Examination papers and Examiners' reports

Statistics 2

2790**04b**, 9900**04b**, 996D**04b**

2003, 2004, 2005

Undergraduate study in
Economics, Management, Finance and the Social Sciences

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UNIVERSITY OF LONDON 18 58 EXTERNAL PROGRAMME

Examiner's report 2005

Zone A

General remarks

Once again, it seems advisable to remind candidates that they could make life easier for examiners by writing on the right hand pages only, and by starting each question on a new page. Scraps of solutions of different questions mixed all together are difficult to mark.

Section A (which is compulsory) is designed to test general knowledge and understanding of the whole syllabus. Here you should give reasoned answers, with some explanation, avoiding one word responses, which will never be given any marks. For instance, the one-word answer "True" will never be awarded a mark. It is very clear that candidates who do badly in Section A have a strong tendency to do badly in Section B. It may be that results would improve if candidates spent more time preparing Section A.

In Section B candidates should answer two, and only two, questions. It is strongly recommended that candidates use the built-in statistics routines on their calculators for means, standard errors, analysis of variance and regression.

Once again this year, candidates presented a large amount of badly memorised algebra, and other stock answers. They should concentrate more on understanding the material of the course, and not so much on past exam questions and their solutions. The exam tries to test understanding, not memory.

It is hard to emphasise enough that negative sums of squares and probabilities outside the range 0 to 1 should not be presented in answers, as they give a clear indication of a lack of understanding by the candidate.

Specific comments on questions

Question 1

(a)
One word answers are not enough here. For i a candidate could say that continuous random variables take a particular value with probability 0.
For ii many candidates did not distinguish between the variance of the measurement error and the error sum of squares.
In iii few candidates noticed that the one-sided confidence interval given did have the correct coverage. Some marks could have been gained by mentioning coverage.
For iv one must remember that to judge the performance of an estimator requires consideration of mean squared error, or of bias and variance together. Just thinking about bias is not enough.

(b)
Only four or five lines for each topic are necessary here. A surprising number of candidates gave the definition of Type II error instead of Type I.

(c)
This is a test and confidence interval for a difference between two proportions. One uses results for the binomial distribution, or for part i a chi-squared test. Not all candidates found the critical value 2.1444 in Table 5.

(d)
To answer this question one needs only to know that probabilities sum to 1, and how to calculate the expected value of X^2 and the mean and variance of a discrete random variable X.

(e)
It is important to be familiar with the Cambridge Statistical Tables. For i one must look at the probability of 6 or more successes for a binomial with 15 trials and probability of success 0.31. For ii one uses the complementary probability for $X \leq 15$.

(f)
Few candidates managed to get this right. A total of 3 is possible only if labels 1, 2, 0 are drawn, or if label 3 and two labels 0 are drawn. The first set of labels has probability $6 \times 1/6 \times 1/5 \times 3/4 = 3/20$. The second set of labels has probability $3 \times 1/6 \times 3/5 \times 2/4 = 3/20$.

(g)
This is a goodness-of-fit test, with an estimated parameter. A suitable estimate of the probability of success is the overall proportion of successes, which is $\pi = 81/150 = 0.54$. The probabilities for x successes from that binomial with $n = 3$ are found with a calculator from the formula $\binom{3}{x}\pi^x(1-\pi)^{3-x}$ The expected values are found by multiplying by 50 the probabilities so obtained. The degrees-of-freedom are 4-1-1=2, as one degree-of-freedom is lost through the estimation of π. Remember to state the Null and Alternative hypotheses, the significance level used, the critical value and the outcome of the test.

Question 2

(a)
Proving Bayes' Theorem is bookwork. The application also uses Bayes' theorem. If π is the proportion well-prepared, many candidates obtained the equation $0.95\pi + 0.10(1-\pi) = 0.75$, but not all solved it correctly to give $\pi = 13/17$.

(b)
In order to find both a and b, one must use the fact that the total probability is 1, as

well as the mean given in the question. Since $a = 1, b = 0$, the rest of the question then becomes fairly easy.

Question 3

(a)
This is bookwork - that is - the material required for a complete answer can be found in the appropriate section of your subject guide.

(b)
A good answer is along the lines of:
The variance of aW is $a^2 \operatorname{var} W = 2a^2\sigma^4/(n-1)$. The mean of aW is $a\sigma^2$. So the mean squared error of aW is

$$2a^2\sigma^4/(n-1) + (a-1)^2\sigma^4$$

By calculus, this is minimised for $a = (n-1)/(n+1)$.

(c)
The data are clearly two independent samples of different sizes. Built-in calculator routines can be used to find the standard deviations and means for each of the two samples. It is best to assume equal population variances for the two samples, and use a pooled estimate of that common variance. Candidates should remember to state the significance level used, and in part ii, the conclusion of the test. The upper 2% point of the t distribution with 17 df is not in Table 10, but it is easy to interpolate in Table 9 to get something very close to 2.224.

Question 4

(a)
This is book-work; candidates should say why cross-product terms sum to 0. Candidates should define any symbols used, and say a few words about variation between rows, between columns and within cells.

(b)
This is bookwork.

(c)
Candidates can use built-in calculator routines for the sums of squares in the analysis of variance table. It is better to carry many decimal places to avoid a residual sum of squares calculated as negative. It is important to state Null and Alternative hypotheses, and significance levels, and to state clearly conclusions from tests. In part iii candidates should have given three simultaneous confidence intervals.

Question 5

(a)
This question asked candidates to minimise

$$(y_1 - 2\beta)^2 + (y_2 - \alpha - \beta)^2 + (y_3 - 2\alpha - \beta)^2$$

with respect to α, β. This gives

$$\hat{\alpha} = \frac{-2y_1 + y_2 + 3y_3}{7}$$

$$\hat{\beta} = \frac{10y_1 + 2y_2 - y_3}{21}$$

(b)
For part i, the intercept and slope can be taken from built-in calculator routines. Part ii is book-work. In part iii, the model says that soft wheat yield goes down by 0.154 for each unit increase in rye, yield of barley held constant, and that soft wheat yield goes up by 1.21 for each unit increase in barley, yield of rye held constant.
The simpler model says that without allowing for barley at all, soft wheat yield goes up by 0.704 for each unit increase in rye.
There is no contradiction between these models.

Examination paper for 2006

There will be no change in the format, style or number of questions in the examination paper for 2006.