BIOS 665: Final Problem Set (Take Home Exam)
Assigned: November 21, 2019
Due: **December 12, 2019 at 12:00 noon (1301 McGavran-Greenberg Hall)**

If you are **not** a BIOS student and are trying to earn a P, complete any 18 problems.

If you **are** a BIOS student or are trying to earn an H, complete all Problems.

****

FOR GRADING PURPOSES,
PLEASE STAPLE AND SUBMIT YOUR SOLUTIONS TO EACH PROBLEM *SEPARATELY*,
AND MAKE SURE YOUR *NAME APPEARS ON EACH PAGE*

****

ALL students: Please complete this checklist, and submit this facesheet with your exam.

BIOS 665 Final Exam Checklist
Fall 2019

NAME _____

PID _____

DEPARTMENT _____

Campus Box (CB) # _____

Check One: I am attempting to earn an 'H' grade: [ ]

I am attempting to earn a 'P' grade: [ ]

Please check the problems you are submitting:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Honor Code: Please remember that the Honor Code is in effect and ***all*** work must be done independently. You may only consult the instructors, and no other individuals.

All submitted work has been completed independently by me, and I am bound by the Honor Code.

Signed _____ Date _____

For all hypothesis tests, please state the method, the null hypothesis, the test statistic, the distribution to which you will compare the test statistic, and the p-value; use a two-sided significance level of 0.05, unless otherwise stated.

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 (___ , ___).

Part I

For Problems 1-18, refer to the following data from a randomized clinical trial to evaluate whether varying doses of an experimental treatment are associated with a particular adverse event (compared to placebo). Participants were randomized to one of three groups: High dose, Low dose, or Placebo. The following table summarizes data by treatment group, sex, and severity of the adverse event.

Treatment	Sex	Severity of Adverse Event				Total
		None	Mild	Moderate	Severe	
High Dose	**Men**	9	12	12	21	54
	Women	6	16	16	21	59
Low Dose	**Men**	7	15	26	17	65
	Women	11	16	12	21	60
Placebo	**Men**	16	16	12	8	52
	Women	20	16	20	10	66

1. Without using a formal statistical model, provide an estimate of the common odds ratio and its 95% confidence interval for the effect of pooled treatment (high dose + low dose) vs. placebo on the severity of the adverse event, dichotomized as (none or mild) vs. (moderate or severe), when controlling for sex.

2. Without using a formal statistical model, statistically test the null hypothesis that the effect of pooled treatment (high dose + low dose) vs. placebo on the dichotomized severity of the adverse event − (none or mild) vs. (moderate or severe) − is the same for each sex. Provide a sentence explaining your results.

3. Under minimal assumptions, conduct a statistical test to determine whether there is a difference in the proportion of moderate or severe adverse event (vs. none or mild) among the three treatment groups, controlling for sex. For this problem, you should consider the treatment groups as <u>nominal</u>. Write a sentence to interpret your findings.

4. Under minimal assumptions, conduct a statistical test to determine whether there is a trend in the proportion of moderate or severe adverse event (vs. none or mild) across the ordered treatment groups, controlling for sex. Write a sentence to interpret your findings.

5. Under minimal assumptions, conduct a statistical test to assess the association of pooled treatment (high dose + low dose) vs. placebo with the severity of adverse event (all four ordered levels managed as distinct), controlling for sex. Justify your method. If you determine that $p < 0.05$, discuss whether pooled treatment is associated with greater severity or lesser severity of the adverse event.

6. Under minimal assumptions, conduct a statistical test to determine whether there is a progressive location shift in the severity of the adverse event (as distinct levels) across high dose, low dose, and placebo, controlling for sex. In a sentence, interpret your findings.

7. Report the Spearman rank correlation coefficients and corresponding 95% confidence intervals separately by sex as measures of association for pooled treatment (high dose + low dose) versus placebo with the severity of adverse event (as distinct levels). Write a sentence indicating whether men or women exhibit a stronger association, and briefly justify your finding.

8. Report the Spearman rank correlation coefficients and corresponding 95% confidence intervals separately by sex as measures of association for ordered treatment groups with severity of the adverse event (as distinct levels). Write a sentence indicating whether men or women exhibit a stronger association, and briefly justify your finding.

9. Separately within each treatment group, test the association between sex and severity of the adverse event (as ordered distinct levels). Also, assess such association under minimal assumptions, and controlling for treatment groups. Write a sentence to interpret your results.

10. Mathematically specify a logistic regression model for the dichotomous response of (moderate or severe) adverse event (vs. none or mild), with main effects for treatment (as distinct levels and considered as nominal, treating placebo as the reference) and sex (treating men as the reference). State assumptions, and mathematically define all variables in the model. Interpret all model parameters. Determine the goodness of fit for this model.

11. Using the model from Problem 10, provide estimates and corresponding 95% confidence intervals for the odds ratios of high dose vs. placebo and of low dose vs. placebo for (moderate or severe) adverse event compared to (none or mild).

12. Using the model from Problem 10, perform a statistical test of whether the treatment groups differ with respect to the (moderate or severe) adverse event (i.e., the overall treatment effect). Provide the test statistic, indicate the number of degrees of freedom, and determine statistical significance through its p-value. If this overall effect is statistically significant, test each pairwise treatment comparison at the α=0.05 level, and indicate which treatment groups are significantly better than others. (Note: you do not need to address any adjustment to the type I error for multiple comparisons for this problem.)

13. Using your model from Problem 10, what are the respective model-predicted probabilities for (moderate or severe) adverse event and for (none or mild) adverse event for men on high dose and also for women on placebo?

14. Mathematically specify a proportional odds regression model for 'more' vs. 'less' severity of the adverse event, with main effects for treatment (as distinct levels and considered as nominal, treating placebo as the reference) and sex (treating men as the reference). State assumptions, and mathematically define all variables in the model. Interpret all model parameters. Determine the goodness of fit for this model, including the test for proportional odds. If the proportional odds assumption is not supported, investigate and briefly describe your findings regarding the potential reason for this.

15. Using the model from Problem 14 that assumes proportional odds, provide estimates and corresponding 95% confidence intervals for the odds ratios of high dose vs. placebo and of low dose vs. placebo for (severe or moderate) adverse event compared to (none or mild).

16. Using the model from Problem 14 that assumes proportional odds, what are the respective model-predicted probabilities for none, mild, moderate, and severe adverse event for men on low dose?

17. Mathematically specify a generalized logit model for all levels of severity of adverse event (treating 'none' as the reference), with main effects for treatment (as distinct levels and considered as nominal, treating placebo as the reference) and sex (treating men as the reference). State assumptions, and mathematically define all variables in the model. Interpret all model parameters.

18. Using your model from Problem 17, provide estimates and corresponding 95% confidence intervals for the odds ratios of high dose vs. placebo and of low dose vs. placebo for each severity level compared to 'none'. Compare and contrast these results to those found in Problem 15~~14~~.

Part II

The hypothetical data for Problems 19-22 represent a randomized, controlled trial of 59 geriatric patients who were randomized to either an experimental treatment or placebo. Prior to randomization, baseline data on the number of falls during the preceding eight-week interval were recorded. Counts of falls during two-week intervals before each of four successive post-randomization visits were recorded. The following variables are contained in the FALLS.SAS7BDAT dataset:

- SUBJ: unique participant identification code
- AGE: the participant's age
- TREATMENT: the experimental group to which the participant was randomized
  - 0 = placebo
  - 1 = experimental treatment
- BASELINE0: baseline falls count in the prior eight weeks
- FALLS2: the falls count for weeks 1-2
- FALLS4: the falls count for weeks 3-4
- FALLS6: the falls count for weeks 5-6
- FALLS8: the falls count for weeks 7-8

19. Under minimal assumptions and controlling for baseline falls count in the prior eight weeks trichotomized as <5 falls, 5-10 falls, or >10 falls, conduct a statistical test to assess the association between treatment group and improvement, where improvement is defined as having fewer falls across the entire eight-week study period than during the preceding eight weeks, and no improvement is defined as having the same number or more falls. Briefly interpret your results in 1-2 sentences.

20. Fit a Poisson regression model to the total falls counts calculated across the four post-randomization clinic visits. (Hint: You should be modeling a single total count for each patient.) Use treatment group (with placebo as the reference), baseline falls count trichotomized as <5 falls, 5-10 falls, or >10 falls (with <5 falls as the reference), and age as main effects. Determine whether there is evidence of over-dispersion. If so, adjust for over-dispersion using a scale parameter. For all parameters in the model, report the estimates, standard errors, test statistics, and p-values, with adjustment for overdispersion if needed. Determine the goodness of fit for this model with adjustment for overdispersion, if needed.

21. Using the model from Problem 20, what is the model-predicted mean falls count during the eight-week post-randomization interval for an individual having 5-10 baseline falls in the prior eight weeks, who is 80 years old, and was:
    a. randomized to receive the experimental treatment
    b. randomized to receive placebo

22. Regardless of your findings regarding over-dispersion in Problem 20, fit a negative binomial model to the total falls counts, with the same main effects as specified in Problem 20. For each parameter in the model, report the estimate, standard error, test statistic, and p-value. Compare these results to those you found in your final model (i.e., adjusted for over-dispersion, if necessary) in Problem 20.

Part III

23. For the FALLS.SAS7BDAT dataset, fit a repeated measures Poisson model to two post-randomization periods (combining Weeks 1-2 with Weeks 3-4, and combining Weeks 5-6 with Weeks 7-8) to describe the counts of falls, with main effects of experimental group, baseline falls counts (trichotomized as in Problems 19 and 20), age, and period. You may assume that the sample size is large enough to fit this model. Use the same reference levels as specified in Problem 20, additionally specifying Week 1-4 as the reference for period. Specify estimating equations based on the Poisson distribution to fit the falls counts during each ~~two~~four-week interval.

    For the purposes of this exam, determine goodness of fit for the main effects model by considering only pairwise interactions involving experimental group. Present your final model by listing the SAS table of GEE parameter estimates. Justify your choice for the final model, incorporating a brief discussion of any assumptions (including your choice of the working correlation structure) or limitations of the model.

24. For this problem, you should restate your final recommended model from Problem 23 at the top of the page for the grader's reference. Using this final model, provide the falls rate ratio comparing experimental treatment and placebo at each post-randomization period (i.e., for Weeks 1-4 and separately for Weeks 5-8), and provide the corresponding 95% confidence intervals for these estimates.

25. For this problem, you should restate your final recommended model from Problem 23 at the top of the page for the grader's reference. Provide the predicted mean falls count at the Week 5-8 interval for an individual having 5-10 baseline eight-week falls, who is 80 years old, and was:
    a. randomized to the experimental treatment arm.
    b. randomized to the placebo arm.