Lesson 42: Sampling Distributions

Terms:

Parameter: A number that describes an aspect of a population

Statistics: A number that is computed from sample data; often used to estimate an unknown parameter.

Notation	Parameter	Statistic
Proportion	p	$\hat{p}$
Mean	μ	$\bar{x}$

Example: A census of all NWHS seniors found that 10% got into college early. An SRS of 30 seniors was also taken and in that sample 12% got into college early.

The 10% is a _____

The 12% is a _____

Example: A census of all NWHS seniors found that the average weight was 155lbs. An SRS of 30 seniors was also taken and in that sample the average was 148lbs.

The average of155lbs is a _____

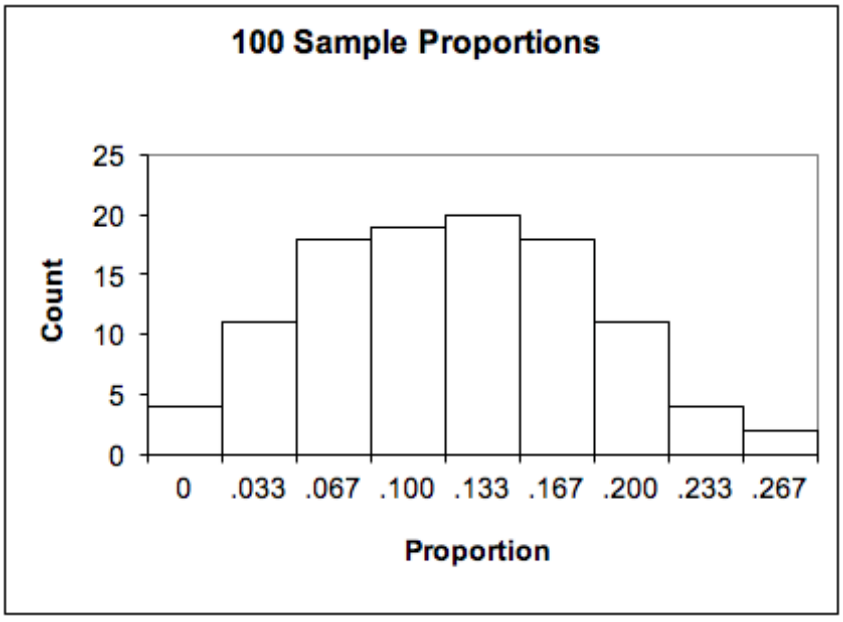
The average of 148lbs is a _____

Sampling distributions and Sampling Variability

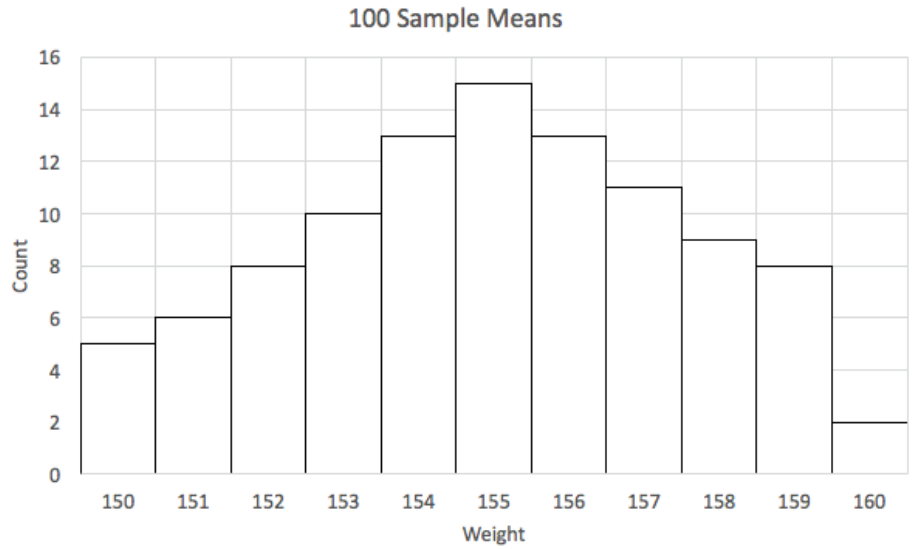
If we take repeated samples from the NWHS senior population and measure the proportion of seniors from those samples that got into college early and their weights, we will see different statistics for the different samples. This is referred to as sample variability.

We can create a distribution of the proportions of all the samples we took and draw a histogram.

The histogram below shows the results of a simulation. 100 samples of size 30 were collected and the percent who got into college early was recorded for each sample. In each sample, the true proportion (parameter) of seniors who got into college early was p = 0.10. The histogram shows that the sample proportion (statistic) will not always match the true proportion. This is because the sample proportion depends on who is in the sample.



The histogram below shows the results of a simulation. 100 samples of size 30 were collected and the average weight was recorded for each sample. In each sample, the true mean (parameter) of seniors was 155. The histogram shows that the sample mean (statistic) will not always match the true mean. This is because the sample mean depends on who is in the sample.



Sampling Distribution:

A Sampling Distribution is shape, center, and spread for the collection of **all** possible samples of the same size from the population

Properties:

- The overall shape of the distribution is symmetric and approximately normal. The larger the sample size the closer the shape is to a normal distribution.
- There are no outliers or other important deviations from the main pattern
- The mean (center) of the distribution is equal to the true population parameter
- The variability (spread) of the sampling distribution depends on the sample size. The larger the sample-size the smaller the variability of the sampling distribution.

Daily Data Collection

Every person in class will draw a card from a standard deck. Then average all values.
When the teacher says "group up" everyone will randomly form groups of 5.
Each group will report the average value on their cards for their group of 5. Repeat until 100 sample means are recorded and graphed below.

Count													
	1	2	3	4	5	6	7	8	9	10	11	12	13

Summary:

Bias:
When a sampling distribution does not have its center equal to the true population parameter (consistently too high or too low), the statistic used to create that sampling distribution is said to be biased.

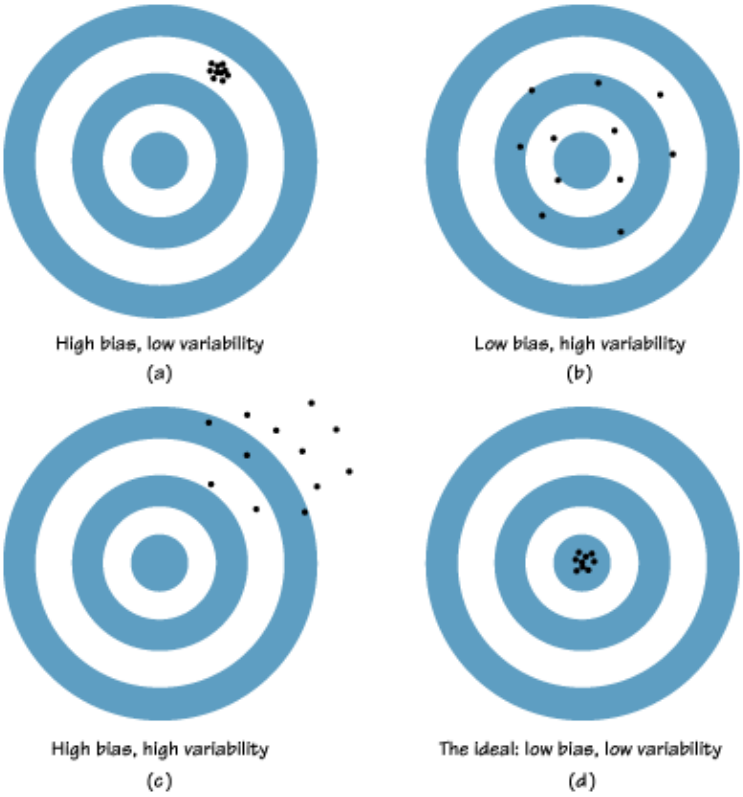
Unbiased:

Proportion $\hat{p}$ is very close to p

Mean $\bar{x}$ is very close to μ

The goal when creating a sampling distribution is to have no bias and low variability. Here is how bias and variability are related →

Note: If the sample is biased, the size of the sample does not matter, so increasing the sample size will NOT improve the results.



- The variability of a sampling distribution is determined by the sampling design and the sample size used to create the sampling distribution. As long as the population is much larger than the sample (at least 10 times as large) The spread of the sampling distribution is the same for any population size.

- Contrary to popular belief and intuition, the behavior of a statistic from random samples is not influenced by the size of the population. To see why, think of taking a sample scoop of m&ms from a well-shuffled 1-pound bag. If the m&ms are well shuffled does the scoop of m&ms really know whether it was surrounded by a one-pound bag of m&ms or a huge bin of m&ms? Clearly it does not.

- The above realization, that variability of a sampling distribution is controlled by the size of the sample, not the size of a population, has major implication for sampling design. It means that a survey of, say, 2000 people is just as accurate if the sample was taken from the population of a small state like Rhode Island as when taken from the population of the entire United States. As long as the sample was an SRS, it can just as easily predict some aspect of the US population as it could from the much smaller Rhode Island population. In other words, the ratio of the sample to the population is NOT important. As a matter of fact, we actually want the ratio of the population to the sample size to be large – more than 10 to 1, in order to be able to conduct most of the statistical analyses we'll be learning about.

HW 42 (Section 7-1): 1, 5, 9, 15, 21-24