STAT 155 Introductory Statistics

Review for the Final Exam

Fall 2006
Final Exam Format

- Closed-book, closed-notes
- One formula sheet (double-sided)
- Calculator
- Scantron and #2 pencils
- Tables will be provided
- 50 multiple-choice questions
  --- at least a half based on materials covered after the 2\textsuperscript{nd} midterm
Suggestions for your review

• Read the *lecture notes* carefully.

• Go over
  – *Homework problems*;
  – *practice midterms and midterms*. *(Solve the problems yourself!)*

• Do the *practice problems* in this final review.
Office hours

- Monday 12/11 --- Wednesday 12/13
  11 am --- 1 pm

Solutions to Hwk 22: available at Blackboard from Thursday 12/7
What have we learned in STAT 155?

- Basic data analysis techniques
- Simple linear regression
- Elementary probability
- Sampling distributions
- Point estimation & confidence intervals
- Hypothesis testing (Z test)
Review of Chapter 1

- Visualization tools:
  - stem-and-leaf, histogram, bar graph, pie chart, boxplot, time plot
- Measures of Location:
  - mean, median, mode
- Measures of Variability:
  - range, interquartile range, standard deviation, variance
- The five-number summary
- Effect of linear transformation
- Normal distribution:
  - Standardization, normal distribution calculation
Review of Chapter 2

• Scatterplots: interpretation.
• Correlation: definition and properties
• Least-Squares Regression:
  – Find the regression equation, prediction and interpretation, the meaning of $r^2$
• Regression diagnostics:
  – Residuals, outliers, influential observations, lurking variables
Review of Chapter 4

- Sample spaces, events, union, intersection, complement, disjoint events, tree diagram, Venn diagram
- Axioms and properties of probability
- Addition rule:
  \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \]
- Conditional probability:
  \[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} \]
- Multiplication rule:
  \[ P(A \cap B) = P(A \mid B) P(B) = P(B \mid A) P(A) \]
- Independence: \[ P(A \cap B) = P(A) P(B) \]
- Bayes rule: no need to memorize … derive it!
- Rules for expected values & variances
Review of Chapter 5

- **Binomial distribution**
  - Sampling distribution for counts and proportions
  - Table C, mean, standard deviation
  - Normal approximation (continuity correction)

- **Sampling distribution of \( \bar{X} \)**

1. \( \mu_{\bar{X}} = \mu_X \); \( \sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n} \)

2. (i). If \( X \) is normal, \( \bar{X} \) is normal.
   (ii). If \( X \) is nonnormal, \( \bar{X} \) is approximately normally distributed for sufficiently large sample size.
Road map for CI and HT (Lectures 18 – 22)

- 1 sample or two samples?
  - 1 sample: \( \mu \) or \( p \) ?
  - 2 samples: \( \mu_1 - \mu_2 \) or \( p_1 - p_2 \) ?
- CI or HT ?
- For \( p_1 - p_2 \): use \( SE_D \) or \( SE_{Dp} \) ?
- \( H_0 \): 1-sided or 2-sided?
- Which \( P \)-value to calculate ?
- Decision: reject or accept \( H_0 \) ?
- CI: what is the margin of error?
- HT: which 4 steps ?
- Relationship between CI and HT ?
Practice Problem 1

• A fair coin is tossed 200 times, what is the probability that the total number of heads is between 95 and 105?

• $X \sim \text{Bin}(200, 0.5)$; Want: $P(95 \leq X \leq 105)$

• Need to use normal approximation

• $\mu_X = 200 \times 0.5 = 100$, $\sigma_X = (200 \times 0.5 \times 0.5)^{1/2} = 7.07$

• $P(95 \leq X \leq 105) = P(X \leq 105) - P(X \leq 94)$

  $\approx P[ (94 + .5 - 100) / 7.07 \leq Z \leq (105 + .5 - 100) / 7.07 ]$

  $= 0.7823 - 0.2177 = 0.5646$. 
Practice Problem 2

• Suppose 9 observations are drawn from a normal population whose standard deviation is 2. The observations are: 15, 9, 13, 11, 8, 12, 13, 8, 10

• At significance level 0.05, you want to determine whether the mean of the population from which this sample \( \bar{x} = 11 \) was taken is significantly different from 10.
  
  – State the null and alternative hypotheses in symbols.
  – Compute the value of the test statistic.
  – Compute the \( p \)-value.
  – Interpret the results.

• Answers:
  
  – \( H_0: \mu = 10 \) vs. \( H_a: \mu \neq 10 \).
  – \( z = 1.5 \) and \( p \)-value = 0.1336
  – We do not reject \( H_0 \) ... Not enough evidence.
• Professors Smith and Jones are both planning studies of serum glucose levels in Martians. The serum glucose levels are known to follow a normal distribution with a standard deviation of 5 units. Each professor plans to measure a random sample of Martians and to construct a 90% confidence interval for the mean Martian serum glucose level \( \mu \). The sample size will be:

\[
\text{Smith: } n = 30; \text{ Jones: } n = 300.
\]

• Assume that the population of Martians is very large.

Q1: Would we expect Jones' confidence interval to be narrower than, wider than, or about the same width as Smith's confidence interval?

Q2: Is the chance that Jones' confidence interval contains \( \mu \) less than, more than, or the same as the chance that Smith's confidence interval contains \( \mu \)?
Practice Problem 4

• A multiple choice exam has 100 questions. Each question has 4 possible answers, of which only one is correct.

(a) What is the expected number of correct answers by sheer guesswork?

(b) What is the probability that sheer guesswork will yield at least 30 correct answers?

• Answer:

(a) \( 100 \times (1/4) = 25 \)

(b) \[
P(X \geq 30) \approx P\left(Z > \frac{30 - 0.5 - 25}{\sqrt{100 \cdot \frac{1}{4} \cdot \frac{3}{4}}} \right) = P(Z > 1.04) = 0.1492.\]
For Lectures 18 -- 22

- Review Hwk 18 --- 22 problems …