"Investing in Africa's Future"

COLLEGE OF HEALTH, AGRICULTURE AND NATURAL SCIENCES

SPH 544: APPLIED BIOSTATISTICS

END OF SECOND SEMESTER SUPPLEMENTARY EXAMINATIONS

AUGUST 2021

LECTURER: MR. E. CHIKAKA

DURATION: 7 HOURS

INSTRUCTIONS

ANSWER **ANY ONE** QUESTION

PLEASE STICK TO THE STANDARD HOUSE STYLE i.e.

- TIMES NEW ROMAN
- FONT SIZE 12
- DOUBLE SPACING
- APA REFERENCING
- SEND YOUR ANSWER AS A PDF DOCUMENT

THE MARK ALLOCATION FOR EACH QUESTION IS INDICATED AT THE END OF THE QUESTION

CREDIT WILL BE GIVEN FOR LOGICAL, SYSTEMATIC AND NEAT PRESENTATION

QUESTION 1

a. "Public Health is the science and art of preventing disease, prolonging life and promoting health through the organized efforts of society." (World Health Organization)
What are the functions of Public Health and the role of BIOSTATISTICS in each and every role? [20]

b. Why is it important to know the type of data that you collect for any decision-making process? What difference does it make whether we measure in terms of a nominal, ordinal, interval or ratio scale? Explain giving examples. [10]

c. Are the following nominal, ordinal, interval or ratio data? Explain your answers.
(i) Temperatures of patients in outpatients dept
(ii) Military ranks.
(iii) Students Registration Numbers.
(iv) Number of patients admitted at Mutare Provincial Hospital.
(v) Size of T-Shirts . [5]

d. The population of male workers at Africa University who have never experienced a major coronary event has a mean systolic blood pressure of 140 mmHG and a mean diastolic blood pressure of 82 mm Hg. You are interested in determining whether these values are the same for the population of workers at Africa University who have suffered a coronary event.

i. A sample of 96 workers who have experienced a major coronary event has a mean systolic blood pressure $x = 14\overline{5}$ mm Hg and standard deviation of s = 14.4 mm Hg. Test the null hypothesis that the mean systolic blood pressure for the population workers who have experienced such an event is the same as the mean for the workers who have not, using a two–sided test at the $\alpha = 0.05$ level. [10]

ii. The same sample of men had a mean diastolic blood pressure of $\bar{x} = 85$ mm Hg and standard deviation s = 11.5 mm Hg. Test the null hypothesis that the mean diastolic blood pressure for the population of workers who have experienced a major coronary event is the same as the mean of workers who have not. [10]

e. Explain how you would work out the following statistical measures often used by statisticians?
(i) Coefficient of variation;
(ii) Kurtosis;
(iii) Coefficient of skewness;
(iv) Regression equation of X on Y;
(v) Coefficient of determination [10]

f. Distinguish between correlation and regression. [5]

g. Give your understanding of non-parametric or distribution free methods explaining their important characteristics. [10]

h. Describe briefly the commonly used sampling distributions and why they are used in biostatistics. [5]

i. What is Chi-square test? Explain its significance in statistical analysis pointing out the following: Additive property of Chi-square; Chi-square as a test of 'goodness of fit'; Precautions in applying Chi-square test and Conditions for applying Chi-square test. [5]

j. Ten young recruits were put through a strenuous physical training programme by the army. Their weights (in kg) were recorded before and after with the following results:

Recruit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Weight before	127	195	162	170	143	205	168	175	197	136
Weight after	135	200	160	182	147	200	172	186	194	141

Using 5% level of significance, should we conclude that the programme affects the average weight of young recruits? (Answer using *t*-test or *A*-test) [10]

QUESTION 2

a. Distinguish with examples the different types of scales of measurement and how each type is presented. [10]
b. A researcher would want to use non-parametric tests to analyze his data. Point out the various tests that he could use and their pros and cons. Also point out the parametric equivalents of the tests you would have discussed. [10]
c. What are the roles of a Biostatistician in Public Health? [5]
d. Clearly explain how you test the equality of variances of two normal populations. [5]
e. What is a *t*-test? When is it used and for what purpose(s)? Explain your answer by means of examples [5]
f. A study is conducted concerning the blood pressure of 50 year old women with glaucoma. In the study 500 60-year old women with glaucoma were randomly selected and the sample mean systolic blood pressure is 139.6 mm Hg and the sample standard deviation is 20 mm Hg.
Calculate a 99% confidence interval for the true mean systolic blood pressure among the population of 60 year old women with glaucoma. [10]

g. An experiment was conducted to test the efficacy of Chloromycetin in checking typhoid. In a certain hospital Chloromycetin was given to 285 out of the 392 patients suffering from typhoid. The number of typhoid cases were as follows:

	Typhoid	**No Typhoid**	**Total**
Chloromycetin	35	250	285
No Chloromycetin	50	57	107
Total	85	307	392

Test at 5 per cent level of significance the effectiveness of Chloromycetin in checking typhoid. [10]

h. What are the properties and assumptions for a normal distribution? [10]
i. What are the assumptions for correlation and regression and how do you check for them if you are given a dataset [10]

j. Suppose you have data on age and systolic blood pressure (SBP) from a random sample of 20 adult females. The data are presented as follows:

Subject	Age (years)	SBP (mmHg)
1	22	131
2	24	116
3	28	114
4	29	123
5	30	117
6	32	122
7	35	121
8	41	171
9	47	111

10	49	133
11	51	130
12	51	133
13	56	145
14	57	141
15	63	155
16	77	217
17	77	156
18	78	155
19	79	149
20	80	156

i. Identify the dependent and independent variable [2]

ii. Draw the scatter plot for this data. Does it suggest anything about the relationship between these variables? [2]

iii. Using systolic blood pressure as the response variable, compute the least squares regression line. [10]

iv. Interpret the slope and the *y*-intercept of the line. [4]

v. Use your model to predict the systolic blood pressure of a female aged 40 years.[2]

j. What is the difference between Cluster and Stratified sampling? [5]

QUESTION 3

a. List and explain all the data sources and data types you know. Why is it important to know the types of data for any decision-making process? [10]

b. What do measures of central tendency and variation indicate? Describe the important measures of central tendency and variation pointing out the situation when one measure is considered relatively appropriate in comparison to other measures. [10]

c. What is Chi-square test? Explain its significance in statistical analysis pointing out the following: Additive property of Chi-square; Chi-square as a test of 'goodness of fit'; Precautions in applying Chi-square test and Conditions for applying Chi-square test. [8]

d. In a cross-sectional survey administered to a random sample of 150 attendees of a local health fair, there were a total 95 smokers and 90 diabetics. 40 of the attendees were diabetic non-smokers.

Current Smoker	Diabetes Yes	No	Totals
Yes			
No			
Totals			

We want to determine, using a level of significance (α) of 0.05, if the risk of having diabetes in the surveyed population is related to smoking. [10]

e. Clearly explain how you will test the equality of variances of two normal populations

f. What is a *t*-test? When is it used and for what purpose(s)? Explain your answer by means of examples. [5]

g. Point out the important limitations of tests of hypotheses. What precaution must a statistician take while drawing inferences as per the results of the said tests? [5]

h. Narrate the various advantages of using non-parametric tests. Also point out their limitations. [5]

k. What are the assumptions of the logistic regression analysis and how do you check for them? [5]

l. This study evaluated a maternal and child health program in which all pregnant women in the intervention area were monitored by village health workers. During the one year study period data were collected during pregnancy, at delivery, and 28 days after delivery. Only single births of 28 weeks or more gestational age were included in the study. The primary outcomes of interest were birth weight and the infant's survival status at 28 days.

The Village Health Worker Programme involved the selection of two health workers from each village in the intervention area. Their function was to provide simple treatment such as iron supplementation for anaemia, to inform pregnant women of services available to them from the health centre, to monitor them throughout pregnancy, to identify high risk cases and arrange for necessary referrals. The pregnancy and delivery were monitored via the completion of the mother's card. This card contained demographic and medical data and information was recorded on the card throughout pregnancy together with the outcomes of delivery.

During the one year study period there were 939 single deliveries in the intervention area and 944 in the control area. For the purpose of this exercise, records for 65 cases of perinatal death and 400 survivors were selected. The variables included in the dataset are:

Variable name	Interpretation	Codes
Status	infant's survival status at 28 days	0= alive 1= dead 1= abnormal
Birthwt	infants birth weight in grams	grams (cont. variable)
Mothage	mother's age in years	years (cont. variable)
DCHILD	no of live births now dead of mother	0 = none 1 = 1+

. logit status birthweight mothagedchild

Logit estimates — Number of obs = 465; LR chi2(3) = 36.86; Prob> chi2 = 0.0000; Log likelihood = -169.69649; Pseudo R2 = 0.0980

status	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
birthweight	-0.0014929	0.0003285	-4.54	0.000	-0.0021368	-0.0008491
mothage	-0.0158665	0.0287126	-0.55	0.581	-0.0721421	0.0404092
dchild	0.7844647	0.2196172	3.57	0.000	0.3540228	1.214907
_cons	2.665993	1.138809	2.34	0.019	0.4339674	4.898018

i. Using the logistic regression separately examine the relationship between each of the independent variables birth weight (as continuous), mothage and dchild and the outcome variable STATUS [6]

ii. From the logistic regression models fitted in part (a) calculate the odds ratios corresponding to:

(a) an increase in birth weight of 100grams
(b) an increase in mother's age of 5 years
(c) one additional dead child [6]

iii. Give an interpretation of each of these odds ratios [6]

iv. Which independent variables are statistically significantly related to infant's survival at 28 days after adjusting for the others? [2]

v. What is the interpretation of Pseudo R2 = 0.0980 in this analysis? What is your comment on the model? [2]

m. i. What are the assumptions for normality? Use the sample of size 48 to check whether it is normally distributed. [15]

14.1	18.5	21.3	26.7	28.8	30.9	32.1	36.6	37.2	38.5	39.9	41.7
43.8	44.9	45.0	46.0	49.3	50.4	52.3	52.9	53.0	54.2	54.7	54.7
54.7	54.8	55.4	55.9	56.0	56.7	57.1	57.6	57.9	58.2	58.3	58.7
58.9	59.2	59.8	60.3	60.8	60.9	61.0	61.1	62.4	62.6	63.0	63.4

ii. What proportion of the data lie within one standard deviation of the mean? How does your answer compare to the empirical rule approximation? [5]