Topics covered on Exam 1: Chapters 1-9, 14, 15 in text.
This material is covered in webassign homework assignments 1 through 5 and worksheets 1-16.
Exam information: materials allowed: calculator (no laptops), one 3 x 5 notecard (2 sided) with notes, definitions, formulas, etc. Normal table will be provided with the exam.

WARNING! The problems below may not cover all topics for which you are responsible on exam 1.

Answers are at the end of the document.

1. The heights of American men aged 18 to 24 are approximately symmetrically mound shaped with mean 68 inches and standard deviation 2.5 inches. Half of all young men are shorter than what height?

2. Use the information in Problem 1. Use the 68-95-99.7 rule to determine the percentage of young men that are taller than 6' 1" ?

3. The grade point averages (GPA) \(x_1, x_2, \ldots, x_7\) of 7 randomly chosen students in a statistics class results in the following value for the sum of the squared differences between each data value \(x_i\) and their overall mean \(\bar{x}\): \[
\sum_{i=1}^{7} (x_i - \bar{x})^2 = 4.51.
\] What is the standard deviation \(s\)?

4. A survey conducted in a college stats class asked students about the number of credit hours they were taking that semester. The number of credit hours for a random sample of 16 students is:

\[
\begin{array}{cccccccc}
10 & 10 & 12 & 14 & 15 & 15 & 15 & 15 \\
17 & 17 & 19 & 20 & 20 & 20 & 20 & 22 \\
\end{array}
\]

The mean and standard deviation are 16.3 and 3.7, respectively.

a. find the median and IQR for the number of credit hours.

b. is it more appropriate to use the mean and standard deviation or the median and IQR to summarize these data?

5. 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

6. 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):

<table>
<thead>
<tr>
<th>Drinks Fluid Milk</th>
<th>Nationwide Food Survey Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>354</td>
</tr>
<tr>
<td>No</td>
<td>226</td>
</tr>
<tr>
<td>Total</td>
<td>580</td>
</tr>
</tbody>
</table>

a. Find the following:

1. What percent of the young girls reported that they drink milk?

2. What percent of the young girls were in the 2002-2003 survey?

3. What percent of the young girls who reported that they drink milk were in the 2002-2003 survey?

4. What percent of the young girls in 2002-2003 reported that they drink milk?

b. What is the marginal distribution of milk consumption?
7. The boxplots show the age of people involved in traffic-related accidents according to their role in the accident.

![Boxplots showing age distribution by role](image)

a. Which role involved the youngest person, and what is the age? ____________, ____________

b. Which role had the lowest median, and what is the age? ____________, ____________

c. Which role had the smallest range of ages, and what is it? ____________, ____________

d. Which role had the largest IQR of ages, and what is it? ____________, ____________

e. Which role generally had the oldest people?

8. A manufacturer of television sets has found that for the sets he produces, the lengths of time until the first repair can be described using a normal model 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?

9. Suppose the amount of tar in cigarettes can be described using a normal model with a mean of 3.5 mg and a standard deviation of 0.5 mg.

   a. What proportion of cigarettes have a tar content that exceeds 4.25 mg?
   
   b. 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.

10. 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 using a normal model, what is the standard deviation of the scores?

11. 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 follow 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.
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:

<table>
<thead>
<tr>
<th>Player</th>
<th>Overall</th>
<th>vs Left-handed pitching</th>
<th>vs Right-handed pitching</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>33 for 103</td>
<td>28 for 81</td>
<td>5 for 22</td>
</tr>
<tr>
<td>B</td>
<td>45 for 151</td>
<td>12 for 32</td>
<td>33 for 119</td>
</tr>
</tbody>
</table>

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 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).

14. Shown below is the normal probability plot for 200 monthly telephone bills.

![Normal Quantile Plot](image1)

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

![Histogram](image2)

15. 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.

a. The table shows summary statistics for the length of service calls for the past month. Fill in the table to find out the cost of the service calls.
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?

16. In 2010 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 2007-2008. 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 Qs = ?

Dollars Spent per Student

<table>
<thead>
<tr>
<th>Western States</th>
<th>Eastern States</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>99</td>
</tr>
<tr>
<td>99</td>
<td>4</td>
</tr>
<tr>
<td>33210</td>
<td>5</td>
</tr>
<tr>
<td>9776</td>
<td>2</td>
</tr>
<tr>
<td>444200</td>
<td>6</td>
</tr>
<tr>
<td>7755</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>567789</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

17. There are 25 pies entered in the pie-baking contest at the state fair. How many ways can the judges select 3 pies to finish 1st, 2nd, and 3rd?

18. Four radar systems are arranged so that they work independently of each other. Each system has a 0.9 chance of detecting an approaching airborne object. Find the probability that at least one radar system will fail to detect an approaching object.
   a. (.9)^4  b. (.1)^4  c. 1−(.9)^4  d. 1−(.1)^4  e. .9 + .9 − (.9)(.9) = .99

19. A copy machine dealer has data on the number x of copy machines at each of 89 customer locations and the number y of service calls in a month at each location. Summary calculations give \( \bar{x} = 8.4, \ s_x = 2.1, \ \bar{y} = 14.2, \ s_y = 3.8, \) and \( r = .86. \) What is the slope of the least squares regression line of number of service calls on number of copiers?

20. In the setting of the previous problem, about what percent of the variation in number of service calls is explained by the linear relation between number of service calls and number of machines?
21. Assume that 70% of teenagers who go to take the written drivers license test have studied for the test. Of those who study for the test, 95% pass; of those who do not study for the test, 60% pass. What is the probability that a teenager who passes the written drivers license test did not study for the test?

22. Each of the following statements contains a blunder. In each case explain what is wrong.
   a. “There is a high correlation between the sex of American workers and their income.”
   b. “We found a high correlation \((r = 1.09)\) between students’ ratings of faculty teaching and ratings made by other faculty members.”
   c. “The correlation between planting rate and yield of corn was found to be \(r = .23\) bushel.”

23. A study of 1,000 families gave the following results:
   \(\bar{x} = 68\) inches; \(\bar{y} = 63\) inches;
   \(\frac{s_y}{s_x} = .925\) wife inches per husband inch; \(r = .25\).
   A. Estimate the height of a wife when her husband is 72 inches tall.
      a. 63 inches  b. 72 inches  c. 64 inches  d. none of these  e. need more information
   B. A sociologist wants to reverse the roles of the variables and use the height of the wife to predict the height of the husband. So wife height is now the \(x\)-variable and husband height is the \(y\)-variable. What is the slope of the new least squares line?

The information below is needed for questions 24 and 25.

In finance, million dollar investments are made with the assistance of the Capital Asset Pricing Model (CAPM). The CAPM uses a least squares line to predict the annual rate of return \((\bar{y})\) of a stock based on the rate of return for the overall stock market \((\bar{x})\). The slope of the line is used to evaluate the risk of investing in the stock:
   - Slope = 1: Average risk (neutral stock)
   - Slope > 1: High risk (aggressive stock)
   - Slope < 1: Low risk (conservative stock)

The data in the accompanying table are the annual rates of return for Disney stock \((\bar{y})\) and the rate of return for the overall stock market \((\bar{x})\) for an 8 year period (where rates of return are measured as a percent). The least squares line for the CAPM model is shown below the table

<table>
<thead>
<tr>
<th>Year</th>
<th>83</th>
<th>84</th>
<th>85</th>
<th>86</th>
<th>87</th>
<th>88</th>
<th>89</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disney rate of return (\bar{y})</td>
<td>2</td>
<td>-10</td>
<td>18</td>
<td>9</td>
<td>12</td>
<td>-1</td>
<td>-12</td>
<td>2</td>
</tr>
<tr>
<td>Overall market rate of return (\bar{x})</td>
<td>1</td>
<td>-5</td>
<td>12</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>-6</td>
<td>2</td>
</tr>
</tbody>
</table>

Results: \(\hat{y} = -.94 + 1.72x\), sum of squares of residuals = 7.996

24. Give an interpretation of the slope of the least squares line:
   a. For a year with an overall market rate of return of 1%, we estimate Disney's rate of return to be 1.72%
   b. For every 1.72% increase in overall market rate of return, we estimate Disney's annual rate of return to increase by 1%.
   c. For every 1% increase in overall market rate of return, we estimate Disney's annual rate of return to increase by 1.72%.
   d. For a year with a market rate of return of 0%, we estimate Disney's rate of return to be -.94%

25. Which of the following is the best interpretation of the sum of the squares of the residuals?
   a. No other line will produce a sum of squares of the residuals greater than 7.996.
   b. The least squares line obtained from these data should yield predictions of Disney's rate of return that are accurate to within \(\pm 2\sqrt{7.996}\).
   c. The small value of 7.996 for the sum of squares of residuals indicates that the straight-line CAPM model is not useful for predicting Disney's rate of return \((\bar{y})\).
   d. No other straight line fit to these data will produce a sum of squares of residuals smaller than 7.996.
Questions 26 and 27 refer to the following:

Data were collected to find the relationship between the labor \( (y \text{ in hours}) \) required to produce lots of custom wood products and the size \( x \) of the lot. The following least squares regression equation was calculated from the data:

\[ \hat{y} = 13.7 + 1.7x. \]

26. What is the predicted hours of labor for a lot size of 55?

27. One of the original data points is \((20, 50.3)\). What is the residual when the lot size is 20?

28. Shown below is a scatterplot with the corresponding least squares line.

![Scatterplot](image)

Choose the residual plot that corresponds to this scatterplot and least squares line.

- a. I
- b. II
- c. III
- d. IV
- e. none

29. Consider the following scatterplot showing miles per gallon (MPG) plotted against the weight of the car (in tons):

![Scatterplot](image)
If the proportion of the variation in MPG explained by differences in car weights is 0.81, what is the correlation between weight and MPG?

30. A federal agency is trying to decide which of two waste dump projects to investigate. An administrator estimates that the probability of federal law violations in the first project is 0.3. She also estimates that the probability of violations in the second project is 0.25. In addition, she believes the occurrence of violations in these two projects are mutually exclusive. The probability of federal law violations in the first project or in the second project or both is

a. .075   b. .05   c. .55   d. none of these   e. can't tell from information given

31. How many four-digit serial numbers can be formed if no digit is to be repeated within any number? (The first digit may be a zero).

a. (10)   b. 10!   c. 10P4   d. 5040   e. 10C4

32. These questions are based on the Peanuts cartoon shown below.

I. In how many ways can seven books be arranged on a bookshelf? (The order in which books are arranged matters).

a. 7C7   b. 7!   c. 7P7   d. 77   e. 77

II. If you have only 3 books to put on a bookshelf, in how many ways can three books be arranged?

a. 3   b. 9   c. 6   d. 27   e. 33

III. Circle your answer to the question posed in the third frame of the cartoon.

a. 5040   b. 4! × 3!   c. 7(3!)   d. 720   e. 7(7!)

PRACTICE PROBLEM ANSWERS

1. 68 inches, the mean; 2. 2.5% 3. .87 4. a. median = 16 credit hours, IQR = Q3 – Q1 = 20 – 14.5 = 5.5 credit hours; b. the histogram is shown below; since the data are not symmetric and mound-shaped, it is more appropriate to use the median and IQR to summarize the data
5. a. 150
6. a1. 56.9% a2. 38.9% a3. 41.1% a4. 60% b. Yes: 56.9%; No: 43.1%
7. a. passenger, less than 1 year; b. passenger, 21 years; c. cyclist, 40 years; d. pedestrian, 44 years; e. Pedestrian; while the oldest person involved in an accident is not a pedestrian, the median for pedestrians is almost 45 years, while the median age in the other groups is between 22 and 35 years old. The oldest 50% of the pedestrian group, from 45 to 87, is generally older than the youngest 75% of two groups - cyclist and passenger, and only the driver group has any of its middle 50% as old. The driver and passenger groups have a few people older than the pedestrian group.
8. z = -1.27; x = 2.595 years. a. .0668 b. z = -0.675; Q = 3.16. 0.51 = (625-600)/49.02. z = (76-55)/8=1.5; z = (72-60)/7.5=1.6. The student in the afternoon section performed better.
9. 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.
10. Yes. Note upward curvature (or “leveling off”) and high density of points in left portion of probability plot (no left tail since bills are not less than $0); plot very “steep” in central portion and somewhat sparse, which means there are not many observations in middle of data (no high point in middle of histogram); downward curvature in right portion (to the right of z = 1) of plot indicates the higher phone bills but no points that you would consider to be outliers.
11. a.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Hours of Service Call</th>
<th>Cost of Service Call</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.5</td>
<td>$210</td>
</tr>
<tr>
<td>Median</td>
<td>3.5</td>
<td>$170</td>
</tr>
<tr>
<td>Stan Dev</td>
<td>1.2</td>
<td>$48</td>
</tr>
<tr>
<td>IQR</td>
<td>2.0</td>
<td>$80</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.5</td>
<td>$50</td>
</tr>
</tbody>
</table>

b. = 1.50
12. a. Eastern b. West has 24 data values; West median = mean of data values in ordered positions 12 and 13: West median = ($5,900 + $6,000) / 2 = $5,950. East has 27 data values; East median is data value in ordered position 14, so East median is $6,800 (last 8 in stem row 6). East Q1 is the median of the lower half; since there are an odd number of East data values, the overall East median (last 8 in stem row 6) is included in both halves when determining the quartiles. East Q1 is the median of the smallest 14 data values, so East Q1 = the mean of the numbers in ordered positions 7 and 8 = ($5,900 + $6,100) / 2 = $6,000. 17. 25%Pb = 13, 800 (order counts!). 18. The probability that no radar system fails to detect an airborne object (i.e. all four radar systems work) is (.9)^4, therefore the probability that at least one fails is 1 - (.9)^4. 19. b1 = r(4 / 3) = .86(3^1 / 21) = 1.56. 20. r^2 = (.86)^2 = .74
21.
Given that a student passed, the probability the student did not study is

\[
P(\text{pass}) = 0.665 + 0.180 = 0.845.\]

Given that a student passed, the probability the student did not study is

\[
\frac{0.180}{0.665 + 0.180} = 0.180 = 0.213.
\]

22. a. The correlation we are studying measures the linear relationship between 2 quantitative variables; sex is a categorical variable.

b. \(-1 \leq r \leq 1\) is violated.

c. \(r\) has no units.

23A  
23B  
23C  
23D  
23E  
23F  
23G  
23H  
23I  
23J  
23K  
23L  
23M  
23N  
23O  
23P  
23Q  
23R  
23S  
23T  
23U  
23V  
23W  
23X  
23Y  
23Z  
24. c. 25. d. 26. 7.2 27. observed \(y\) – predicted \(y\) = 2.6 hours. 28. b

29. \(r^2 = 0.81; r = -\sqrt{0.81} = -0.9\) (note that the correlation \(r\) is negative since the correlation and the slope \(b_1\) have the same sign due to the formula \(b_1 = r \frac{s_y}{s_x}\); the slope of the least squares line for this scatterplot is negative).

30. c 31. c. and d. are both correct since 10(9)(8)(7) = 5040 = 10! \(P_4\).

32. I. b. 7! = 5040  II. c. 3! = 6

III. d. View the problem as filling 5 slots: one slot for each of the four science book and one slot for the group of 3 math books. There are 5! ways to fill these 5 slots since order makes a difference; in addition, the 3 math books can be arranged in 3! ways. So there are 3! \(\times 5! = 720\) ways.