Name:

Analysis of Categorical Data (BIOS 665)
Midterm Exam Date: October 15, 2019
Take-home due date: Hardcopy due on October 24, 2019 at 11:00am

Requirements:

- For estimates, confidence intervals, and tests, simply copying and pasting SAS output without any commentary will not earn full credit. Highlighting is not considered commentary. However, commentary can be as simple as: "The 95% CI for the odds ratio is (___, ___)."
- For the take-home exam, **print each question on a separate page**, and **put your name on every page**. Do not staple the entire exam together, but do staple individual problems together if they span multiple pages. Bring your exam to class on the due date, where you will distribute each problem into the appropriate stack (one per problem). This will facilitate the grading process.
- For the in-class exam, please note that p-values may be reported as ranges based on the table given below. For example, your answer may be '$0.025 < p < 0.05$'. However, for the take-home exam, you should report p-values more precisely (such as 3 decimal places).
- In-class exam: you may use one side of one 8.5 x 11" sheet of paper as a formula sheet. Please be sure your name is on your formula sheet, and submit this along with your exam.
- For each hypothesis test, provide the null hypothesis, test statistic (where appropriate), degrees of freedom (where appropriate), and conclusion.

Honor Pledge: I have neither received nor given unauthorized aid on this exam.

Signed: _____

Chi-Square Distribution: Table of quantiles/critical values ($\chi^2_{df,1-\alpha}$)

df/α	0.10	0.05	0.025	0.01	0.001
1	2.71	3.84	5.02	6.63	10.83
2	4.61	5.99	7.38	9.21	13.82
3	6.25	7.81	9.35	11.34	16.27
4	7.78	9.49	11.14	13.28	18.47
5	9.24	11.07	12.83	15.09	20.52

Z-scores: Quantiles/critical values ($Z_{1-\alpha/2}$)

$Z_{0.8}$=0.842, $Z_{0.9}$=1.282, $Z_{0.95}$=1.645, $Z_{0.975}$=1.960, $Z_{0.99}$=2.326, $Z_{0.995}$=2.576

Name:

Part I

Table 1 contains data from a multi-city study examining survival rates in motor vehicle accidents. There were 421 participants involved in road accidents in a particular year in three cities (A, B, C). Emergency responders requested ambulances (helicopter or road) to transport the victims to the nearest health center. Within 48 hours of the accident, death or survival was recorded. The aim of the study is to determine whether accident survival is associated with the mode of transportation the victim receives.

Table 1

City	Transportation Mode	Accident Survival		
		Yes	No	Total
A	Helicopter	33	47	80
	Road	45	51	96
B	Helicopter	56	136	192
	Road	240	840	1080
C	Helicopter	29	47	76
	Road	4	13	17
	Total	407	1134	1541

1) (10 points) For each city separately, estimate the difference in proportions for accident survival between helicopter transportation and road transportation. For City B only, additionally estimate the 95% confidence interval for this difference in proportions.

2) (10 points) For City A only, provide an estimate of the odds ratio as well as its corresponding two-sided 0.95 confidence interval to compare the mode of transportation (Helicopter vs. Road) in terms of the odds of accident survival (Yes vs. No).

3) (10 points) Under minimal assumptions, test the association between mode of transportation and accident survival, controlling for city. In one sentence, interpret your results.

Name:

Part II

4) (10 points) A randomized trial with two treatment groups is designed to have twice the number of observations in Group A as in Group B with interest in the proportion of events for Group j (i.e., π_j). Determine the sample size required for each group to attain 80% power for testing the hypothesis $H_0: \pi_A = \pi_B$ at the two-sided 0.05 significance level, assuming the true proportion of events in Group A is 0.19 and the true proportion of events in Group B is 0.11.

Name:

Part III

Table 2 provides output from a main effects logistic regression model for data from a randomized study on the level of anxiety before an examination. This logistic model for the probability of no anxiety was fitted to main effects for school, baseline status, and sex, based on reference cell coding as expressed through the descriptions of the parameters in the table. The results include selected estimated parameters, their standard errors and a Wald Chi-Square test.

Table 2

Parameter	Estimate	Standard Error	Wald Chi-Square
Intercept	–4.120	0.237	---
School (A vs. B)	–0.752	---	4.173
Baseline (severe vs. mild)	---*	0.331	0.717
Sex (male vs. female)	3.825	1.347	---

* Note: assume this is a positive value.

5) (10 points) Complete the table so that all values marked as '---' have numeric results.

6) (10 points) Provide an estimate and its two-sided 0.95 confidence interval to describe the extent to which School A has lower odds of no anxiety compared to School B, adjusting for the effects of sex and baseline severity status.

7) (10 points) Conduct a statistical test at 5% significance level to assess whether males have the same odds of no anxiety compared to females, adjusting for the effects of school and baseline severity status.

8) (10 points) Provide model-predicted probabilities of no anxiety for these students:

i). School A, severe baseline status, female.

ii). School B, mild baseline status, male.

Name:

Part IV

Table 3 below classifies patients according to their initial degree of skin damage due to a skin disease and their level of improvement after receiving treatment.

Table 3

Initial degree of skin damage	Levels of Skin Improvement			Total
	None	Slight	Moderate	
High	14	16	22	52
Low	64	42	38	144
Total	78	58	60	196

9) (10 points) Under minimal assumptions, test the association between the initial degree of skin damage in terms of a location shift in the levels of skin improvement using an appropriate test. In one sentence, briefly interpret your results.

Name:

Part V

A pediatric ophthalmology study reported the frequency of eye injuries for each season based on a high-risk sample of 669 children. These data are shown in Table 4.

Table 4

	Eye Injuries		
Season	Yes	No	Total
Winter	45	99	144
Spring	80	119	199
Summer	79	88	167
Fall	87	72	159
Total	291	378	669

10) (10 points) Conduct a statistical test to evaluate whether the occurrence of pediatric eye injuries is the same for level of season ordered from lower risk to higher risk.