This material is covered in webassign homework assignments 5 through 7 and worksheets 7-9. 

Exam information: materials allowed: calculator (no laptops, tablets or cellphone calculators), one 8\(\frac{1}{2}\) x 11 sheet (2 sided) with notes, definitions, formulas, etc..

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.

Answers are at the end of the document.

1. 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 2007-2009 survey?
3. What percent of the young girls who reported that they drink milk were in the 2007-2009 survey?
4. What percent of the young girls in 2007-2009 reported that they drink milk?

b. What is the marginal distribution of milk consumption?

2. 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 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>
</tr>
<tr>
<td>B</td>
<td>45 for 151</td>
<td>12 for 32</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?

3. A medical researcher wanted to examine the relationship between the amount of sunshine (x) in hours, and incidence of melanoma, a type of skin cancer (y). She found data showing the number of melanoma cases detected per 100,000 of population and the average daily sunshine in eight counties around the country. The data are shown below.

<table>
<thead>
<tr>
<th>Average daily sunshine</th>
<th>5</th>
<th>7</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>6</th>
<th>4</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma per 100,000</td>
<td>7</td>
<td>11</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

a. Which scatterplot below is the scatterplot of the above data?

i)
b. Given that $\sum (x_i - \bar{x})(y_i - \bar{y}) = 36$, what is the correlation $r$ between Average daily sunshine and Melanoma per 100,000?

4. Suppose that all major league baseball players had exactly 5 times as many runs batted in as home runs. Describe what information this provides about the correlation between runs batted in and home runs.

5. Outdoor temperature influences natural gas consumption for the purpose of heating a house. The usual measure of the need for heating is heating degree days. The number of heating degree days for a particular day is the number of degrees the average temperature for that day is below 65°F, where the average
temperature for a day is the mean of the high and low temperatures for that day. An average temperature of 20°F, for example, corresponds to 45 heating degree days. A homeowner interested in switching to solar heating panels collects the following data on her natural gas use for the months October through June, where \( x \) is heating degree days per day for the month and \( y \) is gas consumption per day in hundreds of cubic feet.

<table>
<thead>
<tr>
<th>Month</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>15.6</td>
<td>26.8</td>
<td>37.8</td>
<td>36.4</td>
<td>35.5</td>
<td>18.6</td>
<td>15.3</td>
<td>7.9</td>
<td>0</td>
</tr>
<tr>
<td>( y )</td>
<td>5.2</td>
<td>6.1</td>
<td>8.7</td>
<td>8.5</td>
<td>8.8</td>
<td>4.9</td>
<td>4.5</td>
<td>2.5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Calculate the correlation coefficient \( r \) and interpret its value; draw a scatterplot of the data.

6. 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 \( r = 0.23 \) bushel.”
   d. If the correlation between exam grades (\( x \)) and hours spent studying (\( y \)) is 0.83, and the variable names \( x \) and \( y \) are reversed, then the correlation between hours spent studying (\( x \)) and exam grades (\( y \)) is \(-0.83\).

7. Suppose that a PE teacher collected data about the students in his class. Some of these variables included number of pull-ups in 1 minute, number of push-ups in 1 minute, number of sit-ups in 1 minute, and weight. The teacher then calculated the correlation between number of pull-ups and each of the other three variables and found the following correlations: \( r = 0.9, r = -0.5, r = 0.3 \). Which correlation goes with which variable? Explain.

8. The following scatterplots based on data from a recent season show the association between the number of points scored by the teams in the National Basketball Association (NBA) and three different explanatory variables (number of field goals, free throw percentage, and number of free throws). The correlations are (in no particular order) \( r = 0.20, r = 0.61, r = 0.89 \). Match the correlations to the scatterplots.

9. Consider the following scatterplot.
Which of the following is a plausible value for the correlation coefficient between weight and MPG?

a. -0.9   b. -1.0   c. +0.2   d. +0.9   e. +0.7

10. 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. $(0.9)^4$   b. $(0.1)^4$   c. $1 - (0.9)^4$   d. $1 - (0.1)^4$   e. $0.9 + 0.9 - (0.9)(0.9) = 0.99$

11. 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. 0.075   b. 0.05   c. 0.55   d. none of these   e. can't tell from information given

12. A manufacturer of hand soap has introduced a new product. An extensive survey indicates that 40% of the people have seen advertising for the new product. It also showed that 20% of the people in the survey had tried the new product. In addition, 15% of those in the survey had seen it advertised and had tried the product. What is the probability that a randomly chosen person would have seen the advertising for the new product or have tried the product or both?

a. 0.6   b. 0.45   c. 0.08   d. 0.25   e. 0.05

13. 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. $(10)^4$   b. 10!   c. $10 P_4$   d. 5040   e. $\frac{10!}{4!}$

14. Determine which of the following functions is in fact a probability distribution function.

a. $p(x) = \frac{1}{4}, x = 3, 4, 5, 6.$

b. $p(x) = \frac{x^2}{27}, x = 0, 1, 2, 3, 4.$

c. $p(x) = \frac{5-x^2}{6}, x = 0, 1, 2, 3.$
15. In a population of students the number of calculators owned is a random variable $x$ with $p(0) = .2, p(1) = .6, p(2) = .2$. Find the expected value and standard deviation of this probability distribution.

16. An oil firm is to drill three wells, with each well having probability 0.2 of successfully producing oil. It costs the firm $20,000 to drill each well. A successful well will bring in oil worth $750,000. Let the random variable $X_i$ be the oil firm's gain from well $i, i = 1, 2, 3$. The wells are in different geographic areas and so the drilling outcome at any well has no affect on the drilling outcomes at the other wells.
   a. Find the firm's expected gain $G$ from the three wells.
   b. Find the standard deviation of the firm's gain.

17. Let the random variable $X$ denote the displacement in cubic inches $(in^3)$ of the engine in a particular model of automobile. The size of the engine (that is, the cubic inch displacement) varies depending on the options chosen by the buyer of the automobile. It is known that $E(X) = 177$ in$^3$ and $\sigma_X = 22$ in$^3$. If $X^*$ denotes the engine displacement in cubic centimeters $(cm^3)$, determine $E(X^*)$ and $\sigma_{X^*}$. Note that 1 in$^3 = 16.4$ cm$^3$.

18. Let $X$ be the number of accidents per week at a hazardous intersection; $X$ varies with mean 2.2 and standard deviation 1.4. Let $X_1, X_2,$ and $X_3$ be the number of accidents in each of 3 different weeks at this intersection. The number of accidents in a week is not affected by the number of accidents in any other week. What is $\sigma(X_1 + X_2 + X_3)$, the standard deviation of the sum $X_1 + X_2 + X_3$?

19. The probability distribution below describes the number of repair calls that an appliance repair shop may receive during an hour.

<table>
<thead>
<tr>
<th>Repair calls</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

   a. How many calls should the shop expect to receive per hour? What is the standard deviation?
   b. Find the expected value and standard deviation of the number of repair calls the appliance shop should expect during an 8-hour day.

20. A college student on a meal plan reports that the amount of money he spends daily on food varies with a mean (expected value) of $13.50 and a standard deviation of $7.
   a. Find the expected value and standard deviation of the amount he spends on 2 consecutive days (the amounts he spends on different days are independent).
   b. Find the expected value and standard deviation of the amount he spends during a semester that spans 120 days.

### ANSWERS

1. a1. 56.9%  a2. 38.9%  a3. 41.1%  a4. 60%  b. Yes: 56.9%; No: 43.1%  2. 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. 3. a) b. $r = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$ where $s_x$ and $s_y$ are, respectively, the standard deviation of the $x$-data and $y$-data.

Using a calculator to compute the standard deviations, we have $s_x = 1.669$ and $s_y = 3.207$. Therefore

$$ r = \frac{\frac{36}{1}}{\frac{1}{3.207 \cdot 1.669}} = .961 $$

4. Since $Runs\ batted\ in = 5\ \text{home\ runs}$, there is an exact linear relationship between $Runs\ batted\ in$ and $\text{home\ runs}$. Therefore, there is a perfect correlation, so the correlation is $-1$ or $+1$. Since the coefficient (or slope) “5” is positive, the correlation is $+1$.

5. $r = .989$. There is a strong positive linear relationship between heating degree days and gas consumption.

6. 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.
d. the value of the correlation is independent of what variable is called \(x\) and what variable is called \(y\)

7. \(r = 0.9\): number of push-ups. Someone who is good at pull-ups should be good at push-ups since both measure arm strength.
   \(r = -0.5\): weight. You would expect a negative association between pull-ups and weight because it is more difficult to do pull-ups when you are heavier.
   \(r = 0.3\): number of sit-ups. You would expect a positive association between pull-ups and sit-ups because they both measure physical fitness, but the association would not be as strong as the association between pull-ups and push-ups since they are different forms of exercise.

8. 

\[
\begin{array}{ccc}
\text{Points} & \text{Points} & \text{Points} \\
\text{FieldGoals} & \text{FreeThrowPercent} & \text{FreeThrowAttempts} \\
\end{array}
\]

\[
r = 0.89 \quad r = 0.20 \quad r = 0.61
\]

9. a

10. c. since 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 \(1 - (.9)^4\).

11. c

12. b. since \(P(\text{seen advertising} \cap \text{tried the product}) = .4 + .2 - .15 = .45\).

13. c. and d. are both correct since \(10 \times (9)(8)(7) = 5040 = \text{P}^{10}_3\).

14. a. is the only one since in b. \(\sum p(x) > 1\); in c. \(p(3) < 0\).

15. \(\mu = 0(.2) + 1(1.6) + 2(2.2) = 1\)
   \(\sigma = \text{sqrt}[[0 - 1)^2(0.2) + (1 - 1)^2(0.6) + (2 - 1)^2(0.2)] = \text{sqrt}([.4] = .6325

16. a. \(E(X_1) = .2*730,000 + .8*(-20,000) = 130,000; G = X_1 + X_2 + X_3; E(G) = E(X_1) + E(X_2) + E(X_3) = 390,000.\)
   b. \(V ar(X_1) = [(730,000 - 130,000)^2 + (-20,000 - 130,000)^2]*.8 = 9 \times 10^{10};\)
   \(V ar(G) = V ar(X_1) + V ar(X_2) + V ar(X_3) = 27 \times 10^{10};\)
   \(SD(G) = \sqrt{V ar(G)} = \sqrt{27 \times 10^{10}} = 519,615.24.\)

17. \(E(X^2) = 177 \text{in}^3 \times 16.4 \text{cm}^3/\text{in}^3 = 2902.8 \text{cm}^3;\)
   \(\sigma_{X_2} = 22 \text{in}^3/16.4 \text{cm}^3 = 360.8 \text{cm}^3.\)

18. \(V ar(X_1 + X_2 + X_3) = V ar(X_1) + V ar(X_2) + V ar(X_3) = 3*1.96 = 5.88; \text{so}\)
   \(\sigma_{(X_1 + X_2 + X_3)} = \sqrt{5.88} = 2.425.\)

19. a. \(E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 81.7 = 13.6 \text{calls}\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.7 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1\) \(\text{calls}\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.1 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1 = 13.6 \text{calls}\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.1 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1 = 13.6 \text{calls}\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.1 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1 = 13.6 \text{calls}\)

20. a. \(E(X_1 + X_2) = E(X_1) + E(X_2) = 27\)
   \(V ar(X_1 + X_2) = V ar(X_1) + V ar(X_2) = 49 + 49 = 98\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.1 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1 = 13.6 \text{calls}\)
   \(\text{Let } X_i = \text{the number of calls received during hour } i, i = 1, 2, 3, \ldots, 8.\)
   \(E(X_1 + X_2 + \ldots + X_8) = E(X_1) + E(X_2) + \ldots + E(X_8) = 81.1 = 13.6 \text{calls}\)
   \(V ar(X_1 + X_2 + \ldots + X_8) = V ar(X_1) + V ar(X_2) + \ldots + V ar(X_8) = 81.1 = 13.6 \text{calls}\)