المدخل إلى الإحصاء

Introduction to statistics

الطبعة الثانية/ ٢٠٠٠

الفصل الأول

المدرس المساعد

خليل ابراهيم خليل

وزارة التعليم العالي والبحث العلمي .
جامعة الموصل .
كلية الزراعة والغابات .
قسم المحاصيل الحقلية .

Ministry of Higher Education and Scientific Research .
University of Al Mosul .
College of Agriculture and Forestry .
Field crops department .

المدخل إلى الإحصاء

Introduction to statistics

الدكتور خاشع محمود الراوي

الطبعة الثانية/ ٢٠٠٠

الفصل الأول

المقدمة

طبيعة علم الإحصاء

Nature of Statistics

إن كلمة الإحصاء في الماضي كانت تهدف إلى العد والحصر حتى لقد سمي الإحصاء بعلم العد (The science of counting) كما إن لفظ الإحصاء باللغة الانكليزية (Statistics) كانت تستعمل في بلاد أوربا للدلالة على أعمال وحسابات الدولة في شؤون الحرب والضرائب وعدد السكان والمواليد والوفيات والانتاج الزراعي .... الخ .

أما الآن فإن الاحصاء قد تطور كثيراً وخاصتاً في القرن العشرين وأصبح علماً مستقلاً له أهميته كوسيلة وأداة في البحث العلمي لجميع العلوم .

ويمكن تقسيم علم الاحصاء بصورة عامة الى قسمين رئيسيين

1– الاحصاء الوصفي Descriptive Statistics

ويشمل على الطرق الاحصائية المستعملة في وصف مجموعة معينة من البيانات وتتضمن هذه الطرق الاحصائية على أساليب جمع البيانات (Collection of data) في صورة قياسات رقمية (Numerical Measurements) ثم تبويبها وتنظيمها (Organizing) وتلخيصها (Summarizing) وعرضها (Presenting) وحساب بعض المقاييس الاحصائية المختلفة لها .

2- الاحصاء الاستنتاجي أو الاستدلالي Statistical Inference

ويشمل الطرق الاحصائية التي تهدف الى عمل استنتاجات أو استدلالات حول المصدر الذي جمعت منه البيانات . **ويضم هذا القسم فرعين رئيسيين** :

a– التقدير (Estimation)

ويهتم بإيجاد قيم تقديرية للاستدلال منها على القيم الحقيقية لمصدر جمع البيانات . وهذه القيم التقديرية أما أن تكون تقديراً محدداً أي عند نقطة معينة (Point estimation) أو تقديراً في فترة أو مدى (Interval estimation) .

b- اختبار الفرضيات (Test of Hypotheses)

ويتضمن اختبار الفرضيات التي توضع كتفسير أولي للظاهرة المراد دراستها للوصول منها على قرار بقبولها أو رفضها .

ومما تقدم يمكن تعريف علم الاحصاء كالتالي

علم الاحصاء : هو ذلك العلم الذي يعمل على استخدام الاسلوب العلمي في طريق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها على استنتاجات وقرارات مناسبة .

تاريخ موجز عن تطوير علم الاحصاء History of Statistics

إن تاريخ الاحصاء الوصفي Descriptive Statistics يعود الى بداية الحضارة البشرية . ففي عصر البابليين والآشوريين والفراعنة واليونانيين تم استعمال الاحصاء في الحصول على معلومات حول تعداد الرجال للحروب والنتاج الزراعي وتقدير الضرائب .

أما الاحصاء الاستنتاجي أو الاستدلالي Statistical Inference الذي يعتمد اعتماداً كبيراً على نظرية الاحتمال Theory of probability فقد بدأ تطوره منذ القرن السادس عشر كنتيجة لانتشار لعب القمار في أوربا . فقد توجه المقامرون الى علماء الرياضيات لإعطائهم معلومات حول فرص ربحهم أو خسارتهم . ومن أشهر هؤلاء العلماء هم Pascal و Leibnitz و Fermat و Bernoulli مؤسسي نظرية الاحتمال .

وفي عام 1833 اكتشف De Moivre معادلة منحني التوزيع الطبيعي Normal Distribution الذي تعتمد عليه نظرية الاحصاء الاستدلالي . ويسمى المنحني الطبيعي أحياناً بمنحني كاوس Causs (1777 – 1855) الذي اشتق معادلته أثناء دراسته للخطأ الناتج من القياسات المتكررة لنفس الكمية . كما إن دراسات Laplace (1749 – 1827) أدت الى نفس نتائج Gauss . إضافة الى أن تطبيقه للإحصاء في علم الفلك كانت لها أهمية كبرى في ذلك الوقت .

كما تم تطبيق علم الاحصاء من قبل الجيولوجي Charles Lyell (1797 – 1875) والبايولوجي Charles Darwin (1809 – 1882) ومربي النبات Joham Gregor Mendel (1822 - 1884) بالرغم من كونهم غير احصائيين .

وفي القرن التاسع عشر اشتهر العالم البلجيكي Adolph Quetelet (1794 - 1874) بتطبيقه علم الاحصاء بشكل فعال في علمي الاجتماع والتعليم .

ثم جاء العالم Francis Galton (1822 - 1911) واشتهر بتطبيق علم الاحصاء في علم الوراثة والتطور . أما العالم الرياضي الفيزيائي Karl Pearson فقد اشترك مع Galton في إيجاد نظرية الارتباط Correlation والانحدار Regression واليه تعود معظم اساسيات نظرية المعاينة Sampling وهو الذي انشأ مجلة Biometrika ومدرسة كبيرة من الاحصائيين أشهرهم W.S.Gosset الملقب بـ Student وكان يعمل كيمياوياً في معمل للبيرة الذي اشتق توزيع t (t- Distribution) للعينات الصغيرة .

أما اشهر علماء القرن العشرين فهو العالم R.A. Fisher (1890 - 1962) الذي طور علم الاحصاء وطبقه في علوم كثيرة كالزراعة والبايولوجي والوراثة والاقتصاد ووضع اسس تصميم وتحليل التجارب. ومن العلماء الآخرين الذين اسهموا في تطوير علم الاحصاء هم Tchebycheff ، Kolmogorov ، Smirnov ، E.S. Pearson ، J. Neyman ، A. Wald وغيرهم .

تمارين الفصل الأول

1- ما هو علم الاحصاء وما هي استعمالاته ؟

2- ما المقصود بالإحصاء الوصفي والاحصاء الاستدلالي ؟

3- من هم أشهر علماء الاحصاء في القرن العشرين وما هي أشهر أعمالهم ؟

حلول تمارين الفصل الأول

1- ما هو علم الاحصاء وما هي استعمالاته ؟

علم الاحصاء : هو ذلك العلم الذي يعمل على استخدام الاسلوب العلمي في طريق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها على استنتاجات وقرارات مناسبة .

استعمالاته قديماً	استعمالاته حديثاً
كانت تستعمل قديماً في بلاد أوربا للدلالة على أعمال وحسابات الدولة في شؤون الحرب والضرائب وعدد السكان والمواليد والوفيات والنتاج الزراعي .	ويعتبر علماً مستقلاً له أهميته كوسيلة وأداة في البحث العلمي لجميع العلوم .

2- ما المقصود بالإحصاء الوصفي والاحصاء الاستدلالي ؟

a- الاحصاء الوصفي Descriptive Statistics .

ويشمل على الطرق الاحصائية المستعملة في وصف مجموعة معينة من البيانات وتتضمن هذه الطرق الاحصائية على أساليب جمع البيانات (Collection of data) في صورة قياسات رقمية (Numerical Measurements) ثم تبويبها وتنظيمها (Organizing) وتلخيصها (Summarizing) وعرضها (Presenting) وحساب بعض المقاييس الاحصائية المختلفة لها .

b- الاحصاء الاستنتاجي أو الاستدلالي Statistical Inference .

ويشمل الطرق الاحصائية التي تهدف الى عمل استنتاجات أو استدلالات حول المصدر الذي جمعت منه البيانات .

3- من هم أشهر علماء الاحصاء في القرن العشرين وما هي أشهر أعمالهم ؟

إن من اشهر علماء القرن العشرين هم R. A. Fisher ، Tchebycheff ، Kolmogorov ، Smirnov ، E. S. Pearson ، J. Neyman ، A. Wald وغيرهم ، حيث استخدموا علم الاحصاء في علوم كثيرة كالزراعة والبايولوجي والوراثة والاقتصاد وتصميم وتحليل التجارب وغيرها من العلوم الاخرى .

المدخل إلى الإحصاء

Introduction to statistics

الطبعة الثانية/٢٠٠٠

الفصل الثاني

المدرس المساعد

خليل ابراهيم خليل

وزارة التعليم العالي والبحث العلمي .
Ministry of Higher Education and Scientific Research .

جامعة الموصل .
University of Al Mosul .

كلية الزراعة والغابات .
College of Agriculture and Forestry .

قسم المحاصيل الحقلية .
Field crops department .

المدخل إلى الإحصاء

Introduction to statistics

الدكتور خاشع محمود الراوي

الطبعة الثانية/ ٢٠٠٠

الفصل الثاني

طبيعة البيانات والرموز الاحصائية

The Nature of Statistical Data and Symbols

عند جمع بيانات حول ظاهرة ما فإننا نرمز للظاهرة بالرمز (y) وكل مفردة أو مشاهدة منها نرمز لها بالرمز (y_i) فمثلاً عند دراسة أطوال الطلبة في إحدى الجامعات فإننا نرمز لصفة الطول بالرمز (y) وطول أي طالب بالرمز (y_i) وتسمى المشاهدة أو المفردة Observation هذا وأن قيمة (y_i) قد تختلف من طالب إلى أخر ولهذا نقول بأن (y_i) متغير Variable .

<u>**المتغير**</u> : هو أي ظاهرة تظهر اختلافات بين مفرداتها ويرمز له بالرمز y أو أي رمز أخر x أو z .....الـــخ .

<u>**المتغيرات Variables تقسم الى :**</u>

1- متغيرات وصفية أو نوعية Qualitative Variables .

وهي تلك الظواهر أو الصفات التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون العيون (أزرق ، أسود ، بني) والحالة الاجتماعية (غني ، متوسط الحال ، فقير) والجنس (ذكر ، انثى) ......الـــخ .

2- متغيرات كمية Qualitative Variables .

وهي تلك الظواهر أو الصفات التي يمكن قياسها مباشرة بالأرقام عددية مثل صفة الطول والوزن والعمر وكمية المحصول ......الـــخ .

a- متغيرات مستمرة أو متصلة Continuous Variables .

فالمتغير المستمر هو المتغير الذي يأخذ المشاهدة أو المفردة فيه اية قيمة رقمية في مدى معين فلو فرضنا بأن اطوال طلبة جامعة ما تتراوح بين 130.5 و 170 سم فنقول بأن : $(130.5 \leq y \leq 170.0)$. أي أن المتغير ممكن أن يأخذ اية قيمة بين 130.5سم و170 . وكأمثلة اخرى على المتغيرات المستمرة هي (الوزن ، كمية المحصول ، درجة الحرارة ، الزمن) لا نه يمكن قياسها بأجزاء صغير جداً وتأخذ اية قيمة تقع في حدود معينة . وبصورة عامة فإن كل البيانات التي تقاس (Measurements) تعتبر بيانات لمتغير مستمر .

b– متغيرات غير مستمرة أو منفصلة Discrete Variables .

المتغير المنفصل هو المتغير الذي يأخذ المشاهدة أو المفردة فيه قيماً متباعدة أو متقطعة غير مستمرة . فلو فرضنا أن عدد أفراد الاسرة في أربع عوائل هي (2, 3, 4, 5) فنقول بأن $y = 2, 3, 4, 5$.

وكذلك عند رمي زهر النرد (زار الطاولة) نجد أن النتيجة تكون ظهور الوجه 1, 2, 3, 4, 5, 6 فنقول بأن $y = 1, 2, 3, 4, 5, 6$.

وكأمثلة أخرى على المتغيرات غير المستمرة أو المنفصلة هي : عدد الثمار على النباتات أو عدد الوحدات الانتاجية في مصنع ما أو عدد الطلبة في الصفوف الاولى لجامعة ما ... فهي في الغالب تكون أعداداً صحيحة .

وبصورة عامة فإن كل البيانات التي نحصل عليها من العد (Counting's) تعتبر بيانات لمتغير منفصل .

<u>المجتمع والعينة Population and Sample .</u>

<u>المجتمع Population</u>: عبارة عن جميع القيم أو المفردات التي يمكن أن يأخذها المتغير .

فمثلاً إذا كانت دراستنا متعلقة بأطوال طلبة جامعة ما فإن المجتمع في هذه الحالة هو أطوال جميع الطلبة في تلك الجامعة .

والمجتمع إما أن يكون :

a– مجتمعاً محدداً Finite population .

أي ممكن حصر عدد مفرداته كما هو الحال في أطوال طلب جامعة الموصل مثلاً ، أو عدد الوحدات الانتاجية في مصنع ما في يوم معين .

b– مجتمع غير محدود Infinite population .

وهو المجتمع الذي من الصعب أو المستحيل حصر عدد مفرداته مثل مجتمع نوع السمك معين في نهر دجلة وعدد البكتريا في حقل ما .

<u>العينة Sample</u> . العينة جزء من المجتمع .

<u>فالعينة</u> : عبارة عن مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع .

إن دراسة المجتمع ككل قد يكون صعباً أو يحتاج الى وقت وجهد ومال ، لذا فقد استعين عن دراسة المجتمع بدراسة العينة وصفاتها ومنها نستطيع أن نستنتج خواص المجتمع الأصلي الذي أخذت منه هذه العينة.

<u>الرموز الاحصائية Statistical Notations .</u>

فسوف نستعمل الرموز ، والمعادلات اللاتينية كما هي بدون تعريب وذلك لكونها رموزاً عالمية من وجهة ولسهولة الاستفادة والاستنارة بالمراجع الاجنبية ولعدم وجود اتفاق تام في الوقت الحاضر على تعريبها من وجهة أخرى .

وكما ذكرنا سابقاً سنرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i .

مثال

لو كانت أعمار 5 طلاب كالآتي 20, 18, 24, 22, 16 سنة فنكتب. $y_i = 20, 18, 24, 22, 16$

If the ages of 5 students is are follows 20, 18, 24, 22, 16 year we write .

$y_i = 20, 18, 24, 22, 16$

أي أن : $y_1 = 20$ أي القيمة الأولى للمتغير أو المشاهدة الأولى .

و $y_2 = 18$ أي القيمة الثانية للمتغير أو المشاهدة الثانية .

وهكذا الى $y_n = 16$ أي القيمة الاخيرة $(n = 5)$ للمتغير أو المشاهدة الاخيرة .

ويرمز عادةً لمجموع قيم المتغير بالرمز :

$$\sum_{i=1}^{n} yi$$

فالرمز هو حرف إغريقي يسمى (sigma) أي مجموع الـ ..... أو " Summation of" والرقمان 1 و n هما حدا المجموعة .

وعلية فالرمز $\sum_{i=1}^{n} yi$ يقرأ كالاتي

مجموع قيم **y** مبتدأً من المشاهدة الأولى وحتى المشاهدة الأخيرة أي :

$$\sum_{i=1}^{n} yi = y_1 + y_2 + ... + y_n$$

وللاختصار والسهولة قد يكتب الرمز السابق بدون ذكر حدي المجموع أي $(\sum yi)$ فقط إذا لم يكن هناك خوف من الالتباس .

وهناك مجموع جزئي مثل : $\sum_{i=3}^{n} yi$

أي مجموعة المشاهدة الثالثة والرابعة والخامسة : $\sum_{i=3}^{n} yi = y_1 + y_2 + y_3$

ويرمز لمجموع مربعات جميع المشاهدات بالرمز : $(\sum_{i=3}^{n} yi)^2$

$$(\sum_{i=3}^{n} yi)^2 = (y_1 + y_2 + ... + yi)^2$$

كما ويرمز لمجموع حاصل ضرب قيم متغيرين **x** و **y** بالرمز $\sum x_i y_i$

$$\sum x_i y_i = x_1 y_1 + x_2 y_2 + ... + x_n y_n$$

ويرمز لحاصل ضرب مجموعين لقيم متغيرين بالرمز $(\sum x_i)(\sum y_i)$

$$(\sum x_i)(\sum y_i) = (x_1 + x_2 + ...+ x_n)(y_1 + y_2 + ...+ y_n)$$

مثال (1)

نفرض بأن قيم المتغير y هي كالأتي $y_i = 3, 9, 6, 2$ وإن قيم المتغير x هي $x_i = 4, 2, 3, 7$ أوجد قيمة كل مما يأتي .

(A) $\sum_{i=1}^{n} y_i$ (B) $\sum_{i=2}^{3} y_i$ (C) $\sum y_i^2$

(D) $(\sum yi)^2$ (E) $\sum x_i y_i$ (F) $(\sum x_i)(\sum y_i)$

الحـــل :

(A) $\sum y_I = y_I + y_2 + y_3 + y_4 = 3 + 6 + 9 + 2 = 20$

(B) $\sum_{i=2}^{3} yi = y_2 + y_3 = 9 + 6 = 15$

(C) $\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 = (3)^2 + (9)^2 + (26)^2 + (2)^2 = 130$

(D) $\sum (y_i)^2 = (y_1 + y_2 + y_3 + y_4)^2 = (3 + 9 + 6 + 2)^2 = (20)^2 = 400$

(E) $\sum x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = (4)(3) + (2)(9) + (3)(6) + (7)(2)$

$= 62$

(F) $(\sum x_i)(\sum y_i) = (x_1 + x_2 + x_3 + x_4)(y_1 + y_2 + y_3 + y_4)$

$= (4 + 2 + 3 + 7)(3 + 9 + 6 + 2) = (16)(20) = 320$

هذا وفيما يلي بعض القواعد المفيدة في عملية الجمع :

قاعدة (1)

إذا كانت (C) أي عدد ثابت فإن $\longrightarrow$ $\sum_{i=1}^{n} c = nc$

البرهان : $\sum_{i=1}^{n} c = \underbrace{c_1 + c_2 + \ldots + c_n}_{\text{من المرات } n} = nc$

قاعدة (2)

إذا كانت (C) أي عدد ثابت فإن $\longrightarrow$ $\sum c\, y_i = c \sum y_i$

البرهان : $\sum cy_i = cy_1 + cy_2 + \ldots + cy_n = c\,(y_1 + y_2 + \cdots + y_n) = c \sum y_i$

قاعدة (3)

جمع قيم متغيرين أو أكثر هو مجموع جمعهم أي : $\sum (x_i + y_i) = \sum x_i + \sum y_i$

البرهان : $\sum (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \ldots + (x_n + y_n)$

$$\sum x_i + \sum y_i = (x_1 + x_2 + x_3 + x_n) + (y_1 + y_2 + y_3 + y_n)$$

هذا ويجب التفريق بين بعض الرموز الإحصائية مثل :

$$\sum \frac{x_i}{y_i} = \frac{x_1}{y_1} + \frac{x_2}{x_2} + \dots + \frac{x_n}{y_n}$$

بينما

$$\frac{\sum x_i}{\sum y_i} = \frac{x_1 + x_2 + \cdots + x_n}{y_1 + y_2 + \cdots + y_n}$$

فإن كذلك

$$\sum (x_i - 3) = \sum x_i - n(3)$$

تختلف عن

$$\sum x_i - 3$$

مثال (2)

إذا علمت بأن قيم كل من المتغيرين x و y هي كالآتي :

$$x_i = 2, 6, 3, 1 \qquad y_i = 3, 9, 6, 2$$

أوجد قيمة كل مما يأتي :

(A) $\sum (y_i - x_i)^2$ **(B)** $\sum (x_i - 3)(y_i - 5)$ **(C)** $\sum x_i \, y_i^2$

(D) $\sum (y_i - 3)$ **(E)** $\sum y_i - 3$ **(F)** $\sum \frac{x_i + 2}{y_i}$

(G) $\frac{\sum (x_i - 2)}{\sum y_i}$ **(H)** $\sum y_i^2 - \frac{(\sum y_i)^2}{n}$

(I) $\sum x_i \, y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

الـــحــــــل :

(A) $\sum (y_i - x_i)^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 + (y_4 - x_4)^2$.

$$= (3-2)^2 + (9-6)^2 + (6-3)^2 + (2-1)^2$$

$$= (1)^2 + (3)^2 + (3)^2 + (1)^2$$

$$= 1 + 3 + 3 + 1 = 20$$

و هكذا يمكن الوصول الى نفس النتيجة وذلك بفتح القوس ثم التعويض كما يلي :

$$\sum (y_i - x_i)^2 = \sum (y_i^2 - 2\, x_i\, y_i + y_i^2) = \sum y_i^2 - \sum 2\, x_i\, y_i + \sum y_i^2$$

وعلى القارئ أن يعوض فيها للتأكد من النتيجة السابقة.

(B) $\sum (x_1 - 3)(y_1 - 5)$.

$$= (x_1 - 3)(y_1 - 5) + (x_2 - 3)(y_2 - 5) + (x_3 - 3)(y_3 - 5) + (x_4 - 3)(y_4 - 5)$$

$$= (2-3)(3-5) + (6-3)(9-5) + (3-3)(6-5) + (1-3)(2-5)$$

$$= (-1)(-2) + (3)(4) + (0)(1) + (-2)(-3)$$

$$= 2 + 12 + 0 + 6 = 20$$

وهنا أيضاً يمكن الوصول إلى نفس النتيجة بفتح الاقواس ثم التعويض كما يلي .

$$\sum (x_1 - 3)(y_1 - 5) = \sum (x_i y_i - 5x_i - 3y_i + 15)$$

$$= \sum x_i y_i - 5 \sum x_i - 3 \sum y_i + 15)$$

$$= 80 - 5(12) - 3(20) + 60$$

$$= 80 - 60 - 60 + 60 = 20$$

(C) $\sum x_i \, y_i^{\,2} = x_1 \, y_1^{\,2} + x_2 \, y_2^{\,2} + x_3 \, y_3^{\,2} + x_4 \, y_4^{\,2}$

$$= (2)(3)^2 + (6)(9)^2 + (3)(6)^2 + (1)(2)^2$$

$$= (2)(3)^2 + (6)(9)^2 + (3)(6)^2 + (1)(2)^2$$

$$= (2)(9) + (6)(81) + (3)(36) + (1)(4)$$

$$= 18 + 476 + 108 + 4 \; = 616$$

(D) $\sum (y_i - 3) \; = \; \sum y_i - \sum 3 \; = \; \sum y_i - n\,(3)$

$$= \; 3 + 9 + 6 + 2 - 4\,(3)$$

$$= \; 20 - 12 = 8$$

(E) $\sum y_i - \; 3 = \; 3 + 9 + 6 + 2 - \; 3 \; = \; 20 - 3 = 1$

(F) $\sum \frac{x_i + 2}{y_i} \; = \; \frac{x_1 + 2}{y_1} + \frac{x_2 + 2}{y_2} + \frac{x_3 + 2}{y_3} \; + \; \frac{x_4 + 2}{y_4}$

$$= \frac{2+2}{3} + \frac{6+2}{9} + \frac{3+2}{6} + \frac{1+2}{2}$$

$$= \frac{4}{3} + \frac{8}{9} + \frac{5}{6} + \frac{3}{2}$$

$$= 1.33 + \; 0.88 \; + \; 0.83 \; + 1.5 = 4.54$$

(G) $\frac{\sum (x_i + 2)}{\sum y_i} \; = \; \frac{\sum x_i + (n)(2)}{\sum y_i} = \frac{2 + 6 + 3 + 1 + (4)(2)}{3 + 9 + 6 + 2}$

$$= \frac{2 + 6 + 3 + 1 + 8}{20} \; = \; \frac{20}{20} \; = 1$$

(H) $\sum y_i^2 - \frac{(\sum y_i)^2}{n} = y_1^2 + y_2^2 + y_3^2 + y_4^2 - \frac{(y_1 + y_2 + y_3 + y_4)^2}{n}$

$$= (3)^2 + (9)^2 + (6)^2 + (2)^2 - \frac{(3+9+6+2)^2}{4}$$

$$= 9 + 81 + 36 + 4 - \frac{(20)^2}{4} = 130 - \frac{400}{4}$$

$$= 130 - 100 = 30$$

(I) $\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} =$

$$= \sum (x_1y_1) + (x_2y_2) + (x_3y_3) + (x_4y_4) - \frac{(x_1+x_2+x_3+x_4)(y_1+y_2+y_3+y_4)}{n}$$

$$= (2)(3) + (6)(9) + (3)(6) + (1)(2) - \frac{(2+6+3+1)(3+9+6+2)}{4}$$

$$= 6 + 54 + 18 + 2 - \frac{(12)(20)}{4} = = 80 - \frac{240}{4} = 80 - 60 = 20$$

أسئلة الفصل الثاني

1- عين نوع المتغير (مستمر أو منقطع) في كل من الحالات التالية :

a- سرعة السيارة بالأميال في الساعة .

b- عدد الكتب على رفوف مكتبة كلية الزراعة .

c- الدخل السنوي لأستاذ في إحدى الجامعات .

d- عدد الطلبة المقبولين في جامعة ما في عدة سنوات .

e- عدد السيارات المباعة يومياً من الشركة العامة للسيارات .

f- عدد إنجات كمية المطر النازل في مدينة معينة خلال أشهر السنة .

g- درجات الحرارة المقاسة كل نصف ساعة في محطة الانواء الجوية في بغداد .

2- أكتب حدود كل مما يأتي : (Write the limits of each of the following)

A/ $\sum_{i=2}^{5} x_i$ B/ $\sum_{i=1}^{n} c =$ C/ $\sum_{i=1}^{4} (y_i - 3)^2$ D/ $\sum_{i=1}^{3} (X_i - 2y_i + 10)$.

3- أكتب كلاً من الحدود التالية مستعملاً رمز الجمع .

(Write All the limits for the following using the summation sing)

A/ $x_1^2 + x_2^2 + \dots\dots + x_{10}^2 =$

B/ $cx_1^3 + cx_2^3 + \dots\dots + cx_{10}^3 =$

C/ $(x_1 + y_1) + (x_2 + y_2) + \dots\dots + (x_8 + y_8) =$

4- **برهن بأن** : (Prove that) .

$$\sum(ax_i + by_i - cz_i) = a\sum x_i + b\sum x_i - c\sum x_i$$

علماً بأن a و b و c هي أعداد ثابتة .

.knowing that a و b و c

5- **من القيم التالية** : (From the following values) .

$x_1 = 7$, $x_2 = -2$, $x_3 = 4$

$y_1 = 5$, $y_2 = 8$, $y_3 = 2$

أوجد قيمة كل مما يأتي .

A/ $\sum y_i^2 - \frac{(\sum y_i)^2}{n}$.

B/ $\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$.

C/ $\sum (x_i + y_i)(x_i - y_i)$.

D/ $\sum (x_i - 8)$.

E/ $\sum x_i - 8$.

F/ $\sum \frac{(y_i^2 - 10)}{2x_i}$.

G/ $\sum \frac{(y_i^2 - 10)}{\sum 2x_i}$

6- برهن بأن $\sum y_i^2$ لا تساوي $(\sum y_i)^2$

وإن : $\sum x_i y_i$ لا تساوي $(\sum x_i)(\sum y_i)$

أسئلة الفصل الثاني

1- عين نوع المتغير (مستمر أو منقطع) في كل من الحالات التالية :

a- سرعة السيارة بالأميال في الساعة . (مستمر)

b- عدد الكتب على رفوف مكتبة كلية الزراعة . (متقطع غير مستمر)

c- الدخل السنوي لأستاذ في إحدى الجامعات . (متقطع غير مستمر)

d- عدد الطلبة المقبولين في جامعة ما في عدة سنوات . (متقطع غير مستمر)

e- عدد السيارات المباعة يومياً من الشركة العامة للسيارات . (غير مستمر)

f- عدد إنجات كمية المطر النازل في مدينة معينة خلال أشهر السنة . (مستمر)

g- درجات الحرارة المقاسة كل نصف ساعة في محطة الانواء الجوية في بغداد . (مستمر)

2- أكتب حدود كل مما يأتي : (Write the limits of each of the following)

A/ $\sum_{i=2}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$.

B/ $\sum_{i=1}^{n} C = c_1 + c_2 + c_3 \ldots\ldots + c_n = nC$.

C/ $\sum_{i=1}^{4} (y_i - 3)^2 = (y_1 - 3)^2 + (y_2 - 3)^2 + (y_3 - 3)^2 + (y_4 - 3)^2$.

D/ $\sum_{i=1}^{3} (X_i - 2y_i + 10) = (X_1 - 2y_1 + 10) + (X_2 - 2y_2 + 10) + (X_3 - 2y_3 + 10)$.

3- أكتب كلاً من الحدود التالية مستعملاً رمز الجمع .

(Write All the limits for the following using the summation sing)

A/ $x_1{}^2 + x_2{}^2 + \dots\dots + x_{10}{}^2 = \sum_{i=1}^{10} x_i{}^2$

B/ $cx_1{}^3 + cx_2{}^3 + \dots\dots + cx_{10}{}^3 = \sum_{i=1}^{20} x_i{}^3$

C/ $(x_1 + y_1) + (x_2 + y_2) + \dots\dots + (x_8 + y_8) = \sum_{i=1}^{8} x_i$

4- برهن بأن : (Prove that) .

$$\sum (ax_i + by_i - cz_i) = a\sum x_i + b\sum x_i - c\sum x_i$$

علماً بأن a و b و c هي أعداد ثابتة .

نقوم بإدخال $\sum$ إلى داخل القوس ولكل حد وثم نقوم بجعل الاعداد الثابتة قبل $\sum$ لتصبح المعادلة كما يلي .

$$\sum a\,x_i + \sum b\,x_i - \sum c\,x_i = a\sum x_i + b\sum x_i - c\sum x_i$$

٥- من القيم التالية :

$x_1 = 7$, $x_2 = -2$, $x_3 = 4$

$y_1 = 5$, $y_2 = 8$, $y_3 = 2$

أوجد قيمة كل مما يأتي .

A/ $\sum y_i^2 - \frac{(\sum y_i)^2}{n}$

$$= y_1^2 + y_2^2 + y_3^2 - \frac{(y_1 + y_2 + y_3)^2}{n}$$

$$= (5)^2 + (8)^2 + (2)^2 - \frac{(5+8+2)^2}{n} = 25 + 64 + 4 - \frac{(15)^2}{3}$$

$$= 25 + 64 + 4 - \frac{(15)^2}{3} = 93 - \frac{225}{3} = 18 \quad .$$

B/ $\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

$$= (x_1y_1) + (x_2y_2) + (x_3y_3) - \frac{(x_1 + x_2 + x_3)(y_1 + y_2 + y_3)}{n}$$

$$= (7 * 5) + (-2 * 8) + (4 * 2) - \frac{(7 + (-2) + 4)(5 + 8 + 2)}{3}$$

$$= (35) + (-16) + (8) - \frac{(9)(15)}{3} = 27 - 45 = -18 \quad .$$

C/ $\sum (x_i + y_i)(x_i - y_i)$

$$= (x_1 + y_1)\left(x_1 - y_1\right) + (x_2 + y_2)\left(x_2 - y_2\right) + (x_3 + y_3)\left(x_3 - y_3\right)$$

$$= (7+5)\left(7-5\right) + \left(-2+8\right)\left(-2-8\right) + (4+2)\left(4-2\right)$$

$$= (12)(2) + (6)(-10) + (6)(2)$$

$$= 24 + \left(-60\right) + 12 = -24 \quad .$$

D/ $\sum (x_i - 8)$

$$(x_1 - 8) + (x_2 - 8) + (x_3 - 8)$$

$$(7 - 8) + (-2 - 8) + (4 - 8)$$

$$(-1) + (-10) + (- 4) = -2 \quad .$$

E/ $\sum x_i - 8$

$$= x_1 + x_2 + x_3 - 8$$

$$= 7 + (- 2) + 4 - 8 = 9 - 2 = 1 \quad .$$

F/ $\sum \frac{(y_i{}^2 - 10)}{2x_i}$

$$= \frac{y_1{}^2 - 10}{2x_1} + \frac{y_2{}^2 - 10}{2x_2} + \frac{y_3{}^2 - 10}{2x_3}$$

$$= \frac{(5)^2 - 10}{14} + \frac{(8)^2 - 10}{2(-2)} + \frac{(2)^2 - 10}{24}$$

$$= \frac{25 - 10}{14} + \frac{64 - 10}{-4} + \frac{4 - 10}{24}$$

$$= \frac{15}{14} + \frac{54}{-4} + \frac{6}{24}$$

$$= 1.071 + (-13.5) + (- 0.75) = 13.179 \quad .$$

G/ $\sum \frac{(y_i^2 - 10)}{\sum 2x_i}$

$$= \frac{(y_1^2 - 10) + (y_2^2 - 10) + (y_3^2 - 10)}{2 \sum x_I}$$

$$= \frac{(5^2 - 10) + (8^2 - 10) + (2^2 - 10)}{2(7 + (-2) + 4)}$$

$$= \frac{(25 - 10) + (64 - 10) + (4 - 10)}{2(9)}$$

$$= \frac{(15) + (54) + (-6)}{18}$$

$$= \frac{63}{18} = 3.5 \; .$$

6 – برهن بأن $\sum y_i^2$ لا تساوي $(\sum y_i)^2$

$$(y_1^2 + y_2^2 + y_3^2) \neq (y_1 + y_2 + y_3)^2$$

وإن : $\sum x_i y_i$ لا تساوي $(\sum x_i)(\sum y_i)$

$$(x_1 y_1 + x_2 y_2 + x_3 y_3) \neq (x_1 + x_2 + x_3)(y_1 + y_2 + y_3)$$

7– إذا علمت بأن :

$$\bar{x} = \frac{\sum x_i}{n} \quad , \quad \bar{y} = \frac{\sum y_i}{n}$$

برهن بأن

A/ $$\sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= \sum (x_i^2 - 2 x_i \bar{x} + \bar{x}^2)$$

$$= \sum x_i^2 - 2 \sum x_i \bar{x} + \sum \bar{x}^2$$

$$= \sum x_i^2 - 2 \sum x_i * \frac{\sum x_i}{n} + n \frac{\sum x_i^2}{n^2}$$

$$= \sum x_i^2 - 2 \frac{\sum x_i^2}{n} + \frac{\sum x_i^2}{n}$$

$$= \sum x_i^2 - \frac{\sum x_i^2}{n} \quad .$$

B/ $$\sum (y_i - \bar{y}) y_i = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= \sum (y_i^2 - y_i \bar{y})$$

$$= \sum y_i^2 - \sum y_i \bar{y}$$

$$= \sum y_i^2 - \sum y_i * \frac{\sum y_i}{n}$$

$$= \sum y_i^2 - \frac{\sum y_i^2}{n} \quad .$$

$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$ الحل الثاني لفرع **(B)**

$$= \sum (y_i^{\,2} - 2\, y_i\, \bar{y} + \bar{y}_i^{\,2})$$

$$= \sum y_i^{\,2} - 2 \sum y_i\, \bar{y} + \sum \bar{y}_i^{\,2}$$

$$= \sum y_i^{\,2} - 2 \sum y_i * \frac{\sum y_i}{n} + n\, \frac{\sum y_i^{\,2}}{n^2}$$

$$= \sum y_i^{\,2} - 2\, \frac{\sum y_i^{\,2}}{n} + \frac{\sum y_i^{\,2}}{n}$$

$$= \sum y_i^{\,2} - \frac{\sum y_i^{\,2}}{n} \;.$$

C/ $\sum (\bar{x}_i - x)(y_i - \bar{y}) = \sum (x_i - \bar{x})\, y_i = \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

$$= \sum (x_i\, y_i - x_i\, \bar{y} - \bar{x}\, y_i + \bar{x}\, \bar{y})$$

$$= \sum x_i\, y_i - \sum x_i\, \bar{y} - \sum \bar{x}\, y_i + \sum \bar{x}\, \bar{y}$$

$$= \sum x_i\, y_i - \sum x_i * \frac{\sum y_i}{n} - \frac{\sum x_i}{n} * \sum y_i + n\, \frac{\sum x_i}{n} * \frac{\sum y_i}{n}$$

$$= \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n} - \frac{(\sum x_i)(\sum y_i)}{n} + \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n} \qquad .$$

$\sum (x_i - \bar{x})\, y_i = \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n}$ الحل الثاني لفرع **(C)**

$$= \sum (x_i\, y_i - \bar{x}\, y_i)$$

$$= \sum x_i\, y_i - \sum \bar{x}\, y_i$$

$$= \sum x_i\, y_i - \frac{\sum x_i}{n} * \sum y_i$$

$$= \sum x_i\, y_i - \frac{(\sum x_i)(\sum y_i)}{n} \quad .$$

٨- إذا علمت بأن :

$$\sum x_i = -4 \qquad , \qquad \sum x_i^{\,2} = 10$$

احسب كلاً من

A/ $\sum (2\, x_i - 3)$.

$$\sum 2\, x_i + \sum 3$$

$$2 \sum x_i + \sum 3$$

2 – (−4) + 4,3 = −8 + 12 = 4

B/ $\sum x_i\, (x_i - 1)$

$$\sum x_i^{\,2} - \sum x_i$$

10 – (−4) = 14

C/ $\sum (x_i - 5)^2$

$$\sum (x_i^{\,2} - 2\, x_i\, 5 + 5^2)$$

$$\sum x_i^2 - 2\sum x_i\, 5 + \sum 25$$

$$10 - 2\,(-4)(5) + 4(25)$$

$$10 - (-40) + 100 = 150 \quad .$$

(**الجمع و الطرح والضرب والقسمة**) .

الجمع	الطرح	الضرب	القسمة
+ = (+) + (+)	نأخذ إشارة العدد الأكبر = (+) − (+)	+ = (+) × (+)	+ = (+) ÷ (+)
− = (−) + (−)	نأخذ إشارة العدد الأكبر = (−) − (−)	+ = (−) × (−)	+ = (−) ÷ (−)
نأخذ إشارة العدد الأكبر = (−) + (+)	− = (−) − (+)	− = (−) × (+)	− = (−) ÷ (+)
نأخذ إشارة العدد الأكبر = (+) + (−)	+ = (+) − (−)	− = (+) × (−)	− = (+) ÷ (−)

المدخل إلى الإحصاء

Introduction to statistics

الطبعة الثانية/ ٢٠٠٠

الفصل الثالث

المدرس المساعد

خليل ابراهيم خليل

وزارة التعليم العالي والبحث العلمي .
جامعة الموصل .
كلية الزراعة والغابات .
قسم المحاصيل الحقلية .

Ministry of Higher Education and Scientific Research .
University of Al Mosul .
College of Agriculture and Forestry .
Field crops department .

المدخل إلى الإحصاء

Introduction to statistics

الدكتور خاشع محمود الراوي

الطبعة الثانية/ ٢٠٠٠

الفصل الثالث

العرض الجدولي والتمثيل البياني

Tabular presentation and Graphical Presentation

المقدمة Interdiction :

عند جمع البيانات الأولية (Raw data) الخاصة بدراسة ظاهرة ما فإنه عادةً لا يمكن الاستفادة منها وهي بهذه الصورة . لذلك فغالباً ما توضع في جداول مبسطة أو يعبر عنها في صورة أشكال ورسوم بيانية لكي يسهل دراستها وتحليلها .

(1:3) العرض الجدولي : Tabular presentation

هناك نوعان رئيسيان من الجداول الاحصائية وهما :

(1) الجدول البسيط (Simple table):

وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة ويتألف عادةً من عمودين **<u>الأول</u>** يمثل تقسيمات الصفة أو الظاهرة الى فئات أو مجموعات **<u>والثاني</u>** يبين عدد المفردات التابعة لكل فئة أو مجموعة منها جدول(1:3) و (3 : 2) .

جدول (1:3) جدول توزيع عدد من طلبة جامعة ما حسب أوزانها (بالكيلوغرامات) .

فئات الوزن (كغم)	عدد الطلبة
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27
72 – 74	8
المجموع	100

جدول (2:3) جدول توزيع أعضاء البعثات الموفدين إلى الخارج حسب مواد الدراسة لسنة 1970 / 1971.

موضوع البعثة	عدد الطلبة
علوم اساسية	25
علوم زراعية	50
علوم بيطرية	20
علوم هندسية	75
علوم طبية	50
علوم اجتماعية	30
المجموع	250

(2) الجدول المركب (Composite table) :

هو الجدول الذي توزع فيه البيانات حسب صنفين أو ظاهرتين أو أكثر في نفس الوقت .

فمثلاً الجدول المزدوج (لصنفين) يتألف من :

a) الصفوف (the classes) : وتمثل فئات أو مجاميع احدى الصنفين .

b) الأعمدة (columns) : وتمثل فئات أو مجاميع الصفة الاخرى .

أما المربعات التي تقابل الصفوف والاعمدة فتحتوي على عدد المفردات أو التكرارات المشتركة في فئات ومجاميع كلا الصفيين مثل الجدول التالي .

جدول (3:3) جدول توزيع عدد من طلبة كلية ما حسب صفتي الطول والوزن .

الوزن (كغم) / الطول (سم)	51 – 60	61 – 70	71 – 80	المجموع
121 – 140	20	6	4	30
141 – 160	2	40	10	52
161 – 180	2	6	10	18
المجموع	24	52	24	100

هذا وسنشرح الأن بالتفصيل كيفية إنشاء أو تكوين جدول من الجداول البسيطة كثير الاستعمال يدعى بجدول التوزيع التكراري Frequence Table .

(2:3) جدول التوزيع التكراري Frequency Distribution or Frequence Table

جدول التوزيع التكراري : هو جدول بسيط يتكون من عمودين :

<u>الاول</u> : ويقسم فيه قيم المتغير إلى أقسام أو مجموعات تدعى الفئات Classes .

<u>الثاني</u> : يبين مفردات كل فئة ويسمى بالتكرار Frequency .

جدول (4:3) جدول التوزيع التكراري لأطوال ٨٠ نباتاً من القطن (بالسنتمترات) .

فئات الوزن (كغم)	عدد الطلبة
31 – 40	1
41 – 50	2
51 – 60	5
61 – 70	15
71 – 80	25
81 – 90	20
91 – 100	12
المجموع	100

1- بعض التعاريف المهمة :

❖ **البيانات غير المبوبة (Ungrouped data) :**

وهي البيانات الأولية أو الأصلية (Raw data) التي جمعت ولم تبوب .

❖ **البيانات المبوبة (Grouped data) :**

وهي البيانات التي بوبت ونظمت في جدول التوزيع التكراري .

❖ **الفئات (Classes) :**

وهي المجاميع التي قسمت إليها قيم المتغير . وكل فئة تأخذ مدى معين من قيم المتغير فالجدول (4:3) يحتوي على سبع فئات .

❖ **حدود الفئات (Class limits) :**

لكل فئة حدان . حد أدنى Lower class limit وحد أعلى Upper class limit .

❖ **الحدود الحقيقية للفئات (Class boundaries or True class limit) :**

لكل فئة حدان حقيقيان حد أنى حقيقي Lower class boundary حد أعلى حقيقي Upper class boundary

❖ **طول الفئة (Class length or class width) :**

وهو مقدار المدى بين حدي الفئة ، هذا ويستحسن أن تكون أطوال الفئات متساوية لتسهيل العمليات الحسابية . وسنرمز لطول الفئة بالرمز (C) .

❖ **مركز الفئة (Class mark or class mid-point) :**

لكل فئة مركز وسنرمز له بـ (yi) وهو عبارة عن منتصف المدى بين حدي الفئة .

❖ **تكرار الفئة (Class frequency) :**

وهي عدد المفردات أو القيم التي تقع في مدى تلك الفئة وسنرمز لها بـ (fi) هذا ومجموع التكرارات يجب أن يكون دائماً مساوياً للعدد الكلي لقيم الظاهرة .

حيث أن جدول (3 : 5) يوضح ما سبق شرحه بالتفصيل .

جدول (3 : 5) جدول التوزيع التكراري لأطوال نباتات القطن مبيناً فيه الحدود الحقيقية ومراكز الفئات .

تسلسل الفئات	الفئات	الحدود الحقيقية للفئات	مركز الفئة (yi)	التكرار (fi)
1	40 – 31	40,5 – 30,5	35,5	1
2	50 – 41	50,5 – 40,5	45,5	2
3	60 – 51	60,5 – 50,5	55,5	5
4	70 – 61	70,5 – 60,5	65,5	15
5	80 – 71	80,5 – 70,5	75,5	25
6	90 – 81	90,5 – 80,5	85,5	20
7	100 – 91	100,5 – 90,5	95,5	12
المجموع				80

خذ مثلاً الفئة الرابعة = (61 – 70) .

فأحد الأدنى للفئة الرابعة = 61 .

والحد الأعلى للفئة الرابعة = 70 .

طول الفئة

وطول الفئة الرابعة : يمكن حساب طول الفئة من جدول التوزيع التكراري بإحدى الطرق التالية .

<u>الطريقة الأولى</u> : (عندما تكون حدود الفئات أعداداً صحيحة فقط) .

طول الفئة = الحد الأعلى – الحد الأدنى + 1 .

طول الفئة = 70 – 61 + 1 = 10 .

<u>الطريقة الثانية</u> :

طول الفئة = الحد الحقيقي الأعلى – الحد الحقيقي الأدنى لتلك الفئة .

طول الفئة = 70,5 – 60,5 = 10 .

<u>الطريقة الثالثة</u> :

طول الفئة = الفرق بين الحدين الأدنى أو الحدين الأعلى لفئتين متتاليتين .

الفرق بين الحدين الأدنى = 71 – 61 = 10 .

الفرق بين الحدين الأعلى = 80 – 70 = 10 .

<u>الطريقة الرابعة</u> :

طول الفئة = الفرق بين الحدين الحقيقين الأدنى أو الأعلى لفئتين متتاليتين .

طول الفئة = 70,5 – 60,5 = 10

طول الفئة = 80,5 – 70,5 = 10

<u>الطريقة الخامسة</u> :

طول الفئة = الفرق بين مركزي فئتين متتاليتين .

طول الفئة = 75,5 – 65,5 = 10

الحدود الحقيقية

الحدود الحقيقية : ويمكن حساب الحدود الحقيقية لأي فئة بإحدى الطرق التالية :

<u>الطريقة الأولى</u> :

الحد الأدنى الحقيقي لأي فئة = مركز تلك الفئة − $\frac{1}{2}$ (طول تلك الفئة) .

الحد الأدنى الحقيقي للفئة الرابعة = مركز الفئة الرابعة − $\frac{1}{2}$ (طول الفئة الرابعة) .

الحد الأدنى الحقيقي للفئة الرابعة = $65,5 - \frac{1}{\cancel{2}_{1}}(\cancel{10}^{5}) = 60,5$

أما الحد الأعلى الحقيقي = مركز الفئة + $\frac{1}{2}$ (طول الفئة) .

فالحد الحقيقي للفئة الرابعة = $65,5 + \frac{1}{\cancel{2}_{1}}(\cancel{10}^{5}) = 70,5$

<u>الطريقة الثانية</u> :

الحد الأدنى الحقيقي لأي فئة = $\frac{\text{الحد الأدنى لتلك الفئة + الحد الأعلى للفئة السابقة}}{2}$

الحد الأدنى الحقيقي للفئة الرابعة = $\frac{60 + 61}{2} = 60,5$

الحد الأعلى الحقيقي لأي فئة = $\frac{\text{الحد الأعلى لتلك الفئة + الحد الأدنى للفئة التي تليها}}{2}$

الحد الأعلى الحقيقي لأي فئة = $\frac{71 + 70}{2} = 70,5$

ملاحظة

إذا كانت حدود الفئات أعداد صحيحة فإن

الحد الأدنى الحقيقي لأي فئة = الحد الأدنى لتلك الفئة − 0,5

الحد الحقيقي لأي فئة = الحد الأعلى لتلك الفئة + 0,5

ومركز <u>الفئة</u> : وتحسب بإحدى الطريقتين التاليتين :

<u>ك٢الطريقة الأولى</u> :

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى} + \text{الحد الأعلى}}{2} .$$

$$\text{فمركز الفئة الرابعة} = \frac{70 + 61}{2} = 65,5$$

$$\text{مركز الفئة} = \frac{\text{الحد الأدنى} + \text{الحد الأعلى}}{2} .$$

<u>الطريقة الثانية</u> :

$$\text{فمركز الفئة الرابعة} = \frac{70 + 61}{2} = 65,5$$

تكرار الفئة الرابعة = 15 أي أن هناك 15 قيمة من قيم المتغير واقعة في المدى (60 − 70) .

2− الخطوات العامة في إنشاء جدول التوزيع التكراري General Rules Constructing Frequency table

a) استخراج مدى المتغير Range .

b) اختيار وتحديد عدد الفئات Number of classes .

c) إيجاد طول مدى الفئة Class length or width .

d) كتابة حدود الفئات Class limits .

e) استخرج عدد التكرارات لكل فئة Class frequency .

مثال (1)

والمثال التالي يوضح كيفية إنشاء جدول التوزيع التكراري لنباتات القطن .

مثال (1) القيم التالية تمثل أطوال 80 نباتاً من نباتات القطن (مقربة إلى أقرب سنتمتر) والمطلوب إنشاء جدول توزيع تكراري لأطوال هذه النباتات .

جدول (3 : 5) أطوال 80 نباتاً من نباتات القطن مقدرة بالسنتمترات .

80	87	98	81	74	48	79	80
78	82	93	91	70	90	80	84
73	74	81	56	65	92	70	71
86	83	93	65	51	85	68	72
68	86	43	74	73	83	90	35
75	67	72	90	71	76	92	93
81	88	91	97	72	61	80	91
77	71	59	80	95	99	70	74
63	89	67	60	82	83	63	60
75	89	88	66	70	88	76	63

الحـل : نتبع الخطوات التالية :

a) استخراج المدى أو مدى المتغير The Range .

المدى = أعلى قيمة – أقل قيمة

فأطول نبات = 99 سم بينما أقصر نبات = 35 سم .

لذا فالمدى = 99 – 35 = 64 سم .

b) اختيار وتحديد عدد الفئات Number of classes .

هناك عدة طرق حسابية تقريبية لإيجاد عدد الفئات أهمها :

طريقة سترجس (Sturges) .

عدد الفئات = 1 + (3,3 × لوغارتم عدد المفردات) .

طريقة يوك (Yule)

عدد الفئات = 2,5 × $\sqrt[4]{\text{عدد المفردات}}$

ولكل من الطريقتين ميزات وعيوب ولن نستعمل أيا منها هنا بل إننا سنختار عدد الفئات اختياراً على أن لا تقل عن خمسة ولا تزيد عن خمسة عشر فئة وذلك تبعاً لطبيعة البيانات وعدد مفرداتها ومدى التغير فيها .

c) إيجاد طول الفئة Class length .

يجب أن لا يقل طول الفئة عن = $\frac{\text{مدى التغير}}{\text{عدد الفئات}}$ مقربة إلى أقرب عدد صحيح أكبر .

طول الفئة = $\frac{64}{7} = 9\frac{1}{7}$

لذا يستحسن أن يكون طول الفئة = 10

d) كتابة حدود الفئات Class limits .

يجب كتابة حدود الفئات بحيث أن جميع قيم المتغير تقع بين الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة .

يستحسن أن نبدأ بكتابة الحد الأدنى للفئة الأولى بقيمة أصغر مفردة أو أقل من ذلك بقليل وتنتهي بالحد الأعلى للفئة الأخيرة بقيمة أكبر مفردة أو أكثر من ذلك بقليل .

فمثلاً أصغر قيمة من قيم أطوال نباتات القطن هي 35 لذا فمن الممكن أن يكون الرقم 31 يمثل الحد الأدنى للفئة الأولى . وبما أن أطوال الفئة هو 10 لذا فإن حدي الفئة الأولى هما 31 – 40 والفئة الثانية تبدأ من 40 - 50 بينما الفئة السابعة (الأخيرة) هي 91 – 100 . لأحظ بأن الحد الأدنى للفئة الأولى (31) والحد الأعلى للفئة الأخيرة (100) تحتوي على كافة قيم المتغير .

e) استخراج عدد التكرارات لكل فئة Class frequency .

ويتم ذلك بتسجيل القيم الأصلية واحدة بعد الأخرى في الفئة الخاصة به على شكل إشارات أو علامات أولاً ثم ترجمتها إلى أرقام كما مبين في جدول (3 : 6) أدناه .

جدول (3 : 6) توزيع تكراري لأطوال نبات القطن .

الفئات	التكرارات رقماً	التكرارات (بالعلامات)	**الحدود الحقيقية للفئات** (هذا العمود للتوضيح فقط)	**مركز الفئة** (هذا العمود للتوضيح فقط)
40 – 30	1	𝍷	40,5 – 30,5	35,5
50 – 41	2	𝍷𝍷	50,5 – 40,5	45,5
60 – 51	5	𝍸	60,5 – 50,5	55,5
70 – 61	15	𝍸 𝍸 𝍸	70,5 – 60,5	65,5
80 – 71	25	𝍸 𝍸 𝍸 𝍸 𝍸	80,5 – 70,5	75,5
90 – 81	20	𝍸 𝍸 𝍸 𝍸	90,5 – 80,5	85,5
100 – 91	12	𝍸 𝍸 𝍷𝍷	100,5 – 90,5	95,5
المجموع Total	80			

هذا ويجب التأكد بأن المجموع الكلي للتكرارات يجب أن تساوي للعدد الكلي لقيم المتغير

لأحظ أنه في المثال السابق كانت أطوال الفئات متساوية وأرقام صحيحة . والأن سنأخذ مثالاً آخر فيه أطوال الفئات متساوية ولكنها أرقام ذات كسور .

مثال (2)

مثال (2): القيم التالية تمثل كمية المحصول (طن/هكتار) لحنطة المكسيباك في أربعين مزرعة مقدرة بالأطنان ومقربة إلى أقرب رقم عشري واحد .

جدول (3 : 7) كمية المحصول (طن/هكتار) لحنطة المكسيباك في أربعين مزرعة .

2,6	2,2	4,1	3,5	2,0	3,2	3,7	3,0
3,7	3,4	1,6	3,1	3,3	3,8	3,1	2,4
3,1	2,5	4,3	3,4	3,6	2,9	3,3	3,9
3,4	3,3	3,1	3,7	4,4	3,2	4,1	1,9
3,5	2,3	3,8	3,2	2,6	3,9	3,0	4,2

a) ستخرج المدى :

المدى = أعلى قيمة – أقل قيمة

المدى = 4,4 – 1,6 = 2,8 طن .

b) اختيار وتحديد عدد الفئات : (سنختار عدد الفئات هنا 6 فئات)

c) إيجاد طول الفئة :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} \longleftarrow \text{طول الفئة} = \frac{2,8}{6} = 0,466 \simeq 0,467$$

لذا يستحسن أن تكون طول الفئة (0,5) .

d) كتابة حدود الفئات :

بما أن أقل قيمة للمتغير = 1,6 لذا فسنبدأ بكتابة الحد الأدنى للفئة الأولى بـ 1,5 . وبما أن طول الفئة 0,5 لذا فالفئة الأولى ستكون (1,5 – 1,9) والثانية (20 – 2,4) وهكذا الى أن تصل الفئة الأخيرة وهي (4,0 – 4,4) .

e) استخرج عدد التكرارات لكل فئة : نسجل عدد المشاهدات أو المفردات التابعة لكل فئة .

ويجب التأكد بأن مجموع التكرارات الكلي مساوية للعدد الكلي لقيم المتغير وجدول (8:3) يبين التوزيع التكراري لكمية المحصول لحنطة المكسيباك إضافة إلى الحدود الحقيقية ومراكز الفئات .

جدول (3 : 8) جدول التوزيع التكراري لكمية المحصول لحنطة المكسيباك .

التسلسل	حدود الفئات	الحدود الحقيقية للفئات	مركز الفئة	التكرار
1	1,9 − 1,5	1,95 – 1,45	1,7	2
2	2,4 – 20	2,45 – 1,95	2,2	4
3	2,9 – 2,5	2,95 – 2,45	2,7	4
4	3,4 – 30	3,45 – 2,95	3,2	15
5	3,9 – 3,5	3,95 – 3,45	3,7	10
6	4,4 – 40	4,45 – 3,95	4,2	5
	المجموع Total			40

ملاحظة

إذا كانت أعداد قيم المتغير قليلة (أي إذا كان حجم العينة صغير) فليس من الضروري عمل جدول توزيع تكراري لها .

وبالرغم من أن حجم العينة في كلا المثالين صغيراً فالغاية من عمل جدول توزيع تكراري هنا هو فقط توضيح وتبسيط كيفية إنشاء جدول توزيع تكراري بإستخدام أرقام بسيطة وقليلة .

(3:3) جدول التوزيع التكراري النسبي Relative Frequency Distribution

وهو جدول يبين الأهمية النسبية لكل فئة . ويحسب التكرار النسبي لكل فئة بالطريقة التالية :

$$\text{التكرار النسبي لأي فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} = \frac{fi}{\sum fi}$$

ومن الجدول (4:3) فإن :

$$\text{التكرار النسبي للفئة الرابعة} = \frac{\text{تكرار الفئة الرابعة}}{\text{المجموع الكلي للتكرارات}} = \frac{15}{80} = 0,1875$$

وعادةً يوضع التكرار النسبي كنسبة مئوية وذلك بضرب كل تكرار نسبي × 100 كما مبين في جدول (9:3) .

جدول (9:3) جدول توزيع التكراري النسبي والمئوي لأطوال نباتات القطن .

الفئات	التكرار (f_i)	التكرار النسبي	التكرار المئوي
31 – 40	1	0,0125	1,25
41 – 50	2	0,0250	2,50
51 – 60	5	0,0625	6,25
61 – 70	15	0,1875	18,75
71 – 80	25	0,3125	31,25
81 – 90	20	0,2500	25,00
91 – 100	12	0,1500	15,00
المجموع Total	$\Sigma f_i = 80$	1,0000	100,00

(4:3) التوزيعات المتجمعة : Cumulative Distribution

إن جدول التوزيع التكراري العادي الذي سبق شرحه يبين توزيع قيم المتغير على الفئات المختلفة . ولكن في بعض الأحيان قد يكون هناك حاجة إلى معرفة عدد القيم أو المفردات التي تقل أو تزيد عن قيمة معينة . والجداول التي تحوي على مثل هذه المعلومات تدعى بالجداول التكرارية المتجمعة .

وهناك نوعان من هذه الجداول .

(1) جدول التوزيع التكراري التجميعي التصاعدي less than cumulative distribution .

وهو الجدول الذي يعطينا عدد المفردات التي تقل قيمتها عن الحد الأدنى لفئة معينة .

وسنرمز للتكرار المتجمع لأي فئة بـ $\mathbf{F_i}$ وجدول التوزيع التكراري المتجمع التصاعدي يتكون من عمودين :

العمود الأول : نكتب فيه حدود الفئات كما موضح في جدول (10:3) .

العمود الثاني : نكتب فيه التكرار التجميعي التصاعدي بالشكل التالي :

تكرار ما قبل الفئة الأولى = F_0 = صفر

تكرار الفئة الأولى = $F_1 = f_1$

تكرار الفئة الثانية = $F_2 = f_1 + f_2$

تكرار الفئة الثالثة = $F_3 = f_1 + f_2 + f_3$

وهكذا بحيث أن التكرار التجميعي التصاعدي للفئة الأخيرة = $F_n = \Sigma f_i$

جدول (10:3) التوزيع التكراري التجميعي التصاعدي لأطوال نبات القطن .

حدود الفئات	التكرار التجميعي التصاعدي	**ملاحظة** : طريقة إيجاد التكرار التجميعي التصاعدي
أقل من 31	0	F0 = f0 = 0 تكرار ما قبل الفئة الأولى
أقل من 41	1	F1 = f1=1 تكرار الفئة الأولى
أقل من 51	3	F2 = f2+f1=3 تكرار الفئة الثانية
أقل من 61	8	F3 = f3+f2+f1=8 تكرار الفئة الثالثة
أقل من 71	23	F4 = f1+f2+f3+f1=23 و هكذا
أقل من 81	48	F5 = f5+f4+f3+f2+f1=48 و هكذا
أقل من 91	68	F6 = f6+f5+f4+f3+f2+f1 =68 و هكذا
أقل من 101	80	F7 = f7+f6+f5+f4+f3+f2+f1 =80 و هكذا

(2) جدول التوزيع التكراري التجميعي التنازلي "More than" Cumulation distribution .

وهو الجدول الذي يعطينا عدد المفردات التي تزيد قيمتها عن الحد الأدنى لفئة معينة . وهذا الجدول أيضاً يتألف من عمودين .

- العمود الأول : نكتب فيه حدود الفئات .
- العمود الثاني : تكتب فيه التكرارات التجميعية التنازلية بالطريقة التالية :

تكرار الفئة الأولى $= F_1 = \Sigma f_i$

تكرار الفئة الثانية $= F_2 =$ مجموع التكرارات − تكرار الفئة الأولى .

أي : $F_2 = \Sigma f_i - f_1 = F1 - f_1$

تكرار الفئة الثالثة $= F_3 = \Sigma f_i - f_1 - f_2$ أو $F_3 = F_2 - f_2$

وهكذا كما مبين في جدول (11:3) .

جدول (11:3) التوزيع التكراري التجميعي التنازلي لأطوال نباتات القطن .

حدود الفئات	التكرار التجميعي التصاعدي	**ملاحظة** : طريقة إيجاد التكرار التجميعي التنازلي
31 فأكثر	80	تكرار الفئة الاولى $F_1 = \Sigma f_i = 80$
41 فأكثر	79	تكرار الفئة الثانية $F_2 = \Sigma f_i - f_1 = 79$
51 فأكثر	77	تكرار الفئة الثالثة $F_3 = \Sigma f_i - f_2 - f_1 = 77$
61 فأكثر	72	تكرار الفئة الرابعة $F_4 = \Sigma f_i - f_3 - f_2 - f_1 = 72$
71 فأكثر	57	و هكذا $F_5 = \Sigma f_i - f_4 - f_3 - f_2 - f_1 = 57$
81 فأكثر	32	و هكذا $F_6 = \Sigma f_i - f_5 - f_4 - f_3 - f_2 - f_1 = 32$
91 فأكثر	12	و هكذا $F_7 = \Sigma f_i - f_6 - f_5 - f_4 - f_3 - f_2 - f_1 = 12$
101 فأكثر	0	و هكذا $F_8 = \Sigma f_i - f_7 - f_6 - f_5 - f_4 - f_3 - f_2 - f_1 = 0$

هذا وأحياناً يعبر عن التكرار التجميعي التصاعدي أو التنازلي بشكل تكرار تجميعي نسبي أو مئوي .

وفي هذه الحالة فإن التكرار التجميعي النسبي لأي فئة = $\frac{\text{التكرار التجميعي لتلك الفئة}}{\text{المجموع الكلي}}$.

أما التكرار التجميعي المئوي = التكرار التجميعي النسبي × 100 .

(5:3) أمثلة محلولة .

مثال (2)

مثال (2) الجدول التالي يبين التوزيع التكراري للرواتب الشهرية مقدرة بالدينار لـ (65) موظفاً في إحدى الشركات.

فئات الأجور	التكرار (عدد المستخدمين)
50 – 59	8
60 – 69	10
70 – 79	16
80 – 89	14
90 – 99	10
100 – 109	5
110 – 119	2
المجموع	65

المطلوب إيجاد قيمة كل مما يلي :

a) الحد الأدنى للفئة السادسة ؟ **الحل** = 100

b) الحد الأعلى للفئة الرابعة ؟ **الحل** = 89

c) مركز الفئة الخامسة ؟ **الحل** = $\frac{99+90}{2}$ = 94,5

d) طول الفئة الخامسة ؟

الحل / طول الفئة الخامسة = الحد الأعلى للفئة الخامسة – الحد الأدنى للفئة الخامسة +1 .

طول الفئة الخامسة = 99 – 90 + 1 = 10

e) الحد الأدنى الحقيقي للفئة الخامسة ؟

الحل / الحد الأدنى الحقيقي للفئة الخامسة = مركز الفئة الخامسة – $\frac{1}{2}$ (طول الفئة الخامسة) .

الحد الأدنى الحقيقي للفئة الخامسة = $94,5 - \frac{1}{2}(10) = 89,5$

أو الحد الادنى الحقيقي للفئة الخامسة = $\frac{\text{الحد الأدنى للفئة الخامسة + الحد الأعلى للفئة الرابعة}}{2}$.

الحد الادنى الحقيقي للفئة الخامسة = $\frac{89+90}{2} = \frac{179}{2} = 89,5$

f) تكرار الفئة الثالثة ؟ **الحل** = 16

g) التكرار النسبي للفئة الثالثة ؟ **الحل** = $\frac{16}{65} = \frac{16}{65} = 0,246$ (**ملاحظة** نأخذ رقمان أو ثلاث أرقام بعد الفارزة) .

مثال (3)

مثال (3) أكمل جدول التوزيع التكراري التالي .

الفئات	مركز الفئة	الحدود الحقيقية	التكرار	التكرار النسبي	التكرار المئوي
6 – 2	4	6,5 – 1,5	2	0,04	4
11 – 7	9	11,5 – 6,5	5	0,1	10
16 – 12	14	16,5 – 11,5	10	0,2	20
21 – 17	19	21,5 – 16,5	25	0,5	50
26 – 22	24	26,5 – 21,5	8	0,16	16
المجموع			50		

الحل :

طول الفئة = الفرق بين مركزي فئتين متتاليتين .

طول الفئة = $9 - 4 = 5$

الحد الأدنى الحقيقي للفئة الأولى = مركز الفئة الأولى − $\frac{1}{2}$ (طول الفئة) .

الحد الأدنى الحقيقي للفئة الأولى = $4 - \frac{1}{2}(5) = 1,5$.

الحد الأعلى الحقيقي للفئة الأولى = مركز الفئة الأولى − $\frac{1}{2}$ (طول الفئة) .

الحد الأدنى الحقيقي للفئة الأولى = $4 + \frac{1}{2}(5) = 6,5$.

يتم إضافة طول الفئة على الحد الأدنى الحقيقي للفئة الأولى لينتج الحد الأدنى الحقيقي للفئة الثانية وهكذا .

أما الحد الأدنى للفئة الأولى فهو أقرب عدد صحيح للحد الأدنى الحقيقي وهو2 أي بإضافة نصف إلى الحد الأدنى الحقيقي بينما الحد الأعلى فهو بطرح نصف من الحد الأعلى الحقيقي . لذا فحدي الفئة الأولى هما (2 − 6) ثم تضاف طول الفئة بعدئذٍ لكل من الحد الأدنى والحد الأعلى لإيجاد حدود الفئات الأخرى .

أما التكرار النسبي لأي فئة = $\frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}}$.

فمثلاً التكرار النسبي للفئة الأولى = $\frac{2}{50} = 0,4 \longrightarrow \frac{2}{50} = 0,04$.

أما التكرار المئوي = التكرار النسبي × 100

كما مبين ذلك في الجدول أدناه .

الفئات	مركز الفئة	الحدود الحقيقية	التكرار	التكرار النسبي	التكرار المئوي
6 – 2	4	6,5 – 1,5	2	0,04	4
11– 7	9	11,5 – 6,5	5	0,10	10
16 – 12	14	16,5 – 11,5	10	0,20	20
21 – 17	19	21,5 – 16,5	25	0,50	50
26 – 22	24	26,5 – 21,5	8	0,16	16
المجموع			50	100	

مثال (4)

مثال (4) نفرض أن عدد مفردات ظاهرة ما هو 150 مفردة وإن أقل قيمة بينها = 5,18 وأعلى قيمة = 7,44 .

فالمطلوب أيجاد .

a) حدود الفئات b) مراكز الفئات c) الحدود الحقيقية للفئات

التي قد تستعمل في إنشاء جدول توزيع تكراري اهذه القيم .

الحل :

a) المدى = 7,44 − 5,18 = 2,26

لنفرض أن عدد الفئات المناسبة المختارة = 8

طول الفئة = $\frac{2,26}{8}$ = 0,28 (≃ إذا طول الفئة سنعتبرها 0,3)

وبما أن أقل قيمة = 5,18 نبدأ بالحد الأدنى للفئة الأولى بـ 5,10 ثم نظيف طول الفئة للحد الأدنى للفئة الأولى لإيجاد الحد الأدنى للفئة الثانية أي 5,10 + 0,30 = 5,40

أما الحدود العليا ،فبما أن قيم الظاهرة مقرب إلى رقمين عشريين ، لذا فإن الحد الأعلى للفئة الأولى = الحد الأدنى للفئة الثانية − 0,01 . أي الحد الأعلى للفئة الأولى = 5,40 – 0,01 = 5,39

ثم نضيف طول الفئة للحد الأعلى للفئة الأولى لإيجاد الحد الأعلى للفئة الثانية وهكذا .

b) ثم نستخرج طول الفئة = $\frac{\text{الحد الأدنى + الحد الأعلى}}{2}$

فمركز الفئة الأولى = $\frac{5,39 + 5,10}{2}$ = $\frac{10,49}{2}$ = 5,245

ملاحظة : إن عيب مركز الفئة هنا أنها لا تطابق قيم المفردات .

c) أما الحدود الحقيقية فتستخرج بالطريقة التالية :

الحد الأدنى الحقيقي = مركز الفئة − $\frac{1}{2}$ (طول الفئة) .

فمثلا الحد الأدنى الحقيقي للفئة الأولى = 5,245 − $\frac{1}{2}$ (0,3) = 5,095 .

والحد الأعلى الحقيقي = مركز الفئة + $\frac{1}{2}$ (طول الفئة) .

الحد الأعلى الحقيقي للفئة الأولى = 5,245 + $\frac{1}{2}$ (0,3) = 5,395 .

ثم إضافة طول الفئة للحد الأدنى الحقيقي للفئة الأولى لإيجاد الحد الأدنى الحقيقي للفئة الثانية وهكذا بالنسبة للحدود العليا الحقيقية أيضاً كما مبين في الجدول أدناه .

حـدود الفئـــات	الحـدود الحقيقيـــة	مراكز الفئـــات
5,39 – 5,10	5,395 – 5,095	5,245
5,69 – 5,40	5,695 – 5,395	5,545
5,99 – 5,70	5,995 – 5,695	5,845
6,29 – 6,00	5,295 – 5,995	6,145
6,59 – 6,30	5,595 – 6,295	6,445
6,89 – 6,60	6,895 – 6,595	6,745
7,19 – 6,90	7,195 – 6,895	7,045
7,49 – 7,20	7,495 – 7,195	7,345

مثال (5)

إذا علمت بأن عدد المفردات المتغير = 50 أي ($\sum f_i = 50$) فمن جدول التوزيع التكراري النسبي التالي :

الفئـــات	التكرار النسبي
39 – 20	0,12
59 – 40	0,28
79 – 60	0,36
99 – 80	0,20
105 – 100	0,04

أوجـــد ما يلـــي لهـــذا الجـــدول :

a) التكرارات　　b) ومراكز الفئات　　c) والحدود الحقيقية　　d) والتكرار المئوي .

الـــحـــل :

a) التكرار النسبي لأي فئة = $\frac{\text{تكرار الفئة}}{\text{التكرار الكلي}}$.

تكرار الفئة = التكرار النسبي × التكرار الكلي

تكرار الفئة الأولى = $6 = 50 \times 0,12$

تكرار الفئة الثانية = $14 = 50 \times 0,28$　　وهكذا لباقي تكرار الفئة .

b) مركز الفئة = $\frac{\text{الحد الأدنى للفئة + الحد الأعلى للفئة}}{2}$

مركز الفئة الأولى = $\frac{39+20}{2} = \frac{59}{2} = 29,5$

مركز الفئة الثانية = $\frac{59+40}{2} = \frac{99}{2} = 49,5$　　وهكذا لباقي مراكز الفئة .

c) أما طول الفئة = الحد الأعلى للفئة – الحد الأدنى للفئة +1

أو طول الفئة = الفرق بين مركزي فئتين متتاليتين

طول الفئة = $20 = 1 + 20 - 39$

d) أما الحد الأدنى الحقيقي = مركز الفئة − $\frac{1}{2}$ (طول الفئة) .

فالحد الأدنى الحقيقي للفئة الأولى = $19,5 = (20)\ \frac{1}{2} - 29,5$

بينما الحد الأعلى الحقيقي للفئة الأولى = $39,5 = (20)\ \frac{1}{2} + 29,5$　　وهكذا .

e) أما التكرار المئوي = التكرار النسبي × 100

فالتكرار المئوي للفئة الأولى = $12 = 15 \times 0,12$

وهكذا كما مبين في الجدول أدناه

الفئات	التكرار	مركز الفئات	الحدود الحقيقية	التكرار النسبي	التكرار المئوي
39 – 20	6	29,5	39,5 – 19,5	0,12	12
59 – 40	14	49,5	59,5 – 39,5	0,28	28
79 – 60	18	69,5	79,5 – 59,5	0,36	36
99 – 80	10	89,5	99,5 – 79,5	0,20	20
119 – 100	2	109,5	119,5 – 99,5	0,04	4
	50			100	100

مثال (6)

مثال (6) الجدول التالي يبين توزيع التكراري لأوزان (65 طالباً بالكيلومترات) .

فئات الأوزان	التكرار (عدد الطلبة)
45 – 50	8
59 – 55	10
64 – 60	16
69 – 65	14
74 – 70	10
79 – 75	5
84 – 80	2
المجموع	65

المطلوب عمل جدول توزيع تكرار تجميعي تصاعدي وتنازلي ونهما استنتج ما يلي :

a) ما هو عدد الطلبة الذين أوزانهم تقل عن 70 كغم ؟

b) ماهي نسبة الطلبة الذين أوزانهم تقل 70 كغم ؟

c) ما هو عدد الطلبة الذين أوزانهم لا تقل عن 60 كغم ؟

d) ما هو عدد الطلبة الذين أوزانهم لا تقل عن 60 كغم ولكنها أقل من 80 كغم ؟

الحـــل :

جدول توزيع تكراري تجميعي تصاعدي .

حدود الفئات	التكرار التجميعي التصاعدي
أقل من 50	0
أقل من 55	8
أقل من 60	18
أقل من 65	34
أقل من 70	48
أقل من 75	58
أقل من 80	63
أقل من 85	65

جدول توزيع تكراري تجميعي تنازلي .

حدود الفئات	التكرار التجميعي التنازلي
50 فأكثر	65
55 فأكثر	57
60 فأكثر	47
65 فأكثر	31
70 فأكثر	17
75 فأكثر	7
80 فأكثر	2
85 فأكثر	0

a) من جدول التوزيع التكراري التجميعي التصاعدي .

عدد الطلبة الذين أوزانهم أقل من 70 = 48

b) أما نسبة هؤلاء الطلبة $= \frac{48}{65} \times 100 = 73,8$

c) من جدول التوزيع التكراري التجميعي التنازلي .

عدد الطلبة الذين أوزانهم لا تقل عن 60 كغم = 47

d) من جدول التوزيع التكراري التجميعي التنازلي .

عدد الطلبة الذين أوزانهم لا تقل عن 60 ولكنها أقل من 80 كغم = 47 – 2 = 45

(6:3) التمثيل البياني : Graphical Presentation

إن الرسوم والصور والأشكال الهندسية ما هي إلا تعبير وتوضيح للبيانات بطريقة جذابة وسهلة وفعالة تساعد القارئ على فهم واستيعاب قيم الظاهرة ومقارنتها مع بعضها .

ووسائل التمثيل البياني كثيرة ومتنوعة وسنكتفي هنا بشرح العرض البياني للتوزيعات التكرارية فقط . وعادةً نخصص المحور الافقي abscisa أو الاحداثي السيني لتمثيل قيم أو فئات المتغير بينما تخصص المحور العمودي ordinate أو الاحداثي التصاعدي لتمثيل تكرارات هذا المتغير ويجب دائماً أن يبدأ تدريج المحور العمودي من الصفر أما تدريج المحور الافقي فقد لا نبدأ بتدريجه من الصفر . كما أنه ليس من الضروري أن يكون مقياس أو تدريج المحورين من نفس المقياس .

1) التمثيل البياني لجدول التوزيع التكراري .

a) المدرج التكراري Histogram

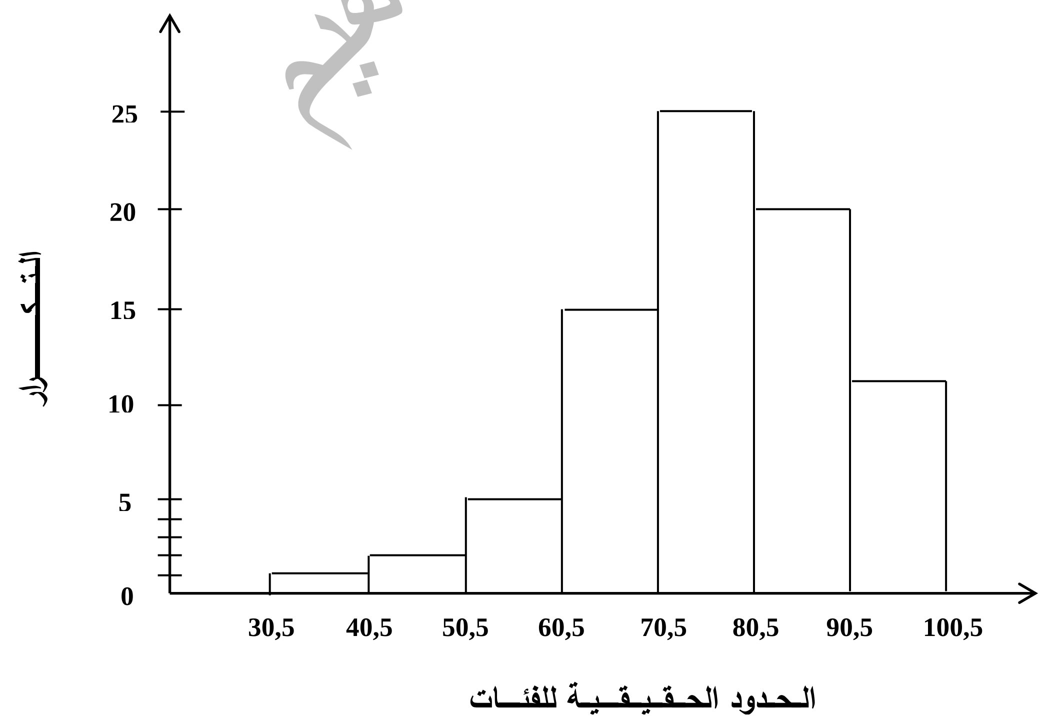
وهو عبارة عن مستطيلات رأسية تمتد قواعدها على المحور الافقي لتمثل أطوال الفئات بينما ارتفاعاتها تمثل تكرارات الفئات .

1- رسم المحور الافقي والمحور العمودي .

2- تدرج المحور الافقي إلى أقسام متساوية بمقياس رسم مناسب بحيث يمثل جميع الحدود الحقيقية للفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الأدنى للفئة الأولى (فيما إذا كانت بداية الفئة الأولى لا تساوي صفر) ويقسم المحور العمودي إلى أقسام متساوية بحيث تشمل على أكبر التكرارات .

3- يرسم على كل فئة مستطيلاً رأسياً تمثل قاعدته طول تلك الفئة وارتفاعه تمثل تكرار تلك الفئة

والشكل (1:3) يمثل المدرج التكراري لجدول (6:3) .



رسم المضلع

b) المضلع التكراري Frequency Polygon

وهو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط كل منها واقعة فوق مركز فئة على إرتفاع يمثل تكرار تلك الفئة . وعادةً يقفل المضلع بأن تصل بداية المضلع بالمحور الافقي بمركز الفئة (خيالية) واقعة إلى يسار أول فئة تكرارها صفراً . وتصل نهاية المضلع بالمحور الافقي بمركز فئة (خيالية) واقعة إلى يمين آخر فئة تكرارها أيضاً صفراً وبذلك تكون مساحة المضلع التكراري مساوية لمساحة المدرج التكراري ولرسم المضلع التكراري نتبع الخطوات التالية :

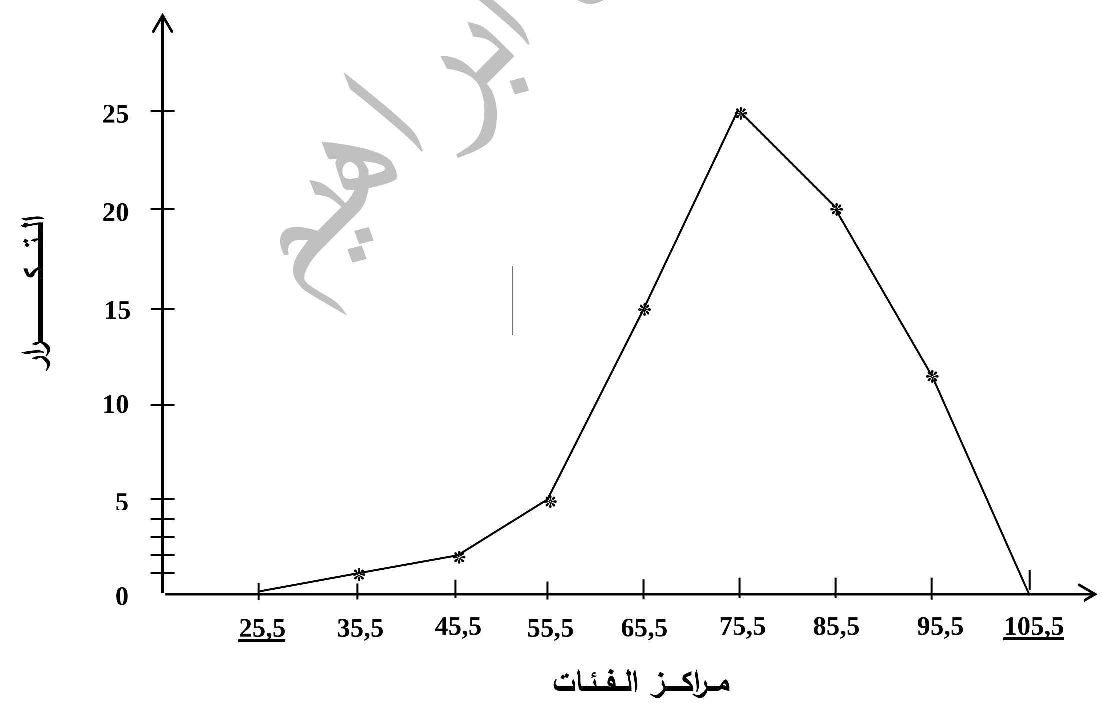
1) رسم المحور الافقي والمحور العمودي .

2) تدريج المحور الافقي إلى أقسام متساوية بحيث يشمل على جميع مراكز الفئات . ويقسم المحور العمودي إلى أقسام متساوية بحيث تشمل على أكبر التكرارات .

3) وضع نقطة أمام مركز كل فئة ارتفاعها يعادل تكرار تلك الفئة .

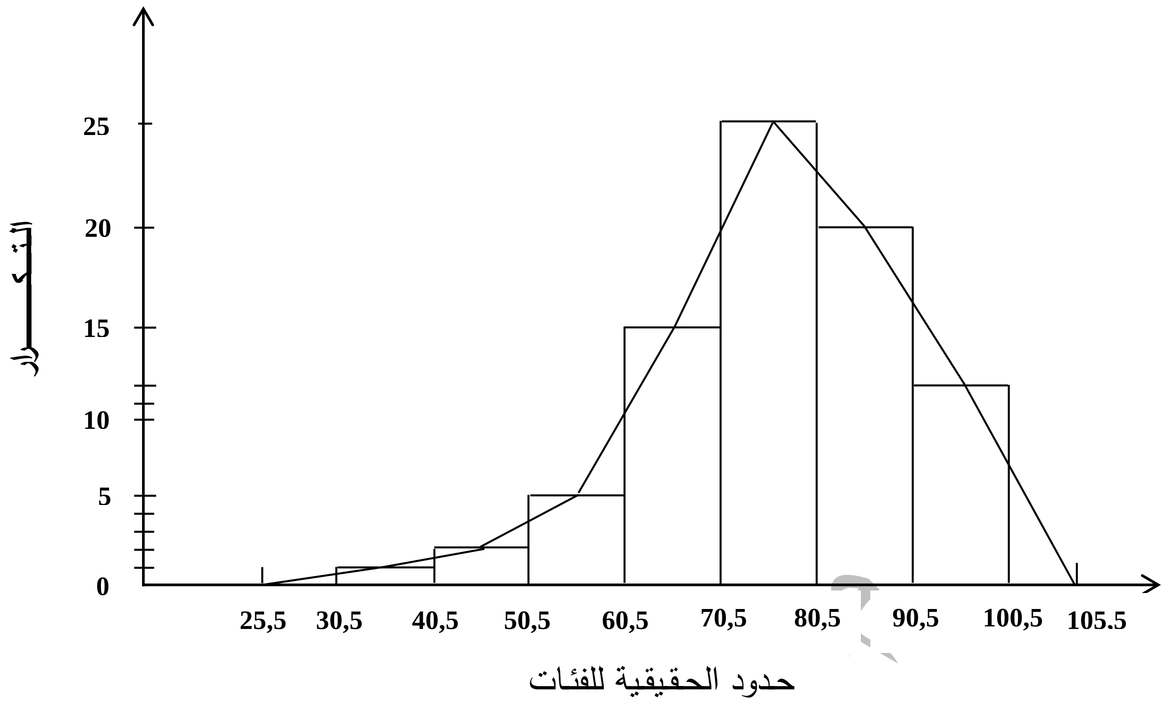
4) توصيل تلك النقاط بخطوط مستقيمة .

والشكل (2:3) يمثل المضلع التكراري لجدول (6:3) .



الشكل (2:3) المضلع التكراري لأطوال نباتات القطن .

هذا ويمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تصنيف القواعد العليا للمستطيلات (والتي تمثل مراكز الفئات) بنقاط ثم توصيل هذه النقاط بمستقيمات كما في الشكل (3:3) .

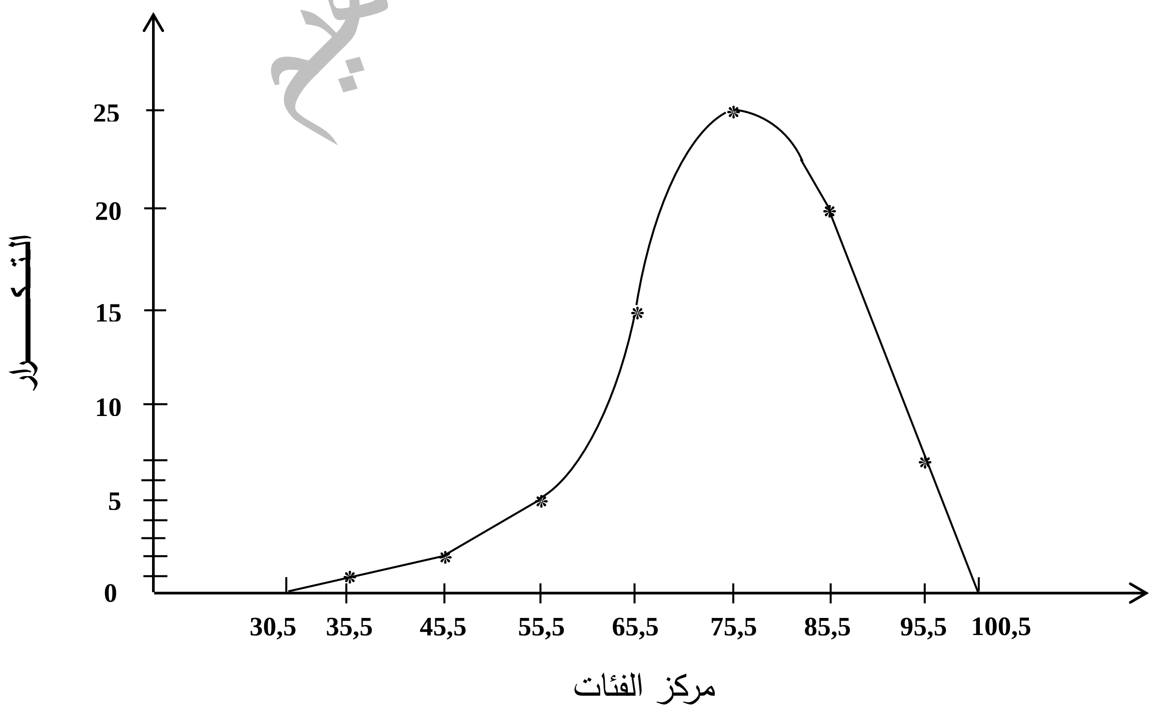


الشكل (3:3) المدرج التكراري والمضلع التكراري لأطوال نباتات القطن .

c) المنحنى التكراري Frequency Curve

وهو عبارة عن منحنى يمر بمعظم النقاط الواقعة على مركز الفئات والتي ارتفاعها يمثل تكرارات تلك الفئات.

وعادةً يقفل المنحنى التكراري بأن نصل بدايته بالحد الأدنى للفئة الأولى ونهايته بالحد الأعلى للفئة الأخيرة . وتكون مساحة المنحنى وكافئه (وليست مساوية) للمضلع التكراري . كما في الشكل (4:3) .



الشكل (4:3) المنحنى التكراري لأطوال نباتات القطن .

ملاحظة

عند مقارنة مجموعتين من نباتات غير متساويتين في عدد مفرداتها بإستخدام المضلع التكراري لهما فيجب إستخدام التكرار النسبي أو المئوي لهما بدلاً من التكرار العادي . والمضلع التكراري في هذه الحالة يسمى المضلع التكراري النسبي Relative frequency polygon أو المضلع التكراري المئوي Percentage polygon .

2) التمثيل البياني لجدول التوزيع التكراري التجميعي .

التمثيل التكراري التجميعي بيانياً تستخدم المضلع التكراري التجميعي . Cumulative frequency polygon or ogive وهو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط واقعة فوق الحدود الحقيقية للفئات وعلى إرتفاع تمثل التكرار التجميعي .

وهناك نوعان من المضلع التكراري التجميعي

1) المضلع التكراري التجميعي التصاعدي Or less Ogive .

ولرسم المضلع التكراري التجميعي التصاعدي نتبع الخطوات التالية .

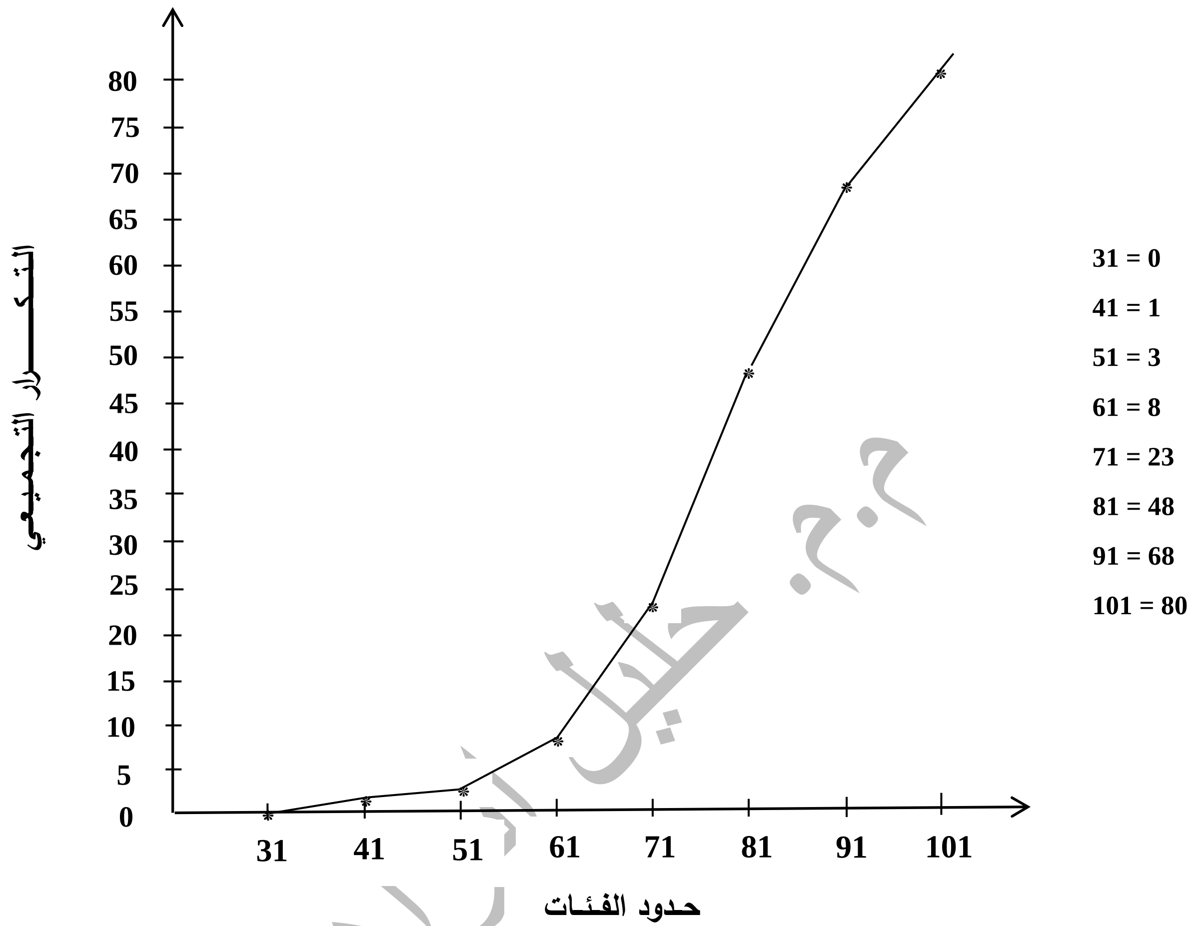
a) رسم المحور الافقي والمحور العمودي .

b) تدريج المحور الافقي إلى أقسام متساوية تشمل على جميع حدود الفئات . ويقسم المحور العمودي إلى أقسام متساوية بحيث تشمل على أكبر التكرارات التجميعية وهي المجموع الكلي للتكرارات .

c) وضع نقطة أمام كل حد فئة ارتفاعها يعادل التكرار التجميعي التصاعدي لذلك الحد .

d) توصيل تلك النقاط بخطوط مستقيمة .

والشكل (5:3) يمثل المضلع التكراري التجميعي التصاعدي لجدول (10:3) .

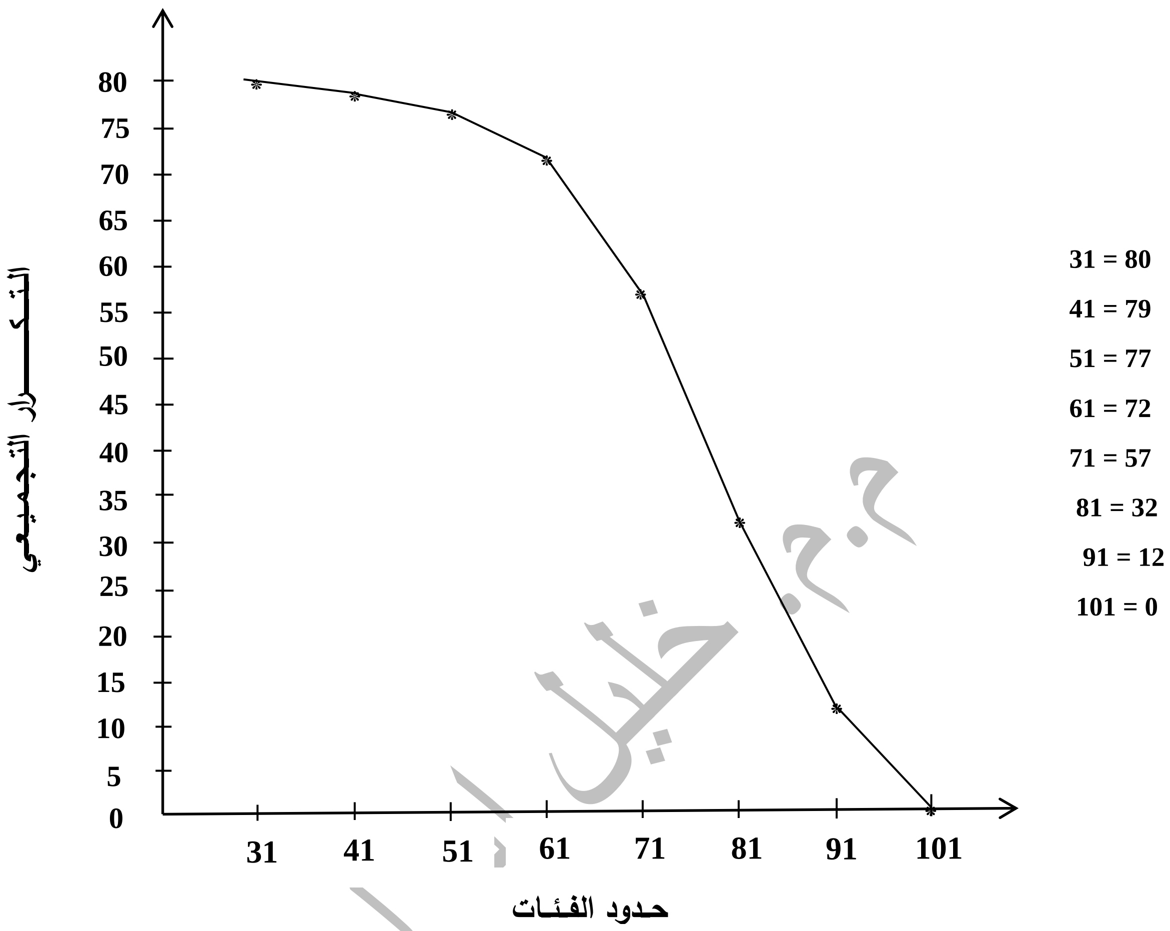


الشكل (5:3) المضلع التكراري التجميعي التصاعدي لأطوال نباتات القطن .

2) المضلع التكراري التجميعي التنازلي .

ويرسم بنفس الطريقة التي رسم فيها المضلع التكراري التجميعي التصاعدي ماعدا كون ارتفاع النقاط هنا هو التكرار التجميعي التنازلي ولذلك فيبدأ المضلع التكراري التجميعي التنازلي من أعلى نقطة (مجموع التكرارات الكلي) وينتهي بالصفر .بعكس المضلع التكراري التجميعي التصاعدي تماماً .

والشكل (6:3) يمثل المضلع التكراري التجميعي التنازلي لجدول (11:3) .



الشكل (6:3) المضلع التكراري التجميعي التنازلي لأطوال نباتات القطن .

ملاحظة

عند رسم التكرار التجميعي النسبي فالمضلع يسمى بالمضلع التجميعي النسبي Relative frequency ogive. وعند رسم التكرار التجميعي المئوي فالمضلع يسمى بالمضلع التجميعي المئوي Percentage frequency ogive . وذلك بإتباع نفس الأساليب السابقة .

هذا وفي كثير من الأحيان يرسم المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد .

كذلك يمكن رسم ما يسمى بالمنحى التكراري التجميعي وذلك برسم منحنِ يمر بمعظم النقاط (بدلاً من الخطوط المستقيمة المتكسرة) .

ما يمكن الاستفادة من منحني التكراري التجميعي التصاعدي أو التنازلي بإيجاد تقديرات معينة مذلك برسم أعمدة من المحور الأفقي لتقطع المنحنى في نقاط ثم نقرأ ما يقابل هذه النقاط على المحور العمودي بالإضافة إلى استخدامها في حساب بعض القيم الحسابية التي سيأتي ذكرها في الفصول القادمة .

مثال (7)

الجدول التالي يبين التوزيع التكراري للرواتب الشهرية لموظفي احدى الشركات .

الفئــــات	التكـــــرار
50 – 59	8
60 – 69	10
70 – 79	16
80 – 89	14
90 – 99	10
100 – 109	5
110 – 119	2
المجموع	65

أوجد كلاً مما يأتي :

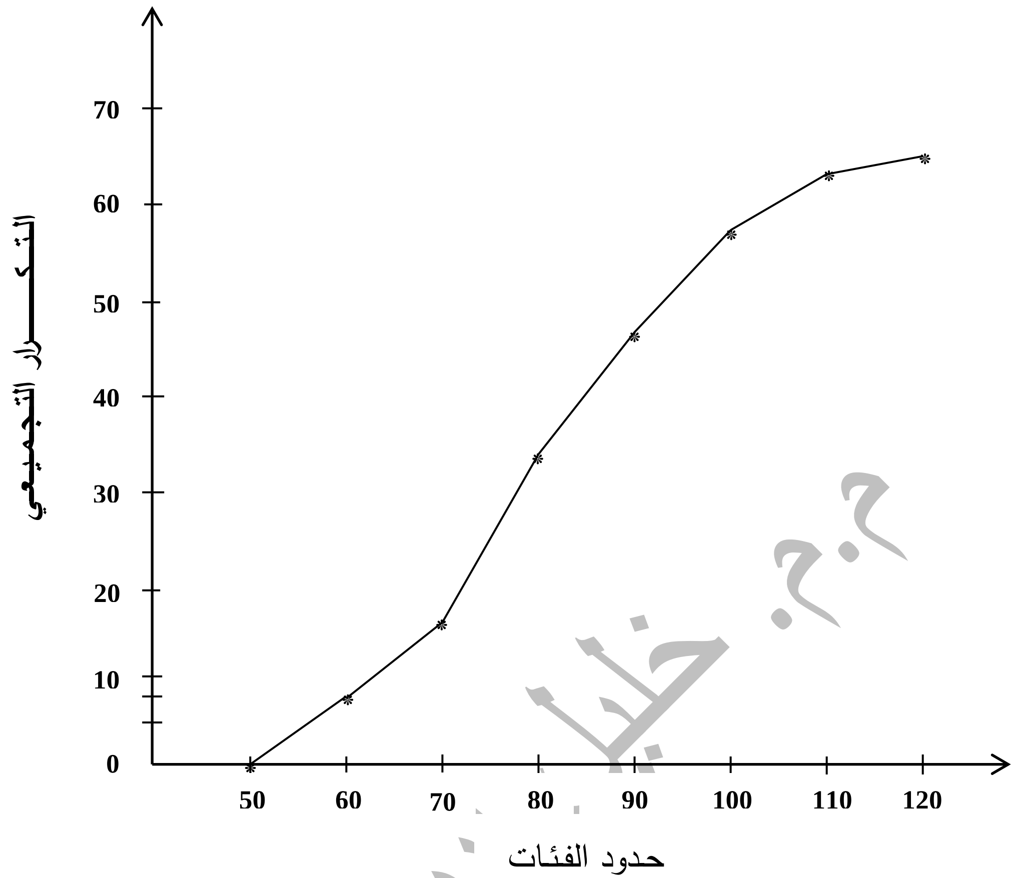
1) ارسم المضلع التكراري التجميعي التصاعدي .

الـــحـــــل :

نجد أولاً جدول التوزيع التكراري التجميعي التصاعدي كما يلي :

حدود فئات	التكرار التجميعي التصاعدي	**ملاحظة** (طريقة إيجاد التكرار التجميعي التصاعدي)
أقل من 50	0	$F_0 = 0$ تكرار ما قبل الفئة الاولى
أقل من 60	8	$F_1 = f_1 = 8$ تكرار الفئة الاولى
أقل من 70	18	$F_2 = f_2+f_1 = 18$ تكرار الفئة الثانية
أقل من 80	34	$F_3 = f_3+f_2+f_1 = 34$ تكرار الفئة الثالثة
أقل من 90	48	$F_4 = f_4+f_3+f_2+f_1 = 48$ تكرار الفئة الرابعة
أقل من 100	58	$F_5 = f_5+f_4+f_3+f_2+f_1 = 58$ و هكذا
أقل من 110	63	$F_6 = f_6+f_5+f_4+f_3+f_2+f_1 = 63$ و هكذا
أقل من 120	65	$F_7 = f_7+f_6+f_5+f_4+f_3+f_2+f_1 = 65$ و هكذا

ومن الجدول أعلاه نرسم المضلع التكراري التجميعي التصاعدي التالي :



الشكل (7:3) المضلع التكراري التجميعي التصاعدي للرواتب الشهرية لموظفي إحدى الشركات .

2) ماهو عدد المستخدمين الذين رواتبهم :

(a) أقل من 88 ديناراً :

<u>**الـحـــل**</u> :

من رسم المضلع التكراري التجميعي التصاعدي .

ارسم من نقطة 88 على المحور الأفقي عموداً ليقطع المضلع في نقطة يقابلها على المحور العمودي الرقم 45.

لذا فإن   = عدد المستخدمين الذين رواتبهم أقل من 88 ديناراً .

(b) 96 ديناراً فأكثر :

الحــــل :

يمكن إستعمال المضلع التكراري التجميعي التصاعدي التنازلي .

من المضلع التصاعدي مثلاً : نرسم من نقطة 96 على المحور الأفقي عموداً ليقطع المضلع في نقطة يقابلها على المحور العمودي الرقم 54 ، وهي تعني أن 54 مستخدماً رواتبهم أقل من 96 ديناراً . وبما أن المجموع الكلي للمستخدمين هو 65 ، لذا فإن 65 – 54 = 11 هو عدد المستخدمين الذين رواتبهم 96 ديناراً فأكثر .

(c) على الأقل 63 ديناراً ولكن أقل من 75 ديناراً .

الحــــل :

العدد المطلوب = عدد المستخدمين الذين رواتبهم أقل من 75 – عدد المستخدمين الذين رواتبهم أقل من 63 = 26 − 11 = 15

ملاحظة

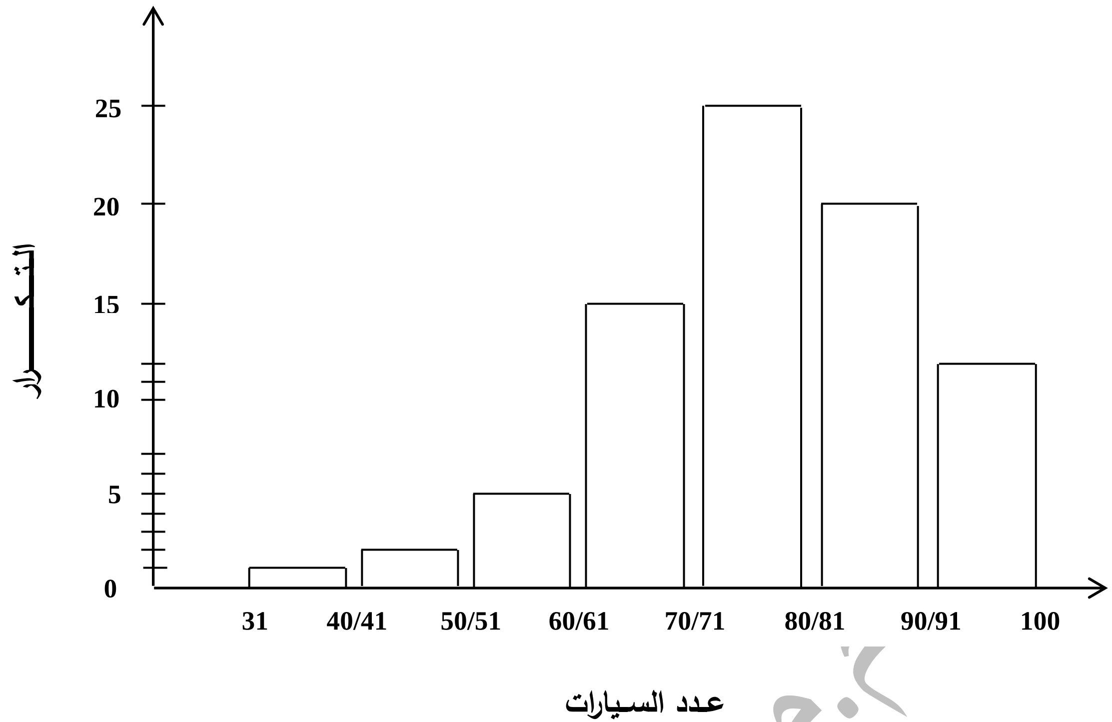
إن جميع الأمثلة التي ذكرت سابقاً هي لمتغيرات مستمرة .

(Continuous variables) في حالة استخدام قيم مقربة إلى أقرب عدد صحيح فقد تم معاملتها على أنها بيانات لمتغير مستمر باستعمال الحدود الحقيقية للفئات فكانت الرسوم البيانية كلها متصلة .

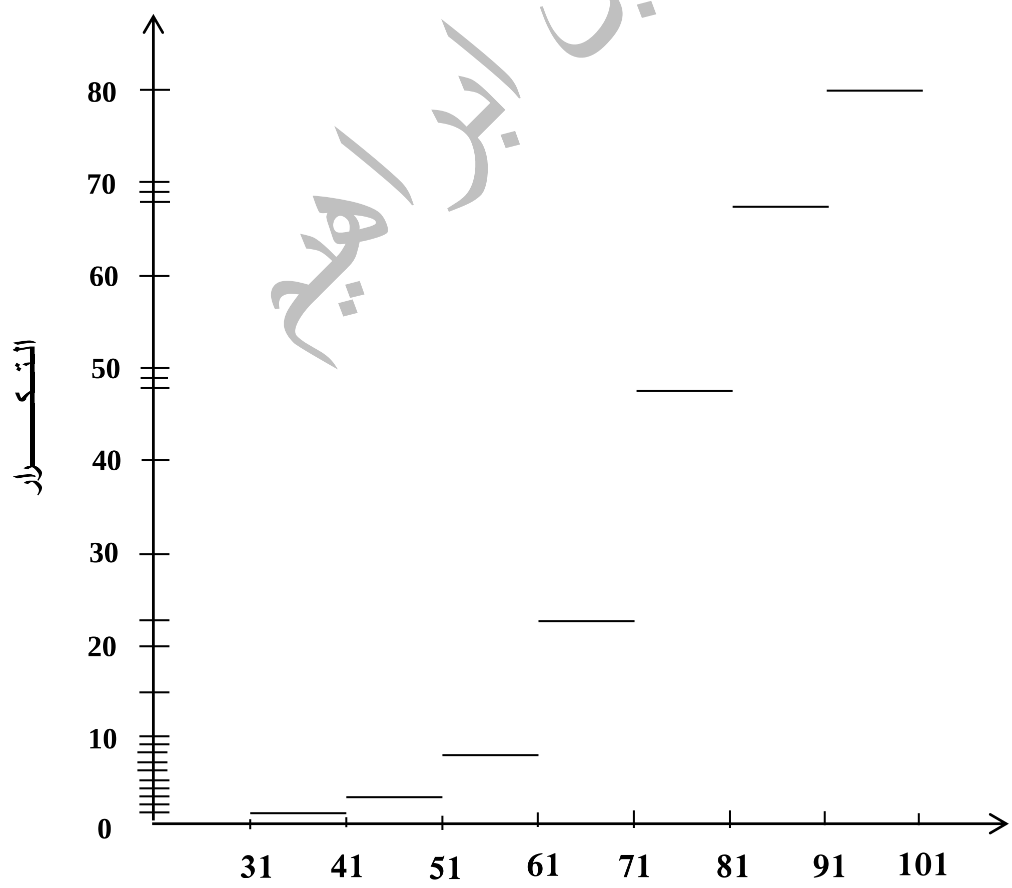
أما في حال إستخدام قيم لمتغير متقطع (Discrete variable) فإن الرسوم البيانية في هذه الحالة تكون متقطعة .

ففي المدرج التكراري مثلاً تكون قواعد المستطيلات منفصلة بعضها عن البعض الآخر .

أما في حال المضلع التكراري التجميعي التصاعدي (مثلاً) فإنه يكون على شكل مضلع متدرج متقطع . ولتوضيح ذلك نفرض بأن البيانات في الجدول (6:3) هي المضلع التكراري التجميعي التصاعدي سيكونان كالآتي .



شـكل (8:3) المدرج التكراري لعدد السيارات



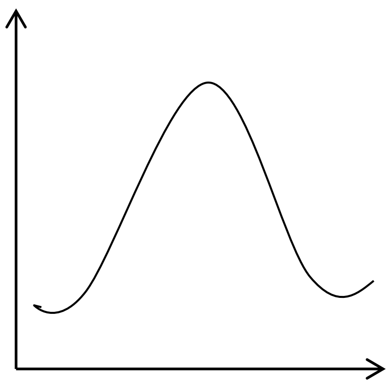
الشكل (9:3) المضلع التكراري التجميعي التصاعدي لعدد السيارات

(7:3) أنواع المنحنيات التكرارية Types of frequency curves

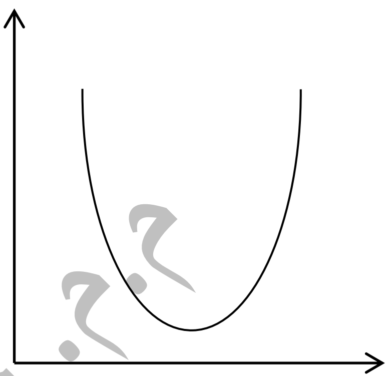
إن من أهم المنحنيات التكرارية التي قد نحصل عليها عملياً هي :

(1) المنحنيات المتماثلة Symmetrical frequency curves :

وهي المنحنيات التي تتصف بأن قيمتها تتوزع بشكل متماثل على خط المنتصف . ومن أشهر أمثلته : المنحني الطبيعي Normal curve (شكل 10:3) والمنحني ذو الشكل U أو المنحنى النوني (شكل 11:3) .



شكل (10:3) منحني الطبيعي



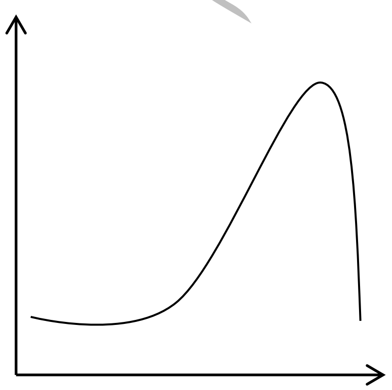
شكل (11:3) منحنى U
المنحني النوني

(2) المنحنيات غير المتماثلة : أو المنحنيات الملتوية (Asymmetrical frequency curves).

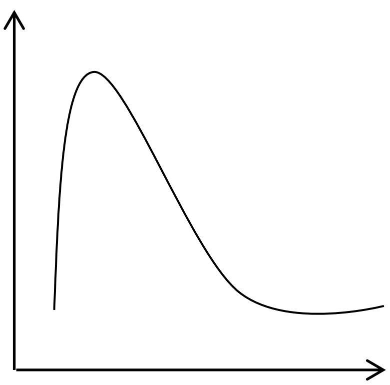
وهي المنحنيات التي تكون إحدى أطرافها أطول من الآخر وتقسم إلى :

(a) منحنيات ملتوية التواء معتدلاً: مثل منحنيات ملتوية التواء موجباً (Positive skewness or skewed to the right)

وهي منحنيات التي طرفها الطويل يقع في الجهة اليمنى (شكل 12:3) .

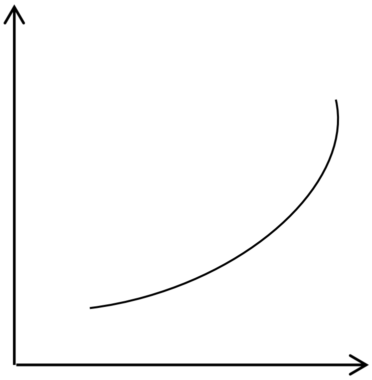


شكل (13:3) منحني ملتوي
إلتواءً سالباً

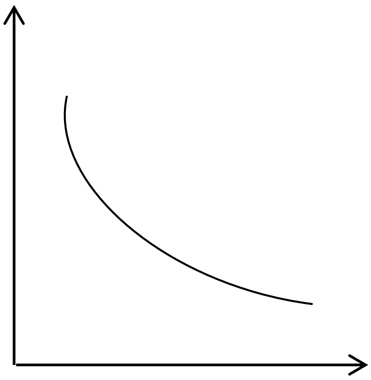


شكل (12:3) منحني ملتوي
إلتواءً موجباً

(ب) منحنيات ملتوية التواءً شديداً مثل المنحنيين التاليين (شكل 3:14 و3:15) .

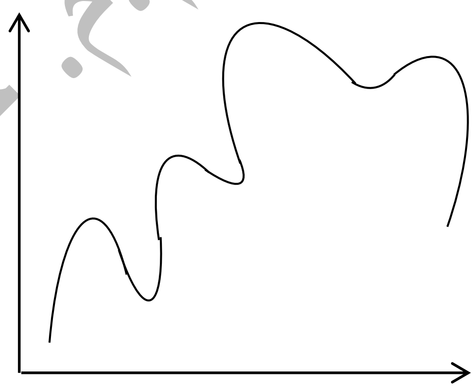


شكل (3:14) منحني على شكل
الحرف (ر) «المنحنى الرائي»

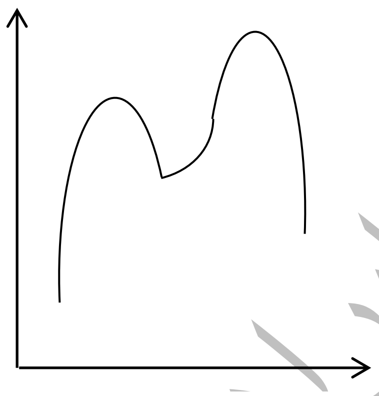


شكل (3:15) منحني على شكل
مقلوب (ر) «المنحنى الرائي المقلوب»

(3) منحنيات متعددة القمم مثل (شكل 3:16 و3:17) .

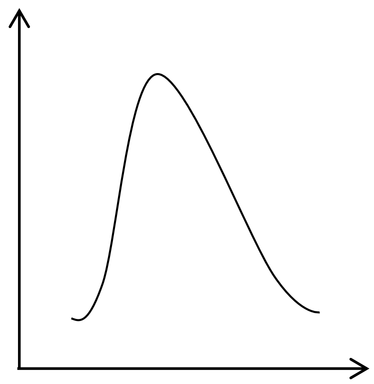


شكل (3:16) منحني متعدد القمم

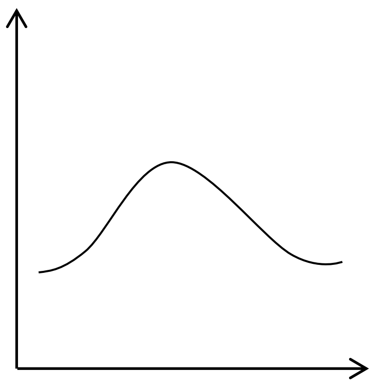


شكل (3:17) منحني ذو قمتين

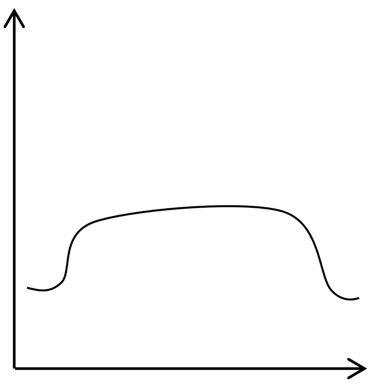
(4) منحنيات متفلطحة Kurtosis من أمثال : منحنيات مدببة القمة (شكل 3:18) ومعتدل القمة (3:19) ومفلطحة (3:20).



شكل (3:18) منحني مدبب



شكل (3:19) منحني معتدل القمة



شكل (3:20) منحني مفلطح

تمارين الفصل الثالث

1) أوجد الحدود الحقيقية ومركز الفئة وطول الفئة لكل من الفئات التالية :

a) 13 – 7

b) (-1) - (-5)

c) 18,7 – 10,4

d) 0,418 – 0,346

e) 1,35 – (-2,75)

f) 86,72 – 78,49 .

2) أوجد طول الفئة المناسبة لكل من التوزيعات التالية على فرض أن عدد الفئات في كل منها = 10

a) أقل قيمة = 7,5 وأكبر قيمة = 18,6 .

b) أقل قيمة = 53 وأكبر قيمة = 149 .

c) أقل قيمة = −15 وأكبر قيمة = zero .

3) إذا علمت بأن مراكز الفئات لأعمار عدد من الطلبة هي (30, 27, 24, 21,18) فما هي :

a) طول الفئة .

b) الحدود الحقيقية للفئات .

c) حدود الفئات لهذا التوزيع .

4) فيما يلي درجات 60 طالباً في الامتحان النهائي لدرس الإحصاء .

81	84	74	75	78	23
74	63	65	54	67	80
15	70	25	76	79	52
85	85	72	82	81	41
36	98	48	57	64	60
76	62	74	41	83	34
67	90	52	78	89	60
43	80	92	64	17	77
79	82	80	84	32	10
61	55	88	69	95	71

<u>المطلوب إيجاد ما يلي :</u>

a) انشاء جدول التوزيع التكراري باستعمال عشر فئات .

b) ارسم المدرج التكراري .

c) ارسم المضلع التكراري .

d) انشاء جدول التوزيع التكراري التجميعي التصاعدي والتنازلي .

e) ارسم المضلع التكراري التجميعي والتنازلي في رسم واحد .

حل تمارين الفصل الثالث

1) اوجد الحدود الحقيقية ومركز الفئة وطول الفئة لكل من الفئات التالية :

a) $13 - 7$ b) $(-1) - (-5)$ c) $18,7 - 10,4$ d) $0,418 - 0,346$

e) $1,35 - (-2,75)$ f) $86,72 - 78,49$.

الحل

a) $13 - 7$

طول الفئة = الحد الاعلى – الحد الادنى + 1 ⟵ $13 - 7 + 1 = 7$

مركز الفئة = $\frac{\text{الحد الادنى + الحد الاعلى}}{2} = \frac{7 + 13}{2} = \frac{20}{2} = 10$

الحد الادنى الحقيقي = مركز الفئة − $\frac{1}{2}$ (طول الفئة) = $10 - \frac{1}{2}(7) = 6,5$

الحد الاعلى الحقيقي = مركز الفئة + $\frac{1}{2}$ (طول الفئة) = $10 + \frac{1}{2}(7) = 13,5$

b) $(-1) - (-5)$

طول الفئة = الحد الاعلى – الحد الادنى + 1 ⟵ $(-1) - (-5) + 1 = 5$

مركز الفئة = $\frac{\text{الحد الادنى + الحد الاعلى}}{2} = \frac{(-5) + (-1)}{2} = \frac{-6}{2} = -3$

الحد الادنى الحقيقي = مركز الفئة − $\frac{1}{2}$ (طول الفئة) = $-3 - \frac{1}{2}(5) = -5,5$

الحد الاعلى الحقيقي = مركز الفئة + $\frac{1}{2}$ (طول الفئة) = $-3 + \frac{1}{2}(5) = -0,5$

c) $18,7 - 10,4$

طول الفئة = الحد الاعلى – الحد الادنى + 1 ⟵ $18,7 - 10,4 + 1 = 8,4$

$$\text{مركز الفئة} = \frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2} = \frac{(18,7)+(10,4)}{2} = \frac{29,1}{2} = 14,55$$

$$\text{الحد الادنى الحقيقي} = \text{مركز الفئة} - \frac{1}{2}\,(\text{طول الفئة}) = 14,55 - \frac{1}{2}\,(8,4) = 10,35$$

$$\text{الحد الاعلى الحقيقي} = \text{مركز الفئة} + \frac{1}{2}\,(\text{طول الفئة}) = 14,55 + \frac{1}{2}\,(8,4) = 18,75$$

d) 0,418 − 0,346

$$\text{طول الفئة} = \text{الحد الاعلى} - \text{الحد الادنى} + 1 \longrightarrow 0,418 - 0,346 + 1 = 0,073$$

$$\text{مركز الفئة} = \frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2} = \frac{(0,418)+(0,346)}{2} = \frac{0,764}{2} = 0,382$$

$$\text{الحد الادنى الحقيقي} = \text{مركز الفئة} - \frac{1}{2}\,(\text{طول الفئة}) = 0,382 - \frac{1}{2}\,(0,073) = 0,3455$$

$$\text{الحد الاعلى الحقيقي} = \text{مركز الفئة} + \frac{1}{2}\,(\text{طول الفئة}) = 0,382 + \frac{1}{2}\,(0,073) = 0,4185$$

e) 1,35 − (−2,75)

$$\text{طول الفئة} = \text{الحد الاعلى} - \text{الحد الادنى} + 1 \longrightarrow 1,35 - (-2,75) + 1 = 4,11$$

$$\text{مركز الفئة} = \frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2} = \frac{1,35 + (-2,75)}{2} = \frac{0,764}{2} = -0,7$$

$$\text{الحد الادنى الحقيقي} = \text{مركز الفئة} - \frac{1}{2}\,(\text{طول الفئة}) = -0,7 - \frac{1}{2}\,(4,11) = 2,755$$

$$\text{الحد الاعلى الحقيقي} = \text{مركز الفئة} + \frac{1}{2}\,(\text{طول الفئة}) = 0,382 + \frac{1}{2}\,(4,11) = 1,355$$

f) 86,72 − 78,49

$$\text{طول الفئة} = \text{الحد الاعلى} - \text{الحد الادنى} + 1 \longrightarrow 86,72 - 78,49 + 1 = 8,24$$

$\text{مركز الفئة} = \frac{\text{الحد الاعلى} + \text{الحد الادنى}}{2} = \frac{86,72 + 78,49}{2} = \frac{165,21}{2} = 82,605$

$\text{الحد الادنى الحقيقي} = \text{مركز الفئة} - \frac{1}{2}\,(\text{طول الفئة}) = 82,605 - \frac{1}{2}\,(8,24) = 78,485$

$\text{الحد الاعلى الحقيقي} = \text{مركز الفئة} + \frac{1}{2}\,(\text{طول الفئة}) = 82,605 + \frac{1}{2}\,(8,24) = 86,725$

2) أوجد طول الفئة المناسبة لكل من التوزيعات التالية على فرض أن عدد الفئات في كل منها = 10

a) أقل قيمة = 7,5 وأكبر قيمة = 18,6

b) أقل قيمة = 53 وأكبر قيمة = 149

c) أقل قيمة = −15 وأكبر قيمة = zero

الحل

a) أقل قيمة = 7,5 وأكبر قيمة = 18,6

$\text{طول الفئة} = \frac{\text{اعلى قيمة} - \text{اقل قيمة}}{\text{عدد الفئات}} = \frac{18,6 - 7,5}{10} = \frac{11,1}{10} = 1,11$

b) أقل قيمة = 53 وأكبر قيمة = 149

$\text{طول الفئة} = \frac{\text{اعلى قيمة} - \text{اقل قيمة}}{\text{عدد الفئات}} = \frac{149 - 53}{10} = \frac{96}{10} = 9,6 \simeq 10$

c) أقل قيمة = −15 وأكبر قيمة = zero

$\text{طول الفئة} = \frac{\text{اعلى قيمة} - \text{اقل قيمة}}{\text{عدد الفئات}} = \frac{zero - (-15)}{10} = \frac{15}{10} = 1,5 \simeq 2$

3) إذا علمت بأن مراكز الفئات لأعمار عدد من الطلبة هي (30, 27, 24, 21,18) فما هي :

a) طول الفئة .   b) الحدود الحقيقية للفئات .   c) حدود الفئات لهذا التوزيع .

الحل

a) طول الفئة

طول الفئة = الفرق بين مركزي فئتين متتاليتين ⟵ 3 = 18 − 21   طول الفئة = ٣

b) الحدود الحقيقية للفئات

الحد الادنى الحقيقي = $18 - \frac{1}{2}(3) = 16,5$

الحد الاعلى الحقيقي = $18 + \frac{1}{2}(3) = 19,5$

c) حدود الفئات لهذا التوزيع

الحد الادنى الحقيقي = حد الفئة − 0,5   16,5 = حد الفئة − 0,5   حد الفئة = −16,5 − 0,5 = 17

الحد الاعلى الحقيقي = حد الفئة + 0,5   19,5 = حد الفئة + 0,5   حد الفئة = −19,5 + 0,5 = 19

* إذاً يصبح جدول التوزيع التكراري بالشكل التالي :

ت	الفئات	الحدود الحقيقية للفئات	مركز الفئة	طول الفئة
1	19 - 17	19,5 - 16,5	18	3
2	22 - 20	22,5 - 19,5	21	3
3	25 - 23	25,5 - 22,5	24	3
4	28 - 26	28,5 - 25,5	27	3
5	31 - 29	31,5 - 28,5	30	3

4) فيما يلي درجات 60 طالباً في الامتحان النهائي لدرس الإحصاء .

81	84	74	75	78	23
74	63	65	54	67	80
15	70	25	76	79	52
85	85	72	82	81	41
36	98	48	57	64	60
76	62	74	41	83	34
67	90	52	78	89	60
43	80	92	64	17	77
79	82	80	84	32	10
61	55	88	69	95	71

المطلوب إيجاد ما يلي :

a) انشاء جدول التوزيع التكراري باستعمال عشر فئات .

b) ارسم المدرج التكراري .

c) ارسم المضلع التكراري .

d) انشاء جدول التوزيع التكراري التجميعي التصاعدي والتنازلي .

e) ارسم المضلع التكراري التجميعي والتنازلي في رسم واحد .

الحل

a) انشاء جدول التوزيع التكراري باستعمال عشر فئات .

1) استخراج المدى $\longleftarrow$ المدى = اعلى قيمة – اقل قيمة $\longleftarrow$ المدى = $98 - 10 = 88$

2) طول الفئة $\longleftarrow$ طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ $\longleftarrow$ طول الفئة = $\frac{88}{10} = 8,8 \simeq 9$

3) بما ان الحد الادنى للفئة موجود في السؤال هو اقل حد = 10 إذاً علينا استخراج الحد الاعلى من القانون التالي: طول الفئة = الحد الاعلى – الحد الادنى +١ = $\longleftarrow$ 9 = الحد الاعلى – 10 + 1 =

الحد الاعلى= $9 + 9 = 18$

إذاً الحد الادنى للفئة هو 10 والحد الاعلى للفئة هو 18 وبهذه الحالة نضيف لكل حد منها العدد 9 .

٤) نوجد الحدود الحقيقية من خلال القانون التالي :

الحد الادنى الحقيقي لأي فئة = الحد الادنى لتكك الفئة − 0,5

الحد الادنى الحقيقي للفئة الاولى = 10− 0,5 = 9,5

الحد الاعلى الحقيقي لأي فئة = الحد الاعلى لتكك الفئة + 0,5

الحد الاعلى الحقيقي للفئة الاولى = 18 + 0,5 = 18,5

* وهكذا لإيجاد بقية الحدود الحقيقية للفئات للحد الحقيقي الادنى والحد الحقيقي الاعلى لكل الفئات .

٥) مركز الفئة = $\frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2} = \frac{10 + 18}{2} = \frac{28}{2} = 14$

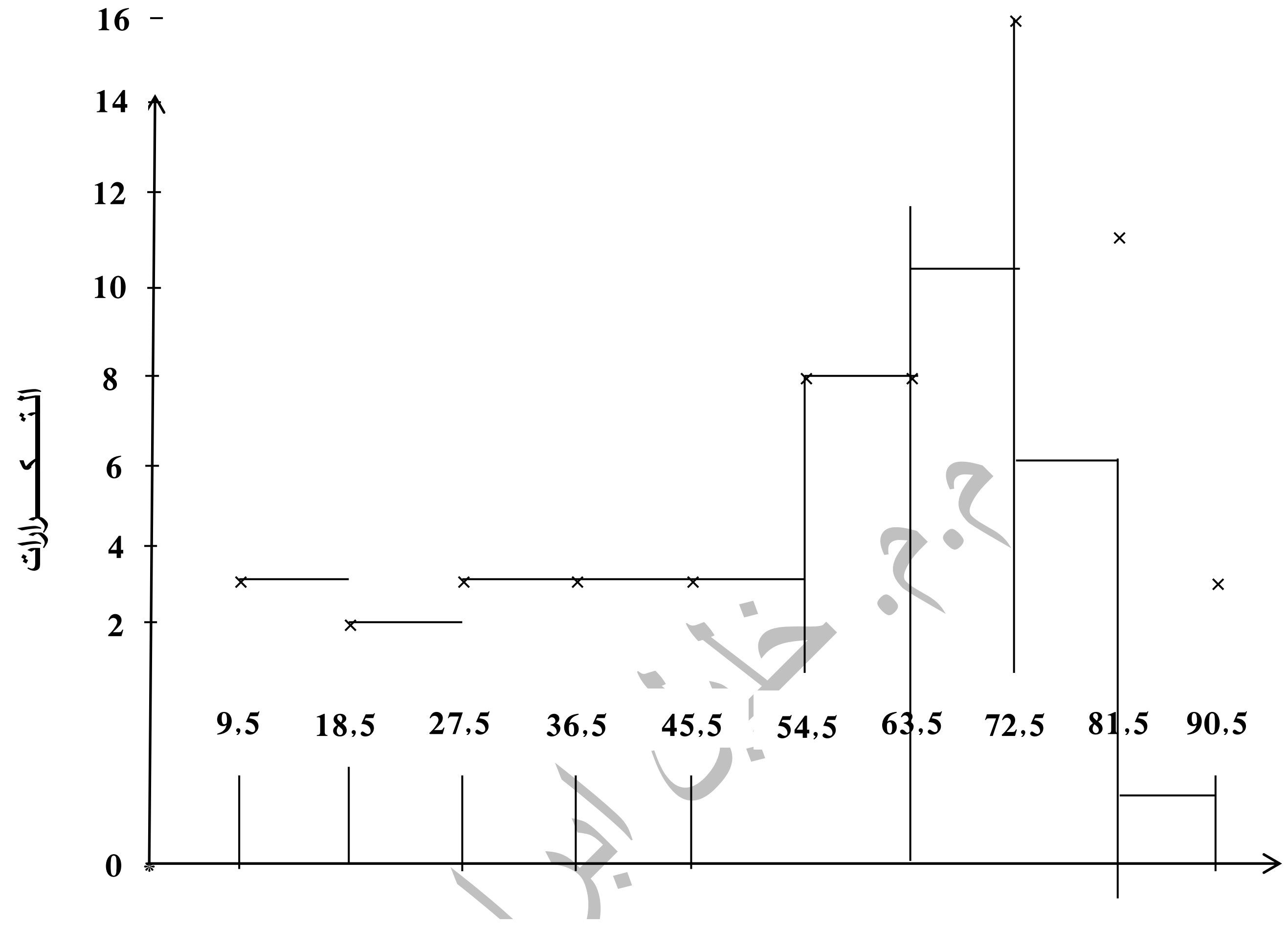
* وهكذا لإيجاد بقية مراكز للفئات من للفئات الاخرى .

٦) يستخرج التكرار من جدول الارقام من خلال الارقام المحصورة بين ارقام كل فئة في جدول التوزيع التكراري

3) جدول التوزيع التكراري

حدود الفئات	الحدود الحقيقية	مركز الفئة	طول الفئة	التكرار
10 − 18	9,5 − 18,5	14	9	3
19 − 27	18,5 − 27,5	23	9	2
28 − 36	27,5 − 36,5	32	9	3
37 − 45	36,5 − 45,5	41	9	3
46 − 54	45,5 − 54,5	50	9	3
55 − 63	54,5 − 63,5	59	9	8
64 − 72	63,5 − 72,5	68	9	8
73 − 81	72,5 − 81,5	77	9	16
82 − 90	81,5 − 90,5	86	9	11
91 − 99	90,5 − 99,5	95	9	3

b) ارسم المدرج التكراري



جدول Table

فئات الوزن (كغم) Weight classes (kg)

عدد الطلبة Number of students

موضوع البعثة ....

عدد الطلبة number of students

الوزن (كغم) weight(kg)

الطول (سم) length(cm)

الفئات Classes

الحدود الحقيقية للفئات Class boundaries or True class limits

مركز الفئة (yi) Cass mark (yi)

التكرار (fi) Frequency (fi)

القوانين المهمة

طول الفئة : هناك عدة طرق لإيجاد طول الفئة

1) طريقة الاولى لإيجاد طول الفئة (عندما تكون حدود الفئات اعداد صحيحة فقط).

طول الفئة = الحد الاعلى – الحد الادنى + 1 .

Class length = Upper limit – Lower limit + 1 .

2) الطريقة الثانية / طول الفئة = الحد الحقيقي الاعلى – الحد الحقيقي الادنى لتلك الفئة .

Class length = Upper class boundary – Lower class boundary of the class .

3) الطريقة الثالثة / طول الفئة = الفرق بين الحدين الادنى أو الحدين الاعلى لفئتين متتاليتين .

4) الطريقة الرابعة / طول الفئة = الفرق بين الحدين الحقيقين الادنى أو الاعلى لفئتين متتاليتين .

5) الطريقة الخامسة) / طول الفئة = الفرق بين مركزي فئتين متتاليتين .

الحدود الحقيقية : يمكن إيجاد الحدود الحقيقية لأي فئة بإحدى الطرق التالية .

1) الطريقة الاولى :

الحد الادنى الحقيقي لأي فئة = مركز تلك الفئة − $\frac{1}{2}$ (طول تلك الفئة) .

الحد الاعلى الحقيقي لأي فئة = مركز تلك الفئة + $\frac{1}{2}$ (طول تلك الفئة) .

2) الطريقة الثانية :

الحد الادنى الحقيقي لأي فئة = $\frac{\text{الحد الادنى لتلك الفئة} + \text{الحد الاعلى للفئة السابقة}}{2}$.

الحد الاعلى الحقيقي لأي فئة = $\frac{\text{الحد الاعلى لتلك الفئة} + \text{الحد الادنى للفئة التي تليها}}{2}$.

مركز الفئة : وتحسب بإحدى الطريقتين التاليتين .

$$\text{مركز الفئة} = \frac{\text{الحد الادنى + الحد الاعلى}}{2} .$$

$$\text{مركز الفئة} = \frac{\text{الحد الادنى الحقيقي + الحد الاعلى الحقيقي}}{2} .$$

ملاحظة :اذا كانت حدود الفئات اعداد صحيحة فإن :

الحد الادنى الحقيقي لأي فئة = الحد الادنى لتلك الفئة – 0,5 .

الحد الاعلى الحقيقي لأي فئة = الحد الاعلى لتلك الفئة + 0,5 .

المدى : لإيجاد المدى من المتغيرات التالية :

المدى = أعلى قيمة – أقل قيمة

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$. أو **طول الفئة** = $\frac{\text{المدى}}{\text{المسافة بين عدد الفئات}}$. أو **طول الفئة** = $\frac{\text{الحد الاعلى – الحد الادنى}}{\text{المسافة بين عدد الفئات}}$.

ملاحظة : أقل قيمة يعني الحد الادنى //// أعلى قيمة يعني الحد الاعلى .

مثال / في جدول التوزيع التكراري لأطوال نباتات القطن مبيناً فيه الحدود الحقيقية ومراكز الفئات .

تسلسل	الفئات	الحدود الحقيقية للفئات	مركز الفئة yi	التكرارات
1	40 – 31	40,5 – 30,5	35,5	1
2	50 – 41	50,5 – 40,5	45,5	2
3	60 – 51	60,5 – 50,5	55,5	5
4	70 – 61	70,5 – 60,5	65,5	15
5	80 – 71	80,5 – 70,5	75,5	25
6	90 – 81	90,5 – 80,5	85,5	20
7	100 – 91	100,5 – 90,5	95,5	12
	المجموع			**80**

الحل / إيجاد طول الفئة

1) طريقة الاولى لإيجاد طول الفئة (عندما تكون حدود الفئات اعداد صحيحة فقط).

طول الفئة = الحد الاعلى – الحد الادنى + 1 .

طول الفئة = 70 − 61 + 1 = 10

2) الطريقة الثانية / طول الفئة = الحد الحقيقي الاعلى – الحد الحقيقي الادنى لتلك الفئة .

طول الفئة = 70,5 − 60,5 = 10

3) الطريقة الثالثة / طول الفئة = الفرق بين الحدين الادنى أو الحدين الاعلى لفئتين متتاليتين .

طول الفئة = 71 − 61 = 10   طول الفئة = 80 − 70 = 10

4) الطريقة الرابعة / طول الفئة = الفرق بين الحدين الحقيقين الادنى أو الاعلى لفئتين متتاليتين .

طول الفئة = 70.5 − 60,5 = 10

طول الفئة = 80.5 − 70,5 = 10

5) الطريقة الخامسة) / طول الفئة = الفرق بين مركزي فئتين متتاليتين .

طول الفئة = 75.5 − 65,5 = 10

الحدود الحقيقية : يمكن إيجاد الحدود الحقيقية لأي فئة بإحدى الطرق التالية .

الحد الادنى الحقيقي لأي فئة = مركز تلك الفئة − $\frac{1}{2}$ (طول تلك الفئة) .

الحد الادنى الحقيقي للفئة الرابعة = 65,5 − $\frac{1}{\cancel{2}}$ ($\overset{5}{\cancel{10}}$) = 60,5

الحد الاعلى الحقيقي لأي فئة = مركز تلك الفئة + $\frac{1}{2}$ (طول تلك الفئة) .

الحد الاعلى الحقيقي للفئة الرابعة = 65,5 + $\frac{1}{\cancel{2}}$ ($\overset{5}{\cancel{10}}$) = 70.5

الحد الادنى الحقيقي لأي فئة = $\frac{\text{الحد الادنى لتلك الفئة + الحد الاعلى للفئة السابقة}}{2}$.

الحد الادنى الحقيقي للفئة الرابعة = $\frac{60 + 61}{2} = 60,5$

الحد الاعلى الحقيقي لأي فئة = $\frac{\text{الحد الاعلى لتلك الفئة} + \text{الحد الادنى للفئة التي تليها}}{2}$.

الحد الاعلى الحقيقي للفئة الرابعة = $\frac{71 + 70}{2} = 70,5$

مركز الفئة : وتحسب بإحدى الطريقتين التاليتين .

مركز الفئة = $\frac{\text{الحد الادنى} + \text{الحد الاعلى}}{2}$.

مركز الفئة الرابعة = $\frac{70 + 61}{2} = 65,5$

مركز الفئة = $\frac{\text{الحد الادنى الحقيقي} + \text{الحد الاعلى الحقيقي}}{2}$.

مركز الفئة الرابعة = $\frac{70,5 + 60,5}{2} = 65,5$

تكرار الفئة الرابعة = ١٥ أي أن هناك 15 قيمة من قيم المتغير واقعة في المدى (61 – 70) .

الخطوات العامة في إنشاء جدول التوزيع التكراري .

1) استخراج مدى المتغير .

2) اختيار وتحديد عدد الفئات .

3) إيجاد طول مدى الفئة .

4) كتابة حدود الفئة .

5) استخراج عدد التكرارات لكل فئة .

مثال / اطوال 80 نباتاً من نباتات القطن مقدرة بالسنتمترات

80	87	98	81	74	48	79	80
78	82	93	91	70	90	80	84
73	74	81	56	65	92	70	71

86	83	93	65	51	85	68	72
68	86	43	74	73	83	90	35
75	67	72	90	71	76	92	93
81	88	91	97	72	61	80	91
77	71	59	80	95	99	70	74
63	89	67	60	82	83	63	60
75	79	88	66	70	88	76	63

الحل / نتبع الخطوات التالية

1) استخراج المدى أو مدى المتغير .

فأطول النبات = 99 سم بينما اقصر نبات = 35 سم .

المدى = أعلى قيمة – أقل قيمة ⟵ لذى فالمدى = 99 – 35 = 64 سم

2) اختيار وتحديد عدد الفئات . هناك عدة طرق حسابية تقريبية لإيجاد عدد الفئات أهمها .

a) طريقة سترجس Sturges

عدد الفئات = 1 + 3,3 × (لوغارتم عدد الفئات) .

b) طريقة يول Yule

عدد الفئات = 2,5 × $\sqrt[4]{\text{عدد المفردات}}$

وكل من الطريقتين مميزات وعيوب ولن نستعمل أياً منها هنا بل إننا سنختار عدد الفئات اختياراً على أن لا تقل عن خمسة ولا تزيد عن خمسة عشر فئة وذلك تبعاً لطبيعة البيانات وعدد مفرداتها ومدى التغير فيها .

3) إيجاد طول الفئة .

يجب أن لا تقل طول الفئة عن : طول الفئة = $\frac{\text{االمدى المتغير}}{\text{عدد الفئات}}$ مقربة إلى أقرب عدد صحيح أكبر .

طول الفئة = $\frac{\text{االمدى المتغير}}{\text{عدد الفئات}} = \frac{64}{7} = 9{,}142 \simeq 10$ لذا يستحسن أن يكون طول الفئة = 10 .

4) كتابة حدود الفئة .

يجب كتابة حدود الفئات بحيث أن جميع قيم المتغير تقع بين الحد الادنى للفئة الاولى والحد الاعلى للفئة الاخيرة .

ويستحسن أن نبدأ بكتابة الحد الادنى للفئة الاولى بقيمة أصغر مفردة أو أقل من ذلك بقليل وتنتهي بالحد للفئة الاخيرة بقيمة أكبر مفردة أو أكثر من ذلك بقليل .

فمثلا اصغر قيمة من قيم أطوال نباتات القطن هي 35 لذا فمن الممكن أن يكون الرقم 31 يمثل الحد الادنى للفئة الاولى . وبما أن طول الفئة هو 10 لذا فإن حدي الفئة الاولى هما 31 – 40 والفئة الثانية تبدأ من 41 – 50 بينما الفئة السابعة (الاخيرة) هي 91 – 100 لأحظ بأن الحد الادنى للفئة الاولى 31 والحد الاعلى للفئة الاخيرة 100 تحتوي على كافة قيم المتغير .

5) استخراج عدد التكرارات لكل فئة .

ويتم ذلك بتسجيل القيم الاصلية واحدة بعد الاخرى في الفئة الخاصة به على شكل اشارات أو علامات أولاً ثم ترجمتها الى أرقام كما مبين في الجدول .

الفئات	**التكرارات رقماً**	**التكرار (بالعلامات)**
31 - 40	1	|
41 - 50	2	||
51 - 60	5	𝍸
61 - 70	15	𝍸 𝍸 𝍸
71 - 80	25	
81 - 90	20	
91 - 100	12	
المجموع	80	

المدخل إلى الإحصاء

Introduction to statistics

الطبعة الثانية/ ٢٠٠٠

الفصل الرابع

المدرس المساعد

خليل ابراهيم خليل

Ministry of Higher Education and Scientific Research .
University of Al Mosul .
College of Agriculture and Forestry .
Field crops department .

وزارة التعليم العالي والبحث العلمي .
جامعة الموصل .
كلية الزراعة والغابات .
قسم المحاصيل الحقلية .

المدخل إلى الإحصاء

Introduction to statistics

الدكتور خاشع محمود الراوي

الطبعة الثانية/ ٢٠٠٠

الفصل الرابع

مقاييس التمركز أو التوسط

Measures of Central Tendency

(1:4) المقدمة

إن معظم القيم لمختلف الظاهر الطبيعية تتمركز عادةً في الوسط أو قريبة منه. ومقاييس التمركز أو التوسط لأي مجموعة من البيانات التابعة لظاهرة ما. هي تلك المقاييس التي تبحث في تقدير قيمة تتمركز حولها أغلبية هذه البيانات وإن هذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر عن أو يمثل جميع بيانات تلك القيمة .

وأهم مقاييس التوسط هي .

أولاً- الوسط الحسابي أو المتوسط	**The Arithmetic Mean**
ثانياً- الوسط الهندسي	**The Geometric Mean**
ثالثاً - الوسط التوافقي	**The Harmonic Mean**
رابعاً- الوسط التربيعي	**The Quadratic Mean**
خامساً- الوسيط	**The Median**
سادساً- المنوال	**The Mode**

هذا وسنشرح كيفية حساب كل مقياس من المقاييس أعلاه في الحالتين :

1- حالة البيانات غير المبوبة .

2- حالة البيانات المبوبة .

أولاً– الوسط الحسابي أو المتوسط The Arithmetic Mean

الوسط الحسابي أو المتوسط لقيم متغير ما هو الناتجة من قسمة مجموع تلك القيم على عددها ويرمز له بالرمز $\bar{y}$.

(a) من البيانات غير المبوبة .

(4:1) إذا كان لدينا n من القيم أو المشاهدات : $y_1 \,,\, y_2 \;\ldots\ldots\ldots\ldots\; y_n$

فإن الوسط الحسابي لها هو :

$$\bar{y} = \frac{\sum y_i}{n}$$

مثال (1)

البيانات التالية تمثل كمية المطر الساقطة سنوياً (بالمليمترات) على مدينة الموصل خلال فترة خمس سنوات 520 ، 350 ، 450 ، 380 ، 400 فما هو متوسط سقوط المطر خلال هذه الفترة .

الحل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{520 + 350 + 450 + 380 + 400}{5} = \frac{2100}{5} = 420 \text{ mm.}$$

أي إن معدل الأمطار خلال تلك الفترة هو 450 ملم .

مثال (2)

أحسب الوسط الحسابي لأطوال نباتات القطن في جدول(5:3) قبل تبويبها

الحل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{80 + 84 + \ldots\ldots\ldots\ldots + 75}{80} = \frac{6126}{80} = 76{,}575 \sim 76{,}58 \text{ cm} .$$

أي إن معدا طول النبات هو 76,58 cm .

(*b*) من البيانات المبوبة .

إذا كانت y_k .... y_1 , y_2 تمثل مركز الفئات في جدول التوزيع التكراري مع تكراراتها f_k .... f_1 , f_2 على التوالي فالوسط الحسابي هو .

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} \qquad \text{أو} \qquad \bar{y} = \frac{\sum_{i=1}^{k} f_i y_i}{\sum_{i=1}^{k} f_i}$$

خطوات إيجاد الوسط الحسابي في بيانات مبوبة هي كالآتي .

1) تعين مراكز الفئات ($\bar{y}$) .

2) ضرب مراكز كل فئة بمقدار تكرار ($f_i\, y_i$) .

3) قيمة مجموع (حاصل ضرب مركز كل فئة × تكرا رها) على مجموع التكرارات .

مثال (3)

استخرج الوسط الحسابي لأطوال النباتات من جدول التوزيع التكراري (6:3) .

الحـــــل : عين مركز الفئات ثم أضرب مركز كل فئة × تكرارها كما في الجدول التالي .

جدول (1:4) .

الفئات	التكرار (f_i)	مركز الفئة (y_i)	التكرار × مركز الفئات
40 – 31	1	35,5	35,5
50 – 41	2	45,5	91,0
60 – 51	5	55,5	277,5
70 – 61	15	65,5	982,5
80 – 71	25	75,5	1887,5
90 – 81	20	85,5	1710,0
100 – 91	12	95,5	1146,0
	$\sum f_i = 80$		$\sum f_i x_i = 6130{,}0$

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6130}{80} = 76{,}625 \text{ cm} \longrightarrow 76{,}62 \text{ cm}$$ نأخذ رقمين بعد الفارز

أي إن معدل طول النبات هو 76,62 سم .

لاحظ أن هذا الرقم يختلف قليلاً عن الوسط الحسابي لنفس البيانات قبل تبويبها ووضعها في جدول توزيع تكراري (76,62 سم) . إن الفرق هذا بين الرقمين يعود إلى فقدان المعلومات عن المفردات أو المشاهدات بسبب وضعها في مجاميع فنحن نفرض بأن طول كل النباتات في فئة معينة مساوياً لمركز تلك الفئة .

خواص الوسط الحسابي

A) مجموع إنحرافات القيم عن وسطها الحسابي تساوي صفر .

أي (للبيانات غير المبوبة) $\sum (y_i - \bar{y}) = 0$

أو (للبيانات المبوبة) $\sum f_i (y_i - \bar{y}) = 0$

البـــرهـان :

للبيانات غير المبوبة

$$\sum (y_i - \bar{y}) = \sum y_i - \sum \bar{y}$$

$$= \sum y_i - n\bar{y}$$

$$= \sum y_i - \cancel{n}\left(\frac{\sum y_i}{\cancel{n}} \right)$$

$$= \sum y_i - \sum y_i = \mathbf{0}$$

للبيانات المبوبة

$$\sum f_i (y_i - \bar{y}) = \sum f_i\, y_i - \bar{y} \sum f_i$$

$$= \sum f_i\, y_i - \left(\frac{\sum f_i y_i}{\cancel{\sum f_i}} \right) \cancel{\sum f_i}$$

$$= \sum f_i\, y_i - \sum f_i\, y_i = 0$$

والجدولان التاليان يوضحان ذلك :

y_i	$(y_i - \bar{y})$
8	0,4
3	– 4,6
5	– 2,6
12	4,4
10	2,4
$\sum y_i = 38$ $\bar{y} = 7,6$	$\sum (y_i - \bar{y}) = 0$

0,4 − (−4,6) − (−2,6) − 4,4 − 2,4 = 0

$$(y_i - \bar{y}) = y_1 - y_2 - y_3 - y_4 - y_5 - \left(\frac{\Sigma y_i}{n}\right) = 0,4 - (-4,6) - (-2,6) - 4,4 - 2,4 - \left(\frac{0}{5}\right) = 0$$

$$\Sigma y_i = y_1 - y_2 - y_3 - y_4 - y_5 = 8 - 3 - 5 - 12 - 10 = 38$$

$$\bar{y} = \left(\frac{\Sigma y_i}{n}\right) = \left(\frac{38}{5}\right) = 7,6$$

$f_i(y_i - \bar{y})$	$(y_i - \bar{y})$	$f_i y_i$	مركز الفئات y_i	التكرار f_i	الفئات
32,25 −	6,45 −	305	61	5	62 – 60
62,10 −	3,45 −	1152	64	18	65 – 63
18,90 −	0,45 −	2814	67	42	68 – 66
68,85 −	2,55 −	1890	70	27	71 – 69
44,40 −	5,55 −	584	73	8	74 – 72
$\Sigma f_i(y_i - \bar{y}) = 0$		$\sum f_i y_i = 6745$ $\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = 67,45$		$\Sigma f_i = 100$	

$$\Sigma f_i = f_1 - f_2 - f_3 - f_4 - f_5 = 5 + 18 + 42 + 27 + 8 = 100$$

$$\sum f_i y_i = (5*61) + (18*64) + (42*67) + (27*70) + (8*73) = 305+1152+2814+1890+584 = 6745$$

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6745}{100} = 67,45$$

B) مجموع مربعات الانحرافات القيم عن الوسط الحسابي هي أقل ما يمكن أي أقل من مجموع مربعات الانحرافات عن أية قيمة غير الوسط الحسابي نفسه أي أن $\sum (y_i - \bar{y})^2$ أقل ما يمكن .

البـــرهـان :

نفرض أن A هو أي قيمة أو وسط فرضي غير الوسط الحسابي فسنبرهن بأن $\sum (y_i - A)^2$ هي أكبر من قيمة $\sum (y_i - A)^2$.

$\sum (y_i - A)^2 =$ القاعدة / مربع الاول هو (y_i^2) – ضعف الاول × الثاني هو $(2Ay_i)$ + مربع الثاني هو (A^2)

$= \sum(y_i^2 - 2Ay_i + A^2)$ نقوم بإدخال $\sum$ الى داخل القوس

$= (\sum y_i^2 - 2A \sum y_i + \sum A^2)$ كل ثابت يضرب في n أي أن A يعتبر كقيمة ثابتة

$= \sum y_i^2 - 2nA\bar{y} + nA^2$

وبإضافة وطرح $n(\bar{y})^2$ من أعلاه ينتج .

$$= \sum y_i^2 - 2nA\bar{y} + nA^2 + n(\bar{y})^2 - n(\bar{y})^2$$

$$= (\sum y_i^2 - n(\bar{y})^2) + nA^2 + 2A\bar{y} + (\bar{y})^2$$

$$= \sum(y_i - \bar{y})^2 + n(A - \bar{y})^2$$

من هذا يتضح بأن مجموع مربعات الانحرافات عن أي قيمة غير الوسط الحسابي هي أكبر من مجموع مربعات الانحرافات عن الوسط الحسابي بمقدار $n(A - \bar{y})^2$ وهو قيمة موجبة .

مثال (4)

من القيم التالية ← $y_i = 9, 8, 6, 5, 7$

أوجد كل من

1) $\bar{y} = \frac{\sum yi}{n} = 7$.

$$\bar{y} = \frac{9+8+6+5+7}{5} = \frac{35}{5} = 7$$

2) $\sum(y_i - \bar{y})^2$.

$$= (9-7)^2 + (8-7)^2 + (6-7)^2 + (5-7)^2 + (7-7)^2$$

$$= (2)^2 + (1)^2 + (-1)^2 + (-2)^2 + (0)^2$$

$$= 4 + 1 + 1 + 4 + 0 = 10$$

فلو طرحنا من القيم هذه أي رقم (غير الوسط الحسابي) وليكن : $A = 10$ فإن مجموع مربعات الانحرافات ستكون :

3) $\sum(y_i - A)^2 = \sum(y_i - 10)^2$

$$= (9-10)^2 + (8-10)^2 + (6-10)^2 + (5-10)^2 + (7-10)^2$$

$$= (1)^2 + (-2)^2 + (-4)^2 + (-5)^2 + (-3)^2$$

$$= 1 + 4 + 16 + 25 + 9 = 55$$

وطبيعي 55 أكبر من 10

ويلاحظ هنا أن الفرق بينهما هو $55 - 10 = 45$

وهو $n(A - \bar{y})^2$ حيث أن $\bar{y}$ هي $= \frac{\sum y_i}{n}$

أي $5(10-7)^2 = 45$

C) عند إضافة عدد ثابت (k) إلى كل قيمة من قيم المشاهدات فإن الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الأصلية + العدد الثابت (k) .

$$x_i = y_i + k$$

$$\bar{x}_i = \bar{y}_i + k$$

البـــرهـان :

$$x_i = y_i + k$$

$$\sum x_i = \sum (y_i + k)$$

$$= \sum y_i + \sum k$$ كل عدد ثابت نضع قبله (n)

$$= \sum y_i + n \sum k$$

$$\frac{\sum x_i}{n} = \frac{\sum y_i}{n} + \frac{nk}{n}$$

$$\bar{x}_i = \bar{y}_i + k$$

مثال (5)

نفرض أن لدينا القيم التالية ← $y_i = 8, 3, 2, 12, 10$

فإن الوسط الحسابي لها هو . $\bar{y} = \frac{\sum y_i}{n} = \frac{8+3+2+12+10}{5} = \frac{35}{5} = 7$

فإذا أضفنا لكل من هذه القيم قيمة ثابتة ولتكن (3) فالقيم الجديدة ستصبح .

$$x_i = 11, 6, 5, 15, 13$$

الوسط الحسابي للقيم الجديدة هو . $\bar{x} = \frac{\sum x_i}{n} = \frac{11+6+5+15+13}{5} = \frac{50}{5} = 10$

$$\bar{x}_i = \bar{y} + 3$$
$$= 7 + 3 = 10$$

مثال (6)

نفرض أن القيم الأصلية هي : ← $y_i = 5, 10, 8, 7, 10$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{5 + 10 + 8 + 7 + 10}{5} = \frac{40}{5} = 8$$

فإذا طرحنا 2 من كل مشاهدة .

فإن الوسط الحسابي للمشاهدات الجديدة سيكون .

$$\bar{y} - 2 = 8 - 2 = 6$$
$$x_i = 3, 8, 6, 5, 8$$
$$\bar{y} = \frac{30}{5} = 6$$

D) إذا ضربت كل قيمة من قيم المشاهدات في قيمة ثابتة (k) فإن :

الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الجديدة × العدد الثابت k .

$$z_i = k\, y_i$$
$$\bar{z} = k\, \bar{y}$$

البــرهــان :

$$z_i = k\, y_i$$
$$\sum z_i = k \sum y_i$$
$$\frac{\sum z_i}{n} = k\, \frac{\sum y_i}{n}$$
$$\therefore\ \bar{z} = k\, \bar{y}$$

مثال (7)

في القيم التالية : ← $y_i = 8, 3, 2, 12, 10$.

نجد أن : ← $\bar{y} = 7$

فإذا كان : ← $\bar{z} = 5\,\bar{y}$

أوجـد قيـمة ($\bar{z}$) .

بما انه عدد المشاهدات 5

إذاً نضرب 5 في كل قيمة من قيم yi

$5 \times 8 = 40$. $5 \times 3 = 15$

$5 \times 2 = 10$. $5 \times 12 = 60$

$5 \times 10 = 50$

الـحــل :

$z_i = 40, 1k5, 10, 60, 50$.

$$\bar{z} = \frac{\sum z_i}{n} = \frac{40 + 15 + 10 + 60 + 50}{5} = \frac{175}{5} = 35$$

$$= 5 \times (\bar{y})$$

$$\bar{z} = 5 \times 7 = 35$$

◆ هذا ويمكن تعميم الخاصيتين السابقتين بالقانون التالي :

$$x_i = a + b\,y_i$$

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

E) الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين .

$$z_i = k\,y_i + y_i$$

$$\bar{z} = \bar{x} + \bar{y}$$

البــرهـان :

$$z_i = x_i + y_i$$

$$\sum z_i = \sum (x_i + y_i)$$

$$\sum z_i = \sum x_i + \sum y_i)$$

$$\frac{\sum z_i}{n} = \frac{\sum x_i}{n} + \frac{\sum y_i}{n}$$

$$\bar{z} = \bar{x} + \bar{y}$$

مثال (8)

اعتبر الجدول التالي .

x_i	y_i	$z_i = x_i + y_i$
2	5	7
4	10	14
4	8	14
8	7	15
5	10	15
$\bar{x} = 5$	$\bar{y} = 8$	$\bar{z} = 8$

ومن هذا يتضح بأن :

$$\bar{z} = \bar{x} + \bar{y}$$

$$= 5 + 8 = 13$$

ملاحظة .

$$\bar{z} = \frac{\sum z_i}{n} = \frac{7 + 14 + 14 + 15 + 15}{5} = 13$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{5 + 10 + 8 + 7 + 10}{5} = 8$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2 + 4 + 4 + 8 + 5}{5} = 5$$

F) إذا كان لكل قيمة من المشاهدات (yi) وزن خاص يتناسب مع أهميتها (wi) فإن الوسط الحسابي الموزون لهذه القيم هو :

مثال (9)

القيم التالية تمثل نتائج امتحان أحد الطلبة في درس الإحصاء علماً بأن لكل امتحان وزناً أو أهمية أو نسبة معينة .

$w_i\ y_i$	أهميتها أو نسبتها أو وزنها (w_i)	الدرجة (yi)	الامتحان
700	% 10	70	الأول
1800	% 30	60	الثاني
750	% 10	75	الثالث
2750	% 50	55	الرابع
$\sum w_i y_i = 6000$	$\sum w = 100$		

فالوسط الحسابي أو معدل الطالب سيكون :

$$\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{700 + 1800 + 750 + 2750}{100} = \frac{6000}{100} = 60$$

مثال (10)

أربع شعب من الطلبة في الصف الأول تتألف من 30 و 35 و 40 و 25 طالباً على التوالي فإذا كان معدل امتحانهم بمادة الإحصاء هو 80 و 75 و 60 و 90 على التوالي فما هو معدل الامتحان في جميع هذه الشُعب ؟

الـحـــل :

$$\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{(30)(80)+(35)(75)+(40)(60)+(25)(90)}{30+35+40+25} = \frac{9675}{130} = 74,4$$

ثانياً– الوسط الهندسي The Geometric Mean

(*a*) بيانات غير المبوبة .

إذا كان لدينا n من القيم أو المشاهدات : $(y_1, y_2, \ldots\ldots\ldots\ldots\ldots y_n)$ فإن الوسط الحسابي لها يرمز له بالرمز $\bar{G}$ هو :

$$\bar{G} = \sqrt[n]{(y_1)(y_2)\ldots\ldots\ldots\ldots(y_n)}$$

ولإيجاد قيمة الوسط الهندسي نستخدم اللوغاريتمات فعند أخذ لوغارتم الطرفين ينتج :

$$\text{Log} = \frac{L}{n} = \frac{\text{Log}\,[(y_1)(y_2)\ldots\ldots\ldots(y_n)]}{n}$$ نقوم بإدخال Log إلى الأقواس

$$\text{Log} = \frac{L}{n} = \frac{\text{Log}(y_1) + \text{Log}(y_2)\ldots\ldots\ldots + \text{Log}\,(y_n)}{n}$$

من ذلك يتضح بأن لوغارتم الوسط الهندسي لمجموعة من القيم هو الوسط الحسابي للوغاريتمات هذه القيم. ولإيجاد قيمة الوسط الهندسي بعد ذلك نستخدم العدد المقابل لـ $\text{Log}\,\bar{G}$.

مثال (11)

أوجد الوسط الهندسي والوسط الحسابي للقيم التالية : ⟵ $y_i = 3, 5, 8, 3, 7, 2$.

الـحـــل :

$$\bar{G} = \sqrt[n]{(y_1)(y_2)\ldots\ldots\ldots\ldots(y_n)}$$

$$\text{Log}\,\bar{G} = \frac{\text{Log}}{n} = \frac{\text{Log}(y_1) + \text{Log}(y_2)\ldots\ldots\ldots + \text{Log}\,(y_n)}{n}$$

$$= \frac{\text{Log } 3 + \text{Log } 5 + \text{Log } 8 + \text{Log } 3 + \text{Log } 7 + \text{Log } 2}{n}$$

$$= \frac{0,4771 + 0,6989 + 0,9030 + 0,4771 + 0,8450 + 0,3010}{n}$$

$$= \frac{3,7024}{6} = 0,6170 \longrightarrow \therefore \ \overline{G} = 4,14$$

أو يمكن حساب الوسط الهندسي بالطريقة التالية :

$$\overline{G} = \sqrt[6]{(3)\ (5)\ (8)\ (3)\ (7)\ (2)} = \sqrt[6]{5760}$$

$$\therefore \text{ Log } \overline{G} = \frac{1}{6} \text{ Log } 5760 = \frac{3,7024}{6} = 0.6171$$

$$\therefore \text{ Log } \overline{G} = 4.14$$

أما الوسط الحسابي فهو :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{3+5+8+3+7+2}{6} = \frac{28}{6} = 4.47$$

من ذلك يتضح بأن الوسط الهندسي لقيم مختلفة موجبة دائماً أصغر من الوسط الحسابي هذا وأكثر ما يستعمل الوسط الهندسي هو في الأرقام القياسية للأسعار أو إيجاد متوسط عدد من النسب أو في إيجاد معدلات التغير في المبيعات أو السكان ...... الخ . كما إنه لا يمكن إيجاد الوسط الهندسي إلا إذا كانت مجموع القيم موجبة .

(*b*) من بيانات المبوبة .

إذا كانت $y_1, y_2 \ldots\ldots y_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $y_1, y_2, \ldots. y_n$ على التوالي فالوسط الهندسي هو :

$$\overline{G} = \sqrt[\sum f_i]{(y_1)^{f_i} \times (y_2)^{f_i} \times \ldots\ldots.. \times (y_k)^{f_i}}$$

وباستخدام اللوغاريتمات فإن :

$$\text{Log}\,\overline{G} = \frac{\sum f_i \,\text{Log}\, y_i}{\sum f_i}$$

$$= \frac{f_1 \,\text{Log}\, y_1 + f_2 \,\text{Log}\, y_2 + \cdots \ldots \ldots + f_k \,\text{Log}\, y_k}{\sum f_i}$$

مثال (12)

أوجد الوسط الهندسي من جدول التوزيع التكراري التالي :

فئات الوزن (كغم)	عدد الطلبة
62 – 60	5
65 – 63	18
68 – 66	42
71 – 69	27
74 – 72	8
المجموع	100

الحل :

الفئات	f_i	y_i	$\text{Log}\, y_i$	$f_i \,\text{Log}\, y_i$
62 – 60	5	61	1.7782	8.8910
65 – 63	18	64	1.8062	32.5116
68 – 66	42	67	1.8261	76.6962
71 – 69	27	70	1.8451	49.8177
74 – 72	8	73	1.8633	14.9064
المجموع	$\sum f_i = 100$	$\sum y_i = 335$		$\sum f_i \,\text{Log}\, y_i = 182.8229$

$$\text{Log}\,\overline{G} = \frac{\sum f_i \,\text{Log}\; y_i}{\sum f_i} = \frac{182.9229}{100} = 1.8282 \qquad \therefore\;\; \overline{G} = 67.3$$

بينما الوسط الحسابي هو :

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{(5*61)+(18*64)+(42*67)+(27*70)+(8*73)}{100}$$

$$= \frac{305 + 1152 + 2814 + 1890 + 584}{100} = = \frac{6745}{100} = 67.45$$

ثالثاً– الوسط التوافقي The Harmonic Mean

a) بيانات غير المبوبة .

إذا كان لدينا n من القيم أو المشاهدات $y_1, y_2 \ldots\ldots y_n$ فإن الوسط التوافقي لها (يرمز له بالرمز $\overline{H}$) هو :

$$\overline{H} = \frac{1}{\dfrac{\sum \frac{1}{y_i}}{n}} = \frac{n}{\sum \frac{1}{y_i}}$$

- فالوسط التوافقي هو مقلوب الوسط الحسابي لمقلوب القيم أو المشاهدات .

مثال (13)

أوجد الوسط التوافقي للقيم التالية : $y_i = 3, 5, 6, 6, 7, 10, 12$

الحـــل :

$$\overline{H} = \frac{n}{\sum \frac{1}{y_i}} = \frac{n}{\frac{1}{y_1}+\frac{1}{y_2}+\frac{1}{y_3}+\frac{1}{y_4}+\frac{1}{y_5}+\frac{1}{y_6}+\frac{1}{y_7}}$$

$$= \frac{n}{\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}}$$

عند القسمة نأخذ رقمين بعد الفارزة

$$= \frac{7}{0,33 + 0,2 + 0,16 + 0,16 + 0,14 + 0,1 + 0,08}$$

$$= \frac{7}{1,17} = 5.98$$

إن الوسط التوافقي أكثر ما يستعمل هو عندما نعطي مجموعة من البيانات منسوبة إلى وحدة ثابتة .

مثال (15)

اشترى مزارع بذور حنطة بـ 100 دينار من كل من الشركات التالية :

الشركة الأولى كان سعر الطن من بذور الحنطة = 20 ديناراً .

والشركة الثانية كان سعر الطن من بذور الحنطة = 25 ديناراً .

أما الشركة الثالثة فكان سعر الطن من بذور الحنطة = 50 ديناراً .

* فما هو سعر الطن من بذور الحنطة :

ملاحظة

البيانات معبرة بـ عدة أطنان بال 100 دينار بينما المطلوب هو متوسط سعر الطن .

الــحـــل :

$$\overline{H} = \frac{n}{\sum \frac{1}{y_i}} = \frac{n}{\frac{1}{20}+\frac{1}{25}+\frac{1}{50}} = \frac{n}{0,05 + 0,04 + 0,02}$$

$$= \frac{3}{0,11} = 27,27$$

a) للبيانات المبوبة .

إذا كان $y_1, y_2 \ldots\ldots y_k$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها : $f_1, f_2 \ldots . f_k$ على التوالي :

فالوسط التوافقي لها هو :

$$\overline{H} = \frac{\sum f_i}{\sum \frac{f_i}{y_i}} =$$

مثال (14)

أوجد الوسط التوافقي للتوزيع التكراري التالي :

y_i	f_i	الفئات
61	5	62 – 60
64	18	65 – 63
67	42	68 – 66
70	27	71 – 69
73	8	74 – 72
$\sum f_i = 335$	$\sum f_i = 100$	المجموع

الــحـــل :

$$\overline{H} = \frac{\sum f_i}{\sum \frac{f_i}{y_i}} = \frac{\sum f_i}{\frac{f_1}{y_1} + \frac{f_2}{y_2} + \frac{f_3}{y_3} + \frac{f_4}{y_4} + \frac{f_5}{y_5}}$$

$$= \frac{100}{\frac{5}{61} + \frac{18}{64} + \frac{42}{67} + \frac{27}{70} + \frac{8}{73}}$$

$$= \frac{100}{0{,}08 + 0{,}28 + 0{,}62 + 0{,}38 + 0{,}10}$$

$$= \frac{100}{0{,}08 + 0{,}28 + 0{,}62 + 0{,}38 + 0{,}10}$$

$$= \frac{100}{1{,}46} = 68.49$$

رابعاً- الوسط التربيعي The Quadratic Mean

a) للبيانات غير المبوبة .

إذا كان لدينا n من القيم أو المشاهدات $y_1, y_2 \ldots\ldots y_n$ فإن الوسط التربيعي لها (ويرمز بـ $\overline{Q}$) هو:

$$\overline{Q} = \sqrt{\frac{\sum {y_i}^2}{n}}$$

أي أن الوسط التربيعي هو الجذر التربيعي للوسط الحسابي لمربعات القيم أو المشاهدات .

مثال (16)

أوجد الوسط التربيعي للبيانات التالية : $y_i = 1, 3, 4, 5, 7$

الحـــل :

$$\bar{Q} = \sqrt{\frac{\sum y_i^2}{n}} = \sqrt{\frac{y_1^2 + y_2^2 + \cdots \ldots\ldots + y_n^2}{n}}$$

$$= \sqrt{\frac{(1)^2 + (3)^2 + (4)^2 + (5)^2 + (7)^2}{5}}$$

$$= \sqrt{\frac{1 + 9 + 16 + 25 + 49}{5}} = \sqrt{\frac{100}{5}}$$

$$= \sqrt{\frac{100}{5}} = \sqrt{20} = 4.47$$

♦ إن الوسط التربيعي يطبق بكثرة في العلوم الفيزيائية .

a) للبيانات المبوبة .

إذا كانت $y_1, y_2 \ldots\ldots y_k$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها : $f_1, f_2 \ldots\ldots f_k$ على التوالي، فالوسط التربيعي لها هو :

$$\bar{Q} = \sqrt{\frac{\sum f_i y_i^2}{\sum f_i}}$$

مثال (17)

أوجد الوسط التربيعي للتوزيع التكراري من الجدول التالي :

الفئات	f_i	y_i	f_i^2	$f_i y_i^2$
62 – 60	5	61	3721	18605
65 – 63	18	64	4096	73728
68 – 66	42	67	4489	188538
71 – 69	27	70	4900	132300
74 – 72	8	73	5329	42632
المجموع	$\sum f_i = 100$	$\sum f_i = 335$	$\sum f_i^2 = 22535$	$\sum f_i y_i^2 = 455803$

الحـــــل :

$$\bar{Q} = \sqrt{\frac{\sum f_i y_i^2}{\sum f_i}} = \sqrt{\frac{455803}{100}} = \sqrt{4558{,}03} = 67.51$$

خامساً– الوسيط The Median

a) للبيانات غير المبوبة .

إذا كان لدينا n من القيم أو المشاهدات $y_1, y_2 \ldots\ldots y_n$ ورتبت ترتيباً (تصاعدياً أو تنازلياً) .

(1) فإذا كانت n عدد فردي فإن الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$ أي الوسيط يرمز له بـ $\overline{Me}$ هو .

$$\overline{Me} = \frac{y\,(\,n+1\,)}{2}$$

(2) أما إذا كانت n عدد زوجي فإن الوسيط هو الوسط الحسابي للقيمتين اللتين ترتيبهما هو :

$$\overline{Me} = \frac{n}{2} + 1, \qquad \overline{Me} = \frac{n}{2}$$

$$\overline{Me} = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$ أي إن

مثال (18)

أوجد الوسيط لدرجات طالب في خمسة امتحانات بدرس الإحصاء إذا كانت الدرجات هي :

= 80, 82, 76, 87, 84

الحـــل :

نرتب الدرجات تصاعدياً . (من أقل قيمة وإلى أعلى قيمة) .

87, 84, 82, 80, 76

وبما أن عدد الأرقام فردي فإن ($n = 5$) . أي إن عدد المشاهدات = 5 .

إذا فالوسيط هو القيمة التي ترتيبها :

$$= \frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$$

$\overline{Me} = y_3 = 82$ أي أن الوسيط = 82

مثال (19)

أوجد الوسيط للقيم التالية : ← $y_i = 5, 4, 8, 7, 3, 12, 9, 2$

الحـــل :

نرتب القيم تصاعدياً

$y_i = 2, 3, 4, 5, 7, 8, 9, 12$

وبما أن عدد القيم زوجي فإن ($n = 8$) . أي إن عدد المشاهدات = 8 .

إذاً فالوسيط الحسابي للقيمتين اللتين ترتيبهما :

$$= \frac{n}{2} + 1, \quad \frac{n}{2} \; .$$

$$= \frac{n}{2} = \frac{8}{2} = 4$$

$$= \frac{n}{2} + 1 = \frac{8}{2} + 1 = 5$$

$$\overline{Me} = \frac{y_4 + y_5}{2} = \frac{5 + 7}{2} = 6$$

a) للبيانات المبوبة .

إذا كانت $y_1, y_2 \ldots\ldots y_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2 \ldots\ldots f_n$ على التوالي فقيمة الوسيط لهذه البيانات (بالاستعانة بجدول التوزيع التكراري المتجمع الصاعد) هو :

$$\overline{Me} = L_1 + \left[\frac{\frac{\sum f_i}{2} - F_i}{f_i} \right] W \; .$$

حيث أن

* الحد الادنى الحقيقي لفئة الوسيط = L_1

* مجموع التكرارات = $\sum f_i$

* التكرار المتجمع عند بداية فئة الوسيط = F_i

* تكرار فئة الوسيط = التكرار المتجمع عند نهاية فئة الوسيط − التكرار المتجمع عند بداية فئة الوسيط
$f_i =$

* طول فئة الوسيط = W

ايجاد قيمتين متتاليتين في التكرار التجميعي التصاعدي يقع بينهما ترتيب الوسيط . يقابل هاتين القيمتين حدا فئة الوسيط الأدنى والأعلى ويستحسن أخذ الحدود الحقيقية لهذه الفئة .

مثال (20)

أوجد الوسيط للتوزيع التكراري من جدول التالي :

فئات الطول	عدد الطلبة (التكرار)
62 – 60	5
65 – 63	18
68 – 66	42
71 – 69	27
74 – 72	8
المجموع	100

الحـــــل :

فئات الطول	التكرار f_i	المتجمع الصاعد		فئة الوسيط
			التكرار التجميعي F_i	
62 – 60	5	أقل من 60	0	نستخرجها كما موضح في الخطوة التالية
65 – 63	18	أقل من 63	5	
68 – 66	42	أقل من 66	23	
71 – 69	27	أقل من 96	65	
74 – 72	8	أقل من 72	92	
المجموع	$\sum f_i = 100$	أقل من 74	100	

ما قبل الفئة الأولى = F_0 .

الفئة الأولى = $F_1 = 5$.

الفئة الثانية = $F_1 + F_2 = 5 + 18 = 23$.

الفئة الثالثة = $F_1 + F_2 + F_3 = 5 + 18 + 42 = 65$.

الفئة الرابعة = $F_1 + F_2 + F_3 + F_4 = 5 + 18 + 42 + 27 = 92$.

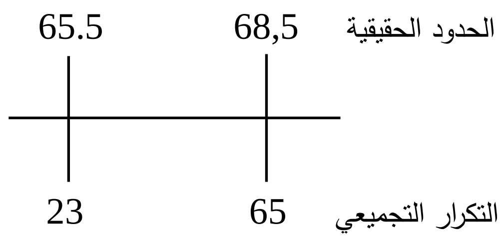
الفئة الخامسة = $F_1 + F_2 + F_3 + F_4 + F_5 = 5 + 18 + 42 + 27 + 8 = 100$.

ترتيب الوسيط $\longleftarrow$ $\therefore \quad \frac{\sum f_i}{2} = \frac{100}{2} = 50$

أي أن الوسيط هو طول الشخص الذي ترتيبه 50 (بعد ترتيب القيم تصاعدياً أو تنازلياً) .

وفي جدول التوزيع التكراري التجميعي التصاعدي نرى أن 50 هي واقعة بين الرقمين (23 و 65) .

إذن فئة الوسيط هي :



الحد الأدنى الحقيقي لفئة الوسيط $\longleftarrow$ $L_1 = 65{,}5$

التكرار التجميعي عند بداية فئة الوسيط $\longleftarrow$ $F_i = 23$

تكرار فئة الوسيط (الفرق بين الفيئتين للتكرار التجميعي) $\longleftarrow$ $f_i = 65 - 23 = 42$

طول فئة الوسيط = (الحد الحقيقي الأعلى – الحد الحقيقي الأدنى) $\longleftarrow$ $W = 68{,}5 - 65{,}5 = 42$

$$\overline{Me} = L_1 + \left[\frac{\frac{\sum f_i}{2} - F_i}{f_i} \right] W .$$

$$\overline{Me} = 65{,}5 + \left[\frac{50 - 23}{42} \right] (3) = 67{,}42 \text{ inch}$$

العملية تجري كالتالي (الطرح ثم القسمة ثم الضرب ثم الجمع) .

ويمكن تلخيص خطوات إيجاد الوسيط للبيانات المبوبة كالآتي :

1) عمل جدول توزيع تكراري تجميعي تصاعدي .

2) إيجاد ترتيب الوسيط $= \frac{\sum f_i}{2}$.

3) إيجاد فئة الوسيط .

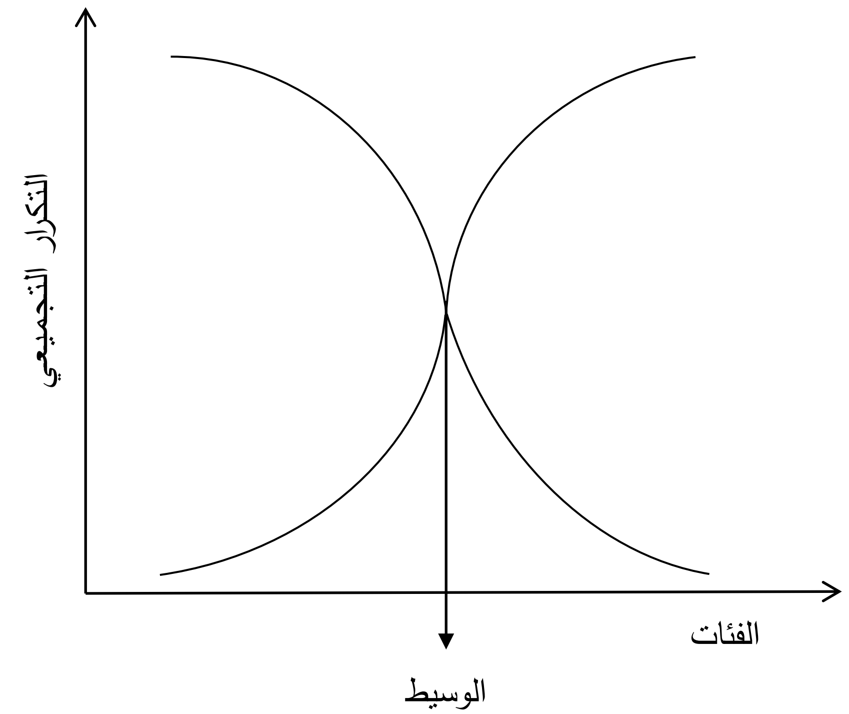
a) إيجاد الحدود الحقيقية .

b) كتابة التكرار التجمعي التصاعدي أمام كل منها .

ملاحظة

من الممكن إيجاد ترتيب الوسيط بـ ($\frac{\sum f_i}{2}$) إذا كان عدد المفردات فردياً أو بـ ($\frac{\sum f_i}{2} + 1$) إذا كان عدد المفردات زوجياً . ولكن نظراً لكون المفردات كبيرا في التوزيعات التكرارية فتستخدم ($\frac{\sum f_i}{2}$) لإيجاد ترتيب الوسيط .

هذا ويمكن إيجاد قيمة الوسيط بإستخدام الرسم البياني للمنحنيين التصاعدي والتنازلي وذلك بإنزال عمود من نقطة تقاطعهما إلى المحور السيني ليقطعه في نقطة هي قيمة الوسيط كما في الشكل :



المضلع التكراري التجميعي التصاعدي والتنازلي مع موفع الوسيط .

سادساً- المنوال أو القسمة The Mode

a) للبيانات غير المبوبة .

إذا كان لدينا n من المشاهدات $y_1, y_2 \ldots\ldots y_n$ فإن المنوال لهذه المشاهدات هو المشاهدة أو القيمة الأكثر تكراراً بين هذه المشاهدات ويرمز لها بـ ($\overline{\mathrm{Mo}}$) .

ومن هذا يتضح بأن قد يكون هنالك منوالاً وأحداً (قمة واحدة) اهذه المشاهدات وعندها يسمى التوزيع وحيد القمة unimodal أو يكون لها منوالان (قمتان) وعندها يسمى التوزيع ذو القمتين bimodal وقد يكون لهأجد المنوال لكل من البيانات التالية :

مثال (21)

A) 3, 5, 2, 6, 5, 9, 5, 2, 8, 6,

B) 48,9 . 49,5 . 50,3 . 48,7 . 51,6

الـحـل :

A) المفردة 5 هي أكثر المفردات تكراراً فهي المنوال ⟵ $\overline{Mo} = 5$.

B) لا يوجد منوال لهذه المفردات .

b) للبيانات المبوبة .

إذا كانت القيم $y_1, y_2, ...y_k$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2,$ $........f_k$ على التوالي فإن المنوال هو :

$$\overline{Mo} = L_1 + [\frac{d_1}{d_1 + d_2}] \times W$$

حيث أن

فئة المنوال هي تلك الفئة التي تملك أكبر التكرارات وإن :

1) الحد الأعلى الحقيقي لفئة المنوال = L_1

2) الفرق بين تكرار فئة المنوال والفئة السابقة لها = d_1

3) الفرق بين تكرار فئة المنوال والفئة اللاحقة لها = d_2

4) طول الفترة = W

مثال (21)

أوجد المنوال للجدول التوزيع التكراري التالي :

فئات الطول	عدد الطلبة (التكرار)
62 – 60	5
65 – 63	18
68 – 66	42
71 – 69	27
74 – 72	8
المجموع	100

فئة المنوال : إن الفئة (66 - 68) لها أكبر التكرارات (42) ف هي فئة المنوال .

الحد الأدنى الحقيقي لفئة المنوال = حد الفئة – 0,5

$66 - 0,5 = 65,5$

$L_1 = 65,5$

$d_1 = 42 - 18 = 24$

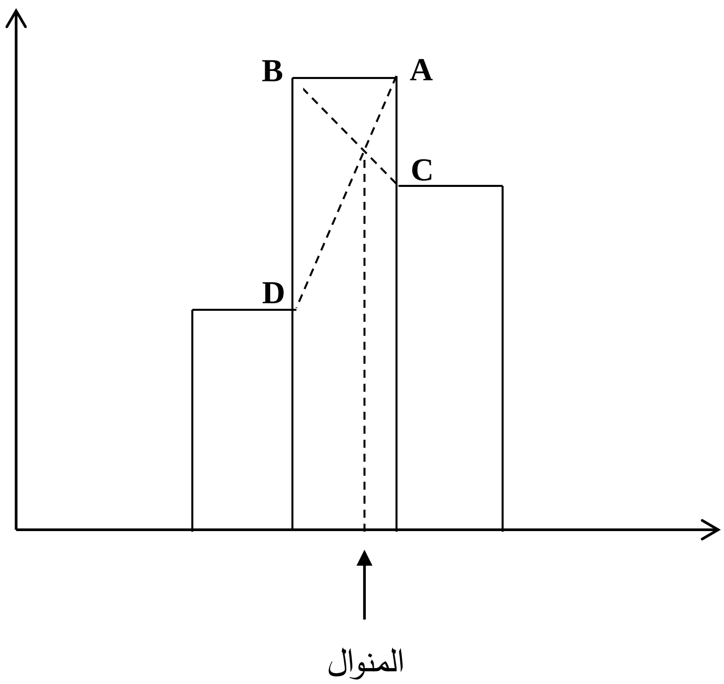
$d_2 = 42 - 27 = 15$

$W = 3$

$$\overline{\mathrm{Mo}} = 65,5 + \left[\frac{24}{24+15} \right] \times (3) .$$

$$\overline{\mathrm{Mo}} = 67,34 \simeq 67,35$$

هذا ويمكن تقدير المنوال بطريقة الرسم البياني للمدرج التكراري وذلك باستعمال مستطيل الفئة المنوالية (وهو أعلى مستطيل لأنه يمثل أكثر التكرارات) والمستطيلان المجاوران له كما في الشكل التالي .



الشكل يمثل المدرج التكراري وموقع المنوال .

حيث نصل A مع D و B مع C ومن نقطة تلاقيهما ننزل عموداً على المحور السيني يقطعه في نقطة هي قيمة المنوال .

تمارين الفصل الرابع

1- البيانات التالية تمثل متوسط لمحصول الدونم من الذرة الصفراء في عينة مكونة من 40 مزرعة في العراق :

588	1023	920	650
796	890	1230	500
800	980	158	420
680	1270	840	750
320	1261	960	720
1050	713	895	350
860	930	368	560
390	660	793	620
495	760	395	480
820	490	1056	630

المطلوب :

1) إيجاد كل من ما يأتي :

A) حساب الوسط الحسابي والوسيط والمنوال لهذه البيانات .

B) إعرض هذه البيانات في جدول توزيع تكراري . 200 ومبتدأ بالفئة الأولى (300 – 499) .

1) الوسط الحسابي .

2) الوسيط .

3) المنوال .

وقارن بين النتيجتين في (A) و (B) .

2) البيانات التالية تمثل عدد الجوز على النباتات (30 ,22 ,28 ,18) .

a) المتوسط الحسابي Arithmetic mean .

b) المتوسط الهندسي Geometric mean .

c) الوسط التوافقي Harmonic mean .

d) الوسط التربيعي Quadratic mean .

e) الوسيط Medium .

f) المنوال Mode .

g) متوسط المدى .

3) من جدل التوزيع التكراري التالي :

الفئات	**التكرار**
33 – 37	10
38 – 42	12
43 – 47	51
48 – 52	30
53 – 57	8

احسب بطريقة الرسم :

a) الوسيط .   b) المنوال .

4) من جدول التوزيع التكراري السابق أوجد :

a) الوسط الهندسي .   b) الوسط التوافقي .   d) الوسط التربيعي .

5) إذا علمت بأن $\bar{y} = 25$ أوجد الوسط الحسابي لكل من :

a) $x_i = yi + 5$.

b) $z_i = 2yi + 20$.

c) $u_i = \frac{3}{5} yi + 10$.

6) إذا علمت بأن : $x_i = 5y_i + 20$ وإن $\bar{x} = 100$ فما هو الوسط الحسابي لقيم y ؟

7) الجدول التالي يبين نتائج امتحان ثلاثة شعب في الصف الأول بمادة الإحصاء :

اسم الشعبة	عدد الطلبة	معدل درجاتهم
A	35	78
B	25	75
C	30	82

احسب الوسط الحسابي لجميع الشعب بمادة الإحصاء هذه :

8) الجدول التالي يبين توزيع عدد الأسر تبعاً لعدد الأفراد بالأسرة :

عدد أفراد الأسرة	التكرار
1	620
2	1000
3	1420
4	2100
5	800
6	460
المجموع	6400

احسب :

a) الوسط الحسابي b) الوسيط . c) المنوال .

9) من جدول التوزيع التكراري التالي لمحصول صنفين من القطن :

فئات المحصول	تكرار الصنف الأول	تكرار الصنف الثاني
25 – 20	5	1
31 – 26	18	25
37 – 32	25	50
43 – 38	40	25
49 – 44	3	20
55 – 50	22	19
المجموع	140	140

المطلوب :

a) مقارنة الوسط الحسابي لكلاً الصنفين. b) حساب الوسيط للصنف الأول بطريقة الرسم.

c) حساب المنوال للصنف الثاني بطريقة الرسم .

10) البيانات التالية يمثل الاجور الاسبوعية لعمال أربعة مصانع :

المصنع	عدد العمال	متوسط الاجور الاسبوعية (دينار)
1	200	12
2	250	8
3	150	15
4	400	7

المطلوب :

ايجاد متوسط الأجر الاسبوعي للعمال في جميع المصانع .

المدخل إلى الإحصاء

Introduction to statistics

الطبعة الثانية/ ٢٠٠٠

الفصل التاسع

المدرس المساعد

خليل ابراهيم خليل

وزارة التعليم العالي والبحث العلمي .
جامعة الموصل .
كلية الزراعة والغابات .
قسم المحاصيل الحقلية .

Ministry of Higher Education and Scientific Research .
University of Al Mosul .
College of Agriculture and Forestry .
Field crops department .

المدخل إلى الإحصاء

Introduction to statistics

الدكتور خاشع محمود الراوي

الطبعة الثانية/ ٢٠٠٠

الفصل التاسع

التوزيعات ألاحتمالية المتقطعة

Discrete Probability Distributions

(1:9) التوزيع ذي الحدين Binomial Distribution

يعتبر التوزيع ذي الحدين من أهم التوزيعات المتقطعة وسنشرح بعض الأمثلة لتسهيل مفهوم هذا التوزيع .

تأمل حدوث تجربة ما بحيث أن جميع النتائج Outcomes يمكن تصنيفها إلى ظهور حادث ما (وليكن A) أو عدم ظهوره. وعادةً يطلق على ظهور الحادث أو النتيجة بالنجاح Success وعدم ظهوره بالفشل Failure وكلمة النجاح هنا تستعمل فقط لتسهيل وصف ظهور الحادث. وهذه التجربة تكرر عدد من المرات وليكن (n).

ولنفرض بأن المتغير العشوائي (y) يمثل عدد النجاحات أي عدد ظهور الحادث الذي يظهر في تكرار التجربة (n) من المرات. إن هذا النوع من المتغيرات يسمى متغير ذو الحدين فهو مصنف إلى صنفين نجاح الحادث أو فشله وهو متقطع لأنه يأخذ قيماً عددية Counts من (الصفر إلى n).

إن تكرار التجربة هنا يكون تكراراً لأصل التجربة في كل مرة أي بمعنى آخر بأن التجارب المتكررة تكون مستقلة.

ففي حال تجربة رمي قطعة نقود ثلاث مرات $\therefore$ $(n = 3)$.

نفرض بأن النجاح هنا هو الحصول على صورة (H) .

وبذلك فإن (y) يمثل عدد الصور الناتجة في الثلاث رميات .