

Final Exam Study Guide 115

Vinny Paris

May 2026

The guide is broken into two parts. The “new” stuff and the “old” stuff which reflects the exam’s makeup. To be upfront there will be a test of a regression coefficient. You will not be tested on ANOVA’s nor the required readings.

1 Two-sample t-test

- When would this situation come up?
 - with two distinct subpopulations which we want to compare means
- Often the null hypothesis is in the form of
 - $H_0: \mu_1 = \mu_2$
 - $H_0: \mu_1 \geq \mu_2$
 - $H_0: \mu_1 \leq \mu_2$
- It’s unusual to say something like
- $H_0: \mu_1 = \mu_2 + 3$
- Assumptions are still the big three
 - IID (Independent and Identically Distributed)
 - Random
 - PopulationSSSS ARE normal or nSSSS ARE large
 - * Strong emphasis on the plural words
 - * Both subpopulations need to pass this assumption
 - * Large sample size from Group A doesn’t do anything for Group B in this situation
- Often the issue with this test is *lurking variables* which are unaccounted-for, important variables
- Those lurking variables violate the IID assumption (specifically identically distributed)

2 Testing Regression Coefficients

- Explain why we test coefficients at all
 - What does it imply if our p-value is large?
 - What does it imply if our p-value is small?
- Be able to write out the null and alternative hypothesis
- Check the assumptions (more in multiple regression section)
- Interpret R's output on testing regression coefficients.
 - What does a low p-value mean?
 - What does a p-value mean generally?
- Explain when testing regression coefficients can be better than a standard t-test
 - Lurking variables aren't accounted for in t-tests if they exist
- More heavily emphasized on the exam than the two-sample t-test

3 Multiple Linear Regression (MLR)

This is not “new” but it ties so heavily to the coefficient notes that I feel the need to keep this here

- Explain why we use MLR over simple linear regression
 - SLR only allows for one explanatory variable which is too limiting
 - MLR allows us to better understand the system as a whole since we can account for differing effects
- How do we interpret the coefficient of a quantitative (numeric) explanatory variable?
- How do we interpret the coefficient of an indicator (eg a nominal/categorical explanatory variable)?
- Assumptions
 - Random
 - Population is normal or n is large
 - IID
 - * There is where homoskedasticity assumption is (need same spread to be identically distributed)
 - * Also where “linear” from SLR went...if the model doesn't fit the residuals won't have the same scattering around the 0 line

- NOTE: Both IID and the “pop is normal” reference the residuals!
- Justify/explain why we color residuals by explanatory variables

4 ANOVA

- You will not be tested on ANOVA’s outside of maybe a bonus point or two
 - Understanding the null hypothesis and what p-value’s imply would be the only meaningful thing
 - IMPORTANT: Traditionally the ANOVA is used before coefficient testing as it acts as a “first pass” on finding any meaningful relationships

5 Old Stuff

- What is a p-value?
- What is the difference between standard deviation and standard error?
- What does a confidence of 95% mean?
- What are the four main challenges colorblind people face?
- How can we make graphics more accessible?
- Be sure to know how to read the different graphs we have seen early in the semester
- Permutation test basics
- How does a hypothesis test work?
- Read and interpret a confidence interval for some statistic.
- Make predictions using SLR or MLR models
 - Not as scary as it sounds I promise
- Be comfortable using indicators in regression
 - Eg make a prediction that uses an indicator variable
 - Eg make indicator variables in a data set
- Correlation
 - Pearson vs Spearman
 - Interpretations