True/False

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

(I) You can identify range of the upper 50% of the data from a histogram.

FALSE. This is true for a boxplot, since it gives the values of the median, but the median is not represented on a histogram in general. Further, what is represented on histograms are intervals of the data, but not the exact value of the data.

(II) For a continuous random variable, if the mean is larger than the median, you may conclude that the probability that a value is less than the mean is over 50%.

TRUE. The median represents the center of the data, and so a value being above the median implies that more than 50% of the data lies below that value.

(III) For two events A and B, where $Pr\{A\} > 0$ and $Pr\{B\} > 0$, $Pr\{A|B\} + Pr\{B|A\}$ must equal 1.

FALSE. This is because the first probability is conditioning on event A, and the second on event B. Since they are using different subsets of all possible outcomes, this equation does not hold.

(IV) For a normal random variable, the Z score of a value X that is below the mean will always be negative.

TRUE. The numerator of the Z-score will then be negative $(Y - \mu_Y)$, and the denominator is always positive (σ_Y).

Full Detail

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

1. A random sample of 45 children (and their parents) asked "How many hours do you spend Trick or Treating for Halloween?". Let Y_i = the number of hours the children (and their parents) spent Trick or Treating. The responses are below, with Y_i and the frequency of Y_i:

y_i	2	4	5	6	10
Frequency of y_i	15	12	13	3	2

For example, 15 children (and their families) spend 2 hours Trick or Treating, 12 spent 4 hours, etc.

(a) Find the average number of hours spend trick or treating.

The mean may be calculated as:
$\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{45}(2*15+4*12+5*13+6*3+10*2) = \frac{1}{45}(181) = 4.0222$

(b) Identify any outliers, being sure to show your work and completely justify your answer.

Since $0.25(n+1) = 0.25(45+1) = 11.5$, we average the $(11)^{th}$ and $(12)^{th}$. Thus, Q_1 is equal to: 2.
Since $0.75(n+1) = 0.50(45+1) = 34.5$, we average the $(34)^{th}$ and $(35)^{th}$. Thus, Q_3 is equal to: 5
The cutoff value is:
lower cutoff: $Q_1 - 1.5(Q_3 - Q_1) = 2 - 1.5*(5-2) = -2.5$
The cutoff value is:
upper cutoff: $Q_3 + 1.5(Q_3 - Q_1) = 5 + 1.5*(5-2) = 9.5$
Thus, the value/s which are outliers are: 10, 10

(c) If the standard deviation for this data set is 1.8888, interpret the standard deviation in terms of the problem.

A typical deviation or distance from the average number of hours of 4.0222 to spend Trick or Treating is 1.8888 hours.

(d) Estimate the probability that a childrens' family falls **within** one standard deviation from the mean.

$(\bar{y} - s, \bar{y} + s) = (2.1334, 5.911)$, so that we have 25 observations that are within this interval (the 4 hours and the 5 hours), which would be an estimated probability of $\frac{25}{45} = 0.5555556$.

2. Suppose that 40% of students go to sleep "on time". If a student went to sleep "on time", the probability that they felt "well rested" was 0.80. Further, the probability that a student both did not go to sleep "on time" and felt "well rested" was 0.20.

(a) Find the probability that a student felt "well rested".

$$Pr\{W\} = Pr\{W \text{ and } O\} + Pr\{W \text{ and } O^C\} = Pr\{W|O\}Pr\{O\} + 0.2 = 0.8(0.4) + 0.2 = 0.32 + 0.2 = 0.52$$

(b) Find the probability that a student did not go to sleep "on time", or did feel "well rested", or both.

$$Pr\{O^C \text{ or } W\} = Pr\{O^C\} + Pr\{W\} - Pr\{O^C \text{ and } W\} = (1 - 0.4) + 0.52 - 0.2 = 0.92$$

(c) A fellow student tells you they feel "well rested". What is the probability the fellow student went to sleep "on time"?

$$Pr\{O|W\} = \frac{Pr\{O \text{ and } W\}}{Pr\{W\}} = \frac{Pr\{W|O\}Pr\{O\}}{Pr\{W\}} = \frac{0.8(0.4)}{0.52} = \frac{0.32}{0.52} = 0.6153846$$

(d) What is the probability a student both does not feel "well rested", and did go to sleep "on time"?

$$Pr\{W^C \text{ and } O\} = Pr\{W^C|O\}Pr\{O\} = (1 - Pr\{W|O\})Pr\{O\} = (1 - 0.8)(0.4) = 0.08$$

Initial: ____________

3. Suppose the length a cat takes a nap (the length of a "cat nap") has a population mean of 3 hours, with population standard deviation 0.75 hours. Assume the length of a cat nap is normally distributed.

   (a) Find the probability that a randomly selected cat takes a nap over 4 hours in length.

   $Pr\{Y > 4\} = Pr\{Z > 1.33\} = 1 - Pr\{Z \leq 1.33\} = 1 - 0.9082 = 0.0918$

   (b) If a cat has been napping for 2 hours, what is the probability that they nap at most an additional 1.5 hours?

   $Pr\{Y < 3.5|Y > 2\} = \frac{Pr\{(Y<3.5) \text{ and } (Y>2)\}}{Pr\{Y>2\}} = \frac{Pr\{2<Y<3.5\}}{Pr\{Y>2\}} = \frac{P\{-1.33<Z<0.67\}}{1-Pr\{Z<-1.33\}} = \frac{Pr\{Z\leq 0.67\}-Pr\{Z\leq -1.33\}}{0.0918} = \frac{0.7486-0.0918}{0.9082} = 0.7232$

   (c) A cat sleeps 4.35 hours. What percentile does this value represent?

   Since $Pr\{Y < 4.35\} = Pr\{Z < 1.8\} = 0.9641$
   It represents the 96.41^{th} percentile.

   (d) Suppose we sample 1000 cats from this population. What is the expected number of cats who sleep under four hours, and the standard deviation of the number of cats who sleep under four hours (out of the 1000)? You may assume all cats are independent.

   This is a binomial distribution, with $n = 1000$, and $p = 1 - 0.0918 = 0.9082$.
   Thus, $\mu_Y = np = 1000(0.9082) = 908.2$, and $\sigma_Y = \sqrt{np(1-p)} = \sqrt{1000(0.9082)(1-0.9082)} = \sqrt{83.37276} = 9.1309$

Scratch Work