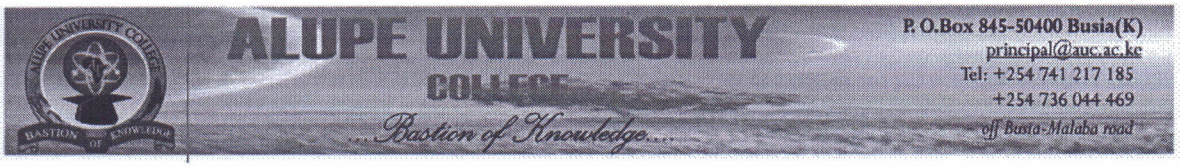

OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2020 /2021 ACADEMIC YEAR

FOURTH YEAR SECOND SEMESTER REGULAR EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH COMPUTING)

COURSE CODE: **STA 423**

COURSE TITLE: **BIOMETRY METHODS**

DATE: 22/7/2021 **TIME: 1300-1600HRS**

INSTRUCTION TO CANDIDATES

- **SEE INSIDE**

THIS PAPER CONSISTS OF 5 PRINTED PAGES **PLEASE TURN OVER**

REGULAR – MAIN EXAM
STA 423: BIOMETRY METHODS
STREAM: **DURATION: 3 hours**

INSTRUCTION TO CANDIDATES
Answer **ALL** questions from section A and any **THREE** from section B.

SECTION A [31 Marks] Answer All questions]

QUESTION ONE [16 Marks]

a) Define clearly the following terms [5 Marks]
   i) Clinical Trials
   ii) Blinding
   iii) Cross-sectional study
   iv) Longitudinal study
   v) Randomization
b) Discuss any four types of endpoints in clinical trials [8 Marks]
c) Identify three objectives of Phase II trials. [3 Marks]

QUESTION TWO [15 Marks]

a) Discuss any two advantages of a case-control study [4 Marks]
b) Consider a placebo control trial investigating lung function in Asthmatic patients, the primary endpoint is the change from baseline in lung function at 16 weeks, at baseline we observe slight imbalances in characteristics such as age and baseline lung function, FEV. With the continuous and normally distributed endpoint of change from baseline in FEV, a simple linear regression analysis to assess treatment effect was conducted. Each patient, *i*, has a value for the outcome or dependent variable Y, FEV change while X is the explanatory variable is treatment which is binary being either active drug or placebo. The following table provides summary output of the simple linear regression model.

```
Call:
lm(formula = fev.change ~ group, data = asthma.trial)

Residuals:
     Min       1Q   Median       3Q      Max
-1.45445 -0.24530  0.01849  0.24354  0.89847

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.01500    0.03613  -0.415   0.6784
groupActive Drug   0.10200    0.05148   1.982   0.0489 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3649 on 199 degrees of freedom
Multiple R-squared:  0.01935,   Adjusted R-squared:  0.01442
F-statistic: 3.926 on 1 and 199 DF,  p-value: 0.04891
```

i) Interpret the coefficients of the model [3 marks]
ii) Make relevant conclusions [2 marks]

c) For the same dataset in Question (b) multivariate linear regression was conducted to adjust for age effects. The regression model output is given below

```
Call:
lm(formula = fev.change ~ group + age, data = asthma.trial)

Residuals:
     Min       1Q   Median       3Q      Max
-1.48008 -0.24273 -0.00455  0.23823  0.84782

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.590232   0.188586   3.130  0.00201 **
groupActive Drug 0.079707   0.050729   1.571  0.11773
age             -0.009250   0.002831  -3.267  0.00128 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3563 on 198 degrees of freedom
Multiple R-squared:  0.06951,   Adjusted R-squared:  0.06011
F-statistic: 7.395 on 2 and 198 DF,  p-value: 0.0007991
```

i) Interpret the coefficients of the model [3 marks]
ii) Could there be any change from the conclusions made in Question (b)? Explain [3 marks]

SECTION B [39 Marks] Answer any THREE questions]

QUESTION THREE [13 Marks]

a) State some reasons a clinical trial may be stopped early. [3 Marks]

b) Use the snapshot below of the structure of Acupuncture dataset from a randomized controlled trial investigating the effects of acupuncture on headache scores compared to a control intervention.

```
str(Acupuncture)

'data.frame':   396 obs. of  18 variables:
 $ id                    : num  100 101 104 105 108 112 113 114 126 130 ...
 $ age                   : num  47 52 32 53 56 45 45 49 47 46 ...
 $ sex                   : Factor w/ 2 levels "Female","Male": 1 1 1 1 1 1 1 1 1 1 ...
 $ migraine              : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ chronicity            : num  35 8 14 10 40 27 30 49 42 3 ...
 $ treatment.group       : Factor w/ 2 levels "Acupuncture",..: 1 2 2 2 2 1 1 1 2 1 ...
 $ score.baseline        : num  10.8 9.5 16 32.5 16.5 ...
 $ score.baseline.4      : Factor w/ 4 levels "[6.75,15.2]",..: 1 1 2 3 2 1 4 3 2 3 ...
 $ age.group             : Factor w/ 4 levels "18-34","35-44",..: 3 3 1 3 4 3 3 3 3 3 ...
 $ score.month3          : num  NA NA NA 44 17.5 ...
 $ score.month12         : num  NA NA 15.3 NA 23.2 ...
 $ withdrawal.reason     : Factor w/ 7 levels "adverse effects",..: 5 7 NA 7 NA NA NA NA NA NA ...
```

For the data, write the R codes to:

i) Display the numbers of patients randomized to the two treatment groups (Hint: use the **table()** function). [2 marks]

ii) Generate summaries of the variables (**score.baseline**, **age**, and **sex** by **treatment.group**) by treatment group and save results as **baselines** (Hint: use the **compareGroups()** function) [4 marks]

iii) Display the results saved in baselines (Hint: use the **createTable()** function) [2 marks]

iv) Display the created summary table, that is, the numbers of patients randomized to the two treatment groups. [2 marks]

QUESTION FOUR [13 Marks]

a) Identify the USA Institutional Review Board's (IRB) specific prerequisites that human research studies must meet. [6 marks]

b) Suppose you want to conduct a clinical trial whereby simple randomization is used to design the study. You are required to produce a randomization list for the trial with 140 patients and two treatment arms, A and B.

i) Generate a vector to store treatment labels "A" and "B". Set the seed to a preferable number. [2 marks]

ii) Randomly select (with replacement) treatment arm 140 times with the **sample()** function and store in a vector [2 marks]

iii) Display the contents of the vector [1 mark]

iv) Tabulate the numbers assigned to each treatment [2 marks

QUESTION FIVE [13 Marks]

a) Giving examples, discuss any three categories of sources of bias in clinical studies [6 marks]

b) Let $\bar{Y}_A$ be the average response for the sample of individuals assigned to treatment A and $\bar{Y}_B$ the similar quantity for treatment B. Show that under stratified randomization,

$$\bar{Y}_A - \bar{Y}_B = \beta + \alpha\left(\frac{n_{A1}}{n_A} - \frac{n_{B1}}{n_B}\right) + (\bar{\epsilon}_A - \bar{\epsilon}_B).$$

[7 marks]

QUESTION SIX [13 Marks]

a) Identify any 5 purposes of a protocol document [5 Marks]

b) A trial was designed to assess whether supplementing with glutamine or selenium, or both affected the odds of infections in critically ill patients. A factorial, randomized design was used. The data stored as **fact.data** has the head, as shown in the snapshot below;

```
head(fact.data)

  glutamine selenium infection
1       Yes       No       Yes
2        No       No        No
3        No      Yes       Yes
4       Yes       No       Yes
5        No      Yes        No
6        No       No       Yes
```

Write R codes to:

i) Display the numbers with and without infections by supplement combination. [2 marks]
ii) Display the numbers and proportions with/without infections for those given glutamin e. [3 marks]
iii) Display the numbers and proportions with/without infections for those given selenium . [3 marks]

QUESTION SEVEN [13 Marks]

a) Identify three reasons for computing a correct sample size in clinical trials [3 marks]
b) Give any four conditions for deriving the sample size necessary to detect a clinically important difference with some desired power [4 Marks]
c) Suppose the standard treatment of care (treatment 2) has a response rate of about .35 (best guess). After collaborations with your clinical colleagues, it is determined that a clinically important difference for a new treatment is an increase in .10 in the response rate. That is, a response rate of .45 or larger. If we are to conduct a clinical trial where we will randomize patients with equal allocation to either the new treatment (treatment 1) or the standard treatment;
   i) How large a sample size is necessary to detect a clinically important difference with 90% power using a one-sided test at the .025 level of significance? [5 marks]
   ii) How many patients will be assigned to each treatment arm? [1 mark]

****