Study Guide – Math 107, Final Exam

Prof. Adam Loy

General Information

The final will be on Tuesday 6/7, from 8 AM to 10:30 AM. No books, notes, computers, or cell phones are allowed. The material covered will be cumulative since statistics is cumulative; however, I will be giving far more weight to the material introduced after Exam 2.

To study, I recommend carefully going through class notes, homework problems, past exams, and handouts (especially this handout, the midterm review handouts, and the examples for inference) actively (intermixing reading, thinking, solving problems, and asking questions). After reviewing those materials I recommend solving lots and lots practice problems (there are many problems in the textbook that we didn't complete that may be helpful—remember that the odd problems have solutions in the back of the textbook).

Chapter 6 - Inference for Means and Proportions

- Definitions/results

  Confidence interval, confidence level, margin of error, standard error, hypothesis test, null hypothesis, alternative hypothesis, test statistic, p-value, t-distribution, degrees of freedom, one-sample t-test, two-sample t-test, reference distribution, sampling distribution

- Review the concept of inference (this was originally introduced in chapters 3 and 4).
- Know the conditions to check, and how to construct a confidence interval for p, μ, and $\mu_1 - \mu_2$.
- Understand the concept of statistical confidence, and how to interpret a confidence interval.
- Know how to determine the sample size needed to estimate a proportion or mean within a specified margin of error at a given confidence level.
- Know how to conduct a hypothesis test for p, μ, and $\mu_1 - \mu_2$
  - Step 1: State the hypotheses
  - Step 2: Check the conditions necessary to use the normal or t distribution
  - Step 3: Compute the test statistic
  - Step 4: Compute the p-value using the appropriate reference distribution
    - Use the alternative hypothesis to find the appropriate probability
  - Step 5: Make a decision and state the implications of your decision in the context of the problem
- Understand how the t-distribution differs from the normal distribution, and when we need to use the t-distribution.

Chapter 2.6/9.1 - Regression

- Definitions/results
  response variable, explanatory variable, correlation coefficient, least squares regression line, slope, y-intercept, principle of least squares, coefficient of determination (R^2), residual, assumptions for regression model
- Be able to describe the association between two quantitative variables using a scatterplot or the correlation coefficient.
- Be able to make predictions using the least squares regression line.
- Be able to interpret each of the following within the context of the problem: slope, intercept, R^2.
- Understand how to calculate a residual and interpret a residual plot.
- Be able to discuss the appropriateness of a regression model.
- Be aware of the regression cautions discussed in class.
- Be able to conduct inference for the slope of a simple linear model using output from R.

Formulas

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$\text{IQR} = Q_3 - Q_1$$

$$Q_1 - 1.5 \times \text{IQR}$$

$$Q_3 + 1.5 \times \text{IQR}$$

$$z = \frac{x - \mu}{\sigma}$$

$$x = z\sigma + \mu$$

$$r = \frac{1}{n-1}\sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s_x}\right)\left(\frac{y_i - \overline{y}}{s_y}\right)$$

$$e_i = y_i - \widehat{y}_i$$

$$n \geq \left(\frac{z^*}{ME}\right)^2 p(1-p)$$

$$n \geq \left(\frac{z^* s}{ME}\right)^2$$

$$\text{statistic} \pm (\text{critical value}) \times \text{SE}$$

$$\text{test statistic} = \frac{\text{statistic} - \text{null value}}{\text{SE}}$$

Standard errors:

$$SE_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}}$$

$$SE_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$