Statistics:
Unlocking the Power of Data

SECTIONS 3.1
• Sampling Distributions (3.1)

Statistical Inference

Statistical inference is the process of drawing conclusions about the entire population based on information in a sample.

Statistic and Parameter

A parameter is a number that describes some aspect of a population.
A statistic is a number that is computed from data in a sample.

• We usually have a sample statistic and want to use it to make inferences about the population parameter

Parameter versus Statistic

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\mu$</td>
</tr>
<tr>
<td>Proportion</td>
<td>$p$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>$\sigma$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$\rho$</td>
</tr>
</tbody>
</table>
Reese’s Pieces

- What proportion of Reese’s Pieces are orange?
- \( p = \) proportion of Reese’s Pieces that are orange
- \( \hat{p} = \) _____
- Get an estimate from one sample.
  \( p = ??? \)

Point Estimate

- We use the statistic from a sample as a point estimate for a population parameter.
- Point estimates will not match population parameters exactly, but they are our best guess, given the data

Key Question and Answer

- **Key Question:** For a given sample statistic, what are plausible values for the population parameter? How far might the true population parameter be from the sample statistic?
- **Key answer:** It depends on how much the statistic varies from sample to sample!

How far might the population parameter fall from the sample statistic?

\[ \hat{p}, \ p? \]

GOAL: Identify an interval of plausible values. (this is called an interval estimate).

More Samples

- Let’s collect a few more point estimates:
  - Important point: Sample statistics vary from sample to sample, and knowing how much a statistic varies from sample to sample helps us assess uncertainty in the statistic!

Lots of Samples

- To really see how statistics vary from sample to sample, let’s take lots of samples and compute lots of statistics!
  - Enter your sample proportion on in the google form (emailed to you)
  - (if you don’t have a computer, have someone near you enter your number)
  - You just made your first sampling distribution!
**Sampling Distribution**

A *sampling distribution* is the distribution of sample statistics computed for different samples of the same size from the same population.

- A sampling distribution shows us how the sample statistic varies from sample to sample.

**Center and Shape**

**Center:** If samples are randomly selected, the sampling distribution will be centered around the population parameter.

**Shape:** For most of the statistics we consider, if the sample size is large enough, the sampling distribution will be symmetric and bell-shaped.

**Sampling Caution**

- If you take *random samples*, the sampling distribution will be centered around the true population parameter.

- If sampling bias exists (if you do not take random samples), your sampling distribution may give you bad information about the true parameter.

  “The polls have stopped making any sense.”

**Remember the Rectangles!**

We really care about the spread of the statistic…

How much do statistics vary from sample to sample?
**Statistics: Unlocking the Power of Data**

**Standard Error**

The **standard error** of a statistic, SE, is the standard deviation of the sample statistic.

- The standard error measures how much the statistic varies from sample to sample.
- The standard error can be calculated as the standard deviation of the sampling distribution.

**Reese's Pieces**

The standard error for \( \hat{p} \), the proportion of orange Reese's Pieces in a random sample of 10, is closest to:

- a) 0.05
- b) 0.15
- c) 0.25
- d) 0.35

**Distance from parameter to statistic gives distance from statistic to parameter**

SE can be used to determine width of interval!

**Lower SE means statistics closer to true parameter value...**

SE measures “typical” distance between parameter and statistic.

**The larger the SE, the larger the interval**

The larger the SE, the larger the interval.
Sample Size Matters!

As the sample size increases, the variability of the sample statistics tends to decrease and the sample statistics tend to be closer to the true value of the population parameter.

- For larger sample sizes, you get less variability in the statistics, so less uncertainty in your estimates.

Reese's Pieces

- StatKey

Larger \( n \) gives smaller SE

So larger \( n \) means narrower intervals

Small \( n \)

\[ \hat{\theta} \]

Large \( n \)

\[ \hat{\theta} \]

Sample Size

Suppose we were to take samples of size 10 and samples of size 100 from the same population, and compute the sample means. Which sample means would have the higher standard error?

a) The sample means using \( n = 10 \)
b) The sample means using \( n = 100 \)

For the larger sample, statistics are less spread out.

Suppose we were to take samples of size 10 and samples of size 100 from the same population, and compute the sample means. Which sample would yield a wider interval estimate for \( \mu \)?

a) The sample of \( n = 10 \)
b) The sample of \( n = 100 \)
To Do

- Read Section 3.1
- HW 3.1 due MONDAY, 2/23