Homework 2

BSTA 513/613

Your name here - update this!!!!

2024-04-25

Directions

- [Download the `.qmd` file here.](#)

- You will need to download the datasets. Use [this link to download](#) the homework datasets needed in this assignment. If you do not want to make changes to the paths set in this document, then make sure the files are stored in a folder named "data" that is housed in the same location as this homework `.qmd` file.

- Please upload your homework to [Sakai](#). **Upload both your `.qmd` code file and the rendered `.html` file**

- For each question, make sure to include all code and resulting output in the html file to support your answers.

- Show the work of your calculations using R code within a code chunk. Make sure that both your code and output are visible in the rendered html file. This is the default setting.

- If you are computing something by hand, you may take a picture of your work and insert the image in this file. You may also use LaTeX to write it inline.

- Write all answers in complete sentences as if communicating the results to a collaborator. This means including a sentence summarizing results in the context of the research study.

> 💡 Tip
>
> It is a good idea to try rendering your document from time to time as you go along! Note that rendering automatically saves your qmd file and rendering frequently helps you catch your errors more quickly.

Questions

Question 1

A study looked at the effects of oral contraceptive (OC) use on heart disease in women 40 to 44 years of age. The researchers prospectively tracked whether or not the women developed a myocardial infarction (MI) over a 3-year period. The table below summarizes their results with columns indicating whether or not women developed MI and rows indicating their OC use.

```
Warning: 'tidy.numeric' is deprecated.
See help("Deprecated")

# A tibble: 2 x 1
  x[,"Yes"] [,"No"]
      <dbl>   <dbl>
1        13    4987
2         7    9993
```

Part a

Compute the estimated risk difference comparing OC users to non-OC users. Include a 95% CI for the estimate and interpretation of the estimated value.

Part b

Compute the estimated relative risk comparing OC users to non-OC users. Include a 95% CI for the estimate and interpretation of the estimated value.

Part c

Compute the estimated odds ratio comparing OC users to non-OC users. Include a 95% CI for the estimate and interpretation of the estimated value.

Part d

Is the OR a good approximation of the RR? Explain why or why not.

Question 2

I was intrigued by the results from Question 1, so I started researching the relationship between contraception and MI. In a journal review article, the authors present smoking status as a potential confounder for the relationship between OC use and MI. In Table 2 of their paper, they present various risk ratios for smoking and nonsmoking women. Use the below tables, that are recreated from one study's data, to answer the following questions.

Nonsmokers:

```
Warning: 'tidy.numeric' is deprecated.
See help("Deprecated")


# A tibble: 2 x 1
  x[,"Yes"] [,"No"]
      <dbl>   <dbl>
1        36     317
2        72     521
```

Smokers:

```
Warning: 'tidy.numeric' is deprecated.
See help("Deprecated")


# A tibble: 2 x 1
  x[,"Yes"] [,"No"]
      <dbl>   <dbl>
1       210     416
2       236     586
```

Part a

Compute the stratum-specific ORs for each smoking status. Present the estimates with 95% Confidence interval.

Aside (not a homework problem)

If you are interested in a short review of the relationship between oral contraceptive pills (OCP) and MI, see this article abstract. I cannot find the full article online, so if you find it, please pass along. The abstract states:

> Although OCP doses were subsequently reduced, epidemiologic evidence continued to support a smaller, but significant association between OCPs and hypertension.

This single statement packs in some interesting aspects of the research process and influences from society. To start, the OCP doses were reduced from their initial approval. However, the early studies (1980s-90s) that link increased risk of hypertension and MI to OCs were still some of the first articles presented in my Google search. You may also find it interesting to think about how the perception of contraception has changed since the 1980s, and how societal views can impact research questions.

Question 3

The data presented below are from a case-control study of bladder cancer. Subjects with and without bladder cancer were recruited and then questioned about their occupation and cigarette consumption.

High risk occupation	High cigarette consumption	Cases	Controls
Yes	No	48	100
Yes	Yes	180	175
No	No	30	24
No	Yes	120	80

Part a

Please present the data using two 2x2 contingency table for High-risk occupation and Disease status, stratified by Cigarette consumption.

Part b

Please present the data for High-risk occupation and Disease status using a 2x2 contingency table (you should combine the cigarette consumption).

4

Part c

Is it possible, in reference to the study design, to estimate the OR for bladder cancer and occupation from this study? If it is possible, calculate (using formulas) the crude odds ratio for bladder cancer comparing high-risk to other occupations, and estimate (using formulas) the associated 95% confidence intervals.

Part d

Compute the stratum-specific ORs at each of the two levels of cigarette consumption. Present the estimates with 95% Confidence interval.

Part e

Is it possible to estimate the RR for bladder cancer and occupation from this study? If it is possible, calculate the crude relative risk for bladder cancer comparing high-risk to other occupations, as well as the stratum-specific RRs at each of the two levels of cigarette consumption. Present the estimates with 95% Confidence interval. If it is not possible, give your reason.

Part f

We will assume we found that odds ratios are relatively the same using the Breslow-Day test (which is true). Calculate the smoking-adjusted Mantel-Haenszel odds ratio for high-risk versus other occupations. Present the estimates with 95% Confidence interval.

Part g

Are the odds of bladder cancer different for high-risk versus other occupations after we adjusted for smoking status? (Hint: You should be running an inference test.)

Part h

Do you think the smoking-adjusted Mantel-Haenszel odds ratio for high-risk versus other occupations should be reported? Give your reason.

Question 4

One important aspect of medical diagnosis is its reproducibility. Suppose that two different doctors examine 100 patients for dyspnea in a respiratory-disease clinic, and that doctor A diagnosed 15 patients as having dyspnea (while doctor B did not), doctor B diagnosed 10 patients as having dyspnea (while doctor A did not), and both doctor A and doctor B diagnosed 7 patients as having dyspnea.

Part a

Construct a two-way contingency table to summarize the dyspnea diagnoses from doctor A and B.

Part b

Compute the Cohen's kappa (No need to compute the confidence interval.)

Part c

How would you characterize the agreement between doctor A and B? Please refer to any guidelines used.

Part d

Is your computed kappa greater than 0? Run the appropriate test and interpret your results.