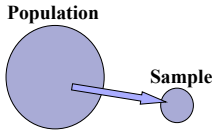
8.1 Sampling distributions

- What we've already covered from 8.1…
  - Distribution of the sample mean $\bar{X}$
- What we're covering now in 8.1…
  - Distribution of the sample proportion $\hat{p}$

1

Estimating the population proportion p using the sample proportion $\hat{p}$

- Recall, we often want to make a statement about the population based on a random sample taken from a population of interest.



2

Population Parameter	Sample Statistic
Population mean μ	Sample mean $\bar{X}$
The mean house value for all houses in Iowa	The mean house value for a sample of n=200 houses in Iowa
Population proportion p	Sample proportion $\hat{p}$
The proportion of all houses in Iowa with lead paint.	The proportion of Iowa houses in a sample of n=200 with lead paint.

↑ Unknown, but estimated from → ↑ Calculated from sample

3

Population proportion p

- The parameter of interest when each subject (or observation) expresses one of two things…
  - In favor of proposition? Yes or no.
  - Alive at age 100? Yes or no.
  - Piece of candy: Orange or Not orange?

4

Population proportion p

- Population
  - The population proportion p of all university of Iowa students who have exactly one sibling.
- Sample of 100 Iowa students
  - The proportion of the sample who have exactly one sibling.

$$\hat{p} = \frac{?}{100}$$

5

Sample-to-sample variability

- All the general concepts we learned about estimating μ using $\bar{x}$ will apply when we estimate p using the sample proportion $\hat{p}$.
- The **sampling error** is the error introduced because a random sample is used to estimate a population parameter… also applies to proportions.

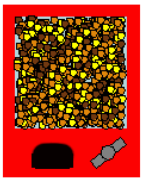
6

What did we learn with $\bar{X}$?

- Different samples will produce different sample proportions $\hat{p}$.
- There is variation from one sample proportion to the next.
- Can we *model* this variation?
  - i.e. Does the $\hat{p}$ statistic vary in a *predictable* way?

7

Simulation



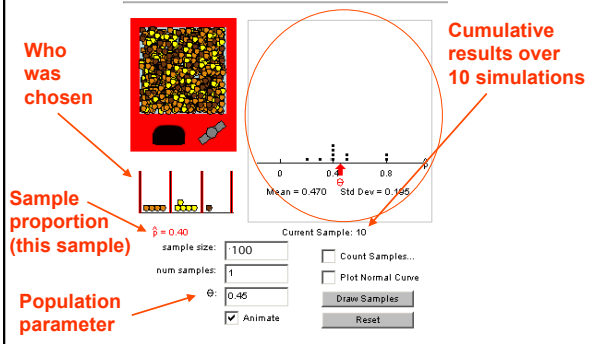
- Population
  - Reeses Pieces
- Population Parameter
  - Proportion of Orange Reeses Pieces

A particular sample of size n = 10.

www.rossmanchance.com/applets/Reeses/ReesesPieces.html

8

Reese's Pieces Samples



9

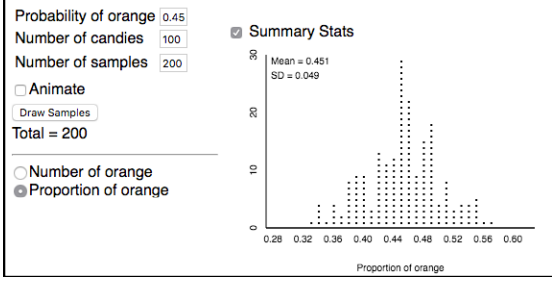
Simulation

- Simple random sample of size n=100.
- Repeat several times.
- Record the sample proportion of orange Reeses Pieces.

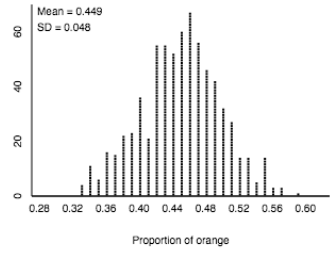
10

Simulation repeated 200 times

This histogram is a summary of the 200 estimates we got from 200 different samples of size 100.



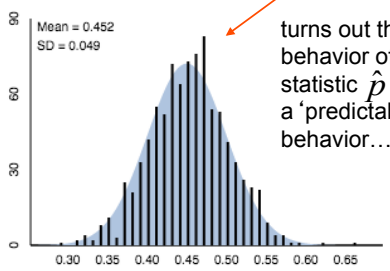
Simulation repeated 500 times



This histogram is a summary of the 500 estimates we got from 500 different samples of size 100.

12

Simulation repeated 900 times

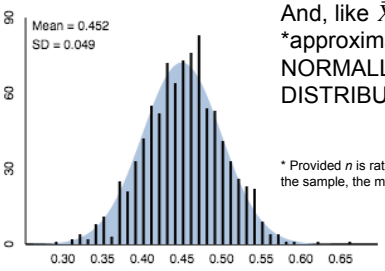


In the long run…

turns out that the behavior of our sample statistic $\hat{p}$ does have a 'predictable' behavior…

13

Simulation repeated 900 times



And, like $\bar{X}$, it's also *approximately NORMALLY DISTRIBUTED!!

* Provided *n* is rather large. And the larger the sample, the more normal it is.

14

Sampling Distribution of $\hat{p}$

- Shape: Approximately Normal
- Center: The mean is p.
- Spread: The standard deviation is

$$\sqrt{\frac{p(1-p)}{n}}$$

15

Exercise 1 (problem 8.1 #18):

- Suppose you know that the distribution of sample proportions of women employees is normal with a mean of p=0.42 and a standard deviation of 0.21. We take a sample and find the sample proportion $\hat{p} = 0.45$.
  - A) How many standard deviations is the sample proportion from the mean of the distribution of all sample proportions?

16

- A)

17

Exercise 1 (problem 8.1 #18):

- Suppose you know that the distribution of sample proportions of women employees is normal with a mean of p=0.42 and a standard deviation of 0.21. We take a sample and find the sample proportion $\hat{p} = 0.45$.
  - B) What is the probability that a second sample selected would have a proportion greater than 0.45?

18

- B)

19

Exercise 2:

- Suppose the proportion of the population in favor of a new law is 0.60. If you take a sample of size n=500, what is the probability that your sample proportion is 0.55 or lower?

20

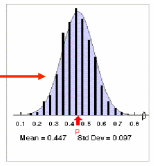
- Answer

21

Note on Sampling Distributions

- Usually, we will only take one sample of a population and compute one statistic (e.g. $\hat{p}$).
  - You get one shot at estimating your parameter p !
- A Sampling Distribution gives us an idea of what we expect to get for our <u>one</u> statistic value.

If we drew an infinite number of random samples and calculated the sample statistic $\hat{p}$ for each, the sampling distribution would look like the resulting histogram of all these $\hat{p}$ values.



22