ALUPE UNIVERSITY COLLEGE

...Bastion of Knowledge....

OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2021 /2022 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: 7TH JUNE 2022   TIME: 1400 – 1700 HRS

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 [15 Marks]

a) State the Hippocratic Oath used in Clinical trials   [6 marks]

b) Discuss any three randomization methods   [9 Marks]

QUESTION TWO [16 Marks]

a) Define the following   [4 marks]

   i)   Bioequivalence trials
   ii) Pharmacokinetics
   iii) Equivalence trials
   iv) Non-inferiority trials

b) Discuss any three advantages of stratified randomization in clinical trials   [6 marks]

c) A treatment difference of 1 unit with a pooled standard deviation of 3 is expected. If there truly is a treatment difference, we want the probability of being able to detect this to be 0.8 (i.e. 80% power). Write a R code to:

   i)   Calculate the sample size needed to detect a treatment difference of 1 unit, with standard deviation of 3 and 80% power and store the results in ss1 (Hint: use the **power.t.test()** function)   [3 marks]
   ii) Round up and display the numbers needed per group   [1 mark]
   iii) Calculate the power given the sample size calculated above   [2 marks]

SECTION B [39 Marks] Answer any THREE questions]

QUESTION THREE [13 Marks]

a) Suppose you want to conduct a placebo-controlled weight loss trial. To determine the sample size, discuss any possible seven requirements.   [7 marks]

b) Discuss any six strategies that can be applied to reduce confounding in clinical trials   [6 marks]

QUESTION FOUR [13 Marks]

In a Prospective study, 40,000 British doctors were followed for 10 years. The following data were collected:

Death Rate from Lung Cancer per 1000 person years.

# cigarettes smoked per day	death rate
0	.07
1-14	.57
15-24	1.39
35+	2.27

For presentation purpose, the estimated rates are multiplied by 1000. Assuming the death rate remains constant over the 10-year period for each group of doctors. Calculate:

i) Probability of dying from lung cancer in one year for British doctors smoking between 15-24 cigarettes per day
[2 marks]

ii) Probability of dying from lung cancer in one year for the heaviest smokers   [2 marks]

iii) Relative risk of death from lung cancer between heavy smokers and non-smokers
[3 marks]

iv) Odds-ratio of dying from lung cancer in one year between heavy smokers and non-smokers [4 marks]

v) Provide the conclusions about the odds ratio and relative risk obtained   [2 marks]

QUESTION FIVE [13 Marks]

a) Define the Intention-to-treat (ITT). What is the benefit and limitation of the principle
[3 marks]

b) Name any five aspects that an informed consent should include   [5 Marks]

c) Suppose we consider patients with Esophageal cancer treated with chemotherapy prior to surgical resection. A complete response is suspected to occur with 35% (guess) probability using a drug under investigation in a phase II study. Obtain a sample that will meet the precision of our estimator to be such that the 95% confidence interval is within 15% of the true π. Comment on the sample obtained.   [5 marks]

QUESTION SIX [13 Marks]

In a randomized controlled trial investigating the effects of acupuncture on headache scores compared to a control intervention, a researcher is interested in exploring the composite endpoints. Specifically, the researcher is interested in examining whether patients visited either a complementary therapist, GP or specialist. As such, a binary composite endpoint from the three variables in the Acupuncture dataset is computed. The structure of the Acupuncture is given in the snapshot below.

```
'data.frame':    396 obs. of  23 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 ...
 $ completedacupuncturetreatment: num  NA NA NA NA NA 1 NA NA NA NA ...
 $ completer                    : num  0 0 1 0 1 1 1 1 1 1 ...
 $ total.therap.visits          : num  NA 2 0 2 0 0 7 0 0 10 ...
 $ total.gp.visits              : num  NA 4 0 0 0 5 1 0 1 0 ...
 $ total.spec.visits            : num  NA 0 0 0 0 0 0 0 0 0 ...
 $ total.days.sick              : num  NA 6 3 NA 23 2 6 9 19 0 ...
 $ diff.month12                 : num  NA NA -0.667 NA 6.75 ...
 $ pct.month12                  : num  NA NA -4.17 NA 40.91 ...
 $ resp35.month12               : Factor w/ 2 levels "greater than 35%",..: NA NA 2 NA 2 2 2 2 2 1 ...
 $ any.gp.visits                : Factor w/ 2 levels "Did not visit GP",..: NA 2 1 1 1 2 2 1 2 1 ...
 $ any.spec.visits              : Factor w/ 2 levels "Did not visit specialist",..: NA 1 1 1 1 1 1 1 1 1 ...
```

Write a R code to:

i) Dichotomize **total.therap.visits** and encode it as a factor. **any.therap.visits** should take the value "**Did not visit CT**" whenever **total.therap.visits** equals 0 or "**Visited CT**" otherwise. [3 marks]

ii) Generate a combined binary endpoint for the event of having at least one visit to any professional. combined should take the value "No visits" whenever **any.therap.visits** equals "**Did not visit CT**" and **any.gp.visits** equals "**Did not visit GP**" and **any.spec.visits** equals "**Did not visit specialist**", or "**At least one visit**" otherwise. Dichotomize **total.therap.visits** into 0 or at least 1 visit, encode it as a factor, and assign it to a column named **any.therap.visits**. [8 marks]

iii) Tabulate the binary endpoint. [2 marks]

QUESTION SEVEN [13 Marks]

A clinician is unsure of the expected treatment difference but believes it would be between 0.5 and 2.0. Using 80% power and standard deviation of 3, provide them with a set of sample sizes required for this range of differences as a table and in a plot. Using R write codes to.

i) Generate a vector containing values between 0.5 and 2.0, incrementing by 0.25 [3 marks]
ii) Generate the sample sizes for each of the treatment differences   [3 marks]
iii)Create a data frame from the deltas and sample sizes   [2 marks]
iv)Plot the patients per group against the treatment differences (Hint: use **ggplot()**) [5 marks]

****