ASSIGNMENT A15.3

HYPOTHESIS TESTING (INDEPENDENCE & GOODNESS OF FIT)

1   For a long time, experts have been trying to explain the complex relations and interactions between leaders and other members in an organisation. A study was carried out to investigate the relevance of a leader's gender in adopting a specific leadership style. A survey was given to a random sample of 79 people holding leadership positions from various organisations and institutions. Their responses to the questionnaire served to identify their dominant leadership style. The results are given in the following contingency table.

	Dominant leadership style		
	Authoritarian	Democratic	Laissez-faire
Male	12	22	9
Female	20	13	3

Carry out a test for the independence of the two factors, gender and the dominant leadership style.

Discuss what the test indicates about the association, if any, between the two factors. You should refer to the *p*-value for your test and t the largest two contributions to the test statistic.   [6]

[EJC2019/Prelim/II/6(modified)]

2 **(a)** After the implementation of the Electronic Road Pricing Scheme, a survey was conducted where the number of vehicles passing the gantry point at Pan-Island Expressway in each period of 20 seconds was recorded. The results for 100 periods are as follows:

Number of vehicles in the period	0	1	2	3	4	5
Frequency	24	36	28	8	3	1

(i) Calculate the mean number of vehicles per period. Perform a goodness of fit test, at the 5% level of significance to determine whether the data could be a sample from a Poisson distribution. [6]

(ii) How would the above test change if it were specified as a Po(1.33) distribution instead? [1]

(b) A survey of 100 families, known to be regular television viewers, was undertaken. They were asked which of the two channels they watched most during an average week. A summary of their replies is given in the following table, together with the region in which they lived.

Region	North	South	East	West
Channel 5	10	17	$10+a$	$23-a$
Channel 8	5	8	$20-a$	$7+a$

To test the hypothesis that there is no association between the channel watched most and the region, show that the χ^2 statistic in terms of a can be simplified as

$$\chi^2 = \frac{10a^2 - 130a + 479}{36}.$$
[3]

Find the set of values of a that would result in the assumption not being rejected at the 5% level of significance. [3]

[JPJC2019/Prelims/II/10(modified)]

3 The proportions of blood types A, B, AB and O in the population of a country are p_1, p_2, p_3 and p_4 respectively, where $\sum_{i=1}^{4} p_i = 1$. In order to test whether the population of a city in the country conforms to these figures, a random sample of size n is selected and the numbers of people with blood types A, B, AB and O are found to be a, b, c and d respectively.

(i) Show that the χ^2 statistic for a goodness of fit test simplifies to

$$\chi^2 = \frac{a^2}{np_1} + \frac{b^2}{np_2} + \frac{c^2}{np_3} + \frac{d^2}{np_4} - n.$$ [3]

(ii) It is given that $p_1 = p_4$, $p_2 = 3p_3$ and the values of a, b c and d are 26, 19, 10 and 45 respectively. Denoting the common value of p_1 and p_4 by p, show that

$$\chi^2 = \frac{2701}{100p} + \frac{661}{75(1-2p)} - 100.$$ [3]

Hence, find the value of p_0 of p for which this value of χ^2 is a minimum. [1]

(iii) Carry out the goodness of fit test at the 10% significance level, with $p = p_0$. [4]

(iv) State, giving your reason, the conclusion of your test for values of p other than p_0. [1]

(v) Using the values of a, b, c and d in **(ii)**, construct a 95% confidence interval for the proportion of people in the country with blood type A. [2]

(vi) Discuss one advantage and one disadvantage of finding a 95% confidence interval instead of a 99% confidence interval. [2]

[TJC2019/Prelim/II/10]