Lab 2

Adam Orr

September 7, 2017

Section: 91973 Friday 9am

Question 1

Script

```
nfemales <- 3 #number of females
total <- 9 #total size of sample
p <- .5 #probability of female

(combinations <- choose(total, nfemales)) #calculate number of combinations
(p_each <- p ^ (nfemales) * (1 - p) ^ (total - nfemales)) #calculate probability of each combination
(tot_p <- combinations * p_each) #probability for 3 females
```

Output

```
(combinations <- choose(total, nfemales)) #calculate number of combinations
```

```
## [1] 84
```

```
(p_each <- p ^ (nfemales) * (1 - p) ^ (total - nfemales)) #calculate probability of each combination
```

```
## [1] 0.001953125
```

```
(tot_p <- combinations * p_each) #probability for 3 females
```

```
## [1] 0.1640625
```

Answers

What are five of the possible combinations with three females and six males? Pick any five you please.

- FFFMMMMMM
- MFFFMMMMM
- MMFFFMMMM
- MMMFFFMMM
- FMFMFMMMM

Question 2

Script

```r
nfemales <- seq(0,9) #number of females
total <- rep(9,10) #total size of sample
p <- rep(.5,10) #probability of female

(combinations <- choose(total, nfemales)) #calculate number of combinations
(p_each <- p ^ (nfemales) * (1 - p) ^ (total - nfemales)) #calculate probability of each combination
(tot_p <- combinations * p_each) #probability for all combinations
```

Output

```r
(combinations <- choose(total, nfemales)) #calculate number of combinations
```

```
##  [1]   1   9  36  84 126 126  84  36   9   1
```

```r
(p_each <- p ^ (nfemales) * (1 - p) ^ (total - nfemales)) #calculate probability of each combination
```

```
##  [1] 0.001953125 0.001953125 0.001953125 0.001953125 0.001953125
##  [6] 0.001953125 0.001953125 0.001953125 0.001953125 0.001953125
```

```r
(tot_p <- combinations * p_each) #probability for all combinations
```

```
##  [1] 0.001953125 0.017578125 0.070312500 0.164062500 0.246093750
##  [6] 0.246093750 0.164062500 0.070312500 0.017578125 0.001953125
```

Answers

Question 3

Script

```r
nfemales <- seq(0,9) #number of females
total <- 9 #total sample size
p <- .5 #probability of female

(prob <- dbinom(nfemales, size = total, .5)) #probability of each outcome
barplot(prob, #
        names.arg = nfemales,
        xlab = "Number of Females",
        ylab = "Probability",
        ylim = c(0, max(prob + .05)),
        main = "Distribution of Females")
(mu <- sum(nfemales * prob)) #calculate mean with general definition
(mu <- total * p) #calculate mean with shortcut formula
(variance <- sum((nfemales - mu)^2 * prob)) #calculate variance with general definition
(sdev <- sqrt(variance)) #take square root to get std. deviation
(variance <- mu * (1 - p)) #calculate variance with shortcut definition
(sdev <- sqrt(variance)) #take square root to get std. deviation

mn <- function(total, p){ total * p } #function to calculate mean
va <- function(mn, p){ mn * (1 - p)} #function to calculate variance
ps <- c(0.1,0.3,0.5,0.7,0.9) #probabilities to test
means <- mn(rep(total,length(ps)),ps) #means of all probabilites
```

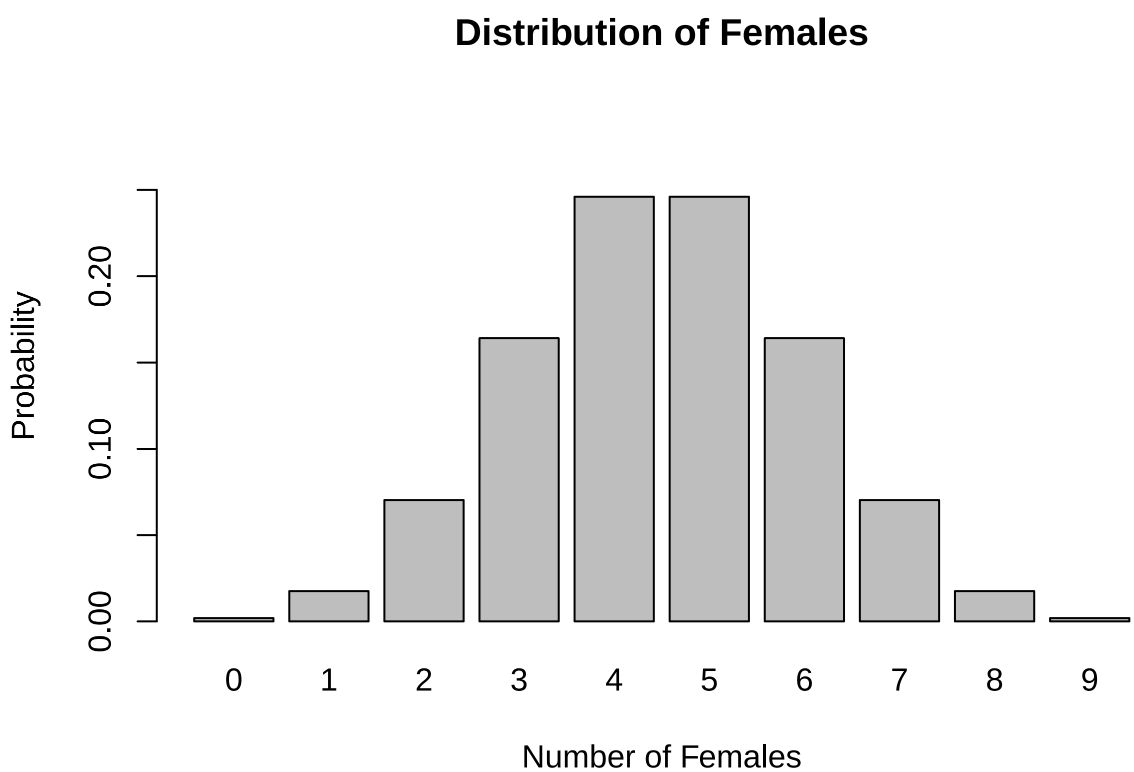
```r
variances <- va(means, ps) #variance for all probabilities
(table <- data.frame(p = ps,mean = means, variance = variances)) #table showing mean and variance for e
```

Output

```r
(prob <- dbinom(nfemales, size = total, .5)) #probability of each outcome
```

```
##  [1] 0.001953125 0.017578125 0.070312500 0.164062500 0.246093750
##  [6] 0.246093750 0.164062500 0.070312500 0.017578125 0.001953125
```

```r
barplot(prob, #
        names.arg = nfemales,
        xlab = "Number of Females",
        ylab = "Probability",
        ylim = c(0, max(prob + .05)),
        main = "Distribution of Females")
```




```r
(mu <- sum(nfemales * prob)) #calculate mean with general definition
```

```
## [1] 4.5
```

```r
(mu <- total * p) #calculate mean with shortcut formula
```

```
## [1] 4.5
```

```r
(variance <- sum((nfemales - mu)^2 * prob)) #calculate variance with general definition
```

```
## [1] 2.25
```

```r
(sdev <- sqrt(variance)) #take square root to get std. deviation
```

```
## [1] 1.5
```

```r
(variance <- mu * (1 - p)) #calculate variance with shortcut definition
```

```
## [1] 2.25
```

```r
(sdev <- sqrt(variance)) #take square root to get std. deviation
```

```
## [1] 1.5
```

```r
(table <- data.frame(p = ps,mean = means, variance = variances)) #table showing mean and variance for e
```

```
##     p mean variance
## 1 0.1  0.9     0.81
## 2 0.3  2.7     1.89
## 3 0.5  4.5     2.25
## 4 0.7  6.3     1.89
## 5 0.9  8.1     0.81
```

Answers

What happens to the mean and variance of the distribution as the probability changes?

As the probability increases, the mean increases as well. On the other hand, the variance until `p = .5`, then decreases again.

At what value is the variance of the distribution highest? Lowest?

The variance of the distribution is highest when `p = .5`. The variance is lowest when `p = .1` or `p = .9`.

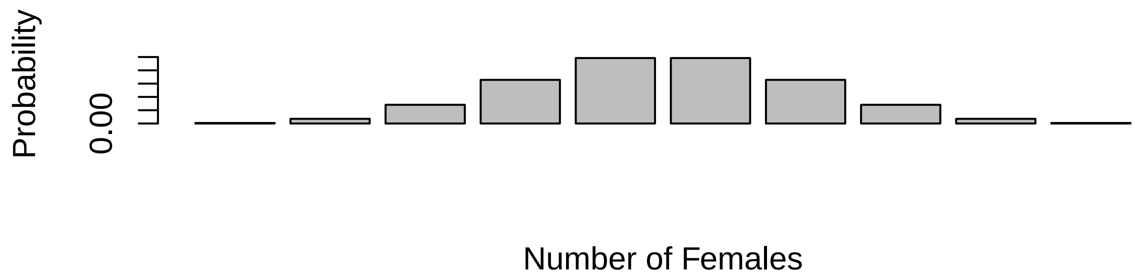
Question 4

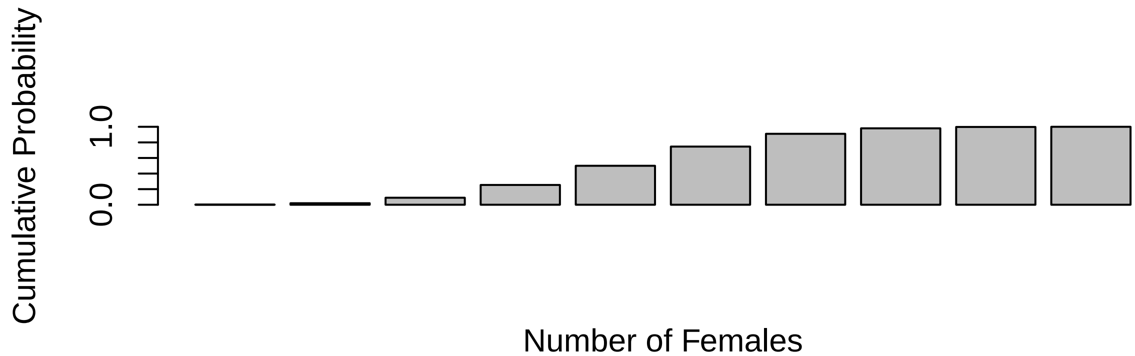
Script

```r
par(mfrow=c(2,1)) #put 2 plots on one figure
barplot(prob, xlab = "Number of Females", ylab = "Probability", ylim = c(0, max(prob + .05))) #plot pro
barplot(cumsum(prob), xlab = "Number of Females", ylab = "Cumulative Probability") #plot cumulative pro
```

Output

```r
par(mfrow=c(2,1)) #put 2 plots on one figure
barplot(prob, xlab = "Number of Females", ylab = "Probability", ylim = c(0, max(prob + .05))) #plot pro
barplot(cumsum(prob), xlab = "Number of Females", ylab = "Cumulative Probability") #plot cumulative pro
```





Answers

What is the relationship between each bar of the cumulative plot and the bars of the distribution plot?

Each bar of the cumulative plot is the height of that bar in the distribution plot added to each bar before that one in the distribution plot.

Question 5

Script

```
p <- .5 #probability of a female
dbinom(4,9,p) # exactly 4 females
dbinom(7,9,p) # exactly 7 females
pbinom(7,9,p) # 7 or fewer females
1 - pbinom(3,9,p) # 4 or more females
pbinom(5,9,p) # 4 or more males
(1 - pbinom(6,9,p)) + pbinom(2,7,p) #7 or more females + 7 or more males
```

Output

```
dbinom(4,9,p) # exactly 4 females
```

```
## [1] 0.2460938
```

```
dbinom(7,9,p) # exactly 7 females
```

```
## [1] 0.0703125
```

```r
pbinom(7,9,p) # 7 or fewer females
```

```
## [1] 0.9804688
```

```r
1 - pbinom(3,9,p) # 4 or more females
```

```
## [1] 0.7460937
```

```r
pbinom(5,9,p) # 4 or more males
```

```
## [1] 0.7460937
```

```r
(1 - pbinom(6,9,p)) + pbinom(2,7,p) #7 or more females + 7 or more males
```

```
## [1] 0.3164063
```

Answers

Question 6

Script

```r
qbinom(.25,9,.5) # first quartile
qbinom(.05,9,.5) # 5th percentile
qbinom(.95,9,.5) # 95th percentile
qbinom(.5,9,.5) # median
```

Output

```r
qbinom(.25,9,.5) # first quartile
```

```
## [1] 3
```

```r
qbinom(.05,9,.5) # 5th percentile
```

```
## [1] 2
```

```r
qbinom(.95,9,.5) # 95th percentile
```

```
## [1] 7
```

```r
qbinom(.5,9,.5) # median
```

```
## [1] 4
```

Answers

Question 7

Script

```r
(vals <- rbinom(100,9,.5)) #100 simulated values
mean(vals) #mean of simulated values
```

Output

```r
(vals <- rbinom(100,9,.5)) #100 simulated values
```

```
##   [1] 8 4 3 4 6 5 6 7 4 3 4 3 5 2 6 6 4 6 6 2 4 5 6 5 5 4 5 9 6 4 7 5 5 6 5
##  [36] 5 5 4 4 5 6 3 6 3 5 4 7 7 5 4 6 4 3 5 6 5 5 4 4 3 4 6 5 4 6 5 4 6 5 6
##  [71] 4 3 5 3 5 4 3 2 2 2 6 4 5 5 6 3 6 5 3 5 4 4 5 4 5 4 5 4 2 4
```

```r
mean(vals) #mean of simulated values
```

```
## [1] 4.66
```

Answers

Compare the average you calculated here with the mean you calculated in question 3. Do you expect them to be the same? Why or why not?

The average I calculated is 4.66, which is very close to the average I calculated in question 3, 4.5. I do not expect these values to be the same due to the stochastic nature of the experiments.