Name: _____ Department: _____

Analysis of Categorical Data (BIOS 665)
Midterm Examination 2017

Exam Date: October 17, 2017
Take-home due date: October 26, 2017 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.
- 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, if using software, 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, degrees of freedom, 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

The data in Table 1 are from a randomized, multi-center, controlled clinical trial for the evaluation of a treatment for a neurological disorder in terms of a favorable response after one year.

Table 1

		Response		
Center	Treatment	Favorable	Unfavorable	Total
1	Test	73	27	100
1	Control	56	44	100
	Total	129	71	200
2	Test	36	14	50
2	Control	24	26	50
	Total	60	40	100

1. (10 points) For Center 1, provide an estimate and corresponding two-sided 95% confidence interval for the difference between the probabilities of favorable outcome for those on test treatment versus control.

Name: _____

2. (10 points) For Center 2, provide an estimate and corresponding two-sided 0.95 confidence interval for the odds ratio corresponding to the relationship between treatment (test vs. control) and response (favorable vs. unfavorable).

Name: _____

3. (10 points) Under minimal assumptions, assess the association between treatment and response (controlling for center) with a statistical test at the two-sided 0.05 level. Interpret your results in one sentence.

Name: _____

The data shown in Table 2 are comprised of a particular subgroup of patients from Center 2 of Table 1 with a history of seizures.

Table 2

	Response		
Treatment	Favorable	Unfavorable	Total
Test	4	0	4
Control	2	2	4
Total	6	2	8

4. (10 points) Assuming the table margins are fixed, calculate the probability of each possible 2×2 table which could have been observed with these margins. Provide a relevant table listing these probabilities.

Name: _____

5. (10 points) Calculate an appropriate one-sided p-value (in favor of the test treatment) for the association between treatment and response. State your method and conclusion in one sentence.

Name: _____

6. (10 points) In designing a follow-up study for the test treatment for a distinct population of patients with the neurological disorder in Table 1, you expect favorable response rates of 0.70 for the test treatment and 0.60 for an appropriate control after one year of follow-up. Using a two-sided 0.05 significance level with twice as many patients in the test treatment group as in the control group, determine the sample size that would be necessary for each treatment group in order to provide 0.90 power for this planned study.

Name: _____

Part II

Scientists are interested in the variety of locations to which the spotted moray eel, *Gymnothorax moringa*, and the purplemouth moray eel, *G. vicinus*, travel in a reef off the coast of Belize. For each tracked eel, they identified the species and classified the locations visited by the eel into three types: grass beds, sand and rubble, or within one meter of the border between grass and sand/rubble. The number of different types of locations each eel traveled is shown in Table 3.

Table 3

Eel Species	Number of Different Locations Eel Travels			Total
	1	2	3	
G. moringa	16	29	15	60
G. vicinus	12	20	28	60
Total	28	49	43	120

7. (10 points) Under minimal assumptions, assess the association between species of eel in terms of a location shift in the number of locations traveled with an appropriate statistical test at the two-sided 0.05 level. Provide the null hypothesis, the test statistic, the distribution to which you will compare it (including degrees of freedom), and your determination of statistical significance. In one sentence, interpret your results.

Name: _____

Part III

A logistic model for the probability of having a major adverse cardiac event (MACE) was fit based on reference cell coding for a multi-center drug safety trial with explanatory variables relating to center, baseline cardiovascular disease (CVD) severity, and treatment group. Selected output from SAS is given below.

The LOGISTIC Procedure

Response Profile

Ordered Value	Response	Total Frequency
1	MACE	96
2	No MACE	198

Probability modeled is Response='MACE'.

Class Level Information

Class	Value	Design Variables
Center	1	0
	2	1
Treatment	Active	1
	Placebo	0
Baseline	Mild	0
	Severe	1

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	373.436	368.806
SC	377.120	383.540
-2 Log L	371.436	360.806

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.6308	3	0.0139
Score	10.5221	3	0.0146
Wald	10.2858	3	0.0163

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Center	1	0.9716	0.3243
Baseline	1	9.4384	0.0021
Treatment	1	0.1940	0.6596

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.0820	0.2430	19.8296	<.0001
Center	2	1	-0.2528	0.2565	0.9716	0.3243
Baseline	Severe	1	0.7879	0.2564	9.4384	0.0021
Treatment	Active	1	0.1119	0.2541	0.1940	0.6596

Name: _____

8. (10 points) Specify the mathematical structure of the main effects model, including mathematical definitions of explanatory variables.

Name: _____

9. (10 points) Calculate an estimate of the odds ratio of MACE (versus no MACE) comparing active treatment versus placebo, while controlling for baseline severity and center. Construct the corresponding two-sided 95% confidence interval and briefly interpret your odds ratio in one sentence.

10. Provide the predicted probabilities of MACE for the subjects described below:

    a. (5 points) A patient from Center 1 with mild baseline cardiovascular disease severity who is assigned to the placebo group.

    b. (5 points) A patient from Center 1 with severe baseline cardiovascular disease severity who is assigned to the active treatment group.