الفصل الثاني

طبيعة البيانات والرموز الاحصائية

(٢ : ١) طبيعة البيانات الاحصائية :

عند جمع بيانات حول ظاهرة ما فإننا نرمز للظاهرة بالرمز (y) وكل مفردة اومشاهدة 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 بين١٣٠,٥ و ١٧٠ سم فنقول بأن :

$$(130{\cdot}5 \leq y \leq 170{\cdot}0)$$

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 ... لأنه 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 الاسرة 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 $(\sum x_i)(\sum y_i)$

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

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

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

وان قيم المتغير x_i 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) $\sum x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4$

$= (4)(3) + (2)(9) + (3)(6) + (7)(2)$

$= 62$

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

$= (16)(20)$

$= 320$

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 $\sum cy_i = c\sum y_i$

البرهان :

$$\begin{aligned}\sum cy_i &= cy_1 + cy_2 + \dots + cy_n \\ &= c(y_1 + y_2 + \dots + y_n) \\ &= c\sum y_i\end{aligned}$$

قاعدة (٣)

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

البرهان :

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

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

هذا ويجب التفريق بين بعض الرموز الاحصائية مثل :

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

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

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

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) $\sum (x_i - 3)(y_i - 5)$

(ج) $\sum x_i y_i^2$ (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 .

(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 :

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

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

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

$$=20$$

(جـ) $$\sum x_iy_i^2 = x_1y_1^2 + x_2y_2^2 + x_3y_3^2 + x_4y_4^2$$

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

$$= 616$$

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

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

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

$$= 20 - 12$$

$$= 8$$

(هـ) $$\sum y_i - 3 = 20 - 3$$

$$= 17$$

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

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

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

$$= 20$$