Page: 124, 125, 127, 128   / /

مقاييس التشتت : تستخدم لتحديد انتشار البيانات.

المدى   القيمة الكبرى   ناقص   Measures of Variation

1. Range: The range is the highest value minus the lowest value. The symbol R is used for the range.

القيمة الصغرى

$R = Max - Min$, R = Highest value − Lowest value

Exp: 60 90 80 70 50

Solution: $R = 90 - 50 = 40$

Exp 3-19, Exp 3-20

التباين

2. Variance: The variance is the average of the squares of the distance each value is from the mean. The symbol for the population variance is σ^2.

$$\sigma^2 = \frac{\Sigma(X-\mu)^2}{N}$$

الانحراف المعياري   الجذر التربيعي

Standard deviation: The standard deviation is the square root of the variance. The symbol for the population standard deviation is σ.

التباين

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(X-\mu)^2}{N}}$$

The formula for the Sample variance, denoted by s^2, is

$$s^2 = \frac{\Sigma(X-\overline{X})^2}{n-1}$$

The " " " Standard deviation of Sample, denoted by s, is

X = individual value $$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n-1}}$$

$\overline{X}$ = Sample mean

n = Sample size

Page: 129, 132

Shortcut or Computational Formulas for S^2 and S.

$$S^2 = \frac{n(\sum X^2) - (\sum X)^2}{n(n-1)} , \quad S = \sqrt{\frac{n(\sum X^2) - (\sum X)^2}{n(n-1)}}$$

Exp 3-23

ملاحظة : ١. مقاييس التشتت دائماً موجبة.
٢. " " تساوي صفر (0) إذا كانت متساوية في قيم البيانات (70 70 70 70)
٣. كلما تقترب للصفر (0) يقل التشتت.

Exp:

$\overset{1}{60} + \overset{2}{90} + \overset{3}{80} + \overset{4}{70} + \overset{5}{50}$

$3600 + 8100 + 6400 + 4900 + 2500$

$n = 5$

$\sum x = 350$

$\sum x^2 = 25500$

1. $S^2 = \frac{(5)(25500) - (350)^2}{(5)(4)}$

$= \frac{127500 - 122500}{20}$

$= \frac{5000}{20} = 250$

2. $S = \sqrt{250}$

$= 15.81$

يوجد طريقتين لمقارنة ومعرفة البيانات الأكثر تشتتاً:
١. الطريقة المباشرة: إذا تحققت الشروط: - أن يكون الوسط الحسابي متساوياً.
- " " لها نفس الوحدات.
٢. استخدام معامل الاختلاف.

ملاحظة : ١. معامل الاختلاف من الممكن أن يكون سالب؛ لأنه ليس من مقاييس التشتت.
٢. القيمة الأبعد عن الصفر (0) هي الأكثر تشتتاً.

Coefficient of variation: The coefficient of variation, denoted by CVar, is the standard deviation divided by the mean. The result is expressed as a percentage.

For Samples: $CVar = \frac{s}{\bar{X}} \cdot 100$

For Populations: $CVar = \frac{\sigma}{\mu} \cdot 100$

Page: 132, 133, 142 / /

Exp 3-25

Exp 3-26

Q. Compare the variations. أيهما أكثر تشتت؟

A. $CVar_1 = 13.1\%$

$CVar_2 = -13.1\%$ } متساوية

B. $\overline{X}_A = 30$ yrs $S_A = 5$ yrs ⇐ الأكثر تشتتاً

$\overline{X}_B = 30$ yrs $S_B = 3$ yrs

مقاييس الموقع : تستخدم لتحديد موقع البيانات بين مجموعة البيانات.

Measures of Position

Standard Scores (Z Score): A z score or standard score for a value is obtained by subtracting the mean from the value and dividing the result by the standard deviation. The symbol for a standard score is z.

note: The z score represents the number of standard deviations that a data value falls above or below the mean.

For Samples: $z = \dfrac{X - \overline{X}}{S}$

For populations: $z = \dfrac{X - \mu}{\sigma}$

Exp 3-29

$X_C = 65$, $\overline{X}_C = 50$, $S_C = 10$

$X_H = 30$, $\overline{X}_H = 25$, $S_H = 5$

Solution:

$z_C = \dfrac{65 - 50}{10} = 1.5$ ⇐ 65 is 1.5 stdvs above the mean ($z > 0$).

$z_H = \dfrac{30 - 25}{5} = 1$

$z = 1.5$ — 5 { 65

10 { 50 The mean

$z = -0.5$ — 5 { 45

Page: 143, 147, 148, 149 / /

note: If the score is positive, the score is above the mean.
If the score is 0, the score is the same as the mean.
If the score is negative, the score is below the mean.

Exp 3-30

Percentiles: divide the data set into 100 equal groups.

$$P = \frac{(\text{number of value below } X) + 0.5}{n} \cdot 100$$

ملاحظة: ترتيب البيانات من الأصغر إلى الأكبر.

Exp 3-32
Exp 3-33

$$c = \frac{n \cdot p}{100}$$

① $c =$ عدد عشري
تقريب إلى أعلى X_c

② $c =$ عدد صحيح
$$\frac{X_c + (X_{c+1})}{2}$$

Exp 3-34
Exp 3-35

الربيعات
Quartiles: divide the distribution into four groups. separated by Q_1, Q_2, Q_3.

note: that Q_1 is the same as the 25th percentile, Q_2 is the same as the 50th percentile or the median, Q_3 corresponds to the 75th percentile.

Exp: 60 90 80 70 50

Solution:
n = odd
1. 50 60 70 80 90
2. $Q_2 = MD = 70$
3. $Q_1 = \frac{50+60}{2} = \frac{110}{2} = 55$
4. $Q_3 = \frac{80+90}{2} = 85$

Page: 150, 151, 152, 162 / /

Exp:
1 2 3 4 5 6
100 60 90 80 70 50

Solution:

n = even

1. 50 60 70 80 90 100
2. $Q_2 = MD = \frac{70+80}{2} = 75$
3. $Q_1 = 60$
4. $Q_3 = 90$

Procedure Table:

Exp 3-36

The Interquartile range (IQR): Is defined as the difference between Q_1 and Q_3 and is the range of the middle 50% of the data.

القيم المتطرفة

Outliers: Is an extremely high or an extremely low data value when compared with the rest of the data values.

Procedure Table:
1. Order then find Q_1, Q_3
2. $IQR = Q_3 - Q_1$
3. $(1.5)(IQR)$
4. $Q_1 - (1.5)(IQR)$

   $Q_3 + (1.5)(IQR)$

Undo Redo

القيم المتطرفة هي الأقل من

و الاكثر من

Exp 3-37

Solution:

n = even

1.
   1 2 3 4 5 6 7 8
   5 6 12 13 15 18 22 50

   $Q_1 = 9$ $Q_2 = 14$ $Q_3 = 20$
2. $IQR = 20 - 9 = 11$
3. $(1.5)(IQR) = (1.5)(11) = 16.5$
4. $Q_1 - (1.5)(IQR)$

   $= 9 - 16.5 = -7.5$

   $Q_3 + (1.5)(IQR)$

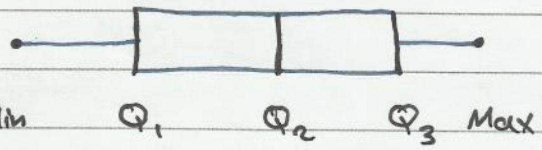
   $= 20 + 16.5 = 36.5$

A boxplot: can be used to graphically represent the data set. These plots involve five specific values:

Page: 162, 164

1. The lowest value of the data set (Minimum).
2. Q_1
3. The Median $= Q_2$
4. Q_3
5. The highest value of the data set (Maximum).

These values are called a five-number Summary of the data set.



Information Obtained from a Boxplot:

A. Using the Box:

1.A If the MD is near the center of the box, the distribution is approximatly Symmetric

2.A " " " falls to the left of the center of the box, the distribution is positively skewed.

3.A If the MD falls to the right of the center of the box, the distribution is negatively skewed.

B. Using the lines:

1.B If the lines are about the same lenght, the distribution is approximately Symmetric.

2.B If the right line is larger than the left line, the distribution is positively skewed.

3.B If the left line is larger than the right line, the distribution is negatively skewed.

Exp 3-39

Page: 534, 535, 536 / /

Correlation: Is a statistical method used to determine weather a linear relationship between variables exists.

Regression: Is a statistical method used to describe the nature of the relationship between variables, that is positive or negative, linear or nonlinear.

The purpose of this chapter is to answer these questions statistically:

Correlation:

1. Are two or more variables linearly related?
2. If so, what is the strength of the relationship?

Regression:

3. what type of relationship exists?
4. what kind of predictions can be made from the relationship?

Simple relationship: There are two variables — an independent variable (explanatory variable) (predictor variable), and a dependent variable (response variable)

Positive relationship: exist when both variables increase or decrease at the same time.

Negative relationship: as one variable increases, the other variable decreases, and vice versa.

10 - 1

Scatter plot: Is a graph of the ordered pairs (x, y) of numbers consisting of the independent variable x and the dependent variable y. The scatter plot is a visual way to describe the nature of the relationship between the independent and dependent variables.

Exp 10 - 1

Page: 537, 538, 539 / /

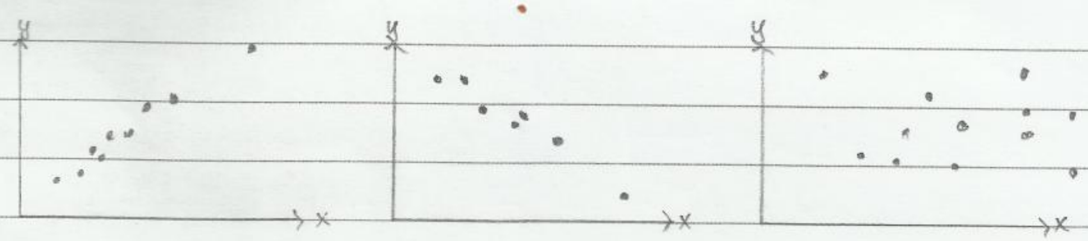
Figure 10-1

~~exp~~ Exp 10-2

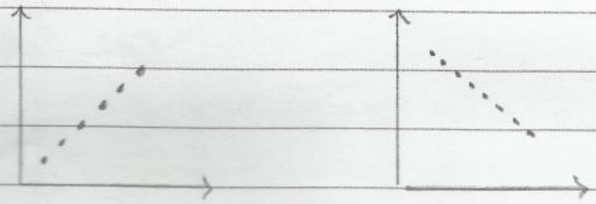
Figure 10-2

Exp 10-3

Figure 10-3



Positive linear relationship - Negative linear relationship - Nonlinear relationship



Complete Positive linear relationship - complet negative linear relationship

علاقة تامة: تامة = الكل في خط مستقيم بدون استثناء.

Correlation Coefficient: Computed from the sample data measures the strenght and direction of a linear relationship between two quantitative variables.

Pearson product moment correlation coefficient (PPMC)

imp: The range of the correlation coefficient is:

$$-1 \leq r \leq +1$$

0	No linear relationship
0.01 ↔ 0.49	weak " "
0.50 ↔ 0.99	Strong " "
1	Complete " "

Formula for the correlation coefficient r:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

Page: 539, 540, 541, 542 / /

Figure 10-5

Exp 10-4

Exp 10-5

Exp 10-6

Exp

$n = 3$

	x	y	xy	x^2	y^2
	1	2	2	1	4
	2	5	10	4	25
	3	6	18	9	36
total ⟹	6	13	30	14	65

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

$$= \frac{(3)(30) - (6)(13)}{\sqrt{[(3)(14) - (6)^2][(3)(65) - (13)^2]}} = \frac{90 - 78}{\sqrt{[42-36][195-169]}} = \frac{12}{\sqrt{[6][26]}}$$

$$= \frac{12}{\sqrt{156}} = \frac{12}{12.49} = 0.96$$

$y = a + bx$

ملاحظة هذه الجزئية من الحل تكون بعد قراءة صفحة ٥١ من الملخص

$$a = \frac{(13)(14) - (6)(30)}{(3)(14) - (6)^2} = \frac{182 - 180}{42 - 36} = \frac{2}{6} = \frac{1}{3}$$

$$b = \frac{(3)(30) - (6)(13)}{(3)(14) - (6)^2} = \frac{90 - 78}{42 - 36} = \frac{12}{6} = 2$$

$\therefore y = a + bx \Longrightarrow y = 0.33 + 2x$ Strong positive linear relationship

$x = 2.5 \Longrightarrow y = 0.33 + 2(2.5)$

$= 0.33 + 5 = 5.33$

Page: 551, 552, 553, 554, 555, 700 / /

If the value of the correlation coefficient is significant, the next step is to determine the equation of the regression line, which is the data's line of best fit.

Line of best fit: Best fit means that the ~~~~ Sum of the Squares of the vertical distances from each point to the line is at a minimum.

The equation of the regression line is written as $y = a + bx$

a = y intercept

b = Slope of the line

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Formulas for the Regression line of the $y = a + bx$ P: 553

Exp 10-9

Exp 10-10

note: The sign of the correlation coefficient and the sign of the slope of the regression line will always be the same.

Exp 10-11

The Spearman Rank correlation coefficient

Formula for Computing the ~~~~ Spearman Rank correlation coefficient:

d = ~~~~ difference ranks.

n = number of data pairs.

$$r_s = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$$

$$d = R_x - R_y$$

Exp 13-7

Solution: $r_s = 1 - \frac{(6)(12)}{(8)(63)}$

$= 1 - \frac{72}{504}$

$= 1 - 0.14$

$= 0.86$

Page: 700 / /

Exp: $n=5$

x	y	R_x	R_y	d	d^2
10	40	1	4	-3	9
20	30	2	3	-1	1
30	10	3	1	2	4
50	20	5	2	3	9
40	50	4	5	-1	1
				0	24

weak Negative linear relationship

Solution: $r_s = 1 - \frac{(6)(\sum d^2)}{n(n^2-1)}$

$= \frac{(6)(24)}{(5)(5^2-1)} = \frac{144}{120}$

$= 1 - 1.20$

$= -0.20$ WNLR

Page: 182, 183, 184, 185 / /

Probability: as a general concept can be defined as the chance of an event occurring.

Basic concepts:

Probability experiment: is a chance process that leads to well defined results called outcome.

محاولة

Outcome: is the result of a single trial of a probability experiment.

Sample space: is the set of all possible outcomes of a probability experiment. The symbol of sample space S.

$S = \{1, 2, 3, 4, 5, 6\}$ $nS = 6$ مثال: ١. مكعب النرد

$S = \{H, T\}$ $nS = 2$ ٢. قطعة نقدية

$S = \{HH, HT, TH, TT\}$ $nS = 4$ ٣. " " (رمي مرتين)

Exp 4-1

Exp 4-2 (الصورة) موجودة بالكتاب

Exp 4-3 $S = \{\underset{1}{BBB}, \underset{2}{BBG}, \underset{3}{BGB}, \underset{4}{BGG}, \underset{5}{GGG}, \underset{6}{GGB}, \underset{7}{GBG}, \underset{8}{GBB}\}$

$nS = 2^3 = 8$

عدد المحاولات

عدد النتائج في المحاولة الواحدة

Tree diagram: طريقة مطولة لكن يمكن إيجاد [illegible].

Exp 4-4

Event: consists of a set of outcomes of a probability experiment.

ملاحظة: الحادثة تساوي (نفس) أو جزء من فضاء العينة ولا يمكن أن تكون أكبر.

$S = \{1, 2, 3, 4, 5, 6\}$ $nS = 6$ مثال: مكعب النرد

$E_1 = \{2, 4, 6\}$ $\cancel{nS = 3}$ $n(E_1) = 3$ أن يكون: عدد زوجي

$E_2 = \{1, 3, 5\}$ $\cancel{nS = 3}$ $n(E_2) = 3$ " فردي

$E_3 = \{5, 6\}$ $\cancel{nS = 2}$ $n(E_3) = 2$ أكبر من 4

$E_4 = \{3\}$ $\cancel{nS = 1}$ $n(E_4) = 1$ عدد 3

Page: 185, 186, 187, 188 / /

أنواع الحادثة:

1. Simple event: event with one outcome.
2. Compound event: Consists of two or more outcomes or simple events

Classical probability: Uses sample space to determine the numerical probability that an event will happen. Classical probability assum that all outcomes in the Simple space are equally likely to occur.

Equally likely event: are events that have the same probability of occuring.

Formula for Classical probability:

$$P(E) = \frac{\text{Number of outcomes in } E}{\text{Total number of outcomes in the Sample space}} = \frac{n(E)}{n(S)}$$

مثال: تابع للمثال السابق (مكعب النرد):

$P(E_1) = \frac{3}{6} = \frac{1}{2}$

$P(E_2) = \frac{3}{6} = \frac{1}{2}$

$P(E_3) = \frac{2}{6} = \frac{1}{3}$

$P(E_4) = \frac{1}{6}$

Rounding Rule for Probability:

1. مثل: 0.667853 = 0.668 التقريب لأقرب 3 خانات
2. مثل: 0.0000079 = 0.000008 إذا كان العدد صغير جداً التقريب لأقرب رقم غير الصفر.

Exp 4-6

There are four basic probability rules:

1. Probability rule 1: The probability of any event E is a number (either a fraction or decimal) between and include 0 and 1. This is denoted by $0 \leq P(E) \leq 1$.

Page: 188, 189 / /

2. probability rule 2: If an event E cannot occur (i.e, the event contains no members in the sample space), it's probability is 0. (impossible). حادثة مستحيلة

مثال: مكعب النرد

$E = \{\ \} = \emptyset \quad n(E) = 0$ احتمال ظهور عدد 9

$P(E) = \frac{0}{6} = 0$

3. Probability rule 3: If an event E is certin. Then the probability of E is 1

Certin = 1

UnCertin $0 < P(E) < 1$

impossible = 0

مثال: مكعب النرد

$E = \{1, 2, 3, 4, 5, 6\} \quad n(E) = 6$

$P(E) = \frac{6}{6} = 1$

4. probability rule 4: The sum of the probabilities of all the outcomes in the sample space is 1.

مثال: مكعب النرد

$E = \{1, 2, 3, 4, 5, 6\}$

$\frac{1}{6} + \frac{1}{6} \ \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$

Complementary events: The Complementary of an event E is the set of outcomes in the sample space that are not included in the outcomes of event E. The complement of E is denoted by $\overline{E}$ (read "E bar").

مثال: مكعب النرد

ظهور عدد زوجي $A = \{2, 4, 6\}$

$\overline{A} = \{1, 3, 5\}$

$B = \{3, 6\}$

Exp 4-10 $\overline{B} = \{1, 2, 4, 5\}$

Page: 190, 191, 192, 193 199 / /

Rule for Compelementary Events

1. $P(\overline{E}) = 1 - P(E)$
2. $P(E) = 1 - P(\overline{E})$
3. $P(E) + P(\overline{E}) = 1 = S$

Exp 4-11

note: Probabilities can be represented pictorially by Venn diagrams.
Figure 4-4

Empirical Probability:

لا يمثل الشرط مثل Classical Probability

Formula for Empirical Probability:

$$P(E) = \frac{\text{frequency for the class}}{\text{total frequency in the distribution}} = \frac{f}{n}$$

هنا الجدول موجود في ⇐ 191

ملاحظة: جدول التوزيع التكراري من المعطاة

Exp 4-12

Method	Frequency
Drive	41
Fly	6
Train or bus	3
Σ	50

Solution: $P(E) = \frac{f}{n} = \frac{6}{50} = \frac{3}{25}$

Exp 4-13
Exp 4-14 "مهم جداً"

The addition rules for probability:

١. الحادثتان A و B شاملتين ومحصورتين، إذا كانا لا يحدثان في نفس الوقت (أي لا يوجد تقاطع بينهما ، أي لا يوجد عناصر مشتركة بينهما).

"mutually exclusive events"

Page: 200, 201

c. الحادثتان A و B غير شاملتين وغير محصورتين ، إذا كانا يحدثان في نفس الوقت (أي يوجد تقاطع بينهما ، أي يوجد عناصر مشتركة بينهما).
"not mutually exclusive events"

Exp 4-15

$S = \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 3, 5\}$ احتمال ظهور عدد فردي
$B = \{2, 4, 6\}$ " " " زوجي
mutually exclusive

$A = \{3\}$ ظهور عدد 3
$B = \{1, 2, 3\}$ احتمال ظهور عدد فردي
not mutually exclusive

$A = \{1, 3, 5\}$ احتمال ظهور عدد فردي
$B = \{1, 2, 3\}$ " " " أقل من 4
not mutually exclusive

$A = \{1, 2, 3\}$ احتمال ظهور عدد أقل من 4
$B = \{5, 6\}$ " " " أكبر من 4

Exp 4-16

The probability of two or more events can be determined by the addition rules.

1. Addition Rule 1: When two events A and B are mutually exclusive, the probability that A or B will occur is:

$$P(A \text{ or } B) = P(A) + P(B)$$

Exp 4-17
Exp 4-18
Exp 4-19