Examination Papers
Examiners' Reports

Examination papers and
Examiners' reports

Statistics 2

Economics, Management, Finance
and the Social Sciences

2003

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

Examiner's report 2003

Zone A

General Remarks

Statistics 2 was examined for the first time this year. The syllabus follows on from and largely assumes knowledge of that for Statistics 1. Questions do not necessarily fall into neat packages based on some small component of the syllabus, so examinees should be cautious when attempting to pick a small number of topics for exam revision. No change in the style of the paper is expected in 2004.

Section A, for 40% of the marks, is a compulsory question, with several parts. It is designed to test general knowledge of the whole syllabus. Though some of the questions are short, candidates should never offer one word answers such as 'yes' or 'no', for such answers will never be rewarded.

Section B has four questions of which no more than two should be presented for marking. They are meant to test a greater depth of knowledge on parts of the syllabus.

All answers should allow the examiner to see that the candidate understands the answers given. For instance, writing out from memory an answer to a slightly different question is likely to lead to random errors that will not show understanding. Candidates should be able to write answers from their own internalised knowledge, and should spend revision time obtaining a firm grasp of methods rather than attempting to memorise detail.

Candidates should avoid presenting probabilities outside the range 0 to 1, and negative sums of squares. If these arise through some error in working, it's best to write a note for the examiner to say that you know they are wrong.

It is very helpful for markers of the scripts if answers are written on right-hand pages only, and if all the pages for the answer to a particular question are contiguous. Most of the addition errors found during our rigorous checking procedures are for answers split into several pieces appearing in widely separated places in the answer booklet.

Specific comments on questions

Question 1

(a)
It is important to give reasons here, not just the answer 'true' or 'false'. Quite often, the judgment as to whether a statement is true or false depends on how it is interpreted, so the examiner must see reasons to assess the worth of an answer. In i A, B can be disjoint without being empty - one can give an example or a Venn diagram. For ii it is necessary to know the formula

$$\operatorname{var}(X+Y) = \operatorname{var} X + \operatorname{var} Y + 2\operatorname{cov}(X,Y)$$

which implies

$$\operatorname{var}(X+Y) \leq \operatorname{var} X + \operatorname{var} Y$$

if $\operatorname{cov}(X,Y) < 0$. The answer to iii is true if the intervals are independent, for each interval covers the true parameter value with probability 0.95, and of 100 independent intervals one would expect 5 would not cover the true value. For iv it is enough to give the definition of an unbiased estimator.

(b)
Only four or five lines for each topic are necessary here. One must be careful not to describe Type I error when wishing to define Type II error.

(c)
The proportion of correct forecasts is 55/80. It is possible to make different sets of assumptions for part ii, but full marks were given for a confidence interval for a binomial proportion based on 80 trials each with the same probability of a correct forecast.

(d)
The other value for X must be negative because $EX = 0$. To answer this question one needs only to know that probabilities sum to 1, and how to calculate the mean and variance of a discrete random variable.

(e)
For this one needs familiarity with the statistical tables used in the examination room. It is best to use them when revising for the exam. Typical calculators in use by candidates this year can't give a correct answer as easily as the tables. For the binomial probability one needs the complementary probability to that of no more than 5 successes. For the Poisson probability one should look for the complementary probability to that for no more than 6. Remember that the tables show cumulative probabilities.

(f)
Notice that a conditional probability is asked for here, so one can use Bayes' Theorem, and may wish to draw a tree of the possible outcomes. Remember that when finding the probability of an intersection of two events one should not usually assume they are independent.

(g)
This is a goodness-of-fit test. The expected values are found by multiplying probabilities obtained from the binomial tables for $n = 20$, $\pi = 0.2$ by 30. Remember to state the null and alternative hypotheses, the significance level used, the critical value and the outcome of the test.

Question 2

(a)
One should explain why

$$P(0 \text{ empty}) = \frac{3}{4} \times \frac{2}{4} \times \frac{1}{4} = \frac{6}{64}$$

$$P(3 \text{ empty}) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{64}$$

$$P(1 \text{ empty}) = 12 \times \frac{4!}{2!1!1!0!} \times \left(\frac{1}{4}\right)^2 \times \frac{1}{4} \times \frac{1}{4} = \frac{36}{64}.$$

The probability that $m = 2$ can be found by subtraction as $\frac{21}{64}$.

(b)
This question requires the use of the formulae

$$EX = \int_0^2 x(2-x)^3/4dx,$$

$$EX^2 = \int_0^2 x^2(2-x)^3/4dx,$$

$$\text{var } X = EX^2 - (EX)^2.$$

The integrations can be done by writing $(2-x)^3 = 8 - 12x + 6x^2 - x^3$.

Question 3

(a)
This question requires the calculation of the expected value and variance of $\bar{X} + 1$ so that the mean squared error can be found from the formula

$$\text{MSE} = (\text{bias})^2 + \text{variance}.$$

One must use $E\bar{X} = \mu$, $\text{var } \bar{X} = \sigma^2/n$.

(b)
One can reduce the amount of work by using symmetry. One must remember to find ES, $\text{var } S$, ET, $\text{var } T$ for both possible populations (either 2 red balls and 1 green ball or 1 red ball and 2 green balls).

(c)
The data shows repeated observations in each of 10 weeks, so one should use the differences between Mondays and Thursdays for paired t procedures. Built-in calculator routines can be used to find the standard deviation and mean for the 10 differences. For part ii, remember to state the null hypothesis, say

$$\mu_{\text{Monday}} - \mu_{\text{Thursday}} = 2,$$

to give your level of significance, say 5% and to state the conclusion of your test.

Question 4

(a)
This is book-work, but you should say why cross-product terms sum to 0. It is best to define any symbols that you use, such as $\hat{\alpha}$.

(b)
One might take logarithms to induce an additive model for the table, and to make the variances of entries in the table approximately equal.

(c)
One can use built-in calculator routines to obtain quickly and accurately the analysis of variance table. It is best not to use 'Treatments' and 'Blocks' as labels in your table, but instead to put in 'years' and 'ages'. In ii, the null hypothesis is that the year effects are all zero. It should not be 'There is no significant evidence that year effects are not all 0', or anything similar. Remember to give the significance level used in the test. Here the MSS for differences between years is 0, so there is no significant evidence that year effects are different from 0. For iii, one should note that the entries appear to be percentages which sum to nearly 100 in each column rather than number in millions (in fact the Irish population is not large enough for that). There is no way that these percentages can be thought of as independent observations from normal distributions with the same variance.

Question 5

(a)
Part i just needs the formulae for the slopes and the correlation coefficient. Part ii is book-work.

(b)
For part i, the intercept and slope can be taken from built-in calculator routines. Part ii is book-work, though one should note that the interval is for the slope, and not for a fitted value. In part iii, one should notice that the price for 1 of 10 is always exactly 84p less than the price for 1, so when the price for 1 in 10 is used as an explanatory variable it will explain the price of 1 exactly. The model will be

$$\text{Price for } 1 = 0.85 + 0 \times \text{price for 1 of 5} + \text{Price for 1 of 10}.$$