**NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**FACULTY OF HEALTH, NATURAL RESOURCES AND APPLIED SCIENCES**

**SCHOOL OF AGRICULTURE AND NATURAL RESOURCES SCIENCES**

**DEPARTMENT OF AGRICULTURAL SCIENCES AND AGRIBUSINESS**

| **QUALIFICATION: BACHELOR OF SCIENCE IN AGRICULTURE (AGRIBUSINESS MANAGEMENT)** | |
|---|---|
| **QUALIFICATION CODE: 07BAGA** | **LEVEL:** 7 |
| **COURSE CODE:** BEA621S | **COURSE NAME:** BASIC ECONOMETRICS FOR AGRICULTURE |
| **DATE:** JANUARY 2025 | **PAPER** 2 |
| **DURATION:** 3 HOURS | **MARKS:** 100 |

| **SECOND OPPORTUNITY/SUPPLEMENTARY EXAMINATION QUESTION PAPER** | |
|---|---|
| **EXAMINER(S)** | Prof David Uchezuba |
| **MODERATOR:** | Mr Mwala Lubinda |

| **INSTRUCTIONS** |
|---|
| 1. Answer ALL the questions.<br>2. Write clearly and neatly.<br>3. Number the answers. |

**PERMISSIBLE MATERIALS**

1. Examination question paper
2. Answering book

**THIS QUESTION PAPER CONSISTS OF 12 PAGES** (Excluding this front page)

**SECTION 1 MULTIPLE CHOICE (20 MARKS)**

**Question 1**

The relationship between a farmer's consumption expenditure (Y) and income (X) is expressed as follows, $E(y/x_i) = f(x_i)$.Which of the following statements about the equation is **incorrect**?

A) It is known as the conditional expectation function

B) It is known as the population regression function

C) It is known as the population regression

D) It is known as the population distribution function

**Question 2**

Which of the following is **incorrect** about the interpretation of the equation $y_i = \beta_1 + \beta_2 x_1 + \beta_2 x_2 + \mu_i$ ?

A) The expected mean value of $y$ is conditionally related to $x_i$

B) The values of $x$ are unobservable

C) The expected mean value or average response of $Y$ varies with $X$

D) The equation is a linear function

**Question 3**

If an estimable model is given as, $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\mu}_i$. Which of the following statements is **incorrect**?

A) The equation is a sample regression function

B) The $\hat{\beta}_2$ is the estimator for $\beta_2$

C) The value of $\hat{\beta}_2$ is known as the parameter estimate

D) The value of $\hat{\beta}_2$ cannot be negative

**Question 4**

Which of the following parameters in the equation $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\mu}_i$ can be calculated using the formula? $\frac{\sum x_i y_i}{\sum x_i^2}$, where $x_i = (X - \overline{X})$ and $y_i = (Y - \overline{Y})$.

A) $\hat{\beta}_2$

B) $\hat{\beta}_1$

C) $\hat{Y}_i$

D) $\hat{\mu}_i$

**Question 5**

The function $y = f(x)$ is said to be a linear function of ($x$), which of the following statements is **incorrect** about this linear function?

A) $x$ must appear with a power or index of 1 only.

B) $x$ must not be multiplied or divided by any other variable

C) The rate of change of $y$ with respect to $x$ must be independent of the value of $x$

D) $x$ can appear as a square root ($\sqrt{x}$)

**Question 6**

Consider the following models

$E(y/x_i) = \beta_1 + \beta_2 x_i^2 + \mu_i$ ............................................................................ (A)

$E(y/x_i) = \beta_1 + \beta_2^2 x_i + \mu_i$ ............................................................................ (B)

Which of the following statements about equations (A) and (B) is **incorrect**?

A) Equation (A) is linear in parameter and (B) is non-linear in parameter

B) Equation (A) is linear in variable and (B) is non-linear in variable

C) Equation (A) is non-linear in variable and (B) is linear in variable

D) Equation (A) is linear in parameter and (B) linear in variable

**Question 7**

The farmer's consumption function is fitted as

$y_i = \beta_1 + \beta_2 x_i + \mu_i$

Which of the following is **INCORRECT** about why $\mu_i$ was included in the model?

A) We do not know other variables affecting consumption expenditure ( $y$)

B) Even if we know, we may not have information (data) about all factors affecting ( $y$)

C) There may be measurement errors in the way data was collected

D) We include $\mu_i$ because it is a non-random and systematic component of the model

**Question 8**

According to the Gauss-Markov theorem, which of the following statements is **NOT CORRECT**? An estimator says the ordinary least square (OLS) estimator $\hat{\beta}_2$, is said to be the best linear unbiased estimator of $\beta_2$, if the following conditions hold.

A) $\hat{\beta}_2$, must be a linear function of the dependent variable ( $y$)

B) $\hat{\beta}_2$, must be unbiased, i.e, its average or expected value $E(\hat{\beta}_2)$ must be equal to $\beta_2$

C) $\hat{\beta}_2$, must have minimum variance

D) $\hat{\beta}_2$, must have a mean of zero

**Question 9**

An unbiased estimator such as $\hat{\beta}_2$, one with the least (minimum) variance is said to be

A) An inefficient estimator

B) An efficient estimator

C) A random noise

D) An asymptote

**Question 10**

Consider the following regression model estimated using the OLS method.

$$Y = \begin{matrix} 463.5136 - 0.3901x_1 + 0.17925x_2 \\ (91.2835) \quad (0.1213) \quad (0.0477) \end{matrix}$$

(Standard errors are in parenthesis)

Using equation (12.1), calculate the t-statistic for the $x_1$ and $x_2$ variables

A) 3.2159 and 3.710

B) 3.2801 and 3.7578

C) 3.2159 and 3.7578

D) 3.2009 and 3.7011

**Question 11**

Which one of the following **incorrectly** defines the coefficient of correlation between variables?

A. Its value is between -1 and +1.

B. It can be positive or negative

C. It is a measure of association

D. It is the same as $R^2$

**Question 12**

The statistical significance of a parameter in a regression model refers to:

a) The conclusion of testing the null hypothesis that the parameter is equal to zero, against the alternative that it is non-zero.

b) The probability that the OLS estimate of this parameter is equal to zero.

c) The interpretation of the sign (positive or negative) of this parameter.

d) All of the above

**Question 13**

All of the following are possible effects of multicollinearity EXCEPT:

a) the variances of regression coefficient estimators may be larger than expected

b) the signs of the regression coefficients may be opposite of what is expected

c) a significant F ratio may result even though the t ratios are not significant

d) removal of one data point may cause large changes in the coefficient estimates

**Question 14**

Suppose that you estimate the model $Y = 50 + B_1X + u$. You calculate residuals and find that the explained sum of squares is 400 and the total sum of squares is 1200. The R-squared is

a) 0.25

b) 0.33

c) 0.5

d) 0.67

**Question 15**

In linear regression, the assumption of homoscedasticity is needed for

I. unbiasedness

II simple calculation of variance and standard errors of coefficient estimates.

III. The claim that the OLS estimator is BLUE.

a) I only.

b) II only.

c) III only.

d) II and III only.

**Question 16**

Which of the following is/are consequences of over-specifying a model (including irrelevant variables on the right—hand side)?

I. The variance of the estimators may increase.

II. The variance of the estimators may stay the same.

III. Bias of the estimators may increase.

a) I only.

b) II only.

c) III only.

d) I and II only.

**Question 17**

Heteroscedasticity means that

a) Homogeneity cannot be assumed automatically for the model.

b) the observed units have different preferences.

c) the variance of the error term is not constant.

d) agents are not all rational.

**Question 18**

By including another variable in the regression, you will

a) look at the t-statistic of the coefficient of that variable and include the variable only if the coefficient is statistically significant at the 1% level.

b) eliminate the possibility of omitted variable bias from excluding that variable.

c) decrease the regression R2 if that variable is important.

d) decrease the variance of the estimator of the coefficients of interest.

**Question 19**

Which of the following statements is TRUE concerning OLS estimation?

a) OLS minimises the sum of the vertical distances from the points to the line

b) OLS minimises the sum of the squares of the vertical distances from the points to the line

c) OLS minimises the sum of the horizontal distances from the points to the line

(d) OLS minimises the sum of the squares of the horizontal distances from the points to the line.

**Question 20**

The residual from a standard regression model is defined as

a) The difference between the actual value, y, and the mean, y-bar

b) The difference between the fitted value, y-hat, and the mean, y-bar

c) The difference between the actual value, y, and the fitted value, y-hat

d) The square of the difference between the fitted value, y-hat, and the mean, y-bar

**SECTION 2 TRUE OR FALSE QUESTIONS (20 Marks)**

Assume a regression model of consumption $y = \beta_1 + \beta_2 x + u$.

**I).** The *t*-test of significance requires that the sampling distributions of estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ follow the normal distribution. True or False

**II).** Even though the disturbance term in the CLRM is not normally distributed, the OLS estimators are still unbiased. True or False

**III).** If there is no intercept in the regression model, the estimated $u_i (= \hat{u}_i)$ will not sum to zero. True or False

**IV).** The p-value and the size of a test statistic mean the same thing. True or False

**V).** In a regression model that contains the intercept, the sum of the residuals is always zero. True or False.

**VI).** If a null hypothesis is not rejected, it is true. True or False.

**VII).** The higher the value of $\sigma^2$, the larger the variance of $\hat{\beta}_2$. True or False.

**VIII).** The conditional and unconditional means of a random variable are the same things. True or False

**IX)** In the two-variable PRF, if the slope coefficient $\beta_2$ is zero, the intercept $\beta_1$ is estimated by the sample mean $\bar{y}$. True or False

**X).** The conditional variance, $var(y_i|x_i) = \sigma^2$, and the unconditional variance of *Y*, $var\ y = \sigma^2 . y$, will be the same if *X* did not influence *Y*. True or False

**SECTION 3 ESSAY-TYPE QUESTIONS (60 Marks)**

**Question 1 (20 Marks)**

1.1. The average salary of thirteen workers and their level of education was analyzed to determine whether education can explain the variation in salary earned using an ordinary least square method represented by the following model, $y = \beta_1 + \beta_2 x + u$.

Where y = salary earned (million $) x = years of education and u = error term. The sum of the squares of the analysis is given as follows.

$$\sum_{i=1}^{n} y = 112.771 \quad \sum_{i=1}^{n} x = 156, \quad \sum_{i=1}^{n} y_i x_i = 131.7856, \quad \sum_{i=1}^{n} x^2 = 182, \quad \sum_{i=1}^{n} \hat{u}^2 = 9.6928,$$

$$\sum_{i=1}^{n} (y - \bar{y})^2 = 105.12$$

Calculate the value of the following

| | | |
|---|---|---|
| 1.1.1 | The slope parameter | (2 Marks) |
| 1.1.2. | The intercept parameter | (2 Marks) |
| 1.1.3. | The variance of the regression | (2 Marks) |
| 1.1.4. | The variance of slope parameter | (2 Marks) |
| 1.1.5. | The explained sum of squares | (2 Marks) |
| 1.1.6. | The total sum of squares | (2 Marks) |
| 1.1.7. | The coefficient of determination | (2 Marks) |

| | | |
|---|---|---|
| 1.2 . | Fit the equation of the regression line | (2 Marks) |
| 1.3. | Predict the mean inventory value $\hat{y}$ for x = 10.90 | (2 Marks) |
| 1.4. | Interpret the value of the slope parameter | (2 Mrks) |

**Question 2 (20 Marks)**

1.1. Use the output of the regression in question 1.1. to set up an Analysis of Variance (ANOVA) table using the format in Table 1. (10 Marks)

Table 1

| ANOVA TABLE | | | | | |
|---|---|---|---|---|---|
| **Source of variation** | **df** | **SS** | **MSS** | **F-stat** | **Prob** |
| ESS | ( ) | ( ) | ( ) | ( ) | ( |
| RSS | ( ) | ( ) | ( ) | | |
| TSS | ( ) | ( ) | | | |

Note: ESS = Explained sum of squares, RSS = Residual Sum of Squares, TSS = Total sum of squares.

1.2. Determine the 95 % confidence interval for the true value of the salary of workers given their education level. (5 Marks)

1.3. A post-regression diagnostic shows the following distribution of the data, skewness = -0.3937, Kurtosis = 2.0462. Calculate the Jarque-Bera statistics (2 Marks)

1.4. What is the optimal value of the distributions of this test?

1.4.1 Skewness (1 Mark)

1.4.2. Kurtosis (1 Mark)

1.5. What is the null hypothesis of this test? (1 Mark)

**Question 3 (20 Marks)**

3.1. What are the consequences of the following violation of a classical linear regression model?

3.1.1 Non-linearity in regression parameter (2 Marks)

3.1.2 Stochastic regressors in a regression model (2 marks)

3.1.3. Non-zero mean of the disturbance term (2 marks)

3.1.4. Heteroscedasticity in the model (2 marks)

3.1.5. Auto-correlated disturbances (2 marks)

3.1.6. Sample observation is less than the number of regressors (2 marks)

3.1.7. Insufficient variability in regressors (2 marks)

3.1.8. Multicollinearity (2 marks)

3.1.9. Specification bias (2 marks)

3.1.10. Non-normality of disturbances (2 marks)

**Formulas and statistical tables**

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \qquad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2} \qquad Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}, \text{ or } \qquad R^2 = \hat{\beta}_2^2 \left(\frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} y_i^2}\right), \text{ or } \qquad R^2 = \left(\frac{(\sum_{i=1}^{N} x_i y_i)^2}{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}\right)$$

$$R^2 = \left(\frac{(\sum_{i=1}^{N} x_i y_i)^2}{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}\right) \qquad se(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} x_i^2}} = \frac{\sqrt{\text{var}(\hat{\beta}_2)}}{\sqrt{\sum_{i=1}^{N} x_i^2}} \qquad t = \frac{\hat{\beta}_2 - \beta_2}{se(\beta_2)}$$

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \qquad \hat{\beta}_2 = \frac{\sum_{i=1}^{N} (x_i y_i)^2}{\sum_{i=1}^{N} x_i^2} \qquad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 X$$

$$Pr\left(\hat{\beta}_2 - t_{n-2,\alpha/2} Se(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{n-2,\alpha/2} Se(\hat{\beta}_2)\right) \qquad ESS = \hat{\beta}_2^2 \sum_i x_i^2$$

**T-distribution table**

| df \ Pr | 0.25<br>0.50 | 0.10<br>0.20 | 0.05<br>0.10 | 0.025<br>0.05 | 0.01<br>0.02 | 0.005<br>0.010 | 0.001<br>0.002 |
|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.31 |
| 2 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 |
| 3 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.214 |
| 4 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 |
| 5 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 |
| 6 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 |
| 7 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 |
| 8 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 |
| 9 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 |
| 10 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 |
| 11 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 |
| 12 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 |
| 13 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 |
| 14 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 |
| 15 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 |
| 16 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 |
| 17 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 |
| 18 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 |
| 19 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 |
| 20 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 |
| 21 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 |
| 22 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 |
| 23 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 |
| 24 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 |
| 25 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 |
| 26 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 |
| 27 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 |
| 28 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 |
| 29 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 |
| 30 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 |
| 40 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 |
| 60 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 |
| 120 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 |
| $\infty$ | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 |