Stat 221 – Final Exam Preparation

- Logistical details
  - Section 1: Mon Dec 8, 7:10-10am
  - Section 2: Wed Dec 10, 7:10-10am
  - Extra office hours
    - Sun Dec 7, 3-5pm; Mon Dec 8, 2-4pm; Tues Dec 9, 12-2pm
  - 170 minutes
    - But the exam will be only 50% longer (roughly) than a midterm
  - Open-book, open-notes
  - Calculator needed
  - Normal (*z*-), *t*-, chi-square tables needed
- Coverage
  - Roughly one-half on newer material
    - Outlines from days 42, 45-50
    - Quizzes 25-27
    - Investigation 10
  - Roughly one-half on earlier material
    - Focusing on "big ideas"
- Resources available online
  - This preparation sheet
  - Day-by-day outlines
  - Quiz solutions
  - Investigation solutions
  - Midterm exam solutions
  - Optional problem list (solutions on Blackboard)
- Types of questions to expect
  - Short answer
  - Calculations
  - Interpretations and explanations
    - Possibly of Minitab output
  - Similar to in-class examples, quizzes, investigations, earlier exams
- Advice for preparing
  - Prepare and organize your notes carefully
  - Don't study less because it's open-notes/book
  - Plan not to rely on your notes/book too much
  - Re-read the day-by-day notes
  - Re-read highlighted passages, watch-out, wrap-up sections of book
  - Focus on understanding, not memorization
  - Don't forget about fundamental ideas from earlier in course
  - Review and make sure that you can answer the quiz, investigation, optional assignment questions
  - Ask questions during review class session, office hours
- Advice during the exam
  - Show up on time!
  - Be cognizant of time constraint

- Read carefully
- Relate conclusions to context
- Write and explain clearly
- Do not elaborate excessively
- Show details of calculations
- Take advantage of partial information
- Review your work, as time permits

Outline (of most important topics)

- Fundamental Ideas
  - Observational unit, variable
    - Categorical vs. quantitative
    - Explanatory vs. response
  - Design of study
    - Observational study vs. experiment
    - Scope of conclusions, causation
  - Graphical displays, numerical summaries
    - Shape, center, spread, outliers
  - Population, sample
    - Parameter, statistic
  - Sampling distribution
    - Standard error
  - Confidence interval
    - General form
    - Interpretation
    - Effects of sample size, confidence level
  - Significance test
    - Components
    - P-value
    - Interpretation

- Inference for two-way tables
  - Testing equality of proportions or independence
  - Chi-square test
  - Expected counts
  - Test statistic, *p*-value
  - Largest contribution(s) to test statistic

- Bivariate relationships
  - Association
    - Scatterplot
    - Form, direction, strength
    - Correlation coefficient
  - Least squares (regression) lines
    - Residuals

        - Prediction
        - Interpretation of slope coefficient
        - Coefficient of determination (R^2)

- Simple linear regression
    - Inference for slope coefficient
        - Standard error
        - *t*-test
        - Confidence interval
    - Test for correlation coefficient

- Multiple regression
    - Interpretation of coefficients
    - R^2
    - Model utility test (*F*-test)
    - Individual *t*-tests
    - Variable selection

- Analysis of variance (ANOVA)
    - Purpose, need
    - Big idea: compare variation between groups to variation within groups
    - ANOVA table
        - Sums of squares
        - Degrees of freedom
        - Mean squares
        - Inter-relationships
    - *F*-test
    - Technical conditions
    - Multiple comparisons
        - Tukey procedure

Which procedure to use when?

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 in this course include:

A. One-sample *t*-procedures for a mean
B. Two-sample *t*-procedures for comparing means
C. Paired-sample *t*-procedures
D. One-sample *z*-procedures for a proportion
E. Two-sample *z*-procedures for comparing proportions
F. Chi-square goodness-of-fit procedures
G. Chi-square procedures for two-way tables
H. Simple linear regression procedures
I. Multiple regression procedures
J. One-way ANOVA procedures

Suppose that I record the following for each student enrolled in this class:

- Gender
- Major
- Score on first exam
- Number of quizzes taken
- Time spent sleeping last night
- Handedness (left- or right-handed)
- Political inclination (liberal, moderate, or conservative)
- 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) Do the various majors differ with regard to average sleeping time?

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

c) Are the three political inclinations equally represented?

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

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

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

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

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?

j) Is there an association between gender and major?