الفصل الثاني

التوزيعات الوصفية للقياسات

٢ ـ ١ مقدمة : كيف يمكن وصف مجموعة من القياسات ، سواء أكانت قياسات عينة أو مجتمع ؟ واذا كان المجتمع المدروس أمامنا فكيف يمكننا وصف هذه المجموعة الضخمة من القياسات ؟

هناك كتب عديدة مكرسة لطرق الإحصاء الوصفي ، أي طرق وصف مجموعات من المعلومات الرقمية . ويمكن تصنيف هذه الطرق إلى طرق بيانية وطرق عددية . وسنقتصر هنا على مناقشة القليل من الطرق البيانية والعددية المفيدة ليس فقط من أجل الإحصاء الوصفي ، ولكن أيضاً من أجل الاستقراء الإحصائي .

٢ ـ ٢ طريقة بيانية : يمثل البيان الاحصائي في الجدول (٢ ـ ١) معدلات ثلاثين من طلبة السنة الجامعية الأولى . وتبين نظرة سريعة للجدول أن أصغر معدل هو 46.7 وأعلى معدل هو 96.7 . فكيف تتوزع القياسات الثماني والعشرين الباقية ؟ هل تقع قرب 46.7 أو قرب 96.7 أم أنها تتوزع بصورة عادلة فوق

جدول (٢ ـ ١) معدلات ثلاثين من طلبة العام الجامعي الأول

46.7	66.7	46.7	86.7	50.0
53.3	66.7	86.7	70.0	60.0
63.3	66.7	73.3	83.3	80.0
96.7	83.3	63.3	66.7	73.3
80.0	73.3	66.7	76.7	70.0
53.3	73.3	56.7	76.7	73.3

المجال [46.7, 96.7] ؟ وللإجابة على هذا السؤال نقسم هذا المجال الى عدد من المجالات الجزئية المتساوية يتوقف عددها على عدد القياسات التي نصنفها . (وكقاعدة أولية ، يكون عدد المجالات الجزئية بين 5 و 20 ، ويتناسب العدد مع عدد القياسات التي نصنفها) . وعلى سبيل المثال يمكن استخدام المجالات الجزئية (46.65-52.90) ، (52.90-59.15) ، (59.15-65.40) ، (65.4-71.65) ، الخ . ونلاحظ أننا اخترنا نقاط التقسيم بحيث لا يقع أي قياس فوق نقطة تقسيم . ونصنف الآن القياسات الثلاثين كلاً في المجال الجزئي الذي يحويه . ويبين الجدول (٢ ـ ٢) مثل هذا التصنيف . وقد وضعنا في العمود الأول رقم المجال الجزئي (أو الصف كما يدعى عادة في الاحصاء الوصفي) وفي العمود الثاني حدود كل صف وفي العمود الثالث يقابل كل صف عدد من الخطوط يساوي عدد القياسات التي تقع ضمن هذا الصف ، ولسهولة التعداد نبرز كل مجموعة من خمسة خطوط على حدة وبحيث يقطع الخط الخامس الخطوط الأربعة السابقة له . ويسمى عدد القياسات الواقعة ضمن الصف i بتواتر هذا الصف ونرمز له عادة بِ f_i ، وهذه التواترات موجودة في مثالنا في العمود الرابع من الجدول (٢ ـ ٢) ، أما العمود الخامس والأخير فيبين نسبة القياسات الواقعة ضمن كل صف إلى العدد الكلي للقياسات التي نصنفها ، وتسمى مثل هذه النسب عادة بالتواترات النسبية . واذا رمزنا لعدد القياسات الكلي بِ n ، مثلاً ، (وفي مثالنا $n = 30$) يكون التواتر النسبي للصف i مساوياً $\frac{f_i}{n}$

جدول (٢ ـ ٢) تصنيف المعلومات الواردة في الجدول (٢ ـ ١) .

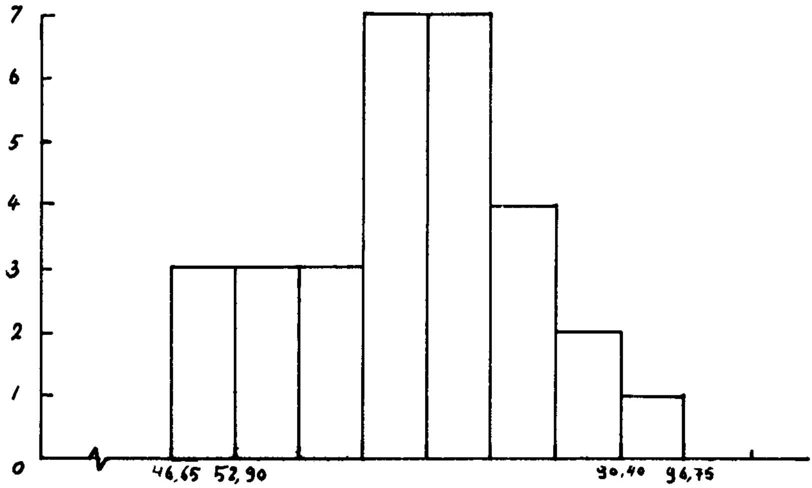
التكرار النسبي	التكرار f_i	التعداد	حدود الصفوف	الصف i
3/30	3	III	46.65-52.90	1
3/30	3	III	52.90-59.15	2
3/30	3	III	59.15-65.40	3
7/30	7	~~IIII~~II	65.40-71.65	4
7/30	7	~~IIII~~II	71.65-77.90	5
4/30	4	IIII	77.90-84.15	6
2/30	2	II	84.15-90.40	7
1/30	1	I	90.40-96.75	8
1	$n = 30$			المجموع

ويمكن تمثيل الجدول الناتج بيانياً على شكل مضلع تكراري (شكل ٢ ـ ١) . حيث نقيم فوق كل مجال جزئي مستطيلاً ارتفاعه متناسب مع عدد القياسات (تواتر الصف) الواقعة ضمن هذا المجال . ونظرة سريعة على المضلع التكراري كافية لكي تعطينا فكرة أولية واضحة عن كيفية توزع المعدلات فوق المجالات الجزئية .

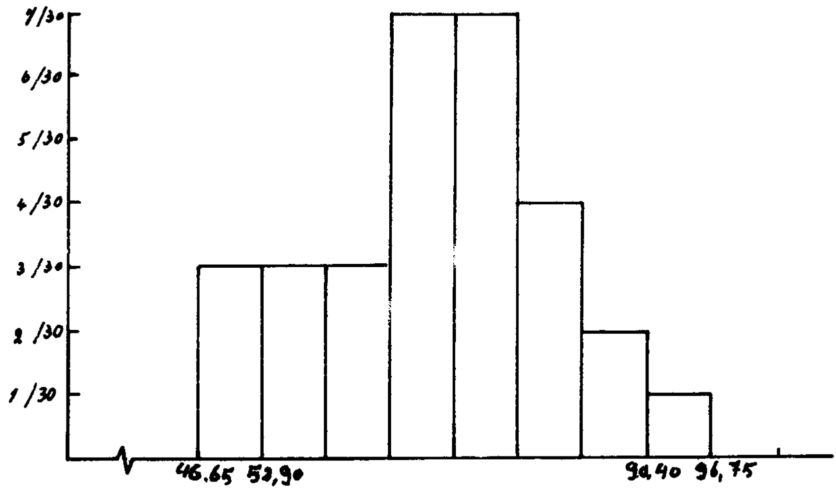
وسيكون من الأنسب على الغالب تعديل المضلع التكراري بحيث يكون ارتفاع المستطيل الذي ننشئه فوق كل مجال جزئي مساوياً للتواتر النسبي الموافق لهذا المجال الجزئي . شكل ٢ ـ ٢) . ونادراً ما يفرق الإحصائي بين مضلع التكرار ومضلع التكرار النسبي ، اذ يبقى الشكل نفسه وتتغير فقط وحدة القياس على المحور الشاقولي .

ومع أن اهتمامنا المباشر ينصب على وصف مجموعة القياسات الثلاثين

الا أننا نهتم أكثر بالمجتمع الذي أخذنا منه العينة . فيمكن النظر الى المعدلات الثلاثين كعينه مأخوذة من مجتمع معدلات طلبة السنة الأولى في جامعة أو عدد من الجامعات التي تخضع لنفس الأنظمة . وفي جميع الأحوال لو توفرت لنا معدلات هذا المجتمع كله لأمكن ، بنفس الطريقة ، إقامة مضلع التكرار النسبي للمجتمع .



شكل (٢ ـ ١) المضلع التكراري



شكل (٢ ـ ٢) مضلع التكرار النسبي

لنعتبر مضلع التكرار النسبي للعينة بتفصيل اكبر . فما هي نسبة الطلاب الذين نالوا معدلاً يساوي 71.7 أو أكثر ؟ وبالعودة الى مضلع التكرار النسبي نرى أن هذه النسبة تتضمن كل الصفوف إلى يمين 71.65 . وبالاستنادة من

الجدول (٢ ـ ٢) نرى أن أربعة عشر من الطلاب لديهم معدلات أكبر أو تساوي 71.65 . أي أن النسبة المطلوبة هي 14/30 أو 47% تقريبا . ونلاحظ أن هذه النسبة هي أيضاً النسبة المئوية للمساحة من المضلع التكراري في الشكل (٢ ـ ٢) التي تقع الى يمين 71.65 . ولنفرض أننا كتبنا كل معدّل على قطعة من الورق . ثم وضعنا قطع الورق الثلاثين في قبعة ، خلطناها بشكل جيد ، ثم سحبنا ورقة من القبعة فما هي امكانية أن تحوي هذه الورقة معدلاً أكبر أو يساوي 71.7 ؟ وبما أن أربعة عشر من بين الأوراق الثلاثين تحوي أعداداً أكبر أو تساوي 71.7 فنقول أن لدينا 14 فرصة من أصل 30. ويمكن القول في هذه الحالة ان احتمال الحصول على مثل هذه النسبة هو 14/30 . ولا بد أن القارئ قد واجه كلمة « محتمل » أو « احتمال » في مناقشاته اليومية العادية ، وسنؤجل تعريف الاحتمال ومناقشة أهميته حتى الفصل القادم .

لنوجه انتباهنا الآن إلى المجتمع الذي سحبنا منه العينة . فما هي نسبة الطلاب في المجتمع الذين نالوا معدلاً أكبر من 71.7 ؟ ولو توفر لنا مضلع التكرار النسبي الخاص بالمجتمع لأمكننا اعطاء جواب دقيق لهذا السؤال بحساب نسبة المساحة الكلية من المضلع الواقعة إلى يمين 71.65 . ولكن عدم توفر مثل هذا المضلع يضطرنا الى القيام باستقراء . ولا بد من تقدير النسبة الحقيقية من المجتمع ، على أساس المعلومات التي تحويها العينة . ومن المعقول أن يكون تقديرنا هنا هو 14/30 أو 47% . ولنفرض الآن أننا نرغب في التعبير عن فرصة أو احتمال أن يكون لطالب انتقيناه بالصدفة من المجتمع معدل أكبر أو يساوي 71.7 . فبدون معرفة مضلع التكرار النسبي للمجتمع ، يمكن أن نستقرئ أن مضلع المجتمع مشابه لمضلع العينة وأن 14/30 من قياسات المجتمع ، على وجه التقريب ،

قد تكون أكبر أو تساوي 71.7 . ومن الطبيعي أن يكون هذا التقدير مخطئاً . وسندرس في فصل قادم كيف نضع حدوداً لمثل هذا الخطأ في التقدير .

وغالباً ما يدعى مضلع التكرار النسبي التوزيع التكراري لأنه يبين الطريقة التي تتوزع فيها المعلومات على طول محور الفواصل . ونلاحظ أن المستطيلات المقامة فوق كل صف خاضعة لتفسيرين . فهي تمثل نسبة الملاحظات الواقعة في صف معين . وأيضاً ، اذا سحبنا، عشوائياً ، قياساً من القياسات فيمكن اعتبار مساحة المستطيل المقام فوق صف معين تمثيلاً لاحتمال وقوع ذلك القياس ضمن هذا الصف . والناحية الأكثر أهمية في مضلع تكرار العينة هو أنه يمدنا بمعلومات حول المضلع التكراري للمجتمع ، ونتوقع أن يكون مضلعا التكرار للعينة والمجتمع متشابهين وتزداد درجة التشابه كلما ازداد حجم العينة ، وعندما تزداد العينة بحيث تتطابق مع المجتمع يتطابق المضلعان التكراريان .

٢ ـ ٣ الطرق الوصفية الرقمية : الطرق البيانية مفيدة للغاية عند تقديم المعلومات الاحصائية ومن أجل نقل وصف عام وسريع للمعلومات الاحصائية . وهذا يؤكد المثل القائل بأن صورة واحدة تساوي ألف كلمة . وتوجد ، على أي حال ، حدود لاستخدام الطرق البيانية من أجل وصف وتحليل المعلومات . وعلى سبيل المثال ، لنفرض أننا نرغب في مناقشة البيان الاحصائي أمام مجموعة من الناس وأنه ليس لدينا طريقة أخرى غير الطريقة الشفهية مما يجعل عرض المضلع التكراري غير متوفر ويضطرنا لاستخدام مقاييس وصفية أخرى قد تنقل إلى المستمعين صورة ذهنية عن المضلع التكراري . والأمر الثاني الذي يضع حداً لاستخدام الطرق البيانية هذه هو صعوبة الاستفادة منها في مجال الاستقراء الاحصائي . وربما اقتصرت فوائدها الاستقرائية على أن يقدم المضلع التكراري للعينة تصوراً عن شكل المضلع التكراري للمجتمع . واستقراؤنا يقف عند الفرض بأنه يوجد تشابه ما بين المضلعين ولكن المشكلة التي نقف عندها هي كيفية قياس مدى الاختلاف بينهما أو بصورة إيجابية قياس درجة التشابه بينهما .

وهكذا نفضل وجود مقاييس وصفية أخرى يمكن استخدامها للتنبؤ بشكل توزيع التواتر الخاص بالمجتمع المدروس . وهذه المقاييس هي مقاييس عددية وبالتالي يمكن استخدامها بسهولة في مجالات التحليل الاحصائي . أي أنه يمكن التغلب على الصعوبة التي تخلقها الآفاق المحدودة للطرق الوصفية بتجاوزها إلى مقاييس وصفية عددية . وهكذا نستخدم معلومات العينة لحساب مجموعة من الأرقام تقدم للاحصائي صورة ذهنية جيدة عن توزيع التواتر وتكون مفيدة للقيام باستقراءات تتعلق بالمجتمع .

٢ ـ ٣ قياسات النزعة المركزية : في طليعة المقاييس الوصفية الهامة تأتي مقاييس النزعة المركزية أي القياس المعبر عن موضع تمركز التوزيع . ونلاحظ في المثال المعطى في الفقرة السابقة أن المعدلات الثلاثين تمتد ما بين 46.7 و 96.9 ، ويتوضع مركز المضلع التكراري في جوار 70.0 . وسنستعرض الآن عدداً من القواعد المحددة لتعيين مركز توزيع مجموعة من المعلومات الاحصائية .

واحدى المقاييس المفيدة الأكثر استخداما للنزعة المركزية لمجموعة من القياسات هي معدّلها الوسطي الحسابي . وهو ما يشار إليه غالبا بالمتوسط الحسابي أو المتوسط لمجموعة من القياسات .

تعريف : المتوسط الحسابي لمجموعة n من القياسات $x_1, x_2, \ldots, x_n$ يساوي مجموع هذه القياسات مقسوماً على n .

ولنتذكر دائماً أننا نهتم بكل من العينة والمجتمع ، ولكل منهما متوسطه الحسابي . ولكي نميز بين المتوسطين فسنرمز بـ $\bar{x}$ لمتوسط العينة ولمتوسط المجتمع بالحرف اليوناني μ (ميو) . ويمكن حساب متوسط العينة وفقا للعلاقة :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (١)$$

مثال ٢ ـ ١ : احسب متوسط مجموعة القياسات 2, 9, 11, 5, 6

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+9+11+5+6}{5} = 6.6$$

والأهم من ذلك هو أن $\bar{x}$ ستُستخدم كتقدير لمتوسط المجتمع μ الذي لا نعرفه . وعلى سبيل المثال فإن متوسط المعدلات المذكورة في الجدول (٢ ـ ١) هو 69.45 . ومتوسط مجتمع كل المعدلات μ غير معروف . واذا أردنا تقديره فمن الممكن أن يكون تقديرنا مساوياً لِ 69.45 وعندها نعتبر $\mu = 69.45$.

والمقياس الثاني للنزعة المركزية هو الوسط .

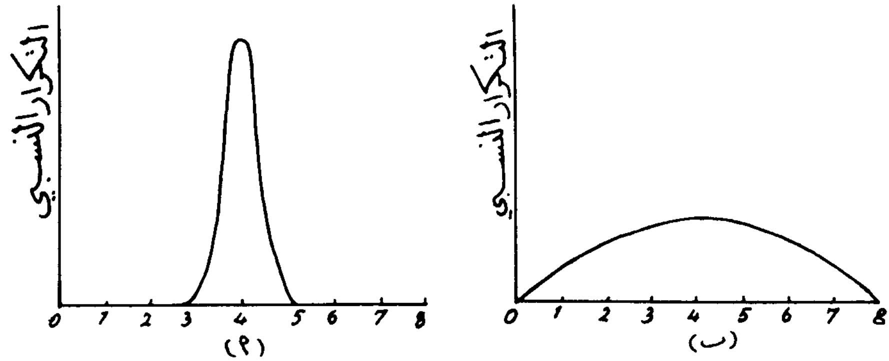
تعريف : نعرف وسط مجموعة n من القياسات $x_1, x_2, \ldots, x_n$ بأنه القيمة الواقعة في الوسط بعد ترتيب القياسات تصاعدياً وفقاً لقيمها .

مثال ٢ ـ ٢ لتكن مجموعة القياسات الخمسة 14 ,11 ,7 ,2 ,9 . فبعد ترتيبها وفقاً لقيمها المتصاعدة نجد 14 ,11 ,9 ,7 ,2 وهكذا يكون الوسط هو القياس 9 . واذا كان عدد القياسات زوجياً فإن الوسط هو معدل القياسين المتوسطين .

مثال ٢ـ٣ لتكن مجموعة القياسات 6 ,14 ,11 ,7 ,2 ,9 فبعد ترتيبها من الأصغر إلى الأكبر نجد 14 ,11 ,9 ,7 ,6 ,2 والوسط هو في منتصف الطريق بين 7 و9 أي

$$\frac{7+9}{2} = 8$$

٢ ـ ٥ قياسات التشتت : بعد أن حددنا تمركز توزيع معلومات إحصائية نقدم الآن قياساً لتباين أو تشتت المعلومات . لنأخذ التوزيعين المبينين في الشكل (٢ ـ ٣) فكلاهما يتمركز حول النقطة $x = 4$ إلا أنه يوجد فرق كبير في تباعد القياسات عن المتوسط في التوزيعين . وبينما تتغير في ٢ ـ ٣(ا) بين 3 و 5 فإنها تتغير في ٢ ـ ٣ (ب) بين الصفر و 8 . ولخاصة التغير هذه أهمية كبيرة في البيان الاحصائي أو المعلومات الإحصائية . وبالاضافة الى أهميته العملية فهو



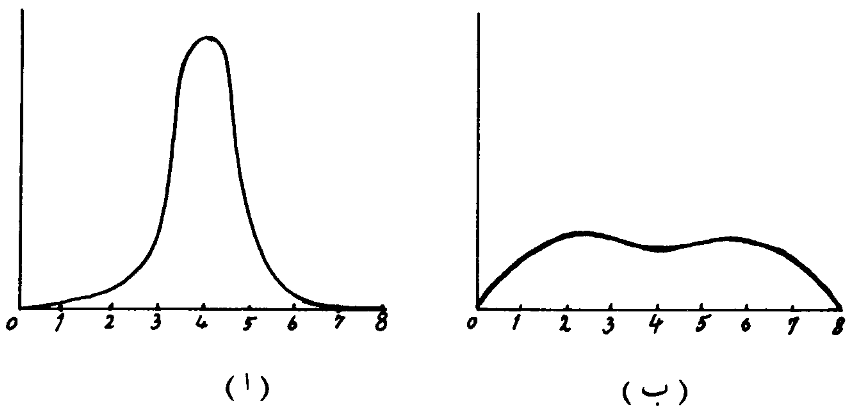
شكل (٢ ـ ٣) تشتت المعلومات الإحصائية .

يشكل قياساً ضرورياً إلى جانب قياس النزعة المركزية لبناء صورة ذهنية لتوزيع التواتر . وتوجد قياسات عديدة للتشتت سنناقش هنا أهمها . ونبدأ بأبسطها وهو المدى .

تعريف : نعرف مدى مجموعة n من القياسات $x_1, x_2, \ldots, x_n$ بأنه الفرق بين أكبر قياس وأصغر قياس .

وفي الجدول (٢ ـ ١) نلاحظ ان القياسات تتغير بين 46.7 و 96.7 وبالتالي يكون المدى $96.7 - 46.7 = 50$.

ولكن المدى ليس قياساً مرضياً تماماً للتشتت . فلنتأمل التوزيعين في الشكل

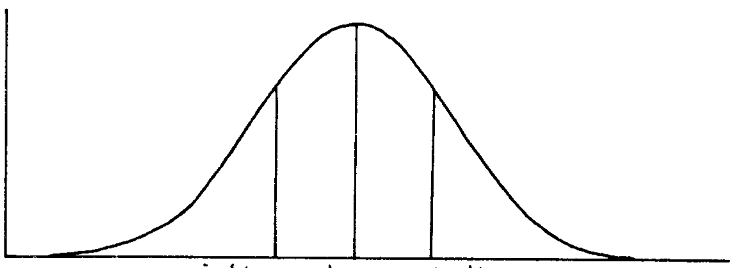


الشكل (٢ ـ ٤) توزيعان لهما نفس المدى ولكن بتشتتين مختلفين

(٢ ــ ٤) فنجد أن كلاهما له نفس المدى إلا أنه من الواضح أن تشتت الشكل (ب) أكبر من تشتت الشكل (أ) . وللتغلب على هذه الصعوبة نلجأ إلى الأرباع والنسب المئوية . ولنتذكر أننا اذا حددنا مجالاً على طول المحور ox في المضلع التكراري فإن النسبة المئوية للمساحة تحت المضلع التكراري وفوق المجال الجزئي يساوي الى النسبة المئوية لعدد القياسات التي تقع في ذلك المجال الجزئي . ومن تعريف الوسط نجد بوضوح أن نصف مساحة المضلع تقع إلى يمينه والنصف الآخر إلى يساره . وبصورة مشابهة يمكن تعريف الأرباع بأنها القياسات التي تقسم مساحة المضلع التكراري إلى أربعة أرباع .

تعريف : لتكن $x_1, x_2, ..., x_n$ مجموعة n من القياسات المرتبة وفقاً لقيمها المتصاعدة . فالربع الأدنى هو القياس الذي يتجاوز في قيمته $\frac{1}{4}$ القياسات وأقل من الـ $\frac{3}{4}$ الباقية من القياسات والربع الثاني هو الوسط أما الربع الأعلى فهو القياس الذي تقع $\frac{3}{4}$ من القياسات إلى يساره والربع الباقي إلى يمينه .

ونلاحظ في الشكل (٢ ــ ٥) أن ربع المساحة يقع على يسار الربع الأدنى و $\frac{3}{4}$ المساحة على يمينه . أما الربع الأعلى فهو قيمة x التي تقع $\frac{3}{4}$ المساحة على يساره و $\frac{1}{4}$ المساحة على يمينه .



شكل (٢ ــ ٥) مواقع الأرباع

وفي بعض التطبيقات نفضل استخدام المئويات .

تعريف : لتكن $x_1, x_2, ..., x_n$ مجموعة n من القياسات المرتبة وفقاً

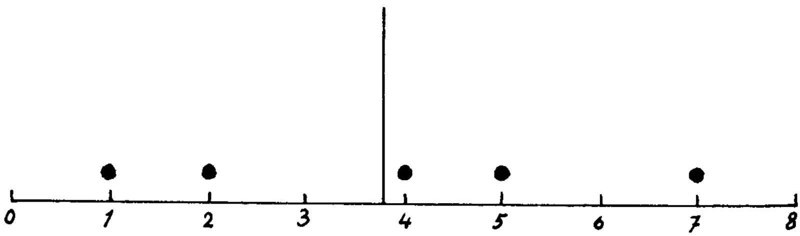
لقيمها المتصاعدة فيكون المئوي p هو بالتعريف قيمة x التي يقع p% من القياسات على يسارها (أي أقل منها) و (1-p)% على يمينها .

وعلى سبيل المثال ، فإن المئوي 90 لمجموعة من المعلومات الإحصائية هو قيمة x التي تتجاوز 90% من القياسات في هذه المجموعة وأقل من الـ 10% الباقية . وكما في حالة الأرباع فإن 90% من مساحة المضلع التكراري تقع على يسار المئوي 90 .

وتتمثل بساطة المدى في أنه يمكن التعبير عنه بعدد واحد . الا أن الأرباع والمئويات ، وهي أكثر فعالية في التعبير عن التشتت ، نحتاج فيها إلى أكثر من عدد واحد لوصف التشتت بصورة مناسبة . فهل يمكن ايجاد قياس للتشتت يمكن التعبير عنه بعدد واحد وأكثر حساسية من المدى ؟

لنأخذ كمثال مجموعة القياسات 4 ,2 ,1 ,7 ,5. فيمكن تمثيل هذه القياسات بيانياً ، كما في الشكل (٢ ـ ٦) ، بوضع نقاط من أجل القياسات على طول محور الفواصل . وبحساب المتوسط نجد : $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{19}{5} = 3.8$

ونمثل المتوسط كنقطة على محور الفواصل . ويمكن النظر الى التشتت الآن من خلال المسافة بين كل من النقاط (القياسات) والمتوسط $\bar{x}$. واذا كانت المسافات كبيرة ، يمكن القول ان المعلومات الاحصائية أكثر تبايناً مما لو كانت المسافات أصغر . ونعرف انحراف قياس x_i بأنه المسافة التي تفصل بين هذا



شكل (٢ ـ ٦) التمثيل البياني النقطي

القياس والمتوسط أي $(x_i - \bar{x})$. ونلاحظ أن انحراف القياسات الواقعة على يمين المتوسط موجبة بينما انحرافات تلك الواقعة على يسار المتوسط سالبة . ويبين الجدول (٢ ـ ٣) الانحرافات في مثالنا السابق .

جدول (٢ ـ ٣) حساب
$\sum_{i=1}^{n} (x_i - \bar{x})^2$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
5	1.2	1.44	25
7	3.2	10.24	49
1	-2.8	7.84	1
2	-1.8	3.24	4
4	.2	.04	16
$\sum_{i=1}^{5} x_i = 19$	0	22.80	95

واذا اتفقنا الآن على أن الانحرافات تحوي معلومات عن التشتت ، فإن الخطوة التالية هي وضع علاقة مبنية على الانحرافات وتمدنا بمقياس جيد للتشتت . وكإمكانية أولى نختار معدل الانحرافات . ولكن من سوء الحظ نجد أن بعض الانحرافات هو دوماً موجب وبعضها الآخر سالب بحيث يكون مجموعها دوماً هو الصفر . اذ لو فرضنا n من القياسات $x_1, x_2, \ldots, x_n$ فنجد :

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

$$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0 \quad .$$

وسيلاحظ القارئ حلاً سهلاً لهذه المشكلة ، فلماذا لا نحسب معدل القيم المطلقة للانحرافات؟ وتُستخدم هذه الطريقة ، في الحقيقة . لقياس التشتت الا أنها ليست مرضية في مجالات الاستقراء الاحصائي نظراً لصعوبة معالجة المقدار الناتج تحليلياً .

تعريف : لتكن $x_1, x_2, \ldots, x_n$ مجموعة من n من القياسات فنعرف الانحراف المتوسط بأنه الكمية :

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

أي معدل القيم المطلقة لانحرافات هذه القياسات عن متوسطها .

ونفضل التغلب على هذه الصعوبة باللجوء الى مربعات الانحرافات بدلاً من قيمها المطلقة .

تعريف : نعرف تشتت مجموعة من n من القياسات $x_1, x_2, \ldots, x_n$ بأنه معدل مربعات انحرافات هذه القياسات عن متوسطها . ونرمز له بِ S'^2 أي أن :

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

وعلى سبيل المثال ، يمكن حساب التشتت لمجموعة القياسات الخمسة المعطاة في الجدول (٢ ـ ٢) فنجد :

$$S'^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2}{5} = \frac{22.80}{5} = 4.56 .$$

ونستخدم S'^2 للدلالة على تشتت عينة . بينما نرمز بِ σ^2 (σ هو الحرف اليوناني سيجما) لتشتت المجتمع .

تعريف : نعرف الانحراف المعياري لمجموعة n من القياسات $x_1, x_2, \ldots, x_n$ بأنه يساوي الجذر التربيعي الموجب للتشتت .

ونقيس التشتت بدلالة مربع الوحدات الأصلية للقياسات . أي أنه اذا كانت وحدات القياس الأصلية بالسنتمتر فإنه يمكن التعبير عن التشتت بالسنتمتر المربع . وبأخذ الجذر التربيعي للتشتت نحصل على الانحراف المعياري بنفس الوحدات الأصلية المستخدمة في القياس . والانحراف المعياري للعينة هو

$$S' = \sqrt{s'^2} = \sqrt{\frac{\sum_{1}^{n}(x_i - \bar{x})^2}{n}} \quad (4)$$

والانحراف المعياري للمجتمع هو $\sigma = \sqrt{\sigma^2}$

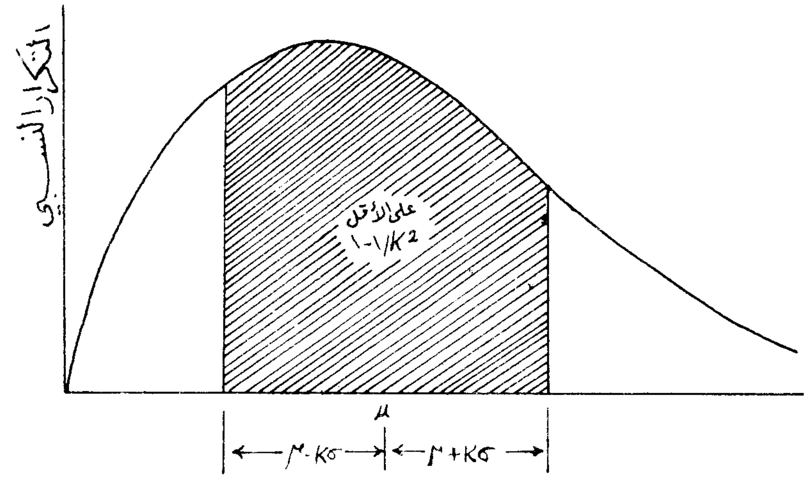
٢ ـ ٦ حول الأهمية العملية للانحراف المعياري :

نقدم الآن نظرية هامة ومفيدة للرياضي الروسي تشيبيشيف وذلك على الشكل المبسط التالي ، ونقبلها بدون برهان .

نظرية تشيبيشيف : ليكن k عدداً أكبر من أو يساوي الواحد ، ومجموعة من n من القياسات $x_1, x_2, \ldots, x_n$. فعلى الأقل يقع $(1 - \frac{1}{k^2})$ من هذه القياسات ضمن k انحرافاً معيارياً على يمين ويسار متوسطها أي ضمن المجال $(\bar{x} - ks')$ و $(\bar{x} + ks')$ حيث x هو المتوسط و s' الانحراف المعياري .

ويوضح الشكل (٢ ـ ٧) الفكرة التي تتضمنها النظرية . حيث يحدد المجال المقام حول متوسط التوزيع μ والذي يمتد بمقدار $k\sigma$ إلى يمين ويسار μ أي المجال $(\mu - k\sigma, \mu + k\sigma)$ ، ويحدد مساحة تساوي على الأقل $1 \frac{1}{k}$ من المساحة الكلية تحت المنحني .

لنختر الآن بعض القيم لِ k ولنحسب النسبة $1 - \frac{1}{k^2}$ (أنظر الجدول ٢ ـ ٤) .



شكل ٢ ـ ٧ توضيح نظرية تشيبيشيف

فعندما تكون $k = 1$ لا تقدم النظرية أية معلومات . ولكن من أجل $k = 2$ يمكن القول أن $\frac{3}{4} = 1 - \frac{1}{2^2}$ القياسات على الأقل واقع ضمن المجال $(\mu - 2\sigma , \mu + 2\sigma)$. وسيقع $8/9$ من القياسات ، على الأقل. ضمن المجال $(\mu - 3\sigma , \mu + 3\sigma)$.

جدول (٢ ـ ٤) قيم توضيحية لِ $1 - \frac{1}{k^2}$

K	$1 - 1/k^2$
1	0
2	3/4
3	8/9

لنعتبر الآن مثالاً نستخدم فيه المتوسط والانحراف المعياري (أو التشتت) لإقامة صورة ذهنية عن توزع مجموعة من القياسات .

مثال ٢ ـ ٤ نعلم أن متوسط وتشتت عينة من $n = 25$ قياساً هما 75 و 100 على الترتيب . والمطلوب استخدام نظرية تشيبيشيف لوصف توزيع القياسات .

الانحراف المعياري هو $S' = \sqrt{100} = 10$. ويتمركز توزيع القياسات

حول $\bar{x} = 75$. ومن نظرية تشيبيشيف نجد أن :

(١) ثلاثة أرباع القياسات الـ 25 ، على الأقل ، واقع ضمن المجال $\bar{x} \pm 2s' = 75 \pm 20$ أي بين 55 و 95 .

(٢) ثمانية أتساع القياسات الـ 25 ، على الأقل ، واقع ضمن المجال $\bar{x} \pm 3s' = 75 \pm 30$ أي بين 45 و 105 .

ويجدر الانتباه الى عبارة « على الأقل » فالنظرية متحفظة باعتبارها تنطبق على أي توزيع للقياسات . وفي معظم الحالات تكون النسبة الفعلية للقياسات الواقعة ضمن المجال المحدد أكبر من $1 - \frac{1}{k^2}$.

ونعرض الآن قاعدة تصف بدقة تشتت توزيعات لها شكل الجرس أو ما يشابهه من الأشكال المقببة للتوزيعات . وتواتر وقوع مثل هذه التوزيعات في الطبيعة ، وبالتالي قابلية تطبيق مثل هذه القاعدة بشكل واسع ، تدعونا لتسميتها بالقاعدة التجريبية .

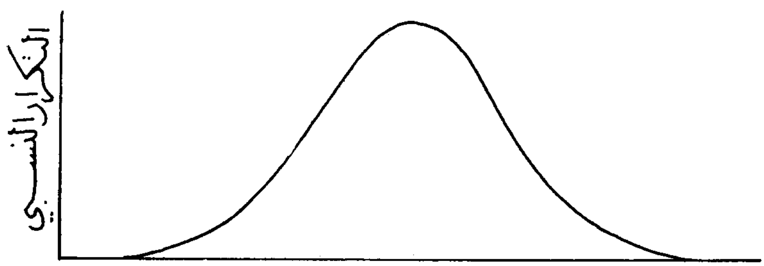
القاعدة التجريبية : ليكن توزيعاً من القياسات على شكل جرس تقريباً فعندئذ :

(١) يحوي المجال $\mu \pm \sigma$ بصورة تقريبية 68% من القياسات .

(٢) يحوي المجال $\mu \pm 2\sigma$ بصورة تقريبية 95% من القياسات .

(٣) يحوي المجال $\mu \pm 3\sigma$ بصورة تقريبية 99.7% من القياسات .

وبصورة عامة يدعى التوزيع الذي يأخذ شكل الجرس (أنظر الشكل ٢ ــ ٨)



شكل ٢ ــ ٨ التوزيع الطبيعي

بالتوزيع الطبيعي وسنناقش هذا التوزيع الفائق الأهمية بالتفصيل في فصل قادم .

مثال ٢ ـ ٥ قمنا بدراسة زمنية لتحديد الوقت الذي يستغرقه انجاز عملية معينة في منشأة صناعية . وقد قسنا الزمن الضروري لانجاز هذه العملية من أجل كل من $n = 40$ عاملاً . ووجدنا أن المتوسط والانحراف المعياري هما 12.8 و 1.7 على الترتيب . ولوصف البيان الاحصائي الذي حصلنا عليه نحسب المجالات :

12.8 ± 1.7	=	$\bar{x} \pm s'$	أو 11.1	إلى	14.5
$12.8 \pm 2\,(1.7)$	=	$\bar{x} \pm 2s'$	أو 9.4	إلى	16.2
$12.8 \pm 3\,(1.7)$	=	$\bar{x} \pm 3s'$	أو 7.7	إلى	17.9

ووفقاً للقاعدة التجريبية يمكن أن نتوقع ، على وجه التقريب ، أن 68% من القياسات واقع ضمن المجال (11.1 14.5) ، وأن 95% من القياسات ضمن المجال (9.4,16.2) و 99.7% من القياسات ضمن المجال (7.7, 17.9)

واذا كنا في شكّ حول شكل التوزيع بأنه ليس جرسياً وإنما على شكل محدب ، أو أردنا لسبب معين أن تكون نتائجنا متحفظة فيمكن عندئذ تطبيق نظرية تشيبيشيف للحصول على نتائج لا ريب فيها وعندها نقول أن 75% على الأقل من القياسات يقع ضمن المجال ((9.4, 16.2)) وعلى الأقل 8/9 من القياسات يقع ضمن المجال (7.7 17.9) .

لنر الآن مدى انطباق القاعدة التجريبية على المعدلات الثلاثين في الجدول (٢ ـ ١) . فلدينا هنا $\bar{x} = 69.45$ $s' = 12.33$. ويبين الجدول (٢ ـ ٥) المجالات الموافقة للقيم 1, 2, 3 للعدد k وعدد القياسات التي تقع ضمن كل مجال ثم نسبة هذا العدد إلى عدد القياسات الكلي $n = 30$

	المجال x + ks′	التواتر ضمن المجال	التواتر النسبي
1	57.12-81.78	19	.63
2	44.79-94.11	29	.97
3	32.46-106.44	30	1.0

٢ ــ ٧ طريقة مختزلة لحساب التشتت : يتطلب حساب التشتت والانحراف المعياري لمجموعة من القياسات الكثير من العمليات الحسابية خاصة اذا اتبعنا التعريف ، أي حسبنا كل انحراف على حدة كما في الجدول (٢ ــ ٣) . وسنستخدم القياسات المبينة في الجدول (٢ ــ ٣) لتوضيح طريقة أقصر في الحساب . ويحوي الجدول (٢ ــ ٦) عمودين أحدهما يتضمن القياسات المفروضة ويتضمن العمود الثاني مربعات هذه القياسات

جدول (٢ ــ ٦)

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

ونحسب الآن

$$\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 95 - \frac{(19)^2}{5} = 95 - \frac{361}{5}$$
$$= 95 - 72.2 = 22.8 .$$

وهذه النتيجة هي نفس ما حصلنا عليه في الجدول (٢ ــ ٣) من اجل $\sum_{i=1}^{n}(x_i-\bar{x})^2$. وبالطبع ليس هذا التوافق من باب الصدفة . فسنبين الآن أن :

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

ومن أجل ذلك نكتب :

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

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

أو :

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

٢ ـ ٨ تقدير تشتت المجتمع : نذكر أننا استخدمنا متوسط العينة $\bar{x}$ كتقدير لمتوسط المجتمع μ. وعلى نفس النهج يبدو من المنطقي أن نعتبر تشتت العينة s'^2 كتقدير لتشتت المجتمع σ^2 . ويمكن البرهان على أن تشتت العينة s'^2 يميل إلى أن يبخس σ^2 بعضاً من قيمتها وأن العلاقة :

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} \qquad (5)$$

تزودنا بتقدير أفضل لِ σ^2 ونلاحظ أن s'^2 و S^2 يختلفان فقط بأن المخرج في الأول هو n بينما يساوي $n-1$ في الثاني . وعندما يكون n كبيراً يكون s'^2 و S^2 متساويين تقريباً . وعلى أي حال فإن الإحصائي يحسب دائماً S^2 وليس s'^2 ويشير عملياً إلى S^2 على أنها تشتت العينة . وتبقى نظرية تشيبيشيف صحيحة ،عند استخدام S بدلاً من s'.وكذلك القاعدة التجريبية ، كقاعدة تقريبية تُطبق على عينات من الحجم المتوسط أو الكبير ، تحتفظ بفائدتها عند استخدام S بدلاً من s' . وحيثما ذكرنا بعد الآن تشتت العينة فإنما نقصد S^2 .

مثال ٢ ـ ٦ : احسب $\bar{x}$ و S لمجموعة القياسات 85, 70, 60, 90, 81

x_i	x_i^2
85	7225
70	4900
60	3600
90	8100
81	6561
386	30386

$$\bar{x}=\frac{386}{5}=77.2$$

$$= 30386 - \frac{(386)^2}{5} = 30386 - 29799.2 = 586.8$$

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

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} = \sqrt{\frac{586.8}{4}} = \sqrt{146.7} = 12.1$$

تمارين

١ ـ قم بالتجربة التالية : اقذف عشر قطع من النقود وسجل x عدد أوجه « النقش » التي تحصل عليها . أعد هذه العملية $n = 50$ مرة ، فتحصل على 50 قيمة لِ x . والمطلوب :

p ـ إقامة مضلع التكرار النسبي لهذه القياسات .

ب ـ احسب $\bar{x}$، s^2، s .

جـ ـ أوجد نسبة القياسات الواقعة ضمن المجال $\bar{x} \pm 2s$. وهل تتفق النتائج مع نظرية تشيبيشيف ؟ وهل تصف القاعدة التجريبية تشتت هذه القياسات الخمسين بصورة مناسبة ؟

د ـ أعد السؤال (جـ) مستخدماً المجال $\bar{x} \pm s$.

٢ ـ تمثل القياسات التالية المعدلات الوسطية لخمس وعشرين من طلبة السنة الجامعية الأولى :

65.0	45.0	65.0	92.5	47.5
50.3	67.5	75.0	60.0	57.5
77.5	65.0	65.0	62.5	67.5
67.5	72.5	85.0	47.5	57.5
82.5	55.0	87.5	75.0	62.5

والمطلوب :

ا ـ إقامة مضلع التكرار النسبي .

ب ـ حساب $\bar{x}$ ، s^2 ، s .

جـ ـ ايجاد نسبة المعدلات الواقعة ضمن المجال $\bar{x} \pm s$ وضمن المجال $\bar{x} \pm 2s$. وهل تتفق هذه النتائج مع نظرية تشيبيشيف والقاعدة التجريبية .

٣ ـ لتكن مجموعة القياسات 0 ,3 ,1 ,2 ,7 ,5 إحسب $\bar{x}$ و s^2 و s

٤ ـ لتكن مجموعة القياسات 4 ,0 ,3 ,5 ,1 ,0 ,3 . أحسب $\bar{x}$. s^2 و s .

٥ ـ لماذا يفضل الإحصائي استخدام s^2 بدلاً من s'^2 لتقدير تشتت المجتمع σ^2 ?

٦ ـ إذا علمنا أن معدلات مادة الرياضيات لـ 400 طالباً لها متوسط يساوي 60 وتشتت يساوي 49 . فاستخدم نظرية تشيبيشيف لوصف توزيع هذه المعدلات . وإذا كان للتوزيع شكل الجرس فكم من المعدلات على وجه التقريب يقع بين 53 و 67 . وعلى وجه التقريب كم من المعدلات نتوقع أن تتجاوز 74 ؟

٧ ـ آلة ميكانيكية تنتج نوعاً من المسامير بقطر يساوي على المتوسط 51 . بوصة وانحراف معياري 01 . بوصة . فإذا كان توزيع أقطار المسامير طبيعياً تقريباً ، فما هي النسبة من الانتاج الكلي التي سيكون لها قطر واقع بين 49. و 53. بوصة ؟

٨ ـ بالإشارة إلى التمرين السابق . لنفرض أن مواصفات المسمار الصحيح تتطلب أن يكون قطره مساوياً لـ 02. ± 5. بوصة . والمسمار الذي لا يحقق هذا الشرط يكون ناقص الصنع . فإذا اشتغلت الآلة وفقاً لما وصفنا في التمرين السابق فما هي النسبة من الانتاج الكلي التي قد تكون ناقصة الصنع ؟

٩ ـ أنتجت مسابقة لمدة ساعة في مادة الإحصاء العلامات التالية :

83	91	88	82	81
84	62	93	50	63
96	68	73	80	97
95	91	38	82	72
88	83	78	93	69
78	91	87	84	91

والمطلوب :

أ ـ رسم مضلع التكرار النسبي لهذه العلامات .

ب ـ حساب $\bar{x}$ ، s^2 و s

جـ ـ إيجاد عدد الدرجات الواقعة ضمن المجال $\bar{x} \pm s$ وضمن المجال $\bar{x} \pm 2s$ وقارن الناتج مع نظرية تشيبيشيف والقاعدة التجريبية .

١٠ ـ في المثال السابق أحسب المدى . وأوجد نسبة المدى إلى s . وإذا كان لدينا عدد كبير من القياسات لها توزيع على شكل الجرس . فكم نتوقع أن تكون نسبة المدى إلى s .

١١ ـ أوجد نسبة المدى إلى s في التمرين الأول .

١٢ ـ أوجد نسبة المدى إلى s في حالة عدد صغير من القياسات . في التمرينين ٣ و ٤ مثلاً . ولاحظ أن هذه النسبة ستكون بصورة عامة أصغر عندما تكون n صغيرة .

١٣ ـ مجموعة من 128 درجة في امتحان التاريخ لها متوسط 92% وانحراف معياري 10% . هل تتوقع أن يكون توزيع التكرار النسبي لهذه الدرجات على شكل جرس ؟ لماذا ؟