# Lecture 10

# Confidence Intervals (for p)

# agenda

| | |
|---|---|
| **Last time on...** | review sampling distributions, the CLT, and bootstrapping |
| **Simulations in R** | see the `bootstrap_ci.Rmd` file located in the Week 7 module |
| **Lecture** | confidence intervals for the population proportion |
| **Back to R** | how to compute confidence intervals in R (see same `bootstrap_ci.Rmd`) |
| **Announcements** | Discuss HW 4 expectations, videos to watch for next week |

# sampling distribution

the distribution of a **statistic** taken from many random samples of the same size from a population

# m&m activity

in this activity each student:

1) had a package of funsize m&m's which contained approximately 55 m&m's,
2) calculated the sample proportion of **blue** m&m's, and
3) plotted the sample proportion on the board.

# What we saw

**shape**

bell-shaped

**center**

$p = 0.18$

**spread**

$se = 0.05$

# Why did we see this?

## The CLT (for proportions)!

Suppose we have a random sample $X_1, X_2, \ldots, X_n$. coming from a Bernoulli distribution with some mean $\mu$ = p and some variance $\sigma^2$ = p(1-p) Then, the sample proportion

$\mu_{\hat{p}} = E[\hat{p}]$

$$\hat{p} \mathrel{\dot\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

$\sigma^2_{\hat{p}} = var[\hat{p}]$

# Back to the m&m activity

we continued with the m&m activity by simulating it online using the Rossman/Chance applets

mean

$$\mathrm{E}[\hat{p}] = \mu_{\hat{p}} = p$$

se

$$\mathrm{SD}[\hat{p}] = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

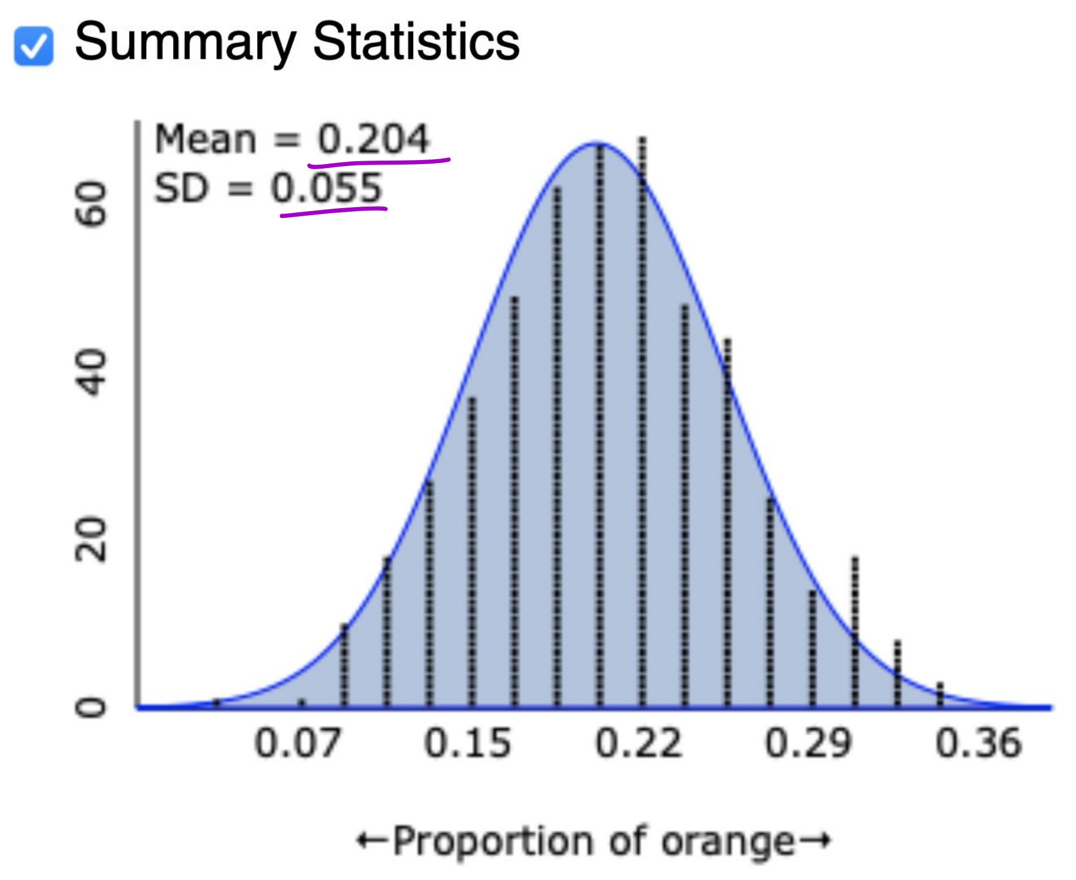
# p = 0.2, n = 55

$$\mu_{\hat{p}} = p = \underline{0.2}$$

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

$$= \sqrt{\frac{0.2 * 0.8}{55}}$$

$$= 0.054$$



# Bootstrapping

a procedure in which a sample is used to create a new distribution, the **bootstrap distribution**, to approximate the sampling distribution

Reference: *Mathematical Statistics with Resampling and R* (available at library.csueastbay.edu)

# Algorithm

## 01

### original sample

we have a sample and want to perform inference on a parameter

## 02

### bootstrap sample

draw B samples of size n with replacement from the original sample

## 03

### bootstrap distribution

calculate the statistic used to estimate the population parameter from each bootstrap sample

# Example in R

go to bootstrap_ci.Rmd

# Confidence intervals

A plausible range of values for a population parameter $(p)$

$\hat{p}$

# Things to note about CI's

## 01
### where
ci's come from the statistic computed from a single sample

## 02
### how
$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

## 03
### why
understand variability in our estimate

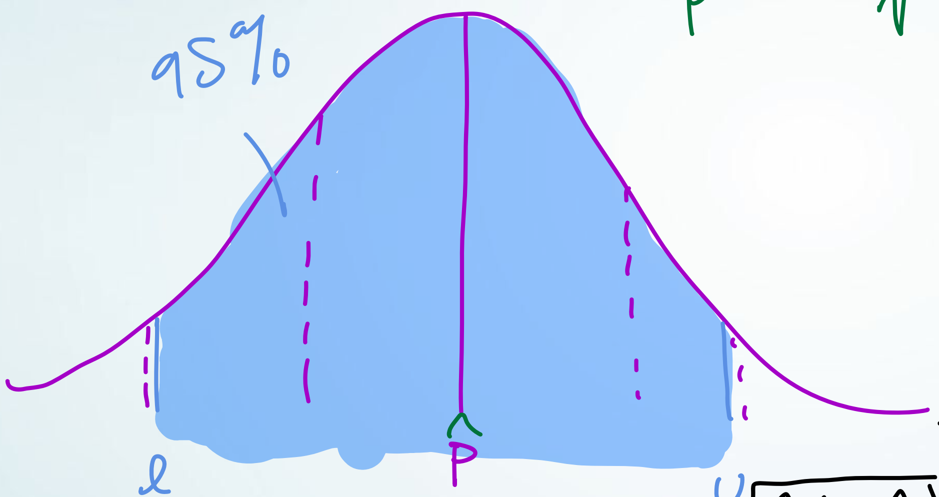
## 04
### random
ci's are **random**! they will vary from sample to sample

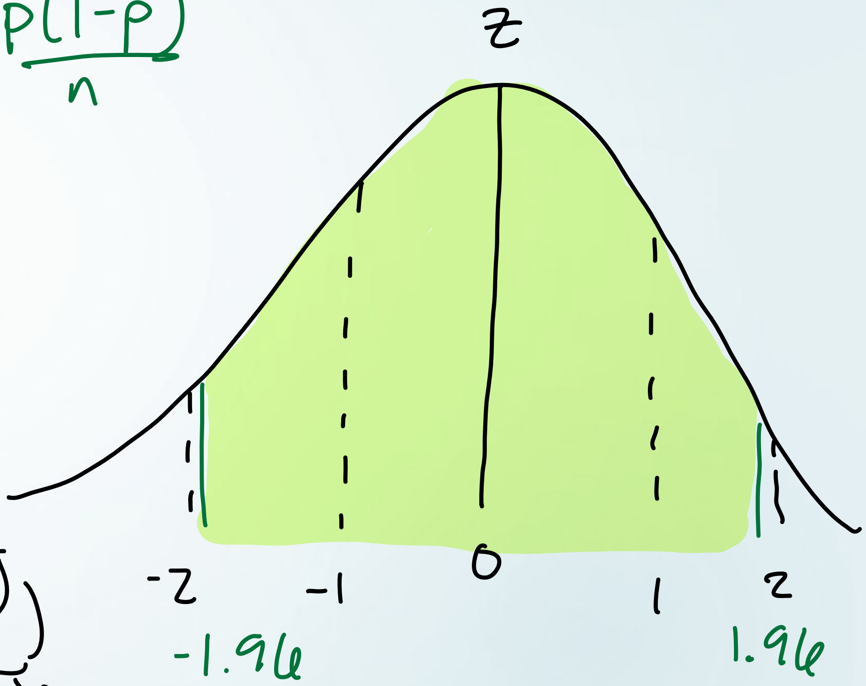
## 05
### interpreting
interpret in terms of the population parameter, not the statistic

## 06
### note
ci's are not probabilities

# Why do we multiply by z*?

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$




$$l = \text{qnorm}\left(0.025, \hat{p}, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$u = \text{qnorm}\left(0.975, \hat{p}, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

# Interpreting a confidence interval

Standard: We are __% confident that the true proportion of _____ is between *l* and *u*.

More complex: If we were to conduct the same study over and over again, __% of the time we <u>expect</u> the true population proportion to fall within *l* and *u*.

# Simulating CI's

First via an applet, then back to bootstrap_ci.Rmd