

EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2016

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 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. (i) Define the following terms that are used in hypothesis testing, explaining how they relate to Type I and Type II errors.
- (a) The *significance level* of the test. (2)
- (b) The *power* of the test. (2)
- (ii) A soft drinks machine at a fast food outlet is regulated so that the amount of drink dispensed is Normally distributed with standard deviation 15 ml. It is intended that the mean amount dispensed should be 200 ml. To test whether the machine is working correctly, it is checked periodically by taking a random sample of 9 drinks and calculating the mean amount dispensed. The machine is considered to be working satisfactorily if this mean lies between 191 ml and 209 ml.
- (a) Write down appropriate null and alternative hypotheses for this testing procedure. (2)
- (b) State the distribution of the sample mean \bar{X} if the null hypothesis is true. (3)
- (c) Calculate the significance level of the test. If you were performing the test at the 5% level of significance, what would you conclude? (5)
- (d) Calculate the power of the test if the true mean amount dispensed is 205 ml. Without performing any further calculations, state how you would expect this value to compare with the power of the test if the true mean amount dispensed is 210 ml, justifying your answer. (6)

2. A commercial vegetable grower sends cabbage seeds to his customers by post in packs containing 10 seeds. In previous years 12% of the seeds delivered fail to grow. This year he is experimenting with a new variety of cabbage, Stalwart, whose seeds may have a different failure rate. A random sample of 60 customers who each received packs of 10 Stalwart seeds gave the information shown in the table below. The vegetable grower wants to examine whether the Stalwart seeds have a different failure rate from his previous variety.

Number of failed Stalwart seeds	0	1	2	3	>3
Frequency	25	28	5	1	1

- (i) State a suitable model for the number of Stalwart seeds in a pack of 10 that fail to grow if the failure rate is the same as for the grower's previous variety. State two assumptions that are required for this model to be appropriate; outline briefly why, in practice, these assumptions may not be valid. (7)
- (ii) You may assume that, if Stalwart has the same failure rate as the previous variety, the expected numbers of the 60 packs with 0, 1, 2, 3 or more than 3 failed seeds are 16.71, 22.79, 13.98, 5.08 and 1.44 respectively. Test whether the expected numbers are consistent with the observed frequencies. Use a 5% significance level for your test and state your hypotheses and conclusions clearly. (11)
- (iii) Briefly discuss the results of your test and the implications for the vegetable grower. (2)

3. When the property market is uncertain, many homeowners choose to improve their homes instead of selling them. The costs of remodelling the kitchens in a random sample of 11 houses are shown below, together with the increases in the valuations of the houses after the remodelling. All values are in thousands of pounds.

<i>House</i>	<i>Kitchen remodelling cost (£000s)</i>	<i>Increase in valuation (£000s)</i>
A	3.4	4.75
B	6.2	7.75
C	2.8	3.5
D	7.8	8.5
E	4.2	5.0
F	3.6	5.5
G	9.2	10.25
H	8.4	9.5
I	5.8	6.5
J	7.1	8.0
K	6.6	7.0

- (i) Calculate a 95% confidence interval for the mean remodelling cost. What distributional assumption have you made? (7)
- (ii) A property advice website states that the increase in valuation will exceed the remodelling cost by more than £1000 on average. Test, at the 5% significance level, the null hypothesis $\mu = 1$ against the alternative hypothesis $\mu > 1$, where μ is the population mean, in units of thousands of pounds, of increase in valuation minus remodelling cost. Justify your choice of test, and state your conclusions clearly. What distributional assumption is required for the validity of your test? (9)
- (iii) Name two tests which could be applied to the data if the assumption in part (ii) is not valid. Give an advantage of each test over the other, but do not perform the tests. (4)

4. A candle manufacturer wants to state an average burning time on the packaging of his perfumed candles. He records the burning times for a random sample of 15 candles and enters the data into a computer package, which tells him that a 90% confidence interval for the mean burning time is from 48.388 hours to 56.012 hours.

(i) What are the mean burning time and its standard error for the sample of 15 candles? (4)

(ii) The manufacturer would like to quote a 95% confidence interval for the mean burning time of his perfumed candles. Calculate this interval and briefly compare it with the previous interval, stating the assumption made about the burning times for both to be valid. (3)

The manufacturer wants to compare the burning times of his identically sized perfumed and unperfumed candles. In particular he suspects that the oils used in his perfumed candles may result in their burning more quickly on average. He lights a random sample of 19 unperfumed candles and finds that their mean burning time is 58.4 hours and the standard deviation of their burning times is 10.2 hours.

(iii) Find the standard deviation of the burning times of the perfumed candles and test whether the variances of the burning times of the two types of candles are equal. Use a 5% significance level for your test. (5)

(iv) Perform a test at the 5% significance level to investigate his suspicions about the mean burning times, stating your hypotheses and conclusions clearly. (8)

BLANK PAGE

BLANK PAGE

BLANK PAGE