Name: ______________________________________

Analysis of Categorical Data (BIOS 665)
Midterm Examination 2018

Exam Date: October 16, 2018
Take-home due date: October 25, 2018 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 **<u>put your name on every page</u>**. Do <u>not</u> 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 ($Z_{1-\alpha/2}$)

$Z_{0.8}=0.842$ $Z_{0.85}=1.036$ $Z_{0.9}=1.282$ $Z_{0.95}=1.645$ $Z_{0.975}=1.960$ $Z_{0.99}=2.326$ $Z_{0.995}=2.576$

Name: ______________________________________

Part I

Table 1 contains data from a randomized, multi-center clinical trial testing a child birth intervention among African women to protect the health of mothers and their children. Researchers recruited 265 expectant mothers in their last trimester from two centers and followed them until 3 months after delivery. The new mothers are randomized to receive either 1) immediate exposure to the intervention within 48 hours of giving birth, or 2) delayed exposure six weeks after giving birth. The researchers want to know whether the immediate exposure results in a greater hypersensitivity reaction in regard to newborn lactation needs.

Table 1

		Hypersensitivity Reaction		
Center	Treatment	Yes	No	Total
1	Immediate	63	37	100
1	Delayed	45	55	100
	Total	108	92	200
2	Immediate	29	6	35
2	Delayed	17	13	30
	Total	46	19	65

1. (10 points) For Center 1, provide an estimate and its corresponding two-sided 95% confidence interval for the difference in proportions of hypersensitivity reaction in newborns for the immediate vs delayed intervention comparison.

Name: ______________________________________

2. (10 points) For Center 2, provide an estimate and corresponding two-sided 95% confidence interval for the odds ratio corresponding to the relationship between timing of intervention (immediate vs. delayed) and response (hypersensitivity reaction vs. not).

Name: ____

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

Name: ______________________________________

4. (10 points) In a follow-up study for immediate versus delayed intervention, the researchers would like to repeat the study summarized in Table 1 in another center, Center 3. They believe that the proportion of newborns with hypersensitivity reaction in the immediate intervention group is 0.85 and is 0.70 for the delayed intervention. Using a two-sided 0.05 significance level and 0.85 power for this planned study, with twice as many mothers in the immediate intervention group as in the delayed intervention group, determine the needed sample size in each group.

Name: ____

Part II

Table 2 is adapted from illustrative data for the drug Keyexelate and its relation to incident colonic necrosis.

Table 2

Group	Response		Total
	Colonic Necrosis	No Necrosis	
Keyexelate	3	114	117
Control	1	861	862
Total	4	975	979

5. (10 points) Assume that the table margins are fixed.

   i. Calculate the incidence (i.e., the proportion) of colonic necrosis for each group (Keyexelate and Control).

Name: ____

ii. Calculate the probability of each possible 2×2 table which could have been observed with these fixed margins. Provide a relevant table listing these probabilities.

Name: ______________________________________

6. (10 points) From 5ii), calculate an appropriate one-sided p-value (in favor of the Keyexelate group) for the association between group and colonic necrosis response. State your method and conclusion in one sentence.

Name: ______________________________________

Part III

FEMA researchers were interested in assessing the level of damage caused by the recent hurricane that affected the coastal communities of North Carolina. Out of the 156 homes completely destroyed by the storm in one of the communities, some of the affected did not have any flood insurance for their homes, others have partial insurance and the some have full insurance coverage for flood damage. Table 3 shows the distribution of the data by income level.

Table 3

Income Level	Levels of flood insurance coverage			Total
	None	Partial	Full	
High income	13	25	63	101
Low income	27	13	15	55
Total	40	38	78	156

7. (10 points) Under minimal assumptions and using 5% significance level, assess the association between the income level in terms of a location shift in the levels of flood insurance coverage with an appropriate statistical test. 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, briefly interpret your results.

Name: ____

Part IV

A logistic model for the probability of being diagnosed with cardiovascular disease (CVD) while participating in a trial of low-salt diet versus and high-salt diet in elderly Americans was fit based on reference cell coding for sex and diet, using female and high-salt as reference. The output below shows partial results of the fitted main effects and interaction models, respectively, in SAS 9.4.

The LOGISTIC Procedure

Response Profile

Ordered Value	*response*	*Total Frequency*
1	CVD	976
2	no CVD	702

Probability modeled is Response='CVD'

MODEL I

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	*Value*	*DF*	*Value/DF*	*P-val*
Deviance	2.1286	1	2.1286	0.1446
Pearson	2.0996	1	2.0996	0.1473

Analysis of Maximum Likelihood Estimates

Parameter		*DF*	*Est.*	*Std Error*	*Wald Chi-Sq*	*P-val*
Intercept		1	-0.569	0.084	45.416	<.0001
sex	*Male*	1	0.465	0.103	20.370	<.0001
treatment	*Low*	1	-0.248	0.128	3.723	0.0537

MODEL II

Joint Tests

Effect	*DF*	*Wald Chi-Square*	*P-val*
Sex	1	12.1710	0.0005
Treatment	1	5.1416	0.0234
*Sex*Treatment*	1	2.0930	0.1480

Analysis of Maximum Likelihood Estimates

Parameter			*DF*	*Est.*	*Std Error*	*Wald Chi-Sq*	*P-val*
Intercept			1	-.526	0.089	34.902	<.0001
Sex	*Male*		1	0.395	0.113	12.171	0.0005
Treatment	*Low*		1	-.508	0.224	5.142	0.0234
*sex*treatment*	*Male*	*Low*	1	0.398	0.275	2.093	0.1480

Note: 'Est.' is estimate, 'P-val' is p-value

Name: ______________________________________

8. (10 points) Specify the mathematical structure of the main effects model; be sure to include mathematical definitions of all explanatory variables.

Name: ______________________________________

9. (10 points)
   i. Which model (main effects or interaction) do you prefer, and why? Justify your choice with a valid statistical test.

   ii. Based on 9i) above, calculate an estimate of the odds ratio(s) of low-salt vs high-salt diet, while controlling for sex. Construct the corresponding two-sided 95% confidence interval(s), and briefly interpret your odds ratio(s) in one sentence.

Name:

10. (10 points) Using the main effects model, provide the model-predicted probabilities of CVD for each of the cases described below:

    a. An elderly woman on the low-salt diet.

    b. An elderly man in the high-salt diet.