STAT 511A Homework 10

Kathleen Wendt

11/27/2019

Load packages

```
library(broom)
library(epitools)
```

Question 1

Bacillus Calmette-Guerin (BCG) is a vaccine for preventing tuberculosis. For this question, we will examine data from 3 studies (Vandiviere et al 1973, TPT Madras 1980, Coetzee & Berjak 1968). The data is summarized below.

A note about the BCG vaccine from Wikipedia: The most controversial aspect of BCG is the variable efficacy found in different clinical trials that appears to depend on geography. Trials conducted in the UK have consistently shown a protective effect of 60 to 80%, but those conducted elsewhere have shown no protective effect, and efficacy appears to fall the closer one gets to the equator.

Data tables

Study 1 data

```
tb_study1 <- matrix(c(619, 10, 2537, 8), nrow = 2, byrow = TRUE)
colnames(tb_study1) <- c("tbneg", "tbpos")
rownames(tb_study1) <- c("ctrl", "trt")
tb_study1
```

```
##      tbneg tbpos
## ctrl   619    10
## trt   2537     8
```

Study 2 data

```
tb_study2 <- matrix(c(87892, 499, 87886, 505), nrow = 2, byrow = TRUE)
colnames(tb_study2) <- c("tbneg", "tbpos")
rownames(tb_study2) <- c("ctrl", "trt")
tb_study2
```

```
##      tbneg tbpos
## ctrl 87892   499
## trt  87886   505
```

Study 3 data

```
tb_study3 <- matrix(c(7232, 45, 7470, 29), nrow = 2, byrow = TRUE)
colnames(tb_study3) <- c("tbneg", "tbpos")
rownames(tb_study3) <- c("ctrl", "trt")
tb_study3
```

```
##      tbneg tbpos
## ctrl  7232    45
## trt   7470    29
```

Part 1A

Calculate the odds ratio (corresponding to TBpos for Trt vs Ctrl) for each study separately. (4 pts)

Study 1 odds ratio

```
tb1_odds <- epitools::oddsratio(tb_study1, method = "wald")
```

```
## Warning in chisq.test(xx, correct = correction): Chi-squared approximation
## may be incorrect
```

```
tb1_odds$measure
```

```
##                         NA
## odds ratio with 95% C.I.  estimate     lower     upper
##                      ctrl 1.0000000        NA        NA
##                      trt  0.1951912 0.0767186 0.4966148
```

Study 2 odds ratio

```
tb2_odds <- epitools::oddsratio(tb_study2, method = "wald")
tb2_odds$measure
```

```
##                         NA
## odds ratio with 95% C.I. estimate     lower    upper
##                      ctrl 1.000000        NA       NA
##                      trt  1.012093 0.8940029 1.145782
```

Study 3 odds ratio

```
tb3_odds <- epitools::oddsratio(tb_study3, method = "wald")
tb3_odds$measure
```

```
##                         NA
## odds ratio with 95% C.I.  estimate     lower     upper
##                      ctrl 1.0000000        NA        NA
##                      trt  0.6239119 0.3907892 0.9961027
```

Part 1B

Use the Breslow-Day test to test for equality of odds ratios across the 3 studies. State your p-value and conclusion. Can we conclude that the odds ratios are equal across the 3 studies? Based on this test, should be combine information across studies? (4 pts)

Question 2

Problem 10.36 involves bomb hits during WWII. Bomb hits were recorded in n = 576 grids in a map of a region of South London.

Bomb Hits 0 1 2 3 ge4 Total

Grids 229 211 93 35 8 576

Part 2A

Find the sample mean (μ) bomb hits per grid.

Part 2B

Use the GOF test to test whether the number of bomb hits per grid follows the Poisson distribution. Calculate the GOF test statistic, df, p-value and give a conclusion using $\alpha = 0.05$. (6 pts)

Question 3

The data "PoissonData.csv" gives observations Y (counts or events) for n = 50 (units) generated from the Poisson distribution (using the rpoiss() function).

Part 3A

Calculate the sample mean and sample standard deviation. Also construct a histogram and qqplot of the data and include them in your assignment. (4 pts)

NOTE: Because the data comes from the Poisson distribution, you should find that the mean and the sample variance (s^2) are close. However, you should also find from the histogram and qqplot that the data looks approximately normal.

Part 3B

Give a standard t-based 95% confidence interval for μ.

Part 3C

Following the example on CH10 Slide 106 (Death by Mule Kick CI), construct a 95% confidence interval for μ based on the normal approximation to the Poisson distribution. (4 pts) In order to do this, you will start by constructing a CI on the total number of events, then divide by the number of units.

NOTE: The CIs from parts B and C should be similar.