



PAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF HEALTH, NATURAL RESOURCES AND APPLIED SCIENCES
SCHOOL OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSHS	LEVEL: 8
COURSE CODE: BIO801S	COURSE NAME: BIOSTATISTICS
SESSION: JUNE 2023	PAPER: THEORY
DURATION: 3 HOURS	MARKS: 100

FIRST OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINER	Dr D. B. GEMECHU
MODERATOR:	Prof L. PAZVAKAWAMBWA

INSTRUCTIONS
<ol style="list-style-type: none">1. There are 5 questions, answer ALL the questions by showing all the necessary steps.2. Write clearly and neatly.3. Number the answers clearly.4. Round your answers to at least four decimal places, if applicable.

PERMISSIBLE MATERIALS

1. Non-programmable scientific calculator

THIS QUESTION PAPER CONSISTS OF 7 PAGES (Including this front page)

Question 1 [21 marks]

- 1.1 Briefly explain the following terminologies as they are applied to Biostatistics.
- 1.1.1 Right-censored observation [2]
 - 1.1.2 Survival function [2]
 - 1.1.3 Hazard function [2]
- 1.2 Briefly discuss the following study designs (your answer should include definition/uses, advantage and disadvantages).
- 1.2.1 Prospective Cohort study [3]
 - 1.2.2 Cross-sectional studies [3]
- 1.3 To investigate the association between Kawasaki syndrome (KS) and carpet shampoo, investigators conducted a case-control study with 100 cases (100 children with KS) and 100 controls (100 children without KS). Among children with KS, 50 gave a history of recent exposure to carpet shampoo. Among those without KS, 25 gave a history of recent exposure to carpet shampoo.
- 1.3.1 Can we compute relative risk of Kawasaki syndrome? Why or why not? [2]
 - 1.3.2 Compute the Odds ratio of Kawasaki syndrome and interpret your result [2]
- 1.4 Suppose you were asked to analyze the data from a small preliminary clinical trial with 20 subjects. All subjects had a comparable degree of knee pain from osteoarthritis, and they were being compared with respect to pain relief after receiving a standard pain medication (Drug B) or a new pain medication (Drug A). The 20 patients were randomly assigned to one drug or the other, and there were ten subjects in each group. After receiving the medication, the investigators checked on the subjects at hourly intervals to see if the subjects had had relief of pain. For each subject, the time at which pain relief occurred was recorded. Results are illustrated in Table 1. Which group appears to have had a greater incidence rate of pain relief? [5]

Table 1: Preliminary clinical trial results, time at pain relief, on 20 subjects. **Key:** 0 = subject did not report relief of pain, x = subject reported pain relief, and — = continued follow-up of a subject.

	Subject	Hours										
		1	2	3	4	5	6	7	8	9	10	
New drug	1	—	x									
	2	—	—	—	—	—	—	—	—	—	—	0
	3	—	x									
	4	—	—	—	—	—	—	—	—	—	—	0
	5	—	—	—	—	—	—	—	—	—	—	0
	6	—	x									
	7	—	—	—	—	—	—	—	—	—	—	0
	8	—	—	x								
	9	—	—	—	x							
	10	—	x									
Old drug	1	—	—	—	—	—	—	—	x			
	2	—	—	—	—	—	—	—	—	—	—	0
	3	—	—	—	—	—	—	—	x			
	4	—	—	—	—	—	—	—	—	—	—	0
	5	—	—	—	—	—	—	—	—	—	—	0
	6	—	—	—	—	—	—	—	x			
	7	—	—	—	—	—	—	—	—	—	—	0
	8	—	—	—	—	—	—	—	—	x		
	9	—	—	—	—	—	—	—	—	—	x	
	10	—	—	—	—	—	—	—	—	x		

Question 2 [20 marks]

2.1 If the random variable Y has Pareto distribution with a parameter θ , then its probability density function is

$$f(y, \theta) = \theta y^{-\theta-1}.$$

- 2.1.1 Show that this distribution belongs to the exponential family and find the natural parameter. [5]
- 2.1.2 Find the score statistics U . [3]
- 2.1.3 Find variance of $a(y)$. [4]
- 2.1.4 Find the information \mathcal{I} [2]
- 2.1.5 If a random sample y_1, y_2, \dots, y_n of size n were selected to estimate the parameter θ numerically, derive the Newton-Raphson approximation estimating equation that will be used obtain the maximum likelihood estimator of θ . [6]

Question 3 [21 marks]

3. Anaemia is a condition in which the number of red blood cells or haemoglobin concentration is reduced below normal levels, thus resulting in reduced oxygen carrying capacity (WHO, 2015). Anaemia is more prevalent among pregnant women and children under five years. Besira (2021) conducted a study to determine the prevalence of anaemia and associated risk factors among pregnant women attending ANC at Katutura Health Centre using the multiple logistic regression model.

The response variable: 1: the pregnant women is anaemic; 0: the pregnant women is non-anaemic. **The explanatory variables:** Age in years; HIV/AIDS status (Positive or negative); Number of live birth also called para (0, 1, ≥ 2); Trimester (1st trimester, 2nd trimester, 3rd trimester); Nutrition status (Malnourished or not-malnourished); number of pregnancies also called gravida (1, 2, ≥ 3). The multiple logistic regression fitted were given in Table 2.

Table 2: Model summary for anaemia among pregnant women attending ANC at KHC in Namibia

Risk Factor	Coeff (bj)	s.e. (bj)	Z-value	P-value	OR	95% CI
Intercept	-23.48	0.888	16.015	< 0.001	0.029	
Age in years	1.360	0.653	4.340	0.037	3.897	(1.084, 14.016)
HIV/AIDS status: Positive	-0.289	0.434	0.443	0.506	0.749	(0.320, 1.755)
Para (ref: 0)						
1	0.279	0.739	0.143	0.706	1.322	(0.311, 5.624)
≥ 2	0.287	0.790	0.132	0.716	1.333	(0.283, 6.267)
Trimester (ref: 1st trimester)						
2nd trimester	1.064	0.761	1.958	0.162	2.899	(0.653, 12.873)
3rd trimester	1.643	0.772	4.534	0.033	5.172	(1.140, 23.469)
Nutrition status: Malnourished	0.865	0.305	8.041	0.005		
Gravida (ref: 1):						
2	-0.107	0.764	0.020	0.889	0.899	(0.201, 4.013)
≥ 3	0.439	0.818	0.288	0.591	1.552	(0.312, 7.715)

- 3.1 Assess the statistical significance of the individual risk factors. [3]
- 3.2 Give brief interpretations of the age in years and gravida coefficients. [4]
- 3.3 Compute and interpret the odds ratios relating the additional risk of anaemia with malnourishment after adjusting for the other risk factors. [2]
- 3.4 Compute and interpret a 95% confidence intervals for the odds ratio in part (3.3) [3]
- 3.5 Find estimated change in the odds ratio when age of the pregnant women increases by 2 years. [2]
- 3.6 Compute odds ratio comparing pregnant women in her third trimester relative to the pregnant women in her second trimester and interpret your answer. [3]
- 3.7 Predict the probability of being anaemic for a first time pregnant 18 years old women who was well nourished and HIV/AIDS negative with no history of live birth and was in her second trimester. [4]

Question 4 [19 marks]

4. Table 3 provides a nominal logistic regression model for the relationship between the level of back pain during work (0=no pain, 1= mild pain and 2 =sever pain) and factors such as age categories (0=18-35 years and 1 = above 35 years) and smoking status (0 = never smoked, 1= ex-smoker and 2=current smoker) of the workers. Answer the questions based the result presented.

Table 3: Model summary for level of back pain during work

Parameter	Estimate	std. error)	Odds ratio (95% CI)
<i>Log(π_2/π_1): mild pain vs. no pain (ref)</i>			
Intercept	-3.3128	0.1909	
Age (older):	0.5380	0.1713	
Smoking status (ex-smoker):	0.7881	0.2588	2.20 (0.2809, 1.2954)
Smoking status (current smoker):	0.8319	0.2140	2.30 (0.4126, 1.2513)
<i>Log(π_3/π_1): sever pain vs. no pain (ref)</i>			
Intercept:	-5.1447	0.4073	
Age (older):	1.3785	0.2855	3.97 (0.8189, 1.9381)
Smoking status (ex-smoker):	0.8223	0.5031	2.28 (-0.1638, 1.8084)
Smoking status (current smoker):	1.3465	0.4164	3.84 (0.5304, 2.1626)
log-likelihood function for the fitted model: -791.3756 (df=8)			
log-likelihood function for the null model: -19.77502 (df=2)			

- 4.1 Express the fitted model using appropriate expression and describe its components. [3]
- 4.2 Test the overall importance of the explanatory variables using likelihood ratio test. [4]
- 4.3 Construct a 95% confidence limit for the odds ratio of older age in the first model. [3]
- 4.4 Assess the statistical significance of the individual explanatory variables. [2]
- 4.5 Comment on the odds ratio of the variable age. [2]
- 4.6 Compute the estimated probability by considering younger age worker who was never smoker. [5]

Question 5 [19 marks]

5.1 COVID-19 is well-known for its rapid spread by asymptomatic carriers, which results in a rapid increase in COVID-19 patients in a short period of time. Even if some patients have minor symptoms and do not require hospitalization, the hospital may become overcrowded because of bed seeking patients, and some of them may require admission to the Intensive Care Unit (ICU) and oxygen. Pietersen and Gemechu (2021) conducted a study to investigate factors impacting length of hospital stay of COVID-19 patients admitted at Katutura state hospital using survival analysis technique. The variables included in the model are:

time: Survival time in days

status: censoring status 1=censored, 2=dead

Factors: sex (male or female), age (< 45 , $45 - 65$, > 65), comorbidities (yes or no), and admission wards (Respiratory unit or other unit). The results of the authors were given in Table 4 and Figure 1.

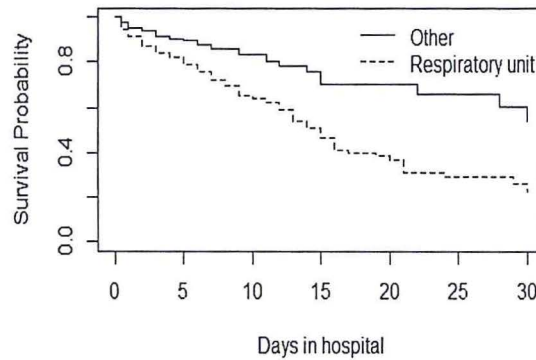


Figure 1: Kaplan Meier survival Curve by admission ward

Call:

```
coxph(formula = Surv(time, status) ~ Sex + 'Age in years' +  
comorbidities + Admission_ward, data = coviddata)
```

Table 4: Factors Associated with survival time by Cox PH Model.

	Estimate	HR	95% CI for HR
sex:Male	0.3819		(1.1474, 1.871)
Age groups, Ref: (45 - 65)			
age <45	-0.5211	0.5939	(0.4096, 0.861)
age > 65	0.4267	1.5321	(1.1722, 2.003)
Comorbidities:Yes	0.5941	1.8114	(1.0072, 3.258)
Ad. Ward: Respiratory Unit	0.8680	2.3822	(1.6777, 3.383)

- 5.1.1 Briefly comment on the Kaplan Meier survival curve. What is the approximate median survival times for patients admitted to respiratory unit? [3]
- 5.1.2 Assess the statistical significance of the individual risk factors. [2]
- 5.1.3 What is the interpretation of the coefficient for the variable "sex" in Table 4? Compute and interpret the hazard ratio. Which gender has a better survival chance?[4]
- 5.2 Let the random variable Y denote the survival time and let $f(y, \lambda, \phi)$ denote its probability density function defined by

$$f(y, \lambda, \phi) = \lambda \phi y^{\lambda-1} \exp(-\phi y^\lambda),$$

where $\phi = \theta^{-\lambda}$.

- 5.2.1 Derive the hazard function of y [8]
- 5.2.2 Find the cumulative hazard function of y [2]

== END OF QUESTION PAPER ==
Total: 100 marks