Comprehensive Exam (2020)
Introduction to Categorical Data Analysis

Name: Stuent id#: Department:

Please show your work in your answer sheet.

1. (22 points) Consider the following $2 \times 2 \times 2$ table, which features joint probabilities for a population of professionals with college degrees. The classification variables are political affiliation (Republican or Democrat), income level (low or high), and educational level (bachelor's or mater's/doctorate).

		Income Level	
Educational Level	Political Affiliation	High	Low
Master's/Doctorate	Republican	.32	.08
	Democrat	.08	.02
Bachelor's	Republican	.08	.12
	Democrat	.12	.18

   (a) Use conditional odds ratios to prove that political affiliation and income level are conditionally independent, given educational level.

   (b) Use a marginal odds ratio to prove that political affiliation and income level are marginal dependent. Give a precise interpretation of this odds ratio in terms of the context of the application.

   (c) Explain why the results of (a) and (b) are different. Which one should you choose? (Why?)

2. (22 points) Woodward et al. (1941) describes a study which examined the relationship between exposure levels of chloracetic acid in mice and survival. At each of 12 levels, 10 mice were administered a dosage. The number of mice who perished as a result of exposure to the dosage was recorded. The data presented in the table below. The dosage are in units of grams per kilogram of body weight.

Dosage	Deaths	Survivals
0.0794	1	9
0.1000	2	8
0.1259	1	9
0.1413	0	10
0.1500	1	9
0.1588	2	8
0.1778	4	6
0.1995	6	4
0.2239	4	6
0.2512	5	5
0.2818	5	5
0.3162	8	2

The SAS output on next page features the results of fitting a linear logit model to the data using PROC GENMOD. An event case is regarded as one where the mouse perishes. Use the SAS output in addressing the following questions.

(a) Perform a hypothesis test based on the likelihood-ratio statistic to determine the goodness-of-it of the linear logit model. (State H_0 and H_a, provide the numerical value of the test statistic, provide the p-value, and state the conclusion).

(b) Perform the Wald test to assess whether the log odds of perishing changes linearly with the dosage. (State H_0 and H_a, provide the numerical value of the test statistic, provide the p-value, and state the conclusion).

(c) We know that an estimated covariance of $\hat{\beta}_0$ (intercept) and $\hat{\beta}_1$ (coefficient of dosage) is 1.5. Using this result and the SAS output, calculate a 95% confidence interval for the logit of the probability of being dead when the dosage is .3162.

SAS output for Problem 3

The GENMOD Procedure

Model Information

Data Set	WORK.MICE
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	Deaths
Response Variable (Trials)	n
Number of Observations Read	12
Number of Observations Used	12
Number of Events	39
Number of Trials	120

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	10	10.2537	1.0254
Scaled Deviance	10	10.2537	1.0254
Pearson Chi-Square	10	8.7421	0.8742
Scaled Pearson X2	10	8.7421	0.8742
Log Likelihood		-63.9447	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-3.5697	0.7053	-4.9522	-2.1873	25.61	<.0001
dosage	1	14.6369	3.3325	8.1053	21.1685	19.29	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
dosage	1	23.45	<.0001

3. (18 points) The logit model may be used for categorical data when there is a binary response variable. Consider a three-way table with variables X, Y and Z of levels I, 2 and K, respectively. Here Y is the binary response variable. The logit is defined as the log of the odds of success for someone having an i, k classification ($i = 1, \ldots, I$; $k = 1, \ldots, K$):

$$\log \{P(Y = 1|X = i, Z = k)/P(Y = 2|X = i, Z = k)\} .$$

Consider two loglinear models (XY, YZ) and (XY, XZ, YZ). Show that under the models the corresponding logit models are equivalent.

4. (18 points) Prove the following result:
For three-way table for three random variables (X, Y, Z), XY marginal and conditional odds ratios are identical if either (1) Z and X are conditionally independent or (2) Z and Y are conditionally independent.