CHUKA UNIVERSITY

UNIVERSITY EXAMINATIONS

EXAMINATION FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS

MATH 858: STATISTICAL METHODS FOR BIOSTATISTICS AND EPIDEMIOLOGY

STREAMS: MSC Y1S2 **TIME: 3 HOURS**

DAY/DATE: FRIDAY 06/12/2019 **2.30 P.M. – 5.30 P.M.**

INSTRUCTIONS:

- **Answer any THREE questions.**
- **Do not write anything on the question paper.**

QUESTION ONE (20 MARKS)

Use the data in table below to answer the following questions:

Age group (years)	Population (x1,000)	Number of HIV deaths	Number of Leukemia deaths
0-4	18,597	12	125
5-14	41,037	25	316
15-24	40,590	178	472
25-34	39,928	1,839	471
35-44	44,917	5,707	767
45-54	40,084	4,474	1,459
55-64	26,602	1,347	2,611
65+	35,602	509	15,277
Not stated		4	0

(a) Compare the mortality rates at all ages, under 65 years, and years of potential life lost for Leukemia and HIV. (16 marks)

(b) Which measure(s) might you prefer if you were trying to support increased funding for leukemia research and for HIV research? Support your choice(s) (4 marks)

QUESTION TWO (20 MARKS)

Discuss any **FOUR** study designs used in clinical studies, arguing for or against application of various epidemiological and statistical tools for each design.

QUESTION THREE (20 MARKS)

(a) The number of deaths from all causes and from accidents by age group in a given region in 2016 are provided on table below. Review the following rates and determine what to call each one, then calculate it. (8 marks)

Age group (years)	All causes	Both sexes Accidents	Estimated Population (x1,000)	All causes	Males Accidents	Estimated Population (x1,000)
0-4	32,892	2,587	19,597	18,523	1,577	10,020
5-14	7,150	2,718	41,037	4,198	1,713	21,013
15-24	33,046	15,412	40,590	24,416	11,438	20,821
25-34	41,355	12,569	39,928	28,736	9,635	20,203
35-44	91,140	16,710	44,917	57,593	12,012	22,367
45-54	172,385	14,675	40,084	107,722	10,492	19,676
55-64	253,342	8,345	35,602	806,431	16,535	14,722
65+	1,811,720	33,641	35,602	806,431	16,635	14,772
Not stated	357	85	0	282	74	0

(b) Using a hypothetical example, differentiate between clinical significance and statistical significance. (7 marks)

(c) The following computer output show a set of results form a poison regression analysis, featuring

(i) The number of accidents per mine in a 3 month period in 44 coal mines in a given region and

(ii) deaths from childhood cancers classified by cytology (lymphoblastic/myeloblastic) and residence (rural/urban) age 6-14. Interpret the output. (5 marks)

(i)

Coefficients:

	Estimate	Std. Error	Z value	P r(> |z|)	
(intercept)	-3.6097078	1.0284740	-3.510	0.000448	***
INB	-0.0014441	0.0008415	-1.716	0.086145	.
EXTRP	0.0622011	0.0122872	5.062	4.14e-07	***
AHS	-0.0017578	0.0050737	-0.346	0.729003	
AGE	-0.0296244	0.0163143	-1.816	0.069394	.

Where, INB: inner burden thickness; EXTRP: percentage of coal extracted from mine; AHS: the average height of the coal seam in the mine, and AGE: the age of the mine

(ii)

Coefficients:	Estimate	Std. Error	Z value	Pr ($>\|z\|$)	
(Intercept)	3.3893	0.1465	23.139	$<2e-16$	***
Cytology M	-1.5983	0.2584	-6.184	$6.24e-10$	***
Age 6 – 14	-0.9821	0.1767	-5.557	$2.75e-08$	***
Residence U	0.3677	0.1546	2.379	0.01736	*
Cytology M: Age 6 – 14	1.0184	0.3500	2.910	0.00362	**

Null deviance: 92:4517 on 7 degrees of freedom
Residual deviance: 5.0598 on 3 degrees of freedom
AIC: 52.858

QUESTION FOUR (20 MARKS)

A laboratory technologist designed an experiment to study the growth of a particular strain of bacteria. It is suspected that the bacteria growth is influenced by temperature and environment and thus the researcher carried out the experiment at four different temperature and three levels of nutrient medium. Due to the length of time required to observe the bacteria growth, the experiment was replicated over five days with the days forming blocks.

(a) Give the design model for the experiment. (2 marks)

(b) Analyse the following results which represents totals over the five days and draw appropriate conclusions give that Total; SS=959.35 and Block SS=421.6. (10 marks)

Temperature (T) Nutrient (N)	T1	T2	T3	T4
N1	74.8	89.0	96.6	102.2
N2	78.4	99.8	109.2	112.5
N3	78.1	94.6	98.6	105.9

(c) Discuss the use of mean separation procedures in data analysis. Apply one of them in the above analysis and make your comment. (8 marks)
