

Statistics Review Problems Units 1 and 2

This is a selection of problems that have come from previous tests, as well as some selected review problems from the book.

1. A sample of 57 college students was asked to report the number of hours devoted to study during a typical week. The stemplot is shown below:

0	3
0	5555667777778999
1	0000000000112222333344444
1	55555668
2	0
2	56
3	00
3	5
4	0

- (a) Explain what the highlighted row represents.
- (b) Is this distribution skewed left, skewed right, or symmetric?
- (c) Would you expect the mean to be less than, equal to, or greater than the median in this distribution? Explain your answer. (You should NOT have to compute anything here.)
2. Students in an Intro Stats class were asked to describe their politics as “Liberal,” “Moderate,” or “Conservative.” The results are below.

		<i>Politics</i>			Total
		Liberal	Moderate	Conservative	
<i>Sex</i>	Female	35	36	6	77
	Male	50	44	21	115
	Total	85	80	27	192

- (a) What percent of the class is male?
- (b) What percent of all females in the class are “Liberals”?
- (c) Find the conditional distribution of *politics* for males.
- (d) Do *politics* and *sex* appear to be independent?

3. The Cornell Lab of Ornithology holds an annual Christmas Bird Count, in which birdwatchers at various locations around the country see how many different species of birds they can spot. Here are some of the counts reported from sites in Texas during the 1999 event.

228	178	186	162	206	166	163	183	181	206	177	175
167	162	160	160	157	156	153	153	152			

- (a) How many sites reported a number of birds within one standard deviation of the mean of the data?
- (b) By hand, draw a well-labeled histogram for these data.
- (c) In one or two sentences, write a brief description of this distribution. Be sure to discuss the overall shape as well as any unusual features.
4. A friend tells you about a recent study dealing with the number of years of teaching experience among current college professors. He remembers the mean but can't recall whether the standard deviation was 6 months, 6 years, or 16 years. Tell him which one it must have been, and why.
5. Human babies can be examined in the womb using ultrasound. Animal studies have suggested that ultrasound examinations can cause low birthweight. As ultrasound became more common, researchers observed an association in humans as well – babies who were exposed more often to ultrasound in the womb had lower birthweight on average than babies who were exposed less often.
- (a) What are the explanatory and response variables? Is this association positive or negative?
- (b) Suggest at least one lurking or confounding variable that could explain the association.
6. Hoping to learn what issues may resonate with voters in the coming election, the campaign director for a mayoral candidate randomly selects one block from each of the city's election districts. Staff members go there and interview all the residents they can find.
- (a) Identify the population.
- (b) Identify the sampling method used.
- (c) Identify any potential sources of bias you can detect.
- (d) The population parameter(s) of interest are not clear. List one reasonable population parameter of interest for this situation.

7. Shortly before a mayoral election, a market research firm took a poll to find out which candidate people were planning to vote for. The results are shown in the frequency table below:

<u>Candidate</u>	<u>Count</u>
Li Fong	2120
Bob Green	2329
Sue Moore	1042
José Alvarez	399

You wish to display the distribution of data with a chart. It should be easy to see from the display which candidate is in the lead, according to the poll. Which type of display would be more useful, a bar graph or a pie chart? Explain.

8. People who eat lots of fruits and vegetables have lower rates of colon cancer than those who eat little of these foods. Since fruits and vegetables are high in vitamins, researchers wondered whether vitamin pills would also help reduce the rates of colon cancer. The clinical trial studied this question with 864 people who were at risk of colon cancer. The subjects were divided into four groups: daily vitamin A, daily vitamins C and E, all three vitamins every day, and daily placebo. After four years, the researchers were surprised to find no significant difference in the occurrence of colon cancer among the groups.

(a) What are the factors and levels in this experiment?

(b) The study was double-blind. What does that mean?

9. The time to complete a standardized exam follows a Normal model with a mean of 75 minutes and a standard deviation of 10 minutes.

(a) What percentage of students will complete the exam in an hour (60 minutes) or less?

Draw a picture illustrating your answer.

(b) If I want 90% of the students to complete the test, how many minutes shall I allow them?

Draw a picture illustrating your answer.

10. Classified ads in the Ithaca Journal offered several used Toyota Corollas for sale. Listed below are the ages of some of the cars with their advertised prices.

Age of Car (in years)	1	2	4	5	6	6	9	9	11	13
Asking price	12995	10495	7990	6995	5990	4995	3200	3995	2995	1750

(a) Which is the explanatory variable and which is the response variable?

(b) Make a scatterplot of the data in your calculator. Describe the form, direction and strength of what you see. Are there any outliers? You do not have to draw the scatterplot for me.

(c) Write the equation for the regression line for predicting the asking price from the age of the car.

(d) What asking price would you predict for a Toyota Corolla that was 8 years old?

11. This is a standard deviation contest. You must choose four numbers from 0 to 10, inclusive, with repeats allowed.

(a) Choose four numbers that have the smallest possible standard deviation.

(b) Choose four numbers that have the largest possible standard deviation.

12. The table below shows amounts of fat and calories in six fast food hamburgers.

Fat (g)	31	34	35	39	39	43
Calories	580	590	570	640	680	660

(a) Assume a linear model is appropriate. How many calories would you expect in a fast food hamburger with 41 grams of fat?

(b) Is a linear model appropriate? Explain clearly in complete sentences with as much supporting evidence as possible.

13. The drug AZT was the first effective treatment for AIDS. An important medical experiment demonstrated that regular doses of AZT delay the onset of symptoms in people in whom HIV is present. The researchers who carried out this experiment wanted to know:

- Does taking either 500 mg of AZT or 1500 mg of AZT per day delay the development of AIDS?
- Is there any difference between the effects of these two doses?

The subjects were 1200 volunteers already infected with HIV but with no symptoms of AIDS when the study started.

Design an experiment that would have answered the questions raised in the above paragraph. Use diagrams to summarize the design. Be sure to address all the elements of experimental design. What are some confounding variables? How might you control for them?

14. According to a 1996 report on full-term pregnancies by the American Medical Association, women pregnant with twins who received intensive prenatal care had an infant mortality rate of 5.4 deaths per thousand live births, while women pregnant with twins who received only adequate care had an infant mortality rate of 3.9 deaths per thousand live births. Does this suggest that a woman pregnant with twins should be wary of seeking too much medical care?

15. To simulate the shoe sizes of a population of females, a student uses $\text{RandInt}(5,13)$ to generate a random number from 5 to 13. Ignoring the issue of half-sizes, identify the one or two main problems with this simulation.

16. One reason to invest abroad is that markets in different countries don't move in step. When American stocks go down, foreign stocks may go up. So an investor who holds both bears less risk. That's the theory.

A recent issue of *The Economist* states, "The correlation between changes in American and European share prices has risen from 0.4 in the mid-1990s to 0.8 in 2000." Explain (briefly) to an investor who knows no statistics why this fact reduces the protection provided by buying European stocks.

17. A city police officer, using radar, checked the speed (in mph) of twelve cars as they were traveling down a city street. The table shows some of the statistics.

Suppose one of the cars in the sample is found to be 53 mph. Can the police officer treat this data value as an outlier? Justify your answer.

statistic	value
mean	30.0
median	29.0
1 st Quartile	23.0
3 rd Quartile	34.5
standard deviation	9.6

18. Black men's heights have a mean of (approximately) 5'10" with a standard deviation of (approximately) 3 inches. White women's heights have a mean of (approximately) 164 cm with a standard deviation of approximately 6 cm. Shaquille O'Neal (a black man) is 7'1" tall; Lauren Jackson (a white woman) is about 196 cm tall. Whose height is more extreme, relative to the appropriate group?

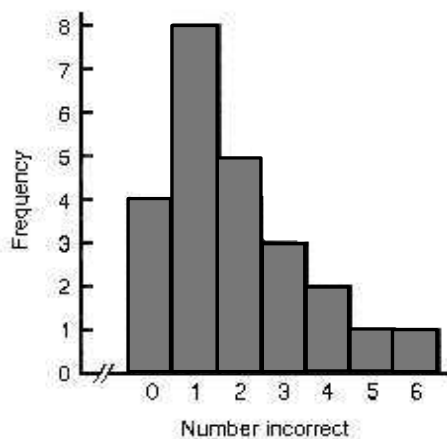
19. A chemistry instructor appropriately created a linear regression equation to predict students' final exam scores from their midterm exam scores. Each exam was scored out of 100 points. The regression equation was $Final = 1.1(Midterm) - 6$. Address parts (a) and (b). Sentence answers not required.

- (a) The mean grade on the midterm exam was 76.4 points. What was the mean grade on the final exam?
- (b) The standard deviation of the final exam scores was 12 points and the standard deviation of the midterm exam scores was 8 points. Find the correlation between the scores of the two exams.

20. Last spring, 1737 qualified applicants applied for admittance into the University of Washington. 926 of these qualified applicants were accepted, 307 were waitlisted, and 504 were turned away for lack of space. 481 of those who were accepted were women, 172 of those who were waitlisted were women, and 276 of those who were turned away were women. Determine the proportion of each of the following (do not convert your fractions to decimals and you do not have to reduce your fractions).

- (a) The proportion of qualified applicants that were accepted.
- (b) The proportion of those accepted applicants that were men.
- (c) The proportion of women applicants that were turned away for lack of space.

21. A frequency histogram for the number of questions answered incorrectly on an eight-question science quiz by each of 24 randomly selected fifth-grade students is given to the right. Use the histogram to address parts (a) – (c).



(a) Determine the mean and standard deviation of this sample's data. Report each of these statistics with the appropriate number of digits and include the appropriate units of measurement.

(b) Determine the actual percentage of the data that is within one standard deviation of the mean? Be sure you clearly communicate how you obtained your answer.

(c) How many standard deviations above the mean is the fifth-grade student who answered 5 of the 8 questions on the quiz incorrectly? Be sure it is clear how you obtained your answer.

22. The lifespan of *BrightLite* light bulbs is Normally distributed with a mean of 54 months and a standard deviation of 6 months.

(a) What percentage of *BriteLite* bulbs have a lifespan that's between 36 months and 50 months? Please include a labeled sketch of this Normal model; your graphic should illustrate what you are calculating.

(b) What percentage of *BriteLite* bulbs last longer than 70 months? No sketch required.

(c) Determine the IQR of *BriteLite* bulbs lifespans? No sketch required.

23. The salaries (in thousands of dollars) of 20 employees of a company for the year 2009 are given below.

28	31	34	35	37	41	42	43	44	49
49	51	52	52	60	61	67	72	75	77

(a) Make a stemplot (i.e. stem-and-leaf display).

(b) Determine the five-number summary.

(c) How many standard deviations above the mean is the \$61,000 salary? Justify your answer.

(d) Is the \$28,000 salary more than one and half IQRs below Q1? Justify your answer.

24. Read each scenario and identify the sampling method being described (simple random sample, convenience sampling, stratified random sampling, systematic sampling, cluster sampling). Explain.
- (a) To determine monthly rental prices of apartments in the Shoreline area, all of the apartments were classified according to the number of bedrooms (1, 2, 3, or 4 bedrooms). Then, ten apartments were chosen from each of these four classifications.
 - (b) Suppose a doctor wants to collect data on a sample of human beings' body temperatures. Since he suspects that the age and gender of the person affects their body temperature, he randomly selects thirty young males, thirty middle-aged males, thirty old males, thirty young females, thirty middle aged females, and thirty old females.
 - (c) A researcher interested in the number of cars Washington state households have registered with the state randomly selects fifty of the over 1400 zip codes in Washington state and records the number of cars registered at every address in those zip codes.
 - (d) The manager of the local Safeway store is interested in the customer wait-times of the most recently-hired cashier. He decides to begin recording the wait-time of her 76th customer and then record the wait-time of every 10th customer thereafter, continuing this process over his employee's entire shift.
25. A company held a blood pressure screening clinic for its employees. The results are summarized in the table below by age group and blood pressure level. Assume all employees participated.

		AGE		
		Under 30	30 – 49	Over 50
BLOOD PRESSURE	Low	27	37	31
	Normal	48	91	93
	High	23	51	73

- (a) What percent of employees are under age 30 and have high blood pressure?
 - (b) What percent of employees with high blood pressure are under age 30?
26. Low interest rates for home mortgages generally encourage the construction of new houses. Data from a Midwestern city over a period of 18 months shows a clear relationship between I , the prevailing mortgage interest rate (as a percent), and N , the number of new housing starts per month. A scatterplot of the data shows that a linear model is appropriate. The equation of the least-squares regression line is $N = 670.89 - 32.35I$, with $R^2 = 0.64$.
- (a) What is the correlation coefficient between interest rate and monthly new housing starts?
 - (b) What does the model predict the number of new housing starts in a month will be when the interest rate is 8.6% (i.e. when $I = 8.6$)?
 - (c) What change in the number of new housing starts would you expect from a 1% hike in the interest rate?

27. Criminal laboratories must sometimes estimate a person's height from the length of the thigh bone. A random sample of eight living adult males is taken; the length of the thigh bone and the man's height are recorded. Address parts (a) – (c).

Thigh bone length (inches), x	Height (inches), y
17.5	70
20	80
21	78
19	73
15.5	63
18.5	71
16	64
22	73

- (a) After verifying that linear regression is appropriate, determine a linear regression equation to predict a man's height from the length of his thigh bone. Report values to the nearest tenth (i.e. with one digit after the decimal point).
- (b) Richard, who is 6 feet tall, has a thigh bone that is 18 inches. How big is his residual?
- (c) Interpret the meaning of the slope in this context. Your sentence answer must contain the slope's value (rounded to the nearest tenth) and its units of measurement.

Review problems from the textbook:

pp. 155 – 163 Exercises 1, 3, 5, 8, 17, 23, 25, 27, 32, 33, 38

pp. 278 – 286 Exercises 1, 2, 6, 8, 9, 10, 12, 16, 17, 27, 29, 30, 36, 37, 41, 42

pp. 357 – 361 Exercises 1, 3, 5, 7, 11, 22, 25, 26, 29, 35, 39, 41, 43, 44