UNIVERSITY OF FORT HARE

STE 706

DEGREE EXAMINATIIONS

NOV 2018

Time: 3 HOURS

Subject: Categorical Data Analysis

Marks: 100

This paper consists of 3 pages including cover page

Internal Examiner

Prof Y. Qin

External Examiner

Prof D. J. de Waal

Instructions

Answer all questions

Statistical tables will be provided

Categorical Data Analysis Paper

Question 1(15 marks). A balanced coin is flipped twice. Let Y = number of heads obtained.

a. Specify the probabilities for the possible values for Y, and give the distribution's mean and variance. [5 marks]

b. Calculate probabilities for the Poisson distribution having the same mean as the distribution in (a). [5 marks]

c. For each flip of a possible unbalanced coin, let π denote the probability of a head. Suppose there are 0 heads in 2 slips. Find the maximum estimator of π. [5 marks]

Question 2(20 marks). In a 2×2 contingency table related to two random variables X and Y with Y being the response. Suppose that X and Y take two values 0 and 1.

a Give out the definition of difference of proportions and odds ratio in above table. [10 marks]

b Show that if the ratio of odds is not 1 then X and Y are not independent. [10 marks]

Question 3(30 marks). A 2 by 2 contingency table to analyze the association between Gender and smoking status is

	Smoke	No smoke
Male	260	568
Female	150	460

Answer the following questions (first write formulas and then give computation results).

(a). Give out the estimator of the relative risk in the 2 by 2 contingency table. [7 marks]

(b). Write out a large-sample confidence interval for the difference of proportions related in the 2 by 2 contingency table. [7 marks]

(c). Write out the estimated odds ratio related in the 2 by 2 contingency table. [6 marks]

(d). Conduct a test of independence between gender and smoking. [10 marks]

Hint: $P(\chi_1^2 \leq 3.8459) = P(\chi_4^2 \leq 9.487729) = 0.95$.

Question 4(15 marks). (a). List four commonly used link functions for generalized linear models and give the ranges for the mean response to use these link functions. [7 marks]

(b). Give out an example to explain the concepts of partial tables and marginal table in a 3-way contingency table. [8 marks]

Question 5(20 marks). Suppose that y and x are two univariate random variables which are the predictor and response respectively. Suppose that y takes two values 0 and 1. Denote $\pi(x) = P(y = 1|x)$. We use the following logistic regression model (1) to fit a set of date $\{(x_i, y_i), 1 \leq i \leq n\}$. Answer the following questions:

$$log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x. \tag{1}$$

(a). Suppose that the estimators of $\alpha = 2.3$ and $\beta = 0.8$. Give out the estimators of $\pi(x)$ and $\pi(0)$. [7 marks]

(b). Suppose that the ASE of $\hat{\beta}$ is 0.1. Construct the 95% confidence interval for β. [7 marks] (Hint $P(N(0,1) > 1.96) = 0.025$).

(c). Test hypothesis $H_0 : \beta = 0$. Use 0.05 as the significant level. [6 marks] (Hint $P(\chi_1^2 > 3.84) = 0.05$).