ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2017

MODULE 3 : Basic statistical methods

Time allowed: One and a half hours

*Candidates should answer **THREE** questions.*

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

The notation log *denotes logarithm to base **e**.*
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* $^{n}C_{r}$.

This examination paper consists of 4 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1.   A new exercise programme to help people lose weight has been devised.   The programme is considered successful if the average weight loss in a month is about 4 kg.   A sample of 90 volunteers undertakes the programme.   Using their data, the programme managers want to test whether weight loss follows a Normal distribution with mean 4 kg and standard deviation 2.5 kg.   They calculate the expected frequencies for 8 weight loss categories.   These are shown below, together with the observed number of volunteers in each category.

Weight loss (kg)	Observed frequency	Expected frequency
less than 1	9	10.36
between 1 and 2	10	8.71
between 2 and 3	13	11.94
between 3 and 4	18	13.99
between 4 and 5	16	13.99
between 5 and 6	9	11.94
between 6 and 7	9	8.71
more than 7	6	10.36

(i)   Explain how the figure 8.71 for the expected frequency for the category 'between 1 and 2' was calculated, showing all your working.

(4)

(ii)   Use a χ^2 test at the 5% significance level to investigate the null hypothesis that weight loss has a $N(4, 2.5^2)$ distribution.   State your conclusions clearly and concisely.

(6)

(iii)   Using only the observed frequencies of the eight weight loss categories given in the table, briefly describe how you would estimate the mean and variance of the weight loss, and what problems you might encounter in doing so.   Without performing the calculations, describe how you would then test for Normality of the data, and which χ^2 distribution you would use.

(4)

(iv)   An independent expert believes that the mean weight loss for people following the programme will be less than the 4 kg claimed by the programme managers. The sample mean and sample standard deviation of the 90 weight losses are estimated to be 3.54 kg and 2.1 kg respectively.   Test at the 5% significance level whether the expert's opinion is upheld.

(6)

2.   In an investigation into the effect of a high fibre diet on vitamin D deficiency, we are told that 8 patients are given a high fibre diet and 8 others are given a diet with a normal amount of fibre for a week. The amount of a form of vitamin D in the blood is recorded in suitable units for each individual at the end of the trial. Summary values are given below.

	High Fibre	Normal Fibre
Mean	220.83	214.68
Standard Deviation	8.59	7.74

(i)   Assume the two samples are drawn at random from Normally distributed populations. Test, at the 5% significance level, whether the variability in the two underlying populations could be equal.

(3)

(ii)   Recent research suggests that a high fibre diet increases the amount of vitamin D in the blood. Perform a suitable test at the 5% significance level to investigate whether the data are consistent with the research, stating your hypotheses and conclusions clearly.

(8)

You now discover that the same 8 patients were used for both diets. Half of them, chosen at random, had the high fibre diet first whilst the rest had the normal fibre diet first. The resulting data are given in the table below.

Patient	High Fibre	Normal Fibre
1	215.8	215.7
2	215.9	206.5
3	219.5	213.7
4	222.5	213.7
5	231.6	227.2
6	235.6	224.6
7	212.2	209.5
8	213.5	206.5

(iii)   Perform a more suitable test, at the 5% significance level, to find whether the high fibre diet produces an increase in vitamin D in the blood. Briefly discuss the reasons for any differences from the results of the previous test.

(9)

3.   A politician is interested in the opinions of her constituents about a new government social care policy. A random sample of 500 constituents who had voted in the last election was selected, and asked whether or not they are in favour of the new policy and which party they voted for in the last election.

	Party voted for at last election		
	Conservative	Labour	Other
In favour of policy	118	73	47
Not in favour of policy	179	57	26

(i)   Test whether there is an association between the party that constituents voted for at the last election and their attitude to the new policy. Use a 5% significance level for your test, state your null and alternative hypotheses clearly and report your conclusions.   (9)

(ii)   The politician is really only interested in whether or not constituents voted Labour in the last election in relation to their attitude to the new policy. Produce a table that summarises the party voted for at the last election as either 'Labour' or 'not Labour'. Without calculating the result, state how you think the result of the test in part (i) would change if this table was used instead, giving a reason for your answer.   (3)

(iii)   Using the table that you produced in part (ii), calculate the proportion of those who voted for Labour who were in favour of the new policy and similarly for those who did not vote for Labour. Test if these proportions differ significantly using a 5% significance level, and relate the outcome of your test to your comments in part (ii).   (8)

4.   Prospective customers applying to open a savings account with a finance company are required to complete an online application form. A manager is interested in the time taken to fill in the form, so he asks fifteen applicants chosen at random to record their completion times. The observed times are in minutes.

5.4	15.0	11.4	34.8	4.2	22.2	27.6	4.8
8.4	16.8	36.6	9.6	13.2	3.6	25.8	

The sample mean and sample deviation of these values are 15.96 and 11.04 minutes respectively, and the times may be regarded as independent observations.

(i)   Calculate a 95% confidence interval for the mean completion time. State clearly how this interval should be interpreted and the assumption made about the data in order for it to be valid.   (9)

(ii)   In a longer run of observations, the manager notices that the data values do not appear to be symmetrically distributed. He would like to test the hypothesis that the median completion time is 12 minutes. State two nonparametric tests that could be used and perform the more powerful of these tests. State your null hypothesis and conclusions clearly.   (11)

4