

Final Exam – Practice problems

1. True or **False**: ANOVA is a statistical measure adopted to analyze differences in group variances.
2. True or **False**: If you do not have a significant interaction, you would still want to conduct post-hoc tests such as the Tukey and Bonferroni to assess hypothesized pairwise comparisons.
3. True or **False**: The only difference between a Tukey and Bonferroni is that a Tukey is divided by the number of comparisons.
4. **True** or False: ANOVA is robust with violations to assumptions of homogeneity of variance so if you fail Levene's test by a small amount, you're still ok.
5. True or **False**: A significant difference among groups is considered large and important.
6. **True** or False: A 3×3 between-subjects design is a 2-way ANOVA.
7. Which of the following is (are) assumptions of the between groups F-test?
 - 1) the sampling distributions are normally distributed
 - 2) the variances of the raw score populations are the same
 - 3) subject responses should be independent
 - 4) a and b
 - 5) **all of the above**
8. True or **False**: A "main effect" is the direct effect of an independent variable on one level of another independent variable
9. Which of the following is true about the F ratio?
 - 1) it has no negative values
 - 2) it is positively skewed
 - 3) the mean of the F distribution equals zero
 - 4) it must be greater than 1 to determine if there is a significant interaction
 - 5) **a, b, and d**
 - 6) a and b only
10. Which of the following is NOT an assumption required to perform a one-way between groups Analysis of Variance?
 - 1) Independence of Observations
 - 2) Homogeneity of Variance
 - 3) **Absence of Multicollinearity**
 - 4) Normality of the Sampling Distribution
 - 5) These are all assumptions which must be met in order to run a one-way ANOVA

11. A simple effect is one that:
- 1) **Determines the effect of one independent variable at one level of another independent variable**
 - 2) Determines the effect of independent variable on another independent variable
 - 3) Determines the effect of one dependent variable on one level of an independent variable
 - 4) Determines the effect of one independent variable on one dependent variable
 - 5) None of the above
12. True or **False**: In order for an interaction to occur the plotted lines must always be cross.
13. Which of the following statements about Omega squared is NOT true?
- 1) It is less biased than Eta Squared
 - 2) Both the numerator and denominator are adjusted for the within groups variance
 - 3) It includes more information than Eta Squared
 - 4) **It is larger than Eta Squared**
 - 5) It is a better estimate of the population parameter than Eta squared
14. What does Eta squared tell you?
- 1) The proportion of the pooled standard deviation compared to the grand mean
 - 2) **The proportion of variance which is shared between the independent variable and the dependent variable**
 - 3) The proportion of variance which is not shared between the independent variable and the dependent variable
 - 4) It doesn't tell me anything. I just work here
15. Which of the following is the simplest factorial design?
- 1) A study with two independent variables which have 1 level each.
 - 2) A study with one independent variable with 2 levels.
 - 3) **A study with two independent variables, each containing 2 levels.**
 - 4) A study with two continuous dependent variables.
 - 5) Anyone who has done a factorial ANOVA by hand knows that no factorial is simple!

16. A researcher performed a study where people watched either Old School, Airplane, or Corky Romano, and rate (on a scale of 1-10) how funny it was. The data follow:

	Old School	Airplane	Corky Romano
1	7	9	3
2	6	10	5
3	8	6	2
4	4	7	7
5	6	7	3
Mean	6.2	7.8	4
Std Dev	1.4832397	1.643168	2

- a. Assuming these observations are independent (meaning that there were a total of 15 people in the study), run a one way analysis of variance on this data.

SSbetween= 36.4 dfbetween=2 MSbetween= 18.2 F= 6.135
SSwithin= 35.6 dfwithin= 12 MSwithin= 2.97 Fcrit= 3.885
SStotal= 72 dftotal= 14

- b. Compute η^2 and ω^2 for this data.

$\eta^2 = 36.4/72 = .506$ $\omega^2 = 36.4 - (2 * 2.97)/ 72 + 2.97 = .406$

- c. Using individual comparisons, where do the differences truly lie?

Old School vs. Airplane = -1.62 No Old school vs. Corky = 1.98 No
(Tcrit = 2.306) Airplane vs. Corky = 3.28 Yes

- d. Why might it be troublesome to run an Analysis of Variance for these data?

The normality assumption has not been met. There are only 5 people per cell.
Conclusions about this F should not technically be made.

17. What are the 3 sources of variability in a within subjects anova?

Subjects, between groups and within groups

18. Consider the following summary ANOVA table for a between subjects experiment with 25 total subjects.

	SS	Df	MS	F
Between	?	4	?	?
Within	?	20	5	
Total	120	?		

What can we conclude about the experiment?

ss-within = 20X5 =100

ss-between must be 20

ms between must be 20/4 = 5;

f ratio = 5/5 = 1.

Null hypothesis cannot be rejected

19. I am testing the effects of vitamins on IQ in a between subjects experiment with two subjects in each group– here is the data.

	K	A	E	
	100	102	104	
	102	104	106	
	----	----	----	
Mean	101	103	105	103 = grand mean

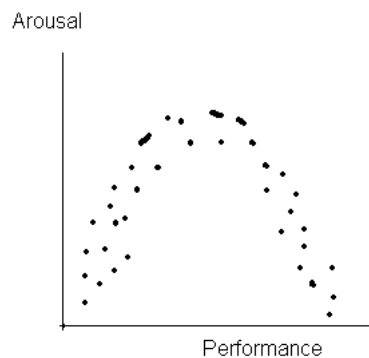
What is the MS_{between} ?

$$MS_{BG} = \frac{\sum n_j (\bar{X}_{.j} - \bar{X}_{..})^2}{k - 1}$$

$$MS_{BG} = \frac{2*[(101-103)^2 + (103-103)^2 + (105-103)^2]}{3-1}$$

$$MS_{BG} = \frac{2*[4 + 0 + 4]}{2} = \frac{16}{2} = 8$$

20. The following graph describes the results of an experiment linking performance under stress to the level of anxiety experienced



If we calculated a correlation coefficient for this data:

The correlation would be close to 0 because there is no linear relationship

21. In the result-section of a paper a researcher claims that knowing a persons height reduces the amount of variance in the prediction of this persons weight by 36%. Based on this information, what was the correlation between height and weight in this particular sample?

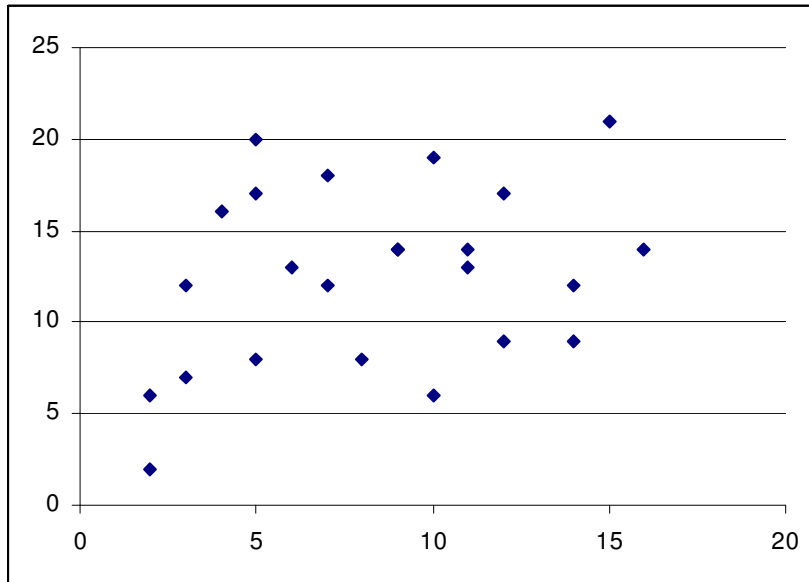
r=.6

22. X=Maximum positive correlation possible. Y= Maximum negative correlation possible.
What is X+Y?

0

23. Weight and IQ are not correlated. Weight has a mean of 150 and a standard deviation of 30 pounds. My score is from a sample with 4 subjects. If I have a Z-score of 1.0 on my IQ, what is your best guess of my weight?

150 pounds



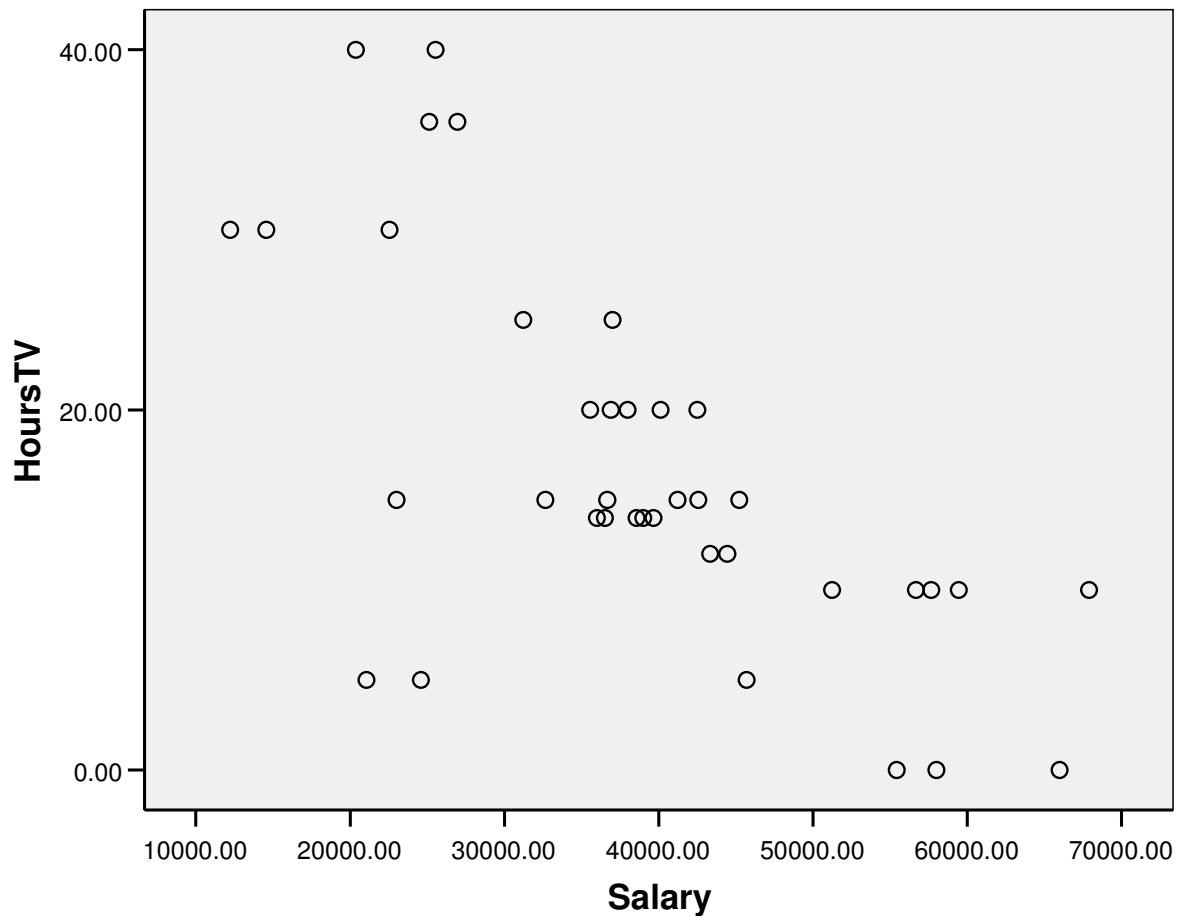
24. Based on the scatterplot above, if I add a score that has $X = 100000$ and $Y = 100000$, what will happen to the R^2 ?

It will get very close to 1

25. Of these three correlation coefficients, $r = .650$, $r = .300$, and $r = -.700$, which is the strongest relationship?
- .650
 - .300
 - .700**
 - Not enough information is given here to determine the strength of these relationships.
26. The following scatterplot shows the relationship between Hours of Television Watched per week (HourTV) and Annual Salary (Salary). What kind of relationship is this?
- A direct relationship
 - An Inverse Relationship**
 - A positive correlation
 - A long distance relationship

Graph

[DataSet0]



27. At the .05 level, what is the minimum number of subjects (n) needed for a correlation of .500 to be considered significant?
- 12
 - 14
 - 16 (.500 is significant for 14 df. Df are determined by N-2, so 16 is the correct answer)**
 - 30
28. What is the difference between standard error and sampling error?
- Sampling error is the standard error squared.
 - There is no difference, they are both the same.
 - Standard error is the expected variability of a statistic over repeated sampling, while sampling error is the standard deviation of a sampling distribution.
 - Sampling error is the expected variability of a statistic over repeated sampling, while the standard error is the standard deviation of a sampling distribution.**

29. The following regression equation predicts a movie's gross profit (Y) as a function of how many big name stars appear in it (X).

$$Y' = 18976234.12(X) + 20225500.00$$

If a movie has 3 big name stars in it, how much should we expect it to make?

- \$56,928,702.36 ← results if they forget to add the intercept
- \$77,154,202.36**
- \$ -1.07 ← results if they plug 3 in for Y instead of X.
- Well, it really depends on WHICH stars, now doesn't it?

30. Short Answer:

If the correlation coefficient (r) for the relationship between Salary and Hours of TV watched per week is $r = -.680$, and the standard deviations for salary (Y) and hours of TV watched per week (X) are 13717.09 and 10.56 respectively, what would the slope (b) of the regression equation be if we wanted to predict Salary from Hours of TV watched per week?

First, take $-.680 = \text{Cov} / (13717.09 * 10.56) \rightarrow 144852.4704$. Then solve for Cov, multiplying 144852.4704 by both sides, and you have Cov = -98499.68

Now, you have all the info you need to find the Slope: $b = \text{Cov} / S^2x$

Soooo, $b = -98499.69 / 10.56^2$, which equals -883.297

31. Answer the questions below based on the following data set:

	X	Y			
	Hours of Sleep	Score	(X - Xbar)	(Y - Ybar)	(X - Xbar) * (Y - Ybar)
	5	82	0.80	11.00	8.8
	7	90	2.80	19.00	53.2
	3	56	-1.20	-15.00	18
	4	74	-0.20	3.00	-0.6
	2	53	-2.20	-18.00	39.6
Mean	4.20	71			
SD	1.924	16.125			
Sum					119

- What is the covariance between X and Y?
 $119/4 = 29.75$
- What is the correlation between X and Y?
 $29.75 / (1.924 * 16.125) = .959$ or roughly **.96**
- What is the slope for Score predicted by Hours of Sleep?
 $29.75 / 1.924^2 = 8.04$
- What is the slope for Hours of Sleep predicted by Score?
 $29.75 / 16.125^2 = .114$

32. Answer the questions below based on the following data set:

	X	Y			
	Hours of Study	Score	Zx	Zy	Zx*Zy
	23	82	1.055	0.682	0.7194
	22	90	0.919	1.178	1.0833
	6	56	-1.244	-0.930	1.1571
	15	74	-0.027	0.186	-0.0050
	10	53	-0.703	-1.116	0.7848
Mean	15.20	71			
SD	7.396	16.125			
Sum					3.7397

- What is the correlation between X and Y?
 $3.7397 / 4 = .935$ or roughly **.94**
- What is the slope for Score predicted by Hours of Study?

$$b = .94 * \left(\frac{16.125}{7.396} \right) = 2.05$$

- What is the slope for Hours of Study predicted by Score?

$$b = .94 * \left(\frac{7.396}{16.125} \right) = .43$$

33. For the following questions indicate the type of correlation you would use.

- If I wanted to correlate the place (1st, 2nd, etc.) that a set of runners finish in 2 separate races.
Spearman
- If I wanted to correlate gender with political party (i.e. Democrat or Republican).
Phi
- If I wanted to correlate gender with IQ.
Point Biserial
- If I wanted to correlate height and intelligence.
Pearson

34. If Hand Size (relative to the size of your face) and Incidence of Cancer are correlated at .74 and Hand Size and Cancer have SDs of 4.5 and 10.2 respectively, what is the standardized regression coefficient predicting Cancer from Hand Size?

.74

35. Answer the questions below based on the following data set:

	X	Y		
	Attractiveness	Dates	Y'	(Y - Y')^2
	5	4	4.779	0.6068
	7	7	5.515	2.2052
	2	3	3.675	0.4556
	5	6	4.779	1.4908
	9	5	6.251	1.5650
Mean	5.60	5		
SD	2.608	1.581		
CorrelationXY			0.606	
Sum				6.3235

- What are the slope and intercept values?
Slope = .368 and the intercept = 2.939
- What is the missing Y' value?
 $Y' = 2(.368) + 2.939 = 3.675$
- What is the sum of (Y-Y')^2?
Once you figure out the missing Y' value you can take $(3 - .675)^2$ which is .4556 and then add them all up to get 6.3235.
- Calculate the standard error of estimate in 2 different ways.

- 6.3235 is the sum of squared residuals so if we divide that by N-2 and take the square root of the whole thing we have the standard error of estimate.**

$$s_{Y-\hat{Y}} = \sqrt{\frac{6.3235}{5-2}} = 1.452$$

- Since we know the correlation and S_y we can use the $s_{Y-\hat{Y}} = s_y \sqrt{(1-r^2) \left(\frac{N-1}{N-2} \right)}$**

$$\text{formula. } s_{Y-\hat{Y}} = 1.581 \sqrt{(1-.606^2) \left(\frac{5-1}{5-2} \right)} = 1.452$$

- Calculate the residual variance in 2 different ways.
 - Since 6.3235 is the sum of squared residual we can divide that by N-2 so get the variance. $6.3235/2 = 2.108$ or 2.11**
 - Since .606 is the correlation we know that r-squared is .367 and this is the percentage they share.**
 - We also know that 1-.367 is .633 and this is the percentage they do now share, therefore**
 - .633 * the Sums of Squares for Y would be SSresidual.**
 - 1.581^2 is the variance on Y so $1.581^2 * 4$ is the SStotal = ,**
 - $.633 * 9.998 = 6.329$**
 - $6.329/\text{df residual} = 6.329 / 3 = 2.11$**

36. For men 35% of the error in predicting Virility can be reduced by knowing their Foot size.

- What is the correlation between virility and foot size?
.35 is r^2 so $r = \text{SQRT}(.35) = .59$

- b. If the total SS for virility is 39.74 what is the SS regression? SSresidual?
.35 * 39.74 = 13.91 for SSregression. 1 - .35 = .65. .65 * 39.74 = 25.83 for SSresidual
- c. What percentage of the error in predicting Foot Size can be reduced by knowing how Virile a man is?
Same, .35

37. Answer the questions based on the following output:

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Number of dates in a week	5.00	1.581	5
Rating of attractiveness out of 10	5.80	2.950	5

Correlations

		Number of dates in a week	Rating of attractiveness out of 10
Pearson Correlation	Number of dates in a week	1.000	.536
	Rating of attractiveness out of 10	.536	1.000
Sig. (1-tailed)	Number of dates in a week	.	.176
	Rating of attractiveness out of 10	.176	.
N	Number of dates in a week	5	5
	Rating of attractiveness out of 10	5	5

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Rating of attractiveness out of 10 ^a	.	Enter

- a. All requested variables entered.
- b. Dependent Variable: Number of dates in a week

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.536 ^a	.287	.050	??...

a. Predictors: (Constant), Rating of attractiveness out of 10

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.874	1	2.874	1.210	.352 ^a
	Residual	7.126	3	2.375		
	Total	10.000	4			

a. Predictors: (Constant), Rating of attractiveness out of 10

b. Dependent Variable: Number of dates in a week

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.333	1.665		2.002	.139
	Rating of attractiveness out of 10	.287	??...	.536		.352

a. Dependent Variable: Number of dates in a week

a. What is the standard error of estimate?

This is simply the square root of the residual variance. $\text{SQRT}(2.375) = 1.541$

b. What is the standard error for the slope?

The formula is $se_b = \frac{s_{Y-\hat{Y}}}{s_X \sqrt{N-1}}$, so we just need to plug in a few values.

$$se_b = \frac{s_{Y-\hat{Y}}}{s_X \sqrt{N-1}} = \frac{1.541}{2.950 * \sqrt{5-1}} = .26$$