Topic Y3 Correlation regression & chi-squared tests (Post-TT B) [42]

1.

Gloria is a market trader who sells jeans. She trades on Mondays, Wednesdays and Fridays. Wishing to investigate whether the volume of trade depends on the day of the week, Gloria analysed a random sample of 150 days' sales and classified them by day and volume (low, medium and high). The results are given in the table below.

		Day		
		Monday	Wednesday	Friday
Volume	Low	15	13	2
	Medium	23	26	23
	High	12	9	27

Gloria asked a statistician to perform a suitable test of independence and, as part of this test, expected frequencies were calculated. These are shown in the table below.

		Day		
		Monday	Wednesday	Friday
Volume	Low	10.00	9.60	10.40
	Medium	24.00	23.04	24.96
	High	16.00	15.36	16.64

(i) Show how the value 23.04 for medium volume on Wednesday has been obtained. [2]

(ii) State, giving a reason, if it is necessary to combine any rows or columns in order to carry out the test. [1]

The value of the test statistic is found to be 21.15, correct to 2 decimal places.

(iii) Stating suitable hypotheses for the test, give its conclusion using a 1% significance level. [4]

Gloria wishes to hold a sale and asks the statistician to advise her on which day to hold it in order to sell as much as possible.

(iv) State the day that the statistician should advise and give a reason for the choice. [2]

(Total 9 marks)

2.

For each of the last five years the number of tourists, x thousands, visiting Sackton, and the average weekly sales, £ y thousands, in Sackton Stores were noted. The table shows the results.

Year	2007	2008	2009	2010	2011
x	250	270	264	290	292
y	4.2	3.7	3.2	3.5	3.0

(i) Calculate the product moment correlation coefficient r between x and y. **[4]**

(ii) It is required to estimate the average weekly sales at Sackton Stores in a year when the number of tourists is 280 000. Calculate the equation of an appropriate regression line, and use it to find this estimate. **[4]**

(iii) Over a longer period the value of r is –0.8. The mayor says, "This shows that having more tourists causes sales at Sackton Stores to decrease." Give a reason why this statement is not correct. **[1]**

(Total 9 marks)

3.

In a test of association of two factors, A and B, a 2×2 contingency table yielded 5.63 for the value of χ^2 with Yates' correction.

(i) State the null hypothesis and alternative hypothesis for the test. **[1]**

(ii) State how Yates' correction is applied, and whether it increases or decreases the value of χ^2. **[2]**

(iii) Carry out the test at the $2\frac{1}{2}\%$ significance level. **[3]**

(Total 6 marks)

4.

Two judges placed 7 dancers in rank order. Both judges placed dancers A and B in the first two places, but in opposite orders. The judges agreed about the ranks for all the other 5 dancers. Calculate the value of Spearman's rank correlation coefficient. **[4]**

(Total 4 marks)

5.

A soft drinks factory produces lemonade which is sold in packs of 6 bottles. As part of the factory's quality control, random samples of 75 packs are examined at regular intervals. The number of underfilled bottles in a pack of 6 bottles is denoted by the random variable X. The results of one quality control check are shown in the following table.

Number of underfilled bottles	0	1	2	3
Number of packs	44	20	8	3

A researcher assumes that $X \sim \mathrm{B}(3, p)$.

(i) By finding the sample mean, show that an estimate of p is 0.2. **[3]**

(ii) Show that, at the 5% significance level, there is evidence that this binomial distribution does not fit the data. **[10]**

(iii) Another researcher suggests that the goodness of fit test should be for $\mathrm{B}(6, p)$. She finds that the corresponding value of χ^2 is 2.74, correct to 3 significant figures. Given that the number of degrees of freedom is the same as in part **(ii)**, state the conclusion of the test at the same significance level. **[1]**

(Total 14 marks)