STAT 477/STAT 577
HW 1 Solutions - 100 points

1. (30 pts; 10 pts for each variable) In lecture, we discussed a course survey that was given to students enrolled in STAT 101 during the Fall 2014 semester. Links to the survey and the data collected can be found in the **OpeningSurvey.pdf** and **OpeningSurveyData.csv** files in Canvas. Select three categorical variables from the survey and use R to summarize each variable with its summary table and bar graph.

   Your answers will vary depending on which three variables you selected. Possible options include: CellPhone, EyeColor, ColorBlind, Gender, HairColor, HairLength, Wet, Lather, Rinse, YearinSchool, IdealEyeColor, IdealHairColor, OffCampusWork, OnCampusWork, FromIowa?, FromUS?, Living

   Here is an example of the code and output for the variable HairLength.

   Read in the data:

```
surveydata <- read.csv(file.choose(), header = T)
```

   Set levels for the Hair Length variable:

```
surveydata$HairLength<- factor(surveydata$HairLength,
                              levels = c("Short", "Medium", "Long"))
```

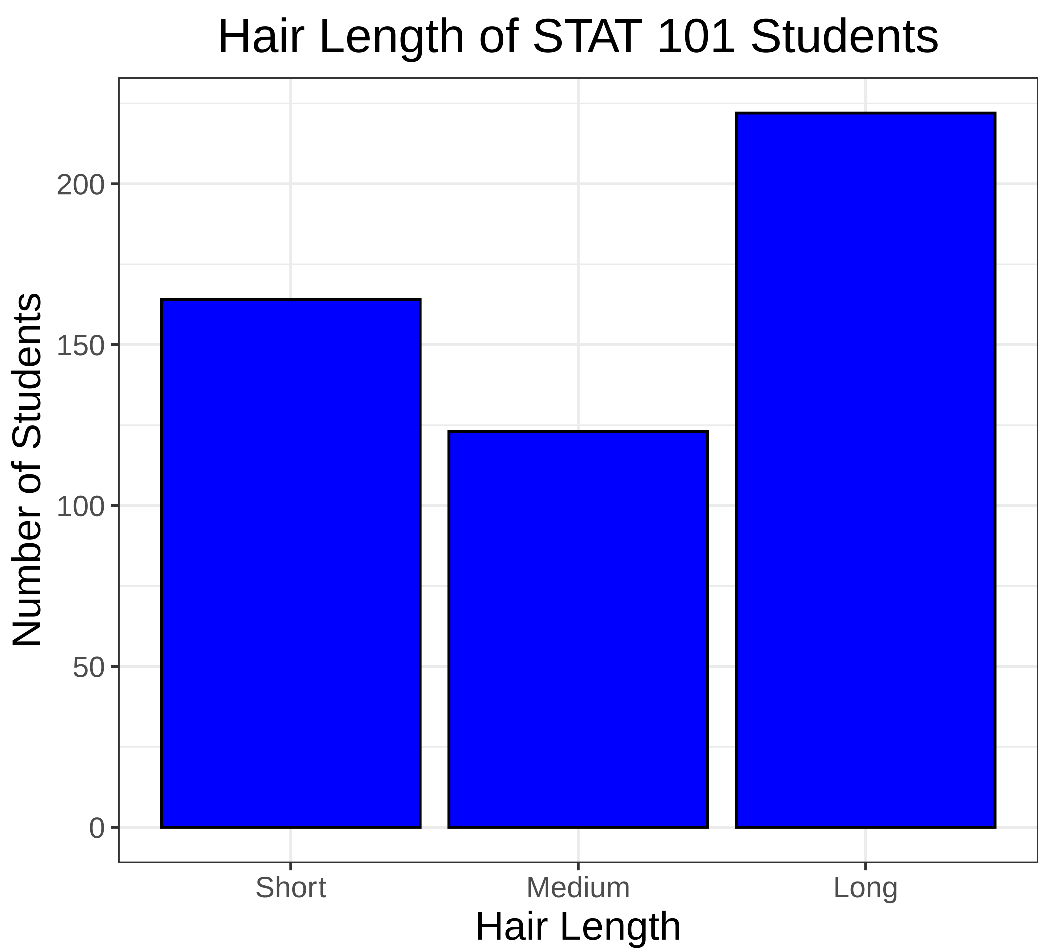
   Make summary table for Hair Length variable:

```
hair.counts<- count(surveydata, var = 'HairLength')
hair.table<- mutate(hair.counts,
                    prop = freq/sum(hair.counts[2]))
hair.table<- rbind(hair.table, data.frame(HairLength='Total',
                                          t(colSums(hair.table[, -1]))))
hair.table

##   HairLength freq      prop
## 1      Short  164 0.3222004
## 2     Medium  123 0.2416503
## 3       Long  222 0.4361493
## 4      Total  509 1.0000000
```

   Make the bar graph for the Hair Length variable:

```
ggplot(surveydata, aes(x=HairLength))+
  geom_bar(fill = "blue", colour = "black")+
  labs(x = "Hair Length",
```

```
      y = "Number of Students",
      title = "Hair Length of STAT 101 Students")+
   theme_bw()+
   theme(axis.title.y = element_text(size = rel(1.4)),
         axis.title.x = element_text(size = rel(1.4)),
         axis.text.x = element_text(size = rel(1.2)),
         axis.text.y = element_text(size = rel(1.2)),
         plot.title = element_text(hjust=0.5, size = rel(1.6)))
```



2. (23 pts) A certain genetic mutation occurs in a population with probability 0.05. A researcher has genetic material from 40 unrelated members of this population and tests for the mutation.

   (a) (2 pts) The number of people in a sample of 40 unrelated members of this population with this genetic mutation has a binomial distribution. What are the values of the parameters for this binomial distribution (n and p)?

   $n = 40$ and $p = 0.05$.

   (b) (4 pts) Use R to calculate the probability that at least 1 person in a sample of 40 unrelated members of this population will have the genetic mutation.

Find $P(Y \geq 1)$. Two equivalent ways to do this in R are:

```
1 - dbinom(0, 40, 0.05)

## [1] 0.8714878

sum(dbinom(1:40, 40, 0.05))

## [1] 0.8714878
```

(c) (3 pts) Use R to calculate the probability that no more than 3 people in a sample of 40 unrelated members of this population will have the genetic mutation.
Find $P(Y \leq 3)$.

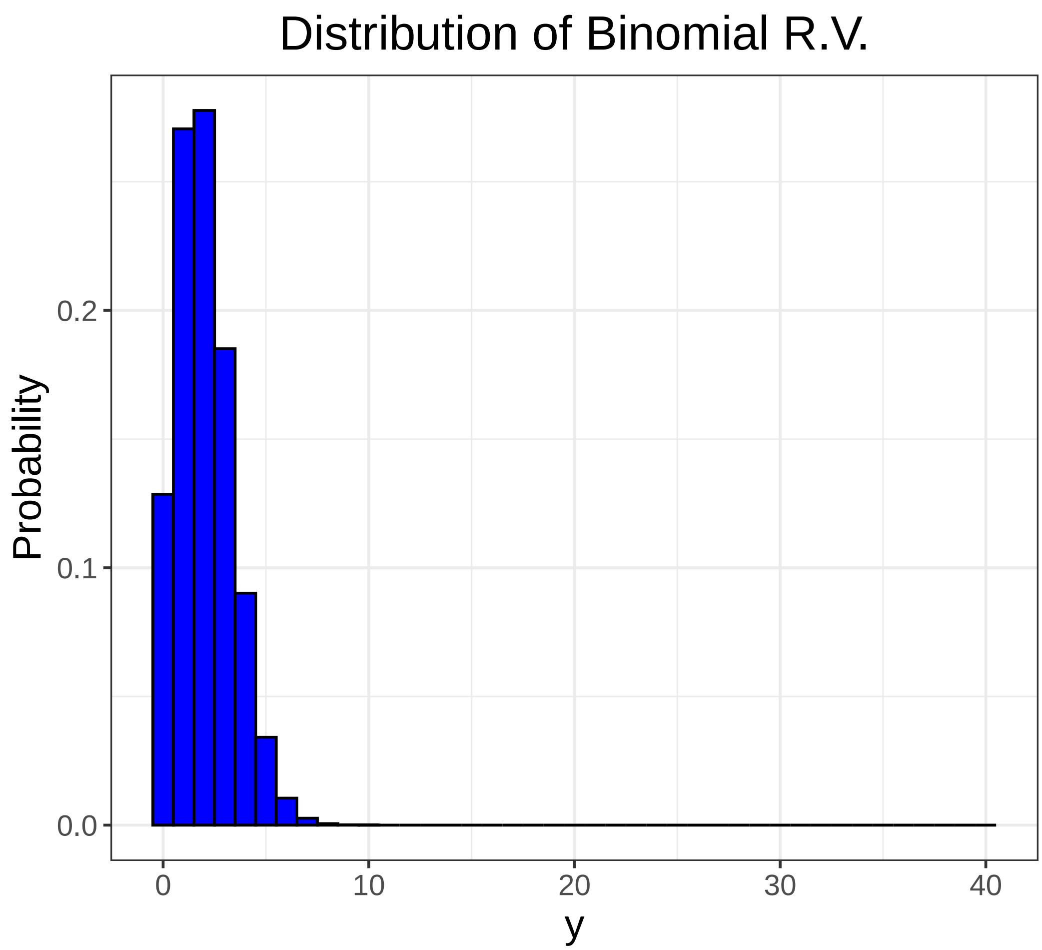
```
sum(dbinom(0:3, 40, 0.05))

## [1] 0.8618502
```

(d) (3 pts) What is the mean number of people with the genetic mutation in a sample of 40 unrelated members of this population?
The expected value is $E(Y) = np = 40(0.05) = 2$.

(e) (5 pts; 3 pts for variance and 2 pts for std. dev.) What is the variance and standard deviation of the number of people with the genetic mutation in a sample of 40 unrelated members of this population?
The variance is $V(Y) = np(1-p) = 40(0.05)(0.95) = 1.9$.
The standard deviation is $\sqrt{1.9} = 1.3784$.

(f) (6 pts; 4 pts for graph and 2 pts for description) Use R to produce a graph of the distribution of the number of people with the genetic mutation in a sample of 40 unrelated members of this population. Describe the shape of the distribution.

```
plot.binom(40, 0.05)
```



This distribution is unimodal and skewed right.

3. (23 pts) Cocker spaniels (a breed of dog) are susceptible to anemia. Suppose that 30% of the population of seven year old cocker spaniels have anemia.

(a) (2 pts) The number of cocker spaniels with anemia in a sample of 40 dogs from this population has a binomial distribution. What are the values of the parameters for this binomial distribution (n and p)?

$n = 40$ and $p = 0.3$.

(b) (4 pts) Use R to calculate the probability that at least 13 of the dogs in a sample of 40 dogs from this population will have anemia.

Find $P(Y \geq 13)$.

```
sum(dbinom(13:40, 40, 0.3))

## [1] 0.4228191
```

(c) (3 pts) Use R to calculate the probability that no more than 8 dogs in a sample of 40 dogs from this population will have anemia.

Find $P(Y \leq 8)$.

```r
sum(dbinom(0:8, 40, 0.3))
```

```
## [1] 0.1110092
```

(d) (3 pts) What is the mean number of dogs with anemia in a sample of 40 dogs from this population?

The expected value is $E(Y) = np = 40(0.3) = 12$

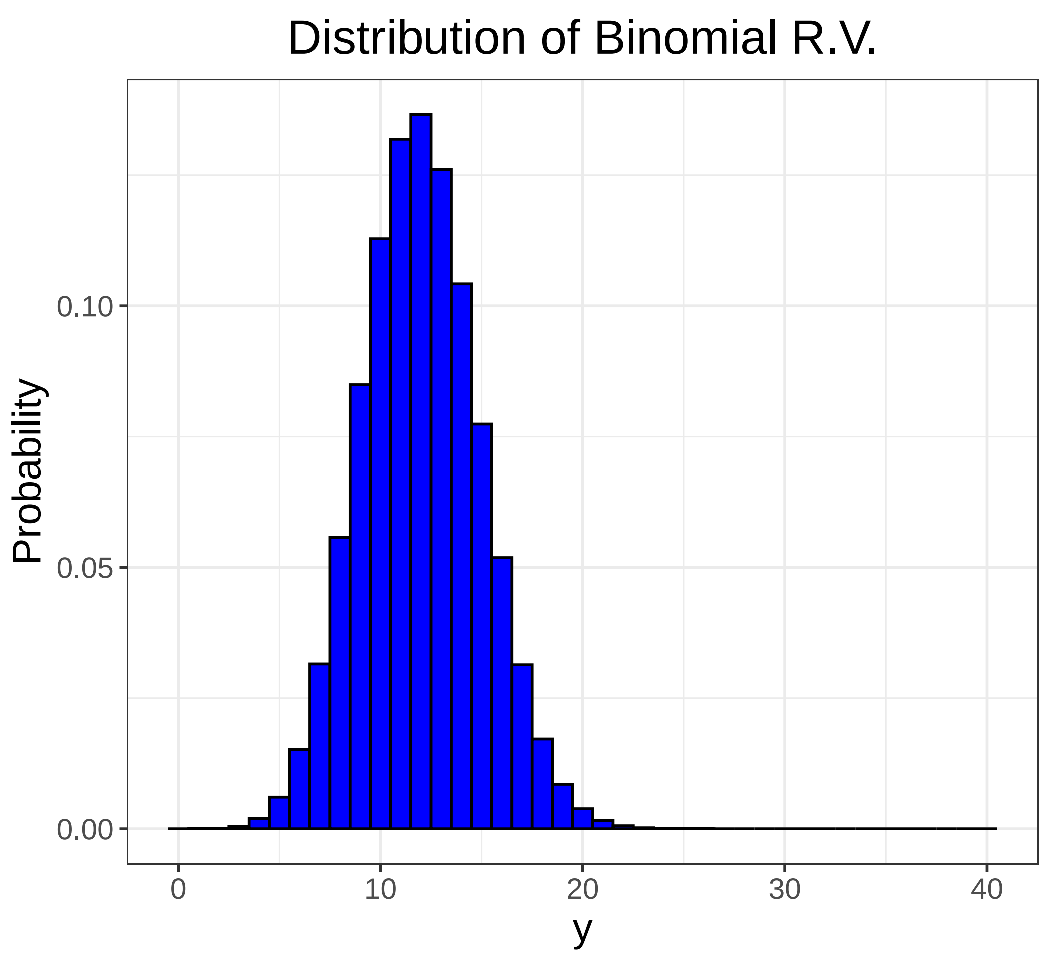
(e) (5 pts; 3 pts for variance and 2 pts for std. dev.) What is the variance and standard deviation of the number of dogs with anemia in a sample of 40 dogs from this population?

The variance is $V(Y) = np(1-p) = 40(0.3)(0.7) = 8.4$.

The standard deviation is $\sqrt{8.4} = 2.8983$.

(f) (6 pts; 4 pts for graph and 2 pts for description) Use R to produce a graph of the distribution of the number of dogs with anemia in a sample of 40 dogs from this population. Describe the shape of the distribution.

```r
plot.binom(40, 0.3)
```



The shape of the distribution is unimodal, symmetric and bell-shaped.

4. (24 pts) Suppose, based on numerous chess games between these two players, it has been determined the probability Player A would win is 0.40, the probability Player B would win is 0.35, and the probability the game would end in a draw is 0.25.

   (a) (6 pts) Find the probability that Player A would win 7 games, Player B would win 2 games, and the remaining 3 games would each end in a draw if they played 12 games.
   Find $P(Y_A = 7, Y_B = 2, Y_T = 3)$

```r
y<- c(7, 2, 3)
p <- c(0.4, 0.35, 0.25)
dmultinom(y, 12, p)

## [1] 0.02483712
```

   (b) (6 pts; 3 pts for each) Find the expected number of games Player A would win and the expected number of games Player B would win if the two players played 12 games.
   Player A: $E(Y_A) = np_A = 12(0.4) = 4.8$
   Player B: $E(Y_B) = np_B = 12(0.35) = 4.2$

   (c) (6 pts; 3 pts for each) Find the variance of the number of games Player A would win and the variance of the number of games Player B would win if the two players played 12 games.
   Player A: $V(Y_A) = np_A(1 - p_A) = 12(0.4)(0.6) = 2.88$
   Player B: $V(Y_B) = np_B(1 - p_B) = 12(0.35)(0.65) = 2.73$

   (d) (6 pts) Find the correlation of the number of games won between Player A and Player B.

$$\rho(Y_A, Y_B) = -\sqrt{\frac{p_A p_B}{(1 - p_A)(1 - p_B)}} = -\sqrt{\frac{(0.4)(0.35)}{(0.6)(0.65)}} = -0.5991447$$