Final Exam Preparation

Logistics:
- Mon Mar 12 or Tues Mar 13 or Fri Mar 16
  - 170 minutes
- Open-notes, open-handouts
  - You may bring anything that I have provided or that you produce yourself
- Bring calculator
  - No computer use
- Roughly one-half to two-thirds on material since most recent exam
  - Handouts 18-25, Quizzes 18-25, HW7-8, Chapter 7
- Roughly one-third to one-half on earlier material
  - Focusing on big ideas

Overview (since most recent exam):
- We have learned how to apply statistical methods to quantitative data, using both descriptive and inferential methods.
  - Comparing two groups with quantitative response
    - Independent Samples
    - Paired data
  - One-sample inference with quantitative variable
  - Association between two quantitative variables

Outline:
- Handout 18: Analyzing Quantitative Data
  - Center, variability, shape, outlier, symmetry, skew, mean, median, standard deviation (SD), interquartile range (IQR), five-number summary, boxplot, outlier test, modified boxplot, histogram, resistance
- Handout 19: Statistical Significance with Quantitative Data
  - Simulating randomization test for comparing two groups with quantitative response, approximate p-value
- Handout 20: Two-Sample $t$-Test, $t$-Interval
  - Test statistic, p-value, technical conditions, confidence interval, effect of sample sizes, effect of within-sample variability
- Handout 21: One-Sample $t$-Test, $t$-Interval
  - One-sample $t$-test, $t$-interval, paired data, paired $t$-test, matched pairs design
- Handout 22: Association with Quantitative Variables
  - Form, direction, strength of association; scatterplot, correlation coefficient
- Handout 23: Least Squares Regression
  - Residual, least squares criterion, slope coefficient, influential observation, coefficient of determination $r^2$, residual standard deviation, calculating coefficients, extrapolation
Handout 24: Inference for Regression
  - Standard error of sample slope, hypothesis test for slope, confidence interval for slope, confidence interval for mean value of response, prediction interval
Handout 25: Multiple Regression
  - Interpretation of coefficients, $R^2$, model utility test, individual coefficients test

Advice:
- Organize notes for efficient retrieval of information/formulas
- Don’t plan to use notes too much
  - Prepare as if exam were closed book/notes
  - Focus on understanding, not memorization
  - Be cognizant of time constraint
- Expect similar questions to what we answer in class every day, in quizzes, on HW
- Be prepared to think/explain/interpret
  - Not just plug into formulas
  - Be ready to explain process of how you would do calculations
- Be ready to interpret computer output
  - Possibly exclude irrelevant output
- Read carefully
  - Be sure to answer the question asked
- Take advantage of information provided
  - Perhaps including computer output
- Relate conclusions to context
- Justify/explain your answers
  - Unless you are explicitly told not to bother
- Prepare as thoroughly as you would for a closed-notes exam
  - Re-read handouts
  - Read online chapters
  - Re-read previous exam preparation documents
  - Re-work in-class examples
  - Re-work quiz questions
  - Re-work HW questions
  - Bring questions to office hours
    - Fri 10:30-11:30, 2:30-3:30, Mon 2:30-3:30, Wed 10:30-11:30

Choice of Procedure:
Now that we have learned many procedures for analyzing and drawing conclusions from data, one of the challenges is deciding which procedure to apply in a given situation. Some of the questions to ask yourself are:
- Is there only a response variable, or is there also an explanatory variable?
- Is the response variable quantitative or categorical?
- Is the explanatory variable quantitative or categorical?
- For categorical variables, are there two categories or more than two?
• When there is a quantitative response variable and a binary categorical explanatory variable, were the data collected in a matched-pairs or independent-samples design?

Some of the statistical inference techniques we have studied include:

A. One-sample $z$-procedures for a proportion
B. Two-sample $z$-procedures for comparing proportions
C. One-sample $t$-procedures for a mean
D. Two-sample $t$-procedures for comparing means
E. Paired-sample $t$-procedures
F. Simple linear regression procedures
G. Multiple regression procedures

Suppose that I consider the students in the class to be a sample from the population of all Cal Poly students, and I record the following for each student enrolled in this class:

- Gender
- Score on first exam
- Number of quizzes taken
- Time spent sleeping last night
- Handedness (left- or right-handed)
- Time spent on the final exam
- Score on the final exam

For each of the following questions, indicate (by capital letter) which procedure is the appropriate one to address the question.

a) Are more than 10% of Cal Poly students left-handed?

b) Is a student’s score on the first exam useful for predicting his/her score on the final exam?

c) Do students tend to score lower on the final exam than on the first exam?

d) Do males and females differ with regard to the average time they spend on the final exam?

e) Do the proportions of left-handers differ between males and females on campus?

f) Is time spent on the final exam useful for predicting final exam score, once exam 1 score is already considered?

g) Is there a significant association between the number of quizzes that a student takes and his/her score on the final exam?

h) How much sleep did Cal Poly students get on average last night?

i) Are sleeping time, exam 1 score, and number of quizzes taken useful for predicting time spent on final exam?