# Unit 6: Homework

John Andrus

October 7, 2020

# 1 ML for an Exponential Distribution

Say that $[X_1, ..., X_n]$ is an i.i.d. sample from an exponential distribution, with common density,

$$f_X(x) = \lambda e^{-\lambda x}$$

Compute the maximum likelihood estimate for $\lambda$.

$$L(\lambda|X_1, ..., X_n) = L(\lambda|X_1)L(\lambda|X_2)...L(\lambda|X_n)$$

$$= \lambda e^{-\lambda X_1} \lambda e^{-\lambda X_2} ... \lambda e^{-\lambda X_n}$$

$$= \lambda^n [e^{-\lambda \sum_{i=0}^{n} X_i}]$$

Maximum likelihood equals $argmax(L)$

$$\frac{d}{d\lambda} L(\lambda|X_1...X_n) = \frac{d}{d\lambda} ln(\lambda^n [e^{-\lambda \sum_{i=0}^{n} X_i}]) = 0$$

$$\frac{d}{d\lambda} nlog(\lambda) - \lambda \sum_{i+1}^{n} X_i = 0$$

$$n = \lambda \sum_{i=1}^{n} X_i$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} X_i}$$

## 2 Did You Say T as in Zebra?

You record the total amount spent for a sample of 15 grocery store shoppers. You compute a 95

$$\left(\bar{X} - \Phi^{-1}(.975)\frac{s}{\sqrt{n}}, \bar{X} + \Phi^{-1}(.975)\frac{s}{\sqrt{n}}\right)$$

Where $\bar{X}$ is the sample mean, $s$ is the sample standard distribution, and $\Phi^{-1}$ is the quantile function (qnorm) for the standard normal distribution. The result you get is [50,70].

1. What is your sample mean and sample standard distribution?

   The sample mean is the midpoint of the confidence interval.

   $$\bar{X} = 60$$

   The sample standard distribution can be calculated by rearranging the endpoint expression

   $$s = (\bar{X} - 50)\frac{\sqrt{n}}{\Phi^{-1}(.975)}$$

   Where,

   $$\bar{X} = 60, n = 15, \Phi^{-1}(.975) = 1.95996$$

   Solving,

   $$s = 19.76$$

2. If you had correctly used a t-distribution, what would your confidence interval have been?

   First compute the T-Score,

   $$df = 15 - 1 = 14$$
   $$p = \frac{1 - .05}{2} = .025$$
   $$T = 2.14479$$

   Next compute the standard error,

   $$\frac{\sigma}{\sqrt{n}} = \frac{19.76}{\sqrt{15}} = 5.102$$

   The confidence interval is equivalent to $\bar{X} \pm T * SE$

   $$60 \pm 10.94$$

3. What is the true confidence level for the interval you computed with a normal distribution? In other words, what fraction of confidence intervals constructed in this way would include the mean?

$$10 = T\frac{\sqrt{\sigma}}{n}$$

Inputting values from part 2

$$10 = 5.099T$$

Which gives a T-Score of approximately 1.96. Using a T-Solver, it is determined that $[50, 70]$ is actually approximately a 93 percent confidence interval.

# 3 Low-Oxygen Statistics

The file expeditions.csv contains data about 10,000 climbing expeditions in the Himalayan Mountains of Nepal. The data was compiled by the Himalayan Database and published in csv format on Tidy Tuesday.

First, navigate to https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-09-22 to read some basic information about the data and examine the codebook.

The variable texthighpoint metres represents the highest elevation reached by each expedition. Your task is to construct a 95 percent confidence interval for the mean highest elevation

1. Using the documentation about the data, your background knowledge, and the data itself, assess whether the assumptions underlying a valid t-based confidence interval are met.

   - Continuous / Ordinal Scale

     Max expedition height, measured in meters, is measured in a continuous scale.
   - Simple Random Sample

     The data is probably closer to a population than a random sample, but for the purpose of this analysis will be treated as a random sample.
   - Normally Distributed

     A quick histogram reveals that this data is not normally distributed. It has a left tail and sizeable peaks at what I assume are the summits of popular mountains.
   - Large Sample Size

     The sample size of approximately 10,000 constitutes a large sample size.

2. Compute the 95 percent confidence interval, by plugging in the values from the data manually into the formula.

   See attached RMD

3. Confirm that your 95 percent confidence interval is correct, by running the t.test command and examining the confidence interval included in the output.

   See attached RMD