

MATH 140 SEMESTER REVIEW PROBLEMS

1. The following data are the scores of 30 students on a statistics examination.

69 72 89 50 84 55 75 62 87 13 96 90 91 75 81
92 33 94 82 78 73 67 64 71 50 85 72 88 67 56

- a. Make a stem plot of the data, and calculate the median, mean, s.d., and quartiles.
- b. Use the results of (a) to draw a box plot of the data, and describe the important features of the distribution.

2. T F The median and quartiles are not strongly affected by outliers; the mean and standard deviation are more strongly affected by outliers.

3. Twenty-five years ago, a sample of the ages of art objects in a museum had mean 486 years, and standard deviation 223 years. If this same particular sample were taken today, what would the mean and standard deviation be?

4. Would you rather score 80 on a test with a mean of 75 and standard deviation of 10 or score 70 on a test with a mean of 65 and a standard deviation of 5? Or are they equivalent? Explain. Assume both tests have normally distributed scores.

5. What is the probability that a normal random variable takes on values between one and 1.5 standard deviations above its mean?

6. In a sample of 1760 hospitals across the country it was found that the average length of stay for patients under 65 years of age was approximately 6 days. Assuming a normal distribution of length of stay for those under 65 with an s.d. of 2.8 days,

- a. what is the probability that a patient will need to stay 7 or more days?
- b. what is the probability that a group of 60 patients, entering under age 65, will consume a total of 420 or more days of hospital resources? (Look at what this means about the average stay for the group)

7. Suppose the grades of students on a statewide examination have a mean of 70 and an s.d. of 12. If these grades are normally distributed what is their 90th percentile?

8a. T F If data on some variable are strongly skewed to the right, then the mean will be much greater than the median.

8b. T F If the correlation coefficient is close to +1, then it can be inferred that increases in one quantitative variable cause increases in another quantitative variable.

8c. T F The correlation measures the strength and direction of the linear association between two quantitative variables.

8d. T F It is difficult to argue a cause-and effect type relation from an observational study; however, a randomized comparative experiment can provide good evidence for a cause-and-effect relation.

9. Which pairs of variables have a negative correlation?

- the weight and height of a student in a Math 140 class
- the age of a child and the number of times it cries each day
- the SAT score and the college GPA of a student
- the outside temperature on a given day and the number of sodas sold on campus that day
- the age of a car and the cost per year for repairs

10. A college newspaper interviews a psychologist about a proposed system for rating the teaching ability of faculty members. The psychologist says, “The evidence indicates that the correlation between a faculty member’s research productivity and teaching rating is close to zero.” Which of the following is the best interpretation of this statement?

- a. Good researchers tend to be poor teachers and vice versa.
- b. Good teachers tend to be poor researchers and vice versa.
- c. Good researchers are just as likely to be good teachers as they are bad teachers; likewise for poor researchers.
- d. Good research and good teaching go hand in hand.

11. Suppose that we are interested in studying the relationship between body weight and the percentage of calories consumed as fat for American adults.

- a. What is the response variable?
- b. What is the explanatory variable?
- c. If the body weight and the percentage of calories consumed as fat were measured for a sample of 20 American adults, what do you think the scatterplot would look like? Make a sketch. Include the variable names, but not numerical values.
- d. In one word, describe the kind of association that you think is present between these two variables.

12. The following data represents the yield y of a chemical reaction at various temperatures x :

$x(F^\circ)$	y
150	77.4
150	76.7
160	78.2
180	84.5
200	83.9
200	83.7
225	85.6
250	88.9
250	90.3
280	94.8

The summary statistics are:

$$\bar{x} = 204.5, \bar{y} = 84.4, s_x = 45.7, s_y = 5.89, \text{ and } r = 0.975.$$

- a. Find the equation of the regression line of y on x .
- b. What yield is predicted for the temperature $x = 190$?
- c. What is the residual associated with the observation (225, 85.6)?
- d. The yield is predicted to increase by _____ for every 1° increase in temperature.

13. A statistics instructor does a linear regression of final exam scores versus midterm scores for the 32 students in his class. The correlation, r , is 0.8. The regression equation is: $\text{Final} = 10 + 0.9 * \text{Midterm}$. One student, Mary, got a 90 on the midterm.
- Predict Mary's final exam score, based on the regression.
 - In fact, Mary's final exam score was 98. What is her residual? (If you could not do part a., assume the answer to that part is 93.)
 - What percent of the variation in final exam scores is explained by its linear relation to midterm scores?

14. A study of the effect of abortions on the health of the subsequent children was conducted as follows. The names of women who had abortions were obtained from medical records in New York City hospitals. Birth records were then searched to locate all women in this group who bore a child within 5 years of the abortion. The hospital records were examined again for information about the health of the newborn child. Was this an observational study or an experiment? Explain your answer.

15. Pick a random sample of 5 students from this list of 12 students using the random digit table below. Explain exactly how you did it.

Annabelle	Barbara	Carlos	Damon			
Ezra	Ferdinand	Gary	Hiawatha			
Isabella	Judy	Keith	Lamont			
03802	29341	29264	80198	12371	13121	54969
43912	77320	35030	77519	41109	98296	18984
87065	74133	21117	70595	22791	67306	28420
52067	87370	88099	89695	87633	76987	85503

16. T F Any statistic that we compute from a random sample will vary from sample to sample and have a sampling distribution.
17. T F Statistical inference draws conclusions about populations, either real or conceptual, on the basis of results from a sample.
18. A simple random sample of 1000 Americans found that 61% were satisfied with the service provided by the dealer from which they bought their car. A simple random sample of 1000 Canadians found that 58% were satisfied with the service provided by the dealer from which they bought their car. The sampling variability associated with these statistics is what?
- about the same
 - much smaller for the sample of Canadians since the population of Canada is much smaller than that of the United States, hence the sample is a larger proportion of the population
 - smaller for the sample of Canadians since the percentage satisfied was smaller than that for the Americans
 - larger for the Canadians, since Canadian citizens are more widely dispersed throughout the country than in the United States, hence have more variable views

19. A telemarketing firm in Los Angeles uses a device that dials residential telephone numbers in that city at random. Of the first 100 numbers dialed, **23** are unlisted. This is not surprising because **38%** of all Los Angeles residential phone numbers are unlisted. Identify whether each underlined-boldface number is a statistic or a parameter.

20. The central limit theorem says that when a simple random sample of size n is drawn from any population with mean μ and standard deviation σ , then when n is sufficiently large:

- the standard deviation of the sample mean is σ / \sqrt{n}
- the distribution of the population is exactly normal
- the distribution of the sample mean is approximately normal
- the distribution of the sample mean is exactly normal

21. At a major university, 60% of the students entering the university as engineering majors either switch to a different major or drop out before graduation. If the records of 45 randomly selected entering engineering students are examined, what is the probability that more than half of them graduate as engineers?

22. G.E. light bulbs are guaranteed for 1000 hours or your money back. Suppose the actual lifetime is normally distributed with mean 1150 hours and standard deviation 100 hours.

- What percent of their light bulbs will fail to meet the guarantee?
- Suppose packs of 4 lightbulbs are sold. What is the probability the average of the 4 lasts less than 1100 hours?

23. A random sample of 101 measurements of uncongested freeway driving speeds is taken. The results are $\bar{x} = 57.3$ m.p.h. and $s = 6.0$ m.p.h. Construct a 95% confidence interval for the mean driving speed.

24. You have measured the systolic blood pressure of a random sample of 35 employees of a company. You use the data to compute the 95% confidence interval for the mean systolic blood pressure, and it is (122, 138). Which of the following are valid interpretations of this interval?

- 95% of the sample have a systolic blood pressure between 122 and 138.
- 95% of all the employees in the company have a systolic blood pressure between 122 and 138.
- If this procedure were repeated many times, 95% of the resulting confidence intervals would cover the mean systolic blood pressure for all employees of this company.
- The true population mean, μ , is very likely to be somewhere in this interval from 122 to 138, but is not certain to be.

25. Suppose that the Gallup organization takes 300 one-question polls in a given year. Each time the sample size is 1571. About how many times would you expect Gallup's 95% confidence interval to contain the true population percentage?

26. A survey was conducted at a movie theater to determine movie-goers' preference for different kinds of popcorn. The results of the survey showed that Brand A was preferred by 65 percent of the people with a margin of error of plus or minus 3 percent. What is meant by the statement "plus or minus 3 percent"?

- a. Three percent of the population that was surveyed will change their minds.
- b. Three percent of the time the results of such a survey are not accurate.
- c. Three percent of the population was surveyed.
- d. The true proportion of the population who preferred Brand A popcorn could be determined if 3 percent more of the population was surveyed.
- e. It would be unlikely to get the observed sample proportion of 65 percent unless the actual percentage of people in the population of movie-goers who prefer Brand A is between 62 percent and 68 percent.

27. You compute a 95% confidence interval and a 99% confidence interval from the same data. Circle the letter of the correct statement:

- a. The intervals have the same width.
- b. The 95% confidence interval is wider.
- c. The 99% confidence interval is wider.
- d. You cannot determine which interval is wider unless you know n .
- e. You cannot determine which interval is wider unless you know the s.d.

28a. A school board wishes to know the current mean reading level of 6th graders throughout a very large school district. Past experience shows that the standard deviation of such reading scores is about 2.5. If they wish to be 95% sure that their result is correct to within .4, how large a sample do they need to have?

28b. In order to avoid less cooperative schools, the school board decides to choose their sample of 6th graders from schools who volunteer to participate in the testing. Criticize this and suggest a better sampling method.

29. A *New York Times* poll interviewed 1025 women and 472 men randomly selected from the U.S. The poll found that 47% of the women said they do not get enough time for themselves.

- a. Why can't we just say that 47% of all adult women do not get enough time for themselves?
- b. What is the margin of error for the percentage of women who say they do not get enough time for themselves? (A 95% confidence level is standard for national polls.)
- c. The margin of error for men announced by the poll 4%. Why is this larger than the margin of error for women?
- d. Explain what "95% confidence" means here.

30. Do people have more mental disorders during the full moon? The average number of admissions per day to the emergency room of a mental hospital in Virginia in 1972 was 11.2. The standard deviation was 5.5. The number of admissions per day were normally distributed. During the 12 days of the year with the fullest moon, the average admission rate was 13.3. (continued on next page)

- a. In words, what is the null hypothesis?
 - b. Draw a picture of the sampling distribution of \bar{x} for a true null hypothesis. Include a scale on the x-axis.
 - c. Compute the z-score.
 - d. Do you have evidence to reject the null hypothesis?
31. T F Very large z or t scores correspond to very strong evidence in favor of the null hypothesis.
32. T F Very small p -values are strong evidence against H_o .
33. T F It is possible to reject a null hypothesis that is true, even with good methodology and a good random sample.
34. When performing a test of significance for a null hypothesis, H_o , against an alternative hypothesis, H_a , the p -value is
- a. the probability that H_o is true.
 - b. the probability that H_a is true.
 - c. the probability that H_o is false.
 - d. the probability of observing a value of a test statistic at least as extreme as that observed in the sample if H_o is true.
 - e. the probability of observing a value of a test statistic at least as extreme as that observed in the sample if H_a is true.
35. A medical report states that the mean time for patients to gain complete mobility from hip replacement surgery where stainless steel joints are used is 115 days. A new nylon joint is being introduced. Test that the mean time for patients to gain mobility with this new joint is less than that with the steel joint if a random sample of 10 patients receiving this new joint produced the mobility recovery times (in days) below. Use $\alpha = .025$ and don't forget to estimate the p -value. You may assume that the population of mobility recovery times has a normal distribution.
- Data: 109, 111, 102, 120, 104, 118, 112, 101, 125, 94
36. A TV station claims that 38% of the 6:00 - 7:00 pm viewing audience watches its evening news program. A consumer group believes this is too high and plans to perform a test at the 5% significance level. Suppose a sample of 830 viewers from this time range contained 282 who regularly watch the TV station's news program. Carry out the test and compute the p -value.

37. The heights (in cm) of an SRS of nine children are given below at their 11th and 12th birthdays:

Child	1	2	3	4	5	6	7	8	9
11 th Birthday	113	116	106	105	106	118	116	103	115
12 th Birthday	119	120	111	109	110	122	119	109	121

Find a 95% confidence interval for the average growth of children between the ages of 11 and 12. State any assumptions you need to make.

38. The knee velocity of a sample of skilled rowers (crew) is compared with a sample of novice rowers in a study attempting to characterize differences between skilled and novice rowers. The results of this comparison are given below:

GROUP	n	\bar{x}	s
Novice	8	3.82	.89
Skilled	10	4.44	.52

Is there significant evidence at the 5% level that the knee velocity of skilled rowers is greater than that of novice rowers?

39. Cuckoos lay their eggs in the nests of other birds. Some biologists speculate that the size of the cuckoo's eggs might vary depending on whether the eggs are laid in warblers' nests or wrens' nests. To check this, biologists searched a wildlife refuge for warblers' or wrens' nests; data on the lengths of the cuckoo's eggs found in these nests are shown below:

Eggs from warblers' nests: $n_1 = 29, \bar{x}_1 = 22.20, s_1 = 0.65$

Eggs from wrens' nests: $n_2 = 35, \bar{x}_2 = 21.12, s_2 = 0.75$

Use a 99% confidence interval to determine if these data support the biologists' speculation that the size of the eggs differs depending on whether they are laid in warblers' or wrens' nests. Remember to draw a conclusion from your calculation!

40. It is claimed that Democrats are more likely than Republicans to favor publicly funded television. 500 Democrats and 400 Republicans are chosen at random. 420 of the Democrats and 300 of the Republicans favor publicly funded television. Test the claim.

ANSWERS

1. 1|3 Median: $(73+75) \div 2 = 74$ Mean: 72.03 S.D.: 18.76
 2| 1st Quartile: 64 3rd Quartile: 87
 3|3
 4|
 5|0 0 5 6
 6|2 4 7 7 9
 7|1 2 2 3 5 5 8
 8|1 2 4 5 7 8 9
 9|0 1 2 4 6

2. T

3. Mean: $486 + 25 = 511$ S.D.: 223 (unchanged)

4. Compare $Z = (80 - 75) \div 10 = 0.5$ to $Z = (70 - 65) \div 5 = 1.0$. The score of 70 will correspond to a higher percentile than the score of 80 since it has a higher z-score.

5. .0919

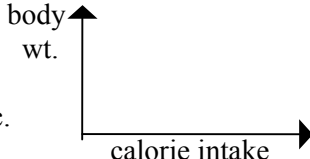
6. a. $P(Z \geq 0.36) \cong .36 = 36\%$.
 b. $\bar{X} = 420 \div 60 = 7$. $P(\bar{X} \geq 7) = P(Z \geq 2.77) = .0028$.

7. 85

8. a. T b. F (shows association, not causation) c. T d. T

9. Age of child and no. times it cries.

10. c

11. a. body weight b. % calories consumed as fat c.  d. positive

12. a. $y = 58.7 + 0.126x$ b. 82.6 c. -1.4 d. 0.126

13. a. 91 b. $98 - 91 = 7$ c. $0.8^2 = 64\%$

14. An observational study. No treatment was imposed on the subjects.

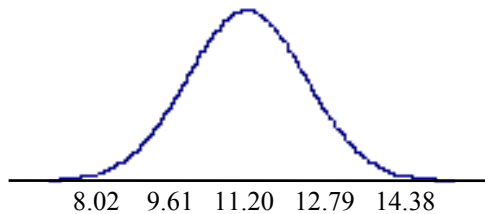
15. (One way) Number the students as follows: 01 = Annabelle, 02 = Barbara, 03 = Carlos, etc., then search the random number table two digits at a time for the first 5 different numbers between 01 and 12. Reading across the rows we find:

03|80|22|93|41|29|26|48|01|98|12|37|11|31|21|54|96|94|39|12|77|32|03|07

which gives a sample of Carlos, Annabelle, Lamont, Keith and Gary.

16. T 17. T 18. a 19. **23** is a statistic, **38%** is a parameter.

20. c 21. 9% 22. a. .0668 \cong 7% b. .1587
23. 57.3 ± 1.18 , or (56.1, 58.5) 24. c and d 25. $95\% \times 300 = 285$
26. e 27. c
28. a. 151
 b. Voluntary response leads to bias. Use a random sample and see to it that every school participates.
29. a. The population proportion will normally differ from the sample proportion. b. 3.1%
 c. Because the sample size for men is smaller.
 d. If this poll were repeated many times, 95% of the time the sample proportion would be within 3.1% of the true proportion of women who say they do not get enough time for themselves.
30. a. The full moon has no effect on the number of mental disorders.
 b. c. 1.32 d. No--Z is not extreme enough (p -value = .0934)



31. F 32. T 33. T 34. d
35. $t = -1.78$ on 9 d.f., which is not significant at $\alpha = .025$ since t does not fall beyond the critical value of -2.262 . The p -value is slightly greater than 5%. There is not sufficient evidence to conclude that the new type of joint will produce a lower mean recovery time.
36. $\hat{p} = 282 \div 830 = .3398$. The p -value is $P(\hat{p} \leq .3398)$ (assuming $H_0: p = .38$ is true) = $P(Z \leq -2.39) = .0084$. There is good evidence that the consumer group is right, as the chance of such a low sample proportion is so small ($p < .01$).
37. 4.67 ± 0.86 , or (3.81, 5.53). We must assume that the amounts that children grow between their eleventh and twelfth birthdays (i.e., the differences in height) are normally distributed.
38. $t = 1.75$. Using 7 d.f., the results are not significant since the t -value does not exceed the 5% critical value of 1.89.
39. The 99% confidence interval for the mean weight difference $\mu_1 - \mu_2$ is 1.08 ± 0.48 (using 28 d.f.), or (0.60, 1.56). Since the interval contains only positive numbers, the data supports the biologists' speculation that there is a difference in egg size according to the type of nest.
40. $H_0: p_D = p_R$; $H_a: p_D > p_R$. $Z = 3.35 \implies$ very strong evidence against H_0 and in favor of H_a .