AFRICA UNIVERSITY
(A United Methodist-Related Institution)

"Investing in Africa's Future"

FACULTY OF HEALTH SCIENCES

FIRST SEMESTER EXAMINATIONS

COURSE CODE:	NSPH541
COURSE TITLE:	HEALTH STATISTICS
DATE:	DECEMBER 2023
TIME:	3 hours

INSTRUCTIONS

Answer **ALL** Questions in **Section A** and **ANY 3** questions from **Section B**

The mark allocation for each question is indicated at the end of the question

Credit will be given for logical, systematic and neat presentations.

SECTION A

QUESTION 1: 20 marks

a) Match the statements below with the corresponding terms from the list.

[5]

R2- adjusted
residual plots
residual
interaction/effect modification
heteroscedasticity
R2
Outliers
multiple regression model
multicollinearity
dummy variables

	Used to check the assumptions of the regression model.
	Used when trying to decide between two models with different numbers of predictors
	Used when the effect of a predictor on the response depends on other predictors
	The proportion of the variability in y explained by the regression model
	Is the observed value of y minus the predicted value of y for the observed x
	A point that lies far away from the rest.
	$y= \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$
	The problem can occur when the information provided by several predictors overlaps.
	Used in a regression model to represent categorical variables.
	The constant variance of fitted values of the regression model

b) State **three** data features suitable for survival analysis [3]

c) When is it suitable for one to fit a logistic regression and not a linear regression [2]

d) Define the following terms [8]

i. Power
ii. Level of significance

iii. Type I error

iv. Type II error

e) In case-control studies, it is difficult to get enough cases, what concept is used in such studies to enrol enough participants [2]

QUESTION 2: 20 marks

A clinician wishes to assess the effect of an intervention drug in curing disease A and reducing the time to death. A total of 48 participants were enrolled in the clinical trial study. There were 20 participants in the placebo arm and 28 participants in the interventions arm. The data description is shown below.

```
  obs:            48                          Patient Survival in Drug Trial
 vars:             8                          3 Mar 2018 02:12
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
studytime       byte    %8.0g                 Months to death or end of exp.
died            byte    %8.0g                 1 if patient died
drug            byte    %8.0g                 Drug type (0=placebo)
age             byte    %8.0g                 Patient's age at start of exp.
```

a) The clinician performed the following analysis to determine the effect of the intervention drug. One of the Stata outputs is shown below

```
     failure event:  died != 0 & died < .
obs. time interval:  (0, studytime]
 exit on or before:  failure

------------------------------------------------------------------------------
         48  total observations
          0  exclusions
------------------------------------------------------------------------------
         48  observations remaining, representing
         31  failures in single-record/single-failure data
        744  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =        39
```

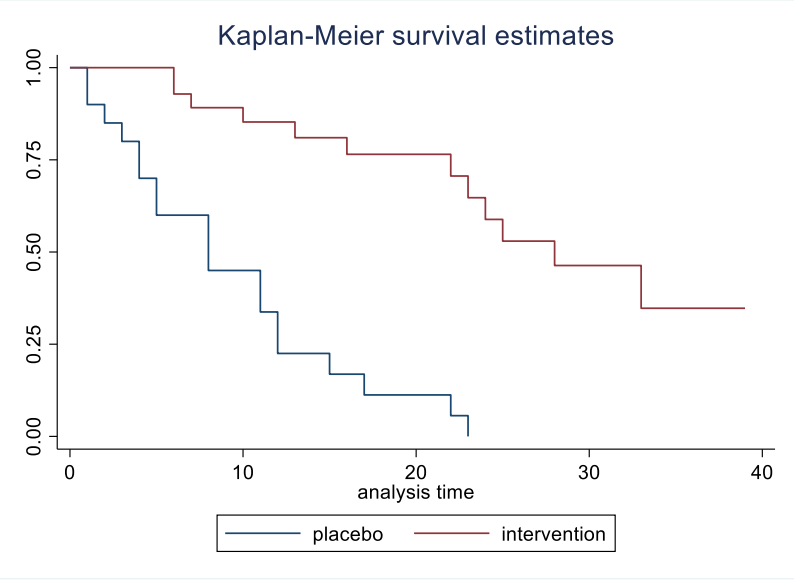
i. Specify the type of analysis the clinician performed [1]

ii. What was the total follow-up time in this study? Specify the time units. [1]

iii. How many participants died at the end of the study? [1]

iv. What is the maximum number of years the last participant stayed in the study? [2]

b) The clinician plotted the graph below. Interpret this plot fully [2]



c) The rate of occurrence of the outcome was estimated for each drug (0=placebo; 1=intervention). Interpret the rate of death in each drug arm fully [4]

```
. strate drug, per(100)

        failure _d:  died
  analysis time _t:  studytime

Estimated failure rates
Number of records  =       48
```

drug	D	Y	Rate	Lower	Upper
0	19	1.8000	10.5556	6.7329	16.5486
1	12	5.6400	2.1277	1.2083	3.7465

```
Notes: Rate = D/Y = failures/person-time (per 100).
       Lower and Upper are bounds of 95% confidence intervals.
```

d) The clinician performed the test to compare if there was a significant difference in the rate of mortality between the two-drug arm and the results are shown below:

i. State the name of the test performed [1]

ii. Interpret the results fully [2]

drug	Events observed	Events expected
0	19	7.25
1	12	23.75
Total	31	31.00

chi2(1) = 28.27
Pr>chi2 = 0.0000

e) The clinician fitted the univariate regression model. Below is the output

No. of subjects = 48
No. of failures = 31
Time at risk = 744
Log likelihood = -88.00019

Number of obs = 48
LR chi2(1) = 23.82
Prob > chi2 = 0.0000

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
drug						
0	1	(base)				
1	.1327581	.0584002	-4.59	0.000	.0560555	.3144157

Interpret the coefficient of the drug (Remember: 0=placebo; 1=intervention). [3]

f) The analysis was further adjusted for age. Below is the output

No. of subjects = 48
No. of failures = 31
Time at risk = 744
Log likelihood = -83.323546

Number of obs = 48
LR chi2(2) = 33.18
Prob > chi2 = 0.0000

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
drug						
0	1	(base)				
1	.1048772	.0477017	-4.96	0.000	.0430057	.2557622
age	1.120325	.0417711	3.05	0.002	1.041375	1.20526

i. Compare the adjusted effect of the drug and the unadjusted effect. [1]

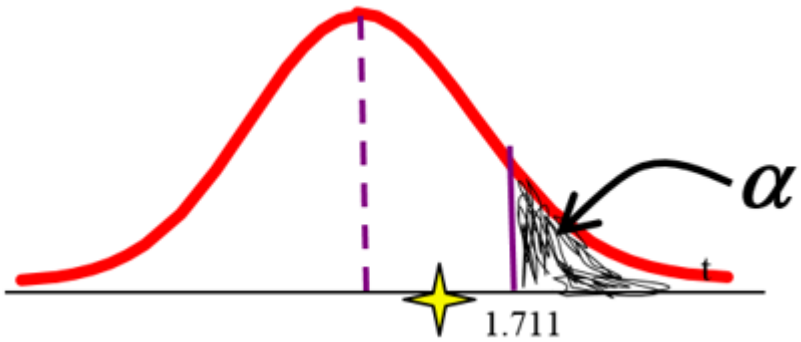
ii. Interpret the effect of age in this model [2]

SECTION B

QUESTION 3: 20 marks

a) List the steps for hypothesis testing for a single mean [5]

b) A study was conducted to determine if the average height of students in a class was **above** 170 cm. A sample of 25 students was taken and the T-test statistic was calculated. The following image shows the position of the critical value (t-crit=1.711) and the test statistic value represented by a star (t-statistic=1.27). Using this information, test the hypothesis that the height is above 170cm. Show all the steps



[7]

c) State the **four** assumptions of linear regression and describe how each of the assumptions is assessed. [8]

QUESTION 4: 20 marks

a) Understanding the amount of serum catecholamine in body circulation has been an emerging topic in managing hypertension. You wish to carry out a study comparing serum catecholamine levels in normotensive patients and patients with essential hypertension. Previous studies have found mean serum catecholamine levels of 0.85mg/mL. (sd = 0.45) in normotensives and 0.72mg/mL. (sd = 0.35) for patients with essential hypertension.

i. Calculate the required sample size at a 5% level of significance and a power of 80%. [5]

ii. Calculate the required sample size at a 5% level of significance and a power of 90%. [4]

iii. Comment on your findings from (i) and (ii) [1]

iv. Maintaining all other factors the same, calculate the required sample size if the **difference** to be detected was

(i) 0.08 mg/m [2]

(ii) 0.0065mg/m [2]

v. Assuming attrition of 15%, using your answer form (i), what will the final sample size be? [2]

b) The researcher considered using secondary data to answer this same research question. What limitations should the researcher acknowledge in their report? [4]

QUESTION 5: 20 marks

a) In a cohort study, a researcher sets out to determine if alcohol intake was associated with developing coronary heart disease (chd). A sample of 205 participants was considered and the following 2x2 table was constructed after reviewing the data:

```
. ta alcohol chd [freq=freq]
```

	chd		
alcohol	Yes	No	Total
High	72	32	104
Low	33	68	101
Total	105	100	205

i. Specify the appropriate measure of association the researcher should estimate [2]

ii. Calculate the measure of association for this study [3]

iii. Calculate the 95% confidence interval of the measure of association calculated in (b). Show all your steps [5]

iv. Interpret the effect of alcohol intake on developing coronary heart disease and state if this was statistically significant [3]

v. The researcher gave this data to a student doing MPH and Africa University to perform an analysis for them using Stata. The student reported the following:

chi2(1) = 27.41
Pr>chi2 = 0.001

vi. How similar or different are these Stata results from your manual calculation? [3]

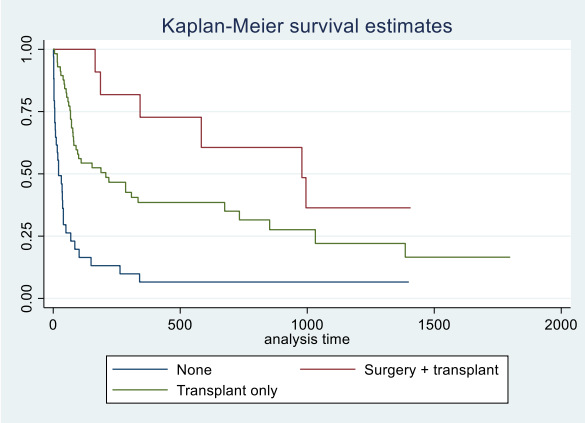
b) Discuss any additional analysis you would recommend the researcher to perform [4]

QUESTION 6: 20 marks

a) These results are part of a study that was done to determine the effects of heart transplants and surgery on survival among patients who were in the ICU. Interpret

the results fully and highlight is these procedures affected the survival of the patients or not. State your recommendations in this study. [10]

Output A



Output B

```
Estimated failure rates
Number of records  =       172
```

	ST	D	Y	Rate	Lower	Upper
None		30	3.2850	9.13242	6.38526	13.06151
Both		6	8.1850	0.73305	0.32933	1.63168
Transplant only		39	20.4681	1.90540	1.39215	2.60788

```
Notes: Rate = D/Y = failures/person-time (per 1000).
       Lower and Upper are bounds of 95% confidence intervals.
```

Output C

```
No. of subjects =          103                 Number of obs    =         172
No. of failures =           75
Time at risk    =      31938.1
                                               LR chi2(2)       =       28.94
Log likelihood  =   -283.84695                 Prob > chi2      =      0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ST						
None	1	(base)				
Both	.1466495	.0669375	-4.21	0.000	.0599451	.3587631
Transplant only	.3025945	.0754613	-4.79	0.000	.1856041	.4933266

b) This study was set to determine the weight of participants who attended clinic A in Europe. The weight is measured in pounds.

```
. reg weight i.sex i.race, base
```

Source	SS	df	MS	Number of obs	=	4,071
				F(3, 4067)	=	229.02
Model	697319.808	3	232439.936	Prob > F	=	0.0000
Residual	4127692.91	4,067	1014.92326	R-squared	=	0.1445
				Adj R-squared	=	0.1439
Total	4825012.72	4,070	1185.50681	Root MSE	=	31.858

weight	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sex						
male	0	(base)				
female	-25.16021	.9999993	-25.16	0.000	-27.12076	-23.19967
race						
White	0	(base)				
Black	8.217199	1.59442	5.15	0.000	5.091263	11.34313
Other	-21.92051	3.872409	-5.66	0.000	-29.51255	-14.32847
_cons	172.5975	.7463242	231.26	0.000	171.1343	174.0607

i. Comment on the variability value for the adjusted model [2]

ii. What is the average adjusted weight in this study? [2]

iii. Interpret the effects of race on weight [4]

iv. Interpret the effect of sex on weight [2]

End of paper!