الفصل الخامس

عروض ووصف البيانات

Presentation and Description of Data

(٥-١) المجتمعات والعينات **Populations and Samples**

تسجل نتيجة كل تجربة إحصائية ،كما ذكرنا في الفصل الثالث ، إما بقيمة رقمية أو تمثيل وصفي. فعلى سبيل المثال عند إلقاء زهرة نرد مره واحدة وإذا كان الإهتمام بعدد النقاط التي تظهر على السطح العلوي للنرد فإننا نسجل قيمة رقمية. بينما عند سؤال مجموعة من العاملين في هيئة ما عن الحالة الإجتماعية لكل منهم، فإن التمثيل الوصفي يكون أكثر فائدة. فالحالة الاجتماعية لأي شخص إما أعزب أو متزوج أو مطلق أو أرمل. عادة يهتم الإحصائى بالقيم الرقمية لذلك فإن التمثيل الوصفي يمكن تحويله إلى قيم عددية. فعلى سبيل المثال عند تصنيف مجموعة من الأشخاص حسب الحالة الإجتماعية فإنه يمكن تخصيص الرقم 1 للأعزب والرقم 2 للمتزوج والرقم 3 للمطلق والرقم 4 للأرمل. القيمة التي تسجل من نتيجة تجربة إحصائية تسمى بيان أو مشاهدة (مقياس) كما ذكرنا في الفصل الثالث. عندما يقوم باحث بتصنيف العاملين في شركة ما حسب الحالة الإجتماعية، في هذه الحالة يكون لديه عدد محدود من المشاهدات. بينما عند إلقاء زهرة نرد عدد لانهائي من المرات وتسجيل عدد النقط التي تظهر في كل مرة فإننا نحصل على فئة لانهائية من القيم. كل المشاهدات تحت الدراسة ، سواء كانت محدودة أو غير محدودة ، تسمى مجتمع **population** . في السنوات الماضية كانت كلمة مجتمع تشير إلى مشاهدات من دراسات إحصائية تشمل أشخاص. أما الآن فإن الإحصائي يستخدم هذه الكلمة لتشير إلى مشاهدات عن أي شيء موضع إهتمامه سواء مجموعة من الأشخاص، حيوانات، نباتات... الخ .

تعريف : يتكون المجتمع من كل الأشياء التي نهتم بها.

عدد المشاهدات في المجتمع تسمى حجم المجتمع وعادة يرمز لحجم المجتمع بالرمز N ، وفي هذه الحالة نقول أن المجتمع محدود. فعلى سبيل المثال عند تصنيف 500 شخصا في شركة ما حسب الحالة الإجتماعية، فإننا نقول أن المجتمع محدود وحجمه $N=500$. الأطوال والأوزان والدخل السنوي لمجموعة من الأشخاص أمثلة لمجتمعات محدودة. في كل حالة العدد الكلى للمشاهدات رقم محدود. في بعض الأحيان يكون حجم المجتمع غير محدود، مثل مجتمع كرات الدم البيضاء التي تسرى في دم إنسان. أيضا المشاهدات التي نحصل عليها من قياس الضغط الجوى كل يوم من الماضي إلى المستقبل تمثل مجتمع غير محدود.

كل مشاهدة في المجتمع تمثل قيمة من قيم المتغير العشوائي X .على سبيل المثال عند إلقاء زهرة نرد عدد لانهائي من المرات وإذا كان X يمثل عدد النقط التي تظهر على النرد كل مرة،

أي أن x=1,2,3,4,5,6 ، فإن كل مشاهدة في المجتمع تمثل قيمة من قيم المتغير العشوائي X.

لمزيد من التوضيح بفرض أن المجتمع التالي يمثل عدد مرات الغياب في السنة المسجلة لكل من عشرين طالبا في كلية ما: 0,0,7,8,9,9,9,8,8,6,5,4,3,2,1,1,6,6,7,8 ٠ وعلى ذلك X متغير عشوائي يأخذ القيم 0,1,2,3,4,5,6,7,8,9 ٠ قيم X مع الإحتمالات المقابلة لها تعرف التوزيع الإحتمالي للمتغير العشوائي X ٠ وبما أن الإحتمالات تمثل تكرارات نسبية، فإن المجتمع أصبح معرف تماما من توزيعه الاحتمالي٠ إن دراسة خواص المتغير العشوائي تكافئ دراسة خواص المجتمع. على سبيل المثال القيمة المتوقعة للمتغير X هي :-

$$E(X) = (0)(\frac{2}{20}) + (1)(\frac{2}{20}) + (2)(\frac{1}{20}) + (3)(\frac{1}{20}) + (4)(\frac{1}{20}) + (5)(\frac{1}{20}) +$$

$$(6)(\frac{3}{20}) + (7)(\frac{2}{20}) + (8)(\frac{4}{20}) + (9)(\frac{3}{20}) = \frac{107}{20} = 5.35.$$

والتي تساوى الوسط الحسابي للمجتمع μ وذلك من المعادلة التالية :-

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{107}{20} = 5.35.$$

تعريف : القيم العددية التي توصف المجتمع تسمى معالم.

يهتم الباحث بالوصول إلى إستنتاجات تخص معالم المجتمع، ولكن عادة يكون من المستحيل أو غير عملي ملاحظة كل قيم الفئة الممثلة للمجتمع. وعلى ذلك لابد من الإعتماد على فئة جزئية من قيم المجتمع لتساعدنا في الوصول إلى إستدلالات عن المعالم، وهذا يأخذنا إلى نظرية المعاينة theory of sampling .

تعريف : العينة sample هي فئة جزئية من المجتمع.

حتى يكون الإستدلال صحيح لابد من فهم العلاقة بين المجتمع والعينة. من المؤكد أن العينة سوف تمثل المجتمع لذلك لابد أن تكون غير متحيزة unbiased أى عينة عشوائية random sample .

تعريف : العينة العشوائية من الحجم n هي عينة تختار بحيث أن كل فئة جزئية حجمها n من مشاهدات المجتمع لها نفس الإحتمال في الاختيار.

قد نرغب في الوصول إلى إستنتاجات تخص نسبة الأشخاص المدخنين في بلد ما. في بعض الأحيان يكون من الصعوبة سؤال كل شخص في هذا البلد وحساب المعلمة التي تمثل نسبة

المدخنين الحقيقية. بدلا من ذلك نختار عينة عشوائية كبيرة ونحسب النسبة من العينة. هذه القيمة تستخدم في عمل بعض الإستدلال الذي يخص النسبة الحقيقية. القيمة المحسوبة من العينة تسمى الإحصاء statistic • وبما أن عينات عشوائية كثيرة يمكن إختيارها من نفس المجتمع فإننا نتوقع أن يختلف الإحصاء من عينة إلى أخرى، وعلى ذلك يعتبر الإحصاء متغير عشوائي •

تعريف : الإحصاء متغير عشوائي يعتمد فقط على قيم العينة المختارة •

عادة، يمثل قيمة أي إحصاء بحرف من الحروف اللاتينية الصغيرة • على سبيل المثال إذا أخذنا عينة عشوائية من الحجم $n=3$ من المجتمع الذي يمثل عدد مرات الغياب لـ 20 طالب و بفرض أن قيم العينة 6,7,8 . يعتبر الوسط الحسابي إحصاء وعادة يمثل بالرمز $\overline{X}$ • قيمة المتغير العشوائي $\overline{X}$ لهذا المثال هي $\overline{x}$، وعلى ذلك :-

$$\overline{x} = \frac{6+7+8}{3} = 7.$$

في التطبيق، قيمة الإحصاء تستخدم في تقدير قيمة معلمة المجتمع. لمعرفة مدى جودة قيمة الإحصاء (التقدير) لابد من معرفة التوزيع الاحتمالي للإحصاء والذي يسمى التوزيع العيني **sampling distribution** • التوزيعات لبعض الإحصاءات المفيدة سوف نتناولها في الفصل السادس.

(٥-٢) التوزيع التكراري **Frequency Distribution**

غالبا ما يكون التوزيع الاحتمالي لمتغير عشوائي غير معروف. تعتبر البيانات الإحصائية التي يجمعها الباحث بكميات كبيرة مفيدة جدا في دراسة سلوك المتغير العشوائي إذا تم عرضها بشكل مناسب. المعلومات الكثيرة يمكن الحصول عليها بتجميع البيانات في فئات **classes** وحساب عدد المشاهدات في كل فئة. مثل هذا التنظيم يسمى التوزيع التكراري • عندما نقوم بتجميع البيانات في فئات فإننا نحصل على أحسن صورة للمجتمع موضع الدراسة ولكننا في المقابل نفقد الكثير من التفصيلات عن المشاهدات في العينة. عدد المشاهدات في فئة خاصة يسمى تكرار الفئة **class frequency** ويرمز له بالرمز **f** • الجدول (٥-١) يمثل التوزيع التكراري لأطوال **22** نبات من نوع ما (المشاهدات معطاة لأقرب عدد صحيح). في هذا المثال لدينا **6** فئات وهم :**60-64 , 55-59 , 50-54 , 45 -49 , 40-44 , 35-39** . يشار الى أصغر وأكبر القيم التي تقع في فئة معطاة بحدود هذه الفئة **class limits** • فعلى سبيل المثال للفئة **55 – 59** أصغر رقم هو **55** ويمثل الحد الأدنى للفئة **lower class limit** وأكبر رقم هو **59** ويمثل الحد الأعلى للفئة **upper class limit** • وحيث أن البيانات الأصلية مسجلة لأقرب رقم

صحيح، فإن 4 مشاهدات تقع في الفئة 59-55 يمثلون كل المشاهدات في العينة التي قيمهم أكبر من أو يساوى 54.5 وأصغر من 59.5 . الأرقام 54.5 و 59.5 تسمى الحدود الفعلية class boundaries للفئة 59−55 . الرقم 54.5 يسمى الحد الأدنى الفعلي lower class boundary والرقم 59.5 يسمى الحد الأعلى الفعلي upper class boundary، أيضا الرقم 59.5 يسمى الحد الأدنى الفعلي للفئة التالية أي الفئة 64-60 . ويلاحظ أنه بالرغم من أن الفئات لها حدود فعلية مشتركة إلا أنه من غير الممكن أن تقع مشاهدة واحدة على أحد هذه الحدود وذلك لأن الحدود الفعلية للفئات تحتوى على خانات عشرية أكبر من تلك الموجودة في البيانات نفسها.

جدول (٥−١)

حدود الفئة	35-39	40-44	45-49	50-54	55-59	60-64
التكرار	1	2	3	4	4	8

يعرف الفرق بين الحد الأعلى الفعلي والحد الأدنى الفعلي للفئة بطول الفئة class width ويساوى أيضا الفرق بين الحد الأعلى والحد الأدنى للفئة زائدا وحدة دقة، أي وحدة من الوحدات التي قربت إليها الأعداد في البيانات (في هذا المثال وحدة الدقة هي الواحد الصحيح لأننا قربنا البيانات لأقرب رقم صحيح). من الناحية العملية يفضل الحصول على فئات ذات أطوال متساوية ما أمكن. سوف نرمز لطول الفئة بالرمز Δ . أطوال الفئات في جدول (٥−١) متساوية وتساوى $\Delta = 5$.

منتصف الفئة midpoint تسمى مركز الفئة class midpoint أو class mark ونحصل عليها بجمع الحد الأدنى الفعلي والحد الأعلى الفعلي للفئة وقسمة المجموع على 2 وذلك تحت فرض أن جميع المشاهدات داخل الفئة تأخذ قيما تتطابق مع مركز الفئة. مثال ذلك إفتراض أن 8 تكرارات في الفئة 64-60 تأخذ القيمة 62 والتي تمثل مركز هذه الفئة. أيضا يمكن الحصول على مركز الفئة بجمع الحد الأدنى والحد الأعلى للفئة وقسمة المجموع على 2 . من جدول (٥−١) مراكز الفئات هم : 62 , 57 , 52 , 47 , 42 , 37 ، يمثل جدول (٥−١) توزيع تكراري من النوع الذي نشاهده في التقارير المنشورة في الصحف. للأغراض الإحصائية يكون من الأفضل الحصول على توزيعات ذات تفصيلات أكثر، كما هو موضح في جدول (٥−٢) لنفس البيانات المعطاة في جدول (٥−١) .

جدول (٥-٢)

حدود الفئة	الحدود الفعلية الفئة	مركز الفئة	التكرار
35-39	34.5-39.5	37	1
40-44	39.5-44.5	42	2
45-49	44.5-49.5	47	3
50-54	49.5-54.5	52	4
55-59	54.5-59.5	57	4
60-64	59.5-64.5	62	8

مثال (٥-١) في عينة عشوائية حجمها n=35 من حيوانات التجارب تم تسجيل كمية مركب ما في الدم لكل حيوان كما في جدول (٥-٣). المطلوب عرض هـذه البيانات في توزيع تكراري.

جدول (٥-٣)

1.1	1.2	1.4	1.1	1.2	1.5	1.6
1.2	1.6	1.5	1.8	1.9	1.8	1.7
1.4	1.5	2.1	2.2	2.2	2.3	2.4
2.5	2.6	2.7	2.8	2.9	2.9	2.9
2.9	2.8	2.8	2.7	2.7	2.6	2.6

الحل . في البداية نقرر عدد الفئات والتي سوف تتوزع فيها البيانات. عادة يفضل أن تكون عدد الفئات من 5 إلى 20 . إذا زاد عدد الفئات عن عشرين خسر الباحث البساطة التي يكسبها عادة عند وضع البيانات في التوزيع التكراري، وإذا قل عدد الفئات عن 5 فإن ذلك يؤدى إلى ضياع الكثير من التفصيلات الموجودة في البيانات. بفرض أننا قررنا أن يكون عدد الفئات 5 ، لحساب طول الفئة فإننا أولا نحسب المدى وهو عبارة عن الفرق بين أكبر وأصغر مشاهدة في العينة وعلى ذلك يكون المدى 2.9-1.1=1.8 . ثانيا نقسم المدى على عدد الفئات المقترحة أي $\frac{1.8}{5} = 0.36$ ويقرب الكسر الناتج من خارج القسمة إلى أقرب رقم عشري (لأن البيانات الخام أصلا مقاس لأقرب رقم عشري). أي أن طول الفئة سوف يكون $\Delta = 0.4$. نحدد بداية الفئة الأولى (الحد الأدنى للفئة الأولى) والذي غالبا ما يكون أصغر رقم في البيانات وهو 1.1 ، وكذلك نحدد الحد الأدنى للفئة الثانية بإضافة طول الفئة إلى الحد الأدنى للفئة الأولى، وهكذا لتعيين الحدود الدنيا لباقي الفئات. أما بالنسبة لتحديد الحد الأعلى للفئة الأولى فإنه يمكن تعيينه بإضافة طول الفئة إلى الحد الأدنى للفئة الأولى ثم نطرح من حاصل الجمع مقدار وحدة دقة من الوحدات التي قربت إليها الأعداد في المشاهدات، أي 0.1 . وكذلك نحدد الحد الأعلى للفئة الثانية بإضافة طول الفئة إلى الحد الأعلى للفئة الأولى، وهكذا لتعيين الحدود العليا لباقي الفئات،

وذلك تحت شرط أن الفئات متساوية الأطوال. أخيرا نقوم بعد عدد المشاهدات التي تقع في كل فئة ويوضع الرقم في عمود التكرار، ولابد أن يكون مجموع التكرارات مساويا لعدد المشاهدات في جدول (٥-٣) • يمثل جدول (٥-٤) التوزيع التكراري للبيانات المعطاة في جدول (٥-٣). التكرار النسبي لكل فئة يمكن الحصول عليه بقسمة تكرار الفئة على مجموع التكرارات. الجدول الذي يحتوى على التكرارات النسبية يسمى التوزيع النسبي the relative frequency •التوزيع النسبي للبيانات المعطاة في جدول (٥-٤) موضح في جدول (٥-٥)• وبضرب كل تكرار نسبي في 100 نحصل على التكرار المئوي percentage distribution•

جدول (٥-٤)

حدود الفئة	الحدود الفعلية الفئة	مركز الفئة	التكرار
1.1-1.4	1.05-1.45	1.25	7
1.5-1.8	1.45-1.85	1.65	8
1.9-2.2	1.85-2.25	2.05	4
2.3-2.6	2.25-2.65	2.45	6
2.7-3.0	2.65-3.05	2.85	10
المجموع			35

جدول (٥-٥)

حدود الفئة	1.1-1.4	1.5-1.8	1.9-2.2	2.3-2.6	2.7-3.0	المجموع
التكرار النسبي	0.2000	0.2286	0.1143	0.1714	0.2857	1

في بعض الأحيان يكون الإهتمام ليس فقط بعدد المشاهدات في فئة معطاة ولكن في عدد المشاهدات الذي يقع فوق أو تحت قيمة معينة. على سبيل المثال في جدول (٥-٤) عدد الحيوانات التي كمية المركب في دمها 2.25 أو أقل هو 19=4+8+7، والذي يمثل التكرار المتجمع cumulative frequency للفئة الرابعة. جدول (٥-٦) يوضح التكرارات المتجمعة والتي تم حسابها من جدول (٥-٤) ، ويسمى التوزيع التكراري المتجمع cumulative frequency distribution .

أما إذا استخدمنا التكرارات النسبية بدلا من التكرارات فإننا نحصل على التكرار المتجمع النسبي relative cumulative frequency. وإذا استخدمنا التكرارات المئوية بدلا من التكرارات فإننا نحصل على التوزيع التكراري المتجمع المئوي• كما يمكن الحصول

التوزيع التكراري المتجمع المئوي من جدول التكرار المتجمع وذلك بقسمة التكرار المتجمع لكل فئة على مجموع التكرارات وضرب الناتج في 100% ، يوضح جدول (٥-٧) التوزيع التكراري المتجمع المئوي للبيانات في جدول (٥-٦).

جدول (٥-٦)

التكرار المتجمع	الحدود العليا للفئات
0	أقل من 1.05
7	أقل من 1.45
15	أقل من 1.85
19	أقل من 2.25
25	أقل من 2.65
35	أقل من 3.05

جدول (٥-٧)

التكرار المتجمع %	الحدود العليا للفئات
00.00	أقل من 1.05
20.00	أقل من 1.45
42.86	أقل من 1.85
54.29	أقل من 2.25
71.43	أقل من 2.65
100.00	أقل من 3.05

مثال (٥-٢) البيانات في جدول (٥-٨) تمثل الدخل اليومي الصافي-لأقرب جنيه- في محل تجارى في مدة 40 يوم. كون جدول تكراري لهذه البيانات على أن يحتوى الجدول على مركز الفئة - الحدود الفعلية للفئة - التكرار المئوي - التكرارالمتجمع .

جدول (٥-٨)

118	124	128	134	135	138	140	142
125	130	136	138	141	143	145	144
144	146	147	150	152	154	155	146
146	147	155	168	157	163	160	163
146	168	170	175	181	181	175	168

الحل . أولا نحسب المدى = أكبر قيمة–أصغر قيمة أي 63=118-181 وحيث أن عدد الفئات المقترحة 13 فإن طول الفئة هو $\frac{63}{13} = 4.846$ أي تقريبا 5 لأن البيانات مقاسة لأقرب عـــدد صحيح. الحد الأدنى للفئة الأولى هو 118 والذي يمثل أصغر قيمة في البيانات.جـــدول (٥–٩) يحتوى على المطلوب.

جدول (٥–٩)

حدود الفئة	الحدود الفعلية	مركز الفئة	التكرار	التكرار المئوي	التكرار المتجمع
118-122	117.5-122.5	120	1	2.5	1
123-127	122.5-127.5	125	2	5.0	3
128-132	127.5-132.5	130	2	5.0	5
133-137	132.5-137.5	135	3	7.5	8
138-142	137.5-142.5	140	5	12.5	13
143-147	142.5-147.5	145	10	25.0	23
148-152	147.5-152.5	150	2	5.0	25
153-157	152.5-157.5	155	4	10.0	29
158-162	157.5-162.5	160	1	2.5	30
163-167	162.5-167.5	165	2	5.0	32
168-172	167.5-172.5	170	4	10.0	36
173-177	172.5-177.5	175	2	5.0	38
178-182	177.5-182.5	180	2	5.0	40
المجموع			40	100	

مثال (٥–٣) أخذت عينة من مزرعة دواجن وكانت أوزان الدجاج مقاسه لأقرب مائة جـــرام كما في جدول (٥–١٠). كون جدول تكراري لهذه البيانات على أن يحتوى الجدول علــــى مركز الفئة – الحدود الفعلية للفئة – التكرار المئوي – التكرارالمتجمع .

الحل . أولا نحسب المدى = أكبر قيمة–أصغر قيمة أي 800=600-1400 وحيـــث أن عـــدد الفئات المقترحة 5 فإن طول الفئة هو $\frac{800}{5} = 160$ أي تقريبا 200 لأن البيانات مقاس لأقرب 100 جرام. الحد الأدنى للفئة الأولى هو 600 والذي يمثل أصغر قيمة في البيانات. جــــدول (٥–١١) يحتوى على المطلوب.

جدول (٥–١٠)

600	600	900	1000	800	800	900
700	800	800	900	700	700	700
600	900	700	1100	1200	1200	1300
1300	1300	1400	1000	1200	1200	1400
1300	1200	1300	1300	1400	1400	1000

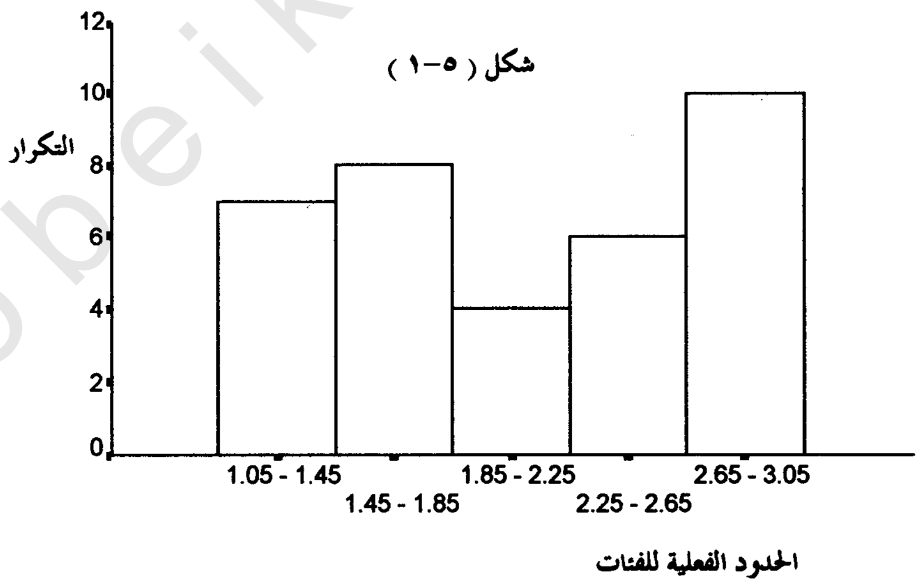
جدول (٥-١١)

التكرار المتجمع	التكرار المئوي	التكرار	مركز الفئة	الحدود الفعلية	حدود الفئة
8	22.86	8	650	550-750	600-700
16	22.86	8	850	750-950	800-900
20	11.43	4	1050	950-1150	1000-1100
31	31.42	11	1250	1150-1350	1200-1300
35	11.43	4	1450	1350-1550	1400-1500
	100	35			المجموع

(٥-٣) التمثيل البياني **Graphic Representation**

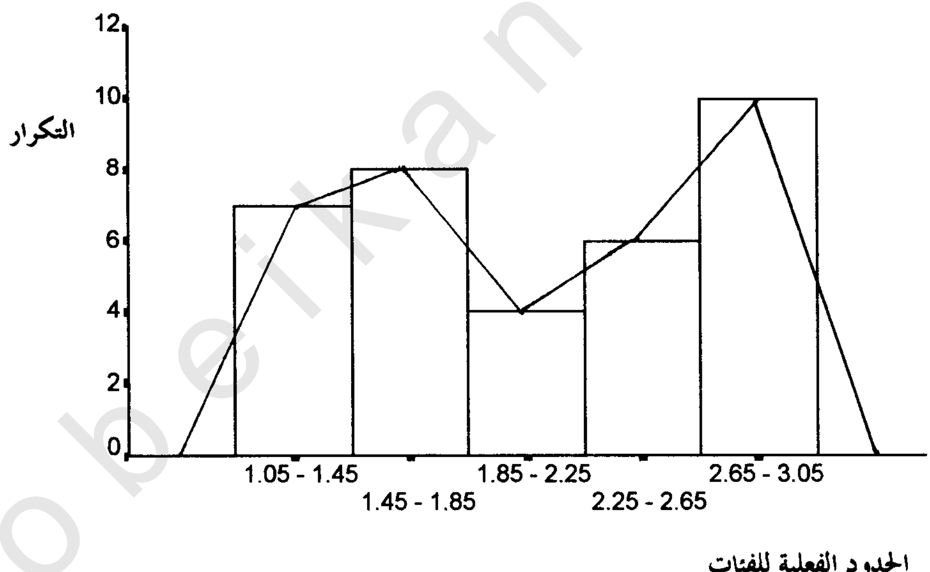
المعلومات التي نحصل عليها من التوزيع التكراري في شكل جدولي تصبح أسهل في الفهم إذا تم عرضها بيانيا ، من أكثر الأشكال البيانية الواسعة الانتشار في تمثيل البيانات الرقمية ما يعرف بالمدرج التكراري **histogram** والذي يناسب البيانات المتصلة. ونحصل عليه بتمثيل تكرار كل فئة من فئات التوزيع بمستطيل قاعدته الحدود الفعلية لتلك الفئة وارتفاعه يساوى تكرار الفئة. ويتم ذلك برسم محورين أحدهما أفقي والآخر رأسي ونرصد على المحور الأفقي الحدود الفعلية لكل فئة من فئات التوزيع التكراري ونقيم على كل فئة مستطيل ارتفاع يساوى تكرار تلك الفئة. المدرج التكراري للتوزيع التكراري المعطى في جدول (٥-٤) موضح في شكل (٥-١) .

شكل (٥-١)



في بعض المشاكل يكون من الأفضل وضع التكرار النسبي أو المئوي على المحور الرأسي. المدرج التكراري في هذه الحالة يسمى المدرج التكراري النسبي **relative frequency histogram** أو المدرج التكراري المئوي **percentage frequency histogram** ، ويكون لهما نفس شكل المدرج التكراري، كما أن مجموع مساحات الأعمدة للمدرج التكراري النسبي تساوى الواحد الصحيح.

الطريقة الثانية المفيدة لتمثيل البيانات الرقمية بيانيا هو استخدام المضلع التكراري **frequency polygon** والذي نحصل عليه بتصنيف الأضلاع العلوية للمستطيلات في المدرج التكراري ثم نوصل هذه النقاط بعضها بالبعض. ولكي نغلق الخط المنكسر الذي حصلنا عليه نحدد على المحور الأفقي مركز الفئة السابقة للفئة الأولى ومركز الفئة اللاحقة للفئة الأخيرة ونغلق المضلع. يوضح شكل (٥-٢) المضلع التكراري للتوزيع التكراري في جدول(٥-٤) .ويتضح من الشكل أن مجموع مساحات المستطيلات يساوى المساحة تحت المضلع التكراري، ويحدث هذا فقط في حالة الفئات المتساوية .

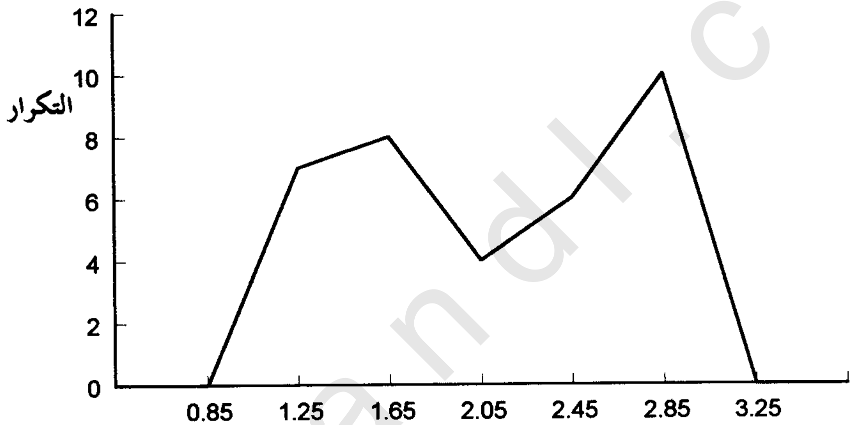


شكل (٥-٢)

وهناك طريقة أخرى لرسم المضلع التكراري وذلك برسم محورين أحدهما أفقي والآخر رأسي. يمثل المحور الأفقي مراكز الفئات ويمثل المحور الرأسي التكرارات. نعتبر مركز كل فئة إحداثيا أفقيا لنقطة ونعتبر تكرار هذه الفئة الإحداثي الرأسي لتلك النقطة. ولكي نغلق الخط

المنكسر الذي حصلنا عليه نحدد على المحور الأفقي مركز الفئة السابقة للفئة الأولى ومركز الفئة اللاحقة للفئة الأخيرة ونغلق المضلع. يوضح شكل (٥-٣) المضلع التكراري للتوزيع التكراري في جدول (٥-٤) وذلك بتحديد النقاط الآتية على الرسم :-
(1.25,7), (1.65,8), (2.05,4), (2.45,6), (2.85,10) والنقطتان الإضافيتان هما (0.85,0), (3.25,0). ثم وصل هذه النقاط بعضها بالبعض.

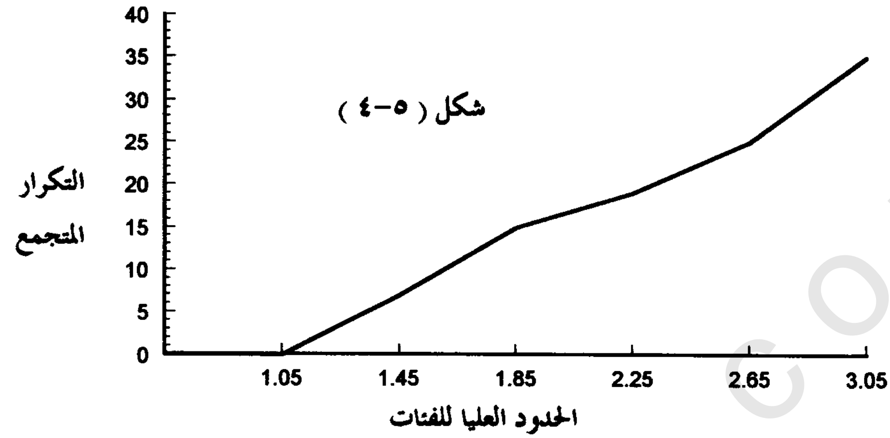
شكل (٥-٣)



عند الرغبة في مقارنة فئتين من البيانات مختلفتين في عدد مفرداتهما فإنه يمكن تمثيل المضلعين التكراريين على نفس الرسم. في هذه الحالة لابد من استخدام التكرارات النسبية أو المئوية.

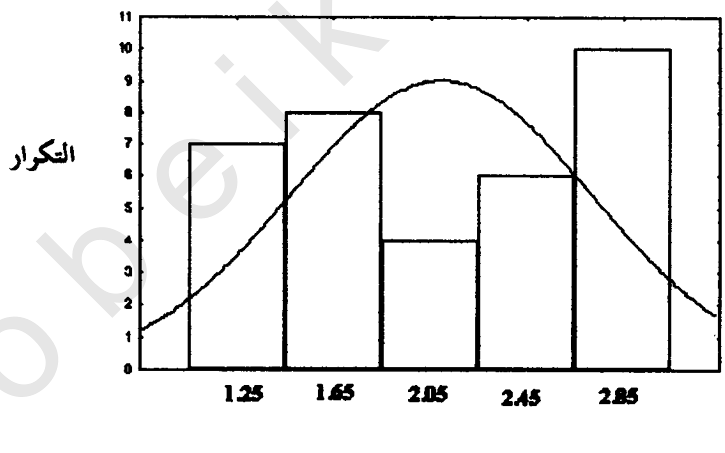
أما بالنسبة للتوزيعات التكرارية المتجمعة فهناك ما يسمى المضلع التكراري المتجمع **cumulative frequency polygon** ونحصل عليه برسم محورين أحدهما أفقي والآخر رأسي وكل نقطة على الرسم إحداثياها الحد الأعلى الفعلي للفئة والتكرار المتجمع وبتوصيل النقاط نحصل على المضلع التكراري المتجمع. شكل (٥-٤) يوضح المضلع التكراري المتجمع للتوزيع التكراري المتجمع في جدول (٥ - ٦).

يمكن تمثيل التوزيع التكراري المتجمع النسبي والتوزيع التكراري المتجمع المئوي بيانيا، بنفس الطريقة التي مثلنا بها التوزيع التكراري المتجمع بيانيا، وذلك باستخدام التكرارات المتجمعة النسبية والتكرارات المتجمعة المئوية.



شكل (٥-٤)

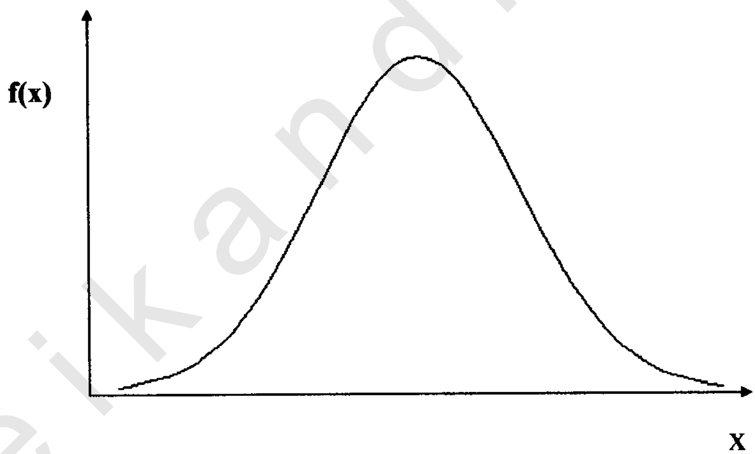
هناك طريقة أخرى لتمثيل التوزيعات التكرارية بيانيا وذلك باستخدام المنحنى التكراري frequency curve . ونحصل عليه برسم المضلع التكراري وتمهيد الخطوط المنكسرة التي تصل بين هذه النقط. وقد يكون التمهيد باليد أو بطرق رياضية ولا يشترط أن يمر المنحنى بجميع رؤوس المضلع التكراري. يبين شكل (٥-٥) المدرج التكراري والمنحنى التكراري معا للتوزيع التكراري في جدول (٥-٤)، عموما كلما ضاقت أطوال الفئات وزاد عدد المشاهدات فإن المضلع التكراري يؤول إلى المنحنى التكراري.



شكل (٥-٥)

عند الرغبة في تقدير التوزيع الاحتمالي لمتغير عشوائي متصل X نقوم بتمـــهيد المنحـــى التكراري النسبي. شكل المنحنى يساعدنا في اقتراح شكل $f(x)$ للمتغير العشـــوائي المتصـــل موضع الدراسة. على الرغم من أننا تمكنا من الحصول على تقدير للدالة $f(x)$ بيانيا فلا نـــزال نجهل الصيغة الرياضية أو المعادلة الخاصة بالدالة $f(x)$ وبالتالي لا نستطيع حســـاب تقديـــرات للاحتمالات. كثير من التوزيعات المتصلة يمكن تمثيلها بيانيا بمنحنى على شـــكل النـــاقوس **bell** كما في شكل (٥-٦). الصيغة أو المعادلة الخاصة بالدالة $f(x)$ معروفـــة (دالـــة كثافـــة الاحتمال) وتعتمد على معلمتين μ , σ. بمجرد الحصول على تقدير لكل منهما من البيانـــات يمكن كتابة المعادلة المقدرة ثم استخدام الجداول المناسبة لحساب أي مساحة تحت المنحنى.

شكل (٥ -٦)

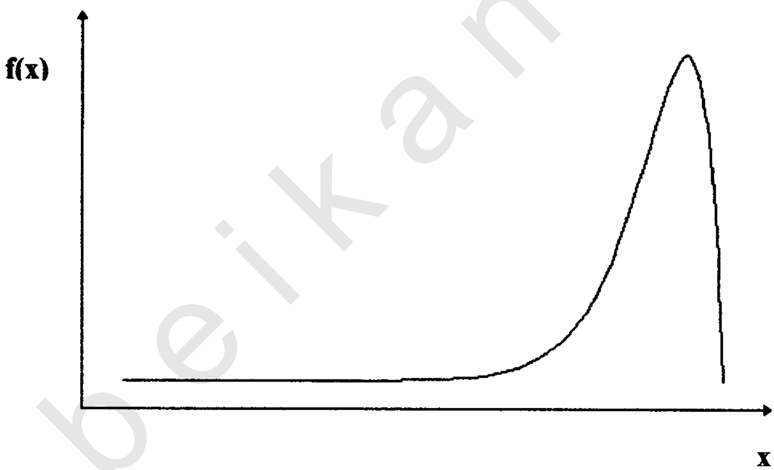


عادة تأخذ المنحنيات أشكال كثيرة ويمكن اشتقاق معادلتها بتقديـــر معالمـــها المجهولـــة. مشاكل التقدير صعبة جدا وتنتمي إلى فرع الإحصاء الرياضي.

يقال للتوزيع أنه متماثل **symmetrical**، كما في شكل (٥-٦) إذا أمكننا إقامـــة عمود على المحور الأفقي بحيث يقسم هذا العمود التوزيع إلى قسمين ينطبقان على بعضهما تمـــام الانطباق. النقطة التي تمكننا من إقامة العمود تسمى نقطة التماثل. أما التوزيعات التي يكون عـــدم التماثل واضحا فتسمى توزيعات ملتوية **skewed**. يكون التوزيع ملتويا إلى اليمين أو موجـــب الالتواء **positive skewed** إذا كان معدل التناقص في المنحنى أسرع جهة اليمين منه جهـــة اليسار بحيث يكون الجانب الأيمن من المنحنى أطول من الجانب الأيســـر كما في شـــكل (٥-٧

)• بينما يكون التوزيع ملتويا إلى اليسار وسالب الالتواء negative skewed إذا كان معدل التناقص في المنحنى أسرع جهة اليسار منه جهة اليمين بحيث يكون الجانب الأيسر من المنحنى أطول من الجانب الأيمن كما في شكل (٥–٨)•

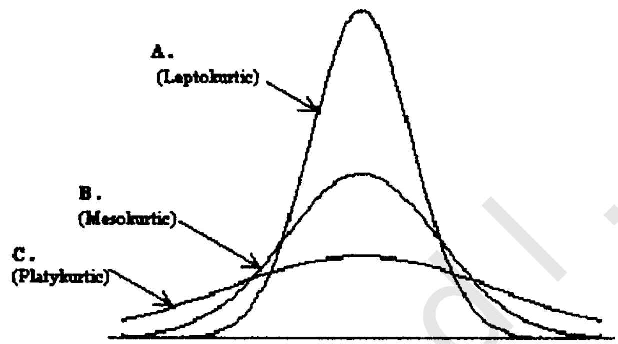
شكل (٥–٧)



شكل (٥–٨)

عند مقارنة المنحنيات وحيدة القمة قد نجدها تختلف من حيث شكل القمة. فقد تكون قمة إحداهما أكثر تدببا أو تفرطحا من بعض القمم الأخرى• ففي شكل (٥–٩) ثلاث منحنيات مختلفة في كمية التفلطح• المنحنى ، A ، ذو القمة المدببة leptokurtic يمثل توزيع بقيم تتركز بشدة حول نقطة الوسط midpoint• المنحنى الثاني ، B ، وهو المعتدل

mesokurtic يكون متوسط التفلطح ويمثل توزيع بقيم تتركز بدرجة أقل حول نقطة الوسط عن المنحنى المدبب . وأخيرا المنحنى ، C ، المفلطح platykurtic والذي يكون منبسطا وتنخفض قمته عن قمة المنحنى المعتدل . وهذا يدل على أن قيمه تقع حول نقطة الوسط في مدى غير ضيق .



شكل (٥-٩)

(٥- ٤) مقاييس النزعة المركزية Measures of Central Tendency

في بعض الأحيان التمثيل البياني وحده لا يمد الباحث بكل المعلومات التي يحتاج إليها من فئة المشاهدات تحت الدراسة، فالبيانات لابد أن توصف وتحلل. واحد من الطرق لوصف فئة من المشاهدات، سواء عينة أو مجتمع ، هو استخدام المتوسطات averages (مقاييس النزعة المركزية) . فالمتوسط هو القيمة التي تتركز حولها معظم المشاهدات . في هذا البند سوف نستعرض أربعة مقاييس للنزعة المركزية .

(٥-٤-١) الوسط الحسابي Arithmetic Mean

يعتبر الوسط الحسابي من أفضل مقاييس النزعة المركزية لتمثيل مركز فئة من المشاهدات.

تعريف : إذا كانت الفئة من المشاهدات $x_1, x_2, \ldots, x_N$ ، ليس من الضروري أن تكون كلها مختلفة ، تمثل مشاهدات مجتمع محدود من الحجم N ، فإن الوسط الحسابي للمجتمع يمكن حساب من الصيغة التالية :-

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}.$$

مثال (٥-٤) أوجد الوسط الحسابي لمجتمع مشاهداته هي :-

8,10,13,9,7,11,10,12,10,9,11.

الحل.

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{8+10+13+9+7+11+10+12+10+9+11}{11}$$

$$= \frac{110}{11} = 10.$$

تعريف : إذا كانت الفئة من المشاهدات $x_1, x_2, \ldots, x_n$ ، ليس من الضروري أن تكون كلها مختلفة ، تمثل مشاهدات عينة من الحجم n ، فإن الوسط الحسابي للعينة يمكن حسابه من الصيغة التالية :-

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

مثال (٥-٥) أوجد الوسط الحسابي لعينة مشاهداتها هي :-

6,7,7,8.

الحل .

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{6+7+7+8}{4}$$

$$= \frac{28}{4} = 7.$$

عند وضع البيانات في توزيع تكراري فإننا نفقد الهوية لأي مشاهدة في العينة. المعلومات التي تبقى هي عدد المشاهدات التي تقع في كل فئة. لحساب الوسط الحسابي من توزيع تكراري نفترض أن كل المشاهدات داخل فئة معطاة تقع عند مركز الفئة.

تعريف : إذا كانت $x_1, x_2, \ldots, x_k$ هي مراكز الفئات لتوزيع تكراري مع تكراراتها المقابلة $f_1, f_2, \ldots, f_k$ (حيث k تمثل عدد الفئات) فإن الوسط الحسابي يحسب من الصيغة التالية :

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}.$$

مثال (٥-٦) أوجد الوسط الحسابي للبيانات المعطاة في جدول (٥-١٢) والتي تمثل التوزيع التكراري لأطوال عينة من الأسماك.

جدول (٥-١٢)

حدود الفئة	التكرار f_i	مركز الفئة x_i	$f_i x_i$
30-39	11	34.5	379.5
40-49	12	44.5	534.0
50-59	16	54.5	872.0
60-69	23	64.5	1483.5
70-79	17	74.5	1266.5
80-89	11	84.5	929.5
90-99	10	94.5	945.0
المجموع	100		6410

الحل . بما أن $k = 7$, $\sum_{i=1}^{k} f_i = 100$, $\sum_{i=1}^{k} f_i x_i = 6410$ وعلى ذلك الوسط الحسابي هو :

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{6410}{100} = 64.1.$$

في بعض الأحيان يكون من المناسب إضافة (أو طرح) ثابت إلى القيم الأصلية ثم حساب الوسط الحسابي. السؤال الآن كيف تكون العلاقة بين الوسط الحسابي الجديد والوسط الحسابي لفئة المشاهدات الأصلية ؟ من نظرية (٤-١) بوضع $a = 1$ ، فإن $E(X + b) = E(X) + b$ ، وعلى ذلك فإن إضافة (أو طرح) ثابت إلى كل القيم سوف يغير الوسط الحسابي بنفس المقدار . على سبيل المثال لإيجاد الوسط الحسابي للقيم $-4,-2,1,4,6$ وبفرض أننا أضفنا 4 وحصلنا على القيم الموجبة التالية : $0,2,5,8,10$ والتي وسطها الحسابي 5. وعلى ذلك فإن الوسط الحسابي للقيم الأصلية $5 - 4 = 1$. أيضا من نظرية (٤-١) و نتيجة (٢) فإن $E(aX) = aE(X)$. وعلى ذلك بضرب (أو قسمة) ثابت إلى فئة من البيانات، فإن المشاهدات الأصلية سوف يكون وسطها الحسابي عبارة عن

الوسط الحسابي للمشاهدات الجديدة مقسوما على (مضروبا في) الثابت. على سبيل المثال الوسط الحسابي للقيم 5,10,15 هو 10، وعلى ذلك بعد قسمة كل المشاهدات على 5، فإن المشاهدات الجديدة هي 1,2,3 والتي وسطها 2 . وعلى ذلك الوسط الحسابي للقيم الأصلية هو $(2)(5) = 10$.

ومن مميزات الوسط الحسابي أنه مألوف وسهل الفهم كما أنه معرف لأى فئة من البيانات وقيمته وحيدة . أيضا كل قيمة في فئة المشاهدات تدخل في حسابه.

أما عيوبه فهي تأثره بالقيم الشاذة ولذلك لا ينصح باستخدامه للبيانات التي منحناها شديد الالتواء. أيضا لا يمكن تقديره من التوزيعات التكرارية التي تحتوى على فئات مفتوحة. وأخيرا لا يمكن حسابه بالرسم.

ومن الخصائص المميزة للوسط الحسابي أن مجموع انحرافات القيم (الانحراف هو بعد أي مشاهدة عن المتوسط) المشاهدة عن الوسط الحسابي يساوى صفرا، أى أن $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ ، أما في حالة التوزيعات التكرارية فإن $\sum_{i=1}^{k}(x_i - \overline{x}) \cdot f_i = 0$. أيضا مجموع مربعات انحرافات القيم حول الوسط الحسابي $\sum_{i=1}^{n}(x_i - \overline{x})^2$ أقل ما يمكن، أي أقل من مجموع مربعات انحرافات القيم عن أى قيمة أخرى غير الوسط الحسابي.

عند حساب الوسط الحسابي يفترض أن كل قيمة لها نفس الأهمية، مثل هذا الفرض قد يكون خاطئ. في الحقيقة، إذا كانت القيم ليس لها نفس الأهمية يكون من الأفضل حساب الوسط الحسابي المرجح. فإذا كانت $x_1, x_2, \ldots, x_n$ تمثل قيم المتغير X، وكانت $w_1, w_1, \ldots, w_n$ الأوزان المناظرة لها فإن الوسط الحسابي المرجح يكون :-

$$\overline{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

مثال (٥-٧) تدفع شركة أجرا قدره 25 جنيها في الساعة لعمالها غير مهرة وعددهم 20 وتدفع 30 جنيها في الساعة للعمال شبه مهرة وعددهم 10 وتدفع 40 جنيها في الساعة للعمال المهرة وعددهم 5 . ما هو الوسط الحسابي المرجح للأجر في الساعة التي تدفعه الشركة ؟

الحل .

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{(20)(25)+(10)(30)+(5)(40)}{20+10+5} = \frac{1000}{35} = 28.57.$$

يعاب على الوسط الحسابي المرجح هو عدم وجود قاعدة لتحديد الأوزان وتخضع للتقدير الشخصي. يعتبر الوسط الحسابي حالة خاصة من الوسط الحسابي المرجح إذا كانت الأوزان متساوية ،

إذا كانت لدينا k من المجموعات، وكانت أحجام عيناها $n_1, n_2, \ldots, n_k$ وأوساطها الحسابية على التوالي $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ فإن الوسط الحسابي المرجح للمجموعة التي حجمها $(n_1 + n_2 + \ldots + n_k)$ الناتجة من دمج k من المجموعات هو :-

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}.$$

ويسمى الوسط الحسابي المرجح للأوساط الحسابية ،

مثال (٥-٨) إذا كان الوسط الحسابي لأطوال 20 حيوان من نوع ما هو $\bar{x}_1 = 10$ وكان الوسط الحسابي لأطوال 30 حيوان من مجموعة أخرى من نفس النوع هو $\bar{x}_2 = 15$ أوجد الوسط الحسابي المرجح للأوساط الحسابية ،

الحل .

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i} = \frac{(20)(10)+(30)(15)}{20+30} = \frac{650}{50} = 13.$$

(٥-٤-٢) الوسيط Median

يعتبر الوسيط هو المقياس الأفضل بعد الوسط الحسابي ، سوف نرمز لوسيط المجتمع بالرمز $\tilde{\mu}$ ووسيط العينة بالرمز $\tilde{x}$ ، الوسيط لفئة من المشاهدات مرتبة تصاعديا (أو تنازليا) هو العدد الأوسط منها إذا كان عددها فرديا وهو الوسط الحسابي للعددين الأوسطين إذا كان عددها زوجيا ،

تعريف : إذا كانت $x_1, x_2, \ldots, x_n$ تمثل فئة من المشاهدات المرتبة تصاعديا (أو تنازليا) فإن الوسيط لهذه الفئة هو العدد $x_{(\frac{n+1}{2})}$ إذا كان n فرديا وهو العدد $\frac{1}{2}[x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}]$ إذا كانت n زوجيا .

مثال (٥-٩) أوجد الوسيط للمجتمع الذي مشاهداته 10,9,8,6,7 .

الحل . بترتيب المشاهدات تصاعديا أي 6,7,8,9,10 فإن وسيط المجتمع يكون $\tilde{\mu} = 8$.

مثال (٥-١٠) أوجد الوسيط للعينة التي مشاهداتها هي 10,9,6,1,2,7 .

الحل . بترتيب البيانات تصاعديا أي 1,2,6,7,9,10 فإن وسيط العينة يكون :

$$\tilde{x} = \frac{6+7}{2} = \frac{13}{2} = 6.50$$

ومن الخصائص المميزة للوسيط أن مجموع القيم المطلقة لانحرافات القيم عن الوسيط أقل ما يمكن. أي أن ، $\sum_{i=1}^{n}|x_i - \tilde{x}|$ أقل ما يمكن . القيمة المطلقة للرقم هي القيمة بدون إشارة، وعلى ذلك فإن 5=|5-| .

من مميزات الوسيط أنه سهل الفهم ولا يتأثر بالقيم الشاذة.كما يمكن استخدامه في التوزيعات التي تحتوى على فئات مفتوحة .

ومن عيوب الوسيط أن كل المشاهدات لا تدخل في حسابه . كما أنه مثل الوسط الحسابي ، في بعض الأحيان ، يكون قيمة صناعية **artificial** ، بمعنى عدم وجود قيمة في فئة المشاهدات ، في الحقيقة ، تمثل الوسيط.

تعريف : لأى توزيع تكراري فإن الفئة التي تحتوى على الوسيط تسمى الفئة الوسيطة **median class** .

مثال (٥-١١) أوجد الوسيط للتوزيع التكراري المعطى في جدول (٥-١٢) .

الحل . لحساب الوسيط نفترض أن البيانات تتوزع بانتظام داخل فئة الوسيط . ومن جدول (٥-١٢) نحصل على جدول (٥-١٣) . لحساب الفئة الوسيطة نحسب موقع الوسيط وهو من جدول (٥-١٣) يساوى $50 = \frac{100}{2} = \frac{n}{2}$ (حيث $\sum_{i=1}^{k} f_i = n$) سواء كان مجموع التكرارات فرديا أو زوجيا . من جدول (٥-١٣) وفي عمود التكرار المتجمع نجد أن 62 مشاهدة قيمهم أقل من 69.5 . أيضا من عمود التكرار المتجمع نجد أن 39 مشاهدة قيمهم أقل من 59.5 . أي أن الوسيط لابد أن يقع في الفئة الرابعة، أي أن فئة الوسيط هي -59.5

69.5، وبما أن الوسيط هو القيمة التي ترتيبها الـ 11 =39-50 داخل الفئة الوسيطية حيث 50 تمثل ترتيب الوسيط و 39 تمثل التكرار المتجمع السابق لفئة الوسيط، وتحت فرض أن 23 مشاهدة تمثل تكرار الفئة الوسيطة موزعة بانتظام على الفئة الوسيطة التي طولها 10 وعلى ذلك تكون المسافة من بداية الفئة الوسيطة وموقع الوسيط هي $\frac{11}{23}\cdot 10 = 4.783$ ،

جدول (٥- ١٣)

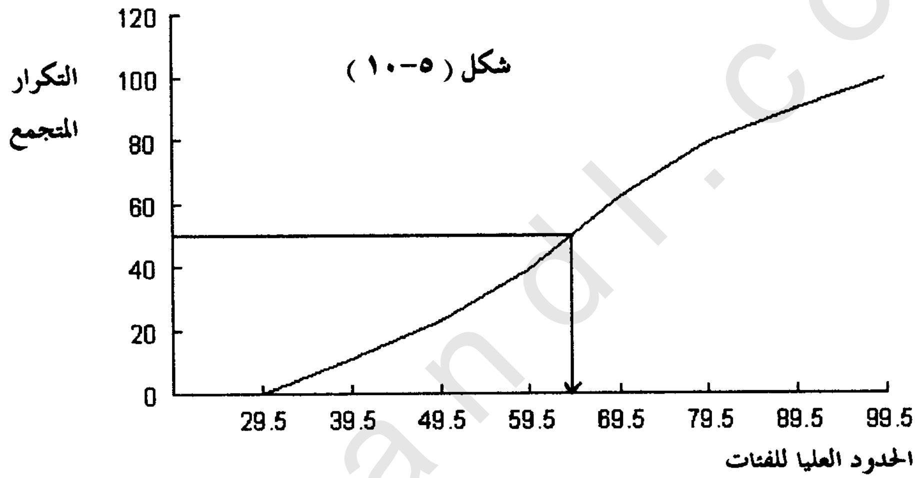
التكرار المتجمع	الحدود العليا للفئات	التكرار f_i	الحدود الفعلية
11	أقل من 39.5	11	29.5-39.5
23	أقل من 49.5	12	39.5-49.5
39	أقل من 59.5	16	49.5-59.5
62	أقل من 69.5	23	59.5-69.5
79	أقل من 79.5	17	69.5-79.5
90	أقل من 89.5	11	79.5-89.5
100	أقل من 99.5	10	89.5-99.5
		100	المجموع

أي أن الوسيط هو $\tilde{x} = 59.5 + 4.783$ أي $\tilde{x} = 64.283$ ، وعلى ذلك يمكن القول أن 50% من الأسماك أطوالها تكون أقل من 64.283 ، ولأن الكثيرون يفضلون حساب الوسيط من صيغة فسوف نقدم لهم الصيغة التالية لحساب الوسيط :-

$$\tilde{x} = L + \left(\frac{\frac{n}{2} - F}{f_m} \right) \Delta.$$

حيث : L =الحد الأدنى الفعلي لفئة الوسيط ، F =التكرار المتجمع السابق لفئة الوسيط ، Δ =طول فئة الوسيط ، f_m = تكرار فئة الوسيط.

يمكن إيجاد الوسيط بالرسم من المضلع التكراري المتجمع. فعلى سبيل المثال للتوزيع التكراري لجدول (٥-١٣) نرسم المضلع التكراري المتجمع كما في شكل (٥-١٠) ويتم تحديد موقع الوسيط على المحور الرأسي ثم نسقط عمود من نقطة موقع الوسيط على المضلع التكراري المتجمع وعند التقائه بالمضلع نسقط عمود على المحور الأفقي فتكون هي قيمة الوسيط. من شكل (٥-١٠) فإن الوسيط يكون $\tilde{x} = 64.5$.



$\tilde{x} = 64.5$

(٥-٤-٣) المنوال **Mode**

يعرف المنوال بأنه القيمة الأكثر شيوعا أو التي تتكرر أكثر من غيرها. في بعض الأحيان لا يوجد منوال لفئة من المشاهدات حيث لا تتكرر القيم أكثر من مرة ، وإذا وجد قد لا يكون وحيدا. على سبيل المثال المنوال للمشاهدات 3,5,5,5,5,5,5,5,7,7,9 هو 5 .

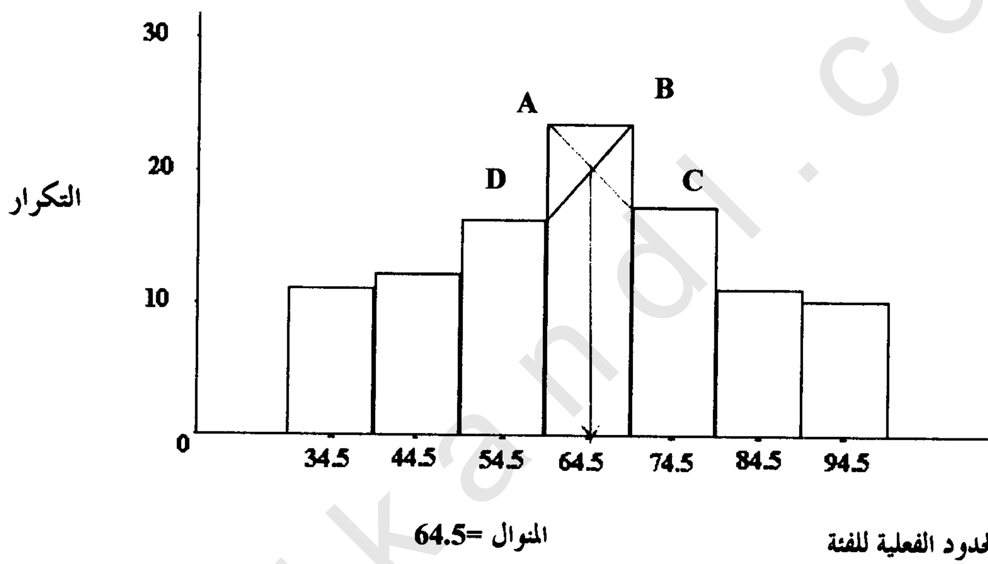
مثال (٥-١٢) أوجد المنوال للمشاهدات 2,4,5,6,6,6,6,6,7,7,7,7,7,8,9 .

الحل. التوزيع في هذه الحالة يكون ثنائي المنوال حيث المنوال هو 6,7 .

يعتبر المنوال أقل مقاييس النزعة المركزية استخداما. للفئات الصغيرة من البيانات لا يكون له فائدة ، فقط يكون له معنى إذا كان حجم البيانات كبيرا. ومن مميزاته أنه لا يحتاج إلى عمليات حسابية، كما يمكن استخدامه للبيانات الوصفية. ويمكن حساب المنوال بالرسم من المدرج التكراري. فعلى سبيل المثال للتوزيع التكراري لجدول (٥-١٣) نرسم المدرج التكراري كما في شكل (٥-١١). نصل الرأس الأيمن العلوي للمستطيل الذي يمثل أكبر

تكرار بالرأس الأيمن للمستطيل السابق له (BD كما في الشكل) • أيضا نصل الرأس الأيسر العلوي لأطول مستطيل بالرأس الأيسر العلوي للمستطيل التالي له (AC) فيتقاطع المستقيمان في نقطة، نسقط عمود من هذه النقطة على المحور الأفقي فيكون هو المنوال • من شكل (٥-١١) المنوال يساوى 64.5 •

شكل (٥-١١)



في حالة البيانات التكرارية فإن المنوال لدالة كثافة الاحتمال $f(x)$ هو قيمة x التي عندها يأخذ المنحنى أعلى قيمة • وعلى ذلك يمكن اعتبار مركز الفئة التي يقابلها أعلى تكرار هو تقدير للمنوال • وعلى ذلك للتوزيع التكراري في جدول (٥-١٢) فإن المنوال التقريبي هو 64.5 •

(٥-٤-٤) الوسط الهندسي <u>The geometric Mean</u>

في الحقيقة ، فإن الوسط الهندسي له استخدامات خاصة في المشاكل الاقتصادية وفي المجال السكاني •

تعريف : إذا كان لدينا الفئة من المشاهدات $x_1, x_2, \ldots, x_n$ ، فإن الوسط الهندسي يمكن حسابه من الصيغة التالية :-

$$G = \sqrt[n]{x_1 \cdot x_2 \cdot \ \ldots \ \cdot x_n}.$$

ولتسهيل حساب الوسط الهندسي تستخدم الصيغة التالية إذا كان $n > 2$:-

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

مثال (٥-١٣) حصل مستثمر على عائد من رأسماله المستثمر قدره 2% للسنة الأولى، 3% للسنة الثانية، 5% للسنة الثالثة أوجد الوسط الهندسي للعائد السنوى.

الحل .

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

$$\log G = \frac{\sum_{i=1}^{n} \log x_i}{n} = \frac{1}{3}(\log 2 + \log 3 + \log 5)$$

$$\frac{1}{3}(0.3010 + 0.4771 + 0.6990)$$

$$= \frac{1.4771}{3} = 0.49237.$$

وعلى ذلك فإن $G = 3.1072$.

دائما يكون الوسط الهندسي أصغر من الوسط الحسابى. كما أن الوسط الهندسي لا يمكن حسابه إذا كانت إحدى القيم مساوية للصفر أو رقم سالب.

يمكن حساب الوسط الهندسي من جداول تكرارية من التعريف التالى.

تعريف : إذا كانت $x_1, x_2, \ldots, x_k$ تمثل مراكز الفئات لتوزيع تكراري مع تكراراها المقابلة $f_1, f_2, \ldots, f_k$ (حيث k تمثل عدد الفئات) فإن الوسط الهندسي G يحسب من الصيغة التالية :

$$\text{Log}G = \frac{\sum_{i=1}^{k} f_i \log x_i}{\sum_{i=1}^{k} f_i}.$$

مثال (٥-١٤) البيانات في جدول (٥-١٤) تمثل التوزيع التكراري لأطوال عينة من خمسين نبات من نوع ما والمطلوب إيجاد الوسط الهندسى.

الحل . من جدول (٥-١٤) فإن :-

$$LogG = \frac{\sum_{i=1}^{k} f_i \log x_i}{\sum_{i=1}^{k} f_i} = \frac{55.1799}{50} = 1.103598.$$

وعلى ذلك فإن الوسط الهندسي هو $G = 12.69399$ ٠

جدول (٥-١٤)

حدود الفئة	التكرار f_i	مركز الفئة x_i	$\log x_i$	$f_i \log x_i$
6-8	6	7	0.8451	5.0706
9-11	10	10	1.0000	10.0000
12-14	15	13	1.1139	16.7085
15-17	12	16	1.2041	14.4492
18-20	7	19	1.2788	8.9516
المجموع	50			55.1799

(٥-٥) الربيعات والمئينات والعشيرات

Quartiles, Percentiles, Deciles

كما ذكرنا سابقا، إذا رتبنا فئة من المشاهدات حسب قيمها تصاعديا فإن القيمة التي تكون في المنتصف والتي تقسم البيانات إلى قسمين متساويين في العدد هي الوسيط ٠ وبتعميم الفكرة وتقسيم البيانات إلى أربعة أجزاء متساوية (بعد ترتيب المشاهدات تصاعديا) فإن نقاط التقسيم يرمز لها بالرموز Q_3 ،Q_2 ، Q_1 حيث Q_1 يسمى الربيع الأول **first quartile** (الربيع الأدنى **lower quartile**) و Q_2 يسمى الربيع الثاني **second quartile** (نفسه الوسيط) و Q_3 يسمى الربيع الثالث **third quartile** (الربيع الأعلى **upper quartile** فالربيع الأول هو القيمة Q_1 التي يسبقها ربع البيانات ويليها ثلاثة أرباع البيانات ٠ والربيع الثاني (وهو أيضا الوسيط) هو القيمة Q_2 التي يسبقها نصف البيانات ويليها نصف البيانات ٠ وفي النهاية الربيع الثالث وهو القيمة Q_3 التي يسبقها ثلاثة أرباع البيانات ويليها ربع البيانات ٠ عند استخدام فئة من المشاهدات فإن الربيعات الثلاثة يتم حسابها بتعين موقعها أولا فموقع الربيع الأول هو $\frac{n+2}{4}$ والربيع الثاني موقعه هو $\frac{n+1}{2}$ والربيع الثالث موقعه هو $\frac{3n+2}{4}$ ٠

للمشاهدات في شكل (٥-١٢) موقع الربيع الأول هو $3.5 = \frac{12+2}{4}$ إما قيمتــه فهي متوسط القيمة الثالثة والرابعة أي $6 = \frac{5+7}{2} = Q_1$ • أيضا موقع الربيع الثاني والثــالث هما على التوالي $6.5 = \frac{12+1}{2}$ ، $9.5 = \frac{36+2}{4}$ وعلى ذلك فإن قيمة الربيع الثاني والثــالث على التوالي هما $12 = \frac{11+13}{2} = Q_2$ ، $18 = \frac{17+19}{2} = Q_3$ •

شكل (٥-١٢)

19 20 22	13 16 17	7 10 11	1 4 5

عند استخدام التوزيعات التكرار يه، فإن قيمة الربيــع الأول والثــالث تحسـب مــن المعادلتين:-

$$Q_3 = L + \left(\frac{\frac{3n}{4} - F}{f_{Q_3}}\right)\Delta \quad ، \quad Q_1 = L + \left(\frac{\frac{n}{4} - F}{f_{Q_1}}\right)\Delta.$$

حيث L = الحد الأدنى الفعلي لفئة الربيع و F =التكرار المتجمع الســابق لفئــة الربيــع و Δ =طول فئة الربيع و f_Q =تكرار فئة الربيع•

مثال (٥-١٥) أوجد الربيع الأول والثالث للبيانات في جدول (٥-١٣)•

الحل• موقع الربيع الأول هو $25 = \frac{100}{4} = \frac{n}{4}$ وفئة الربيع الأول هي **49.5-59.5** وقيمــة الربيع الأول هو :-

$$Q_1 = L + \left(\frac{\frac{n}{4} - F}{f_{Q_1}}\right)\Delta = 49.5 + \frac{(25-23)}{16} \cdot 10 = 49.5 + 1.25 = 50.75.$$

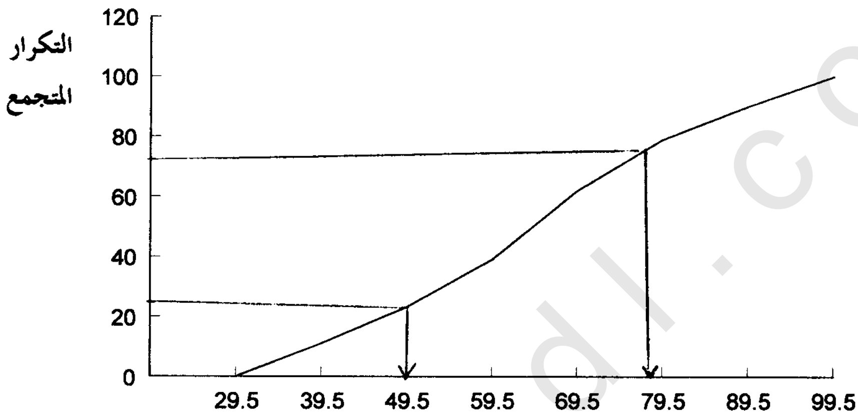
بنفس الشكل موقع الربيع الثالث $75 = \frac{(3)(100)}{4} = \frac{3n}{4}$ وفئة الربيع الثالث هــي - **69.5** **79.5** وقيمة الربيع الثالث هو :-

$$Q_3 = L + \left(\frac{\frac{3n}{4} - F}{f_{Q_3}}\right)\Delta = 69.5 + \frac{(75-62)}{17} \cdot 10 = 69.5 + 7.6471 = 77.1471.$$

ويمكن الحصول على الربيع الأول والثالث بيانيا من المضلع التكراري المتجمع بنفــس

الطريقة التي استخدمت في حساب الوسيط بيانيا مع استخدام موقع الربيع بـدلا مـن موقـع الوسيط . باستخدام البيانات المعطاة في جدول (٥-١٣) . فإن قيم الربيع الأول والثالث بيانيا هي على التوالي , 49.5 و 77.0 وذلك من شكل (٥-١٣) .

شكل (٥-١٣)



$Q_1 = 49.5 \quad Q_3 = 77$

أيضا يمكن إيجاد القيم التي تقسم فئة من المشاهدات (بعد ترتيبها تصاعديا) إلى عشرة أقسـام ونرمز لنقط التقسيم بالرموز $D_1, D_2, \ldots, D_9$ حيث D_1 العشير الأول وهو يمثل القيمة التي يسبقها $\frac{1}{10}$ من البيانات ويليها $\frac{9}{10}$ من البيانات و D_2 العشير الثاني وهو يمثل القيمة الـتي يسبقها $\frac{2}{10}$ من البيانات ويليها $\frac{8}{10}$ من البيانات وهكذا للعشيرات الأخرى . بنفس الشـكل يمكن إيجاد القيم التي تقسم فئة من المشاهدات (بعد ترتيبها تصاعديا) إلى مائة قسم ونرمز لنقـط التقسيم بالرموز $P_1, P_2, \ldots, P_{99}$ حيث P_1 المئين الأول هو يمثل القيمة التي يسبقها $\frac{1}{100}$ من البيانات ويليها $\frac{99}{100}$ من البيانات و P_2 المئين الثاني وهو يمثل القيمة التي يسبقها $\frac{2}{100}$ من البيانات ويليها $\frac{98}{100}$ من البيانات وهكذا لباقي المئينات . في حالة التوزيعات التكرارية يمكـن حساب العشيرات و المئينات بنفس طريقة حساب الوسيط مع استبدال $\frac{n}{2}$ بـــ $\frac{n}{10}$ للعشـير

الأول و $\frac{2n}{10}$ للعشير الثاني وهكذا لباقي العشيرات . أيضا استبدال $\frac{n}{2}$ بـــ $\frac{n}{100}$ للمئين الأول و $\frac{2n}{100}$ للمئين الثاني وهكذا الباقي المئينات .

مثال (٥-١٦) أوجد العشير الأول والمئين التسعين للبيانات في جدول (٥-١٣).

الحل. موقع العشير الأول هو $10 = \frac{100}{10} = \frac{n}{10}$ وفئة العشير الأول هي 29.5-39.5 وقيمة العشير الأول هو :-

$$D_1 = 29.5 + \frac{(10-0)}{11} \cdot 10 = 29.5 + 9.091 = 38.591.$$

حيث 29.5 = الحد الأدنى الفعلي لفئة العشير الأول و0 = التكرار المتجمع السابق لفئة العشير و 10= طول فئة العشير الأول و 11= تكرار فئة العشير الأول.

بنفس الشكل موقع المئين التسعين $90 = \frac{(90)(100)}{100} = \frac{90n}{100}$ وفئة المئين التسعين هي 79.5 -89.5- وقيمة المئين التسعين هو :-

$$P_{90} = 79.5 + \frac{(90-79)}{11} \cdot 10 = 79.5 + 10 = 89.5.$$

حيث 79.5= الحد الأدنى الفعلي لفئة المئين التسعين و 79= التكرار المتجمع السابق لفئة المئين التسعين و 10= طول فئة المئين التسعين و 11= تكرار فئة المئين التسعين.

يمكن الحصول على العشيرات والمئينات بيانيا من المضلع التكراري المتجمع بنفس الطريقة التي استخدمت في حساب الوسيط بيانيا . ومما يجدر الإشارة إليه أن $Q_2 = D_5 = P_{50}$. أيضا كل الربيعات والعشيرات تمثل مئينات فعلى سبيل المثال D_8 هو P_{80} و Q_3 هو P_{75} .

(٥-٦) مقاييس التشتت Measures of Dispersion

مقاييس النزعة المركزية التي تمت مناقشتها في البند السابق لا تكفى لإعطاء وصف كافي لتوزيع فئة من المشاهدات فلا توضح طبيعتها ولا كيفية توزيع مشاهداتها . كما أن الاعتماد فقط على أي مقياس للنزعة المركزية لمقارنة عدة مجموعات لا يكفى لإظهار حقيقة المقارنة، فمن الممكن أن يكون لعدة مجموعات من البيانات نفس الوسط الحسابي والوسيط ولكنهم يختلفوا عن بعضهم تمام الإختلاف . فقد تكون مشاهدات إحدى المجموعات متقاربة بعضها من بعض (متمركزة حول متوسطها) أو مبعثرة (متشتتة) . فعلى سبيل المثال في جدول (٥-١٥) ثلاثة فئات من المشاهدات A_1, A_2, A_3 كل فئة لها نفس الوسط الحسابي(60) والوسيط(60)

، ولكن يختلفوا في التشتت أو الإنتشار ، في الفئة A_1 القيم الستة لهم نفس القيمة ولا يوجد أي تشتت فهم متجانسين تماما ، في الفئة A_2 القيم تختلف من خلال ثلاث قيم وفي الفئة A_3 القيم تختلف من خلال قيمتين ولكن هناك انتشار أكثر في الفئة A_3 حيث لا يوجد قيمة في الفئة A_3 تقترب من المتوسط الحسابي أو الوسيط ، ومن هنا كان من الضروري عند وصف فئة من البيانات بمقياس رقمي أن نَصِفها عن طريق مقياس من مقاييس النزعة المركزية ومقياس آخر يقيس بعد البيانات عن بعضها أو بعدها عن المتوسط ، بعبارة أخرى نصف درجة تشتتها ، في الجزء التالي سوف نقدم بعض مقاييس التشتت الأكثر أهمية ،

جدول (٥-١٥)

A_1	A_2	A_3
60	35	0
60	35	0
60	60	0
60	60	120
60	85	120
60	85	120

(٥-٦-١) المدى ونصف المدى الربيعي

The range and semi interquartile range

المدى لفئة من المشاهدات، هو الفرق بين أكبر وأصغر مشاهدة ، ففي جدول(٥-١٥) المدى للفئة A_1 هو صفر، والمدى للفئة A_2 هو 50 والمدى للفئة A_3 هو 120 ، وعلى ذلك يعتبر المدى مقياس للتشتت من السهل جدا حسابه ويعطى فكرة سريعة جدا عن طبيعة البيانات ويستخدم كثيرا في مراقبة الجودة وكذلك في وصف الأحوال الجوية ، ولكن من عيوبه أنه يتأثر بالقيم الشاذة ويعطى معلومات خاطئة عن الانتشار الحقيقي لمعظم البيانات كما أنه لا يستخدم جميع البيانات في حسابه ،

عند استخدام التوزيعات التكرارية فإن المدى يحسب بعدة طرق سوف نذكر منها الطريقة التالية :

المدى= الحد الأعلى الفعلي للفئة الأخيرة-الحد الأدنى الفعلي للفئة الأولى ،

للبيانات في جدول (٥-١٣) المدى = 70=99.5-29.5 ،

هناك مقاييس أخرى للتشتت يمكن استخدامها بدلا من المدى في حالة وجود قيم شاذة ، تعتمد هذه المقاييس على إهمال جزء من البيانات عند طرفي التوزيع حتى نتخلص من القيم الشاذة وتسمى شبيهات المدى ، فمثلا بحذف أعلى 10% من المشاهدات وأصغر 10% منها

نحصل على المدى المثيني أي $P_{90} - P_{10}$ ، هذا المقياس يستخدم في اختبارات الذكاء في مجال التربية وعلم النفس، أيضا بحذف أعلى 25% من البيانات وأصغر 25% منها نحصل على المدى الربيعي أي $Q_3 - Q_1$ ، وأخيرا هناك مقياس آخر يستنتج من المدى الربيعي وهو نصف المدى الربيعي (الانحراف الربيعي) **semi interquartile range** ونحصل عليه بقسمة المدى الربيعي على 2 فإذا رمزنا له بالرمز MR فإن :-

$$MR = \frac{Q_3 - Q_1}{2}.$$

وعلى ذلك فإن نصف المدى الربيعي لفئة المشاهدات في شكل (٥-١٢) هو :-

$$MR = \frac{Q_3 - Q_1}{2} = \frac{18 - 6}{2} = 6.$$

أيضا نصف المدى الربيعي للتوزيع التكراري في جدول (٥-١٣) هو :-

$$MR = \frac{Q_3 - Q_1}{2} = \frac{77.1471 - 50.75}{2} = 13.198550.$$

يعاب على نصف المدى الربيعي أنه لا يستعمل جميع مشاهدات العينة في حسابه ولتلافي هذا العيب سوف نقدم في البند التالي بعض المقاييس الأخرى للتشتت التي تستخدم جميع البيانات في حسابها .

(٥-٦-٢) الانحراف المتوسط The average Deviation

تمثل $|x_i - \mu|$ أو $|x_i - \overline{x}|$ القيمة المطلقة لانحراف أي قيمة عن الوسط الحسابي للمجتمع أو العينة على التوالى .

تعريف : إذا كانت لدينا الفئة من المشاهدات $x_1, x_2, \ldots, x_n$ ، فإن الانحراف المتوسط يمكن حسابه من الصيغة التالية :-

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

مثال (٥-١٧) أوجد الانحراف المتوسط لفئة المشاهدات : 2,3,5,7,8

الحل . من القيم نجد أن $\overline{x} = 5$ والانحرافات : 3,-2,0,2,3- والقيم المطلقة: 3,2,0,2,3 وعلى ذلك:

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n} = \frac{3 + 2 + 0 + 2 + 3}{5} = 2.$$

تعريف : إذا كانت $x_1, x_2, \ldots, x_k$ تمثل مراكز الفئة لتوزيع تكراري مع تكراراتها المقابلة $f_1, f_2, \ldots, f_k$ فإن الانحراف المتوسط هو :-

$$M.D = \frac{\sum_{i=1}^{k} f_i |x_i - \overline{x}|}{\sum_{i=1}^{k} f_i}.$$

مثال (٥-١٨) أوجد الانحراف المتوسط للبيانات في جدول (٥-١٦)٠

جدول (٥-١٦)

مركز الفئة x_i	التكرار f_i	$x_i - \overline{x}$	$f_i \lvert x_i - \overline{x} \rvert$
34.5	11	-29.6	325.6
44.5	12	-19.6	235.2
54.5	16	-9.6	153.6
64.5	23	0.4	9.2
74.5	17	10.4	176.8
84.5	11	20.4	224.4
94.5	10	30.4	304
المجموع	100		1428.8

الحل. الوسط الحسابي سبق حسابه من جدول (٥-١٢) وهو $\overline{x} = 64.1$ ٠وعلى ذلك فإن الانحراف المتوسط من جدول (٥-١٦) هو :-

$$M.D = \frac{\sum_{i=1}^{k} f_i |x_i - \overline{x}|}{\sum_{i=1}^{k} f_i)} = \frac{1428.8}{100} = 14.288.$$

(٥-٦-٣) التباين **The Variance**

واحد من خصائص الوسط الحسابي ، كما ذكرنا سابقا ، هو أن مجموع مربعات انحرافات قيم المشاهدات عن الوسط الحسابي للعينة ، أي $\sum_{i=1}^{n}(x_i - \overline{x})^2$ ، أقل ما يمكن٠ وإذا كان اهتمامنا بالمجتمع فإن المقدار $\sum_{i=1}^{N}(x_i - \mu)^2$ أيضا يكون أقل ما يمكن، وعلى ذلك إذا أخذنا هذا المقدار الأخير وقسمناه على حجم المجتمع فإننا نحصل على تباين المجتمع٠

تعريف : إذا أعطيت مجتمع محدود $x_1, x_2, \ldots, x_N$ فإن تباين المجتمع هو:-

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

الانحراف المعياري، ويرمز له بالرمز σ يمكن الحصول عليه بأخذ الجذر التربيعي للتباين.

مثال (٥-١٩) أوجد الانحراف المعياري لمشاهدات المجتمع 3,4,5,5,6,7 .

الحل. لتسهيل العمليات الحسابية يمكن استخدام جدول (٥-١٧) للحصول على الوسط الحسابي كما يأتي:-

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{30}{6} = 5.$$

أيضا التباين:-

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{6} = \frac{10}{6} = \frac{5}{3}.$$

والانحراف المعياري هو : $\sigma = \sqrt{\frac{5}{3}} = 1.29$.

جدول (٥-١٧)

x_i	$(x_i - \mu)$	$(x_i - \mu)^2$
3	-2	4
4	-1	1
5	0	0
5	0	0
6	1	1
7	2	4
30	0	10

تباين العينة ، يرمز له بالرمز s^2، ويمثل قيمة من قيم الإحصاء S^2 . عينات عشوائية مختلفة من الحجم n مختارة من نفس المجتمع، عموما ، تؤدي إلى قيم مختلفة من s^2. في كثير من المشاكل

قيمة σ^2 تكون غير معروفة وتقدر بالقيمه s^2. لكي يكون تقديرنا جيد لابد من حساب التباين من صيغة بحيث في المتوسط ينتج القيمة الحقيقية σ^2. بمعنى أننا إذا أخذنا كل العينات الممكنة من الحجم n من المجتمع وحصلنا على قيمة s^2 لكل عينة، فإن متوسط كل قيم s^2 لابد أن تساوي σ^2. الإحصاء الذي يحقق هذا الشرط يسمى إحصاء غير متحيز **unbiased**.

تعريف : إذا سحبت العينة العشوائية $x_1, x_2, \ldots, x_n$ فإن تباين العينة الغير متحيز هو :-

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

والانحراف المعياري للعينة هو $s = \sqrt{s^2}$

مثال (٥-٢٠) أوجد تباين العينة والانحراف المعياري للعينة التي مشاهداتها 12,15,17,20 .

الحل . التباين والانحراف المعياري للعينة يمكن حسابه من جدول (٥-١٨).

جدول (٥-١٨)

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
12	-4	16
15	-1	1
17	1	1
20	4	16
64	0	34

حيث :-

الوسط الحسابي:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{64}{4} = 16.$$

وتباين العينة هو :-

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{34}{3} = 11.33.$$

والانحراف المعياري للعينة هو $\sqrt{11.33} = 3.37$.

هناك صيغة أخرى لحساب تباين العينة تفيد عند استخدام الآلة الحاسبة وهى :-

$$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right].$$

مثال (٥-٢١) أوجد التباين والانحراف المعياري للعينة التي مشاهداتها هي 3,4,5,6,6,7

الحل، بما أن : $\sum_{i=1}^{n} x_i = 31 \; , \; \sum_{i=1}^{n} x_i^2 = 171$ وعلى ذلك :-

$$s^2 = \frac{1}{5}[171 - \frac{(31)^2}{6}] = \frac{13}{6} = 2.1667.$$

الانحراف المعياري للعينة نحصل عليه بإيجاد الجذر التربيعي لتباين العينة، أي أن

$$s = \sqrt{\frac{13}{6}} = 1.47196.$$ ،

يمتاز الانحراف المعياري عن التباين بأنه يعبر عنه باستخدام نفس وحدات القياس كما في البيانات بينما يكون تمييز التباين " وحدات القياس مربعة "،

عرفنا من نظرية (٤-٤) أن $\sigma^2_{X+b} = \sigma^2$ وهذا يعنى أن التباين لفئة من المشاهدات لا يتأثر إذا أضفنا ثابت أو طرحنا ثابت من كل مشاهدة، أيضا عرفنا من نظرية (٤-٥) أن $\sigma^2_{aX} = a^2\sigma^2$ وهذا يعنى أنه إذا ضربنا كل مشاهدة في ثابت (أو قسمنا على ثابت) فإن التباين الأصلي نحصل عليه من التباين الجديد بقسمته على (أو ضربه في) مربع الثابت،

في حالة التوزيعات التكرارية فإن تباين العينة يحسب من المعادلة التالية :-

$$s^2 = \frac{1}{n-1}[\sum_{i=1}^{k} x_i^2 f_i - \frac{(\sum_{i=1}^{k} x_i f_i)^2}{n}].$$ (حيث $\sum_{i=1}^{k} f_i = n$)

مثال (٥-٢٢) أوجد التباين للمشاهدات في جدول (٥-١٩) ،

الحل، بالتعويض في صيغة التباين السابقة فإن :-

$$s^2 = \frac{1}{99}[442265 - \frac{(6410)^2}{100}] = 317.0101.$$

جدول (٥-١٩)

حدود الفئة	مركز الفئة x_i	التكرار f_i	$x_i f_i$	$x_i^2 f_i$
30-39	34.5	11	379.5	13092.75
40-49	44.5	12	534.0	23763.00
50-59	54.5	16	872.0	47524.00
60-69	64.5	23	1483.5	95685.75
70-79	74.5	17	1266.5	94354.25
80-89	84.5	11	929.5	78542.75
90-99	94.5	10	945.0	89302.50
المجموع		100	6410	442265

نظرية شيبيشي **Chebyshev's theorem**

هذه النظرية مهمة في وصف فئة من المشاهدات ، إذا كانت $x_1, x_2, \ldots, x_n$ فئة من المشاهدات فإنه على الأقل $(1-\frac{1}{k^2})$ من المشاهدات سوف تقع ضمن k انحرافات معيارية من وسطها الحسابي ، تعتبر النظرية غير مفيدة عندما k تساوى واحد لأن هذا يعنى أنه على الأقل $0=(1-\frac{1}{1^2})$ من المشاهدات تقع في الفترة من $\mu-\sigma$ إلى $\mu+\sigma$ ، عندما $k=2$ فهذا يعنى أنه على الأقل $\frac{3}{4}=(1-\frac{1}{2^2})$ من المشاهدات تقع في الفترة من $\mu-2\sigma$ إلى $\mu+2\sigma$ ، وأخيرا عندما $k=3$ فهذا يعنى أنه على الأقل $\frac{8}{9}=(1-\frac{1}{2^3})$ من المشاهدات تقع في الفترة من $\mu-3\sigma$ إلى $\mu+3\sigma$ ، بفرض أن فئة من المشاهدات تمثل عينة، وإذا كانت **k=2** فالنظرية تنص على أنه على الأقل $\frac{3}{4}=(1-\frac{1}{2^2})$ من المشاهدات تقع في الفترة من $\bar{x}-2s$ إلى $\bar{x}+2s$ ، بنفس الشكل، النظرية تنص على أنه $\frac{8}{9}=(1-\frac{1}{2^3})$ من المشاهدات تقع في الفترة من $\bar{x}-3s$ إلى $\bar{x}+3s$ ،

مثال (٥-٢٣) إذا كان الوسط الحسابي والانحراف المعياري لعينة من المشاهدات هما على التوالي **60** و **10** ، أستخدم نظرية تشيبيشى لوصف توزيع المشاهدات ،

الحل ، (ا) على الأقل $\frac{3}{4}$ من المشاهدات تقع في الفترة 60 ± 20، أى من **40** إلى **80** ،

(ب) على الأقل $\frac{8}{9}$ من المشاهدات تقع في الفترة 60 ± 30، أى من **30** إلى **90** ،

مثال (٥-٢٤) المشاهدات التالية تمثل مشاهدات عينة :

14,18,19,19,19,19,20,20,20,20,21,25 أثبت أنه على الأقل $\frac{3}{4}$ من المشاهدات تقع ضمن أثنين انحراف معياري من الوسط الحسابي وأنه على الأقل $\frac{8}{9}$ من المشاهدات تقع ضمن ثلاثة انحراف معياري من الوسط الحسابي ،

الحل ، الوسط الحسابي والانحراف المعياري هما : $\bar{x}=19.5, \quad s=2.468$ وعلى ذلك $\bar{x}\pm 2s = 19.5 \pm (2)(2.468) = 19.5 \pm 4.94$ ، يوضح العد الحقيقي أن الفترة من

14.56 إلى 24.44 تحتوى على 10 مشاهدة من 12 مشاهدة، تمثل تقريبا 83% من المشاهدات، أي أكثر من $\frac{3}{4}$ القيم .أيضا

$\bar{x} \pm 3s = 19.5 \pm (3)(2.468) = 19.5 \pm 7.4 = 21.96$. يوضح العدد الحقيقي أن الفترة من 12.1 إلى 26.9 تحتوى على كل القيم في العينة وبالتأكيد تحتوى على الأقل على $\frac{8}{9}$ منهم.

الآن سوف نقدم قاعدة تصف بدقة الانتشار للتوزيع الناقوسى. يعرف التوزيع الناقوسى عادة بالتوزيع الطبيعي والذي سوف نناقشه بالتفصيل في الفصل السادس. كما تعطى هذه القاعدة نتائج جيدة للتوزيعات القريبة الشكل من التوزيع الطبيعى.

قاعدة تجريبية Empirical Rule

أعتبر توزيع فئة من المشاهدات لها التوزيع الناقوس، الفترة :-

(ا) $\bar{x} \pm s$ سوف تحتوى تقريبا على68% من المشاهدات.

(ب) $\bar{x} \pm 2s$ سوف تحتوى تقريبا على 95% من المشاهدات.

(جـ) $\bar{x} \pm 3s$ سوف تحتوى تقريبا على 99.7% من المشاهدات.

مثال (٥-٢٥) المشاهدات في جدول (٥-٢٠) تمثل التوزيع التكراري لدرجات مجموعة من الطلبة في مادة الإحصاء ، والمطلوب إيجاد التوزيع التكراري ووصف البيانات منه.

الحل. للبيانات في جدول (٥-٢١). الوسط الحسابي والانحراف المعياري على التوالي هما $\bar{x} = 66.1$ ، $s = 13.7$ وعلى ذلك يمكننا حساب الفترات:

(ا) $\bar{x} \pm s = 66.1 \pm 13.7$.

(ب) $\bar{x} \pm 2s = 66.1 \pm 27.4$.

(جـ) $\bar{x} \pm 3s = 66.1 \pm 41.1$.

جدول (٥-٢٠)

35	41	44	45	40	51	48
54	56	55	53	58	59	60
60	61	62	63	67	64	64
67	65	66	68	69	66	70
73	75	74	72	71	76	81
79	80	78	82	83	85	86
50	62	68	72	80	88	51
91	92					

جدول (٥-٢١)

حدود الفئة	التكرار
35-39	1
40-44	3
45-49	2
50-54	5
55-59	4
60-64	8
65-69	8
70-74	6
75-79	4
80-84	5
85-89	3
90-94	2

القاعدة التجريبية السابقة سوف تفيدنا كثيرا في وصف المشاهدات في جدول (٥-٢١) ٠ تبعا لهذه القاعدة فإننا نتوقع أن تقريبا 68% من المشاهدات تقع في الفترة 52.4 إلى 79.8٠ يوضح العدد الحقيقي من جدول (٥-٢٠) أن 34 من 51 مشاهدة تقع في هذه الفترة وتمثل 67.0% من المشاهدات٠ أيضا نتوقع أن تقريبا 95% من المشاهدات تقع في الفترة 38.7 إلى 93.5 ٠ يوضح العدد الحقيقي أن 49 من 51 مشاهدة تقع في هذه الفترة وتمثل 96% من المشاهدات٠ وأخيرا نتوقع أن تقريبا 99.7% من كل المشاهدات تقع في الفترة 25 إلى 107.2٠ العدد الحقيقي يوضح أن كل المشاهدات تقع في هذه الفترة وتمثل 100% من كل المشاهدات٠

(٥-٦-٤) معامل الاختلاف **Coefficient of Variation**

تعتبر كل مقاييس التشتت السابقة مقاييس مطلقة لأنها تأخذ تمييز الوحدات الأصلية ولذلك لا تصلح للمقارنة بين مجموعتين وحدات القياس بينهما مختلفة٠ لذلك سوف نناقش مقياس نسبي يسمى معامل الاختلاف والذي يحول الانحراف المعياري إلى مقياس نسبي باعتبار أنه نسبة مئوية من الوسط الحسابي٠ ويمكن حساب معامل الاختلاف من إحدى المعادلتين التاليتين :

$$V = (\frac{\sigma}{\mu})\,100 \quad \text{أو} \quad V = (\frac{S}{\bar{X}})\,100$$

لتسهيل استخدام معامل الاختلاف نقدم المثال التالي والذي يوضح كمية الإنتاج في شركة ما خلال 80 شهرا ثم كمية الإنتاج خلال فترة ثانية مقدارها 15 شهرا وقد تم حساب الوسط الحسابي والانحراف المعياري ومعامل الاختلاف لكل مجموعة والنتائج في جدول (٥-٢٢)٠ يلاحظ من النتائج أن الإنتاج خلال 15 شهرا له وسط حسابي أكبر ومعامل اختلاف أقل والذي يعتبر نتائج جيدة لمدير الإنتاج والذي يهتم بزيادة الإنتاج وانخفاض معامل الإختلاف٠ وبالرغم

من أن الانحراف المعياري قد زاد من 13.2 إلى 15 إلا أنه يمكن القول بناء على معامل الاختلاف أن الفترة الثانية أقل تشتتا من الفترة الأولى .

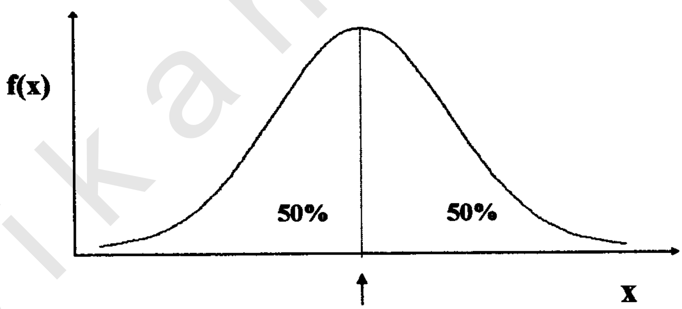
جدول (٥-٢٢)

الفترة	$\bar{x}$	s	$V = (\frac{s}{\bar{x}})\ 100$
80	125	13.2	10.56
15	160	15	9.375

(٥-٧) الالتواء والعلاقة بين الوسط الحسابي والوسيط والمنوال
Skewness and the Relation of the Mean , Median , and Mode

عرفنا مما سبق أن الالتواء هو بعد التوزيع التكراري عن التماثل . فإذا كان التوزيع متماثلا فسوف نجد أن 50% من القيم تقع على كل جانب من المنوال كما في شكل (٥-١٤) .

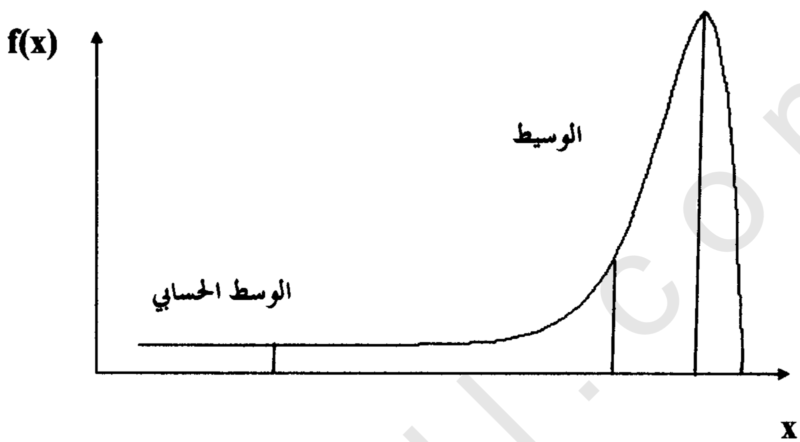
شكل (٥- ١٤)

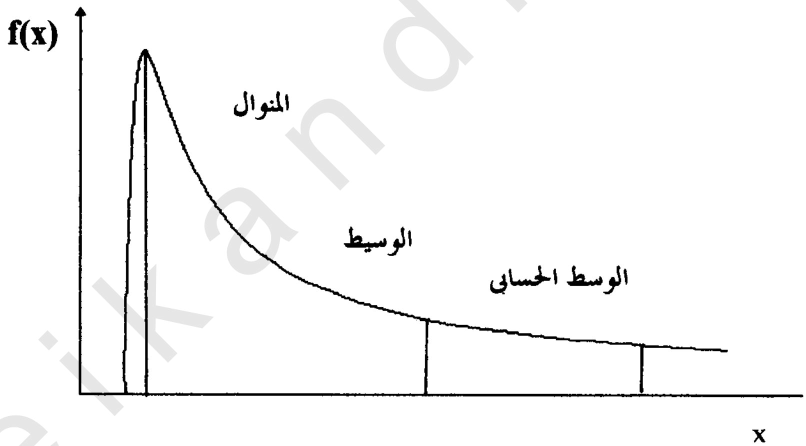


المنوال=الوسيط=الوسط الحسابي

أيضا نلاحظ من شكل (٥-١٤) أن التوزيع له منوال واحد **unimodal** (وحيد المنوال) وأن الوسط الحسابي = الوسيط= المنوال . بينما في شكل (٥-١٥) نجد أن هناك علاقة بين الوسط الحسابي والوسيط والمنوال حيث الوسط الحسابي < الوسيط < المنوال وذلك لأن التوزيع ملتويا جهة اليسار . بينما في شكل (٥-١٦) نجد أن الوسط الحسابي > الوسيط> المنوال وذلك لأن التوزيع ملتويا جهة اليمين . وفى كلتا الحالتين فإن الوسيط يقع بين الوسط الحسابي والمنوال كما أن الوسط الحسابي يقع دائما في اتجاه القيم الشاذة .

شكل (٥-١٥)





شكل (٥-١٦)

(٥-٨) بعض مقاييس الالتواء ولتفلطح

Some Measures of Skewness and Kurtosis

أولا بالنسبة لمقاييس الالتواء ، سوف نتناول مقياسين للالتواء الأول ويسمى معـــامل بيرســـون للالتواء **Pearsonian coefficient for skewness**. تعرف معادلـــة معـــامل بيرسون للالتواء كالتالي :-

$$S_k = \frac{3(\bar{x} - \tilde{x})}{s}.$$

حيث $\overline{x}$ الوسط الحسابي و $\tilde{x}$ الوسيط و s الانحراف المعياري للعينة. ينحصر قيمة معامل بيرسون بين -3 إلى $+3$ ، عندما $S_k = 0$ فهذا يعنى أن التوزيع متماثل، وإذا كانت قيمة S_k موجبة فهذا يعنى أن الوسط الحسابي أكبر من الوسيط ومن المنوال وبذلك يكون المنحنى ملتويا وله ذيل ناحية اليمين ويكون الالتواء موجبا، وأخيرا وإذا كانت قيمة S_k سالبة فهذا يعنى أن الوسط الحسابي أصغر من الوسيط ومن المنوال وبذلك يكون المنحنى ملتويا وله ذيل ناحية اليسار ويكون الالتواء سالبا ،من جدول (٥-٢٢) وإذا كان الوسيط للبيانات في الفترة التي تمثل 80 شهرا هي **115.3** فإن معامل الالتواء هو :-

$$S_k = \frac{3(\overline{x} - \tilde{x})}{s} = \frac{3(125 - 115.3)}{13.2} = +2.2045.$$

وهذا يعنى أن التوزيع به كمية من الالتواء الموجب.

يعتمد المقياس السابق للالتواء على أنه في التوزيعات الملتوية فإن الوسيط يقع تقريبا في $\frac{1}{3}$ المسافة بين الوسط الحسابي والمنوال في اتجاه الوسط الحسابي كما في شكل (٥-١٥) وشكل (٥-١٦) وهذا غير صحيح دائما، ولذلك سوف نتناول مقياس آخر للالتواء يعتمد على العزم المقدر من بيانات العينة.

تعريف : العزم r حول المتوسط لفئة المشاهدات $x_1, x_2, \ldots, x_n$ هو :-

$$m^r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^r}{n}.$$

في حالة التوزيعات التكرارية وإذا كانت $x_1, x_2, \ldots, x_k$ تمثل مراكز الفئات لتوزيع تكراري مع تكراراتها المقابلة $f_1, f_2, \ldots, f_k$ (حيث k تمثل عدد الفئات) فإن العزم r يحسب من الصيغة التالية :-

$$m^r = \frac{\sum_{i=1}^{k} f_i (x_i - \overline{x})^r}{\sum_{i=1}^{k} f_i}.$$

المقياس الثاني للالتواء و الذي يعتمد على العزم الثالث حول الوسط الحسابي هو:

$$a_1 = \frac{m^3}{s^3}.$$

إذا كان التوزيع متماثل ، فهذا يدل على أن $a_1 = 0$ وإذا كان $a_1 > 0$ يكون التوزيع موجب الإلتواء، وإذا كان $a_1 < 0$ يكون التوزيع سالب الإلتواء.

ثانيا بالنسبة لمقاييس التفلطح سوف نتناول مقياس يعتمد على العزم الرابع حول المتوسط معادلته هي :-

$$a_2 = \frac{m^4}{s^4}.$$

إذا كانت $a_2 = 3$ ، فذلك يعنى أن التوزيع متوسط التفلطح، وإذا كان $a_2 > 3$ فذلك يعنى أن التوزيع له قمة مدببة وإذا كان $a_2 < 3$ فهذا يدل على أن التوزيع مفلطحا.

مثال (٥-٢٦) أوجد مقياس الالتواء a_1 ومقياس التفلطح a_2 لفئة المشاهدات $2,4,6,8,13,15$.

الحل، الجدول (٥-٢٣) يسهل عملية الحساب كالتالي :-

جدول (٥-٢٣)

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
2	-6	36	-216	1296
4	-4	16	-64	256
6	-2	4	-8	16
8	0	0	0	0
13	5	25	125	625
15	7	49	343	2401
48	0	130	180	4594

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{48}{6} = 8 \quad , \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{130}{5} = 26 \quad ،$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^3 = 180 \quad ، \quad s^4 = 676 \quad , \quad s^3 = (26)(5.09902) = 132.5745 \quad ،$$

وعلى ذلك العزم الثالث حول الوسط الحسابي هو :-

$$m^3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n} = \frac{180}{6} = 30.$$

ومقياس الالتواء :-

$$a_1 = \frac{m^3}{s^3} = \frac{30}{132.5745} = 0.226288.$$

ويمكن إيجاد مقياس التفلطح a_2 وذلك بحساب القيم التالية من جدول (٥-٢٣) :-

$$\sum_{i=1}^{n}(x_i - \bar{x})^4 = 4594 \text{ ،} \quad m^4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n} = \frac{4594}{6} = 765.667$$

وعلى ذلك نحصل على مقياس التفلطح :-

$$a_2 = \frac{m^4}{s^4} = \frac{765.667}{676} = 1.13264.$$

وهذا يدل على أن توزيع المشاهدات مفلطح ٠

مثال (٥-٢٧) أوجد مقياس للتفلطح ومقياس للالتواء للمشاهدات في جدول (٥-٢٤) ٠

جدول (٥-٢٤)

المجموع	22-26	17-21	12-16	7-11	2-6	حدود الفئة
	24	19	14	9	4	مركز الفئة x_i
20	7	3	5	3	2	التكرار f_i
330	168	57	70	27	8	$x_i f_i$
	7.5	2.5	-2.5	-7.5	-12.5	$(x_i - \bar{x})$
	52.5	7.5	-12.5	-22.5	-25	$(x_i - \bar{x})f_i$
925	393.75	18.75	31.25	168.75	312.5	$(x_i - \bar{x})^2 f_i$
-2250	2953.1	46.9	-78.1	-1265.6	-3906.3	$(x_i - \bar{x})^3 f_i$
80781.7	22148.3	117.3	195.3	9492.0	48828.8	$(x_i - \bar{x})^4 f_i$

الحل ٠ من جدول (٥-٢٤) فإن الوسط الحسابي هو :

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{330}{20} = 16.5$$

والتباين يساوى :

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^2 f_i}{(\sum_{i=1}^{k} f_i - 1)} = \frac{925}{19} = 48.68421.$$

$$s^3 = 339.6896 \;,\; s^4 = 2370.1523 \;,\; \sum_{i=1}^{k}(x_i - \overline{x})^3 f_i = -2250.$$

$$m^3 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^3 f_i}{(\sum_{i=1}^{k} f_i)} = \frac{-2250}{20} = -112.5.$$

وعلى ذلك يمكن الحصول على مقياس الالتواء كما يلي :-

$$a_1 = \frac{m^3}{s^3} = \frac{-112.5}{339.6896} = -0.331185.$$

ويمكن إيجاد مقياس التفلطح a_2 وذلك من جدول (٥-٢٤):

$$\sum_{i=1}^{k}(x_i - \overline{x})^4 f_i = 80781.7,$$

$$m^4 = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^4 f_i}{\sum_{i=1}^{k} f_i} = \frac{80781.7}{20} = 4039.085.$$

وعلى ذلك يمكن الحصول على مقياس للتفلطح كما يلي :-

$$a_2 = \frac{m^4}{s^4} = \frac{4039.085}{2370.1523} = 1.704146.$$

تمارين

- ١ - إذا أعطيت التوزيع الاحتمالي التالي :

x	1	2	3
P(X=x)	0.25	0.25	0.5

(ا) أوجد قيم فئة المشاهدات من الحجم $N = 100$ التي مثلت هذا التوزيع .

(ب) أوجد الوسط الحسابي μ للمجتمع بطريقتين.

- ٢ - أوجد المعلمة μ للمجتمع 4,6,5 ثم ضع قائمة بكل العينات الممكنة من الحجم $n = 2$ مع إمكانية تكرار القيمة الواحدة في العينة ، واحسب لكل عينة القيمة $\bar{x}$ للإحصاء $\bar{X}$ و أثبت أن $E(\bar{X}) = \mu$.

- ٣ - في تمرين - ٢ - أحسب s^2 لكل عينة وذلك تحت فرض (١) μ معروفة (ب) μ غير معروفة وأثبت أن $E(S^2) = \sigma^2$ لكل حالة.

- ٤ - المطلوب إلقاء 10 عملات 100 مرة وتسجيل قيم x التي تمثل عدد مرات ظهور الصورة ثم إيجاد المدرج التكراري للتوزيع التكراري الذي يمثل عدد مرات ظهور الصورة.

- ٥ - أوجد للفئات التالية الحدود الفعلية للفئة ومركز الفئة وطول الفئة :

(١) 8-12 (ب) 11.5-16.7 (جـ) 2.2-3.3 (د) 0.345-0.416 .
(ز) 77.45-86.12 .

- ٦ - فيما يلي التوزيع التكراري للدرجات التي حصل عليها 80 شخصا في دورة تدريبية في مجال الحاسب الآلي استمرت 6 شهور والمطلوب إيجاد : (١) عدد الفئات (ب) طول الفئة (جـ) مركز الفئة (د) التكرار النسبي والتكرار المئوى.

90-99	80-89	70-79	60-69	50-59	حدود الفئة
5	10	50	10	5	التكرار

- ٧ - إذا كانت مراكز الفئات للتوزيع التكراري الذي يمثل أوزان مجموعة من النباتات (بالرطل) هي : 17,20,23,26,29,32,35 . أوجد حدود الفئة والحدود الفعلية وطول الفئة لهذا التوزيع.

-٨- في عينة من الزجاجات سعة كل منها لترا واحدا تم قياس ما تحتويه من سائل بالملليلتر وتم وضع البيانات في الجدول التالي:-

940-949	930-939	920-929	910-919	900-909	حدود الفئة
15	24	8	8	5	التكرار

أوجد : الوسط الحسابي لما تحتويه الزجاجات من سائل.

- ٩ - في عينة عشوائية من 20 طالب في كلية ما تم تسجيل عدد أيام الغياب لكل طالب خلال الفصل الدراسي الأول وكانت كالتالي : 1,0,3,4,5,4,1,1,1,1,1,0,0,0,0,3,3,2,2,1

أحسب كلا من :

الوسط الحسابي - الوسيط - المنوال - الانحراف المتوسط - الانحراف المعيارى.

– ١٠ – فيما يلي السرعة (بالأميال لكل ساعة) التي سجلها رادار على الطريق الزراعي لعينة عشوائية من 50 سيارة مرت عند نقطة المراقبة خلال ساعة :

60	66	54	54	49	74	71	65	56	47
59	70	71	66	65	70	65	64	63	55
55	54	60	63	61	65	45	53	54	61
54	54	53	48	47	70	74	63	62	61
54	66	70	64	65	64	63	68	66	70

(ا) ضع هذه البيانات في شكل توزيع تكرارى٠

(ب) أوجد التوزيعات التكرارية النسبية والمئوية والمتجمعة٠

(جـ) ارسم المدرج التكراري والمضلع التكراري لهذا التوزيع٠

(د) ارسم المنحنى التكراري النسبي والمنحنى التكراري المتجمع النسبى٠

– ١١ – إذا كان عدد أسماك السالمون التي تم صيدها بواسطة 10 صيادين في اليوم الأول من الموسم هي : 5,6,7,7,7,8,9,10,3,7 أوجد : الوسيط – الانحراف المتوسط – الانحراف المعيارى٠

– ١٢ – الجدول التالي يمثل التوزيع التكراري لأعمار مجموعة من أباء طلبة كلية

حدود الفئة	35-39	40-44	45-49	50-54	55-59
التكرار	6	15	40	25	14

والمطلوب :

(ا) رسم المدرج التكراري والمضلع التكراري لهذا التوزيع٠

(ب) إيجاد التكرار النسبي لهذا التوزيع ورسم مضلعة التكراري ٠

(جـ) إيجاد التكرار المئوي لهذا التوزيع ورسم مضلعة التكراري ٠

(د) إيجاد الوسط الحسابي والانحراف المعيارى٠

(ز) ما نسبة الآباء الذين أعمارهم : ضمن انحراف معياري واحد من الوسط الحسابي أي الواقعة في الفترة $\bar{x} \pm s$ ٠

(ر) كرر (ز) للفترة $\bar{x} \pm 2s$ وأيضا للفترة $\bar{x} \pm 3s$ ٠

– ١٣ – اختيرت عينة عشوائية من 10 مسامير لتقدير كمية الضغط الضروري لكسر المسمار وكانت النتائج كالتالي : 18,22,26,25,27,26,19,17,22,20 أحسب كلا من :

الوسط الحسابي – الوسيط – المنوال – الانحراف المتوسط – الانحراف المعياري – مقياس الالتواء a_1 ومقياس التفلطح a_2 – مقياس الالتواء لبيرسون٠

– ١٤ – فيما يلي التوزيع التكراري لعدد الدقائق التي يتأخرها مجموعة من الطلبة في دخولهم المحاضرة بعد دخول الأستاذ :

التأخير بالدقائق	0	1	2	3	4	5	6
التكرار	180	1	2	3	5	6	6

أوجد الوسط الحسابي و الانحراف المعياري لهذا التوزيع .

– ١٥ – تمتلك شركة ما 10 قوارب للصيد، قامت الشركة بتسجيل تكاليف صيانة كل قارب (بالدولار) وكانت كما يلي : 500,505,460,470,530,506,994,880,600,460

أوجد :

(ا) الوسيط والوسط الحسابي والمنوال . (ب) الربيع الأول والربيع الثالث .

(ج المدى الربيعي والمدى .

(د) مقياس للالتواء وآخر للتفلطح .

– ١٦ – الجدول التالي يمثل التوزيع التكراري لتكاليف تجهيز عينة عشوائية من نبات للتصدير .

حدود الفئة	1.00-1.02	1.03-1.05	1.06-1.08	1.09-1.11	1.12-1.14
التكرار	5	25	57	40	41

المطلوب إيجاد (ا) الوسيط والوسط الحسابي والمنوال . (ب) الربيع الأول والربيع الثالث .

– ١٧ – قامت شركة متخصصة في بيع البرامج الجاهزة بتسجيل حصيلة البيع الشهري بالدولار وذلك في مدة 28 شهرا وكانت المشاهدات كما يلي :

757.34	990.16	118.01	871.54	858.29	820.54	710.99
1150.5	1280.3	723.06	876.09	1230.9	657.98	1018.6
1140.6	997.05	657.90	999.98	1140.8	800.75	345.89
1234.8	1280.3	723.06	887.09	1209.0	670.01	678.98

(ا) كون توزيع تكراري مناسب (ب) أوجد الوسط الحسابي لهذا التوزيع .

(جـ) أوجد الوسيط لهذا التوزيع حسابيا وبيانيا .

(د) ارسم المدرج التكراري وأحسب من المنوال .

–١٨ – إذا كانت درجة الحرارة المئوية في إحدى المدن خلال أيام إحدى الأسابيع هي : 22,9,13,12,18,15,9 أحسب ما يلي : (ا) الوسيط والوسط الحسابي والمنوال .

(ب) المدى الربيعي والمدى .

– ١٩ – فيما يلي 25 قيمة لمتغير عشوائي :–

34	50	44	54	33
34	32	34	27	22
25	12	60	50	13
40	34	35	35	34
23	34	6	7	18

(ا) كون توزيع تكراري مناسب ٠

(ب) ارسم المدرج التكراري والمضلع التكراري لهذا التوزيع ٠

(جـ) أوجد العزوم الأربعة الأولى حول الوسط الحسابي ٠

– ٢٠ – مجتمع له وسطه الحسابي $\mu = 11$ وانحرافه المعياري $\sigma = 2$ ٠ باستخدام نظرية تشيبيشى أوجد (ا) نسبة القيم التي تقع بين 7 و 15 (ب) نسبة القيم التي تقع بين 5 و 17 ٠

– ٢١ – فيما يلي التوزيع التكراري للضرائب التي تم تحصيلها من مجموعة من الموظفين في شركة لتسخين البترول في عام 1990 ٠

حدود الفئة	400-500	501-601	602-702	703-803	804-904
التكرار	17	25	29	25	28

(ا) ارسم المضلع التكراري المتجمع ثم من الرسم حدد الوسيط والربيع الأول والربيع الثالث ٠

(ب) أحسب مقياس للنزعة المركزية ٠

(جـ) ما هو الرقم الذي يقسم التوزيع بحيث يكون عدد العاملين التي الضرائب المستحقة عليهم الأكبر منه يساوى عدد العاملين الأقل منه ؟

(د) ما نسبة العاملين الذين يدفعون ضرائب أقل من 702.5 ٠

– ٢٢ – المشاهدات التالية تمثل التوزيع التكراري للأجور الشهرية (بالدولار) لمجموعة من العمال في شركة لصناعة الغاز الطبيعى ٠

الأجور	التكرار
900-950	9
951-1001	15
1002-1052	21
1053-1103	25
1104-1154	27
1155-1205	19
1206-1256	14
1257-1307	9
1308-1358	6

(ا) كون توزيع تكراري مناسب .

(ب) ارسم المدرج التكراري والمضلع التكراري لهذا التوزيع .

(جـ) أوجد العزوم الأربعة الأولى حول الوسط الحسابي .

-٢٣ - الآتي يمثل أعمار 30 عاملة في مصنع لتعبئة الحلوى :

43	58	21	20	30	48
50	32	61	29	34	20
24	49	50	31	32	25
23	21	47	34	35	41
34	36	34	32	33	42

المطلوب إيجاد التوزيع التكراري ورسم المدرج و المضلع و المنحنى التكراري .

-٢٤ - الجدول التالي يمثل التوزيع التكراري لأعمار مجموعة من الذكور والإناث في كلية ما .

العمر	15-16	17-18	19-20	21-22	23-24
الإناث	100	125	130	160	190
الذكور	110	131	150	146	200

(ا) ارسم المضلع التكراري لكل من الذكور والإناث .

(ب) أوجد الوسط الحسابي والانحراف المعياري لكل من الذكور والإناث .

(جـ) أوجد معامل الاختلاف لكل من الذكور والإناث وأي المجموعتين أكثر تشتتا .

- ٢٥ - إذا كان معدل التضخم لشعب ما هو %3 في السنة الأولى و%4 في السنة الثانية و%8 في السنة الثالثة . أوجد الوسط الهندسي لمعدلات التضخم .

- ٢٦ - حصل مستثمر على عائد على رأس ماله قدره %3 للسنة الأولى و%5 للسنة الثانية %12 للسنة الثالثة أوجد الوسط الهندسي للعائد السنوى .

- ٢٧ - يشترى شخص 4 قمصان من الشركة A بسعر الواحد $22 و4 قمصان من الشركة B بسعر الواحد $25 و7 قمصان من الشركة C بسعر الواحد $30 . أوجد متوسط سعر القميص .

- ٢٨ - أسرة لديها 8 أطفال، أعمارهم كالتالي : 8,10,6,14,14,12,18,20 أوجد :

(ا) مقاييس النزعة المركزية لهذه البيانات .

(ب) المدى - المدى الربيعي - الانحراف المتوسط - الانحراف المعيارى .

– ٢٩ – أشترى شخص 10 كيلو من السمك من نوع (A) بسعر الكيلو 4 جنيها و 5 كيلو من السمك من نوع (B) بسعر الكيلو 2.5 جنيها و3 كيلو من السمك من نوع (C) بسعر الكيلو 1.5 ، باستخدام الوسط الحسابي المرجح أوجد متوسط سعر السمك الذي أشترى به،

– ٣٠ – يقوم رجل أعمال بتأجير الشقق التي يمتلكها شهريا بأسعار مختلفة، فالشقة من النوع الممتاز بسعر 1500 جنيها والشقة من النوع الفاخر بسعر 1000 جنيها والشقة من النوع المتوسط بسعر 500 جنيها، فإذا كان الرجل يمتلك 50 شقة من النوع الممتاز و 30 شقة من النوع الفاخر و 15 شقة من النوع المتوسط، أوجد الوسط الحسابي للإيجار الشهري الذي يتحصل علية،

– ٣١ – لدى رجل أعمال أربعة حسابات في بنك ما، ففي الحساب A لـه 4,000 دولار وفى الحساب B له 3,000 دولار وفى الحساب C له 10,000 دولار، فإذا كان الحساب A يعطى له أرباح قدرها 5.5% كل سنة والحساب B يعطى له أرباح قدرها 4% كل سنة والحساب C يعطى له أرباح قدرها 6% كل سنة، أوجد الوسط الحسابي المرجح للربح السنوى،

– ٣٢ – فيما يلي توزيع عدد المشاريع المنفذة شهريا خلال عام 1995 في شركة بترول:

15,11,7,6,8,10,12,6,8,9,6,13 أحسب :

(ا) الوسط الحسابي – الوسيط – المنوال – الانحراف المعياري – المدى – المدى الربيعى،

– ب– مقياس الالتواء a_1 ومقياس للتفلطح a_2 – مقياس الالتواء لبيرسون،

– ٣٣ – يقوم العاملين في شركة صغيرة بالتوقيع في صفحات الزمن للدلالة على زمن مغادرة الشركة، المشاهدات التالية تمثل أزمنة المغادرة للعاملين وذلك في يوم عشوائي :

5:33	5:30	5:40	1:10	3:50	5:15
4:30	5:16	5:12	4:30	5:59	4:45
3:30	3:33	3:30	5:40	2:40	5:30
4:30	5:41	5:50	3:22	4:56	5:30
5:12	5:35	5:44	4:32	4:23	5:30

(ا) أوجد الوسط الحسابي لهذا التوزيع،

(ب) أوجد الوسيط لهذا التوزيع ،

(جـ) أوجد الانحراف المعيارى لهذا التوزيع،

–٣٤ – فيما يلي توزيع عدد مرات احتراق الطعام خلال أسبوع لعينة من 10 سيدات:

3,2,4,1,1,0,0,0,4,1 والمطلوب إيجاد الوسط الحسابي والانحراف المعياري ووصف البيانات باستخدام نظرية تشيبيشى،

– ٣٥ –الجدول التالي يمثل التوزيع التكراري للودائع في بنك ما في نهاية عام 1993 :

عدد العملاء	حجم الودائع
7100	0-500
8300	501-1001
9345	1002-1502
9945	1503-2003
6257	2004-2504
2003	2505-3005
14	3006-3506
1445	3507-4007

أوجد : الوسط الحسابي – المنوال – الانحراف المتوسط – الانحراف المعياري

– ٣٦ – إذا كانت جامعة ما تمثل مجتمع عناصره عدد المدرسين في كل كلية . بفرض أن الوسط الحسابي لعدد المدرسين في الكلية الواحدة هو $\mu = 16$ بانحراف معياري $\sigma = 15$. أستخدم نظرية تشيبيشى في وصف هذا المجتمع .

– ٣٧ – أجرى اختبار في مادة الإحصاء لمجموعة من الطلبة عددهم 400 وقد وجد أن الوسط الحسابي هو $\bar{x} = 77$ والانحراف المعياري $\sigma = 5$ أستخدم نظرية تشيبيشى لوصف هذه الفئة من البيانات .

– ٣٨ – وجد أن تلوث البحار يؤدى إلى نمو أنواع مختلفة من البكتريا . فإذا كان عدد البكتريا لكل 100ملليمتر في 10 أماكن من مياه البحر هي : 49,69,60,41,69,70,51,68,67,66

أوجد: الوسط الحسابي – الوسيط – المنوال – الانحراف المتوسط – الانحراف المعيارى – مقياس للالتواء وآخر للتفلطح .

– ٣٩ – تم سؤال عينة عشوائية من 10 عمال عن المسافة (بالأميال) التي يقطعونها في الذهاب إلى المزرعة التي يعملون بها وكانت إجابتهم كما يلي : 25,6,1,2,4,8,5,6,5,4 . أوجد : الوسيط والربيع الأول والربيع الثالث والمدى الربيعى– مقياس للالتواء وآخر للتفلطح .

– ٤٠ – المشاهدات التالية تمثل عدد المرضى الذين يتم الكشف عليهم يوميا في مستشفى خاص من قبل 10 أطباء : 15,8,6,9,15,18,21,39,5,7 : أوجد الوسط الحسابي والانحراف المعياري واستخدم نظرية تشيبيشى لوصف هذه الفئة من المشاهدات وأوجد مقياس للالتواء وآخر للتفلطح .

– ٤١ – في جامعة ما يتقاضى الأستاذ في المتوسط $15,000 دولار بانحراف معيـــاري $5,000 بينما في جامعة أخرى يتقاضى الأستاذ أجر قدره $10,000 بانحراف معيـــاري $3,000 أوجـــد معامل الاختلاف لكل جامعة وأي الجامعتين أكثر تشتتا ؟

– ٤٢ – تمتلك شركة أربع مزارع ، بعض البيانات عن هذه المزارع في الجدول التالي :

الانحراف المعياري	متوسط الأجر السنوي	عدد العمال	المزرعة
$75	$8,300	100	1
1200	11,200	180	2
900	9,000	200	3

أوجد معامل الاختلاف لكل مزرعة وما هي المزرعة التي لها أكبر تشتت في الأجر السنوى ؟

– ٤٣ – إذا كان الوسط الحسابي للأجر الشهري للموظف في شركتين B , A متساوي وهـــو $10,000 بانحراف معياري $500 للشركة A وانحراف معياري $550 للشركة B ، أوجـــد معامل الاختلاف لكل شركة وأي الشركتين أكثر تشتتا في الأجر ؟

– ٤٤ – قام مزارع بوزن نوعين من ثمار البرتقال في المزرعة التي يمتلكها وقـــد حصـــل علـــى المشاهدات التالية : $\overline{x}_1 = 15$, $s_1 = 1$, $\overline{x}_2 = 14$, $s_2 = 2$ أوجد معامل الاختلاف لكل نوع وأي النوعين أكثر تشتتا ؟