**Murang'a University College**
*(A Constituent College of Jomo Kenyatta University of Agriculture and Technology)*
*University Examination*
**School of Pure and Applied Sciences**
**Supplementary/Special Examination for the degree of;**
**Bachelor of Science in Mathematics & Computer Science,**
**Bachelor of Science in Mathematics & Economics**
**AMS 2101/SMA 2103 PROBABILITY & STATISTICS I.**

Date: July/August 2016 2 Hours
INSTRUCTIONS: Attempt Question **One** and any other **Two** Questions.

## Question One (30 Marks)

(a) Find the constant $c$ such that the function

$$f(x) = \begin{cases} cx^2 & \text{for} \quad 0 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

is a density function. **(5 Marks)**

(b) Find the mean for a random variable X defined by

$$X = \begin{cases} 2 & \text{probability} \quad 1/3 \\ -1 & \text{probability} \quad 2/3. \end{cases}$$

**(3 Marks)**

(c) Suppose that a game is to be played with a single die assumed fair. In this game a player wins \$20 if a 2 turns up, \$40 if a 4 turns up; loses \$30 if a 6 turns up; while the player neither wins nor loses if any other face turns up. Find the expected sum of money to be won. **(5 Marks)**

(d) Consider the random variable $X$ with probabilities

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P[X = x]$ | 0.05 | 0.10 | 0.20 | 0.40 | 0.15 | 0.10 |

Compute the variance of $X$. **(5 Marks)**

(e) Derive an expression for $\widehat{\beta}_0$ from the linear model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where the error terms $\epsilon_{i's} \sim N(0, \sigma^2)$. **(5 Marks)**

(f) Define the following terms:

(i) Statistic. **(1 Mark)**

(ii) Parameter. **(1 Mark)**

(g) In an attempt to determine the relationship between the daily mid-day temperature (measured in degrees Celsius) and the number of defective parts produced during that day, a company recorded the data presented in the table below. For this data set, $x_i$ represents the temperature in degrees Celsius and $y_i$ the number of defective parts produced on day $i$.

| Day | Temperature | Number of Defects |
|---|---|---|
| 1 | 24.2 | 25 |
| 2 | 22.7 | 31 |
| 3 | 30.5 | 36 |
| 4 | 28.6 | 33 |
| 5 | 25.5 | 19 |
| 6 | 32.0 | 24 |
| 7 | 28.6 | 27 |
| 8 | 26.5 | 25 |
| 9 | 25.3 | 16 |
| 10 | 26.0 | 14 |
| 11 | 24.4 | 22 |
| 12 | 24.8 | 23 |
| 13 | 20.6 | 20 |
| 14 | 25.1 | 25 |
| 15 | 21.4 | 25 |
| 16 | 23.7 | 23 |
| 17 | 23.9 | 27 |
| 18 | 25.2 | 30 |
| 19 | 27.4 | 33 |
| 20 | 28.3 | 32 |
| 21 | 28.8 | 35 |
| 22 | 26.6 | 24 |

Find the sample correlation coefficient for the data presented in the table above. **(5 Marks)**

**Question Two (20 Marks)**

a) (i) Prove that $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$. **(4 Marks)**

(ii) When fitting the simple linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ to a set of data using the least squares method, prove that the sum of the ordinary least squares residuals is zero. **(6 Marks)**

(b) Using the data in the table below

| Row | Minutes | Units | Row | Minutes | Units |
|---|---|---|---|---|---|
| 1 | 23 | 1 | 8 | 97 | 6 |
| 2 | 29 | 2 | 9 | 109 | 7 |
| 3 | 49 | 3 | 10 | 119 | 8 |
| 4 | 64 | 4 | 11 | 149 | 9 |
| 5 | 74 | 4 | 12 | 145 | 9 |
| 6 | 87 | 5 | 13 | 154 | 10 |
| 7 | 96 | 6 | 14 | 166 | 10 |

compute $\text{Cor}(Y, X)$. **(10 Marks)**

**Question Three (20 Marks)**

(a) Let $Y$ and $X$ denote the labour force participation rate of women in 1972 and 1968, respectively, in each of 19 cities in the United States. The regression output for this data set is shown in the table below. It was also found that SSR $= 0.0358$ and SSE $= 0.0544$. Suppose that the model $Y = \beta_O + \beta_1 X + \varepsilon$ satisfies the usual regression assumptions.

| Variable | Coefficient |
|---|---|
| Constant | 0.203311 |
| $X$ | 0.656040 |
| $n$ | 19 |

(i) Compute $\text{Var}(Y)$. **(4 Marks)**

(ii) Suppose that the participation rate of women in 1968 in a given city is 45%. What is the estimated participation rate of women in 1972 for the same city? **(3 Marks)**

(iii) Compute $R^2$, where $R = \text{Cor}(Y, X)$. **(3 Marks)**

(b) The following are the activity values (micromoles per minute per gram of tissue) of a certain enzyme measured in the normal gastric tissue of 35 patients with gastric carcinoma:

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.360 | 1.189 | 0.614 | 0.788 | 0.273 | 2.464 | 0.571 |
| 1.827 | 0.537 | 0.374 | 0.449 | 0.262 | 0.448 | 0.971 |
| 0.372 | 0.898 | 0.411 | 0.348 | 1.925 | 0.550 | 0.622 |
| 0.610 | 0.319 | 0.406 | 0.413 | 0.767 | 0.385 | 0.674 |
| 0.521 | 0.603 | 0.533 | 0.662 | 1.177 | 0.307 | 1.499 |

Calculate the

(i) mean $\overline{x}$ **(3 Marks)**

(ii) variance $s^2$, and **(4 Marks)**

(iii) standard deviation $s$. **(3 Marks)**

**Question Four (20 Marks)**

(a) (i) The probability is 0.038 that a person reached on a "cold call" by a telemarketer will make a purchase. If the telemarketer calls 40 people, what is the probability that at least one sale will result? **(5 Marks)**

(ii) A random variable $X$ has a density function given by

$$f(x) = \begin{cases} 2e^{-2x} & \text{for} \quad x \geq 0 \\ 0 & x < 0. \end{cases}$$

Find the moment generating function about the origin.

**(5 Marks)**

(b) Let $X$ be a random variable with pdf $f(x) = kx^2$ where $0 \leq x \leq 1$.

(i) Find the value of the constant $k$. **(3 Marks)**

(ii) Find $E(X)$ and $\mathrm{Var}(X)$. **(7 Marks)**

**Question Five (20 Marks)**

(a) Given the random variable $X$ whose probability density function is given by

$$f(x) = \begin{cases} e^{-x} & \text{for} \quad x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

find the coefficient of skewness for this distribution. **(10 Marks)**

(b) Let $X$ be a random variable that can take on the values 2, 1, and 3 with respective probabilities 1/3, 1/6, and 1/2. Find

(i) the mean, **(5 Marks)**

(ii) the third moment about the mean. **(5 Marks)**