



## KENYATTA UNIVERSITY

### UNIVERSITY EXAMINATIONS 2011/2012

#### SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE

#### SMA 364: APPLIED STATISTICAL METHODS

**DATE:** Monday 26<sup>th</sup> March, 2012

**TIME:** 8.00 a.m. – 10.00 a.m.

---

#### **INSTRUCTIONS**

#### **Question 1 (30 marks)**

- (a) A straight line that passes through some data points is given by a linear regression model

$$Y_i = \beta_0 + \beta_1 (x_i - \bar{x}) + e_i, \quad i = 1, 2, \dots, n$$

- (i) State the usual assumptions on the error term,  $e_i$ , to allow for least squares estimator and inference on the parameters. [2 marks]
- (ii) Determine the least squares estimators of  $\beta_0$  and  $\beta_1$ . [4 marks]
- (iii) Show that the least squares estimators obtained in (ii) are unbiased for the respective parameters. [6 marks]
- (iv) Show that  $\frac{1}{n-2} \left\{ \sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \right\}$  is unbiased for  $\sigma^2$ . [10 marks]
- (b) Form an Analysis of Variance (ANOVA) table for the model in part (a), giving the degrees of freedom for each line, and give a Test of significance of level and for the overall regression based on the components of the ANOVA table. [3 marks]
- (c) In an experiment to compare three diets on a group of cows, both the weight gain  $Y$  and the initial weight (covariate)  $X$ , of each cow is available. For each diet four cows are assigned at random. Write down a statistical model for analysing the data in this situation. Remember to define each component of the model and the necessary assumptions. [5 marks]

**Question 2 (20 marks)**

- (a) A psychiatrist wanted to know whether the level of pathology (Y) in psychotic patients six months after treatment could be predicted with reasonable accuracy from knowledge of pre-treatment symptom ratings of thinking disturbance ( $x_1$ ) and hostile suspiciousness ( $x_2$ ). A computer output is given below:

| Source                    | d.f | ss     |
|---------------------------|-----|--------|
| Regression ( $x_1, x_2$ ) | 2   | 2,784  |
| Residual                  | 250 | 11,008 |

| Source    | d.f. | Ss     | Source    | d.f. | ss     |
|-----------|------|--------|-----------|------|--------|
| $x_1$     | 1    | 1,546  | $x_2$     | 1    | 161    |
| $x_2/x_1$ | 1    | 1,238  | $x_1/x_2$ | 1    | 2,623  |
| Residual  | 50   | 11,008 | Residual  | 50   | 11,008 |

- (i) Carry out the overall regression significance at 0.05 level. Compute  $R^2$  and interpret in light of the data. [3 marks]
- (ii) Provide variable-added in-order test for both variables with thinking disturbance added first use 0.05 level of significance. [6 marks]
- (b) In a simple regression of variable Y on variable X the following information is available.  
 $n = 8$ ,  $\sum x_i = 22$ ,  $\sum x_i^2 = 71$ ,  $\sum Y_i = 248$ ,  $\sum Y_i^2 = 7800$ ,  $\sum X_i Y_i = 707$
- (i) Determine the least squares regression line of Y on X for translated model. [3 marks]
- (ii) Construct an Analysis of Variance (ANOVA) table. [5 marks]
- (iii) Construct a 90% prediction interval (PI) for Y when  $x = 1.5$ . [3 marks]

**Question 3 (20 marks)**

- (a) Give some three reasons that would require the use of Analysis of Covariance (ANACOVA). [3 marks]
- (b) In an experiment involving three treatments (A,B,C), the response variable (Y) and a covariate variable (X) were recorded for 15 subjects. The sums of square and cross-products table was obtained as follows:

Sums of squares and cross-products

| Source of Variation | d.f. | xx     | xy     | yy     |
|---------------------|------|--------|--------|--------|
| Treatment           | 2    | 66.13  | 96.00  | 140.40 |
| Error               | 12   | 195.60 | 186.60 | 206.00 |
| Total               | 14   | 261.73 | 282.60 | 346.40 |

- (i) Form the Analysis of Covariance (ANACOVA) table adjusted for regression and complete the analysis (use  $\alpha = 0.10$ ).

[Show clearly how you compute the necessary components of the table]

[12 marks]

- (ii) Estimate the regression coefficient  $\beta$  and test the hypothesis  $H_0: \beta = 0$  against  $H_1: \beta \neq 0$  at  $\alpha = 0.05$  level of significance. What do you conclude? [5 marks]

**Question 4 – (20 marks)**

The response time (in milliseconds) was determined for three different brands of circuit breakers as follows:

| Brand             | Time                   |
|-------------------|------------------------|
| Northern Electric | 9, 12, 10, 8, 5        |
| General Electric  | 20, 21, 23, 17, 30, 23 |
| Pacific Electric  | 6, 5, 8, 16            |

- (a) If the aim is to compare the response times of the three brands,
- (i) Give the statistical model to be used in the analysis together with necessary assumptions. [3 marks]
- (ii) State clearly the null and alternative hypothesis. [2 marks]
- (b) Compute the necessary components of the ANOVA table, showing your work clearly in each case; hence form a complete ANOVA table. [9 marks]
- (c) Carry out the test for the hypothesis in part (a) at  $\alpha = 0.01$  and state your conclusion. Indicate how you would determine which brand differ significantly.

**Question 5 – (20 marks)**

- (a) A city highway utilizing four lanes in each direction was studied to see whether drivers preferred to drive on the inside lanes. A total of 1,000 automobiles were observed during the heavy early morning traffic and their respective lanes recorded. The results are as shown in the table

|                |     |     |     |     |
|----------------|-----|-----|-----|-----|
| Lane           | 1   | 2   | 3   | 4   |
| Observed count | 294 | 276 | 238 | 192 |

- (i) State the appropriate null hypotheses and alternative hypothesis for testing where there is preference for some lanes. [2 marks]
- (ii) Carry out the test for the hypothesis in part (i) using 0.05 level of significance and state your conclusion. [4 marks]
- (b) A printer is interested in examining the relationship between the number of printing errors and the font size used. He selects 3 different books recently printed by his company, each printed using a different font size. For each book he records the number of error-free pages. The results are shown in the table.

|                      | Font | Size |     |
|----------------------|------|------|-----|
|                      | A    | B    | C   |
| Pages with errors    | 23   | 17   | 41  |
| Pages without errors | 241  | 183  | 210 |

- (i) State the appropriate null and alternative hypotheses for this type of problem. [2 marks]
- (ii) Carry out the necessary test at 0.05 level of significance. [5 marks]
- (c) A study of purchase decision for 3 stock portfolio managers, A, B and C was conducted to compare the rates of stock purchases that resulted in profits over a time period that was less than or equal to one year. One hundred randomly selected purchases obtained for each of the managers showed the results given in the table:

| Result                       | Manage |     |     |
|------------------------------|--------|-----|-----|
|                              | A      | B   | C   |
| Purchases show profit        | 63     | 71  | 55  |
| Purchases do not show profit | 37     | 29  | 45  |
| Total                        | 100    | 100 | 100 |

- (i) State the appropriate null and alternative hypotheses for this situation. [2 marks]
- (ii) Carry out the test for the hypothesis in (i) at 0.05 level of significance and state your conclusion. [5 marks]

.....