BIOS 665: Problem Set 2
Assigned: Sep 12, 2019
Due: Sep 26, 2019

Note: For all hypothesis tests, please state the statistical method you are using, the null hypothesis, the test statistic, the p-value, and the interpretation of the test using significance level of 0.05, unless otherwise stated. Please round all estimates to **three decimal** places.

Helpful hints: For estimates and tests, simply copying and pasting SAS output without any commentary will not earn full credit, especially on exams. Highlighting is not considered commentary. However, commentary can be as simple as: The 95% CI for the odds ratio is (___, ____).

1. The following data are from a randomized clinical trial to assess whether an experimental treatment drug reduces severity of a rash after 2 weeks of treatment (compared to placebo). Participants were randomized to either High dose, Low dose, or Placebo groups. The following table contains data on treatment (High/Low/Placebo), Gender (Male/Female), and severity of rash after 2 weeks:

Treatment	Gender	Severity of Rash after 2 weeks				Total
		None	Mild	Moderate	Severe	
High Dose	Male	6	10	12	20	48
	Female	7	14	19	18	58
Low Dose	Male	9	7	30	19	65
	Female	10	17	11	16	54
Placebo	Male	19	15	17	5	56
	Female	21	18	10	5	54

a) Fit a logistic model that include main effects for treatment and gender, as well as their interaction.
   i. Specify the mathematical structure for a logistic regression model to describe the relationship of treatment and gender to severity of rash after 2 weeks (categorized as 'Moderate or Severe' versus 'None or Mild'), including main effects for treatment and gender as well as their interaction. Define all the explanatory variables. Use "placebo" as the reference group for treatment and "male" as the reference group for gender.
   ii. State at least 3 assumptions for using logistic regression to model these data.
   iii. Fit the model specified in (a) to the data. Conduct a statistical test with $\alpha=0.05$ to assess whether the relationship between treatment and severity of rash after 2 weeks is the same for both males and females.

b) Regardless of your answer in a) iii, fit a main effects logistic regression model to the data to describe the relationship of treatment and gender to severity of rash after 2 weeks

(categorized as 'Moderate or Severe' versus 'None or Mild'). Use "placebo" as the reference group for treatment and "male" as the reference group for gender.

i. State the mathematical specification of the main effects logistic regression model.
ii. Provide a predicted probability of Moderate or Severe rash after 2 weeks for females on Low Dose of treatment.
iii. Provide an estimate and 95% confidence interval for the odds ratio comparing the odds of 'Moderate or Severe' rash after 2 weeks on Low dose to the odds of 'Moderate or Severe' rash after 2 weeks on High dose.

2. The following data arise from a health policy study that included interviews with commuters from both East Coast and West Coast geographic regions. Use logistic regression to describe the relationship of favorable opinion response (vs. unfavorable response) to stressful commuting and geographical region. Use reference groups of "East" for geographic region and "Low" for stress level during commuting time.

Region	Stress Level	Favorable	Unfavorable	Total
East	Low	53	20	73
East	Medium	109	97	206
East	High	85	76	161
Total		247	193	440
West	Low	51	37	88
West	Medium	67	59	126
West	High	118	92	210
Total		236	188	424

a) Mathematically specify the main effects model, and interpret all parameters.

b) Fit the above model and provide a quantity that expresses the effect of high stress (as compared to low stress) on favorable response (vs. unfavorable response), and provide a 95% two-sided confidence interval for this quantity. How would you estimate this quantity and its 95% confidence interval by hand, given the computer output? Please show details of your calculations.

c) Provide a quantity that expresses the effect of high stress (as compared to medium stress) on favorable response (vs. unfavorable response), and provide a 95% two-sided confidence interval for this quantity.

d) Formally test the hypothesis that region has no effect on opinion response.

e) Provide predicted probabilities for favorable response for each of the following:

a. An individual from an East Coast area with low stress.

b. An individual from a West Coast area with medium stress.

c. An individual from a West Coast area with high stress.

f) Test the hypothesis that the model fit is adequate. Briefly justify your choice for this test.

3. The table below shows two logistic regression models which were fit to describe the relationship between the probability of no Crohn's disease (versus otherwise) at post-treatment and explanatory variables for treatment (Placebo, High dose, Low dose), baseline severity (Moderate, Severe), and (for Model 2) their interaction are presented below. The results for the two models include estimated parameters, their standard errors, and the maximized log-likelihood for the fit of the model.

	Model 1		**Model 2**	
Parameter	Estimate	Standard Error	Estimate	Standard Error
Intercept	-3.221	0.583	-2.527	0.890
Severe (Baseline Severity)	-1.793	0.262	-1.933	1.687
High dose (Treatment)	2.267	0.522	2.005	0.732
Low dose (Treatment)	1.938	0.863	3.812	0.763
Severe*High	NA	NA	0.969	1.191
Severe*Low	NA	NA	0.064	2.027
Log-likelihood	-159.743		-157.872	

* Explanatory variables equal 1 if the category applies, and 0 if otherwise.
** NA = Not Applicable

a) For Model 1, state the relevant assumptions for the application of logistic regression to these data. Specify the mathematical structure of the model, including mathematical definitions for all explanatory variables.

b) Numerically calculate the odds ratios of no Crohn's disease (vs. otherwise) for each of the following conditions:

	Model 1	**Model 2**
Odds ratio for Severe to Moderate, for Placebo		
Odds ratio for Severe to Moderate, for Low dose		
Odds ratio for High dose to Placebo, for Moderate baseline		
Odds ratio for High dose to Placebo, for Severe baseline		

c) Comment on the calculations in b), particularly in regard to the comparisons between rows 1 and 2, and between rows 3 and 4.

d) For Model 1:

i). Interpret the parameter corresponding to Low dose in Model 1.

ii). Provide a 95% confidence interval and estimate for the odds ratio of no Crohn's disease (vs. otherwise) for High dose versus Placebo, controlling for baseline severity.

For Model 2:

i). Provide a predicted probability of no Crohn's disease for an individual on low dose who had severe pain at baseline. Please provide a numeric solution.

e) Apply a statistical test at the $\alpha = 0.05$ significance level to assess the hypothesis that Model 1 has satisfactory goodness of fit in the sense that any association between baseline severity and probability of no Crohn's disease post-treatment is homogeneous across the placebo, low dose, and high dose groups.

4. A study examined the effect of disinfectant on acute cutaneous complication (ACCs) during insulin-pumping treatment. The following table shows the distribution of patients according to the occurrence of ACC before and after disinfectant use for males and females.

Gender	Before Disinfectant	After Disinfectant	
		ACCs	No ACCs
Male	ACCs	10	18
	No ACCs	2	10
Female	ACCs	22	37
	No ACCs	15	19

a) Do the data provide evidence that the proportion of diabetic patients with ACCs after using the disinfectant differs between diabetic patients with and without ACC before using the disinfectant, ignoring gender?

b) For each gender separately, test the association between the ACC status at baseline (i.e., before using the disinfectant) and the occurrence of ACC vs not after disinfectant use.

c). Is there evidence that the association in (b) differs between male and female diabetic patients?

d). Repeat the test in (b), but now controlling for gender rather than conducting separate analysis.

e). For each test above, briefly justify your choice of method.