Homework 6

Point and Interval Estimation

Instructions

Please answer the following questions in the order in which they are posed. Add a few empty lines below each and write your answers there. **Focus on answering in complete sentences and show work whether we ask for it or not**. You will also need scratch paper/pen to work out the answers before typing it.

For help with formatting documents in RMarkdown, please consult R Markdown: The Definitive Guide. Another option is to search using Google.

Exercises

1. (Measurement error) A ph-meter is known to have systematic error[1] of size δ_0. In order to estimate δ_0, six measurements are made from a solution with pH **known** to be 4.84.

   The measurement model is that $X_1, X_2, \ldots, X_6$ are independently drawn from a probability distribution with mean $\mu_0 = 4.84 + \delta_0$ and some standard deviation σ_0. In other words, you are being told that

   $$E[X] = 4.84 + \delta_0, \quad Var[X] = \sigma_0^2.$$

a. Find the method of moments **estimator** of δ_0. Show your work.

b. Is your estimator from part a. unbiased for δ_0? Show your work.

c. Is your estimator from part a. consistent? Show your work.

2. (CLT) Suppose that the time (in days) until a component fails has a gamma distribution with shape $k = 5$ and rate $\lambda = \frac{1}{10}$. When a component fails, it is immediately replaced by a new component. Use the Central Limit Theorem to estimate the probability that 40 components will together be sufficient to last for at least 6 years. (You may assume a year has exactly 365.25 days)

   You may use R to perform the calculations but be sure to set up the problem mathematically first, and show your work and code.

3. The `MIAA05` basketball data contains statistics on 134 players in the MIAA 2005 Men's Basketball season. The following code chooses 100 different samples of size 15 from the dataset. From each sample, the mean and standard deviation of `PTSG` is calculated.

a. From each sample, calculate a 90% confidence interval for the mean PTSG (points per game) of MIAA players. Add the lower and upper limits of the confidence interval as additional columns called `lower` and `upper` in the `sample_summary` dataframe. Then print the first 10 rows of the dataframe. (Show your code and output)

b. The following code represents the confidence intervals you calculated in part a.as horizontal line segments. Fill in the `labs` layer. Also add a vertical line corresponding to the true mean `PTSG` in red. (You will need to calculate this from the `MIAA05` data.)

[1] this means it gives readings that are systematically higher or lower than what they should be

```
#remove eval=F once you have the previous code chunk entered
ggplot(data=sample_summary)+
  geom_segment(mapping = aes(x = lower,
                             xend = upper,
                             y = 1:nsamp,
                             yend = 1:nsamp)) +
  labs(x = " ",
       y = " ",
       title = " ") +
  theme(axis.text.y=element_blank(),
        axis.ticks.y=element_blank())
```

c. Of the confidence intervals you calculated in part a., how many contain the true mean `PTSG`? Write code to calculate this and show your code and answer below.

d. Suppose you bump up the sample size from 15 to 25. Would you *expect* more intervals to cover the true mean `PTSG`? Why or why not?

4. Suppose you want to estimate the mean shoe size of adults in a city. You would like to have a 95% confidence interval that is no wider than 0.5 shoe sizes (the margin of error would be at most 0.25). How large a sample must you get?

   a. This calculation will require that you make a guess about what approximately the standard deviation will be. What are the implications of guessing too high or too low? Should you guess on the low side or the high side?

   b. Should you include men and women in your sample or just one or the other? Why?

   c. Suppose you guess that the standard deviation of the population will be approximately 2. How large must your sample be to get the desired confidence interval?