Lecture 29

Review Session

Here is the Plan for Final Weeks

- This week
  - Tuesday: Last class, review session and some announcement
  - Thursday: No class, Q & A session in Room 236, Cate Center 1 during regular class time.
- Next Week
  - Tuesday: No class, Q & A session in Room 236, Cate Center 1 from 1:30pm - 2:30pm.
  - Wednesday 12:30pm - 6:30pm, window for the optional final exam.

Clarification on Optional Final Exam

It is truly optional!

- Calculate your grades in the following manner:
  - Take the arithmetic mean of your highest 15 out of the 18 assignments.
  - For assignment with technical issues, make sure confirm with me through email about excluding it.
  - For assignments with technical issues, you can calculate the grade by taking the arithmetic mean of your highest $15 - x$ out of the $18 - x$ assignments, where x is number of assignments with issues (that you have confirmed with me).

Clarification on Optional Final Exam

Take the optional final if

- You are not happy with the current cumulative grade.
- You are interested in taking it.
- In the end, I will take the higher one between your assignments average and optional final.

About the Optional Final Exam Format

- It will be given between 12:30pm - 6:30pm on December 11th, which is next Wednesday.
- It has 30-35 T/F and multiple choices questions.
- You have 100 minutes to complete it.
- Covers material until multiple linear regression.
- You can start at any time between 12:30pm - 4:50pm, after 4:50pm might be too late.
- Open book exam, it is similar to the quizzes except the length.

Curve and Attendance Policy

Curving

- As mentioned in syllabus, I will curve the median grade to 90 if it is below 90.
- I will do the curving by moving the bell shape grade distribution toward right.

Attendance

- I won't deduct your marks for missing classes.
- I will add marks in calculating letter grades for those show up in classes.

The Course Reflection Survey

- I'd like to ask you to do me a favor. If possible, could you please help me with the Course Reflection Survey? It is due by this Sunday.
- Then I will know where I should improve.
- Thank you!

Data Types and Statistics

- Data
  - Categorical
    - nominal
    - ordinal
  - Numerical
    - discrete
    - continuous
- Population - parameters (unknown)
  - mean
  - variance
- Sample - statistics (calculated)
  - mean
  - variance

Hypothesis Testing

- Define your null (H_0) and alternative (H_1) hypotheses.
- Calculate an appropriate test statistic.
- Based on the sampling distribution of the test statistic under H_0, reject H_0 if the observed test statistic is extreme (using rejection regions or p-values).
- Reject H_0 if your test statistic falls in the rejection region or if your p-value is less than α.

Single Population

- Testing μ when σ^2 known:
  - Calculate Z-statistic.
  - Compare to $N(0,1)$ distribution.
  - One or two-tailed depending on H_1.
- Testing μ when σ^2 unknown:
  - Calculate T-statistic.
  - Compare to t-distribution with $n-1$ degrees of freedom.
  - One or two-tailed depending on H_1.
- Testing a population proportion p:
  - Calculate Z-statistic.
  - Compare to $N(0,1)$ distribution.
  - One or two-tailed depending on H_1.

Comparing Two Populations

Independent Samples

- Testing $\mu_1 - \mu_2$ when σ_1^2, σ_2^2 known:
  - Calculate Z-statistic.
  - Compare to $N(0, 1)$ distribution.
  - One or two-tailed depending on H_1.
- Testing $\mu_1 - \mu_2$ when σ_1^2, σ_2^2 unknown and $\sigma_1^2 = \sigma_2^2$:
  - Calculate pooled sample variance s_p^2.
  - Calculate T-statistic.
  - Compare to t-distribution with $n_1 + n_2 - 2$ degrees of freedom.
  - One or two-tailed depending on H_1.

Comparing Two Populations

Independent Samples

- Testing $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$:
  - Calculate F-statistic (put larger sample variance on top).
  - Compare to F-distribution with $n_1 - 1$ numerator degrees of freedom and $n_2 - 1$ denominator degrees of freedom (n_1 is the sample size corresponding to the larger sample variance).
  - Two-tailed, but only need to look at upper tail of F-distribution.

Comparing Two Populations

Paired Samples

- Testing μ_D:
  - Calculate paired differences (remember to set up hypotheses appropriately).
  - Calculate T-statistic.
  - Compare to t-distribution with $n-1$ degrees of freedom.
  - One or two-tailed depending on H_1.

Comparing Two Populations

- Testing $H_0 : p_1 - p_2 = D_0$ for $D_0 \neq 0$:
  - Calculate Z-statistic.
  - Compare to $N(0, 1)$ distribution.
  - One or two-tailed depending on H_1.
- Testing $H_0 : p_1 - p_2 = 0$:
  - Calculate combined proportion $\hat{p}$.
  - Calculate Z-statistic.
  - Compare to $N(0, 1)$ distribution.
  - One or two-tailed depending on H_1.

ANOVA

- Calculate sums of squares, degrees of freedom and mean squares for each source of variation.
- Calculate F-statistic.
- Compare to F-distribution with numerator degrees of freedom equal to the factor or interaction degrees of freedom, and denominator degrees of freedom equal to the error degrees of freedom.
- One-tailed, reject when F-statistic is too large.

Simple Linear Regression

- Testing overall significance of model:
  - That is, testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.
  - Calculate T-statistic.
  - Compare to t-distribution with $n - 2$ degrees of freedom.
  - Two-tailed.
  - Can also test overall significance of model by testing $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$.

Simple Linear Regression

- General tests for β_0 and β_1:
  - Calculate T-statistic.
  - Compare to t-distribution with $n-2$ degrees of freedom.
  - One or two-tailed depending on H_1.
  - May not be able to use p-value given in computer output, since this is the p-value for testing $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, for $j = 0, 1$.

Multiple Linear Regression

- Testing overall significance of model:
  - Calculate sums of squares, degrees of freedom and mean squares.
  - Calculate F-statistic.
  - Compare to F-distribution with k numerator degrees of freedom and $n - k - 1$ denominator degrees of freedom, where k is the number of independent variables in the model.
  - One-tailed, reject when F-statistic is too large.

Multiple Linear Regression

- Testing individual coefficient parameters:
  - Calculate T-statistic.
  - Compare to t-distribution with $n - k - 1$ degrees of freedom.
  - One or two-tailed depending on H_1.
  - Note that the p-value given in the computer output is for testing $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, for $j = 0, \ldots, k$.
  - The conclusion of each individual test is conditional on the fact that the other independent variables have already been included in the model.

Using Probability Tables

- Binomial tables:
  - Can only use the binomial tables if the required values of n and p are listed in the table.
  - For values of n and p which are not in the table, you must use the binomial formula.
- Normal tables:
  - When looking up a Z-value to find the corresponding probability, round the Z-value to 2 decimal places.
  - When looking up a probability to find the corresponding Z-value, choose the closest probability.
  - When probability falls exactly in the middle between two Z-values, choose mid-point (e.g., $z_{0.05} = 1.645$).

Using Probability Tables

- t-table:
  - To find a lower critical value, put a negative on the corresponding upper critical value.
- F-tables:
  - Remember to use the correct F-table corresponding to the value of α.
- t-table and F-tables:
  - If you can't find the exact degree of freedom in the table, chose the closest degree of freedom.