HW1 Key

Name: Will Schievelbein
Each question is worth 1 point. No partial credit on the questions on this HW, including propagation of mistakes.

Q1A: D and D' are discrete variables. This is because the points in between outcomes is not possible, one cannot roll a 3.4258.

Q1B:

	Expected Value (E)	Variance (Var)
D	3.5	2.917
D'	5.4	1.84

Q1C: E(D') is larger because the dice is weighted to roll the largest number more often.

```
#1 Working with Distributions

#In R, hashtags represent comments and will not run. They are simply notes

#We create both the outcomes and probabilities as vectors. A vector is simply and ordered set of
numbers.
Outcomes <- c(1,2,3,4,5,6)
prob1 <- c(1/6,1/6,1/6,1/6,1/6,1/6)
prob2 <- c(.04,.04,.04,.04,.04,.8)

#Find expected value by multiplying outcomes by probabilities and sum.
E1 <- sum(Outcomes*prob1)
E2 <- sum(Outcomes*prob2)

#Find variance summing the square difference of the outcomes and the expected value, weighted by
probability
#if variance is confusing, google is your friend.
V1 <- sum( ((Outcomes-E1)^2) * prob1)
V2 <- sum( ((Outcomes-E2)^2) * prob2)
```

Q2: To work through these, please first look at what are known:

$$P(M) = .5; P(B) = .3; P(B|M) = .4$$

This reads: The probability of Male is .5, the probability of supporting the ballot is .3, and the probability of supporting the ballot given male is .4.

Q2A: The proportion of residents that are both male and support the ballot is .2

$$P(M \cap B) = P(B|M)P(M) = .4 * .5$$

Q2B: The proportion of residents that are either male or support the ballot is .6

$$P(M \cup B) = P(M) + P(B) - P(M \cap B) = .5 + .3 - .2$$

Q2C: The proportion of ballot supporter that are male is .667

$$P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{.2}{.3}$$

Q3A: The population is the mall as this is the pool of possible people that the sample is drawn. The sample is the 100 selected people.

Q3b: The null hypothesis is that the population mean score for Texacola is the same as that for Oklepsi. The alternative hypothesis is that population mean score for Texacola is the greater than that for Oklepsi. Note that the test is designed so that we are "trying to disprove the null".

Q3C: The low p-value signifies that it is a statistically unlikely that the mean score for Texacola is less than or equal to that of Oklepsi.

Q3D: The p-value would be lower as the higher sample size (n) reduces the variance of the estimator (β).

Q4A: The population in this example is the students at the school. The sample is the randomly selected 25 students.

Q4B: The 95% confidence interval represents an interval in which the procedure has a 95% chance of containing the true mean.

Q4C: This question is open to interpretation, but there is a sound possibility that the true average score of the school is under the state mandated proficiency of 75.

Q4D: The 99% confidence interval will have the same center (the sample mean), but a wider width.