# Supplementary Review Problems for the Final

**Comment:** To prepare for the final exam, you should study these problems as well as the problems from quizzes 1 – 4.

1. There are about 2,700 institutions of higher learning in the United States (including junior colleges and community colleges). In 1976, as part of a continuing study of higher education, the Carnegie Commission took a simple random sample of these institutions. The average enrollment in the 256 sample schools was 3,500, with an SD of 4,000. A histogram for the enrollments was plotted and did not follow the normal curve.

   Say whether each of the following statements is *true* or *false*, and explain why.

   (a) An approximate 95%-confidence interval for the average enrollment of all 2,700 institutions runs from 3,000 to 4,000.

   (b) It is estimated that 95% of the institutions of higher learning in the United States enroll between $3,000$ and $4,000$ students.

   (c) There is an approximately 95% chance that the average enrollment at all institutions of higher learning in the United States is between 3000 and 4000 students.

   (d) The normal curve can't be used to figure confidence levels here at all, because the data doesn't follow the normal curve.

2. A math instructor at a community college wants to teach his class the benefits of practicing. To do this he divides the class into two groups. Students who have practiced hitting a baseball off a tee for more than a year (in high school or little league, for example) go into group A, and the rest of the students go into group B. The class goes out to the school's practice field and every student hits three balls off a tee. The average distance the ball travels for the 37 students in group A is 151 feet, while for the 68 students in group B the average distance the ball travels is 83 feet. At the end of the afternoon, the instructor says "See — practice makes perfect!"

   (a) Is the teacher's study observational or is it a controlled experiment?

   (b) Are there any *confounding* variables?

   (c) Does the baseball-hitting study confirm the instructor's conclusion? Explain.

3. Chapter 14, review question 9.

4. Chapter 26, exercise set F, problems 4 and 5.

5. In a calculus class with 180 students, the final exam score contributed 50% of the course score, the midterm score contributed 30% of the course score and the average homework score contributed 20% of the course score.

   **a.** After the course was over, the instructor computed three correlation coefficients based on the class data:
   - $r_1$ = correlation between average homework score and midterm exam score,
   - $r_2$ = correlation between average homework score and final exam score,
   - $r_3$ = correlation between average homework score and score in the class.

   The three numbers she computed were 0.3521, 0.5582 and 0.4112, but she forgot to label them. Match each number with the appropriate correlation coefficient and explain your choices.

**b.** The same instructor also divided the class into six groups based on average homework scores, using the ranges

$$(0\% - 49\%), \ (50\% - 59\%), \ (60\% - 69\%), \ (70\% - 79\%), \ (80\% - 89\%) \text{ and } (90\% - 100\%).$$

She the computed the average homework score and average final exam score in each group, and computed the correlation coefficient, $r_4$, for the resulting six data points. Do you think that $r_4$ was higher than, lower than or about the same as $r_2$? Explain your answer.

6. Investigators studying the relationship between cigarette smoking and blood pressure in adult men collected data from 6235 U.S. men aged 20 - 40, and generated the following statistics:

$$\overline{X} = 24 \quad SD_X = 5.5$$
$$\overline{Y} = 135 \quad SD_Y = 9 \qquad r = 0.7$$

where $X$ = number of cigarettes per day, and $Y$ = systolic blood pressure, measured in mmHG.

(a) Use the *regression method* to estimate the average systolic blood pressure for U.S. men, aged 20 - 40 who smoke 20 cigarettes per day. *Show your work.*

(b) What is the predicted systolic blood pressure of a 28-year old man who smokes 30 cigarettes per day? Include a '*give-or-take*' number with your estimate. *Show your work.*

(c) Joseph is a 60-year old man who smokes about 40 cigarettes a day. Is it reasonable to predict that his systolic blood pressure is somewhere between 147 and 160 mmHG, based on the given information? *Explain your answer.*

7. John Smith is running for office. One week before the election, his campaign manager hires a Polling firm to survey likely voters. The firm surveyed a simple random sample of 2700 likely voters and found that 51% favor Smith. They also found that of the 1250 women in the survey, 54% favor Smith.

You may assume that the survey was based on a simple random sample, that the population is in the millions and that to win the office, the candidate needs to win more than 50% of the votes cast.

**a.** What percentage of the men in the survey favor Smith?

**b.** Compute 95% confidence intervals for the percentage of women who favor Smith, the percentage of men who favor Smith and the percentage of likely voters who favor Smith.

*(\*)* *The following two questions are based on the NAEP (National Assessment of Educational Progress) 2009 survey of science proficiency in 4th, 8th and 12th grades. The scores were scaled so that the national average in all three grade levels was set to 150 (out of 300).[†]*

8. The survey averages for females and males in all three grade levels are summarized in the table below (with standard errors in parentheses).

| Gender | Fourth grade | Eighth grade | 12 grade |
|--------|--------------|--------------|----------|
| Male   | 151 (0.3)    | 152 (0.4)    | 154 (0.9) |
| Female | 149 (0.3)    | 148 (0.3)    | 146 (0.9) |

**a.** Construct 95%-confidence intervals for the average score in each category in the nation as a whole.

**b.** In which grade is the difference between the average score of males and and average score of females most **significant**, statistically speaking? Explain you answer.

---

[†]To serve as a benchmark for future tests.

**9.** The 2009 NAEP science proficiency survey also breaks down the sample data by race and ethnicity. A researcher sees that the survey reveals that the average score of *white* 12th graders is 159 and the average score of *Asian/Pacific Island* 12th graders is 164.

He concludes that the national *percentage* of Asian/Pacific Islander 12th graders who are proficient or above in science is likely higher than the national percentage of White 12th graders who are proficient or above in science.

To test this conclusion, he breaks down the 2009 survey data for 12th graders by proficiency levels and finds that

- the sample percentage of white 12th grade students who are proficient or above in science is 27%, with a standard error of 0.8%;
- the sample percentage of Asian/Pacific Island 12th grade students who are proficient or above in science is 36%, with a standard error of 3.8%.

**a.** What are the appropriate null and alternative hypotheses the researcher should use to test his belief?

**b.** What is the test statistic?

**c.** What is the p-value?

**d.** What do you conclude? Are the results significant? Highly significant? Neither?

**e.** What, if anything, is wrong with this study/test of significance?

**10.** As part of a class project, a statistics student at a large university (15,000 students — 9000 men and 6000 women), went to the central plaza of the campus at noon one day, approached 100 students and asked them where they went to high school. His sample included 51 women and 49 men. Is it likely that the student's sampling procedure was like taking a simple random sample? Justify your answer as precisely as possible (using numbers, probability, etc.).

**11.** According to the 1999 census, the median household income in the city of San Diego was $46,500. In 2004, a high-end grocery chain hires a statistical research firm to corroborate their marketing consultant's claim that median household income has gone up since 1999. The research firm takes a simple random sample of 600 San Diego households and finds that 55% of the sample households have incomes above $46,500.

Was the consultant right?

**12.** Suppose that a fair die is rolled 600 times.

**a.** What is the expected number of ⚀s?

**b.** What is the probability that a ⚀ is observed between 95 and 105 times?

**c.** What is the probability that more than 110 ⚀s are observed?

**13.** There are about 25,000 high schools in the United States and each high school has a principal. These 25,000 high schools also employ a total of about one million teachers. As part of a national survey of education, a simple random sample of 625 high schools is chosen.

(a) In 505 of the sample high schools the principal has an advanced degree. If possible, find an approximate 95% confidence interval for the percentage of all 25,000 high school principals who have advanced degrees. If this is not possible, explain why not.

(b) As it turned out, the 625 sample high schools described above employed a total of 12,000 teachers, of whom 6,500 had advanced degrees. If possible, find an approximate 95% confidence interval for the percentage of all one million high school teachers with advanced degrees. If this is not possible, explain why not.