Review Sheet for the Final Exam

1. Know what the null and research hypotheses are; be able to state them both with words and with symbols for different types of analysis. Understand the difference between directional and nondirectional research hypotheses and when each is selected.
2. Understand the meaning of statistical significance. Know what significance levels we usually use, and what they mean. Understand the difference between statistical significance and practical significance (meaningfulness) as well as the concept of effect size. Be able to assess practical significance for differences in means (by calculating effect size), correlation coefficients, regression coefficients, contingency tables (by assessing the size of percentage differences).
3. Know what a test statistic is and understand how to compare it to a critical value to conclude whether we reject or fail to reject the null hypothesis. Understand what a p-value is and know how to compare it to alpha to conclude whether we reject or fail to reject the null hypothesis.
4. Know and understand the two types of error in hypothesis testing (Type I and Type II) and the ideas of "false positive" and "false negative." Understand how the probability of Type I and Type II error are related, how we make decisions about Type I and Type II error, and how these are related to the level of confidence and the sample size. Know how to evaluate the size of Type I and Type II error after hypothesis testing, whether done by hand or in Stata. Know what statistical power is and understand how statistical power is related to Type II error.
5. Know the difference between the one-tailed and two-tailed tests and how it is related to the notion of a rejection region; understand when each is used. Know that two-tailed test requires a larger effect in the sample (e.g., larger difference between the sample means) to achieve statistical significance. Know that one-tailed test is more powerful, but two-tailed test is more conservative, and why.
6. Understand the issues underlying the reproducibility crisis in social sciences, and especially those related to p-hacking; know why it can be problematic to rely solely on the $p<0.05$ cutoff.
7. Be able to test hypotheses for a single sample mean (comparing it to a number derived from theory or another population) by hand (state the hypotheses, decide if one or two-tailed test, set the significance level, calculate an appropriate test statistic, find the critical value, and make a conclusion).
8. Be able to test hypotheses of mean difference for two means of independent samples both by hand and using Stata. Understand the difference between comparing sample means, and making a conclusion whether there is a difference between the means in the larger population. Understand the difference between independent samples and paired (dependent) samples, be able to identify them based on a description.
9. Be able to test hypotheses of mean difference for three or more means (ANOVA) using F ratio, both by hand and using Stata. Understand how we decompose the variance into two components (BSS and WSS) and what they mean; understand what each part of ANOVA table means.
10. Know why doing pairwise mean comparisons instead of ANOVA is problematic; understand the idea of inflated alpha. Understand the idea behind Bonferroni corrections and how they are applied by hand as well as when using Stata.
11. Know how to use a Pearson's correlation coefficient to test whether there is a relationship between two interval/ratio variables in the larger population. Using Stata, be able to calculate correlation coefficients and obtain significance tests; know how to do a one-tailed test by dividing the p-value by 2. Be able to interpret the strength and direction of the relationship in the sample. Know when it is applicable to use Pearson's correlation coefficient. Understand the underlying assumption about the linear relationship between the two variables.
12. Know how to calculate coefficients of determination and alienation. Be able to explain what each of them tells you about the relationship between two variables.
13. Understand the difference between correlation and causation; know the conditions necessary for a proof that a causal link exists. Understand what a spurious correlation is.

14. Be able to construct a scatterplot using Stata. Know how to determine from a scatterplot whether there is a positive, negative, or no relationship between two variables, and be able to visually distinguish a strong vs a weak relationship. Be able to add a regression line and/or a lowess smoother to a scatterplot in Stata; know what they show us and be able to evaluate whether a relationship looks linear.
15. Understand the basic idea of Ordinary Least Squares regression, know what dependent and independent variables are in regression analysis, and what it means to regress Y on X.
16. Know how to use regression analysis to test whether there is a relationship between two interval/ratio variables in the larger population (based on the significance test for the regression slope). Using Stata, be able to find and interpret both the intercept and the slope (a and b) for your sample, as well as R-squared (coefficient of determination). Know how to use confidence intervals to discuss the size of the slope in the population. Distinguish when we should use correlation vs. regression analysis.
17. Understand what predicted values are, and be able to calculate a predicted value of Y given X using a regression equation. Be able to calculate the error of estimate for a specific value of X. Know what a standard error of estimate is. Be able to evaluate practical significance of a regression slope.
18. Recognize potential pitfalls of regression analysis and know what they mean; this includes pitfalls related to nonlinear relationships, correlations among independent variables (collinearity), outliers and skewed distributions, "fishing" for effects, assuming causation based on regression coefficients, and extrapolation. Understand the difference between extrapolation and interpolation.
19. Be able to interpret a contingency table (with either column or row percentages) to examine a relationship between two categorical variables. Know how to construct a contingency table with column percentages.
20. Understand the idea of variable independence and know the difference between observed and expected frequencies. Be able to use contingency tables and chi-square test to evaluate whether there is a relationship between two categorical variables in the larger population, both by hand and in Stata.
21. Know what residuals are in chi-square test calculations and what large vs small residuals mean. Be able to conduct post-hoc analysis of residuals to pinpoint the specific differences after the overall chi-square test indicates that there is an overall relationship.
22. Know how to decide what type of statistical analysis (two independent samples t-test, ANOVA, correlation, regression, chi-square) to use for a given problem or for two specific variables in Stata (based on identifying their level of measurement).