

Comment: To prepare for the final exam, you should study these problems, the homework from chapters 26 - 27, as well as reviewing the material from quizzes 1 - 4. The final exam will be comprehensive, with an emphasis on untested material.

- Chapter 26, exercise set F, problems 4 and 5.
- 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.
 - What does '*...significantly... at the 1% significance level*' mean?
 - State the null and alternative hypotheses for this test in terms of the appropriate parameter.
 - 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 is the p-value?
 - What do you conclude?
 - What additional assumptions, if any, are needed to justify the methods (and conclusions) of this test of significance?
- A marketing firm wants to survey a sample of 1400 adults from a large state to estimate the percentage of adults in the state who prefer streaming movies to watching them on cable TV. Are they more likely to use (a) a *simple random sample*, (b) a *convenience sample* or (c) a *multistage cluster sample* to choose the sample?

Explain your answer.

- 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.
- 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:

$$\begin{aligned} \bar{X} &= 24 & SD_X &= 5.5 \\ \bar{Y} &= 135 & SD_Y &= 9 & r &= 0.7 \end{aligned}$$

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

- Use the *regression method* (or regression equation) to estimate the average systolic blood pressure for U.S. men, aged 20 - 40 who smoke 20 cigarettes per day. *Show your work.*
 - 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.*
 - 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.*
- 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.
7. 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.).
8. 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.
9. Suppose that a fair die is rolled 3 times.
- a. What is the probability that a \square is observed *at least once*?
- b. What is the probability that a \square is observed *exactly once*?
- c. What is the probability of that the *sum* of the three rolls is 4 or 5?
10. Suppose that a fair die is rolled 600 times.
- a. What is the expected number of \square s?
- b. What is the probability that a \square is observed between 95 and 105 times?
- c. What is the probability that more than 110 \square s are observed?
- d. What is the probability that the *sum* of the 600 rolls is between 2070 and 2130?
11. 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 you can, find an approximate 95% confidence interval for the percentage of all one million high school teachers with advanced degrees. If you cannot, explain why not.
12. A researcher studying the media consumption habits of U.S. adults suspects that women watch more 'reality' shows than men. To test this hypothesis, she surveys a simple random sample of 1225 U.S. men and a simple random sample of 1444 U.S. women. The men surveyed watched an average of 4.36 hours per week of 'reality' shows, with an SD of 1.8 hours per week. The women 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) Is the researcher right? In what sense? Explain.
13. Chapter 27, Review problem 7.