GX1589

OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SUMMER EXAMINATIONS 1999

THIRD UNIVERSITY EXAMINATION

MA338 - *STATISTICS*

Dr. M. Kenward
Professor T. Hurley
Dr. J. Ward

Time allowed: *Two* hours
Answer three questions

Statistical tables and formula sheets are provided.

1. Discuss briefly various methods of sampling from a population.
A mathematics class is divided into three strata:
(a) 60 commerce students; (b) 40 arts Students; (c) 50 science students.
For the corresponding class in the preceeding year, the mean mathematics marks and their standard deviations for the strata are tabulated below:

	Commerce	**Arts**	**Science**
Mean	65	55	60
St. dev.	10	12	8

The standard deviation for the class based on these figures is 10.7. To estimate the progress of the class, a test is given to a random sample of 50 students. Find the proportional and optimum allocations of the sample to the strata and estimate the standard deviation of the estimate for each of these allocations.

Estimate the standard deviation of the estimate of the mean obtained from a simple random sample of size 50 chosen from the class. If the sample stratum means obtained in the survey are 62%, 55% and 58% respectively, estimate the class mean.

p.t.o.

2. In the general context of the analysis of time series, write brief notes on each of the following:

(i) Principles of decomposition of time series;

(ii) Forecasting and trend curves;

(iii) Adaptive forecasting.

The following table gives the demand Y_t for television sets at a retail outlet during month t:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	113	101	109	98	87	95	103	98	106	88	84	92

Smooth the series using an EWMA with starting value M_0=103 and A=0.2. Calculate the moving average of length 4 and illustrate the original data and both smoothed series graphically.

3. For a 1-way analysis of variance with unequal replications show that

$$\sum_{j=1}^{N}\sum_{i=1}^{k_j}(x_{ij}-\bar{x}..) = \sum_{j=1}^{N}k_j(\bar{x}._j-\bar{x}..)^2 + \sum_{j=1}^{N}\sum_{i=1}^{k_j}(x_{ij}-\bar{x}._j)^2,$$

where $\bar{x}._j = \frac{1}{k_j} = \sum_{i=1}^{k_j}x_{ij}$ and $\bar{x}.. = \frac{1}{n}\sum_{j=1}^{N}\sum_{i=1}^{k_j}x_{ij}$ with $n=(k_1+k_2+\ldots+k_N)$.

The cuckoo (Cuculus canorus) has the distinctive habit of laying its eggs in nests of birds of other species. Measurements (in mm.) of cuckoo eggs found in nests of 3 different species are given below:

Hedge Sparrow	Robin	Wren
22.0	23.9	20.3
23.1	23.0	22.1
20.9	22.9	22.0
23.5	22.4	20.9
25.0	22.6	20.8
23.0	23.0	21.0
21.7	21.1	
	23.0	

Is the difference significant? State the assumptions underlying the test.

4. (a) Given that $Y_i = a + bX_i + \varepsilon_i, 1 \le i \le n$ where $E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2$, derive the least squares estimates $\hat{a}, \hat{b}$ of the parameters a and b.

(b) Now assume the ε_i are independent and normally distributed.

(i) Given the fact that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\left\{Y_i - \left(\hat{a} + \hat{b}X_i\right)\right\}^2$$

is independed of $\hat{b}$ and has a χ^2_{n-2} distribution, derive a suitable statistic for a test of the hypothesis $b = b_o$.

(ii) In the case of the following random sample of pair (X, Y) values, calculate the least squares regression line of Y on X and (given that the relevant assumptions are justified) test whether the slope of this line differs significantly from 1.2.

X:	10	14	18	22	26	30	34	38
Y:	19	22	25	31	33	39	44	45

(iii) Calculate the sample correlation coefficient and comment on your results in (i) and (ii).

5. In a multiple linear reg. analysis 4 independent variables X_1, X_2, X_3, X_4 were regressed against Y. There were 25 data points and a computer package was used to perform the regression. Part of the output was

Source	***df***	**Sum of Squares**	**Mean Sq.**	***F***
Regression	5	6044.34	z	f
Error	19	61.59	w	
Total	x	y		

(a) (i) Calculate the missing entries x, y, z, w and f in the table.

(ii) Calculate the coefficient of determination R^2.

(iii) Taking $\alpha = 0.01$ perform an overall F test of the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ versus H_1: not all the βs are zero. Note that one of the following critical points is relevant

$$F_{.01,5,19,} = 4.171 \quad F_{.01,19,5,} = 9.643 \quad F_{.01,5,24} = 3.895.$$

p.t.o.

(b) Consider using a forward selection procedure to explain the dependent variable Y using at most three of the four variables X_1, X_2, X_3, X_4, above.

Information about all possible *15 (2^4-1)* models is given below.

(i) Using $\alpha = .01$, show that the best single variable model $(Y = \beta_0 + \beta_i X_i)$ is for X_1.

(ii) Show that the increase in sum of squares by adding X_2 to the model already containing X_1 is 104.9191 (1 d.f.), but that the increase in sum of squares by adding X_4 to the model aready containing X_1 is 116.4725. Calculate the corresponding partial F statistics:

$$\frac{SS_R(\beta_2 \mid \beta_1, \beta_o)/1}{MS_E(X_2, X_1)} \quad \text{and} \quad \frac{SS_R(\beta_4 \mid \beta_1, \beta_0)/1}{MS_E.(X_4, X_1)}.$$

$[[F_{.01,1,22} = 7.945]]$.

(iii) Retaining the 2-variable model $Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4$, decide by calculation which of X_2 or X_3 yields the larger partial F-value by comparing

$$\frac{SS_R(\beta_2 \mid \beta_4, \beta_1 \beta_0)}{MS_E(X_2, X_4, X_1)} \quad \text{and} \quad \frac{SS_R(\beta_3 \mid \beta_4, \beta_1, \beta_0)}{MS_E(X_3, X_4, X_1)}.$$

$[[F_{.01,1,21} = 8.017]]$.