TRINITY COLLEGE DUBLIN
THE UNIVERSITY OF DUBLIN

Faculty of Engineering, Mathematics and Science

School of Computer Science & Statistics

Integrated Computer Science Programme **Trinity Term 2015**
Year 3 Annual Examinations

Statistical Analysis

Professor Myra O'Regan

Saturday 16th May, 2015 Luce Lower 14:00 – 16:00

Instructions to Candidates:

Answer all questions. All questions carry equal marks. Appendix A contains some useful formulae and normal tables.

Calculators may be used.

Question 1

a) The number of customers visiting a site in an hour is known to have a Poisson distribution with an average of 3.4. What is the probability of having 5 visitors to the site in an hour? What is the probability of having at least 1 visitor to the site in an hour?

(17 marks)

b) Seventy percent of light aircraft that disappear while in flight in a certain country are subsequently discovered. Of the aircraft that are discovered, 60% have an emergency locator. Of the aircraft which were not discovered 90% do not have an emergency locator. Suppose that a light aircraft has disappeared. If it has an emergency locator, what is the probability that it will be discovered?

(17 marks)

c) The time to complete a job is normally distributed with a mean of 200ms and a standard deviation of 40ms. What is the probability of a job being completed in

i) Less than 180ms **(8 marks)**

ii) Between 170 and 210 ms **(8 marks)**

Question 2

You have been asked to compare the lifetime of similar batteries from two companies A and B. You pick a random sample of 50 batteries from each company and measure their lifetime. The following summary statistics were obtained for the variable lifetime. Lifetimes were measured in hours.

Company	Mean	SD	max	min	25th ile	50th ile	75th ile	n
A	213.7	20.0	255.1	177.2	195.3	216.7	229.1	50
B	200.6	17.5	241.2	156.1	192.1	198.7	213.1	50

You also obtained a bootstrapped confidence interval (6.09 to 20.20) using the percentile method for the difference in population means.

a) Draw an appropriate graph for the above data. **(5 marks)**

b) Explain how the bootstrapped interval is computed. **(15 marks)**

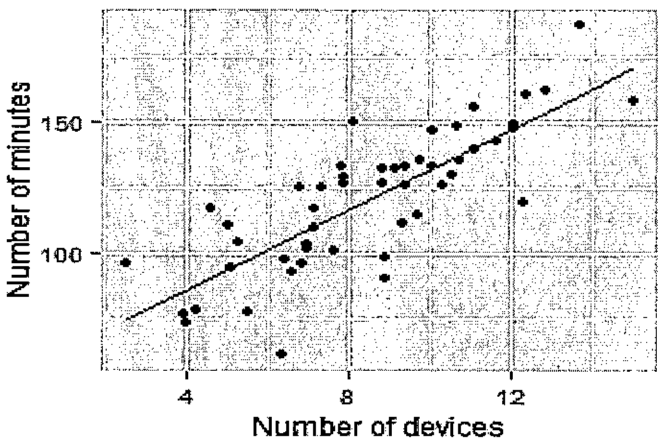
c) Interpret the interval. **(10 marks)**

d) How does this confidence interval relate to the classical hypothesis testing approach? **(10 marks)**

e) You have been asked for advice on sample size for a study with a similar design, explain how you would approach the problem **(10 marks)**

Question 3

A company that provides a preventative maintenance and repair for computing devices has carried out a study of the times taken on service calls. Data were gathered for the last 50 calls on the number of devices serviced and the total number of minutes taken. A plot of the data together with the output from a simple linear regression model are shown below.




```
lm(formula = Minutes ~ Devices)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.051  -9.297   0.039  10.861  32.744 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  55.3492     7.2846   7.598 8.96e-10 ***
Devices       7.6994     0.8147   9.451 1.56e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.82 on 48 degrees of freedom
Multiple R-squared:  0.6504,
```

a) Explain what a linear simple regression model is. **(7 marks)**

b) How is the line in the above graph derived? **(8 marks)**

c) Explain each of the figures in the Coefficients table. **(20 marks)**

d) Explain how the above can be used to predict how long it would take to service 10 devices. **(10 marks)**

e) In the context of simple linear regression, explain what a residual is? **(5 marks)**

Appendix for Section A

Poisson distribution P(x)=$e^{-\lambda}\frac{\lambda^x}{x!}$

Binomial distribution P$(x) = \binom{n}{x}p^x(1-p)^{n-x}$

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - hypothesised\ value}{SE(\overline{x}_1 - \overline{x}_2)}$$

$$SE(\overline{x}_1 - \overline{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Chi-square= $\chi^2 = \sum_{cells} \frac{(\boldsymbol{observed\ frequency - Expected\ frequency})^2}{\boldsymbol{Expected\ Frequency}}$

Proportion of Area to left of point for Standardised Normal Distribution

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50	0.50	0.51	0.51	0.52	0.52	0.52	0.53	0.53	0.54
0.1	0.54	0.54	0.55	0.55	0.56	0.56	0.56	0.57	0.57	0.58
0.2	0.58	0.58	0.59	0.59	0.59	0.60	0.60	0.61	0.61	0.61
0.3	0.62	0.62	0.63	0.63	0.63	0.64	0.64	0.64	0.65	0.65
0.4	0.66	0.66	0.66	0.67	0.67	0.67	0.68	0.68	0.68	0.69
0.5	0.69	0.69	0.70	0.70	0.71	0.71	0.71	0.72	0.72	0.72
0.6	0.73	0.73	0.73	0.74	0.74	0.74	0.75	0.75	0.75	0.75
0.7	0.76	0.76	0.76	0.77	0.77	0.77	0.78	0.78	0.78	0.79
0.8	0.79	0.79	0.79	0.80	0.80	0.80	0.81	0.81	0.81	0.81
0.9	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.84	0.84
1.0	0.84	0.84	0.85	0.85	0.85	0.85	0.86	0.86	0.86	0.86
1.1	0.86	0.87	0.87	0.87	0.87	0.87	0.88	0.88	0.88	0.88
1.2	0.88	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90	0.90
1.3	0.90	0.90	0.91	0.91	0.91	0.91	0.91	0.91	0.92	0.92
1.4	0.92	0.92	0.92	0.92	0.93	0.93	0.93	0.93	0.93	0.93
1.5	0.93	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
2.5	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995
2.6	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999
3.0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999