

# Final Exam

**Directions:** Fill out each question, showing your work/reasoning (yes, even for multiple choice questions as it gives you an opportunity to get partial credit in the rare situation where you select the wrong answer). In each case, pick the best answer. You may use your notes and homework on this exam (no textbook). The entire exam is worth 100 points. Have fun (or at least try to enjoy it)!

## Section I: Multiple Choice (70 points possible, 5 per question)

1. You are doing a study about the influence of having children on depression levels. You do a regression and get an  $r^2 = .50$ . What is an appropriate interpretation of this result?
  - a. 50% of the variability in depression is accounted for by having children
  - b. Parents are 50% more likely to have depression than non-parents
  - c. Having children causes 50% of the depression found in parents
  - d. The results are not significant
2. We hypothesize that both time spent hiking and shoe-type (Jordans or Nike) predict resting heart rate. We want to test this hypothesis. What approach can we use?
  - a. Chi-square test of independence
  - b. Pearson's correlation
  - c. Multiple Regression
  - d. ANOVA
3. When do you use an ANOVA?
  - a. When you don't know how to use t-tests.
  - b. When you have 3 or more groups.
  - c. When your sample sizes are more than 100.
  - d. When you have a categorical outcome.
4. What should we use if we want to find the influence of a third variable on a relationship between two variables?
  - a. Mediation
  - b. Moderation
  - c. Either a or b or both
  - d. Neither a nor b
5. Which of the following is not an assumption for Chi Square?
  - a. Independence
  - b. Appropriated measurement
  - c. Homogeneity of variance
  - d. Expected frequency 5+
6. A correlational study between the number of shark attacks and the amount of ice cream purchased found an  $r$  of .361 and an  $R^2$  of 0.10. What size of an effect would this be classified as?
  - a. Small
  - b. Moderate
  - c. Large
  - d. Not enough information
7. You decide to look at exam scores for students (with permission, of course). You found a positive association between note-taking and high performance on the exam. This association may be explained by hours studying. What type of variable is "hours studying"?
  - a. Predictor
  - b. Confounding
  - c. Moderating
  - d. None of the above.

8. In simple linear regression, the line of best fit goes through:
  - a. Average of the predictor
  - b. Average of the outcome
  - c. Both a and b
  - d. Neither a or b
9. Which scale(s) of measurement is/are used for Pearson's r (select all that apply)?
  - a. Nominal
  - b. Ordinal
  - c. Interval
  - d. Ratio

Given this output, answer questions 10 – 12.

#### Contingency Tables

sex	natsci				Total
	No Opinion	Not Enough	Just Right	Too Much	
Male	1229 (37%)	920 (28%)	946 (28%)	233 (7%)	3328
Female	1588 (39%)	971 (24%)	1190 (29%)	302 (8%)	4051
Total	2817	1891	2136	535	7379

#### $\chi^2$ Tests

	Value	df	p
$\chi^2$	13.2	3	0.004
N	7379		

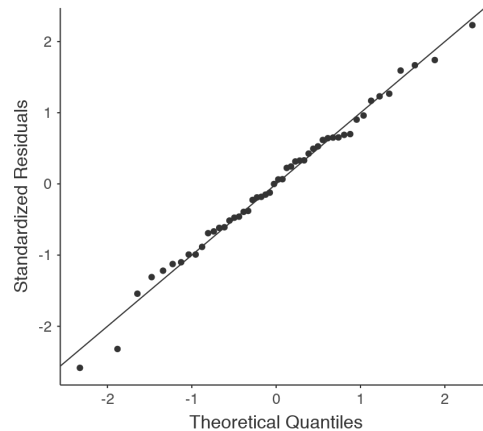
10. What analysis was run?
  - a. Chi Square Goodness of Fit
  - b. Chi Square Test of Independence
  - c. Correlation (Pearson's r)
  - d. Contingency Correlation Analysis
11. Is the effect significant?
  - a. Yes, p value is below .05.
  - b. Yes, p value is above .004
  - c. No, p value is below .05
  - d. No, p value is above .004
12. Where are the differences?
  - a. There are no significant differences.
  - b. Females tend to not have an opinion compared to males; males are more likely to think there is not enough spending than females.
  - c. Females tend to be much more evenly split between all options.
  - d. Males tend to think there is too much spending compared to females.
13. What section of the course seems to be most relevant to your work?
  - a. Descriptive statistics
  - b. T-tests and ANOVA
  - c. Regression
14. Just circle option a.
  - a. Option a.
  - b. Not option a.

## Section II: Applied Statistics (25 points possible)

*You are interested in whether the teen birth rate (continuous) is influenced by poverty (continuous), violent crime (continuous), and the region of the United States (categorical with three levels).*

### Step 1: Examine Variables to Assess Statistical Assumptions

In this example, let's say you know you satisfy all assumptions other than no omitted influences and normality. Below we look at a Q-Q plot.



Do the residuals look like they are normal?

Do you think we have a problem with having an omitted influence (i.e., some variable that relates to both teen birth rates and any of the predictors)? If so, provide an example. If not, explain why.

### Step 2: State the Null and Research Hypotheses (two-tailed)

### Step 3: Define Critical Regions (I did it for you below)

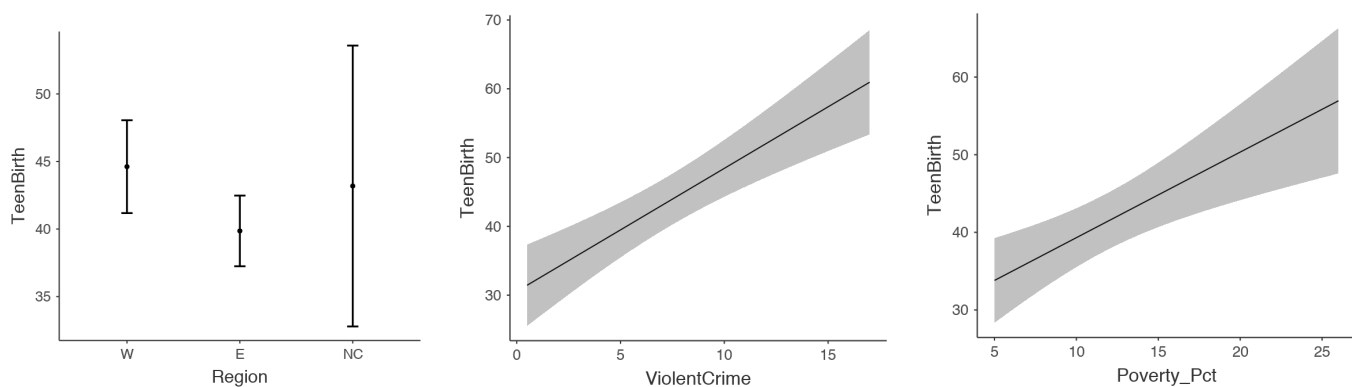
$$\alpha = .05$$

*If the p-value is below .05, we will reject the null hypothesis.*

Step 4: Compute the Test Statistic & Compare to Critical Region and Step 5: Compute an Effect Size and Describe It

**Model Coefficients**

Predictor	Estimate	SE	t	p	Stand. Estimate
Intercept	16.29	3.456	4.713	<.001	
Poverty_Pct	1.10	0.300	3.675	<.001	0.385
ViolentCrime	1.79	0.338	5.291	<.001	0.548
Region:					
E – W	-4.76	2.170	-2.192	0.034	-0.222
NC – W	-1.43	5.508	-0.260	0.796	-0.499



Step 6: Interpret the results (include the significance, the interpretation of the estimates and the standardized estimates, and what the plots show us)