

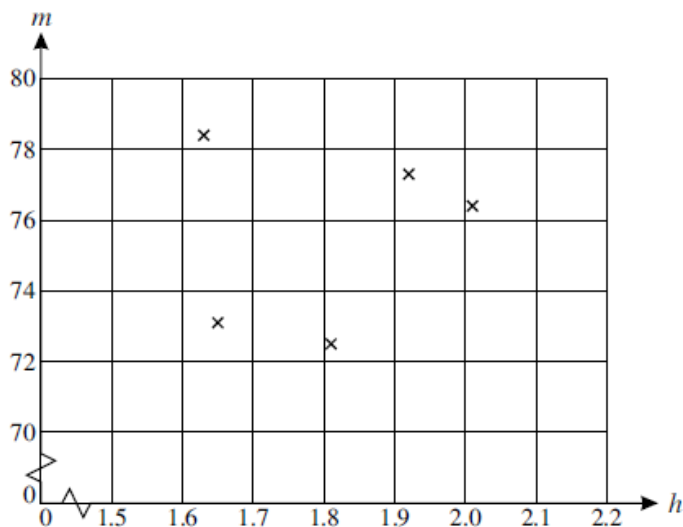
Topic Y3 Correlation regression & chi-squared tests (Pre-TT A) [48]

1.

In a study of the inheritance of skin colouration in corn snakes, a researcher found 865 snakes with black and orange bodies, 320 snakes with black bodies, 335 snakes with orange bodies and 112 snakes with bodies of other colours. Theory predicts that snakes of these colours should occur in the ratios 9:3:3:1. Test, at the 5% significance level, whether these experimental results are compatible with theory. [6]

2.

The heights, h m, and weights, m kg, of five men were measured. The results are plotted on the diagram.



The results are summarised as follows.

$$n = 5 \quad \Sigma h = 9.02 \quad \Sigma m = 377.7 \quad \Sigma h^2 = 16.382 \quad \Sigma m^2 = 28\,558.67 \quad \Sigma hm = 681.612$$

- (i) Use the summarised data to calculate the value of the product moment correlation coefficient, r . [3]
- (ii) Comment on your value of r in relation to the diagram. [2]
- (iii) It was decided to re-calculate the value of r after converting the heights to feet and the masses to pounds. State what effect, if any, this will have on the value of r . [1]
- (iv) One of the men had height 1.63 m and mass 78.4 kg. The data for this man were removed and the value of r was re-calculated using the original data for the remaining four men. State in general terms what effect, if any, this will have on the value of r . [1]

3.

Students at a science department of a university are offered the opportunity to study an optional language module, either German or Mandarin, during their second year of study.

From a sample of 50 students who opted to study a language module, 31 were female.

Of those who opted to study Mandarin, 8 were female and 12 were male.

Test, using the 5% level of significance, whether choice of language is independent of gender.

The sample of students may be regarded as random.

[8 marks]

4.

In a newspaper article a journalist claimed that in football matches, teams that have more shots at goal also score more goals.

A football fan wanted to see if the journalist's claim was correct, so collected data for a random sample of 10 teams from the previous season. She collected, for each team, the number of shots at goal and the number of goals scored that season.

Her results are given in the table below.

Team	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>
Number of shots at goal	902	1017	932	945	996	947	943	984	997	917
Number of goals scored	118	112	119	120	121	83	106	89	122	94

- (a) Calculate the value of Spearman's rank correlation coefficient between the number of shots at goal and the number of goals scored. Show your working clearly. (5)
- (b) Stating your hypotheses clearly, test whether the value that you calculated in part (a) supports the journalist's claim. Use a 5% level of significance. (4)

The football fan discovers that the number of shots at goal for team *A* is in fact 802 not 902

- (c) Without further calculations, state how this effects the answer to part (a). (1)

(Total 8 marks)

5.

A chemical solution was gradually heated. At five-minute intervals the time, x minutes, and the temperature, y °C, were noted.

x	0	5	10	15	20	25	30	35
y	0.8	3.0	6.8	10.9	15.6	19.6	23.4	26.7

$$[n = 8, \Sigma x = 140, \Sigma y = 106.8, \Sigma x^2 = 3500, \Sigma y^2 = 2062.66, \Sigma xy = 2685.0.]$$

- (i) Calculate the equation of the regression line of y on x . [4]
- (ii) Use your equation to estimate the temperature after 12 minutes. [2]
- (iii) It is given that the value of the product moment correlation coefficient is close to +1. Comment on the reliability of using your equation to estimate y when
- (a) $x = 17$,
- (b) $x = 57$.

[2]

(Total 8 marks)

6.

A dairy industry researcher, Robyn, decided to investigate the milk yield, classified as low, medium or high, obtained from four different breeds of cow, A, B, C and D.

The milk yield of a sample of 105 cows was monitored and the results are summarised in contingency **Table 1**.

		Yield			
		Low	Medium	High	Total
Breed	A	4	5	12	21
	B	10	6	4	20
	C	8	17	7	32
	D	5	20	7	32
Total		27	48	30	105

The sample of cows may be regarded as random.

Robyn decides to carry out a χ^2 -test for association between milk yield and breed using the information given in **Table 1**.

- (a) Contingency **Table 2** gives some of the expected frequencies for this test.

Complete **Table 2** with the missing expected values.

[2 marks]

		Yield		
		Low	Medium	High
Breed	A			6
	B	5.14	9.14	5.71
	C			
	D	8.23	14.63	9.14

- (b) (i) For Robyn's test, the test statistic $\sum \frac{(O - E)^2}{E} = 19.4$ correct to three significant figures.

Use this information to carry out Robyn's test, using the 1% level of significance.

[5 marks]

- (b) (ii) By considering the observed frequencies given in **Table 1** with the expected frequencies in **Table 2**, interpret, in context, the association, if any, between milk yield and breed.

[2 marks]