

# Sampling Distributions Chapter 18

## Parameter vs Statistic

Example: Identify the population, the parameter, the sample, and the statistic in the given settings.

- a) The Gallup Poll asked a random sample of 515 US adults whether or not they believe in ghosts. Of the respondents, 160 said "Yes".

Try this:

- b) During the winter months, the temperatures outside a cabin in Colorado can stay well below freezing for weeks at a time. To prevent the pipes from freezing, the cabin owner sets the thermostat at 50 degree F. She wants to know how low the temperature actually gets in the cabin. A digital thermometer records the indoor temperature at 20 randomly chosen times during a given day. The minimum reading is 38 degrees F.

## Sampling Variability

How can the sample mean of only a few thousand of the 121 million American households be an accurate estimate of the population mean? After all, a second random sample might produce very different results. The basic fact is called SAMPLING VARIABILITY. The value of a statistic varies in repeated random sampling.

To make sense of sampling variability we can:

- Take a large number of samples from the same population
- Calculate the statistic for each sample
- Make a graph
- Examine the distribution

An applet to do so...<http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1>

In this example, we took 100 samples of size 25 from the population. There are many many possible simple random samples of size 25 from this population. If we looked at all of the sample s of this size and calculated the mean proportion for each we would have the SAMPLING DISTRIBUTION.

### AP EXAM TIP:

The population distribution and the distribution of sample data describe individuals. A sampling distribution describes how a statistic varies in many samples from the population. Be careful with wording.

## Bias and unbiased estimator

A statistic used to estimate a parameter is an UNBIASED ESTIMATOR if the mean of its sampling distribution is equal to the value of the parameter being estimated.

### ACTIVITY: SAMPLING HEIGHTS

In this activity, we will use a population of quantitative data to estimate whether a given statistic is an unbiased estimator of its corresponding population parameter.

1. Each student should write his or her height in inches on a sticky note and fold the paper.
2. After the notes have been mixed, each student will pick four notes. Calculate the sample mean and range of these values.
3. Write all three values on the board. An example is here:

Height	Sample Mean	Sample Range
62, 75, 68, 73	67	75-62-13

4. We will plot the values of our sample mean and sample range on dotplots below.

5. When everyone has finished, find the population mean and population range.

6. Which statistic appears to be unbiased? Which biased?

From before....so, why do we divide by  $n - 1$  to calculate sample variance and standard deviation? Because variance is a biased estimator (tends to underestimate the population variance) we have to adjust to make it unbiased.

Lower variability is better! Larger samples give smaller spreads and will give better a more trustworthy estimate of the parameter.

What can we say about the shape of the distribution (candy simulation)? What happens if I increase the number of samples?

### Sample Proportion Distribution Model (categorical variable)

We notice that as the sample size gets larger and larger, the mean of the simulated samples  $\hat{p}$  gets closer and closer to the actual population probability of .50 ( $p$ ). It is an unbiased estimator. We saw in the last chapter that for a Binomial model the standard deviation for the number of successes =  $\sqrt{npq}$ . We want the standard deviation for the proportion of successes, so we need to divide that value by the number of trials  $n$ . It simplifies to:  $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}}$ .

This value is called the **STANDARD ERROR**. It is not really an error but it represents the variability you would expect to see from one sample to another.

If the number of samples is large enough will a normal model always be a good model? If these conditions are true, the answer is generally yes:

1. Randomization: The sample should be a SRS of the population.
2. 10% condition: The individual sample size should be no larger than 10% of the population otherwise they would not be considered independent.
3. Success/Failure: The sample size has to be big enough so that  $np$  and  $nq$  are at least 10. We need to have at least 10 successes and 10 failures to have enough data to make a conclusion.

### Sample Mean Distribution Model (quantitative data):

The standard error for this distribution is  $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

**CENTRAL LIMIT THEOREM:** States that the sampling distribution of the sample mean (and proportion) is approximately Normal for large  $n$ , regardless of the distribution of the population, as long as the observations are independent.

Let's do p 428 #1 and 3 together.

Homework #4 p 428 #7, 9, 11