MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS

2022/ 2023 ACADEMIC YEAR

FIRST YEAR FIRST SEMESTER EXAMINATIONS FOR THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION

MBA 8104: QUANTITATIVE METHODS OF MANAGEMENT

DATE: APRIL 2023 **TIME: 3 HOURS**

INSTRUCTIONS: Answer Question ONE [Compulsory] and Any Other Two Questions

Question One [40 Marks]

a) For a study of the effect of work hours on drinking, suppose that if we estimate the simple linear regression model we obtain the following Statistical output:

```
Source     | SS           df       MS                    Number of obs = 7489
-------------+----------------------------------         F( 1, 7487) = 26.51
Model      | 20239.281    1        20239.281             Prob > F = 0.0000
Residual   | 5716797.56   7487     763.563185            R-squared = 0.0035
-------------+----------------------------------         Adj R-squared = 0.0034
Total      | 5737036.84   7488    766.164109             Root MSE = 27.633
------------------------------------------------------------------------------
drmo       | Coef.        Std. Err.      t       P>|t|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
hwk        | .0809038     .0157143       5.15    0.000   0.0500995      .1117082
_cons      | 8.92635      .619701        14.40   0.000   7.711562       10.14114
------------------------------------------------------------------------------
```

Suppose also that if we estimate the multiple linear regression model (excluding income), we obtain the following Statistical output:

```
Source     | SS           df       MS                    Number of obs = 7488
-------------+----------------------------------         F( 7, 7480) = 52.17
Model      | 266836.037   7       38119.4338             Prob > F = 0.0000
Residual   | 5465529.92   7480    730.685819             R-squared = 0.0465
-------------+----------------------------------          Adj R-squared = 0.0457
Total      | 5732365.96   7487    765.642575             Root MSE = 27.031
------------------------------------------------------------------------------
drmo       | Coef.        Std. Err.      t       P>|t|   [95% Conf. Interval]
-------------+----------------------------------------------------------------
hwk        | .0302414     .0161589       1.87    0.061   -.0014345      .0619174
age        | -.3381229    .1397713      -2.42    0.016   -.612114       -.0641318
educ       | -.2849688    .1265659      -2.25    0.024   -.5330735      -.0368641
black      | -3.643506    .7159052      -5.09    0.000   -5.046882      -2.240131
other      | -3.100133    1.358236      -2.28    0.022   -5.762659      -.4376079
married    | -3.096517    .6613095      -4.68    0.000   -4.392869      -1.800164
female     | -10.73453    .6449336     -16.64    0.000   -11.99878      -9.470277
_cons      | 38.07319      6.498279      5.86    0.000   25.33474       50.81164
------------------------------------------------------------------------------
```

i. Is hwk statistically significant at the 5% level, 10% in the simple regression? What about in the multiple regressions? (3 marks)

ii. Explain in words exactly what the coefficient estimate for hwk in the multiple regression model is telling us. (3 marks)

iii. Does it appear that our estimator for the effect of work hours on drinking is biased if we do not include any control variables? (3 marks)

b) The data below shows the daily number of loan applications at a microfinance institution over a period of 100 days. Using the first class as 40 - 49, group the data and present its frequency distribution.

45	63	60	45	57	73	67	63	65	61
50	45	55	70	56	47	64	54	69	51
51	60	48	69	48	60	60	56	48	57
55	57	62	71	50	70	68	65	53	65
56	74	57	65	63	68	64	48	52	72
64	70	58	57	55	62	53	71	53	60
61	71	62	48	56	54	60	58	63	59
59	40	61	41	44	60	63	67	63	66
55	56	63	53	78	73	73	66	60	59
65	49	65	65	57	67	61	76	53	68

(7 marks)

c) Given the following data

Class	21 - 27	28 - 34	35 - 41	42 - 48	49 - 55	56 - 62	63 - 69
Frequency	2	9	11	17	15	10	6

Compute the

i. Mean (2 marks)

ii. Median (2 marks)

iii. Mode (2 marks)

iv. Standard deviation (3 marks)

v. Pearson's first and second skewness coefficient and comment on your results (5 marks)

d) A production plant has two fabricating systems: one uses automated equipment, the other is manually operated. Since the automated system costs more to install, we wish to know whether it provides increased production in terms of the mean

number of finished products fabricated per day. Consider the accompanying data below on the production for 32 days.

Automated	56	36	67	88	86	97	18	57	84
	28	48	28	25	42	47	26	51	
Manual	42	22	28	22	37	52	21	44	18
	22	42	37	16	51	41			

Test the hypothesis that the daily mean of production for the automated system is higher than the manual one. (6 marks)

e) A researcher reports that the average salary of assistant professors is more than $42,000. A sample of 30 assistant professors has a mean salary of $43,260. At α = 0.05, test the claim that assistant professors earn more than $42,000 a year. The standard deviation of the population is $5230 (4 marks)

Question Two (20 Marks)

a) The weight of a certain product has a mean weight of 38kg. An observation was made that in products of a given firm were underweight. As evidence, officials sampled 20 of such products from the firm. Test an appropriate hypothesis at $\alpha = 0.05$.

31.9	42.3	25.2	37.4	40.4	31.1	29.2	31.7	31.6	46.4
42.8	40.5	32.7	29.4	34.5	41.5	46.6	33.9	42.4	34.5

(10 marks)

b) It is hypothesized that the location of a business affects its sales. A small company has branches in two locations trading on similar items. The given data are the daily sales in thousands of KES. Test the hypothesis of different mean sales at the two sites.

(10 marks)

Day	1	2	3	4	5	6	7	8
Location A	5.51	7.14	5.45	3.95	7.74	5.22	4.39	4.55
Location B	4.71	5.47	6.12	3.69	5.43	5.20	2.16	4.34

Day	9	10	11	12	13	14	15	16
Location A	5.74	6.20	3.29	6.62	4.01	5.55	4.60	7.08
Location B	3.17	4.44	4.99	2.11	4.17	5.15	6.13	4.09

Question Three (20 Marks)

The data given in Appendix R.1 represents the impact of three advertising media- YouTube, Facebook and Newspaper, on sales. The values for all the four columns are given in millions of Kenyan shillings (KSHs). The advertising experiment was repeated 200 times, hence there are 200 data values for each column but Appendix 1 shows the first 6 observations.

a) Based on the scatter plots in Appendix R.2, comment on the relationship between (both strength and direction) each of the three media channels with sales. (4 marks)

b) Using the R Output in Appendix R.3 answer the following questions

(i) Write the equation for the fitted model (3 mark)

(ii) Interpret the p-value corresponding to the F- statistic (2 mark)

(iii) Give an interpretation of the estimates and p-values of the intercept, and the coefficients of the three advertising media (4 marks)

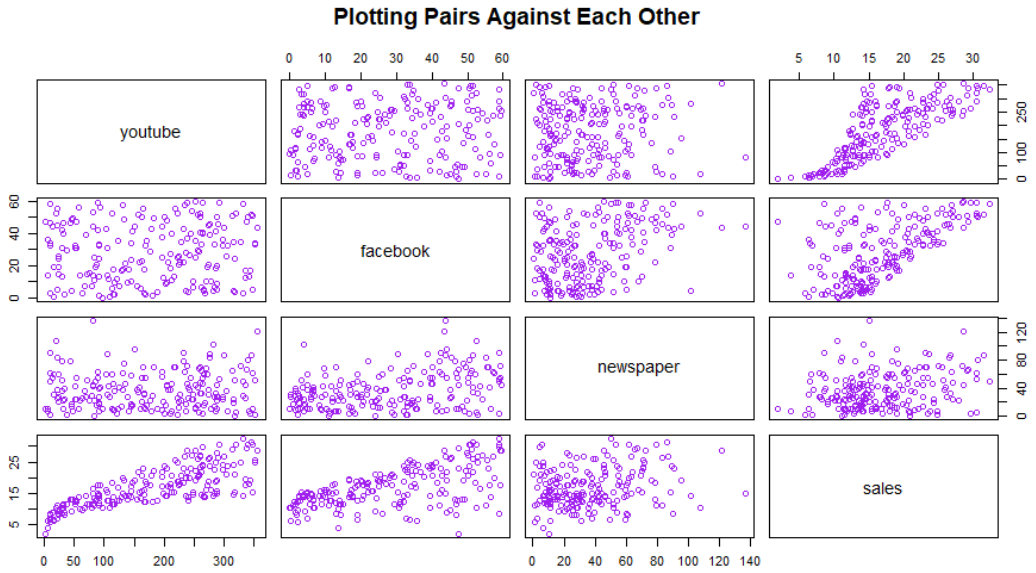
(iv) Based on the output, do you think one or more of the independent variables can be removed? If yes, which one and why? If No, why? (4marks)

c) What patterns or problems do you see in the diagnostic plots in Appendix R.4. Is the multiple linear regression a good fit for the marketing dataset? (3 marks)

Appendix R.1 Marketing Data

	youtube	facebook	newspaper	sales
1	276.12	45.36	83.04	26.52
2	53.40	47.16	54.12	12.48
3	20.64	55.08	83.16	11.16
4	181.80	49.56	70.20	22.20
5	216.96	12.96	70.08	15.48
6	10.44	58.68	90.00	8.64

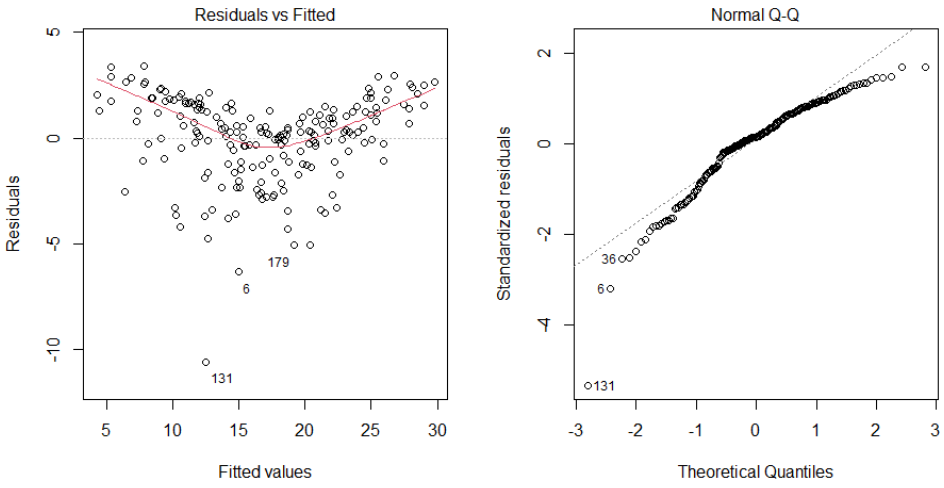
Appendix R.2 Scatter Plots

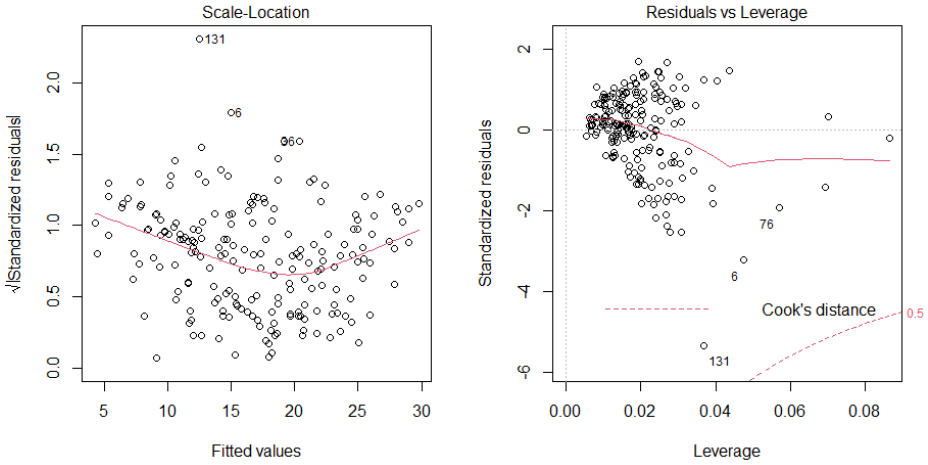


Appendix R.3 Model Summary

```
Call:
lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)

Residuals:
     Min       1Q   Median       3Q      Max
-10.5932  -1.0690   0.2902   1.4272   3.3951

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.526667   0.374290   9.422   <2e-16 ***
youtube      0.045765   0.001395  32.809   <2e-16 ***
facebook     0.188530   0.008611  21.893   <2e-16 ***
newspaper   -0.001037   0.005871  -0.177     0.86
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.023 on 196 degrees of freedom
Multiple R-squared:  0.8972,    Adjusted R-squared:  0.8956
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

Appendix R.4 Diagnostic plots





Question Four (20 Marks)

a) Four different strategies of marketing were analyzed to determine their effectiveness in boosting sales at various locations. The data are shown below. Perform an appropriate test at $\alpha = 0.05$.

Strategy	Percentage Increase of Sales				
A	13.5	13.4	14.1	14.2	
B	13.2	12.7	12.6	13.9	
C	16.8	17.2	16.4	17.3	18.0
D	18.1	17.2	18.7	18.4	

(14 marks)

b) A national magazine claims that the average college student watches less television than the general public. The national average is 29.4 hours per week, with a standard deviation of 2 hours. A sample of 30 college students has a mean of 27 hours. Is there enough evidence to support the claim at $\alpha = 0.01$? (6 marks)

Question Five (20 Marks)

a) From a given data set it was established that there was some relationship between the number of advertisement of a certain product and the volumes of sales made for 20 days, use the summary statistics given below to establish;

$$\sum_{i=1}^{10} x_i = 40, \quad \sum_{i=1}^{10} y_i = 60, \quad \sum_{i=1}^{10} x_i y_i = 7800, \quad \sum_{i=1}^{10} {x_i}^2 = 180, \quad \sum_{i=1}^{10} {y_i}^2 = 400,$$

i) The equation relating number of advertisement x and volume of sales y (4 Marks)

ii) Pearson correlation coefficient between x and y (3 Marks)

b) Consider the following regression model for wage with some predictor variables. Use it to answer the question that follows:

$$\hat{Y} = -1.56577 + 0.558018X_1 + 0.313286X_2 + 0.262683X_3 + 0.330714X_4$$
$$t = (-4.662) \quad (3.319) \quad (2.830) \quad (2.048) \quad (2.885)$$
$$S.E = (0.0060) \quad (0.1857) \quad (0.09343) \quad (0.6062) \quad (0.0736)$$

$$T = 90 \quad R^2 = 0.7134 \quad F(4,85) = 56.389 \quad \hat{\sigma} = 0.25821$$

Where

$$Y = Log(Wage); \; X_1 = Education\ Level, \quad X_2 = Experience,$$
$$X_3 = Term\ of\ Employment, X_4 = Age$$

i) Test the Hypothesis : Experience improves (increases) log(wage) of an individual at $\alpha = 0.05$ (3 Marks)

ii) What does $R^2 = 0.7134$ represent? (2 Marks)

iii) Test whether Education level, Experience, Term of Employment collectively improve the log(wage) significantly at $\alpha = 0.05$ (2 Marks)

c) The following model with two repressors including the constant is estimated over 8 observations $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ and the following data have been calculated from the original data, use it to calculate the coefficients of the estimates and obtain Durbin Watson values.

$$(XY)^{-1} = \begin{bmatrix} 5.43403 & -0.085930 & -0.118856 \\ -0.085930 & 0.00147070 & 0.0016094 \\ -0.118856 & 0.0016094 & 0.00359276 \end{bmatrix}$$

And

$$X'Y = \begin{bmatrix} 81.7 \\ 3413.11 \\ 1157.4 \end{bmatrix}$$

The matrix given above is obtained from the data given below. Use it to calculate the coefficients of the estimates and obtain Durbin Watson values. (6 Marks)

$$\begin{bmatrix} X_1 & 28 & 28 & 32 & 14 & 16 & 20 & 21 & 25 & 13 & 42 & 21 \\ X_2 & 28 & 23 & 30 & 22 & 23 & 25 & 22 & 27 & 14 & 32 & 24 \\ Y & 50 & 47 & 32 & 40 & 52 & 54 & 53 & 50 & 40 & 53 & 55 \end{bmatrix}$$