

Old Sample Final

Math 125 Spring 2006

To get full credit you **must show your work**. No work, no credit.

- Using the data below, draw the histogram for the distribution of men by number of books read last month. Label the units of the vertical scale.

Number of books	0	1	2-3	4-6	7-10	11-15	16 or more
Percent of men	29	15	24	15	12	5	0

- It was found that 74 percent of the freshman class at U Mass/Boston scored over 450 points on the verbal section of the SAT. If the verbal SAT scores for the entire class have an SD of 80 points and follow the normal curve, what is the average?

- Find the correlation coefficient for the data set below.

x	y
10	15
8	25
7	19
5	13
6	17
9	23
4	21

- A statistical analysis is made of the midterm and final scores in a large course, with the following results:

average midterm score ≈ 60 , SD ≈ 15 ,
 average final score ≈ 65 , SD ≈ 20 , $r \approx 0.50$.

The scatter diagram is football-shaped.

- About what percentage of students scored over 80 on the final?
 - Of the students who scored 80 on the midterm, about what percentage scored over 80 on the final?
- For the 988 men age 18-24 in the HANES sample

average height ≈ 70 inches SD ≈ 3 inches
 average weight ≈ 162 pounds SD ≈ 30 pounds
 correlation ≈ 0.47

One man in the sample was 66 inches tall and weighed 140 pounds. In comparison with the other men in the sample of the same height, this man would be

- (i) a little light (ii) a little heavy

Circle one option and explain your choice.

6. In 1976 baseball instituted what was called the “free agent draft.” In essence, this gave players the right to negotiate a contract with any team they chose to, rather than belonging to a particular team forever or until traded. A few of the wealthiest clubs (like the Yankees) promptly paid enormous sums to obtain a few players who had very high batting averages the previous year. During the course of the following year, sports writers had lots of fun pointing out what bad judgement the owners had shown, for almost none of these high-priced players did as well that year as they had done the year before. Do you think all that money made the players fat and lazy? Or do you have another explanation?
7. 60% of the pupils in a particular university are male. What is the probability that a random sample of 400 pupils will contain
 - (a) fewer than 220 males?
 - (b) exactly 220 males?
8. A simple random sample of 1,000 persons is taken to estimate the percentage of Democrats in a large population. It turns out that 543 of the people in the sample are Democrats. The sample percentage is $(543/1,000) \times 100\% = 54.3\%$. The SE for the sample percentage of Democrats is figured as 1.6%. True or false, and explain:
 - (a) $54.3\% \pm 3.2\%$ is a 95%-confidence interval for the percentage of Democrats in the population.
 - (b) $54.3\% \pm 3.2\%$ is a 95%-confidence interval for the percentage of Democrats in the sample.
 - (c) There are about two chances in three for the percentage of Democrats in the population to be in the range $54.3\% \pm 1.6\%$.
 - (d) If 2,000 such samples were taken of the same large population and estimates were made in each case, about 91 of the 2,000 estimates would be off the actual percentage of Democrats in the large population by more than 3.2 percentage points.
9. The unconditional probability of event A is $1/3$; the unconditional probability of event B is $1/10$. True or false, and explain:
 - (a) If A and B are independent, they must also be mutually exclusive.
 - (b) If A and B are mutually exclusive, they cannot be independent.
10. A poker hand of five cards is selected at random without replacement. Find the chance that the hand contains (keep in mind that a picture card is a Jack, Queen, or King):
 - (a) All picture cards.
 - (b) No picture cards.
 - (c) At least one picture card.

11. Los Angeles has about four times as many registered voters as San Diego. A simple random sample of registered voters is taken in each city, to estimate the percentage who will vote for school bonds. Other things being equal, a sample of size 4,000 taken in Los Angeles will be about

- (i) four times as accurate
- (ii) twice as accurate
- (iii) as accurate

as a sample of size 1,000 taken in San Diego. Choose one option and say why.

12. Twenty-five measurements are made on the speed of light. These average out to 300,007 and the SD is 10, the units being kilometers per second. Fill in the blanks in part (a), then say whether each of (b–f) is true or false; explain your answers briefly. (You may assume the Gauss model, with no bias.)

- (a) The speed of light is estimated as _____; this estimate is likely to be off by _____ or so.
- (b) Each measurement is off 300,007 by 10 or so.
- (c) The average of all 25 measurements is off 300,007 by 2 or so.
- (d) If a 26th measurement were made, it would be off the exact value for the speed of light by 2 or so.
- (e) A 95%-confidence interval for the speed of light is $300,007 \pm 4$.
- (f) A 95%-confidence interval for the average of the 25 measurements is $300,007 \pm 4$.

13. A survey organization takes a simple random sample of 625 households from a city of 80,000 households. On the average, there are 2.30 persons per sample household, and the SD is 1.75. Say whether each of the following statements is true or false, and explain.

- (a) The 2.30 is 0.07 or so off the average number of persons per household in the whole city.
- (b) A 95%-confidence interval for the average household size in the sample is 2.16 to 2.44.
- (c) A 95%-confidence interval for the average household size in the city is 2.16 to 2.44.
- (d) 95% of the households in the city contain between 2.16 and 2.44 persons.
- (e) Household size in the city follows the normal curve.
- (f) The 95%-confidence level is about right because household size follows the normal curve.

14. A butter machine makes sticks of butter that average 4.0 ounces in weight, with an SD of 0.16 ounces. There is no trend or pattern in the weights. There are 4 sticks to the package.
- (a) A package weighs _____, give or take _____ or so.
 - (b) A store purchases 100 packages. Estimate the chance that it gets 100 pounds of butter, to within 8 ounces.
15. In a long series of trials, a computer program is found to take on average 23.7 seconds of CPU time to execute, with an SD of 1.23 seconds. There is no trend or pattern in the times. It will take about _____ seconds of CPU time to execute the program 400 times, give or take _____ seconds or so. (The CPU is the “central processing unit,” where the machine does logic and arithmetic.)
16. A group of 10,000 tax forms shows an average gross income of \$17,000, with an SD of \$10,000. Furthermore, 10% of the forms show a gross income over \$30,000. A group of 900 forms is chosen at random for audit. To estimate the chance that somewhere between 9% and 11% of the forms chosen for audit show gross incomes over \$30,000, a box model is needed.
- (a) Should the number of tickets in the box be 900, or 10,000?
 - (b) Each ticket in the box shows
a zero or a one a gross income
 - (c) True or false: The SD of the box is \$10,000.
 - (d) True or false: The number of draws is 900.
 - (e) Find the chance (approximately) that between 9% and 11% of the forms chosen for audit show gross incomes over \$30,000.
 - (f) A supervisor randomly chooses seven of the audited forms for double-checking. Find the chance that exactly three of the seven chosen by the supervisor show gross incomes over \$30,000.
 - (g) Find—if possible from the information given—the chance that a tax return selected at random will show a gross income of \$27,000 or more.
17. A certain town has 25,000 families. These families own 2.1 cars, on the average, with an SD of 0.80. And 10% of them have no cars at all.
- (a) If one draws a simple random sample of six families from the 25,000, what is the chance that exactly two of the six sample families will not own cars?
 - (b) If one draws a simple random sample of 1,600 families from the 25,000, what is the chance (approximately) that somewhere between 9% and 11% of the sample families will not own cars? (Any correction factors may be ignored.)

18. A survey organization wants to take a simple random sample in order to estimate the percentage of people who have seen a certain television program. To keep the costs down, they want to take as small a sample as possible. But their client will only tolerate chance errors of $1/4$ of 1 percentage point or so in the estimate. Should they use a sample of size 100, 2,500, 10,000, 40,000, or 1,000,000? (See the Table in Review Exercise 1 on p. 371 of the text; you may assume the population to be very large.)
19. Would taking the average of 225 measurements divide the likely size of the chance error by a factor of 5, or 10, or 15, or 25?
20. (a) One day, upon tossing a single die 60 times, I got:

5 ones, 7 twos, 17 threes, 16 fours, 8 fives, and 7 sixes.

Compute χ^2 and find P for this experiment.

(b) Another day, upon tossing the same single die 600 times, I got:

90 ones, 110 twos, 100 threes, 80 fours, 120 fives, and 100 sixes.

Compute χ^2 and find P for this experiment.

(c) Now, compute the pooled χ^2 using the combined degrees of freedom, and find the pooled P -value.

Is the die biased, based on the combined evidence?
21. Four hundred draws are made at random with replacement from a box. The average of the draws is 1084.7, with an SD of 600. Someone claims that the average of the box equals 1000. Is this plausible?
22. According to one investigator's model, his data are like 400 draws made at random from a large box. The null hypothesis is that the average of the box equals 80; the alternative is that the average of the box is more than 80. In fact, his data averaged out to 85.4, with an SD of about 45. Compute z and P . What should he conclude?
23. A vocabulary test for six-year-old children is standardized on a large nationwide sample to have an average score of 50 out of 100 and an SD of 11.6. School authorities in Maine choose a statewide random sample of 400 six-year-olds. These children average 51.3 on the test, with an SD of 12. Is it safe to infer that if the test had been administered to all six-year-olds in Maine, they would have averaged above 50? Or can the 1.3 point difference be explained as a chance variation?
 - (a) Does this call for a one-sample z -test or a two-sample z -test?
 - (b) Formulate the null and alternative hypotheses as statements about a box model.
 - (c) Compute z and P .
 - (d) Was it a chance variation?

24. There are about 2,500 colleges and universities in the United States. The National Center for Educational Statistics calculated the average undergraduate enrollment in 1974 to have been around 3,700 students per institution. Current data is hard to get, so one organization takes a simple random sample of 225 institutions. Among the sample institutions, the average enrollment in the current year is 3,500 students per institution, with an SD of 6,000. Does this show that the average enrollment among all colleges and universities has gone down? Or can the difference be explained as a chance variation? Formulate the null and alternative hypotheses in terms of a box model before deciding.
25. You are taking a plane trip and have heard that the chances of someone bringing a bomb on board are 1 in 1000. You are a little worried, but then you read that the chances of *two* people (independently) bringing a bomb on the same plane are 1 in 1,000,000. What should you do? (Bring a bomb on board?)