UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Departments of Statistics and Computer Science

B.A. (Mod) Computer Science
J. S. Examination

Trinity Term 2002

3BA1 Statistics and Numerical Analysis

Thursday 30th May **MANSION HOUSE** **09.30 - 12.30**

Mr. Eamonn 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) I estimate that there are ten spam messages, on average, to be discarded when I open my email account each morning. If the number of messages may be described by a Poisson distribution, what is the probability that on any given morning I will receive less than four? (4 marks)

   (b) If I keep a tally of the number of spam messages I receive for ten days, what is the probability that I will receive fewer than four on two or more of those days? (4 marks)

(contd.)

(c) On a particular day I start counting the number of days until I receive fewer than four spam messages when I open my email account. Write down an expression which can be used to compute the probability of this occurring on the X^{th} day. (4 marks)

(d) Give examples of when the Binomial and Hypergeometric models may be applied to a practical situation. Explain clearly in each case why the model is applicable to the example you describe. (4 marks)

(e) What is Bayes' theorem/formula and for what is it used? Give an example to illustrate your answer. (4 marks)

2. As part of a market research study interviewees were asked whether they had ever heard of a particular software product, had heard of but not bought it or had bought it at least once. The results classified by region are presented below.

Table of Responses Classified by Region

Region	Never Heard	Heard, but did not buy	Bought	Totals
A	36	55	109	200
B	45	56	49	150
C	54	78	168	300
Totals	135	189	326	650

(a) A chi-squared test was carried out on the table: it gave an X^2 value of 24.6. Explain what 'null hypothesis' is being tested, state the degrees of freedom and the critical value for the appropriate chi-square distribution, and interpret the result of the test in practical terms. (6 marks)

(b) Draw a rough sketch which will illustrate clearly the patterns in the data. (4 marks)

(c) Obtain a 95% confidence interval for the difference between the proportions of people in the target populations in regions B and C who had heard of the product but had not bought it. (5 marks)

(contd.)

(d) Are the results consistent with the hypothesis that 20% of the target population in Region A had never heard of the software product? (5 marks)

3. Experimental determinations of spring constants were made for three spring types; Type 1 (4 inch design with a theoretical spring constant (tsc) of 1.95), Type 2 (6 inch, tsc = 2.63), Type 3 (4 inch, tsc = 2.12) purchased from one manufacturer. The following summary statistics were obtained; six springs were tested in each case. You may assume Normality.

	Spring Type		
	1	**2**	**3**
$\overline{X}$	2.03	2.55	2.34
S	.068	.084	.064

(a) Formally test that the long run average tsc-value for springs of Type 1 equals the theoretical value 1.95. (4 marks)

(b) Obtain a 99% confidence interval for the long run average tsc-value for springs of Type 3. (4 marks)

(c) Formally test whether the long run standard deviations for the tsc-values for spring Types 2 and 3 are the same. (4 marks)

(d) Obtain a 95% confidence interval for the long run difference in the mean tsc-values of Types 2 and 3. (4 marks)

(e) Discuss the issues that need to be considered when determining appropriate sample sizes for studies of this type. (4 marks)

4. A company that provides a preventive maintenance and repair service for workstations carried out a study of the times taken on service calls. Data were recorded for the eighteen most recent preventive maintenance calls and a simple linear regression line was fitted. The regression output is shown below; X is the number of devices serviced and Y is the total number of minutes spent by the service man or woman on the call.

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	-1.947	2.431	-0.80	0.435
X	14.6930	0.5069	28.99	0.000

S = 4.419   R-Sq = 98.1%

Analysis of Variance

Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	1	16407.2	16407.2	840.19	0.000
Residual Error	16	312.4	19.53		
Total	17	16719.6			

Assume that the simple linear regression model is appropriate for these data.

(a) Obtain a 95% confidence interval for the change in the mean response when the number of devices requiring service increases by one. (4 marks)

(b) The workstation manufacturer suggests that the mean time required should not increase by more than 13 minutes for each additional workstation that is given routine preventive maintenance on a service call. Carry out a test to decide whether this standard is being satisfied by the service company. Use a significance level of 0.05. (4 marks)

(c) Estimate the mean service time on calls during which seven devices are serviced (note: $\bar{X} = 4.33$ and $\sum (X_i - \bar{X})^2 = 76.0$). (4 marks)

(d) Obtain a prediction interval for the service time on the next call during which seven devices are to be serviced. (4 marks)

(e) Write down the model on which the regression analysis is based. Explain the elements of the model and the assumptions being made. (4 marks)

5. A computer user counted the number of emails she received on each day of the working week for ten weeks. The totals are recorded in the table below.

Day	Monday	Tuesday	Wednesday	Thursday	Friday
No. of emails	706	674	495	527	478

(a) Carry out a formal statistical test of the hypothesis that for this user the rate of receipt of emails is the same for each day of the week. (7 marks)

(b) Calculate and interpret a 95% confidence interval for the proportion of her week's emails that are received on the first two days of the week. (5 marks)

(c) Explain the types of error that can arise in carrying out statistical tests. Use the test from part (a) to illustrate your answer. (8 marks)

6. A company sells three software products A, B, C and provides hotline assistance for each product. The distributions of the daily numbers of minutes of helpline calls for each product can be represented by a Normal distribution with the following characteristics (in minutes):

A: mean = 1200 , standard deviation = 100
B: mean = 2500 , standard deviation = 200
C: mean = 500 , standard deviation = 50

The demand for assistance for each product is independent of the demands for the other products and the demands on different days are independent of each other. The total service provided on any day, T, is the sum of the three demands.

(a) Derive the expectation and variance of T, clearly stating any results which are assumed in your derivation. (4 marks)

(b) What is the probability that the total service demand will exceed 4500 minutes? (4 marks)

(c) What is the probability that the average number of minutes demanded for product A in a five-day week is greater than 1300 minutes? (4 marks)

(d) What is the probability that in any one week the demand for assistance with product B exceeds 2800 minutes on at least one day? (4 marks)

(e) A Normal probability plot might be used to verify the assumption of Normality assumed in this question. Explain the rationale underlying such plots. (4 marks)

Section B

7. Define the Simple Iterative Method for finding a root of a single non-linear equation. (4 marks)

Show using a Taylor Series that it has linear convergence. Under what condition will it converge? (6 marks)

Use the method to find a root of the equation:

$$2x^2 - 3x - 2 - e^{-x} = 0$$

to a tolerance of 1×10^{-3}.

(10 marks)

8. What is Lagrangean Interpolation? (4 marks)

Write down the interpolation formula. What is its complexity? (6 marks)

Derive the barycentric form of the method and derive its complexity. (10 marks)

ST3BA11

University of Dublin
Trinity College

Department of Statistics

Statistical Tables for Examinations

151
3BA1
Dip. in Stats

ST3BA11

Areas under the standard Normal curve

The table gives the area to the left of z. For example, if z = 1.32, the area is .9066. If z = -1.32, the area is .0934.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Selected critical values of the t-distribution

α	.25	.10	.05	.02	.01	.002	.001
ν = 1	2.41	6.31	12.71	31.82	63.66	318.32	636.61
2	1.60	2.92	4.30	6.96	9.92	22.33	31.60
3	1.42	2.35	3.18	4.54	5.84	10.22	12.92
4	1.34	2.13	2.78	3.75	4.60	7.17	8.61
5	1.30	2.02	2.57	3.36	4.03	5.89	6.87
6	1.27	1.94	2.45	3.14	3.71	5.21	5.96
7	1.25	1.89	2.36	3.00	3.50	4.79	5.41
8	1.24	1.86	2.31	2.90	3.36	4.50	5.04
9	1.23	1.83	2.26	2.82	3.25	4.30	4.78
10	1.22	1.81	2.23	2.76	3.17	4.14	4.59
12	1.21	1.78	2.18	2.68	3.05	3.93	4.32
15	1.20	1.75	2.13	2.60	2.95	3.73	4.07
20	1.18	1.72	2.09	2.53	2.85	3.55	3.85
24	1.18	1.71	2.06	2.49	2.80	3.47	3.75
30	1.17	1.70	2.04	2.46	2.75	3.39	3.65
40	1.17	1.68	2.02	2.42	2.70	3.31	3.55
60	1.16	1.67	2.00	2.39	2.66	3.23	3.46
120	1.16	1.66	1.98	2.36	2.62	3.16	3.37
∞	1.15	1.64	1.96	2.33	2.58	3.09	3.29

Note that this table gives the percentage split between the two tails. For example, 95% of the area under a t-distribution with 1 degree of freedom lies between -12.71 and 12.71

Selected upper tail critical values for the F distribution
with ν_1 numerator and ν_2 denominator degrees of freedom

5% critical values for the F distribution

ν_2 \ ν_1	1	2	3	4	5	6	7	8	10	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	241.9	243.9	249.1	254.3
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.5	19.5
3	10.1	9.6	9.3	9.1	9.0	8.9	8.9	8.8	8.8	8.7	8.6	8.5
4	7.7	6.9	6.6	6.4	6.3	6.2	6.1	6.0	6.0	5.9	5.8	5.6
5	6.6	5.8	5.4	5.2	5.1	5.0	4.9	4.8	4.7	4.7	4.5	4.4
6	6.0	5.1	4.8	4.5	4.4	4.3	4.2	4.1	4.1	4.0	3.8	3.7
7	5.6	4.7	4.3	4.1	4.0	3.9	3.8	3.7	3.6	3.6	3.4	3.2
8	5.3	4.5	4.1	3.8	3.7	3.6	3.5	3.4	3.3	3.3	3.1	2.9
9	5.1	4.3	3.9	3.6	3.5	3.4	3.3	3.2	3.1	3.1	2.9	2.7
10	5.0	4.1	3.7	3.5	3.3	3.2	3.1	3.1	3.0	2.9	2.7	2.5
12	4.7	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.8	2.7	2.5	2.3
15	4.5	3.7	3.3	3.1	2.9	2.8	2.7	2.6	2.5	2.5	2.3	2.1
20	4.4	3.5	3.1	2.9	2.7	2.6	2.5	2.4	2.3	2.3	2.1	1.8
30	4.2	3.3	2.9	2.7	2.5	2.4	2.3	2.3	2.2	2.1	1.9	1.6
40	4.1	3.2	2.8	2.6	2.4	2.3	2.2	2.2	2.1	2.0	1.8	1.5
120	3.9	3.1	2.7	2.4	2.3	2.2	2.1	2.0	1.9	1.8	1.6	1.3
∞	3.8	3.0	2.6	2.4	2.2	2.1	2.0	1.9	1.8	1.8	1.5	1.0

2.5% critical values for the F distribution

ν_2 \ ν_1	1	2	3	4	5	6	7	8	10	12	24	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	968.6	976.7	997.3	1018.3
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.5	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.4	14.3	14.1	13.9
4	12.2	10.6	10.0	9.6	9.4	9.2	9.1	9.0	8.8	8.8	8.5	8.3
5	10.0	8.4	7.8	7.4	7.1	7.0	6.9	6.8	6.6	6.5	6.3	6.0
6	8.8	7.3	6.6	6.2	6.0	5.8	5.7	5.6	5.5	5.4	5.1	4.8
7	8.1	6.5	5.9	5.5	5.3	5.1	5.0	4.9	4.8	4.7	4.4	4.1
8	7.6	6.1	5.4	5.1	4.8	4.7	4.5	4.4	4.3	4.2	3.9	3.7
9	7.2	5.7	5.1	4.7	4.5	4.3	4.2	4.1	4.0	3.9	3.6	3.3
10	6.9	5.5	4.8	4.5	4.2	4.1	3.9	3.9	3.7	3.6	3.4	3.1
12	6.6	5.1	4.5	4.1	3.9	3.7	3.6	3.5	3.4	3.3	3.0	2.7
15	6.2	4.8	4.2	3.8	3.6	3.4	3.3	3.2	3.1	3.0	2.7	2.4
20	5.9	4.5	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.7	2.4	2.1
30	5.6	4.2	3.6	3.2	3.0	2.9	2.7	2.7	2.5	2.4	2.1	1.8
40	5.4	4.1	3.5	3.1	2.9	2.7	2.6	2.5	2.4	2.3	2.0	1.6
120	5.2	3.8	3.2	2.9	2.7	2.5	2.4	2.3	2.2	2.1	1.8	1.3
∞	5.0	3.7	3.1	2.8	2.6	2.4	2.3	2.2	2.0	1.9	1.6	1.0

Selected critical values of the chi-squared distribution

α	.2	.1	.05	.025	.01	.005
ν = 1	1.64	2.71	3.84	5.02	6.64	7.88
2	3.22	4.61	5.99	7.38	9.21	10.60
3	4.64	6.25	7.82	9.35	11.35	12.84
4	5.99	7.78	9.49	11.14	13.28	14.86
5	7.29	9.24	11.07	12.83	15.09	16.75
6	8.56	10.65	12.59	14.45	16.81	18.55
7	9.80	12.02	14.07	16.01	18.48	20.28
8	11.03	13.36	15.51	17.54	20.09	21.96
9	12.24	14.68	16.92	19.02	21.67	23.59
10	13.44	15.99	18.31	20.48	23.21	25.19
12	15.81	18.55	21.03	23.34	26.22	28.30
15	19.31	22.31	25.00	27.49	30.58	32.80
20	25.04	28.41	31.41	34.17	37.57	40.00
24	29.55	33.20	36.42	39.36	42.98	45.56
30	36.25	40.26	43.77	46.98	50.89	53.67
60	68.97	74.40	79.08	83.30	88.38	91.96
120	132.81	140.23	146.57	152.21	158.95	163.65

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:

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

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

(Contd.)

Linear Regression

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

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

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

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