Stat 225 - Summer 2012 Exam 2

Your Name: ____________________

Your Section (circle one):
Sveinn (08:40)  Glen (09:50)  Mike (11:00)

Instructions:

- Show your work on ALL questions. Unsupported work will NOT receive full credit.
- Decimal answers should be exact, or to exactly 2 non-zero decimal places. (Examples: if it is .25 use .25, if it is say .00891234 then use .0089.)
- You are responsible for upholding the Honor Code of Purdue University. This includes protecting your work from other students.
- Please write legibly. If a grader cannot read your writing, NO credit will be given.
- You are allowed the following aids: two 8.5" x 11" handwritten (in your handwriting) cheat sheets, a scientific calculator, and pencils or pens.
- 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), turn in your cheat sheet and sign the class roster 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 Earned</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>15</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. A Stat 225 instructor's daily beverage consumption can be described as follows: He drinks anywhere between 0 and 2 liters of Diet Coke, between 0 and 2 liters of Powerade, and before he goes to bed he gulps down 1 to 2 liters of water. For each beverage he is equally likely to drink any amount within the corresponding interval. Let C, P, and W denote his Coke, Powerade, and water consumption, respectively. (3 points each)

a) State the distributions and the parameters for C, P and W

\[ C \sim \text{Unif} \left( 0, 2 \right) \]
\[ P \sim \text{Unif} \left( 0, 2 \right) \]
\[ W \sim \text{Unif} \left( 1, 2 \right) \]

b) What is the probability that on any given day, he consumes at least 1.5 liters of Diet Coke?

\[ P(C > 1.5) = \frac{2 - 1.5}{2 - 0} = 0.25 \]

c) Given that on July 1st he had at least a liter of Powerade, what is the probability he had at most 1.4 liters of Powerade that day?

\[ P(P \leq 1.4 \mid P \geq 1) = \frac{P(1 \leq P \leq 1.4)}{P(P \geq 1)} = \frac{1.4 - 1}{2 - 1} = 0.4 \]

d) What is the expected value of his total daily consumption?

\[ E[C + P + W] = E[C] + E[P] + E[W] = \frac{2 + 0}{2} + \frac{2 + 0}{2} + \frac{2 + 1}{2} = 3.5 \]
2. Let \( X \) be a continuous random variable with the following PDF (probability density function):

\[
f(x) = \begin{cases} 
\frac{x-2}{c} & 2 < x < 6 \\
0 & \text{otherwise}
\end{cases}
\]

a) Find the number \( c \) such that \( f(x) \) is a legitimate PDF (3 points)

\[
\int_{2}^{6} f(x) \, dx = 1 \quad \Rightarrow \quad \int_{2}^{6} \frac{x-2}{c} \, dx = \left[ \frac{x^2}{2c} - \frac{2x}{c} \right]_{2}^{6} = \frac{36}{2c} - \frac{12}{c} = \frac{4}{2c} - \frac{4}{c}
\]

\[
\Rightarrow \quad \frac{36}{2c} - \frac{12}{c} - \frac{4}{2c} + \frac{4}{c} = 1
\]

\[
\Rightarrow \quad c = 8
\]

b) Find the CDF (cumulative distribution function) for the random variable \( X \) (3 points)

\[
F(a) = \int_{2}^{a} f(x) \, dx = \int_{2}^{a} \frac{x-2}{8} - \frac{1}{4} \, dx = \left[ \frac{x^2}{16} - \frac{1}{4}x \right]_{2}^{a} = \left( \frac{a^2}{16} - \frac{a}{4} \right) - \left( \frac{1}{4} - \frac{1}{2} \right) = \frac{a^2}{16} - \frac{a}{4} + \frac{1}{4}
\]

\[
\Rightarrow \quad F(a) = \begin{cases} 
\frac{a^2}{16} - \frac{a}{4} + \frac{1}{4} & a < 2 \\
1 & 2 \leq a \leq 6 \\
1 & a > 6
\end{cases}
\]

c) Find \( P(2 < X < 4 \mid X < 4) \) (3 points)

\[
P(2 < X < 4 \mid X < 4) = \frac{P(2 < X < 4)}{P(X < 4)} = \frac{F(4) - F(2)}{F(4)} = \frac{\left( \frac{1}{16} \cdot 4^2 - \frac{1}{4} \cdot 4 + \frac{1}{4} \right) - \left( \frac{1}{16} \cdot 2^2 - \frac{1}{4} \cdot 2 + \frac{1}{4} \right)}{\frac{1}{16} \cdot 4^2 - \frac{1}{4} \cdot 4 + \frac{1}{4}} = \frac{\frac{1}{4} - \frac{1}{4} + \frac{1}{2} - \frac{1}{4}}{\frac{1}{4}} = 1
\]

d) Find \( E[X-2] \) (3 points)

\[
E(X) = \int_{2}^{6} x \cdot f(x) \, dx = \int_{2}^{6} x \left( \frac{x}{8} - \frac{1}{4} \right) \, dx = \int_{2}^{6} \frac{x^2}{8} - \frac{1}{4}x \, dx = \left[ \frac{1}{24}x^3 - \frac{1}{8}x^2 \right]_{2}^{6} = \left( \frac{6^3}{24} - \frac{6^2}{8} \right) - \left( \frac{2^3}{24} - \frac{2^2}{8} \right)
\]

\[
= 4.67
\]

\[
\]
3. The average amount of time Purdue students sleep each night is normally distributed with mean 7 hours and standard deviation 2 hours.

a) What is the probability a student will sleep more than 400 minutes? (3 points)

\[ X = \text{average amount of time Purdue students sleep} \]
\[ X \sim N(\text{mean}=7, \text{SD}=2) \]
\[ P(X > 400) = P\left( Z > \frac{400 - 420}{120} \right) = P(Z > -0.17) = 1 - P(Z < -0.17) = 1 - 0.4325 = 0.5675 \]

b) On average, 20% of students sleep less than how many hours? (3 points)

Let \( m \) be the cutoff

\[ P(X < m) = 0.2 \Rightarrow P\left( Z < \frac{m - 7}{2} \right) = 0.2 \Rightarrow \frac{m - 7}{2} = -0.84 \]
\[ \Rightarrow m = (-0.84) \cdot 2 + 7 = 5.32 \]

For c and d suppose there are 31% of Purdue students sleep more than 8 hours a night. Now there are 200 STAT 225 students. Let \( X \) denote the number of STAT 225 students that sleep more than 8 hours a night.

c) What is the exact probability 60 STAT 225 students sleep more than 8 hours a night? (2 points)

\[ X \sim \text{Bin}(n, p) \implies \text{Bin}(200, 0.31) \]
\[ P(X = 60) = \binom{200}{60} (0.31)^{60} (0.69)^{140} \]

d) State the appropriate approximate distribution of \( X \), and explain why approximation is valid. (4 points)

\[ n = 200, \quad p = 0.31 \]
\[ \text{Since } n \cdot p > 5 \text{ and } n \cdot (1-p) > 5, \quad X \sim N(\text{mean}=n \cdot p, \text{SD}=\sqrt{n \cdot p \cdot (1-p)}) \]
\[ X \sim N(\text{mean}=62, \text{SD}=6.5406) \]

e) What is the approximate probability more than 80 STAT 225 students sleep more than 8 hours a night? (3 points)

\[ P(X > 80) = P(X > 81) = P\left( Z > \frac{80.5 - 62}{6.5406} \right) = P(Z > 2.83) \]
\[ = 1 - P(Z < 2.83) = 1 - 0.9977 = 0.0023 \]
4. Suppose the number of people calling the Advising Office between 8:00 am and 5:00 pm follows a Poisson Process. On average, there are three calls in one hour. Let $Y$ denote the number of calls that the Advising Office receives between 1:00 pm and 4:00 pm.

a) Identify the distribution of $Y$ and its parameter(s). (2 points)

\[
Y \sim \text{POI}(\lambda) = \text{POI}(9)
\]

b) What is the probability that the life time for her new laptop will be between 3 and 5 years? (3 points)

\[
P(Y = 10) = \frac{9^{10} \cdot e^{-9}}{10!} = 0.12
\]

\[
\text{Poisson} = \frac{(\lambda^y)(e^{-\lambda})}{y!}
\]

c) Assume the first call today comes at 8:30 am, what is the probability the second call comes before 8:40 am? (3 points)

\[
T \sim \text{interarrival time} \sim \text{Exp}(\text{mean}=20 \text{ minutes})
\]

\[
P(T < 10) = 1 - e^{-\frac{10}{20}} = 0.39
\]

d) Given there were 7 calls between 8:00 am and 11:00 am, what is the probability 5 calls came between 9:00 am and 11:00 am? (3 points)

\[
X = \# \text{ of calls came between 9:00 am and 11:00 am}
\]

\[
X \sim \text{Bin}(n, p) = \text{Bin}(7, \frac{2}{3})
\]

\[
P(X = 5) = \binom{7}{5} \cdot \left(\frac{2}{3}\right)^5 \cdot \left(1-\frac{2}{3}\right)^2 = 0.31
\]
5. The Statistics Department at Purdue consists of 67 graduate students from many different nations. 32 of them are Chinese, 18 are American, and the rest are of other nationalities. The yearly department picnic is coming up and a 6 member student committee needs to be formed to plan the event.

a) How many different committees are possible? (3 points)

\[
\binom{67}{6} = 99795696
\]

b) What is the probability that all the committee members are from the same group (Chinese, American, Others)? (3 points)

\[
\frac{\binom{32}{6} + \binom{18}{6} + \binom{17}{6}}{\binom{67}{6}} = 0.0094
\]

c) What is the probability that each of the three groups are equally represented on the committee? (3 points)

\[
\frac{\binom{32}{2} \cdot \binom{18}{2} \cdot \binom{17}{2}}{\binom{67}{6}} = 0.10
\]
6. Suppose Vivian only uses one laptop. On average, the lifetime for her laptop is 3 years. Let $X$ denote the life time for her new laptop.

a) Identify the distribution of $X$ and its parameter(s). (2 points)

$$X \sim \text{Exp}(\text{mean}=3) = \text{Exp}(\text{mean}=\lambda)$$

b) What is the probability that the life time for her new laptop will be between 3 and 5 years? (3 points)

$$P[3 < X < 5] = P[X > 3] - P[X > 5] = e^{-\frac{3}{3}} - e^{-\frac{5}{3}} = 0.18$$

c) Given that her laptop has lasted for more than 3 years, what is the probability that the life time for her laptop will be more than 7 years? (3 points)

Lack of Memory Property

$$P[X > 7 | X > 3] = P[X > 4] = e^{-\frac{4}{3}} = 0.26$$

d) Suppose Vivian buys a new Mac and Apple will offer a free replacement for Macs with a lifetime in the lowest 5% of all Mac lifetimes. What is the cutoff point for when Vivian can receive a free replacement if hers breaks? (3 points)

Let $m$ be the cutoff

$$P[X < m] = 0.05$$

$$1 - e^{-\frac{m}{3}} = 0.05 \Rightarrow e^{-\frac{m}{3}} = 0.95$$

$$\Rightarrow -\frac{m}{3} = \ln(0.95)$$

$$\Rightarrow m = (-3) \cdot \ln(0.95) = 0.15 \text{ (Year)}$$
7. A poker hand consists of 5 cards drawn without replacement from a standard deck of 52 cards. A game consists of Mike being dealt a poker hand.

a) What is the probability that Mike will get two Aces during a particular game? (3 points)

\[
\frac{\binom{4}{2} \cdot \binom{48}{3}}{\binom{52}{5}} = 0.04
\]

b) Let X denote the number of games needed until Mike gets the first poker hand where there are two Aces. Identify the distribution of X and its parameter(s). (2 points)

\[ X \sim \text{Geo}(p) \equiv \text{Geo}(0.04) \]

d) What is the probability that it takes Mike between 25 and 27 games (inclusive) to get the first poker hand where there are two Aces? (3 points)

\[
P(X = 25) + P(X = 26) + P(X = 27) = (1 - 0.04)^{24}(0.04) + (1 - 0.04)^{25}(0.04) + (1 - 0.04)^{26}(0.04) = 0.043
\]

e) What is the probability that it takes Mike more than 8 games to get the first poker hand where there are two Aces? (3 points)

\[
P(X > 8) = (1 - 0.04)^8 = 0.72
\]

e) Given that the first poker hand where there are two Aces comes at the 3rd game, in which game can we expect the 5th poker hand where there are two Aces? (3 points)

\[
Y = \# \text{ of poker hands needed until getting } 4 \text{ more two-Ace poker hand}
\]

\[
Y \sim \text{NB}(r, p) \equiv \text{NB}(4, 0.04)
\]

\[
E[Y] = \frac{r}{p} = \frac{4}{0.04} = 100
\]

\[
100 + 3 = 103 \Rightarrow \text{In 103rd poker hand}
\]
8. The number of calories consumed daily by a college student is normally distributed with mean equal to 2650 and standard deviation equal to 500.

a) 95% of students eat between what two calorie amounts? (3 points)

By Empirical Rule

\[ \mu = 2650 \quad \sigma = 500 \]

\[ \mu - 2 \sigma = 2650 - 2 \cdot 500 = 1650 \]

\[ \mu + 2 \sigma = 2650 + 2 \cdot 500 = 3650 \]

\[ \Rightarrow 95\% \text{ of students eat between 1650 and 3650 calories.} \]

For parts b, c, and d; assume 50 college students are randomly sampled.

b) State the approximate distribution of the mean calorie intake and compute the approximate probability the mean calorie intake is above 2725? (4 points)

\[ \bar{X} \sim N(\text{mean}=2650, \text{var} = \frac{500^2}{50}) \Rightarrow N(\text{mean}=2650, \text{var} = 5000) \]

\[ P(\bar{X} > 2725) = P\left( Z > \frac{2725-2650}{\sqrt{5000}} \right) = P(Z > 1.06) \]

\[ = 1 - P(Z < 1.06) \]

\[ = 1 - 0.8554 = 0.1446 \]

c) State the approximate distribution of the total calorie intake and compute the approximate probability the total calorie intake is between 130,000 and 137,000? (4 points)

\[ S = X_1 + X_2 + \ldots + X_{50} \]

\[ S \sim N(\text{mean}=50 \cdot 2650, \text{var} = 50 \cdot (500^2)) \Rightarrow N(\text{mean}=132500, \text{var} = 1350000) \]

\[ P(130000 < S < 137000) = P(-0.71 < Z < 1.27) \]

\[ = P(Z < 1.27) - P(Z < -0.71) = 0.8980 - 0.2389 = 0.6591 \]

d) 16% of college students eat more than 3150 calories per day. Now we sample 50 students randomly.

Let \( \hat{p} \) denote the proportion of these 50 students eat more than 3150 calories per day. State the approximate distribution of \( \hat{p} \), and compute the approximate probability of these 50 students sampled, more than 20% eat more than 3150 calories per day? (4 points)

\[ p = 0.16 \]

\[ \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.16 \cdot 0.84}{50}} = 0.052 \]

\[ P(\hat{p} > 0.2) = P\left( Z > \frac{0.2 - 0.16}{0.052} \right) \]

\[ = P(Z > 0.77) = 1 - P(Z < 0.77) \]

\[ = 1 - 0.7794 = 0.2206 \]
### Areas under the Standard Normal Density: Pr(Z ≤ z)

<table>
<thead>
<tr>
<th>z</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.4</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>-3.2</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>-3.0</td>
<td>0.0010</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
<tr>
<td>-2.9</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>-2.8</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0014</td>
<td>0.0014</td>
</tr>
<tr>
<td>-2.7</td>
<td>0.0022</td>
<td>0.0022</td>
<td>0.0022</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0019</td>
</tr>
<tr>
<td>-2.6</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0026</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0024</td>
<td>0.0024</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.0033</td>
<td>0.0033</td>
<td>0.0033</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0030</td>
<td>0.0029</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0026</td>
</tr>
<tr>
<td>-2.4</td>
<td>0.0040</td>
<td>0.0040</td>
<td>0.0040</td>
<td>0.0039</td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0036</td>
<td>0.0035</td>
<td>0.0034</td>
<td>0.0033</td>
</tr>
<tr>
<td>-2.3</td>
<td>0.0048</td>
<td>0.0048</td>
<td>0.0048</td>
<td>0.0047</td>
<td>0.0046</td>
<td>0.0045</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0042</td>
<td>0.0041</td>
</tr>
</tbody>
</table>

### Pr(Z ≤ z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.

**Symmetry:**

\[ Pr(Z > z) = Pr(Z < -z)\]

\[ = 1 - Pr(Z ≤ z) \]

**Usage:** To find Pr(Z ≤ 1.26) look in row 1.2 under the column 0.06: Pr(Z ≤ 1.26) = 0.8962.

---

### Example

- **Pr(Z ≤ 0.8):**
  - From the table, Pr(Z ≤ 0.8) = 0.7881.

---

**Normal-Dist.tex, Johnson (2004)