AFRICA UNIVERSITY
(A United Methodist-Related Institution)

"Investing in Africa's Future"

FACULTY OF HEALTH SCIENCES

COURSE CODE: SPH541

COURSE TITLE: HEALTH STATISTICS

DUE DATE: NOVEMBER-DECEMBER 2015

INSTRUCTIONS

Answer **ALL** Questions in **SECTION A** and **ANY 3** 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

A pediatrician designed a study to test his hypothesis that children who consume large amounts of red food coloring are more likely to be hyperactive than those who do not. He interviewed the parents of 100 hyperactive children (identified from the medical records of six psychiatrists) and compared their diets to those of 100 children undergoing treatment by one of the same six psychiatrists for other behavioral problems to determine the dietary habits of their children over the preceding two year period.

i) What is the research design used in this study?

ii) Of the 100 children diagnosed as being hyperactive, 42 were assessed as consuming high levels of red food coloring, as compared to 24 children being seen for other behavioral problems. What is the risk of hyperactivity associated with red food coloring?

iii) What is the population attributable risk (aetiologic fraction)?

iv) What biases may be operating in this study? [10]

QUESTION 2

a. Discuss how the sample size of a study can influence the results of inferential statistical tests with regard to clinical significance. [5]

b. Describe type I and type II errors, and how they are related to power and significance level of a study. [5]

QUESTION 3

a. List and briefly discuss the criteria you would apply if you wanted to establish causality between an exposure and a disease. [5]

b. Explain some of the issues to consider when estimating the required sample size for a study. [5]

QUESTION 4

In an outbreak of tuberculosis among prison inmates in South Carolina in 1999, 28 of 157 inmates residing on the East wing of the dormitory developed tuberculosis, compared with 4 of 137 inmates residing on the West wing.

a) Summarize the information above

b) Calculate the appropriate measure of association

c) Interpret the results [10]

SECTION B

QUESTION 5

You conduct a case-control study to examine the relationship between eating margarine and depression. You find that among the 185 patients who suffer from depression, 65 eat margarine while 50 of the 230 controls eat margarine.

a). Draw a 2x2 table and calculate the crude OR [4]

b). Interpret the crude OR [2]

c). Among the 100 women in this group who suffer from depression, 25 eat margarine. Among the 50 female controls, 5 eat margarine. Among the 85 male cases, 40 eat margarine. Among the 180 male controls, 45 eat margarine. Draw out the stratified 2x2 tables and calculate their respective ORs. [8]

d). interpret the stratum –specific ORs. [2]

e). Is this an example of effect modification or confounding? [2]

f). Explain in layman terms what this conclusion means. [2]

QUESTION 6

The results below are from an occupational health study of factors related to the development of byssinosis (a lung disease) in workers employed in a textile industry ***[Higgins and Koch, 1977. Source: Kleinbaum Kupper and Muller. Applied Regression Analysis and Other Multivariate Methods. PWS-KENT Publishing Company, 1988]***. The variables of interest are:

Byssinosis (outcome) 1: yes 0: no

Workplace (type of place worked in) 1: dusty 0: not dusty

Years of employment	Year 1	Year 2
<10	0	0
10-20	1	0
>20	0	1

Smoker 1: yes 0: no

STATA output for the results on the next page

```
logit byssinosis workplace smoker year1 year2 [fweight=count]
Iteration 0: log likelihood = -738.5961
Iteration 1: log likelihood = -624.03854
Iteration 2: log likelihood = -600. 92294
Iteration 3: log likelihood = -599.45822
Iteration 4: log likelihood = -599.44488
Iteration 5: log likelihood = -559.44488
```

Log Estimates — No of Obs = 5419; Chi2(4) = 278.30; Prob > chi2 = 0.0000

Log Likelihood = -599.44488 — Pseudo R2 = 0.1884

Byssinosis	coef.	Std Error	z	P>|z|	95% Conf.	Interval
Wrkplace	2.669849	.1696234	15.74	0.000	2.337393	3.002305
Smoker	.6193696	.1907348	3.247	0.001	.2455187	.9932205
Year1	.5018791	.2488369	2.017	0.044	.0141678	.9895904
Year2	.6707002	.1813142	3.699	0.000	.315331	1.02607
_cons	-.5121706	.2170068	-23.602	0.000	-5.547032	-4.696381

Logistic byssinosis wrkplace year1 year2 [fweight=count]

Log Estimates — No of Obs = 5419; Chi2(4) = 278.30; Prob > chi2 = 0.0000

Log Likelihood = -599.44488 — Pseudo R2 = 0.1884

Byssinosis	Odds Ratio	Std Error	z	P>|z|	95% Conf.	Interval
Wrkplace	14.43779	2.448986	15.74	0.000	10.35421	20.13188
Smoker	1.857757	.3543555	3.247	0.001	1.278284	2.269916
Year1	1.651822	.4110343	2.017	0.044	1.014269	2.690132
Year2	1.955606	.3545791	3.699	0.000	1.370713	2.790078

a. What variables are statistically significantly associated with development of byssinosis? Why? [8]

b. Write down the regression model [4]

c. What does the coefficient of 2.67 for the variable “workplace” mean? [4]

d. Provide an explanation of the interpretation of the 95% confidence for the odds ratio for the variable “smoker”. [4]

QUESTION 7

6 a). In a study conducted by Familiar et al (1989) two drugs for the treatment of peptic ulcer were compared. The percentage of ulcers healed by the drug pirenzepine was 75% and 60% of ulcers were healed by trithiozone.

(i) How many subjects would be required for a randomised trial to have 80% chance of detecting this difference (75% and 60%) at the 0.05 (two-tailed) level of significance? [2]

(ii) How many subjects would be required to increase the power to 90%? [2]

(iii) Taking the healing rate for trithiozone to be 60%, what sample size would be required to detect an increase of 10% in the healing rates due to pirenzepine? (Use 80% power) [3]

(iv) If only 20% of the subjects are to be given trithiozone, what sample size is now required to detect the difference described in part (i) [3]

6 b). The role of circulating catecholamines in essential hypertension has been a subject of intense study. You wish to carry out a study comparing serum catecholamines levels in normotensive patients and patients with essential hypertension. Previous studies have found mean serum catecholamines levels of 0.218mg/mL. (sd = 0.14) in normotensives

(i) If the clinically important difference to be detected in catecholamines levels in hypertensive patients is an increase by 0.1 mg/m. How many subjects would you sample? [2]

(ii) How many subjects would you need if the difference to be detected was
(a) 0.05 mg/mL [2]
(b) 0.025 mg/mL [2]

(iii) The study to investigate catecholamines levels will be under taken on patients attending their general practitioner. Since 20 subjects will be selected for each GP enrolled in the study, the issue of clustering may be a problem. What sample size is required to detect a difference of 0.1 mg/mL with 90% power if the intra-cluster correlation coefficient is estimated to be:
(a) 0.05 [2]
(b) 0.10 [2]

QUESTION 8

Data from 37 patients receiving a non-depleted allogeneic bone marrow transplant were examined to see which variables were significantly related to survival time. Backward stepwise Cox regression analysis using diagnosis (CML or not), recipient's age and sex, donor's age and sex, whether the donor had been pregnant, the index of mixed epidermal cell-lymphocyte reactions and whether or not patient developed graft vs host disease (GvHD) to predict survival yields the following model:

Variable	**Regression Coefficient**	**Standard Error**
GvHD (0=No, 1= Yes)	2.603	0.5898
CML (0=No, 1= Yes)	-2.805	0.80956

a. What is the interpretation of the opposite signs for the regression coefficients? [2]

b. Calculate the relative risks of dying (hazard ratios) for the following patients relative to non-GvHD non-CML patients:

   (i) With GvHD but not CML

   (ii) CML but without GvHD

   (iii) CML and GvHD [9]

c. Calculate the 95% confidence interval for the hazard ratio associated with GvHD [6]

d. Comment on the reliability of the Cox regression model in view of the sample size (37) and the number of deaths (18). [2]

QUESTION 9

Ten health care workers who were accidentally stuck with needles contaminated with HIV were followed for onset of AIDS. Six of the workers developed AIDS at 36, 65, 90, 110, 140 and 121 months. Four of the workers had still not developed AI DS at the time they were last contacted which was 40, 75, 130 and 160 months after the needle stick occurred.

a. What statistical method would you use to find the median incubation period (time from HIV exposure to AIDS) ? [2]

b. Complete the table below and hence draw the Kaplan-Meir graph [15]

c. Estimate the median incubation period. [3]

Event Number	Time of Aids (t)	# who fail at time t	# at risk at time t	q	$p=1-q$	$S(t+)$	
0	0	0	10	0	1	1	
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							