Assignment 7 - Inference about a Population Rate (λ). Due November 9, 11:59pm 2018

EPIB607 - Inferential Statistics[a]

[a] Fall 2018, McGill University

This version was compiled on October 31, 2018

In this assignment you will practice conducting inference for a one sample rate. Answers should be given in full sentences (DO NOT just provide the number). All figures should have appropriately labeled axes, titles and captions (if necessary). Units for means and CIs should be provided. State your hypotheses and assumptions when applicable. All graphs and calculations are to be completed in an R Markdown document using the provided template. You are free to choose any function from any package to complete the assignment. Concise answers will be rewarded. Be brief and to the point. Please submit both the compiled HTML report and the source file (.Rmd) to myCourses by November 9, 2018, 11:59pm. Both HTML and .Rmd files should be saved as 'IDnumber_LastName_FirstName_EPIB607_A7'.

Poisson distribution | Rates | Intensity | Bootstrap

Template

The `.Rmd` template for Assignment 7 is available here

1. Cancers near Pickering nuclear station

The observed number of cases of leukemia was 18 in the Pickering nuclear station experience. We want to test if this observed number is compatible with the expected number of 12.8 cases, or whether it is greater than expected.

a. Calculate the standardized incidence ratio (SIR).
b. State the null and alternative hypotheses. Calculate the p-value for this test and provide a statement about the evidence.
c. Compute a 95% CI for the SIR. Does your statement in part (b) agree with the 95% CI?

2. Self-reported percutaneous injuries in interns

Read Section 6.3 of JH notes on rates. Reproduce the CIs for the 3 P's (Pediatrics Psychiatry Pathology) in Table 1. Show your calculations. You may not use canned functions.

3. RCT of HPV vaccine

This question is based on Table 3 of the article A Controlled trail of a Human Papillomavirus Type 16 Vaccine. Laura A. Koutsky et al., for The Proof of Principle Study Investigators. The New England Journal of Medicine Vol 347 Nov 21, 2002, p1645 (also available on myCourses).

a. For the primary per-protocol efficacy analysis, calculate a 95% CI to accompany the point estimates of infection rate (per 100 woman years) for both the HPV-16 vaccine and placebo groups. Show your calculations. You may not use canned functions.

b. Based on your results in part (a), is there evidence to suggest that the HPV-16 vaccine reduced infection rates? Explain.

4. Deaths by Horsekicks in the Prussian Army

This question is in reference to the Deaths by Horsekicks in the Prussian Army example presented in the slides on one sample rates. The following dataset gives the number of deaths by horsekicks, by year and army corps:

```
horsekicks <- read.csv("horsekicks.csv")
head(horsekicks)
```

```
#    year corps1 corps2 corps3 corps4 corps5 corps6 corps7 corps8 corps9
#  1 1875      0      0      0      0      0      0      0      1      1
#  2 1876      2      0      0      0      1      0      0      0      0
#  3 1877      2      0      0      0      0      0      1      1      0
#  4 1878      1      2      2      1      1      0      0      0      0
#  5 1879      0      0      0      1      1      2      2      0      1
#  6 1880      0      3      2      1      1      1      0      0      0
#    corps10 corps111 corps12 corps13 corps14 total
#  1       0        0       0       1       0     3
#  2       0        0       0       1       1     5
#  3       0        1       0       2       0     7
#  4       0        1       0       1       0     9
#  5       0        0       2       1       0    10
#  6       2        1       4       3       0    18
```

a. Create a table of observed and expected frequencies, like the one shown in the slides on one sample rates

b. Calculate a 95% CI for the rate parameter.

c. (**BONUS**) Compute a goodness of fit statistic that measures the discrepancy between the observed and expected frequencies. Be sure to group the categories that have low expected values. Include in your answer, the sources you used to help you with this calculation. Write a sentence about your findings.

5. Crimes on full moon

This question is based on Tables I and II of the article: Full moon and crime.Br Med J (Clin Res Ed) 1984; 289 (also available on myCourses). The incidence of crimes reported to three police stations in different towns (one rural, one urban, one industrial) was studied to see if it varied with the day of the lunar cycle. The period of the study covered 1978-1982.

a. Compare crime rates on FullMoon vs. NotFullMoon days (Table I) statistically by calculating a confidence interval for the daily rates in the two segments of time and seeing if they overlap.

b. We don't recommend this way of testing differences; instead we recommend computing a single CI for the rate difference or rate ratio. Following the example of the comparison of bone mineral losses (A4, Q3), compute a CI for the ratio and difference of daily rates by bootstrapping.