Project #2

STAT 875

Spring 2014

Complete the following problems below. Within each part, include your R program output with code inside of it and any additional information needed to explain your answer. Your R code and output should be formatted in the same manner as in the lecture notes.

1. (23 total points) The results of Rerks-Ngram et al. (*New England Journal of Medicine*, 2009, volume 361, p. 2209-2220) gave new hope for finding a HIV vaccine. This paper lead to expanding the efforts of another HIV vaccine clinical trial, known as HVTN 505. More on this study and its discontinuation can be found at the following locations:
* Expansion of the trial (see bottom): [http://www.niaid.nih.gov/topics/HIVAIDS/Research/
vaccines/clinical/Pages/highlights.aspx](http://www.niaid.nih.gov/topics/HIVAIDS/Research/vaccines/clinical/Pages/highlights.aspx)
* Discontinuation of the trial: [http://www.niaid.nih.gov/news/newsreleases/2013/Pages/
HVTN505April2013.aspx](http://www.niaid.nih.gov/news/newsreleases/2013/Pages/HVTN505April2013.aspx)
* Questions and Answers: <http://www.niaid.nih.gov/news/QA/Pages/HVTN505qa2013.aspx>
* NY Times article about discontinuation of the trial: [http://www.nytimes.com/2013/04/26/
health/agency-halts-trial-for-aids-vaccine.html](http://www.nytimes.com/2013/04/26/health/agency-halts-trial-for-aids-vaccine.html)
* History of HIV vaccine research: [http://www.niaid.nih.gov/topics/HIVAIDS/Research/
vaccines/Pages/history.aspx#](http://www.niaid.nih.gov/topics/HIVAIDS/Research/vaccines/Pages/history.aspx)

Below is a contingency table summarizing the results of the trial for those who were enrolled for at least 28 weeks (thus, some individuals may have been part of the trial for more than 28 weeks).

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Response |  |
|  |  | HIV | No HIV |  |
| Treatment | Vaccine | 27 | 1,223 | 1,250 |
| Placebo | 21 | 1,223 | 1,244 |
|  |  | 48 | 2,446 | 2,494 |

Based on this and other examinations of the clinical trial’s data, vaccinations for this trial were discontinued in 2013 due to lack of efficacy. The purpose of this problem then is for you to use the appropriate methods from our course with this data to reach the same conclusion.

* 1. (3 points) A description of who was in the sample is given in some of the web links above. Based on this sample, what was the population? Is there perhaps a larger intended population that would be of interest? If so, what assumptions need to be made?
	2. (12 points) Examine the benefits to using the vaccine rather than the placebo using the following methods (α = 0.05):
		1. Agresti-Caffo confidence interval for the difference in probabilities
		2. Wald confidence interval for the relative risk
		3. Wald confidence interval for the odds ratio
		4. Pearson chi-square test for independence

With each method, provide the details for the calculations and FULL interpretations of the numerical results (statistical and practical interpretations).

* 1. (5 points) How much better than the placebo did the vaccine need to be in order to show the vaccine worked? For the n1+, n2+, and w2 given in the data, find what values of w1 would have been needed in order to conclude that the vaccine worked. Provide the estimated odds ratio and relative risk associated with these values of w1. Discuss the potential consequences for the study design.
	2. (3 points) At the end of a 28 week period of time from first vaccination, the following HIV infection data was observed:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Response |  |
|  |  | HIV | No HIV |  |
| Treatment | Vaccine | 14 | 1,236 | 1,250 |
| Placebo | 9 | 1,235 | 1,244 |
|  |  | 23 | 2,471 | 2,494 |

Discuss if this version of the data or the previous version may be more appropriate to examine. Are there any assumptions then that need to be made with the previous version of the data to make it appropriate?

1. (16 points) Testing error also occurs in contingency table settings where the response categorical variable Y can be misclassified. The grouping variable, X, is recorded correctly. Below are two contingency tables given on p. 844 of Neuhaus (*Biometrika*, 1999) showing the true and misclassified responses, denoted as  and Y, respectively.

|  |  |  |
| --- | --- | --- |
| Table #1 |  | Table #2 |
|  |  |  |  |  |  |  | Y |  |
|  |  | 1 | 0 |  |  |  |  | 1 | 0 |  |
| X | 1 | 40 | 60 | 100 |  | X | 1 | 38 | 62 | 100 |
| 2 | 20 | 80 | 100 |  | 2 | 24 | 76 | 100 |
|  |  | 60 | 140 | 200 |  |  |  | 62 | 138 | 200 |

Normally, one would not know the  responses, but we will assume they are known here to demonstrate some important points for this problem. For this data, η = P(Y = 1 |  = 1) = 0.8 and δ = P(Y = 0 |  = 0) = 0.9. Using these η and δ values directly, one can see how Table #2 is formed from Table #1. From project #1, . Using this expression in terms of estimated probabilities, we have



where I use a ^ on P(⋅) to denote an estimated probability. Also, we can use the same type of expression for row #1:



For Table #2, w+ is found using 200×0.31 = 62 and w1 = 100×0.38 = 38. The rest of Table #2 can be constructed from this information. Thus, Table #2 is found using “estimated” expected values with the help of information in Table #1.

To make sure this is clear, Table #1 would not be known in a real setting. Also, even if Table #1 was known, there would be variability in the construction of Table #2, which would lead to potentially different observed counts than those given above. What we are using here for Table #2 is an estimate of what we would expect to happen on average.

* 1. (2 points) Find the estimated odds ratios for Table #1 and Table #2 using the usual equations (don’t adjust the equations to account for testing error). Compare the values of these odds ratios.
	2. (2 points) Find the estimated standard deviation for the estimated log odds ratio constructed for Table #1. Also, find the 95% Wald confidence interval for the odds ratio.
	3. Normally, only Table #2 would be known, but the goal would be to estimate the true odds ratio, say . Using Table #2 only, complete the following:
		1. (5 points) The correct estimated odds ratio for the true relationship between X and  given only data on X and Y is



Justify this formula and find the estimated odds ratio using the data. Hint: When justifying the formula, start with



where  and  are the true probabilities.

* + 1. (2 points) The standard deviation of the log odds ratio, , can be derived through using likelihood-based methods and the delta method. This standard deviation is



Find the estimated standard deviation.

* + 1. (2 points) State the formula for a Wald confidence interval for the odds ratio  and compute it for the data.
	1. (3 points) Compare the standard deviations found for b) and c)ii) and provide intuitive justification for why one is smaller than the other. Describe the effect this can have on inferences.