




## Reminders

- **Homework assignments are SOLO work!**
- **Please stop moving decimals arbitrarily** ☹
  - 2% = 2/100 = 0.02 …. But 2 years isn't 0.02; it's just 2
- **Don't make inferences *about the sample***
  - You are NEVER going to *estimate* data you already *have* ☺
  - You are 100% confident about what the sample statistic is
    - That's how CI is calculated – it's the *1st step*, that number +/- something
    - It's the *middle* of the interval, the *dead center*.
- **Don't get creative about the Null Hypothesis**
  - The null *always* has an "equals" sign, & *only* =, and *only* the null has it
  - For one variable tests, Ha & Ho have the same value
  - For *two* variable tests (starting today), that value (in both hypos) is zero



SOC424 – Statistics w/ Dr. Ellis Godard

Lecture 20 of 27! ☺

# A Difference of Means



## Claims of Causality

- Most claims made about the world imply a causal relation of the form X ⟶ Y
  - But it's probabilistic and tentative, *not* simply "if x then y"
- **Consider a claim of the form "X explains Y"**
  - When X,Y is more likely
  - Y is more likely if/when X
  - Y happens more often under the *condition* of X
- Y is the "**dependent variable**"
  - What's being predicted or explained
  - It's values *depend* on the values of the *other* variable
- X is the "**independent variable**"
  - Is the predictor or explainer
  - The value of the *other* variable *depends* on what this one is





## Outline for Today…

- **Reminders**
- **Causation & Dependence**
- **Introduction to Bivariate Statistics**
- **Review of a Means Test**
  - Connect w/ Confidence Intervals
- **Difference of Means Test**
  - Same Elements & Steps; Different test statistic & standard error
  - One example of the math & ideas
  - How to do it in SPSS
    - **Independent** Samples t-test w/ Levine's Test *(that's the weird part!)*
  - One example using SPSS





## Simple Rules about Causality

If a causal relation exists, there must be

1. an *association* between two variables;

   But not *just that* – "correlation is not causation"

2. appropriate *time order:* causes b4 effects;

   but not *just that* – *starting* the car doesn't *get* me anywhere

3. *Elimination of alternative explanations*

   e.g. possibility of a *spurious* relationship (more in Elaboration lecture)

## Statistical Differences are Key

- It is *impossible* to establish statistically that the relation between variables *is* a causal relation.
- The best we can do is show that our sample is very different from what we would expect *if* there was *no* relationship – that is, that the data differs from the null.
- We want our data to be *so* different from the null, that the difference is *too* large to just be due to chance.
- We want to see a *significant* difference, which suggests that there's actually a difference in the population.



## Is it just sampling error?

- **There are lots of concerns about empirical claims**
  - Research Methods (e.g. SOC 497) addresses many of them
  - Choices in methods, concerns about bias, etc.
  - Data used to support an idea might be bad data – that's 497
- **A key one is whether <u>inferences</u> are sound**
  - Claims about data might be exaggerated – that's SOC424
  - Small differences don't matter, don't help us generalize
    - Small differences might just be sampling error from random samples
    - Sound inferences require *clear* differences (esp. big ones)
    - An apparent association might be due to sampling variability
  - Need to statistically assess (test) whether a difference is sufficiently large (i.e. far enough from *no* difference) to statistically justify an inference about a population
  - If the difference is far from zero, it is *statistically significant*



## Pre-Bivariate Statistics

- **We have focused on univariate statistics**
  - Single variables, analyzed one at a time
  - HWs 1 thru 6, Report 1, and *most* of Report 2
- **Now, start thinking about a 2^nd^ variable**
  - Examine patterns of values in one variable, in relation to patterns of values on another variable, for each case
  - For example, does the mean of __ differ if we compare two groups (__ and __) of some other variable?
    - Income for men vs women; GPA for Asian vs. Latinx



## Bivariate Statistics

- **Rest of semester is *bivariate* then *multivariate* relations**
  - Basic idea is the same: Is the difference between our data and the null large enough to think there really is a difference in the population?
- **First step is comparing subgroups**
  - Eg look at Ybar for Males & Ybar for Females
- **Next step is comparing groups**
  - One variable to **<u>separate</u>** the cases into groups
    - That's the *grouping variable*, a categorical IV such as gender or race
  - another variable to **<u>compare</u>** those groups
    - That's the test variable, the DV
    - May be proportional (next lecture) or a mean (today)



## Pre-Bivariate Statistics, cont'd

- ***Could* do that using univariate statistics**
  - Will do in later lab:
    - Data / Select Cases / If Condition Satisfied… for 1 group
    - Get Univariate info (shape, center, dispersion)
    - Repeat for a 2nd group
  - Compare pt. estimates of central tendency & dispersion
- ***Could* do that and make interval inferences**
  - Confidence intervals
    - Calculate 2 confidence intervals for the mean, 1 for each group
    - Compare them – do they overlap? If so, what's that mean?
  - Hypothesis test
    - Use a two-samples t-test to compare two means…
- **Or *could* do that with a single inference…**



## Test Statistic

<u>To test a hypothesis</u>, we collect a sample, measure it's mean or proportion, and <u>compute a *test statistic*</u>.

<u>All test statistics measure</u> the same thing: how much observations <u>differ from the (*null*)</u> hypothesized *value*

<u>All tests ask</u> the same question:

<u>is that difference large</u> enough in the sample, that we expect it to be large in the population too?

or have we seen a difference (distance) that is small, negligible, possibly just a result of chance?

The way the test statistic is calculated depends on the *sample size*, the *scale of measurement* of our variables, and sometimes the *shape of the population distribution* – each statistic assumes something about each of those – but all have the same logic!

## Testing a Mean

- **Z & T use same basic test score formula**

$$\frac{\textit{some \# - someother \#}}{\textit{someerror}}$$

  - Use t if n<30; use z if n>29
  - t distribution is approximately std normal (i.e. same as z) at 30
    - higher kurtosis below 30 - e.g. df=6 or df=2





## One-tailed hypotheses

For a one-tail test,

$$H_o: \mu_1 = \mu_2 \qquad H_o: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 < \mu_2 \quad \text{so} \quad H_a: \mu_1 - \mu_2 < 0$$

Alternatively, switching the direction of the inequality,

$$H_o: \mu_1 = \mu_2 \qquad H_o: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 > \mu_2 \quad \text{so} \quad H_a: \mu_1 - \mu_2 > 0$$





## Testing a *Difference* of Means

- **Same idea as other tests**
  - Test the null of no difference
  - Get a t value associated w/ some p value
  - P = probability of being wrong in rejecting the null
  - We want that probability to be a small as possible
    - We want to reject the null hypothesis
    - We don't want to be wrong when we do that
    - So we want the risk of being wrong to be really small
- **One statistical difference:**
  - Sampling distribution is of *differences in means*
    - Rather than a sampling distribution of means (t or z)
    - Different sampling distribution, so different standard error
    - Table in the book is like the chi-square table (later)

New (but not really)





## Example: Education and Children

Whom would you expect to have more kids – college-educated adults or adults without a college education?

Formally, test this hypothesis…

$H_o: \mu_{Coll} = \mu_{noColl}$ **(No difference in family size)**

$H_a: \mu_{Coll} < \mu_{noColl}$ **(College-educated have fewer children)**





## Setting up the Hypotheses

Usually, we are interested in whether or not the difference between two parameters is equal to zero.

For a two-tail test,

$$H_o: \mu_1 = \mu_2 \qquad H_o: \mu_1 - \mu_2 = 0$$
$$H_a: \mu_1 \neq \mu_2 \quad \text{But that's really} \quad H_a: \mu_1 - \mu_2 \neq 0$$





## Data pieces that we need…

| | SampleSize | Mean | StdDev | (Est. Std. Error)$^2$ |
|---|---|---|---|---|
| First | $n_1$ | $\overline{Y}_1$ | $s_1$ | $s_1^2/n_1$ |
| Second | $n_2$ | $\overline{Y}_2$ | $s_2$ | $s_2^2/n_2$ |
| **Difference** | | $\overline{Y}_1 - \overline{Y}_2$ | | $\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$ |

## GSS Results
(College Educated = 16 or more years of schooling)

| | SampleSize | Mean | StdDev | (EstStdErr)$^2$ |
|---|---|---|---|---|
| Not College | 1168 | 2.05 | 1.800 | $(1.800)^2/1168$ |
| College | 334 | 1.35 | 1.487 | $(1.487)^2/334$ |
| Difference | | 0.7045 | | $\frac{(1.800)^2}{1168} + \frac{(1.487)^2}{334}$ |





## Test of the Difference

$$\text{standard error} = \sigma_{\overline{Y}_2 - \overline{Y}_1} = \sqrt{\frac{(1.800)^2}{1168} + \frac{(1.487)^2}{334}} = 0.096$$

test statistic

$$z = \frac{\overline{Y}_2 - \overline{Y}_1}{\sigma_{\overline{Y}_2 - \overline{Y}_1}} = \frac{0.7045}{0.096} = 7.27$$

The p-value that corresponds to this test is so small that it is not even listed in Table A. We can reject $H_0$.and find support for the claim that college-educated individuals have fewer children than non-college educated individuals.





## Dependent Groups t-test

NO – NOT THIS ONE!

- Sometimes subjects in the two groups will be matched in some way.
  - It may be the same subjects at two points in time.
  - It may be individuals matched on several significant variables.
  - It may be twins, etc.
- In that case, the distribution of differences between means should have less variance. This is because the linkages between subjects makes the extreme differences you get with independent subjects unlikely to occur.
- This makes it easier to detect differences if they exist.





## Dependent Groups t-test

NO – NOT THIS ONE!

- The formula contains a new term in the denominator

$$t_{df} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2_{\bar{x}_1 - \bar{x}_2} - 2r_{12} s_{\bar{x}_1} s_{\bar{x}_2}}}$$

- When there is a correlation between the two sets of subjects, the standard error of the distribution of differences is reduced.
- When the correlation equals 0, the term drops out and you are back to the original equation.
- Regardless, don't compute it yourself. Let SPSS do the work!





## What's an "Independent" Sample?

Yes – this one ☺

Samples are independent when the likelihood of selecting a particular member of one sample is not dependent on who or what is chosen in the other sample.

Importantly, when two groups -- e.g. North and South -- are selected for a comparative analysis, if the overall sample was randomly selected, then the subsamples are independent random samples.





*SPSS note*

## Levine's F Test for Variance

Don't forget this!

- **Difference of Means output has 2 rows**
  - 2 formulas for the test statistic
  - Differ in whether treat group variances as equal or not
- **If the p value is above .05, use 1st row's t & p**
  - Can't reject the null that they are equal
  - Must assume that the variances are equal
- **If the p value is below .05, use 2nd row's t & p**
  - Can reject the null that they are equal
  - Needn't assume the variances are equal
- **Summary:**
  - Above .05? Use above line
  - Below .05? Use below line

*using rshows.sav*

## T-Test Example: RSHOWS x PCTYPE

- **Hypotheses:**
  - $H_A$: Laptop owners watch more reality shows than desktop owners
  - $H_0$: Mean for laptop owners and desktop owners is equal
- **ANALYZE – COMPARE MEANS – Independent…**
  - Test variable (DV) is RSHOWS, an index of 8 nominal variables, that measures how many of eight reality shows respondents watch
  - Grouping variable (IV) is PCTYPE – 1 = laptop, 2 = desktop
  - You could click "Options" to set Confidence Level (95% by default) for the Confidence Interval (which is also in the output)
  - In recent versions of SPSS (e.g. v30), you'll need to click the box for "Homogeneity of Variance Test" – that's the Levine's test, which will be in a separate table below the t test results. (In earlier versions, it was 2 columns in that larger table of all the t test results.)

26 SOC424 @ CSUN - Ellis Godard



*using rshows.sav*

## T-Test Example: Use of Output

- Laptop owners watch an average of 1.0000 reality shows
- Desktop owners watch an average of 1.2917 reality shows
- The t score for that difference is 0.832, which is quite small
- The p value associated w/ that distance is 0.410
  - There's a 41% chance of being wrong in rejecting the null hypothesis
- That's too great a risk, so we can't reject the null hypothesis
  - We can't conclude that laptop users watch *significantly* more reality shows than desktop PC owners

27 SOC424 @ CSUN - Ellis Godard



*using musicB.sav*

## T-Test Lab: MUSIC x GRASS

- **Hypotheses:**
  - $H_A$: Those who support legalization of marijuana listen to more kinds of music than those who do not support legalization
  - $H_0$: The mean number of kinds of music is not likely to change with regard to whether a respondent supports legalization or not
- **ANALYZE – COMPARE MEANS – Independent…**
  - Test variable (DV) is MUSIC, an index of 14 nominal variables, that measures how many genres of music each of you listen to
  - Grouping variable (IV) is Grass – 1 = Yes, 2 = No
  - See slide 26 if you have a recent version of SPSS – you need the Levine's test!!
- **Read slide 25 if you haven't yet ☹**
- **Things to do with the output:**
  - What's the mean number of genres legalization supporters like?
  - What's the mean number of genres legalization opponents like?
  - What is the t score for the difference in those means?
  - What is the p value ("Sig.") for that t score?
  - Interpret that p value – what does it mean?
  - Make a conclusion – can we reject the null hypothesis? Is Ha supported?

28 SOC424 @ CSUN - Ellis Godard