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>16</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Cheat Sheet</td>
<td>1</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 completes the sentences below. Each phrase should be used no more than once. (2 points each)

(A) Experiment  (B) Event  (C) With Replacement
(D) Without Replacement  (E) Intersection  (F) Union
(G) Mutually Exclusive  (H) Variance  (I) Covariance
(J) Correlation coefficient  (K) Outlier(s)  (L) Sample Space
(M) Simple Random Sample  (N) Convenience Sample
(O) Stratified Sample  (P) Interquartile Range
(Q) Census  (R) Pie Chart

1. You are drawing 5 cards from a deck of 52 and get a full house (3 of a kind and a pair). That is called a (an) ________ in probability.

2. A researcher uses social economic status (SES) as the only factor to predict people’s income. He wants to know how much variation in people’s income is explained by SES and we would suggest one quantity, __________, will help him answer this question.

3. Yong wanted to get his students’ idea of his teaching, so he decided to design a questionnaire to have all his students fill out. Therefore, he is planning to take a(an) _________. However, he found that only half of his students showed up in class the next day. He decided to let the students fill out the survey anyway. He finally used a(an)_______ to collect data.

4. All the possible lottery numbers constitute a(an)_______ for a lottery drawing. When someone buys a lottery ticket, he/she is sampling ____C/D______.

5. In stat225, 60 students are business majors and 20 students are liberal arts majors. It is believed that students from different majors have different expectations from this course. If we want to take a survey of 40 students, the best sampling scheme is to use a(an)_______.

6. After the first midterm, you want to know your ranking in this class and you will be happy if you are in the middle 50%. Therefore, in addition to your own grade and the 25th percentile, you also need to know __________.

7. A university official is interested in the starting salary for class 2010 and he found that the mean salary is very different when he included and excluded a few student athletes. A STAT225 student on this project informed him that those athletes should be considered as ________.

8. __________ describes how data points in a sample are distributed about the mean.
2. Suppose there are 100 scholars in attendance at the 2010 Smithsonian History Convention.

22 are experts in both Ancient and Natural History;
19 scholars are experts in both Natural and Modern History;
23 scholars are experts in both Ancient and Modern History;
51 scholars are experts in Ancient History;
52 are experts in Natural History;
48 are experts in Modern History;
3 are reporters who have no expertise in any of the three histories.

(A: Ancient History; N: Natural History; M: Modern History)

a. Complete the following Venn diagram (4 points).

Assume an attendee is selected at random and find the following probabilities:

b. The attendee is an expert in Ancient or Modern History. (2 points)

\[ P(A+M) = P(A) + P(M) - P(AM) = \frac{51}{100} + \frac{48}{100} - \frac{23}{100} = \frac{76}{100} \]

c. The attendee is an expert in Ancient or Natural History but not Modern History. (2 points)

\[ \frac{16 + 21 + 12}{100} = \frac{49}{100} \]

d. The attendee has expertise in all three of the histories. (2 points)

\[ \frac{10}{100} \]
3. A card is selected at random from a standard 52 card deck. Let A denote the event the card is red and B denote the event the card is “Royal” (Jack, Queen, or King). Find the probability of the following events (2 points each):

a. A and B
\[ P(A \cap B) = \frac{6}{52} = \frac{3}{26} = 0.115 \]

b. A^c
\[ P(A^c) = \frac{26}{52} = 0.5 \]

c. A or B
\[ P(A) + P(B) - P(AB) = \frac{26}{52} + \frac{12}{52} - \frac{6}{52} = \frac{32}{52} = 0.615 \]

d. A and B^c
\[ P(A \cap B^c) = P(A) - P(AB) = \frac{26}{52} - \frac{6}{52} = \frac{20}{52} = 0.385 \]

4. Given the data in the table, answer the following questions:

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>15</td>
<td>19</td>
<td>27</td>
<td>33</td>
<td>37</td>
<td>44</td>
<td>52</td>
<td>64</td>
<td>86</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>23</td>
<td>27</td>
<td>34</td>
<td>38</td>
<td>46</td>
<td>52</td>
<td>68</td>
<td>89</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>24</td>
<td>29</td>
<td>35</td>
<td>39</td>
<td>49</td>
<td>53</td>
<td>71</td>
<td>95</td>
</tr>
<tr>
<td>11</td>
<td>19</td>
<td>25</td>
<td>31</td>
<td>35</td>
<td>42</td>
<td>51</td>
<td>59</td>
<td>75</td>
<td>104</td>
</tr>
<tr>
<td>13</td>
<td>19</td>
<td>26</td>
<td>32</td>
<td>35</td>
<td>42</td>
<td>52</td>
<td>60</td>
<td>75</td>
<td>112</td>
</tr>
</tbody>
</table>

a. Construct a boxplot and label ALL relevant numbers on the plot. (4 points)

UL: 53 + 1.5*29 = 96.5
Q3: 53
Median: 36
Q1: 24
LL: 24 - 1.5*29 = -19.5
Outliers: 104 and 112

b. Is the data left skewed, right skewed or symmetric? Justify your answer. (2 points)

Right skewed. Since there are two high outliers.

c. Are there any outliers? If yes, what are they? Justify your answer. (3 points)

Outliers: 104, 112. Since UL=96.5 and LL=-19.5.
5. The following table gives counts of the dual classification of 1332 preschoolers by their fathers’ educational attainment and whether both, one, or neither of the parents smoke. (3 points each)

<table>
<thead>
<tr>
<th></th>
<th>Both Parents Smoke</th>
<th>One Parent Smokes</th>
<th>Neither Parent Smokes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>University Education</td>
<td>42</td>
<td>68</td>
<td>90</td>
<td>200</td>
</tr>
<tr>
<td>Intermediate Education</td>
<td>47</td>
<td>69</td>
<td>75</td>
<td>191</td>
</tr>
<tr>
<td>High School Education</td>
<td>183</td>
<td>281</td>
<td>273</td>
<td>737</td>
</tr>
<tr>
<td>Primary Education or None</td>
<td>69</td>
<td>73</td>
<td>62</td>
<td>204</td>
</tr>
<tr>
<td>Total</td>
<td>341</td>
<td>491</td>
<td>500</td>
<td>1332</td>
</tr>
</tbody>
</table>

a. If we randomly choose a preschooler whose father has attained high school education, what is the probability neither of her parents smoke? Is this a joint, marginal or conditional probability?

\[
\frac{273}{737} = 0.370, \text{ Conditional}
\]

b. What is the probability a randomly chosen preschooler has at least one parent smoke? Is this a joint, marginal, or conditional probability?

\[
\frac{(341+491)}{1332} = 0.625, \text{ Marginal}
\]

c. What is the probability a randomly chosen preschooler has one parent smoke and the father has completed no higher than intermediate level education? Is this a joint, marginal, or conditional probability?

\[
\frac{(69+281+73)}{1332} = 0.318, \text{ Joint}
\]
6. Use the data below to answer the following questions

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3</td>
<td>-2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>-1</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

a. Calculate the sample variance of X ($s^2_X$). (3 points)

b. First: the mean of X is: $\bar{X} = \frac{-3 + 2 + 1 + 7 + 3}{4} = \frac{3}{4} = 0.75$

$s^2_X$ is: $\frac{(-3 - 0.75)^2 + (-2 - 0.75)^2 + (1 - 0.75)^2 + (7 - 0.75)^2}{4 - 1} = \frac{60.75}{3} = \frac{81}{4} = 20.25$

b. Given:
   i. the sample variance of Y ($s^2_Y$) = 10/3
   ii. the sample correlation of X and Y ($r_{XY}$) ≈ -0.933, find the sample covariance of X and Y ($s_{XY}$). (3 points)

$$r_{XY} = \frac{s_{XY}}{s_x s_y} \Rightarrow s_{XY} = r_{XY} * s_x s_y = r_{XY} * \sqrt{s^2_x} * \sqrt{s^2_y} \approx (-0.933) * \sqrt{20.25} * \sqrt{\frac{10}{3}} \approx -7.66$$

c. Given:
   i. the sample variance of Z ($s^2_Z$) = 14,
   ii. the sample covariance of X and Z ($s_{XZ}$) = -5/3
   iii. the sample covariance of Y and Z ($s_{YZ}$) = 1/3

Which set of variables (XY, XZ, or YZ) has the weakest correlation? (3 points)

$$r_{XY} \approx -0.933$$

$$r_{XZ} = \frac{s_{XZ}}{s_x s_z} = \frac{s_{XZ}}{\sqrt{s^2_x s^2_z}} = \frac{\left( -\frac{5}{3} \right)}{\sqrt{20.25} \sqrt{14}} \approx -0.099$$

$$r_{YZ} = \frac{s_{YZ}}{s_y s_z} = \frac{s_{YZ}}{\sqrt{s^2_y s^2_z}} = \frac{\left( \frac{1}{3} \right)}{\sqrt{\frac{10}{3} \sqrt{14}}} \approx 0.0488$$

So Y and Z have the weakest correlation.
7. 18 players on a soccer team warming up before a game are wearing various colored soccer socks: 8 white, 7 black, and 3 blue. (3 points each)

a. How many possible ways are there to line up the 18 players?

\[ 18! = 6.402374 \times 10^{15} \]

b. What is the probability that players wearing socks of the same color will stand together in a lineup?

\[ \frac{3!8!7!3!}{18!} = 1.142648 \times 10^{-6} \]

c. What is the probability that all the players wearing blue socks will stand together in a line up?

\[ \frac{16!3!/18!}{0.02} \]

d. In a random selection of three players, what is the probability that all three players are wearing white socks?

\[ \frac{8C_3}{18C_3} = 0.06862745 \]

e. In a random selection of two players, what is the probability that they are wearing different colored socks?

\[ \frac{8C_1 \cdot 7C_1 + 8C_1 \cdot 3C_1 + 7C_1 \cdot 3C_1}{18C_2} = 0.6601307 \]
8. The following graph measures the body weight (bwt, in lb) of a group of subjects and the distance they can run before taking a break (distbreak, in feet). The estimated trend line is distbreak=-4.979*bwt+2664.311 with a $R^2$ of 0.6388.

a. Interpret the slope in the context of the problem. (3 points)

For every extra pounds in body weight, the subject runs about 5 feet less before taking a break. (answers may vary)

b. If somebody’s body weight is 220 lbs, how long would we predict he can run before taking a break? How about someone whose body weight is 400 lbs? Which of these two predictions is more reliable, why? (4 points)

Distbreak=-4.979*220+2664.311=1568.931 (feet);
Distbreak=-4.979*400+2664.311=672.711 (feet). The first one is more reliable since 220 lb is in the range of data while 400 is not.

c. What is the Pearson (sample) correlation coefficient between body weight (bwt) and the running distance before taking a break (distbreak)? Is it a strong, moderate or weak correlation? (3 points)

$\sqrt{0.6388}=-0.8$, strong or moderate.

d. What does the $R^2$ mean in this problem? (2 points)

$R^2$ means among the total variation in distbreak, about 64% is explained by bodyweight.
9. A fast food restaurant, Becker K, is running a promotion which allows you to choose a mix and match of 3 items from cheeseburger, grill-chicken sandwich, French fries, onion rings, coffee and fountain drink, all for $5.99. Assuming all the items on the menu have an equal chance of being selected, answer the following questions. (A customer can choose the same item more than once but the total number of items can not exceed three).

a. How many possible ways can a customer place an order? (3 points)

\[ C_1^6 + P_2^6 + C_3^6 = 56 \]

b. What is the probability that a customer orders a meal (one entrée, one side and one drink)? (4 points)

\[ \frac{2 \times 2 \times 2}{56} = \frac{1}{7} = 0.1428571 \]

c. The store manager found in the morning that their fountain machine is broken and their store policy requires that they fix it if it will be needed by at least 20% of all orders? Do you think the store manager should fix the fountain machine? (State your answer and show the supporting work). (4 points)

Here, we need to find the probability that an order includes at least one fountain drink (which mean 1, 2 or 3 fountain drinks), which is,

\[ \frac{(1+5+C_2+C_3)}{56} = 0.375 \].

Therefore, the manager should have the fountain drink machine fixed.

(Note: if a customer orders 1 fountain drink, he/she may order another 2 different items \( C_2 \) or another 2 same entries \( C_1 \).)