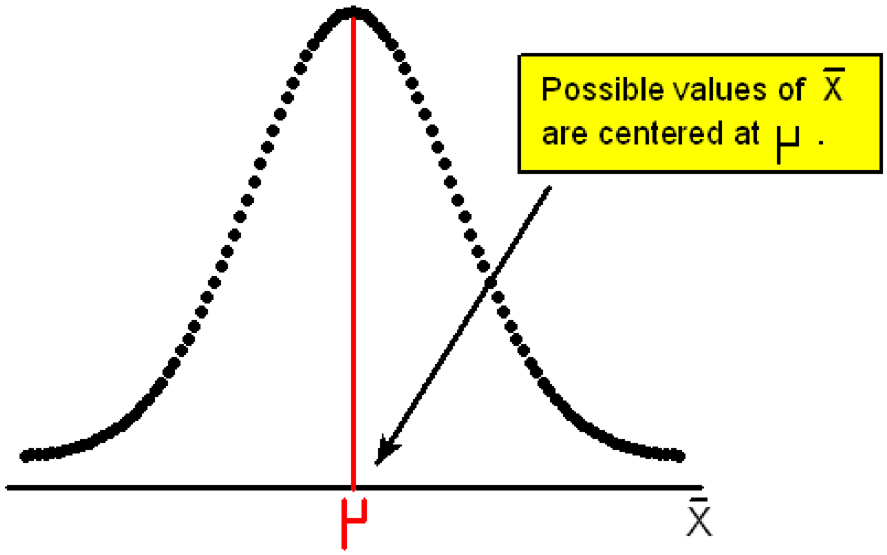
Point Estimation: Unbiased Estimators

Learning Objective: Determine point estimates in simple cases, and make the connection between the sampling distribution of a statistic, and its properties as a point estimator.

Comment 1

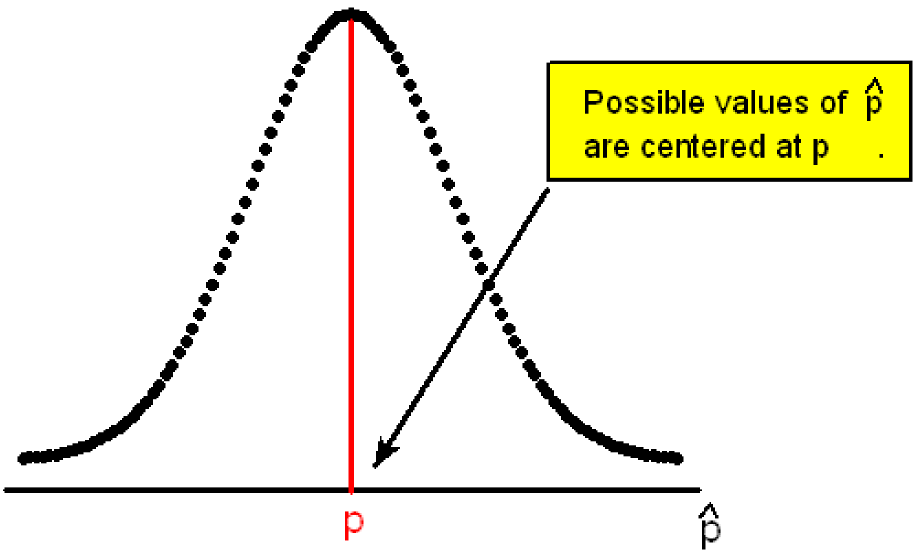
You may feel that since it is so intuitive, you could have figured out point estimation on your own, even without the benefit of an entire course in statistics. Certainly, our intuition tells us that the best estimator for μ should be $\overline{x}$, and the best estimator for p should be $\hat{p}$.

Probability theory does more than this; it actually gives an explanation (beyond intuition) **why** $\overline{x}$ and $\hat{p}$ are the good choices as point estimators for μ and p, respectively. In the Sampling Distributions module of the Probability unit, we learned about the sampling distributions of $\overline{X}$ and found that **as long as a sample is taken at random**, the distribution of sample means is exactly centered at the value of population mean.



$\overline{X}$ is therefore said to be an **unbiased estimator** for μ . Any particular sample mean might turn out to be less than the actual population mean, or it might turn out to be more. But in the long run, such sample means are "on target" in that they will not underestimate any more or less often than they overestimate.

Likewise, we learned that the sampling distribution of the sample proportion, $\hat{p}$, is centered at the population proportion p (as long as the sample is taken at random), thus making $\hat{p}$ an **unbiased estimator** for p.



As stated in the introduction, probability theory plays an essential role as we establish results for statistical inference. Our assertion above that sample mean and sample proportion are unbiased estimators is the first such instance.

Comment 2

Notice how important the principles of sampling and design are for our above results: if the sample of U.S. adults in (example 2 on the previous page) was not random, but instead included predominantly college students, then .56 would be a biased estimate for p, the proportion of all U.S. adults who believe marijuana should be legalized. If the survey design were flawed, such as loading the question with a reminder about the dangers of marijuana leading to hard drugs, or a reminder about the benefits of marijuana for cancer patients, then .56 would be biased on the low or high side, respectively. Our point estimates are truly unbiased estimates for the population parameter only if the **sample is random and the study design is not flawed.**

Did I Get This

1/1 point (graded)

A researcher wanted to estimate μ, the mean number of hours that students at a large state university spend exercising per week. The researcher collects data from a sample of 150 students who leave the university gym following a workout.

Which of the following is true regarding $\bar{x}$, the average number of hours that the 150 sampled students exercise per week?

- It is an unbiased estimate for μ.
- It is not an unbiased estimate for μ and probably underestimates μ.
- It is not an unbiased estimate for μ and probably overestimates μ. ✔

Answer

Correct:

It is not an unbiased estimator for μ because the sample was not a random sample of 150 students from the entire student body. In addition, students who leave the university gym following a workout are likely students who exercise on a regular basis and therefore tend to exercise more, on average, than students in general.

Submit

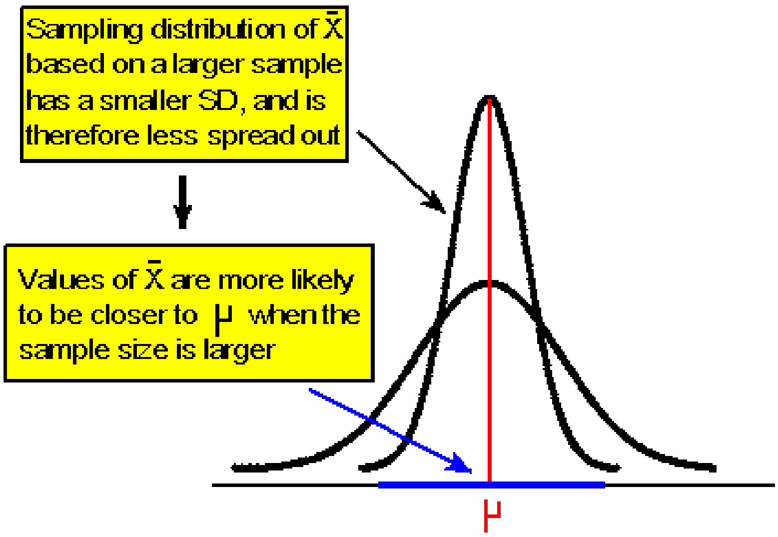
Comment 3

Not only are sample mean and sample proportion on target as long as the samples are random, but their accuracy improves as sample size increases. Again, there are two "layers" here for explaining this.

Intuitively, larger sample sizes give us more information with which to pin down the true nature of the population. We can therefore expect the sample mean and sample proportion obtained from a larger sample to be closer to the population mean and proportion, respectively. In the extreme, when we sample the whole population (which is called a census), the sample mean and sample proportion will exactly coincide with the population mean and population proportion.

There is another layer here that, again, comes from what we learned about the sampling distributions of the sample mean and the sample proportion. Let's use the sample mean for the explanation.

Recall that the sampling distribution of the sample mean $\overline{X}$ is, as we mentioned before, centered at the population mean μ and has a standard deviation of $\frac{\sigma}{\sqrt{n}}$. As a result, as the sample size n increases, the sampling distribution of $\overline{X}$ gets less spread out. This means that values of $\overline{X}$ that are based on a larger sample are more likely to be closer to μ (as the figure below illustrates):



Similarly, since the sampling distribution of $\hat{p}$ is centered at p and has a standard deviation of $\sqrt{\frac{p(1-p)}{n}}$, which decreases as the sample size gets larger, values of $\hat{p}$ are more likely to be closer to p when the sample size is larger.

Did I Get This

1/1 point (graded)
In May 2015, two opinion polls were conducted regarding same-sex marriage. In particular, both polls were conducted in order to estimate *p*, the proportion of all U.S. adults who believe that same-sex marriage should be legal.

• The Gallup poll was based on a random sample of 1,024 U.S. adults and estimated the proportion of all U.S. adults who support gay marriage to be 0.60.

• The Pew Research Center poll was based on a random sample of 2,002 U.S. adults and estimated the proportion of all U.S. adults who support gay marriage to be 0.57.

Which of the two polls estimates *p* (the proportion of all U.S. adults who support gay marriage) more accurately? That is, which of the two point estimates (0.57 or 0.60) is more likely to be closer to the actual value of *p*?

- ○ The Gallup poll
- ◉ The Pew Research Center poll ✔
- ○ Both polls estimate *p* with the same accuracy because both are based on a random sample.

Answer
Correct:
The larger the sample the point estimate is based on, the closer it is likely to be to the parameter it estimates.

Submit

Comment 4

Another example of a point estimate is using sample variance, $s^2 = \frac{(x_1-\overline{x})^2+\ldots+(x_n-\overline{x})^2}{n-1}$, to estimate population variance, σ^2 .

In this course, we will not be concerned with estimating σ^2 for its own sake, but since we will often substitute s for σ when standardizing the sample mean, it is worth pointing out that s^2 is an unbiased estimator for σ^2. If we had divided by n instead of n - 1 in our estimator for population variance, then in the long run our sample variance would be guilty of a slight underestimation. Division by n - 1 accomplishes the goal of making this point estimator unbiased. Making unbiased estimators a top priority is, in fact, the reason that our formula for s, introduced in the Exploratory Data Analysis unit, involves division by n - 1 instead of by n.

Did I Get This

1/1 point (graded)
Based on survey results, the proportion of U.S. adults who use the Internet on a daily basis is 0.37. This point estimate would be unbiased and most accurate if the survey was based on:

- ○ a random sample of 1,000 U.S. adults.
- ◉ a random sample of 2,500 U.S. adults. ✔

- a random sample of 1,000 college students.
- a random sample of 2,500 college students.

Answer
Correct:
The estimate is based on a random sample (and is therefore unbiased) and is also based on a larger sample, which makes it more accurate.

Submit

Did I Get This

1/1 point (graded)
A study estimated that the mean number of children per family in the the United States is 1.3. This point estimate would be unbiased and most accurate if it were based on which of the following?

- A random sample of 10,000 U.S. families with children from the state of Utah
- A random sample of 500 U.S. families with children
- A random sample of 5,000 U.S. families with children
- A random sample of 1,000 U.S. families ✔

Answer
Correct:
The estimate is based on a random sample (and is therefore unbiased) and is also based on a large sample, which makes it more accurate.

Submit

Let's Summarize

We use $\hat{p}$ (sample proportion) as a point estimator for p (population proportion). It is an unbiased estimator: its long-run distribution is centered at p as long as the sample is random.

We use $\overline{x}$ (sample mean) as a point estimator for μ (population mean). It is an unbiased estimator: its long-run distribution is centered at μ as long as the sample is random.

In both cases, the larger the sample size, the more accurate the point estimator is. In other words, the larger the sample size, the more likely it is that the sample mean (proportion) is close to the unknown population mean (proportion).