0.1 Information Theory

Question 61. (27 marks)

A source language has 5 symbols A, B, C, D and E. The associated probabilities of these symbols are given in the table below:

Symbol	Probability
A	0.60
B	0.30
C	0.05
D	0.03
E	0.02

  a. (5 marks) Calculate the entropy of the source language.

  b. (10 marks) Define a Huffman binary code for the source language.

  c. (3 marks) Calculate the efficiency of the code in (b) above.

  d. (2 marks) Calculate the redundancy of the code in (b) above.

  e. (2 marks) Briefly state what is meant by the Prefix Condition.

Question 62. (3 marks)

For each of the following codes state whether they are

  a. (1 mark) nonsingular

  b. (1 mark) uniquely decodable

  c. (1 mark) instantaneous

	Code 1	Code 2	Code 3
A	1	10	1
B	01	01	00
C	001	10	000

Question 63. (10 marks)

Consider a data source drawn from an alphabet (A, B, C, D) with probability distribution (0.3, 0.4, 0.2, 0.1).

  a. Derive a fixed length binary code and a Huffman code for the source.

  b. Comment on the difference in shape of a binary tree representing the fixed length code and the Huffman tree for the source.

  c. Discuss which code is closer to the optimal. Show all your work to justify your answer.

Confidence Intervals

- A researcher takes a random sample of 500 urban residents and finds that 122 have fibre-optic broadband access.

- Calculate a 90% Confidence Interval for the true percentage of residents who have fibre-optic broadband access.

Question 10.

a. (1 Mark) The observation of the air pressure at a volume of 5 cubic metres was 19.87 bars. Calculate the residual from the regression model corresponding to this observation.

b. (3 marks) Using the table above to justify your conclusion, test the null hypothesis that there is no monotonic (systematic) relationship between volume and pressure. State the null and alternative hypotheses clearly.

c. (2 marks) Briefly explain why the use of linear regression to describe pressure as a function of volume is inappropriate.

MA4413 Tutorials for Week 9 Q1. An IT competency test, used for staff recruitment, is devised so as to give a normal distribution of scores with a mean of 100. A random sample of 49 experienced IT users who are given the test achieve a mean score of 121 with a standard deviation of 14.

Perform a hypothesis test to assess whether this group of IT Users is unusual (i.e. have a different mean from the general population). Compute a 95

Q2. A claim has been made that the mean body temperature of healthy adults is equal to 98.6 degrees. A sample of 106 people has produced a mean body temperature of 98.2 degrees and a standard deviation of 0.62. Test the claim using a 0.05 significance.

Q3. A manufacturer of computer monitors has for many years used a process giving a mean life of 4700 hours and a standard deviation of 1460 hours. A new process is tried to see if it will increase the life significantly. A sample of 100 monitors gave a mean life of 5000 hours. Does the new process make a difference at the 5

Q4. In a study of store checkout scanners, 1234 items were checked and 20 of them were overcharges. Use a 0.05 significance level to test the claim that with scanners, 1

Q5. The quality control manager at the Telektronic Company considers the production of telephone answering machines to be out of control when the overall rate of defects exceeds 4Testing of a random sample of 150 machines revealed that 9 are defective. The production manager claims that production is not out of control and no corrective action is necessary. Use a 0.05 significance level to test the production managers claim.

Q6. In a study of company salaries, salaries paid by 2 different IT companies were randomly selected. For 40 Deltatech employees the mean is 23,870 and the standard deviation is 2960. For 35 Echelon employees , the mean is 22,025 and the standard deviation is 3065.

At the 0.05 level of significance, test the claim that Deltatech employees earn the same as their Echelon counterparts.

Q7 Does it pay to take preparatory courses for standardised tests such as the Comptia Exams? Using the sample data in the following table, compute the case-wise differences, the mean of the case-wise differences and the standard deviation of the case wise differences for the following data set.

```
Student A B C D E F G H I J
Score Before 700 840 830 860 840 690 830 1180 930 1070
Score After 720 840 820 900 870 700 800 1200 950 1080
```

1. A study was carried out in which researchers collected crime data. Of those convicted of arson, 50 were drinkers and 43 abstained. Of those convicted of fraud, 63 were drinkers and 144 abstained. Use a 0.01 level of significance to test the claim that the proportion of drinkers among convicted arsonists is greater than the proportion of drinkers convicted of fraud.

2. A bank is concerned about the amount of debt being accrued by customers using its credit cards. The board of directors voted to institute an expensive monitoring system if the mean for all the banks customers is greater than 2000. The bank randomly selected 50 credit-card holders and determined the amounts they charged. For this sample group, the mean is 2177 and the standard deviation is 1257. Using a 0.05 level of significance, test the claim that the mean amount charged is greater than 2000. Will the monitoring system be implemented?