# KABARAK UNIVERSITY

## UNIVERSITY EXAMINATIONS

### 2009/20010 ACADEMIC YEAR

## FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE

**COURSE CODE: COMP 328**

**COURSE TITLE: RESEARCH METHODS IN COMPUTER**

**STREAM: Y3S2**

**DAY: FRIDAY**

**TIME: 2.00 – 4.00 P.M.**

**DATE: 06/08/2010**

---

**INSTRUCTIONS:**

1. **Answer question ONE and any other two questions**
2. **Show all your working and be neat**

**PLEASE TURNOVER**

**QUESTION ONE (30 marks)**

a) Outline research procedure cycle one has to follow for a credible study as taught to you in this course **[5 marks]**

b) Assume you have conducted a research for your academic dissertation/project, briefly outline what should appear in each chapter of the university research booklet **[5 marks]**

c) Mention three types of sample survey designs normally used for collecting data. State the advantages and disadvantages of the design methods you have mentioned. **[5 marks]**

d) Two types of correlation were covered in this course can you describe each one of them and when they are used. **[ 5 marks]**

e) Industrial study was conducted by comparing tool wear when treated and untreated. Ten specimens were picked from both treated and untreated. Conduct a suitable hypothesis and draw the conclusion. Estimate a 95% confidence intervals also.( $t_{0.025,9}$)=2.262; $t_{0.025,18}$)=2.101)

**[10 marks]**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Untreated | 0.56 | 0.50 | 0.69 | 0.59 | 0.47 | 0.42 | 0.45 | 0.45 | 0.47 | 0.50 |
| Treated | 0.13 | 0.13 | 0.18 | 0.23 | 0.18 | 0.31 | 0.35 | 0.23 | 0.31 | 0.33 |

**QUESTION TWO (20 marks)**

a) The following are burning times of flares of two different types

| Type 1 | | Type 2 | |
|---|---|---|---|
| 65 | 82 | 64 | 56 |
| 81 | 67 | 71 | 69 |
| 57 | 59 | 83 | 74 |
| 66 | 75 | 59 | 82 |
| 82 | 70 | 65 | 79 |

Test the hypothesis of equal mean burning times. Use α=0.05 and tabulated $t_{(0.025,18)}$=2.101.**[5 marks]**

b) A new filtering device was installed in purifying water at KABU. Before installation, a random sample yielded the following information about the percentage of impurity: $\overline{y_1}$ =12.5, $S_1^2$=101.17 and $n_1$=8. After installation, a random sample yielded $\overline{y_2}$ =10.2, $S_2^2$=94.73, and $n_2$=9. Can you conclude that the two variances are equal? Has the filtering devices reduced the percentage of impurity significantly? (i.e. test $H_0$:$\mu_1=\mu_2$ against $H_1$:$\mu_1\neq\mu 2$). {$F_{0.05\ (7,8)=}$3.23}, $t_{(0.025,15)}$=2.131 **[10 marks]**

c) A study was conducted to determine performance of security forces by gender. Given the following data, is there any difference in the mean number of incidents handled by men and women police officers? Construct a 95% confidence interval of the differences between the gender means. Tabulated $t_{(0.025,37)}$=2.027.

| Men | Women |
|---|---|
| $\overline{X_1}=3.2$ | $\overline{X_2}=3.0$ |
| $S_1$=0.48 | $S_2$=0.62 |
| $n_1$=21 | $n_2$=18 |

**[5 marks]**

**QUESTION THREE (20 marks)**

a) Environmental causes of mortality (per 100,000 of males population) from 1958 to 1964 due to water-borne diseases in major towns in England and Wales were sampled. Average water hardness due to calcium level was sampled. About 36 towns in northern England were sampled and a sample of the results presented in the table below. Calculate Pearson correlation coefficient and test the appropriate hypothesis and comments about the result. $t_{(0.025,33)}=2.036$ **[10 marks]**

| Mortality (Y) | Water Hardness (X) | YX | $X^2$ | $Y^2$ |
|---|---|---|---|---|
| 1668 | 17 | 28356 | 289 | 2782224 |
| 1800 | 14 | 25200 | 196 | 3240000 |
| | | | | |
| . | . | . | . | . |
| . | . | . | . | . |
| 1627 | 20 | 32540 | 400 | 2647129 |
| 1378 | 71 | 97838 | 5041 | 1898884 |
| ∑Y=58580 | ∑X=1101 | ∑XY=1745214 | $\sum X^2$=56937 | $\sum Y^2$=96012380 |

b) The same data was collected from 25 towns in southern England as in (a) above. Calculate spearman's rank correlation and test the appropriate hypothesis and comments about the result. $t_{(0.025,24)}=2.064$ **[10 marks]**

| Mortality | Rank | Hardness | Rank | d | $d^2$ |
|---|---|---|---|---|---|
| 1466 | 9 | 5 | 24.5 | | |
| 1486 | 6.5 | 5 | 24.5 | | |
| 1625 | 2 | 13 | 23 | | |
| 1581 | 3 | 14 | 22 | | |
| 1260 | 18 | 21 | 20.5 | | |
| 1519 | 5 | 21 | 20.5 | | |
| 1257 | 20 | 50 | 19 | | |
| 1627 | 1 | 53 | 18 | | |
| 1309 | 15 | 59 | 17 | | |
| 1527 | 4 | 60 | 16 | | |
| 1369 | 12 | 68 | 15 | | |
| 1392 | 11 | 73 | 14 | | |
| 1299 | 17 | 78 | 12.5 | | |
| 1307 | 16 | 78 | 12.5 | | |
| 1485 | 8 | 81 | 11 | | |
| 1359 | 13 | 84 | 10 | | |
| 1456 | 10 | 90 | 9 | | |
| 1254 | 21 | 96 | 8 | | |
| 1236 | 23 | 101 | 7 | | |
| 1247 | 22 | 105 | 6 | | |
| 1175 | 24 | 107 | 5 | | |
| 1318 | 14 | 122 | 3.5 | | |
| 1486 | 6.5 | 122 | 3.5 | | |
| 1259 | 19 | 133 | 2 | | |
| 1096 | 25 | 138 | 1 | | |

**QUESTION FOUR (20 marks)**

a) A simple linear regression model is presented by the model shown below. Define each of its components as much as possible. **[4 marks]**

$Y_i = \alpha_i + \beta X_i + e_i$ ; i >2,

b) What hypotheses do we test in the simple linear regression? **[4 marks]**

c) Some of the advantages of using the methods indicated in conducting a survey. Fill in with "Yes" or "No"

| | Personal Interview | Group Interview | Postal Survey | Telephone Survey |
|---|---|---|---|---|
| Are sure that correct respondent is interviewed | | | | |
| Can be sure respondent is part of sampling frame | | | | |
| Subjects selected independently | | | | |
| Interviewers trained to deal with problems, obtain information for particular questionnaire | | | | |
| Can make explanations and answer questions | | | | |
| Can probe for details | | | | |
| Opportunity to motivate respondents | | | | |
| Opportunity to edit response | | | | |
| Opportunity to evaluate responses | | | | |
| Some control over respondent's task performance | | | | |
| Opportunity for direct quality control of interviewing/data entry process | | | | |
| Opportunity to reflect on responses | | | | |
| Data can be entered directly into computer | | | | |
| Can use visual aids | | | | |
| Possibility of recording errors by interviewer | | | | |
| Invasion of privacy | | | | |

**[12 marks]**

**QUESTION FIVE (20 marks)**

a) Define in a sample survey a sampling frame, respondents, enumerators, secondary and primary data

**[5 marks]**

b) You are required to interview KABU students and staff on their satisfaction on the use of the available computer facilities using simple random sampling design. Describe where you will get sampling frame for respondents and computers. **[5 marks]**

c) You have gone to Kericho district and wish to get information on the performance of tea industry by small scale and large scale farmers. Where will you source your information that will assist you apply random sampling method to pick respondents for interview? Describe briefly how you will go about creating a sampling frame where it does not exist and finally organizing a credible sample survey.

**[10 marks]**

You may use formulae relevant to your case

A (1- α)100 Confidence Interval

$$((\overline{x_1}-\overline{x_2})-ts_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \text{ to } ((\overline{x_1}-\overline{x_2})+ts_p\sqrt{\frac{1}{n_1}+\frac{2}{n_2}}$$

$$t_0=\frac{(\overline{x_1}-\overline{x_2})}{\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}}$$

$${S_p}^2=\frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2}$$

$$\rho=1-\frac{6\sum d^2}{n(n^2-1)}$$

$$t_0=\rho_c\sqrt{\frac{n-2}{1-\rho^2}}$$

$$t_0=\hat{r}_c\sqrt{\frac{n-2}{1-\hat{r}^2}}$$

$$r=\frac{\sum X_iY_i-\frac{(\sum X_i)(\sum Y_i)}{n}}{(\sqrt{(\sum X^2-\frac{(\sum X)^2}{n})(\sum Y^2-\frac{(\sum Y)^2}{n})}}$$

$$S_d=\left[\frac{\sum_{i=1}{d_i}^2-\frac{1}{n}\left(\sum_{i=1}d_i\right)^2}{n-1}\right]^{1/2}$$

$$t_0=\frac{\overline{d}}{s_d/\sqrt{n}}$$