Sampling Distributions

The Distribution of a Sample Mean: Shape

In this exercise, you will use the Shiny app *Sampling Distribution of the Mean* to explore the shape of the distribution of possible sample means when the probability distribution of the individual measurements is skewed and how this shape varies with sample size.

The app shows two plots. The first plot is labelled the "Most Recent Trial". The black curve on this plot is the probability density function that describes the behaviour of the individual measurements. Blue point(s) at the bottom are randomly generated observations from this probability distribution, that is, they are measurements whose random value is described by this probability distribution. The blue vertical bar is the average (or "sample mean") of this sample of observations. When the sample size is one, the sample mean is equal to the one observation. These sample means are plotted in the histogram in the second plot. Each time a different random sample of observations is generated, the resulting sample mean is added to the histogram; in the first plot, only the last random sample of observations and the corresponding last sample mean is shown.

The app has three adjustable parameters:

- Sample Size per Trial – The number of sample observations generated when clicking on the Generate Samples button.
- Skew – Controls how right-skewed (if positive) or left-skewed (if negative) the probability distribution of the individual measurements is. When the skew is 0, the probability distribution is symmetric.
- Number of Trials – The number of samples, each of which has the number of observations indicated by the sample size, and the corresponding number of resulting sample means that are generated when clicking on the Generate Samples button. Adjusting this parameter allows you to generate many sample means quickly, to observe the resulting behaviour of the sample means in the histogram.

When carrying out a study, the researcher only gets one sample, or one set of observations, and thus gets one sample mean. But a key idea in reasoning statistically about the results of the study is this:

> Every study that is carried out should be viewed as a single outcome from a population of studies that could have been performed under the same conditions.

So the particular sample mean that the researcher got in his or her one study that was carried out is one of many possible sample means that could have been observed under the same

conditions. To carry out statistical analysis on the results, it is important to understand the probability distribution of this hypothetical population of possible sample means.

Follow the instructions below to explore how the shape of the distribution of sample means compares to the shape of the distribution of the individual measurements.

1. Set the parameters of the app as follows:
   Sample Size per Trial = 1; Skew = a positive value between 4 and 6; Number of Trials = 1
   Repeatedly generate a sample by pressing the Generate Samples button many times. Notice how the values of each of the random observations vary. What do you observe?

2. Increase the Number of Trials to get a large number of samples. Observe the distribution of the sample means in the histogram. How does the shape and the location of the peak(s) of the histogram compare to the probability distribution of the individual observations?

3. Change Sample Size per Trial to a number between 5 and 10 and Number of Trials to 1. Note that the first plot now shows each of the individual measurements in the sample and the sample mean, that is, the average of these observations. The sample mean is also plotted in the histogram in the second plot. Generate many samples and observe how the samples behave and how the sample means behave. What do you observe?

4. Generate many samples by adjusting Number of Trials. How does the shape of the distribution of the sample means in the second plot compare to the shape of the probability distribution of the individual observations in the first plot?

5. Experiment with larger values of skew and various sample size, repeating 3. and 4. How do your observations differ with larger skew? Does sample size matter?

6. Fill in the blanks:
   Even if the distribution of the individual observations is _____, the distribution of the sample mean is approximately _____ and becomes even more so as the sample size _____.