UNIVERSITY OF NEWCASTLE UPON TYNE

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2

2001/2002

MAS272/672

Foundations of Statistics

Time: 1 hour 30 minutes

Credit will be given for ALL answers to questions in Section A and the best TWO answers to questions in Section B.

No credit will be given for other answers and students are strongly advised not to spend time producing answers for which they will receive no credit.

There are FOUR questions in Section A and THREE questions in Section B. Marks allocated to each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

Statistical tables (Neave and studentised range) are provided.

SECTION A

A1. Suppose that $X_1, X_2, ..., X_n$ are a sample of independent and identically distributed random variables with finite mean μ and variance σ^2. What, according to the Central Limit Theorem, will be the approximate distribution of the sample mean $\bar{X}$?

The distribution of a discrete random variable X is given below. Evaluate the mean and variance of X and determine the approximate probability that the mean of 20 independent observations on X will exceed 2.5.

x	0	1	2	3	4	5
$p(X = x)$	0.05	0.25	0.3	0.25	0.1	0.05

[10 marks]

A2. (a) Explain carefully what is meant by a *confidence interval*. Why is the alternative name *"interval estimate"* sometimes thought more appropriate?

(b) A random sample from a Normal population has the following summary statistics:

$$n = 15, \quad \bar{x} = 14.8, \quad s = 4.2.$$

(i) Calculate a 99% confidence interval for the population mean. Would a 95% confidence interval be wider or narrower? [Do not calculate this interval.]

(ii) Calculate a 90% confidence lower bound for the population standard deviation. If the same value of s were to be obtained in a larger sample, would the corresponding bound be lower or higher?

[12 marks]

A3. Independent random samples from two Normal populations have the following summary statistics:

$$n_1 = 20, \quad \bar{x}_1 = 14.9, \quad s_1 = 3.6$$
$$n_2 = 16, \quad \bar{x}_2 = 15.2, \quad s_2 = 6.8.$$

Test the null hypothesis that the population *standard deviations* are equal.

[6 marks]

A4. (a) Describe, giving one illustrative example in each case, pitfalls in the interpretation of correlation coefficients which can lead to:

(i) A high correlation between apparently unrelated variables.

(ii) A low correlation between strongly related variables.

(b) Assuming a Normal distribution for the underlying observations, test the null hypothesis $\rho = 0$ for the following correlation coefficients and sample sizes.

(i) $n = 8, \quad r = 0.59$

(ii) $n = 12, \quad r = 0.59$

(iii) $n = 20, \quad r = 0.59$.

Explain briefly why the same correlation coefficient leads to different conclusions in these three cases.

[12 marks]

SECTION B

B5. (a) Define the moment generating function (mgf) of a random variable in terms of the expectation of a function of X.

Suppose that X and Y are independent random variables with mgf's $M_X(t)$ and $M_Y(t)$ respectively. Show that the mgf of the sum $S = X + Y$ is given by $M_S(t) = M_X(t) \times M_Y(t)$.

(b) You are given that the mgf for a χ^2_ν random variable is $M(t) = (1-2t)^{-\nu/2}$.

(i) Derive the mean and variance of a χ^2_ν random variable.

(ii) Show that the sum of two independent χ^2 random variables also follows a χ^2 distribution.

(c) Write down an expression for the sampling distribution of S^2, the random variable associated with the sample variance. From this expression, and using the results in part (b) above, find the mean and variance of S^2 in terms of the sample size n and the population variance σ^2.

(d) Explain why the Central Limit Theorem can be used to given an approximation to the distribution of S^2 in large samples and write down the form of the approximating distribution. Hence obtain an approximate 95% confidence interval for σ^2 when $n = 250$ and $s^2 = 4.8$.

[30 marks]

B6. (a) You may **assume** (without proof) the following results:

Let Z be a standard Normal random variable and let U be a χ^2_ν random variable, with Z and U independent. Then

$$T = Z \Big/ \sqrt{\frac{U}{\nu}}$$

follows Student's t-distribution on ν degrees of freedom.

Use this result to show that, in Normal random samples, the statistic $\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ follows a t-distribution on $(n-1)$ degrees of freedom.

[**NB**: You may use, without proof, standard results for the sampling distributions of $\bar{X}$ and S^2 in Normal random samples.]

(b) Explain briefly the different circumstances in which the paired-sample t-test and the two-sample t-test are used.

In a comparative study of two groups of patients, group A were the treatment group and group B were the control group. Each patient was assessed at the end of the investigation and their scores (higher indicates greater improvement) were as follows:

A: 2.6, 2.0, 1.7, 2.7, 2.5, 2.6, 2.5, 3.0

B: 1.2, 1.8, 1.8, 2.3, 1.3, 3.0, 2.2, 1.3, 1.5, 1.6.

Sketch a comparative plot of the data (in your answer book) and carry out an appropriate test of the null hypothesis that the means of the two populations are equal. Give a 95% confidence interval for the difference in the population means.

[30 marks]

B7. (a) State the null and alternative hypotheses used in a one-way Analysis of Variance (ANOVA). What assumptions are usually made?

(b) In an experiment on plant growth, fruit bushes were grown in four different types of soil in standard conditions. The resulting yields (kg) were:

Soil	Crop yield						Total
A	15.9	17.2	16.6	16.9	15.9	17.1	99.6
B	16.5	16.3	16.1	15.7	16.9	15.2	96.7
C	13.8	15.6	14.3	14.4	14.7	13.7	86.5
D	17.1	16.6	17.0	16.4	17.6	17.2	101.9

You are given that the *corrected* Total Sum of Squares (TSS) is 29.870. Give a rough comparative dotplot of the data (in your answer book, not on a separate sheet). Construct the ANOVA table and use it to test whether the mean yield varies between the soils. State your conclusions clearly.

Use LSD and Tukey methods to identify specific differences in the soil means and comment on your findings.

[30 marks]