

In activity #13, we explored the concept of *maximum likelihood estimation* using the binomial distribution. In this assignment, you'll find maximum likelihood estimators for the exponential and geometric distributions.

Situation: Suppose that after my office hours begin, I sit and wait in my office for students to arrive. The times (in minutes) I wait for each subsequent student to arrive are: 5, 5, 10, 10, 20, 40.

If the time I wait can be modeled by an exponential distribution, I need to know a value of the parameter λ .

1. The average of those six waiting times is 15 minutes. Based on what we learned in activity #10, what would be your best guess for the value of λ ?

Answer: _____

2. As we learned in activity #10, the pdf for an exponential distribution is: $f(x) = \lambda e^{-\lambda x}$. If the times waiting for subsequent students are independent of one another, then:

$$f(x_1 \cap x_2 \cap \dots \cap x_6) = f(x_1)f(x_2)\dots f(x_6) = \prod_{i=1}^6 f(x_i)$$

The likelihood function to estimate λ , then, is:

$$L(\lambda; x_1, x_2, \dots, x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

Let's write out this likelihood function, substituting our six waiting times for x . Complete the following...

$$\prod_{i=1}^6 \lambda e^{-\lambda x_i} = (\lambda e^{-5\lambda})(\lambda e^{-5\lambda})(\lambda e^{-10\lambda})(\lambda e^{-10\lambda})(\text{_____})(\text{_____})$$

3. We want to find the value of λ that will maximize that function. In other words, we want to find the value of λ that has the best chance of producing the waiting times we actually observed.

Before we find this maximum likelihood, let's simplify our likelihood function. As we learned in class, taking the natural log of a likelihood function can make it easier to deal with. Let's start with our likelihood function:

$$(\lambda e^{-5\lambda})(\lambda e^{-5\lambda})(\lambda e^{-10\lambda})(\lambda e^{-10\lambda})(\lambda e^{-20\lambda})(\lambda e^{-40\lambda})$$

We can combine like terms and use properties of exponents to simplify this to:

$$\lambda^6 e^{-90\lambda}$$

We can then take the natural log to convert this to:

$$\ln(L) = \ln(\lambda^6 e^{-90\lambda}) = \ln(\lambda^6) + \ln(e^{-90\lambda}) = 6\ln(\lambda) - 90\lambda$$

4. Now that we have a simplified (natural log) form of our likelihood function, let's find the maximum.

First, let's find this maximum graphically using [desmos.com](https://www.desmos.com) (the website we also used in assignment #12).

a) Go to <http://www.desmos.com> and click **LAUNCH CALCULATOR**

b) On the left, you can enter functions that will be graphed to the right.

c) We want to graph our likelihood function. Since desmos treats x as the unknown independent variable, we'll type "x" instead of λ :

$$y = 6 \ln(x) - 90x$$

d) Once you type this in, you'll notice... nothing. We need to set the scales for the x- and y-axes to get the graph of our function to appear.

To change the scales of the axes, click the  icon at the top-right of the screen. That will open a pop-up menu that will allow you to type in new limits for your x- and y-axes. Set the x-axis to go from 0 to 2. Set the y-axis to range from -100 to 0. You should end up with a graph similar to what appears below.

e) Now, click the graph near that maximum to have desmos tell you the coordinates of that point. What value of x (λ) maximizes our likelihood function?



Answer: _____

Does that match your answer to question #1? _____

5. We can also find the maximum of our likelihood function by taking the first derivative and setting it equal to zero.

Do that below. Find the derivative of your likelihood function, set it equal to zero, and demonstrate you get the same answer you wrote above.

$$\text{Likelihood function} = 6 \ln(x) - 90x$$

6. In activity #10, we learned the best value to substitute for λ in an exponential function is $1 / E(x)$, or one divided by the average of our sample wait times.

It's worth noting that this is the maximum likelihood estimator. See if you can follow this derivation:

Our maximum likelihood function, for the exponential distribution, is:

$$L(\lambda; x_1, x_2, \dots, x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

We can simplify this to:

$$\prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

Taking the natural log, we obtain:

$$\ln(L) = \ln\left(\lambda^n e^{-\lambda \sum x_i}\right) = \ln(\lambda^n) + \ln\left(e^{-\lambda \sum x_i}\right) = n \ln(\lambda) - \lambda \sum x_i$$

To maximize this function, we take the first derivative...

$$\frac{d}{d\lambda} (n \ln(\lambda) - \lambda \sum x_i) = \frac{n}{\lambda} - \sum x_i$$

... and set it equal to zero to find:

$$\lambda = \frac{n}{\sum x_i}$$

Take a look at that last result. To find the value of λ that maximizes the likelihood of observing a set of data, we count up the number of observations we have and divide it by their sum. That's exactly the reciprocal of the formula we use to calculate an average. Therefore, we just showed the best estimate for λ is to take $1 / E(x)$.

Situation: Suppose I'm interested in the first question students answer correctly on an exam. In a small class of 5 students, the first question answered correctly by each student are questions #1, 3, 5, 10, 11.

If the number of questions needed to find the first correct answer is modeled by a geometric distribution, I need to know a value of the parameter p .

7. In activity #8, we learned the pdf for a geometric distribution is: $P(x = k) = (1 - p)^{k-1} p$. If we have independent observations, then:

$$L(p) = [(1-p)^{1-1} p] [(1-p)^{3-1} p] [(1-p)^{5-1} p] [(1-p)^{10-1} p] [(1-p)^{11-1} p]$$

We can simplify this likelihood function to:

$$L(p) = p^5 (1-p)^{0+2+4+9+10} = p^5 (1-p)^{25}$$

Taking the natural log, we find:

$$\ln(L) = \ln(p^5 (1-p)^{25}) = \ln(p^5) + \ln((1-p)^{25}) = 5 \ln(p) + 25 \ln(1-p)$$

Maximize this function.

First, use [desmos.com](https://www.desmos.com) to maximize it graphically. Write the value of p that maximizes this function here: _____

Then, find the derivative and set it equal to zero. Make sure you get the same answer that you just wrote above.

8. Let's generalize these results to find the maximum likelihood estimator for p in any geometric distribution.

Our maximum likelihood function, for the geometric distribution, is:

$$L(p) = \prod_{i=1}^n (1-p)^{k-1} p = p^n (1-p)^{\sum x_i - n}$$

Taking the natural log, we obtain:

$$\ln(L) = \ln(p^n (1-p)^{\sum x_i - n}) = n \ln(p) + (\sum x_i - n) \ln(1-p)$$

To maximize this function, we take the first derivative...

$$\frac{d}{dp} [n \ln(p) + (\sum x_i - n) \ln(1-p)] = \frac{n}{p} - \frac{\sum x_i - n}{1-p}$$

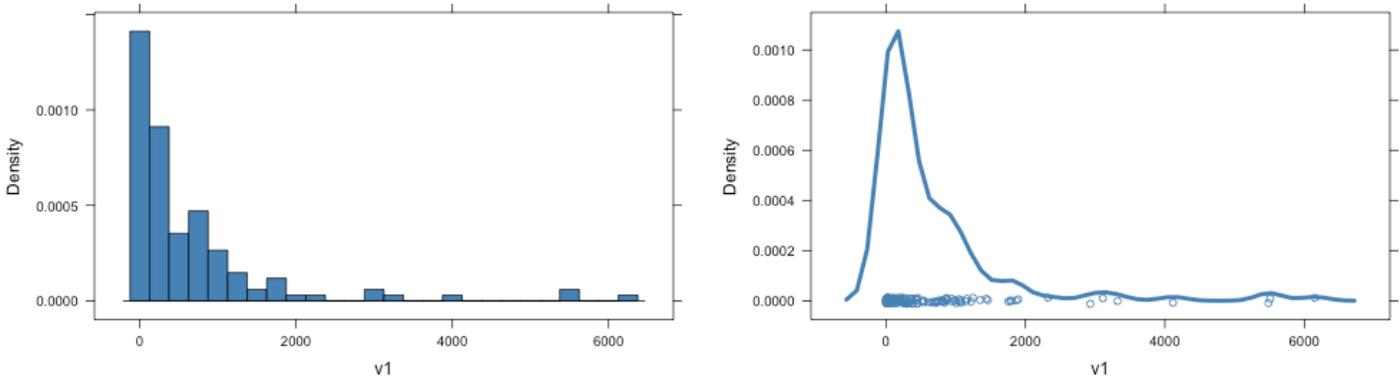
... and set it equal to zero to find:

$$p = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}$$

Take a look at that last result. To find the value of p that maximizes the likelihood of observing a set of data, we count up the number of observations we have and divide it by their sum. That's exactly the reciprocal of the formula we use to calculate an average. Therefore, we just showed the best estimate for p is to take 1 / E(x).

Situation: Suppose I've developed a new computer hard drive. Before I start selling this hard drive, I'd like to know the average amount of time it will take before the hard drive crashes. I want to know this for the entire population of hard drives I may sell, but I only have collected a sample of data from 136 hard drives.

The following histogram and density plot display the distribution of my sample data (measured in days).



9. Here are summary statistics calculated from these 136 observations:

min	Q1	median	Q3	max	mean	sd	n
0	80.5	272	831.75	6150	649.1397	1036.02	136

From this, you can see the average is 649.1397. For now, let's pretend these 136 observations are our entire population. We can then think of our average as a population average.

Ok, now that we have our population, let's forget it exists. We want to sample some hard drives, record how long it takes them to crash, and use that data to estimate the population average (which we wouldn't know is 649.1397).

Suppose we sample 50 hard drives and observe the following crash times:

```

75 386 2 143 1247 330 1351 108 245 281 790 1897 816 369 670 4116 193
232 828 242 724 30 748 50 36 138 983 1800 97 600 197 325 180 1435
1222 255 1082 21 1064 0 843 10 446 422 15 729 1160 810 3110 148

```

From this data, we might decide to calculate:

- Sample mean = 660.02
- 10% trimmed mean = 514.875
- 25% trimmed mean = 459.6923
- Median = 377.5

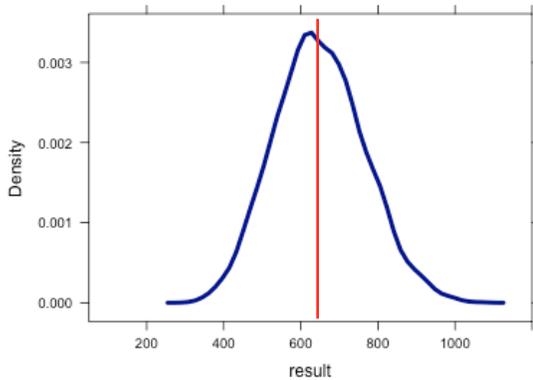
Obviously, the sample mean (660.02) was closest to our population mean (649.1397), but we wouldn't know that if we didn't know our population. Likewise, the sample mean was closest for this single sample of 50 observations. What if we were to repeatedly sample 50 observations and calculate the mean, trimmed means, and median. On average, which one would give us the best estimate of the population mean?

Go ahead and guess. Which sample statistic do you think would be the best estimator of our population mean?

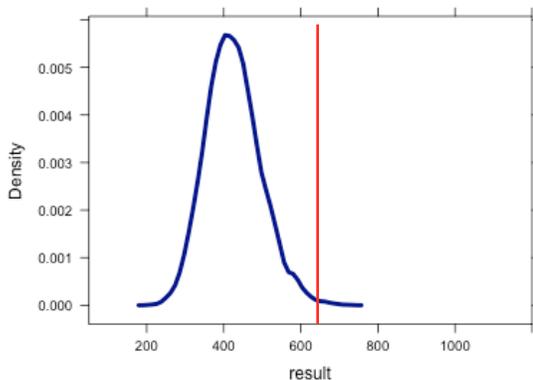
Circle one of the following: **SAMPLE MEAN** **10% TRIMMED MEAN** **25% TRIMMED MEAN** **MEDIAN**

10. As we learned in class, we may be interested in selecting an estimator that is unbiased with a small variance.

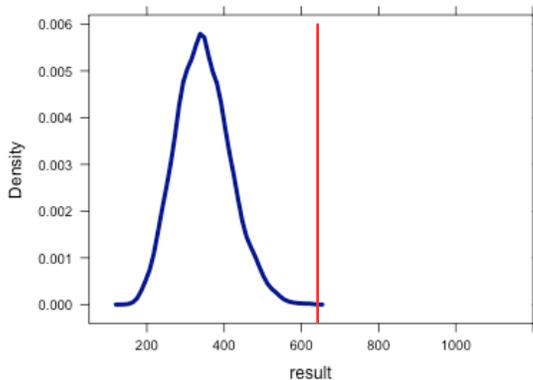
To investigate the bias and variance of these possible estimators of the population mean, I had a computer take 10,000 samples of size $n=50$ from our hard drive crash data. For each sample, the computer calculated the sample mean, the trimmed means, and the median. The following histograms and sample statistics were found:



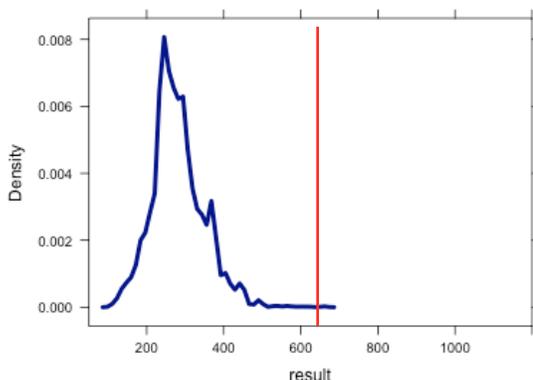
The graph to the left shows sample means from the 10,000 samples.
The mean of these sample means is 648.8458
The standard deviation of these sample means is 116.3446
The vertical line shows the location of the population mean.



This graph shows 10% trimmed means from the 10,000 samples.
The mean of these 10% trimmed means is 424.767
The standard deviation of these 10% trimmed means is 70.59756
The vertical line shows the location of the population mean.



This graph shows 25% trimmed means from the 10,000 samples.
The mean of these 25% trimmed means is 345.9071
The standard deviation of these 25% trimmed means is 69.71196
The vertical line shows the location of the population mean.



This graph shows medians from the 10,000 samples.
The mean of these medians is 281.0999
The standard deviation of these medians is 66.74158
The vertical line shows the location of the population mean.

11. Circle your answers to the following:

a) Which estimator appears to have the least bias?

SAMPLE MEAN 10% TRIMMED MEAN 25% TRIMMED MEAN MEDIAN

b) Which estimator appears to have the most bias?

SAMPLE MEAN 10% TRIMMED MEAN 25% TRIMMED MEAN MEDIAN

c) Which estimator appears to have the smallest variance?

SAMPLE MEAN 10% TRIMMED MEAN 25% TRIMMED MEAN MEDIAN

d) Which estimator appears to have the largest variance?

SAMPLE MEAN 10% TRIMMED MEAN 25% TRIMMED MEAN MEDIAN

e) Which estimator would you choose as being the best estimator of the population mean?

SAMPLE MEAN 10% TRIMMED MEAN 25% TRIMMED MEAN MEDIAN

Why? _____

== Here's the R code I used to generate the examples in this assignment ==

```
library(mosaic)
crashes <- read.csv(url("http://bradthiessen.com/html5/stats/m300/failures.csv"))
histogram(~v1, data=crashes, width=250, col="steelblue")
densityplot(~v1, data=crashes, auto.key=TRUE, lw=4, col="steelblue")
favstats(~v1, data=crashes)

meanCrashes <- do(10000) * mean(~v1, data = sample(crashes, 50))
trim_mean_10 <- do(10000) * mean(~v1, trim=0.10, data = sample(crashes, 50))
trim_mean_25 <- do(10000) * mean(~v1, trim=0.25, data = sample(crashes, 50))
medianCrashes <- do(10000) * median(~v1, data = sample(crashes, 50))

densityplot(~result, data=meanCrashes, plot.points = FALSE, xlim=c(150,1100), col="darkblue", lwd=4)
densityplot(~result, data=trim_mean_10, plot.points = FALSE, xlim=c(150,1100), col="darkblue", lwd=4)
densityplot(~result, data=trim_mean_25, plot.points = FALSE, xlim=c(150,1100), col="darkblue", lwd=4)
densityplot(~result, data=medianCrashes, plot.points = FALSE, xlim=c(150,1200), col="darkblue", lwd=4)
```