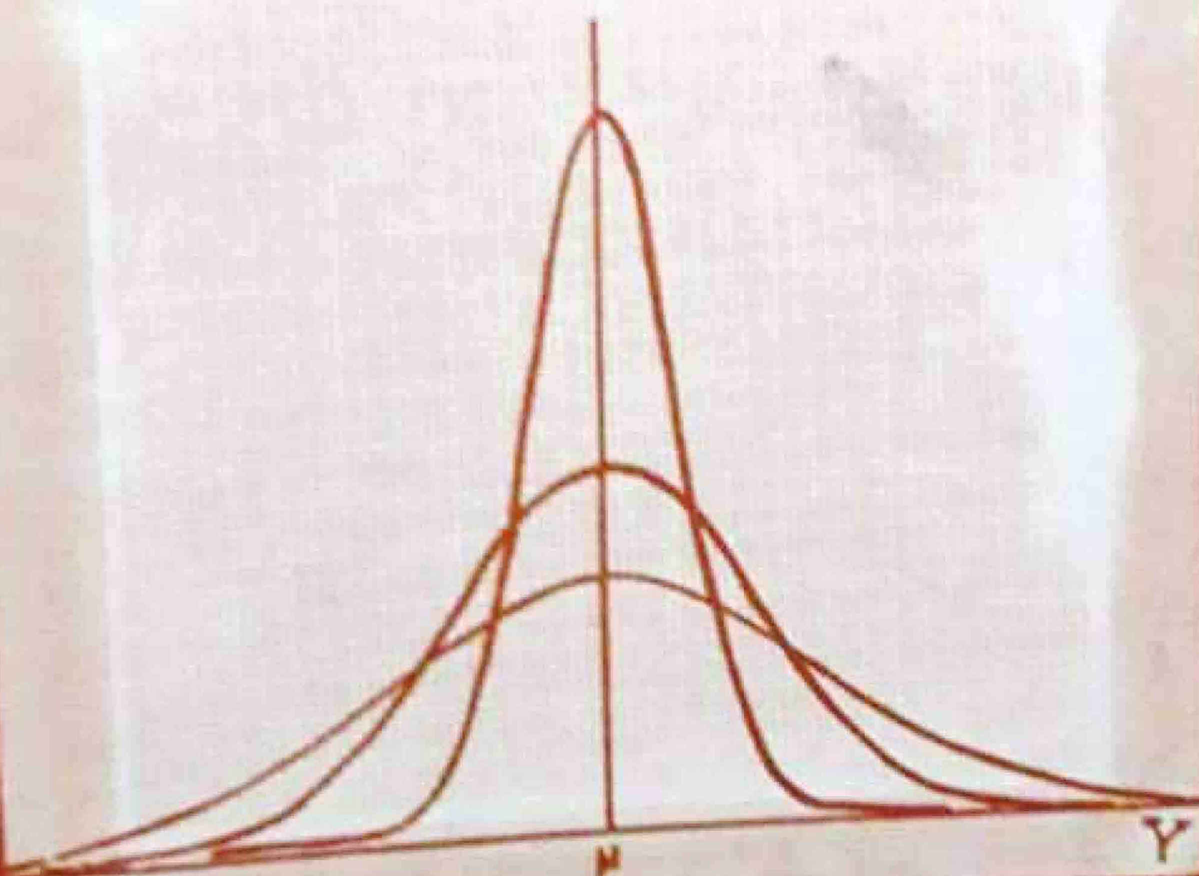
وزارة التعليم العالي والبحث العلمي
جامعة الموصل

المدخل الى الأحصاء



الدكتور خاشع محمود الراوي
كلية الزراعة والغابات / جامعة الموصل

طبعة ثانية

الفصل الثاني

طبيعة البيانات والرموز الاحصائية

(٢ : ١) طبيعة البيانات الاحصائية :

عند جمع بيانات حول ظاهرة ما فإننا نرمز للظاهرة بالرمز (y) وكل مفردة او مشاهدة منها نرمز لها بالرمز (y_i) . فمثلا عند دراسة اطوال الطلبة في احدى الجامعات فإننا نرمز لصفة الطول بالرمز (y) وطول اي طالب بالرمز (y_i) (وتسمى المشاهدة او المفردة (Observation)

هذا وان قيمة y_i قد تختلف من طالب الى آخر ولهذا نقول بأن y متغير « Variable ».

تعريف (٢ : ١) :

المتغير هو اي ظاهرة تظهر اختلافات بين مفرداتها ويرمز له بالرمز y (أو أي رمز آخر مثل x أو z ......).

والمتغيرات Variables تنقسم الى :

(١) متغيرات وصفية او نوعية (Qualitative variables)

وهي تلك الظواهر او الصفات التي لايمكن قياسها مباشرة بالارقام العددية مثل صفة لون العيون (ازرق ، اسود ، بني) والحالة الاجتماعية (غني ، متوسط الحال ، فقير) والجنس (ذكر ، انثى) .... الخ .

(٢) متغيرات كمية (Quantitative variables)

وهي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بأرقام عددية مثل : صفة الطول والوزن والعمر وكمية المحصول .... الخ .

هذا وتنقسم المتغيرات الكمية الى قسمين هما :

(أ) متغيرات مستمرة (او متصلة) (Continuous variables)

فالمتغير المستمر هو المتغير الذي تأخذ المشاهدة او المفردة فيه اية قيمة رقمية في مدى معين. فلو فرضنا بأن اطوال طلبة جامعة ما تتراوح بين ١٣٠,٥ و ١٧٠ سم فنقول بأن :

$$(130.5 \leq y \leq 170.0)$$

اى ان المتغير y ممكن ان يأخذ اية قيمة بين ١٣٠,٥ سم و ١٧٠ سم . وكأمثلة اخرى على المتغيرات المستمرة هي : الوزن وكمية المحصول ودرجة الحرارة والزمن ... لانه يمكن قياسها بأجزاء صغيرة جدا وتأخذ اية قيمة تقع في حدود معينة .

* وبصورة عامة فان كل البيانات التي تقاس (Measurements) تعتبر بيانات لمتغير مستمر .

(ب) متغيرات غير مستمرة (او منفصلة) (Discrete variables)

المتغير المنفصل هو المتغير الذي تأخذ المشاهدة او المفردة فيه قيماً متباعدة او متقطعة غير مستمرة .

فلو فرضنا ان عدد افراد الأسرة في اربع عوائل هي : ٢ ، ٣ ، ٤ ، ٥ فنقول بأن :

$$y = 2, 3, 4, 5 .$$

كذلك عند رمي زهر النرد (زار الطاولة) نجد ان النتيجة تكون ظهور الوجه ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ فنقول بأن

$$y = 1,2,3,4,5,6 .$$

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

* وبصورة عامة فان كل البيانات التي نحصل عليها من العد (Countings) تعتبر بيانات لمتغير منفصل .

(٢:٢) المجتمع والعينة Population and sample

(١) المجتمع Population

> **تعريف (٢:٢) :**
>
> المجتمع عبارة عن جميع القيم او المفردات التي يمكن ان ياخذها المتغير

فمثلا اذا كانت دراستنا متعلقة بأطوال طلبة جامعة ما فان المجتمع في هذه الحالة هو اطوال جميع الطلبة في تلك الجامعة .

والمجتمع اما ان يكون :

(أ) مجتمعاً محدودا (Finite population) :

أي ممكن حصر عدد مفرداته كما هو الحال في اطوال طلبة جامعة الموصل مثلا ، او عدد الوحدات الانتاجية في مصنع ما في يوم معين .

(ب) مجتمعاً غير محدود.(Infinite population)

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

(٢) العينة (Sample)

> **تعريف (٢ : ٣) :**
>
> العينة جزء من المجتمع.

فالعينة عبارة عن مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع . ان دراسة المجتمع ككل قد يكون صعباً او يحتاج الى وقت وجهد ومال، لذا فقد استعيض عن دراسة المجتمع بدراسة العينة وصفاتها ومنها نستطيع ان نستنتج خواص المجتمع الاصلي الذي اخذت منه هذه العينة .

(٢ : ٣) الرموز الاحصائية Statistical notations

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

وكما ذكرنا سابقا سنرمز للمتغير بالرمز y ولكل قيمة له بالرمز y_i

فلو كانت أعمار ٥ طلاب كالآتي : 20, 18, 24, 22, 16 سنة فنكتب

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

أي ان $y_1 = 20$ أي القيمة الاولى للمتغير أو المشاهدة الاولى.

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

وهكذا ... الى:

$y_n = 16$ أي القيمة الأخيرة $(n=5)$ للمتغير أو المشاهدة الأخيرة .

ويرمز عادة لمجموع قيم المتغير بالرمز

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

فالرمز $\sum$ هو حرف اغريقي يسمى (Sigma) أي مجموع الـ ... أو "Summation of"

والرقمان 1 و n هما حدا المجموع .

وعليه فالرمز $\sum_{i=1}^{n} y_i$ يقرأ كالآتي :

مجموع قيم y مبتدأ من المشاهدة الاولى وحتى الأخيرة أي :

$$\therefore \quad \sum_{i=1}^{n} y_i = y_1 + y_2 + \dots + y_n$$

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

وهناك مجموع جزئي مثل $\sum_{i=3}^{5} y_i$

أي مجموع المشاهدة الثالثة والرابعة والخامسة :

$$\sum_{i=3}^{5} y_i = y_3 + y_4 + y_5$$

ويرمز لمجموع مربعات جميع المشاهدات بالرمز $\sum_{i=1}^{n} y_i^2$ ويساوي :

$$\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + \dots + y_n^2$$

ويرمز لمربع مجموع المشاهدات بالرمز $\left(\sum_{i=1}^{n} y_i \right)^2$

$$(\sum y_i)^2 = (y_1 + y_2 + \dots + y_n)^2$$

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

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

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

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

مثال (١) نفرض بأن قيم المتغير y هي كالآتي :

$$y_i = 3, 9, 6, 2$$

وان 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) $(\sum y_i)^2$ (e) $\sum x_i y_i$ (f) $(\sum x_i)(\sum y_i)$

الحـــل :

(a) $\sum y_i = y_1 + y_2 + y_3 + y_4$

$$= 3 + 9 + 6 + 2 = 20$$

(b) $\sum_{i=2}^{3} y_i = y_2 + y_3$

$$= 9 + 6 = 15$$

(c) $\sum y_i^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2$

$$= (3)^2 + (9)^2 + (6)^2 + (2)^2 = 130$$

(d) $(\sum y_i)^2 = (y_1 + y_2 + y_3 + y_4)^2$

$$= (3 + 9 + 6 + 2)^2 = (20)^2$$

$$= 400$$

$$(e)\quad \sum x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$$
$$= (4)(3) + (2)(9) + (3)(6) + (7)(2)$$
$$= 62$$

$$(f)\quad (\sum x_i)(\sum y_i) = (x_1 + x_2 + x_3 + x_4)(y_1 + y_2 + y_3 + y_4)$$
$$= (16)(20)$$
$$= 320$$

هذا وفيما يلي بعض القواعد المفيدة في عملية الجمع

قاعدة (١)

اذا كانت (c) أي عدد ثابت فإن :

$$\sum_{i=1}^{n} c = nc$$

البرهان :

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

قاعدة (٢)

اذا كانت (c) أي عدد ثابت فإن $\sum c y_i = c \sum y_i$

البرهان :

$$\sum c y_i = c y_1 + c y_2 + \dots + c y_n$$
$$= c(y_1 + y_2 + \dots + y_n)$$
$$= c \sum y_i$$

قاعدة (٣)

جمع قيم متغيرين او اكثر هو مجموع جمعهم أي $\sum (x_i + y_i) = \sum x_i + \sum y_i$

البرهان :

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

$$\sum (y_i - 3) = (y_1 - 3) + (y_2 - 3) + (y_n - 3)$$

$$\sum y_i - 3 = y_1 + y_2 + \dots y_n - 3$$

$$= (x_1 + x_2 + \dots + x_n) + (y_1 + y_2 + \dots + y_n) \qquad \sum (y_i - 3) = \sum y_i - \sum 3$$

$$= \sum x_i + \sum y_i \qquad \sum y_i - nc$$

هذا ويجب التفريق بين بعض الرموز الاحصائية مثل :

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

بينما

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

كذلك فإن

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

تختلف عن

$$\sum x_i - 3$$

مثال (٢) اذا علمت بأن قيم كل من المتغيرين x و y هي كالآتي :

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

اوجد قيمة كل مما يأتي :

(أ) $\sum (y_i - x_i)^2$ (ب) $\sum (x_i - 3)(y_i - 5)$

(جـ) $\sum x_i y_i^2$ (د) $\sum (y_i - 3)$ (هـ) $\sum y_i - 3$

(و) $\sum \frac{x_i + 2}{y_i}$ (ز) $\frac{\sum (x_i + 2)}{\sum y_i}$

(ح) $\sum y_i^2 - \frac{(\sum y_i)^2}{n}$ (ط) $\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

الحل :

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

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

$$= 20$$

هذا ويمكن الوصول الى نفس النتيجة وذلك بفتح القوس ثم التعويض كما يلي :

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

$$= \sum y_i^2 - 2\sum x_i y_i + \sum x_i^2$$

وعلى القارىء ان يعوض فيها للتأكد من النتيجة السابقة .

$$(\text{ب})\sum (x_i-3)(y_i-5) = (x_1-3)(y_1-5) + (x_2-3)(y_2-5)$$

$$+ (x_3-3)(y_3-5) + (x_4-3)(y_4-5)$$

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

$$= 20$$

وهنا ايضاً يمكن الوصول الى نفس النتيجة بفتح الاقواس ثم التعويض كما يلي :

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

$$= \sum x_iy_i - 5\sum x_i - 3\sum y_i + (4)(15)$$

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

$$= 20$$

$$(\text{ج})\sum x_iy_i^2 = x_1y_1^2 + x_2y_2^2 + x_3y_3^2 + x_4y_4^2$$

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

$$= 616$$

$$(\text{د})\sum (y_i-3) = \sum y_i - \sum (3)$$

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

$$= \sum y_i - (4)(3)$$

$$= 20 - 12$$

$$= 8$$

$$(\text{هـ})\sum y_i - 3 = 20 - 3$$

$$= 17$$

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

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

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

$$= \frac{164}{36}$$

(ز) $$\frac{\sum(x_i + 2)}{\sum y_i} = \frac{\sum x_i + (n)(2)}{\sum y_i}$$

$$= \frac{12+8}{20}$$

$$= 1$$

(ح) $$\sum y_i^2 - \frac{(\sum y_i)^2}{n} = (y_1^2 + y_2^2 + y_3^2 + y_4^2) - \frac{(y_1 + y_2 + y_3 + y_4)^2}{4}$$

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

$$= 130 - \frac{(20)^2}{4}$$

$$= 130 - 100$$

$$= 30$$

(ط) $$\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = (x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4) - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= (2)(3) + (6)(9) + (3)(6) + (1)(2) - \frac{(12)(20)}{4}$$

$$= 80 - \frac{(12)(20)}{4}$$

$$= 20$$

covariance

حل الاسئلة

تمارين الفصل الثاني

(١) عين نوع المتغير (مستمر او متقطع) في كل من الحالات التالية :

(أ) عدد السيارات المباعة يوميا من الشركة العامة للسيارات .

(ب) درجات الحرارة المقاسة كل نصف ساعة في محطة الانواء الجوية في بغداد.

(ج) الدخل السنوي لاستاذ في احدى الجامعات .

(د) عدد الكتب على رفوف مكتبة كلية الزراعة .

(هـ) عدد انجات كمية المطر النازل في مدينة معينة خلال اشهر السنة .

(و) سرعة السيارة بالاميال في الساعة .

(ز) عدد الطلبة المقبولين في جامعة ما في عدة سنوات .

(٢) اكتب حدود كل مما يأتي :

(أ) $\sum_{i=2}^{5} x_i$ (ج) $\sum_{i=1}^{4} (y_i - 3)^2$

(ب) $\sum_{i=1}^{n} c$ (د) $\sum_{i=1}^{3} (x_i - 2y_i + 10)$

(٣) اكتب كلا من الحدود التالية مستعملا رمز الجمع :

(أ) $x_1^2 + x_2^2 + \dots + x_{10}^2$

(ب) $cx_1^3 + cx_2^3 + \dots + cx_{20}^3$

(ج) $(x_1 + y_1) + (x_2 + y_2) + \dots + (x_8 + y_8)$

(٤) برهن بأن :

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

علما بأن a و b و c هي اعداد ثابتة .

(٥) من القيم التالية :

$$x_1 = 7,\ x_2 = -2,\quad x_3 = 4$$
$$y_1 = 5,\ y_2 = 8,\quad y_3 = 2$$

اوجد قيمة كل مما يأتي :

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

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

(جـ) $\sum (x_i + y_i)(x_i - y_i)$

(د) $\sum (x_i - 8)$

(هـ) $\sum x_i - 8$

(و) $\sum \frac{(y_i^2 - 10)}{2x_i}$

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

(٦) برهن بان : $\sum y_i^2$ لاتساوي $(\sum y_i)^2$

وان : $\sum x_i y_i$ ، لاتساوي $(\sum x_i)(\sum y_i)$

(٧) اذا علمت بأن

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

برهن بأن

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

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

A ② B ①

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

C ③

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

(٨) اذا علمت بأن

$$\sum x_i = -4 \quad , \quad \sum x_i^2 = 10$$

احسب كل من :

(أ) $\sum (2x_i + 3)$

(ب) $\sum x_i(x_i - 1)$

(ج) $\sum (x_i - 5)^2$

$A = C$

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

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

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

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

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

$A = C$

$A = B = C$

الفصل الثالث

منها

العرض الجدولي والتمثيل البياني

(٣ : ١) مقدمة

عند جمع البيانات الاولية (Raw data) الخاصة بدراسة ظاهرة ما فإنه عادة لايمكن الاستفادة منها وهي بهذه الصورة . لذلك فغالبا ما توضع في جداول مبسطة او يعبر عنها في صورة اشكال ورسوم بيانية لكي يسهل دراستها وتحليلها .

(٣ : ١) العرض الجدولي :. Tabular presentation

هناك نوعان رئيسيان من الجداول الاحصائية وهما :

(١) الجدول البسيط : وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة . ويتألف عادة من عمودين : الاول يمثل تقسيمات الصفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة مثل جدول (٣ : ١) و (٣ : ٢) .

جدول (٣ : ١) . جدول توزيع عدد من طلبة جامعة ما حسب اوزانهم (بالكيلوغرامات)

فئات الوزن (كغم)	عــدد الطلبــة
٦٠ – ٦٢	٥
٦٣ – ٦٥	١٥
٦٦ – ٦٨	٤٥
٦٩ – ٧١	٢٧
٧٢ – ٧٤	٨
المجـــموع	١٠٠

جدول (٣:٢) جدول توزيع اعضاء البعثات الموفدين الى الخارج حسب مواد الدراسة لسنة ١٩٧١/١٩٧٠

موضوع البعثة	عدد الطلبة
علوم اساسية	٢٥
علوم زراعية	٥٠
علوم بيطرية	٢٠
علوم هندسية	٧٥
علوم طبية	٥٠
علوم اجتماعية	٣٠
المجموع	٢٥٠

(٢) الجدول المركب : وهو الجدول الذي توزع فيه البيانات حسب صفتين أو ظاهرتين أو أكثر في نفس الوقت .
فمثلاً الجدول المزدوج (لصفتين) يتألف من :
الصفوف : وتمثل فئات أو مجاميع احدى الصفتين ،
والاعمدة : وتمثل فئات أو مجاميع الصفة الاخرى .
اما المربعات التي تقابل الصفوف والاعمدة فتحتوي على عدد المفردات أو التكرارات المشتركة في فئات ومجاميع كلا الصفتين مثل جدول (٣:٣).

جدول (٣:٣) جدول توزيع عدد من طلبة كلية ما حسب صفتي الطول والوزن

الوزن (كغم) الطول (سم)	٥١—٦٠	٦١—٧٠	٧١—٨٠	المجمـــوع
١٢١ – ١٤٠	٢٠	٦	٤	٣٠
١٤١ – ١٦٠	٢	٤٠	١٠	٥٢
١٦١ – ١٨٠	٢	٦	١٠	١٨
المجموع	٢٤	٥٢	٢٤	١٠٠

هذا وسنشرح الان بالتفصيل كيفية انشاء او تكوين جدول من الجداول البسيطة كثير الاستعمال يدعى بجدول التوزيع التكراري Frequence Table. :.

(٣ : ٢) جدول التوزيع التكراري

Frequency Distribution or Frequency Table

تعريف (٣ : ١)

جدول التوزيع التكرارى : هو جدول بسيط يتكون من عمودين :

الاول : وتقسم فيه قيم المتغير الى اقسام او مجموعات تدعى بالفئات (Classes)

والثاني : يبين مفردات كل فئة ويسمى بالتكرار Frequency

كما في جدول (٣ : ٤)

جدول (٣ : ٤) جدول توزيع تكراري لاطوال ٨٠ نباتا من القطن (بالسنتمترات)

فئات الطول	التكرار (عدد النباتات)
٣١ – ٤٠	١
٤١ – ٥٠	٢
٥١ – ٦٠	٥
٦١ – ٧٠	١٥
٧١ – ٨٠	٢٥
٨١ – ٩٠	٢٠
٩١ – ١٠٠	١٢
المجموع	٨٠

(٩) بعض التعاريف المهمة :

البيانات غير المبوبة : Ungrouped data

وهي البيانات الاولية او الاصلية (Raw data) التي جمعت ولم تبوب .

البيانات المبوبة : Grouped date

وهي البيانات التي بوبت ونظمت في جدول توزيع تكراري .

الفئات : Classes

وهي المجاميع التي قسمت اليها قيم المتغير . وكل فئة تأخذ مدى معين من قيم المتغير . فجدول (٣ : ٤) يحوي 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 limits

لكل فئة حدان حقيقيان حد أدنى حقيقي Lower class boundary وحد أعلى حقيقي

طول الفئة : Class length or class width

وهو مقدار المدى بين حدي الفئة هذا ويستحسن ان تكون اطوال الفئات متساوية

لتسهيل العمليات الحسابية . وسنرمز لطول الفئة بالرمز (c)

مركز الفئة : Class mark or class mid-point

لكل فئة مركز وسنرمز له بـ y_i 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 ان يكون دائما مساويا للعدد الكلي لقيم الظاهرة .

هذا وجدول (٣ : ٥) يوضح ما سبق شرحه بالتفصيل :

جدول (٣:٥) جدول توزيع تكراري لاطول نباتات القطن مبينا فيه الحدود الحقيقية ومراكز الفئات

تسلسل الفئات	الفئات Classes.	الحدود الحقيقية للفئات Class boundaries	مركز الفئة yi Class mark	التكرار fi Frequency
١	٣١–٤٠	٣٠,٥–٤٠.٥	٣٥.٥	١
٢	٤١–٥٠	٤٠,٥–٥٠.٥	٤٥.٥	٢
٣	٥١–٦٠	٥٠,٥–٦٠,٥	٥٥.٥	٥
٤	٦١–٧٠	٦٠,٥–٧٠.٥	٦٥.٥	١٥
٥	٧١–٨٠	٧٠.٥–٨٠,٥	٧٥,٥	٢٥
٦	٨١–٩٠	٨٠,٥–٩٠,٥	٨٥.٥	٢٠
٧	٩١–١٠٠	٩٠,٥–١٠٠,٥	٩٥.٥	١٢
	المجموع			٨٠

خذ مثلا الفئة الرابعة = (٦١–٧٠) :

فالحد الأدنى للفئة الرابعة = ٦١

والحد الأعلى للفئة الرابعة = ٧٠

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

الطريقة الاولى (عندما تكون حدود الفئات اعدادا صحيحة فقط)

أ) طول الفئة = الحد الأعلى–الحد الأدنى + ١

= ٧٠ – ٦١ + ١ = ١٠

الطريقة الثانية

ب) طول الفئة = الحد الحقيقي الأعلى – الحد الحقيقي الأدنى لتلك الفئة

= ٧٠,٥ –٦٠,٥= ١٠

الطريقة الثالثة:

ج) طول الفئة = الفرق بين الحدين الأدنى (أو الحدين الاعلى) لفئتين متتاليتين

الفرق بين الحدين الادنى = ٧١–٦١ = ١٠

الفرق بين الحدين الاعلى = ٨٠–٧٠ = ١٠

الطريقة الرابعة:

د) طول الفئة = الفرق بين الحدين الحقيقيين الأدنى (أو الاعلى) لفئتين متتاليتين .

= ٧٠,٥–٦٠,٥ = ١٠

= ٨٠,٥–٧٠,٥ = ١٠

الطريقة الخامسة :

هـ) طول الفئة = الفرق بين مركزي فئتين متتاليتين

= ٧٥,٥–٦٥,٥ = ١٠

الحدود الحقيقية يمكن حساب الحدود الحقيقية لأي فئة باحدى الطرق التالية :

الطريقة الاولى :

(١) الحد الأدنى الحقيقي لأي فئة = مركز تلك الفئة – $\frac{١}{٢}$ (طول تلك الفئة)

فالحد الادنى الحقيقي للفئة الرابعة = مركز الفئة الرابعة – $\frac{١}{٢}$ (طول الفئة الرابعة)

= ٦٥,٥ – $\frac{١}{٢}$ (١٠)

= ٦٠,٥

أما الحد الأعلى الحقيقي = مركز الفئة + $\frac{١}{٢}$ (طول الفئة)

فالحد الحقيقي للفئة الرابعة = ٦٥,٥ + $\frac{١}{٢}$ (١٠)

= ٧٠,٥

الطريقة الثانية :

الحد الادنى الحقيقي لأي فئة = $\frac{\text{الحد الادنى لتلك الفئة + الحد الأعلى للفئة السابقة}}{٢}$

فالحد الأدنى الحقيقي للفئة الرابعة = $\frac{٦١ + ٦٠}{٢}$ = ٦٠,٥

الحد الاعلى الحقيقي لأي فئة = $\frac{\text{الحد الأعلى لتلك الفئة + الحد الأدنى للفئة التي تليها}}{٢}$

فالحد الأعلى الحقيقي للفئة الرابعة = $\frac{٧٠ + ٧١}{٢}$ = ٧٠,٥

ملاحظة : اذا كانت حدود الفئات اعداد صحيحة فان :

الحد الأدنى الحقيقي لأي فئة = الحد الأدنى لتلك الفئة − ٠,٥

والحد الحقيقي لأي فئة = الحد الأعلى لتلك الفئة + ٠,٥

مركز الفئة : وتحسب باحدى الطريقتين التاليتين :

الطريقة الاولى:

مركز الفئة = $\frac{\text{الحد الأدنى + الحد الأعلى}}{٢}$

فمركز الفئة الرابعة = $\frac{٦١ + ٧٠}{٢}$ = ٦٥,٥

الطريقة الثانية :

مركز الفئة = $\frac{\text{الحد الأدنى الحقيقي + الحد الأعلى الحقيقي}}{٢}$

مركز الفئة الرابعة = $\frac{٧٠.٥ + ٦٠.٥}{٢}$ = ٦٥.٥

تكرار الفئة الرابعة = ١٥ أي أن هناك ١٥ قيمة من قيم المتغير واقعة في المدى (٦١–٧٠) .

(٢) الخطوات العامة في انشاء جداول التوزيع التكرارية

General Rules for Constructing Frequency Table

لتكوين إنشاء جدول توزيع تكراري يجب اتباع الخطوات التالية :

(آ) استخراج مدى المتغير Range

(ب) اختيار وتحديد عدد الفئات Number of classes

(ج) إيجاد طول مدى الفئة Class length or width

(د) كتابة حدود الفئات Class limits

(هـ) استخراج عدد التكرارات لكل فئة Class frequency

والمثال التالي يوضح كيفية انشاء جدول التوزيع التكراري لنباتات القطن .

مثال (١) القيم التالية تمثل أطوال ٨٠ نباتا من نباتات القطن (مقربة الى أقرب سنتمتر) والمطلوب انشاء جدول توزيع تكراري لاطوال هذه النباتات.

جدول (٣ : ٥) اطوال ٨٠ نباتا من نباتات القطن مقدرة بالسنتمترات

٨٠	٧٩	٤٨	٧٤	٨١	٩٨	٨٧	٨٠
٨٤	٨٠	٩٠	٧٠	٩١	٩٣	٨٢	٧٨
٧١	٧٠	٩٢	٦٥	٥٦	٨١	٧٤	٧٣
٧٢	٦٨	٨٥	٥١	٦٥	٩٣	٨٣	٨٦
٣٥	٩٠	٨٣	٧٣	٧٤	٤٣	٨٦	٦٨
٩٣	٩٢	٧٦	٧١	٩٠	٧٢	٦٧	٧٥
٩١	٨٠	٦١	٧٢	٩٧	٩١	٨٨	٨١
٧٤	٧٠	٩٩	٩٥	٨٠	٥٩	٧١	٧٧
٦٠	٦٣	٨٣	٨٢	٦٠	٦٧	٨٩	٦٣
٦٣	٧٦	٨٨	٧٠	٦٦	٨٨	٧٩	٧٥

الحـل : نتبع الخطوات التالية :

(أ) استخراج المدى (او مدى المتغير) The Range

المدى = (أعلى قيمة – أقل قيمة)

فأطول نبات = ٩٩ سم بينما أقصر نبات = ٣٥ سم .

لذا فالمدى = ٩٩ – ٣٥ = ٦٤ سم

(ب) اختيار وتحديد عدد الفئات Number of classes .

هناك عدة طرق حسابية تقريبية لإيجاد عدد الفئات أهمها :

طريقة سترجس Sturges

عدد الفئات = ١ + (٣.٣ × لوغارتم عدد المفردات)

وطريقة يول Yule

عد الفئات = ٢.٥ × $\sqrt[٤]{\text{عدد المفردات}}$

ولكل من الطريقتين ميزات وعيوب ولن نستعمل أيا منها هنا بل إننا سنختار عدد الفئات اختيارا على ان لا تقل عن خمسة ولاتزيد عن خمسة عشر فئة وذلك تبعا لطبيعة البيانات وعدد مفرداتها ومدى التغير فيها .

ولنفرض إننا اخترنا ٧ فئات .

(جـ) إيجاد طول الفئة : Class length

يجب أن لا يقل طول الفئة عـن : $\frac{\text{مدى التغيـــر}}{\text{عدد الفئات}}$ مقربة الى اقرب عدد صحيح اكبر

طول الفئة = $\frac{\text{مـدى التغير}}{\text{عدد الفئات}} = \frac{٦٤}{٧} = ٩ \frac{١}{٧}$

لذا يستحسن ان يكون طول الفئة = ١٠

وكما ذكرنا يستحسن ان تكون أطوال الفئات متساوية

(د) كتابة حدود الفئات Class limits

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

ويستحسن 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 بقليـل .

فمثلاً اصغر قيمة من قيم أطوال نباتات القطن هي ٣٥ لذا فمن الممكن أن يكون الرقم ٣١ يمثل الحد الأدنى للفئة الأولى . وبما أن طول الفئة هو ١٠ لذا فإن حدي الفئة الأولى هما ٣١–٤٠ والفئة الثانية تبدأ من ٤١–٥٠ بينما الفئة السابعة (الأخيرة) هي ٩١–١٠٠ . لاحظ بأن الحد الأدنى للفئة الأولى (٣١) والحد الأعلى للفئة الأخيرة (١٠٠) تحوي على كافة قيم المتغير .

● استخراج عدد التكرارات لكل فئة Class frequency

وتتم ذلك بتسجيل القيم الأصلية واحدة بعد الأخرى في الفئة الخاصة به على شكل شخطات أو علامات أولاً ثم ترجمتها الى أرقام كما مبين في جدول (٣ : ٦) أدناه :

جدول (٣ : ٦) جدول توزيع تكراري لاطوال نباتات القطن

الفئات	التكرار (بالعلامات)	التكرار رقما
٣١–٤٠	|	١
٤١–٥٠	||	٢
٥١–٦٠	~~||||~~	٥
٦١–٧٠	~~||||~~ ~~||||~~ ~~||||~~	١٥
٧١–٨٠	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	٢٥
٨١–٩٠	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~	٢٠
٩١–١٠٠	~~||||~~ ~~||||~~ ||	١٢
المجموع		٨٠

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

لاحظ انه في المثال السابق كانت اطوال الفئات متساوية وأرقام صحيحة . والان سنأخذ مثالاً آخراً فيه اطوال الفئات متساوية ولكنها ارقام ذات كسور .

مثـال (٢) : القيم التالية تمثل كمية المحصول (طن / هكتار) لحنطة المكسيباك في اربعين مزرعة مقدرة بالاطنان ومقربة الى أقرب رقم عشري واحد .

جدول (٣ : ٧) كمية المحصول (طن / هكتار) لحنطة المكسيباك في أربعين مزرعة

٢,٦	٢,٢	٤,١	٣,٥	٢,٠	٣,٢	٣,٧	٣,٠
٣,٧	٣,٤	١,٦	٣,١	٣,٣	٣,٨	٣,١	٢,٤
٣,١	٢,٥	٤,٣	٣,٤	٣,٦	٢,٩	٣,٣	٣,٩
٣,٤	٣,٣	٣,١	٣,٧	٤,٤	٣,٢	٤,١	١,٩
٣,٥	٢,٣	٣,٨	٣,٢	٢,٦	٣,٩	٣,٠	٤,٢

(أ) استخرج المدى :

المدى = أعلى قيمة – أقل قيمة

= ٤,٤ – ١,٦ = ٢,٨ طن

(ب) اختيار وتحديد عدد الفئات :

سنختار عدد الفئات هنا ٦ فئات

(ج) ايجاد طول الفئة :

طول الفئة = المدى / عدد الفئات

طول الفئة = ٢,٨ / ٦

= ٠,٤٦٧

لذايستحسن ان تكون طول الفئة ٠,٥

(د) كتابة حدود الفئات :

بما أن أقل قيمة للمتغير = ١,٦ لذا فسنبدأ بكتابة الحد الأدنى للفئة الأولى ١,٥ . وبما أن طول الفئة ٠,٥ لذا فالفئة الأولى ستكون (١,٥ – ١,٩) والثانية (٢,٠ – ٢,٤) وهكذا الى أن تصل الفئة الأخيرة وهي (٤,٠ – ٤,٤) .

(هـ) استخراج عدد التكرارات لكل فئة : نسجل عدد المشاهدات أو المفردات التابعة لكل فئة .

ويجب التأكد بأن مجموع التكرارات الكلي مساوية للعدد الكلي لقيم المتغير وجدول

(٣:٨) يبين التوزيع التكراري لكمية المحصول لحنطة المكسيباك اضافة الى الحدود الحقيقية ومراكز الفئات .

جدول (٣:٨) جدول التوزيع التكراري لكمية المحصول لحنطة المكسيباك

تسلسل الفئات	حدود الفئات	الحدود الحقيقية للفئات	مركز الفئة	التكرار
١	١,٩–١,٥	١,٩٥–١,٤٥	١,٧	٢
٢	٢,٤–٢,٠	٢,٤٥–١,٩٥	٢,٢	٤
٣	٢,٩–٢,٥	٢,٩٥–٢,٤٥	٢,٧	٤
٤	٣,٤–٣,٠	٣,٤٥–٢,٩٥	٣,٢	١٥
٥	٣,٩–٣,٥	٣,٩٥–٣,٤٥	٣,٧	١٠
٦	٤,٤–٤,٠	٤,٤٥–٣,٩٥	٤,٢	٥
	المجموع			٤٠

ملاحظة : اذا كانت أعداد قيم المتغير قليلة (أي اذا 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 لتوضيح وتبسيط كيفية أنشاء جدول التوزيع التكراري بأستخدام أرقام بسيطة وقليلة .

(٣:٣) جدول التوزيع التكراري النسبي Relative Frequency Distribution

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

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

$$\boxed{= \frac{f_i}{\sum f_i}}$$

ومن جدول (٣ : ٤) فإن :

$$\text{التكرار النسبي للفئة الرابعة} = \frac{\text{تكرار الفئة الرابعة}}{\text{المجموع الكلي للتكرارات}}$$

$$= \frac{١٥}{٨٠} = ٠.١٨٧٥$$

وعادة يوضع التكرار النسبي كنسبة مئوية وذلك بضرب كل تكرار نسبي × ١٠٠ كما مبين في جدول (٣ : ٩)

جدول (٣ : ٩) جدول التوزيع التكراري النسبي والمئوي لاطوال نباتات القطن

الفئات	التكرار	التكرار النسبي	التكرار المئوي
٣١–٤٠	١	٠.٠١٢٥	١.٢٥
٤١–٥٠	٢	٠.٠٢٥٠	٢.٥٠
٥١–٦٠	٥	٠.٠٦٢٥	٦.٢٥
٦١–٧٠	١٥	٠.١٨٧٥	١٨.٧٥
٧١–٨٠	٢٥	٠.٣١٢٥	٣١.٢٥
٨١–٩٠	٢٠	٠.٢٥٠٠	٢٥.٠٠
٩١–١٠٠	١٢	٠.١٥٠٠	١٥.٠٠
المجموع	٨٠	١.٠٠٠٠	١٠٠.٠٠

(٣ : ٤) التوزيعات المتجمعة Cumulative Distribution

ان جدول التوزيع التكراري العادي الذي سبق شرحه يبين توزيع قيم المتغير على الفئات المختلفة . ولكن في بعض الأحيان قد يكون هناك حاجة الى 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

(١) 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 F_i وجدول التوزيع التكراري المتجمع التصاعدي يتكون من عمودين :

العمود الأول : نكتب فيه حدود الفئات كما موضح في جدول (٣ : ١٠)

العمود الثاني : نكتب فيه التكرار التجميعي التصاعدي بالشكل التالي :

تكرار ماقبل الفئة الأولى = F_0 = صفر

تكرار الفئة الأولى = $f_1 = F_1$

تكرار الفئة الثانية = $f_1 + f_2 = F_2$

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

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

جدول (٣ : ١٠) التوزيع التكراري التجميعي التصاعدي لاطوال نباتات القطن

حدود الفئات	التكرار التجميعي التصاعدي
أقل من ٣١	٠
أقل من ٤١	١
أقل من ٥١	٣
أقل من ٦١	٨
أقل من ٧١	٢٣
أقل من ٨١	٤٨
أقل من ٩١	٦٨
أقل من ١٠١	٨٠

(٢) 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 ايضاً يتألف من عمودين :

العمود الأول : تكتب فيه حدود الفئات

العمود الثاني : تكتب فيه التكرارات التجميعية التنازلية بالطريقة التالية :

تكرار الفئة الاول : $= F_1 = \sum f_i$

تكرار الفئة الثانية $= F_2 =$ مجموع التكرارات - تكرار الفئة الاولى

أي :

$$F_2 = \sum f_i - f_1 = F_1 - f_1$$

تكرار الفئة الثالثة $= F_3$

$$F_3 = \sum f_i - f_1 - f_2$$

$$= F_2 - f_2$$

وهكذا كما مبين في جدول (٣:١١)

جدول (٣:١١) التوزيع التكراري التجميعي التنازلي لاطوال نباتات القطن

حدود الفئات	التكرار التجميعي التنازلي
٣١ فأكثر	٨٠
٤١ فأكثر	٧٩
٥١ فأكثر	٧٧
٦١ فأكثر	٧٢
٧١ فأكثر	٥٧
٨١ فأكثر	٣٢
٩١ فأكثر	١٢
١٠١ فأكثر	٠

٨٠ - ١ = ٧٩

هذا واحياناًيعبر عن التكرار التجميعي التصاعدي أو التنازلي بشكل تكرار تجميعي نسبي اومئوي . وفي هذه الحالة فان التكرار التجميعي النسبي لأي فئة $= \frac{\text{التكرار التجميعي لتلك الفئة}}{\text{المجموع الكلي}} = \frac{F_i}{\sum f_i}$

أما التكرار التجميعي المئوي = التكرار التجميعي النسبي × ١٠٠

(٣:٥) أمثلة محلولة

مثال (٢) الجدول التالي يبين التوزيع التكراري للرواتب الشهرية مقدرة بالدينار(٦٥) موظفا في احدى الشركات :

فئات الأجور	التكرار (عدد المستخدمين)
٥٠ – ٥٩	٨
٦٠ – ٦٩	١٠
٧٠ – ٧٩	١٦
٨٠ – ٨٩	١٤
٩٠ – ٩٩	١٠
١٠٠ – ١٠٩	٥
١١٠ – ١١٩	٢
المجمـــوع	٦٥

والمطلوب إيجاد قيمة كل مما يلي :

(أ) الحد الأدنى للفئة السادسة ؟

الحـل = ١٠٠

(ب) الحد الاعلى للفئة الرابعة

الحـل : ٨٩

(جـ) مركز الفئة الخامسة

الحـل : $\frac{٩٩ + ٩٠}{٢} = ٩٤,٥$

(د) طول الفئة الخامسة

الحـل : طول الفئة الخامسة = الحد الأعلى للفئة الخامسة – الحد الأدنى للفئة الخامسة+١

=٩٩–٩٠+١=١٠

(هـ) الحد الادنى الحقيقي للفئة الخامسة ؟

الحـل : الحد الأدنى الحقيقي = مركز الفئة الخامسة − $\frac{١}{٢}$ (طول الفئة الخامسة

$$= ٩٤,٥ - \frac{١}{٢}(١٠) = ٨٩.٥$$

$$\text{أو} = \frac{\text{الحد الادنى للفئة الخامسة + الحد الأعلى للفئة الرابعة}}{٢}$$

$$= \frac{٨٩ + ٩٠}{٢} = ٨٩,٥$$

(و) تكرار الفئة الثالثة

الحـل : ١٦

(ز) التكرار النسبي للفئة الثالثة

الحل : $\frac{١٦}{٦٥} = ٠,٢٤٦$

مثال (٣) أكمل جدول التوزيع التكراري التالي :

الفئات	مركز الفئات	الحدود الحقيقية	التكرار	التكرار النسبي	التكرار المئوي
	٤	١,٥ - ٦,٥	٢		
	٩		٥		
	١٤		١٠		
	١٩		٢٥		
	٢٤		٨		
المجموع			٥٠		

الحـــل : طول الفئة = الفرق بين مركزي فئتين متتاليتين

= ٩ − ٤ = ٥

الحد الأدنى الحقيقي للفئة الأولى = مركز الفئة الأولى − $\frac{١}{٢}$ (طول الفئة)

= ٤ − $\frac{١}{٢}$ (٥)

= ١,٥

الحد الأعلى الحقيقي للفئة الأولى = مركز الفئة الأولى + $\frac{١}{٢}$ (طول الفئة)

= ٤ + $\frac{١}{٢}$ (٥)

= ٦,٥

ثم تضاف طول الفئة على الحد الأدنى الحقيقي للفئة الأولى لينتج الحد الأدنى الحقيقي للفئة الثانية وهكذا

ثم تضاف طول الفئة على الحد الأعلى الحقيقي للفئة الأولى لينتج الحد الأعلى الحقيقي للفئة الثانية وهكذا .

أما الحد الأدنى للفئة الأولى فهو أقرب عدد صحيح للحد الأدنى الحقيقي وهو٢ أي باضافة نصف الى الحد الأدنى الحقيقي بينما الحد الأعلى فهوبطرح نصف من الحد الأعلى الحقيقي . لذا فحدي الفئة الأولى هما (٢−٦) ثم تضاف طول الفئة بعدئذٍ لكل من الحد الأدنى والحد الأعلى لأيجاد حدود الفئات الأخرى .

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

فمثلاً التكرار النسبي للفئة الأولى = $\frac{٢}{٥٠}$ = ٠,٠٤

أما التكرار المئوي = التكرار النسبي × ١٠٠

كما مبين ذلك في الجدول أدناه

الفئات	مركز الفئات	الحدود الحقيقية	التكرار	التكرار النسبي	التكرار المئوي
٢ – ٦	٤	١,٥ – ٦,٥	٢	٠,٠٤	٤
٧ – ١١	٩	٦,٥ – ١١,٥	٥	٠,١٠	١٠
١٢ – ١٦	١٤	١١,٥ – ١٦,٥	١٠	٠,٢٠	٢٠
١٧ – ٢١	١٩	١٦,٥ – ٢١,٥	٢٥	٠,٥٠	٥٠
٢٢ – ٢٦	٢٤	٢١,٥ – ٢٦,٥	٨	٠,١٦	١٦
			٥٠	١,٠٠	

مثال (٤) نفرض أن عدد مفردات ظاهرة ما هو ١٥٠ مفردة وان أقل قيمة بينها = ٥,١٨
وأعلى قيمة = ٧,٤٤
فالمطلوب أيجاد:
(أ) حدود الفئات
(ب) مراكز الفئات
(جـ) الحدود الحقيقية للفئات

التي قد تستعمل في انشاء جدول توزيع تكراري لهذه القيم .

الحل :

(أ) المدى = ٧,٤٤ – ٥,١٨ = ٢,٢٦
لنفرض ان عدد الفئات المناسبة المختارة = ٨

$$\text{طول الفئة} = \frac{٢,٢٦}{٨} = ٠,٢٨$$

اذن طول الفئة سنعتبرها = ٠,٣
وبما أن أقل قيمة = ٥,١٨
نبدأ بالحد الأدنى للفئة الاولى بـ ٥,١٠
ثم نضيف طول الفئة للحد الأدنى للفئة الاولى لايجاد الحد الأدنى للفئة الثانية أي
٥,١٠ + ٠,٣٠ = ٥,٤٠
أما الحدود العليا ، فبما ان قيم الظاهرة مقرب الى رقمين عشريين ، لذا فان

لذلك يجب أخذ مرتبتين عشريتين

الحد الاعلى للفئة الاولى = الحد الأدنى للفئة الثانية − ٠,٠١

أي الحد الأعلى للفئة الاولى = ٥,٤٠ − ٠,٠١ = ٥,٣٩

ثم نضيف طول الفئة للحد الاعلى للفئة الاولى لايجاد الحد الأعلى للفئة الثانية وهكذا ..

(ب) ثم نستخرج مركز الفئة = $\frac{\text{الحد الأدنى + الحد الأعلى}}{٢}$

فمركز الفئة الاولى = $\frac{٥,٣٩ + ٥,١٠}{٢}$

= ٥,٢٤٥

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

(ج) أما الحدود الحقيقية فتستخرج بالطريقة التالية :

الحد الأدنى الحقيقي = مركز الفئة − $\frac{١}{٢}$ (طول الفئة)

فمثلا الحد الادنى الحقيقي للفئة الاولى = ٥,٢٤٥ − $\frac{١}{٢}$ (٠,٣٠)

= ٥,٠٩٥

والحد الاعلى الحقيقي = مركز الفئة + $\frac{١}{٢}$ (طول الفئة)

فالحد الاعلى الحقيقي للفئة الاولى = ٥,٢٤٥ + $\frac{١}{٢}$ (٠,٣)

= ٥,٣٩٥

ثم إضافة طول الفئة للحد الادنى الحقيقي للفئة الاولى لايجاد الحد الادنى الحقيقي للفئة الثانية وهكذا بالنسبة للحدود العليا الحقيقية ايضا كما مبين في الجدول ادناه

مراكز الفئات	الحدود الحقيقية	حدود الفئات
٥,٢٤٥	٥,٣٩٥ – ٥,٠٩٥	٥,٣٩ – ٥,١٠
٥,٥٤٥	٥,٦٩٥ – ٥,٣٩٥	٥,٦٩ – ٥,٤٠
٥,٨٤٥	٥,٩٩٥ – ٥,٦٩٥	٥,٩٩ – ٥,٧٠
٦,١٤٥	٥,٢٩٥ – ٥,٩٩٥	٦,٢٩ – ٦,٠٠
٦,٤٤٥	٥,٥٩٥ – ٦,٢٩٥	٦,٥٩ – ٦,٣٠
٦,٧٤٥	٦,٨٩٥ – ٦,٥٩٥	٦,٨٩ – ٦,٦٠
٧,٠٤٥	٧,١٩٥ – ٦,٨٩٥	٧,١٩ – ٦,٩٠
٧,٣٤٥	٧,٤٩٥ – ٧,١٩٥	٧,٤٩ – ٧,٢٠

مثال (٥) إذا علمت بأن عدد مفردات المتغير = ٥٠ (أي $\sum f_i = 50$)
فمن جدول التوزيع التكراري النسبي التالي :

التكرار النسبي	الفئات
٠,١٢	٣٩ – ٢٠
٠,٢٨	٥٩ – ٤٠
٠,٣٦	٧٩ – ٦٠
٠,٢٠	٩٩ – ٨٠
٠,٠٤	١١٩ – ١٠٠

اوجد التكرارات ومراكز الفئات والحدود الحقيقية والتكرار المئوي لهذا الجدول .

الحل :

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

تكرار الفئة = التكرار النسبي × التكرار الكلي

تكرار الفئة الاولى = ٠,١٢ × ٥٠ = ٦

تكرار الفئة الثانية = ٠,٢٨ × ٥٠ = ١٤

وهكذا

$$\text{أما مركز الفئة} = \frac{\text{الحد الأدنى للفئة + الحد الأعلى للفئة}}{٢}$$

$$\text{مركز الفئة الأولى} = \frac{٢٠ + ٣٩}{٢} = ٢٩,٥$$

$$\text{مركز الفئة الثانية} = \frac{٤٠ + ٥٩}{٢} = ٤٩,٥$$

وهكذا

أما طول الفئة = الحد الأعلى للفئة − الحد الأدنى للفئة + ١

أو = الفرق بين مركزي فئتين متتاليتين

طول الفئة = ٣٩ − ٢٠ + ١ = ٢٠

$$\text{أما الحد الأدنى الحقيقي} = \text{مركز الفئة} - \frac{١}{٢}\ (\text{طول الفئة})$$

$$\text{فالحد الأدنى الحقيقي للفئة الأولى} = ٢٩,٥ - \frac{١}{٢}(٢٠) = ١٩,٥$$

$$\text{بينما الحد الأعلى الحقيقي للفئة الأولى} = ٢٩,٥ + \frac{١}{٢}(٢٠) = ٣٩,٥$$

وهكذا

أما التكرار المئوي = التكرار النسبي × ١٠٠

فالتكرار المئوي للفئة الأولى = ٠,١٢ × ١٠٠ = ١٢

وهكذا كما مبين في الجدول أدناه :

الفئات	التكرار	مركز الفئات	الحدود الحقيقية	التكرار النسبي	التكرار المئوي
٢٠ – ٣٩	٦	٢٩,٥	١٩,٥ – ٣٩,٥	٠,١٢	١٢
٤٠ – ٥٩	١٤	٤٩,٥	٣٩,٥ – ٥٩,٥	٠,٢٨	٢٨
٦٠ – ٧٩	١٨	٦٩,٥	٥٩,٥ – ٧٩,٥	٠,٣٦	٣٦
٨٠ – ٩٩	١٠	٨٩,٥	٧٩,٥ – ٩٩,٥	٠,٢٠	٣٠
١٠٠ – ١١٩	٢	١٠٩,٥	٩٩,٥ – ١١٩,٥	٠,٠٤	٤
	٥٠			١,٠٠	١٠٠

مثال (٦) الجدول التالي يبين التوزيع التكراري لأوزان (٦٥ طالباً بالكيلوغرامات)

فئات الوزن	التكرار (عـدد الطلبة)
٥٠ – ٥٤	٨
٥٥ – ٥٩	١٠
٦٠ – ٦٤	١٦
٦٥ – ٦٩	١٤
٧٠ – ٧٤	١٠
٧٥ – ٧٩	٥
٨٠ – ٨٤	٢
	٦٥

والمطلوب عمل جدول توزيع تكراري تجميعي تصاعدي وتنازلي ومنهما استنتج مايلي :

(أ) ماهو عدد الطلبة الذي اوزانهم تقل عن ٧٠ كغم ؟

(ب) ماهي نسبة الطلبة الذي اوزانهم تقل ٧٠ كغم ؟

(جـ) ماهو عدد الطلبة الذي أوزانهم لايقل عن ٦٠ كغم ؟
(د) ماهو عدد الطلبة الذي اوزانهم لاتقل عن ٦٠ كغم ولكنها أقل من ٨٠ كغـم ؟

الحـل :

جدول توزيع تكراري تجميعي تصاعدي

حدود الفئات	التكرار التجميعي التصاعدي
أقل من ٥٠	٠
أقل من ٥٥	٨
أقل من ٦٠	١٨
أقل من ٦٥	٣٤
أقل من ٧٠	٤٨
أقل من ٧٥	٥٨
أقل من ٨٠	٦٣
أقل من ٨٥	٦٥

جدول توزيع تكراري تجميعي تنازلي

حدود الفئات	التكرار التجميعي التنازلي
٥٠ فأكثر	٦٥
٥٥ فأكثر	٥٧
٦٠ فأكثر	٤٧
٦٥ فأكثر	٣١
٧٠ فأكثر	١٧
٧٥ فأكثر	٧
٨٠ فأكثر	٢
٨٥ فأكثر	٠

(أ) من جدول التوزيع التكراري التجميعي التصاعدي
عدد الطلبة الذين اوزانهم أقل من ٧٠ كغم = ٤٨

(ب) أما نسبة هؤلاء الطلبة = $\frac{٤٨}{٦٥} \times ١٠٠ = ٧٣{,}٨$

(جـ) من جدول التوزيع التكراري التجميعي التنازلي
عدد الطلبة الذين أوزانهم لا تقل عن ٦٠ كغم = ٤٧

(د) من جدول التوزيع التكراري التجميعي التنازلي
عدد الطلبة الذين أوزانهم لاتقل عن ٦٠ ولكنها أقل من ٨٠ كغم
٤٧ − ٢ = ٤٥

(٣ : ٦) التمثيل البياني Graphical Presentation

ان الرسوم والصور والأشكال الهندسية ما هي الا تعبير وتوضيح للبيانات بطريقة جذابة وسهلة وفعالة تساعد القارىء على فهم واستيعاب قيم الظاهرة ومقارنتها مع بعضها .

ووسائل التمثيل البياني كثيرة ومتنوعة وسنكتفي هنا بشرح العرض البياني للتوزيعات التكرارية فقط . وعادة نخصص المحور الافقي (abscisa) او الاحداثي السيني لتمثل قيم أو فئات المتغير بينما نخصص المحور العمودي (ordinate) أو الاحداثي التصاعدي لتمثل تكرارات هذا المتغير ويجب دائما ان يبدأ تدريج المحور العمودي من الصفر أما تدريج المحور الأفقي فقد لا نبدأ بتدريجه من الصفر . كما انه ليس من الضروري ان يكون مقياس او تدريج المحورين من نفس المقياس .

(١) التمثيل البياني لجدول التوزيع التكراري

(أ) المدرج التكراري Histogram

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

ولرسم مدرج تكراري نتبع الخطوات التالية :

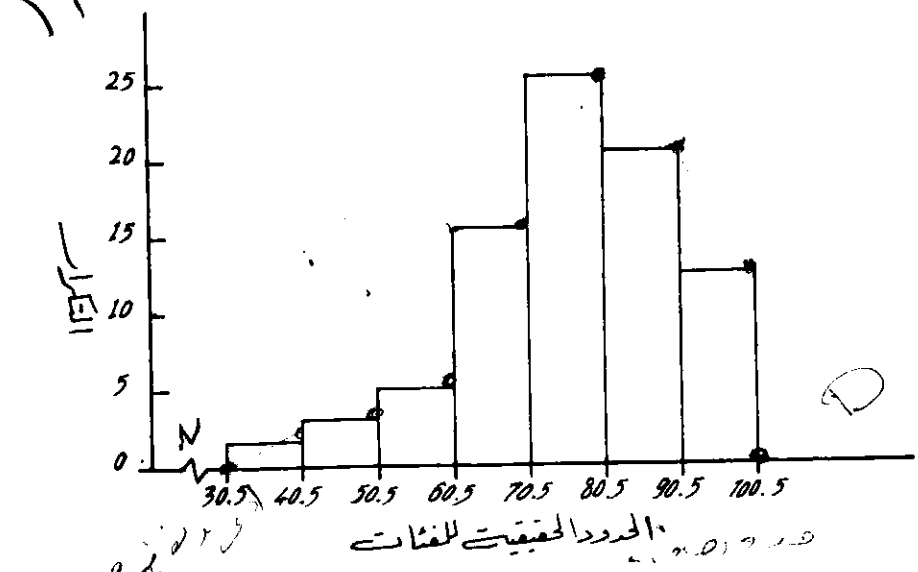
١ - رسم المحور الافقي والمحور العمودي .

٢ - تدريج المحور الافقي الى أقسام متساوية بمقياس رسم مناسب بحيث يشمل جميع الحدود الحقيقية للفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الأدنى للفئة الاولى (فيما اذا كانت بداية الفئة الاولى لا تساوي صفر) . ويقسم المحور العمودي الى أقسام متساوية بحيث تشمل على أكبر التكرارات.

٣ - يرسم على كل فئة مستطيلا رأسيا تمثل قاعدته طول تلك الفئة وارتفاعه تمثل تكرار تلك الفئة .

والشكل (٣ : ١) يمثل المدرج التكراري لجدول (٣ : ٦) .

والشكل (٣ : ١) يمثل المدرج التكراري لجدول (٣ : ٦)



شكل (٣ : ١) المدرج التكراري لاطوال نباتات القطن

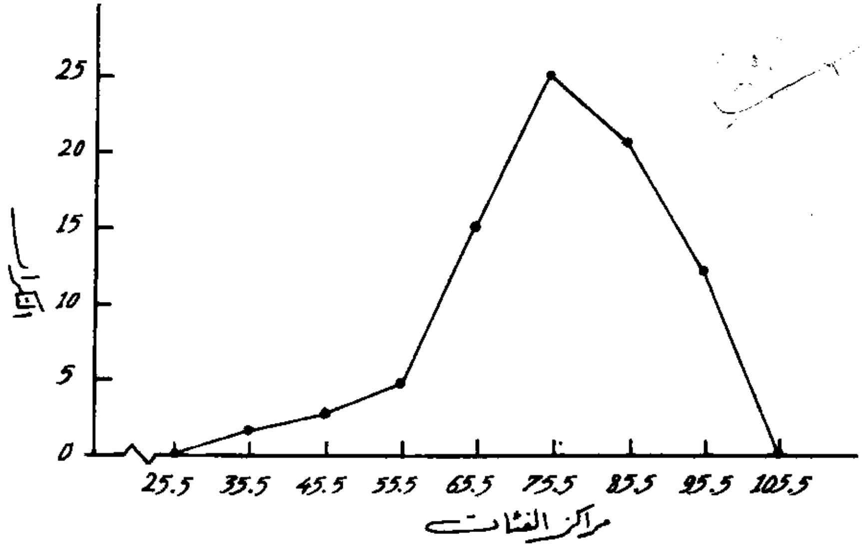
(ب) المضلع التكراري Frequency Polygon

وهو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط كل منها واقعة فوق مركز فئة على ارتفاع يمثل تكرار تلك الفئة . وعادة يقفل المضلع بأن نصل بداية المضلع بالمحور الافقي بمركز فئة (خيالية) واقعة الى يسار أول فئة تكرارها صفراً . ونصل نهاية المضلع بالمحور الافقي بمركز فئة (خيالية) واقعة الى يمين آخر فئة تكرارها أيضاً صفراً وبذلك تكون مساحة المضلع التكراري مساوية لمساحة المدرج التكراري .

ولرسم المضلع التكراري نتبع الخطوات التالية :

١- رسم المحور الافقي والمحور العمودي.
٢- تدريج المحور الافقي الى أقسام متساوية بحيث يشمل على جميع مراكز الفئات . ويقسم المحور العمودي الى أقسام متساوية بحيث تشمل على أكبر التكرارات .
٣- وضع نقطة أمام مركز كل فئة ارتفاعها يعادل تكرار تلك الفئة .
٤- توصيل تلك النقاط بخطوط مستقيمة .

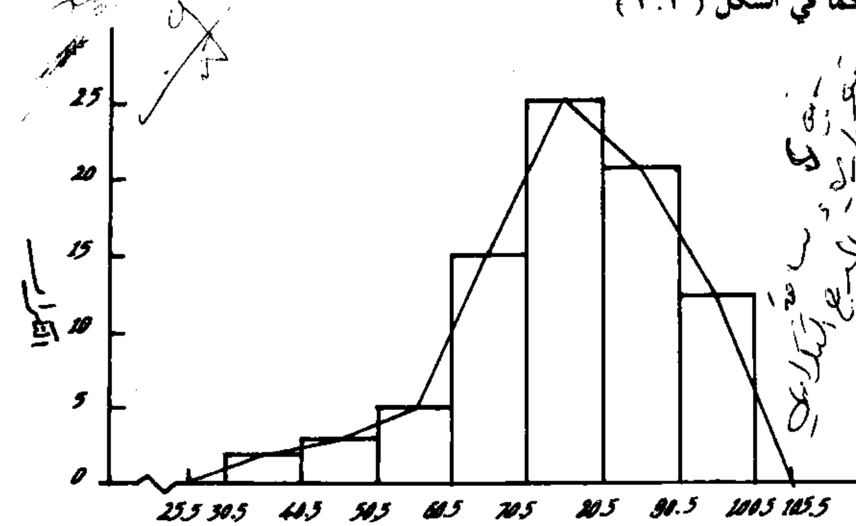
والشكل (٣ : ٢) يمثل المضلع التكراري لجدول (٣ : ٦).



شكل (٣ : ٢) المضلع التكراري لاطوال نباتات القطن

كيف يمكن ذلك

هذا ويمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك بعد تنصيف القواعد العليا للمستطيلات (والتي تمثل مراكز الفئات) بنقاط ثم توصيل هذه النقاط بمستقيمات كما في الشكل (٣ : ٣)

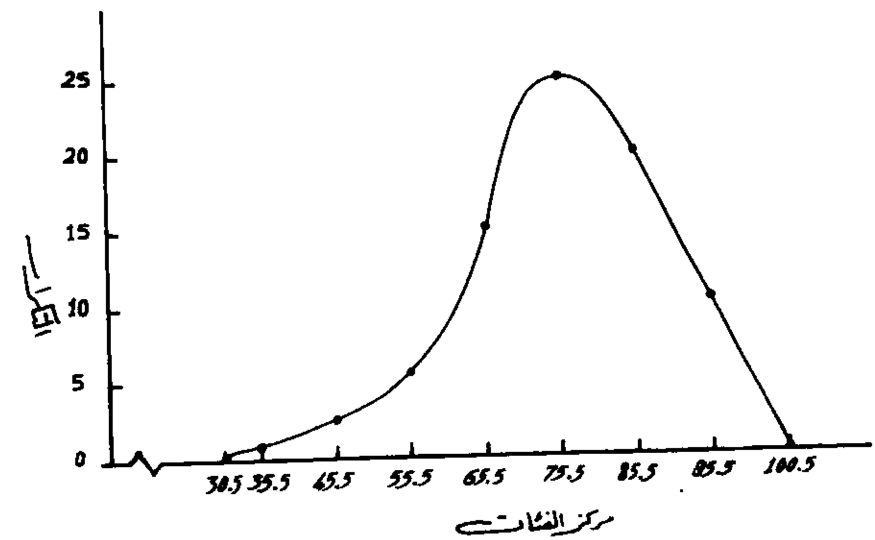


شكل (٣ : ٣) المدرج التكراري والمضلع التكراري لاطوال نباتات القطن .

(جـ) المنحني التكراري Frequency Curve

وهو عبارة عن منحني يمر بمعظم النقاط الواقعة على مراكز الفئات والتي ارتفاعها يمثل تكرارات تلك الفئات .

وعادة يقفل المنحني التكراري بأن نصل بدايته بالحد الأدنى للفئة الأولى ونهايته بالحد الأعلى للفئة الأخيرة . وتكون مساحة المنحني مكافئة (وليست مساوية) للمضلع التكراري . كما في شكل (٣ : ٤).



شكل (٣:٤) المنحني التكراري لأطوال نباتات القطن

ملاحظة : عند مقارنة مجموعتين من البيانات غيرمتساويتين في عدد مفرداتها باستخدام المضلع التكراري لهما فيجب استخدام 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.

(٢) التمثيل البياني لجدول التوزيع التكراري التجميعي :

لتمثيل التكرار التجميعي بيانياً نستخدم المضلع التكراري التجميعي . Cumulative frquency polygon or ogive وهو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط واقعة فوق الحدود الحقيقية للفئات وعلى ارتفاع تمثل التكرار

التجميعي . وهناك نوعان من المضلع التكراري التجميعي :

(أ) المضلع التكراري التجميعي التصاعدي Or less Ogive.

ولرسم المضلع التكراري التجميعي التصاعدي نتبع الخطوات التالية :

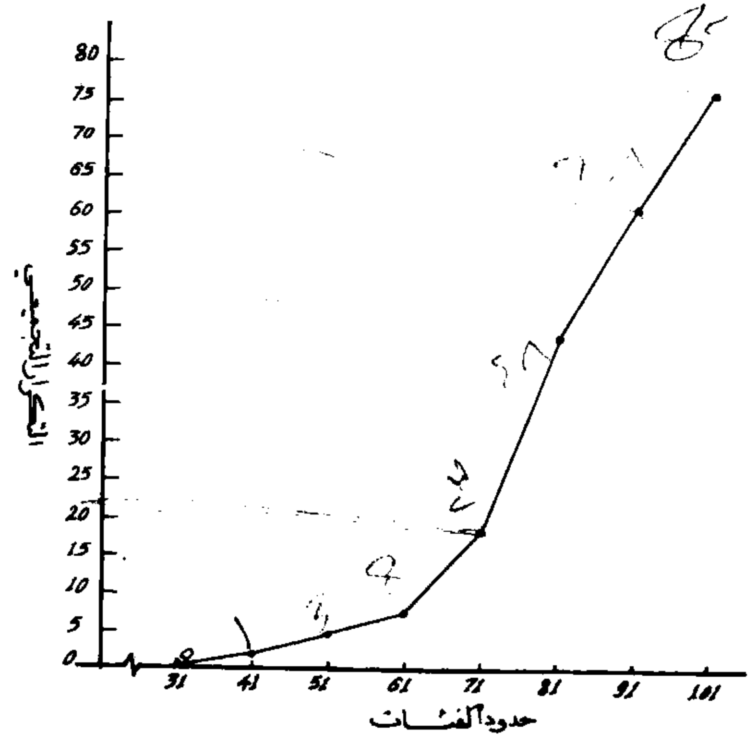
١ . رسم المحور الأفقي والمحور العمودي .

٢ . تدريج المحور الأفقي الى أقسام متساوية تشمل على جميع حدود الفئات . ويقسم المحور العمودي الى أقسام متساوية بحيث تشمل على أكبر التكرارات التجميعية وهي المجموع الكلي للتكرارات .

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

٤ . توصيل تلك النقاط بخطوط مستقيمة .

والشكل (٣ : ٥) يمثل المضلع التكراري التجميعي التصاعدي لجدول (٣ : ١٠)

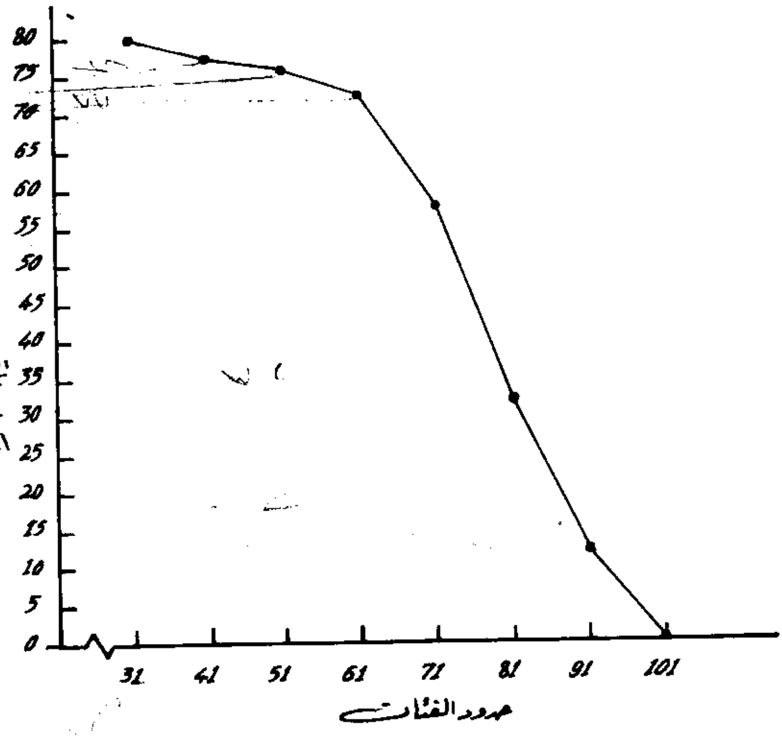


شكل (٣ : ٥) المضلع التكراري التجميعي المتصاعدي لاطوال نباتات القطن

(ب) 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 (٣ : ٦) يمثل المضلع التكراري التجميعي التنازلي لجدول (٣ : ١١)



شكل (٣ : ٦) المضلع التكراري التجميعي التنازلي لاطوال 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 frequenty ogive وذلك باتباع نفس الأساليب السابقة .

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

وكذلك يمكن رسم مايسمى بالمنحني التكراري التجميعي وذلك برسم منحنٍ يمر بمعظم النقاط (بدلاً من الخطوط المستقيمة المتكسرة) .

كما يمكن الاستفادة من المنحني التكراري التجميعي التصاعدي أو التنازلي بايجاد تقديرات معينة وذلك برسم أعمدة من المحور الأفقي لتقطع المنحني في نقاط ثم نقرأ مايقابل هذه النقاط على المحور العمودي بالأضافة الى استخدامهما في حساب بعض القيم الحسابية التي سيأتي ذكرها في الفصول القادمة .

مثال (٧) الجدول التالي يبين التوزيع التكراري للرواتب الشهرية لموظفي احدى الشركات

الفئات	التكـــــرار
٥٠ – ٥٩	٨
٦٠ – ٦٩	١٠
٧٠ – ٧٩	١٦
٨٠ – ٨٩	١٤
٩٠ – ٩٩	١٠
١٠٠ – ١٠٩	٥
١١٠ – ١١٩	٢
المجموع	٦٥

أوجد كلاً مما يأتي :

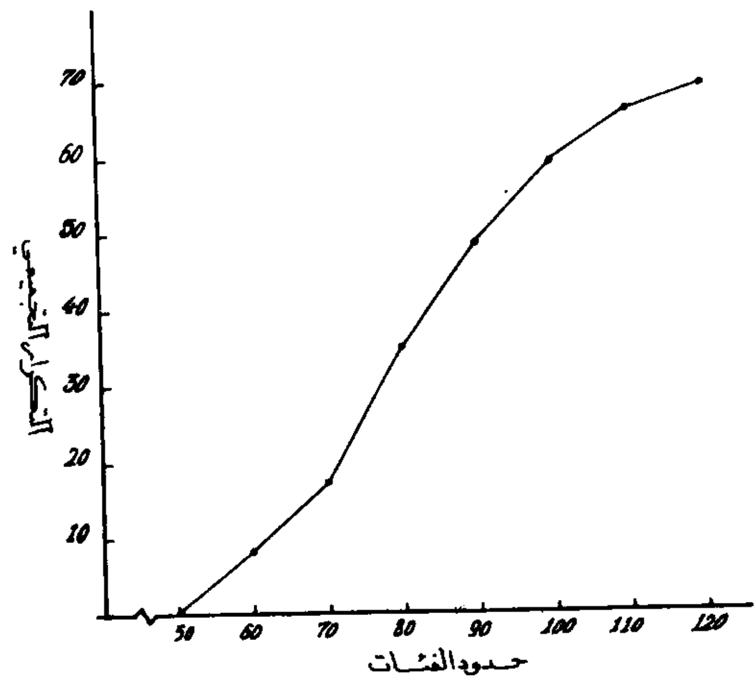
(١) ارسم المضلع التكراري التجميعي التصاعدي

الحـل :

نجد اولاً جدول التوزيع التكراري التجميعي التصاعدي كما يلي :

التكرار التجميعي التصاعدي	حدود الفئات
٠	أقل من ٥٠
٨	أقل من ٦٠
١٨	أقل من ٧٠
٣٤	أقل من ٨٠
٤٨	أقل من ٩٠
٥٨	أقل من ١٠٠
٦٣	أقل من ١١٠
٦٥	أقل من ١٢٠

ومن الجدول اعلاه نرسم المضلع التكراري التجميعي التصاعدي التالي :



شكل (٣:٧) المضلع التكراري التجميعي التصاعدي للرواتب الشهرية لموظفي احدى الشركات .

(٢) ماهو عدد المستخدمين الذين رواتبهم :

(أ) أقل من ٨٨ ديناراً :

الحــل : من رسم المضلع التكراري التجميعي التصاعدي .

ارسم من نقطة ٨٨ على المحور الأفقي عموداً ليقطع المضلع في نقطة يقابلها على المحور العمودي الرقم ٤٥ .

لذا فإن ٤٥ = عدد المستخدمين الذين رواتبهم أقل من ٨٨ ديناراً .

(ب)٩٦ ديناراً فأكثر :

الحــل : يمكن استعمال المضلع التكراري التجميعي التصاعدي أو التنازلي.

من المضلع التصاعدي مثلاً : نرسم من نقطة ٩٦ على المحور الأفقي عموداً ليقطع المضلع في نقطة يقابلها على المحور العمودي الرقم ٥٤ . وهي تعني أن ٥٤ مستخدماً رواتبهم أقل من ٩٦ ديناراً . وبما أن المجموع الكلي للمستخدمين هو٦٥ ، لذا فإن ٦٥−٥٤ = ١١ هو عدد المستخدمين الذين رواتبهم ٩٦ ديناراً فأكثر .

(جـ) على الأقل ٦٣ ديناراً ولكن أقل من ٧٥ ديناراً

الحــل :

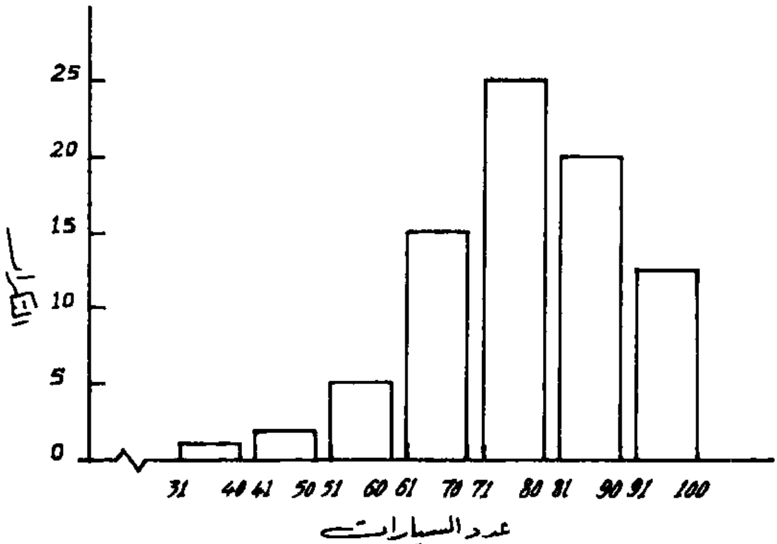
العدد المطلوب = عدد المستخدمين الذين رواتبهم أقل من٧٥ـ عدد المستخدمين الذين رواتبهم أقل من ٦٣ = ٢٦−١١=١٥

ملاحظـــة : ان جميع الأمثلة التي ذكرت سابقاً هي لمتغيرات مستمــرة . (Continuous variables) ففي حالة استخدام قيم مقربة الى أقرب عدد صحيح فقد تم معاملتها على أنها بيانات لمتغير مستمر باستعمال الحدود الحقيقية للفئات فكانت الرسوم البيانية كلها متصلة .

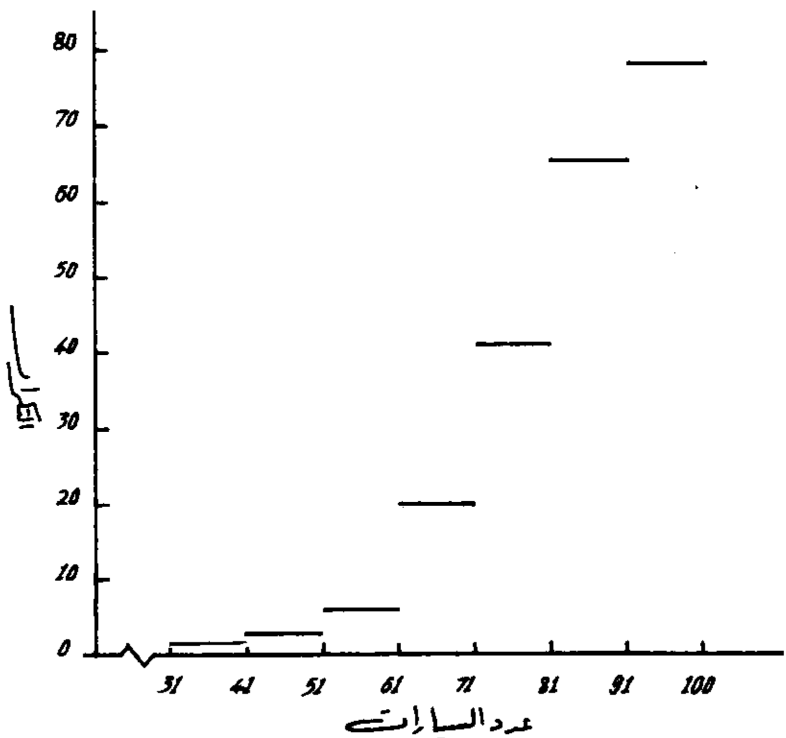
أما في حالة استخدام قيم لمتغير متقطع (Discrete variable) فإن الرسوم البيانية في هذه الحالة تكون متقطعة .

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

أما في حالة المضلع التكراري التجميعي التصاعدي (مثلاً) فإنه يكون على شكل مضلع متدرج متقطع . ولتوضيح ذلك نفرض بأن البيانات في جدول (٣:٦) هي تمثل قيم متغير متقطع (عدد السيارات في مؤسسة ما في العراق) فالمدرج التكراري والمضلع التكراري التجميعي التصاعدي سيكونان كالآتي :



شكل (٣،٨) المدرج التكراري لعدد السيارات



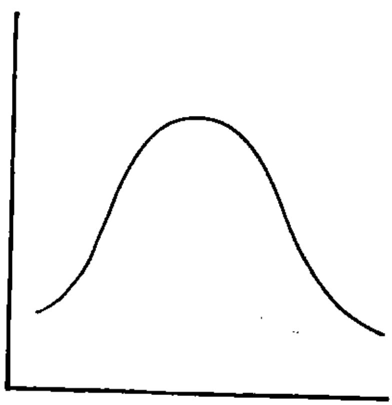
شكل (٣،٩) المضلع التكراري التجميعي الصاعدي لعدد السيارات

(٣ : ٧) أنواع المنحنيات التكرارية Types of frequency curves

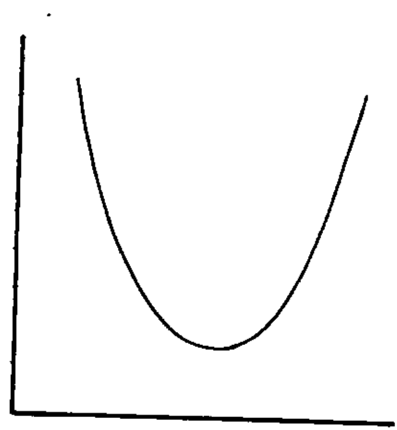
ان من أهم المنحنيات التكرارية التي قد نحصل عليها عمليا هي :

(١) 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 ٣ : ١٠)والمنحني ذو الشكل U أوالمنحني النوني (شكل ٣:١١).



شكل (٣ : ١٠) المنحني الطبيعي



شكل (٣:١١) منحني U
= المنحني النوني =

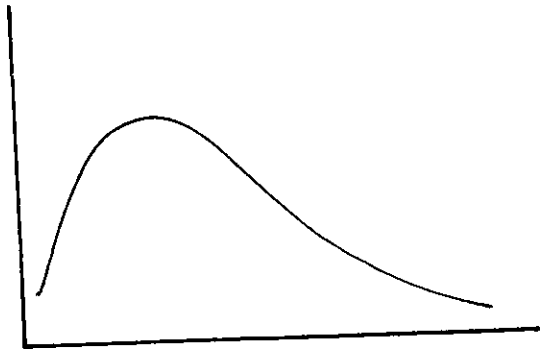
(٢) المنحنيات غير المتماثلة :أوالمنحنيات الملتوية Asymmetrical frequency curves.

وهي المنحنيات التي تكون احد اطرافها اطول من الآخروتنقسم الى :

(أ) منحنيات ملتوية التواء معتدلاً :

مثل : منحنيات ملتوية التواء موجباً
Positive skewness or skewed to the right

وهي المنحنيات التي طرفها الطويل يقع في الجهة اليمنى (شكل ٣:١٢)

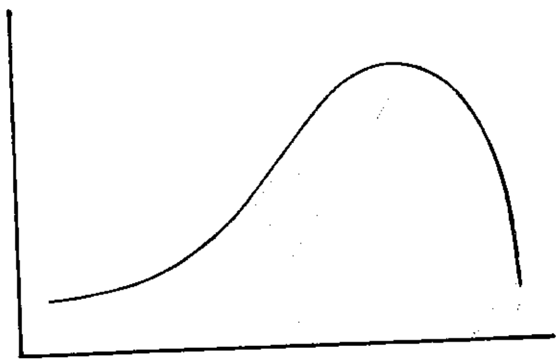


شكل (٣ : ١٢) منحني ملتوي التواءً موجباً

وكذلك منحنيات ملتوية التواء سالبا

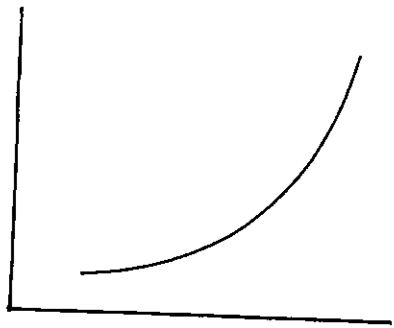
Negative skewness or skewed to the left

وهي المنحنيات التي طرفها الطويل يقع في الجهة اليسرى (شكل ٣ : ١٣)



شكل (٣ : ١٣) منحني ملتوي التواءً سالباً

(ب)منحنيات ملتوية التواءً شديداً مثل المنحنيين التاليين (شكل٣ : ١٤ ، ٣ : ١٥)



شكل (٣ : ١٤) منحني على شكل الحرف J
= المنحني الرائي =

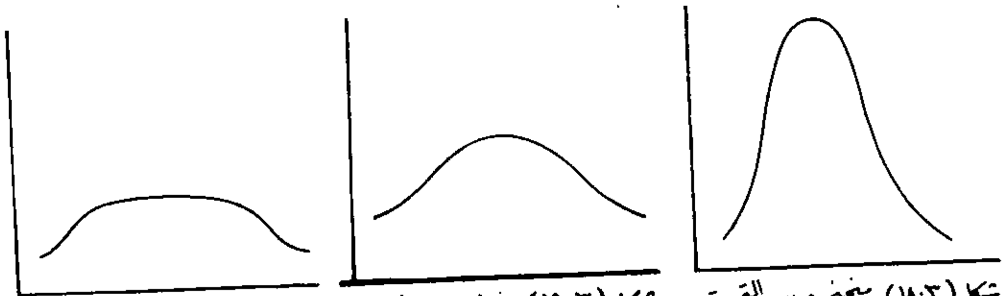
شكل (٣ : ١٥) منحني على شكل مقلوب J
= المنحني الرائي المقلوب =

(٣) منحنيات متعددة القمم مثل :(شكل ٣ : ١٦ و ٣ : ١٧)

شكل (٣ : ١٧) منحني متعدد القمم Multimodal

شكل (٣ : ١٦) منحني ذو قمتين Bimodal

(٤) منحنيات متفلطحة Kurtosis من امثال : منحنيات مدببة القمة (شكل ٣ : ١٨) ومعتدلة القمة (شكل ٣ : ١٩) ومفلطحة (شكل ٣ : ٢٠) .



شكل (٣:١٨) منحني مدبب القمة Leptokurtic

شكل (٣:١٩) منحني معتدل القمة Mesokurtic

شكل (٣:٢٠) منحني مفلطح Platykurtic

Lepta

تمارين الفصل الثالث

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

(أ) ٧ – ١٣ (ب) (–٥) – (–١)

(ج) ١٠,٤ – ١٨,٧ (د) ٠,٣٤٦ – ٠,٤١٨

(هـ) (–٢,٧٥) – ١,٣٥ (و) ٧٨,٤٩ – ٨٦,٧٢

(٢) أوجد طول الفئة المناسبة لكل من التوزيعات التالية على فرض أن عدد الفئات في كل منها = ١٠

(أ) أقل قيمة = ٧,٥ وأكبر قيمة = ١٨,٦

(ب) أقل قيمة = ٥٣ واكبر قيمة = ١٤٩

(ج) أقل قيمة = –١٥ اكبر قيمة = صفر

(٣) اذا علمت بأن مراكز الفئات لاعمار عدد من الطلبة هي : ١٨ ، ٢١ ، ٢٤ ، ٢٧ ، ٣٠ ، فما هي :

(أ) طول الفئة (ب) الحدود الحقيقية للفئات

(ج) حدود الفئات لهذا التوزيع ؟

(٤) فيما يلي درجات ٦٠ طالباً في الأمتحان النهائي لدرس الأحصاء

٧٣	٧٨	٧٥	٧٤	٨٤	٨١
٨٠	٦٧	٥٤	٦٥	٦٣	٧٤
٥٢	٧٩	٧٦	٧٥	٧٠	٦٥
٤١	٨١	٨٢	٧٢	٨٥	٨٥
٦٠	٦٤	٥٧	٤٨	٩٨	٣٦
٣٤	٨٣	٤١	٧٤	٦٢	٧٦
٦٠	٨٩	٧٨	٥٢	٩٠	٦٧
٧٧	٦٧	٦٤	٩٢	٨٠	٤٣
١٠٠	٣٢	٨٤	٨٠	٨٢	٧٩
٧١	٩٥	٦٩	٨٨	٥٥	٦١

١٠

والمطلوب ايجاد مايلي :

(أ) انشاء جدول التوزيع التكراري بأستعمال عشر فئات.

(ب) ارسم المدرج التكراري .

(ج) ارسم المضلع التكراري .

(د) انشاء جدول التوزيع التكراري التجميعي التصاعدي والتنازلي .

(هـ) ارسم المضلع التكراري التجميعي والتنازلي في رسم واحد .

(٥) اكمل كلاً من الجداول التالية (علماً بأن اطوال الفئات متساوية وأنها ارقام صحيحة).

الفئات	مراكز الفئات	الحدود الحقيقية
١ – ٧	٤	
–		
–		
–		
–		
–		

ب)

الفئات	مراكز الفئات	الحدود الحقيقية
٣ - ٦	٤,٥	٢,٥ - ٦,٥
٧ - ١٠	٨,٥	٦,٥ - ١٠,٥
١١ - ١٤	١٢,٥	١٠,٥ - ١٤,٥
١٥ - ١٨	١٦,٥	١٤,٥ - ١٨,٥
١٩ - ٢٢	٢٠,٥	١٨,٥ - ٢٢,٥
٢٣ - ٢٦	٢٤,٥	٢٢,٥ - ٢٦,٥

اضافة سنة طول الفئة

+٠,٥   -٠,٥

(جـ)

طول الفئة = الفرق بين مركزي فئتين متتاليين / عدد فئات بينها

الفئات	مراكز الفئات	الحدود الحقيقية
٥,٥ - ١٠,٥	٨	٥ - ١١
١١,٥ - ١٦,٥	١٤	١١ - ١٧
١٧,٥ - ٢٢,٥	٢٠	١٧ - ٢٣
٢٣,٥ - ٢٨,٥	٢٦	٢٣ - ٢٩
٢٩,٥ - ٣٤,٥	٣٣	٢٩ - ٣٥

+٠,٥   -٠,٥

(٦) الجدول التالي يبين التوزيع التكراري لاعمار المصابيح الكهربائية من انتاج شركة ما :

فئات العمر (بالساعات)	التكرار (عدد المصابيح)
٣٠٠ – ٣٩٩	١٤
٤٠٠ – ٤٩٩	٤٦
٥٠٠ – ٥٩٩	٥٨
٦٠٠ – ٦٩٩	٧٦
٧٠٠ – ٧٩٩	٦٨
٨٠٠ – ٨٩٩	٦٢
٩٠٠ – ٩٩٩	٤٨
١٠٠٠ – ١٠٩٩	٢٢
١١٠٠ – ١١٩٩	٦

والمطلوب ايجاد كل مما يلي :

(أ) التكرار النسبي للفئة السادسة .

(ب) نسبة المصابيح التي عمرها لايزيد عن ٦٠٠ ساعة .

(ج) نسبة المصابيح التي عمرها مساو أو أكثر من ٩٠٠ ساعة.

(د) نسبة المصابيح التي عمرها على الأقل ٥٠٠ ساعة ولكن أقل من ١٠٥٠ ساعة.

(هـ) ارسم المدرج التكراري .

(و) ارسم المضلع التكراري التجميعي التصاعدي والتنازلي في رسم واحد .

(ز) نسبة المصابيح التي عمرها:

(١) أقل من (٥٦٠) ساعة .

(٢) ٩٧٠ ساعة أو أكثر .

(٣) بين ٦٢٠ الى ٨٩٠ ساعة .

الفصل الرابع

مقاييس التمركز أو التوسط

Measures of Central Tendency

(٤ : ١) مقدمة

ان معظم القيم لمختلف الظواهر الطبيعية تتمركز عادة 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 وان هذه القيمة المتوسطة أو المتمركزة هي رقم واحد يعبر عن أو يمثل جميع بيانات تلك المجموعة .

وأهم مقاييس التوسط هي :

(١) الوسط الحسابي (أو المتوسط) $\bar{y}$ — The Arithmetic Mean

(٢) الوسط الهندسي $\bar{G}$ — The Geometric Mean

(٣) الوسط التوافقي $\bar{H}$ — The Harmonic Mean

(٤) الوسط التربيعي $\bar{Q}$ — The Quadratic Mean

(٥) الوسيط Me — The Median

(٦) المنوال Mo — The Mode

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

١. حالة البيانات غير المبوبة

٢. حالة البيانات المبوبة

(٤ : ٢) الوسط الحسابي — The Arithmetic Mean

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

طرق حسابه :

(أ) من بيانات غير مبوبة :

تعريف (٤ : ١) :

اذا كان لدينا n من القيم أو المشاهدات : $y_1 , y_2 , \dots\dots\dots , y_n$

فإن الوسط الحسابي لها هو : $\bar{y} = \frac{\sum^{n} y_i}{n}$

مثال (١) : البيانات التالية تمثل كمية المطر الساقطة سنويا (بالمليمترات) على مدينة الموصل خلال فترة خمس سنوات ٥٢٠ ، ٣٥٠ ، ٤٥٠ ، ٣٨٠ ، ٤٠٠

فا هو متوسط سقوط المطر خلال هذه الفترة ؟

الحــل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{520 + 350 + 450 + 380 + 400}{5}$$

$$= \frac{2100}{5}$$

$$= 420 \text{ mm.}$$

اي ان معدل سقوط الأمطار خلال تلك الفترة هو ٤٢٠ ملم .

مثال (٢) : أحسب الوسط الحسابي لأطوال نباتات القطن في جدول (٣ : ٥) قبل تبويبها .

الحــل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{80 + 84 + \dots\dots\dots + 75}{80}$$

$$= \frac{6126}{80} = 76{\cdot}58 \quad \text{cm.}$$

اي ان معدل طول النبات هو ٧٦,٥٨ سم .

(ب) من بيانات مبوبة :

تعريف (٤ : ٢) :

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

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

خطوات إيجاد الوسط الحسابي في بيانات مبوبة هي كالآتي :

(١) تعيين مراكز الفئات y_i.

(٢) ضرب مركز كل فئة بمقدار تكرارها $(f_i y_i)$.

(٣) قسمة مجموع (حاصل ضرب مركز كل فئة × تكرارها) على مجموع التكرارات .

مثال (٣) : إستخرج الوسط الحسابي لأطوال النباتات من جدول التوزيع التكراري (٣ : ٦).

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

جدول (٤ : ١)

الفئـات	التكـرار	مركز الفئـات	التكرار × مركز الفئات
٣٦ – ٤٠	١	٣٥ ,٥	٣٥ ,٥
٤١ – ٥٠	٢	٤٥ ,٥	٩١ ,٠
٥١ – ٦٠	٥	٥٥ ,٥	٢٧٧ ,٥
٦١ – ٧٠	١٥	٦٥ ,٥	٩٨٢ ,٥
٧١ – ٨٠	٢٥	٧٥ ,٥	١٨٨٧ ,٥
٨١ – ٩٠	٢٠	٨٥ ,٥	١٧١٠ ,٠
٩١ – ١٠٠	١٢	٩٥ ,٥	١١٤٦,٠
	$\Sigma f_i = 80$		$\Sigma f_i x_i = 6130{\cdot}0$

$$\therefore \quad \bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6130}{80} = 76.62 \text{ cm.}$$

أي ان معدل طول النبات هو ٧٦،٦٢ سم .

لاحظ بأن هذا الرقم يختلف قليلا عن الوسط الحسابي لنفس البيانات قبل تبويبها ووضعها في جدول توزيع تكراري (٧٦،٥٨ سم) . ان الفرق هذا بين الرقمين يعود الى فقدان المعلومات عن المفردات او المشاهدات بسبب وضعها في مجاميع فنحن نفرض بأن طول كل النباتات في فئة معينة مساوياً لمركز تلك الفئة .

مهم جداً (٣) خواص الوسط الحسابي

(أ) مجموع إنحرافات القيم عن وسطها الحسابي تساوي صفرا

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

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

البرهان :

$$\begin{aligned}
\sum (y_i - \bar{y}) &= \sum y_i - \sum \bar{y} \\
&= \sum y_i - n\bar{y} \\
&= \Sigma y_i - \Sigma y_i \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\sum f_i(y_i - \bar{y}) &= \sum f_i y_i - \bar{y}\sum f_i \\
&= \sum f_i y_i - \left(\frac{\sum f_i y_i}{\sum f_i}\right)\sum f_i \\
&= \sum f_i y_i - \sum f_i y_i \\
&= 0
\end{aligned}$$

والجدولان التاليان يوضحان ذلك :

جدول (٤ : ٢)

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

جدول (٤ : ٣)

$f_i(y_i - \bar{y})$	$(y_i - \bar{y})$	$f_i y_i$	مركز الفئات y_i	التكرار f_i	الفئات
− ٣٢,٢٥	− ٦,٤٥	٣٠٥	٦١	٥	٦٠ − ٦٢
− ٦٢,١٠	− ٣,٤٥	١١٥٢	٦٤	١٨	٦٣ − ٦٥
− ١٨,٩٠	− ٠,٤٥	٢٨١٤	٦٧	٤٢	٦٦ − ٦٨
+ ٦٨,٨٥	+ ٢,٥٥	١٨٩٠	٧٠	٢٧	٦٩ − ٧١
+ ٤٤,٤٠	+ ٥,٥٥	٥٨٤	٧٣	٨	٧٢ − ٧٤
$\sum f_i(y_i - \bar{y}) = 0$		$\sum f_i y_i$ = ٦٧٤٥ $\bar{y} = \frac{\sum f_i y_i}{\sum f_i}$ = ٦٧,٤٥		$\sum f_i$ = ١٠٠	

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

البرهان :

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

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

وباضافة وطرح $n(\bar{y})^2$ ، من اعلاه ينتج :

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

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 (٤) من القيم التالية: $y_i = 9, 8, 6, 5, 7$

$$\therefore \bar{y} = \frac{\sum y_i}{n} = 7$$

$$\sum(y_i - \bar{y})^2 = (9 - 7)^2 + (8 - 7)^2 + (6 - 7)^2 + (5 - 7)^2 + (7 - 7)^2 = 10$$

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

$$\sum (y_i - A)^2 = \sum(y_i - 10)^2$$
$$= (9 - 10)^2 + (8 - 10)^2 + (6 - 10)^2 + (5 - 10)^2 + (7 - 10)^2$$
$$= 55$$

وطبيعي ٥٥ أكبر من ١٠
ويلاحظ هنا أن الفرق بينهما هو : $55 - 10 = 45$

وهــو $n(A - \bar{y})^2$

أي $5(10 - 7)^2 = 45$

بكير

(ج) عند إضافة عدد ثابت (k) الى كل قيمة من قيم المشاهدات فإن :
الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الاصلية + العدد الثابت (k)

$x_i = y_i + k$

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

البرهان :

$$x_i = y_i + k$$

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

$$= \sum y_i + nk$$

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

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

مثال (٥) نفرض ان لدينا القيم التالية :

$$y_i = 8, 3, 2, 12, 10$$

فالوسط الحسابي لها هو :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{35}{5} = 7$$

فإذا أضفنا لكل من هذه القيم قيمة ثابتة ولتكن ٣
فالقيم الجديدة ستصبح :

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

والوسط الحسابي للقيم الجديدة هو :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50}{5} = 10$$

الذي هو في الحقيقة :

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

مثال (٦) نفرض ان القيم الأصلية هي :

$$y_i = 5, 10, 8, 7, 10$$

$$\therefore \bar{y} = \frac{40}{5} = 8$$

فإذا طرحنا ٢ من كل مشاهدة
فإن الوسط الحسابي للمشاهدات الجديدة سيكون :

أي :

$$\bar{y} - 2 = 8 - 2 = 6$$

$$x_i = 3, 8, 6, 5, 8$$

$$\therefore \bar{x} = \frac{30}{5} = 6$$

(د) إذا ضربت كل قيمة من قيم المشاهدات في قيمة ثابتة (k) فإن :
الوسط الحسابي للقيم الجديدة = الوسط الحسابي للقيم الاصلية × العدد الثابت k

$$z_i = k y_i$$

$$\bar{z} = k \bar{y}$$

البرهان :

$$z_i = k y_i$$

$$\sum z_i = k \sum y_i$$

$$\frac{\sum z_i}{n} = k \frac{\sum y_i}{n}$$

$$\therefore \bar{z} = k \bar{y}$$

مثال (٧) في القيم التالية :

$$y_i = 8, 3, 2, 12, 10$$

نجد أن

$$\bar{y} = 7$$

فإذا كان

$$z_i = 5 y_i$$

أوجد قيمة $\bar{z}$

الحل :

$$\therefore z_i = 40, 15, 10, 60, 50$$

$$\therefore \bar{z} = \frac{\sum z_i}{n} = \frac{175}{5} = 35$$

وهي تساوي

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

$$\therefore \quad \bar{z} = 5(7) = 35$$

هذا ويمكن تعميم الخاصيتين السابقتين بالقانون التالي

إذا كان : $x_i = a + by_i$

فـإن : $\therefore \quad \bar{x} = a + b\bar{y}$

(هـ) الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين أي :

إذا كان $z_i = k\,y_i + y_i$

فإن $\bar{z} = \bar{x} + \bar{y}$

البرهان :

$$z_i = x_i + y_i$$

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

$$= \sum x_i + \sum y_i$$

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

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

مثال (٨) اعتبر الجدول التالي :

$z_i = x_i + y_i$	y_i	x_i
٧	٥	٢
١٤	١٠	٤
١٤	٨	٤
١٥	٧	٨
١٥	١٠	٥
$\bar{z}$ = ١٣	$\bar{y}$ = ٨	$\bar{x}$ = ٥

من هذا يتضح بأن :

$$\bar{z} = \bar{x} + \bar{y} = 5 + 8 = 13$$

(٩) إذا كان لكل قيمة من المشاهدات (y_i) وزن خاص يتناسب مع أهميتها (w_i) فإن الوسط الحسابي (الموزون) لهذه القيم هو :

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

مثال (٩) القيم التالية تمثل نتائج إمتحان أحد الطلبة في درس الأحصاء علما بأن لكل أمتحان وزنا أو أهمية أو نسبة معينة .

الامتحان	الدرجة y_i	أهميتها أو نسبتها أو وزنها w_i	$w_i y_i$
الأول	٧٠	١٠ ٪	٧٠٠
الثاني	٦٠	٣٠ ٪	١٨٠٠
الثالث	٧٥	١٠ ٪	٧٥٠
الرابع	٥٥	٥٠ ٪	٢٧٥٠
		$\sum w_i =$ ١٠٠	$\sum w_i y_i =$ ٦٠٠٠

فالوسط الحسابي أو معدل الطالب سيكون :

$$\bar{y} = \frac{\sum w_i y_i}{\sum w_i} = \frac{6000}{100} = 60$$

مثال (١٠) : أربع شعب من الطلبة في الصف الأول تتألف من ٣٠ و٣٥ و٤٠ و٢٥ طالبا على التوالي فإذا كان معدل أمتحانهم بمادة الأحصاء هو ٨٠ و٧٥ و٦٠ و٩٠ على التوالي فما هو معدل الامتحان في جميع هذه الشعب ؟

الحـل :

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

$$= 74{\cdot}4$$

(٤ : ٣) الوسط الهندسي The Geometric Mean

(آ) بيانات غير مبوبة :

تعريف (٤ : ٣) :

إذا كان لدينا n من القيم أو المشاهدات :

$y_1 , y_2, \ldots\ldots\ldots\ldots, y_n$

فإن الوسط الهندسي لها (ويرمز له بالرمز $\bar{G}$) هو :

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

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

$$\text{Log}\ \bar{G} = (1/n) \log[(y_1)(y_2)\ldots\ldots(y_n)]$$

$$\boxed{\therefore\ \text{Log}\ \bar{G} = \frac{\log y_1 + \log y_2 + \ldots. + \log y_n}{n}}$$

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 (١١) : أوجد الوسط الهندسي والوسط الحسابي للقيم التالية :

$y_i = 3, 5, 8, 3, 7, 2$

الحـل :

(١) الوسط الهندسي

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

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

$$= \frac{\log 3 + \log 5 + \log 8 + \log 3 + \log 7 + \log 2}{6}$$

$$= \frac{0{\cdot}4771 + 0{\cdot}6990 + \dots\dots + 0{\cdot}3010}{6}$$

$$= \frac{3{\cdot}7024}{6}$$

$$\therefore \log \bar{G} = 0{\cdot}6171$$

$$\therefore \bar{G} = 4{\cdot}14$$

أو يمكن حساب الوسط الهندسي بالطريقة التالية :

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

$$\therefore \log \bar{G} = \frac{1}{6}\log 5760 = \frac{3{\cdot}7024}{6}$$

$$= 0{\cdot}6171$$

$$\therefore \bar{G} = 4{\cdot}14$$

أما الوسط الحسابي فهو :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{3+5+8+3+7+2}{6} = \frac{28}{6} = 4{\cdot}67$$

من ذلك يتضح بأن الوسط الهندسي لقيم مختلفة موجبة دائما 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 او السكان ... الخ . كما إنه لا يمكن إيجاد الوسط الهندسي الا إذا كانت مجموع القيم موجبة .

(ب) من بيانات مبوبة

تعريف (٤ : ٤) :

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

$$\bar{G} = \sqrt[\Sigma f_i]{(y_1)^{f_1} (y_2)^{f_2} \dots (y_k)^{f_k}}$$

مركز الفئة تكرار الفئة

الوسط الحسابي ≥ الوسط الهندسي

الوسط الحسابي = الوسط الهندسي إذا تساوت البيانات

وباستخدام اللوغاريتمات فإن :

$$\text{Log } \bar{G} = \frac{\sum f_i \log y_i}{\sum f_i}$$

$$= \frac{f_1 \log y_1 + f_2 \log y_2 + ... + f_k \log y_k}{\sum f_i}$$

مثال (١٢) أوجد الوسط الهندسي لجدول التوزيع التكراري (٣ : ١)

الحـل :

$f_i \log y_i$	$\log y_i$	y_i	تكرارها f_i	الفئـات
٨,٨٩١٠	١,٧٧٨٢	٦١	٥	٦٠ – ٦٢
٣٢,٥١١٦	١,٨٠٦٢	٦٤	١٨	٦٣ – ٦٥
٧٦,٦٩٦٢	١,٨٢٦١	٦٧	٤٢	٦٦ – ٦٨
٤٩,٨١٧٧	١,٨٤٥١	٧٠	٢٧	٦٩ – ٧١
١٤,٩٠٦٤	١,٨٦٣٣	٧٣	٨	٧٢ – ٧٤
١٨٢,٨٢٢٩				

$$\log \bar{G} = \frac{\sum f_i \log y_i}{\sum f_i} = \frac{182{\cdot}8229}{100} = 1{\cdot}8282$$

$$\therefore \bar{G} = 67{\cdot}3$$

بينما الوسط الحسابي =٦٧,٤٥

(٤ : ٤) الوسط التوافقـي The Harmonic Mean

(آ) بيانات غير مبوبة

تعريف (٤ : ٥) :

إذا كان لدينا n من القيم او المشاهدات : $y_1, y_2, ..., y_n$

فإن الوسط التوافقي لها (ويرمز له بالرمز $\bar{H}$) هو :

$$\bar{H} = \frac{1}{(\sum \frac{1}{y_i})/n} = \boxed{\frac{n}{\sum \frac{1}{y_i}}}$$

فالوسط التوافقي هو مقلوب الوسط الحسابي لمقلوب القيم او المشاهدات .

مثال (١٣) : أوجد الوسط التوافقي للقيم التالية :

$y_i = 3, 5, 6, 6, 7, 10\ 12,$

الحـل :

$$\bar{H} = \frac{n}{\sum 1/y_i} = \frac{n}{1/y_1 + 1/y_2 + ... + 1/y_n}$$

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

$$= 5{\cdot}87$$

ان الوسط التوافقي أكثر ما يستعمل هو عندما تعطى مجموعة من البيانات منسوبة الى وحدة ثابتة .

مثال (١٤) : إشترى مزارع بذور حنطة بـ ١٠٠ دينار من كل من الشركات التالية :

الشركة الأولى كان سعر الطن من بذور الحنطة = ٢٠ دينارا

والشركة الثانية كان سعر الطن من بذور الحنطة = ٢٥ دينارا

أما الشركة الثالثة فكان سعر الطن من بذور الحنطة = ٥٠ دينارا

فما هو متوسط سعر الطن من بذور الحنطة .

(ملاحظة : البيانات معبرة بـ عدة أطنان بالـ ١٠٠ دينار بينما المطلوب هو متوسط سعر الطن)

الحـل :

$$\bar{H} = \frac{3}{\frac{1}{20} + \frac{1}{25} + \frac{1}{50}} = 27{\cdot}27$$

(ب) لبيانات مبوبه

تعريف (٤ : ٦) :

إذا كانت $y_1, y_2, ..., y_k$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكرارتها :

$f_1, f_2, ..., f_k$ على التوالي

فالوسط التوافقي لها هو :

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

مثال (١٥) : أوجد الوسط التوافقي للتوزيع التكراري التالي :

الفئات	f_i	y_i
٦٠ - ٦٢	٥	٦١
٦٣ - ٦٥	١٨	٦٤
٦٦ - ٦٨	٤٢	٦٧
٦٩ - ٧١	٢٧	٧٠
٧٢ - ٧٤	٨	٧٣
المجموع	١٠٠	

الحــل :

$$\bar{H} = \frac{\sum f_i}{\sum \frac{f_i}{y_i}} = \frac{\sum f_i}{\frac{f_1}{y_1} + \frac{f_2}{y_2} + \dots\dots + \frac{f_k}{y_k}}$$

$$= \frac{100}{\frac{5}{61} + \frac{18}{64} + \frac{42}{67} + \frac{27}{70} + \frac{8}{73}} = \frac{100}{1{\cdot}4855} = 67{\cdot}3$$

(٤ : ٥) الوسط التربيعي The Quadratic Mean

(آ) لبيانات غير مبوبة :

تعريف (٤ : ٧) :

اذا كان لدينا n من القيم او المشاهدات :

$y_1, y_2, \dots\dots, y_n$

فإن الوسط التربيعي لها (ويرمز بـ $\bar{Q}$) هو :

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

* واسع الاستعمال في العلوم الفيزيائية
* الوسط الحسابي
* الوسط التوافقي H

أي ان الوسط التربيعي هو الجذر التربيعي للوسط الحسابي لمربعات القيم او المشاهدات .

مثــال (١٦) : أوجد الوسط التربيعي للبيانات التالية : $y_i = 1, 3, 4, 5, 7$

الحــل :

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

$$= \sqrt{\frac{(1)^2 + (3)^2 + (4)^2 + (5)^2 + (7)^2}{5}} = \sqrt{20} = 4.47$$

ان الوسط التربيعي يطبق بكثرة في العلوم الفيزيائية .

(ب) لبيانات مبوبة

* تعريف (٤ : ٨) :

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

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

مثــال (١٧) : أوجد الوسط التربيعي للتوزيع التكراري فـي جـدول (٣ : ١) .

الحــل :

الفئــات	f_i	y_i	y_i^2	$f_i y_i^2$
٦٠ − ٦٢	٥	٦١	٣٧٢١	١٨٦٠٥
٦٣ − ٦٥	١٨	٦٤	٤٠٩٦	٧٣٧٢٨
٦٦ − ٦٨	٤٢	٦٧	٤٤٨٩	١٨٨٥٣٨
٦٩ − ٧١	٢٧	٧٠	٤٩٠٠	١٣٢٣٠٠
٧٢ − ٧٤	٨	٧٣	٥٣٢٩	٤٢٦٣٢
	١٠٠			٤٥٥٨٠٣

$$\bar{Q} = \sqrt{\frac{\sum f_i y_i^2}{\sum f_i}} = \sqrt{\frac{455803}{100}} = \sqrt{4558{\cdot}03} = 67{\cdot}51$$

(٤ : ٦) الوسيط The Median

(آ) لبيانات غير مبوبة

تعريف (٤ : ٩) :

اذا كان لدينا n من القيم أو المشاهدات $y_1, y_2, \ldots, y_n$

ورتبت ترتيبا تصاعديا (أو تنازليا)

(١) فاذا كانت n عدد فردي

فان الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2}$

أي الوسيط (ويرمز له بـ $\bar{M}e$) هو $\bar{M}e = y_{(n+1)/2}$

(٢) أما اذا كانت n عدد زوجي

فان الوسيط هو الوسط الحسابي للقيمتين اللتين ترتيبهما

أي : $\frac{n}{2}$ ، $\frac{n}{2}+1$

$$\bar{M}e = \frac{y_{n/2} + y_{(n/2)+1}}{2}$$

مثال (١٨) : أوجد الوسيط لدرجات طالب في خمسة امتحانات بدرس الاحصاء اذا كانت الدرجات هي :

٨٠ ، ٨٢ ، ٧٦ ، ٨٧ ، ٨٤

الحل :

نرتب الدرجات تصاعديا

٨٧ ، ٨٤ ، ٨٢ ، ٨٠ ، ٧٦

وبما ان عدد الأرقام فردي $(n = 5)$

اذن فالوسيط هو القيمة التي ترتيبها : $\frac{n+1}{2} = \frac{5+1}{2} = 3$

أي الوسيط = ٨٢ $\bar{M}e = y_3 = 82$

مثال (١٩) أوجد الوسيط للقيم التالية

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

نرتب القيم تصاعديا :

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

وبما أن عدد القيم هو زوجي $(n = 8)$

اذن فالوسيط الحسابي للقيمتين اللتين ترتيبهما

$$(n/2) + 1, \ (n/2)$$

$$(n/2) = \frac{8}{2} = 4$$

$$(n/2) + 1 = 5$$

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

$\overline{Me} = L + \frac{(\sum f_i/2) - F_i}{f_i} w$

(ب) لبيانات مبوبة

تعريف (١٠ : ٤) :

اذا 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

فقيمة الوسيط لهذه البيانات (بالاستعانة بجدول التوزيع التكراري المتجمع الصاعد) هو

$$\overline{Me} = L_1' + \left[\frac{(\sum f_i/2) - F_i}{f_i} \right] w$$

حيث أن

- الحد الأدنى الحقيقي لفئة الوسيط $= L_1'$
- مجموع التكرارات $= \sum f_i$
- التكرار المتجمع عند بداية فئة الوسيط $= F_i$
- تكرار فئة الوسيط $= f_i$

= التكرار المتجمع عند نهاية فئة الوسيط – التكرار المتجمع عند بداية فئة الوسيط

طول فئة الوسيط $= w$

وخطوات ايجاد الوسيط هي :

١) عمل جدول توزيع تكراري تجميعي تصاعدي (أو تنازلي)

٢) ايجاد ترتيب الوسيط وهو $\frac{\sum f_i}{2}$

٣) نحدد فئة الوسيط وهي الفئة التي تقع قيمة الوسيط بين حديها وذلك عن طريق

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

٤) نطبق القانون ....

مثـال (٢٠) اوجد الوسيط للتوزيع التكراري في جدول (٣ : ١)

الحل :

فئات الطول	تكرارات f_i	المتجمع الصاعد		
			F_i	
		أقل من ٦٠	٠	
٦٠ – ٦٢	٥	أقل من ٦٣	٥	
٦٣ – ٦٥	١٨	أقل من ٦٦	٢٣	فئة الوسيط
٦٦ – ٦٨	٤٢	أقل من ٦٩	٦٥	
٦٩ – ٧١	٢٧	أقل من ٧٢	٩٢	
٧٢ – ٧٤	٨	أقل من ٧٤	١٠٠	
	١٠٠			

ترتيب الوسيط

$$\frac{\sum f_i}{2} = \frac{100}{2} = 50$$

ان قيمة الوسيط هو طول الشخص الذي ترتيبه ٥٠ (بعد ترتيب القيم تصاعديا او تنازليا) .

من جدول التوزيع التكراري التجميعي التصاعدي نرى بأن ٥٠ هي واقعة بين الرقمين ٢٣ ، ٦٥

فئة الوسيط هي

الحدود الحقيقية	68.5	65.5
التكرار التجميعي	65	23

الحد الأدنى الحقيقي لفئة الوسيط $\therefore \quad L_1 = 65.5$

التكرار المتجمع عند بداية فئة الوسيط $\therefore \quad F_i = 23$

تكرار فئة الوسيط $f_i = 65 - 23 = 42$

طول فئة الوسيط $w = 68.5 - 65.5 = 3$

$$\therefore \quad \bar{Me} = L_1 + \left[\frac{(\sum f_i/2) - F_i}{f_i} \right] w$$

$$= 65.5 + \left[\frac{50 - 23}{42} \right] (3)$$

$$= 67.43 \text{ inch.}$$

ويمكن تلخيص خطوات إيجاد الوسيط لبيانات مبوبة كالآتي :

١) عمل جدول توزيع تكراري تجميعي تصاعدي .

٢) إيجاد ترتيب الوسيط = $\frac{\sum f_i}{2}$

٣) إيجاد فئة الوسيط :

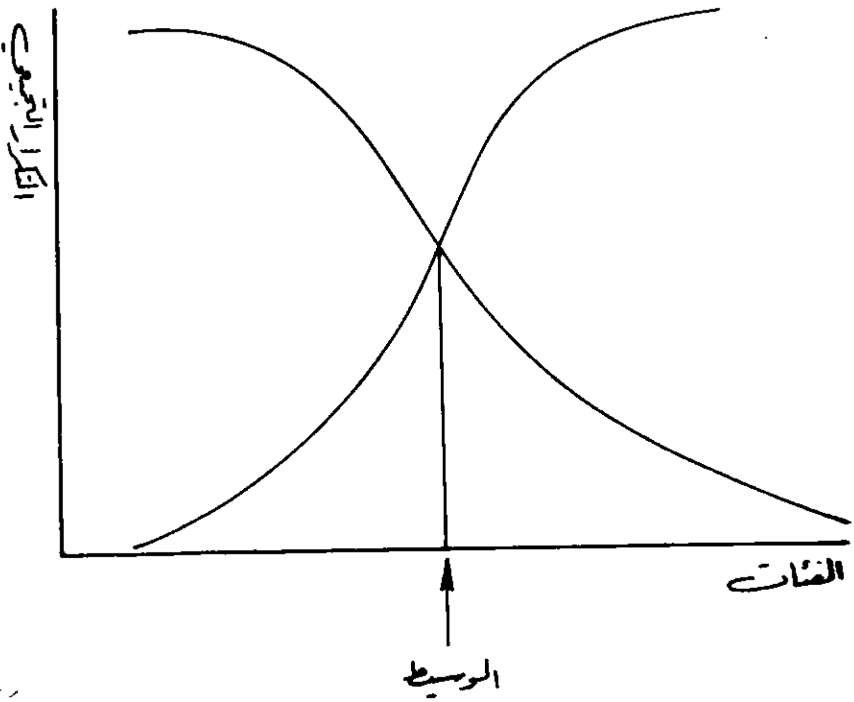
أ) إيجاد حدودها الحقيقية

ب) كتابة التكرار التجميعي التصاعدي أمام كل منها .

٤) تطبيق القانون .

ملاحظة : من الممكن إيجاد ترتيب الوسيط بـ ($\frac{\sum f_i}{2}$) إذا كان عدد المفردات فرديا أو بـ$\left(\frac{\sum f_i}{2} + 1 \right)$ إذا كان عدد المفردات زوجيا . ولكن نظرا لكون المفردات كبيرا في التوزيعات التكرارية فتستخدم$\left(\frac{\sum f_i}{2} \right)$ لإيجاد ترتيب الوسيط .

هذا ويمكن إيجاد قيمة الوسيط بإستخدام الرسم البياني للمنحنيين التصاعدي والتنازلي وذلك بانزال عمود من نقطة تقاطعهما الى المحور السيني ليقطعه في نقطة هي قيمة الوسيط كما في الشكل .



The Mode

شكل (٤ : ١) المضلع التكراري التجميعي التصاعدي والتنازلي مع موقع الوسيط .

(٤ : ٧) المنوال او القمة The Mode

أ) لبيانات غير مبوبة

تعريف (٤ : ١١) :

اذا كان لدينا n من المشاهدات $y_1, y_2, \ldots\ldots, y_n$
فإن المنوال لهذه المشاهدات هو المشاهدة او القيمة الأكثر تكرارا بين هذه المشاهدات ويرمز له بـ Mo

ومن هذا يتضح بأنه قد يكون هناك منوالا واحدا (قمة واحدة) لهذه المشاهدات وعندها يسمى التوزيع وحيد القمة unimodal او يكون لها منوالان (قمتان) وعندها يسمى التوزيع ذو قمتين bimodal وقد يكون لها أكثر من منوالين كما انه قد لا يوجد منوال للمشاهدات .

مثـال (٢١) اوجد المنوال لكل من البيانات التالية :

(آ) $3, \underline{5}, 2, 6, \underline{5}, 9, \underline{5}, 2, 8, 6$

(ب) $51{\cdot}6, 48{\cdot}7, 50{\cdot}3, 49{\cdot}5, 48{\cdot}9$ لا يوجد منوال .

الحـل : (آ) المفردة ٥ هي أكثر المفردات تكرارا فهي المنوال

$\therefore\ \bar{M}o = 5$

(ب) لا يوجد منوال لهذه المفردات

(ب) لبيانات مبوبة

اذا كانت القيم $y_1, y_2, \ldots\ldots, y_k$

تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها

$f_1, f_2, \ldots\ldots f_k$

على التوالي

فان المنوال :

$$\bar{M}o = L_1 + \left(\frac{d_1}{d_1 + d_2}\right) w$$

حيث أن

* مهم لكل بيانات المنوال

فئة المنوال : تلك الفئة التي تملك اكبر التكرارات وأن :

الحد الأدنى الحقيقي لفئة المنوال $= L_1$

الفرق بين تكرار فئة المنوال والفئة السابقة لها $= d_1$

الفرق بين تكرار فئة المنوال والفئة اللاحقة لها $= d_2$

طول الفترة $= w$

مثـال (٢٢) اوجد المنوال للجدول التوزيع التكراري (٣ : ١)

* يجب تحديد اول مرة الفئة ثم بعد ذلك تحديد الحد الحقيقي الأدنى لفئة المنوال .

من ضلال الاحتمالات لتوزيع المنوالات .
١- ممكن تكون توزيع بدون منوال .
٢- توجد منوال واحد . او منوالين او اكثر .

الحـل :

الفئات	f_i
٦٠ – ٦٢	٥
٦٣ – ٦٥	١٨
٦٦ – ٦٨	٤٢
٦٩ – ٧١	٢٧
٧٢ – ٧٤	٨
	١٠٠

$Mo = 65.5 + \frac{(42-18)}{(42+15)} = 15$

فئة المنوال :

ان الفئة (٦٦ – ٦٨) لها اكبر التكرارات (٤٢) فهي فئة المنوال

الحد الأدنى الحقيقي لفئة المنوال

$L_1 = 65.5$

$d_1 = 42 - 18 = 24$

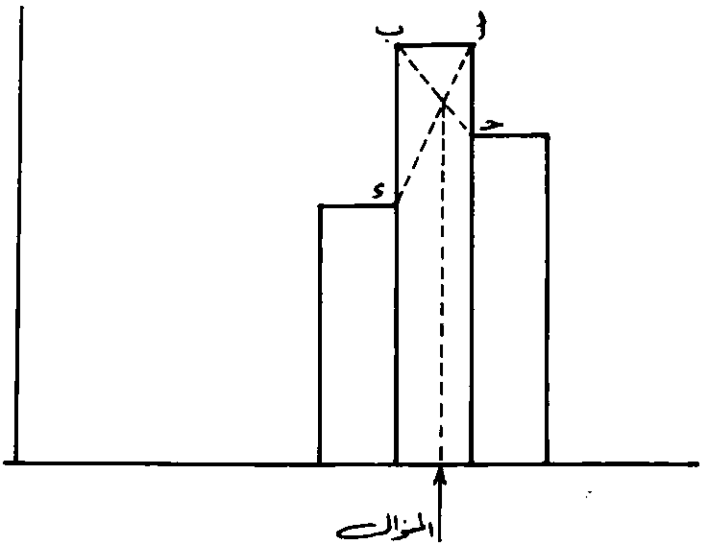
$d_2 = 42 - 27 = 15$

$w = 3$

$$\bar{M}o = 65.5 + \left(\frac{24}{24 + 15}\right)(3)$$

$$= 67.35$$

هذا ويمكن تقدير المنوال بطريقة الرسم البياني للمدرج التكراري وذلك باستعمال مستطيل الفئة المنوالية (وهو أعلى مستطيل لأنه يمثل اكثر التكرارات) والمستطيلان المجاوران كما في الشكل التالي :



شكل (٤ : ٢) المدرج التكراري وموقع المنوال

حيث نصل ا مع د وب مع جـ ومن نقطة تلاقيهما ننزل عمودا على المحور السيني يقطعه في نقطة هي قيمة المنوال .

(٤ : ٨) مقاييس أخرى للتوسط أو التمركز :

(١) الربيع الأدنى والربيع الاعلى : Upper and Lower Quartiles

الربيع الادنى (او الربيع الاول) : هو قيمة المفردة التي تقسم المفردات (بعد ترتيبها تصاعديا او تنازليا) الى قسمين بحيث يسبقها ٢٥٪ من المفردات ويليها ٧٥٪ من المفردات .

والربيع الاعلى (او الربيع الثالث) : هو قيمة المفردة التي تقسم المفردات (بعد ترتيبها تصاعديا او تنازليا) الى قسمين بحيث يسبقها ٧٥٪ من المفردات ويليها ٢٥٪ من المفردات .

ملاحظة : يمكن إعتبار الوسيط بالربيع الاوسط (اوالربيع الثاني) لأنه يتوسط المفردات أي يقسم المفردات الى قسمين (بعد ترتيبها تصاعديا او تنازليا) بحيث يسبقه ٥٠٪ ويليه ٥٠٪ من المفردات .

(٢) العشريات Deciles والمئينيات Percentiles

العشريات والمئينيات هي مواقع أيضا على التوزيع تقسم المفردات (بعد ترتيبها

تصاعديا او تنازليا) فمثلا العشير الثالث : هو قيمة المفردة او المشاهدة التي يسبقها

$\frac{٣}{١٠}$ المفردات ويليها $\frac{٧}{١٠}$ المفردات ....

أما المئين السبعون فهو قيمة المفردة التي يسبقها ٧٠٪ من المفردات ويليها ٣٠٪ من المفردات .

(٣) منتصف المدى او المدى المتوسط Mid – Range

وهو الوسط الحسابي لأصغر وأكبر قيمة بين المفردات ويرمز له بـ M.R. .

$$M.R. = \frac{y_{min} + y_{max}}{2}$$

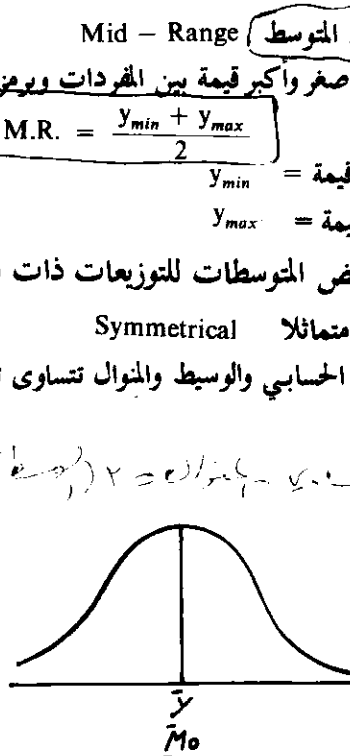
حيث ان y_{min} = أصغر قيمة

y_{max} = أكبر قيمة

(٤ : ٩) العلاقة بين بعض المتوسطات للتوزيعات ذات منوال أو قيمة واحدة unimodal

(١) إذا كان التوزيع متماثلا Symmetrical

فإن قيمة الوسط الحسابي والوسيط والمنوال تتساوى تماما .



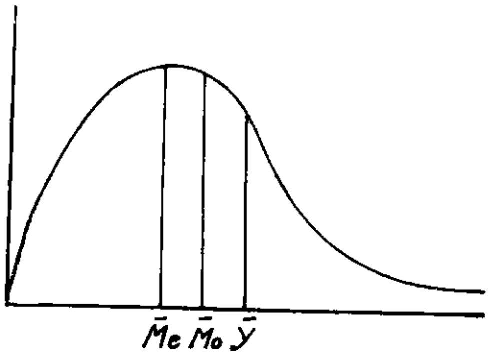
منحني متماثل

(٢) أما اذا كان التوزيع غير متماثل Asymmetrical
(ملتوى التواء معتدلا)

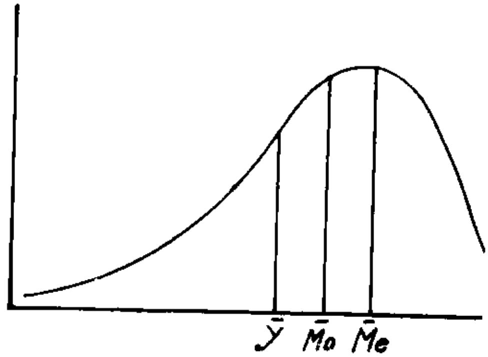
فإن :

الوسط الحسابي – المنوال = ٣ (الوسط الحسابي – الوسيط)

$$\bar{y} - \bar{M}o = 3(\bar{y} - \bar{M}e)$$



التواء موجب



التواء سالب

(٣) في حالة جميع المفردات موجبة فإن :

$$\bar{Q} \leq \bar{H} \leq \bar{G} \leq \bar{y}$$

وحالة التساوي هذه تظهر عندما تتساوى قيم جميع المفردات .

ومن هذا يتضح بأن الوسط الحسابي يكون أكبر من الوسط الهندسي والتوافقي والتربيعي بينما الوسط التربيعي فيكون أقل قيمة من الوسط الحسابي والهندسي والتوافقي بينما الوسط الهندسي يكون اكبر من التوافقي .

تمارين الفصل الرابع

(١) البيانات التالية تمثل متوسط لمحصول الدونم من الذرة الصفراء في عينة مكونة من ٤٠ مزرعة في العراق :

٦٥٠	٩٢٠	١٠٢٣	٥٨٨
٥٠٠	١٢٣٠	٨٩٠	٧٩٦
٤٢٠	٣٥٨	٩٨٠	٨٠٠
٧٥٠	٨٤٠	١٢٧٠	٦٨٠
٧٢٠	٩٦٠	١٢٦١	٣٢٠
٣٥٠	٨٩٥	٧١٣	١٠٥٠
٥٦٠	٣٦٨	٩٣٠	٨٦٠
٦٢٠	٧٩٣	٦٦٠	٣٩٠
٤٨٠	٣٩٥	٧٦٠	٤٩٥
٦٣٠	١٠٥٦	٤٩٠	٨٢٠

والمطلوب :

(آ) حساب الوسط الحسابي والوسيط والمنوال لهذه البيانات

(ب) إعرض هذه البيانات في جدول توزيع تكراري بخمس فئات طول كل منها ٢٠٠ ومبتدأ بالفئة الأولى (٣٠٠ – ٤٩٩) ومنه اوجد

١- الوسط الحسابي

٢- الوسيط

٣- المنوال

وقارن بين النتيجتين في (آ) و (ب) .

(٢) البيانات التالية تمثل عدد الجوز على ٨ نباتات من القطن ٢٢ ، ١٥ ، ١٢ ، ٢٥ ، ٣٠ ، ٢٢ ، ٢٨ ، ١٨ .

والمطلوب حساب :

(آ) المتوسط الحسابي

(ب) ~~الوسط الهندسي~~

(ج) ~~الوسط التوافقي~~

(د) ~~الوسط التربيعي~~

(و) الوسيط

(ز) المنوال

(هـ) متوسط المدى

(٣) من جدول التوزيع التكراري التالي :

الفئات	التكرار
٣٣–٣٧	١٠
٣٨–٤٢	١٢
٤٣–٤٧	٥١
٤٨–٥٢	٣٠
٥٣–٥٧	٨

احسب بطريقة الرسم

(أ) الوسيط

(ب) المنوال

(٤) من جدول التوزيع التكراري السابق اوجد

(أ) الوسط الهندسي

(ب) الوسط التوافقي

(جـ) الوسط التربيعي

(٥) اذا علمت بأن $\bar{y} = 25$

اوجد الوسط الحسابي لكل من :

(a) $x_i = y_i + 5$

(b) $z_i = 2y_i + 20$

(c) $u_i = (3/5)y_i + 10$

(٦) اذا علمت بأن :

$$X_i = 5y_i + 20$$

وان

$$\bar{x} = 100$$

فما هو الوسط الحسابي لقيم y ؟

(٧) الجدول التالي يبين نتائج امتحان ثلاثة شعب في الصف الأول بمادة الاحصاء

اسم الشعبة	عدد الطلبة	معدل درجاتهم
أ	٣٥	٧٨
ب	٢٥	٧٥
جـ	٣٠	٨٢

احسب الوسط الحسابي لجميع الشعب بمادة الاحصاء هذه .

(٨) الجدول التالي يبين توزيع عدد الاسر تبعا لعدد الأفراد بالاسرة :

عدد أفراد الأسرة	التكرار
١	٦٢٠
٢	١٠٠٠
٣	١٤٢٠
٤	٢١٠٠
٥	٨٠٠
٦	٤٦٠
	٦٤٠٠

احسب :

(أ) الوسط الحسابي

(ب) الوسيط

(ج) المنوال

(٩) من الجدول التوزيع التكراري التالي لمحصول صنفين من القطن :

فئات المحصول	تكرار الصنف الأول	تكرار الصنف الثاني
٢٠-٢٥	٥	١
٢٦-٣١	١٨	٢٥
٣٢-٣٧	٢٥	٥٠
٣٨-٤٣	٤٠	٢٥
٤٤-٤٩	٣	٢٠
٥٠-٥٥	٢٢	١٩
	~~١٤٠~~ ١١٣	١٤٠

والمطلوب :

(أ) مقارنة الوسط الحسابي لكلا الصنفين

(ب) حساب الوسيط للصنف الأول بطريقة الرسم

(ج) حساب المنوال للصنف الثاني بطريقة الرسم

(١٠) البيانات التالية تمثل الاجور الاسبوعية لعمال اربعة مصانع

المصنع	عدد العمال	متوسط الاجورالاسبوعية (دينار)
١	٢٠٠	١٢
٢	٢٥٠	٨
٣	١٥٠	١٥
٤	٤٠٠	٧

والمطلوب : ايجاد متوسط الاجر الاسبوعي للعمال في جميع المصانع .

الفصل الخامس

مقاييس التشتت أو الاختلاف

Measures of Dispersion or Variation

(٥ : ١) مقدمــة

يقصد بالتشتت أو الاختلاف بأنه التباعد أو التقارب الموجود بين قيم المشاهدات التابعة لمتغير ما، ومقاييس التشتت هي مقاييس لمدى تشتت قيم المشاهدات عن وسطها .

هذا وكلما كان مقياس التشتت كبيرا دلّ ذلك على عدم التجانس بين القيم . ويكون مقياس التشتت صغيرا عندما تكون الاختلافات بين قيم المشاهدات قليلة .وقد سبق لنا ان ذكرنا بأن مقاييس التوسط أو التمركز السابقة تعطينا فكرة عن مكان تمركز قيم المشاهدات بينما نلاحظ ان مقاييس التشتت تعطينا فكرة عن مدى تجانس اوتباين هذه القيم حول مركزها . أي درجة انتشارها .

ان لمقاييس التشتت أهميتها في وصف التوزيعات ومقارنتها مع بعضها.حيث ان مقاييس التوسط وحدها لا تكفي لهذا الغرض،فقد يتساوى الوسط الحسابي لمجموعتين من القيم مثلا بينما يختلف مدى انتشار قيم المجموعة الأولى عن انتشار قيم المجموعة الثانية كما يتضح من مقارنة المجموعتين التاليتين :

المجموعة الأولى : ١٧ ، ٢٠ ، ٢٣ ، ١٨ ، ١٩ ، ٢١ ، ٢٢ .

المجموعة الثانية : ٣٥ ، ١٥ ، ٧ ، ٥ ، ٤٥ ، ٢٠ ، ١٣ .

فالوسط الحسابي لكل من المجموعتين هو ٢٠ ولكن المجموعة الأولى تبدو أكثر تجانسا .

ولمقاييس التشتت أهميتها في تطبيق نظرية العينات والاستنتاج الاحصائي واختبار الفرضيات كما سيأتي شرحه في الفصول القادمة .

هذا وهناك عدة مقاييس للتشتت أهمها :

(أولا) مقاييس التشتت المطلق :

أي التي تكون وحداتها نفس وحدات القيم الاصلية وأهمها :

تشتت مطلق

(١) المدى The Range.

(٢) الانحراف المتوسط The Mean Deviation.

(٣) التباين والانحراف القياسي The Variance and The Standard Deviation

(ثانيا) مقاييس التشتت النسبي

أي التي تكون خالية من وحدات القياس وأهمها:

تشتت نسبي

معامل الاختلاف Coefficient of Variation

خالية من وحدات لا قياس

(٥ : ٢) مقاييس التشتت المطلق

(١) المدى The Range

تعريف : (٥ : ١)

المدى لمجموعة من القيم هو الفرق بين أعلى قيمة وأقل قيمة في تلك المجموعة ويرمز له بـ R

$$R = y_{max} - y_{min}.$$

مثال (١) اوجد المدى لكل من المجموعات التالية :

(a) $y_i = 12, 6, 7, 3, 15, 10, 18, 5$

(b) $y_i = 9, 3, 8, 8, 9, 8, 9, 18$

الحـل :

(a) $R = y_{max} - y_{min}.$

$= 18 - 3 = 15$

(b) $R = 18 - 3 = 15$

ان المدى في كلا المجموعتين متساوٍ ولكننا نلاحظ حقيقة أن الاختلاف في المجموعة (a) أكبر منه في المجموعة (b) لأن قيم المجموعة (b) تتألف معظمها من ٨ و ٩.

لذلك فإن (المدى) يكون احياناً مضللا لأنه يعتمد فقط على القيمتين الطرفيتين (بعد ترتيب القيم تصاعدياً أو تنازلياً) اللتين كثيراً ماتكونان شاذتين .

هذا ومن الصعب حساب المدى الحقيقي من جدول توزيع تكراري لعدم معرفة القيمتين الطرفيتين .

(٢) الانحراف المتوسط : The Mean Deviation

(أ) البيانات غير مبوبة :

تعريف (٥ : ٢)

اذا كان لدينا n من المشاهدات $y_1, y_2, \ldots\ldots, y_n$ فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة (أي باهمال الاشارة) عن وسطها الحسابي ويرمز له بـ M.D. أي أن :

$$M.D. = \frac{\sum|y_i - \overline{y}|}{n}$$

وان السبب في أخذ الانحرافات المطلقة هو ان ابقاء الاشارات الموجبة والسالبة يجعل مجموع الانحرافات صفرا حيث اننا ذكرنا سابقاً بأن: $\sum(y_i - \overline{y}) = 0$ دائماً .

مثال (٢) أوجد الانحراف المتوسط للقيم التالية :

$$y_i = 9, 8, 6, 5, 7$$

الحل :

y_i	$y_i - \overline{y}$	$\|y_i - \overline{y}\|$
9	2	2
8	1	1
6	-1	1
5	-2	2
7	0	0
$\sum y_i = 35$ $\overline{y} = 7$	0	6

$$\therefore \quad M.D. = \frac{\sum|y_i - \overline{y}|}{n} = \frac{6}{5}$$

$$= 1.2$$

تعريف : (٣ : ٥)

اذا 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
فان الانحراف المتوسط هو :

$$\text{M.D.} = \frac{\sum f_i |y_i - y|}{\sum f_i}$$

مثال (٣) اوجد الانحراف المتوسط لجدول التوزيع التكراري (٣ : ١)

الحـــل :

الفئات	f_i	y_i	$f_i y_i$	$\lvert y_i - \bar{y} \rvert$	$f_i \lvert y_i - \bar{y} \rvert$
60 – 62	5	61	305	6·45	32·25
63 – 65	18	64	1152	3·45	62·10
66 – 68	42	67	2814	0·45	18·90
69 – 71	27	70	1890	2·55	68·85
72 – 74	8	73	584	5·55	44·40
	100		6745		226·50

$$\bar{y} = \frac{\sum f_i y_i}{\sum f_i} = \frac{6745}{100} = 67.45$$

$$\text{M.D.} = \frac{\sum f_i |y_i - \bar{y}|}{\sum f_i} = \frac{226.5}{100} = 2.265$$

(٣) التباين والانحراف القياسي Variance and Standard Deviation

لكي يمكننا التغلب على مشكلة الاشارات عند جمع الانحرافات والتي تؤدي دائماً لأن يكون مجموع انحرافات أي عينة عن وسطها الحسابي يساوي صفراً ، وبدلاً من أخذ القيم المطلقة للانحرافات أي بدون اشارات كما سبق في الجزء السابق فاننا نستطيع أن نتغلب على ذلك بطريقة اخرى وهي بتربيع قيم الانحرافات وبذلك تصبح جميعها موجبة ، أي نحصل على مجموع مربعات الانحرافات Sum of Squares والتي يرمز لها (SS) وعلى ذلك فان :

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

ولكي نأخذ في الاعتبار حجم العينة حتى يمكن مقارنة العينات المختلفة الأحجام فاننا نقسم مجموع مربعات الانحرافات على درجات الحرية (n − 1) . وبذلك نحصل على ما يسمى بالتباين (S^2) .

تعريف (٥ : ٤)

اذا كان لدينا n من المشاهدات $y_i = y_1, y_2, \ldots\ldots, y_n$ فإن التباين (ويرمز له S^2) يكون:

$$S^2 = \frac{\sum(y_i - \bar{y})^2}{n-1} = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$$

ويلاحظ ان القانون السابق هو لحساب تباين العينة أما اذا كانت قيم المشاهدات تمثل المجتمع كله فإن التباين (ويرمز له في هذه الحالة σ^2 وتلفظ باللاتيني Sigma Square) يحسب كما يلي :

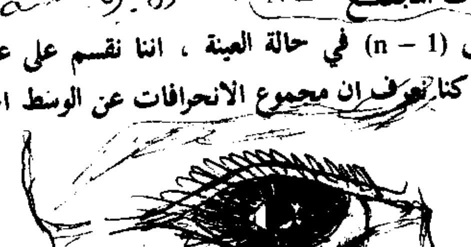
$$\boxed{\sigma^2 = \frac{\sum(y_i - u)^2}{N}}$$

حيث أن :

الوسط الحسابي للمجتمع = μ

عدد مفردات المجتمع = N

والسبب اننا نقسم على (n − 1) في حالة العينة ، اننا نقسم على عدد القيم الحرة أي على درجات الحرية ولما كنا نعرف ان مجموع الانحرافات عن الوسط الحسابي يساوي



صفراً لذلك فعند سحب عينة فان $(n - 1)$ من المشاهدات هي قيم حرة. أما المشاهدة الأخيرة فلابد ان يكمل انحرافها مجموع الانحرافات عن الوسط الحسابي الى الصفر أي ان احدى المشاهدات تثبت بمعرفة انحرافات $(n - 1)$ من المشاهدات ، وعلى ذلك فان عدد القيم الحرة في أية عينة هي $(n - 1)$ وهي ما سميناه بدرجات الحرية .

ونظراً لاننا عند حساب التباين قد قمنا بتربيع الانحرافات ، فان قيمة التباين تكون مقاسة بمربع الوحدات المستخدمة في قياس المشاهدات ، فاذا كانت المشاهدات مقاسة بالسنتيمتر فان التباين يكون مقاساً بالسنتيمتر المربع . ولا توجد مشكلة في ذلك ، ولكن المشكلة تظهر عندما يكون مربع الوحدات غير ذي معنى او غير مقبول .

فمثلاً اذا كانت المشاهدات عبارة عن أوزان بالكيلوغرام او عبارة عن مبالغ بالدينار او عبارة عن أعداد عمال مثلاً او عدد الاطفال في الأسر المختلفة فان التباين يكون عندئذ مقاساً بالكيلو غرام المربع او الدينار المربع او العامل المربع او الطفل المربع وهذه كلها غير ذات معنى .

وكحل لذلك ولكي نرجع وحدات القياس الى أصلها فاننا نأخذ الجذر التربيعي للتباين لنحصل على قيمة (S) (أي ان $S = \sqrt{S^2}$) وهو ما يسمى بالانحراف القياسي والذي يكون مقاساً بالوحدات الاصلية أي في الحالات السابقة يكون مقاساً بالسنتيمتر او الكيلوغرام او الدينار او العامل او الطفل وهكذا .

تعريف (٥ : ٥) :

الانحراف القياسي S لعينة ما هو الجذر التربيعي لتباين تلك العينة :

$$S = \sqrt{\frac{\Sigma (y_i - \bar{y})^2}{n - 1}} = \sqrt{\frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n - 1}}$$

ويكون الانحراف القياسي للمتجمع σ (Sigma) هو :

$$\sigma = \sqrt{\frac{\sum (y_i - u)^2}{N}}$$

ولكي نثبت ان معادلتي التباين (أو الانحراف القياسي) السابق ذكرهما متساويتان نتبع الخطوات التالية :

$$S^2 = \frac{\sum(y_i - \overline{y})^2}{n - 1}$$

حيث أن :

مجموع مربعات الانحرافات $= \sum(y_i - \overline{y})^2$

وتسمى للاختصار مجموع المربعات Sum of Squares ويرمز لها SS

$$\therefore SS = \sum(y_i - \overline{y})^2$$
$$= \sum(y_i^2 - 2\overline{y}y_i + \overline{y}^2)$$
$$= \sum y_i^2 - 2\overline{y}\sum y_i + n(\overline{y})^2$$
$$= \sum y_i^2 - 2\left(\frac{\sum y_i}{n}\right)\left(\sum y_i\right) + n\frac{(\sum y_i)^2}{n^2}$$
$$= \sum y_i^2 - \frac{2(\sum y_i)^2}{n} + \frac{(\sum y_i)^2}{n}$$
$$= \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

وعليه فان التباين يساوي :

$$S^2 = \frac{SS}{n-1} = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$$

وهذه هي الطريقة المختصرة لايجاد التباين (أو نأخذ جذرها لايجاد الانحراف القياسي).

مثـال(٤) البيانات التالية تبين كمية المحصول / للقطعة (كغم) للقطن في خمس مزارع. احسب الانحراف القياسي لها

$$y_i = 9, 8, 6, 5, 7$$

الحـل :

(١) الطريقة المطولة

y_i	$y_i - \overline{y}$	$(y_i - \overline{y})^2$
9	2	4
8	1	1
6	−1	1
5	−2	4
7	0	0
$\sum y_i = 35$ $\overline{y} = 7$	0	10

$\overline{Y} = \frac{\sum Y_i}{n}$

$$S = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{10/4} = \sqrt{2\cdot 5} = 1\cdot 58 (kgm)$$

(٢) الطريقة المختصرة

y_i	y_i^2
9	81
8	64
6	36
5	25
7	49
$\sum y_i = 35$	$\sum y_i^2 = 255$

$$SS = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= 255 - \frac{(35)^2}{5} = 10$$

$$\therefore S = \sqrt{\frac{10}{4}} = \sqrt{2\cdot 5} = 1\cdot 58 \text{ kgm.}$$

أما التباين لهذه القيم فهومربع الانحراف القياسي أي نرفع الجذر :

$$\therefore S^2 = \frac{10}{4} = 2\cdot 5 \, (kgm)^2$$

(ب) البيانات مبوبـة

تعريف (٥ : ٦) :

اذا كانت $y_1, y_2, \ldots, y_k$ تمثل مراكز الفئات في جدول التـوزيع التكراري وان تكراراتها هي $f_1, f_2, \ldots, f_k$ على التوالي فان الانحرافالقياسي لهـا هو :

$$S = \sqrt{\frac{\sum f_i (y_i - \bar{y})^2}{\sum f_i - 1}} = \sqrt{\frac{\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{\sum f_i}}{\sum f_i - 1}}$$

مثـال(٥) أحسب الانحراف القياسي والتباين لجدول التوزيع التكراري (٣ : ١) :

الحـل :

(١) الطريقة المطولة

الفئات	f_i	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$f_i(y_i - \bar{y})^2$
60 – 62	5	61	– 6·45	41·6025	208·0125
63 – 65	18	64	– 3·45	11·9025	214·2450
66 – 68	42	67	– 0·45	0·2025	8·5050
69 – 71	27	70	2·55	6·5025	175·5675
72 – 74	5	73	5·55	30·8025	246·4200
	100				852·7500

فمجموع المربعات SS هو :

$$SS = \sum f_i(y_i - \bar{y})^2 = 852{\cdot}7500$$

أما التباين فهو :

$$S^2 = \frac{SS}{\sum f_i - 1} = \frac{852{\cdot}7500}{99} = 8{\cdot}6$$

أما الانحراف القياسي فهو :

$$S = \sqrt{S^2} = \sqrt{8{\cdot}6} = 2{\cdot}9$$

(٢) الطريقة المختصرة

الفئات	f_i	y_i	$f_i y_i$	y_i^2	$f_i y_i^2$
60 – 62	5	61	305	3721	1860[illegible]
63 – 65	18	64	1152	4096	73728
66 – 68	42	67	2814	4489	188538
69 – 71	27	70	1890	4900	132300
72 – 74	5	73	584	5329	42632
	100		6745		4558[illegible]

$$SS = \sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{\sum f_i} = 455803 - \frac{(6745)^2}{100} = 852{\cdot}75$$

$$\therefore S^2 = \frac{SS}{\sum f_i - 1} = \frac{852{\cdot}75}{99} = 8{\cdot}6$$

$$\therefore S = \sqrt{S^2} = \sqrt{8{\cdot}6} = 2{\cdot}9$$

(جـ) أهم خواص التباين والانحراف القياسي

(١) عند اضافة (أو طرح) عدد ثابت (k) الى كل قيمة من قيم المشاهدات فان قيمة التباين والانحراف القياسي لايتغيران أي:

التباين للقيم الجديدة = التباين للقيم الاصلية

الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية

أي اذا كان

$$x_i = (y_i + k) \quad , \quad x_i = (y_i - k)$$

فان

$$S_x^2 = S_y^2$$

$$\therefore S_x = S_y$$

مثال (٦) احسب التباين والانحراف القياسي للقيم التالية ثم إضف لكل منها ٣ واحسب التباين والانحراف القياسي للقيم الجديدة

القيم الاصلية: $y_i = 8, 3, 2, 12, 10$

القيم الجديدة بعد اضافة ٣ لكل قيمة : $x_i = 11, 6, 5, 15, 13$

y_i	y_i^2
8	64
3	9
2	4
12	144
10	100
35	321

x_i	x_i^2
11	121
6	36
5	25
15	225
13	169
50	576

بسم الله الرحمن الرحيم

$$SS_x = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= 576 - \frac{(50)^2}{5}$$

$$= 76$$

$$\therefore S_x^2 = \frac{SS_x}{n-1}$$

$$= \frac{76}{4} = 19$$

$$\therefore S_x = \sqrt{S_x^2}$$

$$= \sqrt{19}$$

$$SS_y = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= 321 - \frac{(35)^2}{5}$$

$$= 76$$

$$\therefore S_y^2 = \frac{SS_y}{n-1}$$

$$= \frac{76}{4} = 19$$

$$\therefore S_y = \sqrt{S_y^2}$$

$$= \sqrt{19}$$

أي ان التباين أو لانحراف القياسي لم يتأثر باضافة العدد الثابت الى كل قيمة من القيم الاصلية .

هذا واذا طرحنا (5) من كل من القيم الاصلية في المثال السابق فان التباين والانحراف القياسي لن يتأثرا أيضاً .

> (٢) اذا ضربت كل قيمة من قيم المشاهدات بعدد ثابت (k) فان :
> التباين للقيم الجديدة = التباين للقيم الاصلية × مربع العدد الثابت
> الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية × العدد الثابت

أي اذا كان : $x_i = k y_i$

فان : $S_x^2 = k^2 S_y^2$

كما ان : $\therefore S_x = K S_y$

هذا ويمكن التعبير عن الخاصيتين السابقتين بما يلي

فاذا كان : $x_i = k + c y_i$

$$③\ \bar{x} = \frac{\sum k - \sum c y_i}{n} = \frac{nk - c\sum y_i}{n}$$

$$④\ \bar{x} = k + c\frac{\sum y_i}{n}$$

$$⑤\ \bar{x} = k + c\bar{y}$$

حيث ان k , c ثوابت ؛

فان :

$$S_x^2 = c^2 S_y^2$$
$$S_x = c S_y$$

البرهان :

$$①\ x_i = k + c y_i \qquad ②\ \frac{\sum x_i}{n} = \frac{\sum (k + c y_i)}{n}$$

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

$$\therefore x_i - \bar{x} = (k + c y_i) - \bar{x}$$

بالتعويض

$$= (k + c y_i) - (k + c\bar{y})$$
$$= c y_i - c\bar{y}$$
$$\therefore (x_i - \bar{x}) = c(y_i - \bar{y})$$

وبتربيع كلا الطرفين فان :

$$(x_i - \bar{x})^2 = c^2(y_i - \bar{y})^2$$

وعند ادخال $\sum$ على كلا الطرفين ينتج :

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

وبقسمة كلا الطرفين على $n-1$ ينتج :

$$\frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{c^2\sum(y_i - \bar{y})^2}{n-1}$$

$$\therefore S_x^2 = c^2 S_y^2$$

وكذلك فان :

$$S_x = cS_y$$

(٣) اذا كان كل من x و y متغيرين مستقلين وكان المتغير z يساوي مجموعهما أي :

$$z_i = x_i + y_i$$

فان تباين z = تباين x + تباين y أي :

$$S_z^2 = S_x^2 + S_y^2$$

QUART2 ② $S_p^2 = \dfrac{(n_1-1)\times\frac{SS_1}{(n_1-1)} + (n_2-1)\frac{SS_2}{(n_2-1)}}{(n_1-1)+(n_2-1)}$

(٤) اذا كانت مجموعتان من القيم مؤلفة من n_1 و n_2 من المشاهدات ولها تباين S_1^2 و S_2^2 على التوالي فان التباين المتجمع لجميع المشاهدات (pooled variance) هـو :

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

وهذا ما يسمى بالتباين الموزون أو المرجح ويمكن كتابته بالصيغة التالية :

$$S_p^2 = \frac{SS_1 + SS_2}{n_1+n_2-2}$$

العلاقة بين الانحراف القياسي والانحراف المتوسط

اذا كان التوزيع غير متماثل (ملتوي التواء بسيطاً)

فان :

الانحراف المتوسط = $\frac{٤}{٥}$ (الانحراف القياسي) أي :

$$M.D. = \frac{4}{5}S$$

ملاحظة : عند قياس مدى تشتت متوسطات العينات التابعة لمجتمع ما فانه يستخدم ما يسمى بالخطأ القياسي Standard Error او الانحراف القياسي للمتوسطات Standard Deviation of the Mean ويرمز له بـ $S_{\bar{y}}$ ويحسب بالقانون التالي :

$$S_{\bar{y}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

وللخطأ القياسي أهمية كبرى في الاستنتاج الاحصائي كما سيأتي ذكره في الفصول القادمة .

(٤) مقاييس اخرى للتشتت المطلق :

هناك بعض المقاييس الأخرى لقياس التشتت المطلق ولكنها قليلة الاستعمال وتسمى شبيهات المدى (Quasi–ranges)

مثـــل :

(أ) نصف المدى الربيعي Semi–interquartile Range

(أو الانحراف الربيعي) ويرمز له بـ Q :

$$Q = \frac{Q_3 - Q_1}{2}$$

حيث ان

الربيع الأول $= Q_1$

الربيع الثالث $= Q_3$

(ب) نصف المدى العشيري أو المئيني

فمثلا :

$$\text{The Semi (10–90)\% Range} = \frac{P_{90} - P_{10}}{2}$$

(٥ : ٣) مقاييس التشتت النسبي

$C.V = \frac{S}{\bar{x}} \times 100$

ان مقاييس التشتت النسبي لها أهميتها عند مقارنة تشتت مجموعتين أو أكثر تختلف في وحدات القياس لقيمها . لأن مقاييس التشتت النسبي تكون خالية من وحدات القياس وأهم مقاييس التشتت النسبي هي :

(١) معامل الاختلاف Coefficient of Variation

> **تعريف (٥ : ٧) :**
>
> اذا كان S و $\bar{y}$ هما الانحراف القياسي والوسط الحسابي لمجموعة من القيم على التوالي فان معامل الاختلاف لها (ويرمزله بـ C.V.)
>
> هـو :
>
> $$C.V. = \frac{S}{\bar{y}} \times 100$$

مثـال (٧) :نتائج الامتحانات النهائية لدرسي الاحصاء والكيمياء للصف الأول كانا كالآتـي :

معامل الاختلاف $C.V = \frac{S}{\bar{X}} \times 100$

	الاحصاء	الكيمياء
الوسط الحسـابي	٧٨	٧٣
الانحراف القياسي	٨	٧٫٦

ففي اي الموضوعين كان تشتت الدرجات أكثر ؟

الحـل :

$$C.V. = \frac{S}{\bar{y}} \times 100$$

للاحصاء $= \frac{8}{78} \times 100 = 10{\cdot}25\ \%$

للكيمياء $= \frac{7{\cdot}6}{73} \times 100 = 10{\cdot}41\ \%$

أي أن التشتت لدرجات الكيمياء كان أكثر

لاحظ بأنه لو قارنا التشتت بمقياس الانحراف القياسي لكان التشتت أكبر في الاحصاء عنه في الكيمياء

مثـال (٨) : أجريت تجربة لدراسة طول النبات (سم) وكمية المحصول (كغم) لـ ١٥٠ نباتا من الذرة فكانت النتائج كالآتـي :

	الطول	كمية المحصول
الوسط الحسابي :	٢٠٠	٨٠٠
الانحراف القياسي :	١٦,	٣٦

قارن بين تشتت الصفتين :

الحـل :

$$C.V. = \frac{S}{\bar{y}} \times 100$$

بالنسبة للطول

$$C.V. = \frac{16}{200} \times 100 = 8\%$$

بالنسبة للمحصول

$$C.V. = \frac{36}{800} \times 100 = 4.5\%$$

اذن التشتت كان أكبر في صفة الطول .

(٢) وهناك مقاييس اخرى للتشتت النسبـي لأنه في الحقيقة

$$\text{معامل الاختلاف} = \frac{\text{مقياس من المقاييس المطلقة}}{\text{مقياس من مقاييس التوسط}} \times ١٠٠$$

فمثلا :

(أ) في حالة استخدام الانحراف القياسي فان

$$C.V. = \frac{S}{\bar{y}} \times 100$$

كما ذكر سابقا

(ب) في حالة استخدام المدى الربيعي فان

$$C.V. = \frac{Q_3 - Q}{Q_3 + Q_1}$$

ويسمى بمعامل الاختلاف الربيعي

(جـ) وفي حالة استخدام الانحراف المتوسط

$$C.V. = \frac{M.D}{\bar{y}} \times 100$$

أو

$$C.V. = \frac{M.D.}{Me} \times 100$$

$Z_i = \frac{X - \bar{X}}{\sigma}$

(٥ : ٤) الدرجة القياسية Standardized Scores

في كثير من الاحيان نحتاج الى مقارنة مفردتين من مجموعتين مختلفتين . وفي هذه الحالة يجب تحويل وحدات كل مفردة الى وحدات قياسية حتى تكون المقارنة ذات معنى وذلك باستخدام الوسط الحسابي والانحراف القياسي لكل مجموعة .

تعريف (٥ : ٨) :
تسمى القيمة Z_i درجة قياسية اذا كانت تساوي $Z_i = \frac{y_i - \bar{y}}{S}$

ومن هذا يتضح بأن الدرجات القياسية خالية من الوحدات المستخدمة في القياس .

هذا واذا حولنا جميع قيم مجموعة ما الى درجات قياسية فان الوسط الحسابي لهذه الدرجات القياسية يساوي صفرا وان تباينها يساوي ١ :

أي أن $Z_i \sim N(0,1)$

أي أن Z_i تتوزع طبيعيا بوسط حسابي لها = صفر وتباين = ١ .

مثـال (٩): حصل طالب على درجة ٨٤ في الامتحان النهائي بالرياضيات علما بان الوسط الحسابي في امتحان الرياضيات لجميع الطلبة كان ٧٦ وبانحراف قياسي قدره (١٠) . أما في امتحان الفيزياء فقد حصل نفس الطالب على درجة ٩٠ حيث كان الوسط الحسابي في امتحان الفيزياء لجميع الطلبة = ٨٢ والانحراف القياسي = ١٦ .

ففي اي الموضوعين كانت قابلية هذا الطالب أعلى ؟

$Z_i = \frac{X - \bar{X}}{\sigma}$

الحـل :

عند مقارنة درجة الامتحانين مباشرة نجد أن درجته في الفيزياء (٩٠) أعلى من درجته في الرياضيات (٨٤) .

ولكن عند تحويل هاتين الدرجتين الى درجات قياسية نجد أن :

$$Z_i = \frac{y_i - \bar{y}}{S}$$

بالنسبة للرياضيات :

$$Z = \frac{84 - 76}{10} = 0.8$$

أما بالنسبة للفيزياء

$$Z = \frac{90 - 82}{16} = 0.5$$

ومن هذا يتضح بأن قابليته في الرياضيات أعلى مما في الفيزياء وهو عكس ما توصلت اليه المقارنة السابقة .

تمارين الفصل الخامس

(١) لكل من المجموعات التالية أوجد :

(أ) المدى

(ب) التباين

(ج) الأنحراف القياسي

(د) الأنحراف القياسي للوسط الحسابي

(هـ) الأنحراف المتوسط

(و) معامل الأختلاف

(a) $y_i = 2, 5, 9, 11, 13$
(b) $y_i = 4, 10, 2, 8, 4, 14, 10, 12, 8$
(c) $y_i = -4, 2, -6, 0, -4, 6, 2, 4, 0$
(d) $y_i = 16, 2, 22, 8, 6, 20, 24, 14$

(٢) البيانات التالية يبين جدول التوزيع التكراري للأجور الاسبوعية لعمال مصنع السمنت في الموصل :

الفئات	التكرار
١٦–٢١	١٥
٢٢–٢٧	١٦
٢٨–٣٣	٢٥
٣٤–٣٩	٢٠
٤٠–٤٥	١٠

والمطلوب ايجاد :

(١) التباين والأنحراف القياسي
(٢) الأنحراف المتوسط
(٣) الخطأ القياســي
(٤) معامل الأختلاف

(٣) في كل من الأفرع التالية أوجد الكمية المفقودة :

	$\bar{y}$	S	C.V.
(a)		10	20%
(b)	50	30	
(c)	25		5%

(٤) اذا علمت بأن قيم y_i لها وسط حسابي قدره ٦ بتباين قدره ١٠ ما هو الوسط الحسابي والتباين لكل مما يأتي

(a) $x = y + 5$
(b) $x = 3y$
(c) $x = 3y + 5$

(٥) اذا علمت بأن

$z = 5y + 20$
$\bar{z} = 100$
$S_z^2 = 40$

فما هو الوسط الحسابي والتباين لقيم y ؟

(٦) من البيانات التالية :

(a) $y_i = 2, 5, 8, 11, 14$
(b) $x_i = 2, 8, 14$

المطلوب أيجــاد :

(أ) الوسط الحسابي والتباين لكل مجموعة
(ب) الوسط الحسابي لجميع قيم x و y مجتمعة
(ج) التباين للمجموعتين مجتمعة

(٧) (أ) برهن بأن التباين للقيم التالية (متوالية عددية)

$y_i = a, a + d, a + 2d, \ldots\ldots, a + (n-1)d$

هــو

$$S^2 = \frac{1}{12}(n^2 - 1)d^2$$

(ب) أستخدم النتيجة أعلاه لأيجاد التباين للقيم التالية

$y_i = 4, 10, 16, 22, \ldots\ldots, 154$

(٨) (أ) اذا كانت قيم مجموعة من الأرقام مؤلفة من ١ أو الصفر فاذا كان نسبة الـ ١ في المجموعة $= p$

ونسبة الصفر في المجموعة $= q$

فان الأنحراف القياسي لهذه المجموعة هو $S = \sqrt{pq}$

(ب) أوجد التباين والأنحراف القياسي للقيم التالية :

$y_i = 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1,$

باستخدام

(١) الطريقة العادية $S = \sqrt{\dfrac{SS}{n-1}}$

(٢) أو $S = \sqrt{pq}$

(٩) قارن بين تشتت درجات الأحصاء بالنسبة للطلاب والطالبات في الجدول التالي :

فئات الدرجات	عدد الطلاب	عدد الطالبات
٣١ - ٤٠	١	١
٤١ - ٥٠	٦	٤
٥١ - ٦٠	٨	٧
٦١ - ٧٠	٣٧	١٦
٧١ - ٨٠	٣٣	٤٠
٨١ - ٩٠	١٢	٢٢
٩١ - ١٠٠	٣	١٠

(١٠) (أ) حول القيم التالية الى درجات قياسية

$y_i = 6, 2, 8, 7, 5$

(ب) برهن بأن الوسط الحسابي للدرجات القياسية = صفر

وأن التباين لها = ١

(١١) (أ) حول درجات الأمتحان في الجدول التالي الى درجات قياسية

الدرجات	عدد الطلبة
٣٠-٣٩	١
٤٠-٤٩	٣
٥٠-٥٩	١١
٦٠-٦٩	٢١
٧٠-٧٩	٤٣
٨٠-٨٩	٣٢
٩٠-٩٩	٩

(ب) أرسم بيانيا هذه الدرجات القياسية مع رسم Relative frequency curve معاً .

الفصل السادس

مقاييس الالتواء Measures of Skewness

ومقاييس التفلطح Measures of Kurtosis

(٦-١) مقدمـــة :

تكلمنا في الفصول السابقة عن مقياسين من المقاييس الوصفية التي تصف التوزيعات التكرارية وهما مقاييس التوسط (لايجاد القيمة المتوسطة التي تتمركز حولها مفردات التوزيع) ومقاييس التشتت أو الاختلاف (لايجاد درجة تشتت هذه المفردات حول تلك القيمة المتوسطة) .

والآن سنأخذ مقاييس اخرى تحدد شكل المنحني من حيث التماثل أو الالتواء Skewness من جهة وتدبب القمة أو تفلطحها Kurtosis من جهة أخرى . وقبل شرح هذه المقاييس سنتكلم باختصار عن ما يسمى بالعزوم Moments لأنها تدخل في تقدير هذه المقاييس .

(٦-٢) العـــزوم Moments

(١) العزم الرائي حول الصفر The rth moment about zero

(أ) بيانات غير مبوبة :

> **تعريف : (٦ : ١)**
>
> اذا كان لدينا n من المشاهدات التابعة للمتغير y
> فأن العزم الرائي حول الصفر هو
>
> $$\overline{y^r} = \frac{\sum y_i^r}{n}$$

فالعزم الأول حول الصفر هو الوسط الحسابي أي :

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

أما العزم الثاني حول الصفر فهو :

$$y^2 = \frac{\sum y_i^2}{n}$$

وهكـــذا

(ب) بيانات مبوبــة

تعريف : (٦ : ٢)
اذا كانت $y_1, y_2 \dots y_n$ تمثل مراكز الفئات في جدول التوزيع التكراري مع تكراراتها $f_1, f_2 \dots f_n$ على التوالي فأن العزم الرائي حول الصفر هو

$$\bar{y}^r = \frac{\sum f_i y_i^r}{\sum f_i}$$

(٢) العزم الرائي حول الوسط الحسابي The rth moment about the mean

(أ) بيانات غير مبوبة

العزم الرائي حول الوسط الحسابي هو :

$$m_r = \frac{\sum (y_i - \bar{y})^r}{n}$$

فإذا كانت $r = 1$ فأن $m_1 = 0$

$$m_1 = \frac{\sum (y_y - \bar{y})}{n} = 0$$

أما اذا كانت $r = 2$ فأن ...

$$m_2 = \frac{\sum (y_y - \bar{y})^2}{n} =$$

(ب) بيانات مبوبة :

العزم الرائي حول الوسط الحسابي هو

$$m_r = \frac{\sum f_i (y_i - \bar{y})^r}{\sum f_i}$$

مثال (١) : اوجد العزم الأول والثاني والثالث (حول الصفر) للبيانات التالية

$$y_i = 4, 7, 5, 9, 8, 3, 6$$

الحـــل :

$$\bar{y}^r = \frac{\Sigma y_i^r}{n}$$

فالعزم الأول هو

$$\bar{y} = \frac{\sum y_i}{n} = \frac{4+7+5+9+8+3+6}{7} = \frac{42}{7} = 6$$

والعزم الثاني هــو

$$\bar{y}^2 = \frac{\sum y_i^2}{n} = \frac{4^2+7^2+5^2+9^2+8^2+3^2+6^2}{7} = 40$$

والعزم الثالث هــو :

$$\bar{y}^3 = \frac{\sum y_i^3}{n} = \frac{4^3+7^3+5^3+9^3+8^3+3^3+6^3}{7} = 288$$

مثــال(٢) : أوجد العزم الأول والثاني والرابع حول الوسط الحسابي للبيانات أعــلاه : -

الحـــل :

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

فالعزم الأول حول الوسط الحسابي = صفر لأن

$$m_1 = \frac{\sum(y_i - \bar{y})}{n} = \frac{0}{n} = 0$$

أما العزم الثاني حول الوسط الحسابي فهــو

$$m_2 = \frac{\sum(y_i - \bar{y})^2}{n} =$$

$$= \frac{(4-6)^2+(7-6)^2+(5-6)^2+(9-6)^2+(8-6)^2-(3-6)^2+(6-6)^2}{7} = 4$$

أما العزم الرابع حول الوسط الحسابي فهو

$$m_4 = \frac{\sum(y_i - \bar{y})^4}{n}$$

$$= \frac{(4-6)^4+(7-6)^4+(5-6)^4+(9-6)^4+(8-6)^4+(3-6)^4+(6-6)^4}{7}$$

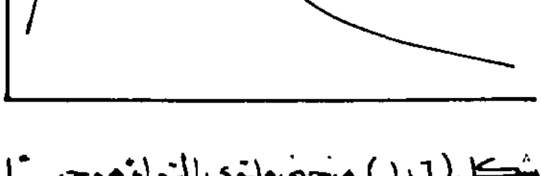
$$= 25 \cdot 86$$

(٦-٣) مقاييس الالتواء Measures of skewness

تعريف : (٦ : ٣)

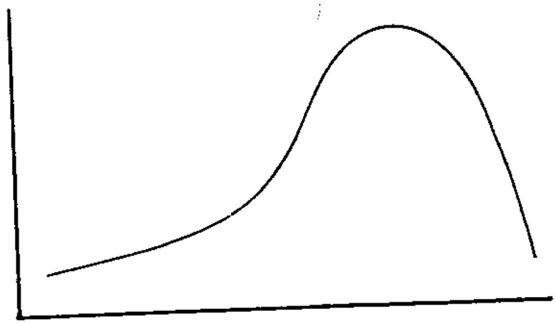
الالتواء هو انحراف منحني التوزيع التكراري عن التماثل وقد يكون الالتواء موجب (أي الالتواء الى اليمين) أو سالب (أي الالتواء الى اليسار) .

ومنحني التوزيع ذو الالتواء الموجب تكون مفرداته متمركزة في الجهة اليسرى (عند الفئات الدنيا) وطرفه يمتد الى اليمين كما في الشكل التالي :



شكل (٦ ، ١) منحني ملتوي التواء موجباً
= نحو اليمين =

أما منحني التوزيع ذو الالتواء السالب فأن مفرداته تتمركز في الجهة اليمنى (عند الفئات العليا) بينما طرفه يمتد الى اليسار كما في الشكل التالي (شكل ٦ : ٢) .



شكل (٦ : ٢) منحني ملتوي التواء سالباً (نحو اليسار)

وأهم مقاييس الالتواء هي :

(١) بأستخدام المنوال .

(٢) بأستخدام الوسيط .

(٣) باستخدام العزوم .

هذا وفي جميع هذه الطرق يكون الالتواء موجباً عندما يكون معامل الالتواء موجبا وسالبا عندما يكون معامل الالتواء سالبا ومتماثلا عندما يكون معامل الالتواء صفرا .

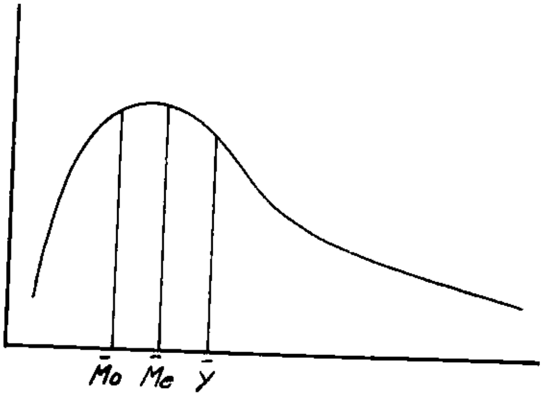
(١) معامل الالتواء الأول بأستخدام المنوال (ويرمز له بـ α_1)

$$\text{فمعامل الالتواء الأول} = \frac{\text{الوسط الحسابي} - \text{المنوال}}{\text{الانحراف القياسي}}$$

أي

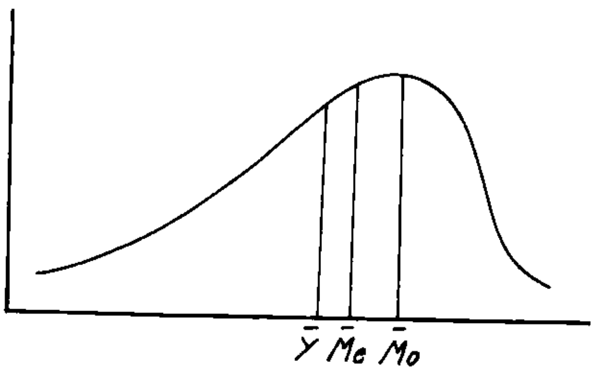
$$\alpha_1 = \frac{(\bar{y} - \bar{M}o)}{S}$$

ومن هذا يتضح بأنه اذا كانت قيمة الوسط الحسابي أكبر من المنوال كانت النتيجة موجبة وبالتالي كان الالتواء موجبا (شكل (٦ : ٣) .



شكل (٦ : ٣) موقع الوسط الحسابي والمنوال والوسيط لمنحني ملتوي التواءً موجبـاً

أما اذا كانت قيمة الوسط الحسابي أقل من المنوال كانت النتيجة سالبة وبالتالي كان الالتواء سالبا (شكل ٦ : ٤)



شكل (٦ : ٤) موقع الوسط الحسابي والمنوال والوسيط لمنحني ملتوي التواءً سالبـاً

مع تحيات د. سلام حسين عويد الهلالي

https://scholar.google.com/citations?user=t1aAacgAAAAJ&hl=en

salamalhelali@yahoo.com

فيس بك... كروب... رسائل وأطاريح في علوم الحياة

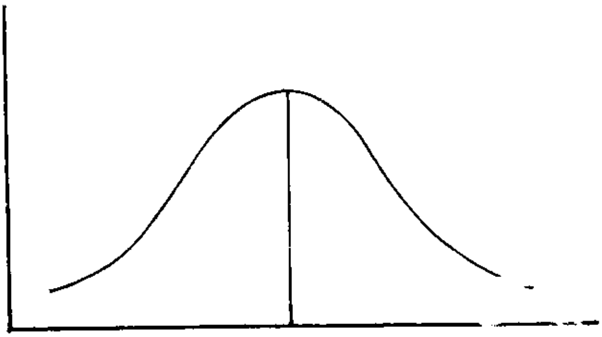
https://www.facebook.com/groups/Biothesis/

https://www.researchgate.net/profile//Salam_Ewaid

https://orcid.org/0000-0001-9734-7331

07807137614

أما اذا تساوت قيمة الوسط الحسابي والمنوال كانت النتيجة تماثل التوزيع (أي $\alpha_1 = 0$) (شكل ٦ : ٥) .



شكل (٦ : ٥) الوسط الحسابي والمنوال لمنحني متماثل (طبيعي)

(٢) معامل الالتواء الثاني باستخدام الوسيط (ويرمز له بـ α_2)

$$\text{معامل الالتواء الثاني} = \frac{\text{الوسط الحسابي} - \text{الوسيط}}{\text{الانحراف القياسي}}$$

أي

$$\alpha_2 = \frac{\bar{y} - \bar{M}e}{S}$$

فاذا كانت قيمة الوسط الحسابي أكبر من الوسيط كانت نتيجة المعامل الالتواء الثاني موجبة وبالتالي كان الالتواء موجبا واذا كانت قيمة الوسط الحسابي أقل من الوسيط كان الالتواء سالبا . أما اذا كان الوسط الحسابي مساويا للوسيط تماما فأن معامل الالتواء الثاني يكون صفرا وبالتالي يكون التوزيع متماثلا

(٣) معامل الالتواء الثالث باستخدام العزوم (ويرمز له α_3)

$$\text{معامل الالتواء الثالث} = \frac{\text{العزم الثالث حول الوسط الحسابي}}{\text{مكعب الجذر التربيعي للعزم الثاني حول الوسط الحسابي}}$$

أي
$$\alpha_3 = \frac{m_3}{\left(\sqrt{m_2}\right)^3} = \frac{m_3}{\sqrt{m_2^3}}$$

فإذا كانت $\alpha_3 = 0$ دلّ ذلك علَى وجود تماثل في التوزيع .

أما اذا كانت قيمة m_3 موجبة فالتوزيع ملتوي التواء موجباً أما اذا كانت سالبة فالتوزيع ملتو التواء سالبا .

مثـال (٣) : احسب معامل الالتواء الأول والثاني في جدول التوزيع التكراري التالي :

الفئـــات	التكـــرار
٥٩,٩٩ – ٥٠,٠	٨
٦٩,٩٩ – ٦٠,٠	١٠
٧٩,٩٩ – ٧٠,٠	١٦
٨٩,٩٩ – ٨٠,٠	١٤
٩٩,٩٩ – ٩٠,٠	١٠
١٠٩,٩٩ – ١٠٠,٠	٥
١١٩,٩٩ – ١١٠,٠	٢
المجمــوع	٦٥

٥٤,٩٩٥
٦٤,٩٩٥
٧٤,٩٩٥
٨٤,٩٩٥
٩٤,٩٩٥
104,995
114,995

الحـــل :

الوسط الحسابي = ٧٩,٧٦ = $\bar{y}$

الـوسيــط = ٧٩,٠٦ = $\bar{Me}$

المـنـــــوال = ٧٧,٥٠ = $\bar{Mo}$

الانحراف القياسي = ١٥,٦٠ = S

لذا فأن

$$\alpha_1 = \frac{\bar{y} - \bar{Mo}}{S} = \frac{79.76 - 77.5}{15.6} = 0.14$$

$$\alpha_2 = \frac{\bar{y} - \bar{Me}}{S} = \frac{79.76 - 79.06}{15.6} = 0.13$$

وبما ان α_1 و α_2 موجبتان لذا فأن المنحني ملتوي التواء موجبا (أي الى اليمين).

مثـــال (٤) احسب معامل الالتواء الثالث لجدول التوزيع التكراري التالي :

الفئـــــات	التكـــــرار
٦٠–٦٢	٥
٦٣–٦٥	١٨
٦٦–٦٨	٤٢
٦٩–٧١	٢٧
٧٢–٧٤	٨
المجمـــوع	١٠٠

الحــل :

$m_2 = 8.5275$

$m_3 = -2.6932$

$$\alpha_3 = \frac{m_3}{(\sqrt{m_2}\ )^3} = \frac{-2.6932}{(\sqrt{8.5275}\ )^3} = -0.14$$

لذا فأن المنحني ملتوٍ إلتواءً سالباً (الى اليسار) لان α_3 سالبة .

(٦–٤)مقاييس التفلطح Measures of Kurtosis

تعريف : (٦ : ٤)

التفلطح او التدبب هو انحراف قمة منحني التوزيع التكراري عن قمة المنحني الطبيعي .

فالقمة العالية والضيقة حول الوسط الحسابي تسمى قمة مدببة- Lepto kurtic-أما القمة المنخفضة والمتسعة حول الوسط الحسابي فتسمى قمة مفلطحة Platy kurtic-أما قمة منحني التوزيع الطبيعي فتسمى قمة معتدلة Meso-kurtic

الأشكال التالية توضح ذلك

قمة مفلطحة   قمة معتدلة   قمة مدببة

وأهم مقياس للتفلطح هو :

$$\text{معامل التفلطح} = \frac{\text{العزم الرابع حول الوسط الحسابي}}{\text{مربع العزم الثاني حول الوسط الحسابي}} - ٣$$

أي

$$\beta = \frac{m_4}{m_2^2} - 3$$

فإذا كانت $\beta = 0$

سميت القمة معتدلة

واذا كانت $\beta > 0$

سميت القمة مدببة

واذا كانت $\beta < 0$

سميت القمة مفلطحة

مثال (٥) احسب معامل التفلطح (β) لجدول التوزيع التكراري في المثال السابق (مثال (٤)) .

الحـــــل :

$$m_2 = 8{\cdot}5275$$

$$m_4 = 199{\cdot}3759$$

$$\beta = \frac{m_4}{m_2^2} - 3$$

$$= \frac{199{\cdot}3759}{(8{\cdot}5275)^2} - 3 = -0{\cdot}26$$

لذا فأن المنحني له قمة مفلطحة (Platy-kurtic) لأن $\beta < 0$

تماريـن الفصـل السادس

(١) من البيانات التاليـة :

$$y_i = 2, 3, 7, 8, 10$$

أوجد ما يـلي :

آ- العزم الأول والثاني والثالث والرابع حول الصفر .

ب- العزم الأول والثاني والثالث والرابع حول الوسط الحسابي .

(٢) أحسب العزم الثالث حول الوسط الحسابي لجدول التوزيع التكراري التالي :

الفئـات	التكـرار
٦٠-٦٢	٥
٦٣-٦٥	١٨
٦٦-٦٨	٤٢
٦٩-٧١	٢٧
٧٢-٧٤	٨

(٣) أحسب معامل الألتواء الأول والثاني في جدول التوزيع التكراري التالي :

الفئـات	التكـرار
١١٨-١٢٦	٣
١٢٧-١٣٥	٥
١٣٦-١٤٤	٩
١٤٥-١٥٣	١٢
١٥٤-١٦٢	٥
١٦٣-١٧١	٤
١٧٢-١٨٠	٢

(٤) اذا كان العزم الثاني حول الوسط الحسابي لتوزيعين ما هما ٩ ، ١٦ . بينما العزم الثالث حول الوسط الحسابي للتوزيعين هما (-8.1) (-12.8) على التوالي . فأي التوزيعين أكثر التواء لليسار من الآخر ؟

(٥) أحسب معامل التفلطح (β) لجدول التوزيع التكراري في التمرين الثالث أعـلاه .

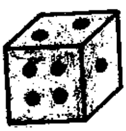
الفصل السابع

مبادئ نظرية الاحتمال

Elementary Probability Theory

(٧ : ١) المقدمـــة

ان نظرية الاحتمال تلعب دورا هاما في نظريات وتطبيقاتِ علم الاحصاء . ونظرية الاحتمال تعني بدراسة التجارب العشوائية . هذا ومعظم أمثلة الاحتمال مبنية على التجارب التالية :



(١) تجارب في زار الطاولة = زهر النرد = Dice

ويتألف الزار من ٦ وجوه كل وجه يأخذ رقما من واحد الى ستة التي هي عدد النقاط على ذلك الوجه .

(٢) تجارب قطعة النقود = Coin

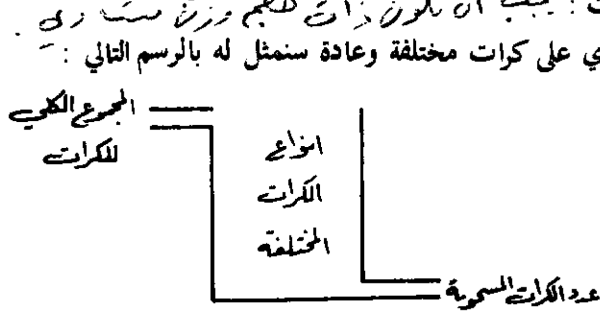
ان لقطعة النقود وجهان :

الصـــورة Head = (H)

والكتابـــة Tail = (T)

(٣) تجارب صندوق الكرات : يجب ان تكون ذات الحجم ووزن متساوي .

وصندوق الكرات يحتوي على كرات مختلفة وعادة سنمثل له بالرسم التالي :



(٤) تجارب مجموعة أوراق اللعب Deck cards

تتألف مجموعة اوراق اللعب من ٥٢ ورقة مقسمة الى أربعة مجاميع كل مجموعة بها ١٣ ورقة .

مجموعة الـ Heart (قلـب)

مجموعة الـ Spade (سنك)

مجموعة الـ Club (ماجـة)

مجموعة الـ Diamond (دينار)

وكل مجموعة تحتوي على ٤ أوراق صور هي Ace وKing وQueen وJack و ٩ أوراق تحمل ارقاماً من ٢ الى ١٠ .

كما ان مجموعة أوراق اللعب تتكون من لونين أسود وأحمر كل لون له ٢٦ ورقة .

المجموعة \ اللون	الأحمر	الأسود	المجموع
أرقام	١٨	١٨	٣٦
صور	٨	٨	١٦
المجموع	٢٦	٢٦	٥٢

ان استعمال هذه الأمثلة في نظرية الاحتمالات لا يعني بأن نظرية الاحتمالات لاتطبق الا في هذه المجالات والحقيقة أن نظرية الاحتمالات لها تطبيقات في شتى مجالات الحياة وهنالك سببان لاستعمال مثل هذه الأمثلة بكثرة في كتب الاحتمالات . السبب الأول تأريخي يعود الى أن المقامرين كانوا أول من أهتم بتطبيقات نظرية الاحتمالات في الألعاب والثاني هو سهولة مثل هذه الأمثلة من قبل جميع القراء حيث أن الأمثلة المتخصصة في علم معين تكون سهلة الفهم فقط على القارىء المتخصص في هذا العلم وصعبة على سواه وسوف يجد القارىء في هذا الكتاب الى جانب الأمثلة التوضيحية أمثلة من مختلف مجالات الحياة .

(٧ : ٢) مصطلحات وتعاريف

(١) التجربة العشوائية The random experiment

تعريف (٧ : ١) :

التجربة العشوائية هي التجربة التي لايمكن معرفة نتيجتها لخضوعها لقوانين الاحتمال

ان رمي زار الطاولة هي تجربة عشوائية لأن النتائج الممكنة لهذه التجربة تخضع لقوانين الاحتمال . واذا اراد كيمياوي أن يقدر كمية الزيت في بذور القطن فان العينة التي سيبني عليها تقديره والنتائج التي سيحصل عليها ستخضع لقوانين الاحتمال ولهذا فان هذه التجربة هي تجربة عشوائية .

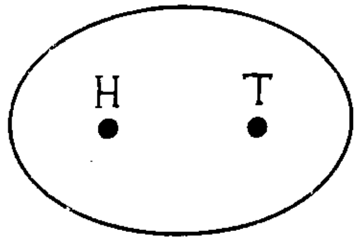
هذا وان نسبة عالية من التجارب العلمية هي تجارب عشوائية .

(٢) فضاء العينة : Sample space

تعريف (٧ : ٢) :

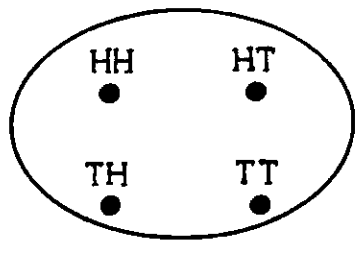
فضاء العينة هو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ماحيث أن كل نتيجة Outcome تمثل بنقطة point أو عنصر element في فضاء العينة .

فعند رمي قطعة نقود مرة واحدة فان فضاء العينة يتكون من نتيجتين ممكنتين : T و H



أما اذا رمينا قطعتين من النقود فان فضاء العينة عندئذ سيتكون من أربعة نتائج ممكنة هي :

HH   HT   TH   TT



وعليه فيمكن كتابة فراغ العينة في المثال الأخير كما يلي :

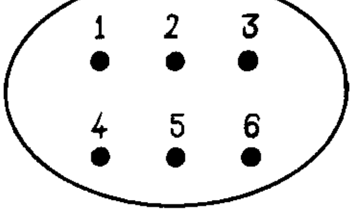
$A = \{HH, HT, TH, T\}$

حيث أن HH تعني ظهور الوجه الذي فيه الصورة على قطعة النقود الأولى والثانية . بينما HT فتمثل ظهور الصورة على قطعة النقود الأولى والكتابة على الثانية وهكذا ....

وعند رمي زار الطاولة (زهر النرد) مرة واحدة فان فضاء العينة يتكون من ٦ نتائج ممكنة هي :

$S = \{1,2,3,4,5,6\}$

حيث ان الرقم ٤ مثلاً هو ظهور الوجه الذي عليه ٤ نقاط وهكذا ...



هذا ويمكن ايجاد فضاء عينة آخر لكل من الأمثلة السابقة فمثلاً في تجربة رمي قطعتي نقود مرة واحدة قد يكون فضاء العينة هو عدد الصور التي يمكن الحصول عليها أي :

$B = \{0,1,2\}$

فنقاط فضاء العينة هي : الحصول على صفر من المرات صورة . الحصول على صورة واحدة . والحصول على صورتين على التوالي .

(٣) الحادث (أو الحدث) The event

> **تعريف (٣ : ٧) :**
>
> الحادث هو نقطة أو عدة نقاط في فضاء العينة ويرمز له ب (E_i).

فالحصول على الصورة (H) من رمي قطعة النقود مرة واحدة يسمى حادثاً وهو يتكون من نقطة واحدة $\{H\}$ من مجموعة نقاط فضاء العينة $\{H,T\}$.

وكذلك فان الحصول على (عدد زوجي) في رمية زار الطاولة يسمى أيضاً حادثاً وهو يتكون من النقاط [٢ . ٤ . ٦] من مجموعة نقاط فضاء العينة [١ . ٢ . ٣ . ٤ . ٥ . ٦] .

لذا فان الحادث قد يكون بسيطاً Simple event اذا تكون من نقطة واحدة في فضاء العينة (أي حالة واحدة من الحالات التي تظهر نتيجة التجربة) أو يكون حادثاً مركباً (Compound event) اذا شمل حالتين أو أكثر من الحالات التي تظهر نتيجة التجربة .

(٤) الحوادث المتنافية (المستبعدة) Mutually (exclusive) events

تعريف (٧ : ٤) :

يقال عن الحادثتين E_1 , E_2 انهما متنافيان (مستبعدان) اذا استحال حدوثهما معاً .

فمثلاً عند رمي قطعة نقود فمن المستحيل الحصول على صورة وكتابة في نفس الوقت .

(٥) الحوادث المستقلة Independent events :

تعريف (٧ : ٥) :

الحوادث المستقلة هي الحوادث التي اذا وقع أحدها لا يمنع أو يؤثر على وقوع الأحداث الأخرى .

فمثلاً عند رمي قطعتي نقود فالحصول على صورة في القطعة الأولى مثلاً لا يؤثر في نتيجة القطعة الثانية .

كذلك في حالة صندوق به عدد معين من الكرات البيضاء والسوداء (المتماثلة وزناً وحجماً) . فعند سحب كرتين بحيث تعاد الأولى قبل سحب الثانية فان نتيجة السحبة الأولى لا تؤثر في نتيجة السحبة الثانية لذا فالحادثتان مستقلان .

(٦) الحوادث غير المستقلة Non independent events :

تعريف (٧ : ٦) :

الحوادث غير المستقلة هي الحوادث التي اذا وقع أحدها يؤثر في وقوع الأحداث الأخرى .

فمثلاً في حالة صندوق به كرات بيضاء وسوداء فعند سحب كرتان على التوالي بحيث لا تعاد الكرة الاولى فان نتيجة السحبة الثانية تتأثر بنتيجة السحبة الأولى لذا فالحادثان غير مستقلين .

(٧) الحالات الممكنة Possible cases

تعريف (٧ : ٧)
الحالات الممكنة هي جميع الحالات المختلفة التي يمكن أن تظهر في تجربة معينة.

فعند رمي قطعة النقود فعدد الحالات الممكنة هنا حالتين (صورة أو كتابة) وعند رمي زهرة النرد فعدد الحالات الممكنة هي ٦ حالات. أما عندما نرمي زهرتي نرد فعدد الحالات الممكنة هي ٦ × ٦ = ٣٦ حالة . من ذلك نرى بأن الحالات الممكنة هي نفسها فضاء العينة .

(٨) الحالات المواتية Favourable cases

تعريف (٧ : ٨)
الحالات المواتية هي الحالات التي تحقق ظهور الحادث المراد دراسته وتسمى ايضاً بحالات النجاح .

فمثلاً عند رمي زهر الطاولة فإذا كان الحادث هو الحصول على عدد زوجي فالحالات التي تحقق ظهور هذا الحادث هي الحصول على (٢) أو (٤) أو (٦) فهذه الحالات الثلاثة تسمى الحالات المواتية.

(٩) الحالات المتماثلة Equally likely cases

تعريف (٧ : ٩)
الحالات المتماثلة هي الحالات المتكافئة والمتساوية في امكانية حدوثها .

فمثلاً عند رمي قطعة النقود فإن الظروف المهيأة للحصول على أي وجه (صورة أو كتابة) تكون متكافئة فيقال بأن الحالتين التي تنتج عن تجربة رمي قطعة النقود حالتان متماثلتان .

(۱۰) $\boxed{\text{مضروب } n}$: n factorial

> تعريف (۷ : ۱۰)
>
> مضروب n (ويرمز له بـ $n!$) يعرف بأنه :
>
> $$\boxed{n! = n(n-1)(n-2).......1}$$

فمثلاً مضروب ٥ هو

$$5! = 5(4)(3)(2)(1) = 120$$

لاحظ بأن :

$$0! = 1$$

$$1! = 1$$

$$n! = n[(n-1)!]$$

$$\frac{n!}{n} = (n-1)!$$

(۱۱) $\boxed{\text{التباديل}}$: Permutation

> تعريف (۷ : ۱۱)
>
> يقصد بالتباديل بأنها عدد طرق الاختيار المرتب التي يمكن تكوينها من عدة اشياء يأخذ ها كلها أوبعضها ويرمز له بـ nPr أي تباديل r من n وقانونه هو
>
> $$\boxed{nPr = \frac{n!}{(n-r)!}}$$

مثال (۱)

اذا كان لدينا أربعة حروف A, B, C, D واختير منها حرفان . فما هي عدد الطرق التي يمكن بها اختيار هذين الحرفين ؟

الحـــل :

$$4P2 = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12$$

أي أن عدد الطرق = ١٢ وهذه الطرق هي :

AB AC AD BC BD CD

BA CA DA CB DB DC

حيث ان كلاً منهما يمثل ترتيباً مختلفاً للحرفين .

مثال(٢) : كتبت الأرقام من ١ الى ٩ على بطاقات ووضعت في صندوق ثم سحبت منه ٥ بطاقات (الواحدة بعد الأخرى) فكم عدداً خماسياً (أرقامه مختلفة) يمكن تكوينه ؟

الحـل :

ان المرتبة الأولى من الرقم الخماسي هذا يمكن اختيار رقماً لها بتسعة طرق وبعد اختيار هذا الرقم يبقى ثمانية أرقام نختار أحدها للمرتبة الثانية وسبعة أرقام نختار أحدها للمرتبة الثالثة وستة أرقام نختار أحدها للمرتبة الرابعة . وخمسة أرقام نختار أحدها للمرتبة الخامسة لذا فان :

$$ {}_9P_5 = \frac{9!}{(9-5)!} = \frac{(9)(8)(7)(6)(5)4!}{4!} $$

$$ = 15120 $$

أي أن هناك ١٥١٢٠ رقماً خماسياً يمكن تكوينه .

هذا واذا كانت $r = n$ في قانون التباديل فان القانون لايزال صحيحاً من الوجهة الرياضية

$$ nPn = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! $$

ومن الوجهة العلمية . اذا كانت $r = n$ فان عدد التباديل هو عدد الطرق التي يمكننا ترتيب n من الأشياء على خط مستقيم . فمثلاً اذا أراد طالب أن يرتب أربعة كتب مختلفة المواضيع على رف مكتبته فانه يمكنه اختيار الكتاب الأول بأربعة طرق والثاني بثلاثة طرق والثالث بطريقتين والرابع بطريقة واحدة وهذه هي عدد الطرق التي يمكننا فيها اختيار أربعة أشياء من أربعة أشياء وهي تساوي

$$ {}_4P_4 = 4! $$
$$ = (4)(3)(2)(1) = 24 $$

التباديل في حالة وجود مجموعات متشابهة :

لندرس الآن عدد الطرق المختلفة التي يمكننا فيها ترتيب حروف كلمة«(باب) . قد تبدو لأول وهلة أن عدد هذه الطرق هو6 = !3 الا أن تكرار حرف ب مرتين يمنع استعمال هذه القاعدة لأنه لا يمكن التمييز بين ب الموجودة في أول الكلمة و ب الموجودة في آخر الكلمة وفي الحقيقة أن عدد الطرق هو ٣ وليس ٦ وهي :

باب و أبب و ببا .

فعدد الحرف هنا ٣ ، ٢ منها متشابهة لذا فان عدد الطرق الممكن بها ترتيب حروف باب هو :

$$\frac{3!}{2!1!} = 3$$

ويمكن القول بأنه اذا كانت لدينا n من الأشياء ، m منها متشابهة فان عدد الطرق لترتيب هذه الأشياء على خط هو

$$\frac{n!}{m!}$$

فاذا كانت هناك m_1 من الأشياء المتشابهة و m_2 من الأشياء المتشابهة الأخرى فان عدد الطرق الممكنة لترتيب هذه الأشياء على خط هو

$$\frac{n!}{m_1!\,m_2!}$$

فمثلاً عدد الطرق الممكنة لترتيب حروف كلمة السلسلة هو

$$\frac{7!}{2!3!1!1!} = 420$$

وبالامكان الآن تعميم هذه القاعدة فنقول أن التباديل في حالة وجود مجموعات متشابهة هو

$$\frac{n!}{m_1!\,m_2!\,m_3!\,.....}$$

علماً بأن

$$m_1 + m_2 + ... = n$$

مثال(٣) ماهو عدد الترتيبات التي يمكن تكوينها من أحرف كلمة Statistics ؟

الحل :

عدد الأحرف الكلية = ١٠

وان الحرف s تكرر ٣ مرات

وان الحرف t تكرر ٣ مرات

وان الحرف a تكرر مرة واحدة

وان الحرف i تكرر مرتان

وان الحرف c تكرر مرة واحدة

لذا فان

$$\boxed{\frac{n!}{m_1!\,m_2!\,m_3!\,m_4!\,m_5!}} = \frac{10!}{3!\,3!\,1!\,2!\,1!} = 50400$$

(١٢) التوافيق : Combinations

تعريف (٧ : ١٢)

يقصد بالتوافيق بأنها عدد طرق الاختيار غير المرتب التي يمكن تكوينها من عدة أشياء بأخذها كلها أو بعضها ويرمز له بـ nC_r أو $\binom{n}{r}$ وقانونه هــــو :

$$\boxed{\binom{n}{r} = \frac{n!}{r!\,(n-r)!}}$$

أي أن الترتيب في حالة التوافيق غير مهم.

لاحظ بأن
$$\binom{n}{r} = \binom{n}{n-r}$$

مثال (٤) ماعدد طرق الاختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من ٥ اشخاص من مجموع ٩ اشخاص ؟

الحـــل : لاحظ بأن ترتيب الاشخاص هنا غير ضروري لأن اختيار عمرو قبل زيد أو العكس هي نتيجة واحدة .

$$\therefore \quad \binom{9}{5} = \frac{9!}{5!\,(9-5)!}$$

$$= \frac{9!}{5!\,4!}$$

$$= 126$$

طريقة

هذا وهناك قاعدتان اساسيتان يعتمد عليهما كل من التباديل والتوافيق وهما :

(١) اذا كان عدد الطرق الممكنة لوقوع الحادث E_1 هو n . وان عدد الطرق الممكنة لوقوع الحادث E_2 هو m وكان E_1 و E_2 حادثان متنافيان فان عدد الطرق الممكنة لوقوع الحادث E_1 أو E_2 هو $(n + m)$ من الطرق.

$$\binom{4}{2} = \frac{4\times3\times2}{2!\,2!} = \frac{12}{2} = 6$$

مثال (٥) : من المعلوم أن عدد أوراق اللعب هو ٥٢ ورقة وان ورقة (Spade) يمكن أن تحدث بـ ١٣ طريقة وان ورقة (Heart) يمكن أن تحدث بـ ١٣ طريقة أيضاً فعند سحب ورقة عشوائيا من أوراق اللعب فإن عدد الطرق الممكنة لاختيار Spade أو Heart = ١٣ + ١٣ = ٢٦ طريقة .

(٢) اذا كان عدد الطرق الممكنة لوقوع الحادث E_1 هو n وان عدد الطرق الممكنة لوقوع E_2 هو m وكان E_1 و E_2 حادثان مستقلان فان عدد الطرق الممكنة لوقوع الحادثان E_1 و E_2 هو (nm) من الطرق .

مثال(٦) اذا سحبت ورقتان من مجموعة اوراق اللعب بحيث أن احداهما تكون spade. والأخرى Heart فإن هناك ١٣ × ١٣ = ١٦٩ طريقة لعمل ذلك .

مثال(٧) صندوق به ٦ كرات حمراء و ٤ سوداء و ٢ بيضاء فبكم طريقة يمكن اختيار ٥ كرات بحيث تكون ٣ منها حمراء و ٢ سوداء ؟

الحـــل :

عدد طرق اختيار ٣ كرات حمراء هو

$$\binom{8}{1} = \frac{8!}{1!\,(8-1)!} = \frac{8\times7!}{1!\,7!}$$

$$\binom{6}{3} = 20$$

وعدد اختيار ٢ كرة سوداء هو

$$\binom{4}{2} = 6$$

اذن عدد الطرق لاختيار ٣ كرات حمراء و ٢ سوداء هو

$$\binom{6}{3}\binom{4}{2} = (20)\ (6) = 120$$

مثال(٦) اذا كان لدى مدرب فريق الشباب العراقي ٢٠ لاعباً واراد تشكيل منهم فريقاً (١١ شخصاً) يتألف من ١ حامي هدف و ٢ للدفاع و ٣ للوسط و ٥ للهجوم فإذا كان ٢ من الشباب يمكنهما أن يكونا حماة للهدف و ٥ يمكنهم أن يلعبوا في موقف الدفاع و ٦ يمكنهم أن يلعبوا في موقف الوسط و ٧ يمكنهم أن يلعبوا في موقف الهجوم فما هو عدد الفرق الممكن تشكيلها ؟

الحــل :

فريق
$$\binom{2}{1}\binom{5}{2}\binom{6}{3}\binom{7}{5} = 8400$$

(٧ : ٣) الاحتمال Probability

التعريف الكلاسيكي للاحتمال (٧ : ١٣)

لنفرض أن حادثا معينا (E_i) يمكن أن يتحقق (أو يحدث أو ينجح) في n من الحالات (وتسمى الحالات المواتية Favourable cases) من مجموع N من الحالات الممكنة (All possible cases) وعلى فرض أن جميع الحالات الممكنة هذه متساوية الحدوث (Equally likely cases) فإن درجة احتمال ظهور الحادث E_i ويرمز له بـ $p(E_i)$ هو

$$P(E_i) = \frac{\text{عدد الحالات المواتية للحادث}}{\text{عدد الحالات الممكنة}} = \frac{n}{N}$$

هذا وان احتمال عدم ظهور هذا الحادث (أي فشله) ويرمز له بـ $P(\bar{E}_i)$ هــو

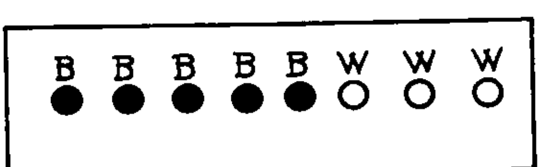
$$P(\bar{E}) = 1 - P(E_i)$$

ويسمى $P(\bar{E}_1)$ باحتمال الحادث المكمل (او احيانا يسمى بالاحتمال العكسي) .

مثـال(٧) صندوق به ٣ كرات بيضاء و٥ كرات سوداء (وان الكرات جميعها متماثلة من جميع الوجوه ماعدا اللون) فإذا سحبت كرة من هذا الصندوق عشوائيا فما هو احتمال ان تكون سوداء ؟

الحــل :

ان فضاء العينة لهذه التجربة هو .



لذا فإن عدد الحالات الممكنة = ٨ اي من الممكن اختيار كرة بـ ٨ طرق .

وان عدد الحالات المواتية = ٥ اي من الممكن اختيار كرة سوداء بـ ٥ طرق .

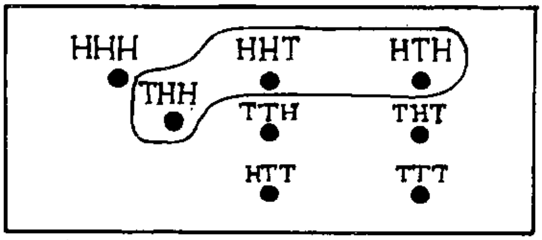
لذا فإن احتمال ان تكون هذه الكرة سوداء هو .

$$P(B) = \frac{5}{8}$$

مثـال (٨) : رميت قطعة نقود ٣ مرات فما هو احتمال الحصول على الصورة مرتين ؟

الحـل :

ان فضاء العينة لهذه التجربة هو :



لذا فعدد الحالات الممكنة = ٨ حالات (لان ٢×٢×٢=٨) وهي :

HHH, HHT, HTH, THH, H TT , THT, TTH, TTT

أما الحالات الممكنة = ٣ حالات وهي :

(HHT, HTH, THH)

لذا فاحتمال الحصول على الصورة مرتين هو

$$P(2H) = \frac{3}{8}$$

ان التعريف الكلاسيكي للاحتمال يفترض بأن تكون الحالات الممكنة متماثلة (Equally likely) لذا فالاحتمالات التي تحسب باستعمال التعريف الكلاسيكي تسمى الاحتمالات القَبْلية (Apriori probability)

(٢) التعريف النسبي او التجريبي (٧ : ١٤) :

اذا اجريت تجربة ثم اعيدت n من المرات فإن احتمال حدوث الحادث E_i هو عبارة عن نسبة حدوثه اي

$$P(E_i) = \frac{\text{عدد ظهور الحـــادث}}{\text{عدد مرات اجراء التجربـــة}}$$

مثال(٩) القي زار الطاولة ١٠٠ مرة فكان عدد مرات ظهور الوجه الذي عليه ٣ نقاط هو ٢٠ لذا فإن احتمال ظهور الوجه الذي عليه ٣ نقاط هو $= \frac{٢٠}{١٠٠}$

من هذا نرى بأن الاحتمال هنا يحسب بعد اجراء التجربة ومشاهدة نتائجها لذا فالاحتمالات هذه تسمى احتمالات بعدية (A posteriori probability)، هذا وان جميع القوانين التي ستذكر بعد الان هي مستنتجة من الاحتمالات القبلية ولكنها ايضا صحيحة بالنسبة للاحتمالات النسبية ايضا .

(٣) بعض خواص الاحتمال :

خاصية (١) اذا رمزنا لاحتمال حدوث حادث ما بالرمز $P(E_i)$ واحتمال عدم حدوث هذا الحادث بالرمز $P(\bar{E}_i)$
فإن

$$P(E_i) + P(\bar{E}) = 1$$

لأن

$$\boxed{P(E) = \frac{n}{N}}$$

$$\boxed{P(\bar{E}) = \frac{N-n}{N}}$$

$$\boxed{P(E) + P(\bar{E}) = 1}$$

مثال (١٠) صندوق يحتوي على ٦ كرات حمراء و٤ بيضاء و٥ صفراء فإذا سحبت منه كرة واحدة عشوائيا فما هو درجة احتمال ان تكون هذه الكرة :

(أ) حمراء

(ب) غير حمراء

الحـل :

(أ) احتمال كونها حمراء هو

$$P(R) = \frac{n}{N}$$

$$= \frac{6}{15}$$

(ب) احتمال كونها غير حمراء هو

$$P(\bar{R}) = \frac{N-n}{N}$$

$$= \frac{9}{15}$$

من هذا يتضح بأن

$$P(R) + P(\bar{R}) = 1$$

خاصية (٢) ان درجة احتمال أي حادث تتراوح بين الصفر والواحد أي $0 \leq P(E_i) \leq 1$

فاذا كان درجة احتمال حدوث الحادث E_i = ١ سمي الحادث (E_i) بأنه أكيد. أما اذا كان درجة احتمال حدوث الحادث E_i = صفر سمي الحادث (E_i) بأنه مستحيل
وبديهي فان $p(E_i) = 1$ عندما تكون $n = N$.
وان $p(E_i) = 0$ عندما تكون $n = 0$

مثـال (١١) صندوق يحتوي على ٢٠ كرة بيضاء فاذا سحبت منه كرة واحدة عشوائية فما هو احتمال :

(أ) ان تكون بيضاء ؟

(ب) ان تكون حمراء ؟

الحـل :

$$P(W) \frac{n}{N} = \frac{20}{20} = 1$$

(ب) $$P(R) = \frac{n}{N} = \frac{0}{20} = 0$$

ومن هذا يتبين أيضاً بأن قيمة درجة الاحتمال لا يمكن ان تكون أقل من الصفر (سالبة) مطلقاً .

خاصية (٣) اذا كانت $E_1, E_2, \ldots E_n$ هي عناصر أو نقاط فضاء العينة فان مجموع درجات احتمالاتها = ١ أي $P(E_1) + P(E_2) + \ldots + P(E_n) = 1$ $\sum P(E_i) = 1$

مثـال(١٢) صندوق يحتوي على ٦ كرات حمراء و ٤ بيضاء و ٥ صفراء فإذا سحبت منه كرة عشوائيا فاحتمال ان تكون حمراء هو

$$P(R) = \frac{6}{15}$$

واحتمال ان تكون بيضاء هو

$$P(W) = \frac{4}{15}$$

واحتمال ان تكون صفراء هو

$$P(Y) = \frac{5}{15}$$

من هذا يتضح بأن

$$P(R) + P(w) + P(y) = 1$$

(٤) قوانين الاحتمال Laws of Probability

لقد وضعت القوانين التالية لتسهيل حساب درجة الاحتمال عند وقوع حادثين او أكثر بدلاً من ايجادها عن طريق تعريف الاحتمال الذي يكون من الصعوبة في مثل هذه الحالات حساب عدد الحالات المواتية والممكنة .

وقبل شرح قوانين الاحتمالات نفرض ان هناك حادثان E_1 و E_2 فالتعابير التالية يقصد بها ما يلي :

*معرفة حدوث حدثين في نفس ا

احتمال وقوع الحدث E_1 أو الحدث E_2 = $P(E_1 + E_2)$

(اي احتمال وقوع ايا منهما فقط)

احتمال وقوع الحادث E_1 والحادث E_2 معا = $P(E_1 E_2)$

احتمال حدوث E_2 علماً بأن الحادث E_1 قد وقع = $P(E_2 | E_1)$ ويسمى بالاحتمال الشرطي (Conditional probability)

(أ) قانون الجمع Addition law

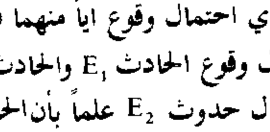
١ – اذا كانت الاحداث متنافية

تعريف (٧ : ١٥) الحصول اما على الحدث الاول E_1 او الحدث الثاني E_2

اذا كانت E_1 و E_2 حادثان متنافيان .

فإن احتمال حدوث ايا منهما (أي E_1 أو E_2) هو حاصل جمع احتمال كل منهما اي

$$P(E_1 + E_2) = P(E_1) + P(E_2)$$



وبصورة عامة اذا كانت $E_1, E_2 \ldots E_n$ احداثاً متنافية فإن

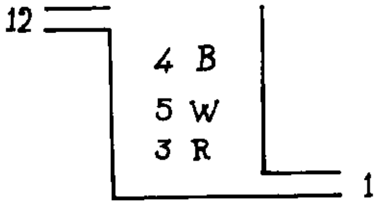
$$P(E_1 + E_2 + \ldots + E_n) = P(E_1) + p(E_2) + \ldots + P(E_n)$$

مثـال(١٣) صندوق يحتوي على ٤ كرات سوداء و٥ بيضاء و٣ حمراء ، فإذا سحبت كرة واحدة فما هو احتمال أن تكون أما سوداء أو بيضاء ؟

الحـل :

$$P(B + W) = P(B) + P(W)$$

$$= \left(\frac{4}{12}\right) + \left(\frac{5}{12}\right) = \frac{9}{12}$$



مثـال (١٤) في حالة رمي زارين (زهرتي النرد) ماهواحتمال الحصول عـلى عدد زوجي

الحـل :

$$P(2 + 4 + 6) = P(2) + P(4) + P(6)$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

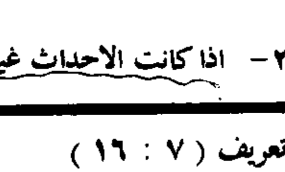
$$= \frac{3}{6} = \frac{1}{2}$$

٢- اذا كانت الاحداث غير متنافية

تعريف (٧ : ١٦)

اذا كانت E_1 و E_2 حادثان غير متنافيين فإن احتمال حدوث أي منهما (E_1 او E_2) هو حاصل جمع احتمال كل منهما ناقصا احتمال حدوثهما معا اي

$$P(E_1 + E_2) = P(E_1) + P(E_2) - P(E_1 E_2)$$



وبصورة عامة اذا كانت $E_1 , E_2 \dots E_n$ احداثاً غير متنافية فإن :

$$P(E_1 + E_2 + \dots + E_n) = P(E_1) + (E_2) + \dots + P(E_n)$$
$$- P(E_1 E_2) - P(E_1 E_3) \dots$$
$$+ P(E_1 E_2 E_3) + \dots$$

بحيث تكون اشارة + للحالات الفردية و – للحالات الزوجية .
هذا وقد تكون الاحداث غير المتنافية مستقلة أو غير مستقلة .

مثال (١٥) في احدى الكليات ، ٢٥ ٪ من الطلبة رسب بالرياضيات و ١٥ ٪ من الطلبة رسب في الكيمياء و ١٠ ٪ رسب في كلا الرياضيات والكيمياء فاذا انتخب طالب منهم عشوائياً فما هو احتمال أن يكون راسباً في الرياضيات والكيمياء .

الحــل :

نرمز للرياضيات بالرمز M والكيمياء بالرمز C

$$P(M + C) = P(M) + P(C) - P(MC)$$
$$= \cdot 25 + \cdot 15 - \cdot 10 = \cdot 30$$

مثال (١٦) اذا كان الرجل (A) يصيب هدفاً ما باحتمال $\frac{1}{4}$ وان الرجل B يصيب نفس الهدف بأحتمال $\frac{2}{5}$. ماهو احتمال اصابة الهدف اذا صوب A و B نحو الهدف ؟

الحــل :

المقصود هنا ماهو احتمال A أو B أو كلاهما يصيب الهدف

$$P(A + B) = P(A) + P(B) - P(AB)$$

$$P(A) = \frac{1}{4}$$

$$P(B) = \frac{2}{5}$$

وبما أن A و B حادثين مستقلان

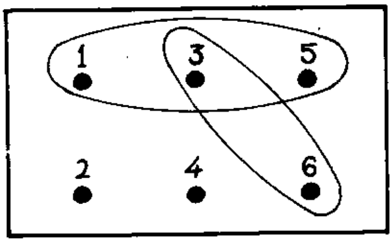
$$\therefore P(AB) = P(A) \cdot P(B)$$
$$= (\frac{1}{4})(\frac{2}{5})$$
$$= \frac{1}{10}$$

$$\therefore P(A+B) = P(A) + P(B) - P(AB)$$
$$= \frac{1}{4} + \frac{2}{5} - \frac{1}{10}$$
$$= \frac{11}{20}$$

مثال (١٧) اذا القى (زار) مرة واحدة فما هو احتمال ظهور عدد يكون فرديا او يقبل القسمة على ٣ ؟

الحـل :

ان فضاء العينة لهذه التجربة هو :



(A) = الحادث (فردي) وعدد حالاته الممكنة ٣ (وهي الاوجه ١ ، ٣ ، ٥)

$$\therefore P(A) = \frac{3}{6}$$

(B) = الحادث يقبل القسمة على ٣ وعدد حالاته الممكنة ٢ (وهي ٣ ، ٦)

$$\therefore P(B) = \frac{2}{6}$$

(AB) الحادث فردي ويقبل القسمة على ٣ هو فقط ٣ فقط (عدد الحالات الممكنة = ١)

$$\therefore \quad P(AB) = \frac{1}{6}$$

$$\therefore \quad P(A+B) = P(A) + P(B) - P(AB)$$

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

$$= \frac{4}{6}$$

مثـال(١٨) اذا كان احتمال ان الطالب (A) يستطيع حل مسألة ما هو $\frac{٤}{٥}$ وان احتمال الطالب (B) يستطيع حل نفس المسألة هو $\frac{٢}{٣}$ واحتمال ان الطالب (C) يستطيع حلها هو $\frac{٣}{٧}$. فإذا ثلاثتهم حاولوا حل المسألة . فما هو احتمال ان المسألة تحل ؟

الحـل :

الطريقة الاولى

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC$$

وبما ان هذه الحوادث مستقلة لذا فإن :

$$P(A+B+C) = P(A) + P(B) + P(C) - P(A)\cdot P(B) - P(A)\cdot P(C) - P(B)\cdot P(C) + P(A)P(B)\cdot P(C)$$

$$= (4/5) + (2/3) + (3/7) - (4/5)(2/3) - (4/5)(3/7) - (2/3)(3/7) + (4/5)(2/3)(3/7)$$

$$= \frac{101}{105}$$

الطريقة الثانية :

احتمال ان A لا يحلها هو : $P(\overline{A}) = 1 - (4/5) = (1/5)$

احتمال ان B لا يحلها هو : $P(\overline{B}) = 1 - (2/3) = (1/3)$

احتمال ان C لا يحلها هو :

$$P(\bar{C}) = 1 - (3/7) = (4/7)$$

احتمال أنهم جميعا لا يحلونها هو

$$\therefore\ P(\bar{A}\ \bar{B}\ \bar{C}) = P(\bar{A}) \cdot P(\bar{B}) \cdot P(\bar{C})$$

$$= \left(\frac{1}{5}\right) \cdot \left(\frac{1}{3}\right) \left(\frac{4}{7}\right) = \frac{4}{105}$$

∴ احتمال ان واحداً منهم على الاقل يحلها هو ●

P (يحلها واحد منهم على الأقل) $= 1 - P(\bar{A}\ \bar{B}\ \bar{C}) = 1 - \frac{4}{105} = \frac{101}{105}$

(ب) قانون الضرب Multiplication law

١. اذا كانت الاحداث مستقلة

تعريف (٧ : ١٧)

اذا كان E_1 و E_2 حادثين مستقلين فأن احتمال حدوثهما معا هو حاصل ضرب احتمال كل منهما أي :

$$P(E_1 E_2) = P(E_1) \cdot P(E_2)$$

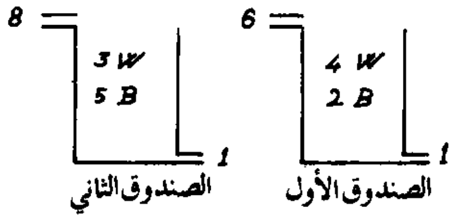
وبصورة عامة اذا كانت $E_1, E_2, ..., E_n$ احداثا مستقلة فأن احتمال حدوثهم معا هو حاصل ضرب احتمال كل منهم

$$P(E_1 E_2 .... E_n) = P(E_1) \cdot P(E_2) ... \cdot P(E_n)$$

مثال(١٩) صندوقان . الأول يحوي على ٤ كرات بيضاء و٢ سوداء والثاني يحوي على ٣ بيضاء و٥ سوداء فاذا سحبت كرة من كل منهما فما هو احتمال ان يكونا سوداوين ؟

الحـــل :

نرمز لاحتمال الكرة السوداء من الصندوق الأول بـ $P(B_1)$ واحتمال الكرة السوداء من الصندوق الثاني بـ $P(B_2)$ ،



$$\therefore \quad P(B_1 B_2) = P(B_1) \cdot P(B_2)$$
$$= (\frac{2}{6})(\frac{5}{8})$$
$$= \frac{10}{48}$$

مثـال(٢٠) عند رمي قطعتي نقود ما هو احتمال الحصول صورة في كليهما ؟

الحـل : نرمز لاحتمال الحصول على الصورة من قطعة النقود الأولى بـ $P(H_1)$ ولاحتمال الحصول على الصورة من قطعة النقود الثانية بـ $P(H_2)$

$$\therefore \quad P(H_1 H_2) = P(H_1) \cdot P(H_2) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$$

مثـال (٢١) اذا كان احتمال اصابة الطائرة الأولى (A) لهدف معين يساوي $\frac{١}{٣}$ واحتمال اصابة الطائرة الثانية (B) لنفس الهدف = $\frac{١}{٥}$ فاذا كان عمل كل منهما مستقلا عن الاخرى واطلقت كل من الطائرتين قنبلة في آن واحد تجاه الهدف فما هو احتمال

(أ) ان تصيب الطائرتان معا بالهدف

(ب) ان لاتصيبا الهدف

الحـل :

(a) $P(AB) = P(A) \cdot P(B)$

$$= (\frac{1}{3}) \cdot (\frac{1}{5}) = \frac{1}{15}$$

احتمال الأولى ان لا تصيب الهدف :

(b) $P(\overline{A}) = 1 - P(A) = \frac{2}{3}$

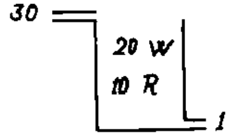
احتمال الثانية ان لا تصيب الهدف :

$P(\overline{B}) = 1 - P(B) = \frac{4}{5}$

$$\therefore \quad P(\overline{A}\,\overline{B}) = P(\overline{A}) \cdot P(\overline{B}) = (\frac{2}{3})(\frac{4}{5}) = \frac{8}{15}$$

مثال (٢٢) صندوق به ٢٠ كرة بيضاء و ١٠ كرة حمراء سحبت كرة ثم اعيدت ثم سحبت كرة اخرى ، ما هو احتمال ان يكونا حمرا .

الحـل :



نرمز لاحتمال الاولى حمراء = $P(R_1)$ ولاحتمال الثانية حمراء بـ $P(R_2)$

$$\therefore \quad P(R_1 R_2) = P(R_1) \cdot P(R_2) = \left(\frac{10}{30}\right)\left(\frac{10}{30}\right) = \frac{1}{9}$$

٢. اذا كانت الاحداث غير مستقلة :

تعريف : (٧ : ١٨)

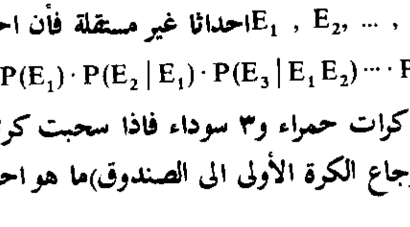
اذا كان E_1 و E_2 حدثين غير مستقلين فأن احتمال حدوثهما معا يساوي حاصل ضرب احتمال وقوع الحادث الأول في احتمال وقوع الحادث الثاني مشروطا بحدوث الأول أي :

$$P(E_1 E_2) = P(E_1) \cdot P(E_2 \mid E_1)$$

وبصورة عامة اذا كانت $E_1 , E_2, \dots , E_n$ احداثا غير مستقلة فأن احتمال حدوثهم معا هو

$$P(E_1 E_2 \dots E_n) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1 E_2) \cdots P(E_n \mid E_1 E_2 \dots E_{n-1})$$

مثـال(٢٣) :صندوق به ٥ كرات حمراء و٣ سوداء فاذا سحبت كرتان سوية (او سحبت كرتان على التوالي بدون ارجاع الكرة الأولى الى الصندوق)ما هو احتمال ان تكون كلتاهما سوداً ؟

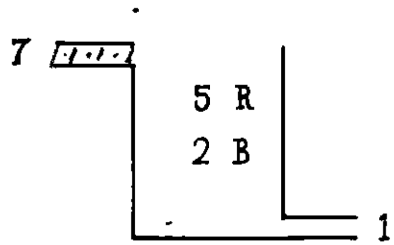


الحـــل :

احتمال الحصول على كرة سوداء في السحبة الأولى هو

$$P(B_1) = \frac{3}{8}$$

أما في السحبة الثانية (بدون ارجاع الكرة المسحوبة الى الصندوق) فأن احتمال أن تكون الكرة سوداء هو :



$$P(B_2 | B_1) = \frac{2}{7}$$

$$\therefore \quad P(B_1 B_2) = P(B_1) \cdot P(B_2 | B_1) = \left(\frac{3}{8}\right)\left(\frac{2}{7}\right) = \frac{6}{56}$$

مثــال(٢٤) شعبة أ من الصف الأول تتألف من ٢٥ طالبا و ١٠ طالبات . فاذا اختيرت ٣ اسماء عشوائيا فما هو احتمال ان يكونوا من الذكور ؟

الحل : نرمز لاحتمال ان يكون الأول طالبا لـ $P(B_1)$ وهكذا ...

$$P(B_1 B_2 B_3) = P(B_1) \cdot P(B_2 | B_1) \cdot P(B_3 | B_1 B_2)$$

$$= \left(\frac{25}{35}\right)\left(\frac{24}{34}\right)\left(\frac{23}{33}\right)$$

(٥) الاحتمال الشرطي Conditional probability

تعريف : (١٩ : ٧)

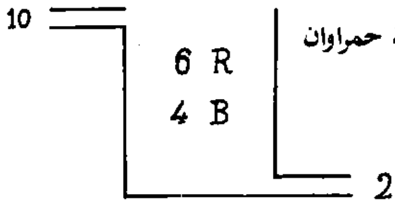
اذا كان A و B حادثتين في فضاء العينة فان احتمال وقوع الحادث A علما بأن الحادث B قد وقع (ويرمز له بـ $P(A | B)$) هو :

$$P(A | B) = \frac{P(AB)}{P(B)} = \frac{AB}{B} \; \frac{\text{عدد الحالات المواتية لوقوع الحادث}}{\text{عدد الحالات المواتية لوقوع الحادث}}$$

على ان يكون احتمال وقوع B اكبر من صفر أي $(P(B) > 0)$.

مثــال(٢٥) صندوق يحوي على ٦ كرات حمراء و ٤ سوداء فاذا سحبت كرتان على التوالي (بدون ارجاع) ما هو احتمال ان تكون الكرة الثانية حمراء علماً بأن الكرة الأولى كانت حمراء أيضاً ؟

الحــل : $P(R_2 | R_1) = \dfrac{P(R_1 R_2)}{P(R_1)}$



حيث ان $P(R_1 R_2)$ ؛ هو احتمال الكرة الأولى والثانية حمراوان

ان عدد اختيار كرتان حمرا $= \begin{pmatrix} 6 \\ 2 \end{pmatrix}$

وعدد اختيار كرتان من الصندوق $= \begin{pmatrix} 10 \\ 2 \end{pmatrix}$

$$\therefore p(R_1 R_2) = \frac{\begin{pmatrix} 6 \\ 2 \end{pmatrix}}{\begin{pmatrix} 10 \\ 2 \end{pmatrix}} = \frac{1}{3}$$

أما احتمال ان تكون الكرة الأولى حمراء هي : $P(R_1) = (6/10)$

$$\therefore \quad P(R_2 | R_1) = \frac{P(R_1 R_2)}{P(R_1)} = \frac{(1/3)}{(6/10)} = \frac{5}{9}$$

ملاحظة : يمكن حساب الاحتمال الشرطي هذا مباشرة :
فبما ان الكرة الأولى كانت حمراء فان عدد الكرات الحمر الباقية هي ٥ ومجموع الكرات الكلية أصبحت ٩ لذا فاحتمال الكرة الثانية أيضاً حمراء هي $\frac{٥}{٩}$ أي

$$P(R_2 | R_1) = \frac{5}{9}$$

مثـال(٢٦) صنف الشباب في احدى المدن كالآتي :

	له وظيفة (E)	ليست له وظيفة (V)	المجموع
ذكور (M)	٤٦٠	٤٠	٥٠٠
اناث (F)	١٤٠	٢٦٠	٤٠٠
المجمــوع	٦٠٠	٣٠٠	٩٠٠

فاذا أخترنا شابا بصورة عشوائية فما هو احتمال ان يكون ذكرا موظفا

الحل :

نرمز للذكــر بــ = M

والمــوظـف بــ = E

$$\therefore \quad P(M | E) = \frac{P(ME)}{P(E)} = \frac{(460)}{(600/900)} = \frac{460}{600} = \frac{23}{30}$$

ملاحظة : ان الاحتمال الشرطي لأكثر من حادثين يمكن استنتاجه بسهولة فمثـــلا

$$P(AB|C) = \frac{P(ABC)}{P(C)}$$

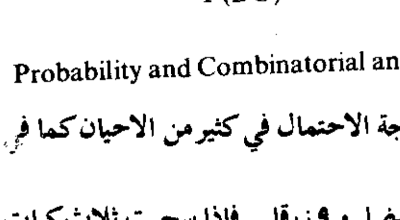
وكذلك

$$P(A|BC) = \frac{P(ABC)}{P(BC)}$$

(٧ : ٤) الاحتمال والتحليل التوافقي Probability and Combinatorial analysis

ان التحليل التوافقي يسهل كثيرا في حساب درجة الاحتمال في كثير من الاحيان كما في الأمثلة التالية :

مثـال(٢٧) صندوق يحتوي على ٨ كرات حمرا و٣ بيضا و٩ زرقا . فاذا سحبت ثلاث كرات عشوائيا . احسب احتمال :



(أ) ثلاثتهم حمر

(ب) ثلاثتهم بيض

(جـ) ٢ حمر و ١ بيضاء

(د) على الأقل ١ بيضاء

(هـ) واحدة من كل لون

(و) الكرات سحبت بالترتيب التالي حمراء ثم بيضاء ثم زرقاء

الحل :

(أ) نرمز R_1 , R_2 , R_3 للحوادث الثلاثة (سحب الكرة الأولى حمراء والكرة الثانية حمراء والكرة الثالثة حمراء)

الطريقة الأولى: ان هذه الحوادث غير مستقلة لذلك فأن قانون الضرب :

$$P(R_1 R_2 R_3) = P(R_1) \cdot P(R_2|R_1) \cdot P(R_3|R_1 R_2)$$

$$= (\frac{8}{20})(\frac{7}{19})(\frac{6}{18}) = \frac{14}{285}$$

الطريقة الثانية: باستخدام التحليل التوافقي :

$$P(3\,R) = \frac{\text{عدد اختيار ٣ كرات حمر من ٨ كرات}}{\text{عدد اختيار ٣ كرات من ٢٠ كـــرة}}$$

$$= \frac{\binom{8}{3}}{\binom{20}{3}} = \frac{14}{285}$$

(ب) $$P(3W) = \frac{\binom{3}{3}}{\binom{20}{3}} = \frac{1}{1140}$$

(جـ) $$P(2R\,1W) = \frac{\binom{8}{2}\binom{3}{1}}{\binom{20}{3}} = \frac{7}{95}$$

(د) الطريقة الأولى :

احتمال الكرة غير بيضاء

$$P(\bar{W}) = \frac{\binom{17}{3}}{\binom{20}{3}} = \frac{34}{57}$$

$\therefore$ P(على الأقل واحدة بيضاء) $= 1 - P(\bar{W}) = \dfrac{23}{57}$

الطريقة الثانية

P(على الأقل واحدة بيضاء) $= P(1W2\bar{W}) + P(2W1\bar{W}) + p(3W)$

$$= \frac{\binom{3}{1}\binom{17}{2}}{\binom{20}{3}} + \frac{\binom{3}{2}\binom{17}{1}}{\binom{20}{3}} + \frac{\binom{3}{3}}{\binom{20}{3}}$$

$$= \frac{23}{57}$$

(هـ)

$$P(1R\,1W\,1B) = \frac{\binom{8}{1}\binom{3}{1}\binom{9}{1}}{\binom{20}{3}} = \frac{18}{95}$$

(و)

$$P(R_1\,W_2\,B_3) = P(R_1)\cdot P(W_2|R_1)\cdot P(B_3|R_1\,W_2)$$

$$= \left(\frac{8}{20}\right)\left(\frac{3}{19}\right)\left(\frac{9}{18}\right) = \frac{3}{95}$$

مثـال(٢٨) اذا سحبت ورقتان عشوائيا من مجموعة اوراق اللعب ما هو احتمال ان تكون واحدة من كل لون ؟

الحل :

$$P(BA) = \frac{\binom{26}{1}\binom{26}{1}}{\binom{52}{2}}$$

مثـال(٢٩) سحبت ٣ أوراق عشوائيا من مجموعة أوراق اللعب ما هو احتمال :

(أ) ان تكون ثلاثتها Aces

(ب) ان تكون ثلاثتها Spades

الحل :

(أ)

$$P(3A) = \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{1}{5525}$$

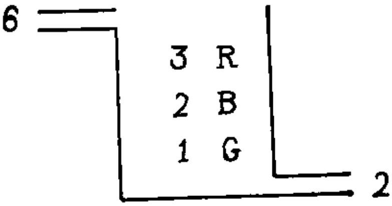
(ب)

$$P(3S) = \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{11}{850}$$

(٧ : ٥) الاحتمال بطريقة الرسم Diagrammatic approach

كثيراً ما تستعمل طريقة الرسم لحساب درجة احتمال الحوادث المركبةCompound events

مثـال(٣٠) : صندوق يحتوي على ٣ كرات حمراء و ٢ زرقاء و ١ خضراء



وسحبت منه كرتان عشوائيا :

ان هذه التجربة تسمى تجربة ذات مرحلتين Two–Stage experiment فحوادث المرحلة الأولى تمثل بخطوط رئيسية ، أما حوادث المرحلة الثانية فتمثل بخطوط فرعية.

فالخطوط الرئيسية هـي :

(١) احتمال الكرة الأولى حمراء $P(R) = 3/6$

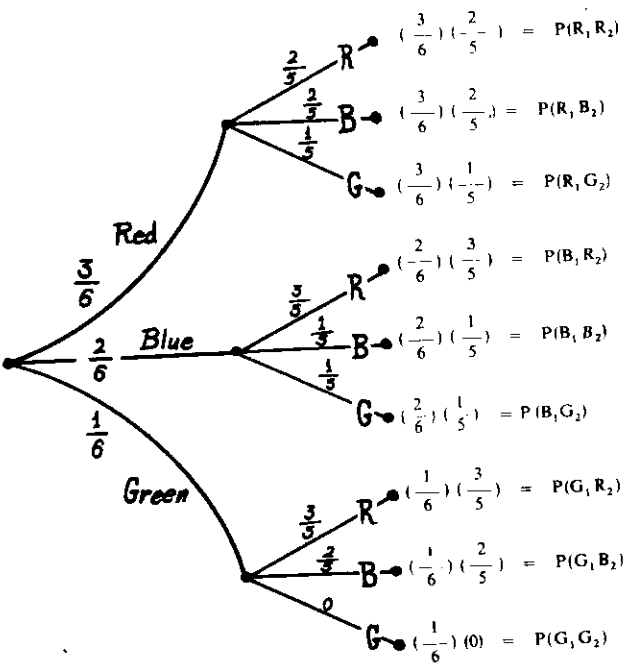
(٢) احتمال الكرة الأولى زرقاء $P(B) = 2/6$

(٣) احتمال الكرة الأولى خضراء $P(G) = 1/6$

أما الخطوط الفرعية فتمثل احتمال شرطي :

فمثلا الفرع الأول للخط الرئيسي الأول هو : احتمال الكرة الثانية حمراء علما بأن الكرة الأولى كانت حمراء .

أما الفرع الثاني للخط الرئيسي الأول هو : احتمال الكرة الثانية زرقاء علما بأن الكرة الأولى كانت حمراء وهكذا



ومن هذه الشجرة : يمكن حساب درجة الاحتمال فمثلا :

(١) ما هو احتمال كون الكرتان زرقا ؟

الحل :

$$P(B_1 B_2) = \left(\frac{2}{6}\right)\left(\frac{1}{5}\right) = \frac{1}{15}$$

(٢) ما هو احتمال كون الأولى زرقاء والثانية خضراء :

الحل :

$$P(B_1 G_2) = \left(\frac{2}{6}\right)\left(\frac{1}{5}\right) = \frac{1}{15}$$

وهكـــذا.....

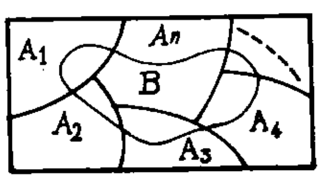
(٧ : ٦) نظرية بيز Bay s theorm

تعتمد هذه النظرية على حساب درجة الاحتمال التقديري الذي يحسب بعد اجراء التجربة ومشاهدة نتائجها .

تعريف (٧ : ٢٠)

اذا كان . $A_1 , A_2 .... A_n$ تمثل حوادث متنافية (مستبعدة) Mutually exclusive وشاملة Exhaustive وان B هو احتمال حادث معين منها بحيث أن احتمال B لايساوي صفراً $(P(B) \neq 0)$ فإن

$$P(A_i \mid B) = \frac{P(A_i) \cdot P(B \mid A_i)}{P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2) + ... + P(A_n) \cdot P(B \mid A_n)}$$



مثــال(٣١) مصنع به ثلاثة أقسام A_1, A_2, A_3 ينتج A_1 . ٥٠٪ من الانتاج الكلي من المصابيح ومتوسط نسبة المعيب له ٣٪ وينتج القسم الثاني (A_2). ٣٠٪ من الانتاج

الكلي ونسبة المعيب ٤٪ بينما القسم الثالث (A_3) فينتج ٢٠٪ من الأنتاج الكلي ونسبة معيب ٥٪.

فاذا أخذنا مصباحاً عشوائياً ودون تحيز من انتاج المصنع كله ووجد بأنه معاب. أوجد أحتمال ان هذا المصباح المعاب هو من انتاج (A_1).

الحل :

نفرض أن الحادث D هو أن المصباح معاب

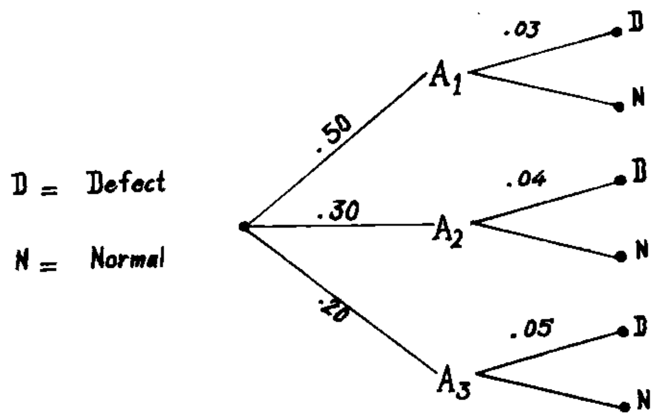
فأحتمال أن المصباح معاب هو

$$P(D) = P(A_1)\cdot P(D|A_1) + P(A_2)\cdot P(D|A_2) + P(A_3)\cdot P(D|A_3)$$

$$\therefore \quad P(A_1|D) = \frac{P(A_1)\cdot P(D|A_1)}{P(A_1)\cdot P(D|A_1) + P(A_2)\cdot P(D|A_2) + P(A_3)\cdot P(D|A_3)}$$

$$= \frac{(\cdot 50)(\cdot 03)}{(\cdot 50)(\cdot 03) + (\cdot 30)(\cdot 04) + (\cdot 20)(\cdot 05)}$$

$$= \frac{15}{37}$$



D = Defect مصاب

N = Normal طبيعي

$$\therefore \quad P(A_1|D) = \frac{(\cdot 50)(\cdot 03)}{(\cdot 50)(\cdot 03) + (\cdot 30)(\cdot 04) + (\cdot 20)(\cdot 05)} = \frac{15}{37}$$

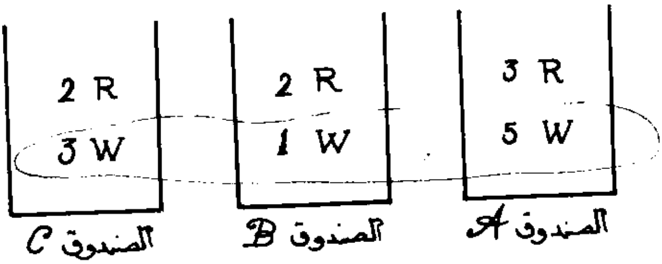
مثال(٣٢) ٣ صناديق تحوي كرات على الشكل التالي :

الصندوق A يحوي على ٣ حمراء و ٥ بيضاء

الصندوق B يحوي على ٢ حمراء ١ بيضاء

الصندوق C يحوي على ٢ حمراء ٣ بيضاء

فإذا انتخب أحد الصناديق عشوائيا وسحبت منه كرة . فإذا كانت الكرة حمراء ما هو احتمال أن تكون من الصندوق A .



الحــــل :

احتمال انتخاب أي صندوق = $\frac{١}{٣}$

$$P(A|R) = \frac{P(A)\cdot P(R|A)}{P(A)\cdot P(R|A) + P(B)\cdot P(R|B) + P(C)\cdot P(R|C)}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{3}{8}\right)}{\left(\frac{1}{3}\right)\left(\frac{3}{8}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{2}{5}\right)}$$

تمارين الفصل السابع

(١) أوجد نتيجة كل مما يأتي :

i) (a) $_{20}P_2$ (b) $_8P_5$ (c) $_7P_5$ (d) $_7P_7$

ii) (a) $_{20}C_2$ (b) $_8C_5$ (c) $_7C_5$ (d) $_7C_7$

(٢) كم عدداً ثلاثياً يمكن تكوينها من الأرقام صفر ، ١ ، ٢ ، ٣ ، ٤ ، ٥ اذا لم يسمح بتكرار الأرقام ؟

(٣) أمتحان مكون من ١٠ أسئلة ومطلوب الأجابة منه على ٦ أسئلة منها فقط فبكم طريقة يمكن للطالب الأجابة على هذا الأمتحان ؟

(٤) كم لجنة سباعية يمكن أختيارها من ٦ كيمياويين و ٥ أطباء على أن تضم ٤ كيمياويين ؟

(٥) رتب ٦ كتب مختلفة من علوم الحيوان biology و ٥ كتب مختلفة من الكيمياء و ٢ كتب مختلفة من الفيزياء على رف مكتبتك بحيث أن كتب كل مجموعة ترتب لوحدها . فكم طريقة يمكنك ترتب ذلك ؟

(٦) صندوق يحوي على ٣ كرات حمراء و ٢ بيضاء و ٤ زرقاء فاذا سحبت منه كرة عشوائياً ماهو أحتمال :

(أ) أن تكون حمراء

(ب) أن تكون غير حمراء

(ج) أن تكون بيضاء

(د) أن تكون حمراء أو زرقاء

(٧) في رمية واحدة لزارين : ماهو احتمال الحصول على مجموع ٧ أو ١١ ؟

(٨) صندوق يحتوي على ٣ كرات حمر و ٧ سود :

فاذا سحبت كرتان فما هو أحتمال

أ- أن يكونا سوداً

ب- أن يكونا بالترتيب التالي : حمراء ثم سوداء

جـ- أن يكونا متعاقبتين باللون ؟

(٩) صندوقان يحتوي الأول على ٥ كرات بيض و٦ حمر و٧ سود والثاني يحتوي على ٥ بيض و٤ حمر و٣ سود . فاذا سحبت كرة واحدة من كل منهما فما هو احتمال :

أ- أن يكونا بيضاء　　　ب- أن يكونا من نفس اللون

(١٠) عند تلقيح نبات هجين طويل مع نبات هجين طويل . ما هوأحتمال الحصول على نبات طويل نقي أو طويل هجين ؟

(١١) عند تزاوج نبات هجين طويل وأحمر مع آخر هجين طويل وأحمر ما هو احتمال الحصول على نبات طويل وأحمر ؟

(١٢) لعب الفريق العراقي لكرة القدم مع الفريق اللبناني ١٥ مرة فربح ٨ منها وخسر ٧ فاذا أتفق أن يلعب الفريقان ٣ مباريات أخرى في دمشق فما هو احتمال :

١) أن يربحا بالتعاقد ؟

٢) أن يربح الفريق العراقي مباراتين ؟

(١٣) احتمال رجل يعيش ٢٥ سنة القادمة = $\frac{٢}{٥}$

واحتمال أن زوجته تعيش نفس المدة = $\frac{٢}{٣}$

فما هو احتمال :

١) أن كليهما يبقيان حيا في هذه المدة ؟

٢) أن يبقى الرجل فقط حيا ؟

٣) أن يموت كلاهما خلال هذه المدة ؟

(١٤) عند رميك زاري الطاولة فما هو احتمالات حصولك على :

(آ) وجهين متماثلين ؟

(ب) وجهين مختلفين ؟

(١٥) صندوق به ٣ كرات بيضا و٧ سودا وسحبت منه كرتان على التوالي . فما هو احتمال حصولك على كرتين من اللون الأبيض اذا :

(آ) اعيدت الكرة الاولى قبل ان تسحب الثانية .

(ب) لم تعد الكرة الأولى قبل أن تسحب الثانية .

(١٦) سحبت ورقتان من مجموعة أوراق اللعب بدون اعادة . فما هو احتمال أن تكونا سبعتان ؟

(١٧) مجموعة من طلبة الصف الأول لكلية الزراعة جامعة بغداد تتألف من ٣٠./ٍ من الطالبات و٧٠./ٍ من الطلاب .
وكانت نسبة الرسوب بينهم في درس الأحصاء هو ٢٠./ٍ و١٠./ٍ على التوالي .
فاذا اختير أحدهم وتبين بأنه من الراسبين فما هو احتمال أن يكون من الطلاب ؟

الفصل الثامن

التوزيع الإحتمالي

Probability Distribution

(٨ : ١) مقدمة

في الفصول السابقة تكلمنا عن التوزيع التكراري للعينات مع ذكر الخواص الوصفية التي تصف هذه التوزيعات .

وفي هذا الفصل سيكون اهتمامنا بالتوزيع التكراري للمجتمع مع ذكر الخواص الوصفية له . ان التوزيع التكراري للعينة هو توزيع تقديري للتوزيع التكراري للمجتمع العائدة له تلك العينة . وعادة يطلق عليه بالتوزيع الأعتباري أوالوصفي Empirical Distribution . وكلما زاد عدد مفردات العينة كان هذا التقدير أقرب الى الواقع . والتوزيع الأحتمالي عبارة عن نموذج رياضي (a Mathematical Model) للتوزيع التكراري الحقيقي للمجتمع .

(٨ : ٢) المتغير العشوائي (المتغير الاحصائي)

Random variable or Statistical variable

في الفصل السابق استخدمنا كلمة تجربة Experiment لوصف أية عملية تؤدي الى ظهور النتائج (Outcomes) كما استعملنا كلمة فضاء العينة (Sample space) للدلالة على جميع النتائج الممكنة للتجربة وقلنا أن كل نتيجة تمثل بنقطة point في فضاء العينة . ولكن غالباً لا نهتم كثيراً بالتفاصيل المتعلقة بكل نقطة أو نتيجة بل يكون اهتمامنا منصباً حول الوصف العددي للنتيجة .

مثال (١) : اذا رمينا قطعة نقود ثلاث مرات (أو رمينا ثلاث قطع نقود مرة واحدة) فهنا يكون فراغ العينة مكوناً من ٢×٢×٢=٨ عناصر أو نقاط كالآتي :

• HHH	• HHT	• HTH	• THH	• HTT	• THT	• TTH	• TTT

ولكن قد يكون اهتمامنا هو عدد مرات ظهور الصورة (H) ففي هذه الحالة فان عدد الصور هي اما صفراً أو ١ أو ٢ أو ٣ فقط فاذا رمزنا لكـل قيمة بالرمز y_i فان :

$$y_i = 0, 1, 2, 3$$

فالرقم صفر مقترن مع النقطةTTTفي فضاء العينة وان ١ مقترن مع النقطةHTTوTHTوTTH وهكذا . من هذا يتضح بأن هناك علاقة تبين كل قيمة من y ونقاط فضاء العينة (ويرمز له بـ E) هي $y = f(E)$

فضاء العينة f(E)	قيمة y
HHH	3
HHT	2
HTH	2
THH	2
HTT	1
THT	1
TTH	1
TTT	0

أي أن كل قيمة من y تعتمد على نقاط معينة من فضاء العينة أو بعبارة أخرى فان y هي دالة نقاط من فضاء العينة.هذه الـ y تسمى المتغير العشوائي .

وكمثال افرض بأن صندوقاً يحوي على ٤ كرات حمرا و٣سوداوسحبت منه كرتان . فاذا رمزنا بالرمز y لعدد الكرات الحمر فان قيم y ستكون كالآتي .

f(E)	y
RR	2
RB	1
BR	1
BB	0

$n + 1 =$ [illegible]

فـ y هذه تسمى المتغير العشوائي . فكلمة متغير تعني بأن y تأخذقيما عددية مختلفة في كل حالة . وكلمة عشوائية (Random) ، تعني بأن قيمة المتغير في أي تجربة تعتمد على نتائج التجربة التي بدورها تعتمد على الصدفة .

تعريف (٨ : ١)

المتغير العشوائي (أو المتغير الاحصائي) هو دالة (Function) قيمها حقيقية Real تحدد بكل عنصر من عناصر فضاء العينة .

مما سبق اتضح بأن اهتمامنا هنابقيم المتغير العشوائي لتجربة بدلا من جميع النتائج الممكنة للتجربة . لذلك فيمكن وضع فضاءعينة خاص به بدلا من فضاء العينة للتجربة.

فمثلاً تجربة رمي قطعة النقود ٣ مرات وان y هي عدد ظهور الصورة (H) فإن :

$$y = 0, 1, 2, 3$$

وان احتمال عدم ظهور صورة،أي P(TTT) هي :

$$P(TTT) = P(T)\cdot P(T)\cdot P(T) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

وان احتمال ظهور صورة واحدة أي :

$$P(HTT) + P(THT) + P(TTH) = \frac{3}{8}$$

وان احتمال ظهور صورتين أي

$$P(HHT) + P(HTH) + P(THH) = \frac{3}{8}$$

وان احتمال ظهور ٣ صور هي :

$$P(HHH) = \frac{1}{8}$$

هذا ويمكن كتابة الاحتمالات السابقة كالآتي

$$P(y = 0) = \frac{1}{8}$$

$$P(y = 1) = \frac{3}{8}$$

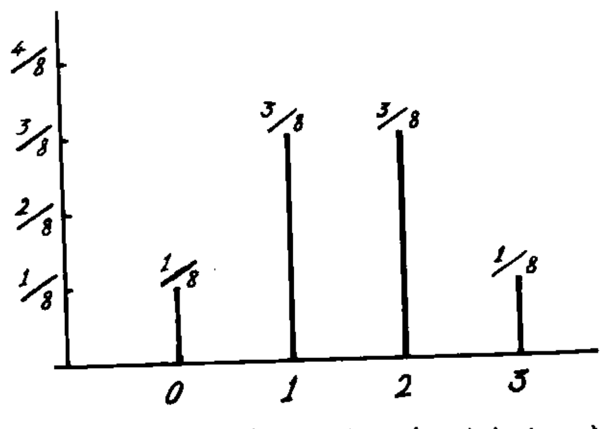
$$P(y = 2) = \frac{3}{8}$$

$$P(y = 3) = \frac{1}{8}$$

اذن فهذه الحوادث المركبة يمكن الآن أن نعاملها وكأنها حوادث بسيطة في فضاء العينة الجديدة المكونة من ٤ نقاط فقط وكل نقطة مقترنة مع قيمة للمتغير العشوائي y . وكل قيمة للمتغير العشوائي مقترنة معه ايضاً بدرجة احتمال المحسوبة اعلاه كما في الجدول التالي :

y	0	1	2	3
P(y)	1/8	3/8	3/8	1/8

والان وبعد أن عينا فضاء العينة للمتغير العشوائي مع درجة احتمال كل قيمة من قيمه فان التوزيع الاحتمالي لهذا المتغير العشوائي قد عرف وبمعرفة هذا القانون يمكن حل جميع المسائل المرتبطة بهذا المتغير العشوائي فالتوزيع الاحتمالي للتجربة السابقة هو كما مبين في الشكل التالي :



شكل (١:٨) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الصورة

مثال (٢): تأمل تجربة رمي زارين : فضاء العينة هنا هو

2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	10
6	7	8	9	10	11
7	8	9	10	11	12

فإذا كان اهتمامنا بمجموع القيم او النقاط على وجهي الزارين عند رميهما فان y تأخذ القيم التالية

$$y = 2, 3, 4, \ldots\ldots, 2$$

وبذلك يمكن وضع فضاء عينة جديد لهذا المتغير العشوائي مع الاحتمال المقترن بكل من هذه القيم :

1	2	3	4	5	6	7	8	9	10

احتمال الحصول على مجموع اثنان هو

$$P(y = 2) = \frac{1}{36}$$

واحتمال الحصول على مجموع ٣ هو

$$P(y = 3) = \frac{2}{36}$$

وهكذا كما مبين ادناه :

$$p(y = 4) = \frac{3}{36}$$

$$P(y = 5) = \frac{4}{36}$$

$$P(y = 6) = \frac{5}{36}$$

$$P(y = 7) = \frac{6}{36}$$

$$P(y = 8) = \frac{5}{36}$$

$$P(y = 9) = \frac{4}{36}$$

$$p(y = 10) = \frac{3}{36}$$

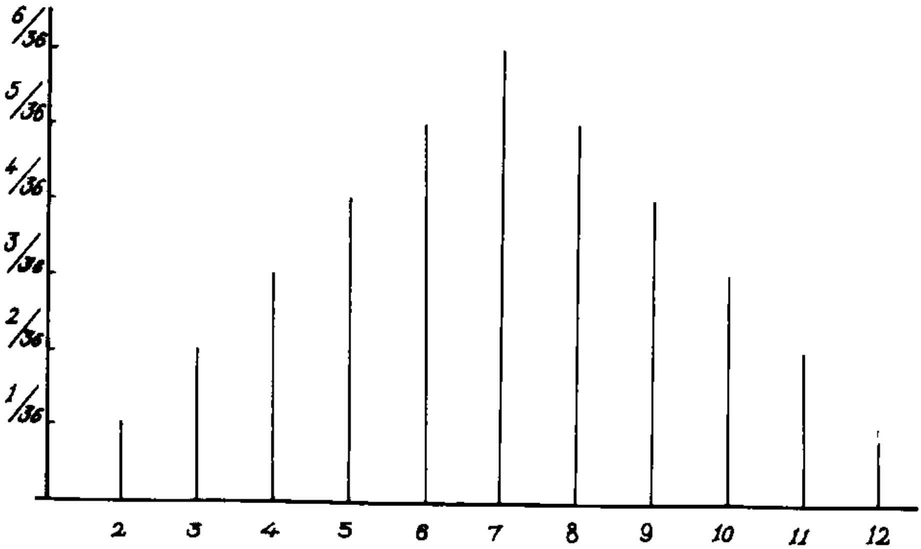
$$p(y = 11) = \frac{2}{36}$$

$$P(y = 12) = \frac{1}{36}$$

فالتوزيع الاحتمالي لـ y هوكالآتي :

y	2	3	4	5	6	7	8	9	10	11	12
P(y)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

اما التمثيل البياني للتوزيع الاحتمالي لهذا المتغير العشوائي فهو



شكل (٨ : ٢) التمثيل البياني للتوزيع الاحتمالي لمجموع النقاط عند رمي زارين .

(٣:٨) التوزيعات الاحتمالية المتقطعة او المنفصلة

Discret Probability Distribution

المتغير العشوائي الذي تختلف قيمه الواحدة عن الأخرى بكميات محدودة معينة يسمى بالمتغير العشوائي المتقطع . وبصورة عامة فإن هذه القيم هي عددية (Countings) وليست قياسية (Measures).

تعريف (٤:٨) :

التوزيع الاحتمالي المتقطع هو جدول او قانون يعطي جميع قيم المتغير العشوائي المتقطع مع جميع الاحتمالات المقترنة مع كل قيمة من قيم المتغير المتقطع . ويرمز لها بـ $f(y)$ بحيث ان :

$$f(y) \geq 0$$

$$\sum f(y_i) = 1$$

وتسمى ايضا Probability density function (p.d.f)

مثال (٣) نفرض ان y تمثل عدد الصور في تجربة رمي قطعتي نقود . فنقاط فضاء العينة مع قيم y هي كالآتي :

قيمة y	فضاء العينة
2	HH
1	HT
1	TH
0	TT

لذا فان y له فضاء العينة التالي $y = 0, 1, 2$

ويمكن حساب احتمال كل قيمة من قيم المتغير العشوائي كالآتي :

$$P(y = 0) = P(TT) = P(T)\cdot P(T) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$$

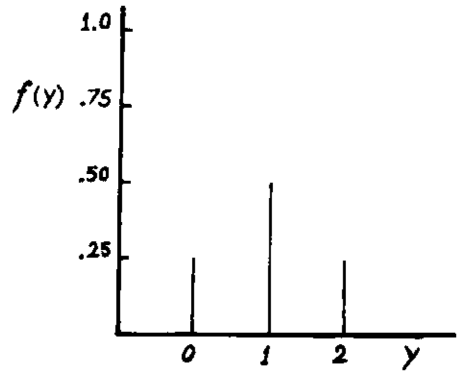
$$P(y = 1) = P(HT + TH) = P(H)\cdot P(T) + P(T)\cdot P(H) = \frac{1}{2}$$

$$P(y = 2) = P(HH) = P(H)\cdot P(H) = \frac{1}{4}$$

فالتوزيع الاحتمالي لـ y هو كالآتي :

y	P(y) = f(y)
0	·25
1	·50
2	·25
المجموع	1·00

والتمثيل البياني لهذا التوزيع هو :



شكل (٨ : ٣) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الصورة عند رمي قطعتي نقود

هذا ومن الجدير بالملاحظة بأن التوزيع الاحتمالي يمكن ان يلخص بإيجاد دالة التوزيع المتجمع (c.d.f) ،

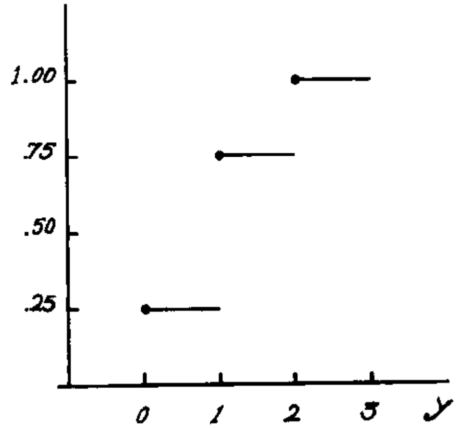
(Cumulative distribution function) = F(y)

التي تعطي الاحتمالات $(Y \leq y_i)$ حيث ان : y_i تمثل أي قيمة من قيم المتغير Y في فضاء العينة فدالة التوزيع المتجمع (c.d.f) للمثال السابق هو كما في الجدول التالي

y	F(y) = c.d.f
0	0·25
1	0·75
2	1·00

$$F(y) = \begin{cases} 0 & \text{عندما تكون} \quad y < 0 \\ \cdot 25 & \text{عندما تكون} \quad 0 \leq y < 1 \\ \cdot 75 & \text{عندما تكون} \quad 1 \leq y < 2 \\ 1\cdot 00 & \text{عندما تكون} \quad y \geq 2 \end{cases}$$

وان التمثيل البياني لدالة التوزيع المتجمع هو



شكل (٨ : ٤) التمثيل البياني لدالة التوزيع المتجمع
لعدد ظهور الصورة عند رمي قطعتي نقود

ومن هذا يتضح بأن التمثيل البياني هذا ماهو الا قفزات غير متصلة كما ان (c.d.f) يقابل (التكرار النسبي المتجمع) في العينة .

مثال (٤) نرمز y لعدد النقاط على وجه الزار

$$\therefore \quad y = 1, 2, 3, 4, 5, 6$$

واحتمال ظهور اية قيمة = $\frac{1}{6}$ أي

$$f(1) = f(2) = f(3) = f(4) = f(5) = f(6) = \frac{1}{6}$$

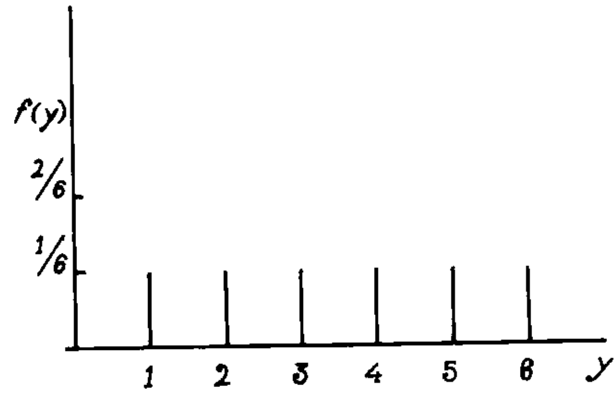
هي p.d.f لان

$$f(y_i) > 0$$

ولأن

$$\sum f(y_i) = 1$$

والرسم البياني لها هو

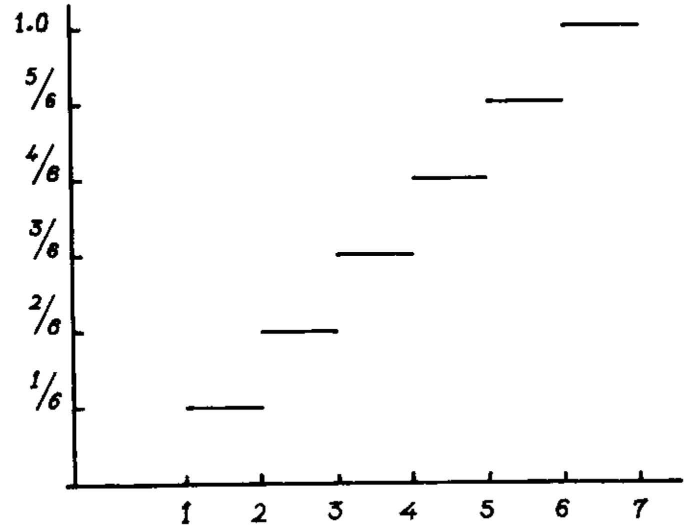


شكل (٨ : ٥) التمثيل البياني للتوزيع الاحتمالي لعدد النقاط على وجه الزهر عند رميه

وال c.d.f لها هو

$$F(y) = \begin{cases} 0 & \text{عندما تكون } y < 1 \\ \frac{1}{6} & \text{عندما تكون } 1 \leq y < 2 \\ \frac{2}{6} & \text{عندما تكون } 2 \leq y < 3 \\ \frac{3}{6} & \text{عندما تكون } 3 \leq y < 4 \\ \frac{4}{6} & \text{عندما تكون } 4 \leq y < 5 \\ \frac{5}{6} & \text{عندما تكون } 5 \leq y < 6 \\ 1 & \text{عندما تكون } y \geq 6 \end{cases}$$

والرسم البياني للـ c. d. f. هو



شكل (٨ : ٦) التمثيل البياني لدالة التوزيع المتجمع لعدد النقاط على وجه الزار عند رميه

مثال (٥) أوجد قانونا (أي p.d.f) للتوزيع الاحتمالي لعدد الصور عند رمي قطعة نقود ٤ مرات

الحل :

بما ان عدد نقاط فضاء العينة هو $16 = 2 \times 2 \times 2 \times 2$

اذن المقام لجميع الاحتمالات هو 16

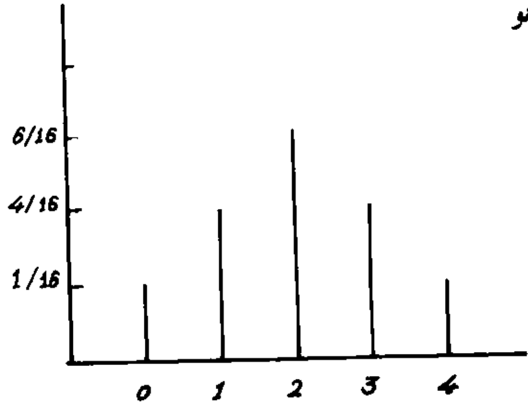
ولإيجاد عدد الطرق للحصول على ٣ وجوه مثلا هو $\binom{4}{3}$

لذلك فبصورة عامة نفرض ان y هي عدد الصور حيث ان

$$y = 0, 1, 2, 3, 4$$

$$\therefore \quad f(y) = \frac{\binom{4}{y}}{16} \qquad y = 0, 1, 2, 3, 4$$

والرسم البياني لها هو



شكل (٨ : ٧) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الصور عند رمي قطعة النقود اربعة مرات

(٨ : ٤) التوزيعات الاحتمالية المستمرة أو دالة كثافة الاحتمال المستمرة

Continuous probability distributions
or probability density function

المتغير العشوائي الذي يأخذ أية قيمة تقع داخل حدود معينة يسمى بالمتغير العشوائي المستمر وبصورة عامة فان هذه القيم هي قياسية (Measures).

> **تعريف (٨ : ٣)**
>
> التوزيع الاحتمالي المستمر هو التوزيع الذي يأخذ فيه المتغير العشوائي قيما بين حدين ودالته (f(y)) تكون موجبة لجميع قيم y بين $-\infty < y < \infty$:
>
> ولأي حدث (A) فان
>
> $$P(A) = P(y \text{ is in } A) = \int_A f(y)\,dy$$

ان دالة الاحتمال هذه تسمى دالة كثافة الاحتمال Probability density function (p.d.f)

ومن خواص هذه الدالة هو :

أن دالة كثافة الاحتمال موجبة (1) $f(y) \geq 0$

ان مجموع المساحة تحت المنحني تساوي واحد . (2) $\int_{-\infty}^{\infty} f(y) = 1$

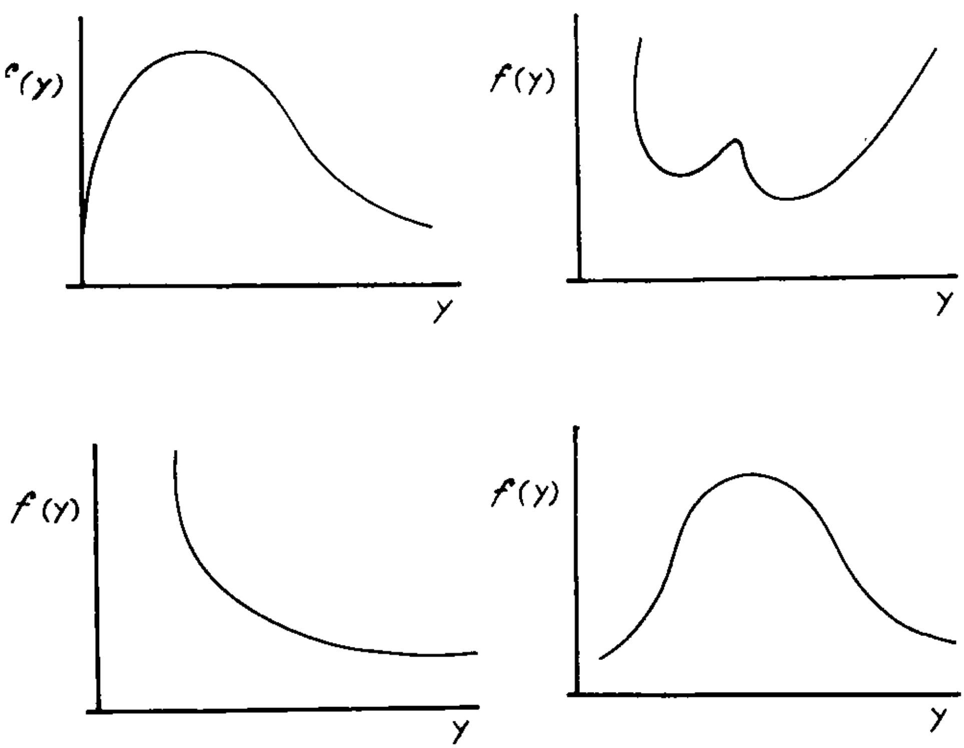
وان المساحة المحصورة بين a و b من منحني دالة الاحتمال هو

$$(3) \qquad P(a < y < b) = \int_a^b f(y)\,dy$$

وأن احتمال y تساوي قيمة معينة = صفر أي :

$$(4) \qquad P(y = a) = \int_a^a f(y)\,dy = 0$$

ان التوزيع الاحتمالي للمتغير العشوائي المستمر له قانون (الذي سميناه p.d.f م) ولكن لايمكن تمثيل التوزيع الاحتمالي المستمر على هيئة جدول كما في التوزيع الاحتمالي للمتغير العشوائي المقطع .
هذا وبما أن y مستمر فان التمثيل البياني f(y) هو مستمر كما في الامثلة التاليـة :

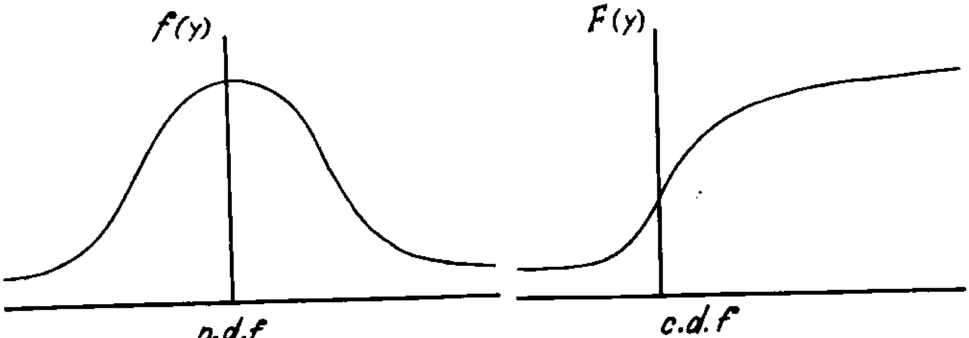


شكل (٨ : ٨) التمثيل البياني لبعض التوزيعات الاحتمالية المستمرة

أما دالة التوزيع المتجمع للمتغير العشوائي المستمر Cumulative distribution function التي تعطي الاحتمالات $(y \le y_i)$ حيث أن : y تمثل أي قيمة من قيم المتغير y في فضاء العينة فهو

$$F(y) = P(y \le y_i) = \int_{-\infty}^{y} f(t)dt$$

والذي يقابل التكرار النسبي المتجمع في العينة .



شكل (٨ : ٩) التمثيل البياني لدالة التوزيع المتجمع لمتغير عشوائي مستمر

هذا ولا يوجد مجال هنا للتوسع في دالة كثافة الاحتمال ودالة التوزيع المتجمع لأنه ليس من اختصاص هذا الكتاب .

(٨ : ٥) القيمة المتوقعة Expected value

او التوقع الرياضي Mathematical expectation

تعريف (٨ : ٤) :

اذا كان المتغير العشوائي y يأخذ القيم $y_1, y_2, \ldots\ldots y_n$ باحتمال $f(y_1), f(y_2), \ldots. f(y_n)$ على التوالي فإن القيمة المتوقعة ل y ويرمز لها بـ $E(y)$ هو

$$E(y) = y_1 f(y_1) + y_2 f(y_2) + \ldots + y_n f(y_n)$$

$$= \sum_{i=1}^{n} y_i f(y_i)$$

فإذا كان التوزيع الاحتمالي منفصلا فإن القيمة المتوقعة هي

$$E(y) = \sum y f(y_i)$$

أما اذا كان التوزيع الاحتمالي مستمرا فإن القيمة المتوقعة هي

$$E(y) = \int_{-\infty}^{\infty} y f(y_i) dy$$

ان القيمة المتوقعة هي في الحقيقة قيمة المتوسط النظري للمجتمع μ اي ان

$$E(y) = \mu$$

أما تباين المجتمع σ^2 فهو

$$\sigma^2 = E(y - \mu)^2$$

مثال (٦) من التوزيع الاحتمالي المتقطع التالي :

y	0	1	2	3	4
f(y)	0·15	0·15	0·35	0·25	0·10

أحسب متوسط المجتمع وتباينه .

الحـل :

متوسط المجتمع هو

$$\mu = E(y) = \sum_{y=0}^{4} y_i f(y)$$

$$= (0)(\cdot 15) + (1)(\cdot 15) + (2)(\cdot 35) + (3)(\cdot 25) + (4)(\cdot 10) = 2$$

أما التباين فهو

$$\sigma^2 = E(y - \mu)^2 = E(y-2)^2$$

$$= \sum_{y=0}^{4} (y-2)^2 f(y)$$

$$= (0-2)^2(\cdot 15) + (1-2)^2(\cdot 15) + (2-2)^2(\cdot 35) + (3-2)^2(\cdot 25) + (4-2)^2(\cdot 10)$$

$$= 1\cdot 4$$

مثال (٧) اذا اشترى رجل بطاقة يانصيب وكان احتمال فوزه بالجائزة الاولى وقدرها ٥٠٠٠ دينار هو ٠,٠٠١ واحتمال فوزه بالجائزة الثانية وقدرها ٢٠٠٠ هو ٠,٠٠٣ . فما هي القيمة العادلة التي يشتري بها بطاقة اليانصيب هذه ؟

الحل :

نرمز للجائزة الاولى بـ y_1 واحتمالها $P(y_1)$

نرمز للجائزة الثانية بـ y_2 واحتمالها $P(y_2)$

$$
\begin{aligned}
E(y) &= y_1(P(y_1) + y_2(P(y_2)) \\
&= (5000)(\cdot 001) + (2000)(\cdot 003) \\
&= 11 \quad \text{دينارا}
\end{aligned}
$$

مثال (٨) اوجد العدد المتوقع من الاولاد للجنة مؤلفة من ٣ انتخبت عشوائيا من ٤ اولاد و٣ بنات .

الحل :

نفرض ان y هو عدد الاولاد في اللجنة . فالتوزيع الاحتمالي لـ y هو

$$
f(y) = \frac{\binom{4}{y}\binom{3}{3-y}}{\binom{7}{3}} \qquad y = 0, 1, 2, 3
$$

$$
f(0) = \frac{1}{35} \qquad f(1) = \frac{12}{35} \qquad f(2) = \frac{18}{35} \qquad f(3) = \frac{4}{35}
$$

$$
\begin{aligned}
\therefore \quad E(y) &= \sum y\, f(y) \\
&= (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = 1\cdot 7
\end{aligned}
$$

لذا فإنه اذا انتخبت لجنة مؤلفة من ٣ لمرات ومرات من ٤ اولاد و٣ بنات فإن هذه اللجنة تحوي في المتوسط على ١,٧ ولد .

بعض خواص التوقع Some properties of expectation

هناك بعض الخواص او القوانين الخاصة بالتوقع ندرجها فيما يلي :

$$
(1) \qquad E(g(y)) = \sum g(y)\, f(y)
$$

واذا كانت a و b ثوابت فإن

(2) $E(a) = a$

(3) $E(ay) = a E(y)$

(4) $E(a y + b) = a E(y) + b$

مثـال(٩) من التوزيع الاحتمالي التالي :

y	8	12	16	20	24
P(y)	$\frac{1}{8}$	1/6	3/8	1/4	1/12

اوجد

(أ) $E(y)$

(ب) $E(y^2)$

(جـ) $E(y-\mu)^2$

الحـــل :

(أ) $E(y)$ هو الوسط الحسابي للتوزيع الاحتمالي اعلاه وهو

$$\mu = E(y) = \Sigma y p(y)$$

$$= (8)\left(\frac{1}{8}\right) + (12)\left(\frac{1}{6}\right) + (16)\left(\frac{3}{8}\right) + (20)\left(\frac{1}{4}\right) + (24)\left(\frac{1}{12}\right)$$

$$= 16$$

(ب) $E(y^2)$ يمثل العزم الثاني حول الصفر أي

$$E(y^2) = \Sigma y^2 p(y)$$
$$= (8)^2\left(\frac{1}{8}\right) + (12)^2\left(\frac{1}{6}\right) + (16)^2\left(\frac{3}{8}\right) + (20)^2\left(\frac{1}{4}\right) + (24)^2\left(\frac{1}{12}\right)$$

$E(y-\mu)^2$ (جـ).

$$E(y-\mu)^2 = \Sigma (y-\mu)^2 p(y)$$

$$= (8-16)^2\left(\frac{1}{8}\right) + (12-16)^2\left(\frac{1}{6}\right) + (16-16)^2\left(\frac{3}{8}\right)$$

$$+ (20-16)^2\left(\frac{1}{4}\right) + (24-16)^2\left(\frac{1}{12}\right)$$

$$= 20$$

وهو تباين التوزيع الاحتمالي

مثـال(١٠) اذا علمت بان

$$f(y) = \binom{3}{y}\left(\frac{1}{2}\right) \qquad y = 0,1,2,3$$

اوجد $E(y^2+1)$

الحـل :

نجد كلاً من f(0) و f(1) و f(2) و f(3) كما في الجدول التالي

y	0	1	2	3
f(y)	1/8	3/8	3/8	1/8

$$\therefore E(y^2+1) = \sum_{y=0}^{3} (y^2+1) f(y)$$

$$= (0^2+1)\left(\frac{1}{8}\right) + (1^2+1)\left(\frac{3}{8}\right) + (2^2+1)\left(\frac{3}{8}\right)$$

$$+ \left(3^2+1\right)\left(\frac{1}{8}\right)$$

تمارين الفصل الثامن

(١) عرف كلاً مما يلي :

(أ) المتغير العشوائي (ب)التوزيع الأحتمالي للمتغير المنفصل

(ج) التوزيع الاحتمالي للمتغير المستمر (د) القيمة المتوقعة .

(٢) صندوق يحوي على ٤ كرات سوداء و ٢ خضراء سحبت ٣ كرات على التوالي وكل كرة مسحوبة ترجع قبل سحب الأخرى .
اوجد التوزيع الاحتمالي لعدد الكرات الخضر .

(٣) اوجد التوزيع الأحتمالي للمتغير العشوائي y الذي يمثل عدد ظهور الصور في تجربة رمي خمسة قطع من النقود مرة واحدة .

(٤) اذا كان التوزيع الأحتمالي للمتغير y هوكما مبين في الجدول التالي :

y	0	1	2	3	4
f(y)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

(أ) أرسم الرسم البياني لهذه الدالة f(y)

(ب)أوجد الوسط الحسابي والتباين ومعامل الاختلاف .

(٥) اذا كان التوزيع الأحتمالي للمتغير العشوائي المستمر y الذي يأخذ قيماً بين ١ = y و y = ٣ هو

$$f(y) = \frac{1}{2}$$

(أ) برهن بأن المساحة تحت المنحني = ١

(ب)أوجد $P(2 < y < 2\cdot5)$

(ج) اوجد $P(\ y \leq 1\cdot6)$

(٦) أوجد الوسط الحسابي للتوزيع الاحتمالي التالي :

$$f(y) = \binom{3}{y}\left(\frac{1}{4}\right)^{y}\left(\frac{3}{4}\right)^{3-y} \qquad y = 0,1,2,3$$

(٧) من التوزيع الأحتمالي التالي :

y	-3	6	9
f(y)	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

احســب مايلي :

(أ) $E(y)$

(ب) $E(y^2)$

(جـ) $E(2y + 1)$

الفصل التاسع

التوزيعات الاحتمالية المتقطعة

Discrete Probability Distributions

(٩ : ١) توزيع ذي الحدين Binomial Distribution

(١) تعريف توزيع ذي الحدين :

يعتبر توزيع ذي الحدين من أهم التوزيعات المتقطعة وسنشرح بعض الأمثلة لتسهيل مفهوم هذا التوزيع .

تأمل حدوث تجربة ما بحيث أن جميع النتائج Outcomes يمكن تصنيفها الى ظهورحادث ما (وليكن A) أو عدم ظهوره . وعادة 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 التي تظهر في تكرار التجربة n من المرات . ان هذا النوع من المتغيرات يسمى متغير ذو حدين فهو مصنف الى صنفين نجاح الحادث أو فشله وهو متقطع لأنه يأخذ قيما عددية (Counts) من صفر الى n .

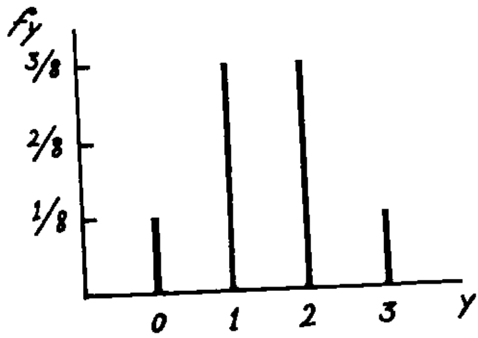
ان تكرار التجربة هنا يكون تكراراً لأصل التجربة في كل مرة أي بمعنى آخر بأن التجارب المتكررة تكون مستقلة .

ففي حالة 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 .

فالتوزيع الاحتمالي للمتغير ذي الحدين هذا هو كما يلي :

p (y)	احتمال كل حالة ممكنة	الحالات الممكنة	عدد ظهور الحادث y
$\frac{1}{8}$	$\frac{1}{8}$	TTT	0
$\frac{3}{8}$	$\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$	HTT THT TTH	1
$\frac{3}{8}$	$\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$	HHT HTH THH	2
$\frac{1}{8}$	$\frac{1}{8}$	HHH	3
1·0	1·0		المجموع

والتمثيل البياني لهذا التوزيع هوكما يلي :



شكل (٩:١) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الصورة (H) عند رمي قطعة النقود ثلاث مرات

وكمثال آخر افترض التجربة التالية :

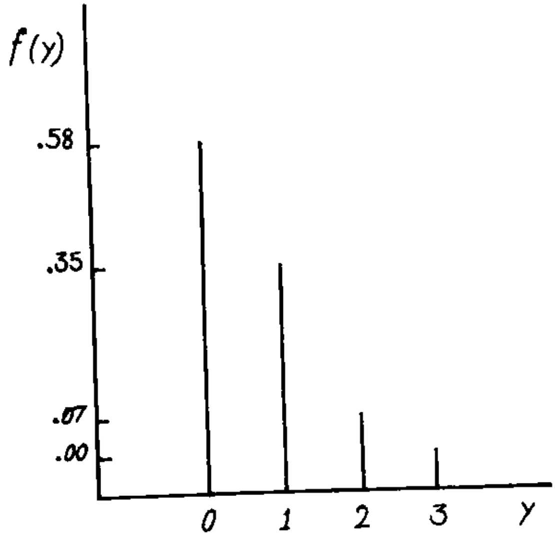
رمي زار ثلاث مرات $\therefore n = 3$

نفرض أن النجاح هنا هو الحصول على الوجه الذي عليه نقطتان ونرمز له بـ S . فالمتغير العشوائي y هو يمثل عدد الاوجه التي تحمل (نقطتان) التي تظهر في الثلاث رميات . فإذا رمزنا بـ F (للفشل) اي الأوجه التي لا يظهر عليها (نقطتان) .

فالتوزيع الاحتمالي لهذا المتغير ذي الحدين كما يلي :

عدد ظهور الحادث Y	الحالات الممكنة	احتمال كل حالة ممكنة	P(y)
0	FFF	$\left(\frac{5}{6}\right)^3$	$\left(\frac{5}{6}\right)^3$
1	SFF	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$	$3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$
	FSF	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$	
	FFS	$\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$	
2	SSF	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	$3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$
	SFS	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	
	FSS	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	
3	SSS	$\left(\frac{1}{6}\right)^3$	$\left(\frac{1}{6}\right)^3$

والتمثيل البياني للتوزيع الاحتمالي هذا هو :



شكل (٩ : ٢) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الوجه الذي يحمل نقطتين في تجربة رمي زار ثلاث مرات

من هذا يتضح بان التكتيك المستخدم في المثالين السابقين ممكن استخدامه لأي مسألة خاصة بتوزع ذي الحدين ويمكن تلخيص ذلك :

(١) اولاً : بايجاد فضاء العينة للتجربة كاملة .
فاذا كان هناك ٥ تكرارات للتجربة فسيكون هناك $(٢)^{٥}$ = ٣٢ نقطة في فضاء العينة .

(٢) ثانيا : إيجاد احتمال كل نتيجة من النتائج المحتملة .

(٣) ثالثا : تعيين القيمة الخاصة للمتغير العشوائي المقترنة بكل نقطة من فضاء العينة .

(٤) رابعا : جمع جميع الاحتمالات للنقاط التي تمثل تلك القيمة الخاصة بالمتغير العشوائي .

(٥) خامسا : هذه الاحتمالات تعطي التوزيع الاحتمالي للمتغير العشوائي ذي الحدين .

ان من الجدير بالملاحظة أنه أحيانا يكون من الصعوبة عمل جميع هذه الحسابات لكل مسألة من مسائل توزيع ذي الحدين لذلك فن المستحسن إيجاد قانون عام لهذا التوزيع الذي يعطي أية احتمال من احتمالات توزيع ذي الحدين .

تعريف (٩ : ١) :

في التجارب المتكررة n من المرات والمستقلة والتي تصنف نتائجها الى صنفين :

نجاح (ظهور الحادث) أو الفشل (عدم ظهور الحادث) .

فإذا رمزنا لاحتمال وقوع النجاح بـ p والفشل بـ q بحيث : $p + q = 1$

ورمزنا للمتغير العشوائي y بعدد النجاح

فإن احتمال ظهور الحادث y عدد من المرات في n من التجارب أو المحاولات يمكن حسابه بقانون توزيع ذي الحدين التالي :

$$\boxed{P(y = y_0) = \binom{n}{y} p^y q^{n-y}} \qquad y = 0, 1, 2 \ldots, n$$

والمتغير y يقال له بأنه يتوزع توزيعا ذي حدين .

فتجارب ذي الحدين تتميز بما يلي :

(١) ان التجربة تتكرر n من المرات.

(٢) التجارب المتكررة هذه تكون لتجربة الاصل اي مستقلة .

(٣) نتيجة كل تجربة أما أن الحادث ينجح (يظهر) أو يفشل (لا يظهر) .

(٤) احتمال نجاح الحادث يرمز بـ p (وفشله بـ q) ويبقى ثابتا من تجربة لأخرى .

مثـال(١) في تجربة رمي قطعة النقود ٣ مرات :

$$n = 3 \quad , \quad p = 1/2 = q$$

فاحتمالات $y = 3 \ , \ y = 2 \ , \ y = 1, y = 0$

يمكن ايجادها مباشرة بتطبيق القانون كالآتي :

$$P(y = 0) = \binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$P(y = 1) = \binom{3}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

$$P(y = 2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8}$$

$$P(y = 3) = \binom{3}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \frac{1}{8}$$

مثـال(٢) كذلك في تجربة رمي الزار ثلاث مرات فان :

$$p = \frac{1}{6}$$

$$q = \frac{5}{6}$$

$n = 3$

$y = 0, 1, 2, 3$ عدد ظهور الوجه الذي يحمل نقطتان

فيمكن ايجاد احتمالات المتغير العشوائي مباشرة بالقانون كالآتي :

$$P(y = 0) = \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 = \left(\frac{5}{6}\right)^3$$

$$P(y = 1) = \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = 3\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2$$

$$P(y = 2) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1$$

$$P(y = 3) = \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 = \left(\frac{1}{6}\right)^3$$

مثـال(٣) اذا كان احتمال اصابة لاعب كرة السلة A الهدف في رمية حرة هو $\frac{٣}{٤}$ فما هو احتمال اصابته الهدف مرتين من ٤ رميات حرة ؟

الحـــل :

$p = \frac{3}{4}$ ، $q = \frac{1}{4}$

$n = 4$ ، $y = 2$

$$P(y = 2) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = 0.211$$

0.0625

0.5625

مثــالَ(٤) : في عائلة مكونة من ٥ أطفال :
احسب احتمال أن يكون بينهم ٣ ذكورعلماً بأن نسبة الذكورالى الاناث ١ : ١
الحــل :

$$p = \frac{1}{2} \quad , \quad q = \frac{1}{2}$$

$$n = 5 \quad , \quad y = 3$$

$$P(y = 3) = \binom{5}{3}\left(\frac{1}{2}\right)^{3}\left(\frac{1}{2}\right)^{2}$$

مثــال(٥) : وجد عند مفترق طريقيــن بــأن $\frac{٢}{٣}$ السيارات تتجه الى اليمين و $\frac{١}{٣}$ السيارات تتجه الى اليسار .
فاذا وصلت ٤ سيارات الى المفترق فما هو احتمال اتجاه ٣ منها الى اليســـار ؟
الحــل :

$$p = \frac{1}{3} \quad , \quad q = \frac{2}{3}$$

$$n = 4 \quad , \quad y = 3$$

$$P(y = 3) = \binom{4}{3}\left(\frac{1}{3}\right)^{3}\left(\frac{2}{3}\right)^{1}$$

مثال (٦) : وجد في أحد المصانع بأن نسبة العلب التالفة (المعابة) في معجون الطماطة التي ينتجها المصنع هي ٥ ٪ : فاذا أخذت عينة مؤلفة من ١٠ علب ، أحسب احتمال :
(أ) أن تكون العينة كلها تالفة.
(ب) أن تكون العينة كلها جيدة.
(ج) أن يكون بالعينة ٣ علب تالفة فقط .
الحــل :
(أ)

$$p = 0.05 \quad , \quad q = 0.95$$

$$n = 10 \quad , \quad y = 10$$

$$\therefore P(y = 10) = \binom{10}{10}(.05)^{10}(.95)^{0} \longrightarrow 10 - 10 = 0$$

(ب)

$p = \cdot 95 \quad q = \cdot 05$

$n = 10 \quad y = 10$

$$P(y = 10) = \binom{10}{10}(\cdot 05)^{10}(\cdot 95)^{0}$$

(جـ)

$p = \cdot 05 \qquad n = 10$

$q = \cdot 95 \qquad y = 3$

$$\therefore \quad P(y = 3) = \binom{10}{3}(\cdot 05)^{3}(\cdot 95)^{7}$$

مثال (٧) : في احدى تجارب مندل الوراثية وجد بأن احتمال الحصول على نبات طويل يساوي $\frac{٣}{٤}$ وعلى نبات قصير يساوي $\frac{١}{٤}$ في الجيل الثاني فاذا فحصت عينة مؤلفة من ٤ نباتات فما هو احتمال :

(أ) ان تكون كلها طويلة .

(ب) نبات واحد قصير فقط.

الحل :

(أ)

$p = \frac{3}{4} \quad , \quad q = \frac{1}{4}$

$n = 4 \quad , \quad y = 4$

$$\therefore P(y = 4) = \binom{4}{4}\left(\frac{3}{4}\right)^{4}\left(\frac{1}{4}\right)^{0}$$

(ب)

$p = \frac{1}{4} \quad , \quad q = \frac{3}{4}$

$n = 4 \quad , \quad y = 1$

$$\therefore P(y = 1) = \binom{4}{1}\left(\frac{1}{4}\right)^{1}\left(\frac{3}{4}\right)^{3}$$

(٢) دالة التوزيع المتجمع c. d. f للتوزيع ذي الحدين :

ان اكثر مايستخدم توزيع ذي الحدين في الاستنتاج الاحصائي هوعلى شكل تراكمي ، لذا فدالة التوزيع المتجمع لـ c. d. f هي :

$$\boxed{P(y \leq y_0) = \sum_{y=0}^{n}\binom{n}{y}p^{y}q^{n-y}}$$

وهو احتمال أن y تساوي أوأصغر من قيمة معينة y_o
كما يمكن رسم تمثيل بياني لـ c.d.f

مثال (٨) : في تجربة رمي قطعة النقود ٤ مرات
فاذا رمزنا y لعدد ظهور الصورة في كل رمية ،
فان :

$$n = 4$$

$$p = \frac{1}{2} \quad , \quad q = \frac{1}{2}$$

$$y = 0, 1, 2, 3, 4$$

فالتوزيع الاحتمالي لجميع قيم y هوكالآتي :

$$P(y=0) = \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$$P(y=1) = \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 = \frac{4}{16}$$

$$P(y=2) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{6}{16}$$

$$P(y=3) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{4}{16}$$

$$P(y=4) = \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0 = \frac{1}{16}$$

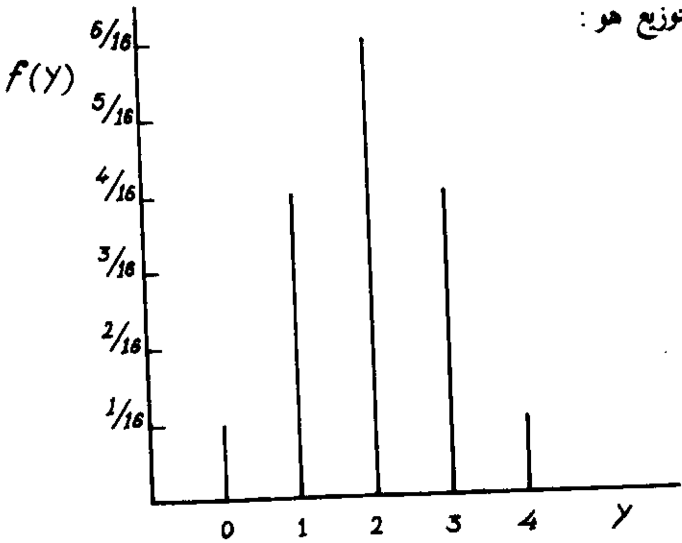
أولاً: لاحظ بأن مجموع الاحتمالات هذه يجب أن = ١ لان مجموع جميع الاحتمالات للتوزيع = ١
أي ،

$$\sum_{y=0}^{4} \binom{4}{y}\left(\frac{1}{2}\right)^y\left(\frac{1}{2}\right)^{4-y} = ١$$

كما مبين في الجدول التالي :

y	P(y)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16
المجموع	1·00

والتمثيل البياني لهذا التوزيع هو :



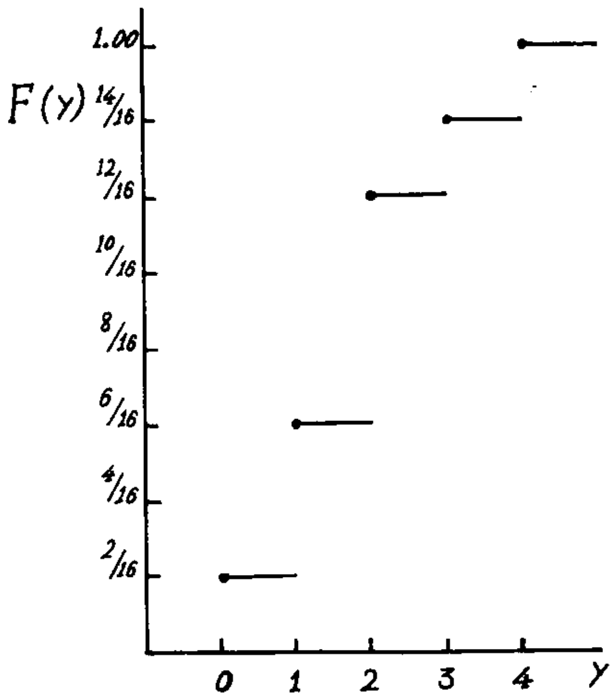
شكل (٩ : ٣) التمثيل البياني للتوزيع الاحتمالي لعدد ظهور الصورة عند رمي قطعة النقود اربعة مرات

أما دالة التوزيع المتجمع c. d. f لهذه التجربة هو :

y	c.d.f
0	1/16
1	5/16
2	11/16
3	15/16
4	1·00

$$F(y) = \begin{cases} 0 & \text{when} \quad y \leq 0 \\ 1/16 & \text{when} \quad 0 \leq y \leq 1 \\ 5/16 & \text{when} \quad 1 \leq y \leq 2 \\ 11/16 & \text{when} \quad 2 \leq y \leq 3 \\ 15/16 & \text{when} \quad 3 \leq y \leq 4 \\ 1.0 & \text{when} \quad y \geq 4 \end{cases}$$

والتمثيل البياني لدالة التوزيع المتجمع هو :



شكل (٤:٩) التمثيل البياني لدالة التوزيع المتجمع لعدد ظهور الصورة عند رمي قطعة النقود اربعة مرات

والآن يمكن تلخيص بعض الحالات المهمة كما يلي :

نفرض بأن : $n = 5$

1. $P(y=2) = \binom{5}{2} p^2 q^3$
2. $P(y \leq 2) = P(y=2) + P(y=1) + P(y=0)$
3. $P(y > 2) = P(y=3) + P(y=4) + P(y=5)$
4. $P(2 \leq y < 4) = P(y=2) + P(y=3)$

(٣) أمثلة متنوعة :

مثال (٩) : في عائلة مؤلفة من ٤ أطفال احسب أحتمال :

(ا) على الأقل فيها طفل ذكر واحد

(ب) على الأكثر ٢ ذكر

الحل :

(أ)

$$p = \frac{1}{2} \quad , \quad q = \frac{1}{2}$$

$$n = 4$$

$$\therefore \quad P(y \geq 1) = P(y=1) + P(y=2) + P(y=3) + P(y=4)$$

$$= \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 + \binom{4}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^0$$

$$= \frac{1}{4} + \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{15}{16}$$

(ب)

$$P(y \leq 2) = P(y = 2) + P(y = 1) + P(y = 0)$$

$$= \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + \binom{4}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^3 + \binom{4}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^4$$

$$= \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{11}{16}$$

مثال (١٠) : اذا كان احتمال مريض يشفى من مرض ما = ٠.٤

فاذا دخل المستشفى ٥ مرضى مصابين بهذا المرض فما هو احتمال :

(أ) لايشفى منهم احد

(ب) يشفى واحد منهم فقط

(ج) يشفى واحد منهم على الأقل

الحل :

(أ)

$p = \cdot 4 \quad , \quad q = \cdot 6$

$n = 5 \quad , \quad y = 0$

$$\therefore P(y = 0) = \binom{5}{0} (\cdot 4)^0 (\cdot 6)^5 = \cdot 08$$

(ب)

$p = \cdot 4 \quad , \quad q = \cdot 6$

$n = 5 \quad , \quad y = 1$

$$P(y = 1) = \binom{5}{1} (\cdot 4)^1 (\cdot 6)^4 = \cdot 26$$

(جـ)

$p = \cdot 4 \quad , \quad q = \cdot 6$

$n = 5 \quad y \geq 1$

$$P(y \geq 1) = P(y=1) + P(y=2) + P(y=3) + P(y=4) + P(y=5)$$

$$= \binom{5}{1}(\cdot 4)^1 (\cdot 6)^4 + \binom{5}{2}(\cdot 4)^2 (\cdot 6)^3 + \binom{5}{3}(\cdot 4)^3 (\cdot 6)^2 +$$

$$\binom{5}{4}(\cdot 4)^4 (\cdot 6)^1 + \binom{5}{5}(\cdot 4)^5 (\cdot 6)^0$$

$$= \cdot 92$$

لاحظ أن في توزيع ذي الحدين فان عدد المرات التي نسميها نجاحاً زائداً عدد المرات التي نسميها فشلاً يساوي n وان احتمال عدد مرات النجاح y هو :

$$\binom{n}{y} p^y q^{n-y}$$

وهو يساوي عدد مرات الفشل : $x = n - y$

$$\binom{n}{n-y} p^y q^{n-y} = \binom{n}{y} p^y q^{n-y}$$

(٤) خواص توزيع ذي الحدين :

(أ) الوسط الحسابي :

الوسط الحسابي لتوزيع ذي الحدين ويرمز له بـ μ أو $E(y)$ هو عبارة عن عدد النجاحات

المتوقعة والتي يمكن الحصول عليها في n من الحالات :

$$\boxed{\mu = E(y) = \sum_{y=0}^{n} y\,P(y)}$$

مثال (١١) : الوسط الحسابي لتجربة رمي قطعة النقود ٤ مرات :

y	P(y)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

هو

$$\mu = E(y) = \sum_{y=0}^{4} y\,P(y)$$

$$= 0\left(\frac{1}{16}\right) + 1\left(\frac{4}{16}\right) + 2\left(\frac{6}{16}\right) + 3\left(\frac{4}{16}\right) + 4\left(\frac{1}{16}\right)$$

$$= 2$$

وطبعاً هذه الطريقة طويلة ومملة اذا كان عدد النقاط في فضاء العينة كبيراً جداً .

ان من السهولة برهنة أن الوسط الحسابي لتوزيع ذي الحدين هو : $\boxed{\mu = np}$

البرهان :

$$\mu = \sum_{y=0}^{n} y\,p(y)$$

$$= \sum_{y=0}^{n} y\binom{n}{y} p^{y} q^{n-y}$$

$$= \sum_{y=1}^{n} y\binom{n}{y} p^{y} q^{n-y} + \underbrace{0\binom{n}{0} p^{0} q^{n-0}}$$

$$= \sum_{y=1}^{n} y\,\frac{n!}{y!\,(n-y)!}\,p^{y} q^{n-y}$$

$$= \sum_{y=1}^{n} y \frac{n(n-1)!}{y(y-1)!(n-y)!} (p)(p)^{y-1}(q)^{n-y}$$

$$= np\sum \frac{(n-1)!}{(y-1)(n-y)!} p^{y-1} q^{n-y}$$

فاذا فرضنا بأن $y = k + 1$ لذا فان $k = y - 1$،

أي أن

$$\mu = np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k q^{n-1-k}$$

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k}$$

$$= np(1)$$
$$= np$$

لذلك ففي المثال السابق :

$$n = 4$$

$$p = \frac{1}{2}$$

$$\therefore \mu = np = (4)\left(\frac{1}{2}\right) = 2$$

كالسابق .

مثال (١٢) : اذا كان نسبة المعيب في وحدة انتاج مصنع ما هو ١٠ ./. أوجد الوسط الحسابي للمعيب في ٤٠٠ وحدة انتاج .

الحل :

$$p = 0{\cdot}10$$
$$n = 400$$
$$\therefore \mu = np$$
$$= (400)({\cdot}10)$$
$$= 4$$

أي نتوقع أن تكون من بين ٤٠٠ وحدة انتاج،انتاج ٤٠ وحدة معيبة .

(ب) التباين

التباين بصورة عامة هو :

$$\sigma^2_y = E(y^2) - [E(y)]^2.$$

$$\boxed{= \sum_{y=0}^{n} y^2 p(y) - \mu^2}$$

مثال (١٣) : التباين لتجربة رمي قطعة النقود ٤ مرات هو :

$$\sigma^2_y = (0)^2\left(\frac{1}{16}\right) + (1)^2\left(\frac{4}{16}\right) + (2)^2\left(\frac{6}{16}\right) + (3)^2\left(\frac{4}{16}\right) + (4)^2\left(\frac{1}{16}\right) - (2)^2 = 1$$

هذا ويمكن ايجاد التباين لتوزيع ذي الحدين بأنه :

$$\boxed{\sigma^2_y = npq}$$

وللمثال السابق :

$n = 4$

$p = \frac{1}{2} \qquad q = \frac{1}{2}$

$$\therefore \sigma^2_y = npq = (4)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 1$$

تعريف (٩:٢) :

الوسط الحسابي والتباين لمتغير يتوزع توزيعاً ذي حدين هوعلى التوالي :

$$\mu = np$$
$$\sigma^2 = npq$$

هذا وهناك بعض الخواص الأخرى لتوزيع ذي الحدين مثلاً :

معامل الالتواء الثالث α_3 هو :

$$\boxed{\alpha_3 = \frac{q-p}{\sqrt{npq}}}$$

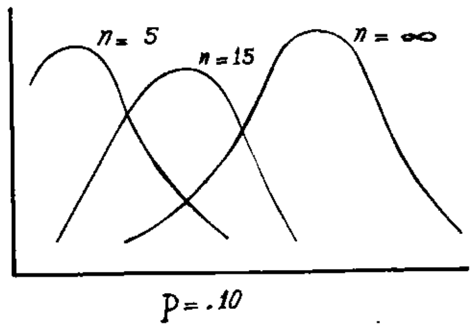
ومعامل التفلطح β هو :

$$\boxed{\beta = 3 + \frac{1-6pq}{npq}}$$

(٥) الشكل العام لمنحني توزيع ذي الحدين

يعتمد شكل منحني توزيع ذي الحدين على n و p.
فاذا كانت $p = q = .5$ فان المنحني يكون متماثلاً (بغض النظر عن قيمة n).
أما اذا كانت $p \neq q$ فان المنحني يكون غير متماثل .
فعندما تكون $p<q$ فمنحني التوزيع يكون ملتويا التواءً موجبا الى اليمين .

وعندما تكون $p \geq q$ فمنحني التوزيع يكون ملتوياً التواءً سالباً الى اليسار .
وعندما تكون n قريبة من المالانهاية ($n \to \infty$)
فان المنحني يقترب من التماثل بغض النظر عن قيمة كل من p و q .



شكل (٩:٥) ثلاث منحنيات لتوزيع ذي حدين

(٩ : ٢) التوزيع الاحتمالي المتعدد الحدود
Multinomial Probability Distributions

تعريف : (٩ : ٣)

في التجارب المتكررة n من المرات والمستقلة والتي تصنف نتائجها الى k من النتائج $(E_1, E_2, \dots\dots E_k)$ وباحتمالات $p_1, p_2, \dots\dots p_k$ على التوالي فان التوزيع الاحتمالي للمتغيرات العشوائية $y_1, y_2 \dots\dots, y_k$ والتي تمثل عدد ظهور الـ $E_1, E_2 \dots\dots, E_k$ هو :

$$P(y = y_1, y_2, \dots y_k) = \binom{n}{y_1, y_2, \dots y_k} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$$

علماً بأن :

$$\Sigma y_i = n$$
$$\Sigma p_i = 1$$

مثال (١٤) : احدى نظريات الوراثة تقول بأنه عند تهجين معين بين بعض الحيوانات ينتج اللون الأحمر والاسود والأبيض في الجيل الثاني بنسبة ٨ : ٤ : ٤ على التوالي . فاذا تم اختيار ١٠ حيوانات من الجيل الثاني فما هو احتمال ظهور ٥ حمراء و٣ سوداء و٢ بيضاء

الحل

$$p(y_1 = 5, y_2 = 3, y_3 = 2) = \frac{10!}{5!\,3!\,2!} \left(\frac{8}{16}\right)^5 \left(\frac{4}{16}\right)^3 \left(\frac{4}{16}\right)^2$$

مثال (١٥) : صندوق يحتوي على ٥ كرات حمر و٦ كرات بيضاء و ٨ كرات سوداء .
فاذا سحبت كرة وسجل لونها ثم اعيدت الى الصندوق وكررت هذه العملية ٥ مرات .
ماهو احتمال ان تحصل كرتين حمراء وواحدة بيضاء و ٢ سوداء ؟

الحـــل :

$$p_1 = P(R) = \frac{5}{20}$$

$$p_2 = P(W) = \frac{6}{20}$$

$$p_3 = P(B) = \frac{9}{20}$$

$$p(y_1 = 2, y_2 = 1, y_3 = 2) = \frac{5!}{2!1!2!} \left(\frac{5}{20}\right)^2 \left(\frac{6}{20}\right)^1 \left(\frac{9}{20}\right)^2$$

تعريف : (٩ : ٤)
الوسط الحسابي والتباين للتوزيع الاحتمالي المتعدد الحدود هو على التوالي:

$$\mu_i = np_i$$

$$\sigma^2 y_i = np_i q_i$$

(٩ : ٣) التوزيع الهندسي الزائدي (او المفرط)

Hypergeometric Distribution

في هذا النوع من التوزيع يكون احتمال وقوع الحادث متغيرا لأن التجارب المتكررة هنا غير مستقلة .

فمثلاً عند سحب كرات من صندوق بدون اعادة فمثل هذه التجارب تكون غير مستقلة فاحتمال وقوع الحادث يتغيرفي كل تجربة أومن حالة الى أخرى .

تعريف : (٩ : ٥)

في التجارب المتكررة غير المستقلة وحجمها N والحاوية على K نجاح و$N-K$ فشل فالتوزيع الاحتمالي للمتغير العشوائي الهندسي الزائدي (عدد النجاحات في عينة من n) هو

$$P(y) = \frac{\binom{K}{y}\binom{N-K}{n-y}}{\binom{N}{n}} \qquad y = 0, 1, 2, \ldots\ldots n$$

مثـال(١٦) : اذا سحبت ٥ أوراق من مجموعة ورق اللعب (٥٢ ورقة) ما هو احتمال ان يكون بينها ٣ قلب Hart ؟

الحـل :

$$N = 52$$
$$n = 5$$

عدد اوراق القلب في مجموعة ورق اللعب = ١٣

$$K = 13$$
$$y = 3$$

$$\therefore P(y = 3) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} = 0.0815$$

تعريف : (٩ : ١٠)

الوسط الحسابي والتباين للتوزيع الهندسي الزائدي هو على التوالي :

$$\boxed{\mu = \frac{nk}{N}}$$

$$\sigma^2 = \frac{N-n}{N-1}\, n\, \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

هذا ويمكن تعميم التوزيع الهندسي الزائدي الى

$$P(y_1, y_2, \dots y_k) = \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}\binom{n_3}{y_3}\dots\binom{n_k}{y_k}}{\binom{N}{n}}$$

علما بأن

$$\sum y_i = n$$
$$\sum n_i = N$$

(٩ : ٤) التوزيع البواسوني Poisson Distribution

في بعض التجارب قد يحدث المتغير العشوائي في جزء من فترةأووقت محدودة. أو في منطقة صغيرة فمثلاً عدد المكالمات التلفونية في الساعة المستعملة من قبل دائرة ما ، عدد أيام تعطيل المدارس بسبب سقوط الثلوج ، أوعدد اللعب المؤجلة في كرة القدم بسبب الأمطار ،

أو عدد الفئران / الدونم في حقول الذرة أو عدد البكتريا في محلول معين . أو عدد الاخطاء عند طبع صفحة معينة وهكذا ...

فالتجارب التي تؤدي الى ظهور مثل هذا المتغير العشوائي تسمى تجارب بواسون Poisson experiments.

وهذه التجارب تتصف بما يأتي :

(١) متوسط عدد ظهور « النجاحات » μ معروف.

(٢) احتمال نجاح واحد خلال فترة قصيرة وفي منطقة صغيرة يتناسب مع طول الوقت وحجم المجال .

(٣) احتمال أكثر من نجاح واحد في مثل هذه الفترة القصيرة والمنطقة الصغيرة هو احتمال نادر .

تعريف (٩ : ١١)

المتغير العشوائي البواسوني هو يمثل عدد النجاحات في التجربة البواسونية.

تعريف (٩ : ١٢)

التوزيع الاحتمالي للمتغير العشوائي البواسوني y الذي يمثل عدد النجاحات التي تحدث في فترة وقت محددة ومنطقة معينة هو

$$P(y) = \frac{e^{-\mu}\,\mu^{y}}{y!} \qquad y = 1,2,.......$$

حيث ان

$$\text{متوسط عدد النجاحات} = \mu$$

$$e = 2.71828$$

مثال (١٧) : اذا كان متوسط عدد الأيام التي تعطل فيها الدراسة في مدرسة معينة في مدينة ما بسبب نزول الثلوج في فصل الشتاء هو ٤ .
ماهو احتمال أن المدارس في هذه المدينة ستعطل فيها الدراسة لمدة ٦ أيام خلال الشتاء ؟

الحـــل :

$$y = 6$$
$$\mu = 4$$
$$P(y=6) = \frac{e^{-\mu}\,\mu^{y}}{y!}$$
$$= \frac{e^{-4}\,4^{6}}{6!}$$
$$= 0.1042$$

مثال (١٨) : اذا كان متوسط عدد الفئران / دونم في حقل الذرة المؤلف من ٥٠ دونما هو ١٠ . احسب احتمال دونم معين يحوي على أكثر من ١٥ فارة .

$$P(y > 15) = 1 - P(y \le 15)$$
$$= 1 - \sum_{y=0}^{15} P(y)$$
$$= 0.04\,87$$

تعريف : (٩ : ١٣)

الوسط الحسابي أو التباين للتوزيع البوسوني هو μ .

(٩ : ٥) توزيع ذي الحدين السالب Negative Binomial Distribution

أفرض بأن تجربة لها نفس خواص تجربة توزيع ذو الحدين ماعدا أن المحاولة تكرر حتى ظهور عدد ثابت في النجاح . لذلك فبدلاً من ايجاد الاحتمال لـ y من النجاحات في n من المحاولات فنحن الآن نهتم في ايجاد الاحتمال لظهور النجاح الـ k^{th} الذي ظهر في المحاولة الـ y^{th} .

هذه الانواع من التجارب تدعى تجارب ذي الحدين السالب.

مثال (١٩) : نفرض بأن لاعب كرة نسبة اصابته للهدف هو ٦٠ ./. . نحن الان نهتم في ايجاد الاحتمال بأن اصابته الخامسة للهدف ستحدث في محاولته السابعة .
فاذا رمزنا لاصابته الهدف بالنجاح S وللفشل بـ F فاحدى الترتيبات هي SFSSSFS ونستطيع ان نعمل كل الترتيبات Ss′ , Fs′ ماعدا الحادث الاخير S في المحاولة السابقة .
أي يجب ان نختار ٤ من (S) من المجموع الكلي ٦ وهذه ممكن عملها بـ :
طريقة متنافية $\binom{6}{4} = 15$

لذا فان y يمثل اصابة الهدف على ان تكون الاصابة الخامسة في المحاولة السابعة.

$$P(y=7) = \binom{6}{4}(\cdot 6)^5 (\cdot 4)^2 = 0\cdot 1866$$

تعريف : (٩ : ١٤)

اذا كررت محاولات مستقلة عدة مرات وكان الناتج بالنجاح باحتمال p، والفشل باحتمال q لذلك فالاحتمال التوزيعي للمتغير العشوائي y (الذي يمثل عدد المحاولات لظهور النجاح k^{th}) قد حدث هو :

$$P(y=k) = \binom{y-1}{k-1} p^k q^{y-k} \qquad y = k, k+1, k+2, \ldots$$

مثال (٢٠) : عند رمي ٣ قطع من النقود ما هو احتمال الحصول على كلهم صوراً أوكلهم كتابة للمرة الثانية في الرمية الخامسة ؟

الحـــل :

$$p = P(3H + 3T) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

$$y = 5, k = 2$$

$$\therefore P(y=k) = \binom{y-1}{k-1} p^k q^{y-k}$$

$$= \binom{4}{1}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3$$

$$= 0{\cdot}1055$$

تمارين الفصل التاسع

(١) ما هو التوزيع الاحتمالي لظهور الصورة عند رمي خمس قطع نقود مرة واحـدة . احسب الوسط الحسابي والتباين لهذا التوزيع .

(٢) اذا علمت بأن نسبة العلب التالفة في مصنع كربلاء لمربى المشمش هو ٩٠٪ . فاذا اخذت عينة مؤلفة من ٦ علب والمطلوب :

(آ) ايجاد احتمال ان تحتوي هذه العينة على ٤ علب تالفة .

(ب) ايجاد احتمال ان تحتوي هذه العينة على الأقل على علبتين تالفتين ؟

(ج) ايجاد الوسط الحسابي والانحراف القياسي للعلب التالفة في هذه العينة .

(٣) في احدى البساتين الكبيرة كانت نسبة اصابة ثمار التفاح هي ٢٠٪ . فاذا اخترت اربع تفاحات عشوائيا فما هو احتمال :

(آ) ان تكون واحدة مصابة فقط .

(ب) ان تكون هناك تفاحة على الأقل مصابة .

(ج) ان تكون هناك على الأكثر ثلاث تفاحات مصابة ؟

(٤) اذا علمت بأن نسبة الطلبة العرب في كلية الزراعة هو ٤٪ وان توزيع الطلبة على الأقسام الداخلية يكون عشوائيا . فاذا كان عدد الطلبة في كل غرفة هـو ٣ . فما هو احتمال ان يكون على الأكثر ٢ منهم من الطلبة العرب ؟

(٥) في عائلة مؤلفة من ٥ أطفال :

(آ) ما هو احتمال ان يكون بينهم ذكر واحد .

(ب) ما هو احتمال ان يكون بينهم على الأكثر ٣ بنات .

(٦) انتخبت اربع بذور من الجيل الناتج من التلقيح التالي :

$$AaBb \times aabb$$

ما هو احتمال ان البذور تنتج :

(آ) اربعة نباتات من النوع AaBb

(ب) على الأكثر ثلاثة منهم من النوع Aabb

(٧) عند رمي ٦ زارات : احسب احتمال ظهور الاعداد الستة .

(٨) من تجارب مندل على البزاليا وجد بأنه عند تهجين نباتات خليطة في زوجين من الجينات (ولنفرض صفتي الطول والبذور المستديرة) . ينتج اربع مجاميع :

طويلة ومستديرة

طويلة ومجعدة

قصيرة ومستديرة

قصيرة ومجعدة

بنسبة ١ : ٣ : ٣ : ٩ على التوالي

فاذا تم اختبار خمسة نباتات عشوائياً فما هو احتمال الحصول على :

نباتين طويلة ومستديرة

نبات واحد طويل ومجعدة

نبات واحد قصير ومستديرة

نبات واحد قصير ومجعدة

(٩) صف من الصفوف يحتوي على ٢٠ طالبا وخمس طالبات فاذا اخذت عينة عشوائية حجمها ٣ ، فما هو احتمال أن تتكون من طالبين وطالبة ؟

(١٠) اذا علمت بأن متوسط عدد الأشخاص المدخنين المصابين في مدينة ما بمرض السرطان هو ٠,٤٪ . فاذا اخذت عينة عشوائية مكونة من ١٠٠ شخص مدخن ما هو احتمال ان تحتوي العينة على صفر ، ١ ، ٢ ، ٣ ، ٤ ، اكثر من اربعة أشخاص مصابين بالسرطان ؟

(١١) اوجد احتمال الحصول على مجموع نقاط ٧ للمرة الثانية في المحاولة الثامنة عند رمي زارين .

الفصل العاشر

التوزيعات الاحتمالية المستمرة او المتصلة

Continuous Probability Distribution

التوزيع الطبيعي

Normal Distribution

(١٠ : ١) مقدمة

يعتبر التوزيع الطبيعي من أهم التوزيعات المستمرة التي تستخدم في جميع مجالات الاحصاء. (ومن التوزيعات المستمرة التي سيأتي ذكرها بعدئذ هي توزيع t وتوزيع F وتوزيع χ^2 .)

(ان أهمية التوزيع الطبيعي ترجع الى أربعة اعتبارات مهمة):

(١) ان كثيراً من المتغيرات تتوزع توزعا طبيعيا فمعظم الصفات البيولوجية والصفات النفسية والاجتماعية وغيرها من الصفات المهمة يكون توزيعها مشابها للتوزيع الطبيعي أو مقاربا له .

(٢) توزيعات المعاينة لمتوسطات العينات تكون مقاربة للتوزيع الطبيعي ويزداد هذا التقارب كلما زاد حجم العينة .

(٣) امكانية تحويل توزيعات كثيرة الى توزيع طبيعي .

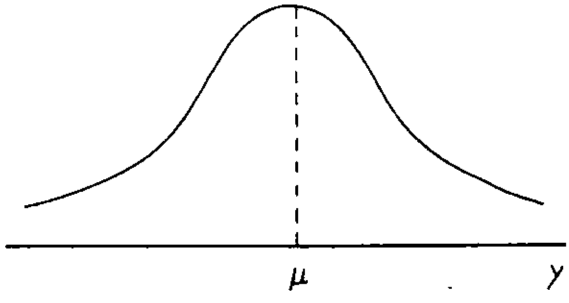
(٤) ان معظم الاختبارات المستخدمة في الاستنتاج الاحصائي مبنية على كون المتغير يتوزع توزعاً طبيعياً .

(وفي أغلب الاحيان تكون هذه النتائج صحيحة أو قريبة من الصحة حتى اذا لم يتوفر شرط التوزيع الطبيعي .)

(٥) تلعب السهولة دورا مهما في اختيار التوزيع الطبيعي .

هذا وفي سنة ١٧٣٣ اشتق De Moivre المعادلة الرياضية للمنحني الطبيعي . واحيانا يسمى

المنحني الطبيعي بمنحني كاوس Gaussian على شرف Gauss (1777-1855) الذي اشتق معادلته عند دراسته الخطأ في القياسات المتكررة . وقد يسمى كذلك(كاوس -لابلاس) ايضا .



شكل (١٠:١) المنحني الطبيعي

(١٠ : ٢) المنحني الطبيعي Normal Curve

تعريف (١٠ : ١)
اذا كان المتغير العشوائي y يتوزع توزعا طبيعيا وله وسط حسابي μ وتباين σ^2 فان معادلة المنحني الطبيعي هي :

$$f(y) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}$$

$$-\infty < y < \infty$$

حيث ان

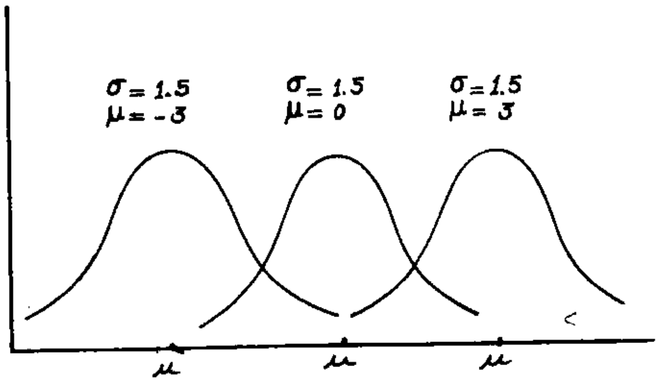
$\pi = 3{\cdot}14159$

$e = 2{\cdot}71828$

هذا ويطلق على $f(y)$ دالة التوزيع الاحتمالي p.df وهي تمثل المحور الصادي وقيم y تمثل المحور السيني . وان مجموع المساحة الكلية الواقعة تحت المنحني يساوي واحد .
إن دالة التوزيع تعتمد على شيئين :

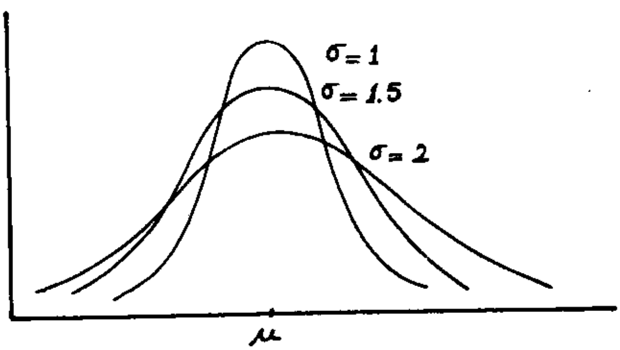
① الوسط الحسابي μ والتباين σ^2 وان قيمة μ , σ^2 يحددان موقع وشكل المنحني الطبيعي . فالشكل التالي يبين ثلاثة توزيعات طبيعية لها نفس الانحراف القياسي (σ) ولكن أوساطها الحسابية مختلفة . (لاحظ شكل (١٠ : ٢))

عمليتي الجمع والطرح لا يؤثران في التباين
بينما عمليتي الضرب تؤثران في التباين



شكل (١٠ : ٢) توزيعات طبيعية اوساطها الحسابية مختلفة بينما انحرافاتها القياسية متساوية

بينما يبين الشكل التالي ثلاثة توزيعات طبيعية لها نفس الوسط الحسابي ولكن انحرافاتها القياسية مختلفة (شكل ١٠ : ٣)



شكل (١٠ : ٣) توزيعات طبيعية لها نفس الوسط الحسابي ولكن انحرافاتها القياسية مختلفة

قاعدة الجمع $y = x + 3$

$\bar{y} = \bar{x} + 3$

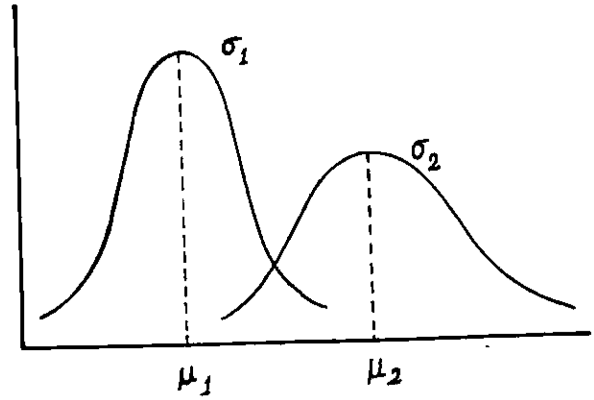
قاعدة الضرب $y = 2x + 3$

$\bar{y} = 2\bar{x} + 3$

حيث يرمز μ = متوسط X

σ^2 = التباين X

والشكل التالي يبين توزيعان طبيعيان لهما وسطان حسابية وانحرافان قياسيان مختلفان (لاحظ شكل (١٠ : ٤))



شكل (١٠ : ٤) توزيعان طبيعيان لهما وسطان حسابيان وانحرافان قياسيان مختلفان

فمما سبق يمكن تلخيص خواص المنحني الطبيعي كما يلي :

(١) شكل المنحني يكون على هيئة ناقوس Bell

(٢) تتركز المشاهدات حول الوسط الحسابي ويكون المنحني متماثلا حول الوسط الحسابي بحيث يقسمه الى قسمين متساويين ولذلك فان ارتفاع المنحني حول

$y = \mu + 2\sigma$

مثلا يكون بالضبط مساويا لأرتفاع المنحني حول

$y = \mu - 2\sigma$

وكنتيجة لهذا التماثل فان الوسط الحسابي والوسيط والمنوال لهم نفس القيمة .

(٣) ان طرفي المنحني يتناقصا بالارتفاع كلما ابتعدنا عن الوسط الحسابي ولكنهما لا يلتقيان بالمحور السيني ابدا وعمليا فأن المساحة الموجودة بعد $\mu \pm 3\sigma$ ليس لها أهمية . وبمعنى آخر فان المساحة المهمة هي المحصورة بين $\mu - 3\sigma , \mu + 3\sigma$ حيث أن : (شكل ١٠ : ٥)

(أ) المساحة بين $\mu - \sigma$ و $\mu + \sigma$ هي ٦٨,٢٧ ./. من مجموع المساحة أو بعبارة أخرى ان ٦٨,٢٧ ./. من المشاهدات تقع بين $\mu - \sigma$ و $\mu + \sigma$ أي :

$$P(\mu - \sigma < y < \mu + \sigma) = 68{\cdot}27 \ \%$$

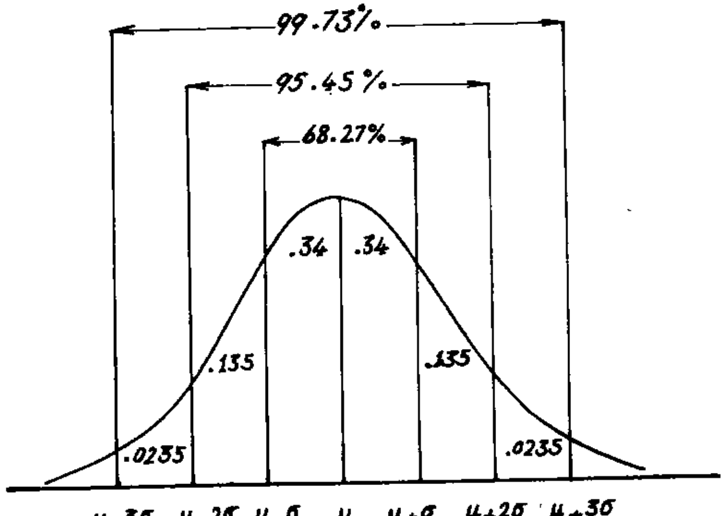
(ب) والمساحة بين $\mu - 2\sigma$ و $\mu + 2\sigma$ هي ٩٥,٤٥ ./. من مجموع المساحة أي ٩٥,٤٥ ./. من

المشاهدات تقع بين $\mu - 2\sigma$ و $\mu + 2\sigma$ أي :

$$P(\mu - 2\sigma < y < \mu + 2\sigma) = 95.45\,\%$$

(جـ) والمساحة بين $\mu - 3\sigma$ و $\mu + 3\sigma$ هي ٩٩,٧٣٪ من مجموع المساحة أي أن ٩٩,٧٣٪ من المشاهدات تقع بين $\mu - 3\sigma$ و $\mu + 3\sigma$ أي :

$$P(\mu - 3\sigma < y < \mu + 3\sigma) = 99.73\,\%$$



شكل (١٠ : ٥) المساحات تحت المنحني الطبيعي

(٤) ان مجموع المساحة الكلية تحت المنحني الطبيعي = ١ ، ان المتغير الذي يتوزع توزعا طبيعيا نرمز له بـ $y \sim N(\mu, \sigma^2)$ أي أن y يتبع التوزيع الطبيعي بوسط حسابي قدره μ وتباين قدره σ^2 وبما أن f(y) هي p.d.f .

لذا فإن

$$\int_{-\infty}^{\infty} f(y)\, dy = 1$$

(١٠ : ٣) المساحة تحت المنحني الطبيعي

Areas Under the Normal Curve

ان الاحتمال في التوزيعات المستمرة تمثل بالمساحات . فمثلا لايجاد احتمال :

$$P(y_1 < y < y_2)$$

فاننا نحسب المساحة المحصورة بين y_2 , y_1 فقيمة هذه المساحة هي درجة الاحتمال . ومن المعروف انه يمكن حساب هذه المساحة باستخدام التكامل

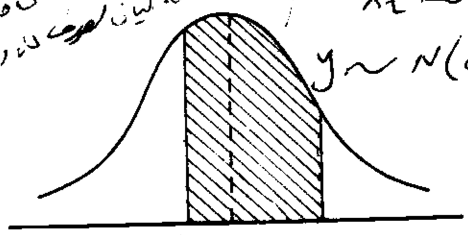
$y = a_1 x_1 + a_2 x_2$

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$y \sim N(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2)$

$y \sim N(\sum_{i=1}^{n} a_i \mu_i,\ \sum_{i=1}^{n} a_i^2 \sigma_i^2)$



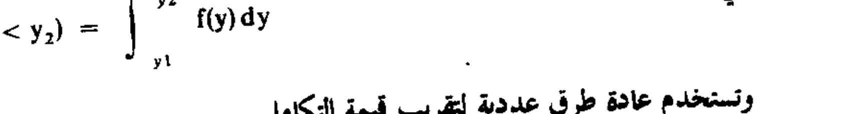
شكل (١٠ : ٦) احتمال ان قيمة y تقع بين القيمتين y_1 ، y_2

أي

$$P(y_1 < y < y_2) = \int_{y1}^{y2} f(y)\, dy$$

وتستخدم عادة طرق عددية لتقريب قيمة التكامل

وكما ذكرنا سابقا بأن المنحني يعتمد على معرفة μ و σ^2 كذلك فان المساحة تحت هذا المنحني المحصورة بين حدين هي ايضا تعتمد على كل من μ و σ^2 وهذا واضح في الشكل التالي :



شكل (١٠ : ٧) احتمال ان قيمة y تقع بين القيمتين y_1 ، y_2 لمنحنيين طبيعيين مختلفين

فالمساحة المضللة هي تمثل احتمال $P(y_1 < y < y_2)$ للمنحنيين اللذين لهما وسطان حسابيان وانحرافان قياسيان مختلفان . وطبيعي فان المساحتين المضللتين تختلفان عن بعضهما ولذلك فان احتمال كل منهما سيكون مختلفا أيضا . هذا ومن الصعب وغير العملي وضع جداول للمنحنيات الطبيعية لكل قيمة من μ و σ . وحتى نتحاشى استعمال التكامل فقد تم وضع جدول واحد محسوب للمساحات المختلفة

$Z = \frac{y_i - \mu}{\sigma}$ μ = المتوسط الحسابي للمتغير σ = = = = التباين

لتوزيع طبيعي ذي متوسط حسابي يساوي صفرا وتباين يساوي ١ ، ويطلق على هذا التوزيع بالتوزيع الطبيعي القياسي Standard Normal Distribution

تعريف : (١٠ : ٢)

التوزيع الطبيعي القياسي هو توزيع طبيعي له متوسط حسابي = صفرا وانحراف قياسي = ١ ويرمز للمتغير العشوائي فيه بالرمز Z أي

$$Z \sim N(0, 1)$$

$$\boxed{f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}$$

والآن أصبح من السهل تحويل المتغيرات العشوائية y التي تتوزع توزعا طبيعيا الى متغيرات عشوائية Z والتي تتوزع توزعا طبيعيا قياسيا وذلك بالطريقة التالية :

$$\boxed{Z = \frac{y - \mu}{\sigma}}$$

فمثلا اذا كانت y بين الحدين y_1 ، y_2 فان Z لها ستكون بين Z_1 ، Z_2

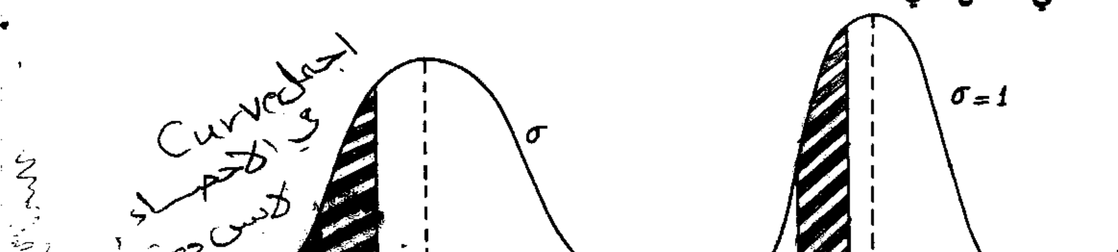
حيث ان :

$$Z_1 = \frac{y_1 - \mu}{\sigma}$$

وان :

$$Z_2 = \frac{y_2 - \mu}{\sigma}$$

كما في الشكل التالي :



شكل (١٠ : ٨) المنحني الطبيعي الاصلي والمنحني القياسي لـــه

ولذلك فالمساحة الموجودة بين الحدين y_1, y_2 تساوي المساحة الموجودة بين Z_1 و Z_2

أي أن : $P(y_1 < y < y_2) = P(Z_1 < Z < Z_2)$

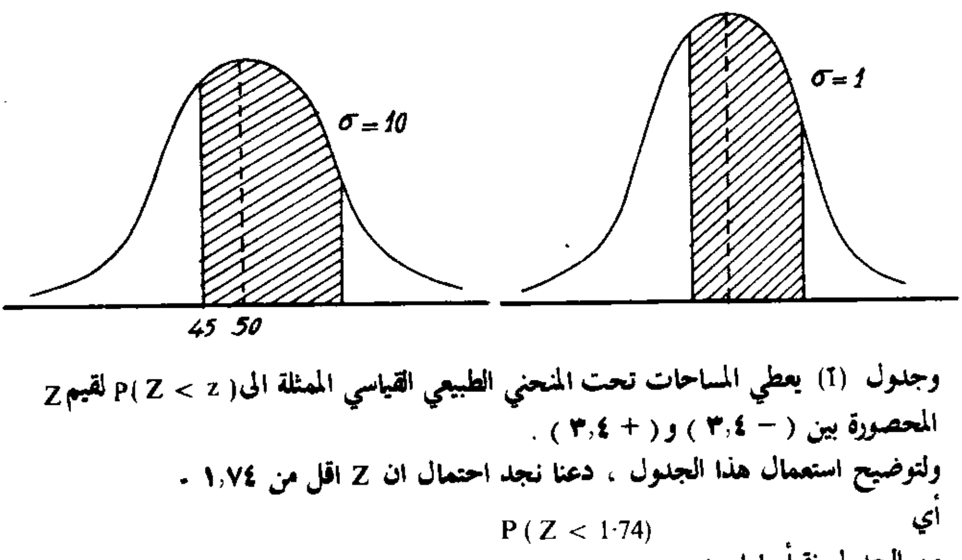
مثـال(١) : اذا كان الوسط الحسابي لتوزيع طبيعي هو$\mu = 50$والانحراف القياسي له هو $\sigma = 10$فاوجد قيمة Z_2, Z_1 بحيث أن : $P(45 < y < 62) = P(Z_1 < Z < Z_2)$

الحـل :

$$Z_1 = \frac{y_1 - \mu}{\sigma} = \frac{45 - 50}{10} = -0.5$$

$$Z_2 = \frac{y_2 - \mu}{\sigma} = \frac{62 - 50}{10} = 1.2$$

$$\therefore P(45 < y < 62) = P(-0.5 < Z < 1.2)$$



وجدول (آ) يعطي المساحات تحت المنحني الطبيعي القياسي الممثلة الى $P(Z < z)$ لقيم Z المحصورة بين (− ٣,٤) و (+ ٣,٤) .

ولتوضيح استعمال هذا الجدول ، دعنا نجد احتمال ان Z اقل من ١,٧٤ .

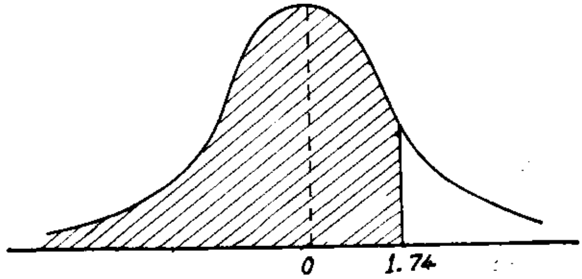
أي $P(Z < 1.74)$

من الجدول نقرأ ما يلي :

من العمود الأيسر Z نعين الرقم ١,٧ ومن العمود الذي يحوي ٠,٠٤ نتنزل عموداً الى الخط الافقي الممتد من الرقم ١,٧ فيلتقيا في الرقم ٠,٩٥٩١ .

تصاميم اكيد

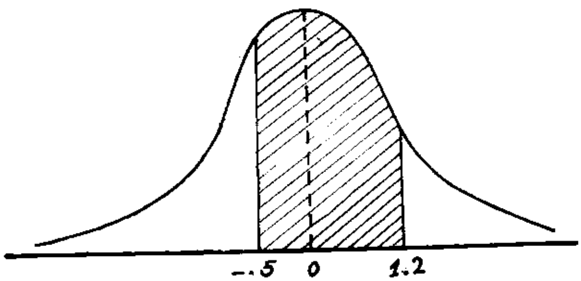
$$\therefore P(Z < 1.74) = 0.9591$$



$P(Z < 1.74) = 0.9591$

ولايجاد قيمة :

$P(-\cdot 5 < Z < 1\cdot 2)$



فهي عبارة عن المساحة المظللة في الشكل اعلاه وهي عبارة عن المساحة الكلية التي عن يسار $Z = 1\cdot 2$ مطروحا منها المساحة التي عن يسار $Z = -\cdot 5$ أي أن :

$$P(-\cdot 5 < Z < 1\cdot 2) = P(Z < 1\cdot 2) - P(Z < -\cdot 5)$$

$$= 0\cdot 8849 - 0\cdot 3085$$

$$= 0\cdot 5764$$

مثـال (٢) : اوجد الاحتمال التالي :

$$P(\mu - \sigma < y < \mu + \sigma)$$

الحـل :

$$Z_1 = \frac{(\mu - \sigma) - \mu}{\sigma} = -1$$

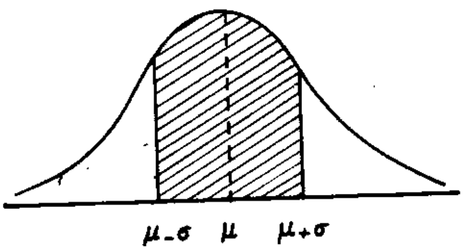
$$Z_2 = \frac{(\mu + \sigma) - \mu}{\sigma} = +1$$

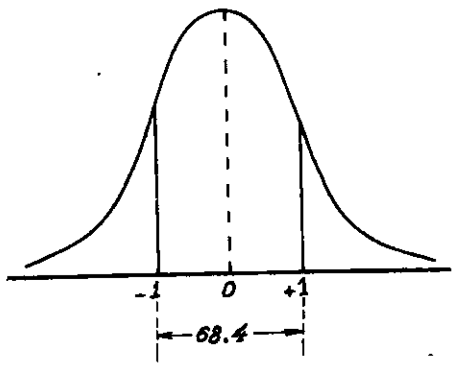
$$\therefore P(\mu - \sigma < y \leq \mu + \sigma) = P(-1 < Z < +1)$$

$$= P(Z < 1) - P(Z < -1)$$

$$= 0.8413 - 0.1587$$

$$= 0.6826$$

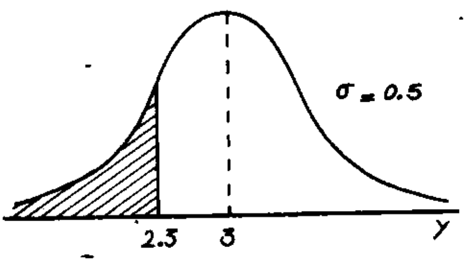




(١٠ : ٤) أمثلة متنوعة

مثـال(٣) نوع معين من بطارية سيارات متوسط مدة استهلاكه = ٣ سنوات وانحرافه القياسي = ٠,٥ سنة . فاذا كانت مدة استهلاكه تتبع التوزيع الطبيعي . ما هو احتمال ان بطارية معينة ستستهلك بأقل من ٢,٣ سنة .

الحـل :



$$Z = \frac{y - \mu}{\sigma}$$

$$= \frac{2.3 - 3}{0.5} = -1.4$$

$$\therefore P(y < 2.3) = P(Z < -1.4)$$

$$= 0.0808$$

وهذا احتمال ان بطارية معينة ستستهلك بأقل من ٢,٣ سنة أو بعبارة أخرى فان نسبة البطاريات التي ستستهلك بأقل من ٢,٣ سنة تعادل ٨./. تقريبا .

مثـال(٤) : اذا كان المتغير y يتبع التوزيع الطبيعي بوسط حسابي قدره $\mu = 20$ وانحراف قياسي قدره $\sigma = 5$.

والمطلوب :

(أ) ايجاد قيمة y_1 بحيث ان :

$$P(y < y_1) = 0{\cdot}2514$$

(ب) ايجاد قيمة y_2 بحيث أن :

$$P(y > y_2) = 0{\cdot}0655$$

الحـل :

(a) $P(y < y_1) = 0{\cdot}2514 = P(Z < z)$

ومن جدول (Z)، نجد ان قيمة Z التي لهذه المساحة هي :

$$\therefore Z = -0{\cdot}67$$

$$\therefore Z = \frac{y - \mu}{\sigma}$$

$$\therefore -0{\cdot}67 = \frac{y_1 - 20}{5}$$

$$\therefore y_1 = 16{\cdot}65$$

(b) $p(y > y_2) = 0{\cdot}0655$

$$\therefore P(y < y_2) = 1 - 0{\cdot}0655 = 0{\cdot}9345$$

وهذه المساحة يقابلها قيمة لـ Z تعادل : $1{\cdot}51$

$$\therefore Z = \frac{y - \mu}{\sigma}$$

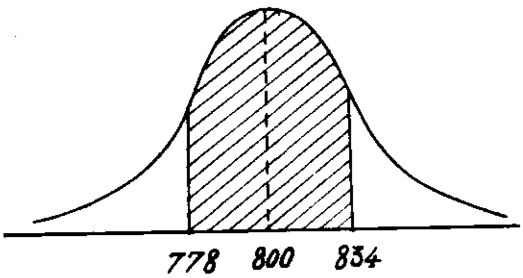
$$\therefore 1{\cdot}51 = \frac{y_2 - 20}{5}$$

$$\therefore y_2 = 27{\cdot}55$$

مثـال(٥) : اذا كان متوسط انتاج الدونم من الذرة الصفراء هو ٨٠٠ كغم وبانحراف قياسي قدره ٤٠ كغم . وعلى فرض ان كمية المحصول تتبع التوزيع الطبيعي . ما هو احتمال ان

نباتا يعطي محصولا بين ٧٧٨ و ٨٣٤ كغم (أو بعبارة أخرى ما هي نسبة النباتات التي تعطي كمية محصول بين ٧٧٨ و ٨٣٤ كغم) ؟

الحـل :



$$\therefore Z = \frac{y-\mu}{\sigma}$$

$$Z_1 = \frac{778 - 800}{40} = -0.55$$

$$Z_2 = \frac{834 - 800}{40} = 0.85$$

$$\therefore P(778 < y < 834) = P(-0.55 < Z < 0.85)$$

$$= P(Z < 0.85) - P(Z < -0.55)$$

$$= 0.8023 - 0.2912$$

$$= 0.5111$$

أي أن ٥١ ./. من النباتات تعطي محصولا بين ٧٧٨ و ٨٣٤ كغم / دونم .

مثـال(٦) : اذا كان متوسط طول ٥٠٠ طالب في احدى المدارس الثانوية هو ١٥١ سم بانحراف قياسي قدره ١٥ سم . افرض بأن الاطوال تتوزع توزيعاً طبيعياً . أوجد القيمة المتوقعة للطلبة الذين :

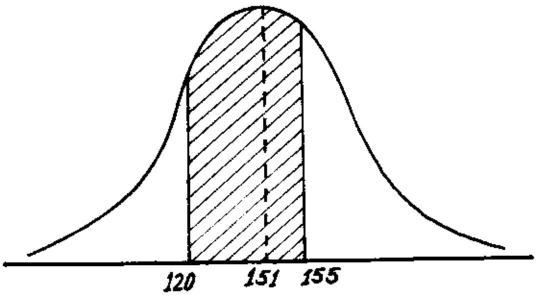
(١) أطوالهم بين ١٢٠ و ١٥٥ سم

(٢) أطوالهم اكثر من ١٨١ سم

(٣) أطوالهم أقل من ١٢٨ سم

الحـل :

(a) $P(120 < y < 155)$



$$Z = \frac{y - \mu}{\sigma}$$

$$Z_1 = \frac{120 - 151}{15} = \frac{-31}{15} = -2{\cdot}07$$

$$Z_2 = \frac{155 - 151}{15} = \frac{4}{15} = 0{\cdot}27$$

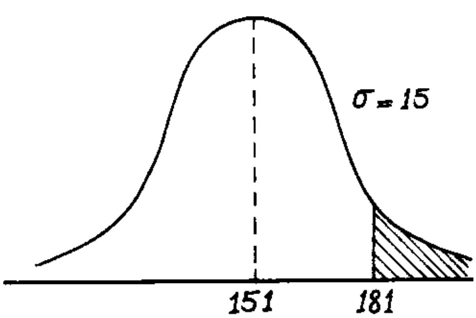
$$\therefore P(120 < y < 155) = P(-2{\cdot}07 < Z < 0{\cdot}27)$$

$$= P(Z < 0{\cdot}27) - P(Z < -2{\cdot}07)$$
$$= 0{\cdot}6064 - 0{\cdot}0192$$
$$= 0{\cdot}5872$$

أي حوالي ٦٠ ./· من الطلبة أطوالهم تقع بين ١٢٠ و ١٥٥ سم
اذن عدد الطلبة = ٠,٦٠ × ٥٠٠ = ٣٠٠ طالب

(b) $P(y > 181)$

$$Z = \frac{181 - 151}{15} = \frac{30}{15} = 2$$



$$\therefore P(y > 181) = P(Z > 2)$$

$$= 1 - P(Z < 2)$$

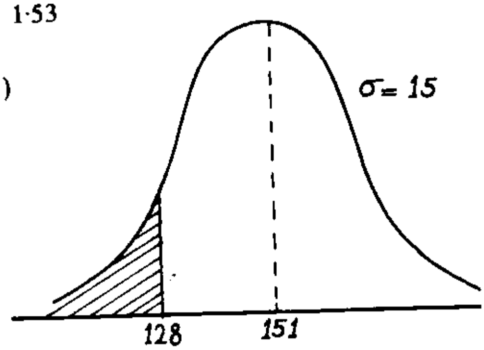
$$= 1 - 0{\cdot}9772$$

$$= 0{\cdot}0228$$

أي أن حوالي ٢ ./. من الطلبة لهم اطوال اكثر من ١٨١ سم
اذن عدد الطلبة = ٠,٠٢ × ٥٠٠ = ١٠ طلاب

(c) $P(y < 128)$

$$Z = \frac{128 - 151}{15} = \frac{-23}{15} = -1{\cdot}53$$



$$\therefore P(y < 128) = P(Z < -1{\cdot}53)$$

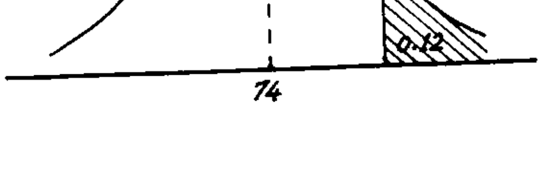
$$= 0{\cdot}0630$$

اذن ٦,٣ ./. من الطلبة أطوالهم أقل من ١٢٨
اذن عدد الطلبة = ٠,٠٦٣ × ٥٠٠ = ٣٢ طالبا

مثـال(٧) : في احدى امتحانات الاحصاء كان معدل الدرجات = ٧٤ بانحراف قياسي قدره ٧ .

فاذا كان ١٢ ./. من الطلبة قد حصلوا على امتياز وكانت الدرجات تتوزع توزيعا طبيعيا فما هي أقل درجة للامتياز وأعلى درجة لجيد جدا ؟

الحـل :



$$P(Z > z) = 0{\cdot}12$$

$$\therefore P(Z < z) = 1 - P(Z > z)$$

$$\therefore P(Z < z) = 0{\cdot}88$$

ومن جدول Z نجد أن :

$$P(Z < 1{\cdot}175) = 0{\cdot}88$$

$$\therefore Z = \frac{y-\mu}{\sigma}$$

or

$$y = \mu + Z\sigma$$
$$= 74 + (1{\cdot}175)(7)$$

$$= 82{\cdot}225$$

لذا فان درجة امتياز هي ٨٣ واعلى درجة جيدجدا هي ٨٢ .

مثـال(٨) : اذا كان معدل طول (١٠,٠٠٠) نبات من نباتات القطن هو ٨٠ سم . فاذا وجد بان ١٥٨٧ نباتا يقل طولهم عن ٦٠ سم . فما هي عدد النباتات التي يزيد طولها عن ٩٠ سم ؟

الحـل :

ان نسبة النباتات التي يقل اطوالها عن ٦٠ سم هي $\frac{١٥٨٧}{١٠٠٠٠}$ = ٠,١٥٨٧ وهذه النسبة تعادل المساحة تحت المنحني الطبيعي ومنها نجد أن قيمة Z التي تقابل هذه المساحة هي $Z = -1$

أي أن :

$$P(y < 60) = P(Z < -1)$$

وبتطبيق القانون التالي نجد قيمة σ :

$$Z = \frac{y-\mu}{\sigma}$$

$$-1 = \frac{60-80}{\sigma}$$

$$\therefore \sigma = 20$$

$$\therefore P(y \geq 90) = 1 - P(y \leq 90)$$
$$= 1 - P\left(Z < \frac{90-80}{20}\right)$$
$$= 1 - P(Z < 0{\cdot}5)$$
$$= 1 - 0{\cdot}6915$$
$$= 0{\cdot}3085$$

لذا فعدد النباتات التي يزيد طولها عن ٩٠ سم هو ٠,٣٠٨٥ × ١٠٠٠٠
أي ٣٠٨٥ نباتا .

(١٠ : ٥) بعض الخواص الاخرى للتوزيع الطبيعي

(١) اذا كان x و y متغيرين عشوائيين وكان :

$$y = ax + b$$

بحيث ان : $a \neq 0$

وان : $X \sim N(\mu, \sigma^2)$

فان : $y \sim N(a\mu + b, a^2\sigma^2)$

أي أن y تتوزع توزيعا طبيعيا بمتوسط قدره $(a\mu + b)$ وتباين يساوي $a^2\sigma^2$

مثـال(٩) : افرض بأن الوسط الحسابي للمتغير العشوائي x هو ١٠ وتباينه = ٢

فاذا علمت بأن x يتوزع توزيعا طبيعيا اي أن : $X \sim N(10, 2)$

وكان $y = 3x + 5$

ففي هذه الحالة فأن y ايضا تتوزع طبيعيا بوسط حسابي قدره :

$$\begin{aligned} \mu_y &= 3\mu_x + 5 \\ &= 3(10) + 5 \\ &= 35 \end{aligned}$$

وتباين قدره :

$$\begin{aligned} \sigma_y^2 &= (3)^2\sigma_x^2 \\ &= 9(2) \\ &= 18 \end{aligned}$$

$$\therefore y \sim N(35, 18)$$

(٢) اذا كان كل من x_1 و x_2 يتوزعان توزيعا طبيعيا كالآتي :

$$X_1 \sim N(\mu_1, \sigma^2{}_1)$$

$$X_2 \sim N(\mu_2, \sigma^2{}_2)$$

وكان x_1 و x_2 مستقلين ، فاذا كان المتغير العشوائي y هو

$$y = a_1x_1 + a_2x_2$$

فان y يتوزع توزيعا طبيعيا أيضا كالآتي :

$$\mu_y = a_1\mu_1 + a_2\mu_2$$

$$\sigma^2{}_y = a^2{}_1\sigma^2{}_1 + a^2{}_2\sigma^2{}_2$$

$$\therefore y \sim N(a_1\mu_1 + a_2\mu_2, a^2{}_1\sigma^2{}_1 + a^2{}_2\sigma^2{}_2)$$

مثـال(١٠) : اذا علمت بأن :

$$X_1 \sim N(10, 3)$$
$$X_2 \sim N(12, 4)$$

فاذا كان $y = 2x_1 + 3x_2$ وكان كل من x_1 و x_2 مستقلين
فان y يتوزع طبيعيا أيضا بوسط حسابي قدره :

$$\mu_y = 2\mu_1 + 3\mu_2 = 56$$

وتباين قدره :

$$\sigma^2_y = 4\sigma^2_1 + 9\sigma^2_2 = 48$$

$$\therefore y \sim N(56, 48)$$

وبصورة عامة :

اذا كان $y = \sum_{i=1}^{n} a_i x_i$ ، وان $X_i \sim N(\mu_i, \sigma^2_i)$
وان $x_1, x_2, \ldots\ldots\ldots, x_n$ متغيرات مستقلة
فان :

$$y \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \ \sum_{i=1}^{n} a^2_i \sigma^2_i \right)$$

(١٠ : ٦) العلاقة بين التوزيع الطبيعي وتوزيع ذي الحدين

Relation Between Normal and Binomial Dist.

ان الاحتمالات المتعلقة بتجارب توزيع ذي الحدين يمكن حسابها من قانونه أو من جدول خاص عندما تكون n = عدداً قليلا . ولكن اذا كانت n كبيرة وقيمة p أو q قريبتين من النصف أو بعيدتين نوعاً ما من الصفر ، فان توزيع ذا الحدين يقترب من التوزيع الطبيعي . ويكون التوزيع الطبيعي القياسي هو :

$$\boxed{Z = \frac{y - np}{\sqrt{npq}}}$$

وبذلك يمكن استخدام جدول التوزيع الطبيعي القياسي (جدول Z) لايجاد الاحتمالات لتوزيع ذي الحدين .

مثـال(١١) : لاعب كرة سلة نسبة اصابته للهدف هي ٦٠٪ ، ما هو احتمال تهديفه اقل من ٥٠ هدفا اذا رمى ١٠٠ رمية ؟

$n = 100$
$y = 50$
$p = 0.6$
$q = 0.4$

الحـل :

$$\mu = np = (100)(\cdot 6) = 60$$

$$\sigma = \sqrt{npq} = \sqrt{(100)(\cdot 6)(\cdot 4)} = 4{\cdot}9$$

$$\therefore \; P(y < 50) = P(Z < \frac{50-60}{4{\cdot}9})$$
$$\simeq P(Z < -2{\cdot}04)$$
$$= 0{\cdot}0207$$

خاص
مهم

مثـال(١٢) : في السنين السابقة تبين بأن نسبة النجاح في درس الانكليزي في امتحان البكلوريا لاحدى المدارس الثانوية هو ٣٦٪ . فاذا شارك ١٠٠ طالب لهذه السنة فما هو احتمال ان ينجح ما بين ٢٤ الى ٤٢ طالبا في درس الانكليزي ؟

$p = 0.36$ $\quad n = 100$ $\quad y_2 = 42$
$q = 0.64$ $\quad y_1 = 24$

الحـل :

$P(24 \leqslant y \leqslant 42)$

$$\mu = np = (100)(\cdot 36) = 36$$

$$\sigma = \sqrt{npq} = \sqrt{(100)(\cdot 36)(\cdot 64)} = 4{\cdot}8$$

$$Z_1 = \frac{24 - 36}{4{\cdot}8} = -2{\cdot}5$$

$P(y_1 \leqslant y \leqslant y_2) = P(Z_1 \leqslant Z \leqslant Z_2)$
$P(Z_1 \leqslant Z \leqslant Z_2) = P(Z_2) - P(Z_1)$

$$Z_2 = \frac{42 - 36}{4{\cdot}8} = 1{\cdot}25$$

$$\therefore \; P(24 \leq y \leq 42) \simeq P(-2{\cdot}5 < Z < 1{\cdot}25)$$
$$\simeq P(Z < 1{\cdot}25) - P(Z < -2{\cdot}5)$$
$$= 0{\cdot}8882$$

لاحظ بأنه باستخدام قانون توزيع ذي الحدين فان :

$$P(24 < y < 42) = \sum_{24}^{42} \binom{100}{y} (\cdot 36)^y \, (\cdot 64)^{100-y}$$
$$= 0{\cdot}90739$$

بينما باستخدام التقريب الى التوزيع الطبيعي كان $P(24 < y < 42) = 0{\cdot}8882$ وبالامكان عمل تحسين الى هذا التقريب وذلك بطرح نصف من القيمة الاقل وزيادة نصف الى القيمة الأعلى .

ملاحظة مهمة : $P(Z>) = 1 - P(Z< P(y>50) =$

[illegible]

أي :

$$Z_1 = \frac{23.5 - 36}{4.8} = -2.60$$

$$Z_2 = \frac{42.5 - 36}{4.8} = 1.35$$

$$\begin{aligned} \therefore \quad P(24 < y < 42) &\simeq P(-2.6 < Z < 1.35) \\ &= P(Z < 1.35) - P(Z < -2.6) \\ &\simeq 0.90683 \end{aligned}$$

مثال (١٣) : افرض ان في توزيع ذي الحدين :

$$n = 80$$
$$p = 0.16$$

ماهو احتمال ان :

$$\underline{y = 20}$$

الحــل :

$$\mu = np = (80)(.16) = 12.8$$

$$\sigma = \sqrt{npq} = \sqrt{80(.16)(.84)} = 3.279$$

$$\begin{aligned} \therefore \quad P(y = 20) &\simeq P\left(\frac{19.5 - 12.8}{3.279} < Z < \frac{20.5 - 12.8}{3.279}\right) \\ &\simeq P(2.04 < Z < 2.35) \\ &= 0.01129 \end{aligned}$$

بينما باستخدام قانون ذي الحدين فأن :

$$P(y = 20) = \binom{80}{20}(.16)^{20}(.84)^{60} = 0.012234$$

من هذه الأمثلة يتضح بان استخدام جدول Z لايجاد احتمال توزيعات ذي الحدين هو تقريب ممتاز .

ملاحظة :

في التوزيع الطبيعي فان قيم y تتراوح من ∞ – الى ∞ ولكن في التوزيع ذي الحدين فهناك حد اقل وحد اعلى ولذلك يجب الانتباه الى هذه النقطة لحساب احتمال y اكبر من قيمة معينة . كالمثال التالي :

مثال (١٤) : افرض بأن توزيعا ذا حدين فيه :

$n = 95$
$p = 0.91$

والمطلوب ايجاد احتمال بأن y أكبر او تساوي ٨٠ أي : $P(y \geq 80)$

الحـــل :

بما ان y لا يمكن ان تكون أكبر من 95 (لأن n = 95) لذا فان :

$$P(y \geq 80) = P(80 \leq y \leq 95)$$

$$\mu = np = (95)(.91) = 86.45$$

$$\sigma = \sqrt{npq} = \sqrt{(95)(.91)(.09)} = 2.789$$

$$\therefore \; P(80 \leq y \leq 95) = P\left(\frac{79.5 - 86.45}{2.789} < Z < \frac{95.5 - 86.45}{2.789} \right)$$

$$= P(-2.49 < Z < 3.24)$$

$$= P(Z < 3.24) - P(Z > -2.49)$$

$$= 0.99301$$

بينما الاحتمال الحقيقي باستخدام قانون توزيع ذي الحدين هي :

$$P(y \geq 80) = \sum_{80}^{95} \binom{95}{y} (.91)^y (.09)^{95-y} = 0.989417$$

تمارين الفصل العاشر

(١) في احدى الامتحانات كان الوسط الحسابي لدرجات الطلبة هو (٦٢) والانحراف القياسي كان (١٥) . فاذا علمت بأن الدرجات توزعت توزيعاً طبيعياً ، فالمطلوب :

أ- تحويل الدرجات التالية الى وحدات طبيعية قياسية :

(a) 50 (b) 83 (c) 72 (d) 62

ب- حول الوحدات الطبيعية القياسية التالية الى درجات طبيعية :

(a) -1 (b) $1{\cdot}6$ (c) 2 (d) 1

(٢) استخدم جدول التوزيع الطبيعي القياسي Z لايجاد :

أ- $P(0 \le Z \le 1)$ ب- $P(Z \ge 1)$ جـ- $P(-1 \le Z \le 1)$

د- $P(Z \ge -1)$ هـ- $P(Z \le {\cdot}75)$ و- $P({\cdot}25 \le Z \le {\cdot}75)$

(٣) أستخدم جدول التوزيع الطبيعي القياسي Z لايجاد :

أ- $P(-1{\cdot}96 \le Z)$ ب- $P({\cdot}49 \le Z \le 1{\cdot}05$ جـ- $P(Z \ge -{\cdot}62)$

د- $P(-{\cdot}72 \le Z \le 1{\cdot}89)$ هـ- $P(Z \ge 1{\cdot}17)$ و- $P(Z \le -2{\cdot}12)$

(٤) باستخدام جدول التوزيع الطبيعي القياسي Z اوجد قيمة K اذا علمت :

أ- $P(Z \le K) = {\cdot}025$ ب- $P(Z \ge K)\,{\cdot}03216$ جـ- $P(Z \le K) = {\cdot}95543$

د- $P(-K \le Z \le K) = {\cdot}95$ هـ- $P(Z \le K) = {\cdot}30854$

و- $P(1 \le Z \le K) = {\cdot}12193$

(٥) اذا علمت بأن y يتوزع توزيعاً طبيعياً بحيث كانت $\mu = 50$ $\sigma = 10$ فاوجد :

أ- $P(y \le 65)$ ب- $P(y \ge 20)$ جـ- $P(y \le 13)$

د- $P(19 \le y \le 40)$

(٦) اذا علمت بأن y تتوزع طبيعيا وكانت $\mu = {\cdot}130$ $\sigma^2 = {\cdot}000625$ فاوجد

أ- $P(y \geq \cdot 126)$ ب- $P(y \leq \cdot 152)$ جـ- $P(\cdot 110 \leq y \leq \cdot 165)$

(٧) اذا علمت بأن y تتوزع طبيعياً وكانت $\mu = 50$ و $\sigma^2 = 100$

أ- اوجد قيمة c اذا علمت بأن : $P(y \leq c) = 0.8406$

ب- اوجد قيمة العددين A و B اذا علمت بأنهما متساويان في بعدهما عن μ وبأن :

$P(A \leq y \leq B) = \cdot 966$

(٨) اذا علمت بأن عمر تشغيل نوع معين من مصابيح الكهرباء يتوزع توزيعاً طبيعياً . فاذا كان ٩٢.٥ ./. منها لها عمر تشغيل اطول من ٢١٦٠ ساعة بينما ٣.٩٢ ./. لها عمر تشغيل أطول من ١٧٠٤٠ ساعة . فما هو متوسط عمر التشغيل والانحراف القياسي ؟

(٩) في أحد البساتين الكبيرة كانت نسبة اصابة ثمار التفاح هي ١٠ ./. . فاذا أخترت أربع تفاحات عشوائياً فما هي احتمالات :

أ- أن تكون واحدة فقط مصابة ؟

ب- أن تكون جميع الثمار سليمة ؟

جـ- أن تكون هناك ثمرة على الأقل مصابة ؟

(١٠) اذا قمت برمي زار طاولة متزن مائتي مرة فما هو احتمال :

أ- ان تحصل على الوجه واحد ما بين ٣٠ و ٥٠ مرة

ب- ان تحصل على الوجه(واحد)أقل من ٢٠ مرة .

(١١) اذا كانت نسبة التالف في أنتاج أحد المصانع تبلغ ٥ ./. وأخذت عينة من ١٠٠ فما هو الاحتمال بأن يكون التالف بها أقل من ٥ ؟

الفصل الحادي عشر
نظرية المعاينة
Sampling Theory

(١١ : ١) مقدمــة :

ذكرنا سابقاً بأن علم الاحصاء يمكن تقسيمه الى قسمين اساسيين هما الاحصاء الوصفي -- Descriptive Statistics والأحصاء الاستنتاجي Statistical Inference. وقلنا ايضاً بأن المجتمع عبارة عن جميع المفردات التي يمكن أن يأخذها المتغير بينما العينة هي جزء من المجتمع . والاحصاء الاستنتاجي يعتمد اعتماداً كلياً على دراسة العينات وصفاتها ومنها يستنتج أو يستدل على خواص المجتمعات التي أخذت منها هذه العينات لأن دراسة المجتمع ككل قد يكون مستحيلاً أو صعباً اضافةً الى أنه يحتاج الى وقت وجهدومال

تعريف (١١ : ١)

المعلمة : Parameter هي القيمة التي تحسب من المجتمع وتكون هذه القيمة ثابتة للمجتمع الواحد .

فإذا كان حجم المجتمع هو N من المفردات فإن الوسط الحسابي للمجتمع ويرمز له بـ μ هو :

$$\mu = \frac{\sum y_i}{N}$$

وهي قيمة ثابتة لهذا المجتمع .

أما التباين لهذا المجتمع فهو متوسط مجموع مربعات الانحرافات . أي :

$$\sigma^2 = \frac{\sum (y_i - \mu)^2}{N}$$

محمد نادر محمد محمد محمد محمد قادر علي

تعريف (١١ : ٢)

الاحصائية statistic : عبارة عن كل قيمة تحسب من العينة أو بعبارة أخرى هي عبارة عن متغير عشوائي قيمته تعتمد على العينة .

والاحصائية قيمة متغيرة لأنها تختلف من عينة الى أخرى داخل المجتمع الواحد .

فمثلاً في عينة حجمها n فإن الوسط الحسابي لها هو :

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

والتباين هو :

$$S^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

وطبيعي فإن التغير في قيمة الاحصائية يعتمد على :

(١) حجم المجتمع
(٢) حجم العينات
(٣) طريقة اختيار العينات

(١١ : ٢) تصاميم العينات Sample Designs

تعريف (١١ : ٣)

تصميم العينة : هو خطة أو طريقة اختيار العينة من مجتمع معين .

وهناك عدة طرق لاختيار العينة بعضها بسيط والبعض الآخر معقد كما قد تستعمل عدة طرق سوية لاختيار اجزاء مختلفة من العينة من نفس المجتمع .

وفيما يلي مختصر لأهم التصاميم المستخدمة في اختيار العينات :

(١) المعاينة العشوائية Random Sampling

تعريف (١١ : ٤)

المعاينة العشوائية البسيطة Simple Random Sampling هي طريقة اختيار عينة بصورة عشوائية بحيث يكون لجميع وحدات المعاينة Sampling units في العشيرة نفس الفرصة أو الاحتمال في الاختيار .

فاذا كان عدد مفردات المجتمع هو N فان احتمال اختيار أي مفردة منه هو 1/N .
هذا واختيار العينة قد يكون بالارجاع Sampling with replacement أي ارجاع وحدة المعاينة المسحوبة قبل سحب الوحدة التي تليها .
أو اختيار العينة بدون ارجاع Sampling without replacement . أي بدون ارجاع وحدة المعاينة المسحوبة قبل سحب الوحدة التي تليها .

هذا وان أبسط أنواع اختيار عينة عشوائية حجمها n هي بأن تسجل وحدة المعاينة لجميع المجتمع على بطاقات متشابهة تماماً ثم نسحب عدداً من البطاقات (عددها n) وتخلط هذه البطاقات بعد كل سحبة .

واذا كانت وحدات المعاينة للمجتمع كبيرة جداً فيستحسن استعمال جدول الاعداد العشوائية .

(٢) المعاينة المنتظمة Systematic Sampling

تعريف (١١ : ٥) :

المعاينة المنتظمة : Systematic Sampling .

هي طريقة اختيار عينة منتظمة وذلك عن طريق اختيار وحدة المعاينة المرقمة K (والتي تسمى نسبة المعاينة وهي حجم المجتمع الى حجم العينة) ثم اختيار رقم عشوائي بين ١ و K ليكون رقم العينة الاولى ثم اضافة K ومضاعفاتها على رقم العينة الأولى الى ان يكمل حجم العينة .

فمثلاً لوكان المجتمع يتكون من ١٠,٠٠٠ وحدة معاينة وان حجم العينة هو ٥٠٠ وحدة معاينة فأن K تساوي :

$$K = \frac{10000}{500} = 20$$

ثم نختار رقماً عشوائياً بين ١ و ٢٠ وليكن ٨ ، فهذا يكون رقم العينة الأولى ثم نضيف ٢٠ ومضاعفاتها الى رقم العينة الأولى لنحصل على وحدات المعاينة وهي ٨ ، ٢٨ ، ٤٨ ، ٦٨ ، وهكذا الى أن يكمل حجم العينة المؤلفة من ٥٠٠ وحدة معاينة .

٨ - ٢٨ - ٤٨ - ٦٨
٨٨ - ١٠٨ - 128 - 148
168 — 188 — 200 — 28

(٣)المعاينة الطبقية Stratified Sampling

تعريف (١١ : ٦) :

المعاينة الطبقية Stratified Sampling

وهي طريقة اختيار عينة طبقية عن طريق تقسيم المجتمع الى أقسام متجانسة تعرف بالطبقات (Strata) ثم اختيار عينة عشوائية فرعية Sub-Sample بصورة عشوائية من كل طبقة وهذه العينات الفرعية مجتمعة تكون العينة الطبقية.

وعادة حجم العينة الفرعية يكون متناسباً مع حجم الطبقة وهذه الطريقة تسمى طريقة التخصيص النسبي Proportionale وأحياناً أخرى يكون حجم العينة الفرعية متساوياً لجميع طبقات المجتمع .

(٤) المعاينة المتعددة المراحل Multi Stage Sampling

تعريف (١١ : ٧)

المعاينة المتعددة المراحل Multi-Stage Sampling هي طريقة لاختيار عينة متعددة المراحل وذلك عن طريق اجراء الاختيار على مراحل متعددة فاذا كان المجتمع مقسماً الى أقسام فإننا في المرحلة الأولى نختار عشوائيا عينة من هذه الأقسام وفي المرحلة الثانية نختار عينة عشوائية من العينة التي اختيرت في المرحلة الأولى وهكذا حتى نحصل على حجم العينة المقررة .

(١١ : ٣) توزيع المعاينة الحسابي : Sampling distribution of the mean

تعريف (١١ : ٨)

التوزيع الاحتمالي للاحصائية يدعى بتوزيع المعاينة لتلك الاحصائية.

تعريف (١١ : ٩)

الانحراف القياسي لتوزيع المعاينة للاحصائية يدعى بالخطأ القياسي للاحصائية .

فالتوزيع الاحتمالي لـ $\bar{y}$ يسمى بتوزيع المعاينة للوسط الحسابي ، والانحراف القياسي للوسط الحسابي هو الخطأ القياسي لتوزيع المعاينة لـ $\bar{y}$.

مثال (١) نفرض بأن مجتمعاً متقطعاً متماثلاً Discrete uniform distribution يتألف من القيم التالية :

$$y = 0, 1, 2, 3$$

فالوسط الحسابي لهذا المجتمع هو :

$$\mu = \frac{\sum y_i}{N} = \frac{0+1+2+3}{4} = \frac{3}{2}$$

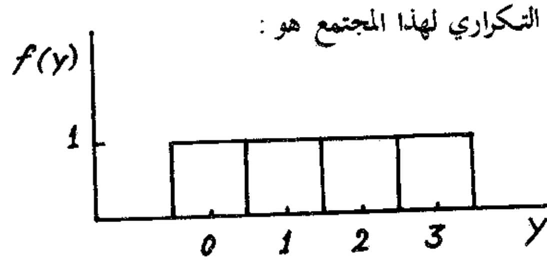
والتباين لهذا المجتمع هو :

$$\sigma^2 = \frac{\sum (y_i - \mu)^2}{N}$$

$$= \frac{(0-\frac{3}{2})^2 + (1-\frac{3}{2})^2 + (2-\frac{3}{2})^2 + (3-\frac{3}{2})^2}{4}$$

$$= \frac{5}{4}$$

والمدرج التكراري لهذا المجتمع هو :



شكل (١١ : ١) المدرج التكراري لمجتمع التوزيع المنتظم

والآن نفرض بأنه يراد أخذ كل العينات الممكنة بحجم (n = 2) من هذا المجتمع ، (تذكر بأنه في العادة لايتسنى لنا سوى الحصول على عينة واحدة فقط ومنها نستنتج

بعض خصائص المجتمع الذي تعود اليه تلك العينة . ولكن لكون المثال بسيطاً ولان الغاية هو دراسة التوزيع الاحتمالي للوسط الحسابي :

فإننا نفرض بأنه تم الحصول على جميع العينات الممكنة ذات حجم($n = 2$)من هذا المجتمع (بطريقة الارجاع (with replacement) وعددها $N = 16$ عينة) .
فجميع العينات الممكنة مع اوساطها الحسابي هي كما في جدول (١١ : ١)

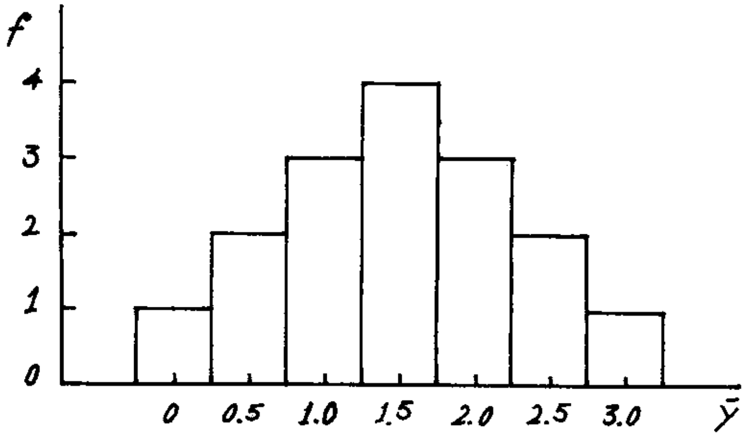
جدول (١١ : ١) الاوساط الحسابية للعينات العشوائية (بطريقة الارجاع)

رقم العينة	العينة		الوسط الحسابي للعينة
	y_1	y_2	$\bar{y}$
1	0	0	0
2	0	1	0·5
3	0	2	0·1
4	0	3	1·5
5	1	0	0·5
6	1	1	1·0
7	1	2	1·5
8	1	3	2·0
9	2	0	1·0
10	2	1	1·5
11	2	2	2·0
12	2	3	2·5
13	3	0	1·5
14	3	1	2·0
15	3	2	2·5
16	3	3	3·0

فالاحصائية $\bar{y}$ يتخذ لنفسه القيم $\bar{y}$ التي تتراوح من صفر الى ٣ والتوزيع التكراري لمجتمع الأوساط الحسابية للعينات هوكما في جدول (١١ : ٢)

جدول (١١ : ٢) توزيع المعاينة لـ y (بارجاع)

$\bar{y}$	f_i
0	1
0·5	2
1·0	3
1·5	4
2·0	3
2·5	2
3·0	1



شكل (١١ : ٢) المدرج التكراري للوسط الحسابي ($\bar{y}$)

والمدرج التكراري لهذه الأوساط الحسابية هو :

من الرسم أعلاه يتضح بأن توزيع المعاينة لـ $\bar{y}$ يقترب من المنحني الطبيعي بوسط حسابي قدره :

$$\mu_{\bar{y}} = \frac{\sum f\bar{y}_i}{\sum f_i} = \frac{24}{16} = \frac{3}{2} = \mu$$

وتباين قدره :

$$\sigma^2_{\bar{y}} = \frac{\sum f_i(\bar{y}_i - \mu_{\bar{y}})^2}{\sum f_i} = \frac{5}{8} = \frac{(5/4)}{2} = \frac{\sigma^2_y}{n}$$

$$\sigma^2_{\bar{y}} = \frac{\sigma^2}{n}$$

ان الوسط الحسابي والتباين لمجتمع الأوساط الحسابية قد حسبت من جدول التوزيع التكراري أعلاه .

من هذا نستنتج بأن الوسط الحسابي لـ $\bar{y}$ (أي الوسط الحسابي لجميع الأوساط الحسابية للعينات الممكنة) هو دائماً يساوي الوسط الحسابي للمجتمع الذي أخذت منه هذه العينات بينما التباين له (أي لـ $\bar{y}$) فيعتمد على تباين المجتمع وعلى حجم العينة أي $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$ وهو بذلك أقل من تباين المجتمع .

وبالنتيجة فانه كلما كبر حجم العينة قل الخطأ القياسي لـ $\bar{y}$ وقرب وسط تلك العينة من الوسط الحسابي للمجتمع . ولهذا فأن $\bar{y}$ ممكن استخدامه كتقدير لـ μ .

مثال (٢)

افرض أن لدينا نفس المجتمع (0, 1, 2, 3) وانه يراد أخذ كل العينات الممكنة وبحجم $n = 2$ ولكن بطريقة عدم الأرجاع without replacement :

فالعينات الممكنة مع أوساطها الحسابية هي كما في جدول (١١ : ٣)

جدول (١١ : ٣) الأوساط الحسابية للعينات العشوائية (بدون ارجاع)

رقم العينة	y_1	y_2	$\bar{y}$
1	0	1	0·5
2	0	2	1·0
3	0	3	1·5
4	1	2	1·5
5	1	3	2·0
6	2	3	2·5
7	1	0	0·5
8	2	0	1·0
9	3	0	1·5
10	2	1	1·5
11	3	1	2·0
12	3	2	2·5

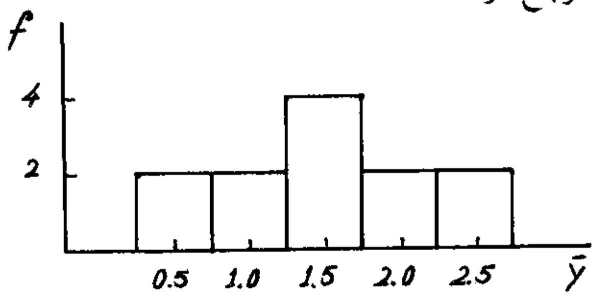
فالاحصائية $\bar{y}$ قيمة متغيرة من ٠,٥ الى ٢,٥ .

والتوزيع التكراري لمجتمع الأوساط الحسابية للعينات هو كما في جدول (١١ : ٤)

جدول (١١ : ٤) توزيع العاينة لـ $\bar{y}$ (بدون ارجاع)

$\bar{y}$	f
0·5	2
1·0	2
1·5	4
2·0	2
2·5	2

والمدرج التكراري لـ $\bar{y}$ بدون ارجاع هو :



شكل (١١ : ٣) المدرج التكراري للوسط الحسابي ($\bar{y}$)

$$\mu_{\bar{y}} = \frac{\Sigma f_{\bar{y}}}{\Sigma f} = \frac{18}{12} = \frac{3}{2} = \mu$$

$$\sigma_{\bar{y}}^2 = \frac{\Sigma f\left(\bar{y} - \frac{3}{2}\right)^2}{\Sigma f} = \frac{5}{12} = \frac{5/4}{2}\left(\frac{4-2}{4-1}\right) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

هذا واذا كانت N كبيرة نسبة الى n فان النتيجة $\frac{N-n}{N-1}$ تقترب من ١ . ولذا فان التباين سيصبح $\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}$ لذا ففي المجتمع الكبير أو غير المحدود سواء كان مستمراً أو متقطعاً فالنظرية التالية تنطبق عليه وهي من أهم النظريات التي تتعلق بتوزيعات المعاينة وتسمى نظرية النهاية المركزية Centeral limit theorm

نظرية (١١ : ١) :

اذا سحبت عينة عشوائية ذات حجم n من مجتمع كبير أو غير محدود (infinite) (له وسط حسابي μ وتباين σ^2) فان توزيع المعاينة للوسط الحسابي $\bar{y}$ يقترب من التوزيع الطبيعي بوسط حسابي قدره : $\mu_{\bar{y}} = \mu$

وانحراف قياسي : $\sigma_{\bar{y}} = \dfrac{\sigma}{\sqrt{n}}$

وبذا فان : $Z = \dfrac{\bar{y} - \mu}{\sigma/\sqrt{n}}$

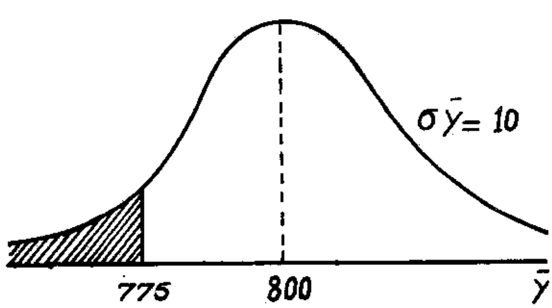
هي نتيجة من قيم المتغير الطبيعي القياسي Z الذي له وسط حسابي = صفراً وتباين قدره واحد .

مثال(٣) اذا كان توزيع عمر مصابيح احدى الشركات يقترب من التوزيع الطبيعي بوسط حسابي قدره ٨٠٠ ساعة وانحراف قياسي قدره ٤٠ ساعة أوجد احتمال : أن عينة عشوائية مؤلفة من ١٦ مصباحاً لها وسطاً حسابياً أقل من ٧٧٥ ساعة ؟

الحل :

$$\mu_{\bar{y}} = \mu = 800$$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{16}} = 10$$



$$Z = \frac{775 - 800}{10} = -\ 2.5$$

$$\therefore P(\bar{y} < 775) = P(Z < -\ 2.5)$$
$$= 0.006$$

(١١ : ٤) توزيع المعاينة للفروق بين الاوساط الحسابية :

Sampling Distribution of the Differences of Means

نفرض أن لدينا المجتمعين التاليين :

المجتمع الأول : وسطه الحسابي μ_1 وتباينه $\sigma^2{}_1$

المجتمع الثاني : وسطه الحسابي μ_2 وتباينه $\sigma^2{}_2$

ونفرض أن y_1 يمثل الوسط الحسابي لمجتمع الأوساط الحسابية للعينات العشوائية ذات حجم n_1 التي سحبت من المجتمع الأول.

وان $\bar{y}_2$ يمثل الوسط الحسابي لمجتمع الأوساط الحسابية للعينات العشوائية ذات حجم n_2 والتي سحبت من المجتمع الثاني مستقلة عن العينات من المجتمع الأول .

فتوزيع الفروق بين $\bar{y}_1 - \bar{y}_2$ للمجموعتين المستقلتين في الاوساط الحسابية يسمى بتوزيع المعاينة للاحصائية $(\bar{Y}_1\text{-}\bar{Y}_2)$

مثال (٤) : نفرض بأن المجتمع الاول يتألف من : 3,4,5

فالوسط الحسابي له :

$$\mu_1 = \frac{\Sigma Y_{1i}}{N_1}$$

$$= \frac{3+4+5}{3} = \frac{12}{3} = 4$$

والتباين له :

$$\sigma_1^2 = \frac{\Sigma (Y_{1i}-\mu_1)^2}{N_1}$$

$$= \frac{(3-4)^2 + (4-4)^2 + (5-4)^2}{3} = \frac{2}{3}$$

والمجتمع الثاني يتألف من قيمتين : 0,3

فالوسط الحسابي له :

$$\mu_1 = \frac{0+3}{2} = \frac{3}{2}$$

والتباين له :

$$\sigma_2^2 = \frac{\left(0-\frac{3}{2}\right)^2 + \left(3-\frac{3}{2}\right)^2}{2} = \frac{9}{4}$$

مثال (٥)

نفرض بأنه قد سحبت جميع العينات الممكنة وذات حجم $n_1 = 2$ بطريقة الارجاع من المجتمع الاول وحسبت $\bar{Y}_1$. وأنه قد سحبت جميع العينات الممكنة وذات الحجم $n_2 = 3$ بطريقة الارجاع أيضاً من المجتمع الثاني وحسبت $\bar{Y}_2$.

فجميع العينات الممكنة في كلا المجتمعين مع اوساطها الحسابية هي كما في جدول (١١ : ٥)

جدول (١١ : ٥) الأوساط الحسابية للعينات العشوائية (بطريقة الارجاع) من مجتمعين محدودين :

المجتمع الاول				المجتمع الثاني				
رقم العينة	العينات y_{11}	العينات y_{12}	$\bar{y}_1$	رقم العينة	العينات y_{21}	العينات y_{22}	العينات y_{23}	$\bar{y}_2$
1	3	3	3.0	1	0	0	0	0
2	3	4	3·5	2	0	0	3	1
3	3	5	4·0	3	0	3	0	1
4	4	3	3·5	4	3	0	0	1
5	4	4	4·0	5	0	3	3	2
6	4	5	4·5	6	3	0	3	2
7	5	3	4·0	7	3	3	0	2
8	5	4	4·5	8	3	3	3	3
9	5	5	5·0					

اما جميع الفروقات الممكنة بين $\bar{y}_1 - \bar{y}_2$ فهي كما في جدول (١١ : ٦)

جدول (١١ : ٦) الفروقات بين الأوساط الحسابية المستقلة .

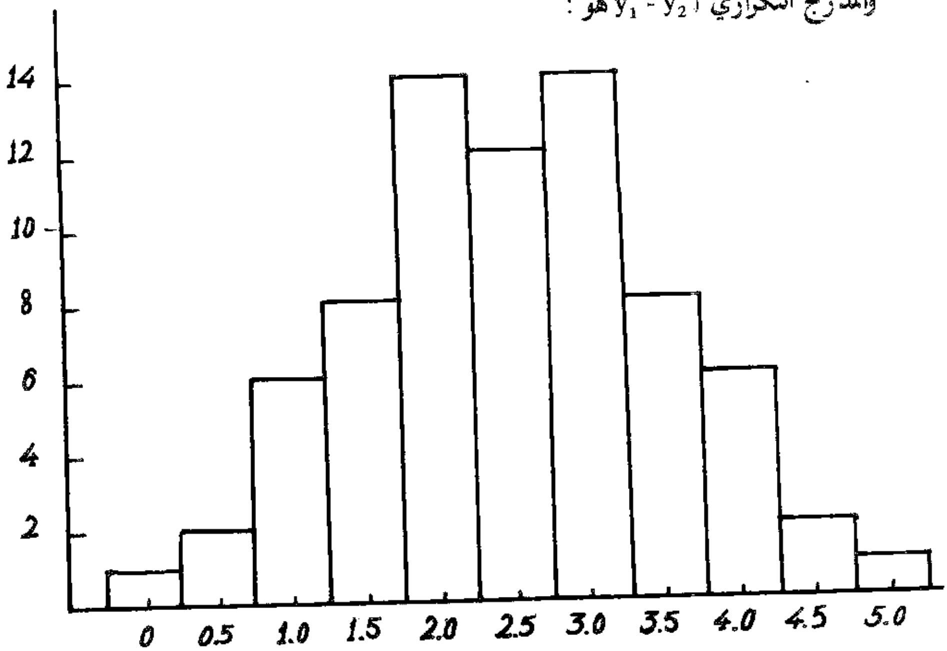
$\bar{y}_2$	$\bar{y}_1$								
	3·0	3·5	4·0	3·5	4·0	4·5	4·0	4·5	5·0
0	3·0	3·5	4·0	3·5	4·0	4·5	4·0	4·5	5·0
1	2·0	2·5	3·0	2·5	3·0	3·5	3·0	3·5	4·0
1	2·0	2·5	3·0	2·5	3·0	3·5	3·0	3·5	4·0
1	2·0	2·5	3·0	2·5	3·0	3·5	3·0	3·5	4·0
2	1·0	1·5	2·0	1·5	2·0	2·5	2·0	2·5	3·0 ...
2	1·0	1·5	2·0	1·5	2·0	2·5	2·0	2·5	3·0
2	1·0	1·5	2·0	1·5	2·0	2·5	2·0	2·5	3·0
3	0·0	0·5	1·0	0·5	1·0	1·5	1·0	1·5	2·0

والتوزيع التكراري لـ $\bar{y}_1 - \bar{y}_2$ هوكما في جدول (١١:٧)

جدول (١١:٧) : توزيع المعاينة لـ $(\bar{y}_1 - \bar{y}_2)$ (بطريقة الارجاع)

$\bar{y}_1 - \bar{y}_2$	f
0·0	1
0·5	2
1·0	6
1·5	8
2·0	13
2·5	12
3·0	13
3·5	8
4·0	6
4·5	2
5·0	1

والمدرج التكراري لـ $\bar{y}_1 - \bar{y}_2$ هو :



شكل (١١-٤) المدرج التكراري لـ ($\bar{y}_1 - \bar{y}_2$) بطريقة الارجاع

من هذا يتضح بأن المتغير العشوائي ($\bar{y}_1 - \bar{y}_2$) يقترب توزيعه من التوزيع الطبيعي. ويتحسن هذ التقرب كلما زادت قيمة n_2 و n_1 فالوسط الحسابي :

$$\mu_{\bar{y}_1 - \bar{y}_2} = E(\bar{y}_1 - \bar{y}_2) = E(\bar{y}_1) - E(\bar{y}_2) = \mu_1 - \mu_2$$

ويمكن حسابه ايضاً من البيانات كالاتي :

$$\mu_{\bar{y}_1 - \bar{y}_2} = \frac{\Sigma f_i (\bar{y}_1 - \bar{y}_2)}{\Sigma f_i} = 2{\cdot}5 = 4 - 1{\cdot}5 = \mu_1 - \mu_2$$

والتباين :

$$\sigma^2_{\bar{y}_1 - \bar{y}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$= \frac{2/3}{2} + \frac{9/4}{3} = \frac{13}{12}$$

ويمكن حسابه ايضاً من البيانات في الجدول اعلاه .

ان النتيجة التي حصلنا عليها لتوزيع المعاينة لـ $(\bar{y}_1 - \bar{y}_2)$ بطريقة السحب بالارجاع من مجتمع محدود هي نفسها لو سحبناها من مجتمع محدود مستمر أو متقطع وهي نفسها أيضاً لوكان المجتمع محدوداً والمعاينة بطريقة عدم الارجاع على أن يكون حجم المجتمعين N_2 و N_1 كبيراً نسبة الى n_2 و n_1 .

وفي هذا الكتاب سيكون اهتمامنا بتوزيع المعاينة للفرق بين وسطين حسابيين مستقلين في حالة كون حجمي المجتمعين اللذين سحبت منهما العينتان كبيرين .

نظرية (١١ : ٢)

اذا كانت العينات مستقلة وذات حجم n_2 و n_1 وسحبت من مجتمعين كبيرين أو غير محدودين (مستمر أو متقطع) بوسطين حسابيين μ_2 و μ_1 وتباينين σ_2^2 و σ_1^2 على التوالي فان توزيع المعاينة للفرق بين الوسطين الحسابيين $(\bar{y}_1 - \bar{y}_2)$ يقترب من التوزيع الطبيعي بوسط حسابي قدره :

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2$$

وانحرف قياسي قدره

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

لذا فإن :

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

هي قيمة من قيم المتغير الطبيعي القياسي Z

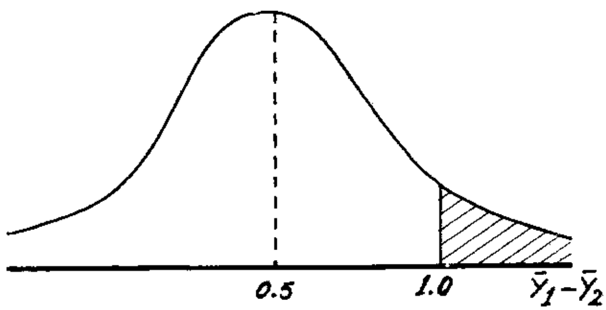
مثال(٦) اذا كانت شركة A تنتج شاشة تلفزيون متوسط عمرها ٦,٥ سنة وبإنحراف قياسي قدره ٠,٩ سنة . بينما شركة B تنتج شاشة تلفزيون متوسط عمرها ٦ سنة وبإنحراف قياسي ٠,٨ سنة . احسب احتمال أن عينة عشوائية ذات حجم ٣٦ شاشة من انتاج شركة A لها متوسط عمر على الأقل سنة أكثر من متوسط عمر عينة مؤلفة من ٤٩ شاشة من انتاج شركة B .

الحل : المعلومات التي اعطيت هي :

المجتمع الثاني (B)	المجتمع الاول (A)
$\mu_2 = 6{\cdot}0$	$\mu_1 = 6{\cdot}5$
$\sigma_2 = 0{\cdot}8$	$\sigma_1 = 0{\cdot}9$
$n_2 = 49$	$n_1 = 36$

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2 = 6{\cdot}5 - 6{\cdot}0 = 0{\cdot}5$$

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0{\cdot}81}{36} + \frac{0{\cdot}64}{49}} = 0{\cdot}189$$



$$Z = \frac{1{\cdot}0 - 0{\cdot}5}{0{\cdot}189} = 2{\cdot}646$$

$$\begin{aligned} \therefore P(\bar{y}_1 - \bar{y}_2 \geq 1{\cdot}0) &= P(Z > 2{\cdot}646) \\ &= 1 - P(Z < 2{\cdot}646) \\ &= 1 - 0{\cdot}9959 \\ &= 0{\cdot}0041 \end{aligned}$$

نظرية : (١١ : ٣)

اذا كان المتغيران العشوائيان x و y مستقلين ويتوزعان توزيعا طبيعيا بوسط حسابي μ_X و μ_Y وتباين σ_x^2 و σ_y^2 على التوالي فأن الفرق $x - y$ يتوزع ايضا توزيعا طبيعيا بوسط حسابي قدره : $\mu_{X-Y} = \mu_X - \mu_Y$

وتباين :

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

(١١ : ٦) توزيعات المعاينة للنسب

Sampling Distributions of Proportions

نعود الآن الى تجربة ذي الحدين التي فيها احتمال النجاح = p، واحتمال الفشل = q حيث ان $p + q = 1$

للحصول على عينة حجمها n فإنه يجب اعادة التجربة n من المحاولات.

ان توزيع المعاينة للمتغير العشوائي y الذي هو عدد النجاحات في العينات ذات الحجم n يمكن ان يكون قريبا من التوزيع الطبيعي بوسط حسابي قدره

$$\mu = np$$

وانحراف قياسي :

$$\sigma = \sqrt{npq}$$

على شرط ان لا تكون قيمة p قريبة من الصفر أو الواحد.

هذا ولكل عينة ذات حجم n من مجتمع ذي الحدين نستطيع تحديد النسبة $\hat{p}$ للنجاحات.

ان القيمة $\hat{p} = \frac{y}{n}$ تختلف من عينة لأخرى ولذا فيمكن اعتبارها قيمة من قيم الاحصائية $\hat{p}$.

فتوزيع المعاينة لـ $\hat{p}$ هو قريب من التوزيع الطبيعي بوسط حسابي قدره :

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{y}{n}\right) = \frac{np}{n} = p$$

وتباين قدره :

$$\sigma_{\hat{p}}^2 = \sigma^2{}_{y/n} = \frac{1}{n^2}\sigma_y^2$$

$$= \frac{npq}{n^2} = \frac{pq}{n}$$

نظرية (١١ : ٤)

اذا تم الحصول على عينات عشوائية ذات حجم n من مجتمع ذي الحدين الذي وسطه الحسابي $\mu = np$ وتباينه $\sigma^2 = npq$ فتوزيع المعاينة لـ $\hat{p}$ (نسبة النجاحات) هو قريب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\hat{p}} = p$$

وانحراف قياسي قدره :

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

وبذا فأن :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

(حيث ان $\hat{p} = \frac{y}{n}$)

هو قيمة من قيم المتغير الطبيعي القياسي Z

نظرية : (١١ : ٥)

اذا سحبت عينتان مستقلتان حجمهما n_1 و n_2 من مجتمعين من ذي الحدين وسطاهما

$$\mu_1 = n_1 p_1$$

$$\mu_2 = n_2 p_2$$

على التوالي

وتباينهما :

$$\sigma_1^2 = n_1 p_1 q_1$$

$$\sigma_2^2 = n_2 p_2 q_2$$

فتوزيع المعاينة لفرق النسبة $(\hat{p}_1 - \hat{p}_2)$ يقترب من التوزيع الطبيعي بوسط حسابي $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ وانحراف قياسي قدره

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

ولـذا فأن :

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

هي قيمة من قيم المتغير الطبيعي القياسي Z .

(١١ : ٧) توزيع المعاينة للتباين Sampling Distribution of the Variance

اذا اخترنا عينات عشوائية كل منها ذات حجم n من مجتمع طبيعي ذي تباين σ^2 ثم اعيد الاختيار لعدة مرات وحسب تباين كل عينة S^2 فأننا سنحصل على الاحصائية S^2 . ان توزيع المعاينة S^2 كما هو ليس ذا فائدة عملية في الاحصاء . ولكن بدلا من ذلك سوف ندرس توزيع المعاينة لـ $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$ الذي يدعى مربع كاي Chi - Square والذي قيمته تحسب في كل عينة بالقانون التالي :

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

ان توزيع χ^2 يطلق عليه توزيع مربع كاي Chi - Square distribution بدرجه حرجة تساوي $(v = n - 1)$

> **نظرية (١١ : ٦)**
>
> اذا كان S^2 هو تباين عينة عشوائية ذات حجم n اخذت من مجتمع طبيعي له تباين σ^2 فأن
>
> $$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$
>
> هي قيمة من قيم المتغير العشوائي χ^2 الذي له توزيع مربع كاي بدرجة حرجة $(v = n - 1)$

ان الوسط الحسابي للتباين هو :

$$\mu_{s^2} = \sigma^2$$

أما التباين للتباين

$$\sigma^2_{s^2} = \frac{2\sigma^4}{n-1}$$

تمارين الفصل الحادي عشر

(١) افترض بأن مجتمعاً يتألف من القيم التالية

$$y_i = 2,4,6$$

(أ) ارسم المدرج التكراري لـ $\bar{y}$ اذا اخذت جميع العينات الممكنة ذات حجم ($n = 4$) و (بطريقة الارجاع)

(ب) بين بأن

$$\mu_{\bar{y}} = \mu$$

$$\sigma^2_{\bar{y}} = \frac{\sigma^2}{n}$$

(٢) افترض المجتمع التالي :

$$y_i = 1, 1, 1, 2, 2, 3, 4$$

(أ) اعمل جميع العينات الممكنة ذات حجم 2 ($n = 2$) التي يمكن أخذها (بدون ارجاع) من هذا المجتمع .

(ب) بين بأن

$$\mu_{\bar{y}} = \mu$$

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

(٣) اذا كانت أطوال ١٠٠٠ طالب تتوزع طبيعياً بوسط حسابي قدره ٦٨,٥ أنج وانحراف قياسي قدره ٢,٧ انج واخذت ٢٠٠ عينة عشوائية ذات حجم ٢٥ من هذا المجتمع احسب :

(أ) الوسط الحسابي والانحراف القياسي المتوقع لتوزيع المعاينة لـ $\bar{y}$.

(ب) عدد العينات التي أوساطها الحسابية تقع بين : ٦٧,٩ و٦٤,٢

(ج) عدد الأوساط الحسابية التي تقع دون ٦٧,٠ .

(٤) اعطيت عينة حجمها $n = 100$ في المجتمع الطبيعي بوسط حسابي قدره (١٠) وتباين = ١٦ . ما هو الوسط الحسابي والتباين لمتوسط العينة ؟

(٥) افترض المجتمعين التاليين :

المجتمع	μ	n	σ
1	20	64	4
2	25	81	9

(أ) ما هو متوسط الفرق بين الوسطين الحسابيين .

(ب) احسب التباين لمتوسط الفرق بين الوسطين الحسابيين .

(٦) ما هو توزيع المعاينة لكل من :

(أ) $\hat{p}$

(ب) $\hat{p}_1 - \hat{p}_2$

(جـ) s^2

الفصل الثاني عشر

نَظَرِيَّةُ التَّقْدِيرْ

Estimation Theory

(١٢ : ١) مقدمـــة

ذكرنا سابقاً بأن العينة sample هي جزء من المجتمع Population وطريقة اختيار هذا الجزء يسمى طريقة المعاينة Sampling method والغاية الرئيسية من دراسة العينات هو للاستدلال منها على خواص المجتمع الذي تعود اليه هذه العينات .

فاذا كان لدينا المتغير العشوائي y فان توزيعه الاحتمالي (أو كثافة احتماله) تعتمد على ثابت θ (ثيتا theta) واحد أو أكثر لا تعرف قيمتها . وهذه الثوابت تسمى معالم Parameters وفي هذا الفصل سندرس طرق تقدير معالم المجتمع من مقاييس الاحصائيات Statistics التي تحسب من العينة حيث اننا نحتاج لحساب قيمة أو احصائية Statistic من العينة لكل معلمة من معالم المجتمع (أو دالة كثافة الاحتمال) .

هذا وكل قيمة تحسب من العينة تسمى تقديراً Estimate أما الطريقة التي استخدمت في التقديرفتسمى مقدراً Estimator فالتقدير غير ثابت من عينة الى أخرى عند استخدام نفس الطريقة بينما المقدر يكون ثابتاً الا اذا تغيرت طريقته .

هذا والتقدير اما ان يكون :

(١) تقدير المعلمة بنقطة Point Estimation

(٢) تقدير المعلمة بفترة Interval Estimation

(١٢ : ٢) تقدير النقطة Point Estimation

تعريف : (١٢ : ١)

اذا حسبت قيمة مفردة من العينة كتقدير لمعلمة من المجتمع فالطريقة تسمى تقدير النقطة Point Estimation، لأن نقطة واحدة فقط من فضاء العينة قد استخدم تقدير للمعلمة .

مثال(١) ان قيمة الوسط الحسابي للعينة $\bar{y}$ هو تقدير نقطة .Point est للمعلمة μ (الوسط الحسابي للمجتمع الذي تعود اليه هذه العينة) كما أن التباين S^2 للعينة هو تقدير نقطة لتباين المجتمع σ^2 وكذلك النسبة $\hat{p}$ هو تقدير لنسبة المجتمع P

هذا وبصورة عامة اذا رمزنا للمعلمة غير المعروف قيمتها بالرمز θ للتوزيع الاحتمالي للمتغير العشوائي y واذا كانت $y_1, y_2 \cdots y_n$ هي عينة عشوائية حجمها n سحبت من هذا التوزيع لذا فان أي تقدير للمعلمة θ (ويطلق عليه $\hat{\theta}$) يعتمد على مشاهدات العينة .

أي أن

$$\hat{\theta} = f(y_1 , y_2 .... y_n)$$

لذا فان $\hat{\theta}$ هو متغير (لانه يختلف من عينة الى أخرى) وله توزيع معاينة بوسط حسابي وتباين خاص به .

هذا وكما ذكرنا سابقاً فأن الوسط الحسابي والمنوال والوسيط التي تحسب من العينات ما هي الا تقديرات للوسط الحسابي للمجتمع μ الا ان خصائص المقدر الجيد هي :

(١) عدم التحيز Unbiasedness

تعريف (١٢ : ٢)

ان المقدر $\hat{\theta}$ يعتبر مقدراً غير متحيز اذا كان توقعه يساوي قيمة المعلمة θ أي

$$E(\hat{\theta}) = \theta$$

مثال(٢) : افرض بأن $y_1 , y_2 .... y_n$ هي عينة عشوائية من مجتمع له وسط حسابي μ

لذا فأن $E(\bar{y}) = \mu$

وكـــذلك $E(y_i) = \mu$

لذا فإن كلا من $\bar{y}$ و y_i هي مقدرا غير متحيز للوسط الحسابي للمجتمع μ .

(٢) الاتساق Consistency

تعريف : (١٢ : ٣)

يكون المقدر متسقاً اذا كانت قيمته لا تختلف اختلافاً جوهرياً عن قيمة المعلمة الحقيقية للمجتمع بزيادة حجم العينة أي أن

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

حيث ان $\in$ هو الفرق بين المقدر والمعلمة .

(٣) الكفـــاءة Efficiency

تعريف : (١٢ : ٤)

ان كفاءة المقدر غير المتحيزة $\hat{\theta}_1$ الى المقدر غير المتحيز $\hat{\theta}_2$ هو نسبة تباين المقدر $\hat{\theta}_2$ الى تباين المقدر $\hat{\theta}_1$ أي

$$e(\hat{\theta}_1) == \frac{V(\hat{\theta}_2)}{V(\theta_1)}$$

ومن التعريف أعلاه يتضح بأن المقدر الأقل تبايناً هو الأعلى كفاءة .

مثـال(٣) : الوسط الحسابي للعينة $\bar{y}$ هو مقدر غير متحيز لـ μ وكذلك الوسيط للعينة $\bar{\mu}$ هو مقدر غير متحيز لـ μ ولكن تباين $\bar{y}$ هو $\frac{\sigma^2}{n}$ بينما تباين الوسيط هو

$$\frac{\pi}{2}\ \frac{\sigma^2}{n}$$

لذا فان كفاءة الوسط الحسابي هـو

$$e(\bar{y}) = \frac{V(\bar{\mu})}{V(\bar{y})} = \frac{\frac{\pi}{2}\ \frac{\sigma^2}{n}}{\sigma^2/n} = \frac{\pi}{2}$$

أي أن الوسط الحسابي $\bar{y}$ هو أكفأ تقديراً من الوسيط .

(٤) الكفـــاية : Sufficiency

تعريف : (١٢ : ٥)

يكون المقدر $\hat{\theta}$ مقدراً كافياً للمعلمة θ اذا كان قد شمل كل المعلومات ذات العلاقة بـ θ المتوافرة في العينة .

فعند سحب عينة عشوائية في مجتمع يتوزع توزعاً طبيعياً فأن $\bar{y}$ هو مقدر كاف لـ μ لأنه لا يمكن اضافة أي شيء على $\bar{y}$ لجعله مقدراً أحسن لـ μ لأن $\bar{y}$ يحوي على جميع المعلومات المتعلقة بـ μ من العينة .

هذا وهناك ٤ طرق مهمة لايجاد تقدير المعلمة بنقطة .Point est ولعدم وجود مجال لشرحها هنا فسنذكرها فقط وهي :

١. طريقة الامكان الاكبر

Maximum Likelihood Method

٢. طريقة العزوم

Method of Moments

٣. طريقة مربع كاي المصغرة

Minimum Chi – Square Method

٤. طريقة المربعات الصغرى

Method of Least Square

(١٢ : ٣) تقدير فترة

Interval Estimation

تعريف : (١٢ : ٦)

ان طريقة تحديد فترة (a, b) التي تضم معلمة Parameter المجتمع θ، باحتمال قدره $(1 - \alpha)$ تسمى تقدير فترة

حيث ان:

هو احتمال ان الفترة لا تضم المعلمة = α -

لذا فنحن نقول بأن

$$P(a < \theta < b) = 1 - \alpha$$

حيث ان a,b هما متغيران عشوائيان يعتمدان على المقدر θ للمعلمة θ وان

الحد الأدنى للفترة = a

الحد الأعلى للفترة = b

وان:

فترة الثقة = (a,b)

وان:

هي قياس لدقة التقدير = $b - a$

وان:

هي قياس الثقة = $1 - \alpha$

(١) تقدير فترة الثقة للوسط الحسابي للمجتمع (μ)

(أ) عندما يكون الانحراف القياسي للمجتمع σ معلوم

ذكرنا سابقاً بأنه حسب نظرية النهاية المركزية Central limit theory :

اذا كان y أي متغير عشوائي بمتوسط μ وتباين σ^2 فان توزيع المعاينة

للوسط الحسابي $\bar{y}$ لعينة كبيرة حجمها n يقترب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\bar{y}} = \mu$$

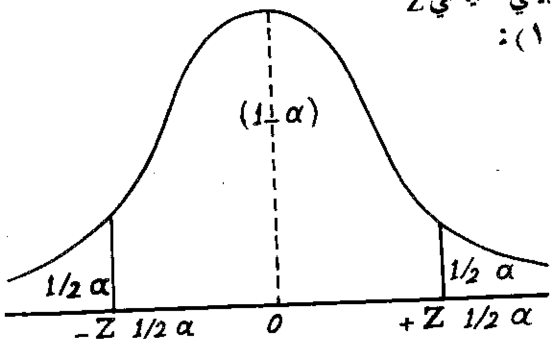
وانحراف قياسي قدره

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

وبذا فان .

$$Z = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

هي قيمة من قيم المتغير الطبيعي القياسي Z
ومن النظر الى الشكل (١٢: ١) :



شكل (١٢ : ١) احتمال ان تقع قيمة Z بين القيمتين
$Z_{1/2\,\alpha}$, $-Z_{1/2\,\alpha}$ هو $(1-\alpha)$

فان احتمال ان Z تقع بين القيمتين $Z_{\alpha/2}$, $-Z_{\alpha/2}$ هي

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

فاذا عوضنا عن Z بما يساويها وهي $\frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ينتج

$$P\left(-Z_{\alpha/2} < \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\alpha/2}\right) = 1 - \alpha$$

وبضرب كل حد بـ $\frac{\sigma}{\sqrt{n}}$ ينتج

$$P(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{y}-\mu < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

وبطرح $\bar{y}$ من كل حد ثم الضرب بـ (-١) نحصل على

$$P(\bar{y}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{y}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

فاذا أردنا انشاء فترة ثقة ٩٥./· فأن

$$1-\alpha = \cdot 95$$

$$\therefore \alpha = \cdot 05$$

$$\therefore \quad +Z_{\alpha/2} = Z_{(\cdot 025)} = 1\cdot 96$$

$$-Z_{\alpha/2} = -Z_{(\cdot 025)} = -1\cdot 96$$

$$\therefore \; P(\bar{y}-1\cdot 96\frac{\sigma}{\sqrt{n}} < \mu < \bar{y}+1\cdot 96\frac{\sigma}{\sqrt{n}}) = \cdot 95$$

وبنفس الطريقة فأن فترة ثقة ٩٩./· هي

$$P(\bar{y}-2\cdot 58\frac{\sigma}{\sqrt{n}} < \mu < \bar{y}+2\cdot 58\frac{\sigma}{\sqrt{n}}) = \cdot 99$$

تعريف : (١٢ : ٧)

ان فترة ثقة 100 ($1-\alpha$)لـ μ عندما تكون σ^2 معلومة هي

$$\bar{y}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{y}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

حيث ان

$\bar{y}$ هو الوسط الحسابي لعينة حجمها n عن مجتمع تباينه σ^2 معلوم وان

$Z_{\alpha/2}$ هي قيمة المتغير الطبيعي القياسي الذي يترك مساحة $\alpha/2$ الى اليمين

مثال(٤) : سحبت عينة عشوائية حجمها ٣٦ مفردة من مجتمع انحرافه القياسي ٠,٣ فوجد ان الوسط الحسابي للعينة كان ٢,٦ .

احسب فترة ثقة ٩٥./· للوسط الحسابي للمجتمع الذي سحبت منه هذه العينة .

الحـل :

$$(1-\alpha)100 = \cdot 95$$

$$\therefore \alpha = \cdot 05$$

$$\therefore Z_{\alpha/2} = Z_{\cdot 025} = 1\cdot 96$$

لذا فان فترة ثقة ٩٥./ هو

$$\bar{y} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{y} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

أي

$$2\cdot 6 - (1\cdot 96)\left(\frac{\cdot 3}{\sqrt{36}}\right) < \mu < 2\cdot 6 + (1\cdot 96)\left(\frac{\cdot 3}{\sqrt{36}}\right)$$

$$2\cdot 50 < \mu < 2\cdot 70$$

أي أن هناك احتمال قدره ٠,٩٥ بأن الوسط الحسابي للمجتمع الذي سحبت منه هذه العينة يقع بين القيمتين ٢,٧٠ و ٢,٥٠ .

(ب) عندما يكون الانحراف القياسي للمجتمع (σ) غير معلوم .

اذا كان الانحراف القياسي للمجتمع الذي سحبت منه العينة غير معلومٍ وكان حجم العينةكبيراً أي($n > 30$) فانه يمكن استعمال S (الانحراف القياسي المحسوب من العينة) بدلاً من σ وبذلك تكون فترة ثقة ٩٥./ للوسط الحسابي للمجتمع هي

$$\left(\bar{y} - 1\cdot 96\frac{S}{\sqrt{n}} < \mu < \bar{y} + 1\cdot 96\frac{S}{\sqrt{n}}\right)$$

وفترة الثقة ٩٩./ هي

$$\left(\bar{y} - 2\cdot 58\frac{S}{\sqrt{n}} < \mu < \bar{y} + 2\cdot 58\frac{S}{\sqrt{n}}\right)$$

مثـال(٥) : سحبت عينة عشوائية حجمها ٣٦ طالباً من جامعة ما في العراق فكانت معدل أوزانهم ١٦٠ باوند بانحراف قياسي قدره ٣٠ باوند اوجد فترة ثقة ٩٥./ للوسط الحسابي لأوزان جميع طلبة تلك الجامعة

$$\bar{y} - 1\cdot 96\frac{S}{\sqrt{n}} < \mu < \bar{y} + 1\cdot 96\frac{S}{\sqrt{n}}$$

$$160 - 1.96\frac{30}{\sqrt{36}} < \mu < 160 + 1.96\frac{30}{\sqrt{36}}$$

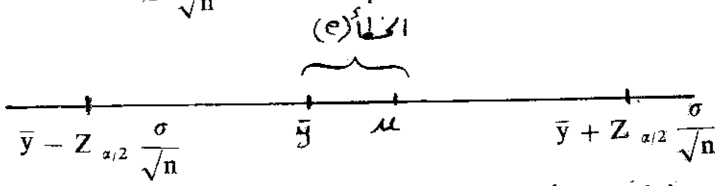
$$150.2 < \mu < 169.8$$

أي ان

$$P(150.2 < \mu < 169.8) = .95$$

ملاحظة : اذا وقعت μ في منتصف الفترة معنى ذلك ان $\bar{y}$ يعطى تقديراً لـ μ بدون خطأ (error) . ولكن في معظم الاحيان فان $\bar{y}$ لايكون مساوياً لـ μ بل يختلف عنه . هذا الاختلاف بين قيمة $\bar{y}$ و μ هو مقياس للخطأ .

لذلك فان $\bar{y}$ يختلف عن μ بكمية اقل من $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$



وفي كثيرمن الأحيان يكون اهتمامنا بحجم العينة التي تعطينا كمية معلومة من الخطأ (e). وفي هذه الحالة فاننا نختار n بحيث ان

$$e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\therefore n = \left(\frac{Z_{\alpha/2}\cdot\sigma}{e}\right)^2$$

وفي حالة σ غير معلومة فاننا نعوض عنها بـ (S) أي

$$n = \left(\frac{Z_{\alpha/2}\,S}{e}\right)^2$$

مثال(٦) : ما هو حجم العينة التي نختارها لنكون على ثقة ٩٥ ./· بان الوسط الحسابي للمجتمع μ يختلف عن $\bar{y}$ بأقل من ٠,٠٦ علماً بأن الانحراف القياسي للمجتمع ٠,٣ ؟

الحــل :

$$1 - \alpha = .95$$
$$\therefore \alpha = .05$$
$$\therefore Z_{\alpha/2} = 1.96$$
$$\sigma = .3$$
$$e = .06$$

$$\therefore n = \left(\frac{Z_{\alpha/2}\ \sigma}{e}\right)^2 = \left(\frac{(1.96)(.3)}{0.06}\right)^2 = 96.04$$

لذلك بثقة ٩٥٪ فان العينة العشوائية ذات حجم ٩٦ تعطي تقدير $\bar{y}$ يختلف عن μ بكمية اقل من ٠.٠٦ .

(٢) تقدير فترة ثقة للفرق بين الوسطين الحسابيين لمجتمعين

(أ) في حالة تباين المجتمعين معلومين .

اذا كان لدينا مجتمعين وسطهما الحسابي μ_1 و μ_2 وتباينهما σ_1^2 و σ_2^2 فكما ذكرنا سابقا (نظرية ١١ : ٢) فان توزيع المعاينة لـ $(\bar{y}_1 - \bar{y}_2)$ يقترب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2$$

وانحراف قياسي قدره

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

لذا فان

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

هي قيمة المتغير الطبيعي القياسي Z

واحتمال Z تقع بين القيمتين $Z_{\alpha/2}$ و $-Z_{\alpha/2}$ هي

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

فاذا عوضنا عن Z بما يساويها اعلاه ينتج

$$P\left(-Z_{\alpha/2} < \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z_{\alpha/2}\right) = 1 - \alpha$$

وبضرب كل حد بـ $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ ثم بطرح $(\bar{y}_1 - \bar{y}_2)$ من كل حد ثم ضرب كل حد بـ (‎-‎١) ينتج

$$P\left[(\bar{y}_1 - \bar{y}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{y}_1 - \bar{y}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha$$

فاذا اردنا انشاء فترة ثقة ٩٥ ٪ فان :

$$1 - \alpha = \cdot 95$$
$$\therefore \alpha = \cdot 05$$
$$+ Z_{\alpha/2} = Z_{\cdot 025} = 1\cdot 96$$
$$- Z_{\alpha/2} = - Z_{\cdot 025} = - 1\cdot 96$$

$$\therefore P\left((\bar{y}_1 - \bar{y}_2) - 1\cdot 96 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{y}_1 - \bar{y}_2) + 1\cdot 96 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 1 - \alpha$$

وبنفس الطريقة فان فترة ثقة ٩٩ ٪ هي

$$P\left(\bar{y}_1 - \bar{y}_2 - 2\cdot 58 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{y}_1 - \bar{y}_2 + 2\cdot 58 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 1 - \alpha$$

تعريف : (١٢ : ٨)

ان فترة ثقة $100(1 - \alpha)$ ٪ لـ $(\mu_1 - \mu_2)$ عندما تكون σ_1^2 و σ_2^2 معلومتان هي

$$\left(\bar{y}_1 - \bar{y}_2 \right) - Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{y}_1 - \bar{y}_2 \right) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

حيث ان

$\bar{y}_1$ و $\bar{y}_2$ هما وسطان حسابيان لعينتين عشوائيتين مستقلتين ذات حجم n_1 , n_2 على التوالي من مجتمعين لهما تباين σ_1^2 و σ_2^2 على التوالي وان $Z_{\alpha/2}$ هي قيمة المتغير الطبيعي القياسي الذي يترك مساحة $\alpha/2$ الى اليمين .

مثال (٧) : في احدى الامتحانات القياسية في الكيمياء سحبت عينة مؤلفة من ٥٠ طالبة من مجتمع له انحراف قياسي ٦ فكان معدل درجاتهن هو ٧٦ .
وسحبت عينة مؤلفة من ٧٥ طالباً من مجتمع له انحراف قياسي ٦ فكان معدل درجاتهم ٨٢ .
اوجد فترة ثقة ٩٦ ٪ للفرق $(\mu_1 - \mu_2)$ حيث ان:

معدل جميع الطلاب (مجتمع الطلاب) = μ_1

معدل جميع الطالبات (مجتمع الطالبات) = μ_2

الذين قد اخذوا هذا الامتحان .

الحـل : نطبق القانون التالي

$$(\bar{y}_1 - \bar{y}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{y}_1 - \bar{y}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$82{\cdot}76 - 2{\cdot}054\sqrt{\frac{64}{75} + \frac{36}{50}} < (\mu_1 - \mu_2) < 82{\cdot}76 + 2{\cdot}054\sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$3{\cdot}42 < (\mu_1 - \mu_2) < 8{\cdot}58$$

(ب) عندما يكون التباين لكلا المجتمعين (σ_2^2 , σ_1^2) غير معلومين :

اذا كان الانحراف القياسي لكلا المجتمعين الذين سحبت منهما العينات غير معلومين وكان حجم العينات كبيراً أي ($n_1,n_2 > 30$) فانه يمكن استعمال S_2 , S_1 (الانحراف القياسي المحسوب من كل عينة) بدلاً من σ_1 و σ_2 على التوالي وبذلك تكون فترة الثقة

$$(\bar{y}_1 - \bar{y}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < (\mu_1 - \mu_2) < (\bar{y}_1 - \bar{y}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

مثـال(٨) : سحبت عينة عشوائية مؤلفة من ٤٥ عاملاً من مصنع ما من الحاصلين على شهادة السادس الثانوي فكان معدل اجرهم الشهري هو ٥٠ ديناراً بتباين قدره ٤٨ . وسحبت عينة عشوائية مؤلفة من ٦٠ عاملاً من الحاصلين على شهادة الثالث ثانوي فكان معدل أجرهم الشهري هو ٤٣ ديناراً بتباين قدره ٥٦ . احسب فترة ثقة ٩٥./٠ للفرق بين الوسطين الحسابيين لمجتمعيهما .

الحـل :

$$(50 - 43) - 1{\cdot}96\sqrt{\frac{48}{45} + \frac{56}{60}} < \mu_1 - \mu_2) < (50 - 43) + 1{\cdot}96\sqrt{\frac{48}{45} + \frac{56}{60}}$$

$$4{\cdot}23 < (\mu_1 - \mu_2) < 9{\cdot}77$$

أي باحتمال ٩٥٪ فإن الفرق بين وسطيهما الحقيقيين يقع بين ٩,٧٧ و٤,٢٣ ديناراً .

(٣) تقدير فترة ثقة للنسبة في مجتمع :

ذكرنا سابقاً في نظرية (١١ : ٤) بان توزيع المعاينة لـ $\hat{p}$ (نسبة النجاحات) هو قريب من التوزيع الطبيعي بوسط حسابي قدره

$$\mu_{\hat{p}} = p$$

وانحراف قياسي قدره

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

لذا فان

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

حيث أن

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

فبالتعويض عن Z ينتج

$$P\left(-Z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} < Z_{\alpha/2}\right) = 1 - \alpha$$

وبضرب كل حد بـ $\sqrt{\frac{pq}{n}}$ ثم طرح $\hat{p}$ من كل حد وبعدها ضرب كل حد بـ − ١ ينتج

$$P\left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

وعندما تكون n كبيرة فيمكن استبدال p تحت الجذر بـ $\hat{p} = \frac{y}{n}$ فتصبح

$$P\left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \simeq 1 - \alpha$$

ولعينة معينة ذات حجم n فترة ثقة $100(1-\alpha)$ لـ p هي

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

تعريف (١٢ : ٩) :

ان فترة ثقة $100(1-\alpha)$ لـ p في توزيع ذو الحدين هو تقريباً

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

حيث أن $\hat{p}$ = نسبة النجاح في العينة العشوائية ذات حجم n

$$\hat{q} = 1 - \hat{p}$$

وأن $Z_{\alpha/2}$ هي قيمة المتغير الطبيعي القياسي الذي يترك مساحة $\alpha/2$ الى اليمين .

مثال (٩) : في عينة عشوائية ذات حجم (n = 500) عائلة تملك التلفزيون في بغداد . وجد أن (y = 160) عائلة تملك تلفزيون ملون . احسب فترة ثقة ٩٥ ٪ للنسبة الحقيقية لمالكي التلفزيون الملون في بغداد .

الحـل :

$$\hat{p} = \frac{160}{500} = 0{\cdot}32 \qquad \therefore \hat{q} = {\cdot}68$$

$$Z_{\alpha/2} = Z_{\cdot 025} = 1{\cdot}96$$

$$\left(\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

$${\cdot}32 - 1{\cdot}96\sqrt{\frac{({\cdot}32)({\cdot}68)}{500}} < p < {\cdot}32 + 1{\cdot}96\sqrt{\frac{({\cdot}32)({\cdot}68)}{500}}$$

$${\cdot}28 < p < {\cdot}36$$

(٤) تقدير فترة ثقة للفرق بين نسبتين لمجتمعين .

لانشاء فترة ثقة لـ $p_1 - p_2$ نرجع الى نظرية (١١ : ٥) التي تقول بأن توزيع المعاينة لـ $\hat{p}_1 - \hat{p}_2$ عندما تكون n كبيرة يكون قريباً من التوزيع الطبيعي بوسط حسابي قدره $\mu_{\bar{p}_1 - \bar{p}_2} = p_1 - p_2$

وبانحراف قياسي قدره

$$\sigma_{\bar{p}_1 - p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

لذا فان

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1-\alpha$$

حيث أن

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}$$

وبالتعويض يكون

$$P\left(-z_{\alpha/2} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}} < z_{\alpha/2}\right) = 1-\alpha$$

وبضرب كل حد بـ $\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$ وطرح $\hat{p}_1 - \hat{p}_2$ من كل حد ثم بضرب كل حد بـ (−١) ينتج

$$P\left[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right] = 1-\alpha$$

وعندما تكون n_2, n_1 كبيرة نعوض عن p_1 و p_2 تحت الجذر بـ $p_1 = \dfrac{y_1}{n_1}$ ، $\hat{p}_2 = \dfrac{y_2}{n_2}$ فينتج

$$P\left(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}\right) = 1-\alpha$$

ولأي عينتين عشوائيتين مستقلتين فان فترة الثقة $100(1-\alpha)$ هي

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

تعريف (١٢ : ١٠) :

إن فترة ثقة $100(1-\alpha)$ لـ $(p_1 - p_2)$ عندما تكون $(n_1 , n_2 > 30)$ هي تقريباً مساوية الى

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

حيث ان

$\hat{p}_1$, $\hat{p}_2$ هي نسبة النجاح في العينتين ذات حجم n_1 , n_2 على التوالي
وان $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{q}_2 = 1 - \hat{p}_2$
و $z_{\alpha/2}$ قيمة المتغير الطبيعي القياسي الذي يترك مساحة $\alpha/2$ الى اليمين .

مثال (١٢) : اخذت عينة عشوائية حجمها ٥٠٠٠ رجلاً في أرياف مدينة أ فكان ٢٤٠٠ منهم من الاميين واخذت عينة عشوائية حجمها ٢٠٠٠ رجلٍ في ارياف مدينة ب فكان ١٢٠٠ منهم أمياً أحسب فترة ثقة ٩٠٪ للفرق بين نسبة الاميين الحقيقيين في ارياف المدينتين .

الحـــل :
نفرض ان نسبة الاميين الحقيقية في المدينة أ هي p_1
وللمدينة ب هي p_2

$$\hat{p}_1 = \frac{2400}{5000} = \cdot 48$$

$$\hat{p}_2 = \frac{1200}{2000} = \cdot 60$$

$$z_{\alpha/2} = z_{\cdot 05} = 1\cdot 645$$

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$-0.12-1.645\sqrt{\frac{(.48)(.52)}{5000}+\frac{(.6)(.4)}{2000}}<(p_1-p_2)<-0.12$$

$$+1.645\sqrt{\frac{(.48)(.52)}{5000}+\frac{(.60)(140)}{2000}}$$

$$-.1414<p_1-p_2<-0.0986$$

وبما ان كلا طرفي الفترة سالباً لذا فإننا نستنتج بأن نسبة الاميين في محافظة ب أعلى من نسبة الأميين في محافظة أ .

تمارين الفصل الثاني عشر

١- اذا توفرت لديك البيانات التالية لعينات من مجتمعات طبيعية فما هي أحسن التقديرات لكل من المتوسط ، التباين ، تباين المتوسط ، الانحراف القياسي ، والإنحراف القياسي للمتوسط ؟

آ- $n=9, \sum y_i=36, \sum(y_i-\bar{y})^2 = 288$

ب- $n=9, \bar{y}=50, \sum(y_i-\bar{y})^2 = 32$

جـ- $n=16, \sum y_i=320, \sum y_i^2=664$

٢- لكل من العينات التالية المأخوذة من مجتمعات طبيعية أوجد أحسن التقديرات لما يأتي : $\sigma_{\bar{y}}, \sigma^2_{\bar{y}}, \sigma, \sigma^2, \mu$

آ- $12, 6, 15, 3, 12, 6, 21, 15, 18$

ب- $2, 5, 9, 11, 13$

جـ- $-4, 2, -6, 0, -4, 6, 2, 4, 0$

٣- بإفتراض عينات عشوائية من مجتمعات معروف تبايناتها . أوجد فترات الثقة حول المتوسطات عند درجات الثقة المبينة أمام كل منها :

آ- فترة ثقة $n=36, \bar{y}=20, \sigma^2=9, 95\%$

ب- فترة ثقة $n=49, \bar{y}=52, \sigma^2=64 \quad 99\%$

٤- بافتراض عينات من مجتمعات طبيعية معروف تبايناتها . أوجد :

أ- درجة الثقة المستخدمة اذا علمت بأن $\sigma=8, n=36$. وبأن المدى الكلي لفترة الثقة حول المتوسط هو ٣,٢٩ وحدة .

ب – حجم العينة اذا علمت بأن $\sigma^2 = 100$ وان فترة الثقة حول المتوسط عند درجة ثقة ٩٥٪ هي من ١٧٫٢ الى ٢٢٫٨ وحدة .

جـ – التباين المعروف عندما $n = 100$ فترة الثقة حول المتوسط عند درجة ثقة ٩٨٪ بمدى يساوي ٢٣٫٢٦ وحدة .

٥ – في تجربة لمقارنة صنفين من الذرة ، زرع ٥٠ ايكر لكل صنف في ظروف متشابهة . فأعطى الصنف (A) . كمعدل . ٧٨٫٣ بشل / ايكر مع انحراف قياسي قدره ٥٫٦ بشل / ايكر . بينما الصنف B أعطى . كمعدل ، ٨٧٫٢ بشل / ايكر بانحراف قياسي قدره ٦٫٣ بشل /ايكر .
والمطلوب أنشاء فترة ثقة ٩٥٪ للفرق بين الوسطين الحسابين لمجتمعيهما .

٦ – في عينه عشوائية ذات حجم ٤٠٠ شخصاً وجد بأن ١٠٪ منهم لايؤيدون اعادة انتخاب السيد × كرئيس للجمهورية للمرة الثانية أوجد فترة ثقة ٩٩٪ للنسبة الحقيقية للأشخاص الذين لايؤيدون السيد × .

٧ – أخذت عينة عشوائية مؤلفة من ٢٠٠ شخصاً فوجد بأن ٤٢ منهم يفضلون سيكاير (بغداد) . وأخذت عينه أخرى عشوائية حجمها ١٥٠ شخصاً فوجد بأن ١٨ شخصاً يفضلون سيكاير (ريم) .
احسب فترة ثقة ٩٥٪ للفرق بين النسبتين الحقيقيتين للذين يفضلون سيكاير بغداد وريم .

(٨ – عرف مايلي :
(أ) تقدير نقطة (ب) تقدير فترة
(جـ) الكفاءة (د) الكفاية