

Activity 15

Population vs My Sample

Under Result Type choose "Simulation", set the seed to your Net ID. Run 1 repetition with a sample size of 15. Report the population parameters and your sample statistics by switching the 'Variable' and Sample Statistic of interest.

will be
different
for everyone

	Price Mean	Price Variance	Duration Mean	Duration Variance	Wait time Mean	Wait time Variance	Slope	Intercept
Population	19.50	20.8	30.10	97.6	4.51	9.10	0.45	5.94
Sample	20.63	20.34	31.87	92.80	3.60	5.26	0.461	5.946

Are your sample values close to the parameter values?

The means are all relatively close to the population parameters. The variances tend to be slightly underestimated.

Class Data Sample

Under Result Type choose "Class Data" and enter your Net ID. Use variable: Price, Sample statistic: mean, and sample size 15. Run the simulation.

This shows the distribution of each member of the classes sample statistics. That are 114 repetitions of sample size 15.

Report the overall mean of each sample statistic (ie: our class sample average).

	Price Mean	Price Variance	Duration Mean	Duration Variance	Wait time Mean	Wait time Variance	Slope	Intercept
Class Sample	19.63	20.83	30.38	97.78	4.54	9.13	0.45	5.95

What do you notice about the means of these sample statistics (Hint: compare them to the parameter values)?

All the means of the sample statistics are fairly close to their respective parameters.

Inspect the histogram of the following variables and statistics. What do you notice about these distributions? Where are they centering? What about their shape? Did your sample happen to fall close to the truth (population) or far?

- sample means of price:

The distribution appears to be centering around the respective parameter value 19.50.

The distribution is unimodal and symmetric.

My sample was a bit away and was an overestimate

- sample variances of price:

The distribution appears to be centering around the respective parameter value 20.8.

The distribution is unimodal and right skewed.

My sample happened to be very close to the population

- sample estimates of slope coefficient:

The distribution appears to be centering around the respective parameter value 0.45.

The distribution is unimodal and symmetric.

My sample happened to be quite close to the population

Simulate MANY samples

Under Result Type choose "Simulation", set the seed to your Net ID. Run 20000 repetition with a sample size of 15. Report the population parameters and your sample statistics by switching the 'Variable' and Sample Statistic of interest.

	Price Mean	Price Variance	Duration Mean	Duration Variance	Wait time Mean	Wait time Variance	Slope	Intercept
20,000 Samples	19.50	20.72	30.12	97.15	4.51	9.07	0.45	5.94

What do you notice about the means of these sample statistics (Hint: compare them to the parameter values)?

The values are VERY close to their respective parameter values.

Inspect the histogram of the following variables and statistics. What do you notice about these distributions? Where are they centering? What about their shape?

- Similar to what we saw with the class sample statistics, each distribution appears to be centering around their respective parameter value. Shape is easier to examine in this case because we have 20,000 data points instead of only 113 data points.
- sample means of price:

The sampling distribution of the sample means of price looks to be centering around \$19.50. The distribution is unimodal and symmetric.

- sample variances of price:

The sampling distribution of the sample variance of price looks to be centering around \$20.80. The distribution is unimodal and right skewed.

- sample estimates of slope coefficient:

The sampling distribution of the sample slope estimates looks to be centering around 0.45. The distribution is unimodal and symmetric.

Click on the population tab. Look at the distribution for the variable **price** (this is census data for ride shares).

How does this population histogram compare to the histogram of sample means of price (i.e. sampling distribution with 20000 repeated samples)? Can you explain the difference?

(Might help to flip back and forth between the Simulation tab with 20,000 repetitions of sample size 15 and the Population tab).

The distributions have the same shape (unimodal, no skew) and center (\$19.50). The DIFFERENCE is that the spread of the sampling distribution is less than the spread of the distribution of price.

Each data point/observation in the distribution of price represents ONE ride.

Each data point/observation in the sampling distribution represents a mean of 15 randomly selected ride prices.

When we take the mean of values the extremes work to cancel each other out, resulting in less spread in means when compared to spread of individual values.