Comment: *To prepare for the final exam, you should study these problems, the homework and the three midterms. The final exam will be comprehensive, with an emphasis on untested material (including the material of Midterm Exam 2).*

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

2. Every day, the quality control engineer for ACME Dairies randomly selects 25 half-gallon (64 fl oz) cartons of whole milk from the day's production run and carefully measures the quantity of milk in each one. If the average amount of milk in this sample differs *significantly* from 64 fl oz, at the 1% significance level, she recalibrates the carton-filling apparatus.

(a) What does '*...significantly... at the 1% significance level*' mean?

(b) State the null and alternative hypotheses for this test in terms of the appropriate parameter.

(c) Today's sample of 25 cartons has an average of 64.21 fl oz with a standard deviation of 0.37 fl oz. What is the test statistic? What probability distribution does it follow? What is the p-value?

(d) What do you conclude?

(e) What additional assumptions, if any, are needed to justify the methods (and conclusions) of this test of significance?

3. A marketing firm wants to survey a sample of adults from a large state to estimate the percentage of adults in the state who prefer streaming movies to watching them on cable TV. Describe the advantages and disadvantages of each of the following sample types that they might use: (a) a *simple random sample*, (b) a *convenience sample* or (c) a *multistage cluster sample*.

4. A researcher claims to have found a strong correlation ($r = 0.88$) between a person's blood alcohol content (BAC), one hour after drinking, and the *type* of alcohol they consume: beer, wine or hard liquor. What is wrong with the researcher's claim? What would make more sense here? Explain.

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. 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.

6. Investigators studied the relationship between screen time (measured in hours/day) and obesity (measured with *body-mass-index* BMI) in adults age 20 - 40. They surveyed 5178 U.S. adults in this age group, and generated the following summary statistics:

$$\overline{X} = 10 \quad SD_X = 6$$
$$\overline{Y} = 27 \quad SD_Y = 9 \qquad r = 0.6$$

where X = hours of screen-time per day, and Y = BMI.

(a) Use the *regression method* (or regression equation) to estimate the BMI for U.S. adults, aged 20 - 40 who have 7 hours of screen time per day. *Show your work.*

(b) What is the predicted BMI of a 28-year old woman who has 15 hours of screen-time per day? Include a '*give-or-take*' number with your estimate. *Show your work.*

(c) Ivan is a 24-year old Russian graduate student at UCLA, who has about 12 hours of screen-time per day. Is it reasonable to predict that his BMI is somewhere between 21.6 and 36, 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 and that the population is in the millions.

   (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.
   (c) Based on your answer to (b), how would you suggest that Smith focus his advertising money in the last few days before the election? Explain.

8. 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.).

9. 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? Frame your answer in terms of an appropriate test of significance.

10. Suppose that a fair die is rolled 3 times.

    a. What is the probability that a ⚀ is observed ***at least once***?
    b. What is the probability that a ⚀ is observed ***exactly once***?
    c. What is the probability of that the ***sum*** of the three rolls is 4 or 5?

11. 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?
    d. What is the probability that the ***sum*** of the 600 rolls is between 2070 and 2130?

12. 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 you can, find an approximate 95% confidence interval for the percentage of all 25,000 high school principals who have advanced degrees. If you cannot, explain why not.
    (b) As it turned out, there were 250,000 students enrolled in the 625 sample high schools described above. These 250,000 students spent an average of 10.7 hours per week on homework, with a standard deviation of 3.5 hours. If you can, find an approximate 95% confidence interval for the average number of hours per week spent on homework of all U.S. high school students. If you cannot, explain why not.

13. A researcher studying the media consumption habits of U.S. adults suspects that Californians watch more 'reality' shows than New Yorkers. To test this hypothesis, she surveys a simple random sample of 1225 Californians and a simple random sample of 1444 New Yorkers. The New Yorkers surveyed watched an average of 4.36 hours per week of 'reality' shows, with an SD of 1.8 hours per week. The Californians watched an average of 4.43 hours per week of 'reality' shows, with an SD of 1.7 hours per week.

(a) Formulate appropriate null and alternative hypotheses in terms of a box model to test the researcher's hypothesis at the 5% significance level.

(b) Find the test statistic and the P-value.

(c) What is your conclusion? Is the researcher right? In what sense? Explain.

14. Chapter 27, Review problem 7.