Statistical Inference

Learning about a population

Download the section 10 .Rmd handout to `STAT240/lecture/10-inference`.

Material in this section is coverd by Chapter 11 on the notes website.

Random variables represent theoretical populations.

In statistics, we study a real population.

- We can't measure every object

We have to take a sample, which adds *uncertainty*.

A **parameter** is a value describing the population.

- What % of citizens support a ballot measure?
- What is the true average price of apples?
- What is the median age of homeowners?

Study the corresponding value from a sample.

We can study a proportion p, or a difference in proportions:

$$p_1 - p_2$$

Example: Loan approval

We can study a single mean μ, or a difference in two means:

$$\mu_1 - \mu_2$$

Example: Sleep hours

Linear regression models two variables:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

β_1 describes the linear relationship.

Example: Test scores

Suppose 53% of 100 Madisonians surveyed support a ballot measure.

0.53 is $\hat{p}$, an estimate for the true rate of support p. This is a realization of a RV.

What are all of the possible values of $\hat{p}$?

Each parameter has a corresponding statistic:

- Sample proportion $\hat{p}$ estimates p
- Sample mean $\bar{X}$ estimates μ
- Observed slope $\hat{\beta}_1$ estimate β_1

Each statistic has its own **sampling distribution**.

Goal: use observed statistic + theoretical sampling distribution to learn about parameter.

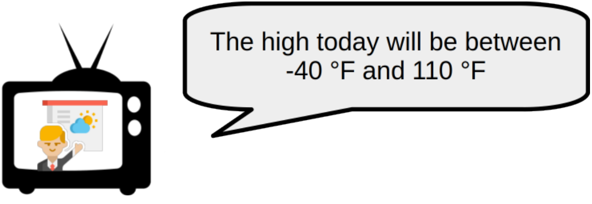
Two ways:

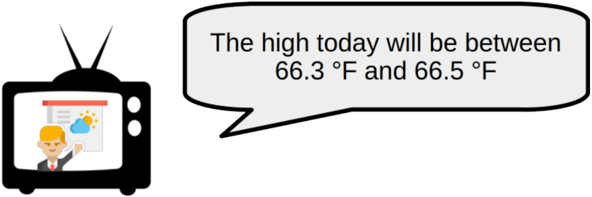
- Confidence intervals
- Hypothesis testing

0.53 is a **point estimate** for p.

It can be useful to instead find an **interval estimate** that represents a range of “guesses”.

How big of an interval do we want?


The high today will be between
-40 °F and 110 °F


The high today will be between
66.3 °F and 66.5 °F

Let's center the interval at $\hat{p} = 0.53$.

$$(0.53 \ - \ \text{Margin}, \ 0.53 \ + \ \text{Margin})$$

We need to determine a good margin.

What if we used a margin of 0.2? We guess that the true proportion of supporters is

$$0.53 \pm 0.2 \;=\; (0.33,\; 0.73)$$

More generally:

$$\hat{p} \pm 0.2 \;=\; (\hat{p} - 0.2,\; \hat{p} + 0.2)$$

We can find the probability that this "margin 0.2" interval covers p.

$$P(\hat{p} - 0.2 \ < \ p \ < \ \hat{p} + 0.2) \ = ?$$

An arbitrary margin of 0.2 won't always work.

Instead, we *choose* a coverage probability, and solve for the margin.

$$P(\hat{p} - ? \ < \ p \ < \ \hat{p} + ?) \ = \ 1 - \alpha$$

A **confidence interval (CI)** is an interval estimate with a specific coverage probability $1 - \alpha$.

Common choice: 95% CI, $\alpha = 0.05$.

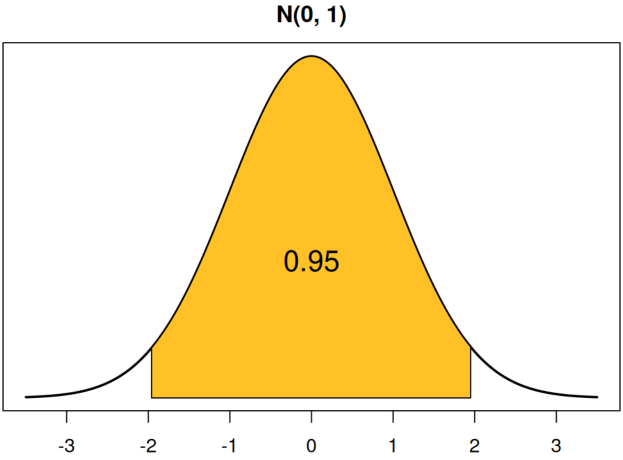
A 95% CI will look like

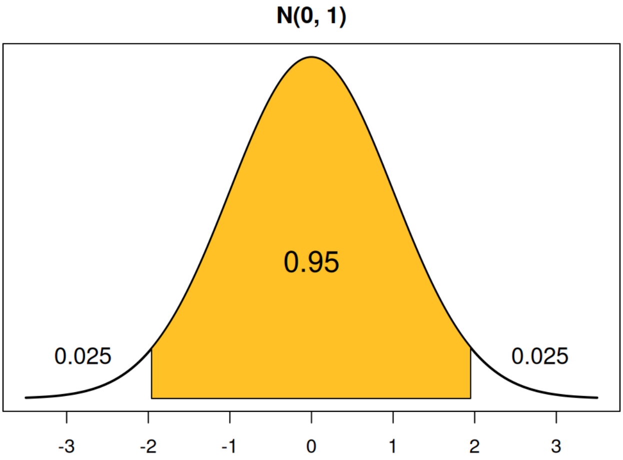
$\hat{p} \pm$ (value related to 0.95) $\times$ (estimation error of $\hat{p}$)

The estimation error of $\hat{p}$ is

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

Use a z-score for a coverage probability of 0.95.





The area in the middle is 0.95, and the area outside is 0.05. Each tail has area 0.025.

The values that correspond to a coverage probability of 0.95 are the 2.5 and 97.5 percentiles of $N(0, 1)$.

These are the **critical values**.

Because of symmetry, we can just use 1.96.

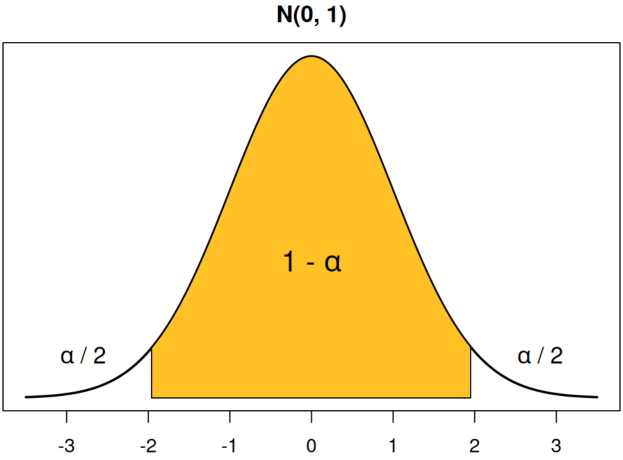
Our 95% CI for p is

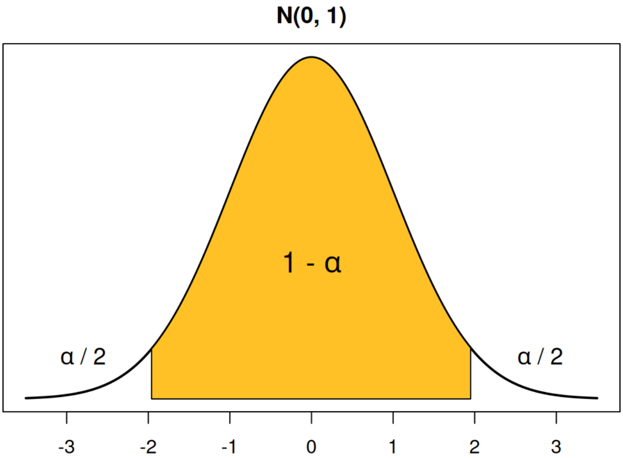
$$0.53 \pm 1.96\sqrt{\frac{0.53(0.47)}{100}}$$

$$(0.432, 0.628)$$

We are 95% confident that the true proportion of supporters is within (0.432, 0.628).

We can generalize this to other α values.





Each tail has area $\alpha/2$. What if we wanted 90% confidence?

General CI formula:

$$\text{point estimate} \;\pm\; \text{critical value} \;\times\; \text{standard error}$$

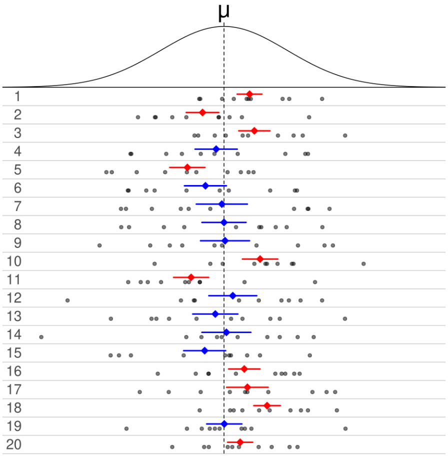
The margin is the **margin of error**.

Once a CI is realized, it either covers *p*, or not.

Across all samples, 95% of the CIs would cover *p*.
The CI procedure has a 0.95 success rate.

We have 95% *confidence* ($100(1-\alpha)\%$ confidence) that an interval covers *p*.

50% CIs, from Wikipedia:



What is a good confidence level?

A standard choice is 95% ($\alpha = 0.05$). Other common choices are 90%, 98%, and 99%.

A smaller α results in a wider interval.

In statistical **testing**, we make a guess about the value of a parameter.

Then, see if the data is consistent with our guess.

In contrast, CIs use the data to form an estimate.

Researcher Muriel Bristol claimed she could tell whether tea had been added before or after milk to her cup, just by tasting it.

Is her ability to identify milk-first versus tea-first cups better than random guessing? ($p = 0.5$).

Background: Lady tasting tea

Let's make some assumptions:

- She gets each cup correct with fixed p
- Each cup is independent
- The number of cups is specified as $n = 8$

Under these assumptions, the number of correct guesses is $X \sim Binom(8,\ p)$.

We formalize our question with **hypotheses**. The **null hypothesis** (H_0) is the “baseline” result.

$$H_0 : p = 0.5$$

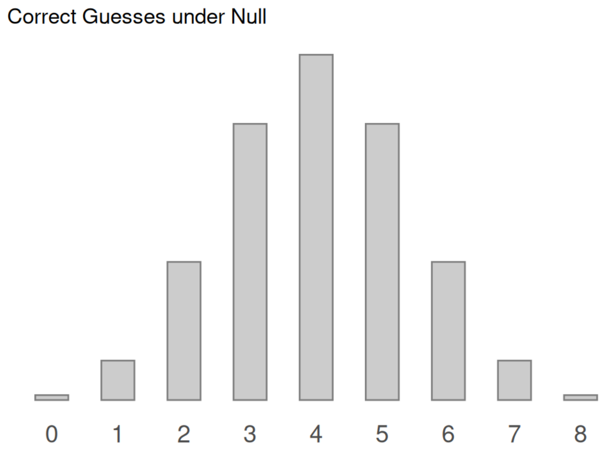
The **alternative hypothesis** (H_A) is the “interesting” result.

$$H_A : p > 0.5$$

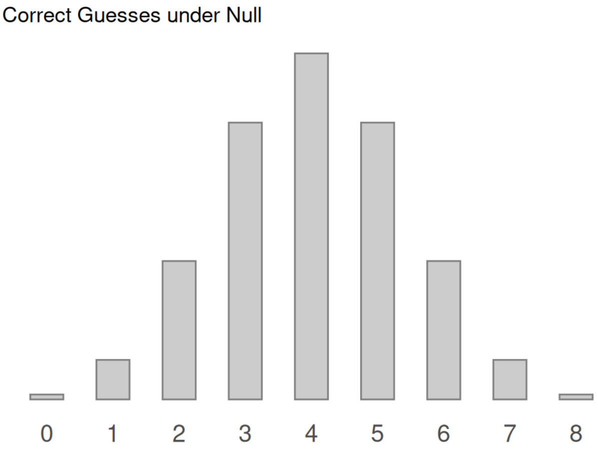
Start by assuming H_0 is true (Bristol is guessing).

We then use our data to collect evidence *against* the null.

Does our data contradict $H_0 : p = 0.5$?



If the null is true, then the number of correct guesses is $Binom(8, 0.5)$.

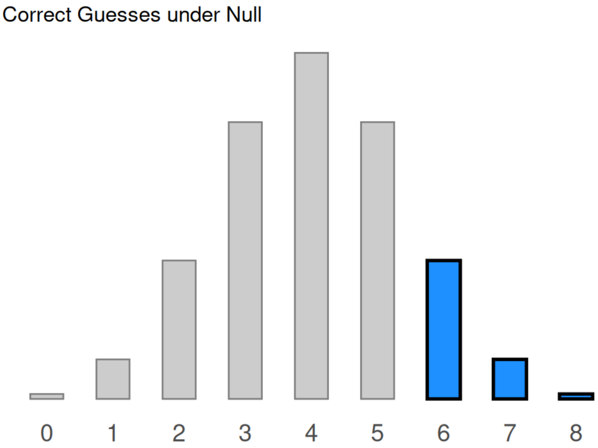


This is the **null distribution**, representing H_0.

Suppose Bristol got 6 guesses right out of 8.

What is the probability that she gets 6 or more right if $p = 0.5$, i.e. by random chance?

This is a probability on our null distribution.

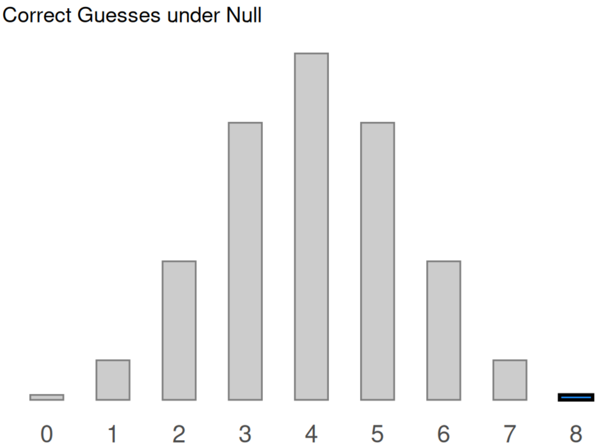


$$P(X \geq 6) = 0.145$$

This probability is a **p-value**. What is the probability of observing our data *or something more extreme*, under H_0?

Smaller p-value $=$ more evidence against H_0.

Our threshold for rejecting H_0 is the significance level α. A common choice is $\alpha = 0.05$.



In reality, Bristol got all 8 cups correct.

Our p-value is $P(X \geq 8) = 0.004$, which is much smaller than $\alpha = 0.05$.

We have strong evidence that Bristol can tell how the tea is made better than random guessing.

We have strong evidence against $H_0 : p = 0.5$.

If the test indicates that the null is likely false, we **reject** H_0.

If we don't have enough evidence against the null, then we **fail to reject** H_0.

It is NOT accurate to "accept" the null hypothesis, since the test does not guarantee that H_0 is true.

$$H_0 : p = 0.5$$

Either H_0 is true, or it is not. Also, we either reject H_0 or fail to reject H_0.

	Fail to reject	Reject
H_0 True		
H_0 False		

Each cell is either an error or the correct decision.

If H_0 is true, but we reject it, we mistakenly found the "interesting" result. This is a false positive.

$$\alpha = P(\text{Reject } H_0, \text{ but } H_0 \text{ true})$$

In stats, we refer to a "Type I error" or "α" error.

If H_0 is false, but we fail to reject it, we mistakenly found the "boring" result. This is a false negative.

$$\beta \;=\; P(\text{Not reject } H_0\text{, but } H_0 \text{ false})$$

In stats, we refer to a "Type II error" or "β" error.

Type I	Type II
α	β
Reject H_0 but H_0 true	Don't reject H_0 but H_0 false
False positive	False negative

There is a tradeoff between α and β - we cannot have them both be small. Choose α that works best.