Mid-semester Review

3/4/2020

Quiz 7 - Difference of Means

Main ideas so far

1. What is sampling? Why can we use samples to reason about populations? How does the process of sampling change our inferences?
2. Descriptives statistics as compression. Deciding what statistic is appropriate when.
3. What is null hypothesis testing? What do the two outcomes mean?
4. The Central Limit theorem: When it holds and what it means
5. The Normal Distribution: Critical values, Z-scores, confidence intervals
6. The t-distribution and t-tests: Proportions, means, paired vs. unpaired
7. ANOVA: When you use it, what the values mean, and follow-up tests

What is sampling?

100 1915-2015 PSYCHOLOGY

Support : 8%
Don't Support: 92%

24 votes cast

Quinnipiac University / Poll

Support : 45%
Don't Support: 41%

1562 votes cast

DEFCON Warning System

Support : 58%
Don't Support: 34%

2510 votes cast

Each of these polls is a **sample**

But I want to make an inference to the **population**

When I draw a conclusion about the population from a sample, I make an **inference**.

The way I collect my sample can lead me to different inferences.

Which of these samples is the best?

What is sampling?

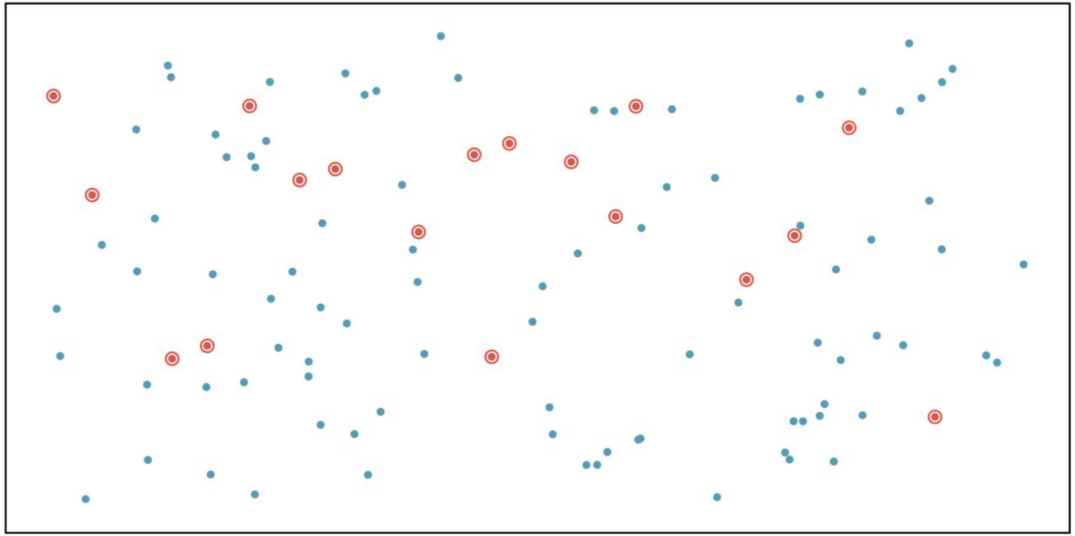
Why is bigger better?

Small samples are more **variable**.

There are 100 dots here, and 18 of them are red.

If I draw 3 dots, **more than half** the time 0 will be red.

If I draw 50 dots,
less than **1 out of 100 billion times** 0 will be red



For random samples, larger samples are more **representative**

Why can we use samples to reason about populations?

When you taste a spoonful of soup and decide the spoonful you tasted isn't salty enough, that's **exploratory analysis**

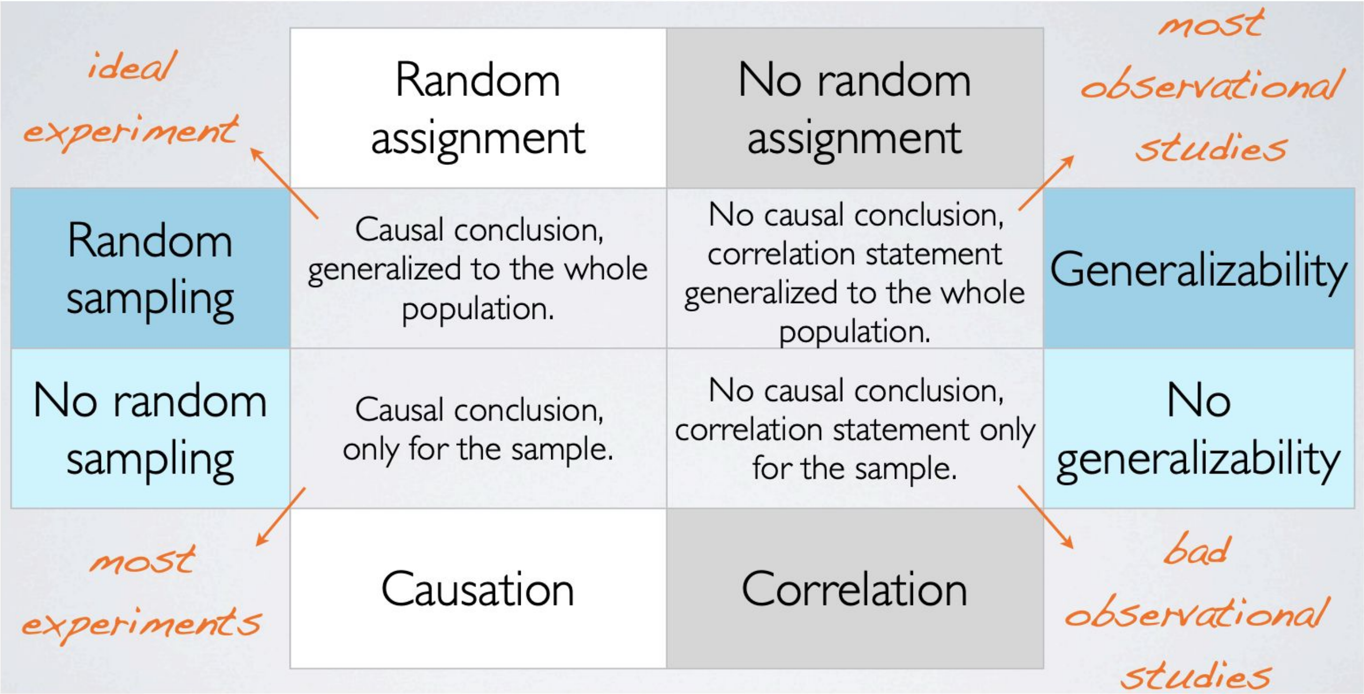
If you generalize and conclude that your entire soup needs salt, that's an **inference**

For your inference to be valid, the spoonful you tasted (the **sample**) needs to be **representative** of the entire pot (the **population**)

If the soup is not well stirred, it doesn't matter how large a spoon you have, it will still not taste right. If the soup is well stirred, a small spoon will suffice to test the soup.

Thanks Mine Çetinkaya-Rundel

How does the process of sampling change our inferences?



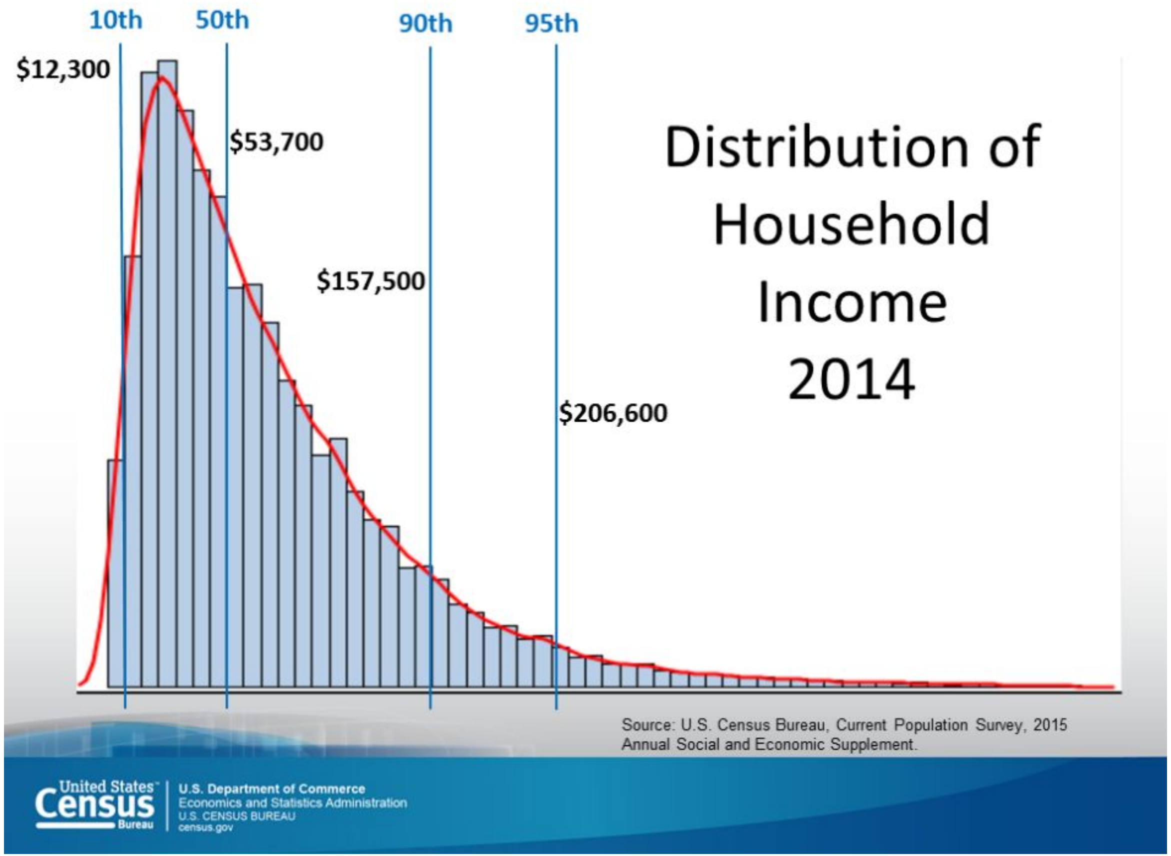
Descriptives statistics as compression

What's the difference between .mp3 and .FLAC?
.jpeg and .png?

.mp3 and .jpeg are **lossy compression** -- they make data smaller by throwing some of it away.

Central tendency is a kind of lossy compression: **What one number is the most representative of my data**?

Deciding what statistic is appropriate when



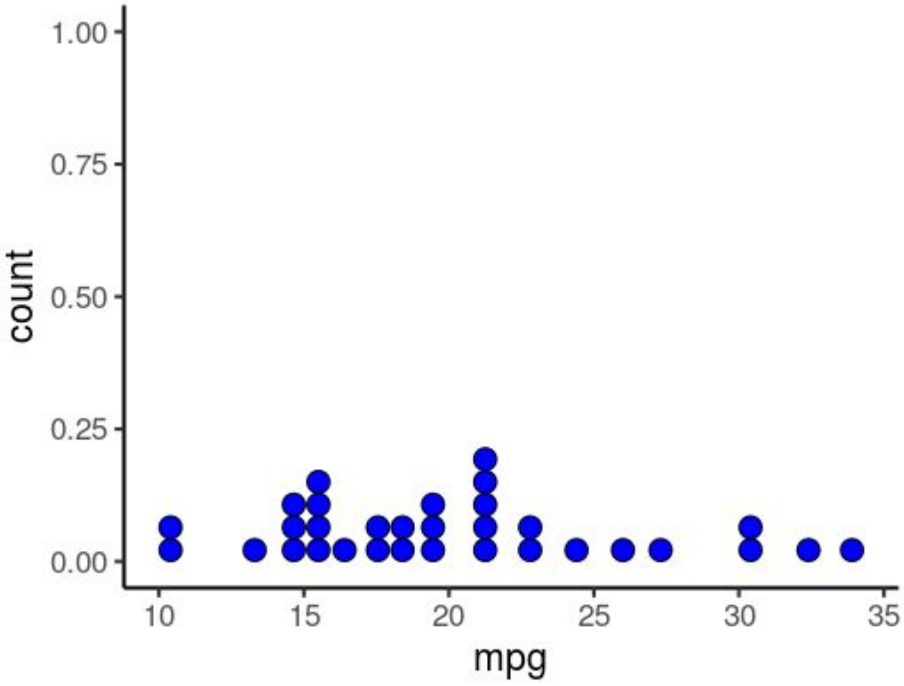
Median: $53,700

Mean: $75,738

Deciding what statistic is appropriate when

A good visualization makes your intuitions when seeing the data match the results of your statistical analyses

Dot plots make it easy to see where most of the data is.



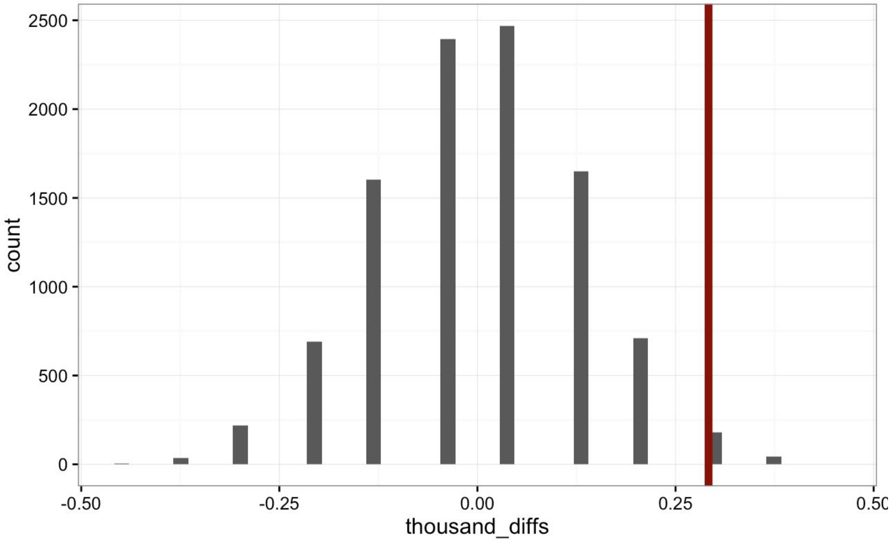

```
mtcars %>%
  ggplot(aes(x = mpg)) +
  geom_dotplot(fill = "blue", color = "black")
```

What is null hypothesis testing?

1. "There is nothing going on" (**Null Hypothesis**)
   The *process* of promotion is independent of gender
   We observed results that *look* dependent due to chance

2. "There is something going on" (**Alternative Hypothesis**)
   The *process* of promotion is dependent of gender
   We observed results that *look* dependent because they *are dependent*

What is null hypothesis testing?



If promotion is independent of gender, we should see a difference like the one we observed l*ess than 1% of the time*.

What do the two outcomes mean?

Inference

Truth

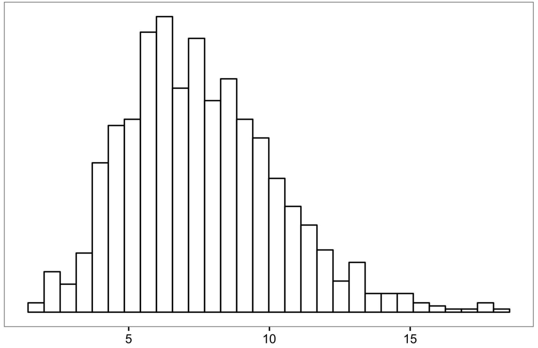
	Do not reject H_0	**Reject H_0 in favor of H_A**
H_0 True	Correct	Type I Error
H_A True	Type II Error	Correct

Increasing our standard of evidence yields fewer Type I Errors, but more Type II Errors.

You can’t avoid this!

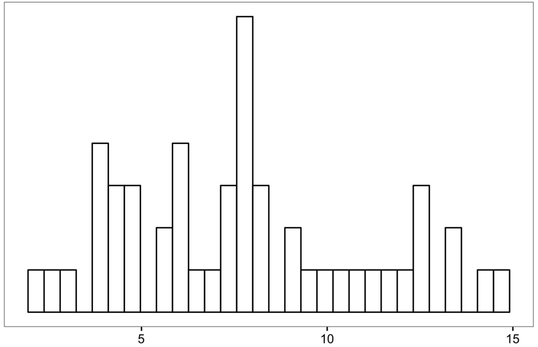
You just have to decide how important each type of error is.

The Central Limit theorem: When it holds and what it means

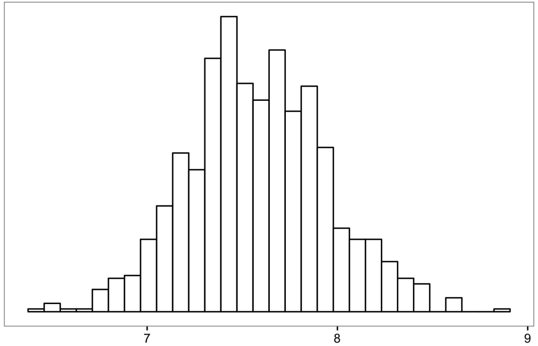


When I draw **independent samples** from the population, as sample size **approaches infinity,** the distribution of means approaches normality

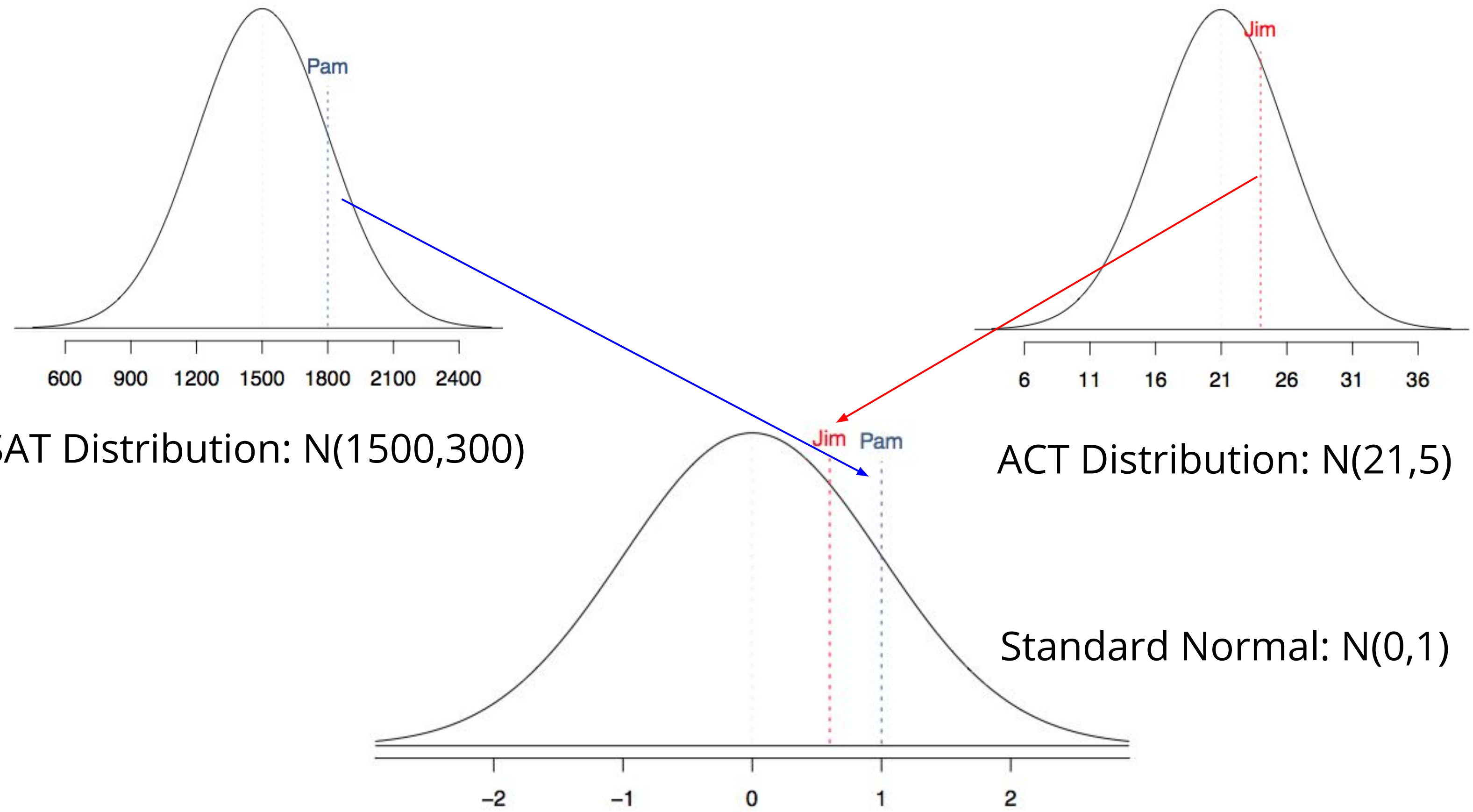
Many statistical methods we use leverage this relationship
(t-test, linear regression, ANOVA, etc)



Take the mean,
Repeat many times...



The Normal Distribution: Z-scores



SAT Distribution: N(1500,300)

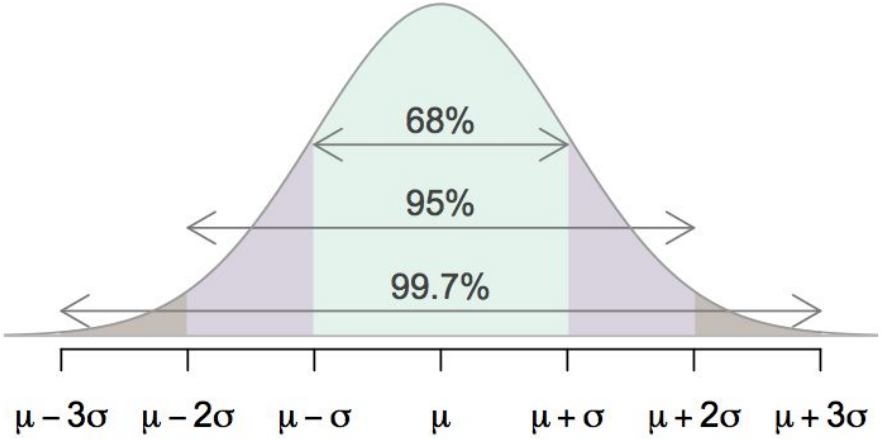
ACT Distribution: N(21,5)

Standard Normal: N(0,1)

The Normal Distribution: Critical values

For nearly normally distributed data,

- about 68% falls within 1 SD of the mean,
- about 95% falls within 2 SD of the mean,
- about 99.7% falls within 3 SD of the mean.



It is possible for observations to fall 4, 5, or more standard deviations away from the mean, but these occurrences are very rare if the data are nearly normal.

The Normal Distribution: Confidence intervals

A plausible range of values for the population parameter is called a *confidence interval*.

Using only a sample statistic to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.

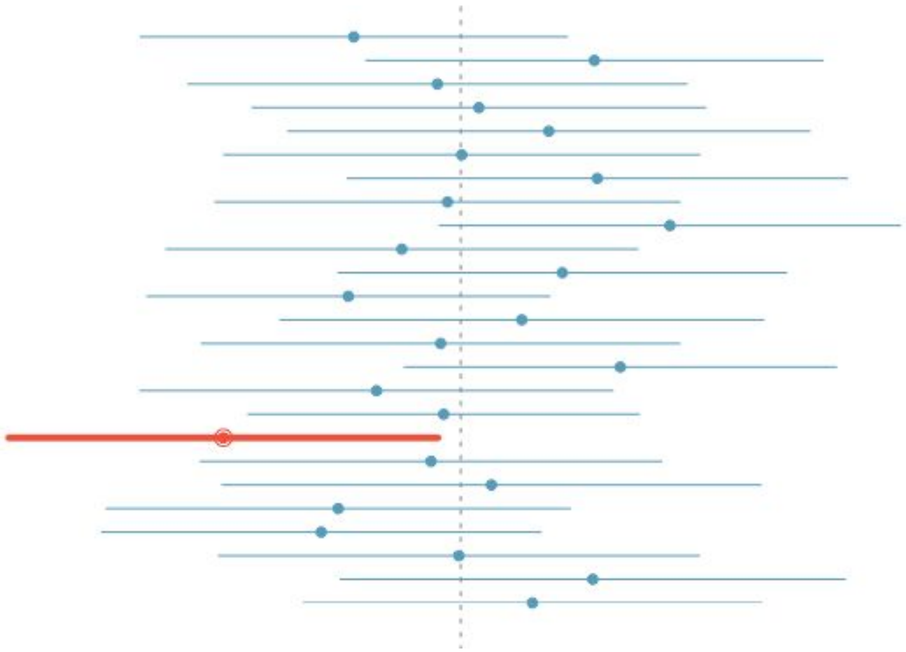
We can throw a spear where we saw a fish, but we'll probably miss. If we toss a net, we have a good chance of catching it.

If we report a point estimate, we probably won't hit the exact population parameter. If we report a range of plausible values we have a good shot at capturing the parameter.

The Normal Distribution: Confidence intervals

Suppose we took many samples and built a confidence interval from each sample using the equation *point estimate ± 2 x SE*.

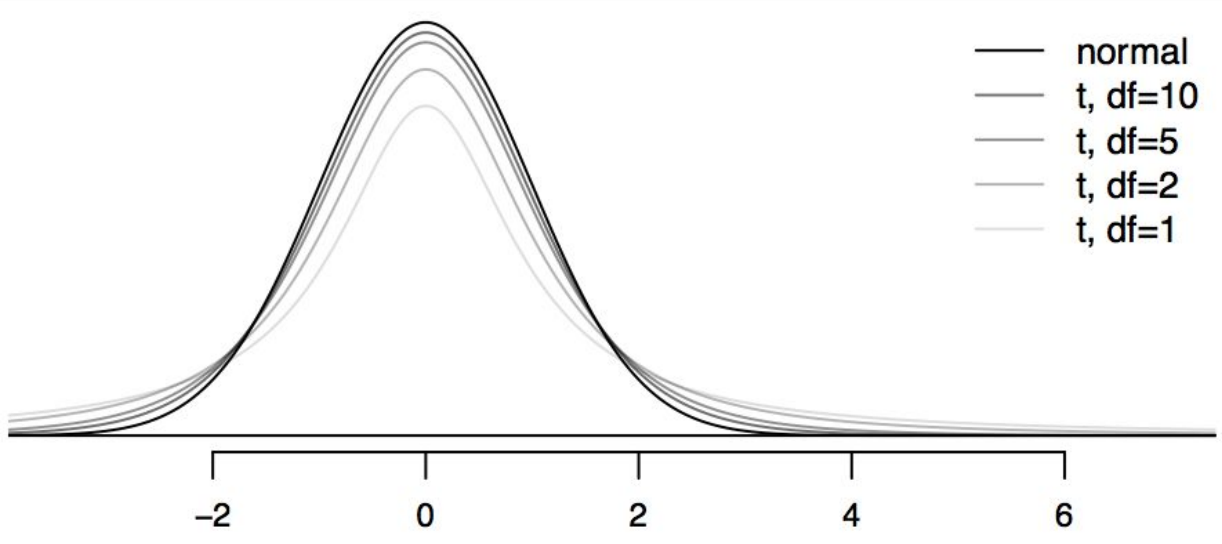
Then about 95% of those intervals would contain the true population mean (*μ*).



The t-distribution and t-tests

Centered at zero like the standard Normal (*z*-distribution).

Has only one parameter: **degrees of freedom (df)**



What happens as df increases? **Approaches the Normal (z)**

The t-distribution and t-tests: paired vs. unpaired

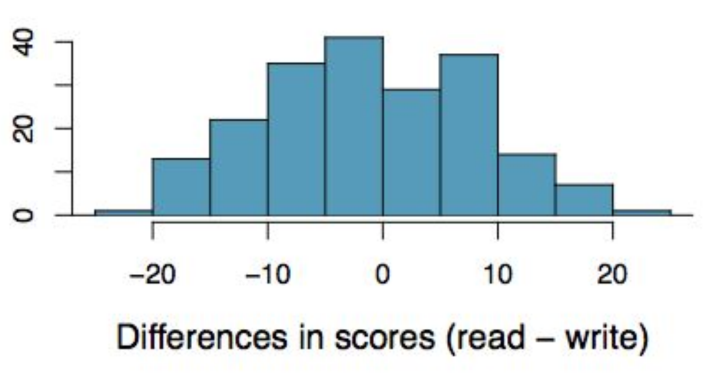
Two sets of data are ***paired*** if each data point in one set depends on a particular point in the other set.

To analyze paired data, we first compute the difference between in outcomes of each pair of observations.

diff = read - write

Note: It’s important that we always subtract using a consistent order.

	id	read	write	diff
1	70	57	52	5
2	86	44	33	11
3	141	63	44	19
4	172	47	52	-5
⋮	⋮	⋮	⋮	⋮
200	137	63	65	-2



The t-distribution and t-tests: paired vs. unpaired

The test statistic for inference on the difference of two small sample means ($n_1 < 30$ and/or $n_2 < 30$) mean is the *T* statistic.

$$T_{df} = \frac{\text{point estimate} - \text{null value}}{SE}$$

$$\text{point estimate} = \bar{x}_1 - \bar{x}_2$$

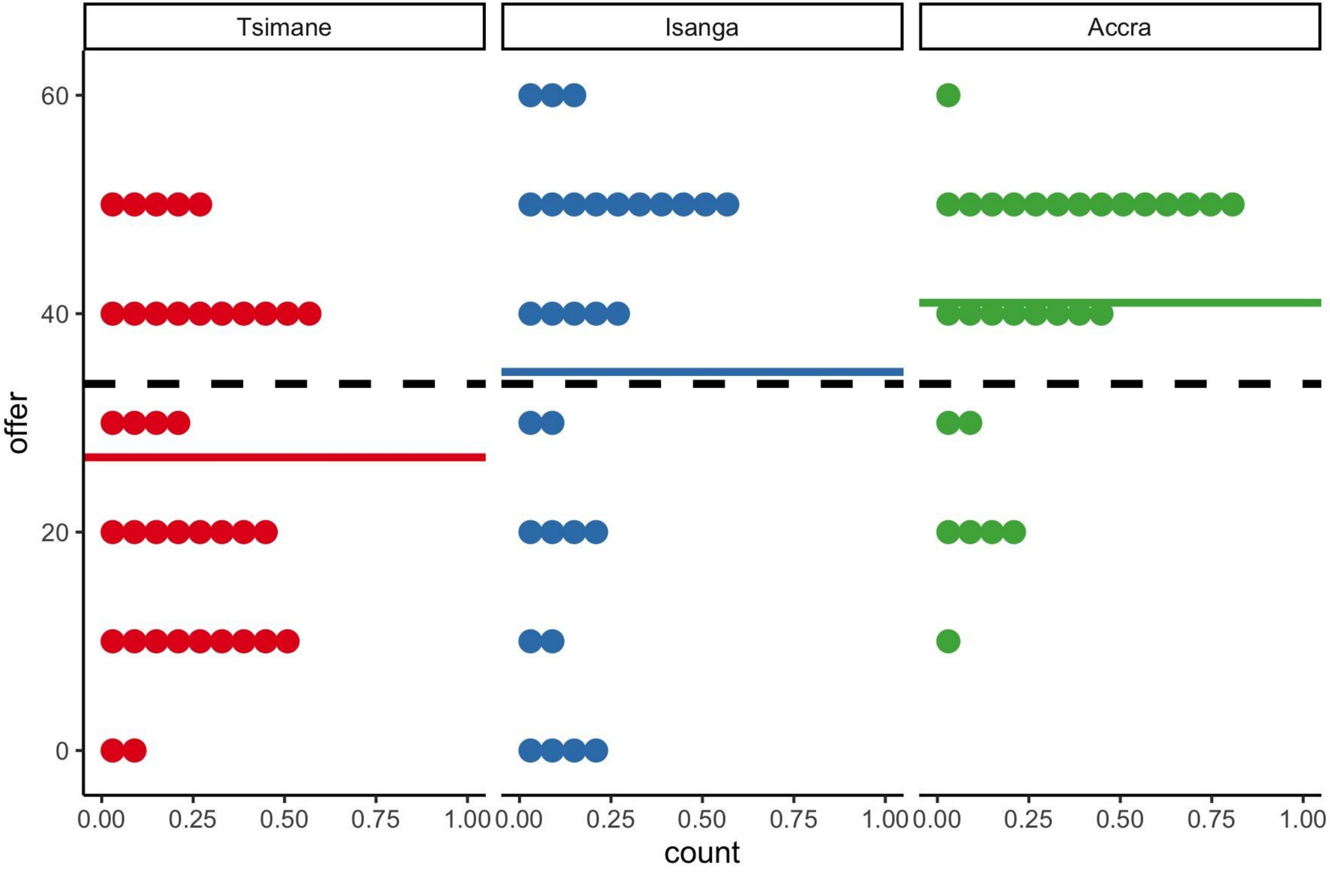
$$\text{null value} = 0$$

where

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and} \quad df = min(n_1 - 1, n_2 - 1)$$

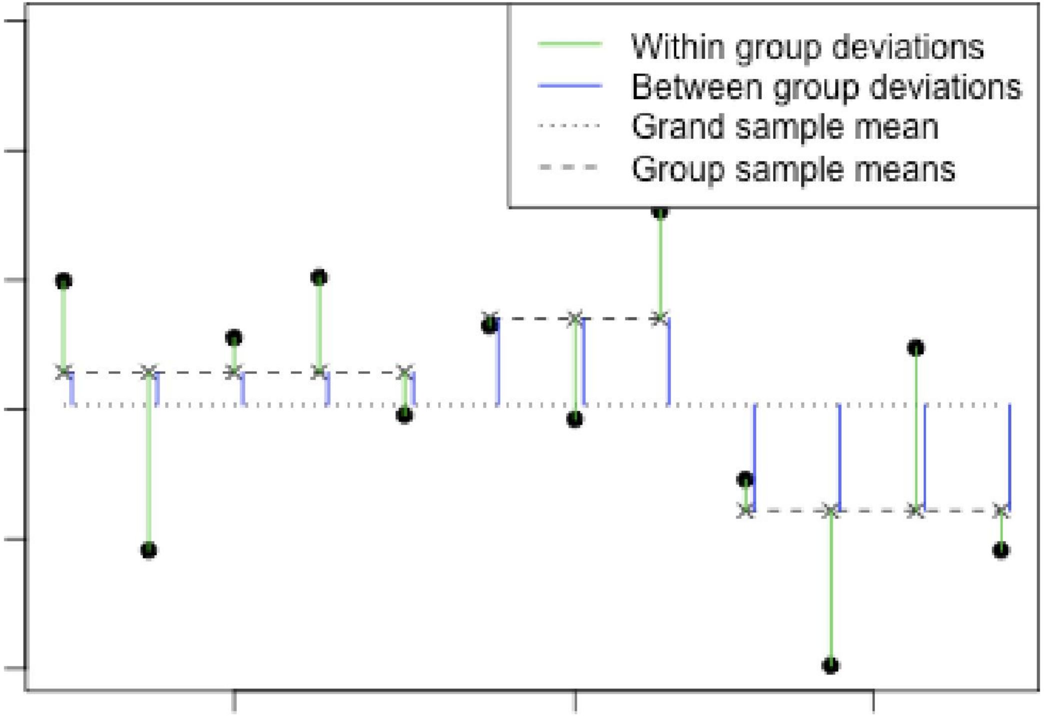
Note: the true *df* is actually different and more complex to calculate (it involves the variance in each estimate relative to its size). But this is close.

Within and between group variance



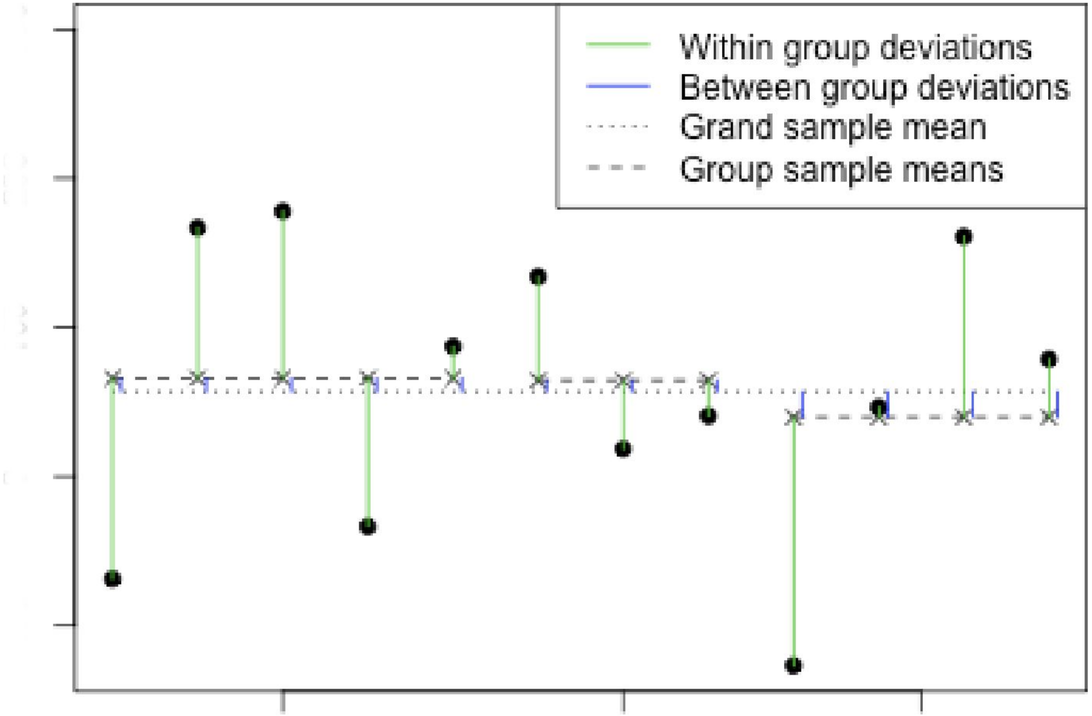
ANOVA compares between group variation to within group variation

$$\sum |^2 \Big/ \sum |^2$$



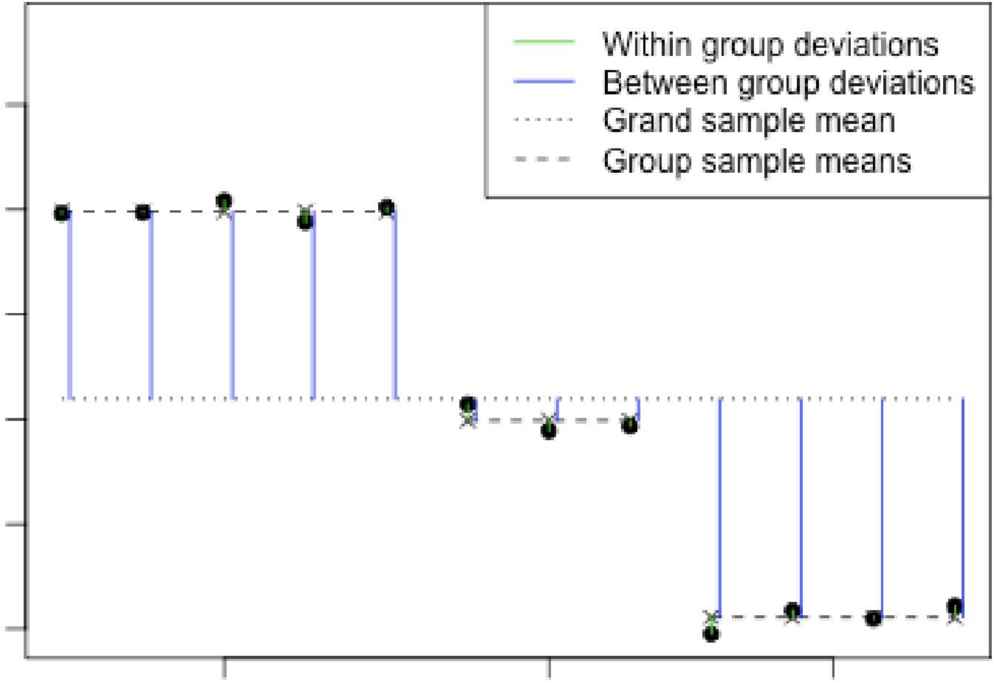
Relatively large WITHIN group variation: little apparent difference

$$\sum |^2 \Big/ \sum |^2$$



Relatively large BETWEEN group variation: there may be a difference

$$\sum |^2 \Big/ \sum |^2$$



Post-hoc tests

If we *knew* we wanted to test only Tsimane vs. Accra, we're only doing one test. But then why did we gather all of this other data?

If we are doing our analyses post-hoc, we are implicitly saying something like "I want to compare the groups that look most different", which is like doing all of those other tests and then rejecting them.

In that case, we are actually doing $\frac{K(K-1)}{2}$ tests.

So our $\alpha^* = \frac{.05}{(16 \cdot 15)/2} = 0.0004$

Main ideas so far

1. What is sampling? Why can we use samples to reason about populations? How does the process of sampling change our inferences?
2. Descriptives statistics as compression. Deciding what statistic is appropriate when.
3. What is null hypothesis testing? What do the two outcomes mean?
4. The Central Limit theorem: When it holds and what it means
5. The Normal Distribution: Critical values, Z-scores, confidence intervals
6. The t-distribution and t-tests: Proportions, means, paired vs. unpaired
7. ANOVA: When you use it, what the values mean, and follow-up tests