنا لخمسين مستجـدا من طلبة الجامعـة اختبـارا لقياس «حـاصل الذكـاء» وكانت درجاتهم كما يلي :

جدول (١ ـ ٤). قياسات حاصل الذكاء لخمسين مستجداً

97	110	105	96	109	94	108	117	107	110	82	99	93
116	126	124	108	90	118	116	124	114	101	112	120	113
110	101	103	115	107	102	123	106	105	106	120	100	107
119	120	112	92	103	88	104	97	101	109	105		

إذا قمنا بتصفيف قياسات هذا البيان فسنجد الجدول (١ ـ ٥).

جدول (١ ـ ٥). تصفيف القياسات الواردة في الجدول (١ ـ ٤)

الرقمان الأول والثاني	الرقم الأخير	التعداد
08	2 8	2
09	3 9 4 6 7 0 7 2	8
10	7 8 9 5 1 8 7 0 6 5 6 2 7 3 1 5 9 1 4 3	20
11	0 7 0 3 2 4 6 8 6 5 0 2 9	13
12	0 4 4 6 0 3 0	7

تـوزعت القيـاسـات علـى الفئات الخمس في الجدول (١ ـ ٥) فكان نصيب الفئـة الأولى 2، وهــي تتضمن جميع القيـاســات التي تنتمي إلى الفــترة (90 ,80]، وكان نصيب الفئـة الثانية 8، وهي تتضمن جميــع قيـاســات البيان الإحصــائي التي تنتمـي إلى الفــترة (100 ,90]. وكـان نصيب الفئـة الثالثـة 20، وهــي تتضمن جميـع قياسات البيان التي تنتمي إلى الفــترة (110 ,100]. وكــان نصيب الفئــة الرابعــة 13، وهــي تتضمن جميع قياسات البيان التي تنتمي إلى الفترة(120 ,110]. وكان نصيب الفئة الخامسة والأخــيرة 7، وهـي تتضمـن جميـع قياســات البيان الإحصــائي التي تنتمي إلى الفترة (130 ,120].

بصورة عامة، لماذا لا نختار طول الفئة وبالتالي عدد الفئات بالشكل الذي نراه مناسبا للحالة المدروسة بدلا من أن تُفرض علينا كما هو الحال هنا؟ ولماذا لا نزيد من مقدار التضحية بمعلومات البيان الأصلي، ذات النفع البسيط للنواحي التي يتركز عليها اهتمامنا لقاء مزيد من توفير الجهود وسهولة العرض والحساب؟ فنحن مثلا قد لا نحتاج إلى الاحتفاظ بمفردات البيان الإحصائي، وإنما يقتصر اهتمامنا على معرفة كيفية توزعها على فئات نحددها سلفا تحديدا لا لبس فيه.

وإن أول ما تجدر معرفته هو مدى تغير القياسات في البيان الإحصائي. وباستعراض بسيط للقياسات نجد أن أصغر قياس هو 82، وأن أكبر قياس هو 126. ونقول إن مدى البيان الإحصائي هو الفرق بين أكبر قياس وأصغر قياس فيه، أي:

$$\text{المدى} = 126 - 82 = 44$$

لنقسم هذا المدى إلى عدد من الفئات نختاره بصورة كيفية. ولنأخذ هنا، مثلا، تسع فئات طول كل منها خمسة، فتكون الفئات كما يلي:

$$82 - 86 , 87 - 91 , 92 - 96 , \; . \; . \; . \; . \; , 117 - 121 , 122 - 126$$

ولنرتب جدولا مثل الجدول (١ ـ ٦). حيث نضع في العمود الأول حدود الفئات، وفي العمود الثاني، وسميناه عمود الفرز، نضع حذاء الفئة خطا مائلا في مقابل كل قياس في البيان ينتمي إلى هذه الفئة. ولسهولة التعداد تظهر كل حزمة من خمسة خطوط على حدة، ويقطع الخط الخامس الخطوط الأربعة السابقة له. ويسمى عدد القياسات التي تنتمي إلى الفئة i، مثلا، تكرار الفئة i، ونرمز له عادة بـ f_i. (f_1 تكرار الفئة الأولى، f_2 تكرار الفئة الثانية، . . .، وهكذا). وتظهر هذه التكرارات في العمود الثالث، وهي ناتجة عن تعداد الخطوط المقابلة للفئة في عمود الفرز. ونجد في العمود الرابع، التكرار النسبي، وهو يساوي التكرار مقسوما على العدد الكلي للقياسات $n = 50$. ونلاحظ أن مجموع عمود التكرار يجب أن يساوي 50، وأن مجموع عمود التكرار النسبي يجب أن

يساوي الواحد تماما. وإذا كان عمود التكرار يعطي عدد القياسات في البيان الإحصائي التي تنتمي إلى الفئة المقابلة فإن عمود التكرار النسبي هو تعبير آخر عن الفكرة نفسها، إذ يُقدِّم ذلك العدد على شكل نسبي (منسوبا إلى عدد القياسات الكلي) وسنلمس فيما بعد فائدة التعبير عن التكرار بالشكل النسبي.

جدول (١ ـ ٦). التوزيع التكراري لقياسات حاصل الذكاء لخمسين مستجدا

حدود الفئات	الفرز	التكرار f_i	التكرار النسبي
82 - 86	/	1	1/50
87 - 91	//	2	2/50
92 - 96	~~///~~	4	4/50
97 - 101	~~///~~ //	7	7/50
102 - 106	~~///~~ ~~///~~	9	9/50
107 - 111	~~///~~ ////	10	10/50
112 - 116	~~///~~ //	7	7/50
117 - 121	//// /	6	6/50
122 - 126	////	4	4/50
المجموع		50	1

وترتيب القياسات كما في الجدول (١ ـ ٦) يسمى توزيعا تكراريا للقياسات. وبصورة عامة، التوزيع التكراري هو أي ترتيب للبيان الإحصائي يظهر توزّع قياساته على فترات معرفة بصورة اختيارية تسمى «فئات».

وقد اختيرت الفئات بصورة كيفية. ومن أجل البيان الإحصائي نفسه يمكن أن يختلف جدول التوزيع التكراري باختلاف تعريف الفئات وعددها، وليس هناك جدول يمكن القول إنه صحيح وما عداه من الجداول التي كان يمكن الوصول إليها غير

صحيحة ، ولكن بعض هذه الجداول أفضل من بعض من حيث مقدرتها على تبيان النواحي المهمة في البيان الإحصائي دون الاحتفاظ بكثير من التفاصيل .

وبصورة عامة ، يستحسن ألا يقل عدد الفئات عن خمس ولا يزيد على عشرين ، تفادياً لظهور فئات خالية عند استكمال عملية الفرز، ذلك لأننا قد نخسر أكثر مما يجب من المعلومات إذا قل عدد الفئات عن خمس ، وقد نحتفظ بما لا ضرورة له من التفاصيل عندما يزيد عدد الفئات على عشرين .

ويجب تعريف حدود الفئات بصورة واضحة لا تترك أي لبس في عملية الفرز، وتضمن انتماء كل قياس في البيان الإحصائي إلى فئة واحدة وواحدة فقط .

(١ ـ ٢) أنواع البيانات الإحصائية

تنقسم البيانات الإحصائية العددية إلى نوعين ، أحدهما منفصل وتكون قياساته ناتجة عن عملية عد أو تعداد ، مثل عدد حوادث المرور اليومي خلال فترة زمنية محددة ، أو العدد السنوي لحالات الولادة ، أو الزواج ، أو الوفاة ، أو الطلاق ، في بلد معين ، وتكون مثل هذه القياسات ، دائما ، أعدادا صحيحة . والنوع الآخر هو النوع المتصل (أو المستمر) ، وتكون قياساته ناتجة عن استخدام جهاز أو أداة للقياس ، مثل بيانات تتضمن قياسات طول ، أو وزن ، أو د رجة حرارة ، أو مستوى التحصيل الدراسي ، أو حاصل الذكاء ، الخ .

وفي البيانات المستمرة ، نفهم من العدد المقدم لنا شيئين ، أولهما تصور عن مقدار الشيء المقيس ، وثانيهما درجة الدقة التي سمح بها جهاز القياس المستخدم . والقول بأن طول شخص هو 167.5سم ، يعطينا فكرة عن ارتفاع قامة الشخص ، ويعطينا أيضا أن القياس جرى بدقة تصل إلى أقرب ملليمتر. أي أن آخر رقم معطى على اليمين ، هو رقم مشكوك فيه . ولو أننا استخدمنا جهازا أكثر دقة ، لحصلنا على قياس واقع في مكان ما بين 167.45 و 167.549 . وأينما وقع هذا القياس فسيؤدي التدوير إلى الرقم العشري

الأول إلى العدد 167.5سم . وللسهولة جرت العادة على القول بأن عددا مثل 167.5سم يعني أي شيء بين 167.45 و 167.55سم .

ولأسباب عدة، نأخذ في الغالب، الحدود المضبوطة للفئة بعين الاعتبار، ونسميها «الحدود الحقيقية للفئة» أو «نهايتي الفئة». لنأخذ الفئة 86 - 82 في الجدول (١ ـ ٦)، فالقياس 82 يعني أي شيء بين 81.5 و 82.5، ويعني الـ 86، أي شيء بين 85.5 و 86.5. وهكذا يتراوح المدى الحقيقي للفئة بين 81.5 و 86.5. ويسمى هذان العددان «الحدان الحقيقيان للفئة» أو «نهايتا الفئة». وتجدر ملاحظة أن تطابق نهاية فئة مع بداية الفئة التي تليها، لا يؤدي إلى أي التباس في عملية الفرز، فالعدد 86.5 الذي يشكل حدا أعلى للفئة الأولى وحدا أدنى للفئة الثانية لا يمكن أن يظهر كقياس في البيان الاحصائي الأصلي ما دامت القياسات جميعها أعدادا صحيحة .

وهناك طرق أخرى يمكن استخدامها للتعبير عن حدود الفئة فمثلا يمكن، في المثال (١ ـ ٢ ـ ٣)، كتابة الفئات على الشكل :

82-, 87-, 92-, 97-, 102-, 107-, 112-, 117-, 122-

ونقصد بـ-82 جميع الأعداد الواقعة ضمن الفترة [82-87)، أي الأعداد بدءا من 82 إلى أقل من 87، وهكذا .

أو يمكن كتابتها على الشكل :

-87, -92, -97, -102, -107, -112, -117, -122, -127

ونقصد بـ 92- جميع الأعداد الواقعة ضمن الفترة [87-92)، أي الأعداد بدءا من 87 إلى أقل من 92. وهكذا .

وسنستخدم في هذا الكتاب الحدود الحقيقية للفئات في جميع البيانات سواء أكانت مستمرة أم منفصلة. واستخدامها في البيانات المنفصلة يضمن استمرارية

الأشكال التي تمثل الجدول التكراري بيانيا كما سنرى في الفقرة التالية، وهذا أمر مستحسن. كما سنستفيد منه في أكثر من مكان في الفصول المقبلة.

ونلاحظ بوضوح أن طول الفئة مساو للفرق بين حديها الحقيقين. أما مركز الفئة فهو منتصف المسافة بين حديها، ولحساب قيمته نأخذ نصف مجموع الحدين، ولو استخدمنا الحدود (87-82]، (92-87] الخ. فإن مركز الفئة الأولى سيكون 84.5 والثانية 89.5 الخ. وعند استخدام الحدود الحقيقية [81.5-86.5]، [86.5-91.5] الخ. فإن مركز الفئة الأولى سيكون 84 والثانية 89 إلخ. واستخدام الحدود الحقيقية إلى جانب أنه يعالج مشكلة وجود فراغات بين الفئات المتتالية ويضمن تطابق نهاية فئة مع بداية الفئة التي تليها بطريقة منطقية وعادلة، فإنه يؤدي أيضا إلى حسابات أكثر دقة بصورة عامة.

وإذا أضفنا طول الفئة إلى مركز الفئة الأولى حصلنا على مركز الفئة الثانية التي تليها وهكذا. وسنتصور وجود فئة على يسار الفئة الأولى، وبالطبع سيكون تكرارها مساويا للصفر ولذلك سنسميها الفئة الصفرية على اليسار، كما سنتصور وجود فئة على يمين الفئة الأخيرة، وبما أن تكرارها صفر فسنسميها أيضا الفئة الصفرية على اليمين.

ولقد ذكرنا أن تصنيف القياسات في فئات، يهدف إلى تيسير عرض البيانات، وسهولة القيام بحساب معايير إحصائية مفيدة في وصف وتحليل البيان الإحصائي، واستنتاج معلومات عامة منه. وننطلق في هذا من نوعين من الافتراضات المتعلقة بكيفية توزع القياسات ضمن الفئة الواحدة.

(i) عند حساب بعض المعايير الإحصائية، أو عند استخدام الطرق البيانية لعرض معلومات إحصائية، نفترض أن القياسات الواقعة ضمن فئة واحدة تتوزع بانتظام على الفترة الممتدة بين نهايتي الفئة. وفي الجدول (١ ـ ٦)، مثلا، ينتمي عشر قياسات إلى الفئة 107-111. والفترة الممتدة بين نهايتي الفئة هي الفترة (106.5, 111.5). ونفترض أن القياسات العشرة تتوزع بانتظام فوق الوحدات الخمس التي تتألف منها الفئة. أي نفترض، كما يبين الجدول (١ ـ ٧)، قياسين بين 106.5 و 107.5، وقياسين بين 107.5 و 108.5، الخ.

جدول (١ ـ ٧). توزع القياسات بانتظام ضمن الفئة الواحدة

الفئة الجزئية	110.5-111.5	109.5-110.5	108.5-109.5	107.5-108.5	106.5-107.5
التكرار	2	2	2	2	2

(ii) والافتراض الثاني الذي نستخدمه عند حساب بعض المعايير الإحصائية هو اعتبار مركز كل فئة ممثلا لجميع القياسات التي تنتمي إليها، أي نفترض أن كل قياس من القياسات التي تنتمي إلى فئة مساوٍ لمركز الفئة.

والجدير بالذكر أن هناك بيانات إحصائية غير عددية تتضمن أصنافا معبرا عنها على شكل كلمات وصفية أو رموز، (علمي، أدبي)، (ذكر، أنثى)، (ممتاز، جيد جدا، جيد، مقبول، ضعيف) وبحيث ينتمي كل عنصر يخضع للتصنيف إلى صنف واحد منها فقط. ومثل هذه البيانات تسمى بيانات وصفية. وإذا أمكن تعريف ترتيب على هذه الأصناف أو الرموز يسمى البيان عندئذ بيانا ترتيبيا. فمثلا، يمكن القول أن «ممتاز» يمثل الصنف الأعلى يليه «جيد جدا» يليه «جيد» إلخ. مما يجعل أي بيان يتضمن تقديرات ممتاز. . . إلى ضعيف بيانا ترتيبيا. ونلاحظ أن ذلك غير ممكن في التصنيف (علمي، أدبي) أو (ذكر، أنثى).

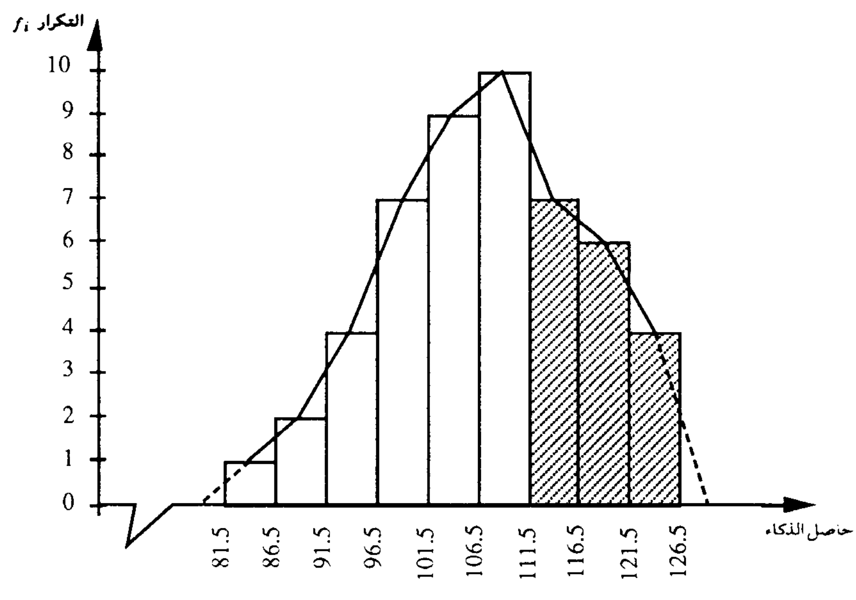
(١ ـ ٣) التمثيل البياني لتوزيع تكراري

يقدم لنا التمثيل البياني للمعلومات الإحصائية عونا كبيرا، فهو يسمح بإدراك الخواص الأساسية للتوزيع التكراري، ومقارنة توزيع تكراري بآخر. والتمثيل البياني هو صورة هندسية لجملة القياسات. وفي العديد من الحالات يسهّل رد مجموعة المعلومات الرقمية إلى صورة هندسية، فهم طبيعة المسألة الإحصائية، واستنباط الحلول المناسبة لها. وقد أصبح التمثيل البياني ممارسة شبه يومية في حياتنا. فالصحف والمجلات، والنشرات التجارية، وتقارير الأعمال والمشاريع، والدوريات العلمية المختلفة،

والتقارير الحكومية، تستخدم جميعها، وعلى نطاق واسع، التمثيل البياني. وهناك تفرعات كثيرة لوسائل التعبير البياني عن جملة من المعلومات الإحصائية، وسنقتصر هنا على ذكر أكثرها أهمية وفائدة في مجالات الاستقراء الإحصائي. ويمكن لمن أراد الاستزادة العودة إلى بعض المراجع المذكورة في نهاية الكتاب.

(١ ـ ٣ ـ ١) المدرج التكراري

لرسم المدرج التكراري، نتخذ المحور الإحداثي السيني لتمثيل الفئات، ونحدد عليه النقاط التي تمثل نهايات الفئات (حدودها الحقيقية). ونتخذ المحور الإحداثي الصادي لتمثيل التكرار f_i . ثم نرسم فوق الفترة الممتدة بين نهايتي كل فئة مستطيلا يرتفع بمقدار التكرار المقابل لهذه الفئة. ونجد في الشكل (١ ـ ١) المدرج التكراري الموافق للتوزيع التكراري المعطى في الجدول (١ ـ ٦).



شكل (١ ـ ١). مدرج التكرار ومضلع التكرار لقياسات حاصل الذكاء لخمسين مستجدا

وتمثل المساحة تحت مدرج التكرار نوعا من أنواع التمثيل الهندسي لجملة القياسات في البيان الإحصائي. فإذا نظرنا إلى تكرار كل فئة بأنه مساهمة الفئة في تركيبة

البيان الإحصائي، إذا جاز التعبير، فإن مساحة المستطيل المقام فوق الفئة يتناسب مع هذه المساحة. وكلما كان التكرار أكبر ارتفع المستطيل وزادت مساحته. ولو تساءلنا في المثال (١ ـ ٣) عن نسبة الطلبة الذين نالوا درجات أعلى من 111.5 لوجدنا أن هذه النسبة تساوي نسبة المساحة تحت مدرج التكرار الواقعة على اليمين من 111.5 (وهي مساحة المستطيلات الثلاثة الأخيرة المظللة في الشكل (١ ـ ١) إلى المساحة الكلية تحت مدرج التكرار. وما دامت الفئات جميعها بالطول نفسه، أي ما دامت قواعد المستطيلات المرسومة متساوية وتبقى ثابتة من فئة إلى أخرى، فإن مساحة كل مستطيل تتناسب مع ارتفاعه (مع تكرار الفئة) ونسبة المساحة المظللة إلى المساحة الكلية هي في الواقع نسبة مجموع التكرارات الموافقة للفئات الثلاث الأخيرة إلى العدد الكلي للقياسات، أي $\frac{17}{50}$ أو 34% وهي النسبة المطلوبة بالضبط.

ويمكن اللجوء إلى هذا المبدأ في تمثيل المدرج التكراري سواء أكانت أطوال الفئات متساوية أم لا. ففي حالة رسم مدرج تكراري لتوزيع تكراري لا تتساوى فيه أطوال الفئات، إما كنتيجة لطبيعة التفاصيل التي رؤي أن يحتفظ بها الجدول التكراري، أو نتيجة لدمج عدة فئات، تكراراتها صغيرة نسبيا، بعضها مع بعض لتشكل فئة واحدة؛ لا بد من القيام بتعديلات مناسبة تأخذ في الاعتبار أطوال الفئات، وتجعل المساحة المقامة فوق فئة، متناسبة مع تكرار هذه الفئة. ويكون رسم مستطيلات ارتفاعاتها مساوية لتكرار الفئة غير صحيح. وبذلك نتجنب رسم مدرج تكرار يعطي انطباعات مضللة إلى حد بعيد. ونوضح الفكرة وطريقة العمل من خلال المثال التالي.

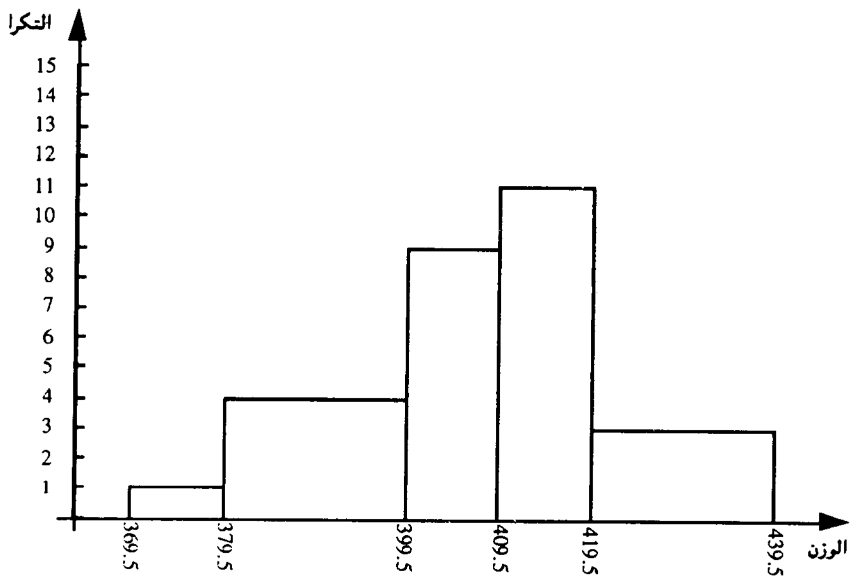
مثال (١ ـ ٤)

فيما يلي جدول توزيع تكراري لأوزان 35 فأرا مقاسة إلى أقرب غرام. ارسم المدرج التكراري.

الرسم مبين في الشكل (١ ـ ٢) حيث عدّلنا في ارتفاع المستطيل المقام فوق كل فئة بحيث نحفظ تناسب المساحة المرسومة فوق الفئة مع التكرار الموافق، والفئة الثانية والخامسة لهما أطوال مضاعفة ولذلك رسمنا فوق كل منها مستطيلا ارتفاعه يساوي

جدول (١ ـ ٨): التوزيع التكراري لأوزان35 فأرا

حدود الفئات	370-379	380-399	400-409	410-419	420-439
التكرار	1	8	9	11	6



شكل (١ ـ ٢). المدرج التكراري لأوزان 35فأرا

نصف التكرار الموافق للفئة (4 في الفئة الثانية و 3 في الفئة الخامسة). إن المساحة الكلية للمدرج هي:

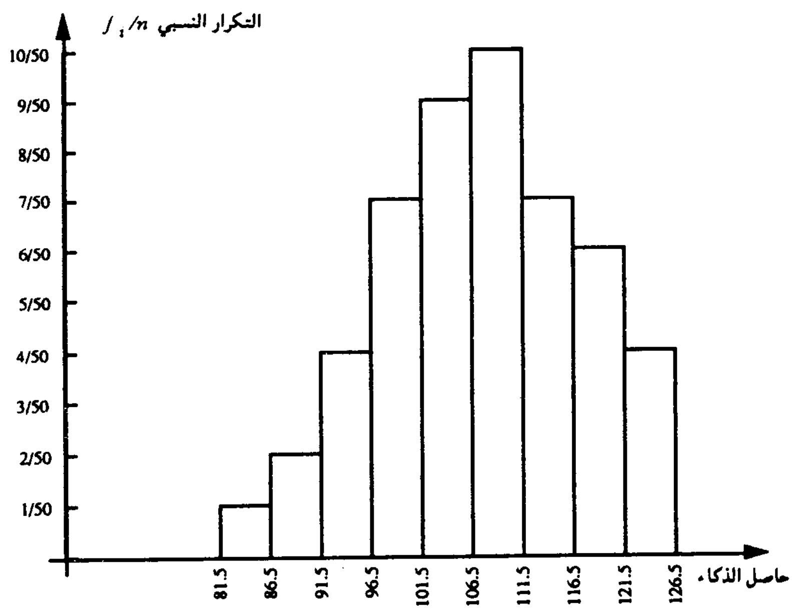
$$10 \times 1 + 20 \times 4 + 10 \times 9 + 10 \times 11 + 20 \times 3 = 350$$

ونسبة مساحة المستطيل الموافق لكل فئة إلى المساحة الكلية تساوي تماما نسبة تكرار الفئة إلى مجموع التكرارات. فمثلا نسبة مساحة المستطيل الثاني إلى المساحة الكلية هي $\frac{80}{350}$ وهي تساوي $\frac{8}{35}$.

(١ ـ ٣ ـ ٢) مدرج التكرار النسبي

لا تختلف طريقة رسم مدرج التكرار النسبي عن مدرج التكرار سوى أن المستطيل الموافق لكل فئة يرتفع الآن بما يساوي التكرار النسبي للفئة. ولكي نحافظ على ارتفاع ووضوح مناسبين للصورة الناتجة، لا بد أن تكون وحدة الطول على المحور الصادي أكبر بصورة مناسبة مما كانت عليه على المحور الصادي لمدرج التكرار. ولو كان لدينا n قياسا، وكبرنا وحدة الطول على المحور الصادي n مرة، لحصلنا على صورة لمدرج التكرار النسبي مطابقة تماما لصورة مدرج التكرار. وكل ما في الأمر أن التدريج 1 على المحور الرأسي أصبح الآن $\frac{1}{n}$ ، والتدريج 2 أصبح $\frac{2}{n}$ ، وهكذا. ونجد في الشكل (١ ـ ٣) مدرج التكرار النسبي لقياسات حاصل الذكاء لخمسين مستجدا.

تجدر ملاحظة أنه ليس من الضروري، عند رسم شكل بياني، أن تكون وحدة الأطوال نفسها على المحورين. ونتخذ وحدة الطول على كل من المحورين لتشغل الصورة الناتجة الحيز المخصص لها، وتتخذ موقعا مناسبا في الاتجاهين الأفقي والرأسي،



شكل (١ ـ ٣). مدرج التكرار النسبي لقياسات حاصل الذكاء لخمسين مستجدا

تماما كما تتخذ الصورة المخرجة في التليفزيون موقعها على الشاشة المخصصة لها، فلا هي منحازة إلى يمين الشاشة ولا إلى يسارها، ولا هي مرتفعة أو منخفضة أكثر مما ينبغي. وتكبير وحدة الطول على المحور السيني يؤدي إلى توسيع الصورة في الاتجاه الأفقي يمينا ويسارا. وتكبير وحدة الطول على المحور الصادي يؤدي إلى تمدد الصورة في الاتجاه الرأسي علوا وهبوطا. وأفضل ترتيب لوحدتي الطول هاتين، هو ذلك الذي يكفل وضوح الصورة، ويُخرجها بحيث تشغل الحيز المخصص لها بشكل مناسب.

ويبقى هذا كله صحيحا في حالة فئات غير متساوية أيضا، فصورة مدرج التكرار النسبي تتطابق مع صورة مدرج التكرار عندما نجعل التدريج 1 على المحور الصادي مساويا $1/n$، والتدريج 2 مساويا لـ$2/n$، وهكذا. والشكل (١ ـ ٢)، يصبح صورة لمدرج التكرار النسبي للتوزيع التكراري لأوزان 35 فأرا المعطى في الجدول (١ ـ ٨)، إذا اعتبرنا التدريج 1 على المحور الصادي مساويا الآن لـ 1/35 والتدريج 2 مساويا لـ 2/35، إلخ.

ومع أن اهتمامنا المباشر، في المثال (١ ـ ٣)، ينصب على وصف القياسات الخمسين، إلا أننا نهتم أكثر بالمجتمع الذي أخذنا منه هذه القياسات. ويمكن النظر إلى القياسات الخمسين كعينة مأخوذة من مجتمع طلبة السنة الأولى في جامعة أو عدد من الجامعات. وفي جميع الأحوال، لو توفرت لنا قياسات حاصل الذكاء لعناصر المجتمع كلها، لأمكن، بالطريقة ذاتها، إقامة المدرج التكراري للمجتمع.

لندرس الآن مدرج التكرار النسبي في الشكل (١ ـ ٣) بتفصيل أكثر. فلو افترضنا أن طول الفئة (وهي تساوي خمس وحدات) أصبحت وحدة قياس جديدة، أي أن طول الفئة بالوحدة الجديدة هو الواحد، فستصبح مساحة المستطيل المقام فوق الفئة مساوية للتكرار النسبي الموافق لهذه الفئة، وستصبح المساحة الكلية تحت مدرج التكرار النسبي مساوية للواحد تماما. ولنسأل الآن، ما هي نسبة الطلاب الذين يزيد حاصل ذكائهم على 111.5 مثلا؟ بالعودة إلى مدرج التكرار النسبي نرى أن هذه النسبة تشمل كل الفئات على اليمين من 111.5 . وبالاستفادة من الجدول (١ ـ ٦) نرى أن

سبعة عشر مستجدا حصلوا على أكثر من 111.5. أي أن النسبة المطلوبة هي 17/50 أو 34%. ونلاحظ أن هذه النسبة هي أيضا المساحة تحت مدرج التكرار النسبي في الشكل (١ ـ ٣) التي تقع على يمين 111.5 .

(١ ـ ٣ ـ ٣) مضلع التكرار

نأخذ منتصفات القواعد العليا للمستطيلات في مدرج التكرار، ونصل بينها بخطوط مستقيمة، فنحصل على ما يسمى بمضلع التكرار، أي أننا لو حددنا من أجل كل فئة نقطة إحداثيها السيني هو مركز الفئة، وإحداثيها الصادي هو تكرار الفئة، ثم وصلنا بين هذه النقاط بقطع مستقيمة لحصلنا على مضلع التكرار. ويمكن رسم مدرج التكرار ومضلع التكرار على الشكل نفسه، أو في شكلين منفصلين. ونجد في الشكل (١ ـ ١) مضلع التكرار للتوزيع التكراري في الجدول (١ ـ ٦) .

ويمكن إغلاق مضلع التكرار على الجانبين بوصل أول نقطة منه بمركز الفئة الصفرية على اليسار، ووصل آخر نقطة منه بمركز الفئة الصفرية على اليمين. (انظر الشكل (١ ـ ١) حيث رسمنا هاتين الوصلتين بخط منقط .)

ونلاحظ وجود فرق بسيط بين المساحة تحت مضلع التكرار والمساحة تحت مدرج التكرار ويتناقص هذا الفرق كلما ازداد عدد الفئات وصغر طول الفئة. هذا بصورة عامة، أما إذا كانت أطوال الفئات متساوية فالمساحتان متساويتان.

(١ ـ ٤) مضلع التكرار المتجمع الصاعد

من الخصائص المهمة للبيان الإحصائي معرفة العدد الذي تقل عنه نسبة معينة من القياسات، أو معرفة النسبة من القياسات التي تقل عن قيمة معينة، أو نسبة القياسات التي تتجاوز قيمة معينة. ففي بيان من الدرجات في مسابقة عامة، يمكن أن نعتبر العدد الذي يقل عنه تسعون بالمائة من القياسات الحد الفاصل بين تقدير الممتاز وما دون الممتاز. وفي بيان يمثل مستويات الهيموغلوبين في الدم تمثل نسبة القياسات، التي تقل عن قيمة معينة، نسبة المصابين بفقر الدم. وفي بيان يمثل

معدلات التوتر الشرياني (ضغط الدم) تمثل نسبة القياسات التي تزيد على قيمة معينة نسبة المصابين بمرض فرط التوتر الشرياني (ارتفاع معدل ضغط الدم).

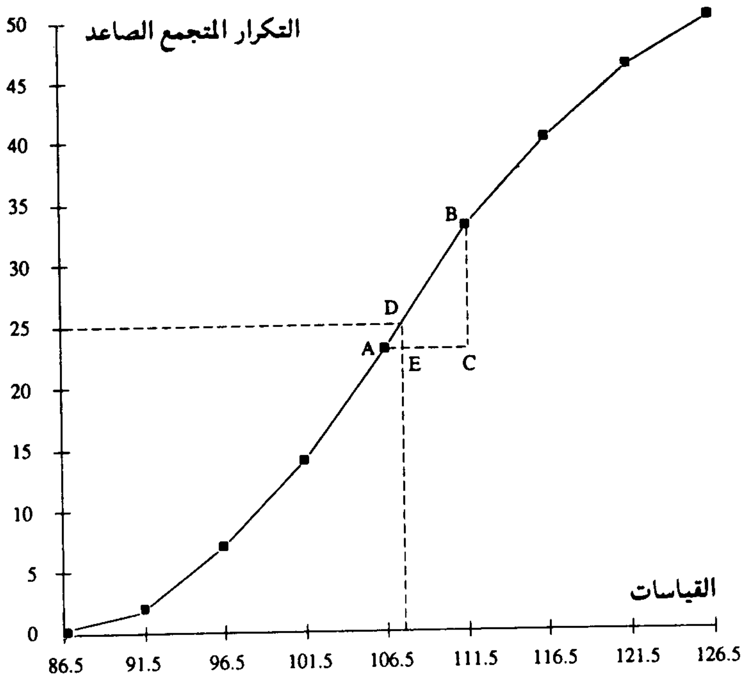
لنعد إلى الجدول (١ ـ ٦) فما هو العدد الذي يقل عنه خمسون بالمائة من القياسات؟ أو ما هي النقطة التي يقع إلى اليسار منها خمس وعشرون بالمائة من القياسات؟ أو ما هي نسبة القياسات التي تقل عن 101.5؟ إلخ.

وللجواب على مثل هذه التساؤلات، بصورة تقريبية وسريعة، نقيم جدول التكرار المتجمع الصاعد كما في الجدول (١ ـ ٩)، حيث نضع في العمود الأول، وعنوانه «أقل من»، الحدود الحقيقية العليا للفئات، ونضع في العمود الثاني، وعنوانه «التكرار المتجمع الصاعد» عدد القياسات الموافق.

جدول (١ ـ ٩). جدول التكرار المتجمع الصاعد لقياسات حاصل الذكاء لخمسين مستجدا

التكرار المتجمع الصاعد	أقل من
1	86.5
3	91.5
7	96.5
14	101.5
23	106.5
33	111.5
40	116.5
46	121.5
50	126.5

ولتمثيل الجدول بيانيا نعتمد المحور السيني محورا للقياسات، والمحور الصادي محورا للتكرار المتجمع الصاعد. ونرسم لكل فئة نقطة في مستوى الإحداثيات، إحداثيها السيني هو الحد الأعلى الحقيقي للفئة، وإحداثيها الصادي هو التكرار المتجمع الصاعد المقابل. ثم نصل بين النقاط الناتجة المتتالية بقطع مستقيمة فنحصل على مضلع يدعى «مضلع التكرار المتجمع الصاعد». (انظر الشكل (١ ـ ٤)).



شكل (١ ـ ٤). مضلع التكرار المتجمع الصاعد لقياسات حاصل الذكاء لخمسين مستجدا

وبـالطريقـة نفسها يمكن رسم مضلع التكـرار النسبي المتجمع الصاعـد وذلك باستخدام المحور الصادي لتمثيل التكرارات النسبية المتجمعة. ولو أن العمود الأول في الجدول (١ ـ ٩) تضمن الحدود الحقيقية الدنيا للفئات وكـان عنوانه «أكثر من» لحصلنا على جدول تكرار متجمع نـازل، ورسمه البياني بالطريقـة السالفة ذاتها سيعطي مضلع التكرار المتجمع النازل. وسنترك ذلك تمرينا للطالب.

ولإيجاد القياس الذي يقع على اليسار منه 50% من القياسات، نحسب أولا رتبة القياس المطلوب $\frac{50}{100} \times n$ ، حيث n عدد القياسات، فنجد:

$$25 = \frac{50}{100} \times 50 = \text{رتبة القياس المطلوب}$$

أي أن القياس المطلوب ينتمي إلى الفترة [106.5, 111.5[.

وسنحسب القياس المطلوب مفترضين أن القياسات التي تنتمي إلى فئة تتوزع بانتظام فوق الفترة التي تمتد بين نهايتي الفئة، أو بعبارة أعم مفترضين أن العلاقة بين القياس والتكرار فوق الفترة [106.5, 111.5] هي علاقة خطية تتمثل في معادلة المستقيم الواصل بين النقطتين A و B.

من تشابه المثلثين ADE و ABC نجد:

$$\frac{AE}{AC}=\frac{ED}{CB}$$

$$\frac{AE}{5}=\frac{2}{10}$$

$$AE=\frac{5\times 2}{10}=1$$

ويكون القياس المطلوب، وهو الاحداثي السيني للنقطة D، مساويا الإحداثي السيني لـ A مضافا إليه AE، وهكذا نجد:

$$107.5 = 1 + 106.5 = \text{القياس المطلوب}$$

ويمكن القيام بهذه الحسابات معتمدين على جدول التكرار المتجمع الصاعد، ودون الحاجة إلى رسم مضلع التكرار، حيث نتبع المحاكمة التالية:

نرى من جدول التكرار المتجمع الصاعد أن 23 قياسا من القياسات الخمسين أقل من 106.5، وأن 33 قياسا أقل من 111.5. وبتطبيق التناسب الطردي نقول إنه عندما زاد التكرار المتجمع بمقدار 10، (من 23 إلى 33) زاد القياس بمقدار 5، (من 106.5 إلى 111.5). فما هي قيمة الزيادة في القياس عندما يزداد التكرار المتجمع بمقدار 2 فقط (من 23 إلى 25)؟

زيادة القياس	زيادة التكرار المتجمع
5	10
x	2

$$x = \frac{2 \times 5}{10} = 1$$ (الزيادة المطلوبة في القياس)

ويكون القياس المطلوب:

$$106.5 + 1 = 107.5$$

وإذا توفر ورق ميلليمتري نرسم عليه مضلع التكرار المتجمع الصاعد، فيمكن استخدام الرسم البياني لإيجاد القياس المطلوب، وهذا القياس ليس إلا الإحداثي السيني لنقطة على مضلع التكرار المتجمع الصاعد إحداثيها الصادي 25 . ولذلك نرسم من النقطة 25 على المحور الصادي خطا أفقيا يقطع مضلع التكرار المتجمع الصاعد في نقطة ننزل منها عمودا على المحور السيني فيقطعه في النقطة المطلوبة، وهي على الشكل (١ ـ ٤) حوالي 107.5 .

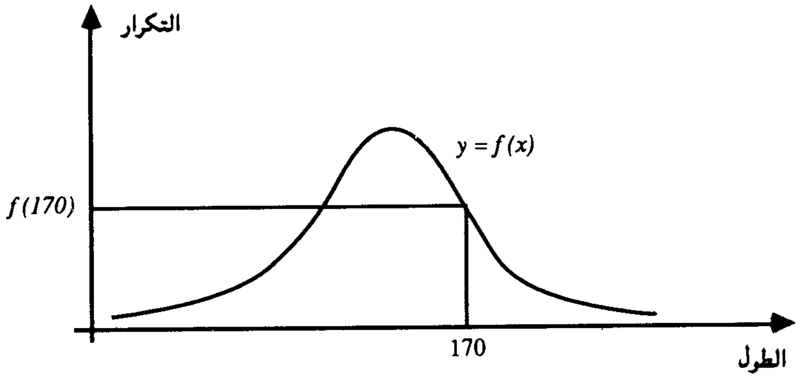
وسنجد فيما بعد أن هذا القياس يسمى الوسيط. وقد لخصنا هنا الطريقتين الحسابية والبيانية للحصول على الوسيط. ومن الواضح أنه يمكن تطبيق الطريقة نفسها لحساب القياس الذي يقع على اليسار منه 25% من القياسات، وبصورة عامة القياس الذي يقع على اليسار منه p بالمائة من القياسات، حيث p أي عدد بين الصفر والمائة، ويسمى مثل هذا القياس المئين p.

ولمعرفة نسبة القياسات التي تقل عن 101.5، مثلا، نرفع من النقطة 101.5 على المحور السني عمودا يقطع مضلع التكرار المتجمع في نقطة نرسم منها موازيا للمحور السيني فيقطع المحور الصادي في النقطة 14، وتكون النسبة المطلوبة $\frac{14}{50} = 28\%$.

(١ ـ ٥) منحنى التكرار

لنعد إلى مضلع الكرار في الفقرة (١ ـ ٣ ـ ٣). ولنفترض أننا صغرنا طول الفئة إلى نصف ما هو عليه. أي ضاعفنا عدد الفئات، ثم رسمنا مضلعا تكراريا، فسيتضاعف عندئذ عدد رؤوس هذا المضلع، وستقترب رؤوس المضلع بعضها من بعض. ولكن العدد البسيط من القياسات لا يسمح لنا بالمضي في مثل هذه العملية، لأنه قد يترك العديد من الفئات خالية وتكرارها صفر، مما يصيب المضلع بانقطاعات في أكثر من مكان، الأمر الذي لا يقلق كثيرا عندما يصف المضلع التكراري «مجتمعا» يتضمن عددا هائلا من القياسات. فلنتصور إذا، أن لدينا معينا لا ينضب من القياسات، أي لنتصور ظرفا يمكننا معه جعل طول الفئة أصغر فأصغر، وفي الوقت ذاته، زيادة عدد القياسات التي تخضع للتصنيف لتصبح أكبر فأكبر، ولندفع الآن مثل هذا التصور إلى نهاياته القصوى ليصبح طول الفئة صغيرا بلا حدود، ويصبح معه عدد القياسات الكلي كبيرا بلا حدود، فسنصل عندئذ إلى خط ناعم مستمر، لا انكسارات فيه ولا زوايا، يسمى منحنى التكرار. وعندئذ يقابل كل قياس على المحور السيني إحداثي صادي يتناسب مع تواتر ظهور هذا القياس في المجتمع الذي يصفه منحنى التكرار.

ولنفرض، على سبيل المثال، أن منحنى التكرار في الشكل (١ ـ ٥) يصف ظاهرة توزع الطول في مجتمع من الذكور البالغين يتضمن عشرات الملايين، فالإحداثي الصادي للنقطة 170سم، مثلا، يمثل أو يتناسب مع تواتر ظهور الطول 170سم في هذا المجتمع. وبصورة عامة، نعتمد منحنيات التكرار كنماذج رياضية (نظرية) لتمثيل ظواهر عامة في حياتنا العملية. وعلى سبيل المثال، سنعرض فيما يلي إحصائيات (Kendall and Stuart, 1977) لثلاث ظواهر مختلفة تتناول عددا كبيرا من الأفراد. وسنجد أن مضلع التكرار لكل ظاهرة يوحي بشكل معين لمنحنى التكرار (أو النموذج) الذي يمكن اعتماده لوصف هذه الظاهرة. وسنتعرف في الفصل الرابع وما بعده على ما نقصده بكلمة «نموذج»، والدور الذي تلعبه النماذج في التطبيقات العملية للإحصاء.



شكل (١ ـ ٥). منحنى التكرار لتوزيع الطول في مجتمع من الذكور البالغين

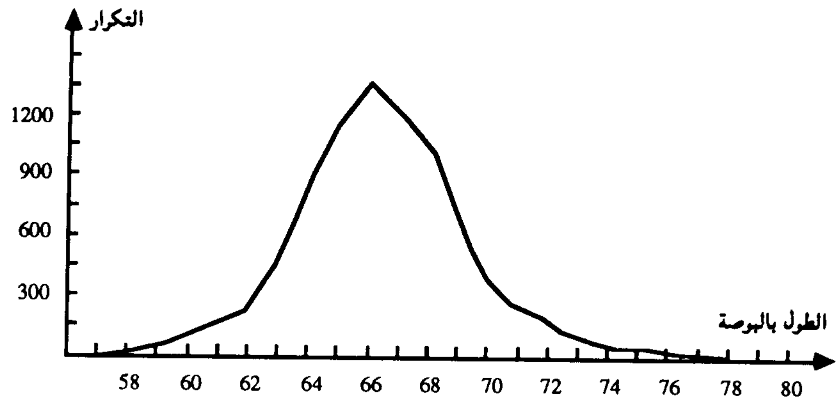
يبين الجدول (١ ـ ١٠) توزيع التكرار لأطوال 8585 ذكرا بالغا ممن ولدوا في المملكة المتحدة. وباعتبار أن دقة القياس كانت إلى أقرب $\frac{1}{8}$ من البوصة، فالحدود الحقيقية للفئات هي من $57\frac{15}{16} - 56\frac{15}{16}$ ، $58\frac{15}{16} - 57\frac{15}{16}$ ، وهكذا . . .

جدول (١ ـ ١٠). التوزيع التكراري لـ 8585 ذكرا بالغا ممن ولدوا في المملكة المتحدة

التكرار	الطول (بدون حذاء)	التكرار	الطول (بدون حذاء)
1230	68 -	2	57 -
1063	69 -	4	58 -
646	70 -	14	59 -
392	71 -	41	60 -
202	72 -	83	61 -
79	73 -	169	62 -
32	74 -	394	63 -
16	75 -	669	64 -
5	76 -	990	65 -
2	77 -	1223	66 -
		1329	67 -

8585 = مجموع التكرارات

وفي الشكـل (١ ـ ٦) نجد مضلـــع التكرار، ومن الواضح أن هذا المضلع يقترح بقوة أن نموذجا على شكل الجرس (انظر الشكل (١ ـ ٥)) هو النموذج المناسب لتمثيل ظاهرة توزع الطول في مجتمع من الذكور البالغين في بيئة معينة .



شكل (١ ـ ٦) . مضلع التكرار للبيان الإحصائي في الجدول (١ ـ ١٠)
(القيم على محور السينات هي بدايات الفئات)

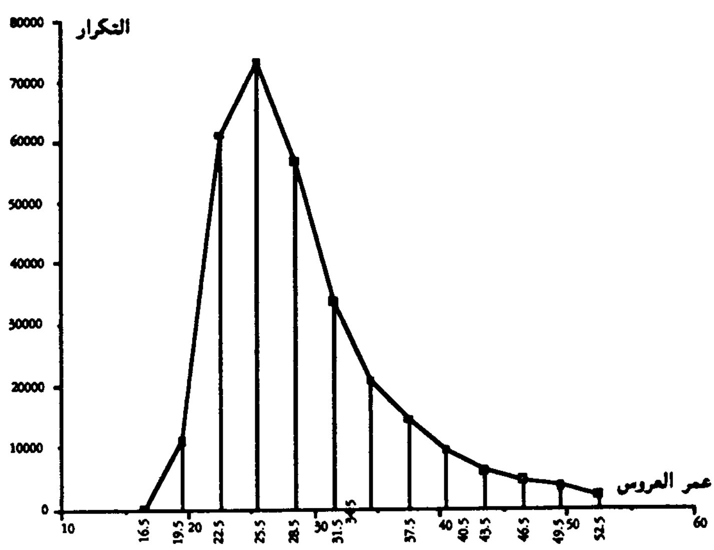
يبين الجدول (١ ـ ١١) توزيع التكرار لـ 301785 عقد زواج في استراليا بين 1907 و 1914، مصنفة وفقا لعمر العروس في فئات طول كل منها 3 سنوات .

والمضلع التكراري يقترح بـوضوح نموذجا يعـرف بنموذج «جامـا» وهو منحنى تكرار غير متماثل يتـزايد بسرعـة إلى قمة ثم ينحـدر منها بسرعة (سرعـة التزايـد وسرعة الانحدار تختلف من حالة إلى أخـرى) ليتهادى بعد ذلك متناقصا باطـراد تناقصا بطيئا مقتربـا من محور السينـات . ونقـول عن نمـوذج كهذا أنـه ملتـو إلى اليمين أو مـوجب الالتواء [انظـر الشكل (١ ـ ٧)]. ونجـد في الشكـل (١ ـ ٨) منحنى تكـرار من النوع «جاما» .

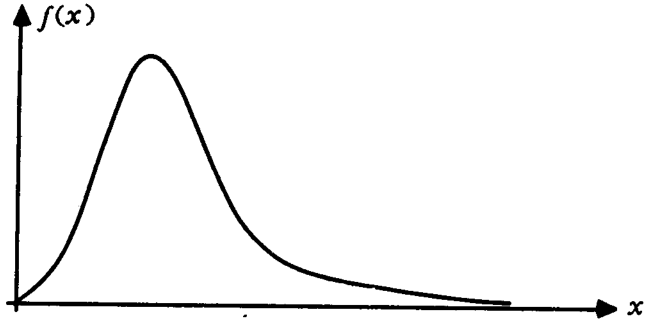
جدول (١ ـ ١١). التوزيع التكراري لـ 301785 عقد زواج في استراليا مصنفة وفق عمر العروس.

التكرار	مركز الفئات	التكرار	مركز الفئات
1655	55.5	294	16.5
1100	58.5	10995	19.5
810	61.5	61001	22.5
649	64.5	73054	25.5
487	67.5	56501	28.5
326	70.5	33478	31.5
211	73.5	20569	34.5
119	76.5	14281	37.5
73	79.5	9320	40.5
27	82.5	6236	43.5
14	85.5	4770	46.5
5	88.5	3620	49.5
		2190	52.5

301785 = مجموع التكرارات



شكل (١ ـ ٧). مضلع التكرار للبيان الإحصائي في الجدول (١ ـ ١١)



شكل (١ ـ ٨). منحنى تكرار من أسرة النموذج جاما

تمــريـــن

ارسم منحنيا ملتويا إلى اليسار.

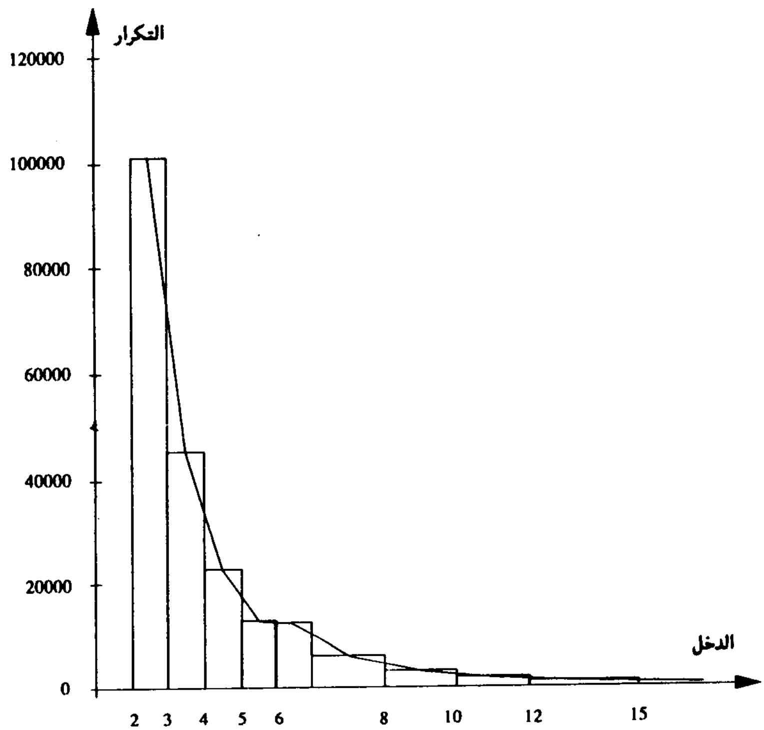
يبين الجدول (١ ـ ١٢) توزيع التكرار لـ 213938 شخصا في المملكة المتحدة مصنفين وفقا لشرائح الدخل مقدرة بآلاف الجنيهات.

جدول (١ ـ ١٢). توزيع التكرار وفق فئات الدخل لـ 213938 شخصا في المملكة المتحدة

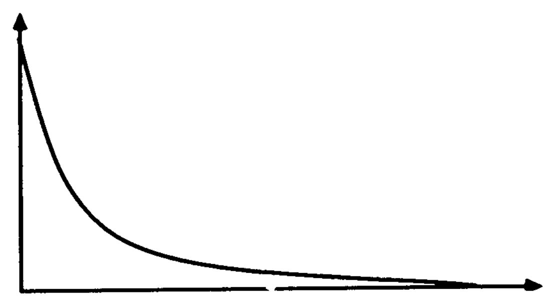
التكرار (عدد الأشخاص)	فئات الدخل بآلاف الجنيهات
101369	2 -
45532	3 -
23263	4 -
13475	5 -
13456	6 -
6419	8 -
3551	10 -
2926	12 -
2007	15 -
820	20 -
399	25 -
376	30 -
134	40 -
128	50 -
45	75 -
38	100 -

المجموع = 213938

ويقترح مضلع التكرار منحنى تكرار مناسب لهذه الظاهرة (توزع فئات الدخل في المملكة المتحدة) من النوع J. وتسمى هذه الأسرة من النماذج بأسرة النماذج الأسية. وهي تبدأ بقمتها ثم تنحدر بسرعة متقاربة إلى محور السينات. ونجد في الشكل (١ ـ ١٠) منحنى تكرار من أسرة النموذج الأسي.



شكل (١ ـ ٩). مدرج التكرار ومضلع التكرار للبيان الإحصائي في الجدول (١ ـ ٨)



شكل (١ ـ ١٠). منحنى تكرار من أسرة النموذج الأسي

تمارين (١ ـ ١)

١) تتغير أوزان خمسين طالبا مقاسة إلى أقرب «باوند» من 177 إلى 265. إذا أردت تصنيف هذه الأوزان في عشر فئات فاكتب حدود الفئات، والحدود الحقيقية للفئات؛ ومراكز الفئات. ما طول الفئة؟

٢) كانت مراكز الفئات لتوزيع تكراري لمجموعة من قياسات درجة الحرارة مأخوذة إلى أقرب درجة مئوية، كما يلي:

16, 25, 34, 43, 52, 61

أوجد:

أ ـ حدود الفئات؛ ب ـ الحدود الحقيقية للفئات.

٣) فيما يلي عدد الأميال التي قطعتها كل من أربعين سيارة إسعاف بجالون واحد من البنزين:

24.5	23.6	24.1	25.0	22.9	24.7	23.8	25.2	24.9
24.1	23.7	24.4	24.7	23.9	25.1	24.6	23.3	24.3
24.8	22.8	24.6	23.9	24.1	24.4	24.5	25.7	23.6
24.0	24.7	23.1	23.9	24.2	24.7	24.9	25.0	24.8
24.5	23.4	24.6	25.3					

أ ـ لخص هذا البيان الإحصائي في جدول توزيع تكراري متخذا الفئات:

22.5 - 22.9; 23.0 - 23.4, . . ., 25.5 - 25.9

ب ـ ارسم مدرج التكرار، ومدرج التكرار النسبي، ومضلع التكرار.

جـ ـ اكتب جدول التكرار المتجمع الصاعد.

د ـ ارسم مضلع التكرار المتجمع الصاعد

هـ ـ ما عدد القياسات التي هي أقل من 23.75؟ ، أكثر من 23.45؟ ، أقل من 24.3؟ ، وأقل من 25.2؟

و ـ ما القياس الذي يقل عنه خمسون بالمئة من القياسات؟ خمس وعشرون بالمئة من القياسات؟ وخمس وسبعون بالمئة من القياسات؟

٤) فيما يلي درجات 40 طالبا في اختبار ١٠٦ إحص :

42	88	37	75	98	93	73	62	96	80
52	76	66	54	73	69	83	62	53	79
69	56	81	75	52	65	49	80	67	59
88	80	44	71	72	87	91	82	89	79

أ ـ اكتب جدول التوزيع التكراري لهذا البيان الإحصائي مستخدما الفئات :

35 - 39; 40 - 44; . . .

ب ـ ارسم مدرج التكرار، ومدرج التكرار النسبي ، ومضلع التكرار، مستخدما ورقة بيانية .

جـ ـ اكتب جدول التكرار المتجمع الصاعد وارسم مضلعه .

د ـ ما القياس المقابل لتكرار متجمع يساوي العشرين بالطريقتين الحسابية والبيانية؟

*٥) يتولى الإشراف الصحي على عدد من مدارس تعليم البنات 44 وحدة صحية منتشرة في أنحاء المملكة . وفيما يلي عدد المدارس المرتبطة بكل من هذه الوحدات الصحية (لا يتضمن البيان مدارس الرياض وجدة والإحساء ومكة المكرمة) :

23,	46,	20,	30,	28,	12,	35,	50,	33,	65,	85,
24,	40,	50,	23,	40,	30,	50,	23,	20,	38,	68,
58,	15,	15,	100,	105,	6,	59,	36,	22,	89,	21,
35,	100,	42,	38,	58,	32,	62,	48,	32,	19,	56

* مأخوذ من التقرير الصحي السنوي الصادر عن وزارة الصحة لعام ١٤٠٦هـ، ص ٢٨٧ .

متخذا الفئات 105 - 90 ,89 - 78 ,...,41 - 30 ,29 - 18 ,17 - 6

أ ـ ارسم مدرج التكرار النسبي .
ب ـ ارسم مضلع التكرار المتجمع الصاعد
جـ ـ أوجد حسابيا وبيانيا المئين 10 والمئين 90 .

٦) عند تلخيص بيان إحصائي حصلنا على التوزيع التكراري التالي :

حدود الفئات	10 - 24	25 - 39	40 - 54	55 - 69	70 - 84	85 - 99
التكرار	15	25	42	50	38	30

أ ـ اكتب التكرار النسبي معبرا عنه في نسبة مئوية .
ب ـ اكتب جدول التكرار المتجمع الصاعد ، والتكرار النسبي المتجمع الصاعد .
جـ ـ ارسم مضلع التكرار المتجمع الصاعد .

٧) فيما يلي أوزان ستين فأرا (مقاسة إلى أقرب غرام) استخدمت في دراسة تجريبية تتعلق بنقص الفيتامين :

125	128	106	111	116	123	119	114	117	143
136	92	115	121	118	137	132	120	104	125
119	115	101	87	129	108	110	133	135	126
127	103	110	118	126	82	104	137	120	95
146	126	119	105	119	132	126	118	100	113
106	125	102	146	117	129	124	113	95	148

أ ـ لخص هذا البيان الإحصائي في جدول توزيع تكراري متخذا الفئات :

80 - 89; 90 - 99; . . .; 140 - 149

ب ـ ارسم مدرج التكرار، ومدرج التكرار النسبي ، ومضلع التكرار.
جـ ـ اكتب جدول التكرار المتجمع الصاعد . والتكرار المتجمع الصاعد النسبي .
د ـ ارسم مضلع التكرار المتجمع الصاعد مستخدما ورقة بيانية .
هـ ـ ما نسبة القياسات التي هي أقل من 109.5؟ ، أكثر من 89.5؟ ، أقل من 133؟ ، وأقل من 105؟
و ـ ما هو القياس الذي يقل عنه ستون بالمئة من القياسات؟ خمس وثلاثون بالمئة من القياسات؟ خمس وعشرون بالمئة من القياسات؟ وخمس وسبعون بالمئة من القياسات؟

٨) مستخدماً فئات طولها 2 مم، اكتب توزيع التكرار وتوزيع التكرار النسبي لسماكة الجلد المعطاة في البيان التالي. (القياسات تمثل سماكة الجلد بالمليمتر في منتصف عضلة الذراع لـ 121 ذكرا).

11.4	15.3	9.1	18.4	10.9	4.7	9.6	20.6	10.4	20.5	22.4	14.3
11.7	11.4	12.7	18.2	15.1	14.6	25.3	11.5	13.2	7.9	12.6	13.9
16.8	11.4	27.3	16.3	13.9	13.2	11.9	20.0	13.2	9.4	18.9	10.7
14.8	17.8	10.8	16.0	15.7	17.7	13.5	11.5	11.1	9.6	15.1	13.6
13.6	8.6	6.9	19.1	18.7	10.1	16.1	20.4	7.9	16.6	18.5	16.2
17.4	18.8	12.6	22.0	9.6	11.1	15.7	23.7	13.3	4.9	8.3	20.1
15.5	23.1	10.2	10.7	15.8	17.6	21.3	16.2	14.9	9.9	9.1	9.9
9.8	8.6	11.8	9.3	14.8	17.3	9.5	13.6	12.4	9.5	14.3	25.7
12.9	22.7	12.1	10.7	16.8	11.3	11.3	11.4	5.9	10.7	14.6	19.8
25.5	7.7	18.4	7.9	7.6	23.3	9.6	8.4	10.4	8.1	12.5	9.1
30.1											

٩) بالعودة إلى المثال (١ ـ ٢)، استخدم الفئات 13.9 - 12.0 ، 15.9 - 14.0 ، . . . الخ . لوضع جدول توزيع تكراري لقياسات مستوى الهيموغلوبين في الدم لتسعين عاملا يعيشون في مناطق ترتفع ارتفاعا شاهقا عن سطح البحر.

ارسم مـدرج التكـرار، ومضلع التكـرار، ومضلع التكـرار المتجمع الصـاعـد، واحسب النسبة من القياسات التي تقل عن 16.5 .

١٠) فيما يلي جدول تـوزيع تكـراري لمستوى الهيمـوغلوبين في الـدم لـ 122 عاملا ممن يعيشون في مناطق لا ترتفع كثيرا عن سطح البحر.

حدود الفئات	11.0-11.9	12.0-12.9	13.0-13.9	14.0-14.9	15.0-15.9	16.0-16.9	17.0-17.9
التكرار	6	21	29	43	19	3	1

أ ـ ارسم مدرج التكرار النسبي .
ب ـ ارسم مضلع التكرار النسبي المتجمع الصاعد .
جـ ـ احسب النسبة من القياسات التي تقل عن 16.5 .

١١) فيما يلي جدول توزيـع تكراري للعمر عند الـوفاة مقاسا إلى أقـرب سنة لـ 302 من المرضى الذين توفوا وهم مصابون بالحمى القرمزية :

حدود الفئات	0 -	1 -	2 -	3 -	4 -	5 -	6 -	7 -	8 -	9 -	10 -	15-20
التكرار	18	43	50	60	36	24	22	21	6	5	14	3

ارسم مـدرج التكرار ومضلع التكـرار. ما العمـر الذي تقل عنـه نسبة 90% من حالات الوفاة؟

١٢) فيما يلي جدول توزيع تكراري لحالات الوفاة بسرطان الدم عند الأطفال مصنفة وفقا للعمر مقاسا إلى أقرب سنة . (الولايات المتحدة عام ١٩٧٠م).

حدود الفئات	(0-0.5)	[0.5-1.5)	[1.5-2.5)	[2.5-3.5)	[3.5-4.5)	[4.5-9.5)	[9.5-14.5)
التكرار	68	82	98	137	196	684	434

ارسم مدرج التكرار النسبي .

١٣) فيما يلي قياسات معدل الكولستيرول في الدم لخمسمائة رجل في الأربعينات من عمرهم (49 - 40) مقاسة بالمليليغرام لكل 100 ميلليلتر:

289	385	306	278	251	287	241	224	198	287
275	301	249	288	337	263	260	228	190	282
368	291	249	300	268	283	319	284	205	294
257	256	294	253	221	241	372	339	292	294
327	195	305	253	251	229	250	348	280	378

282	311	193	242	304	270	277	312	264	262
268	251	333	300	250	234	264	291	271	284
322	381	276	205	251	270	254	299	273	252
280	411	195	256	387	241	245	325	289	306
232	293	285	250	260	316	352	309	229	261

272	196	317	188	215	265	266	217	223	354
169	278	188	252	264	314	246	335	377	305
249	318	270	261	324	289	215	228	315	253
262	250	361	304	248	202	284	291	305	261
292	259	369	289	320	287	230	259	321	268

208	386	298	325	262	326	265	281	262	214
277	248	314	279	279	223	202	188	276	261

318	272	245	285	301	234	420	299	255	285
271	283	260	300	308	319	226	235	318	304
291	388	242	277	235	262	176	226	289	247
389	349	210	241	230	260	324	214	296	279
256	260	250	308	294	320	343	312	224	259
305	286	264	209	233	167	272	274	316	291
289	288	175	260	334	248	287	247	222	300
307	269	311	275	273	272	309	307	233	258
263	293	211	263	281	248	349	225	226	388
332	223	186	190	256	321	297	262	380	337
309	227	164	275	283	268	329	259	247	311
246	253	257	328	242	224	283	249	189	207
312	271	277	311	273	316	360	252	243	311
288	226	329	174	248	305	247	309	323	299
174	215	299	183	187	260	268	293	324	325
282	283	324	284	274	285	299	270	354	290
222	280	210	243	199	262	300	218	224	360
293	221	203	386	282	270	277	227	287	226
262	281	319	279	324	279	178	218	246	274
237	239	251	245	337	249	234	202	341	264
281	243	280	346	245	262	213	312	281	312
261	279	356	329	216	326	269	290	300	338
253	284	306	274	277	353	291	333	280	346
270	289	296	296	269	269	275	217	220	351
260	336	323	246	295	296	285	280	330	258
233	219	225	220	210	308	340	319	217	195

262	219	255	278	359	264	273	238	268	301
260	253	237	271	251	226	281	252	338	310

373	217	204	263	246	334	184	222	294	213
331	354	286	291	223	197	324	367	317	253
367	330	315	260	231	266	286	216	286	353
324	315	271	313	306	287	267	274	290	172
275	262	329	283	300	296	238	325	256	244

لاحظ أن كل جزء من الأجزاء العشرة في هذا البيان يتضمن خمسين قياسا .

أ ــ اختر جزءا من الأجـزاء العشرة وقم بتصنيفـه . ثم ارسم لـه مدرج تكـرار نسبي مستخدما الفئات 189 - 160 ، 219 - 190 ، . . . الخ .

ب ــ ليقم كل اثنين أو ثلاثـة من طلاب الفصل بتنفيـذ السؤال في جـزء محدد من الأجـزاء العشرة من البيـان وبحيث يتم رسـم مـدرج تكـرار نسبي لكل جـزء منها .

جــ ـ قم بضم نتـائج الأجزاء العشرة بعضهـا إلى بعض وارسم مدرج تكـرار نسبي للبيان بكـامله ، ثم انظر نظـرة مقارنة بين مـدرجات التكرار النسبي لـلأجزاء ومدرج التكرار النسبي للبيان بكامله .

١٤) فيما يلي معـدل الـولادة الخام ومعـدل الـوفـاة الخام في انكلترا وويلـز بين 1926 إلى 1976 . وكـذلك الفـرق بين المعـدلين ، ويسمى معدل الـزيـادة الطبيعيـة . اكتب جدول التـوزيع التكراري لكل منها ، وارسم المضلع التكـراري . انظر نظرة مقـارنة بين المضلعـات الثلاثـة . (يمكنك أخذ تسع فئـات طول كل منهـا 1 في معدلات الولادة ومعدلات الزيادة ، وخمس فئات طول كل منها 0.7 في معدلات الوفاة) .

معدل الولادة			معدل الوفاة			معدل الزيادة		
17.8	15.8	14.8	11.6	12.3	12.1	6.2	3.5	2.7
16.7	15.3	14.9	12.3	12.0	12.4	4.4	3.3	2.5
16.7	14.4	15.1	11.7	12.3	11.6	5.0	2.1	3.5
16.3	14.8	14.6	13.4	11.8	12.1	2.9	3.0	2.7
16.3	14.7	14.1	11.4	11.7	14.4	4.9	3.0	0.3
13.9	19.2	15.5	13.5	12.0	12.5	0.4	7.2	3.0
15.6	20.5	15.3	12.3	12.3	11.3	3.3	8.2	4.0
16.2	17.8	15.5	13.0	11.0	11.4	3.2	6.8	4.1
17.2	16.7	15.2	12.7	11.8	11.3	5.0	4.9	3.9
15.9	15.8	15.0	12.6	11.6	11.7	3.3	4.2	3.3
15.7	17.6	17.8	11.7	11.9	11.7	4.0	5.7	6.1
16.1	18.0	17.3	11.5	11.9	11.2	4.6	6.1	6.1
16.4	18.2	16.9	11.7	12.2	11.9	4.7	6.0	5.0
16.5	18.6	16.4	11.6	11.3	11.9	4.9	7.3	4.5
17.1	18.1	16.1	11.5	11.5	11.7	5.6	6.6	4.4
16.0	11.9		11.6	12.0		4.4	-0.1	
14.8			12.0			2.8		
13.7			11.8			1.9		
13.0			11.8			1.2		
12.2			11.7			0.5		

١٥) فيما يلي أوزان 18645 طفلا مولـودا في جنوب غرب انكلترا (أحياء أو أمـوات) عام ١٩٦٥م مستخدما فئات طولها 1 باوند، اكتب التوزيـع التكراري وتوزيع التكرار النسبي. ارسم مدرجا تكراريا ومضلعا تكراريا لتوضيح البيان.

أونزة / باوند	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
1	6	1	1	1	3	0	2	2	3	1	3	4	8	2	2	1
2	18	4	2	2	6	2	4	2	10	4	4	2	8	7	4	3
3	14	6	8	5	9	6	8	9	14	2	6	6	7	5	14	7
4	22	14	16	19	16	14	15	19	47	17	23	15	39	30	26	32
5	66	37	42	46	60	41	67	59	106	78	98	68	135	92	106	81
6	323	101	183	157	337	160	205	172	504	215	299	222	496	256	315	228
7	914	225	390	286	697	311	417	291	817	289	369	279	626	246	330	236
8	920	195	292	220	508	200	230	166	485	147	198	110	288	122	146	78
9	395	83	118	72	142	53	69	45	145	35	42	22	91	18	25	10
10	88	12	26	9	23	11	6	4	18	8	7	2	16	4	2	4
11	17	1	3	2	3	1	0	2	2	0	4	1	2	0	1	0
12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

١٦) تم تنفيذ برنامج استئصال للملاريا في إحدى القرى. وفيما يلي جدول توزيع يعطي النسبة المئوية لقياس الهيموغلوبين في عينة من سكان هذه القرية قبل تنفيذ برنامج الاستئصال. وفي البيان الإحصائي قياسات الهيموغلوبين في عينة أخذت بعد تنفيذ برنامج الاستئصال. اكتب توزيعا مماثلا للبيان الإحصائي الخاص بعينة ما بعد تنفيذ البرنامج. استعن بالرسوم التي تجدها مناسبة.

نسبة الهيموغلوبين	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109	المجموع
التكرار	2	7	14	10	8	2	2	0	45
التكرار النسبي (مئويا)	4.4	15.6	31.1	22.2	17.8	4.4	4.4	0	99.9

البيان الإحصائي لعينة ما بعد تنفيذ البرنامج

43	63	63	75	95	75	80	48	62	71	76	90	51	61	74
103	93	82	74	65	63	53	64	67	80	77	60	69	73	76
91	55	65	69	84	78	50	68	72	89	75	57	66	79	85
70	59	71	87	67	72	52	35	67	99	81	97	74	61	72

١٧) استـدعت الدراسـات التفصيليـة لأحد الأمـراض في إحـدى القرى إجـراء حصر شامـل للسكان. وفيما يلي التـوزيع التكراري لعـدد الذكـور مصنفين وفقا لشرائح العمر في هذه القرية:

العمر	عدد الذكور	النسبة المئوية ٪
0 - 4	154	18.6
5 - 9	135	16.3
10 - 14	107	12.9
15 - 19	72	8.7
20 - 29	112	13.5
30 - 39	97	11.7
40 - 49	67	8.1
50 - 59	47	5.7
60 - 79	39	4.7
المجموع	830	100.2

أ ـ ارسم مدرج التكرار النسبي لهذا التوزيع.

ب ـ اكتب جدول التكرار النسبي المتجمع الصاعد. وارسم مضلعه.

جـ ـ من الرسم البياني حدد العمر الذي يقسم المجتمع بنسبة 50 - 50. أي ما العمر الذي يمكن القول أن 50% من المجتمع أعمر منه؟

*١٨) فيما يلي عـدد الأطبـاء العـاملين وعـدد الأسرة في كل من واحـد وعشرين من المستشفيات في منطقة الرياض:

* مأخوذ عن التقرير الصحي السنوي الصادر عن وزارة الصحة لعام ١٤٠٦هـ، صفحة ٧٤.

عدد الأطباء	59	85	44	12	12	18	85	51	34	16	28	50
عدد الأسرة	200	266	263	200	160	124	230	205	187	30	72	222

عدد الأطباء	32	34	25	35	43	33	15	30	24
عدد الأسرة	130	115	45	124	146	110	15	100	31

أ ـ اكتب جـدول توزيع تكراري لعدد الأطبـاء متخذا الفئات 12 - 26, 27 - 41, . . . ,
وجدول توزيع تكراري لعدد الأسرة متخذا الفئات 15 - 64, 65 - 114, . . . , .

ب ـ ارسم مدرج التكرار لكل منهما.

(١ ـ ٦) استخدام بعض الرموز الإحصائية

إن استخدام الرمـوز للتعبير عن بيان إحصائي، ومعرفـة القواعد التي تخضع لها هذه الـرموز، يسـاعد على التعبير بـاختصار عن خصـائص مهمة للبيـان الإحصائي، واستنباط خطوات العمل الحسـابي للوصول إلى القيم العددية لهذه الخصـائص. والرمز الأكثـر استخدامـا في الإحصاء هـو رمز المجمـوع Σ [انظـر البنـد (٥) من الملحق]. والقياسات في بيان إحصائي هي، بصورة عامة، قيم عددية لمتغير نعبر عنه بحرف x أو y أو z أو أي حـرف آخـر، وهـو يقيس الصفـة أو الخاصيـة التي يـدور حـولها البيـان الإحصائي، كأن نقيس، مثلا، وزنا أو طولا، أو نسجل عمرا أو معدلا، أو عدد مرات وقـوع شيء معين خلال فترة معينـة إلخ، ويمكننا التعبير رمـزيـا عن بيان إحصـائي لم نحصل عليه بعد، وإنما نخطط للحصول عليه، بحروف $x_1, x_2, \ldots, x_n$ حيث n عدد القياسات التي نـريد الحصول عليها، و x_1 هو رمـز لأول قياس سنحصـل عليه، و x_2 رمز للقياس الثاني، وهكذا . . . ، بينما x_n هو رمز لآخر قيـاس سنسجله. ولو حصل

أن كان العدد الأول الـذي نسجله (عند تنفيذ التجربـة أو جمع البيان الإحصائي) 181، مثلا، فعندئذ نقول إن $x_1 = 181$، وهكذا . . . ، ومن الطبيعي أن يتكـرر حصولنا على القمية نفسها أكثر من مرة. فلو فرضنا، مثلا، أن $x_1 = x_3 = x_6 = 181$، لقلنا إن القيمة 181 مكـررة ثلاث مـرات. وتجنبا لـلالتبـاس يمكن أن نستخدم حـرفا آخـر y، مثلا، للدلالة على القيم المختلفة التي ورد ذكـرها في البيان، ونكتب في هذه الحالة $y_1 = 181$، ونقول إن y_1 مكررة ثلاث مرات.

وكما نعلم فإن التـوزيع التكراري هـو أي تـرتيب للبيان الإحصـائي يظهر تـوزع قيـاساتـه على فترات معرفـة بصورة اختيـارية تسمى «فئـات». وإذا كان كل عـدد من الأعداد المختلفـة في بيان إحصـائي يمثل فئـة بحد ذاتـه، فسنقول إننـا في حالـة «بيان مرتب» وفيما عـدا ذلك سنقول للتمييـز إننا في حالـة «بيان مصنف» أو «بيـان مبوب». وإذا قمنا بترتيب جملة من القياسات فستأخذ بعد الترتيب الشكل التالي:

جدول (١ ـ ١٣) بيان مرتب

y_i (القيم المختلفة)	y_1	y_2	.....	y_m
f_i (التكرار)	f_1	f_2	.....	f_m

أي أن هناك m قيمة مختلفـة فقط في البيان الإحصـائي الذي يتضمن n قيمة $(n > m)$.
ومن الواضح أنه يمكن التعبير عن مجموع قيم البيان الإحصائي بشكلين متكافئين:

$$\sum_{i=1}^{n} x_i = \sum_{j=1}^{m} f_j \; y_j$$

والطرف الأيمن تعبير عن عمليات جمع مكرر للعدد نفسه. فإذا كان القياس 181 مكررا ثلاث مرات، فسيكـون من الأيسر، عند حساب مجموع القياسـات، كتابة 181×3 بدلا من $181 + 181 + 181$. وبصورة عامة، إذا كان أحد القياسات في الطرف الأيسر مكررا f_j

مرة، فقد رمزنـا لهذا القياس المكرر بـ y_j وبدلا من جمع y_j عددا مـن المرات يساوي f_j، كتبنا في الطرف الأيمن $f_j\ y_j$.

أما البيان المصنف فسيأخذ، لأغراض حسابية، الشكل التالي:

جدول (١ ـ ١٤). بيان مصنف

(مركز الفئة) y_i	y_1	y_2	.....	y_m
(التكرار) f_i	f_1	f_2	.....	f_m

وهذا يشير إلى أننـا صنفنا (أو بوبنا) قيم البيـان الإحصائي في m فئة، واعتبرنا مركز كل فئة ممثلا لجميع القياسات التي تنتمي إلى هـذه الفئة، وبذلك استعضنا عن الـ f_j قياسا في الفئة الأولى بمركـز الفئة y_1 واعتبرناه مكررا f_1 مرة واستعضنـا عن الـ f_2 قياسا في الفئة الثانية بمركز هذه الفئة y_2 واعتبرناه مكررا f_2 مرة . . . وهكذا. وعلى سبيـل المثال، لو عدنـا إلى الجدول (١ ـ ٦)، وهـو جدول التوزيع التكـراري لقياسـات حاصل الـذكاء لخمسين مستجدا، وأخـذنا القياسات الفعليـة السبعة ضمن الفئة الـرابعة 101 - 97، لوجدنا أنها:

97, 99, 101, 101, 100, 97, 101

ومجموعها الفعلي هو 696. ولكن الجدول (١ ـ ٦) تلخيص للبيان الإحصائي يغنينا عن العودة إلى مفرداته، حتـى في الحسابات العددية. وإذا أردنا حسـاب مجموع القياسات ضمن هذه الفئة فإننا نتخذ مـركز الفئة، وهو هنا 99، ممثلا لجميع القياسات السبعة، أي نفترض القياسات السبعة في هذه الفئة كأنها:

99, 99, 99, 99, 99, 99, 99

ونعتبر مجموع الفئة مساويا لـ $693 = 99 \times 7$. ونلاحظ أننا ارتكبنا خطأ بالنقصان قدره 3، وهو الثمن الذي ندفعه في مقابل كفاءة العرض وسهولة وسرعة الحسابات. ومن حسن الحظ أن الأخطاء في الفئات المختلفة لا تكون، عادة، في اتجاه واحد، فلا تكون جميعا أخطاء بالنقصان أو أخطاء بالزيادة، بل تكون في بعض الفئات أخطاء بالنقصان، وفي بعضها الآخر أخطاء بالزيادة، وبذلك يعدل بعضها بعضا، ويكون الخطأ الإجمالي تافها بالمقارنة مع الوفر الكبير الذي حققناه في عملية تصنيف أو تلخيص البيان في هيئة توزيع تكراري، ناهيك عن وضوح العرض وكفاءته سواء في جدول التوزيع التكراري نفسه، أم فيما ينبثق عنه من جداول ورسوم بيانية.

وينبغي أن يكون هذا كافيا لإيضاح نقطة، وهي أن جدولا كالجدول (١ ـ ١١)، نعتمده لحساب خصائص معينة لبيان إحصائي لا يعطي قيم هذه الخصائص بالضبط، كما لو كنا استخدمنا في الحسابات مفردات البيان نفسه، وإنما يعطي تلك القيم بصورة تقريبية، وبفارق زهيد يمكن إغفاله. وفي الجدول (١ ـ ١١) لو رمزنا بـ $\sum_{j=1}^{n} x_j$ لمجموع قياسات البيان الإحصائي الأصلي (قبل التصنيف)، فسيكون المجموع $\sum_{i=1}^{m} f_i y_i$، كما نأخذه من الجدول، قيمة تقريبية للقيمة المضبوطة $\sum_{j=1}^{n} x_j$.

تمرين

احسب مجموع القياسات الخمسين في الجدول (١ ـ ٤) وقارنه مع المجموع الناتج عن استخدام التوزيع التكراري في الجدول (١ ـ ٦).

(١ ـ ٧) مقاييس النزعة المركزية

لا شك في أن الطرق البيانية مفيدة للغاية عند تقديم المعلومات الإحصائية، وأنها تنقل وصفا عاما وسريعا لتلك المعلومات، مما يتفق مع المثل القائل بأن صورة

واحـدة تساوي ألف كلمـة. إلا أن هنـاك حدودا، على أي حـال، لاستخدام الطـرق البيانيـة في مجال وصف وتحليل المعلـومات. وعلى سبيل المثـال، لنفرض أننا نـرغب في مناقشـة البيان الإحصائي أمام مجمـوعة من الناس، وأنـه ليس لدينا طريقـة أخرى غير الطريقـة الشفهية، مما يجعل عـرض المضلع التكراري غير ممكن، ويضطرنـا لاستخدام مقـاييس وصفيـة أخـرى يمكنهـا أن تنقـل إلى المستمعين صـورة ذهنيــة عن المضلع التكراري. والأمر الثاني الذي يضع حدا لاستخدام الطرق البيانية هو صعوبة الاستفادة منهـا في مجال الاستقراء الإحصـائي. وربما اقتصرت فوائدهـا الاستقـرائية على أن يقـدم المضلع التكـراري لعينـة من القيـاسـات نقـوم بتلخيصهـا، تصـورا عن شكل المضلع التكراري للمجتمع من القياسات الذي جاءت منه العينة.

وإذا كنا أمام جملة من القياسات فإن أول مـا تجدر معرفته هو القيمة التي تتمركز عندها القياسات. ومن الملاحظ، مثلا، أنـه في كثير من الظواهر السلوكية والاجتماعية تنزع معظم القيـاسات إلى التمـركز حـول قيمة وسطيـة، فأولئك الذين يتصفـون بحدة شديـدة في المزاج هم قلة وفي المقابل نجـد ذوي المزاج المفرط في برودته قلـة أيضا وذلك قيـاسا على الجمهـرة من النـاس التي تقع بين بين. وأولئك الذين يتصفـون بـالنحافـة الشديـدة يقابلهم أولئك المتصفون بسمانـة مفرطة هم قليلون بـالقياس إلى عامـة الناس التي تحتل مواقعها بين بين. والملاحظة نفسها نجدها سائدة في مجال توزيع الأطوال بين عمالقة وأقزام. فمعظم النـاس في مجتمع بشري معين تميل أطوالها إلى اتخاذ موقع وسط، وقس على ذلك. ولو طبقنا اختبـارا لقياس حاصل الذكـاء على طلاب الجامعة بأسرهم لوجـدنا أن المتفوقين الموهـوبين قلة والمبتلين بالبـلادة قلة، وينـزع حاصل الذكـاء عند معظم الطلبة إلى التمركز حول الوسط.

وفي حياتنـا اليومية، كثيرا ما نستخـدم كلمة «في المتوسط» فنتحـدث عن الرجل «متوسـط الدخل»، والشـاب «متوسط الثقـافة». وقـد يقول أحدنـا: «نادرا مـا أصل متأخرا إلى مقر عملي ونادرا مـا أصل إليه مبكرا، وفي المتوسط يتفق موعـد وصولي تقريبا مع بداية الدوام الرسمي». كما نقول: «إن استهلاكي اليومي من القهوة (أو الشاي) هو في المتوسط كـذا» إلخ. وهذه الاستخـدامات الشائعـة لكلمة متـوسط تعبر عن شعور

داخلي معين يحسه ويفهمه كل منا ولا يستطيع ترجمته بدقة. ومقاييس النزعة المركزية هي محاولة لترجمة هذا الشعور بطريقة دقيقة ومحددة تماما.

وفي لغة الإحصاء يعبر مقياس النزعة المركزية عن القيمة (أو الموضع أو النقطة) التي يتمركز عندها التوزيع التكراري لجملة من القياسات. وعادة ما تحتشد بقية القياسات أكثر ما تحتشد حول ذلك الموضع. وإذ نهتم عادة بمقياس نزعة مركزية لمجتمع من القياسات نلجأ في الغالب إلى عيّنة من المجتمع ونحسب قيمة ذلك المقياس من أجل قياسات العينة ثم نعتبر هذه القيمة التي حصلنا عليها تقديرا أو تخمينا لقيمة المقياس التي نجهلها والخاصة بالمجتمع الذي جاءت منه العينة. وسنستعرض هنا ثلاثة أشكال لقياس النزعة المركزية لجملة من القياسات هي المتوسط والوسيط والمنوال.

(١ ـ ٧ ـ ١) المتوسط (الوسط الحسابي)

والمقياس الأكثر فائدة والأكثر استخداما للنزعة المركزية لجملة من القياسات هو معدلها الحسابي. ويشار إليه غالبا بالوسط الحسابي أو المتوسط.

تعريف المتوسط

متوسط n من القياسات $x_1, x_2, \dots, x_n$ هو مجموع هذه القياسات مقسوما على عددها. وبصورة رمزية نكتب:

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

وإذا لم يكن هناك خشية التباس يمكن كتابة هذه العلاقة على الشكل:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1}{n}\Sigma x$$

حيث Σx يعني مجموع القيم التي يأخذها المتغير x كافة وعددها n.

مثال (١ ـ ٥)

احسب متوسط القيم 1, 12,14, 6, 3

الحل

$$\bar{x} = \frac{1 + 12 + 14 + 6 + 3}{5} = 7.2$$

ونلاحظ من التعريف مباشرة أن :

$$\sum_{i=1}^{n} x_i = n\bar{x}$$

وفي حالة بيان مرتَّب نعبر عن مجموع القياسات على الشكل : (انظر الجدول ١ ـ ١١).

$$f_1\ y_1 + f_2\ y_2 + \ldots + f_m\ y_m = \sum_{i=1}^{m} f_i\ y_i$$

وتصبح العلاقة المذكورة في التعريف السابق للمتوسط كما يلي :

$$\bar{y} = \frac{\sum_{i=1}^{m} f_i\ y_i}{\sum_{i=1}^{m} f_i}$$

أما في البيانات المصنفة (أو المبوبة) فنفترض أن جميع القياسات التي تنتمي إلى فئة مساوية لمركز هذه الفئة. والخطوات الحسابية ليست إلا تطبيقا للعلاقة الأخيرة من أجل بيان مرتب حيث y_i الآن هي مركز الفئة i، و f_i التكرار الموافق لهذه الفئة. وللتوضيح نأخذ المثال التالي :

مثال (١ ـ ٦)

احسب متوسط حاصل الذكاء في المثال (١ ـ ٣) مستخدما جدول التوزيع التكراري (١ ـ ٦).

الحل

لحساب المتوسط ننظم الجدول التالي :

جدول (١ ـ ١٥). حساب متوسط البيان المصنف في الجدول (١ ـ ٦)

مركز الفئة y_i	التكرار f_i	$f_i y_i$
84	1	84
89	2	178
94	4	376
99	7	693
104	9	936
109	10	1090
114	7	798
119	6	714
124	4	496
المجموع	50	5365

$$\bar{y} = \frac{\sum_{i=1}^{9} f_i \; y_i}{\sum_{i=1}^{9} f_i} = \frac{5365}{50} = 107.30$$

لاحظ أنك عندما تحسب المتوسط من البيان الإحصائي الأصلي في الجدول (١ ـ ٤) ستجد :

$$\bar{x} = \frac{\sum_{i=1}^{50} x_i}{50} = \frac{5364}{50} = 107.28$$

والفرق بين النتيجتين لا يذكر في مقابل الوفر في الجهود الحسابية اللازمة .

(١ ـ ٧ ـ ٢) خواص المتوسط

١ ـ مجموع انحرافات جملة من القياسات عن متوسطها يساوي الصفر.

ولبيان ذلك لنرمز بـ d_i للإنحراف $x_i - \bar{x}$ أي انحراف القياس x_i عن المتوسط $\bar{x}$. ولنحسب مجموع الإنحرافات Σd_i فنجد:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = n\bar{x} - n\bar{x} = 0$$

وهذه الخاصة توضح الدور المركزي الذي يلعبه المتوسط.

٢ ـ يكون مجموع مربعات انحرافات القياسات عن قيمة a أصغر ما يمكن* عندما يكون $a = \bar{x}$.

لنأخذ مجموع مربعات انحرافات القياسات عن قيمة ما a ، أي $\sum_{i=1}^{n} (x_i - a)^2$، فيمكن كتابة:

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

$$= \sum_{i=1}^{n} \left[(x_i - \bar{x}) + (\bar{x} - a)^2 \right]$$

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

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

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - a)^2 \geq \sum_{i=1}^{n} (x_i - \bar{x})^2$$

* البرهان للقراءة فقط.

ذلك لأن $n(\bar{x}-a)^2$ كمية غير سالبة . أي أن مجموع مربعات الانحرافات عن قيمة ما (a) هو دائما أكبر من مجموع مربعات الانحرافات عن المتوسط $\bar{x}$ أو يساويه .

مثال (١ ـ ٧)

في المثال (١ ـ ٥) احسب مجموع الانحرافات عن المتوسط و $\sum_{i=1}^{5}(x_i-7)^2$ و $\sum_{1}^{5}(x_i-7.3)^2$ ثم تحقق من الخاصتين ١ و٢ .

الحل

ننظم جدولا كما يلي :

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	(x_i-7)	$(x_i-7)^2$	$x_i-7.3$	$(x_i-7.3)^2$
1	−6.2	38.44	−6	36	−6.3	39.69
12	4.8	23.04	5	25	4.7	22.09
14	6.8	46.24	7	49	6.7	44.89
6	−1.2	1.44	−1	1	−1.3	1.69
3	−4.2	17.64	−4	16	−4.3	18.49
المجموع	0	126.80	1	127	-0.5	126.85

ونلاحظ أن مجموع العمود الثاني صفر بما يتفق مع الخاصية ١ ، وأن كلا من مجموعي العمودين الخامس والسابع أكبر من مجموع العمود الثالث بما يتفق مع الخاصية ٢ .

٣ ـ لنأخذ العلاقة :

$$\bar{y} = \frac{\sum_{i=1}^{m} f_i\, y_i}{\sum_{i=1}^{m} f_i}$$

ولنكتب، للاختصار، n بدلا من $\sum_{i=1}^{m} f_i$. فيمكن إعادة كتابة هذه العلاقة كما يلي :

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{m} f_i\, y_i = \frac{f_1}{n} y_1 + \frac{f_2}{n} y_2 + \ldots + \frac{f_m}{n} y_m$$

$$= \omega_1 y_1 + \omega_2 y_2 + \ldots + \omega_m y_m$$

$$= \sum_{i=1}^{m} \omega_i\, y_i$$

حيث $\omega_i = \frac{f_i}{n}$. ويسمى ω_i الوزن الموافق للقياس الـ i، ومجموع هذه الأوزان يساوي الواحد تماما لأن :

$$\sum_{i=1}^{m} \omega_i = \sum_{i=1}^{m} \frac{f_i}{n} = \frac{1}{n} \sum_{i=1}^{m} f_i = \frac{1}{n} \times n = 1$$

ومن الواضح أن كل قياس قد أعطي وزنا يتناسب مع تكرار ظهوره في البيان الإحصائي. ويسمى مثل هذا المتوسط «المتوسط المرجح». ومنه التعريف التالي :

تعريف المتوسط المرجح

لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسات. ولتكن $\omega_1, \omega_2, \ldots, \omega_n$ أعدادا موجبة مجموعها الواحد تماما. فعندئذ يسمى المقدار

$$\bar{x} = \sum_{i=1}^{n} \omega_i\, x_i$$

المتوسط المرجح لهذه الجملة من القياسات. ويسمى ω_i الوزن الموافق للقياس x_i .

مثال (١ ـ ٨)

لنفرض أن درجات طالب في الشهادة الثانوية (الفرع العلمي) منسوبة إلى 100 كانت كما يلي : التربية الإسلامية 87، واللغة العربية 94، واللغة الإنكليزية 97،

والرياضيات 94، والفيزياء 92، والكيمياء 97، والأحياء 98. وأن لكل من التربية الإسلامية والرياضيات ثلاثة أمثال، أما اللغة العربية فلها مثلان، ولكل من المواد الباقية مثل واحد. فاحسب المعدل العام لهذا الطالب؟

الحل

$$\bar{x} = \frac{3\times 87 + 2\times 94 + 1\times 97 + 3\times 94 + 1\times 92 + 1\times 97 + 1\times 98}{3+2+1+3+1+1+1}$$

$$= \frac{3}{12}\times 87 + \frac{2}{12}\times 94 + \frac{1}{12}\times 97 + \frac{3}{12}\times 94 + \frac{1}{12}\times 92 + \frac{1}{12}\times 97 + \frac{1}{12}\times 98$$

$$= 92.92$$

والأوزان $\omega_1, \omega_2, \ldots, \omega_n$ هي، على الترتيب، $\frac{1}{12}$ ، $\frac{1}{12}$ ، $\frac{1}{12}$ ، $\frac{3}{12}$ ، $\frac{1}{12}$ ، $\frac{2}{12}$ ، $\frac{3}{12}$ ومجموعها الواحد.

وتجدر ملاحظة أن تعريف المتوسط هو حالة خاصة من تعريف المتوسط المرجح، حيث الأوزان متساوية، وكل منها يساوي $\frac{1}{n}$ ، ومن الواضح عندئذ أن:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}\frac{1}{n}x_i = \sum_{i=1}^{n}\omega_i x_i$$

حيث $\omega_i = \frac{1}{n}$. ومجموع الأوزان هو:

$$\sum_{i=1}^{n}\omega_i = \sum_{i=1}^{n}\frac{1}{n} = n\times\frac{1}{n} = 1$$

٤ ـ ليكن $\bar{x}_1$ متوسطا لمجموعة من n_1 قياسا، و $\bar{x}_2$ متوسطا لمجموعة من n_2 قياسا، . . . ، و $\bar{x}_m$ متوسطا لمجموعة من n_m قياسا. ولنحسب المتوسط العام لهذه القياسات بعد دمجها في مجموعة واحدة. ولهذه الغاية نطبق تعريف المتوسط فنقول إن المتوسط العام هو مجموع كل القياسات مقسوما على عددها. وإذا لاحظنا أن مجموع المجموعة الأولى هو $n_1\bar{x}_1$ ومجموع المجموعة الثانية هو $n_2\bar{x}_2$ ، . . . . ، ومجموع المجموعة الأخيرة هو $n_m\bar{x}_m$ ، يكون:

$$\bar{x} = \frac{n_1\ \bar{x}_1 + n_2\ \bar{x}_2 + \ldots + n_m\ \bar{x}_m}{n_1 + n_2 + \ldots + n_m} = \frac{\sum_{i=1}^{m} n_i\ \bar{x}_i}{\sum_{i=1}^{m} n_i}$$

وهو نـوع من المتـوسط المرجح حيـث $\omega_i = \dfrac{n_i}{\sum_{i=1}^{m} n_i}$. ونـلاحظ أن الـوزن المعطى لكل متوسط يتناسب مع حجم المجموعة التي يمثلها .

وفي الحالة الخاصة التي يكون فيها $n_1 = n_2 = \ldots = n_m = n$ نجد :

$$\bar{x} = \frac{n\ \bar{x}_1 + n\ \bar{x}_2 + \ldots + n\ \bar{x}_m}{mn}$$

$$= \frac{n(\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_m)}{mn}$$

$$= \frac{\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_m}{m}$$

وهو متوسط المتوسطات .

واللجوء إلى متوسط المتوسطات عند حسـاب متوسط عام ، هو خطأ شائع ، ولا يصح إلا في حـالـة واحـدة ، هي عنـدمـا يكون كـل منها متـوسطـا للعـدد نفسـه من القياسات .

مثال (١ ـ ٩)

يتألف مقرر الإحصـاء من ثلاث شعب . وقد حسبنـا متوسط عدد أيـام الغياب خلال شهر رجب فكان كما يلي :

الشعبة	الأولى	الثانية	الثالثة
المتوسط	4	5	3

إذا علمت أن أعداد الطلاب في الشعب الثلاث كان 36 في الأولى، و26 في الثانية، و34 في الثالثة، فاحسب متوسط عدد أيام الغياب في مقرر الإحصاء بشعبه الثلاث؟

الحـــــل

مجموع عدد أيام الغياب في الشعبة الأولى = $30 \times 4 = 120$ يوما،
مجموع عدد أيام الغياب في الشعبة الثانية = $26 \times 5 = 130$ يوما،
مجموع عدد أيام الغياب في الشعبة الثالثة = $34 \times 3 = 102$ من الأيام.

المتوسط العام لكافة الشعب = $\frac{120+130+102}{30+26+34} = \frac{352}{90} = 3.91$ يوما.

ونلاحظ أن متوسط المتوسطات $\frac{3+5+4}{3} = 4$ أيام وهو جواب غير صحيح.

(١ ـ ٧ ـ ٣) أثر الانسحاب وتغيير سلم القياس في المتوسط

١ ـ لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسات، متوسطها $\bar{x}$. إذا أضفنا لكل قياس فيها عددا ثابتا c فإن المتوسط يصبح $\bar{x} + c$. ولبيان ذلك نرمز لـ $x_i + c$ بـ y_i، فيكون متوسط القياسات y_i حسب التعريف:

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{\sum_1^n (x_i + c)}{n} = \frac{\sum_1^n x_i}{n} + \frac{\sum_1^n c}{n}$$
$$= \bar{x} + \frac{nc}{n} = \bar{x} + c$$

(كتبنا تسهيلا للطباعة $\sum_1^n$ بدلا من $\sum_{i=1}^n$). ومنه $\bar{x} = \bar{y} - c$. وتسمى مثل هذه العملية أي إضافة عدد ثابت c (قد يكون موجبا أو سالبا) إلى كل قياس، عملية انسحاب [انظر البند (٨) من الملحق ١].

٢ ـ لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسات متوسطها $\bar{x}$. إذا ضربنا كل قياس بعدد ثابت c فإن المتوسط يضرب بالعدد نفسه. ولبيان ذلك، نرمز للعدد ex_i y_i فيكون متوسط القياسات y_i حسب التعريف •

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{\sum_1^n c\, x_i}{n} = \frac{c \sum_1^n x_i}{n} = c\, \bar{x}$$

ومنه $\frac{\bar{y}}{c} = \bar{x}$. وتسمى عملية ضرب كل قياس بعدد ثابت، عملية تغيير في سلم القياس [انظر الفقرة (٨) من الملحق ١].

٣ ـ لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسات متوسطها $\bar{x}$. إذا خضعت هذه القياسات لتحويل وفق العلاقة الخطية:

$$y_i = a x_i + b$$

أي خضعت لعملية تغيير في سلم القياس (الضرب بعدد ثابت a)، ولعملية انسحاب (إضافة عدد ثابت b)، [انظر البند (٨) من الملحق ١]، فالعلاقة نفسها تربط بين المتوسط $\bar{x}$ والمتوسط الجديد $\bar{y}$ أي

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

ولبيان ذلك يكفي أن نكتب:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{\sum_{i=1}^n (a x_i + b)}{n}$$

$$= \frac{\sum_{i=1}^n a x_i + \sum_{i=1}^n b}{n} = \frac{a \sum_{i=1}^n x_i}{n} + \frac{n b}{n}$$

$$= a \bar{x} + b$$

وتستخدم عمليتا الإنسحاب والتغيير في سلم القياس لتسهيل الحسابات. ونوضح الفكرة بالمثال التالي:

مثال (١ ـ ١٠)

يبين الجدول التالي عدد العمال والأجر الأسبوعي الذي يتقاضاه العامل في مستشفى بالريال .

الأجر الأسبوعي	400	600	800	1000	1200	1400	1600	1800	2000
عدد العمال	6	6	4	9	5	3	3	2	1

احسب متوسط الأجر الأسبوعي للعامل في هذا المستشفى .

الحل

إرمز للأجر الذي يدفعه المستشفى بـ x_i وقم بالتحويل التالي من x_i إلى y_i :

$$y_i = \frac{x_i - 1000}{200} = \frac{1}{200}\, x_i - 5$$

تحصل على الجدول التالي :

الأجر الأسبوعي x_i	التكرار f_i	$y_i = \frac{x_i - 1000}{200}$	$f_i y_i$
400	6	-3	-18
600	6	-2	-12
800	4	-1	-4
1000	9	0	0
1200	5	1	5
1400	3	2	6
1600	3	3	9
1800	2	4	8
2000	2	5	10
المجموع	40		4

$$\bar{y} = \frac{\sum f_i\ y_i}{\sum f_i} = \frac{4}{40} = \frac{1}{10} = 0.1$$

ولحساب المتوسط المطلوب نطبق العلاقة :

$$\bar{y} = \frac{\bar{x} - 1000}{200}$$

فنجد :

$$\bar{x} = 200\ \bar{y} + 1000$$
$$= 20 + 1000 = 1020$$

تمارين (١ ـ ٢)

١) احسب المتوسط لكل مما يلي :

أ ـ 5,2,0,-3,-1

ب ـ 0.004, -0.002, 0.003, 0.001

جـ ـ 2, 2, 3, 7, 10, 100 (لاحظ أثر القيمة 100 على المتوسط) .

٢)* فيما يلي عدد المراكز الصحية والمستوصفات والمستشفيات التي أقيمت في المملكة في كل من الأعوام الثلاثة عشر بين ١٣٩١ هـ و ١٤٠٣ هـ : 4, 14, 36, 47, 26, 92, 127, 48, 31, 67, 79, 11, 22 احسب المتوسط للسنة الواحدة .

٣) متوسط 23 قياسا يساوي 14.7 فما هو مجموع هذه القياسات؟

٤) ابتعنا ستة أنواع من الحاجيات اليومية لمستشفى من كل من ثلاثة مخازن : ا ، ب ، جـ . (الحاجة نفسها من كل مخزن) وكانت الأسعار كما يلي :

* مأخوذ من كتاب منجزات خطط التنمية الصادر عن وزارة التخطيط في المملكة . ص ٢٥٤ .

السادسة	الخامسة	الرابعة	الثالثة	الثانية	الأولى	الحاجة
38.50	27.75	23.00	65.75	33.75	16.00	المخزن أ
40.25	29.50	27.50	66.75	40.50	15.00	المخزن ب
41.25	31.50	24.50	68.00	34.00	19.25	المخزن جـ

ما المخزن الذي توصي إدارة المستشفى بالتعامل معه؟

٥) ثلاث مجموعات من القياسات لها متوسطات $\bar{x}_1 = 25$ ، $\bar{x}_2 = 20$ ، $\bar{x}_3 = 22$.
وهي تتضمن 20 ، 25، و 30 قياسا ، على الترتيب .

ما هو متوسطها بعد ضمها في مجموعة واحدة؟

٦) معدل أجر الساعة وعدد المستخدمين في مستشفى عند كل من خمس مستويات للأجور كانا كما يلي :

5	4	3	2	1	مستوى الأجور
6.5	6	5.5	5	4.5	معدل أجر الساعة
25	20	15	10	5	عدد العمال

ما معدل أجر الساعة للعامل في هذا المستشفى؟

٧) في المثال (١ ـ ٦) اطرح من مركز كل فئة y_i العدد 107 ، أي اكتب عمودا جديدا $z_i = y_i - 107$. احسب المتوسط $\bar{z}$ ثم تحقق أن $\bar{y} = \bar{z} + 107$.

لاحظ أن طرح 107 من مركز كل فئة جعل العمليات الحسابية أسهل، ويسمى العدد الذي نطرحه «المتوسط الافتراضي». اتخذ العدد 97 متوسطا افتراضيا وأعد العمليات نفسها مستخدما 97 بدلا من 107. هل تجد أنه كلما كان المتوسط الافتراضي أقرب إلى المتوسط الفعلي أصبحت الحسابات أسهل؟

٨) إذا أضفنا 1.4 لكل من القياسات في التمرين١ فما أثر ذلك على المتوسط؟ أحسب المتوسط الجديد

٩) إذا ضربنا كل قياس في التمرين ١(ب) بألف فما أثر ذلك على المتوسط؟ أحسب المتوسط الجديد.

١٠) المعلومات التالية مأخوذة من سجل لغياب العاملين في عدد من المؤسسات الصحية خلال شهر شعبان:

عدد أيام الغياب y_i	0	1	2	3	4	5	6	7
التكرار f_i	5	15	23	22	17	10	6	3

احسب متوسط عدد أيام الغياب للعامل الواحد.

١١) فيما يلي السجل الدراسي لأحد الطلاب المستجدين في نهاية العام الدراسي ١٤٠٣ ـ ١٤٠٤هـ.

المقرر	١٠٢سلم	١٥١ريض	١٢١ريض	١٠١حين	١٠١ريض	١٠٢عرب	١٠١حص	١٠١سلم	١٠١فيز	١٠١كيم
عدد الساعات	2	3	3	4	3	2	3	2	4	4
التقدير	4.5	3.5	4	3	3	3	4.5	3.5	4	2.5

احسب المعدل التراكمي لهذا الطالب .

١٢) فيما يلي جدول التوزيع التكراري لأعمار خمسين عاملا في إحدى المستشفيات إلى أقرب سنة .

حدود الفئات	15-24	25-34	35-44	45-54	55-64
التكرار	8	11	25	4	2

احسب متوسط العمر للعمال الخمسين في هذا المستشفى .

١٣) في كل من التمارين ٣ إلى ١٧ من مجموعة التمارين (١ ـ ١)، احسب المتوسط .

١٤)* تناولت أنشطة فحص الدم لطفيل الملاريا لعام ١٤٠٦هـ ثماني عشرة منطقة في أنحاء المملكة وكان عدد العينات الإيجابية في كل منها كما يلي :

401, 119, 36, 779, 88, 80, 386, 180, 535,
64, 531, 565, 576, 64, 248, 246, 4331, 81

احسب متوسط عدد العينات الإيجابية للمنطقة الواحدة .

١٥)** فيما يلي عدد المراكز الصحية وعدد الأطباء في كل من المناطق الأربع عشرة في المملكة .

عدد المراكز	232	69	55	90	72	101	26	157	214	45	104	119	69	78
عدد الأطباء	613	234	156	193	145	293	44	355	317	95	180	306	81	130

* مأخوذة من التقرير الصحي السنوي الصادر عن وزارة الصحة لعام ١٤٠٦هـ، ص٢٠٣ .

** مأخوذة من التقرير الصحي السنوي الصادر عن وزارة الصحة لعام ١٤٠٦هـ، ص٤٤ .

أ - احسب لكل منطقة متوسط عدد الأطباء في المركز الواحد.
ب - احسب متوسط عدد المراكز للمنطقة الواحدة.
جـ - احسب متوسط عدد الأطباء للمنطقة الواحدة.
د - احسب متوسط عدد الأطباء للمركز الواحد على مستوى المملكة.

(١ - ٧ - ٤) الوسيط

نلاحظ من دراستنا للمتوسط أنه إذا كان أحد قياسات البيان الإحصائي كبيرا جدا، أو صغيرا جدا، بالمقارنة مع بقية القياسات، تأثر المتوسط كثيرا بهذه القيمة القاصية، ومال إليها، مما يفقد المتوسط الموقع المركزي الذي يفترض أن يشغله. وبالإضافة إلى ذلك فقد رأينا في ختام الفقرة (١ - ٢)، أن بعض البيانات يمكن أن تكون وصفية أو ترتيبية ولا يوجد أي معنى لكلمة متوسط، كما عرفناها، في مثل هذه البيانات. وسنعرّف الآن مقياسا للنزعة المركزية يمكن حسابه في كل من البيانات العددية والترتيبية، ومع وجود قيمة قاصية في بيان عددي يمكن أن لا يتأثر أبدا، وفي حال وجود أثر فإنه يكون أثرا طفيفا. ويسمى هذا المقياس الوسيط.

فوسيط n من القياسات هو القياس الواقع في الوسط عند ترتيب هذه القياسات. أي القياس الذي رتبته $\frac{n+1}{2}$ إذا كان عدد القياسات n فرديا، ومتوسط القياسين اللذين رتبتاهما $\frac{n}{2}$ و $\frac{n}{2}+1$ إذا كان عدد القياسات زوجيا.

ملاحظة

في بيان ترتيبي يكون الوسيط هو الصفة المقابلة للقياس الذي رتبته $\frac{n+1}{2}$ في حالة n فردي، أما إذا كان n زوجيا وكان للقياسين الذين رتبتاهما $\frac{n}{2}$ و $\frac{n}{2}+1$ الصفة نفسها فهذه الصفة هي الوسيط، وإذا كانا من صفتين مختلفتين، مثلا أحدهما جيد والذي يليه مقبول، قلنا اصطلاحا إن الوسيط هو بين الجيد والمقبول.

مثال (١ - ١١)

ما هو وسيط القياسات

8, 4, 10, 16, 9, 2, 7

الحل

نرتب هذه القياسات فنجد:

2, 4, 7, 8, 9, 10, 16

والوسيط هو القياس الذي رتبته $4 = \frac{7+1}{2} = \frac{n+1}{2}$. أي القياس الرابع. ولكن القياس الرابع في القياسات المرتبة أعلاه هو 8، وبالتالي تكون قيمة الوسيط المطلوبة 8.

ونلاحظ أن 8 يتوسط مجموعة القياسات، إذ يقع من القياسات على اليمين منه بقدر ما يقع منها على اليسار. كما نلاحظ أننا لم نحتج لأي عمليات حسابية، إذ قمنا بعملية ترتيب تلتها عملية اختيار.

مثال (١ ـ ١٢)

في فصل يتضمن 9 طلاب كانت التقديرات في أحد الاختبارات كما يلي:

جيد، ضعيف، مقبول، جيد، جيدجدا، ممتاز، مقبول، جيد، جيدجدا
احسب الوسيط.

الحل

نرتب التقديرات فنجد:
ضعيف، مقبول، مقبول، جيد، جيد، جيد، جيدجدا، جيدجدا، ممتاز

والتقدير المقابل للقياس الذي رتبته $5 = \frac{9+1}{2}$ ، أي للقياس الخامس هو جيد، وهكذا يكون الوسيط في هذا البيان «جيد».

مثال (١ ـ ١٣)

لدينا القياسات 25 ,22 ,25 ,26 ,37 ,16 ,32 ,26. ما الوسيط؟

الحل

نرتب هذه الأعداد فنجد:

16, 22, 25, 25, 26, 26, 32, 37

وبما أن عدد القياسات $n = 8$ زوجي، نأخذ متوسط القياسين اللذين رتبتاهما $\frac{8}{2}=4$ و $5 = 1 + 4$. أي العدد الرابع والعدد الذي يليه وهو الخامس. ولكن العدد الرابع هو 25 والعدد الخامس 26، فقيمة الوسيط تساوي:

$$\frac{25+26}{2} = 25.5$$

ولحساب الوسيط في حالة بيان مصنف، ولنرمز للوسيط بـ M، نتبع الخطوات التالية بعد كتابة جدول التكرار المتجمع الصاعد:

١ ـ نحسب رتبة الوسيط $\frac{n}{2}$ ، وذلك سواء أكان عدد القياسات n زوجيا أم فرديا.

٢ ـ تحدد رتبة الوسيط الفئة التي ينتمي إليها. ونسميها الفئة الوسيطية، كما تحدد بالطبع الفئة السابقة للفئة الوسيطية، وسنسميها اختصارا الفئة السابقة.

٣ ـ لنرمز بـ F_M للتكرار المقابل للفئة الوسيطية في جدول التكرار المتجمع الصاعد، وبـ F_p للتكرار المقابل للفئة السابقة. وبـ ω لطول الفئة، وبـ L للحد الأعلى الحقيقي المقابل للفئة السابقة. فنجد بعملية تناسب طردي بسيط أن:

زيادة القياس	زيادة التكرار
ω	$F_M - F_p$
?	$\frac{n}{2} - F_p$

زيادة القياس المطلوبة لبلوغ الوسيط $= \dfrac{\frac{n}{2} - F_p}{F_M - F_p} \times \omega$

ويكون الوسيط إذا:

$$M = L + \frac{\frac{n}{2} - F_p}{F_M - F_p} \times \omega$$

وتجدر ملاحظة أن تصنيف بيان إحصائي يتضمن نوعا من الترتيب لعناصره. ومع أن هذا الترتيب لا يتناول كل قياس بمفرده، إلا أن هناك نوعا من الترتيب الفئوي، إذا صح التعبير. فكل قياس ينتمي إلى فئة هو حتما أصغر من أي قياس ينتمي إلى فئة لاحقة، وأكبر من أي قياس ينتمي إلى فئة سابقة.

(مثال ١ ـ ١٤)

احسب وسيط البيان الإحصائي المصنف المعطى في الجدول (١ ـ ٣).

الحل

نكتب جدول التكرار المتجمع الصاعد فنجد (انظر الجدول ١ ـ ٥).

التكرار المتجمع الصاعد	أقل من	
1	86.5	
3	91.5	
7	96.5	
14	101.5	
23	106.5	الفئة السابقة
33	111.5	الفئة الوسيطية
40	116.5	
46	121.5	
50	126.5	

وتسير الخطوات الحسابية كما يلي:

١ ـ رتبة الوسيط هي $25 = \frac{50}{2} = \frac{n}{2}$ وأول فئة يزيد التكرار المتجمع المقابل لها على 25 تكون الفئة الوسيطية.

٢ ـ نطبق قاعدة التناسب الطردي فنكتب:

زيادة القياس	زيادة التكرار
5	33-23
?	25-23

$$1 = 5 \times \frac{2}{10} =$$ الزيادة المطلوبة لبلوغ الوسيط

$$M = 106.5 + 1 = 107.5$$ (الوسيط)

أو نطبق الصيغة التي استخرجناها للوسيط فنجد من الجدول أن :
$F_M = 33$ ، $F_P = 23$ ، $w = 111.5 - 106.5 = 5$ ، $L = 106.5$.

وبالتعويض نجد :

$$M = 106.5 + \frac{25 - 23}{33 - 23} \times 5 = 106.5 + 1 = 107.5$$

لاحظ أن الحساب من بيان مصنف هو دائما تقريبي ، ولذلك ترانا تجاوزنا الدقة التامة في حساب رتبة الوسيط فاتخذناها على الدوام $\frac{n}{2}$ سواء أكان n زوجيا أم فرديا . وذلك توخيا للاقتصاد في الجهود الحسابية .

وتجدر ملاحظة أننا إذا رسمنا مضلع التكرار المتجمع الصاعد ومضلع التكرار المتجمع النازل فإن الإحداثي السيني لنقطة تقاطعهما سيكون الوسيط .

تمرين

ارسم على الورقة البيانية نفسها مضلعي التكرار المتجمع الصاعد والنازل للبيان الإحصائي المصنف المعطى في الجدول (١ ـ ٣) واستنتج الوسيط بيانيا .

(١ ـ ٧ ـ ٥) المنوال

رأينا أن المتوسط لا يمكن حسابه إلا من بيانات عددية وأن الوسيط يمكن حسابه من بيانات عددية أو بيانات ترتيبية . وسنعرّف الآن مقياسا للنزعة المركزية يمكن

حسابه في جميع أنواع البيـانات سواء أكانت عددية أم ترتيبيـة أم وصفية. وهذا المقياس يعرف بالمنوال. فالمنوال هو القياس الأكثر تكرارا في جملة من القياسات.

مثال (١ـ١٥)

في تصنيف تناول 2000 من المستجدين في الجامعة حصلنـا على البيان الإحصائي التالي:

	يدخن	لا يدخن
يشرب القهوة	389	1483
لا يشرب القهوة	27	101

ما هو المنوال؟

الحل

المنوال هو «يشرب القهـوة ولا يدخن». فهي الصفة السائدة في هـذه الجملة من القياسات لأن تكرارها 1483 أعلى من تكرار كل من الصفات الثلاث الأخرى.

مثال (١ ـ ١٦)

احسب المنوال في المثال (١ ـ ١٢).

الحل

المنوال هو تقدير «جيد» باعتباره القياس الأكثر تكرارا.

ملاحظات مهمة

١ ـ المنوال هو الصفة الغالبة في بيان وصفي أو ترتيبي. والصفة الغالبة تعني أنها الصفة التي تتحقق في عـدد من العناصر التي نصنفهـا يفوق عـدد العناصر المحققـة لأية صفة أخـرى. ولا تعني بالضرورة أنها الصفة التي تتحقق في أغلبيـة العناصر أي في أكثر من 50% منها. وقد لا يكون هناك أي صفة تتحقق في أغلبية العناصر.

٢ ـ المنوال هو الصفة أو الصنف الأكثر تكرارا وليس تكرار ذلك الصنف.

٣ ـ التكرار الأعلى لا يعني التكرار الذي يقع بتواتر أكبر ولكن الصفة ذات التكرار الأعلى هي التي تقع بتواتر أكبر.

٤ ـ قد يوجد في بيان وصفي أو ترتيبي صفتان أو وصفان لهما أعلى تكرار (تكرارهما متساويان وكل منهما يمثل التكرار الأعلى بالنسبة إلى بقية الصفات أو الأصناف) فعندئذ يمثل كل منهما منوالا، ونقول إن البيان الإحصائي ثنائي المنوال. والبيان الذي يتضمن منوالا فريدا يسمى وحيد المنوال. وقد يكون هناك ثلاثة أو أربعة منوالات الخ. إلا أنه إذا كان لكل صفة أو صنف التكرار نفسه فنقول عندئذ بعدم وجود منوال ولا نقول إن كل صفة أو صنف هي في حد ذاتها منوال.

وعندما توجد في بيان إحصائي عددي مصنف فئة تتمتع بتكرار أعلى من تكرار أي فئة أخرى و يتناقص التكرار، أو يبقى ثابتا، من فئة إلى أخرى من الفئات السابقة أو اللاحقة لها، نقول إن هذه الفئة هي الفئة المنوالية، ونعتبر مركزها منوالا للبيان الإحصائي.* كما نقول عن التوزيع التكراري لهذا البيان إنه وحيد المنوال أو أحادي المنوال. والمنوال بهذا المعنى هو قمة فريدة في مدرج التكرار موافقة لفئة غير الفئة الأولى وغير الفئة الأخيرة. ومن المستحسن ألا نتحدث عن المنوال باعتباره مقياسا للنزعة المركزية إلا في هذه الحالة. وقد يتضمن المدرج التكراري عدة قمم نسبية. (كل فئة يزيد تكرارها على تكرار الفئة السابقة لها مباشرة، وعلى تكرار الفئة اللاحقة لها مباشرة، تشكل قمة نسبية) وفي حالة وجود قمتين نقول إن التوزيع التكراري ثنائي المنوال. وتكون الفئة الموافقة للقمة الأعلى الفئة المنوالية الرئيسة، ومركزها المنوال الرئيس. وتسمى الفئة الأخرى الفئة المنوالية الثانوية، ومركزها هو المنوال الثانوي. ولا يلعب المنوال، بصورة عامة، دورا كبيرا في علم الإحصاء. ويهتم بالمنوال عادة أصحاب الأعمال

---

* توجد في بعض الكتب طرق حسابية وصفية لحساب المنوال في مثل هذه الحالة. ولكن حساسيته للتغير في عدد الفئات أو حدودها لا يترك مسوغا قويا لتلك الطرق.

التجارية، والتسويق والدعاية، والقيمة الأكثر تكرارا لها مغزى خاص بالنسبة إليهم فالنوع الأكثر رواجا في صناعة معينة يجذب اهتمام أصحاب هذه الصناعة زيادة في إنتاجه ومزيدا من الدعاية له. كما يهتم به أحيانا الباحثون في العلوم السلوكية باعتباره قابلا للحساب في جميع أنواع البيانات.

مثال (١ ـ ١٧)

احسب منوال البيان الإحصائي المصنف في الجدول (١ ـ ٦).

الحل

نلاحظ أن أكبر تكرار، وهو 10، يقابل الفئة [107-111]. وأن التكرار يتناقص عندما نبتعد عن هذه الفئة في كلا الاتجاهين. فهذه الفئة هي إذا الفئة المنوالية، ومركزها 109 هو المنوال.

(١ ـ ٧ ـ ٦) مقارنة بين المتوسط والوسيط والمنوال

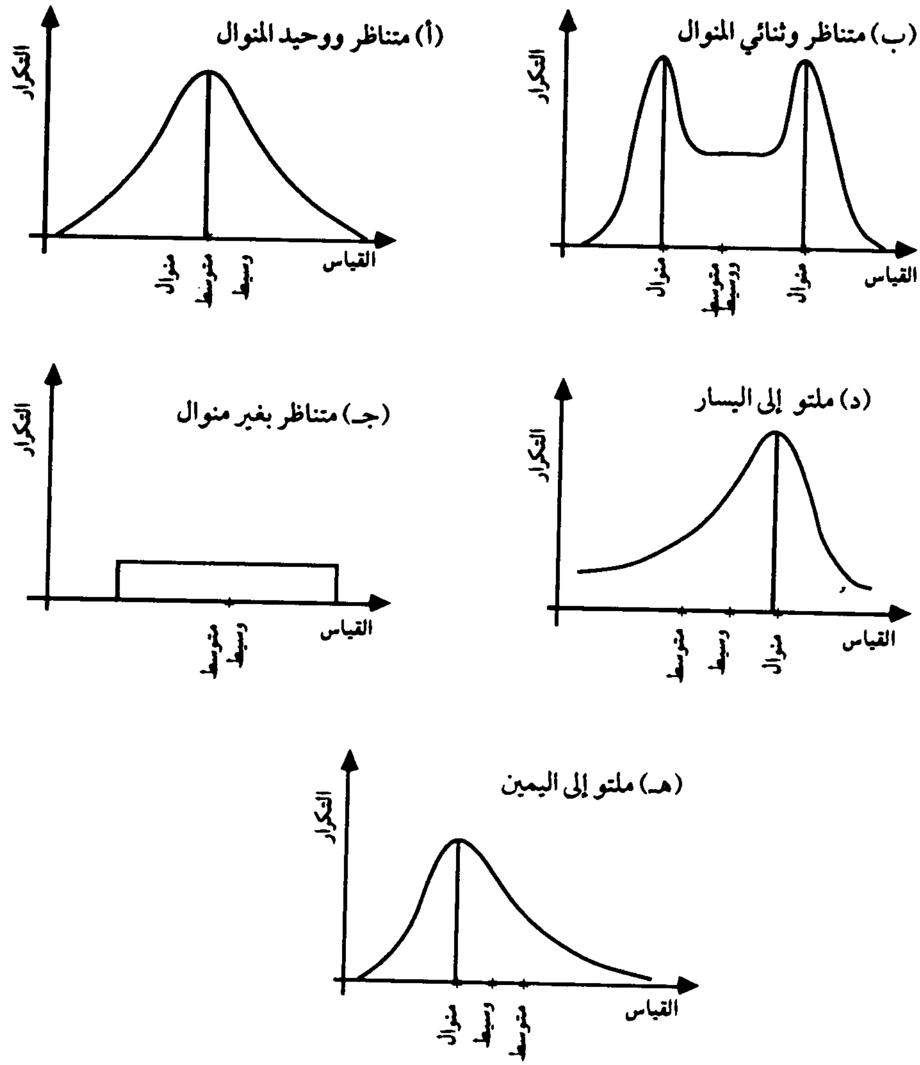
في كل من هذه المقاييس الثلاثة محاولة للتعبير عن الموضع الذي يتمركز عنده التوزيع التكراري، ولذلك سميت مقاييس النزعة المركزية. ويتضح من تعريف المتوسط أن قيمة كل قياس من قياسات بيان إحصائي تسهم في تشكيل قيمة متوسط هذا البيان. ولذلك فقد يتأثر تأثرا بالغا بالقيم المتطرفة. أما الوسيط فيتحدد من خلال المواقع النسبية للقياسات بعضها من بعض، أي أنه يتحدد من خلال رتب هذه القياسات. ولنأخذ، على سبيل المثال، القياسات 5 ,4 ,3 ,2 ,1 فمتوسطها ووسيطها 3، وإذا أضفنا إليها قياسا سادسا كبيرا جدا بالمقارنة مع بقية القياسات، وليكن، مثلا، 69، نجد أن المتوسط أصبح $14 = \frac{84}{6}$، بينما أصبح الوسيط 3.5. فالوسيط زاد بمقدار نصف في حين زاد المتوسط بمقدار 11، والجدير بالذكر أن إضافة القياس السادس لن تزيد الوسيط إلا بمقدار نصف، مهما كانت قيمته، ولكن زيادة المتوسط ستصبح أكبر فأكبر كلما زادت قيمة القياس السادس الذي أضفناه. أما المنوال فلا يتحدد من خلال قيم القياسات، ولا من خلال رتبها، ولكنه يتحدد من خلال تكرار ظهورها في البيان الإحصائي.

لنرسم مدرج التكـرار بعناية على ورق مقوى متجانس، ولنـرسم خطا رأسيا من النقطة التي تمثل المتوسط، ثم لنقص الورقة بـدقة على طول محيط المدرج التكراري. ولو أسندنا القطعة الناتجة، وعلى طول الخط الرأسي المرسوم من المتوسط، إلى حرف مستقيم وحاد كحرف سكين لتوازنت. وهذا يعني أن المتوسط هو مركز ثقل التوزيع. ولو رسمنا من القيمة المقابلـة للوسيط خطا رأسيا لقسم المساحة الـواقعة تحت المدرج التكراري إلى نصفين.

وإذا كان المدرج التكراري متناظرا، (متماثلا) تطابقت المقاييس الثلاثة، المتوسط والوسيط والمنـوال. وبهذا المعنى يكون اختلافهـا البيّن كاشفـا عن عدم تناظـر أو التواء حاد في مدرج التكرار أو في مضلع التكرار. وعلى الوجه الآخر، يشير اقترابها من بعضها إلى درجة عالية من التناظر في التوزيع التكراري.

والسؤال الوجيه هنا أي المقاييس الثـلاثة نختاره للتعبير عن الموضع الذي يتمركز عنده التوزيع التكـراري في حال اختلافها عن بعضها؟ والجواب يتـوقف على نوع البيان الإحصائي وعلى شكل التوزيع وعلى الاستخدام الذي نبغيـه للمقياس. ففي حالة بيان وصفي ليس لدينـا إلا المنوال كما ذكرنـا سابقا. وفي بيـانات ترتيبيـة يمكن أن نختار بين المنوال والوسيط أمـا في البيانات العددية فيمكن اختيار أي من المقـاييس الثلاثة. وإذا كان التـوزيع التكراري متناظـرا ووحيد المنوال [انظر الشكـل (١ ـ ١١)أ]، فلا تـوجد مشكلة لأن المقاييس الثلاثـة متطابقة. أما إذا كان التوزيع متنـاظرا وثنائي المنوال كما في الشكـل (١ ــ ١١)ب، فمن الأفضـل أن نقدّم عنـد وصف البيان الإحصـائي كلا من المنوالين، فقد يحجب تقديم القيمة المشتركة للمتـوسط والوسيط نواحٍ مهمة من البيان الإحصائي. فلنفرض مثـلا أننا سألنا 26 من ذوي الـدخل المحـدود عن الحجم الأمثل الذي يتمنـاه لأسرته (عدد الأطفـال مضافـا إليهم الوالدان)، وقـد حصلنا على الجدول التالي:

الحجم الأمثل للأسرة	10	9	8	7	6	5	4	3	2	1
التكرار	1	2	6	3	2	1	2	6	2	1



شكل (١ ـ ١١) أنواع من التوزيعات

ونجد هنا أن المتوسط $\bar{x} = 5.58$ ، وأن الوسيط $M = 6$. وهما تقريبا متساويان وإذا قلنا إن الحجم الأمثل هو في المتوسط ، فإننا نحجب بذلك وجود تيارين بارزين بين المستجيبين الستة والعشرين الذين سألناهم ، يمثلها المنوالان فقيمة أحد المنوالين 3 وقيمة المنوال الآخر 8. والتياران الرئيسان ينقسمان بين من يريد طفلا واحدا وبين من

يريد عددا من الأطفال يبلغ ستة. وهاتان الناحيتان لا تفصح عنهما القيمة 6 (أي أربعة أطفال). ولا توجد مشكلة في حالة بيان متناظر ليس له منوال كما في الشكل (١ ـ ١١)جـ، فالمنوال غير موجود والمتوسط والوسيط متطابقان.

وفي حالة توزيعات ملتوية نجد أن القياسات في البيان الإحصائي يحتشد بعضها إلى جانب بعض في جانب من المتوسط وتنتشر بعيدا على شكل ذيل في الجانب الآخر منه. ويكون اتجاه الذيل هو اتجاه الالتواء، فإذا كان الذيل على اليسار قلنا إن التوزيع ملتو إلى اليسار كما في الشكل (١ ـ ١١)د. وإذا كان الذيل على اليمين قلنا إن التوزيع ملتو إلى اليمين كما في الشكل (١ ـ ١١)هـ. وفي التوزيعات الملتوية يقع الوسيط دائما بين المنوال والمتوسط. وبما أن المنوال بالطبع عند القمة فالمتوسط يأخذ موقعه في الجانب الآخر أقرب إلى الذيل. وهذا يرشح الوسيط مقياسا أكثر استقرارا وأفضل تعبيرا عن الموقع الذي يتمركز عنده التوزيع. فالمتوسط كما رأينا شديد الحساسية للقيم المتطرفة، ولذلك نراه مائلا إلى اتجاه الذيل. أما المنوال فهو دائما في جانب القمة وشديد الحساسية، في البيانات المصنفة، للتغير في عدد الفئات أو حدودها مما يجعله أيضا خارج الإعتبار. وهكذا نفضل الوسيط في البيانات التي تتصف بالتواء واضح. ولتوضيح هذه الميزة للوسيط لنفرض أن مؤسسة تدفع رواتب سنوية لموظفيها ومستخدميها بالريال كما يلي:

180000, 72000, 30000, 18000, 3000, 3000, 3000, 3000

فنجد في هذا البيان أن المتوسط = 39000 ريال، وأن الوسيط = 10500 ريال، وأن المنوال = 3000 ريال. ومن الواضح أن الأرقام الثلاثة بتعبيرها عن متوسط الرواتب السنوية في هذه المؤسسة تقدم انطباعات مختلفة تماما. وأن كلا من المنوال والمتوسط لا يعبران بموضوعية عما يجري في المؤسسة. ولو أن مراقبا من وزارة الشؤون الاجتماعية والعمل أراد أن يظهر المؤسسة بمظهر الذي يدفع رواتب متدنية جدا في المتوسط لاختار المنوال مقياسا للنزعة المركزية. وفي المقابل فإن مدير المؤسسة سيختار المتوسط وهو 39000 ريال ليثبت أن رواتب الشركة مرتفعة. أما الباحث الاجتماعي الذي يرغب في التعبير

بموضوعية أكثر عما يجري فعلا في الشركة فسيختار الوسيط وهو 10500 ريال مقياسا للنزعة المركزية . وهذا المثال يوضح أيضا سبب قولنا إن الاختيار بين المقاييس الثلاثة يتوقف أحيانا على الغرض الذي نريده من المقياس .

وإلى جانب هذه الميزة للوسيط في التوزيعات الملتوية يمكن أن نضيف أنه بصورة عامة سهل الحساب وغير حساس للقيم المتطرفة ويبقى حسابه ممكنا في بيانات ناقصة سقطت منها قيم بعض المفردات المتطرفة التي نعرف مواقعها النسبية . وعلى سبيل المثال ، لنأخذ البيان التالي عن درجات سبعة طلاب :

71, 68, – , 75, – , 77, –

ففي هذا البيان ثلاث درجات غير معروفة. ولكن لنفرض أننا نعلم عن الطلاب الثلاثة الذين لا نعلم بالتحديد درجاتهم أن أحدهم راسب ، والآخر ناجح بتقدير مقبول ، والثالث ناجح بتقدير ممتاز. فيمكننا معرفة الوسيط ، ويمكن ترتيب معلوماتنا كما يلي :

ممتاز ,77 ,75 ,71 ,68 , مقبول , راسب

واستنتاج أن الوسيط هو 71. لا بل أكثر من ذلك لو علمنا أن ثلاثة منهم بين راسب ومقبول وثلاثة نالوا «جيد مرتفع» أو أفضل ، وأن أحدهم نال 71، لكانت هذه المعلومات كافية لاستنتاج أن الوسيط 71 .

ويبقى المتوسط مقياسا للنزعة المركزية يتمتع بخصائص مهمة تجعله مستخدما على نطاق واسع في علم الإحصاء وستتضح هذه النقطة للقارىء عبر هذا الكتاب ، ولو أخذنا عينات مختلفة بالحجم نفسه من مجتمع وحسبنا لكل عينة المتوسط والوسيط لوجدنا أن التغير من عينة إلى أخرى هو أقل في قيم المتوسط منه في قيم الوسيط ، ونعبر عن ذلك بقولنا إن المتوسط أكثر استقرارا من الوسيط عبر عينات متكررة نسحبها من مجتمع معيّن .

تمارين (١ ـ ٣)

١) أوجد الوسيط لكل من المجموعات التالية من القياسات :

أ - 6, 4, -1, 5, 1, 2

ب - 17.2, 16.9, 17.5, 16.4, 17.1

جـ - 2, 2, 1, 1, 1, 3, 1, 2, 1, 1

٢) في التمرين ٩ من مجموعة التمارين (١ ـ ٢)، احسب الوسيط والمنوال لعدد أيام الغياب.

٣) في التمرين ١١ من مجموعة التمارين (١ ـ ٢)، احسب الوسيط والمنوال لأعمار العمال الخمسين في المستشفى. أي المقاييس الثلاثة تفضل؟

٤) صنفنا عينة من محصول التفاح وفقا لوزن التفاحة مقاسا بالأونزة، فحصلنا على التوزيع التكراري التالي :

الوزن	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70
التكرار	31	45	36	23	11

احسب المتوسط والوسيط والمنوال. أيهما تفضل للتعبير عن النزعة المركزية؟

٥) احسب المتوسط والوسيط إذا علمت توزيع التكرار النسبي التالي :

الفئة	21 - 40	41 - 60	61 - 80	81 - 100
التكرار النسبي	0.24	0.36	0.28	0.12

٦) يبين الجدول التالي توزع فترة الإقامة في المستشفى لأطفال تحت سن الخامسة عشرة

من العمر ممن أجروا عمليات استئصال اللوز والزوائد الأنفية، وذلك في كل من أربع مجموعات من المستشفيات.

احسب المتوسط والوسيط والمنوال لطول فترة الإقامة في كل مجموعة من المستشفيات.

مجموعة المشافي	فترة الإقامة (بالأيام)											المجموع
	0	1	2	3	4	5	6	7	8	9	10	
A	-	-	16	113	36	5	4	2	1	-	1	178
B	-	1	1	2	2	-	27	-	-	-	-	33
C	-	-	12	33	20	28	35	7	1	4	6	146
D	-	97	6	2	6	28	11	27	2	1	4	184

٧) يبين الجدول على الصفحة التالية جزءا من دراسة لحجم رد الفعل لاختبار الليبرومين في ثلاث جماعات من الأطفال، وقد أعطي الأطفال في الجماعة الأولى لقاح الـ B.C.G. عند ولادتهم عن طريق الفم، وفي الجماعة الثانية أعطي الأطفال اللقاح عن طريق أدمة الجلد، ولم يعط أطفال الجماعة الثالثة لقاح الـ B.C.G..

احسب لكل جماعة المتوسط، الوسيط والمنوال لقياسات نصف قطر رد الفعل.

٨) حشرات من نوع العت تتغذى على فئران مصابة بدودة الفيلاريا. وقد أخذت هذه الحشرات بعد فترة وأحصي عدد الميكروفيلاريا في كل عت. ويمثل البيان التالي نتائج التعداد لخمسين عتا. احسب المتوسط، الوسيط، والمنوال لهذه القياسات وعلق على الفروق بين هذه المقاييس الموضعية الثلاثة.

7	12	3	3	1	8	0	7	2	0
10	15	3	19	1	2	2	15	3	4
7	0	9	0	18	4	6	6	10	1
1	9	14	3	7	5	7	5	14	20
6	1	2	14	3	3	5	1	4	3

٩) فيما يلي مستـوى السكـر في الـدم مأخـوذا في الصبـاح قبل تنـاول الفطـور لعشرة أطفال :

56, 62, 63, 65, 65, 65, 68, 70, 72

احسب الوسيط والمنوال .

١٠) يبين الجدول التالي التكرار النسبي المتجمع لعمر العروس وفي أربع عينات من النساء مأخوذة من أربع جماعـات ، تاريخ الميلاد في الجماعة الأولى يعـود إلى ما قبل 1925 ، وفي الثانية بين 1925 إلى 1934 ، وفي الثالثة من 1935 إلى 1944 ، وفي الرابعة من 1945 إلى 1954 .

ارسم مضلع التكرار النسبي المتجمع الصـاعد لكل من العينـات الأربع ، وقدر العمر الوسيط للعروس في كل جماعة .

عدد الأطفال			نصف قطر رد الفعل (بالميلليمتر)
لم يعط .B.C.G	B.C.G. تحت الجلد	B.C.G. عن طريق الفم	
7	2	-	1
2	3	-	2
39	53	36	3
-	22	22	4
9	42	29	5
2	15	18	6
-	4	10	7
2	4	8	8
-	-	3	9
-	2	3	10
-	-	-	11
-	-	3	12
-	-	-	13
-	1	2	14
-	-	1	15
-	-	1	16
61	150	136	المجموع

العمر عند الزواج	قبل 1925 N = 61	1925-1934 N = 83	1935-1944 N = 90	1945-1954 N = 106
	%	%	%	%
9-10	3.4	6.0	5.6	9.6
11-12	6.9	13.3	18.0	21.1
13-14	39.7	27.7	38.2	39.4
15-16	58.6	56.2	68.5	73.1
17-18	63.8	74.7	77.5	90.3
19-20	74.1	80.7	85.4	98.8
21-22	79.3	86.7	89.9	99.0
23-24	82.8	88.0	95.5	100
25-26	87.9	90.4	97.7	
27-28	89.7	92.8	97.7	
20-30	93.1	96.4	98.8	
> 30	100	100	100	

١١) فيما يلي أوزان عشرة حيوانات تجربة وذلك بعد مداخلة جراحية (مقاسة بالكغ) :

13.2, 15.4, 13.0, 16.6, 16.9, 14.4, 13.6, 15.0, 14.6, 13.1

احسب الوسيط .

١٢) فيما يلي المسافة (إلى أقرب ميل) التي قطعها كل من خمسة عشر مريضا حتى وصلوا إلى أقرب مستوصف :

5, 9, 11, 3, 12, 13, 12, 6, 13, 7, 3, 15, 12, 15, 5

ما وسيط المسافة التي يقطعها المريض حتى يصل إلى أقرب مستوصف؟

١٣) كـانت فترة الإقامـة بالأيـام لأول أحـد عشر مريضـا أدخلوا إلى جنـاح للأمـراض النفسية افتتح حديثا في أحد المستشفيات كما يلي :

29, 14, 11, 24, 14, 14, 28, 14, 18, 22, 14

احسب الوسيط والمنوال لعدد أيام الإقامة في المستشفى .

١٤) فيما يلي جدول توزيع تكراري يلخص بيانا إحصائيا عن درجة تلوث الهواء (مقاسة بالميكروجرام في المتر المكعب) في 57 مدينة كبيرة .

الفئات	10-19	20-29	30-39	40-49	50-59	60-69	70-79
التكرار	5	19	10	13	4	4	2

احسب المتوسط والوسيط والمنوال .

١٥) في كل من التمارين ٣ إلى ١٧ من مجموعة التمارين (١ ـ ١)، احسب الوسيط .

١٦) بعد افتتاح مركـز حديث للتسوق قرب ضاحية معينـة ، ازدادت حركة المرور فيها ، وقد حددت إدارة المرور السرعة القصوى في شـارع الضاحية بـ 35 كم/ سا. وبعد شكاوي عن عدم التزام السيارات بهذا الحد قامت دورية مرور خلال 15 دقيقة من المراقبة برصـد سرعات 25 سيـارة مـرت من ذلك الشـارع. وحصلت على البيـان التالي :

15, 40, 47, 25, 37, 23, 20, 38, 29, 40,
35, 28, 37, 38, 35, 37, 27, 36, 30, 38,
40, 43, 25, 20, 42

أ ـ إذا كنت من سكان الضاحية الذين يرغبون في استخدام هذا البيان لإثبات أن السيارات بصورة عامة لا تتقيد بحد السرعة المفروض ، فهل تستخدم المنوال ، أم الوسيط أم المتوسط؟

ب ـ إذا كنت ممن يعارضون فرض حد للسرعة وتريد استخدام هذا البيان لدعم وجهة نظرك بأن السيارات ملتزمة بصورة عامة بلوحة المرور، أي المقاييس تختار؟

١٧) تهتم شركة بمعرفة مدى استخدام موظفيها للهواتف في مكالمات شخصية. وفي أحد الأيام راقبت عدد المكالمات الشخصية التي قام بها كل موظف فحصلت على البيان التالي:

عدد المكالمات الشخصية	10	9	8	7	6	5	4	3	2	1	0
عدد الموظفين	1	1	0	0	11	8	6	23	112	65	273

أ ـ أوجد متوسط المكالمات الشخصية للموظف الواحد في ذلك اليوم.

ب ـ آخذا في اعتبارك أولئك الذين استخدموا الهاتف لأغراض شخصية فقط، ما المقياس الذي تجده أفضل تعبيرا عن النزعة المركزية؟ احسب هذا المقياس.

١٨) في التمرين ٥ من مجموعة التمارين (١ ـ ١)، احسب المتوسط والوسيط والمنوال. أي المقاييس الثلاثة تفضل ولماذا؟

١٩) في التمرين ١٨ من مجموعة التمارين (١ ـ ١)، احسب مقياس النزعة المركزية الذي تعتقد أنه مناسب في كل من بيان الأطباء وبيان الأسرّة. وأوضح أسباب تفضيلك.

٢٠) في التمرين ١٣ من مجموعة التمارين (١ ـ ٢) احسب الوسيط. أيهما تفضل المتوسط أم الوسيط ولماذا؟

٢١) فيما يلي بيان بعدد الزيارات التي قام بها المراجعون للعيادات الخارجية بالمستشفيات ومراكز الرعاية الصحية الأولية في المناطق الأربع عشرة في المملكة، وذلك خلال عام ١٤٠٦هـ، بآلاف المراجعين:

11169, 4330, 4870, 3029, 2050, 4802, 1577,
6375, 6034, 1480, 3876, 3465, 2826, 1794.

احسب المتوسط والوسيط أيهما تعتقد أنه الأفضل لقياس النزعة المركزية ولماذا؟

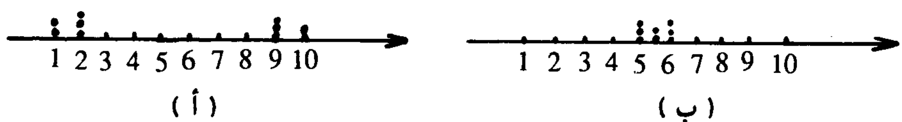
(١ ـ ٨) مقاييس التشتت

ناقشنا في الفقرات السابقة معايير موضعية تهدف إلى تحديد الموضع الذي تتمركز عنده جملة من القياسات الإحصائية. ولكنها لا تكفي وحدها لتشكيل صورة ذهنية متكاملة عن التوزيع التكراري للبيان الإحصائي. وإلى جانب المكان الذي يشكل مركز التوزيع نحتاج إلى معرفة كل ما يمكن معرفته عن خاصية التغير من قياس إلى آخر ضمن البيان الإحصائي، وعن مواقع القياسات بالنسبة إلى مركزها. فالقياسات 1, 2, 3, 4, 5 لها متوسط يساوي 3 وهو بالذات متوسط للقياسات 602 ,209 ,4 ,200 - ,600 - . ولكن شتان ما بين المجموعتين من القياسات من حيث درجة تجمعها حول المركز المشترك لكل منها وهو 3. ومن المعروف أنه لا يمكن لقياسات بيان إحصائي أن تكون متساوية. ولو قسنا، مثلا، أطوال مجموعة من أوراق نبات معين، لاختلف القياس من ورقة إلى أخرى، ولو كان الشخص نفسه هو الذي يكرر قياس ظاهرة معينة مستخدما الجهاز نفسه في كل مرة، فسيختلف القياس الذي يحصل عليه من محاولة إلى أخرى. والتغير ظاهرة ملازمة لكل بيان إحصائي، وإذا كان التوزيع التكراري للبيان الإحصائي يتمركز عند المتوسط الحسابي، فهو ينتشر على جانبي هذا المتوسط، وكلما كان التغير كبيرا من قياس إلى آخر، اتسع انتشار القياسات حول متوسطها. وبصورة عامة. فإن القياسات التي تحتشد وتتجمع حول متوسطها، وقريبا منه، يكون تشتتها صغيرا. بينما يكون تشتت القياسات المبعثرة التي تنتشر بعيدا على جانبي المتوسط، تشتتا كبيرا. وسنحاول فيما يلي تقديم معايير كمية لقياس شدة تبعثر القياسات في بيان إحصائي، أو لقياس درجة انتشار وتشتت القياسات حول متوسطها. وسنبدأ بتعريف المدى.

(١ ـ ٨ ـ ١) تعريف المدى

مـدى بيـان إحصـائي هـو الفـرق بين أكبر قيـاس وأصغـر قيـاس في البيـان الإحصائي.

ومن الـواضح أن المدى يعطي فكرة واضحـة عن المسافة على محور الأعـداد التي يتوضّع فيها البيان الإحصائي. وإذا استثنينا القيمتين المتطرفتين في البيان الإحصائي فإن المدى بمفرده عاجـز عن تقديم أية معلومـات عن أسلوب انتشار بقية القيـاسات حول المتوسط. وعلى سبيـل المثال، إذا كان لـدينا عشرة قيـاسات متـوسطها 5.5 وأحدها 1 وأكبرها 10. فيمكن تصـور هذه القيـاسـات العشرة بأشكال عـديـدة تختلف اختلافـا شديدا في درجـة تبعثرها وتشتتها حول المتوسط، ونجـد في الشكل (١ ـ ١٢) تصورين مختلفين. ففي الشكل (١ ـ ١٢)أ، نجد القياسات 10 ,10 ,9 ,9 ,9 ,2 ,2 ,2 ,1 ,1 وفي الشكـل (١ ـ ١٢)ب، نجـد القياسـات 10 ,6 ,6 ,6 ,5.5 ,5.5 ,5 ,5 ,5 ,1 ومـع أن للمجموعتين المدى نفسه وهو 9 = 1 - 10، إلا أن الفارق كبير بين درجـة تمركز كل منهما حول المتوسط المشترك 5.5 :



شكل (١ ـ ١٢)

وإذا كان المدى يضم بين طـرفيه جميع قياسات البيـان الإحصائي فلماذا لا نفكر بمدى أكثر تواضعـا يضم بين طرفيه نسبة عالية من القياسـات (ثمانين بالمائة منها مثلا) بـدلا من أن يضمهـا جميعهـا. ولـو عـرفنـا مثـلا، القيـاس الـذي يقل عنـه 10% من القيـاسـات، وسنسميه المئين عشرة، والقيـاس الـذي يقل عنـه 90% من القيـاسات، وسنسميـــه المئين تسعين، فبين المئين عشرة والمئين تسعين يقـع ثمانـــون بـــالمائة من القياسات. ولو حسبنا هذين القياسين ووجدناهما قريبين من بعضهما، فسيعطينا ذلك تصـورا مفيـدا تمامـا عن واقع انتشـار أو تشتت البيـان الإحصـائي. وسنعـرف فيما يلي المئينات باعتبارها وسيلة من وسائل التعبير عن تشتت بيان إحصائي.

(١ ـ ٨ ـ ٢) تعريف المئينات

ليكن r أي عـدد صحيح بين الصفر والمائة، نعـرف المئين r لبيان إحصائي بأنه العدد الذي يقل عنه r بالمائة من قياسات البيان الإحصائي .

ونلاحظ من هـذا التعريف أن المئين r ، وسنرمز لـه بـ P_r ، هو القياس المقابل لتكرار متجمع يساوي $n \times \frac{r}{100}$ ، حيث n عدد القياسات . ومن الواضح أن P_{50} هو الوسيط باعتباره القياس المقابل لتكرار متجمع يساوي $n \times \frac{50}{100}$ أو $\frac{n}{2}$. ويسمى P_{25} (المئين 25) الربيع الأدنى ، باعتباره يحصر إلى اليسار منه ثلاثة أرباع القياسات وسنرمز له بـ Q_1 . كما يسمى P_{75} (المئين 75) الربيع الأعلى باعتباره يحصر إلى اليسار منه ثلاثة أرباع القياسات وسنرمز له بـ Q_3 . والمسافة بين هذين القياسين، أي الفرق بين الربيع الأعلى والربيع الأدنى، تسمى المدى الربيعي .

$$\text{المدى الربيعي} = Q_3 - Q_1$$

ويضم المدى الـربيعي بين طرفيـه 50% من القيـاسـات . ويعتبر نصف المدى الربيعي $\frac{Q_3 - Q_1}{2}$ أحد معايير التشتت .

وتجدر ملاحظـة أن أكبر قياس في البيـان الإحصائي هـو المئون 100 (P_{100}) وأن أصغر قياس هو المئون صفر (P_0) . وأن المدى هو $P_{100} - P_0$.

ولقد أوضحنـا عمليـا طريقـة حساب أي مئين في الفقـرة (١ ـ ٤)، ولحسـاب (P_r) ، بصورة عامة ، نكتب أولا جدول التكرار المتجمع الصاعد ، ثم نحسب رتبة المئين r وهـي $n \times \frac{r}{100}$ ، حيـث n عدد القياسات في البيان الإحصائي . وتحدد رتبة المئين الفئة التي ينتمي إليهـا المئون وسنسميها فئـة المئين ، كما تحدد بالطبع الفئة السـابقة لها . لنرمز الآن بـ F_p للتكرار المقابل لفئة المئين في جدول التكـرار المتجمع الصاعد ، وبـ F_b

للتكرار المقابل للفئـة السابقة، وبـ w لطول الفئة، وبـ L للحد الأعلى الحقيقـي المقابل للفئة السابقة. فنجد بعملية تناسب طردي بسيط أن:

زيادة القياس	زيادة التكرار
w	$F_P - F_b$
?	$\frac{nr}{100} - F_b$

ومنه:

$$\text{زيادة القياس المطلوبة لبلوغ المئين } r = \frac{\frac{nr}{100} - F_b}{F_P - F_b} \times w$$

ويكون المئون r المطلوب:

$$P_r = L + \frac{\frac{nr}{100} - F_b}{F_P - F_b} \times w$$

مثال (١ ـ ١٨)

احسب Q_1 (الربيع الأدنى)، و Q_3 (الربيع الأعلى) للتوزيع التكراري في الجدول (١ ـ ٣) واحسب نصف المدى الربيعي.

الحل

١ ـ رتبة الربيع الأدنى هي $n \times \frac{25}{100} = 50 \times \frac{25}{100} = 12.5$

وأول فئة يزيد التكرار المتجمع المقابل لها على 12.5 تكون فئة الربيع الأدنى.

٢ ـ نطبق قاعدة التناسب الطردي فنكتب:

نكتب جدول التكرار المتجمع الصاعد وتسير الخطوات الحسابية كما يلي :

التكرار المتجمع الصاعد	أقل من	
1	86.5	
3	91.5	
7	96.5	— الفئة السابقة
14	101.5	— فئة الربيع الأدنى
23	106.5	
33	111.5	— الفئة السابقة
40	116.5	— فئة الربيع الأعلى
46	121.5	
50	126.5	

زيادة القياس	زيادة التكرار
5	14 - 7
?	12.5 - 7

$\frac{12.5-7}{14-7} \times 5 = 3.93$ = الزيادة المطلوبة لبلوغ الربيع الأدنى

$$Q_1 = 96.5 + 3.93 = 100.43$$

ولحساب الربيع الأعلى (Q_3) نجد بصورة مماثلة أن رتبة الربيع الأعلى هي $50 \times \frac{75}{100} = 37.5$

زيادة القياس	زيادة التكرار
5	40 - 33
?	37.5 - 33

$\frac{37.5-33}{40-33} \times 5 = 3.21$ = الزيادة المطلوبة لبلوغ الربيع الأعلى

$$Q_3 = 111.5 + 3.21 = 114.71$$

أو نطبق الصيغة العامة التي استخرجناها من أجل المئينات فنجد من الجدول ، في حالة الربيع الأدنى أن $r = 25$, $F_b = 7$, $F_p = 14$, $L = 96.5$, $\omega = 5$ ثم نعوض في العلاقة :

$$P_r = L + \frac{\frac{nr}{100} - F_b}{F_Q - F_b} \times \omega$$

وفي حالة الربيع الأعلى يكون $[\ \omega = 5 , L = 96.5 , F_P = 14 , F_b = 7 , r = 25\ ]$.

وأخيرا:

$$\text{نصف المدى الربيعي} = \frac{114.71 - 100.43}{2} = \frac{14.28}{2} = 7.14$$

ولكن لماذا لا نبحث عن مقيـاس للتشتت يسهم في تشكيلـه كل قيـاس من قياسـات البيان الإحصـائي بدلا مـن أن يقتصر على مثينين أو على أكبر قيـاس وأصغر قيـاس؟ ومن الـواضح أن التشتت يعـود في الأسـاس إلى قرب أو بعـد القيـاسـات عن متوسطها. فلنحاول إذاً التعبير عن التشتت بدلالة انحراف كل قياس عن المتوسط، أي بدلالة $d_i = x_i - \bar{x}$. ونعلم من خـواص المتوسط أن مجموع هذه الانحـرافات يساوي الصفر مما لا يترك مجالا للتفكير في متوسط هـذه الانحرافات كمقياس للتشتت. ولكن حل هذه المشكلة سهل طـالما أنه يعود إلى وجود انحـرافات موجبة وانحرافـات سالبة، فلماذا لا نحسب متوسط القيم المطلقة للانحرافات؟

(١ ـ ٨ ـ ٣) تعريف متوسط الانحراف

لتكن $x_1 , x_2 , \ldots, x_n$ جملة من القيـاسات متـوسطها $\bar{x}$. نعـرف متوسط الانحراف لهذه القيـاسات، ونـرمز لـه بـ D ، بأنـه متوسط القيم المطلقـة لانحرافـات القياسات عن متوسطها.

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

وتلافيا للتعقيدات التي يسببهـا وجود القيمة المطلقة عنـد استخدام المعيار D في التحليل الإحصـائي، يمكن اللجـوء إلى حل آخـر لمشكلـة الانحـرافـات الموجبـة والانحرافات السـالبة، وذلك بأخذ مربعـات الانحرافات بدلا من قيمهـا المطلقة، مما يؤدي إلى تعريف مقياس للتشتت يسمى التباين.

(١ ـ ٨ ـ ٤) تعريف التباين

تباين مجتمع من القياسات يتضمن N قياسا. $x_1 , x_2 , ... , x_N$ هو متوسط مربعات انحرافات القياسات عن متوسطها. وسنرمز له بـσ^2 ، وبصورة رمزية نكتب:

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

وتجدر ملاحظة أن التباين موجب دوما لأنه ناشيء عن مجموع مربعات، أي مجموع كميات موجبة. ويكون التباين صفرا إذا وفقط إذا كانت القياسات جميعها متساوية.

(١ ـ ٨ ـ ٥) الانحراف المعياري لمجتمع

الانحراف المعياري σ هو الجذر التربيعي الموجب للتباين

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

ويقاس الانحراف المعياري بوحدة القياس نفسها المستخدمة في البيان الإحصائي.

ملاحظة

الرمز المستخدم σ هو الحرف الأبجدي سيجما في الأبجدية اليونانية بالخط الصغير ويكتب بالخط الكبير على الشكل Σ.

وكما ذكرنا سابقا إذا كان لدينا مجتمع من القياسات وتباينه σ^2 غير معروف أو غير متوفر فيمكن أخذ عينة من هذا المجتمع حجمها n ، مثلا، وحساب تباينها ثم اعتبار هذا التباين تقديرا أو تخمينا لتباين المجتمع الذي نجهله. ومن الطبيعي أن يكون تباين العينة، وفقا لتعريف التباين، مساويا لمتوسط مربعات انحرافات القياسات الـn في العينة عن متوسطها. ولكن يبرهن في الإحصاء الرياضي أن تباين العينة سيكون تقديرا أفضل لتباين المجتمع إذا قمنا بتعديل طفيف جدًا في صيغة التعريف. وهذا التعديل هو أن نقسم مجموع مربعات الانحرافات عن المتوسط $\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]$ على $(n-1)$ بدلا

من قسمتها على (n) . وهكذا سنرمز لتباين عينة بـS^2 ، تمييزا له عن σ^2 تباين المجتمع ، ونعرفه كما يلي :

(١ ـ ٨ ـ ٦) تعريف تباين عيّنة

تباين عينة من القياسات $x_1 , x_2 , \ldots , x_n$ هو:

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

حيث $\bar{x}$ متوسط العيّنة .

(١ ـ ٨ ـ ٧) تعريف الانحراف المعياري لعيّنة

الانحراف المعياري لعيّنة من القياسات $x_1 , x_2 , \ldots , x_n$ هو الجذر التربيعي الموجب لتباين العينة .

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

وفيما تبقى من هذا الفصل سنعتبر تباين أي جملة من القياسات تباين عينة ، ونطبق لحساب التباين التعريف (١ ـ ٨ ـ ٦) وحيثما وردت كلمة التباين أو الانحراف المعياري ، فيما بقي من هذا الفصل ،فسنعني بها تباين العينة (S^2) كما عرفناه في (١ ـ ٨ ـ ٦) ، والانحراف المعياري لعينة (S) كما عرفناه في (١ ـ ٨ ـ ٧) ، إلا إذا ذكرنا ما يخالف ذلك .

مثال (١ ـ ١٩)

لتكن جملة القياسات $4 , 2 , 1 , 7 , 5$ ، احسب متوسط الانحراف ، والتباين ، والانحراف المعياري .

$$\bar{x} = \frac{19}{5} = 3.8$$

الحل

ننظم الجدول التالي بعد حساب المتوسط $\bar{x}$. ثم نطبق التعريف مباشرة لنجد:

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	5	1.2	1.44
	7	3.2	10.24
	1	-2.8	7.84
	2	-1.8	3.24
	4	0.2	0.04
المجموع	19	0	22.8

(متوسط الانحراف) $D = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$

$$= \frac{1}{5}(1.2 + 3.2 + 2.8 + 1.8 + 0.2) = \frac{9.2}{5} = 1.84$$

(التباين) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{22.8}{4} = 4.56$

(الانحراف المعياري) $S = \sqrt{4.56} = 2.14$

والتطبيق المباشر للتعريف يتطلب جهدا حسابيا لا مسوّغ له. (يتضمن $2n + 4$ عملية حسابية) ويعاني، في الغالب، من نقص في الدقة. وإذ نحسب قبل كل شيء المتوسط $\bar{x}$ ، نبدأ بعملية تقسيم، وإذا كانت عملية التقسيم غير منتهية فسيؤثر ذلك على دقة النتائج اللاحقة. وسنقدم الآن صيغة مختزلة لحساب التباين تختصر الجهود الحسابية وتعطي التباين بدقة أكبر.

(١ ـ ٨ ـ ٨) صيغة مختزلة لحساب التباين

باستخدام خواص الرمز Σ المذكورة في البند (٥) من الملحق (١)، ومن تعريف المتوسط يمكن أن نكتب ما يلي:

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

$$=\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}2\bar{x}x_i+\sum_{i=1}^{n}\bar{x}^2$$

$$=\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2$$

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

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

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

ومنه :

$$S^2=\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-\frac{\left(\sum x_i\right)^2}{n}\right]$$

ولم نعد نستهل العمل الحسابي بعملية تقسيم ، بالإضافة إلى أن هذه العبارة تتضمن $(n+6)$ عملية حسابية مما يوفر $(n-2)$ عملية حسابية . وهي بذلك أسرع وأدق من التطبيق المباشر للتعريف .

ونكتب العبارة المختزلة السابقة ، أحيانا ، على الشكل :

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

وهي تتضمن عملية تقسيم واحـدة تشكل خاتمة العمل الحسابي. ومع مـا يبدو للوهلة الأولى من تعقيـد في كتابة الصيغة المختـزلة، إلا أن كل ما نحتاجـه لتطبيقها هو مجموع القياسات ومجموع مربعاتها وعددها.

مثال (١ ـ ٢٠)

احسب تباين القياسات في المثال (١ ـ ٦) بتطبيق الصيغة المختزلة.

الحل

ننظم الجدول المبين جانبا ثم نطبق الصيغة المختزلة فنجد:

	x_i	x_i^2
	5	25
	7	49
	1	1
	2	4
	4	16
المجموع	19	95

$$S^2 = \frac{1}{4}\left[95 - \frac{(19)^2}{5}\right]$$

$$= \frac{1}{4}\left(95 - \frac{361}{5}\right) = 4\ .56$$

(١ ـ ٨ ـ ٩) حساب التباين في بيانات مصنفة

لنعد إلى الفقرة (١ ــ ٥)، وبخاصة إلى الجدولين (١ ـ ٦) و (١ ــ ٧)، ولنحاول تطبيق التعريف (١ ـ ٨ ـ ٦) فالمطلوب إذا هو حساب انحراف كل قياس y_i عن المتوسط $\bar{y}$ ، ثم أخذ مجموع مربعات هذه الانحرافات. وإذا كان القياس y_1 ، مثلا، مكررا f_1 مرة، فسيتضمن مجموع مربعات الانحرافات حدودا متطابقة ومكررة مثل:

$$\underbrace{(.y_1-\bar{y})^2+(y_1-\bar{y})^2+\cdots+(y_1-\bar{y})^2}_{\text{مكرر } f_1 \text{ مرة}}$$

ومن الأفضل بالطبع، كتابة مجموع حدود مطابقة لبعضها مثل هذه الحدود، على الشكل

$$f_1(y_1-\bar{y})^2$$

والأمر نفسه في بقية الحدود، وهكذا تتخذ العلاقة الواردة في تعريف تباين العينة، الصيغة التالية من أجل بيان مرتب:

$$S^2=\frac{1}{\sum_1^m f_i-1}\left[\sum_1^m f_i(y_i-\bar{y})^2\right]$$

حيث:

$$\bar{y}=\frac{\sum_1^m f_i\,y_i}{\sum_1^m f_i}$$

وتصبح الصيغة المختزلة لحساب التباين كما يلي:

$$S^2=\frac{1}{n-1}\left[\sum_1^m f_i\,y_i^2-\frac{\left(\sum_1^m f_i\,y_i\right)^2}{n}\right],$$

حيث $n=\sum_1^m f_i$.

مثال (١ ـ ٢١)

قذفنا حجر نرد مائة مرة فكانت تكرارات النتائج الست الممكنة كما يلي:

جدول (١ ـ ١٦)

y_i	1	2	3	4	5	6
f_i	19	15	15	20	14	17

والمطلوب حساب تباين هذا التوزيع التكراري وانحرافه المعياري .

الحل

لحساب التباين ننظم الجدول التالي

جدول (١ ـ ١٧)

	y_i	f_i	$f_i y_i$	$f_i y_i^2$
	1	19	19	19
	2	15	30	60
	3	15	45	135
	4	20	80	320
	5	14	70	350
	6	17	102	612
المجموع		$100 = \Sigma f_i$	$346 = \Sigma f_i y_i$	$1496 = \Sigma f_i y_i^2$

والتباين المطلوب(S^2) هو:

$$S^2 = \frac{1}{99}\left[1496 - \frac{(346)^2}{100}\right] = 3.02$$

والانحراف المعياري(S) هو

$$S = \sqrt{3.02} = 1.74$$

ولحساب تباين بيان مصنف (أو مبوب) نعتبر أن جميع القياسات التي تنتمي إلى فئة مساوياً لمركز هذه الفئة والخطوات الحسابية هي بالضبط كما في حالة بيان الرتب، حيثy_i هي الآن مركز الفئة، و f_i التكرار الموافق لهذه الفئة. وللتوضيح نأخذ المثال التالي.

مثال (١ ـ ٢٢)

احسب التباين والإنحراف المعياري للتوزيع التكراري في الجدول (١ ـ ٦).

الحل

ننظم الجدول التالي :

جدول (١ ـ ١٨)

	y_i مركز الفئة	f_i التكرار	$f_i y_i$	$f_i y_i^2$
	84	1	84	7056
	89	2	178	15842
	94	4	376	35344
	99	7	693	68607
	104	9	936	97344
	109	10	1090	118810
	114	7	798	90972
	119	6	714	84966
	124	4	496	61504
المجموع		$50 = \Sigma f_i$	$5365 = \Sigma f_i y_i$	$580445 = \Sigma f_i y_i^2$

ويكون التباين المطلوب (S^2) هو

$$S^2 = \frac{1}{49}\left[580445 - \frac{(5365)^2}{50}\right]$$
$$= \frac{1}{49}[580445 - 575664.5]$$
$$= \frac{4780.5}{49} = 97.56$$

والانحراف المعياري (S) هو

$$S = \sqrt{97.56} = 9.88$$

والجدير بالذكر أننا لو حسبنا الإنحراف المعياري من البيان الأصلي المعطى في الجدول (١ ـ ٤) مباشرة لحصلنا على $S = 10.05$.

(١ ـ ٨ ـ ١٠) أثر الانسحاب وتغيير سلم القياس في التباين

لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسات، متوسطها $\bar{x}$ ، وتباينها S_x^2 . إذا أضفنا العدد نفسه، b مثلا، إلى كل قياس، فينبغي ألا يؤثر ذلك على التباين. وإذ تعتمد قيمة التباين على الفروق بين القياسات، فإن الفرق بين أي قياسين لن يتغير

عندما نضيف إلى كل منهما العدد نفسه [انظر البند (٨) من الملحق ١]. أما إذا ضربنا كل قياس بعدد، a مثلا، فسيُضرب التباين بمربع هذا العدد، a^2، ويضرب الانحراف المعياري بالقيمة المطلقة للعدد a. ويمكن بيان ذلك في المحاكمة البسيطة التالية:

لنرمز بـ y_i للقياس الناتج عن ضرب x_i بـ a ثم إضافة b إلى الناتج، أي لنفرض أن:

$$y_i = ax_i + b \ ; \ i = 1, 2, \ldots, n$$

فنعلم من خواص المتوسط (الفقرة ١ ـ ٧ ـ ٢) أن:

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

حيث يرمز $\bar{y}$ للمتوسط الجديد. ومن تعريف التباين وخواص المجموع Σ [انظر البند (٥) من الملحق (١)] يمكن أن نكتب:

$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\left[(ax_i + b) - (a\bar{x} + b)\right]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[a(x_i - \bar{x})\right]^2 = a^2 \times \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = a^2 S_x^2$$

وبأخذ الجذر التربيعي الموجب للطرفين نجد:

$$S_y = |a|\, S_x$$

وهذا يعني أن عملية الانسحاب لا تؤثر في التباين كما توقعنا، ولكن عملية تغيير سلم القياس لها أثر كبير في التباين. وعلى سبيل المثال، إذا كانت القياسات x_i مقاسة بالسنتيمتر، وغيرنا وحدة القياس إلى الميلليمتر، أي ضربنا كل قياس بـ 10، فإن التباين سيضرب بمائة، وسيضرب الانحراف المعياري بعشرة.

(١ ـ ٩) حساب المتوسط والانحراف المعياري من خلال تحويل البيان الإحصائي

سنقدم فيما يلي طريقة لحساب المتوسط والانحراف المعياري توفر الكثير من الجهود الحسابية، وذلك في حالة بيان مصنف أطوال الفئات فيه متساوية. وهي طريقة عامة وسهلة التطبيق، فلنفرض أن طول الفئة ω، وأن y_0 مركز الفئة الواقعة في الوسط تماما إذا كان عدد الفئات فرديا، أو مركز إحدى الفئتين الواقعتين في الوسط إذا كان عدد الفئات زوجيا. ولنطبق على مراكز الفئات التحويل:

$$Z_i = \frac{y_i - y_0}{\omega}$$

أي نطرح من مركـز كل فئة العدد y_0 ثم نقسم النـاتج على ω. ومن علاقة التحويل نستنتج أن :

$$y_i = \omega Z_i + y_0$$

وكما نعلم فإن :

$$\bar{y} = \omega \bar{Z} + y_0$$

و

$$S_y^2 = \omega^2 S_z^2 , S_y = |\omega| S_z = \omega S_z$$

ω هنا موجبة دوما باعتبارها طول فئة .

وسنجد أن المقاديـر Z_i أعداد صحيحـة متناظـرة حول الصفـر. وفي حالـة تسع فئات ، مثلا ، سنجد المقادير Z_i على الشكل $-4, -3, -2, -1, 0, 1, 2, 3, 4$ وبالطبع فإن التعامل مع هذه الأعداد أسهل كثيرا . والآن نعتبر هذه الأعداد مراكز للفئات وننجز الحسابـات تماما كما في الفقـرة السابقـة (١ ـ ٨ ـ ٣) فنحصل على $\bar{Z}$ و S_z^2 و S_Z بسهـولـة ، ومنها نستنتـج المتوسط والتبـاين والانحـراف المعيـاري للبيـان الأصلي قبل التحويل من خلال العلاقات :

$$\bar{y} = \omega \bar{Z} + y_0 \ , S_y^2 = \omega^2 S_z , S_y = \omega S_z$$

مثال (١ ـ ٢٣)

بالعـودة إلى المثـال (١ ـ ٢٢) ، احسب المتـوسط والانحـراف المعيـاري بطريقـة تحويل البيان الإحصائي .

الحل

لديناتسع فئات ، والفئة الواقعة في الوسط هي الفئة الخامسة ومركزها $y_0 = 104$. وطول الفئة $\omega = 5$. وبإجراء التحويل :

$$Z_i = \frac{y_i - 104}{5}$$

تصبح مراكز الفئات

$$Z_1 = \frac{y_1 - 104}{5} = \frac{84 - 104}{5} = -4$$

$$Z_2 = \frac{y_2 - 104}{5} = \frac{89 - 104}{5} = -3$$

وهكذا.

وبدلا من الجدول (١ ـ ١٨) ننظم الجدول (١ ـ ١٩)، التالي :

جدول (١ ـ ١٩)

مركز الفئة yi	التكرار f_i	Z_i	$f_i Z_i$	$f_i Z_i^2$
84	1	-4	-4	16
89	2	-3	-6	18
54	4	-2	-8	16
99	7	-1	-7	7
104	9	0	0	0
109	10	1	10	10
114	7	2	14	28
119	6	3	18	54
124	4	4	16	64
المجموع	50		33	213

$$\bar{Z} = \frac{33}{50} = 0.66$$

$$S_z^2 = \frac{1}{49}\left[213 - \frac{(33)^2}{50}\right] = \frac{191.22}{49}$$

ومنه

$$\bar{y} = \omega\bar{Z} + 104 = 5 \times 0.66 + 104 = 107.3$$

$$S_y^2 = \omega^2 S_z^2 = 25 \times \frac{191.22}{49} = 97.56 \;\; ; \;\; S_y = 9.88$$

وهي الأجوبة ذاتها التي حصلنا عليها في المثال (١ ـ ٢٢).

(١ ـ ١٠) حول الأهمية العملية للمتوسط والإنحراف المعياري

من الطبيعي أن نتساءل عن مدى نجاح التباين S^2 في التعبير عن خاصية التغير في جملة من القياسات. وسنجد الجواب الصريح عن هذا التساؤل في نقطتين نعرضهما فيما يلي:

١ ـ لنأخذ مجموعة القياسات $1, 2, 3, 4$، ولنحسب تباينها:

$$S^2 = \frac{1}{3}\left[\left(1^2 + 2^2 + 3^2 + 4^2\right) - \frac{(1+2+3+4)^2}{4}\right]$$
$$= \frac{1}{3}[30 - 25] = \frac{5}{3}$$

ولنحسب، على الوجه الآخر، الفروق بين كل قياس والقياسات الباقية، كما في الجدول (١ ـ ٢٠)، ثم لنحسب متوسط مربعات هذه الفروق فنجد $\frac{40}{12} = \frac{10}{3}$. أي أن متوسط مربعات الفروق الموجودة بين القياسات كافة يساوي $2S^2$.

جدول ١ ـ ٢٠

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

***وبصورة عامة**

إذا كانت $x_n, \ldots, x_1$ عينة من القياسات، متوسطها $\bar{x}$ وتباينها S^2، فإن عدد الأزواج المختلفة من القياسات التي يمكن تشكيلها هو $n(n-1)$. ومتوسط مربعات الفروق بين العددين في كل زوج منها هو:

*للقراءة فقط.

$$\frac{1}{n(n-1)}\sum_{i}^{n}\sum_{j}^{n}(x_i-x_j)^2$$

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

$$=\frac{1}{n(n-1)}\sum_{i}^{n}\sum_{j}^{n}\left[(x_i-\bar{x})^2+(x_j-\bar{x})^2-2(x_i-\bar{x})(x_j-\bar{x})\right]$$

$$=\frac{1}{n(n-1)}\sum_{i}^{n}\left[n\ (x_i-\bar{x})^2+(n-1)\,S^2+0\right]$$

$$=\frac{1}{n(n-1)}\left[n(n-1)\ S^2+n(n-1)\ S^2\right]=2S^2.$$

وهـذا يوضح أن التبـاين S^2 يلخص بأمـانـة كافـة التغيرات من قيـاس إلى آخـر التي يتضمنها البيان الإحصائي. وبالتالي فإنه يشكل تعبيرا ناجحا عن خاصية التغير ضمن البيان الإحصائي.

٢ ـ هنـاك متباينـة مشهـورة تسمى متباينـة تشيبيشيف، ويمكن التعبير عنها بطـريقة مبسطة كما يلي:

لتكن $x_1, x_2, \ldots, x_n$ جملة من القيـاسات متوسطها $\bar{x}$ وانحـرافها المعياري S . وليكن t عددا أكبر من الواحد أو يساويه، فـالنسبة من هذه القياسات التي تقع ضمن الفترة $(\bar{x}-ts\ ,\bar{x}+ts)$ لا تقل عن $1-\frac{1}{t^2}$.

لنختر الآن بعض القيم لـ t ، ولنحسب النسبة $1-\frac{1}{t^2}$ فنجد:

جدول (١ ـ ٢١)

t	1	2	3
$1-\frac{1}{t^2}$	0	$\frac{3}{4}$	$\frac{8}{9}$

فالمتباينة لا تقدم أية معلومات من أجل $t = 1$. ولكنها تقول، في حالة $t = 2$ ؛ أن ثلاثة أربـاع القياسـات، على الأقل، واقع ضمن فترة تمتـد ضعف الانحراف المعيـاري على جانبي المتوسط. أي تقع ضمن الفترة $(\bar{x} - 2S \, , \, \bar{x} + 2S)$ وتقول في حالة $t = 3$ ، أن ما لا يقل عن ثمانية أتساع القياسات (89% تقريبا) واقع ضمن فترة تمتـد بمقدار ثلاثة انحرافات معيارية على جانبي المتوسط، أي تقع بين العدد $\bar{x} + 3S$ والعدد $\bar{x} - 3S$.

مثال (١ ـ ٢٤)

لنعد إلى البيان الإحصـائي في الجدول (١ ـ ٤)، فقــد حسبنا في المثـال (١ ـ ٦) متـوسطـه فوجـدنـاه $\bar{x} = 107.3$ وحسبنـا في المثـال (١ ـ ٢٢) الانحـراف المعيـاري فوجدناه $S = 9.88$. ولدينا

$$\bar{x} - 2S = 107.3 - 2 \times 9.88 = 87.54$$

$$\bar{x} + 2S = 107.3 + 2 \times 9.88 = 127.06$$

ولو تفقدنا القياسات الخمسين في الجدول (١ ـ ٥)، لوجدنا أن 49 منها واقع بين 87.54 و 127.06، وهي تشكل نسبة $\frac{49}{50} = 98\%$ من القياسات.

ومما تقدم نستنتج بوضوح أن التباين S^2 يشكل مقياسا كميـا ناجحا تماما للتعبير عن خاصية التغير ضمن بيان إحصائي. وأصبح واضحا الآن أن متوسط بيان إحصائي $\bar{x}$ ، وانحرافه المعياري S ، يلخصان بصورة جيدة قياسات ذلك البيان. ومن خلالهما، يمكن تشكيل صـورة ذهنية جيـدة للغايـة عن التـوزيع التكراري للبيـان دون أن نعلم مفردات البيان.

وعلى سبيل المثال، لـو قيل لنـا أن درجـات فصل يتألف من 40 طالبـا في مادة الإحصاء، لها متـوسط يساوي 72 ، وانحراف معيـاري يساوي 8 ، لأمكننا باستخدام هذين الـرقمين فقط، تقديم الوصف التالي لتـوزيع الدرجات، دون أن تكون لـدينا أية معلومات أخرى عن واقع الدرجات نفسها:

تتمركز الدرجات في هـذا الفصل حول القيمة 72 ، وما لا يقل عن ثـلاثين طالبا حصلوا على درجات تتراوح بين $56 = 8 \times 2 - 72$ و $88 = 8 \times 2 + 72$. وما لا يقل عن

36 طالبا من الطلاب الأربعين حصلوا على درجات تتراوح بين $48 = 8 \times 3 - 72$ و $96 = 8 \times 3 + 72$.

ويجدر الانتباه إلى عبارة «ما لا يقل» فمتباينة تشيبيشيف متحفظة، وفي معظم الحالات تكون النسبة الفعلية أكبر من $1 - \frac{1}{t^2}$ خاصة إذا كان البيان الإحصائي قريبا من التناظر.

(١ ـ ١١) معامل التغير

رأينا أن التباين يعبر بنجاح عن خاصية التغير في بيان إحصائي. ومن الطبيعي أن يكون البيان الإحصائي أكثر تجانسا كلما كانت قياساته أقل تغيرا من أحدها إلى الآخر. وكلما زاد التباين استنتجنا أن البيان الإحصائي أقل تجانسا، ولكن هب أننا نريد مقارنة بيانين إحصائيين من حيث أيهما أكثر تجانسا من الآخر، فهل يمكن الاعتماد على مقارنة تباينيهما وإعطاء حكم في هذه المسألة؟ لقد وجدنا في الفقرة (١ ـ ٨ ـ ٣) أن مقدار التباين يعتمد على وحدة القياس المستخدمة في البيان الإحصائي مما يجعله غير صالح للمقارنة بين عينتين من القياسات من حيث درجة التجانس في كل منهما. وهناك عامل آخر، إذ بالرغم من استخدام وحدة القياس نفسها في البيانين اللذين نريد مقارنتهما، إلا أن طبائع الأمور قد تجعل أرقام البيان الأول كبيرة، وأرقام البيان الثاني صغيرة. كأن يتضمن البيان الأول أوزان مجموعة من العجول بالكيلوغرام، ويتضمن البيان الثاني أوزان مجموعة من الفراريج بالكيلوغرام أيضا. ونوضح بالمثال التالي:

مثال (١ ـ ٢٥)

في مزرعة خمسة عجول، وعشرون فروجا، سجلنا الأوزان ضمن كل مجموعة بالكيلوغرام فحصلنا على البيانين التاليين:

العجول :	285.50;	280.40;	283.00;	280.75;	281.40;		
الفراريج :	1.50;	1.40;	0.95;	1.35;	1.45;	1.05;	1.05;
	0.99;	1.45;	1.50;	1.35;	1.45;	1.00;	1.10;
	1.25;	1.35;	1.10;	1.45;	1.00;	1.20;	

احسب المتوسط والتباين والانحراف المعياري لكل منها .

الحل

	المتوسط	التباين	الانحراف المعياري
العجول	282.21	4.37	2.09
الفراريج	1.26	0.036	0.19

ولو استخدمنا ، في المثال السابق ، الانحراف المعياري للمقارنة والحكم على درجة تجانس كل من البيانين ، لاستنتجنا خطأ أن مجموعة الفراريج أكثر تجانسا من مجموعة العجول ، لأن انحرافها المعياري ، وبالتالي تباينها ، أصغر بكثير. ولكن صغر الانحراف المعياري للفراريج ، يعود إلى صغر أوزان الفراريج بالمقارنة مع أوزان العجول ، وليس لكونها أكثر تجانسا .

وسنعرف الآن مقياسا يسمى معامل التغير، وهو لا يعتمد على وحدة القياس المستخدمة ، ولا يتأثر بكون القياسات كبيرة أو صغيرة ، مما يجعله صالحا لمقارنة درجتي التجانس في عينتين من القياسات ، وذلك بصرف النظر عن طبيعة هذه القياسات أو عن وحدات القياس المستخدمة في كل منهما .

تعريف معامل التغير

معامل التغير، ونرمز له بـ $c.v$ ، لجملة من القياسات متوسطها $\bar{x}$ ، وانحرافها المعياري S ، هو بالتعريف :

$$c.v = \frac{S}{\bar{x}}$$

ونعلم من خواص المتوسط وخواص التباين أنه إذا ضربنا كل قياس في جملة من القياسات بعدد معين ، فإن كلا من المتوسط $\bar{x}$ والإنحراف المعياري S ، سيضرب

بالعدد نفسه، (أو يقسم على العدد نفسه) وبالتالي ستبقى النسبة $\frac{S}{\bar{x}}$ بدون تغيير. وإذا كانت أرقام أحد البيانين كبيرة بطبيعتها وأرقام الآخر صغيرة، فإن قسمة S على $\bar{x}$ يعطينا الانحراف المعياري لكل وحدة قياس، مما يخلص معامل التغير من أي أثر لحجم القياسات.

مثال (١ ـ ٢٦)

في المثال السابق (١ ـ ٢٥)، أحسب معامل التغير لكل من جملتي القياسات وقارنهما من حيث درجة التجانس ضمن كل منهما.

الحل

$$c.v.\ (\text{للعجول}) = \frac{S}{\bar{x}} = \frac{2.09}{282.21} = 0.007$$

$$c.v.\ (\text{للفراريج}) = \frac{S}{\bar{x}} = \frac{0.19}{1.26} = 0.15$$

ويتضح الآن أن مجموعة العجول أكثر تجانسا بكثير من مجموعة الفراريج، فمعامل تغيرها 0.007 ، بينما معامل تغير الفراريج 0.15 ، وهو أكبر من معامل تغير العجول بما ينوف على إحدى وعشرين مرة.

(١ ـ ١٢) القيمة المعيارية

وسننتقل الآن إلى مشكلة أخرى، فلنفرض أن لدينا جملتين من القياسات، فكيف يمكن، عند الحاجة، مقارنة قياس من الجملة الأولى بقياس من الجملة الثانية؟

وعلى سبيل المثال

لنفرض أن الجملتين من القياسات هما درجات طلاب الفصل في مادة الرياضيات، ودرجاتهم في مادة اللغة العربية، ونريد مقارنة درجتي طالب معين في المادتين، وإذا فرضنا أن درجته في الرياضيات كانت 70 ، وأنها في اللغة العربية 60 ،

فهل يعني ذلك أن تحصيله في الرياضيات أفضل من تحصيلـه في اللغة العربية؟ المسألة هنا نسبية، فقد يكون معظم طـلاب الفصل نالوا درجات أعلى من 70 في الرياضيات، ولكن قليلا منهم فقط نـال درجات تزيـد على الستين في اللغة العربيـة. وفي مثل هذه الحالة تنعكس الآية فنقول، على عكس ما يوحيه الرقمان، إنه كان من المتفوقين في اللغة العـربية، ومن المقصرين في الـريـاضيات. والطـريقة التي تسمـح لنا بمـراعاة الـواقع النسبي، واتخاذ الحكم الصحيح، هي حساب متـوسط كل جملة وانحرافهـا المعياري. ثم نرد كل درجة إلى ما يسمى بقيمتها المعيارية، بأن نطرح منها المتوسط ثم نقسم الناتج على الانحراف المعياري. وبذلك نحسب كم انحرافا معياريا تبتعد الدرجة عن متوسط الـدرجات؟ أو بعبـارة أخرى، نقيس الفـرق بين الدرجـة والمتوسط بـوحدة قيـاس هي الانحـراف المعياري للـدرجـات. ولنفرض في مثـالنا هنـا أن متـوسط درجات طـلاب الفصـل في مــادة الرياضيـات كـان 75 بانحـراف معياري يسـاوي 5، وأن متـوسط درجـات الطلاب في مادة اللغـة العربيـة كان 52، بانحـراف معياري يسـاوي 6.

فالـدرجة المعياريـة في الرياضيات هي $1- = \frac{70-75}{5}$ والـدر جة المعياريـة في اللغة العـربيـــة هي $1.33 = \frac{60-52}{6}$. وهي أكبر من 1-، أي أن تحصيله في اللغة العربية أفضل.

تعريف القيمة المعيارية

إذا كـان $\bar{x}$ و S متوسط جملـة من القياسات وانحرافهـا المعياري، على الترتيب. فنعرف القيمة المعيارية لأي قياس، x، من هذه الجملة، بأنها:

$$\frac{x-\bar{x}}{S}$$

ونلاحظ أن رد جملـة من القيــاســات إلى الشكل المعيــاري، أو معـايــرة جملـة القياسـات، يجعل متـوسطها مسـاويا للصفـر، وانحرافهـا المعياري مسـاويا للـواحد الصحيح. ولبيان ذلك نكتب:

لتكن $x_1, x_2, \ldots, x_n$ جملة من القياسـات متوسطها $\bar{x}$ وانحرافها المعياري S. ووفقا لتعريف المعايرة يمكن أن نكتب:

$$Z_i = \frac{x_i-\bar{x}}{S} \quad , \quad i = 1, 2, \ldots, n.$$

حيث رمزنا بـ Z_i للقيم المعيارية. لنحسب الآن: $\bar{Z}$ متوسط القيم المعيارية و S_Z^2 انحرافها المعياري فنجد:

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S} = \frac{1}{nS} \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

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

وهذا يعني أن معايرة جملتين من القياسات تردهما إلى جملتين لهما المتوسط نفسه، وهو الصفر، والانحراف المعياري نفسه، وهو الواحد.

بقيت ملاحظة أخيرة، وهي أنه إذا كانت $Z_1, Z_2, \ldots, Z_n$ القيم المعيارية لجملة من القياسات فإن

$$\sum_{i=1}^{n} (Z_i - \bar{Z})^2 = (n-1)\, S_Z^2 = (n-1) \times 1 = n - 1$$

وبما أن $\bar{Z} = 0$ ، فإن

$$\sum_{i=1}^{n} Z_i^2 = n - 1$$

أي أن مجموع مربعات القيم المعيارية لجملة من القياسات يساوي عدد القياسات n مطروحا منه الواحد. وسنستفيد من هذه الخاصة في الفقرة القادمة.

تمارين (١ ـ ٤)

١) فيما يلي الأطوال بالسنتيمتر لعشرة أوراق من نبات منزلي:

10.0; 10.2; 6.5, 7.0, 7.8, 10.8, 6.1, 5.9, 8.9, 10.0

احسب المدى، ومتوسط الانحراف، والتباين، والانحراف المعياري.

٢) استخدمنا سبعة موازين حرارة لقياس درجة حرارة جسم بالتدريج المئوي. فكانت النتائج كما يلي:

$$-4.12, -4.09, -4.10, -4.08, -4.09, -4.13, -4.10$$

احسب التباين والانحراف المعياري.

٣) ماذا يمكن القول عن مجموعة قياسات تباينها يساوي الصفر؟ وإذا حسبت تباين جملة من القياسات فوجدته سالبا فماذا تستنتج؟

٤) في كل مما يلي أحسب المدى والانحراف المعياري:

أ ـ $4, 2, 8, 1, 4, 5, 8, 10, 3$

ب ـ $-5, 3, -1, -4, 3, -8, -2$

جـ ـ $1, 2, 3, 0, -3, 3, 3, -1$

تحقق في (أ) أنك إذا أخذت متوسط مربعات انحرافات كل قياس عن بقية القياسات فإن النتائج يساوي ضعفي التباين.

٥) فيما يلي التوزيع التكراري لعدد القطع المعيبة التي وجدت في 404 صناديق من القطع المصنعة.

عدد القطع المعيبة	0	1	2	3	4	5	6	7	8	9	10	11	12
عدد الصناديق	53	110	81	58	35	20	18	12	9	3	1	2	1

ا ـ احسب المتوسط والتباين ومعامل التغير.

ب ـ احسب الوسيط والمنوال.

٦) إذا كان تباين عينة تتضمن مائة قياس هو 15 ، فاحسب مجموع مربعات انحرافات القياسات عن متوسطها.

٧) تباين عينة من القياسات يساوي 20 . كم يصبح التباين :

ا ـ إذا ضربنا كل قياس بـ 5 ؟

ب ـ إذا قسمنا كل قياس على 5 ؟

٨) أخذنا عينتين من مجتمعين فأعطتا النتائج التالية :

العينة الثانية	العينة الأولى
$\sum_{i=1}^{50} y_i = 400$	$\sum_{i=1}^{30} x_i = 270$
$\sum_{1}^{50} y_i^2 = 3984$	$\sum_{1}^{30} x_i^2 = 2691$

ا ـ احسب تباين كل عينة .

ب ـ أيهما أكثر تجانسا؟

جـ ـ إذا دمجنا العينتين في عينة واحـدة فاحسب متوسط العينة الجديدة ومعامل تغيرها .

٩) احسب نصف المدى الربيعي ومعامل التغير في كل من التمارين ٤ ، ٥ ، ٦ ، ٧ من مجموعة التمارين (١ ـ ١) .

١٠) احسب التباين في كل من التمرينين ٩ ، ١١ من مجموعة التمارين (١ ـ ٢) .

١١) احسب الانحراف المعياري ومعامل التغير لسماكة الجلد في التمرين ٨ من مجموعة التمارين (١ ـ ١) ثم تحقق من أن 95% تقريبا من القياسات واقع في حدود انحرافين معياريين عن يمين ويسار المتوسط .

١٢) بالإشارة إلى التمرين ٧ من مجموعة التمارين (١ ـ ٣)، احسب التباين والانحراف المعياري لنصف قطر رد الفعل لاختبار الليبرومين في كل من التوزيعات الثلاثة .

١٣) احسب الانحراف المعياري ومعامل التغير لكل من البيانات المعطاة في التمارين ٦، ٧، ١٠ من مجموعة التمارين (١ ـ ١).

١٤) فيما يلي بيانات تتعلق بمنطقة معينة لعامي ١٩٥٩ و١٩٦٠. أحسب لكل بيان، المدى، والانحراف المعياري، ومعامل التغير. أي البيانات الثلاثة أكثر تجانسا؟

متوسط الرطوبة النسبية عند التاسعة صباحا (٪)	متوسط درجة الحرارة (بالفهرنهايت)	معدل سقوط المطر (بالبوصة)	الشهر
78	72.1	1.45	يناير
78	72.5	1.44	فبراير
78	72.1	2.69	مارس
77	72.6	5.15	أبريل
79	73.3	7.46	مايو
85	73.2	0.73	يونيو
72	72.8	0.51	يوليو
78	71.9	5.17	أغسطس
78	71.4	4.20	سبتمبر
78	71.7	4.08	أكتوبر
78	71.6	6.68	نوفمبر
79	71.6	2.77	ديسمبر

١٥) في تجربة لتقدير فائدة مضاد للتسمم في معالجة الكزاز، قورنت مجموعة تناولت المضاد مع مجموعة لم تتناوله. وقد تم تخصيص المرضى للمجموعتين بطريقة عشوائية، وفيما يلي بيان بأعمار المرضى. ارسم مضلع التكرار النسبي المتجمع الصاعد لكل من المجموعتين على حدة واستخدمها لتقدير العمر الوسيط ونصف المدى الربيعي لكل مجموعة.

١٦) احسب نصف المدى الربيعي ومعامل التغير في التمرين (٥) من مجموعة التمارين (١ ـ ٣).

لم يتناول مضاد للتسمم (N)		تناول مضاد للتسمم (A)	
33	18	16	41
20	24	28	28
39	19	27	35
36	12	20	40
30	29	17	30
60	14	12	27
17	18	12	50
27	18	16	30
33	50	20	9
14	16	10	40
10	14	11	30
60	52	20	18
12	16	50	31
24	40	29	14
12	30	24	25
10	40	14	27
60	40	17	16
27	27	25	36
8	20	10	25
		24	40
			22

١٧) إذا علمت توزيع التكرار النسبي التالي :

الفئة	19.5 - 39.5	39.5 - 59.5	59.5 - 79.5	79.5 - 99.5
التكرار النسبي	0.12	0.28	0.36	0.24

فاحسب المتوسط ، الوسيط ، المنوال ، Q_1 ، Q_3 ، σ^2 .

١٨) كان متوسط معدلات الطلبة المتقدمين لإحدى الجامعات 20.4 بانحراف معياري 3.1 ، ومتوسط معدلات الطلبة المتقدمين لجامعة أخرى 21.1 بانحراف معياري 2.8 .
إذا تقدم طالب معدله 25 إلى كل من الجامعتين ففي أيهما ستكون فرصة قبوله أفضل؟

١٩) في دراسة قام بها مركز للأغذية تبين أن متوسط مقدار الفيتامين B في عدد من شرائح الخبز هو 0.26 ملغم. بانحراف معياري قدره 0.005 ملغم. استخدم هذه المعلومات لإكمال العبارات التالية:
- ما لا يقل عن 25/36 من هذه الشرائح يحتوي على مقدار من فيتامين B واقع بين (. . . و . . .).
- ما لا يقل عن 63/64 من هذه الشرائح يحتوي على مقدار من فيتامين B واقع بين (. . . و . . .).

٢٠) إذا علمت أن معامل تغير بيان إحصائي يتضمن ثمانين قياسا هو 0.1 وأن مجموع قياساته 800، فاحسب مجموع مربعات القياسات $\sum x_i^2$.

٢١) قمنا بدراسة زمنية لتحديد الوقت الذي يستغرقه إنجاز عملية معينة في مؤسسة صحية. وقد قسنا الزمن الضروري لإنجاز هذه العملية لكل من 40 عاملا، ووجدنا أن المتوسط يساوي 12.8 وحدة زمن بانحراف معياري يساوي 1.7 وحدة زمن. والمطلوب إعطاء وصف للبيان الإحصائي مستخدما متباينة تشيبيشيف.

٢٢) لديك المعلومات التالية عن أسعار مجموعة من مطاعم الدرجة الأولى في مدينة معينة:

الوجبة	متوسط الكلفة	الانحراف المعياري S
لحم	24.25 ر. س	1.50
دجاج	13.72 ر. س	0.94
سمك	33.65 ر. س	1.13

وأحد هذه المطاعم ويسمى «مطعم التـوفير» يقدم وجبة اللحم في مقابل 28 ريالا، ووجبة الدجاج في مقابل 17 ريالا، ووجبة السمك في مقابل 36 ريالا، هل تعتقد أن هذا المطعم يستحق الإسم الذي يدّعيه؟ ولماذا؟

(١ ـ ١٣) الارتباط

(١ ـ ١٣ ـ ١) مقدمة

لدينـا مجموعـة n من الأشخـاص، ولنفرض أننا قمنـا بقياس ظـاهرتين لـدى كل شخص منهـا، ورمزنـا لقياس إحـداهما بـ x، ولقيـاس الأخرى بـ y (مثـلا، x ترمز للطول، y تـرمـز للوزن). فحصلنـا بـذلك على n من أزواج الأعـداد، (x_1, y_1) لأول شخص، (x_2, y_2) للشخص الثاني، . . . ، وأخيرا (x_n, y_n) للشخص الأخير.

لنرتب القيم x_i من الأصغر إلى الأكبر، ثم لنضع أمام كل قيمة لـ x قيمة y الموافقة لها. ولنفرض أننا وجدنـا قيم y مـرتبة أيضـا من الأصغر إلى الأكبر، فأصغـر قيمة لــ x قابلتها أصغر قيمة لـ y (أي أن الشخص ذا الطول الأصغر كـان أيضا ذا الوزن الأصغر بين الأشخاص الـ n الخاضعين للتجربة)، والقيمة بعد الصغرى لـ x قابلتها القيمة بعد الصغرى لـ y، . . . . ، وأخيرا مقابل أكبر قيمة لـ x كان بين قيم y أكبرها.

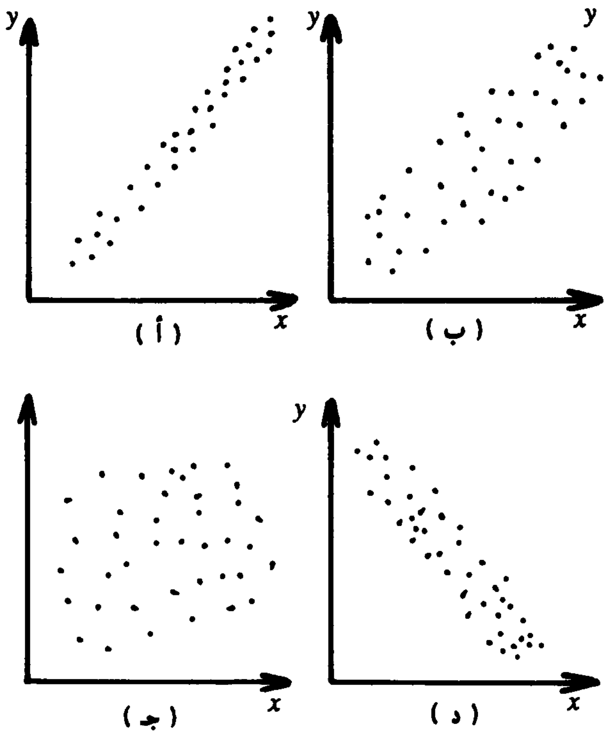
ففي مثل هذه الحالة نقول بوجود ارتباط إيجابي كامل بين المتغيرين x و y، أو بين الظاهـرتين اللتين تقيسانها. وإذا وجدنا عنـد ترتيب القيم أن أصغر قيمـة لـ x قابلتها أكبر قيمة لـ y. والقيمة بعد الصغـرى لـ x قابلتهـا القيمة قبل العظمـى لـ y، . . . ، وأخيرا أكبر قيمة لـ x قابلتها أصغر قيمة لـ y. فعندئذ نقول بوجود ارتباط سلبي كامل بين المتغيرين x، y؛ أو بين الظاهرتين اللتين تقيسـانها. وبين هاتين الحالتين المتطرفتين يمكن أن نتصور ترتيبات تمثل درجات مختلفة من الارتباط في الإتجاه الإيجابي أو في الإتجاه السلبي. ولو أننـا سجلنا قيم y على n ورقة صغيرة، وطـويناها ثم خلطناهـا جيدا في جعبة صغيرة، وسحبنـا عشوائيا ورقـة منها ثم سجلنا القيمـة المذكورة فيها أمـام أصغر قيمة لـ x، وسحبنا ورقـة ثانية عشـوائيا وسجلنا القيمة المذكـورة فيها أمام القيمـة بعد الصغرى لـ x ، وهكذا. . . ، حتى نصل إلى آخر ورقـة بقيت في الجعبة فنسجل القيمة

المذكورة فيها أما أكبر قيمة لـ x، فيمكن القول، مع مثل هذا الترتيب أو التقابل بين قيم x وقيم y، بعدم وجود أي ارتباط بين الظاهرتين. ويمكن تحري وجود صلة بين المتغيرين برسم أزواج القياسات (x_1, y_1)، (x_2, y_2)، . . . ، (x_n, y_n) بيانيا، حيث قيمة x هي الإحداثي السيني، وقيمة y المقابلة هي الإحداثي الصادي. ونحصل بذلك على n نقطة في مستوى الإحداثيات ونسمي الشكل الحاصل «مخطط الإنتشار». والنظر إلى مخطط الإنتشار يولد نوعا من الانطباع البدهي عن درجة الصلة أو الارتباط القائمة بين المتغيرين.

والشكل (١ ـ ١٣ ـ ا) يمثل حالة ارتباط إيجابي مرتفع، ونلاحظ فيه أن النقاط تحدد اتجاها واضحا وفق خط مستقيم إلى حد كبير. ولو وقعت النقاط بالضبط على استقامة واحدة، لكان الارتباط إيجابيا تاما. والشكل (١ ـ ١٣ ـ ب) يمثل ارتباطا إيجابيا منخفضا إذ يكشف المخطط عن نزعة تأخذ، إلى حد ما، شكل الحزمة الخطية. أما الشكل (١ ـ ١٣ ـ جـ) فيمثل حالة عشوائية، ولا تكشف عن أية نزعات أو اتجاهات واضحة، إذ لا يبدو فيها أي نزوع لاقتران قيم عالية لـ x بقيم عالية لـ y، وقيم منخفضة لـ x بقيم منخفضة لـ y. أو العكس، أي قيم عالية لـ x بقيم منخفضة لـ y وقيم منخفضة لـ x بقيم عالية لـ y. ويمثل الشكل (١ ـ ١٣ ـ د) ارتباطا سلبيا مرتفعا إلى حد ما، وهنا أيضا، لو وقعت النقاط على استقامة واحدة لكان الارتباط سلبيا تاما. ومن الواضح أنه بين الحالتين المتطرفتين، حالة ارتباط سلبي تام وحالة ارتباط إيجابي تام. يوجد ما لا حصر له ولا عد من إمكانات ترتيب النقاط التي تمثل ما لا حصر له ولا عد من درجات الارتباط الممكنة بين المتغيرين.

ولا بد من التمييز بوضوح بين وجود ارتباط مرتفع بين ظاهرتين وبين وجود علاقة سببية بينهما. فوجود ارتباط مرتفع لا يعني بالضرورة أن إحدى الظاهرتين هي سبب للأخرى؛ إذ قد يكون الارتباط المرتفع بينهما نتيجة لتأثر كل منهما بظاهرة ثالثة لم تدخل في الحساب.

فمثلا، من المعروف أن هناك ارتباطا مرتفعا بين ظاهرة الابتلاء بعادة التدخين والإصابة بمرض سرطان الرئة. وهناك أيضا ارتباط مرتفع بين ظاهرة الابتلاء بعادة



شكل (١ ـ ١٣)

التدخين وتلون أو اصفرار الأسنان. ولو حصل أن أخذنا بيانا إحصائيا يتضمن درجة تلون الأسنان ونسبة الإصابة بسرطان الرئة، وكان هذا البيان في غالبيته من أفراد تلونت أسنانهم بفعل التدخين فسنجد ارتباطا مرتفعا بين ظاهرة تلون الأسنان وظاهرة الإصابة بسرطان الرئة. وهذا لا يعني بالطبع أن اصفرار الأسنان يؤدي إلى الإصابة بسرطان الرئة أو العكس، وقد لا يوجد أي ارتباط إحصائي فعلي بين الظاهرتين، فالارتباط المرتفع كان نتيجة لوجود عامل ثالث خفي هو عادة التدخين.

وسنستعرض الآن إمكانية إيجاد معيار كمي للتعبير عن درجة الارتباط بين متغيرين نسميه معامل الارتباط.

(١ ـ ١٣ ـ ٢) معامل بيرسون للارتباط

هناك أكثـر من صيغة للتعبير عن معـامل الارتباط بين متغيرين x و y ؛ ولكنها تعرَّف جميعها لتأخذ قيمها بين 1- تعبيرا عن ارتباط سلبي تام، (وعندئذ تقع جميع النقاط (x, y) على خط مستقيم تتناقص معه قيم y عندما تزداد قيم x، وتتزايد y عندما يتناقص x) وبين 1+ تعبيرا عن ارتبـاط إيجابي تـام. (وعنـدئذ تقع جميع النقـاط (x, y) على خط مستقيم يزداد وفقـا له أحد المتغيرين مع زيادة الآخـر ويتناقص مع تناقصه) أمـا القيمة صفر فتعني عدم وجود أي ارتباط أو نزعة أثر أو تأثر بين قيم أحد المتغيرين وقيم المتغير الآخر. ومقياس الارتباط الأكثر استخداما هو معامل بيرسون، ونرمز له عادة بـ R.

تعريف معامل بيرسون

لتكن $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ جملة من n من أزواج القياسات. ولنفرض أن $\bar{x}$ و S_x هما متوسط قيم المتغير x وانحـرافها المعيـاري، وأن $\bar{y}$ و S_y هما متوسط قيم المتغير y وانحرافهـا المعياري. نعـرف معامل بيرسـون للارتبـاط بين قيم المتغير x وقيم المتغير y بأنه:

$$R = \frac{1}{n-1} \sum_{i=1}^{n} Z_i \, Z'_i$$

حيث

$$Z'_i = \frac{y_i - \bar{y}}{S_y} \qquad Z_i = \frac{x_i - \bar{x}}{S_x}$$

وقد رأينا في ختام الفقرة السابقة أن معايرة جملة من القياسات تجعل متوسطها صفرا، وتباينها الواحد، وأن مجموع مربعات القيم بعد معايرتها يساوي عدد القياسات في الجملة مطروحا منه الواحد. وهكذا يمكننا كتابة:

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} Z_i'^2 = n-1$$

لنـأخــذ الآن الحـالة الخـاصـة التي يكون فيها $Z_i = Z'_i$ ، فعندئذ تقع النقاط $(Z_1, Z'_1), \ldots, (Z_n, Z'_n)$ على خـط مستقيم هو منصف الربع الأول، ويكون الارتباط في هذه الحالة إيجابيا وتاما. لنحسب R فنجد:

$$R=\frac{1}{n-1}\sum_{i=1}^{n} Z_i\, Z'_i=\frac{1}{n-1}\sum_{i=1}^{n} Z_i^2=\frac{n-1}{n-1}=1$$

وإذا أخذنا الحالة الخاصة المتطرفة المقابلة حيث Z_i و Z'_i متساويان في القيمة المطلقة ومختلفان في الإشارة، فعندئذ تقع النقاط $(Z_n, Z'_n), \ldots, (Z_1, Z'_1)$ على خط مستقيم هو منصف الربع الثاني، ويكون الارتباط في هذه الحالة سلبيا وتاما، أما قيمة R فهي 1-، ذلك لأن:

$$R=\frac{1}{n-1}\sum_{i=1}^{n} Z_i\, Z'_i=\frac{1}{n-1}\sum_{i=1}^{n} Z_i(-Z_i)=-\frac{1}{n-1}\sum_{i=1}^{n} Z_i^2=-\frac{(n-1)}{n-1}=-1$$

ويمكن البرهان، بصورة عامة، أن معامل بيرسون للارتباط يأخذ قيما بين 1 - و 1+ . ويكون 1+ في حالة ارتباط إيجابي تام و 1 - في حالة ارتباط سلبي تام.

مثال (١ ـ ٢٧)

لتكن أزواج القياسات التالية:

x	1	2	3	4	5
y	11	13	15	17	19

احسب معامل بيرسون للارتباط R .

الحل

ننظم الجدول التالي:

جدول (١ ـ ٢٢)

x_i	y_i	$Z_i = \frac{x_i - \bar{x}}{S_x}$	$Z'_i = \frac{y_i - \bar{y}}{S_y}$	$Z_i \, Z'_i$
1	11	-1.2649	-1.2649	1.60
2	13	-0.6325	-0.6325	0.40
3	0	0	0	0
4	17	0.6325	0.6326	0.40
5	19	1.2649	1.2649	1.60

حيث $S_y = 3.1623 \,,\ \bar{y} = 15 \,,\ S_x = 1.5811 \,,\ \bar{x} = 3$

$$\sum_{i=1}^{5} Z_i \, Z'_i = n - 1 = 4$$

$$R = \frac{1}{4} \sum_{i=1}^{5} Z_i \, Z'_i = \frac{4}{4} = 1$$

والجدير بالذكر أن $y = 2x + 9$ وأن النقاط الخمس:
$(1, 11), (2, 13)$ $(3, 15), (4, 17), (5, 19)$ تقع على استقامة واحدة. ونلاحظ أن $Z_i = Z'_i$

(١ ـ ١٣ ـ ٣) حساب معامل الارتباط R

عند حساب معامل الارتباط يشكل رد القياسات إلى شكلها المعياري جهدا حسابيا مطولا لا مسوغ له. ويمكن تطوير الصيغة المعطاة في تعريف معامل بيرسون بعمليات تعويض بسيطة بحيث تأخذ أشكالا مختلفة.

١ ـ بالتعويض عن Z_i ، Z'_i نجد:

$$R = \frac{1}{n-1} \sum_{i=1}^{n} Z_i \, Z'_i = \frac{1}{n-1} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S_x} \; \frac{y_i - \bar{y}}{S_y}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1) S_x S_y}$$

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

وأخيرا

$$R = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}}$$

حيث $X_i = x_i - \bar{x}$ ، $Y_i = y_i - \bar{y}$. ويمكن استخدام العلاقة الأخيرة في الحسابات .

مثال (١ ـ ٢٨)

لدينا أزواج القياسات التالية :

x	1	7	2	3	4	12	11	5	10	5
y	2	5	6	4	1	5	8	2	6	1

احسب معامل بيرسون للارتباط بين x و y .

الحل

ننظم الجدول التالي :

جدول (١ ـ ٢٣). حساب معامل الارتباط باستخدام الانحرافات عن المتوسط

	x_i	y_i	X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$	
	5	1	-1	-3	1	9	3	
	10	6	4	2	16	4	8	
	5	2	-1	-2	1	4	2	
	11	8	5	4	25	16	20	$\bar{x}=6$
	12	5	6	1	36	1	6	$\bar{y}=4$
	4	1	-2	-3	4	9	6	
	3	4	-3	0	9	0	0	
	2	6	-4	2	16	4	-8	
	7	5	1	1	1	1	1	
	1	2	-5	-2	25	4	10	
المجموع	60	40	0	0	134	52	84	

ولدينا بالتعريف :

$$R=\frac{\sum_{i=1}^{10} X_i\ X_i}{\sum_{i=1}^{10} X_i^2 \sum_{i=1}^{10} Y_i^2}$$

وبالتعويض من السطر الأخير في الجدول (١ ـ ٢٣) نجد :

$$R=\frac{84}{\sqrt{134\times 52}}=0.58$$

٢* ـ ومن المفضل، في الغالب، استخدام صيغة حسابية أخرى تعتمد على القياسات y_i، x_i نفسها. وفي الحقيقة، نجد بسهولة أن :

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}(x_i\ y_i-\bar{x}\ \ y_i-x_i\ \bar{y}+\bar{x}\ \ \bar{y})$$

* التفاصيل للقراءة فقط .

$$= \sum_{i=1}^{n} x_i\ y_i - \bar{x} \sum_{i=1}^{n} y_i - \bar{y} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}\ \bar{y}$$

$$= \sum_{i=1}^{n} x_i\ y_i - n\,\bar{x}\ \bar{y} - n\,\bar{y}\ \bar{x} + n\,\bar{x}\ \bar{y}$$

$$= \sum_{i=1}^{n} x_i\ y_i - n\,\bar{x}\ \bar{y}$$

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

$$= \frac{1}{n}\left[n \sum_{i=1}^{n} x_i\ y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right]$$

ونعلم أنه يمكن كتابة (انظر الفقرة ١ ـ ٨ ـ ٨):

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

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n}\left[n \sum_{1}^{n} y_i^2 - \left(\sum_{1}^{n} y_i\right)^2\right]$$

وبالتعويض في الصيغة الحسابية السابقة نجد:

$$R = \frac{n \sum_{1}^{n} x_i\ y_i - \left(\sum_{1}^{n} x_i\right)\left(\sum_{1}^{n} y_i\right)}{\sqrt{\left[n \sum_{1}^{n} x_i^2 - \left(\sum_{1}^{n} x_i\right)^2\right]\left[n \sum_{1}^{n} y_i^2 - \left(\sum_{1}^{n} y_i\right)^2\right]}}$$

ومع أن مظهر الصيغة معقد، إلا أن جدول الحسابات الضروري لتطبيقها يتضمن خمسة أعمدة فقط، وهي تعتمد كليا على القياسات نفسها، وأسهل صيغة للتطبيق عند توفر آلة حاسبة.

مثال (١ ـ ٢٩)

احسب معامل الارتباط R لأزواج القياسات المذكورة في المثال (١ ـ ٢٦) مستخدما الصيغة التي تعتمد على القياسات مباشرة.

الحل

ننظم الجدول التالي:

وبالتعويض في الصيغة الحسابية التي يمكن أن نكتبها باختصار كما يلي:

$$R=\frac{n\Sigma\ x\ y-(\Sigma x)(\Sigma y)}{\sqrt{\left[n\Sigma x^2-(\Sigma x)^2\right]\left[n\Sigma y^2-(\Sigma y)^2\right]}}$$

$$R=\frac{10\times 288-60\times 40}{\sqrt{\left(10\times 494-60^2\right)\left(10\times 212-40^2\right)}}$$

$$=\frac{480}{\sqrt{1340\times 520}}=0.\ 58$$

(١ ـ ١٣ ـ ٤) معامل سبيرمان لارتباط الرتب

ذكرنا في المقدمة أنه إذا كان لمتغيرين x، y ترتيبان متوازيان أي إذا اتفق ترتيب قيم x مع ترتيب قيم y المقابلة اتفاقا تاما كنا في حالة ارتباط إيجابي تام وإذا كان لهما ترتيبان متعاكسان تماما (أصغر قيمة لـ x قابلتها أكبر قيمة لـ y، والقيمة بعد الصغرى لـ x قابلتها القيمة قبل العظمى لـ y، وهكذا حتى نصل إلى أكبر قيمة لـ x وفي مقابلها أصغر قيمة لـ y) قلنا إن الارتباط سلبي تام. ومعامل سبيرمان لارتباط الرتب يترجم بأمانة هذه الفكرة.

جدول (١ ـ ٢٤): حساب معامل الارتباط باستخدام القياسات نفسها

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	5	1	25	1	5
	10	6	100	36	60
	5	2	25	4	10
	11	8	121	64	88
	12	5	144	25	60
	4	1	16	1	4
	3	4	9	16	12
	2	6	4	36	12
	7	5	49	25	35
	1	2	1	4	2
المجموع	60	40	494	212	288

لنفرض الآن عينة من قيم x تتضمن n قياسا، فكيف نحدد رتب هذه القياسات؟ نكتب في عمود أول الأرقام المتسلسلة من 1 إلى n، وفي عمود مجاور نرتب قيم x من الأصغر إلى الأكبر، وفي عمود ثالث نكتب أمام كل قيمة لـx رتبة تساوي الرقم المتسلسل المقابل لها. ولكن إذا تكررت إحدى قيم x أكثر من مرة فهل نعطي القيمة نفسها رتبا مختلفة؟ وإذا بدا مثل هذا الأمر غير مقبول، وهو في الحقيقة كذلك، فكيف نتصرف؟ والجواب واضح بالبداهة، ففي مثل هذه الحالة نعتبر رتبة كل تكرار لتلك القيمة مساوية للمتوسط الحسابي للأرقام المتسلسلة المقابلة لها. فلنفرض، مثلا، أن الأرقام المتسلسلة 4، 5، 6، 7 في العمود الأول، قابلها في العمود الثاني 70، 70، 70، 70، فتكون الرتبة التي نعطيها لكل من هذه القياسات الأربعة المتساوية هي:

$$\frac{4+5+6+7}{4} = 5.5$$

مثال (١ ـ ٣٠)

لتكن مجموعة القياسات 4، 7، 8، 3، 12، 8، 4، 21، 35، 21، 15، 18، 17، 28، 17، 21 . والمطلوب ترتيب هذه القياسات وتحديد رتبة كل منها .

حدول (١ ـ ٢٥)

رتبة x	قيم x مرتبة	الرقم المتسلسل
1	3	1
2.5	4	2
2.5	4	3
4	7	4
5.5	8	5
5.5	8	6
7	12	7
8	15	8
9.5	17	9
9.5	17	10
11	18	11
13	21	12
13	21	13
13	21	14
15	28	15
16	35	16

وبالطريقة نفسها نرتب قيم y ، وكل رتبة لقيمة من قيم x يوافقها رتبة لقيمة y المقابلة. لنقارن الآن رتب قيم x برتب قيم y المقابلة لها. فلقد كتبنا رُتب x وفق التسلسل الطبيعي ومن الأصغر إلى الأكبر، فما هو الحال بالنسبة إلى تسلسل رُتب y؟ هل حققت ترتيبا موازيا، أي تسلسلا طبيعيا مطابقا لتسلسل رتب x أم طرأ فساد ما على التسلسل الطبيعي لرتب y؟ وما هي درجة أو مدى فساد التسلسل الطبيعي هذا؟ وسنقيس درجة أو مدى فساد التسلسل بالعدد Σd^2 ، حيث d هي رتبة x مطروحا منها رتبة y المقابلة .

وهكذا يمثل Σd^2 مجموع مربعات الفروق بين رتب x ورتب y المقابلة لها. ومن الواضح أن هـذا المقياس لـدرجة فسـاد التسلسل الطبيعي في رتب y سيكـون صفـرا إذا، وفقط إذا تطابقت رتب x مع رتب y المقابلة لها، وعندئذ نكون في حالة ارتباط إيجابي تام. وعندما يكون تسلسل رتب y الناتج بحيث يبدأ بالأكبر وينتهي بالأصغر، أي عكس التسلسل الطبيعي تماما، فإن Σd^2 سيكون أكبر ما يمكن. وهذه الحالة كما أسلفنا هي حالة ارتباط سلبي تام.

ونحتاج الآن إلى تعريف لمعامل ارتباط يعطي القيمة 1+ في الحالة الأولى، والقيمة 1- في الحالة الثانية، ويأخذ القيمة صفرا في حالة عدم وجود أي ارتباط. والمعامل الذي يواجه كل هذه المتطلبات، وسنرمز له بـ τ تمييزا له عن معامل بيرسون للارتباط، هو:

$$\tau = 1 - \frac{2\Sigma d^2}{\text{أكبر قيمة ممكنة لـ } \Sigma d^2}$$

فعندما يتطابق الترتيبان يكون $\Sigma d^2 = 0$ و $\tau = 1$، وعندما يتعاكس الترتيبان تماما يأخذ Σd^2 أكبر قيمة ممكنة له، ويكون:

$$\tau = 1 - \frac{2\,(\text{أكبر قيمة ممكنة لـ } \Sigma d^2)}{\text{أكبر قيمة ممكنة لـ } \Sigma d^2} = 1 - 2 = -1$$

ويمكن برهان أنه في حالة عدم وجود ارتباط يكون

$$2\Sigma d^2 = \text{أكبر قيمة ممكنة لـ } \Sigma d^2$$

أي $\tau = 0$.

وإذا كنا ندرس الارتباط في n من أزواج القياسات، فيمكن البرهان على أن أكبر قيمة ممكنة لـ Σd^2 هي $\frac{n(n^2-1)}{3}$، وبالتعـويض في العلاقة السابقـة نصل إلى معامل سبيرمان لارتباط الرتب، وهو:

$$\tau = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

مثال (١ ـ ٣١)

لتكن مجموعة الأزواج من القياسات :

x	4	4	7	7	7	9	16	17	21	25
y	8	16	8	8	16	20	12	15	25	20

احسب معامل سبيرمان لارتباط الرتب τ .

الحل

(١) نرتب قيم x ، ثم نرتب قيم y في جدولين منفصلين وفق الطريقة الموضحة في المثال (١ ـ ٣٠).

رتبة y	قيم y مرتبة	الرقم المتسلسل
2	8	1
2	8	2
2	8	3
4	12	4
5	15	5
6.5	16	6
6.5	16	7
8.5	20	8
8.5	20	9
10	25	10

رتبة x	قيم x مرتبة	الرقم المتسلسل
1.5	4	1
1.5	4	2
4	7	3
4	7	4
4	7	5
6	9	6
7	16	7
8	17	8
9	21	9
10	25	10

(٢) ننظم الآن جدولا يتضمن عموده الأول رتب x ، ويتضمن عموده الثاني رتب y المقابلة لها (التقابل بين قيم x وقيم y مبين في المثال). ويتضمن العمود الثالث الفرق d ، وهو يساوي الفرق بين رتبة x ورتبة y المقابلة لها. ومجموع هذا العمود يساوي الصفر، ويتضمن العمود الرابع مربعات الفروق d ، ومجموع هذا العمود هو Σd^2 .

رتبة x	رتبة y	d	d^2
1.5	2	-0.5	0.25
1.5	6.5	-5.0	25.0
4	2	2.0	4.0
4	2	2.0	4.0
4	1.5	2.5	6.25
6	8.5	2.5	6.25
7	4	3.0	9.00
8	5	3.0	9.00
9	10	-1.0	1.00
10	8.5	11.5	2.25
المجموع		0	67.00

(٣) نعوض الآن في العلاقة :

$$\tau = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

حيث $\Sigma d^2 = 67$ ، $n = 10$ فنجد :

$$\tau = 1 - \frac{6 \times 67}{10(100 - 1)} = 0.594$$

تمارين (١ ـ ٥)

١) احسب معامل سبيرمان للارتباط في البيان الإحصائي التالي بعد أن ترسم مخطط الانتشار.

x	y	x	y	x	y	x	y
22	18	35	47	19	37	8	18

تابع :

x	y	x	y	x	y	x	y
15	16	46	22	36	42	1	3
9	31	16	25	25	20	9	7
7	8	7	36	10	12	18	28
4	2	6	27	11	17	46	21
45	36	46	45	5	6	9	25
19	12	11	18	26	45		
26	16	27	18	19	30		

٢) لدى مدرس قناعة بأن قائمة من أسئلة «الخطأ والصواب» ستعطيه من المعلومات عن كفاءة الطلاب في مادته، مثل ما تعطيه مجموعة من الأسئلة تتضمن تمارين ومناقشة. ولكي يثبت وجهة نظره، أعد للطلاب امتحانا يتضمن 25 سؤالا من نوع «الخطأ والصواب»، وما تبقى من الامتحان كان تمارين وأسئلة مناقشة. وقسم العلامة التامة، وهي 200، إلى 50 للقسم الأول (خ، ص)، و 150 للقسم الثاني. وفيما يلي درجات طلابه الثلاثين في كل من القسمين، هل تجد معامل ارتباط مرتفع بين المجموعتين من الدرجات؟ وماذا تستنتج؟ ارسم مخطط الانتشار.

(خ، ص)	تمارين	(خ، ص)	تمارين	(خ، ص)	تمارين	(خ، ص)	تمارين
24	150	21	118	23	125	22	135
23	170	19	110	12	102	14	78
24	141	21	129	15	94	15	105
13	84	25	145	16	91	25	141
19	123	16	124	20	127	19	105
17	100	19	108	21	120	17	110
14	92	18	112	16	105		
18	105	16	98	25	149		

٣) فيما يلي قياس الحذاء x، والوزن بالباوند y، لكل من عشرة طلاب جامعيين :

x قياس الحذاء	9.5	9.5	10.5	10.5	11	8.5	8.5	9.5	10	9
y الوزن	140	155	153	150	180	160	155	145	163	150

أ ـ احسب معامل بيرسون للارتباط ،

ب ـ احسب معامل سبيرمان لارتباط الرتب .

٤) فيما يلي سبعة أزواج من القياسات :

x	10	20	30	40	50	60	70
y	-4	-3	-2	0	3	6	7

ارسم مخطط الانتشار ثم احسب معامل الارتباط .

٥) فيما يلي طول الأم بالبوصة ، x ، وطول ابنتها بالبوصة ، y :

x طول الأم	67	64	62	65	69	63	65	66
y طول الابنة	70	69	65	68	66	60	64	66

ارسم مخطط الانتشار واحسب معامل الارتباط بطريقتي بيرسون وسبيرمان .

٦) سجلنا لعشرة عمال طباعة كلا من معدل إنتاجه في الساعة من الوحدات الجيدة ،x؛ ومعدل إنتاجه في الساعة من الوحدات المعيبة ، y ، فوجدنا ما يلي :

x	94	98	106	114	107	93	98	88	103	95
y	4	5	6	7	6	5	6	4	7	5

احسب معامل الارتباط بين x و y.

٧) في معرض فني يتضمن ثلاثين لوحة رتب محكمان اللوحات حسبما يراه عن درجة نجاحها وأعطى كل منهما الرتبة 1 لأفضل لوحة ، و 2 لتلك التي تليها في الأفضلية حسب رأيه ، وهكذا حتى وصلا إلى 30 لأردأ لوحة كل في رأيه . وفيما يلي الرتب التي

أعطاهـا المحكمان لكل من اللـوحات الثلاثين. احسـب معامل سبيرمـان لارتباط الرتب. ماذا تستنتج؟

رتبة المحكم الأول	1	2	3	4	5	6	7	8	9	10	11
رتبة المحكم الثاني	2	4	3	1	5	7	10	17	8	9	14
رتبة المحكم الأول	12	13	14	15	16	17	18	19	20	21	22
رتبة المحكم الثاني	6	15	11	13	12	18	19	21	16	23	30
رتبة المحكم الأول	23	24	25	26	27	28	29	30			
رتبة المحكم الثاني	29	20	22	25	24	28	26	27			

٨) فيما يلي درجة مادة الـرياضيات x ودرجة مـادة العلوم y لكل من عشرة طـلاب في المرحلة الثانوية :

x	90	95	70	70	65	65	65	40	55	60
y	97	97	85	65	70	70	60	55	40	70

أ ـ ارسم مخطط الانتشار.

ب ـ احسب معامل بيرسون للارتباط بين x و y.

جـ ـ احسب معامل سبيرمان لارتباط الرتب بين x و y .

٩)* فيما يلي تطـور انتـاج القمح x في المملكـة بآلاف الأطنـان وتطـور مجموع القـروض الزراعية الممنوحة y ، بملايين الريالات، وذلك بين عامى ١٣٩١هـ و ١٤٠٣هـ.

إنتاج القمح x	42	39	64	153	132	93	125	120	150
مجموع القروض الزراعية y	16.6	16.6	19.6	36.3	145.5	269.4	489.9	585.6	709.1
إنتاج القمح x	142	187	412	741					
مجموع القروض الزراعية y	1128.6	2530.8	2932.9	4166..0					

احسب معامل الارتباط بين x و y .

١٠) يعطي البيان التالي معدلات ما قبل الحرب لامدادات الطعام الصافية x ، ومعدلات وفيات الأطفال y في عدد مختار من الدول :

* مأخوذة من منجزات خطط التنمية الصادر عن وزارة التخطيط في المملكة. ص ٢٠٩ وص ٢١٣ .

y	x	البلاد	y	x	البلاد	معدل وفيات الأطفال لكل 1000 (y)	عدد الحريرات اليومية للشخص الواحد x	البلاد
32.2	3260	نيوزيلاند	64.2	3420	الدانمرك	98.8	2730	الأرجنتين
40.5	3160	النرويج	162.9	2450	مصر	39.1	3300	استراليا
37.4	3010	هولندا	66.1	2880	فرنسا	87.4	2990	النمسا
139.4	2710	بولونيا	63.3	2960	ألمانيا	83.1	3000	بلجيكا
43.3	3210	السويد	113.4	2600	اليونان	202.1	2080	بورما
45.3	3110	سويسرا	42.4	3160	آيسلند	67.4	3070	كندا
55.3	3100	المملكة المتحدة	161.6	1970	الهند	182.8	1920	سيلان
53.2	3150	الولايات المتحدة	69.6	3390	إيرلندا	240.8	2240	شيلي
94.1	2380	أورغواي	102.7	2510	إيطاليا	155.6	1860	كولومبيا
			60.6	2180	اليابان	116.8	2610	كوريا

ارسم مخطط الانتشار واحسب معامل الارتباط بين عدد الحريرات اليومية للشخص الواحد (x)، وبين y معدل وفيات الأطفال لكل 1000 .

١١) في التمرين ١٨ من مجموعة التمارين (١ ـ ١) . معتبرا عدد الأسرة x وعدد الأطباء y . ارسم مخطط الانتشار. واحسب معامل الارتباط بين x و y .

١٢) يتضمن البيان التالي معدل استهلاك الكحول السنوي باللتر للشخص الواحد ممن تزيد أعمارهم عن الرابعة عشرة، x ، ومعدل الوفاة لكل مائة ألف من السكان بمرض تشمع الكبد أو الإدمان، y ، وذلك في مختارات من الدول . ارسم مخطط الانتشار لإيضاح وجود رابطة بين المتغيرين x و y ، ثم احسب معامل الارتباط بينهما .

معدل الوفاة لكل 10^5 من السكان بسبب تشمع الكبد أو الإدمان (y)	معدل استهلاك الكحول السنوي باللتر (x)	البلاد
46.1	24.7	فرنسا
23.6	15.2	إيطاليا
23.7	12.3	ألمانيا الغربية
7.0	10.9	استراليا
12.3	10.8	بلجيكا
14.2	9.9	الولايات المتحدة
7.4	8.3	كندا
3.0	7.2	إنكلترة وويلز
7.2	6.6	السويد
10.6	5.8	اليابان
3.7	5.7	هولندا
3.4	5.6	إيرلندا
4.3	4.3	النرويج
3.6	3.9	فنلندا