STAT 225 – Spring 2010 – Final Solutions

Your Name: _________________________________________________________
Your Instructor: _____________________________________________________
Your class time (circle one):

<table>
<thead>
<tr>
<th>7:30</th>
<th>8:30</th>
<th>9:30</th>
<th>10:30</th>
<th>11:30</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:30</td>
<td>1:30</td>
<td>2:30</td>
<td>3:30</td>
<td>4:30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Show work** for full credit – unsupported work will NOT receive full credit
- All answers should be in **decimal form**: no fractions, permutation, combination, or exponential form. Round all answers to at least 2 decimal places.
- You are responsible for upholding the Honor Code of Purdue University. This includes **protecting your work** from other students.
- You are allowed 2 pages 8.5”x11” handwritten **cheat sheets** and a **calculator**.
- Instructors will **not interpret** questions, tell you if you’re on the right track, or check any answers for you. Only legitimate questions will be answered.
- You must **turn in your Cheat Sheet** at the end of the exam and may be asked to show your student ID.
- **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>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Cheat Sheet(s)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
1. For each of the following scenarios, write the letter corresponding to the appropriate
distribution as stated in each question. Distributions may be used more than once.
(2 points each)

A. Binomial  B. Hypergeometric  C. Poisson
D. Uniform  E. Exponential  F. Normal

___B___ a. You are dealt 7 Uno cards. Let X be the number of “Skips” you have in your hand.
What is the exact distribution of X?

___E___ b. It takes an average of 25 minutes for someone to win a game of chess. A game has
begun; let X be the time until someone wins. What is the distribution of X?

___C___ c. One in every 10,000 eggs contains Salmonella. Let X be the number of eggs with
salmonella in a shipment of 50,000 eggs. What is the approximate distribution of X?

___C___ d. The subway stops at Boston College an average of 5 times every 30 minutes. Let X
be the number of times the subway stops at Boston College in 10 minutes. What is the
distribution of X?

___A___ e. You select 3 cards with replacement from a deck of 52 cards. Let X be the number
of clubs you pick. What is the exact distribution of X.

___A___ f. You select a handful of coins from a 5-gallon container. Let X be the number of
pennies you select. What is the approximate distribution of X?

___D___ g. You are anticipating an email between 2 and 4 pm, all times equally likely. Let X the
be the time you wait for your email to appear in your inbox. What is the distribution of
X?

___F___ h. In a certain population, 40 percent of people prefer chocolate over vanilla ice
cream, all preferences independent of any others. Let X be the number of people in a
crowd of 70 that prefer chocolate over vanilla. What is the approximate distribution of
X?
2. Crocuses are one of the first flowers to bloom in spring, blooming between 10 and 20 days after the first official day of spring with all blooming times equally likely. If crocuses bloom every year around the “Hello Walk” on Purdue’s campus, let \( C \) be the number of days after the first day of spring it takes for the crocuses to bloom.

   a. What are the distribution and parameters of \( C \)? (2 points)

   \[ C \sim \text{Uniform}(10, 20) \]

   b. What is the probability the crocuses bloom within the first two weeks of spring? (2 points)

   \[ P(C < 14) = F(14) = \frac{14 - 10}{10} = 0.4 \]

   c. How long do you expect it to take the crocuses to bloom? (2 points)

   \[ E(C) = \frac{10 + 20}{2} = 15 \text{ days} \]

   d. What is \( E(C^2) \)? (3 points)

   \[ E(C^2) = Var(C) + E(C)^2 \]

   \[ Var(C) = \frac{(20 - 10)^2}{12} = 8.33 \]

   \[ E(C^2) = 8.33 + 15^2 = 233.33 \]
3. The time it takes to housebreak dogs is normally distributed with a mean of 2 weeks and a standard deviation of 4 days. The time it takes to litter train cats is normally distributed with a mean of 1 week and a standard deviation of 2 days. Additionally, you purchase both a dog and a cat and the time to train a dog is independent of the time to train a cat.

a. What is the probability it takes you longer than 10 days to housebreak your dog? (3 points)

\[ P(D > 10) = P( Z > \frac{10 - 14}{4} ) = P(Z > -1) = 1 - \Phi(-1) = 0.8413 \]

b. 88% of the time it will take you less than what number of days to litter train your cat? (3 points)

\[
\Phi(z) = 0.88 \\
z = \frac{x - 7}{2} \\
x = 9.34
\]

c. What is the probability either your dog or cat will be trained in less than 10 days? (4 points)

\[ P(D < 10 \cup C < 10) = P(D < 10) + P(C < 10) - P(D < 10 \cap C < 10) \]

\[ = 0.1587 + 0.9332 - 0.1587 \times 0.9332 = 0.9438 \]

Or

\[ = 1 - P(D > 10 \text{ and } C > 10) = 1 - (1 - 0.1587)(1 - 0.9332) = 0.9438 \]
4. The PDF of a random variable $X$ is:

$$f(x) = \begin{cases} \frac{x^3 + 1}{24} & -1 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

a. Show that this is a legitimate PDF. (2 points)

1. All values are $> 0$

2. \[ \int_{-1}^{3} \frac{x^3 + 1}{24} \, dx = \frac{1}{24} \left( \frac{x^4}{4} + x \right) \bigg|_{-1}^{3} = \frac{1}{24} \left( \left( \frac{3^4}{4} + 3 \right) - \left( \frac{1}{4} - 1 \right) \right) = 1 \]

b. What is the CDF of $x$? (4 points)

\[
F(x) = \begin{cases} 
0 & x < -1 \\
\frac{1}{96} (x^4 + 4x + 3) & -1 \leq x \leq 3 \\
1 & x > 3 
\end{cases}
\]

c. What is the probability $x$ is less than 1? (2 points)

\[
P(x < 1) = F(1) = \frac{1}{196} (1 + 4 + 3) = \frac{8}{96} = 0.083
\]

d. Find the $E(2X)$. (3 points)

\[
E(2X) = \int_{-1}^{3} 2x \cdot \frac{x^3 + 1}{24} \, dx = \frac{2}{24} \left( \frac{x^5}{5} + \frac{x^2}{2} \right) \bigg|_{-1}^{3} = \frac{2}{24} (52.8) = 4.4
\]
5. In 1960 the proportion of left handed people was .08. It has been proposed that there are more left-handed people now than there were 50 years ago because teachers no longer force students to only use their right hand. To test this theory, a sample of 81 people was taken, and 7 of them were found to be left handed. Let \( L \) be the proportion of left handed people in the population in 2010.

a. What are the distribution and parameters of \( L \)? (3 points) *(The solution to this problem is controversial and can be ignored)*

\[
L \sim \text{Normal} \left( \mu = \frac{7}{81} = 0.086, \sigma^2 = \frac{(0.86)(1 - 0.86)}{81} = 0.00097, \sigma = 0.031 \right)
\]

b. There is a 30 percent chance the proportion of left handed people is above what value? (3 points)

\[
P(L > x) = 0.3 \text{ so } P(L < x) = 0.7
\]

\[
\Phi(0.52) = 0.7 \quad x - 0.86
\]

\[
0.52 = \frac{0.031}{x}
\]

\[
x = 0.102
\]

c. Based on our sample proportion, if we were to look at the 420 students currently in Stat 225, how many would we expect to be left handed? (2 points)

\[
420 \times 0.086 = 36.12 \text{ or } 36 \text{ people}
\]

d. If in the 420 students we find that exactly 44 of them are left handed and take a random sample of 10 students. What is the approximate probability there is at least one left handed student in our sample? (3 points)

Exact Dist is \( \text{HG} \left( N = 420, n = 10, p = \frac{44}{420} = 0.1047 \right) \)

Approx Dist \( \text{Bin} \left( n = 10, p = 0.1047 \right) \)

\[
P(\text{at least 1 left handed student}) - 1 - P(\text{no left handed student})
\]

\[
= 1 - (1 - 0.1047)^{10} = 0.669
\]
6. For the following True/False questions, circle the correct answer. Keep in mind that a statement is only true if it is always true. (2 points each)

a. For an Exponential distribution the median is less than the mean  
   T  F

b. For a Uniform distribution the median is less than the mean  
   T  F

c. \( f(x) = x^2 - \frac{5}{6} \) on the interval (0,2), 0 otherwise is a legitimate PDF  
   T  F

d. \( \int_{-\infty}^{\infty} f(x) \, dx = 1 - F(a) \)  
   T  F

e. \( f(a) = P(X = a) \)  
   T  F

f. \( \int_{-\infty}^{\infty} x^3 f(x) \, dx = E(X^3) \)  
   T  F

g. For a given \( \bar{X} \), a 90% confidence interval with \( n = 50 \) is wider than a 90% confidence interval with \( n = 500 \).  
   T  F

h. The variance of \( \sum X_i \) with \( n = 50 \) is larger than the variance of \( \sum X_i \) with \( n = 500 \)  
   T  F
7. A study was conducted to see if eating breakfast resulted in higher test scores, so a sample of 80 Stat 225 students was taken after Exam 2. Of the 80 students sampled, 49 ate breakfast. The average exam scores for those without breakfast was 72 and those with breakfast was 74. Assume test scores have a standard deviation of 10 points, regardless of their breakfast choice.

a. Let B represent the average scores of all stat 225 students that ate breakfast based on our sample. What are the distribution and parameters of B? (3 points)

\[ B \sim \text{Normal} \left( \mu = 74, \sigma = \frac{10}{\sqrt{49}} \right) \]

b. Construct a 90% confidence interval for B. (3 points)

\[ CI = 74 \pm 1.645 \times \left( \frac{10}{7} \right) = (71.65, 76.35) \]

c. Based on your confidence interval, do you think eating breakfast actually results in higher test scores? Why or why not? (2 points)

No because 72 is within the CI it is a possible population mean for those who did eat breakfast, which is the same as those that did not eat breakfast

8. After finals, you decide to take a nice relaxing vacation and spend the day at the beach. While on the beach, you notice that you see an average of 4 people on Jet Ski’s go by every 30 minutes, and all pass by independently of one another. Let J be the time until you see the next person on a Jet Ski pass by.

a. What are the distribution and parameters of J? (2 points)

\[ J \sim \text{Exponential} \left( \mu = \frac{30}{4} \text{ minutes} = 7.5 \text{ min} \right) \]

b. Given it’s been 8 minutes since you saw the last Jet Ski go by and haven’t seen another one, what is the probability it will be at least 15 minutes until you see the next person pass by on a Jet Ski? (3 points)

\[ P(J > 15 | J > 8) = P(J > 7) = e^{-\left(\frac{7}{7.5}\right)} = 0.393 \]

c. Thirty percent of the time it will take you less than how many minutes to see the next Jet Ski? (3 points)

\[ F(x) = .3 = 1 - e^{-\left(\frac{x}{7.5}\right)} \]
\[ x = -7.5(\ln(0.7)) = 2.675 \text{ minutes} \]
9. You have been subscribing to Netflix for some time now and have rated over 100 movies. Netflix will suggest movies it predicts you will like, and you rent each of these movies. The probability you will actually like a movie it suggested is .8, independent of all other movies. Over the course of the year you watch 60 moves Netflix has suggested. Let N be the number of movies Netflix suggested that you actually enjoyed.

   a. What is the distribution of N and what are its parameters? (2 points)

   \[ N \sim Bin (n = 60, p = 0.8) \]

   b. What is the probability you will like exactly 48 movies? (2 points)

   \[ P(N = 48) = \binom{60}{48} (0.8)^{48} (0.2)^{12} = 0.1278 \]

   c. Let \( N^* \) be an approximation of N. State the distribution and parameters of \( N^* \) and why you can use it. (3 points)

   \[ N^* \sim Normal (\mu = 60 \times 0.8 = 48, \sigma^2 = 60 (0.8)(0.2) = 9.6, \sigma = 3.1) \]

   We can use this approximation because:

   \[ np = 48 > 5 \text{ and } n(1-p) = 12 > 5 \]

   d. What is the approximate probability N will be either below 45 or above 50? (3 points)

   \[ P(N < 45 \cup N > 50) = P(N^* < 44.5 \cup N^* > 50.5) \]

   \[ \Phi\left(\frac{44.5 - 48}{3.1}\right) + \left(1 - \Phi\left(\frac{50.5 - 48}{3.1}\right)\right) \]

   \[ \Phi(-1.13) + (1 - \Phi(0.81)) = 0.399 \]