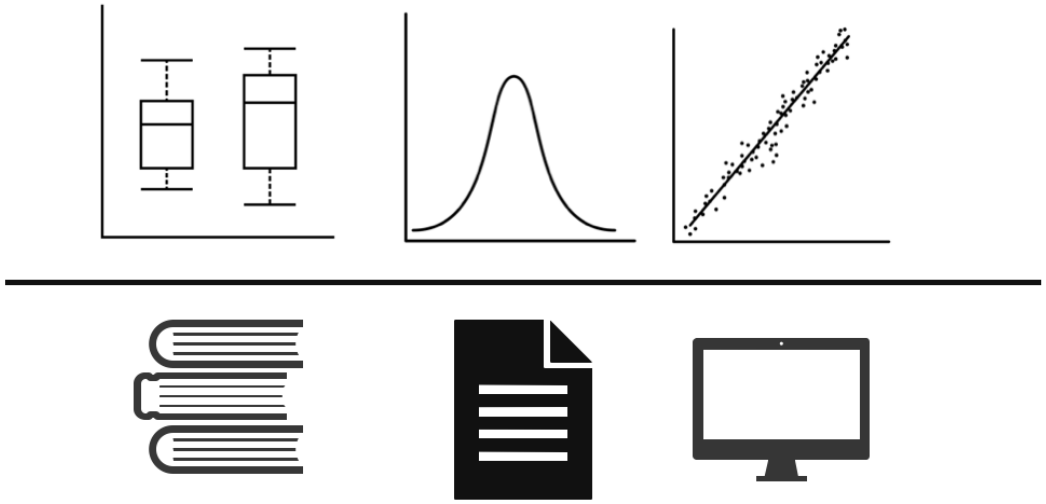
PSYC 2300

Introduction to Statistics



Lecture 16: Final Exam Review

Outline for today

Final Exam Review

- Key Computations
  - Q & A
- Key Concepts
  - Q & A
- Materials and general tips

Key Computations for the Final Exam

Key Computations for the Final Exam

- Independent samples *t*-test
- Dependent samples *t*-test
- One-way ANOVA
- Simple linear regression

Independent samples *t*-test

Independent Samples *t*-test

Independent samples *t*-test

$$t_{(\bar{x}_1 - \bar{x}_2)} = \frac{\overline{x}_1 - \overline{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}}$$

Estimated standard error of the mean difference

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 n_2}\right]}$$

Example: Independent samples *t*-test

No hallway bump

$n_1 = 10$

$\bar{x}_1 = 5.5$

$s_1^2 = 2.5$

Hallway bump

$n_2 = 10$

$\bar{x}_2 = 8.5$

$s_2^2 = 3$

Independent samples *t*-test

$$t_{(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}}$$

Estimated standard error of the mean difference

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 n_2}\right]}$$

Estimated standard error of the mean difference

No hallway bump

$n_1 = 10$

$\bar{x}_1 = 5.5$

$s_1^2 = 2.5$

Hallway bump

$n_2 = 10$

$\bar{x}_2 = 8.5$

$s_2^2 = 3$

$$s_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\left[\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1 n_2}\right]}$$

$$s_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\left[\frac{(10-1)2.5 + (10-1)3}{10+10-2}\right]\left[\frac{10+10}{(10)(10)}\right]}$$

$$s_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\left[\frac{(9)2.5 + (9)3}{18}\right]\left[\frac{20}{100}\right]}$$

Estimated standard error of the mean difference

No hallway bump

$$n_1 = 10$$

$$\bar{x}_1 = 5.5$$

$$s_1^2 = 2.5$$

Hallway bump

$$n_2 = 10$$

$$\bar{x}_2 = 8.5$$

$$s_2^2 = 3$$

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\left[\frac{22.5 + 27}{18}\right]\left[\frac{20}{100}\right]}$$

$$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\left[2.75\right]\left[0.20\right]} = .7416$$

$$s_{(\bar{x}_1 - \bar{x}_2)} = .74$$

Independent samples *t*-test statistic

No hallway bump

$n_1 = 10$

$\bar{x}_1 = 5.5$

$s_1^2 = 2.5$

Hallway bump

$n_2 = 10$

$\bar{x}_2 = 8.5$

$s_2^2 = 3$

$$t_{(\overline{x}_1 - \overline{x}_2)} = \frac{\overline{x}_1 - \overline{x}_2}{s_{(\overline{x}_1 - \overline{x}_2)}}$$

$$t_{(\overline{x}_1 - \overline{x}_2)} = \frac{5.5 - 8.5}{0.74}$$

$$t_{(\overline{x}_1 - \overline{x}_2)} = -4.054$$

Dependent Samples *t*-test

Dependent samples *t*-test

Dependent samples *t*-test

$$t_{\bar{x}_d} = \frac{\bar{x}_d}{s_{\bar{x}_d}}$$

Standard error of the difference scores

$$s_{\bar{x}_d} = \sqrt{\frac{s_d^2}{n}}$$

Dependent samples *t*-test

Dependent samples *t*-test

$$t_{\bar{x}_d} = \frac{\bar{x}_d}{s_{\bar{x}_d}}$$

$\bar{x}_d$ is the mean of the difference scores

$s_{\bar{x}_d}$ is the standard error of the mean difference

Dependent samples *t*-test

Standard error of the difference scores

$$s_{\bar{x}_d} = \sqrt{\frac{s_d^2}{n}}$$

s_d^2 is the variance of the difference scores

n is the size of the sample (the number of difference scores)

Example: Dependent Samples *t*-test

participant	pre_treatment	post_treatment	difference_score
1	5	2	3
2	3	3	0
3	6	2	4
4	9	4	5

$\bar{x}_d = 3$ mean of the difference scores

$s_d^2 = 4.67$ variance of the difference scores

Example: Dependent Samples *t*-test

difference_score
3
0
4
5

$\bar{x}_d = 3$

$s^2_d = 4.67$

$$t_{\bar{x}_d} = \frac{\bar{x}_d}{s_{\bar{x}_d}}$$

$$t_{\bar{x}_d} = \frac{3}{\sqrt{\frac{s^2_d}{n}}}$$

$$t_{\bar{x}_d} = \frac{3}{\sqrt{\frac{4.67}{4}}} = \frac{3}{1.080509} = 2.78$$

One-way ANOVA

$$F = \frac{MS_{between}}{MS_{within}}$$

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

$$SS_{between} = \sum \frac{(\sum x)^2}{n} - \frac{(\sum \sum x)^2}{nT}$$

$$SS_{within} = \sum \sum (x^2) - \sum \frac{(\sum x)^2}{n}$$

$$df_{between} = k - 1$$

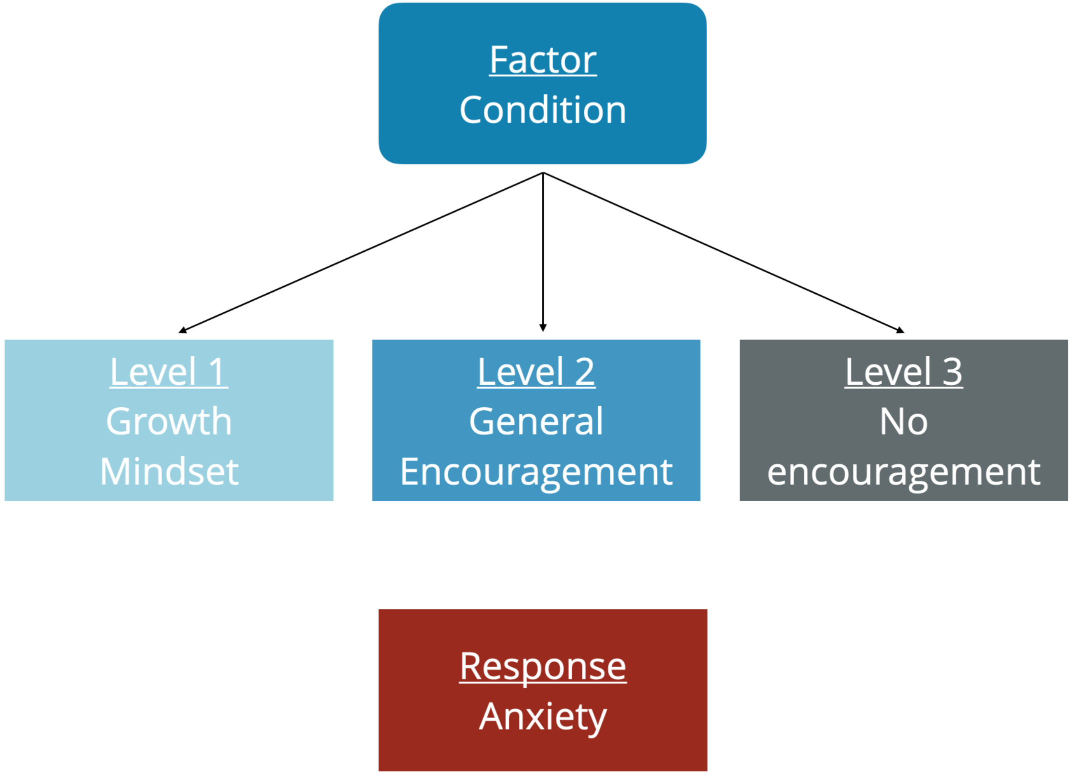
$$df_{within} = nT - k$$

Sum of Squares: One-way ANOVA

For each group, first calculate		Afterwards, calculate across groups
n	Add up all the ns	nT
$\sum x$	Add up all the $\sum x$	$\sum \sum x$
$\bar{x}$	Square the above value and divide by nT	$\frac{(\sum \sum x)^2}{nT}$
$\frac{(\sum x)^2}{n}$	Add up all the $\frac{(\sum x)^2}{n}$	$\sum \frac{(\sum x)^2}{n}$
$\sum (x^2)$	Add up all the $\sum (x^2)$	$\sum \sum (x^2)$

Example Experiment

Research Question: Does teaching students about *growth mindsets* affect students' anxiety more than giving them *general encouragement* or *no encouragement*?



Group Calculations: Group 1

Growth Mindset

Growth	Growth^2
6	36
6	36
5	25
8	64
7	49
6	36
3	9
4	16

$$n_1 = 8$$

$$\bar{x}_1 = 5.625$$

$$\Sigma x_1 = 45$$

$$\frac{(\Sigma x_1)^2}{n_1} = \frac{45^2}{8} = 253.125$$

$$\Sigma(x_1^2) = 271$$

Example Experiment

Repeat the same process for each group.

Growth

$n_1 = 8$

$\bar{x}_1 = 5.625$

$\Sigma x_1 = 45$

$\frac{(\Sigma x_1)^2}{n_1} = 253.125$

$\Sigma(x_1^2) = 271$

General

$n_2 = 8$

$\bar{x}_2 = 6.625$

$\Sigma x_2 = 53$

$\frac{(\Sigma x_2)^2}{n_2} = 351.125$

$\Sigma(x_2^2) = 359$

None

$n_3 = 8$

$\bar{x}_3 = 8.125$

$\Sigma x_3 = 65$

$\frac{(\Sigma x_3)^2}{n_3} = 528.125$

$\Sigma(x_3^2) = 537$

Sum of Squares: One-way ANOVA

For each group,
first calculate

$$n$$

$$\sum x$$

$$\bar{x}$$

$$\frac{(\sum x)^2}{n}$$

$$\sum (x^2)$$

Sum of Squares: One-way ANOVA

For each group, first calculate		Afterwards, calculate across groups
n	Add up all the ns $\longrightarrow$	nT
$\sum x$		
$\bar{x}$		
$\frac{(\sum x)^2}{n}$		
$\sum (x^2)$		

Sum of Squares: One-way ANOVA

For each group,
first calculate

Afterwards,
calculate across groups

n $\xrightarrow{\text{Add up all the } ns}$ nT

$\sum x$ $\xrightarrow{\text{Add up all the } \sum x}$ $\sum \sum x$

$\bar{x}$

$\frac{(\sum x)^2}{n}$

$\sum (x^2)$

Sum of Squares: One-way ANOVA

For each group, first calculate

n

$\sum x$

$\bar{x}$

$\frac{(\sum x)^2}{n}$

$\sum (x^2)$

Add up all the ns

Add up all the $\sum x$

Square the above value and divide by nT

Afterwards, calculate across groups

nT

$\sum \sum x$

$\frac{(\sum \sum x)^2}{nT}$

Sum of Squares: One-way ANOVA

For each group, first calculate		Afterwards, calculate across groups
n	Add up all the ns →	nT
$\sum x$	Add up all the $\sum x$ →	$\sum\sum x$
$\bar{x}$	Square the above value and divide by nT	$\dfrac{(\sum\sum x)^2}{nT}$
$\dfrac{(\sum x)^2}{n}$	Add up all the $\dfrac{(\sum x)^2}{n}$ →	$\sum\dfrac{(\sum x)^2}{n}$
$\sum(x^2)$		

Sum of Squares: One-way ANOVA

For each group, first calculate		Afterwards, calculate across groups
n	Add up all the ns	nT
$\sum x$	Add up all the $\sum x$	$\sum \sum x$
$\bar{x}$	Square the above value and divide by nT	$\frac{(\sum \sum x)^2}{nT}$
$\frac{(\sum x)^2}{n}$	Add up all the $\frac{(\sum x)^2}{n}$	$\sum \frac{(\sum x)^2}{n}$
$\sum (x^2)$	Add up all the $\sum (x^2)$	$\sum \sum (x^2)$

Calculate Terms Across Groups

Calculate all the terms needed for $SS_{between}$ and SS_{within}

$$nT = n_1 + n_2 + n_3 = 8 + 8 + 8 = 24$$

$$\Sigma\Sigma x = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 45 + 53 + 65 = 163$$

$$\frac{(\Sigma\Sigma x)^2}{nT} = \frac{163^2}{24} = 1107.04$$

$$\Sigma\frac{(\Sigma x)^2}{n} = \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} = 253.125 + 351.125 + 528.125 = 1132.38$$

$$\Sigma\Sigma(x^2) = \Sigma(x_1^2) + \Sigma(x_2^2) + \Sigma(x_3^2) = 271 + 359 + 537 = 1167$$

Calculating $SS_{between}$ and SS_{within}

$$SS_{between} = \sum \frac{(\sum x)^2}{n} - \frac{(\sum \sum x)^2}{nT}$$

$$SS_{within} = \sum \sum (x^2) - \sum \frac{(\sum x)^2}{n}$$

$$\frac{(\Sigma\Sigma x)^2}{nT} = 1107.04 \qquad \Sigma\frac{(\Sigma x)^2}{n} = 1132.38 \qquad \Sigma\Sigma(x^2) = 1167$$

$$SS_{between} = 1132.38 - 1107.04 = 25.34$$

$$SS_{within} = 1167 - 1132.38 = 34.62$$

Calculating $df_{between}$ and df_{within}

$$df_{between} = k - 1$$

$$df_{within} = nT - k$$

k = number of groups
nT = total sample size
df = degrees of freedom

$$df_{between} = k - 1 = 3 - 1 = 2$$

$$df_{within} = nT - k = 24 - 3 = 21$$

Calculating $MS_{between}$ and MS_{within}

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

$$SS_{between} = 1132.38 - 1107.04 = 25.34$$

$$df_{between} = k - 1 = 3 - 1 = 2$$

$$MS_{between} = \frac{25.34}{2} = 12.67$$

$$SS_{within} = 1167 - 1132.38 = 34.62$$

$$df_{within} = nT - k = 24 - 3 = 21$$

$$MS_{within} = \frac{34.62}{21} = 1.64$$

Calculating the F statistic

$$F = \frac{MS_{between}}{MS_{within}}$$

$$MS_{between} = \frac{25.34}{2} = 12.67$$

$$MS_{within} = \frac{34.62}{21} = 1.64$$

$$F = \frac{12.67}{1.64} = 7.70$$

Simple Linear Regression

Calculating the regression line

i	x	y	x^2	y^2	xy
1	5	6	25	36	30
2	9	11	81	121	99
3	10	6	100	36	60
4	3	4	9	16	12
5	5	6	25	36	30
6	7	9	49	81	63

Sum each of the columns

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

Calculating the regression line

Slope

$$b = \frac{\Sigma XY - \frac{\Sigma x \Sigma y}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$$

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

$$b = \frac{294 - \frac{(39)(42)}{6}}{289 - \frac{(39)^2}{6}}$$

$$b = \frac{294 - 273}{289 - 253.5} = \frac{21}{35.5} = 0.59$$

Calculating the regression line

Intercept

$$a = \frac{\Sigma Y - b\Sigma X}{n}$$

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

Slope, $b = 0.59$

$$a = \frac{42 - (0.59)(39)}{6} = \frac{18.99}{6} = 3.16$$

Calculating the regression line

Simple linear regression

$$\hat{y} = bX + a$$

Slope: $b = 0.59$

Intercept: $a = 3.16$

$$\hat{y} = 0.59X + 3.16$$

Q & A: Computations

Key Concepts for the Final Exam

Key Concepts

- Identifying when to use a particular test statistic
  - One sample, independent, or dependent *t*-test
  - One-way ANOVA or Simple linear regression
- Identifying design types (within-subjects, between-subjects)
- Difference between factors and levels
- Main effects vs. interaction
- Why ANOVA instead of many *t*-tests?
- Effect size benchmarks

t-tests: one sample

One-sample *t*-test statistic

$$t_{\overline{x}} = \frac{\overline{x} - \mu_{\overline{x}}}{s_{\overline{x}}}$$

Estimated standard error

$$s_{\overline{x}} = \frac{s}{\sqrt{n}}$$

When to use it

- When you only have one sample of participants
- You are comparing a *sample* mean to a population mean or a chance value

Example 1: You want to know if a sample of DU students' math performance is higher than the national average (i.e., a population mean)

Example 2: You want to determine whether a sample's lie detection peformance is above chance, (i.e., 50%)

One sample design

- We know the population mean but not the standard deviation or variance
- We can estimate it with the sample statistics using a one-sample t-test

t-tests: independent samples

Independent samples *t*-test

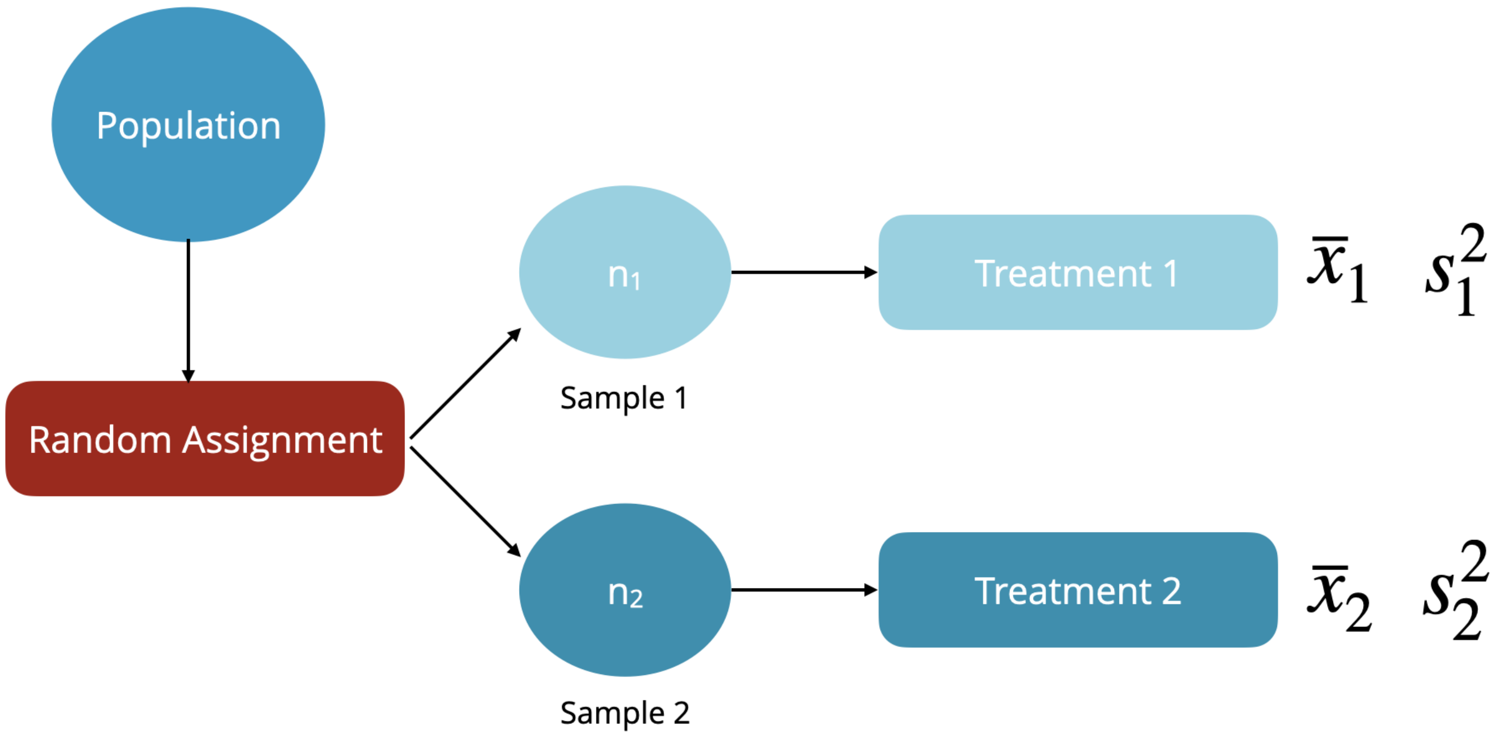
$$t_{(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}}$$

When to use it

- You have two independent samples of participants
- Each sample of participants is in only **one** condition in the study
- Example 1: You are comparing participants who were exposed to cognitive behavioral therapy to those who were not
- Example 2: You are comparing two groups not randomly assigned (e.g., DU students vs. CU students)

Independent samples

- This is also referred to as a **between-subjects** design.
- Why? Because we are comparing the means *between* two groups



t-tests: dependent samples

Dependent samples *t*-test

$$t_{\bar{x}_d} = \frac{\bar{x}_d}{s_{\bar{x}_d}}$$

When to use it

- One sample of participants are in **both** conditions of a study
- You are comparing a pre-treatment measure to a post-treatment measure

Example 1: We are comparing whether participants are faster at identifying short words compared to long words

Exampl 2: We measure anxiety before a treatment, administer a treatment, and measure anxiety post-treatment

Dependent samples

- This is also referred to as a **within-subjects** design.
- Why? Because we are comparing two means *within* groups

Practice Examples

Which type of statistical test?

A researcher is investigating whether the prevalence of drug use on their campus differs from the national prevalence of universities and colleges.

One-sample t-test

What type of design?

A researcher wants to know whether a new memory supplement enhances recall. Participants are given an initial memory test. Then, they are given the new supplement and wait for it to take effect by doing a neutral task. Then, the research gives them a final memory test.

Within-subjects

Practice Examples

Which type of design and statistical test?

A researcher wants to know if people will have better job interview performance if they watch funny videos to calm their nerves before the interview. Participants are randomly assigned to watch either funny videos or neutral videos about waves, birds, and clouds. Participants then do a mock interview. The interviewer rates how likely they would be to hire them.

Between-subjects, *Independent samples t-test*

Practice Examples

The relationship between research designs and *t*-tests

Design	Test
One sample of participants	One-sample t-test
Between-subjects	Independent samples t-test
Within-subjects	Dependent samples t-test

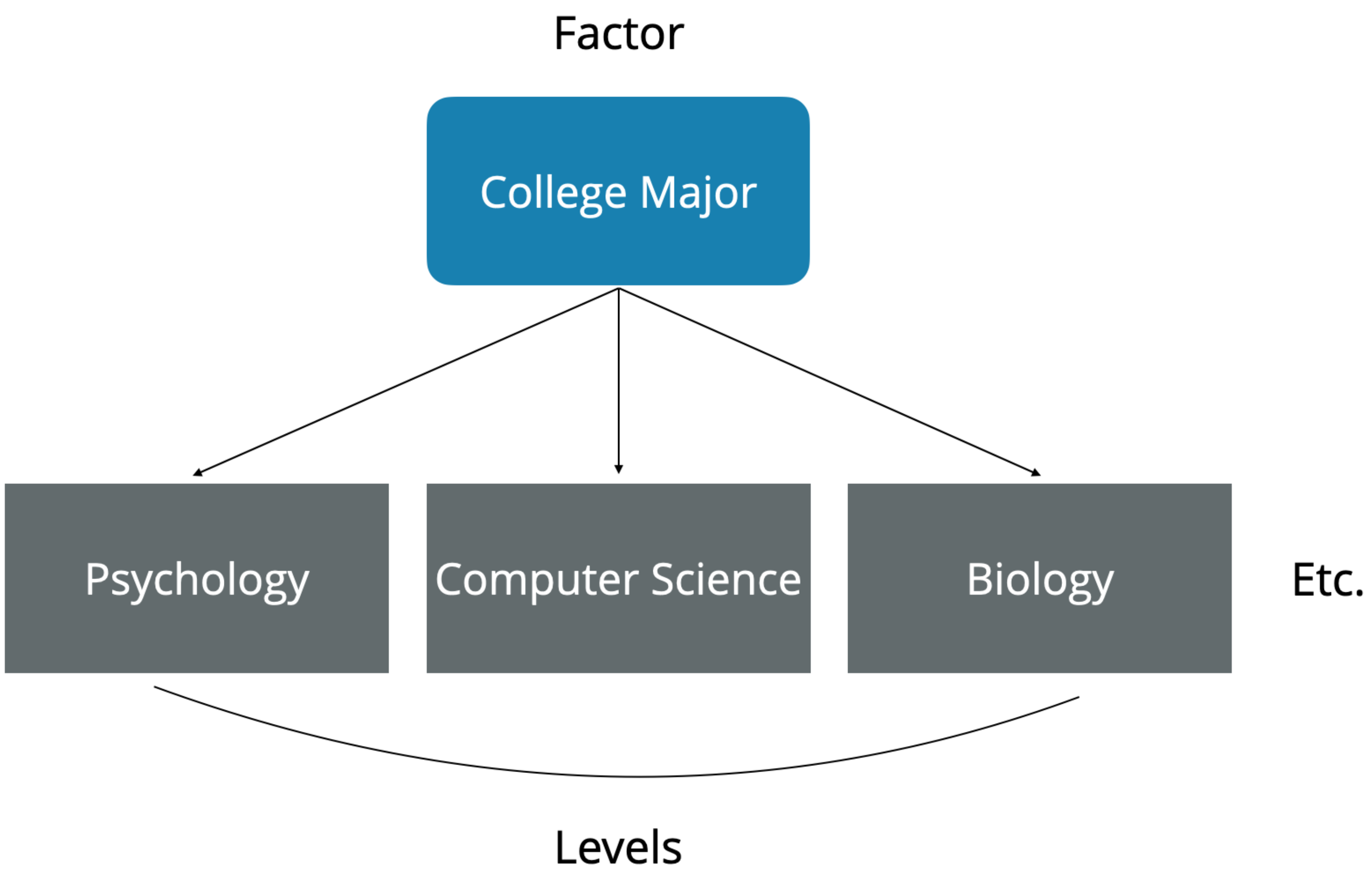
ANOVA: Terminology

Factor: Independent variable(s) in the study

Level(s): Groups within each independent variable

Response: The dependent variable in the study

Example: One factor, several levels



Practice Examples

Factor or level?

A researcher is studying DU students. Specifically, the difference in 1 mile running time among DU athletes, DU students who go to the gym regularly, and DU students who are not athletes or go to the gym regularly. Here, *"DU athletes"* is a:

Level

How many levels are in this factor?

Three levels:

- DU athletes
- DU students who go to the gym regularly
- DU students who are not athletes or go to the gym regularly

Practice Examples

The same researcher does a follow-up study with DU students. Specifically, the researcher is studying difference in 1 mile running time among DU athletes, DU students who go to the gym regularly, and DU students who are not athletes or go to the gym regularly. The researcher also asks students whether they participated in sports during high school (yes, no) to see if prior experience affects the results. Here, *"participated in sports during high school"* is a:

Factor

How many factors and levels are in this factorial design?

Factor 1: Current Fitness status

- DU athlete
- DU gym-goers
- neither

Factor 2: Previous Experience

- Participated in sports in high school
- Did not participate in sports in high school

Main effects and interactions

Main effect: When an analysis of the data reveals a difference between the levels of any factor

Interaction: describes the degree to which the effect of *one factor* depends on the *level* of the other factor

ANOVA vs. Regression

ANOVA

- In ANOVA the factors (or independent variables) *must be* nominal or ordinal (i.e., categorical)
- In ANOVA, we are comparing mean differences

Regression

- In Regression the predictors (or independent variables) *can be* interval or ratio (i.e., continuous)
- In regression, we are making predictions

Q & A: Concepts

Materials to Review

Lectures

Review lectures 10-15 to prepare for the Final Exam

Lecture 09 | Differences from the population

Lecture 10 | Differences between two groups (part I)

Lecture 11 | Differences between two groups (part II)

Lecture 12 | Differences between many groups

Lecture 13 | Differences between many factors

Lecture 14 | Testing relationships with correlations

Lecture 15 | Making predictions using regression

Guides

Review Guides 11-15 to prepare for the Final Exam

Guide 10 | Calculating independent samples t-test

Guide 11 | Effect size for independent samples t-test

Guide 12 | Calculating dependent samples t-test

Guide 13 | Effect size for dependent samples t-test

Guide 14 | Calculating one way anova

Guide 15 | Effect size for one way anova

Readings

Review Ch.11-16 to prepare for the Final Exam

Ch.11 | *t(ea) for Two: Tests Between the Means of Different Groups*

Ch.12 | *t(ea) for Two (Again): Tests Between the Means of Related Groups*

Ch.13 | *Two Groups Too Many? Try Analysis of Variance*

Ch.14 | *Two Too Many Factors: Factorial Analysis of Variance—A Brief Introduction*

Ch.15 | *Testing Relationships Using the Correlation Coefficient: Cousins or Just Good Friends?*

Ch.16 | *Using Linear Regression: Predicting the Future*

Formula Sheet

- Practice identifying each of the equations on the formula sheet
- You may want to make flash cards
- Then, practice with the formula sheet and complete the Final Review Practice Packet

Course Evaluations

The Last Next Time

Application Project

- Due tonight 03/09/2022 at 11:59pm MT

Final Exam

- Thursday 03/17/2022 08:00-09:50am Sturm Hall 187

Extra Credit

- Due 03/17/2022 11:59pm MT
  - SONA Participation
  - Make a meme