EDUR 7130
Presentation 5b

Inferential Statistics

Types of Statistical Procedures

Table 2: Types of Statistical Procedures and Their Characteristics

Statistical Test	*IV*	*DV*	*Special Feature*	*Example Hypothesis*
(1) Pearson's r (correlation coefficient)	Quant.	Quant.		There is a positive relationship between intelligence and mathematics achievement scores.
(2) Two group t-test	Qual.	Quant.	IV has only 2 categories	There will be a difference between boys and girls on mathematics achievement scores.
(3) ANOVA (Analysis of Variance)	Qual.	Quant.	IV may have 2 or more categories.	There will be a difference among Black, Hispanic, and White students in mathematics achievement scores.
(4) ANCOVA (Analysis of Covariance)	1. Qual. 2. Covariate may be either qual. or quan., but usually quan.	Quant.	IV may have 2 or more categories. Covariate used to make adjustments to DV means.	There will be a difference among Black, Hispanic, and White students in mathematics achievement scores after taking into account levels of motivation.
(5) Chi-Square $\chi 2$ Test of Association	Qual.	Qual.		Males will be more likely to drop out of school than females.

Table can be found here
http://www.bwgriffin.com/gsu/courses/edur7130/content/inferential_statistics.htm

Pearson's r (correlation)

- Examines linear relationship between two (or more) variables.
- Typically both variables must be quantitative (although exceptions do exist).

Examples of the type of variables for which Pearson's r would be appropriate to use:

- Math scores and reading scores
- Income and number of years of schooling
- Level of motivation and hours of practice

Relationships with Pearson's r

- ***Positive*** relationship exists when two variables move in the same direction -- high scores on one variable correspond to high scores on the other variable; low scores on one correspond with low scores on the other.
- ***Negative*** means that low scores on one variable are associated with high scores on another variable, so there is an inverse relationship, as one variable increases the other decreases.

Which is this, positive or negative?

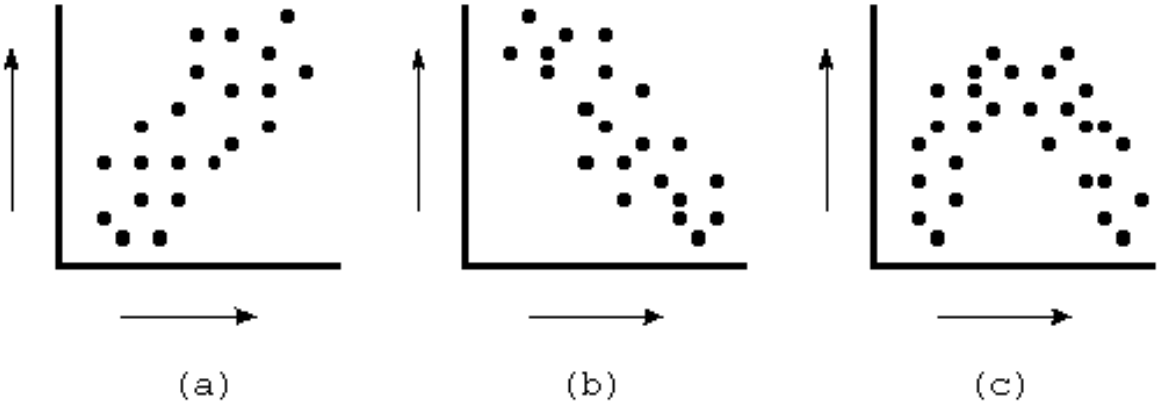
As the number of miles traveled increase, the gallons of gas in the tank decrease.

Negative, so as you get higher numbers on miles traveled, you get lower numbers on gallons in tank.

Numeric Value of Pearson's r

- Measures only linear relationships
- Pearson's r ranges from -1.00 to 1.00 (extremes represent perfect relationships, perfect fits; -1 = perfect negative relation, 1 = perfect positive relation)
- Pearson's r = 0.00 means no linear relationship, but there could be a non-linear relationship

Linear



Which type of relationships are illustrated by (a), (b), and (c)?

(a) Positive,
(b) negative, and
(c) is non-linear. A non-linear (curvilinear) relationship exists when, for example, the relationship between two variables (we will use variables A and B) is positive up to a point, then changes direction and becomes negative. Anytime a scatterplot shows a relationship that does not follow a straight line, but does show a discernible pattern that curves, one has a non-linear relationship.

Non-linear Example

Young children have less dexterity than older children. So as age increases, so does dexterity. This is a positive relationship. At some point dexterity begins to fade in life, so after this point in life, as age increases, dexterity declines. This is now a negative relationship. So age and dexterity are non-linearly related.

r = 0.00; No Linear Relationship

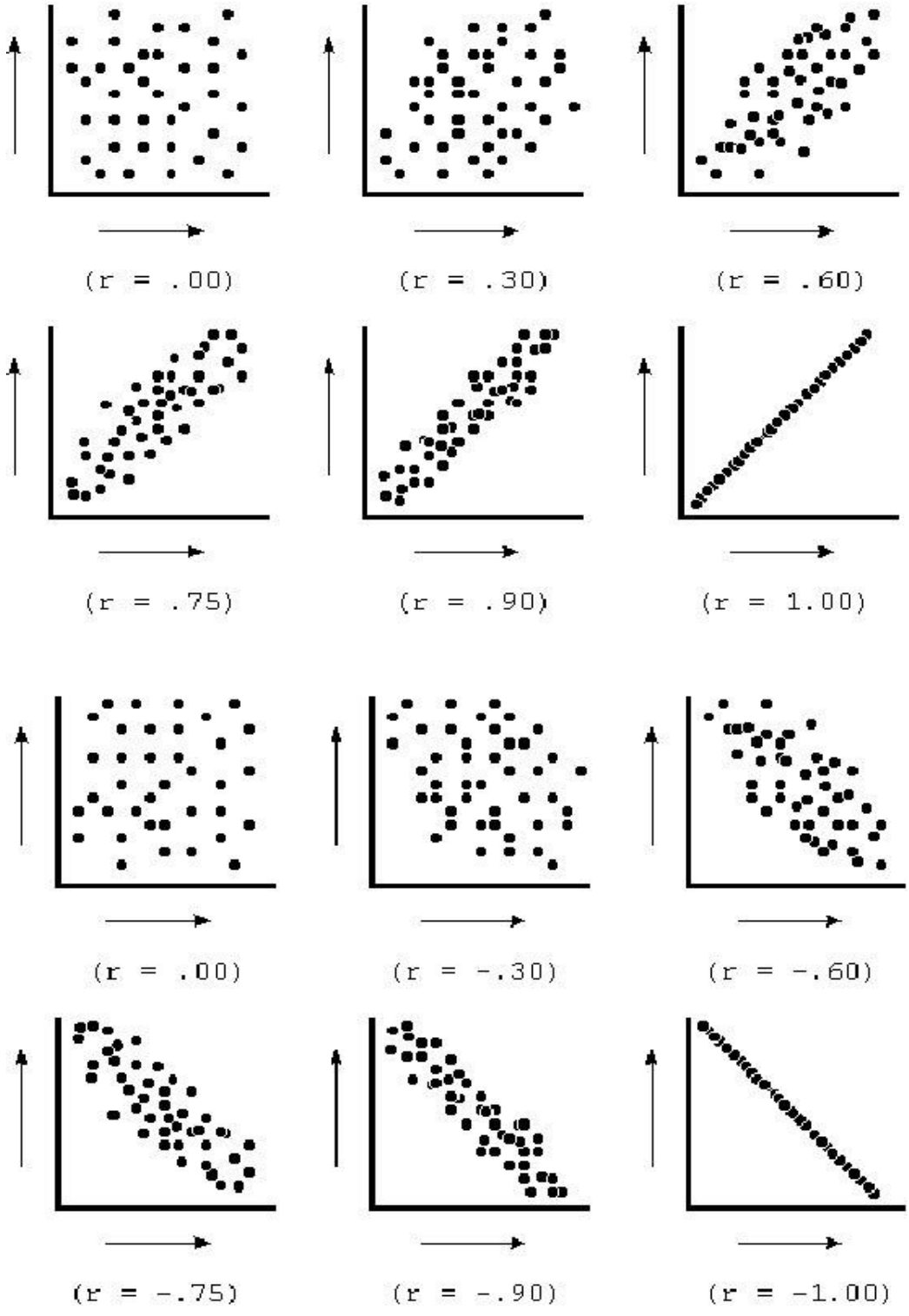
If r = 0, then that means no evidence of a linear relationship, but we cannot rule out a non-linear relationship. So, in the example of age and dexterity, Pearson's r could not provide a good measure of the relationship between age and dexterity since they form a non-linear association. We would have to use some other measure of association besides "r" to measure it.

Also note that "no relationship" is not the same thing as non-linear -- non-linear means there is a relationship, it just does not follow a simple line.

Range of r from -1.00 to 1.00

Here's a graph showing multiple scatterplots with varying degrees of strength:

http://www.bwgriffin.com/gsu/courses/edur7130/images/twelve_scatterplots.jpg



Question 1 About Pearson r

Which of these Pearson r correlations represent the stronger relationship?

-.76, .57, .80, .00, -.88?

Answer

r = -.88 is the stronger relationship. Recall that the correlation coefficient, r, ranges from -1 to 1 and these two extremes represent the same degree of relationship, but only in different directions. So an -.88 is just as strong a relationship as an .88, but the -.88 indicates a negative relationship and the .88 indicates a positive relationship. So, -.88 is a stronger relationship than .80, and whether the correlation coefficient is negative or positive does not matter when determining strength of association.

Question 2 About Pearson r

Which Pearson r correlation represents the stronger relationship?

.3, .5, .8?

r = .8.

Question 3 About Pearson r

Which correlation represents the stronger relationship?

-.3, -.5, -.8?

r = -.8.

Question 4 About Pearson r

Which correlation represents the weakest value possible with Pearson's r (and also the weakest relationship that can be calculated with r)?

.18, .89. -.53, .00, -.13?

r = 0.00

It may seem that 0.00 represents no linear relationship, so it cannot be the weakest linear relationship, rather some value such as -.13 would be weakest of those listed, however, keep in mind that the correlation coefficient is a mathematical index in which the weakest value is 0.00, so that value represents, mathematically, the weakest linear relationship possible. In short, the correct response to which is the weakest relationship possible as indicated by Pearson's r, .00 is the answer.

Question 5 about Pearson's r

For the questions that follow, we will use a table taken from the following linked study.

http://www.bwgriffin.com/gsu/courses/edur7130/content/correlation_ex1_regulation_learning.pdf

In this study the authors examined self-regulated learning behavior and achievement for an undergraduate class which contained some on-line learning applications. We will consider four variables they use:

- *Help Seeking*, which is defined as the ability to know when help is needed and therefore sought
- *Final Grades* -- percentage correct score on the final test in the course;
- *Self-efficacy for learning and performance*, one's belief in one's ability and skills to perform learning tasks;
- *Verbal Ability* which was measured by student performance on a 50-item verbal IQ scale.

Table 2. Pearson Product-Moment Correlations of the Variables. N = 94

	1	2	3	4	5	6	7
1. Intrinsic Goal Orientation	-	.467**	.314**	.118	.009	.072	.157
2. Self-Efficacy for Learning and Performance	.467**	-	.324**	.044	.128	.150	.291**
3. Time and Study Environment	.314**	324**	-	.052	.012	-.045	.146
4. Help Seeking	.118	.044	.052	-	-.086	-.247*	-.102
5. Internet Self-Efficacy	.009	.128	.012	-.086	-	-.043	-.089
6. Verbal Ability	.072	.150	-.045	-.247*	.043	-	.264*
7. Final Grades	.157	.291**	.146	-.102	-.089	.264*	-

Note. *Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Reading the table

1. The first column of this table presents what?

   Variable names

2. What do the first row labels 1, 2, 3, 4, 5, 6, and 7 represent?

   Those numbers correspond to the variables located in the first column. Often those numbers are used to help save space so a large table of correlations can fit more easily on one page; use of numbers as identifying labels is shorter than using variable names.

3. What information is presented in column labeled 1?

   Correlations between variable 1 (Intrinsic Goal Orientation) and all other variables.

4. What is the correlation between variable 1 (Intrinsic Goal Orientation) and variable 3 (Time and Study Environment)?

   r = .314

5. What is the correlation between variables 2 (Self-efficacy for learning and performance) and 4 (Help seeking)?
   r = .044

6. What is the correlation between ***Internet*** Self-efficacy and Verbal Ability?

   r = .043 (or -.043). Note that the two correlations presented for these two variables should be identical, so the authors committed a reporting error in this table by confusing the signs of the correlation.

7. Which two variables correlate most strongly with students' Final Grades in the course, and what are those respective correlation values?

Final Grades correlate .291 with Self-efficacy for learning and performance (variable 2), and Final Grades correlate .264 with Verbal Ability (variable 6).

8. The correlation between Final Grades and Self-efficacy for Learning and Performance is r = .291 --- interpret this correlation.

   The r = .291 indicates that individuals with higher Self-efficacy for learning scores also tend to have higher Final Grades, so positive relation.

9. What is the correlation between Verbal Ability and one's willingness to seek help in the course, and what does this correlation indicate (how is it interpreted)?

   The correlation between Verbal Ability and Help Seeking is -.247 --- those with higher Verbal Ability scores tend to seek less help than those with lower Verbal Ability scores. So as one's Verbal Ability increases, Help Seeking tends to decline.

10. How do we determine, in this table, which correlations are statistically significant?

    The "Note" below the table indicates that correlations denoted with * are significant at the .05 level, and correlations with ** are significant at the .01 level.

11. Which variables demonstrated correlations with Final Grades that were statistically significant?

    Yes, Final Grades had statistically significant correlations with Self-efficacy for learning and performance (variable 2), and with Verbal Ability (variable 6).

12. What does it mean for a correlation to be statistically significant?

    Correct, if a correlation is denoted to be statistically significant, this means the null hypothesis for that correlation is rejected.

13. What is the null hypothesis regarding the correlation between Verbal Ability and Final Grades?

    Null: There is no relationship between Verbal Ability and Final Grades.

14. If this null hypothesis is rejected, as it was for this study, then what conclusion do we draw?

    If the null hypothesis is rejected, then we conclude that Verbal Ability and Final Grades are related. In this case, since we know the correlation is positive (was r = .264), we can conclude that Verbal Ability and Final Grades are positively correlated, that is, the greater one's Verbal Ability, the higher will be one's Final Grade.

15. If the other correlations with Final Grades are not statistical significant, what does that mean?

    Those correlations that are not statistically significant are consistent with the null hypothesis, so our best interpretation for those are "no relationship." For example, the correlation between Final Grades and Help Seeking is r = -.102. Since this correlation is not statistically significant, we can state that our data suggest that Help Seeking is not related to Final Grades.

Independent Samples t-test

Information about t-test:

- The t-test is used when the IV is qualitative and the DV is quantitative.
- A t-test is used to compare 2 groups (the IV) on some quantitative dependent variable.
- Example 1: We will find a difference in test scores between males and females, the IV = student sex, DV = test scores.
- Example 2: Ms. Smith's class will have more disciplinary infractions per student than Ms. White's class. IV = teacher (Smith vs. White) and DV = number of disciplinary infractions per student

Reading t-test Table

Table 7

Results of t-tests and Descriptive Statistics Syst. Blood Pressure, SAT Verbal, SAT Math, and GPA by Sex

Outcome	Sex						95% CI for Mean Difference		
	Male			Female					
	M	SD	n	M	SD	n		t	df
Sys. Blood P.	130.1	20.5	45	102.7	22.2	44	2.3, 52.2	2.38*	87
SAT-Verbal	463.8	98.8	45	532.2	101.2	44	-110.5, -26.2	-3.22*	87
SAT-Math	515.4	99.5	44	483.3	98.9	44	-9.95, 74.2	1.52	86
College GPA	2.7	1.3	45	3.16	1.1	44	-0.97, 0.07	-1.71	87

Note: Sys. Blood P. = Systolic blood pressure. SAT = Scholastic Achievement Test.
* p < .05.

1. What are the IV and DV in the above table?

IV = sex (categories are male and female)
DV =

- Systolic blood pressure
- SAT verbal scroes
- SAT math scores
- College GPA

2. What do M, SD, n, and t represent in Table 7?

M = mean score for males and females for each DV
SD = standard deviation for each DV by sex
N = sample size for each sex
t = calculated value of the t-ratio from a t-test.

3. Who had the higher blood pressure? Is this blood pressure difference statistically significant? How do we know this?

Males higher (130.1 vs. 102.7)

The difference between these blood pressure means is statistically significant (t = 2.38*), we know this because the author marked the t with an asterisks (*). This is common practice – denote statistically significant results with *.

5. For which other DVs in that table do we find significant differences between females and males?

Blood pressure and SAT verbal are the DVs that show significant mean differences between the sexes.

6. What level of probability for making a Type 1 error did the author of this table use for hypothesis testing?

The footnote of the table shows this: * p < .05. This means that the null hypothesis was tested against an alpha of .05 (probability of making a Type 1 error was .05). IF an * is present inside the table, that means that particular test was significant, and if the * is missing, that means that particular t-value was not significant so the means do not appear to be different, statistically.

7. What does it mean, for this study, for there to be a statistically significant difference? What does it mean for there not to be a statistically significant difference?

If the null hypothesis is rejected (results are significant), it means that we believe there to be differences in a given DV between the sexes. If the null is not rejected (not significant), it means we believe the means for that particular DV to be the same or similar between the sexes.

8. For SAT-math and College GPA the results were not significant. Note that males scored M = 515.4 and females scored M = 483.3 on the SAT-math. How do we explain this mean difference of 515.4 – 483.3 = 32.1 if it is not significant? How do we justify claiming the null hypothesis of no difference is not refuted by the data if we can see a 32 difference?

Failure to reject Ho in this case means we don't have enough information in the data to draw a conclusion that a mean difference exists, and any difference we observe could be due to random chance as a result of sampling variation.

ANOVA = Analysis of Variance

What is the distinction between ANOVA and t-test?

- Conceptually ANOVA and t-test are nearly the same
- Mathematically if comparing 2 groups, ANOVA and t-test produce exactly the same results
- ANOVA used to compare 2 or more groups on some quantitative dependent variable.
- The t-test can be used to compare only 2 groups at a time.
- Examples
    - Compare math scores between boys and girls could be analyzed with both t-test and ANOVA.
    - Similarly, comparing the number of skips of rope per minute between boys and girls could also be analyzed with t-test or ANOVA.
    - However, if we wished to compare three auto companies (e.g., Toyota, Honda, Ford) on their MPG for small cars, then we would need to use ANOVA since there are more than 2 groups to compare.
- While conceptually ANOVA and t-test are nearly the same, the test statistics used do differ, t-test forms a t-ratio and ANOVA uses an F-ratio for testing comparisons

Published Study Example 1:

Kuyath, Stephen J., and Susan J. Winter. "Distance education communications: The social presence and media richness of instant messaging." Journal of Asynchronous Learning Networks 10.4 (2006): 67-81.

http://www.bwgriffin.com/gsu/courses/edur7130/content/anova_ex1_social_presence_instant_message.pdf

These researchers examined communication methods in terms of social presence and media richness.

- Social presence has been defined as “the extent to which one feels the presence of a person with whom one is interacting”, and
- Media richness represents “the capacity to allow quick feedback, provide multiple cues, use language variety, and the perception of personal focus,” thus enhancing person communication.

To measure Perception of Social Presence, the authors used responses (ratings) to these three items (see Section D of the study):

(These are three DVs, each examined separately)

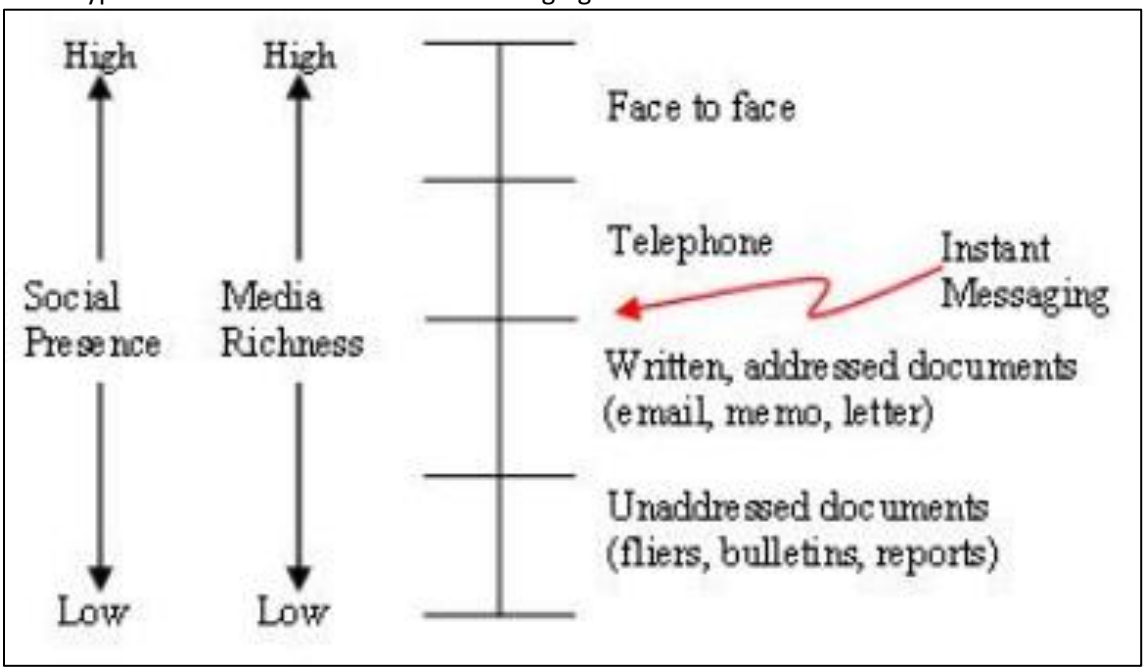
1. Ability to perceive the presence of the person with whom you are communicating;
2. Perception that the person you are communicating with is engaged in the conversation; and
3. Privacy of the conversation.

The authors explained their scoring:

> “All measures consisted of a single item rated on a Likert scale of 0 to 4, with 0 indicating that the media was ineffective and 4 indicating that the media was very effective.”

Note Figure 2 (page 6 of the PDF file) which shows where they hypothesize instant messaging (IM) to lie on the scale -- between what and what?

Figure 2: Hypothesized Location of Instant Messaging on the Social Presence and Media Richness Scales



In terms of social presence, they expect the following ratings:

Top =	face-to-face
	Telephone
	Instant Messaging
	Written, Addressed documents (email, memo, letter)
Bottom =	Unaddressed Documents (fliers, bulletins, reports)

Below is a table of their analysis results

Table 1

Levels of Equivocality	Mean	Std. Dev.		Effect Size
Perception of Presence				
Email	1.64	1.12		
IM	2.47	0.92		d = .81
Phone	3.11	1.15		d = 1.22
Omnibus Test of Differences Among Means			$F_{(1.7, 60.1)} = 29.51^{***}$	Eta-squared = .37
Conversation Engagement				
Email	1.53	1.15		
IM	2.54	0.99		d = .94
Phone	3.08	1.15		d = .51
Omnibus Test of Differences Among Means			$F_{(1.7, 50.9)} 35.24^{***}$	Eta-squared = .41
Perception of Privacy				
Email	1.92	1.33		
IM	2.50	1.14		d = .47
Phone	2.53	1.33		d = .02
Omnibus Test of Differences Among Means			$F_{(1.8, 62)} 6.40^{**}$	Eta-squared = .17

N = 72 for all measures $^{*}p \leq .05$ $^{**}p \leq .01$ $^{***}p \leq .001$

1. Based upon Table 1, what is the IV in this study, and what are the categories examined?
IV = communication medium and it contains three categories, which are email, IM, and phone.

2. How many DV are presented in Table 1, and what are they?
DVs are

- "Perception of Presence" (row 2),
- "Conversation Engagement" (row 9), and lastly
- "Perception of Privacy" (row 16).

3. Let's now focus on "Perception of Presence." Which medium of communication was rated by the study participants to have the highest level of presence?
Phone

4. How do we know phone was rated to have the highest level of presence?
Mean level of "presence" was 3.11 for phone, 2.47 for IM, and 1.64 for email.

5. What is the highest possible score for this variable?
Recall I posted this information earlier: "All measures consisted of a single item rated on a Likert scale of 0 to 4, with 0 indicating that the media was ineffective and 4 indicating that the media was very effective." So 4 is the highest score possible.

6. Which mode of communication was rated highest in terms of "Conversation Engagement?"
Phone with a mean of 3.08.

7. In terms of Privacy, which were rated to be most private?
Phone and IM are rated very similarly, and email rated as low in privacy.

8. Are there statistically significant mean differences for any of the DVs examined? How do we know?
Column 4 presents the ANOVA results. ANOVA uses something called an F-test to determine if mean responses are statistically different.

9. For "Perception of Presence" did the F-test indicate a statistically significant mean difference among the communication modes?
From the other two tables we learned that if something is statistically significant, they are denoted by use of *. In this table the F ratio for "Perception of Presence" is F = 29.51***. This indicates it is statistically significant at the .001 level (see footnote at bottom of table).

10. The null hypothesis for "Perception of Presence" is rejected. What is the null hypothesis here (don't forget that if the IV is qualitative, as it is here, focus wording of hypothesis on group difference)?
Null: There will be no difference in perceptions of presence among users of email, IM, and phone.

11. The null hypothesis was rejected for Perception of Presence. Was the authors' hypothesis presented in Figure 2 (page 6 of the PDF file) supported for this DV?
According to Figure 2, people should rate "presence" highest for phone, next highest for IM, and lowest for email when comparing these three modes of communication. So, given the means reported in Table 1 for "Perception of Presence" it does appear that the authors' hypothesis was supported since the means follow that pattern (lowest for email, next for IM, and higest for phone).

12. What was the order of results for "Conversation Engagement"? Does this order of results support that predicted in Figure 2? Also, was there a statistically significant mean difference for Conversation Engagement?
Yes results support hypothesized order; results from "Conversation Engagement" do support the order predicted: Email (M = 1.53), then IM (M = 2.54), then Phone (M = 3.08).

Yes results are statistically significant; the F ratio was 35.24 which the authors denote as significant at the .001 level given the three asterisks (***).

13. Why use the ANOVA here instead of the t-test?
Since there are three groups to compare (email, IM, and phone), the ANOVA is required to compare the quantitative DVs. Recall the t-test is used when one compares only two groups.

Published Study Example 2:

Espelage, D. L., Aragon, S. R., Birkett, M., & Koenig, B. W. (2008). Homophobic teasing, psychological outcomes, and sexual orientation among high school students: What influence do parents and schools have? School psychology review, 37(2), 202-216.

Parental Communication: Responses to four items - "how often participants talked with at least one parent about sex, drugs, personal issues, and their future...." Responses ranged from 0 = never to 4 = very often.

Parental Support: Two "item scale asks participants how much they feel that their parents care about them and are there when they need them...." Responses ranged from 0 = never to 4 = very often.

Note: See video for this presentation to hear discussion of these tabled results.

Table 1
Sexual Orientation Group Differences on Study Measures

	Heterosexual (N = 11,924)		Questioning (N = 932)		LGB (N = 1065)		ANOVA	
	M	SD	M	SD	M	SD	F	η^2
Homophobic teasing	0.20	0.66	0.84	1.33	0.57	1.13	375.94*	.05
Peer victimization	0.45	0.75	0.95	1.18	0.56	0.90	166.54*	.03
Depression–suicidal ideation	0.63	0.67	1.07	0.95	0.77	0.82	176.48*	.03
Alcohol–marijuana	0.80	0.97	1.36	1.51	1.00	1.16	138.82*	.02
School climate	1.79	0.49	1.63	0.65	1.72	0.56	49.13*	.01
Racism	0.61	0.67	1.03	0.82	0.82	0.76	193.31*	.03
Parent communication	1.89	0.95	1.79	1.13	1.84	1.07	5.63*	.00
Parent support	3.31	0.65	2.83	0.93	3.14	0.80	231.73*	.03

Note. LGB = gay, lesbian, bisexual; ANOVA = analysis of variance.
* $p < .01$.

ANCOVA = Analysis of Covariance

ANCOVA -- how does this differ from ANOVA?

- Both ANOVA and ANCOVA are used to compare groups on a DV that is quantitative, and both use similar tests (F-ratios)
- ANCOVA incorporates a covariate (quantitative IV)
- ANCOVA helps create statistical control due to use of covariates
- Statistical control can be used to help equate groups.

In research when making comparisons it is important to have groups that are as similar as possible to give validity to comparisons and research findings.

ANCOVA provides some control that helps create this equality among groups on whichever variables are controlled in ANCOVA. ANCOVA provides a statistical means of equating groups that are not equated, equivalent, by design.

ANOVA vs. ANCOVA

ANOVA

Groups	~~Mean IQ (Covariate)~~	Observed Posttest Achievement (DV Mean)
Experimental	~~106~~	85
Control	~~98~~	75

The covariate is not incorporated in ANOVA, so it is ignored.

What caused the 10-point difference in achievement between the experimental and control groups? Was it the treatment?

If both groups had an IQ of 100, could IQ be used to explain the 10-point difference?

Hypothetical ANCOVA Adjustment

Groups	Mean IQ (Covariate)	Observed Posttest Achievement (DV Mean)	Adjusted Posttest Achievement (Estimated DV Mean)
Experimental	106	85	82
Control	98	75	78

Because of the ANCOVA adjustments and statistical control, allows us to answer this question: "What would be the predicted achievement means if both groups had the same level of IQ?"

ANCOVA

- Adjusts scores on DV to take into account the imbalanced performance between experimental and control groups on the covariate (IQ). Results, hopefully, provide more realistic picture of performance on the DV.
- Group that starts low on covariate will be adjusted up on the DV, and group that starts high on covariate will be adjusted down on the DV to compensate for differences on the covariate.

If the groups have covariate means that are identical, which direction will any adjustments take?

Hypothetical ANCOVA Adjustment

Groups	Mean IQ (Covariate)	Observed Posttest Achievement (DV Mean)	Adjusted Posttest Achievement (Estimated DV Mean)
Experimental	100	85	85
Control	100	75	75

If the groups are equal on the covariate, there will be no adjustment made on the dependent variable since the groups are already equal.

Why do we use ANCOVA at all?

In research we try to determine causal relationships and ANCOVA helps provide a more realistic assessment of those potential causal relationships.

Example:

> We may be interested in learning whether students using cooperative learning achieve at a higher level than students in lecture classes. In order to make accurate and informative comparisons, it is extremely important that both groups of students be as equal, equivalent, as possible on all variables that predict achievement except for the sole treatment we are studying (cooperative learning vs. lecture). If our two groups had differing levels of intelligence, on average, then we could not explain why achievement scores differ between the two groups until intelligence is controlled (made equal between the two groups). ANCOVA is useful since it statistically controls groups; if groups differ on a covariate, such as intelligence, then ANCOVA will adjust the dependent variable means (achievement levels) to compensate for the group differences in intelligence.

Published Study Example 1:

Bilgin, I., Karakuyu, Y., & Ay, Y. (2015). The effects of project based learning on undergraduate students' achievement and self-efficacy beliefs towards science teaching. Eurasia Journal of Mathematics, Science & Technology Education, 11(3), 469-477.

SEB = self-efficacy beliefs
European number = 70,757 = 70.757 = American number
European number = 1.116,35 = 1,116.35 = American number
European format also used in South American, most of African, and parts of SE Asia

Note: See video for this presentation to hear discussion of these tabled results.

Table 1. The mean and standard deviation for pre-SEBS and post-SEBS

Groups	Dependent Variables	n	Mean	Standard Deviation
Project-based	Pre- SEBS	33	70,757	7,504
Learning	Post- SEBS	33	72,969	9,040
Traditional Method	Pre- SEBS	33	67,757	7,504
	Post- SEBS	33	65,757	7,927

Table 2. Summary of ANCOVA Comparing the Mean Post-SEBS Scores of Students in Treatment and Control Groups

Source Dependent Variable		df	Means Square	F	P	Partial Eta Square
Pre- SEBS	Post- SEBS	1, 63	1116,35	19,808	0,000*	0,239
Group	Post- SEBS	1, 63	485,14	8,608	0,005*	0,120

*n=66, *p<0,05*

Table 3. The Comparison of the Results Obtained in the post-STTAT

Variable	X	SD	t	df	P
Post-STTAT					
Group 1(PBL)	16,969	3,44			
Group 2 (TT)			2,08	64	0,042*
	15,091	3,88			

*n= 66; *P< 0,05*

Published Study Example 2:

Moh'd Al-Migdady, A., & Qatatsheh, F. (2017). The effect of using Crocodile mathematics software on Van Hiele level of geometric thinking and motivation among ninth-grade students in Jordan. INSTRUCTIONAL TECHNOLOGY, 27.

CMG = Crocodile Mathematics Group
NCMG = Non-Crocodile Mathematics Group
9th grade students in Jordan

Note: See video for this presentation to hear discussion of these tabled results.

Table 4

Two way analysis of covariance summary table of students' motivation to learn Mathematics

Source of variation	Sum of Squares	Degrees of Freedom	Mean of Squares	F	P Value	η2
Covariate	11142.437	1	11142.437	56.90	0.00	0.3424
Group	1635.999	1	1635.999	8.354	0.005	0.0502
Gender	64.540	1	64.540	0.330	0.568	0.00198
Group× gender	401.353	1	401.353	2.050	0.156	0.01233
Errors	14686.816	75	195.824			0.4513
Overall	13244.329	79				

*P< 0.05

Table 3

Descriptive statistics for the pre-and the post-tests of students' motivation to learn mathematics

			The Pretest		The Posttest		The Adjusted Mean of the Posttest
Group	Gender	Number	Mean	Standard Deviation	Mean	Standard Deviation	
CMG	Male	19	115	14.04	128	22.11	133.075
	Female	21	132.57	11.79	143.9	17.86	131.127
	Total	40	124.22	15.33	136.9	21.30	131.577
NCMG	Male	19	116.11	19.50	124.21	19.61	127.47
	Female	21	119.52	18.11	120.29	13.67	126.519
	Total	40	117.9	18.62	122.15	16.65	126.97

Note. The maximum possible score = 210.

As noted in Table 3, the CMG had a higher mean posttest score than the NCMG. In order to test whether this difference was significant, a two way analysis of covariance (Two Way ANCOVA) was used. Table 4 summarizes the results of Two Way ANCOVA for motivation to learn mathematics of 9th grade students.

Chi-square Test of Association, Relation between Qualitative Variables

Lastly, when does one use chi-square?

Chi-square is appropriate when both the IV and DV are qualitative.

Examples:

- Sex and pass test (sex is qualitative – female vs. male) and pass test is qualitative (pass or fail).
- Sex and type of car owned (Honda, Ford, Toyota, etc.).

Published Example 1:

White, P., Redford, P., & Macdonald, J. (2019). An example motivated discourse of the chi-squared test of association (2 by 2).

Note: See video for this presentation to hear discussion of these tabled results.

Table 1

	Breast Injury	
	No	Yes
Breast Cancer	32	35
No Breast Cancer	108	16

Table 2a

			Cancer Status		
			Breast cancer	No breast cancer	Total
Injury Status	No Breast Injury	Count	32	108	140
		Percentage	22.9%	77.1%	100.0%
	Breast Injury	Count	35	16	51
		Percentage	68.6%	31.4%	100.0%
Total		Count	67	124	191
		Percentage	35.1%	64.9%	100.0%

Table 2b Chi-Square Tests

	Value	df	Sig. (2-sided)
Pearson Chi-Square	34.388[a]	1	.000

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.89.

Published Example 2:

Salem, A., Hillow, H., Khraisat, A., Smadi, L., & Ryalat, S. (2008). Association between intensity of smoking and periodontal pockets among young university students. Tropical Dental Journal, 31(122), 5.

Note: See video for this presentation to hear discussion of these tabled results.

Table 2 : Chi-square test of association between cigarette smoking and presence/absence of periodontal pockets

	Smoker N = 204 (57.1%)	Non-smoker N = 153 (42.9%)	P value
Pockets N = 100 (28.0%)	82 (40.2%)	18 (11.8%)	<0.0001
No pockets n = 257 (72,0%)	122 (59,8%)	135 (88,2%)	

Practice Test Questions for Identifying Type of Statistical Procedures

Q1. Which statistical test would you use for this example: Students exposed to cooperative learning will score higher on an achievement test than students exposed to lecture?

There are only two variables, IV is type of instruction (lect vs coop) and DV is test scores. So IV is qualitative, and DV is quantitative. So yes, t-test or ANOVA (not ANCOVA) would be ok.

Q2. Students with higher levels of IQ will score higher on a timed test of reading comprehension.

Two variables, IQ and then reading comprehension test scores. These are likely to be quantitative, so correlation (Pearson's r) most appropriate.

Q3. Which group of students achieves highest in mathematics, those with supplemental computer instruction or those with more practice time in class? The outcome of interest is scores on a mathematics achievement test. Students' general understanding of mathematics before entering the class, as measured by a course entry screening test, will be considered to judge similarity of treatment groups.

This is ANCOVA. The IV is qualitative (type of instruction, two groups computer use vs. more practice time), the DV is mathematics achievement, and the covariate will be screening test scores taken at the outset of the study.

Q4. Students with higher levels of test anxiety will score lower on an achievement test. Test anxiety is measured on a scale ranging 10 to 50.

To help visualize how test anxiety could be measured, here's a couple of simple measures of test anxiety:

http://www.mathpower.com/anxtest.htm

Since both variables are quantitative, this is correlation (Pearson's r).

Q5. Students who enter educational research with prior research experience will demonstrate greater achievement than students who enter educational research without research experience.

Note that no covariate is mentioned in the example, so ANCOVA is not possible. The IV is prior experience (yes/no, which makes this variable qualitative) and the DV is achievement level. ANOVA or t-test would work here.

Q6. Students who enter educational research with prior research experience will demonstrate greater achievement than students who enter educational research without research experience once differences in IQ are controlled.

ANCOVA.

Q7. Students exposed to the read, visualize, and draw condition are expected to comprehend more of the biology text, as measured by an achievement test, than are students in the read and visualize condition or the read only condition.

ANOVA is the only method for this example. The IV is qualitative with three groups, and the DV is quantitative (level of comprehension). The t-test will not work here since there are more than two groups to compare (the groups are (a) read, visualize, and draw, (b) read and visualize, and (c) read only). No covariate mentioned, so ANCOVA cannot be used.

Q8. Students exposed to the read, visualize, and draw condition are expected to comprehend more of the biology text, as measured by an achievement test, than are students in the read and visualize condition or the read only condition. On a pre-measure of biology taken before the start of the experiment, students in the three groups showed different average levels of knowledge.

The IV is condition ([a] read/vis/draw, [b] read/vis, and [c] read only). The DV is comprehension. The control variable ---- the covariate ---- will be the scores from the pre-measure of their biology knowledge. So ANCOVA is appropriate with the pre-measure of knowledge serving as the covariate.

If a pre-measure of some sort is used in a study, such as measuring pre-experimental levels of knowledge, anxiety, self-efficacy, interest, or whatever, often those pre-measures are used as covariates in study analyses. So, if groups are compared on some quantitative DV, and a pre-measure of some sort is present, ANCOVA can be used.