EXAMINATIONS OF THE HONG KONG STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2013

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* ${}^{n}C_{r}$.

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. Three new night-sights are being assessed, and an experiment was conducted into the distance (in metres) at which observers using them could detect a man walking towards them. In standardised conditions, 5 observers each made one observation with each sight, with the following results.

	Observer					
Sight	*1*	*2*	*3*	*4*	*5*	*Sum*
A	28	34	30	33	19	144
B	23	28	14	21	20	106
C	29	31	23	32	17	132
Sum	80	93	67	86	56	382

$$\Sigma y^2 = 10\,284$$

(i) Explain the concept of *blocking*, using this experiment as an example. Comment on the importance or otherwise of blocking in this design.

(3)

(ii) Write down an appropriate linear model for these data, clearly stating any assumptions made.

(3)

(iii) Construct the analysis of variance table for these results. State appropriate hypotheses and test at the 5% level whether there is any evidence of a difference in mean detection distance between the sights. Clearly state your conclusions.

Stating appropriate hypotheses, use the analysis of variance table to test at the 5% level whether there is any evidence of a difference in mean detection distance between the observers. Explain how this relates to your answer to part (i).

(12)

(iv) Comment on any potential problems with this experiment and on how it might be improved.

(2)

2. (a) The multiple linear regression model

$$Y = \alpha + \beta x + \gamma z + \varepsilon$$

can be fitted to n data points (x_i, y_i, z_i), $i = 1, 2, \ldots, n$.

(i) Explain what the random variable ε represents and the assumptions about it required by this model.

(2)

(ii) Derive the normal equations for finding the least squares estimators for α, β and γ for this model.

(7)

(iii) Briefly explain how the above model can be modified to fit a quadratic relationship between y and x.

(1)

(b) Define the *residuals* from a fitted multiple regression model. Hence explain how the residuals, and in particular residual plots, can be used to check each of the assumptions made by the model. Use sketch diagrams to illustrate your discussion where appropriate.

(10)

3. An experiment is being run in controlled indoor conditions to assess the effect of external temperature on the time taken by joggers to run 10 km. Seventy male joggers take part, with ten being randomly allocated to each of seven temperatures, 0, 5, 10, 15, 20, 25 and 30 degrees Celsius.

(i) Explain how the results of this experiment could be modelled using either one-way analysis of variance or polynomial regression. Write down an appropriate model in each case and compare and contrast the structure and assumptions for the two models.

(8)

(ii) Polynomial regression models up to the 4th order in the temperature (x) are fitted to the results of this experiment, with the residual sums of squares for each model shown below.

Predictors	*Residual SS*	*Residual d.f.*
x, x^2, x^3, x^4	136.4	21
x, x^2, x^3	143.6	22
x, x^2	184.3	23
x	276.1	24
constant only	383.5	25

Use backwards elimination, with tests at the 1% level, to select the best polynomial regression model for these results, fully explaining and justifying your choice.

(8)

(iii) Comment on the design of this experiment, in particular any potential difficulties. How might the design be improved?

(2)

(iv) If, instead of 70 males, the joggers were 49 men and 21 women, describe how you would change the design of the experiment. Explain how you would modify the analysis in this case.

(2)

4. (a) An industrial process is deemed to be in control when the lengths of the components it produces are Normally distributed with mean 50 mm and standard deviation 3 mm. At regular intervals a batch of four components is removed from the production line and their lengths are measured.

(i) Explain how to construct a Shewhart control chart for the mean length. Define and calculate 95% warning limits and 99.9% action limits. Explain and justify the ideas and assumptions underlying the control chart.

(7)

(ii) Successive batches produce mean lengths (in mm) of:

48.7, 53.9, 50.3, 52.6, 53.5, 54.1, 54.6, 55.8, 56.7.

Explain how the control chart would be used to decide whether the process was out of control.

(2)

(b) Items are delivered in large batches and two methods, A and B, of batch testing are being considered. In each case sampled items are classed simply as acceptable or faulty.

Method A: sample 20 items from the batch, accept the batch if there is at most one faulty item in the sample.

Method B: sample 10 items from the batch, accept the batch immediately if there are no faulty items in the sample, reject immediately if there are two or more. If there is exactly one faulty item, take another sample of 10 items and accept the batch if there are no faulty items in the new sample, reject the batch if there is at least one.

(i) If the proportion of faulty items in a batch is p, find the probability of the batch being rejected for each of methods A and B. State and justify any assumptions you make.

(8)

(ii) Verify algebraically that, for all $0 < p < 1$, method A always has a higher chance of rejecting a batch.

(3)

BLANK PAGE

BLANK PAGE

BLANK PAGE