ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

## HIGHER CERTIFICATE IN STATISTICS, 2015

## MODULE 6 : Further applications of statistics

**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* ${}^nC_r$ .



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.

© RSS 2015

1. A chemical mixing process is used to prepare a certain liquid, and it is thought that the quality of the output liquid may be related to the rotational speed of the mechanical stirrers used for mixing. An experiment is run where the liquid is produced at a wide variety of mixing speeds, and the quality is measured in a standardised way. Polynomial regression models are then used to assess if and how quality depends on mixing speed.

(i) Write down an appropriate fourth order polynomial regression model for the results of this experiment, stating your assumptions about the term that represents random variation.

(3)

(ii) The residual sums of squares for polynomial regression models up to the fourth order are given below.

| Predictors | Residual SS | Residual d.f. |
|---|---|---|
| $x, x^2, x^3, x^4$ | 231.3 | 54 |
| $x, x^2, x^3$ | 246.6 | 55 |
| $x, x^2$ | 288.7 | 56 |
| $x$ | 412.4 | 57 |
| constant only | 615.1 | 58 |

Use backward elimination, testing at the 5% level, to select the best polynomial regression model for these results.

(6)

(iii) Define the *residuals* from a fitted multiple regression model, and state the properties they should possess if the assumptions made by the model are correct.

(2)

(iv) Explain how various types of residual plot can be used as alternative and complementary ways of assessing the appropriateness of the assumptions made by a polynomial regression model. Compare and contrast the behaviour of residual plots and backward elimination in the case of a third order polynomial regression model being fitted when a second order model is most appropriate, and in the case where a second order model is fitted but a third order model is most appropriate.

(9)

2. A researcher asks you for advice in designing an agricultural field trial to assess the effects on a particular crop of a new 'environmentally friendly' pesticide which should affect insects but not birds or mammals. He wishes to see how well the pesticide performs at low, medium and high concentrations in comparison to the standard pesticide at its standard concentration, and also in comparison to no pesticide at all.

He also wants to assess if there is any difference between five different non-lethal methods (nets, fences, etc) of keeping birds and mammals away from the crop.

The researcher has 25 plots of land available for the trial.

(i) Explain to the researcher how this trial could be designed in order to obtain as much useful information as possible about the performance of both this pesticide and the different non-lethal methods.

Concentrate on describing the design of the trial and the reasoning behind it, rather than the analysis, mentioning the analysis only if necessary to explain the design.

Assume that the researcher has little knowledge of statistics or experimental design and define any terms you use, and in particular explain the following terms and how they are important in this case.

(a) Randomisation.

(3)

(b) Factorial design.

(4)

(c) Control group.

(4)

(ii) A few days after your initial meeting the researcher contacts you to say that most of the suppliers of the non-lethal methods have failed to deliver their products, so that now he only wants to test the pesticide. However, on studying the results of similar trials, he is now worried that in recent years some of the plots of land have been better than others for this sort of crop. Explain the concept of *blocking*, why it is important and useful, and how it can be used to improve this trial.

(5)

(iii) After another few days the researcher contacts you again to say that in fact the suppliers have now all delivered their non-lethal methods, so he now wants to test these after all, as well as using the blocks in part (ii). Briefly describe a Latin square design and explain how a design of this type could be used in this trial.

(4)

3. A bank has recorded the number $y$ of businesses that it has driven to bankruptcy by the recalling of loans during a single financial year, classified into five geographical regions (1–5) and four types of business (A–D).

| Type | Region 1 | 2 | 3 | 4 | 5 | Sum |
|---|---|---|---|---|---|---|
| A | 58 | 56 | 51 | 55 | 54 | 274 |
| B | 45 | 39 | 53 | 54 | 50 | 241 |
| C | 45 | 35 | 49 | 37 | 38 | 204 |
| D | 48 | 36 | 41 | 51 | 50 | 226 |
| Sum | 196 | 166 | 194 | 197 | 192 | 945 |

$\Sigma y^2 = 45\,659$

(i) Write down an appropriate linear model for these data, relating $y$ to region and type of business, stating your assumptions about the term that represents random variation. Comment on how appropriate your assumptions are in this case.

(4)

(ii) Construct the two-way analysis of variance table for these data under your model.

(4)

(iii) State appropriate hypotheses and hence use the analysis of variance table to test at the 5% level of significance whether there is any evidence of a difference in mean number of bankruptcies between the types of business and whether there is any evidence of a difference in mean number of bankruptcies between the regions. Carefully state your conclusions.

(6)

(iv) Managers are most interested in seeing how easy it is to force different types of business in different regions into bankruptcy. Comment on what this data set and this analysis show, and do not show, with respect to this question. What other information would be useful?

(3)

(v) A manager comments that businesses of the same type often experience very different business conditions in the different regions. Explain the concept of *interaction* and hence explain if and how it could be modelled in this case, through a different model and/or more data.

(3)

4. (a) A factory makes components that are required to be l00 mm long. Explain how a *cusum chart* is defined and works in general, what is plotted and why it is useful, stating how it would look both for processes which meet the requirements and for those which do not. Hence show how a cusum chart can be used to assess how well this production process is working. Critically comment on the advantages and disadvantages of cusum charts, such as what sort of deviations from the requirement they are good or bad at detecting.

(8)

(b) A company produces items in large batches and tests each batch to assess whether the proportion of faulty items in each batch is tolerably low, where each item is classed as either acceptable or faulty. The company is particularly worried about the risk of rejecting a batch when the true proportion of faulty items is 5% or lower.

The company is considering two different approaches for batch testing.

Scheme A: take a sample of 30 items from the batch and reject the batch if three or more items are faulty, otherwise accept it.

Scheme B: take a sample of 20 items from the batch, accept the batch immediately if there are no faulty items and reject the batch immediately if four or more items are faulty. Otherwise take another sample of 20 items and reject the batch if a total of four or more items (from the two samples combined) are faulty, otherwise accept it.

(i) Find the probability of accepting a batch for each of scheme A and scheme B in terms of $p$, the proportion of faulty items in a batch. State any assumptions you make and comment on their appropriateness.

(7)

(ii) Find the probability of rejecting a batch when $p = 0.05$ under each of schemes A and B. Comment on what the company can and cannot conclude from these probabilities.

(3)

(iii) Comment on any other quantities you would need to take into account in order to decide which scheme is preferable for the company.

(2)

BLANK PAGE

BLANK PAGE

BLANK PAGE