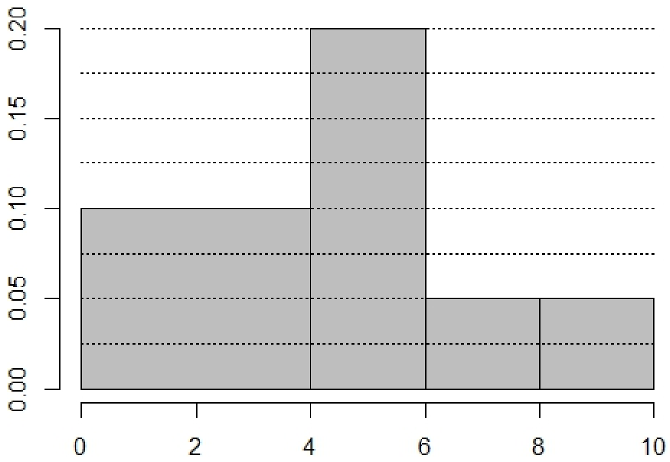
Exam 1
STAT 3010 Spring 2016

Name (Print): ____________________

- **It is a closed-book and in-class exam.**
- **Time: 75 minutes. (2:00 pm—3:15 pm, Tuesday, Februray, 9, 2016)**
- **Show your work to receive full credits. Total points: 65pts +5pts(Bonus)**
- **Basic scientific calculator is allowed, but no laptop/tablet/smartphone or equivalent.**

1	2	3	4	5	6 (part1-4)	6(part5) (Bonus)	Total

Problem 1 (6 pts): The **density** histogram given below is based on 100 measurements.



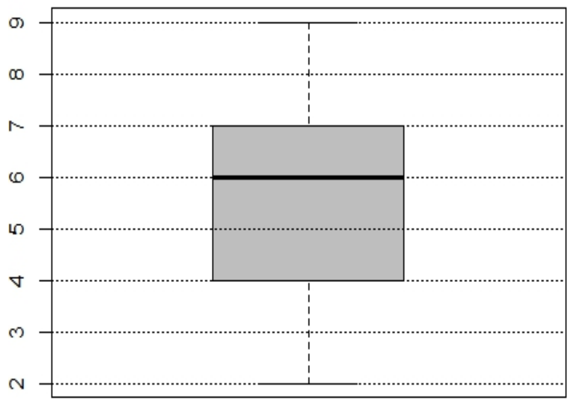
1. (3 pts) What proportion of the data is between 4 and 8?

   Solution: $2(0.2) + 2(0.05) = 0.5$

2. (3 pts) How many measurements are less than 4?

   Solution: $100(4)(0.1) = 40$

Problem 2 (9 pts): The boxplot given below is based on 100 measurements. Fill in the blanks.



1. (3 pts) About 25% of the measurements are greater than 7.

2. (3 pts) About 25 measurements in the data are less than 4.

3. (3 pts) The interquartile range is 3.

Problem 3 (9 pts): Consider the following data: 4, 8, 15, 16, 23, 42, 50, 1668

1. (3 pts) Find the median.

   Solution: $n = 8$, median$=\dfrac{X_{(4)} + X_{(5)}}{2} = \dfrac{1}{2}(16 + 23) = 19.5$

2. (3 pts) Which summary statistics is more appropriate to measure of central tendency of this data? **Mean or median?** Give your reason.

   Solution: Median is more approriate. We have an outlier (1668) in our data. Median is a robust measure of the center of the data, which will not be affected by the outlier. Mean is very sensitive to the outlier.

3. (3 pts) Find the first quartile (Q1).

   Solution: $(n+1)p = (9)(0.25) = \dfrac{9}{4}$, $Q1 = (1 - \dfrac{1}{4})X_{(2)} + \dfrac{1}{4}X_{(3)} = \dfrac{3}{4}(8) + \dfrac{1}{4}(15) =$ 9.75

Problem 4 (15 pts): All the fourth-graders in a certain elementary school took a standardized test. A total of 85% of the students were found to be proficient in reading, 78% were found to be proficient in mathematics, and 65% were found to be proficient in both reading and mathematics. A student is chosen at random.

Let R be the event that a student is proficient in reading, and let M be the event that a student is proficient in mathematics.

1. (5 pts) What is the probability that the student is proficient in mathematics but not in reading? (Hint: You need to find $P(R^c \cap M)$, Venn diagram might be helpful.)

   Solution: $P(M) = P(R \cap M) + P(R^c \cap M)$

   $P(R^c \cap M) = P(M) - P(R \cap M) = 0.78 - 0.65 = 0.13$

2. (5 pts) What is the probability that the student is proficient in reading but not in mathematics? (Hint: You need to find $P(R \cap M^c)$, Venn diagram might be helpful.)

   Solution: $P(R) = P(R \cap M) + P(R \cap M^c)$

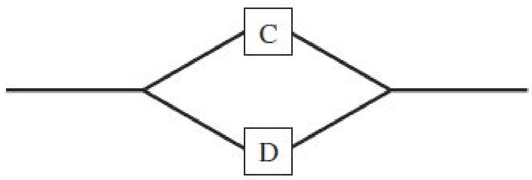
   $P(R \cap M^c) = P(R) - P(R \cap M) = 0.85 - 0.65 = 0.2$

3. (5 pts) What is the probability that the student is proficient in neither reading nor mathematics? (Hint: You need to find $P(R^c \cap M^c)$, Venn diagram might be helpful.)

   Solution: $P(R \cup M) = P(R) + P(M) - P(R \cap M) = 0.85 + 0.78 - 0.65 = 0.98$

   $P(R^c \cap M^c) = 1 - P(R \cup M) = 1 - 0.98 = 0.02$

Problem 5 (6 pts): A system contains two components, C and D, connected in parallel as shown in the diagram.



Assume C and D function independently. For the system to function, either C or D must function.

Let C denote the event that component C functions, and let D denote the event that component D functions. We know the probability that C fails is 0.08 and the probability that D fails is 0.12

Find the probability that the system functions.

Solution: See homework 2, problem 13.

Problem 6: The proportion of people in a given community who have a certain disease is 0.005. A test is available to diagnose the disease. If a person has the disease, the probability that the test will produce a positive signal is 0.99. If a person does not have the disease, the probability that the test will produce a positive signal is 0.01.

Let D represent the event that the person actually has the disease, let $+$ represent the event that the test gives a positive signal and let $-$ represent the event that the test gives a negative signal.

We are given the following probabilities:

$P(D) = 0.005$, $P(+|D) = 0.99$ and $P(+|D^c) = 0.01$. (Hint: Draw a tree diagram might be helpful.)

1. (6 pts) Find the following probability: $P(D^c)$, $P(-|D)$ and $P(-|D^c)$.

Solution:

$P(D^c) = 1 - P(D) = 1 - 0.05 = 0.995$

$P(-|D) = 1 - P(+|D) = 1 - 0.99 = 0.01$

$P(-|D^c) = 1 - P(+|D^c) = 1 - 0.01 = 0.99$

2. (5 pts) Find $P(-)$.

   Solution:

$$\begin{aligned}
P(-) &= P(- \cap D) + P(- \cap D^c) \\
&= P(-|D)P(D) + P(-|D^c)P(D^c) \\
&= (0.01)(0.005) + (0.99)(0.995) \\
&= 0.9851
\end{aligned}$$

3. (5 pts) If a man tests negative, what is the probability that he actually has the disease? Hint: Find $P(D|-)$

   Solution:

$$\begin{aligned}
P(D|-) &= \frac{P(D \cap -)}{P(-)} \\
&= \frac{P(-|D)P(D)}{P(-|D)P(D) + P(-|D^c)P(D^c)} \\
&= \frac{(0.01)(0.005)}{0.9851} \\
&= 5.08 \times 10^{-5}
\end{aligned}$$

4. (4 pts) For many medical tests, it is standard procedure to repeat the test when a positive signal is given.

   If repeated tests are independent, what is the probability that a man will test positive on two successive tests if he has the disease? what is the probability that a man tests positive on two successive tests if he does not have the disease? Hint: Find $P(both + |D)$ and $P(both + |D^c)$.

   Solution:

   $P(both + |D) = 0.99^2 = 0.9801$

   $P(both + |D^c) = 0.01^2 = 0.0001$

5. (Bonus 5 pts) If a man tests positive on two successive tests, what is the probability that he has the disease? Hint: Find $P(D|both+)$.

Solution:

$$\begin{aligned} P(D|both+) &= \frac{P(D \cap both+)}{P(both+)} \\ &= \frac{P(both+|D)P(D)}{P(both+|D)P(D) + P(both+|D^c)P(D^c)} \\ &= \frac{(0.9801)(0.005)}{(0.9801)(0.005) + (0.0001)(0.995)} \\ &= 0.9801 \end{aligned}$$