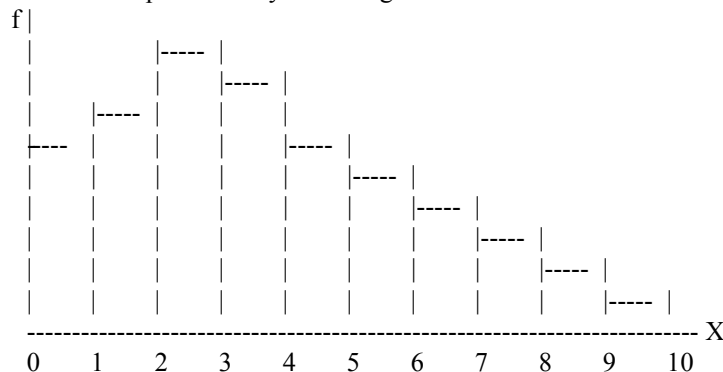


Material covered: Chapters 1-6 in text.
Webassign homework assignments 1 - 4.

Important Note: Most of the questions on this sample exam are in a multiple choice format, but some questions are not. The questions on the exam will be multiple choice; you will use a scantron sheet to indicate your answers.

- The heights of American men aged 18 to 24 can be described by a Normal model with mean 68 inches and standard deviation 2.5 inches. Half of all young men are shorter than
a) 65.5 inches b) 68 inches c) 70.5 inches d) can't tell because the median height is not given
- Use the information in Problem 1 and the **68-95-99.7 rule** to determine the percentage of young men that are taller than 6' 1".
- The grade point averages (GPA) of 7 randomly chosen students in a statistics class are
3.14 2.37 2.94 3.60 1.70 4.00 1.85
The mean GPA for these students is
a) 2.8 b) 2.94 c) 3.6 d) none of the above
- Refer to the information given in the previous problem. If $\sum (y - \bar{y})^2 = 4.51$, what is the standard deviation?
- A standardized test designed to measure math anxiety has a mean of 100 and a standard deviation of 10 in the population of first year college students. Which of the following observations would you suspect is an outlier?
a) 150 b) 100 c) 90 d) 125 e) none of the above
- Which of the following best describes a risk in using the range to measure spread of a data set? Choose one of the following:
a) The range is not in the same unit of measurement as the observations themselves.
b) The largest or smallest observation may be a mistake or an outlier.
c) The range is complicated to compute, which may result in an error.
d) The range might be negative so it doesn't make sense to use the range as a measure of spread.

- The distribution represented by the histogram below is:



Choose one of the following:

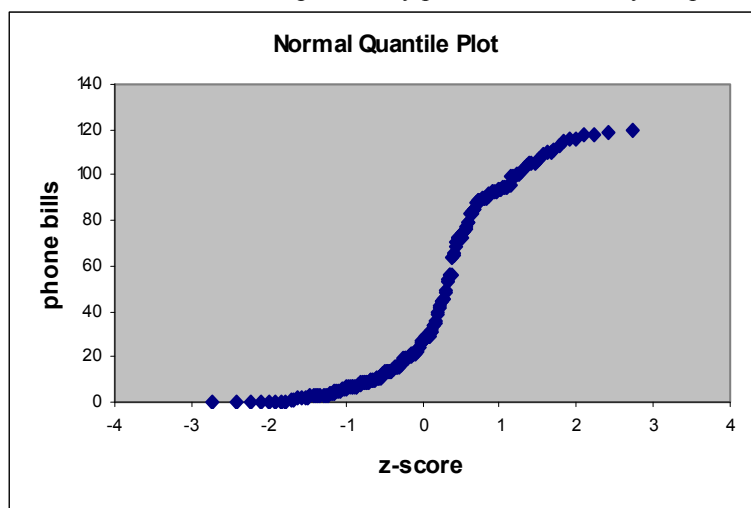
- skewed to the right.
- skewed to the left.
- symmetric.
- normal.

8. Twenty-seven applicants interested in working for the Food Stamp program took an examination designed to measure their aptitude for social work. The following test scores were obtained:
79, 93, 84, 86, 77, 63, 46, 97, 87, 88, 87, 92, 68, 72, 86, 98, 81, 70, 66, 98, 59, 76, 68, 91, 94, 85, 88.
- Find Q_1 .
 - Construct a boxplot for these observations. Do you observe any outliers?
9. A manufacturer of television sets has found that for the sets he produces, the lengths of time until the first repair can be described by a normal distribution with a mean of 4.5 years and a standard deviation of 1.5 years. If the manufacturer sets the warrantee so that only 10.2% of the 1st repairs are covered by the warrantee, how long should the warrantee last?
10. Suppose the amount of tar in cigarettes is described by a normal model with a mean of 3.5 mg and a standard deviation of 0.5 mg.
- What proportion of cigarettes have a tar content that exceeds 4.25 mg?
 - In order to advertise as a low tar brand, a manufacturer must prove that their tar content is below the 25th percentile of the tar content distribution. Find the 25th percentile of the distribution of tar amounts.
11. Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from "Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes" (Shanthy A. Bowman, Journal of the American Dietetic Association, 102(9), pp. 1234-1239):

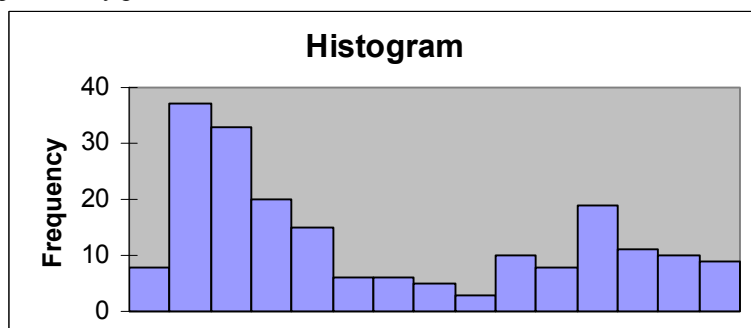
		Nationwide Food Survey Years			
		1987-1988	1989-1991	1994-1996	Total
Drinks Fluid Milk	Yes	354	502	366	1222
	No	226	335	366	927
	Total	580	837	732	2149

- Find the following:
 - What percent of the young girls reported that they drink milk?
 - What percent of the young girls were in the 1989-1991 survey?
 - What percent of the young girls who reported that they drink milk were in the 1989-1991 survey?
 - What percent of the young girls in 1989-1991 reported that they drink milk?
 - What is the marginal distribution of milk consumption?
12. It's the last inning of an important baseball game. The home team is losing by a run, the bases are loaded and the manager needs a pinch hitter. Two batters are available to pinch hit. Here are their statistics:
- | Player | Overall | vs Left-handed pitching | vs Right-handed pitching |
|--------|------------|-------------------------|--------------------------|
| A | 33 for 103 | 28 for 81 | 5 for 22 |
| B | 45 for 151 | 12 for 32 | 33 for 119 |
- Based on their overall batting averages and their batting averages against right-handed and left-handed pitchers, who would you select as the pinch hitter? What is this phenomenon called?
13. The mean SAT verbal score of next year's freshmen entering the local university is 600. It is also known that 69.5% of these freshmen have scores that are less than 625. If the scores can be described by a normal model, what is the standard deviation of the scores?
14. Two students are enrolled in an introductory statistics course at the University of Florida. The first student is in a morning section and the second student is in an afternoon section. If the student in the morning section takes a midterm and earns a score of 76, while the student in the afternoon section takes a midterm with a score of 72, which student has performed better compared to the rest of the students in his respective class? Assume that the test scores can be described by a normal model. For the morning class, the class mean was 64 with a standard deviation of 8. For the afternoon class, the class mean was 60 with a standard deviation of 7.5.

15. Suppose that a Normal model describes the acidity (pH) of rainwater, and the water tested after last week's storm had a z-score of 1.8. This means that the acidity of that rain
- had a pH of 1.8
 - varied with a standard deviation of 1.8
 - had a pH 1.8 higher than the average rainfall
 - had a pH 1.8 times that of average rainwater
 - had a pH 1.8 standard deviations higher than that of average rainwater
16. The highway gas mileage x , measured in miles per gallon (mpg), of 26 models of midsize cars, have the following summary statistics: $\bar{x} = 26.54$ mpg, median = 26 mpg, $s = 3.04$ mpg, $IQR = 3$ mpg. If you convert gas mileage x from miles per gallon to x_{new} which is measured in miles per liter, what are the new values of the summary statistics? (3.785 liters = 1 gallon).
17. Shown below is the normal probability plot for 200 monthly telephone bills.



Shown below is a histogram. Is this a histogram of the same data that was used to construct the normal probability plot?



18. A local plumber makes house calls. She charges \$30 to come out to the house and \$40 per hour for her services. For example, a 4-hour service call costs $\$30 + \$40(4) = \$190$.
- The table shows summary statistics for the past month. Fill in the table to find out the cost of the service calls.

Statistic	Hours of Service Call	Cost of Service Call
Mean	4.5	
Median	3.5	
Stan Dev	1.2	
IQR	2.0	
Minimum	0.5	

- b. This past month, the time the plumber spent on a particular service call had a z-score of -1.50 . What is the z-score for the cost of the service call?
19. In 2000 the Department of Education published the Digest for Education Statistics, a collection of information about education in the United States. They reported the average amount (dollars per student) spent by public schools in each state and Washington, D.C. during the school year 1997-8. The data was recorded according to whether the state lies east or west of the Mississippi River. A back-to-back stem and leaf display of the data is shown below. 6|7 denotes \$6,700.
- a. Which states, Eastern or Western, tend to spend more?
- b. Western states median = ? Eastern states Q_1 = ?

Dollars Spent per Student		
Western States		Eastern States
2	4	
99	4	5
33210	5	12
9776	5	6699
444200	6	114
7755	6	7788
3	7	2
	7	567789
	8	2
	8	6
0	9	22
	9	9
	10	2
	10	

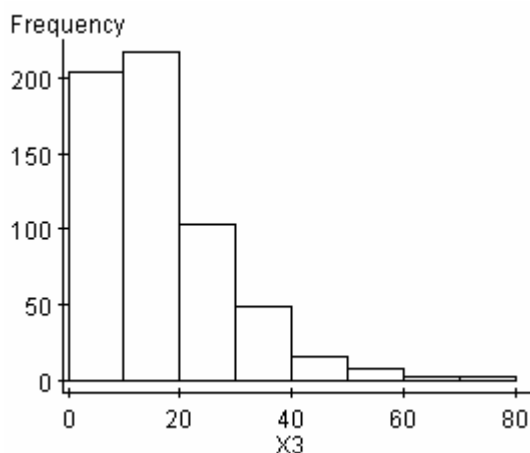
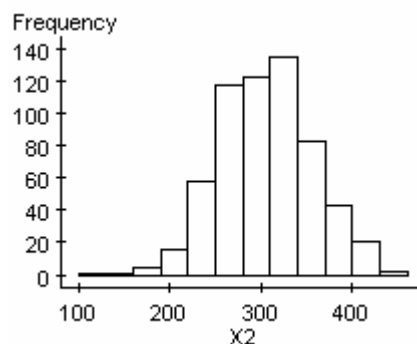
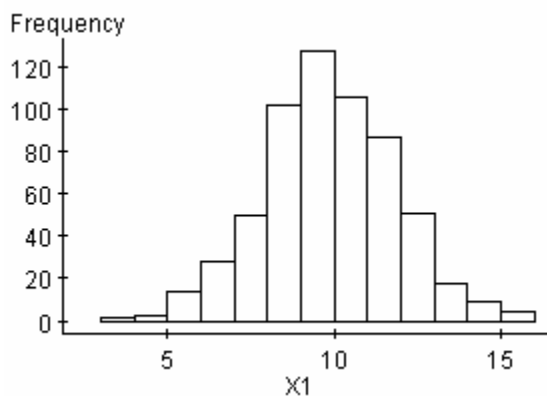
20. The following 2-way table shows students by major and home state for a small private school in the northeast U.S.

Major Program of Study

Home State	Biology	Accounting	History	Education
PA	80	65	55	100
NJ	50	40	65	95
NY	75	50	45	80
MD	65	55	40	40

- a. Find the marginal distribution for home state.
- b. Find the conditional distribution (in percentages) of major distribution for the home state of NJ.
- c. Find the conditional distribution (in percentages) of home state distribution for the biology major.

Consider the following histograms of variables labeled X1, X2 and X3:

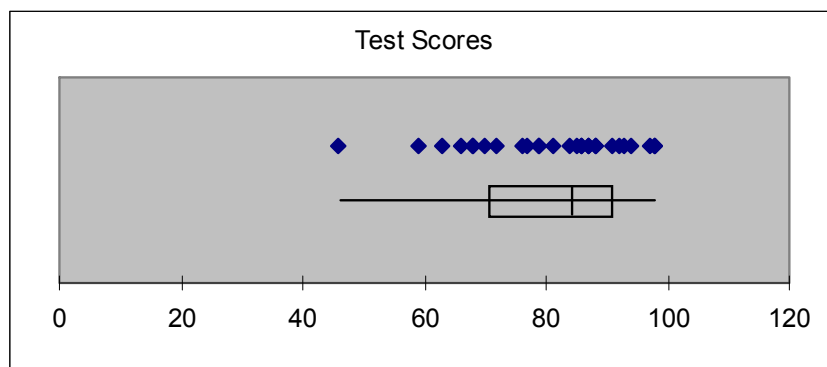


21. The median for variable X2 would be around
a. 10 b. 305 c. 250 d. impossible to tell
22. The third quartile for variable X1 would be around
a. 12 b. 8 c. 5 d. 15
23. The distribution in which the mean and median are most different would be
a. X1 b. X2 c. X3 d. It is impossible to tell.
24. The standard deviation for variable X1 would be
a. About the same as the standard deviation for variable X2.
b. Smaller than the standard deviation for variable X2.
c. Larger than the standard deviation for variable X2.
d. It is impossible to tell.
25. The histograms above are the results of questions asked of a group of undergraduate students. Match the histogram (X1, X2, or X3) above to the appropriate question below.
a. How many hours did you work at a job last week? _____
b. What is your shoe size? _____
c. How much did you spend on textbooks (in dollars) this semester? _____

SAMPLE EXAM 1 ANSWERS

1. b 2. 2.5% 3. a 4. .87 5. a
6. The largest or smallest observation may be a mistake or an outlier.
7. Skewed to the right.

8. a. The first step is to order the data. Then compute the overall median. Since there are 27 observations, the median is the observation in position 14: median = 85. Compute Q_1 : we want the median of the lower half. Since we have an odd number of observations (27), **include the overall median in both halves of the data**. There are 14 observations in the lower half, including the overall median. The median of these lower 14 observations is the mean of the 2 middle observations in positions 7 and 8, so $Q_1 = \frac{70+72}{2} = 71$.
- b. Note that Q_3 is the median of the 14 observations in the upper half, including the overall median. So Q_3 is the mean of the 2 middle observations in positions 20 and 21: $Q_3 = \frac{88+91}{2} = 89.5$.
 $IQR = Q_3 - Q_1 = 89.5 - 71 = 18.5$; $1.5 * IQR = 27.75$.
Boundaries for outliers: $Q_1 - 1.5 * IQR = 71 - 27.75 = 43.25$; $Q_3 + 1.5 * IQR = 89.5 + 27.75 = 117.25$.
 Since the smallest observation is 46 and the largest observation is 98, there are no outliers. See boxplot below (the diamonds above the box represent the individual data values).



9. $z = -1.27$; $x = 2.595$ years. 10. a. .0668 b. $z = -0.675$; $x = 3.16$. 11. a1. 56.9% a2. 38.9% a3. 41.1% a4. 60% b. Yes: 56.9%; No: 43.1%. 12. Player A overall batting avg. = .320; Player B overall batting avg. = .298. **Choose player A.** Player A vs right-handed pitchers = .227, Player B vs right-handed pitchers = .277; Player A vs left-handed pitchers = .346; Player B vs left-handed pitchers = .375. Player B has the higher batting average against both right-handed and left-handed pitchers; **choose Player B.** Simpson's paradox. 13. $0.51 = (625 - 600)/\sigma \Rightarrow \sigma = 49.02$ 14. $z_1 = (76 - 64)/8 = 1.5$; $z_2 = (72 - 60)/7.5 = 1.6$. The student in the afternoon section performed better. 15. e. 16. $\bar{x}_{new} = 7.01$ miles per liter; median_{new} = 6.87 miles per liter; $s_{new} = .803$ miles per liter; $IQR_{new} = .793$ miles per liter. 17. Yes. Note upward curvature on left portion of the normal probability plot (bills are not less than \$0); plot very "steep" in central portion, which means there are not many observations in middle of data; downward curvature in right portion of plot indicates that the data has a shorter right tail than a normal distribution.

18. a.

Statistic	Hours of Service Call	Cost of Service Call
Mean	4.5	\$210
Median	3.5	\$170
Stan Dev	1.2	\$48
IQR	2.0	\$80
Minimum	0.5	\$50

b. - 1.50

19. a. Eastern b. West median = \$5,950, East $Q_1 = \$6,000$

Home State	Number of Students (%)
PA	300 (300/1000 = 30%)
NJ	250 (250/1000 = 25%)
NY	250 (250/1000 = 25%)
MD	200 (200/1000 = 20%)

b. <u>Major</u>	<u>Conditional for NJ</u>
Bio	20%
Acct.	16%
Hist	26%
Educ	38%
c. <u>Home State</u>	<u>Conditional for Biology</u>
PA	29.6%
NJ	18.5%
NY	27.8%
MD	24.1%

21. b 22. a 23. c 24. b 25. a. X3 b. X1 c. X2