Week 5 Confidence interval

1. Introduction to confidence interval

Stat 140 - 04

Mount Holyoke College

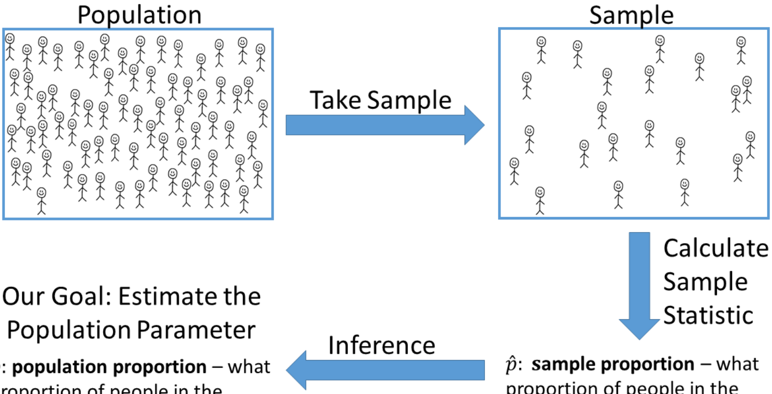
Dr. Shan Shan

Slides posted at *http://sshanshans.github.io/stat140*

1. Questions? (15 minutes)

2. Last week

3. Today: confidence interval



p: **population proportion** – what proportion of people in the **population** support the president?

$\hat{p}$: **sample proportion** – what proportion of people in the **sample** support the president?

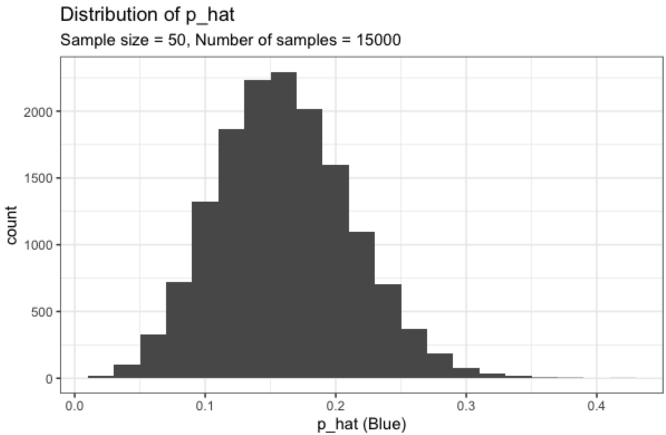
I recently received a gift of 100,000 M&M's. I would like to know what proportion of these M&M's are blue?

Sample once: 10 out of 50 M&M's are blue.

$$\hat{p} = 0.2, \qquad p = ?$$

We use the statistic from a sample as a **point estimate** for a population parameter.

Point estimates will not match population parameters exactly, but they are our best guess, given the data.


Distribution of p_hat
Sample size = 50, Number of samples = 15000
count
2000
1500
1000
500
0
0.0
0.1
0.2
0.3
0.4
p_hat (Blue)

A sampling distribution is the distribution of sample statistics computed for different samples of the same size from the same population.

A sampling distribution shows us how the sample statistic varies from sample to sample

Center

If samples are randomly selected, the sampling distribution will be centered around the population parameter.

Spread

The spread of the distribution measures how much the statistic varies from sample to sample. Also known as **Standard Error**

Shape

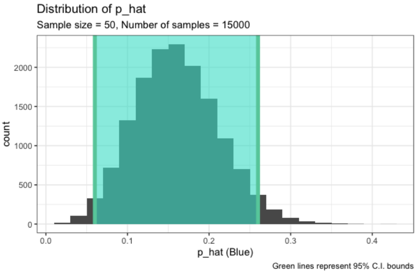
For most of the statistics we consider, if the sample size is large enough the sampling distribution will be symmetric and bell-shaped.

For a sufficiently large sample size, the distribution of sample proportion or sample mean is normal.

For a sufficiently large sample size, the distribution of sample proportion or sample mean is normal.

$$\mathcal{N}\left(\text{Population Parameter}, \text{Standard Error}\right)$$

Green area = middle 95% of the values of sample statistics



Between what two values would you say 95% of sample statistics (proportion) lie?

$$P(? < \hat{p} < ?) = .95$$

1. Questions? (15 minutes)

2. Last week

3. Today: confidence interval

An interval estimate gives a range of plausible values for a population parameter.

Point estimate	Confidence interval
	

One common form for an interval estimate is

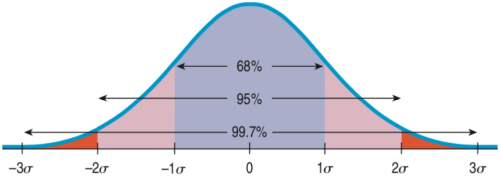
$$\text{Point Estimate} \pm \text{Margin of Error (ME)}$$

where the margin of error reflects the precision of the sample statistic as a point estimate for the parameter.

How would you find the margin of error to ensure that intervals of the form

$$\text{Point Estimate} \pm \text{Margin of Error (ME)}$$

would capture the parameter for 95% of all samples?
(Hint: remember the normal distribution and 95% rule)



Using the central limit theorem, 95% confidence interval can be estimated using

Point Estimate $\pm$ 2 Standard Error (SE)

Poll question

A survey of 1,502 Americans in January 2012 found that 86% consider the economy a "top priority" for the president and congress this year.
The standard error for this statistic is 0.01. What is the 95% confidence interval for the true proportion of all Americans that considered the economy a "top priority" at that time?

1. $(0.85, 0.87)$
2. $(0.84, 0.88)$
3. $(0.82, 0.90)$

In general, a confidence interval is of the form:

$$\text{Point Estimate} \pm \text{Critical value} \times \text{SE}$$

Critical value depends on

- Desired level of confidence (e.g. 90%, 95%, 99%)
- Appropriate sampling distribution (so far we have only learned normal distribution)