Section 8.1b
What Is a Sampling Distribution?

Learning Objectives

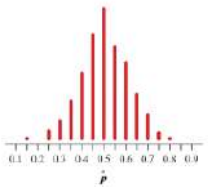
After this section, you should be able to…

- ✓ DISTINGUISH between a parameter and a statistic
- ✓ DEFINE sampling distribution
- ✓ DISTINGUISH between population distribution, sampling distribution, and the distribution of sample data
- ✓ DETERMINE whether a statistic is an unbiased estimator of a population parameter
- ✓ DESCRIBE the relationship between sample size and the variability of an estimator

Describing Sampling Distributions

The fact that **statistics from random samples**

- have **definite sampling distributions**
- allows us to answer the question: **"How trustworthy is a statistic as an estimator of the parameter?"**
- To get a complete answer, we consider the **center**, **spread**, and **shape**.



Note that the center of the approximate sampling distribution is close to 0.5. In fact, if we took ALL possible samples of size 20 and found the mean of those sample proportions, we'd get *exactly* 0.5.

Describing Sampling Distributions

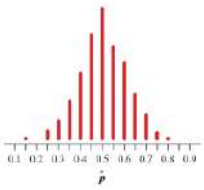
Center: Biased and unbiased estimators

Definition:

A statistic used to estimate a parameter is an **unbiased estimator** if the mean of its sampling distribution is equal to the true value of the parameter being estimated.

In the beans example:

- **We collected many samples** and calculated the **sample proportion of black beans.**
- How well does the **sample proportion estimate (phat)** the **true proportion** of black beans, ***p* = 0.5**?
- Therefore we can say the **sample proportion estimate** is an **unbiased estimator for the population.**



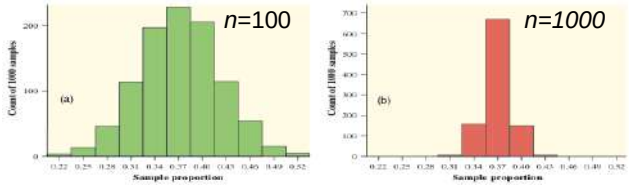
Describing Sampling Distributions

Next Spread: Low variability is better!

- **To get a trustworthy estimate of an unknown population parameter, start by using a statistic that's an unbiased estimator.**
- **This ensures that you won't tend to overestimate or underestimate.**
- **Unfortunately, using an unbiased estimator doesn't guarantee that the value of your statistic will be close to the actual parameter value. So next we look at variability.**

Spread: Low variability is better!

Larger samples have a clear advantage over smaller samples. They are much more likely to produce an estimate close to the true value of the parameter.



Variability of a Statistic

The **variability of a statistic** is described by the spread of its sampling distribution. This spread is determined primarily by the size of the random sample. Larger samples give smaller spread. The spread of the sampling distribution does not depend on the size of the population, as long as the **population is at least 10 times larger than the sample.**

Mathematical Theory Versus the Real World - The "10 percent rule"

- **The "10 percent rule"** is a numerical approximation that we may apply in AP Stats.
- To see the underlying mathematical theory, check out this link:
  http://apcentral.collegeboard.com/apc/members/courses/teachers_corner/39161.html

Describing Sampling Distributions

Bias, variability, and shape

We can think of the true value of the population parameter as the bull's-eye on a target

- **and of the sample statistic as an arrow fired at the target.**

Describing Sampling Distributions

Bias, variability, and shape

Both **bias** **and variability** describe what happens when we take many shots at the target.

Bias means that our aim is off and we consistently miss the bull's-eye in the same direction.

Our sample values do not center on the population value.

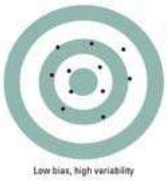
High bias, low variability

(a)

Describing Sampling Distributions

Bias, variability, and shape

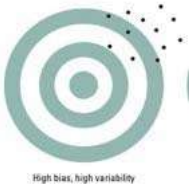
Both bias and **variability** describe what happens when we take many shots at the target.



Low bias, high variability

(b)

High variability means that repeated shots are widely scattered on the target. Repeated samples do not give very similar results.



High bias, high variability

(c)

The lesson about center and spread is clear: **given a choice of statistics to estimate an unknown parameter, choose one with**

1. **No or low bias and**
2. **Minimum variability**.

EXAMPLE

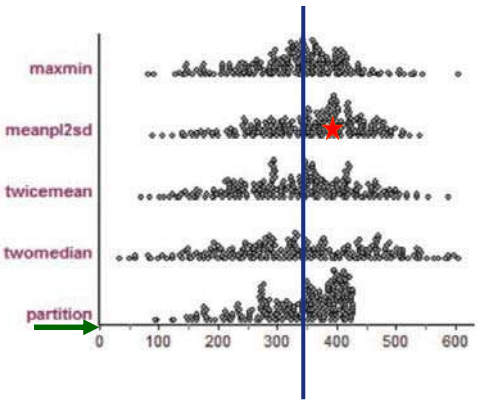
Sampling Distributions - Bias, variability, and shape

- Sampling distributions can take on many shapes.
- The same statistic can have sampling distributions with different shapes depending on the population distribution and the sample size.
- Be sure to consider the shape of the sampling distribution before doing inference.

EXAMPLE:

- Sampling distributions for different statistics used to estimate the number of German Tanks in WW II.
- The blue line represents the true number of tanks.
- The 1st four were presented by groups of students.
- The "Partition Method" was recommended by D.C. mathematicians.

1. Bias or unbias estimators?
2. Describe the different sampling distributions.
3. Which statistic gives the best estimator? Why?



Section 8.1
What Is a Sampling Distribution?

Summary

In this section, we learned that…

- ✓ A **parameter** is a number that describes a population. To estimate an unknown parameter, use a **statistic** calculated from a sample.
- ✓ The **population distribution** of a variable describes the values of the variable for all individuals in a population. The **sampling distribution** of a statistic describes the values of the statistic in all possible samples of the same size from the same population.
- ✓ A statistic can be an **unbiased estimator** or a **biased estimator** of a parameter. Bias means that the center (mean) of the sampling distribution is not equal to the true value of the parameter.
- ✓ The **variability** of a statistic is described by the spread of its sampling distribution. Larger samples give smaller spread.
- ✓ When trying to estimate a parameter, choose a statistic with low or no bias and minimum variability. Don't forget to consider the shape of the sampling distribution before doing inference.

APPENDIX

EXAMPLE: German tanks

Sampling Distributions - Bias, variability, and shape

Which estimators are Bias or Unbias?

- Meanpl2sd is a BIASED estimator. The center of the distribution is too high. This statistic produces consistent overestimates of the number of tanks.
- The other 4 statistics are unbias estimators.

Sampling Distribution – Describe Center, Spread (variability), Shape

- Student sampling distributions Maxmin, Twicemean and Twomedian are roughly symmetric shapes so these statistics are about equally likely to underestimate or overestimate the number of tanks.
- Among the 3 Student sampling distributions: Maxmin has the smallest variability and in general produce estimates that are closer to the actual number of tanks. Maxmin would be the best estimator from the students.
- Partition was developed by Washington DC mathematicians.
  - It is left skewed which means the statistic is more likely to overestimate than underestimate the number of tanks.
  - The math guys felt it would be better to err on the side of caution and give the military commanders an estimate that's slightly too high.