MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS 2022/2023

SCHOOL OF BUSINESS AND ECONOMICS

BACHELOR'S OF SCIENCE IN ECONOMICS AND STATISTICS

FOURTH YEAR SECOND SEMESTER

COURSE CODE: ECS 4204

COURSE TITLE: STATISTICAL INFERENCING

DATE: TIME:

INSTRUCTIONS:
Attempt Question one and any other Three Questions

Question One

a. For the following list of statements state whether the statements are TRUE or FALSE giving arguments to support your decision.

    i.   When a hypothesis fails to be rejected at 95% confidence level it can still be rejected at 99% confidence level.   **(2 marks)**

    ii.   When we carry out a hypothesis test for mean difference at 95% level of confidence and reject the null hypothesis. Then, when we construct a 95% confidence interval for the mean then we expect the confidence interval to contain zero.   **(2 marks)**

    iii.   For any test involving the sample mean as a sample statistic used in computing the test statistic it is mandatory that the distribution of the sample should be approximately symmetric otherwise the test cannot be done using sample mean.   **(2 marks)**

b. The output below shows the results of a statistical test that was used to test a hypothesis about the age of students samples from the university.

```
##
##   Shapiro-Wilk normality test
##
## data:  Age
## W = 0.9956, p-value = 0.8339


##
##   One Sample t-test
##
## data:  Age
## t = 8.3571, df = 199, p-value = 5.48e-15
## alternative hypothesis: true mean is greater than 20
## 95 percent confidence interval:
##   20.92419       Inf
## sample estimates:
## mean of x
##   21.15199
```

    a.   What was the sample size for the study?   **(1 mark)**

    b.   State the null and alternative hypothesis for the test conducted.

        (2 marks)

    c.   What was the assumption of the statistical test conducted in the extract and what was the conclusion regarding the assumption based on the output above at 95% level of confidence?   **(3 marks)**

d. At 99% level of confidence, what was the conclusion regarding the null and alternative hypothesis stated in (ii). **(2 marks)**

c. A sample of 25 students had an average height of 72 and average weight of 165. At 95% level of confidence, test the hypothesis that this sample was taken from a bivariate normal population with average height of 71 and average weight of 172 with a covariance matrix $\Sigma = \begin{pmatrix} 20 & 100 \\ 100 & 1000 \end{pmatrix}$.

(6 marks)

d. Discuss how the violation of each of the following assumptions by a linear regression model affect an estimated regression model. **(5 marks)**
   i. Normality
   ii. Homoscedasticity
   iii. Autocorrelation
   iv. Multicollinearity
   v. Model Specification

Question Two

From time to time, unknown to its employees, the research department at post bank observes various employees for work productivity. Recently, this department wanted to check whether the four tellers at a branch of this bank serve, on average, the same number of customers per hour. The research manager observed each of the four tellers for a certain number of hours. The following table gives the number of customers served by the four tellers during each of the observed hours.

Teller A	Teller B	Teller C	Teller D
19	14	11	24
21	16	14	19
26	14	21	21
24	13	13	26
18	17	16	20
	13	18	

a. At the 5% significance level, test the null hypothesis that the mean number of customers served per hour by each of these four tellers is the same. **(6 marks)**

b. State the assumptions that you made when carrying out the test in part (a) above. **(2 marks)**
c. How would you diagnose each of the assumptions made in part (b) in r package? **(2 marks)**
d. State the type of error you might have committed in part (a) justifying reasons for your conclusion. **(2 marks)**
e. If the company wished to compare the performance of Teller A and B which parametric test would they use and what underlying assumptions must be satisfied by the sample data for the test to be done? **(3 marks)**

Question Three

The manager of a certain sugar company wants to determine the most efficient type of machine to use in production at the company. The manager believe that the efficiency of the machine is measured by the amount of bags of sugar produced per day. The company has 3 types of machines that it uses in producing sugar. For a sample of days, the manager compared the number of bags of sugar produced and the results were as illustrated below;

```
##
##  Shapiro-Wilk normality test
##
## data:  MachineA
## W = 0.55409, p-value = 7.448e-13


##
##  Shapiro-Wilk normality test
##
## data:  MachineB
## W = 0.80196, p-value = 5.475e-07


##
##  Shapiro-Wilk normality test
##
## data:  MachineC
## W = 0.74234, p-value = 1.777e-08

## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value  Pr(>F)
## group   2  4.4395 0.01319 *
##       171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##              Df Sum Sq Mean Sq F value  Pr(>F)
## Machine       2    7.2   3.602    4.83 0.00911 **
## Residuals   171  127.5   0.746
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
##  Kruskal-Wallis rank sum test
##
## data:  Bags by Machine
## Kruskal-Wallis chi-squared = 11.605, df = 2, p-value = 0.00302

##   Tukey multiple comparisons of means
##     99% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = Bags ~ Machine, data = Production)
##
## $Machine
##                          diff        lwr       upr     p adj
## Machine C-Machine B 0.2322470 -0.25856147 0.7230555 0.3445977
## Machine A-Machine B 0.4928531  0.02254435 0.9631618 0.0064875
## Machine A-Machine C 0.2606061 -0.20493589 0.7261480 0.2264870
```

a. State the null and alternative hypothesis for the study.   (2 marks)
b. State the parametric test required to test the hypothesis in (a). (1 mark)
c. State two assumptions required in order to carry out the parametric test in (b).   (2 marks)
d. Based on the output given was the assumption in (c) satisfied, justify your reasoning.   (3 marks)
e. Given the conclusion in (d) what course of action was taken by the analyst next.   (1 mark)
f. Based on the findings in the output what conclusions could be on problem in the study. Justify your conclusion.   (3 marks)
g. Based on the conclusion made in (f) which type of error might have been committed in (f).   (2 marks)
h. Which machine(s) should the company invest its money in?   (1 mark)

Question Four

Let $Y = (y_1, y_2, y_3)$ be a random vector with mean vector $\bar{Y} = \begin{bmatrix} 36.09 \\ 25.55 \\ 34.09 \end{bmatrix}$ and

covariance matrix $S = \begin{bmatrix} 65.09 & 33.65 & 47.59 \\ 33.65 & 46.07 & 28.95 \\ 47.59 & 28.95 & 60.69 \end{bmatrix}$. If the random variables Z and W is

formed through linear combination of Y such that $Z = 3y_1 - 2y_2 + 4y_3$ and
$W = y_1 + 3y_2 - y_3$. Determine;

   a. $\bar{Z}$.   (2 marks)
   b. $\bar{W}$.   (2 marks)
   c. $\text{var}(Z)$.   (3 marks)
   d. $\text{var}(W)$.   (3 marks)
   e. Correlation between Z and W.   (3 marks)
   f. Correlation between y_2 and y_3.   (2 marks)

Question Five

Two Psychological tests were given to 6 Males and 6 Females. The data record was as illustrated in the table below.

Y_1 = Pictorial Inconsistencies

Y_2 = Paper form board

Male		Female	
Y_1	Y_2	X_1	X_2
25	27	13	14
27	25	14	12
25	24	12	19
28	26	17	17
27	20	12	15
29	29	16	14

Using the Hoteling's T² test, test the hypothesis that the psychological test scores for Males is significantly different to that of Females at 95% confidence level.   (15 marks)