Topics for Stat 311 Final Exam

1. **Sampling distributions and the Central Limit Theorem.**
    - Concept of distributions for individual observations of a random variable and a distribution for the sample statistic
    - CLT for $\bar{x}$ or $\sum x$; gives approximate normality when $n \geq 30$
2. **Confidence intervals**
    - For five parameters $(p, p_1 - p_2, \mu, \mu_1 - \mu_2, \mu_d)$
    - Interpretation being based on the method, not probability
    - For $p_1 - p_2$ or $\mu_1 - \mu_2$ you should be able to think about what it means if the CI contains 0; same for μ_d
    - Always use z for p, provided sample size is large enough
    - Use z for μ if σ is known
    - Use t for μ if σ is unknown
    - Remember the degrees of freedom for t when looking up a critical value; for unpooled SE we will use $df = \min(n_1 - 1, n_2 - 1)$.
    - Methods using z assume large samples for p and either large samples or X normally distributed for μ
    - Methods using t assume that X is normally distributed (or at least approximately normally distributed)
    - Know what other assumptions/conditions are needed for CI.
3. **Sample size estimation to determine the minimum number of samples required to achieve a particular ME with a particular level of confidence—always use z as the critical value for this class.**
4. **Hypothesis Testing**
    - For five parameters $(p,\ p_1 - p_2, \mu,\ \mu_1 - \mu_2,\ \mu_d)$ as well as β for regression, and chi-square tests for count data
    - Set α apriori
    - Define competing hypotheses; should be able to do this both in words and using symbols, as appropriate
    - Calculate the appropriate test statistic; same conditions for choice of z versus t as for confidence intervals; use chi-square for count data
    - Find a critical value or a p-value; you need to know how to do both for z and t. I will provide you will p-values or critical values for chi-square
    - Various degrees of freedom apply for t
    - For two-tailed z or t tests, remember to multiply the p-value by 2; chi-square is one-tailed (upper) only
    - Make a decision; choices are reject the null or fail to reject the null
    - Write a one or two sentence conclusion in the context of the problem. Context matters; student's lose the majority of points by being sloppy with context
    - Once you make a decision you risk either a Type I or Type II error (if you were to accept the null)—make sure you know how these apply and that you can describe what the error would be in the context of the problem
    - Know the relationship between the probability of Type I and Type II errors, power, and changes in α and sample size
    - Methods using z assume large samples for p and either large samples or X normally distributed for μ
    - Methods using t assume that X is normally distributed (or at least approximately normally distributed)

- For testing $\mu_1 - \mu_2$ you have the option to use pooled or unpooled variances.
  - Know the rule of thumb for helping to decide
  - Know when it might not be advisable to pool
- Know what other assumptions/conditions are needed for HTs

5. **Inferences for Regression**
   - Be able to write down the null and alternative hypotheses for testing the regression slope
   - Know how to form and interpret a confidence interval for the regression slope
   - Know how to conduct a hypothesis test for the slope of a regression line
   - Know how to provide a one or two sentence conclusion in the context of the problem
   - Know how to calculate and interpret confidence intervals for the mean value of $y \mid x$.
   - Know how to calculate and interpret prediction intervals for a new randomly observed value of $y \mid x$.
   - Know the assumptions, including those with respect to epsilon.
6. **Chi-square Tests for Count Data**
   - Three types of tests: goodness-of-fit, tests for homogeneity, tests for independence
     - Know the hypotheses for each test
     - Know how to compute the test statistic (remember that you are always working with counts, not percentages)
     - Know the degrees of freedom
     - Know that you do not round the expected counts to integers. For the final it is okay to only use one decimal point.
     - Know how to interpret the results of a test and how to make a conclusion in the context of the problem

For calculating confidence intervals and conducting hypothesis tests for p, $p_1 - p_2$, μ, $\mu_1 - \mu_2$, or μ_d, I recommend the following approach:

1. Read the problem and decide if it is a CI or HT problem
2. Figure out what information is given and be clear what you are asked to find
3. Decide which parameter is involved (proportion or mean; one sample or two samples)
4. If p or $p_1 - p_2$ make sure the sample is large enough; if μ, $\mu_1 - \mu_2$, or μ_d with σ known make sure sample is large enough or X is normally distributed; if σ unknown make sure X is approximately normally distributed
5. Calculate the correct SE; for proportions remember that the SE's are likely to be different depending on if the problem is a CI or a HT
6. If critical value or p-value not specified, you can use either method to make a decision; if specified in the problem, you must use that method
7. For tests of $\mu_1 - \mu_2$ I will be clear whether or not I want pooled or unpooled SEs
8. When doing an unpooled test use $\min(n_1 - 1, n_2 - 1)$ for the degrees of freedom
9. All p-values need to be based on the statistical tables. For t this means you may only be able to give a range (i.e., $0.01 < p < 0.05$)
10. All conclusions must be in the context of the problem—very few points given if you fail to add context.