UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Department of Statistics and Department of Computer Science

B.A. (Mod.) Computer Science **Trinity Term 2001**

3BA1 Statistics and Numerical Methods

Thursday 31st May **Mansion House** **9.30 – 12.30**

Mr. E. Mullins and Professor J. G. Byrne

Answer five questions, at least one of which is from Section B. Statistical Tables are available from the Invigilator. Calculators may be used. Use separate answer books for each section. A table of formulae applicable to section A is attached.

SECTION A

1. (a) A computer system is equipped with 3 hard disks, each independently having a 97% probability of functioning correctly. At least 2 of these are required for a working system. What is the probability that the system is not working?

(b) A task involving a large amount of computing time may be delayed due to a computer failure. The probabilities are 0.6 that the computer will fail, 0.85 that the task will be completed on time if the computer does not fail and 0.35 if it does. What is the probability that the task will be completed on time?

(c) In (b), if the task is completed on time, what is the probability that the computer failed?

(d) A customer estimates that when attempting to get connected to an Internet Service Provider the probability of connection at any one attempt is p=.3. Write down an expression to describe the probability that the customer will fail to make the connection three times and get connected at the fourth attempt.

2. A printer manufacturer buys Circuit Card Assemblies (CCAs) from three vendors, A, B and C. Cards are inspected for workmanship non-conformities prior to functional testing in the printers. Cards which are rejected at inspection are classified as having Major or Minor non-conformities. A review of recent inspection records for CCAs gave the following results:

		Vendor		
		A	B	C
Type of Non-Conformity	Minor	87	53	54
	Major	13	9	10
	None	347	190	108
	Total	447	252	172

A standard chi-square test was carried out on the table and gave a test statistic of 14.83

(a) Explain how the test statistic was calculated (calculations not required).

(b) Explain the hypothesis being tested. What is the critical value for a test with a significance level of $\alpha = .05$? Interpret the result of the test.

(c) Calculate a 95% confidence interval for the proportion of CCAs with minor non-conformities from supplier A.

(d) Calculate a 95% confidence interval for the difference between the proportions of CCAs with minor non-conformities from suppliers B and C.

3. The IT manager of a major bank is concerned about the maintenance costs associated with the bank's ATM equipment. As part of a pilot study to get a feel for the problem she selects 15 records at random and fits a regression model to the data. The costs are the total amounts spent on the machines (in IR£) since installation and include both installation and total maintenance. Age is measured in months. The regression analysis, which includes predictions at 36 months, is shown below.

```
                        ATM    Age    Cost

                         1     15    1609.00
                         2     17    1824.80
                         3     20    2157.55
                         4     24    1675.53
                         5     26    1743.97
                         6     30    2212.96
                         7     32    2095.57
                         8     36    2630.93
                         9     38    2553.55
                        10     40    2734.04
                        11     42    2718.49
                        12     44    2793.74
                        13     40    2384.01
                        14     46    2737.75
                        15     48    2468.97

The regression equation is
Cost = 1155 + 34.2 Age

Predictor          Coef         StDev          T          P
Constant         1155.4         184.8       6.25      0.000
Age              34.158         5.311       6.43      0.000

S = 213.5        R-Sq = 76.1%      R-Sq(adj) = 74.2%

Analysis of Variance

Source            DF          SS          MS          F          P
Regression         1     1885947     1885947      41.36      0.000
Residual Error    13      592802       45600
Total             14     2478750


Predicted Values

     Fit  StDev Fit          95.0% CI               95.0% PI
  2385.0       57.1   (  2261.7,   2508.4)  (  1907.5,   2862.6)
```

(a) Write down the model that underlies the regression analysis. In the context of the ATM study interpret the model parameters.

(b) The output contains two t-tests on the regression coefficients. Say what hypotheses are being tested and interpret the results of the tests.

(c) Calculate a 95% confidence interval for the slope parameter and interpret the interval.

(d) Interpret the intervals given by Minitab for time 36 months.

(e) What does R-sq measure and how was it calculated?

4. A time-sharing system has a single CPU to which a number of terminals are connected. Jobs arriving at the CPU are served on a first-come-first-served basis. The following data were collected for 100 one-second intervals.

Jobs per second	0	1	2	3	4	5
Frequency	20	38	26	9	6	1

It has been suggested that the arrival pattern of jobs may be modelled as a Poisson process.

(a) Write down the Poisson model, estimate the arrival rate from the data and from this calculate the distribution of frequencies that would be expected if this were an appropriate model for the system.

(b) Formally test the goodness of fit of the model to the data.

5. (a) The total time to repair a certain piece of computer equipment is the sum of the times required to complete three separate stages. These times (minutes) are each independently subject to variability:

Stage 1:	$N(30, 4^2)$
Stage 2:	$N(15, 3^2)$
Stage 3:	A fixed test time of 10 mins.

(i) Find the mean and variance of the total repair time.

(ii) Find the probability that the total repair time is less than one hour.

(b) In a pilot study of new network software it is assumed that the time it takes to download a test file from a fileserver is Normally distributed. How would you assess the validity of this assumption?

(c) The test file was downloaded from the server 10 times giving a sample mean of 120 seconds and a standard deviation of 10 seconds. Calculate a 95% confidence interval for the true mean time it takes to download the file from the server. Interpret the interval.

(d) Test the hypothesis that the true mean time equals 100 seconds. Use a two tailed test and a 5% level of significance.

(f) Comment on the relationship between your confidence interval and the result of your hypothesis test.

6. The time it takes to load files into RAM is a critical performance measure for computer packages. The times taken to load a standard software package using two different operating systems are recorded below, for 10 PCs of various ages and processor speeds. The study was designed to test whether there was a difference in the average loading times for the two operating systems.

Loading Time in Seconds

PC No.	**Operating System 1**	**Operating System 2**
1	37.2	39.9
2	42.0	48.5
3	45.8	49.5
4	34.6	41.4
5	43.6	45.3
6	36.2	39.3
7	50.3	51.3
8	42.6	40.1
9	37.6	41.6
10	34.6	36.9

Two analyses were carried out using Minitab and some results are shown below.

```
Two sample T for Op-Sys-1 vs Op-Sys-2

             N      Mean     StDev   SE Mean
Op-Sys-1    10     40.45      5.25       1.7
Op-Sys-2    10     43.38      4.93       1.6

T-Test mu Op-Sys-1 = mu Op-Sys-2 (vs not =): T = -1.29
P = 0.21   DF = 18
Both use Pooled StDev = 5.09
```

```
Paired T for Op-Sys-1 - Op-Sys-2

                    N      Mean     StDev   SE Mean
Op-Sys-1           10     40.45      5.25      1.66
Op-Sys-2           10     43.38      4.93      1.56
Difference         10    -2.930     2.680     0.848

T-Test of mean difference = 0 (vs not = 0): T-Value = -3.46
P-Value = 0.007
```

(a) Explain how each t-statistic was calculated.

(b) Assume in turn that each analysis is correct and interpret the result of the t-test.

(c) Which test do you consider more appropriate and why?

(d) Using the summary statistics from the analysis you consider more appropriate, calculate a confidence interval for the long-run mean difference between loading times.

SECTION B

7. (1) Define the finite precision floating point system.

(2) Derive the formula for the relative error in subtracting two floating point numbers.

(3) Develop an algorithm for the solution of quadratic equations which avoids subtraction as much as possible.

8.

(a) Solve the following linear programme by the Simplex method:

$$\text{Maximise } P = 80y + 60x$$

subject to the constraints

$$7y + 10x < 350$$
$$4y + 3x < 120$$
$$y + 30x < 60$$

(b) Derive an expression for the complexity of the Simplex algorithm where m = no. of constraints and n = no. of variables. Give one reason why it is an upper bound.

Some of these formulae may be useful for questions in Section A.

Statistical Distributions:

Name	$p(x), f_X(x)$	E(X)	V(X)
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	np	np(1-p)
Geometric	$p(1-p)^{x-1}$		
Poisson	$e^{-\lambda} \frac{\lambda^x}{x!}$	λ	λ
Hypergeometric	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$		

Rules of probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad , \quad P(A) = \sum_i P(A|B_i)P(B_i) \quad , \quad P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

If A and B independent $P(A|B)=P(A)$.

Statistical Estimation and Testing:

$$\bar{x} = \frac{1}{n}\sum_i x_i \ , \quad s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \quad , \quad z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \quad , \quad t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} \quad , \quad SE(\bar{x}-\bar{y}) = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}, \text{ pooled estimate of } \sigma \quad s = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}}$$

ctd.

Linear Regression

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad , \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad , \quad MSE = \frac{SSE}{n\text{-}2}$$

$$S\hat{E}(\hat{\beta}_1) = \frac{MSE}{\sum (x_i - \bar{x})^2} ,$$

$$\sqrt{MSE(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2})}$$

$$\sqrt{MSE(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2})}$$