STAT 225 – Summer 2009
Exam 1 Solutions

Your Name: ___________________________________________
Your Instructor: ________________________________________

Your class time (circle one):

8:40     9:50     11:00    1:00

Note:

- Show your work on all questions. Unsupported work will not receive full credit.
- All answers should be in decimal form and should be exact, or to at least taken out to two decimal places.
- You are responsible for upholding the Honor Code of Purdue University. This includes protecting your work from other students.
- You are allowed the following aids: a one-page 8 ½” x 11” handwritten (in your handwriting) cheat sheet, a scientific calculator, and pencils.
- Instructors will not interpret questions for you. If you do have questions, wait until you have looked over the whole exam so that you can ask all of your questions at one time.
- You must show your student ID (upon request) and turn in your cheat sheet when you turn in your exam to your instructor.
- Turn off your cell phone before the exam begins!

<table>
<thead>
<tr>
<th>Question</th>
<th>Points Possible</th>
<th>Points Received</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
1. Choose the phrase from the following list that most accurately describes the sentences below. Each response should be used no more than once. (2 points each)

(a) Qualitative (b) Quantitative (c) With Replacement (d) Without Replacement
(e) Intersection (f) Union (g) Population (h) Sample
(i) Elements (j) Event (k) Variables (l) Sample Space

a. In a questionnaire, respondents are asked to mark their marital status. Marital status is an example of _____a______ data.

b. A sample of 5000 envelope lengths was taken. We want to determine whether the mean length is larger than the standard length. The length of an envelope is an example of _____b_____ data.

c. If A and B are disjoint (mutually exclusive), then their _____e_____ is empty.

d. Draw a card at random from a standard deck of cards. The collection of all possible outcomes is referred to as the _____l______.

e. Draw a card at random from a standard deck of cards. B = \{K, Q, J\}, is a subset of all the possible outcomes and is referred to as a(n) _____j______.

f. A player randomly draws 13 cards from a standard deck of cards. This is an example of sampling _____d______.

g. The Hawaii Visitors Bureau collects data on summer visitors to Hawaii. All visitors visiting Hawaii in the summer is referred to as a(n) _____g______.

h. Suppose that a pre-election commercial provides a number a person could call to "vote" for or against a certain candidate. The collection of callers represents a subset of all voters in the election and is referred as a(n) _____h______.

**Fortune Magazine** wanted to present the information regarding the top five Fortune 500 companies in an issue. The data is as follows:

<table>
<thead>
<tr>
<th>Company</th>
<th>Sales $ Millions</th>
<th>Sales Rank</th>
<th>Profits $ Millions</th>
<th>Profits Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Motors</td>
<td>161,315</td>
<td>1</td>
<td>2,956</td>
<td>30</td>
</tr>
<tr>
<td>Ford Motor</td>
<td>144,416</td>
<td>2</td>
<td>22,071</td>
<td>2</td>
</tr>
<tr>
<td>Wal-Mart Stores</td>
<td>139,208</td>
<td>3</td>
<td>4,430</td>
<td>14</td>
</tr>
<tr>
<td>Exxon</td>
<td>100,697</td>
<td>4</td>
<td>6,370</td>
<td>5</td>
</tr>
<tr>
<td>General Electric</td>
<td>100,469</td>
<td>5</td>
<td>9,269</td>
<td>3</td>
</tr>
</tbody>
</table>

i. The five companies are referred to as five _____i____ in the data set.

j. The labels “Sales $ Millions”, “Sale Rank”, “Profits $ Millions”, and “Profit Ranks” are referred to as four _____k______ in the data set.
2. A survey of 100 college students revealed the following facts about their sporting hobbies. Let $S$ be the event that a person swims, $R$ be the event that a person runs, and $B$ be the event that a person plays baseball. Create a Venn Diagram for the following data. (4 points)

- 42 run
- 34 play baseball
- 13 only swim
- 6 swim and run
- 14 swim and play baseball
- 11 run and play baseball
- 22 do no sport
- 2 swim, run, and play baseball

a. What is the probability that a randomly selected student plays baseball and runs? (3 points)

$$\frac{11}{100} = .11$$

b. What is the probability that a randomly selected student plays at least one sport? (3 points)

$$1-\frac{22}{100} = \frac{78}{100} = .78$$

c. What is the probability that a randomly selected student plays exactly 2 sports? (3 points)

$$\frac{12+4+9}{100} = \frac{25}{100} = .25$$
Given the following data, answer the following questions:

14, 15, 16, 13, 14, 14, 12, 13, 15, 14, 2, 2, 16, 10, 18
2, 2, 10, 12, 13, 13, 14, 14, 14, 15, 15, 16, 16, 18

a. draw a box plot of the data (3 points)

b. 25% of the data is above what value? (3 points)

15 (Q3) (the 12th Number)

c. What is the 30th percentile? (3 points)

30/100*15 = 4.5 so we look at the 5th element which is 13. So 13 is the 30th percentile

d. What percent of the data falls between the mode and the mean? (3 points)

Mean is 188/15 = 12.5
Mode = 14
2 data points fall between these two values, so 2/15 = .13
4. Answer the following True or False questions. Remember, a statement is true only if it is ALWAYS true. (2 points each)

   a. If events A and B are mutually exclusive $P(A) + P(B) = 1$  
      
      T  F

   b. In a frequency table, the sum of the cumulative frequency column is 1  
      
      T  F

   c. The value -.7 can be a variance, covariance or correlation  
      
      T  F

5. Start with a standard deck of 52 cards and remove all the hearts and all the spades, leaving 13 red and 13 black cards, including (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K) for each color where J, Q, and K are considered face cards. If you randomly select a card from this deck, find the probability that your card is:

   a. An Ace? (2 points)

      $2/26 = 0.077$

   b. Not a face card? (2 points)

      $20/26 = 0.77$

   c. Neither Red nor an Ace? (3 points)

      $12/26 = 0.46$

   d. Either Black, even-numbered, or a Jack? (3 points)

      $19/26 = 0.73$
6. Each of the following graphs, I, II and III, describes a relationship between two quantitative variables. Which of the following choices correctly orders the strength of the correlation between the two variables in ascending order? (3 points)

a. I, II, III  
b. III, II, I  
c. I, III, II  
d. II, III, I  
e. III, I, II
7. A drug company is trying to promote a new drug, Drug A. In order to show that Drug A is better than Drug B, the drug currently on the market, this company conducted an experiment with the following results. In the following table, the company compared the result of Drug A, Drug B and a Placebo on a total of 1104 patients. The first column in the table shows the three outcomes for the patients after the experiment. Use the information in the table to answer the following questions.

<table>
<thead>
<tr>
<th></th>
<th>Drug A</th>
<th>Drug B</th>
<th>Placebo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovered</td>
<td>276</td>
<td>243</td>
<td>21</td>
<td>540</td>
</tr>
<tr>
<td>Stays the same</td>
<td>67</td>
<td>86</td>
<td>281</td>
<td>434</td>
</tr>
<tr>
<td>worsened</td>
<td>37</td>
<td>39</td>
<td>54</td>
<td>130</td>
</tr>
<tr>
<td>Total</td>
<td>380</td>
<td>368</td>
<td>356</td>
<td>1104</td>
</tr>
</tbody>
</table>

a. Fill in the blank spaces in the chart above. (3 points)

b. Calculate the percentage of people who stayed the same after the experiment. Is this a marginal, conditional or joint probability? (3 points)

\[
\frac{434}{1104} = 0.39 \text{ Marginal}
\]

c. Calculate the percentage of people who took Drug A and recovered. Is this a marginal, conditional or joint probability? (3 points)

\[
\frac{276}{1104} = 0.25 \text{ Joint}
\]

d. If someone is given Drug B, what is the probability they either stay the same or recover? Is this a marginal, conditional or joint probability? (3 points)

\[
\frac{(243+86)}{368} = 0.89 \text{ Conditional}
\]
A university researcher wants to examine the number of hours students work per week during summer semester. He randomly chose 15 students and obtained the following data.

5  8  15  15  18  20  25  25  27  30  30  32  32  35  40

a. Draw a histogram to show the frequency for the data using classes of 5-14, 15-24, 25-34, 35+.

b. Is this right skewed, left skewed, or symmetric?

Left Skewed

c. What is the mean number of working hours? Please show your work.

\[ \bar{x} = \frac{\sum x_i}{n} = 23.8 \]

d. The standard deviation is 10.06. Is this a sample standard deviation or population standard deviation?

The data is a sample, so the standard deviation is computed using a sample Standard Deviation Formula as

\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \]

e. Is 5-hours an outlier of the dataset? Why or why not?

No.

Q1 = 15.

Q3 = 32

The IQR = 32 – 15 = 17. The lower bound for an outlier is 15 – 1.5*17 = -10.5 which is less than 5. So 5 hours of work is not considered as an outlier.
9. Which of the following graph(s) describe(s) a dataset where the mean is greater than the median? (3 points)

- a. I and III
- b. II and IV
- c. III only
- d. IV only
- e. II and III
10. A basic economic model states that in a competitive market, the price of a product will be affected by both supply and demand. The following data represents a sample of daily supply (y, in thousands of units) and unit price (x, in dollars) for a product.

<table>
<thead>
<tr>
<th>Daily Supply (y)</th>
<th>Unit Price (x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
</tr>
</tbody>
</table>

a. What is the standard deviation for Unit Price (x)? Please include the formulas in your work. (3 points)

\[ \bar{x} = 6 \]

\[ s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{148}{7} = 21.142 \]

\[ \text{Sqrt}(21.142) = 4.598 \]

b. The standard deviation of Daily Supply (y) is 11.08. Suppose the covariance has been computed and is -47.86. What is the correlation coefficient? (3 points)

\[ r_{xy} = \frac{s_{xy}}{s_x s_y} = -\frac{47.86}{11.08 \times 4.598} = -0.939 \]

c. Interpret your correlation coefficient. Is the correlation positive or negative? Does it imply a strong or weak linear relationship? How does price (y) change as the supply (x) changes? (3 points)

It is a negative correlation and indicates a strong linear relationship between the two variables. It means that as the supply of the products goes up, the unit price drops down.