Activity 16

We are going to use the ride share data again for this activity. We happen to have the entire census of ride shares (ie: we know the population). This is incredibly RARE in real life.

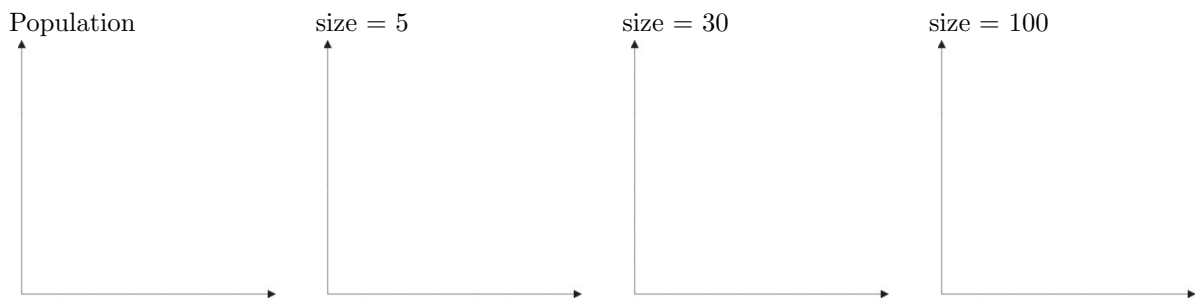## Sampling distribution and sample size

In general, how should a random sampling distribution relate to the population distribution?

Suppose we had a choice of taking a sample of size 10 or 300. Which one would you choose? Why?

Let's explore what happens to the sampling distribution of $\bar{x}$ as we change our sample size. We will use samples of size 5, 30, and 100. The number of repetitions will be 10,000. Sketch the distribution of each of the following and report the means and standard deviations.

Duration

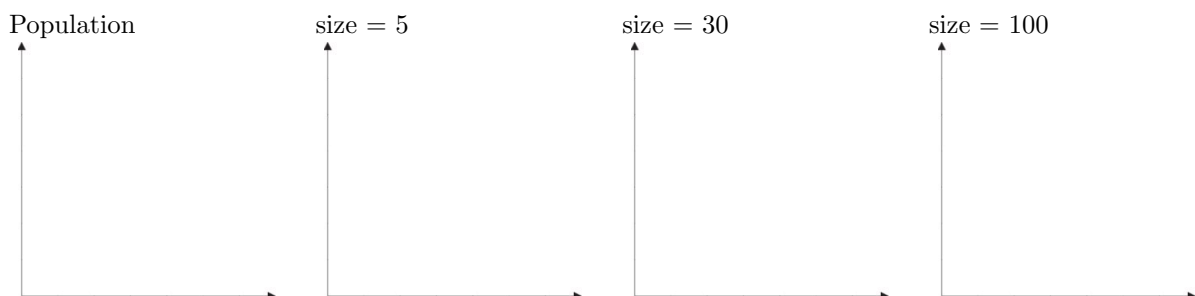Provide a description for the distribution of 'duration' population.

Population	size = 5	size = 30	size = 100

μ : $\bar{x}$: $\bar{x}$: $\bar{x}$:

σ : se : se : se :

Wait time

Provide a description for the distribution of 'wait_time' population.

Population	size = 5	size = 30	size = 100

μ : $\bar{x}$: $\bar{x}$: $\bar{x}$:

σ : se : se : se :

Central Limit Theorem (CLT)

The **CLT** tells us that for a sufficient sample size, the distribution of the sample means will be approximately normally distributed even if the population distribution is not normal! More specifically, for a variable X with mean μ_x and standard deviation σ_x we have $\bar{x} \sim N\left(\mu_{\bar{x}} = \mu_x, SE(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\right)$

Sampling Mean Check

What do you observe for the sampling distribution of the mean for 'wait_time' as the sample size gets larger? What about for 'duration'? Does the CLT hold true?

Standard Error Check

Using the σ's calculated and the standard error formula from the CLT (see Table 9.6 as well), calculate the theoretical standard errors and compare them to the simulated standard errors (s) for the three sampling distributions for each variable. Are the theoretical and simulated values close?
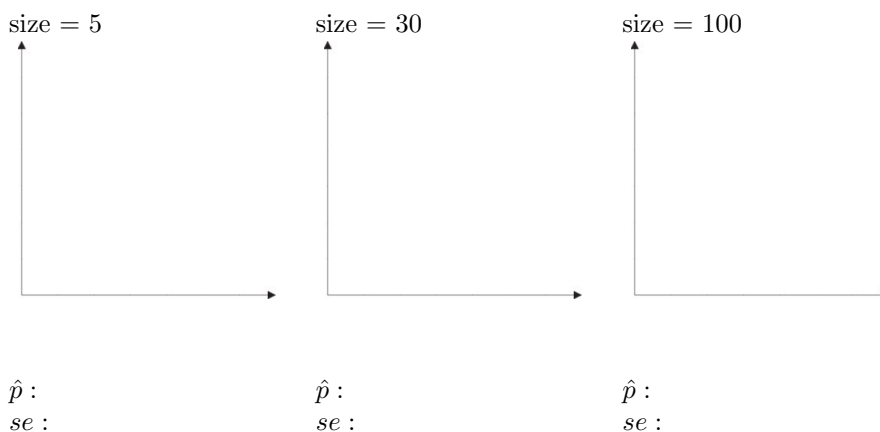
Big Picture

Explain how the sampling distribution is related to a single sample.

Explain how the sampling distribution is related to the population distribution.

Wait time under 5 minutes

The central limit theorem also applies to proportions. What is the population proportion (p) for 'wait_time_under5'?

size = 5   size = 30   size = 100

$\hat{p}:$ $\hat{p}:$ $\hat{p}:$
$se:$ $se:$ $se:$

Standard Error Check

Using the population proportion and the standard error formula from Table 9.6, calculate the theoretical standard errors and compare them to the simulated standard errors (se) for each sampling distribution.