



**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: July 2022  
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]

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

- R<sup>2</sup>
- outliers
- residual
- residual plots
- multiple regression model
- R<sup>2</sup>- adjusted
- heteroscedasticity
- interaction/effect modification
- multicollinearity
- dummy variables

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 reduce 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

```

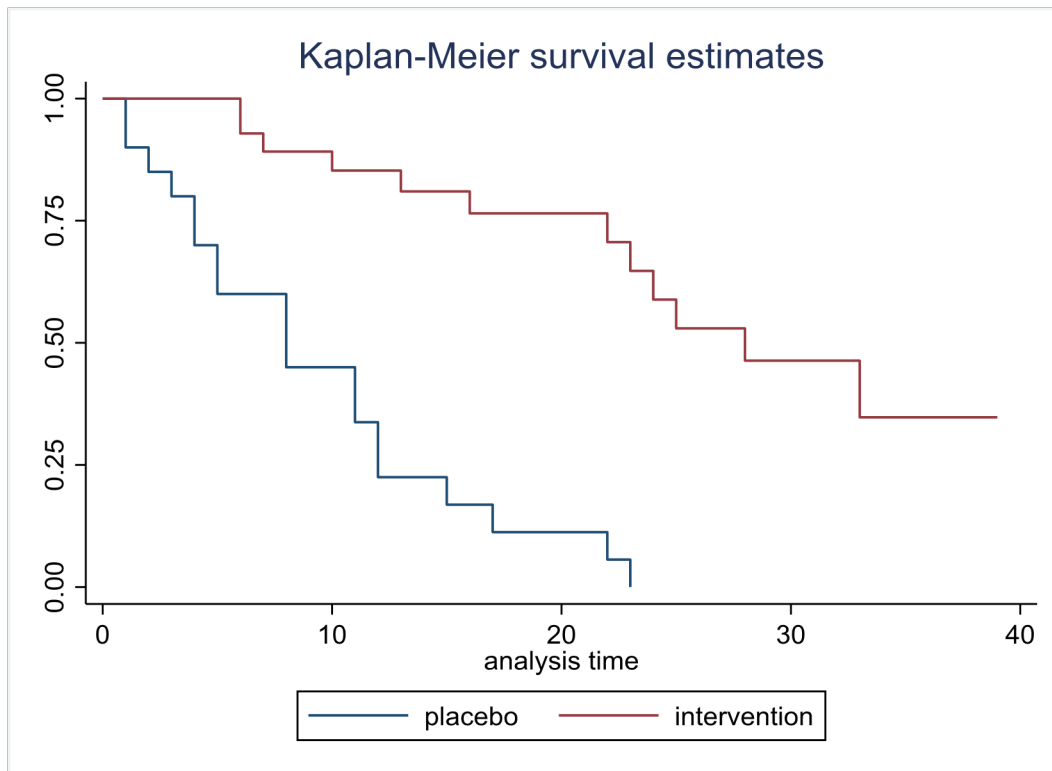
---

```

      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 did 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:
- State the name for the test performed [1]
  - Interpret the results fully [2]

dru	Events expected	Events observed

g	
0	19
1	12
Total	31
l	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	Number of obs	=
48			
No. of failures =	31		
Time at risk =	744		
		LR chi2(1)	=
23.82			
Log likelihood =	-88.00019	Prob > chi2	=
0.0000			

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

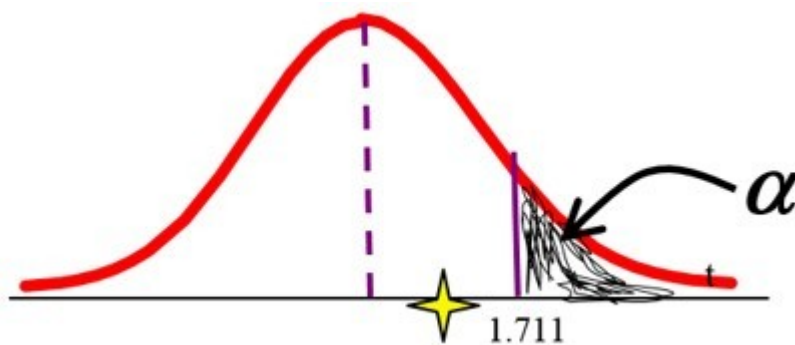
No. of subjects =	48	Number of obs	=
48			
No. of failures =	31		
Time at risk =	744		
		LR chi2(2)	=
33.18			
Log likelihood =	-83.323546	Prob > chi2	=
0.0000			

_t	Haz. Ratio Conf. Interval]	Std. Err.	z	P> z	[95%
drug	1 (base)				
0					
1	.1048772 .2557622	.0477017	-4.96	0.000	.0430057
age	1.120325 1.20526	.0415711	3.05	0.002	1.041375

- 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 ( $t\text{-statistic}=1.27$ ). Using this information, test the hypothesis that the height is above 170cm. Show all the steps



- c) State the **four** assumptions of linear regression and describe how each of the assumptions is assessed. [7]
- [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 0.812mg/mL. (sd = 0.41) in normotensives and 0.762mg/mL. (sd = 0.37) for patients with essential hypertension.
  - i. Calculate the required sample size at 5% eve of significance and a power of 80%. [5]
  - ii. Calculate the required sample size at 5% eve of significance and a power of 90%. [4]
  - iii. Comment on your finding 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 from (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 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:

		chd		Total
alcohol		No	Yes	
High	Hi	32	72	104
Low	L	68	33	101
Total		100	105	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:
- $\chi^2(1) = 27.41$   
 $P > \chi^2 = 0.0000$

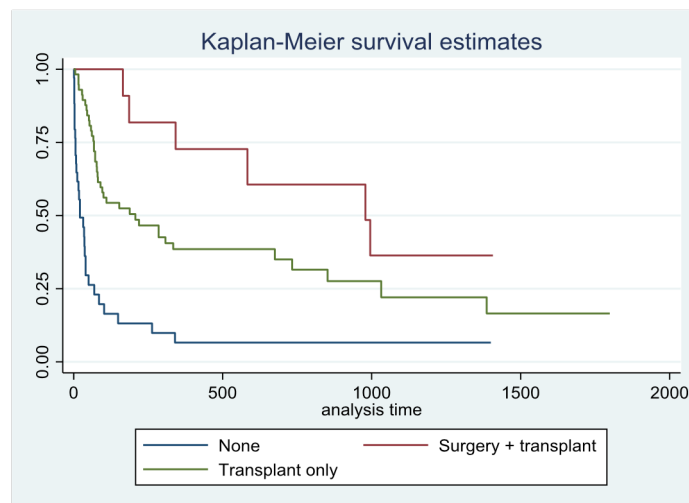
How similar or different are these Stata results from your manual calculation? [3]

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 ICU. Interpret the results fully 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	
Both	6	8.1850	0.73305	0.32933	
Transplant only	39	20.4681	1.90540	1.39215	
Lower					

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 Interval]	Std. Err.	z	P> z	[95% Conf.
		1 (base)				
ST	Non					
e	Bot		.1466495	.0669375		-4.21
h		0.000	.0599451	.3587631		
Transplant			.3025945	.0754613		-4.79
only		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

ce	Sour	MS	SS	df	Number of obs	=	4,071
					F(3, 4067)	=	229.02
					Prob > F	=	
	Mod	697319.808		3	0.0000		
el		232439.936			R-squared	=	
	Residu	4127692.91		4,067	0.1445		
al		1014.92326			Adj R-squared	=	
					0.1439		
	Tot	4825012.72		4,070	Root MSE	=	
al		1185.50681			31.858		
ht	weig	Coef.	Std. Err.	t	P> t	[95% Conf.	
		Interval]					
		0 (base)					
sex							
male							
	femal	-25.16021	.9999993	-25.16	0.000	-27.12076	-
	e	23.19967					
		0 (base)					
race							
White							
	Blac	8.217199	1.59442	5.15	0.000	5.091263	
k		11.34313					
	othe	-21.92051	3.872409	-5.66	0.000	-29.51255	-
r		14.32847					
	_co	172.5975	.7463242	231.26	0.000	171.1343	
ns		174.0607					

- Comment of the adjusted R2 value for the model [2]
- Interpret the average adjusted weight in this study [2]
- Interpret the effects of race on weight [4]
- Interpret the effect of sex on weight [2]

**End of paper!**