THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

# Review of Inference

## Lecture 8

STA 371G

# Updated schedule

- Nothing due this week!
- Quiz 3 next Tuesday
- No Quiz 4
- HW 3 & 4 due due next Monday
- See updated syllabus on Canvas for all details

## About the project

- Build a regression model to answer a research question that is of interest to you, using either the GSS (General Social Survey) or Add Health (National Longitudinal Study of Adolescent to Adult Health) data set

# About the project

- Build a regression model to answer a research question that is of interest to you, using either the GSS (General Social Survey) or Add Health (National Longitudinal Study of Adolescent to Adult Health) data set
- You pick your research question (e.g. predict how much people drink based on their drinking history, personal characteristics, and mental health history)

# About the project

- Build a regression model to answer a research question that is of interest to you, using either the GSS (General Social Survey) or Add Health (National Longitudinal Study of Adolescent to Adult Health) data set
- You pick your research question (e.g. predict how much people drink based on their drinking history, personal characteristics, and mental health history)
- You cannot use the same research question as another group

## About the project

- Build a regression model to answer a research question that is of interest to you, using either the GSS (General Social Survey) or Add Health (National Longitudinal Study of Adolescent to Adult Health) data set
- You pick your research question (e.g. predict how much people drink based on their drinking history, personal characteristics, and mental health history)
- You cannot use the same research question as another group
- You'll be assigned to a group, but can choose to work alone (let me know by EOD tomorrow if you want to work alone)

# About the project

- Build a regression model to answer a research question that is of interest to you, using either the GSS (General Social Survey) or Add Health (National Longitudinal Study of Adolescent to Adult Health) data set
- You pick your research question (e.g. predict how much people drink based on their drinking history, personal characteristics, and mental health history)
- You cannot use the same research question as another group
- You'll be assigned to a group, but can choose to work alone (let me know by EOD tomorrow if you want to work alone)
- The project is divided into multiple components; you'll write a group paper in phases over the remainder of the semester

# Project components

- March 11: Initial proposal (5%)
- March 25: Literature review (10%)
- April 1: Final proposal (5%)
- April 15: Exploratory data analysis (10%)
- April 29: Final paper submission (60%)
- May 3, 5, 18: Presentation & peer evaluations (10%)

## Review of Inference

- Today we'll review key terms around statistical inference: hypothesis tests, confidence intervals, and *p*-values
- We'll also touch on how to run simulations in R (which we unfortunately had to cut from the curriculum due to the UT closure)
- The concepts are the most important; the specific details of the applications here are not critical

1. Minecraft speedrunning controversy
2. Sampling distributions and $t$-tests
3. Using resampling to conduct hypothesis tests

## Minecraft video game speedrunning

- “Speedrunning” in video games is the process of trying to beat video games as fast as possible (e.g., the current world record for Super Mario Bros on Nintendo is under 5 min!)

# Minecraft video game speedrunning

- “Speedrunning” in video games is the process of trying to beat video games as fast as possible (e.g., the current world record for Super Mario Bros on Nintendo is under 5 min!)
- Speedruns are often streamed on YouTube or Twitch and can be big business!

# Minecraft video game speedrunning

- “Speedrunning” in video games is the process of trying to beat video games as fast as possible (e.g., the current world record for Super Mario Bros on Nintendo is under 5 min!)
- Speedruns are often streamed on YouTube or Twitch and can be big business!
- One Minecraft speedrunner, “Dream,” has 18M subscribers and 1.5B views on YouTube, which could result in millions of dollars in ad revenue

## Minecraft video game speedrunning

- “Speedrunning” in video games is the process of trying to beat video games as fast as possible (e.g., the current world record for Super Mario Bros on Nintendo is under 5 min!)
- Speedruns are often streamed on YouTube or Twitch and can be big business!
- One Minecraft speedrunner, “Dream,” has 18M subscribers and 1.5B views on YouTube, which could result in millions of dollars in ad revenue
- But Dream was suspected of cheating—how can you prove it?

## How Minecraft works (sort of)

- To beat the game, you need 12 “eyes of ender”

## How Minecraft works (sort of)

- To beat the game, you need 12 “eyes of ender”
- You can obtain an eye of ender by trading for a pearl, plus some “blaze powder”

# How Minecraft works (sort of)

- To beat the game, you need 12 “eyes of ender”
- You can obtain an eye of ender by trading for a pearl, plus some “blaze powder”
- Every time you attempt a trade, there is a 4.73% chance you will get a pearl

# How Minecraft works (sort of)

- To beat the game, you need 12 “eyes of ender”
- You can obtain an eye of ender by trading for a pearl, plus some “blaze powder”
- Every time you attempt a trade, there is a 4.73% chance you will get a pearl
- As a result, luck plays a big factor in how quickly you can beat the game

## Dream’s luck

- The specific calculations here aren’t important, but let’s get a number anyway

## Dream's luck

- The specific calculations here aren't important, but let's get a number anyway
- Dream attempted 262 trades, and got pearls in 42 of them (16%)—much higher than the 4.73% expected by chance

## Dream's luck

- The specific calculations here aren't important, but let's get a number anyway
- Dream attempted 262 trades, and got pearls in 42 of them (16%)—much higher than the 4.73% expected by chance
- Assuming no cheating, the number of pearls obtained in 262 trials is Binomial $(n = 262, p = 0.0473)$

## Dream’s luck

- The specific calculations here aren’t important, but let’s get a number anyway
- Dream attempted 262 trades, and got pearls in 42 of them (16%)—much higher than the 4.73% expected by chance
- Assuming no cheating, the number of pearls obtained in 262 trials is Binomial ($n = 262$, $p = 0.0473$)
- We can calculate the chance of having luck this good ($P(\text{number of pearls} \geq 42)$) with R:

## Dream's luck

- The specific calculations here aren't important, but let's get a number anyway
- Dream attempted 262 trades, and got pearls in 42 of them (16%)—much higher than the 4.73% expected by chance
- Assuming no cheating, the number of pearls obtained in 262 trials is Binomial ($n = 262$, $p = 0.0473$)
- We can calculate the chance of having luck this good ($P(\text{number of pearls} \geq 42)$) with R:

## Dream's luck

- The specific calculations here aren't important, but let's get a number anyway
- Dream attempted 262 trades, and got pearls in 42 of them (16%)—much higher than the 4.73% expected by chance
- Assuming no cheating, the number of pearls obtained in 262 trials is Binomial ($n = 262$, $p = 0.0473$)
- We can calculate the chance of having luck this good ($P(\text{number of pearls} \geq 42)$) with R:

```r
1 - pbinom(41, 262, 0.0473)

[1] 5.65e-12
```

## Dream’s luck

- In other words, *for someone that is not cheating*, there is about a 1 in 177 billion chance of having luck as good or better than Dream claimed to.

## Dream's luck

- In other words, *for someone that is not cheating*, there is about a 1 in 177 billion chance of having luck as good or better than Dream claimed to.
- This indicates that Dream's claim is implausible (10 runs an hour, 24 hours/day, for 100 years, is just 8.8 million attempts)

## Dream’s luck

- In other words, *for someone that is not cheating*, there is about a 1 in 177 billion chance of having luck as good or better than Dream claimed to.
- This indicates that Dream’s claim is implausible (10 runs an hour, 24 hours/day, for 100 years, is just 8.8 million attempts)
- In other words, we can be confident that Dream was cheating

## Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: Dream is not cheating

## Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: Dream is not cheating
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: Dream is cheating

# Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: Dream is not cheating
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: Dream is cheating
- The $p$-value is the conditional probability of seeing data at least as extreme as what was observed, given that $H_0$ is true: $p = 1$ in 177 billion

# Restating this formally

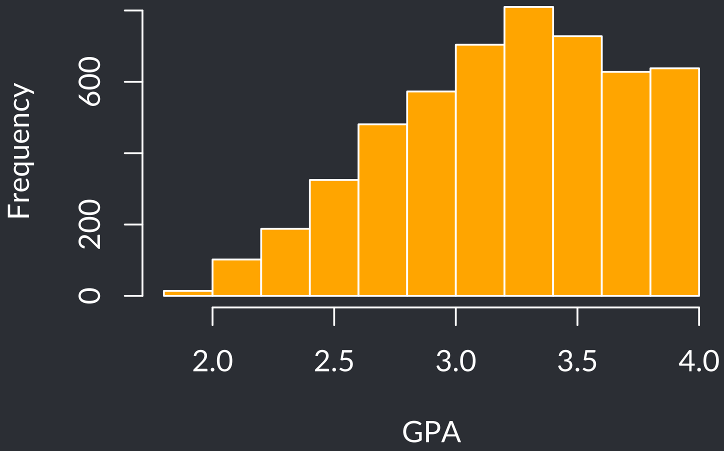
- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: Dream is not cheating
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: Dream is cheating
- The $p$-value is the conditional probability of seeing data at least as extreme as what was observed, given that $H_0$ is true: $p = 1$ in 177 billion
- If $p$ is small, we will reject $H_0$ and believe $H_A$. Otherwise, we continue to believe $H_0$: reject $H_0$

1. Minecraft speedrunning controversy

2. Sampling distributions and *t*-tests

3. Using resampling to conduct hypothesis tests

# GPAs of all UT students entering in Fall 2000



# Let's take a sample

Usually, we only have access to a sample of the data. Let's pretend that we only had a sample of $n = 100$ students:

```
sample.gpas <- sample(ut2000$GPA, 100)
mean(sample.gpas)

[1] 3.2
```

Since we have a random sample, it's a good, but not perfect, estimate of the population GPA (3.212).

## Let’s take a sample

Usually, we only have access to a sample of the data. Let’s pretend that we only had a sample of $n = 100$ students:

```
sample.gpas <- sample(ut2000$GPA, 100)
mean(sample.gpas)

[1] 3.2
```

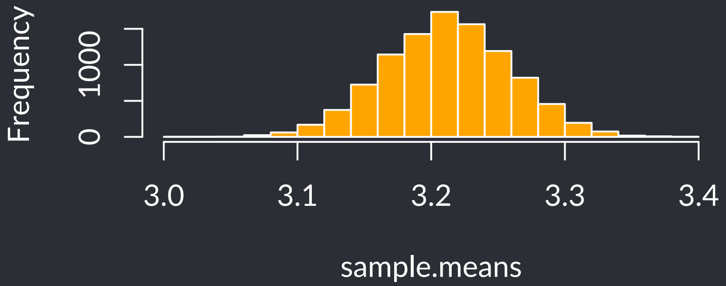
Since we have a random sample, it’s a good, but not perfect, estimate of the population GPA (3.212). But normally we don’t have access to the population, so we don’t know how good our estimate is!

Normally we only have access to one sample.

Normally we only have access to one sample.

Since we have the population, we can simulate what would happen if we repeatedly took samples over and over from the same population.

```r
sample.means <- replicate(10000, {
  the.sample <- sample(ut2000$GPA, 100)
  return(mean(the.sample))
})
hist(sample.means, main="", col="orange")
```



# Sampling distribution of $\overline{GPA}$

The *sampling distribution* of $\overline{GPA}$ is the distribution of sample means, if we took an infinite number of repeated samples:

$$E(\overline{GPA}) = \mu = 3.212$$
$$\mathrm{SD}(\overline{GPA}) = \frac{\sigma}{\sqrt{n}} = \frac{0.48}{\sqrt{100}} = 0.048$$

The last value quantifies how much the sample mean will vary from sample to sample. But we normally can't compute $\sigma$ since we don't have the whole population, so we estimate it by calculating the SD in the *sample* ($\hat{\sigma}$) and dividing by $\sqrt{n}$; this is the *standard error of the mean*.

# Confidence intervals

Given a sample mean, we want to calculate a *confidence interval*, which gives a plausible range of values for the population mean.

# Confidence intervals

Given a sample mean, we want to calculate a *confidence interval*, which gives a plausible range of values for the population mean.

A confidence interval is always of the form

$$\text{sample statistic} \pm (\text{critical value})(\text{standard error}).$$

## Confidence intervals

Given a sample mean, we want to calculate a *confidence interval*, which gives a plausible range of values for the population mean.

A confidence interval is always of the form

$$\text{sample statistic} \pm (\text{critical value})(\text{standard error}).$$

Our sample statistic is $\hat{\mu}$ and our standard error is $\hat{\sigma}/\sqrt{n}$.

# Confidence intervals

Given a sample mean, we want to calculate a *confidence interval*, which gives a plausible range of values for the population mean.

A confidence interval is always of the form

$$\text{sample statistic} \pm (\text{critical value})(\text{standard error}).$$

Our sample statistic is $\hat{\mu}$ and our standard error is $\hat{\sigma}/\sqrt{n}$. What is the critical value?

As it turns out, the sampling distribution (of $\hat{\mu}$) is not *quite* Normal. If we standardize the sample means, the distribution of

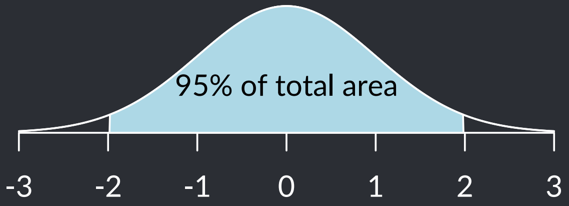
$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}$$

is called a $t$-distribution with $n - 1$ degrees of freedom.

As it turns out, the sampling distribution (of $\hat{\mu}$) is not *quite* Normal. If we standardize the sample means, the distribution of

$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}$$

is called a *t*-distribution with $n-1$ degrees of freedom. The critical value for a 95% confidence interval is $t^* = \pm 1.984$, the value that cuts off 95% of the area under the *t*-distribution:



There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with *df* degrees of freedom.

There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with $df$ degrees of freedom.

There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with $df$ degrees of freedom.
- `qt(y, df)` does the opposite; it figures out what value of $x$ will give $P(t < x) = y$.

There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with *df* degrees of freedom.
- `qt(y, df)` does the opposite; it figures out what value of $x$ will give $P(t < x) = y$.

There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with $df$ degrees of freedom.
- `qt(y, df)` does the opposite; it figures out what value of $x$ will give $P(t < x) = y$.

So the critical value for a 95% confidence interval is `qt(.975, 99)` when $n = 100$.

There are two helpful R functions for calculating values around $t$-distributions (like `pnorm` and `qnorm`):

- `pt(x, df)` calculates $P(t < x)$ if we are looking at a distribution with $df$ degrees of freedom.
- `qt(y, df)` does the opposite; it figures out what value of $x$ will give $P(t < x) = y$.

So the critical value for a 95% confidence interval is `qt(.975, 99)` when $n = 100$.

Using our sample of $n = 100$, we calculate that a 95% confidence interval for the mean population GPA is

$$\hat{\mu} \pm t^* \frac{\hat{\sigma}}{\sqrt{n}}$$

Using our sample of $n = 100$, we calculate that a 95% confidence interval for the mean population GPA is

$$\hat{\mu} \pm t^* \frac{\hat{\sigma}}{\sqrt{n}} \quad \Longrightarrow \quad 3.204 \pm 1.984 \cdot \frac{0.478}{\sqrt{100}}$$

Using our sample of $n = 100$, we calculate that a 95% confidence interval for the mean population GPA is

$$\hat{\mu} \pm t^* \frac{\hat{\sigma}}{\sqrt{n}} \quad \Longrightarrow \quad 3.204 \pm 1.984 \cdot \frac{0.478}{\sqrt{100}} \quad \Longrightarrow \quad (3.109, 3.299)$$

Using our sample of $n = 100$, we calculate that a 95% confidence interval for the mean population GPA is

$$\hat{\mu} \pm t^* \frac{\hat{\sigma}}{\sqrt{n}} \quad \Longrightarrow \quad 3.204 \pm 1.984 \cdot \frac{0.478}{\sqrt{100}} \quad \Longrightarrow \quad (3.109, 3.299)$$

There are two ways to interpret this:

- **Informally,** we are 95% confident that the population mean GPA is between 3.109 and 3.299.

Using our sample of $n = 100$, we calculate that a 95% confidence interval for the mean population GPA is

$$\hat{\mu} \pm t^* \frac{\hat{\sigma}}{\sqrt{n}} \quad \Longrightarrow \quad 3.204 \pm 1.984 \cdot \frac{0.478}{\sqrt{100}} \quad \Longrightarrow \quad (3.109, 3.299)$$

There are two ways to interpret this:

- **Informally,** we are 95% confident that the population mean GPA is between 3.109 and 3.299.
- **Formally,** if we took repeated samples and found the 95% CI within each sample, 95% of the CIs would contain the population mean.

R can do this work for you!

```
t.test(sample.gpas)


One Sample t-test


data:  sample.gpas
t = 67, df = 99, p-value <2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 3.11 3.30
sample estimates:
mean of x
      3.2
```

## Hypothesis tests

- Let's say Mr. Sooner comes along and claims that the average UT GPA is actually 3.15.

# Hypothesis tests

- Let’s say Mr. Sooner comes along and claims that the average UT GPA is actually 3.15.
- Let’s pretend we don’t have the population (which is usually the case), and so we can’t know for sure that he is wrong.

## Hypothesis tests

- Let’s say Mr. Sooner comes along and claims that the average UT GPA is actually 3.15.
- Let’s pretend we don’t have the population (which is usually the case), and so we can’t know for sure that he is wrong.
- We do have some evidence (our sample) that we can bring to bear on the question.

## Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: $\mu = 3.15$

# Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: $\mu = 3.15$
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: $\mu \neq 3.15$

# Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: $\mu = 3.15$
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: $\mu \neq 3.15$
- The *p*-value is the conditional probability of seeing data at least as extreme as what was observed, given that $H_0$ is true: if Sooner is correct, how likely is it that in our sample we would see a sample mean (of $\hat{\mu} = 3.204$) that is so far away from his hypothesized value of $\mu = 3.15$?

R can run hypothesis tests for us:

```
t.test(sample.gpas, mu=3.15)


One Sample t-test

data:  sample.gpas
t = 1, df = 99, p-value = 0.3
alternative hypothesis: true mean is not equal to 3.15
95 percent confidence interval:
 3.11 3.30
sample estimates:
mean of x
      3.2
```

# Restating this formally

- The null hypothesis ($H_0$) is what we assume to be true in the absence of evidence to the contrary: $\mu = 3.15$
- The alternative hypothesis ($H_A$) is what we will believe if it turns out that $H_0$ is false: $\mu \neq 3.15$
- The $p$-value is the conditional probability of seeing data at least as extreme as what was observed, given that $H_0$ is true: $p = 0.3$
- If $p$ is small, we will reject $H_0$ and believe $H_A$. Otherwise, we continue to believe $H_0$: do not reject $H_0$; Mr Sooner's claim is consistent with our sample data

When doing hypothesis testing, we select an $\alpha$ value in advance, and then reject the null hypothesis if and only if $p < \alpha$.

When doing hypothesis testing, we select an $\alpha$ value in advance, and then reject the null hypothesis if and only if $p < \alpha$.

$\alpha = .05$ is a good "default" to use unless you have a reason to set it higher or lower.

1. Minecraft speedrunning controversy

2. Sampling distributions and $t$-tests

3. Using resampling to conduct hypothesis tests

## Comparing meeting strategies

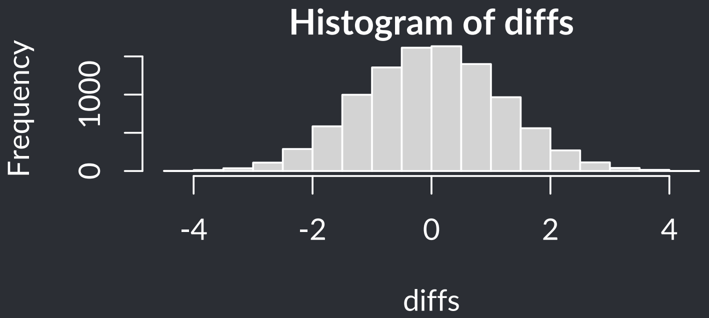
- Renuka holds daily “standup” meetings with her team, under the theory that having the meeting standing up will make it go faster.
- She decides to test this by holding 20 meetings standing up and then 20 meetings sitting down.
- Standing is 3.65 minutes faster, but is that real or just due to random chance?

| Standing | Sitting |
|---|---|
| 27, 29, 30, 30, 31, 31, 31, 32, 33, 33, 33, 34, 34, 35, 35, 35, 36, 37, 37, 40 | 30, 33, 34, 34, 34, 34, 35, 35, 36, 36, 37, 38, 38, 38, 39, 39, 40, 40, 41, 45 |

## Resampling strategy

- Put all 40 meeting times in a blender, and randomly pick out 20 for each group.
- Calculate how often do we get a difference by chance of at least 3.65 minutes?

```r
diffs <- replicate(10000, {
  reordered <- sample(c(standing, sitting), 40)
  group1 <- reordered[1:20]
  group2 <- reordered[21:40]
  return(mean(group1) - mean(group2))
})
hist(diffs)
```



Just by looking at the histogram, we can tell 3.65 is a very unusual difference to see just by chance. So the *p*-value is low and we should reject the null hypothesis.

Just by looking at the histogram, we can tell 3.65 is a very unusual difference to see just by chance. So the *p*-value is low and we should reject the null hypothesis.

To calculate it exactly, we count the number of differences that are larger than $\pm 3.65$:

```
sum(diffs <= -3.65 | diffs >= 3.65) / 10000

[1] 0.0016
```