CSCI 3022-002 Intro to Data Science
Hypotheses

Zach wishes to trash talk the other instructors of his Calculus 2 section when he discovers that his 100 students outperformed the other sections - 300 students total in those sections - by 3 out of 100 points on average, with both groups having a sample standard deviation of 9 points. Can he? Is 0 within the 95% confidence interval for the difference in scores on the tests?

Opening sol:

We are looking for the 95% CI on the difference in the means $\mu_1 - \mu_2$, given : $\bar{X} - \bar{Y} = 3; s_1 = s_2 = 9; n_1 = 100; n_2 = 300$ which is

Opening sol:

We are looking for the 95% CI on the difference in the means $\mu_1 - \mu_2$, given : $\bar{X} - \bar{Y} = 3; s_1 = s_2 = 9; n_1 = 100; n_2 = 300$ which is

Point estimate

distributional part:

C.L.T. says normal

$\mu_1 - \mu_2$:

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

variance: $\bar{X} - \bar{Y}$

: $\text{Var } \bar{X} + \text{Var } \bar{Y}$

$$3 \pm 1.96\sqrt{\frac{9^2}{100} + \frac{9^2}{300}}$$

$$3 \pm 2\sqrt{81/100 + 81/300}$$

Opening sol:

We are looking for the 95% CI on the difference in the means $\mu_1 - \mu_2$, given : $\bar{X} - \bar{Y} = 3; s_1 = s_2 = 9; n_1 = 100; n_2 = 300$ which is

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

$$3 \pm 1.96\sqrt{\frac{9^2}{100} + \frac{9^2}{300}}$$

$$\approx 3 \pm 2 \cdot 9\sqrt{\frac{4}{300}} \approx 3 \pm 3.6/\sqrt{3}$$

$$3 \pm \frac{3.6}{1.7}$$

which doesn't include zero!

interval > 0, so: $\bar{X} - \bar{Y} > 0$

or $\bar{X} > \bar{Y}$

$\sim 3 \pm 2.3$

Opening sol:

We are looking for the 95% CI on the difference in the means $\mu_1 - \mu_2$, given : $\bar{X} - \bar{Y} = 3; s_1 = s_2 = 9; n_1 = 100; n_2 = 300$ which is

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

$$3 \pm 1.96\sqrt{\frac{9^2}{100} + \frac{9^2}{300}}$$

$$\approx 3 \pm 2 \cdot 9\sqrt{\frac{4}{300}} \approx 3 \pm 3.6/\sqrt{3}$$

which doesn't include zero!

Conclusion: Either Zach or Zach's students are the best (??)

Announcements and Reminders

- Exam Tonight!
- NB day Friday.

Right now: Last Second Exam Review! Type your queries into chat!

We did: F 19 MCQs 3-3.

Where we at?

We drew confidence intervals! Random intervals that we really hope cover the true value we're hoping to estimate!

For one large sample, the CI of confidence $100(1-\alpha)\%$ was: for $\underline{\mu}$: population mean

$$\bar{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

(Std. error of $\bar{X}$

Point

distributional

For comparing two samples, we could ask "which mean is larger" by computing a $100(1-\alpha)\%$ CI on the difference in the means $\mu_1 - \mu_2$.

$$\left(\bar{X} - \bar{Y}\right) \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$$

This suggested the possibility of a **decision** rule. More on that, shortly...

The t Distribution

$z_{\alpha/2}$: used if

1) Sample large

2) population was normal for small sample; σ^2 known

Main idea:

With the t-distribution, we're accounting for a second approximation. Not only do ~~we have to~~ approximate

μ (with $\bar{X}$)

We also now have to approximate σ (with s).

t distribution allows us to estimate σ w/ s

The t Distribution

Main idea:
With the t-distribution, we're accounting for a second approximation. Not only do we have to approximate

μ (with $\underline{\bar{X}}$)

We also now have to approximate σ (with $\underline{s}$).

The t Distribution

Main idea:
With the t-distribution, we're accounting for a second approximation. Not only do we have to approximate

μ (with $\underline{\bar{X}}$)

We also now have to approximate σ (with $\underline{s}$).

When our sample size is small, this is often a costly approximation, and as a result we have to *widen* our confidence intervals.

The cost of this approximation scales with n, so as n is smaller, we need to widen our intervals even more.

The t Distribution

Main idea:
With the t-distribution, we're accounting for a second approximation. Not only do we have to approximate

μ (with $\underline{\bar{X}}$)

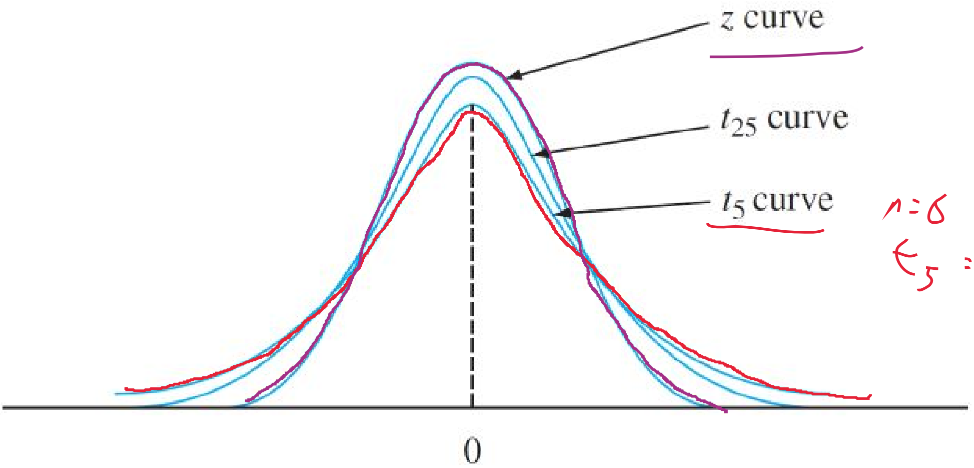
We also now have to approximate σ (with $\underline{s}$).

When our sample size is small, this is often a costly approximation, and as a result we have to *widen* our confidence intervals.

The cost of this approximation scales with n, so as n is smaller, we need to widen our intervals even more.

Intuition: Should t_α be greater or less than z_α?

The t



$n = 6 \Rightarrow$
$t_5 = t_{6-1}$

Properties of the t

n-1

Let t_ν denote the t distribution with ν df. degree of freedom

1. Each t_ν curve is bell-shaped and centered at 0.
   (like $N(0,1)$)
2. Each t_ν curve is more spread out than the standard normal (z) curve.
3. As ν increases, the spread of the corresponding t_ν curve decreases.
4. As ν $\underline{\to \infty}$ the sequence of t_ν curves approaches the standard normal curve (so the z curve is the t curve with df = $\underline{\infty}$)

Properties of the t

Let t_ν denote the t distribution with ν df.

1. Each t_ν curve is bell-shaped and centered at 0.

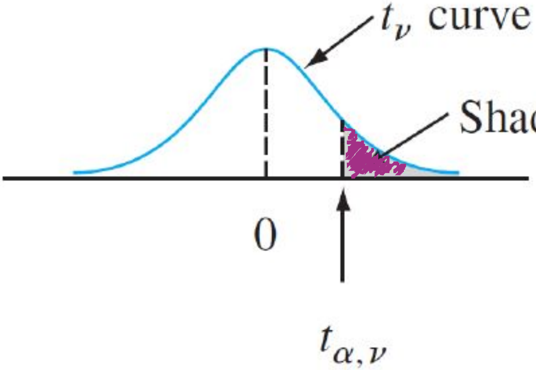
2. Each t_ν curve is more spread out than the standard normal (z) curve.

3. As ν increases, the spread of the corresponding t_ν curve decreases.

4. As ν $\underline{\to \infty\quad}$ the sequence of t_ν curves approaches the standard normal curve (so the z curve is the t curve with df = $\underline{\infty\quad}$)

The t critical values; width multiplier for C.I.

Let $t_{\alpha,\nu}$ = the number on the measurement axis for which the area under the t curve with ν df to the right of t_ν is α;
$t_{\alpha,\nu}$ is called a t critical value.



For example, $t_{.05,6}$ is the t critical value that captures an upper-tail area of .05 under the t

Finding t-values:

The probabilities of t curves are found in a similar way as the normal curve.

Example: obtain $t_{.05,15}$

Finding t-values:

The probabilities of t curves are found in a similar way as the normal curve.

(90% CIs)

$t_{.05,15}$ similar to $z_{.05}$

stats.norm.ppf(.95)

like in cdf.

Example: obtain $t_{.05,15}$

```
stats.t.ppf(.95,15)
```

The t Confidence Interval

Let ____$\bar{X}$____ and ____s____ be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a $100(1-\alpha)\%$ t-confidence interval for the mean μ is

$$\left[\bar{X} - t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}}\right]$$

$t_{\alpha/2,\, n-1}$

or, more compactly:

$$\bar{X} \pm \underset{n-1}{\boxed{t_{\alpha/2,}}}\frac{s}{\sqrt{n}}$$

The t Confidence Interval

Example: Example: Suppose that the GPA measurements for 23 students follow a normal distribution. The sample mean is 3.146. The sample standard deviation is 0.308. Calculate a 90% CI for the mean GPA.

$\overline{X} = 3.146 \quad s = .308 \quad \alpha = .10 \quad n = 23$

$$\overline{X} \pm t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}}$$

$$3.146 \pm t_{.05,\, 22} \cdot \frac{.308}{\sqrt{23}}$$

We'll discuss the two-sample t later.

The t Confidence Interval

Example: Example: Suppose that the GPA measurements for 23 students follow a normal distribution. The sample mean is 3.146. The sample standard deviation is 0.308. Calculate a 90% CI for the mean GPA.

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

We'll discuss the two-sample t later.

The t Confidence Interval

Example: Example: Suppose that the GPA measurements for 23 students follow a normal distribution. The sample mean is 3.146. The sample standard deviation is 0.308. Calculate a 90% CI for the mean GPA.

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$3.146 \pm 1.7171 \cdot \frac{.308}{\sqrt{23}}$$

since `stats.t.ppf(.95,22)`$= t_{.05} = 1.7171$ (compare to $z_{.05} = 1.644!$)

We'll discuss the two-sample t later.

Now what?

$t_{\alpha/2, n-1}$

Decomposing an interval like the interval from our two-sample proportion test

$$\hat{p_1} - \hat{p_2} \pm z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

into a yes or no decision is how we transition into statistical hypothesis testing. Based on our confidence interval on $p_1 - p_2$, we can try to answer whether $p_1 = p_2$, $p_1 < p_2$, etc.

Statistical Hypotheses

Definition: *Statistical Hypothesis*
A *Statistical Hypothesis* is a claim about the value of a parameter or population characteristic.

value = p

Population

data

Examples: "at least $\boxed{30\%}$ of Z's lectures are not boring" $\hat{p}$

"the average midterm score will be > 70" $\bar{X} = 83$

value $\mu = 70$

Statistical Hypotheses

Definition: *Statistical Hypothesis*
A *Statistical Hypothesis* is a claim about the value of a parameter or population characteristic.

Examples:

1. Company A makes parts that last longer than company B.

Statistical Hypotheses

Definition: *Statistical Hypothesis*
A *Statistical Hypothesis* is a claim about the value of a parameter or population characteristic.

Examples:

1. Company A makes parts that last longer than company B.
2. In Boulder, it's usually a colder maximum daily temperature in February than June.

Statistical Hypotheses

Definition: *Statistical Hypothesis*
A *Statistical Hypothesis* is a claim about the value of a parameter or population characteristic.

Examples:

1. Company A makes parts that last longer than company B.
2. In Boulder, it's usually a colder maximum daily temperature in February than June.
3. Students in Zach's sections are generally much more dashing, resourceful, and socially meritorious than students in other sections.

Statistical Hypotheses

One example statisticians often revisit: is a coin fair?
This is a real world question!

`https://www.newscientist.com/article/dn1748-euro-coin-accused-of-unfair-flipping/`

As the Euro was introduced, Polish Mathematicians claimed that the Belgian 1 Euro coin was weighted so that it was more likely ot return a heads!
Suppose I handed you such a coin. How would you decide whether it was fair?

Logic of Hypothesis Testing

Analogy: Jury in a criminal trial.

When a defendant is accused of a crime, the jury (is supposed to) presumes that she is not guilty (not guilty; that's the "null hypothesis").

Then, we gather evidence. If the evidence is seems implausible under the assumption of non-guilt, we might reject non-guilt and claim that the defendant is (likely) guilty.

Logic of Hypothesis Testing

Important Question: Is there strong evidence for the alternative?

The burden of proof is placed on those who believe in the alternative claim.

: Hypothesis

The initially favored claim, the null hypothesis H_0, will not be rejected in favor of the alternative hypothesis, H_a or H_1, unless the sample evidence provides a lot of support for the alternative.

The two possible conclusions:

H_0 null: 0; status quo: boring / baseline / everything is equal

H_1 alternative: one is better, coin is unfair, inequality

Logic of Hypothesis Testing

Important Question: Is there strong evidence for the alternative?

The burden of proof is placed on those who believe in the alternative claim.

The initially favored claim, the null hypothesis H_0, will not be rejected in favor of the alternative hypothesis, H_a or H_1, unless the sample evidence provides a lot of support for the alternative.

The two possible conclusions:
Fail to Reject the null hypothesis if there is insufficient statistical evidence to do so.

Logic of Hypothesis Testing

Important Question: Is there strong evidence for the alternative?

The burden of proof is placed on those who believe in the alternative claim.

The initially favored claim, the null hypothesis H_0, will not be rejected in favor of the alternative hypothesis, H_a or H_1, unless the sample evidence provides a lot of support for the alternative.

The two possible conclusions:
Fail to Reject the null hypothesis if there is insufficient statistical evidence to do so.
Reject the null hypothesis in favor of the alternative if there is statistically *significant* cause to do so.

Logic of Hypothesis Testing

Notation and general process:

Logic of Hypothesis Testing

Notation and general process:

1. Assume the null hypothesis to be true, and state it: we propose that the parameter of interest θ satisfies $H_0 : \theta = \theta_0$.

often μ

Ex: $H_0: \mu = 70$ for exam?

Logic of Hypothesis Testing

Notation and general process:

1. Assume the null hypothesis to be true, and state it: we propose that the parameter of interest θ satisfies $H_0 : \boxed{\theta = \theta_0.}$ one value
   fair coin

2. State the alternative to be tested: H_a :
   $\theta > \theta_0$ **OR** $\theta < \theta_0$ **OR** $\theta \neq \theta_0$ → intervals!
   "more H" ; "more T" unfair coin

3. Draw a decision based on how improbable or probable the actual data looks if the null hypothesis is true. If the observed data is very unlikely, it might be because our hypothesis was wrong!

Why *assume* the null hypothesis?

Logic of Hypothesis Testing

Notation and general process:

1. Assume the null hypothesis to be true, and state it: we propose that the parameter of interest θ satisfies $H_0 : \theta = \theta_0$.

Why *assume* the null hypothesis?

1. Burden of proof
2. We know how to calculate probabilities when we *know* θ!

· know λ, do $\exp(\lambda)$
· know μ, σ^2, do $N(\mu, \sigma^2)$

Logic of Hypothesis Testing

The alternative to the null hypothesis $H_0 : \theta = \theta_0$ will look like one of the following three assertions:

The equality sign is **always** with the null hypothesis.
The alternate hypothesis is the claim for which we are seeking statistical evidence.

Logic of Hypothesis Testing

The alternative to the null hypothesis $H_0 : \theta = \theta_0$ will look like one of the following three assertions:

1. $H_a : \quad \theta \neq \theta_0$

The equality sign is **always** with the null hypothesis.
The alternate hypothesis is the claim for which we are seeking statistical evidence.

Logic of Hypothesis Testing

The alternative to the null hypothesis $H_0 : \theta = \theta_0$ will look like one of the following three assertions:

1. $H_a : \quad \theta \neq \theta_0$

   "unfair" unequal

2. $H_a : \quad \theta > \theta_0$

   greater / less test: is $A > B$

3. $H_a : \quad \theta < \theta_0$

The equality sign is **always** with the null hypothesis.
The alternate hypothesis is the claim for which we are seeking statistical evidence.

Logic of Hypothesis Testing

Example: Suppose a company is considering putting a new type of coating on bearings that it produces.

$H_0: \mu = 1000$

current $\mu = 1000$. $\quad \underbrace{\mu}_{\text{new}} = \underbrace{\mu_0}_{\text{old}}$

The true average wear life with the current coating is known to be 1000 hours. With denoting the true average life for the new coating, the company would not want to make any (costly) changes unless evidence strongly suggested that exceeds 1000.

$H_a: \underbrace{\mu}_{\text{new}} > 1000 \qquad \mu > \mu_0$

Logic of Hypothesis Testing

Example: An appropriate problem formulation would involve testing:

H_0:

H_a:

The conclusion that a change is justified is identified with H_a, and it would take conclusive evidence to justify rejecting H_0 and switching to the new coating.

Scientific research often involves trying to decide whether a current theory should be replaced, or "elaborated upon."

Logic of Hypothesis Testing

Example: An appropriate problem formulation would involve testing:

H_0: New company lifetime average is 1000

$\mu =$ μ_0

H_a: New company lifetime exceeds 1000

μ $>$ μ_0

The conclusion that a change is justified is identified with H_a, and it would take conclusive evidence to justify rejecting H_0 and switching to the new coating.

Scientific research often involves trying to decide whether a current theory should be replaced, or "elaborated upon."

Assume H_0 is true: gives us a prob. distribution

Test Statistics: The Evidence

Definition: *Test Statistic*
A test statistic is a a quantity derived based on sample data and calculated under the null hypothesis. It is used in a decision about whether to reject H_0.

We can think of a test statistic as our evidence. Next, we need to quantify whether we think our evidence is “rare” under the null hypothesis.

Test Statistics: The Evidence

Back to our Belgian Euro: how would you decide whether it was fair?

Test Statistics: The Evidence

Back to our Belgian Euro: how would you decide whether it was fair?

1. State hypothesis: H_0 : fair coin, or $p = .5$.
   H_a : unfair coin, or $p \neq .5$

Test Statistics: The Evidence

Back to our Belgian Euro: how would you decide whether it was fair?

1. State hypothesis: H_0 : fair coin, or $p = .5$.
   H_a : unfair coin, or $p \neq .5$
2. Get to flippin', collect some data

Test Statistics: The Evidence

Back to our Belgian Euro: how would you decide whether it was fair?

1. State hypothesis: H_0 : fair coin, or $p = .5$.
   H_a : unfair coin, or $p \neq .5$
2. Get to flippin', collect some data
3. Compute something from our data. Maybe a sample proportion of heads $\hat{p}$?

Test Statistics: The Evidence

Back to our Belgian Euro: how would you decide whether it was fair?

1. State hypothesis: H_0 : fair coin, or $p = .5$.
   H_a : unfair coin, or $p \neq .5$
2. Get to flippin', collect some data
3. Compute something from our data. Maybe a sample proportion of heads $\hat{p}$?
4. Decide whether $\hat{p}$ is **too far** from $p = .5$, and make a decision accordingly.

Test Statistics: The Evidence

Which test statistic is "best"?

There are an infinite number of possible tests that could be devised, so we have to limit this in some way or total statistical madness will ensue!

In the previous example, we might use $\hat{p}$.

Rejection Regions

How would we know when the test statistic is "sufficiently rare" under the null hypothesis such that we might regard the null as false?

We could define a **rejection region**: a range of values of the test statistic that leads a researcher to reject the null hypothesis.

So when do we reject...?

Suppose we flip our Polish Euro 10 times. How many heads does it take for us to conclude that the coin us unfair?

So when do we reject...?

Suppose we flip our Polish Euro 10 times. How many heads does it take for us to conclude that the coin us unfair?

- What would 10 heads mean?

So when do we reject...?

Suppose we flip our Polish Euro 10 times. How many heads does it take for us to conclude that the coin us unfair?

- What would 10 tails mean?

So when do we reject...?

Suppose we flip our Polish Euro 10 times. How many heads does it take for us to conclude that the coin us unfair?

- What would 6 heads mean?

So when do we reject...?

Suppose we flip our Polish Euro 10 times. How many heads does it take for us to conclude that the coin us unfair?

- Is there a difference between 60% heads if we flip 10 times and 60% heads if we flip 1000 times?

What is extreme: let's compute these!

Bring back α!

Definition: The **Significance level** α of a hypothesis test is the largest *probability* of a test statistic under the null hypothesis that would lead you to reject the null hypothesis.

Equivalently, it's the probability of the entire rejection region!

We thought of α last week during CIs as a term that widened or shrank as our tolerance for error grew, now it's very literally an *error rate*. Specifically, it's the probability of rejecting the null hypothesis when we were not supposed to do so.

Daily Recap

Today we learned

1. Comparing multiple large or normal samples for equivalence of the mean!

Moving forward:

- nb day Friday

Next time in lecture:

- Hypotheses!