



The
University
Of
Sheffield.

MAS370

SCHOOL OF MATHEMATICS AND STATISTICS

**Spring Semester
2012–2013**

**MAS370 Sampling Theory and Design of
Experiments**

2 hours

Restricted Open Book Examination.

*Candidates may bring to the examination lecture notes and associated lecture material
(but no textbooks) plus a calculator which conforms to University regulations.*

*Marks will be awarded for your best **three** answers. 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 A small experiment is being conducted to compare a new treatment against a placebo. There are four participants in the study. Three participants are given the placebo, and one the new treatment. Each observation is subject to a measurement error with mean 0 and variance σ^2 . The following model is proposed.

$$EY_{ij} = \mu + \tau_i,$$

for $i = 1, 2, j = 1, \dots, n_i$, with $n_1 = 3$ and $n_2 = 1$. The constraint $\tau_1 + \tau_2 = 0$ is applied.

- (i) Show that the estimators

$$\hat{\mu} = \frac{Y_{11} + Y_{12} + Y_{13} + 3Y_{21}}{6},$$

$$\hat{\tau}_1 = \frac{Y_{11} + Y_{12} + Y_{13} - 3Y_{21}}{6},$$

are unbiased, and give their variances in terms of σ^2 only. **(3 marks)**

- (ii) What condition must hold for μ and τ_1 to be orthogonal? Show whether this condition is satisfied or not. **(3 marks)**
- (iii) Calculate the standardised prediction variance for a new observation in each group. By considering the maximum of these two variances, explain why the design is not D -optimal. **(7 marks)**
- (iv) Suggest an alternative design with four participants such that μ and τ_1 are orthogonal, verifying orthogonality with a suitable calculation. **(4 marks)**
- (v) If the response was believed to depend on the characteristics of the patient, explain how you would implement blocking to account for this, and state the model you would fit to the data to incorporate block effects. Specify any additional necessary parameter constraints. **(3 marks)**

2 An experiment is to be carried out to investigate the effect of three methods on reading comprehension. There are 6 participants in the study, who will each be given instruction using one of the three methods. After instruction, each participant will be given a reading test, and their scores will be recorded.

- (i) State a balanced incomplete block design, with block sizes of 2, that could be used in the experiment. *(1 mark)*

- (ii) For your design in (i), write down the model you would fit to the data in matrix notation, defining your notation carefully, specifying any parameter constraints, and writing out the observation vector, parameter vector and design matrix in full. *(5 marks)*

- (iii) Suggest how you could organise participants into blocks, assuming you could obtain suitable data before choosing the blocks. *(2 marks)*

- (iv) Suppose a larger study is to be conducted, with more participants. If the block size is to be fixed at 2, what is the smallest sample size that is greater than 6, that could be used in a balanced incomplete block design? Justify your answer. *(2 marks)*

- (v) Suppose now there are 9 participants, three methods of instruction, and three teachers. The aim is now to assess the effect of teacher as well as method of instruction.

(a) If the participants are to be organised into blocks of size 3, give a suitable design using a Latin square. State your design by completing the following table.

participant	block	method	teacher
1			
⋮			
9			

(4 marks)

(b) If it is suspected that some teachers may be more effective when using particular methods, without making any further assumptions, would your design in part (a) be suitable? Briefly justify your answer. What further assumption would you have to make in order to fit a suitable model to the data? *(6 marks)*

- 3** (i) Construct a simplex lattice design for a mixture experiment with three constituents, when the following model is to be fitted.

$$EY = x_1 + x_2 + x_3 + x_1x_2 + x_1x_3 + x_2x_3 + x_1x_2x_3.$$

State, with justification, whether the design is saturated or not.

(4 marks)

- (ii) In an agricultural experiment, two varieties of a crop are being compared. It is also desired to investigate the effects of levels of fertilizer and pesticide, which are measured on a continuous scale. For each choice of crop, fertilizer level and pesticide level, the crop yield in a one acre field will be measured. It is believed that, over the range of feasible choices for fertilizer and pesticide levels, the relationship between these variables and the crop yield will be linear.

- (a) If 8 fields are available in the trial, suggest a suitable experimental design, writing your design in full in an appropriate table. State the most complex model that could be fitted to the data, defining your notation carefully. *(8 marks)*

- (b) Suppose that the fields are located at four different farms, with two fields per farm. The fertility of the fields is suspected to vary more between farms than within farms. Explain how you would modify your design in part (a) and list the design in full in an appropriate table. State the corresponding model that you would fit, defining your notation carefully. For your specified model, state a function of the parameters that represents the effect of the first farm.

(8 marks)

4 (i) In a warehouse containing individual frozen beef lasagnes, it is suspected that a proportion of them may contain horse meat. A simple random sample will be taken, and the proportion will be estimated. How large does the sample need to be to ensure that a 95% confidence interval for the true proportion is no wider than 0.05. It is thought to be very unlikely that the proportion containing horse meat will exceed 0.1. You may ignore the finite population correction. Note that the 97.5th percentile of a standard normal random variable is 1.96. *(5 marks)*

(ii) In a health study, the proportion of adults who are smokers is to be estimated. In a pilot survey, a stratified sample is taken using four strata, and the results are tabulated below.

Stratum	Population size (millions)	Sample size	Number of smokers
1	15	50	25
2	10	50	15
3	5	50	5
4	10	50	20

(a) Estimate the proportion of smokers in the population. *(1 mark)*

(b) Using a normal approximation, calculate a 90% confidence interval for the population proportion of smokers. You may ignore the finite population correction. Note that the 5th percentile of a standard normal random variable is -1.645 . *(5 marks)*

(c) If a larger sample was to be taken, with 1000 participants, suggest a sample size for each stratum using Neyman allocation. *(2 marks)*

(iii) A survey has been conducted to estimate the proportion of adults in a population who have driven a car within two hours of consuming alcohol. Each participant first tosses a coin and rolls a six-sided die in secret. If the outcome of the coin toss is a head, the participant answer question A: “Have you ever driven a car within two hours of consuming alcohol?”. If the outcome of the coin toss is a tail, the participant answers question B: “Was the outcome of the die-roll an even number?” There are 100 participants in the survey.

(a) Out of the 100 participants, 30 answer “Yes”. Estimate the proportion of adults in the population who have driven a car within two hours of consuming alcohol, and estimate the variance of your estimator. *(5 marks)*

(b) Suggest a modification to question B that would reduce the variance of the estimator of the proportion, if the survey were to be repeated, and give the variance of the new estimator. *(2 marks)*

End of Question Paper