

Topic Y3 Correlation regression & chi-squared tests (Post-TT A) [48]

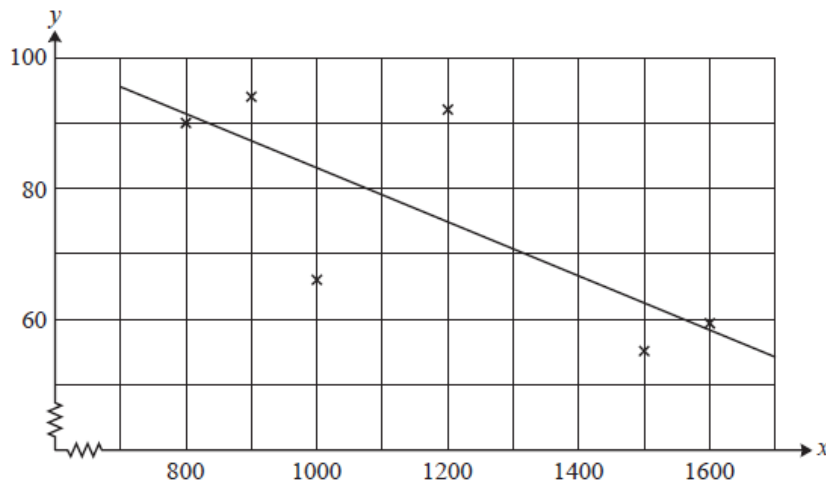
1.

The Gross Domestic Product per Capita (GDP), x dollars, and the Infant Mortality Rate per thousand (IMR), y , of 6 African countries were recorded and summarised as follows.

$$n = 6 \quad \Sigma x = 7000 \quad \Sigma x^2 = 8\,700\,000 \quad \Sigma y = 456 \quad \Sigma y^2 = 36\,262 \quad \Sigma xy = 509\,900$$

(i) Calculate the equation of the regression line of y on x for these 6 countries. [4]

The original data were plotted on a scatter diagram and the regression line of y on x was drawn, as shown below.



(ii) The GDP for another country, Tanzania, is 1300 dollars. Use the regression line in the diagram to estimate the IMR of Tanzania. [1]

(iii) The GDP for Nigeria is 2400 dollars. Give two reasons why the regression line is unlikely to give a reliable estimate for the IMR for Nigeria. [2]

(iv) The actual value of the IMR for Tanzania is 96. The data for Tanzania ($x = 1300, y = 96$) is now included with the original 6 countries. Calculate the value of the product moment correlation coefficient, r , for all 7 countries. [4]

(v) The IMR is now redefined as the infant mortality rate per hundred instead of per thousand, and the value of r is recalculated for all 7 countries. Without calculation state what effect, if any, this would have on the value of r found in part (iv). [1]

2.

A random sample of 80 students who had all studied Biology, Chemistry and Art at a college was each asked which they enjoyed most. The results, classified according to gender, are given in the table.

		Subject		
		Biology	Chemistry	Art
Gender	Male	13	4	11
	Female	37	8	7

It is required to carry out a test of independence between subject most enjoyed and gender at the $2\frac{1}{2}\%$ significance level.

(i) Calculate the expected values for the cells. [3]

(ii) Explain why it is necessary to combine cells, and choose a suitable combination. [2]

(iii) Carry out the test. [8]

3.

Three tutors each marked the coursework of five students. The marks are given in the table.

Student	A	B	C	D	E
Tutor 1	73	67	60	48	39
Tutor 2	62	50	61	76	65
Tutor 3	42	50	63	54	71

(i) Calculate Spearman's rank correlation coefficient, r_s , between the marks for tutors 1 and 2. [5]

(ii) The values of r_s for the other pairs of tutors, are as follows.

$$\text{Tutors 1 and 3: } r_s = -0.9$$

$$\text{Tutors 2 and 3: } r_s = 0.3$$

State which two tutors differ most widely in their judgements. Give your reason. [2]

4.

Desmond believes that a particular six-sided die is biased. He decides to count the number of rolls required until he first rolls a 5 or a 6. He carries out this experiment 100 times. His results are summarised in the table below

Number of rolls to 1st 5 or 6	1	2	3	4	5	6	7 or more
Observed frequency	39	17	17	10	3	3	11

(a) Stating your hypotheses clearly, test, at the 5% significance level, whether or not these results provide evidence that the die is biased. You should show your working clearly, including the expected frequencies and the critical value used.

(10)

Mai suggests recording the results of rolling the die 100 times and using a goodness of fit test for a discrete uniform distribution.

(b) (i) Give 2 reasons why Mai might think that her test is better than Desmond's test.

(ii) Give a reason why Desmond might think that his test is better than Mai's test.

(3)

5.

In a survey, a large number of people were asked whether or not they recycle plastics. In each case the distance, in metres, from their home to the nearest plastic recycling point was measured and recorded. The results are shown in Table 1

Observed frequencies		Recycle plastics	
		Yes	No
Distance to nearest plastic recycling point	Less than 500 m	64	42
	500–1000 m	32	22
	More than 1000 m	14	26

Table 1

The Council's Environment Officer, Barbara, believes that whether or not people recycle plastics is independent of the distance to the nearest recycling point.

Barbara decides to test if the data from the survey supports her belief. Her expected frequencies are shown in Table 2

Expected frequencies		Recycle plastics	
		Yes	No
Distance to nearest plastic recycling point	Less than 500 m	58.3	47.7
	500–1000 m	29.7	24.3
	More than 1000 m	22	18

Table 2

(a) Carry out a hypothesis test, at the 5% significance level, to see if the data from the survey support Barbara's belief. State your hypotheses and the critical value used in this test.

(5)

Barbara explains her results to her assistant, Bill. The data were collected by Bill, who now confesses to Barbara that he actually only collected data from 100 people and then doubled all his results.

(b) With reference to the test statistic and critical value, explain whether or not Barbara's conclusions are still valid.

(2)