Name: ______________________________

True/False

For each of the following questions indicate true or false, then **explain** your answer. You may use examples to illustrate your answer.

(I) The minimum possible value that Y can take on where Y is a binomial random variable is the value 0.

- TRUE. For any value of n, the smallest number of sucecsses that a binomial random variable can take on is zero (or no successes).

(II) For two events A and B (where $Pr\{A\} > 0, Pr\{B\} > 0$), the following equation holds true: $Pr\{A|B\} = Pr\{B|A\}$.

- FALSE. This is not, in general, true. This is because when we switch the event after the conditional, we are changing what subset of the possible events we are looking at.
  Or
  $Pr\{B|A\} = \frac{Pr\{A \text{ and } B\}}{Pr\{A\}}$, which does not equal $\frac{Pr\{A \text{ and } B\}}{Pr\{B\}} = Pr\{A|B\}$ (unless $Pr\{A\} = Pr\{B\}$)

(III) For a sample dataset, it is not possible for the first quartile to equal the median.

- FALSE. If, for example, the entire dataset was the same value repeated, the median and first quartile could equal eachother.

(IV) For a normal random variable Y with mean μ_Y and standard deviation σ_Y, $Pr\{Y > \mu_Y\} = Pr\{Y < \mu_Y\}$.

- TRUE. For a normal random variable, the mean is the median, so 50% lies above it and 50% below.

Name: ____________________

Full Detail

Work out the following problems. **Show your work.**

1. Body mass index (BMI) is a measure of how fit a person is, and is given in units of kg/m^2. BMI for all people in a certain state in the US is believed to be normally distributed, with mean 25.44, and standard deviation 4.88. Use this information to complete the following problems.

   (a) Find the probability that a randomly selected subject has a BMI over 23.

   - $Pr\{Y > 23\} = Pr\{Z > -0.5\} = 1 - Pr\{Z \leq -0.5\} = 1 - 0.3085 = 0.6915$

   (b) If a randomly selected subject has a BMI under 25.44, what is the probability that their BMI is over 20?

   - $Pr\{Y > 20|Y < 25.44\} = \frac{Pr\{(Y>20) \text{ and } (Y<25.44)\}}{Pr\{Y<25.44\}} = \frac{Pr\{20<Y<25.44\}}{Pr\{Y<25.44\}} = \frac{P\{-1.11<Z<0\}}{Pr\{Z<0\}} = \frac{Pr\{Z\leq 0\}-Pr\{Z\leq -1.11\}}{0.5} = \frac{0.5-0.1335}{0.5} = 0.733$

   (c) Suppose subjects in the top 25% of the distribution are considered to be "overweight". What is the cutoff for a subject being considered "overweight"?

   - We would be looking for the 75^{th} percentile, and the 75^{th} percentile (Q_3) for a z-score is 0.67. The corresponding BMI is: $Y^{75} = Z^{75}(\sigma) + \mu = 28.7096$.

   (d) If we randomly sample six subjects, what is the probability that exactly four of them are "overweight"?

   - Let Y = the number of subjects out of 6 which are overweight. Then, Y is binomial, n= 6, and $p = 0.25$. Thus, $Pr\{Y = 4\} = \binom{n}{4}p^4(1-p)^{(n-4)} = \binom{6}{4}(0.25)^4(1-0.25)^{(6-4)} = 15(0.25)^4(0.75)^{(2)} = 0.033$

Name: ____________________

2. For a certain group of people, the probability they have a specific gene is 0.50. If they have this gene, the probability of developing kidney cancer is 0.30. If they do not have this gene, the probability of developing kidney cancer is 0.016. Assume a person was randomly selected from this group.

(a) Find the probability that they develop kidney cancer.

- Let C = Cancer, G = Gene. Then,
$Pr\{C\} = Pr\{C \text{ and } G\} + Pr\{C \text{ and } G^C\} = Pr\{C|G\}Pr\{G\} + Pr\{C|G^C\}Pr\{G^C\}$
$= (0.3)(0.5) + (0.016)(0.5) = 0.15 + 0.008 = 0.158$

(b) Find the probability that if they developed kidney cancer, they have the gene.

- $Pr\{G|C\} = \frac{Pr\{G \text{ and } C\}}{Pr\{C\}} = \frac{Pr\{C|G\}Pr\{G\}}{Pr\{C\}} = \frac{(0.3)(0.5)}{0.158} = \frac{0.15}{0.158} = 0.9494$

(c) Find the probability that they either develop kidney cancer, had the gene, or both (the union).

- $Pr\{C \text{ or } G\} = Pr\{G\} + Pr\{C\} - Pr\{G \text{ and } C\} = 0.158 + 0.5 - 0.15 = 0.508$

(d) Find the probability that if they do not develop kidney cancer, they have the gene.

- $Pr\{G|C^C\} = \frac{Pr\{C^C \text{ and } G\}}{Pr\{C^C\}} = \frac{Pr\{G\} - Pr\{C \text{ and } G\}}{1 - Pr\{C\}} = \frac{0.5 - 0.15}{1 - 0.158} = 0.4157$

Name: ____________________

3. Suppose a random sample of the starting salaries (in thousands of dollars) for 9 veterinarians that have recently completed veterinarian school is:

$$42, 59, 67, 68, 72, 79, 81, 82, 112$$

(a) Calculate the first quartile, the third quartile, and the median.

- Since $0.50(n+1) = 0.50(9+1) = 5$, we average the $(5)^{th}$ and $(5)^{th}$. Thus, the median is equal to: 72.
  Since $0.25(n+1) = 0.25(9+1) = 2.5$, we average the $(2)^{th}$ and $(3)^{th}$. Thus, Q_1 is equal to: 63.
  Since $0.75(n+1) = 0.50(9+1) = 7.5$, we average the $(7)^{th}$ and $(8)^{th}$. Thus, Q_3 is equal to: 81.5.

(b) Identify any outliers, being sure to show all your work.

- The cutoff value is:
  lower cutoff: $Q_1 - 1.5(Q_3 - Q_1) = 63 - 1.5 * (81.5 - 63) = 35.25$
  The cutoff value is:
  upper cutoff: $Q_3 + 1.5(Q_3 - Q_1) = 81.5 + 1.5 * (81.5 - 63) = 109.25$
  Thus, the outliers are: 112

(c) Find the mean. Would you suggest using the mean or median to describe the center of the data? Explain.

- The mean of the data set is : $\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{9}(662) = 73.5556$.
  Since there is an outlier, we should use the median rather than the mean.

(d) If the sample standard deviation is $s = 19.10$, what proportion of the data lies within 1 sample standard deviation from the sample mean?

- The interval which represents one standard deviation from the mean is: (54.4556, 92.6556), so that the proportion is 7/9 or 0.7778.