BIOS 665:   Problem Set 4
Assigned: October 24, 2019
Due: November 07, 2019

Notes: 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. A psychiatric hospital conducted an experiment across three geographical locations to compare three levels of motivation with respect to a 3-level response (low, medium, and high).. At the end of the follow-up period, the patients were evaluated by a team composed of a psychiatrist, a psychologist, a nurse and a social worker. Please use "none" initial level of motivation and "East" geographical location as reference groups.

Initial level of motivation	Location	Response		
		Low	**Medium**	**High**
None	East	62	70	65
	MidWest	67	78	68
	West	35	21	18
Low	East	19	14	15
	MidWest	19	27	24
	West	45	33	24
Average	East	21	22	19
	MidWest	50	38	29
	West	19	14	15

   a) Mathematically specify a proportional odds regression model for response (ordered from low to high) with main effects for location and initial level of motivation. State the model assumptions, and interpret all model parameters. Assess goodness of fit of the proportional odds model, and justify your method.

   b) Conduct a statistical test to assess whether proportional odds across both explanatory variables is a reasonable assumption for these data. Write a sentence explaining the results of your test.

   c) Test whether initial level of motivation at 0.05 significance level has an effect on response. Write a sentence explaining your results.

   d) Provide an estimate and 95% confidence interval for the odds ratio of **high** vs. (**low** or **medium**) response comparing a '**low**' initial level of motivation with those having '**none**';

repeat for 'average' vs. 'none'. What do you conclude about the statistical significance of these effects from their confidence intervals? Briefly discuss how these estimates compare to comparable estimates for (**high** or **medium**) vs. **low**.

2. Use the data from Question 1 above.

   a) Mathematically specify and fit a generalized logits regression model for response, treating (**low**, **medium**, and **high**) as nominal instead of ordinal. Include main effects for initial level motivation and location. Let "**medium**" be your reference for the response. State assumptions, and interpret all model parameters.

   b) From Problem 2a), conduct a statistical test for whether initial level of motivation has an overall effect on response at the 0.05 significance level. Briefly explain the result of your test.

   c) Using this model, provide an estimate and 95% confidence interval for the odds ratio of **high** vs. **medium** response comparing '**low**' initial level of motivation with those having '**none**'; repeat for '**average**' vs. '**none**'..

      i. Repeat for **low** vs. **medium** response, as well as for **high** vs. **low** response.

      ii. What do you conclude about the statistical significance of each effect from these confidence intervals?

3. The table below contains data from a study that evaluated the relationship between doses of two treatments with respect to dichotomous response concerning pain relief.

Treatment	Dose	Favorable	Unfavorable	Sample size
A	1 mg	20	65	85
A	10 mg	31	54	85
A	100 mg	47	38	85
B	2mg	19	66	85
B	20mg	36	49	85
B	200mg	60	25	85

For Treatment A only

a) Using logistic regression, describe the relationship between favorable response (vs. unfavorable response) and ln(dose) as a continuous explanatory variable:

   i. State the assumptions and mathematically specify the model. Evaluate goodness of fit for this model.

ii. Provide estimates and 95% Fiducial Limits (CI) for the dose values corresponding to ED25, ED50, and ED75. In other words, provide estimates and 95% confidence limits for the dose values which produce a response with 0.25, 0.50, and 0.75 probabilities, respectively.

iii. Use a probit analysis to calculate all the estimates and confidence intervals requested in Part 3.a.ii. How do your assumptions change when using a probit model vs a logistic model?

iv. Briefly compare and contrast your results from Problems 3a.ii and iii.

For both treatments
b) Using logistic regression to describe the relationship between favorable response and ln(dose) as a continuous explanatory variable for the data, and allowing for separate effects for each treatment group, as illustrated in class:

i. State the assumptions and mathematically specify the model.

ii. Evaluate goodness of fit of the model.

iii. Provide a point estimate and its 95% confidence interval for the relative potency of Treatment B relative to Treatment A.

4. The table shown below displays the cross-classification of maternal age groups (in years) and the number of births with a particular disorder in a specific geographic area during a specific time period, as well as the corresponding numbers of all births. When necessary, use '20-24 years' as the reference group for maternal age group and '1' as the reference group for birth order.

Birth Order	1	2	3	1	2	3
Maternal Age	Number of Births with Disorder			Total Number of Births		
20-24	128	152	71	329,462	326,735	175,682
25-29	54	112	101	114,987	208,692	207,060
30-34	41	79	109	39,473	83,224	117,312
35-39	38	89	99	14,202	28,478	45,015
40+	22	44	83	3,046	5,381	8,654

a) Specify the mathematical structure of a statistical model to describe the variation in the rates of the disorder per 100,000 live births with respect to maternal age group and birth order.
b) Interpret the estimated parameters of this model, and provide appropriate two-sided 95% confidence intervals for those pertaining to birth order.

c) Use the model from Problem 4.a. to obtain predicted values for the rates of the birth disorder for the respective birth order subpopulations corresponding to '30-34 years' for maternal age group.