Midterm Practice Questions

1. For the following set of 15 data values,

   \[-11, -4, 0, 1, 1, 2, 2, 3, 4, 5, 6, 7, 11, 13, 20,\]

   (a) give the mean, median, standard deviation and IQR and draw a modified boxplot.

   (b) Argue in a sentence whether or not one of the measures of variation, standard
   variation or IQR, is better for these data.

2. In a study of 10000 elderly patients (3000 African American, 7000 white Americans), the
five-year survival rate for lung cancer (the proportion surviving five years after detection
of lung cancer) is 34.1\% for whites and 26.4\% for blacks.

   (a) What is the five-year survival rate overall?

   (b) Given that a patient has survived at least five years, what is the probability that
the patient is black?

3. Assume the five-year survival rate for individuals having surgery for non-small-cell lung
cancer detected early is 40\%.

   (a) Nine patients in a clinic have surgery for the condition described above. What is
the probability that 5 or more survive?

   (b) If you do not have surgery, the five-year survival probability is 0.04. What is the
probability that none of the nine patients survived five years without surgery?

4. In the country of Gatesland, it is known whether or not a citizen has a home computer
and is totally prepared for the year 2000 date change (called the Y2K problem) that
has been in the news (\textit{three years ago}). The probabilities of randomly selecting a citizen
with certain characteristics are given below:

   \[
   \begin{align*}
   P(\text{owns computer and is prepared for year 2000}) &= 0.39 \\
   P(\text{owns computer and is not prepared for year 2000}) &= 0.26 \\
   P(\text{does not own computer and is prepared for year 2000}) &= 0.28 \\
   P(\text{does not own computer and is not prepared for year 2000}) &= 0.07 \\
   \end{align*}
   \]

   (a) What is the probability that a randomly selected citizen owns a home computer?

   (b) What is the probability that a randomly selected citizen who owns a computer is
totally prepared for the year 2000 date change?

   (c) Is owning a home computer independent of being totally prepared for the year 2000
date change? Why or why not?
5. Short Answer:

(a) Suppose that you have a fair coin. The probability on each flip of getting a head \((H)\) equals the probability of getting a tail \((T)\). The outcomes on successive flips are independent. Which of the following sequences is most likely to occur? Which is the least likely?

i. \(HHHHT\)  
ii. \(HHTHT\)  
iii. \(THHHH\)

(b) Describe two random variables, one that has a skewed probability distribution and one that is symmetric. For each distribution, is the mean equal to, higher than, or less than the median? You may describe your random variable and its probability distribution with a well labeled picture or table.

6. The table below reports the number of traffic accidents in which randomly sampled bus drivers for public corporations were involved over a 4-year period. There were a total of 708 bus drivers in the sample.

<table>
<thead>
<tr>
<th>Number of accidents</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>117</td>
</tr>
<tr>
<td>1</td>
<td>157</td>
</tr>
<tr>
<td>2</td>
<td>158</td>
</tr>
<tr>
<td>3</td>
<td>115</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>44</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) For these data, which is higher, the mean or the median? How do you know?

(b) Compute the upper and lower quartiles for these data.

(c) Suppose a 709th driver is randomly selected. If you were asked to evaluate the chance that this driver would be involved in more than 4 accidents in the next four years, what probability would you get?

(d) Suppose the CTA hires a new driver with no prior experience. If you were asked to evaluate the chance that this driver would be involved in more than 4 accidents in the next four years, would your probability be more or less than our answer in the previous part? Why?
7. Decide if the following are true, false, or unknown from the given information. Explain why. Unless stated otherwise, assume the information from each part does not carry to the next part.

(a) \( P(A) = 0.53 \) and \( P(A \cap B) = 0.63 \).
(b) If \( P(G) = 0.4 \) and \( P(G \cap H) = 0.3 \), then \( P(G|H) = 0.75 \).
(c) If events \( A \) and \( B \) are mutually exclusive, and \( P(A) = \frac{1}{4} \), \( P(B) = \frac{1}{2} \), then \( P(A \cap B) = \frac{1}{4} \).
(d) Two mutually exclusive events are independent.
(e) The following is a table of the joint distribution of \( X \) and \( Y \):

<table>
<thead>
<tr>
<th>( x )( \backslash) ( y )</th>
<th>-1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Total</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\( \text{cov}(X, Y) = 0.1 \).

(f) The variables \( X \) and \( Y \) in part (d) are independent.
(g) The variable \( R \) with frequency function below has mean \( \mu_R = -0.5 \).

\[
\begin{array}{c|ccc}
 r & -1 & 0 & 1 \\
 f(r) & 0.8 & -0.1 & 0.3 \\
\end{array}
\]

8. Consider a random variable \( Y \) whose probability density function is given by

\[
f(y) = cy^2 + 1, \quad 0 \leq y \leq 2,
\]

and 0 otherwise, where \( c \) is an unknown constant.

(a) Find \( c \) so that the above is a genuine probability density function.
9. SAT math scores are approximately normally distributed around a mean of 550 with a standard deviation of 100.

(a) What is the probability that an individual gets a score of at least 700?
(b) What is the probability that an individual scores between 520 and 600?
(c) If a person scored in the 97th percentile (in the top 3% of those taking the test), what was their score?
(d) What is the probability that exactly 2 out of any 5 students chosen at random will score above 700? (Assume the individual scores are independent.)

10. The time it takes a University of Chicago student to get to campus can be regarded as a normal random variable with mean of 17 minutes and variance 16 minutes.

(a) What is the probability that it will take a randomly chosen UC student more than twenty minutes to get to campus?
(b) What proportion of students arrive on campus in between 9 and 25 minutes?
(c) Let X = the number of students, out of 100 randomly chosen students, who will take more than 20 minutes to get to campus. What is the distribution of X?
(d) Express $P(25 \leq X < 28)$ in terms of the values of a distribution function. For two bonus points, compute the exact probability.
(e) Use a normal approximation to find the probability in part (d).
(f) For some number $T$, 5% of students will take less than $T$ minutes to get to school. Find $T$. 