Data provided: Formulae and Critical Values Sheet

MAS370

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2015–2016**

Sampling Theory and Design of Experiments **2 hours**

Candidates may bring to the examination a calculator that conforms to University regulations. Answer all questions. Total marks 60.

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 An investigator is studying the dependence of a variable Y on one continuous explanatory variable x_1, which has been scaled to lie between -1 and 1. It is known that $EY = 0$ when $x_1 = 0$, and the following model is proposed.

$$EY = \beta_1 x_1 + \beta_{11} x_1^2.$$

The investigator proposes 2 different designs, both using 4 observations:

Design	Design points
A	$x_1 = \{-1, -0.5, 0.5, 1\}$
B	$x_1 = \{-1, -1, 1, 1\}$

(i) Show that β_1 and β_{11} are orthogonal to each other in design B.

(4 marks)

(ii) If each observation is subject to a measurement error with mean 0 and variance σ^2, give the variances of the least squares estimators $\hat{\beta}_1$ and $\hat{\beta}_{11}$ in terms of σ^2 for design B. ***(2 marks)***

(iii) Experiments are performed and the response (Y) is measured at the 4 points in both designs A and then separately at the 4 points in design B. Describe the geometric shapes of the 95% confidence regions for $(\beta_1, \beta_{11})^T$ for both designs A and B. Justify whether the centres of these shapes coincide.

(4 marks)

(iv) Justify whether design B is G-optimal. ***(5 marks)***

(v) A design is called A-optimal if it minimises the sum of the diagonal elements of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Consider all **orthogonal** designs for the model $EY = \beta_1 x_1 + \beta_{11} x_1^2$ with 2 designs points such that $-1 \leq x_1 \leq 1$. Show that the design $x_1 = \{-1/2, 1/2\}$ is not A-optimal among all the **orthogonal** designs for this model with 2 design points and specify a design that is A-optimal.

(5 marks)

2 (i) Consider a fractional factorial design with 4 factors (x_1, x_2, x_3, x_4) each of which occurs at two levels, denoted $+1$ and -1. Only 4 design points can be used.

(a) Specify the alias structure when the design generators are $x_1x_2 = 1$ and $x_2x_3 = 1$. ***(5 marks)***

(b) Which parameters are confounded with the main effect parameter for x_1? ***(2 marks)***

(c) What is the resolution of this fractional factorial design? Justify your answer. ***(2 marks)***

(d) Suppose that 8 design points are now available with 4 factors (x_1, x_2, x_3, x_4). State the single design generator that allows the intercept, all main effects and the 3 pairwise interactions x_1x_2, x_1x_3 and x_1x_4 to be included in the linear model without confounding. ***(3 marks)***

(e) If 8 design points are still available, construct a fractional factorial design with the design generator $1 = x_1x_2x_3$. ***(3 marks)***

(ii) Two organisations have independently attempted to estimate the number of civilian casualties following a civil war. Each organisation has produced a list of named civilian casualties. The two lists are then cross-checked, to see which names appear on both lists. The following counts are observed.

Number of names on both lists	213
Number of names on first list only	802
Number of names on second list only	410

Estimate the total number of civilian casualties, stating any assumptions you have made. Give an approximate confidence 95% interval for the total number of civilian casualties. ***(5 marks)***

3 (i) A factory has manufactured 100 steel items. A simple random sample of 10 items are selected, and a property known as the Brinell hardnesses is measured for each item. Summary statistics for the ten items are below (where the unit of measurement is the Brinell hardness number, HB).

$$\sum_{i=1}^{10} x_i = 4551, \qquad \sum_{i=1}^{10} x_i^2 = 2,080,125.$$

Suppose the mean hardness is to be estimated for a second batch of 100 items. Another simple random sample is to be taken. Suggest a suitable sample size such that the width of an approximate 95% confidence interval is no more than 20HB, justifying your answer carefully. What assumptions have you made? ***(7 marks)***

(ii) A stratified sample has been taken to estimate annual household expenditure on energy bills in a town. Two strata are chosen based on postcodes. Summary data are below.

Stratum	Stratum size	Sample size	Sample mean (£)	Sample std. dev. (£)
1	20000	50	609.0	245.8
2	10000	50	438.8	116.6

(a) Estimate the mean annual household expenditure on energy bills for the town. ***(1 mark)***

(b) Give an approximate 95% confidence interval for the mean expenditure. ***(3 marks)***

(c) Suppose the survey is to be repeated next year, with the same total sample size. Assuming that costs of sampling within each stratum are the same, suggest sample sizes within each stratum for the new survey. ***(2 marks)***

 Question 3 continued on next page

3 (continued)

(iii) A survey has been conducted to estimate the use of essay-writing services in a cohort of History undergraduate students. Each participant was asked to toss a coin twice. If the two coin tosses produced the same result, the participant answered the question, “Did you observe two heads?”. If the two coin tosses produced different results, the participant answered the question, “Have you ever used a commercial essay-writing service for a coursework assignment?”

(a) If there are 30 “yes” responses from 100 participants, estimate the proportion of students who have used an essay-writing service, showing clearly how your estimate has been obtained. ***(2 marks)***

(b) Calculate the estimated variance of your estimator, ignoring finite population corrections and assuming that the number of “yes” responses is binomially distributed. ***(2 marks)***

(c) Suggest how you might reduce the variance of your estimator, by modifying the experiment (assuming the same number of participants). You do not need to derive the variance for your modified experiment, or prove that it is smaller. Give one potential drawback of your modification. ***(3 marks)***

End of Question Paper