BIOS 665: Problem Set 3
Assigned: September 26, 2019
Due: October 10, 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. Consider the data in the table below. These data are from a clinical trial conducted in two centers for the comparison of two treatments for a gastrointestinal disorder with respect to a dichotomous response (Good vs. Poor).

Center	Treatment	Response		
		Good	Poor	Total
A	Test	22	13	35
	Placebo	33	19	52
B	Test	27	7	34
	Placebo	18	15	33

   a) For each center separately, provide an estimate for the odds ratio (and its 95% confidence interval) describing the relationship between test treatment and placebo for good versus poor response. Briefly justify your methods and interpret your results.

   b) For Center A, provide and interpret the results of a statistical test for the association between treatment and response using the two-sided 0.05 significance level. Repeat separately for Center B.

   c) Under minimal assumptions, assess the association between treatment and response, <u>controlling for center</u>, with a statistical test at the two-sided 0.05 level. Briefly justify your methods and interpret the results.

   d) Assess the homogeneity of the odds ratio across the two centers.

   e) Assuming homogeneity of the effect in (d), provide an estimate of the common odds ratio and a corresponding 95% confidence interval.

2. The following data are from a randomized clinical trial to assess whether an experimental drug treatment 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 (Placebo/High/Low), Gender (Male/Female), and presence of rash after 2 weeks as (Mild or Moderate or Severe combined versus None) response:

Treatment	Gender	Severity of Rash after 2 weeks				Total
		None	Mild	Moderate	Severe	
Placebo	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
High Dose	Male	19	15	17	5	56
	Female	21	18	10	5	54

a) Under minimal assumptions (not involving a formal statistical model), conduct a statistical test to assess the association of pooled test treatments (high or low) vs. placebo with presence of rash after 2 weeks of treatment, controlling for gender.

b) Provide an odds ratio and 95% confidence interval that describe the effect of pooled (high or low dose) treatment vs. placebo on presence of rash after 2 weeks (versus not) for each of the following:
   i. Females only. .3
   ii. Males only. .4
   iii. Controlling for gender. You should address the assumption that the effect of pooled (high or low dose) treatment vs. placebo on presence of rash after 2 weeks treatment is similar in both males and females. common OR .39 - common OR seems to be in between the two. the males has pulled it much more. closer to results for (females?)

mantel-haenszel
OR are homogenous, in stratified they should be similar. if saying they are homogenous, then you are assuming that the difference is not statistically significant. homogeneity exists across gender -- reject the null

c) Fit logistic regression models for the presence of rash after 2 weeks with explanatory variables for pooled (high or low dose) treatment vs. placebo and for female vs. male, using placebo and male as reference groups for the following:

   a). main effects model: provide gender-specific odds ratios and their corresponding 95% confidence intervals.

b). main effects plus two-way interaction model: provide gender-specific odds ratios and their corresponding 95% confidence intervals
---OR for females, and OR for males, and 95% CI for each
--by hand: use contrasts C*B (wald statistic

ignore the difference between part b and c????

d) Do parts (a), (b), and (c) agree? Briefly comment on your findings in 1-2 short sentences.
--Very short answer. what is common between the different part, and the parts that don't say the same things. these parts are similar across, these parts are not similar across

3. A study investigated the relationship between the type of health insurance and tooth extraction (upper and lower jaws) in adults. Patients were randomly assigned to receive medical and dental insurance with differing degrees of coverage, which were classified into three categories: low, intermediate, and high.

Tooth Extraction	Jaw	Type of Insurance		
		Low	Intermediate	High
None	Upper	279	193	373
	Lower	149	83	137
One	Upper	69	21	81
	Lower	29	44	75
Two or more	Upper	45	33	24
	Lower	21	12	19

Note: For parts a) and c), please note the dichotomization.

a) Under minimal assumptions, conduct a statistical test to determine whether there is a location shift across degree of insurance coverage when comparing between no tooth extraction and one or more teeth extraction, while controlling for type of jaw.

not significant
2x3 table, use the percentages

meyer - whitney, similar to log rank test, it is non-parametric-just uses ranking. (plot histogram --) give a measurement of association of comparing the columns of the table. the same association is being tested using a different formulation. no assumption.

b) Please specify which tooth extraction tends to demonstrate a higher degree of insurance coverage if the result in part a) is shown to be statistically significant.

if a is non-significant, just say that a) is not significant

suppose it was significant - then see which direction the significance is going. proportions in the table, give you an idea which direction they are going. can use pearson correlation

since not sig don't have to worry about which tooth is has higher proportion of insurance. you can look at it, but since not significant, don't interpret

c) Under minimal assumptions, conduct a statistical test to determine whether there is a trend in the proportion of degree of high insurance (versus not) across the ordered levels of number of extracted teeth ignoring any effect of jaw.

d) Under minimal assumptions, conduct a statistical test to determine whether there is a trend in the proportion of high insurance coverage (versus not) across the ordered levels of number of extracted teeth, controlling for jaw.

Cochran mantel haenszel test statistic. from the original data can see that the trend is going that way

didn't use Cochran -armitage test for it. tempting to use it. Cochran - highly significant -- does this contradict any of the other test? Cochran mantel haenszel - a 1df of test and 2df test. 2 df test is significant ---- corresponds to Cochran-armitage. this is 3x2 table (sx2 case), looking at trend. 1st output gives you proportions of frequencies. proportion is decreasing from 69% to 30% to 10%. the difference that you are seeing is what you're trying to test, is this sig? for each jaw, you will get the Cochran-armitage

want table with 1df for Cochran mantel haenszel. df: (s-1)(r-1)(2-1)=2df

get that it's not significant.

sx2 --- trend
2xr ---- shift of location

extended mantel haenzsel. a version sx2 ---- use for a, c, d
null hypotheses are similar, just difference

4. Consider data arising from 46 patients, each comparably treated with each of three drugs (labeled A, B, and C). The response is recorded as either favorable (F) or unfavorable (U) to each of the drugs, with the goal to determine whether the three drugs have similar response. These data are explained in more detail in the textbook, *Categorical Data Analysis Using the SAS System, 3rd ed.* on pages 162-163. The dataset can be found on the course web site as drug.sas7bdat with the following variables: PATIENT (uniquely identified patient identification number), DRUG (A,B,C), and RESPONSE (F for favorable, U for unfavorable).

a) Fit a conditional logistic regression model describing the relationship of favorable response to the treatments, conditioning on patient. Use drug A as the reference category. Report a table of all parameter estimates and their standard errors.

b) Provide appropriate measures of association and 99% confidence intervals for the effects of treatments relative to one another for favorable response, while conditioning on the patient effect.

c) Using this model with drug A as reference, perform a hypothesis test to compare drug B to drug C at the 0.05 significance level. Justify your choice of methods and interpret the results.

contrast test. chi square statistic. assumption - H_0 : B-C = 0

(1,1) (B, C)^t = 0, chi sq with one df