

BINF 6310 | Lab 2 | Jon Lee

title: "R Notebook" output: html_notebook — Question 1

In a population, there is $1/3$ chance that a given person has a mutation in some gene

You sample 30 people; what are the odds that exactly 12 of the people have the mutation?

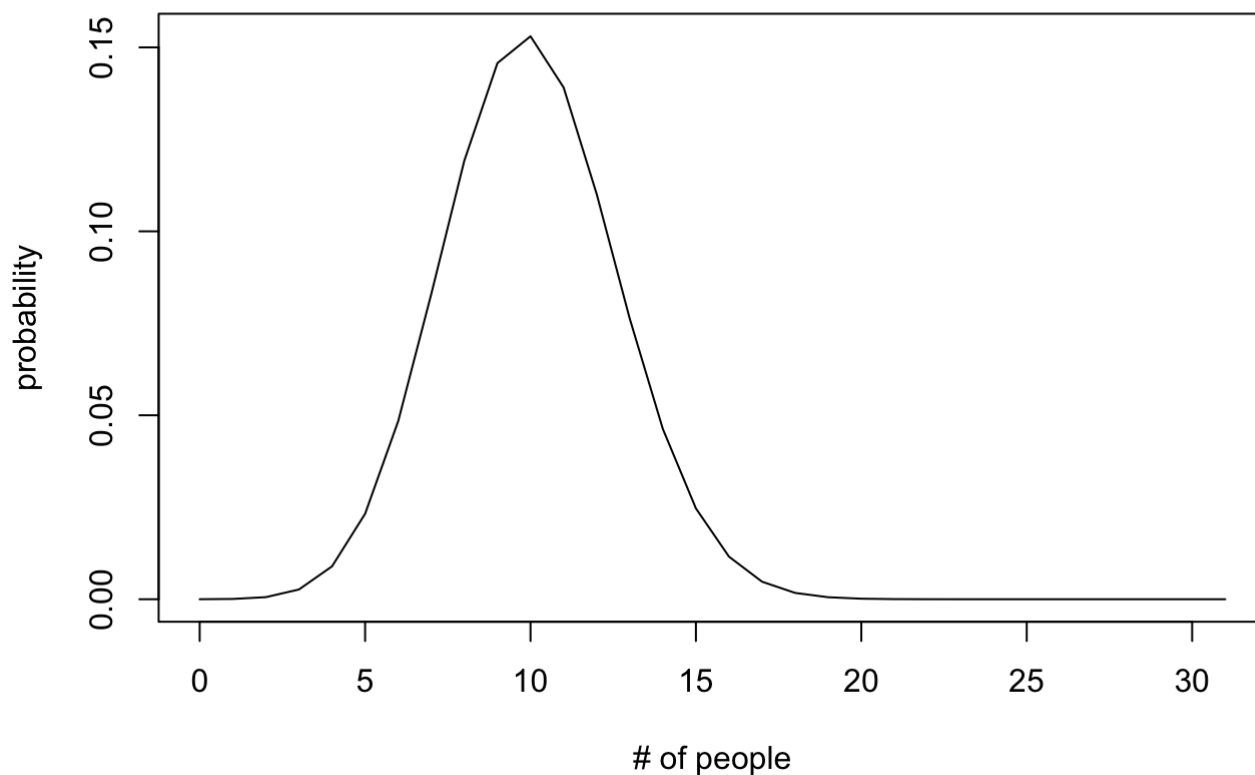
```
dbinom(12, 30, 1/3)
```

```
## [1] 0.1101246
```

In R plot a probability density function (with dbinom) that shows the distribution for observing exactly (0,1,2,...30) people with the mutation.

```
people <- seq(0,31,1)
dist <- dbinom(people, 30, 1/3)
plot(people, dist, type = "l", xlab = "# of people", ylab = "probability", main = "Probability for 'n' number of people having the gene")
```

Probability for 'n' number of people having the gene



What is the mean and variance for the expected number of people with the mutation.

```
#mean = np
#variance = np(1-p)

geneMean <- 30*(1/3)
geneMean
```

```
## [1] 10
```

```
geneVar <- 30*(1/3)*(2/3)
geneVar
```

```
## [1] 6.666667
```

Question 2

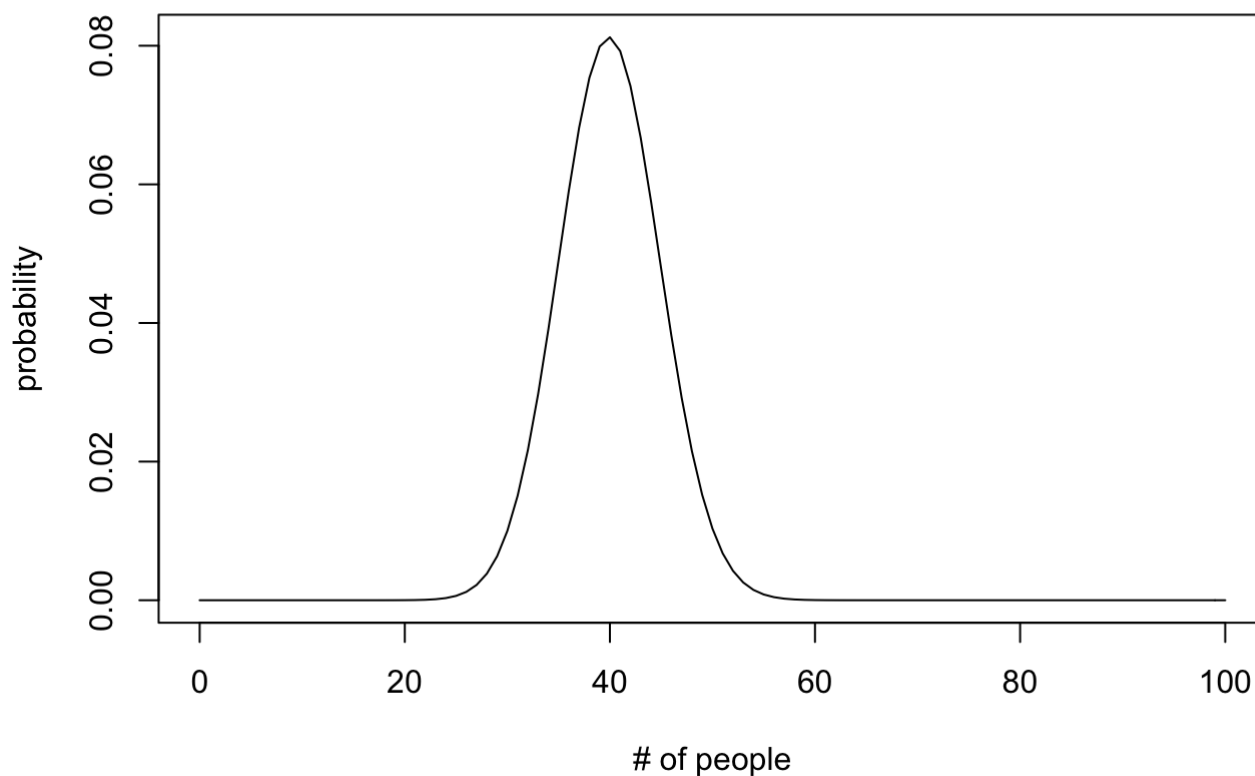
The background expected survival rate for a disease is 0.4. You are running a clinical trial. You have 100 patients on a new drug. 47 patients die (and 53 survive)

From the binomial test: (2A) Plot out the probability density function with the x-axis the number of patients that survive under the null hypothesis.

```
#the null hypothesis is that the survival rate is = 0.4

people <- seq(0,100,1)
dist <- dbinom(people, 100, 0.4)
plot(people, dist, type = "l", xlab = "# of people", ylab = "probability", main = "Distribution such that the null hypothesis is true")
```

Distribution such that the null hypothesis is true



(2B) What is the p-value for a null hypothesis that the drug has no effect. (Show the one line of R code that produces this p-value)

```
#this test is two-sided because the null is that the probability is not equal to 0.4
binom.test(53, 100, p = 0.4, alternative = "two.sided")
```

```
##
##  Exact binomial test
##
## data:  53 and 100
## number of successes = 53, number of trials = 100, p-value = 0.01036
## alternative hypothesis: true probability of success is not equal to 0.4
## 95 percent confidence interval:
##  0.4275815 0.6305948
## sample estimates:
## probability of success
##                0.53
```

(2C) What is the p-value for a null hypothesis that the drug does not improve survival. (show you can get the same answer with `binom.test(...)` and `sum(dbinom(...))`)

#this test is one sided because we want to know if the drug improves survival...that the probability is greater than 0.4

```
binom.test(53, 100, p = 0.4, alternative = "greater")
```

```
##
## Exact binomial test
##
## data: 53 and 100
## number of successes = 53, number of trials = 100, p-value = 0.005761
## alternative hypothesis: true probability of success is greater than 0.4
## 95 percent confidence interval:
## 0.4430853 1.0000000
## sample estimates:
## probability of success
## 0.53
```

```
pbinom(52, 100, 0.4, lower.tail = FALSE)
```

```
## [1] 0.005760648
```

```
sum(dbinom(53:100, 100, 0.4))
```

```
## [1] 0.005760648
```

Question 3

(3A) Use the rbinom function to simulate 1,000 experiments in which 10,000 patients are sampled with a 1/2 chance of seeing a mutation. (You should get 1,000 numbers back with each # the # of patients from the 10,000 that had the mutation...)

(What is the one line of r-code that would produce myVals?)

```
myVals <- rbinom(1000, 10000, 0.5)
```

(3B) What is the expected mean and variance of the vector in (3A). Show that the actual mean and variance are close to the expected mean and variance.

```
mean(myVals)
```

```
## [1] 4998.816
```

```
var(myVals)
```

```
## [1] 2507.345
```

```
exMean <- 10000*0.5  
exMean
```

```
## [1] 5000
```

```
exVar <- 10000*0.5*0.5  
exVar
```

```
## [1] 2500
```

(3C) Take the vector that results from (3A). For each element in that vector, calculate a p-value with `binom.test(...)$p.value` for the null hypothesis that the frequency of the allele in the population for that experiment is 1/2.

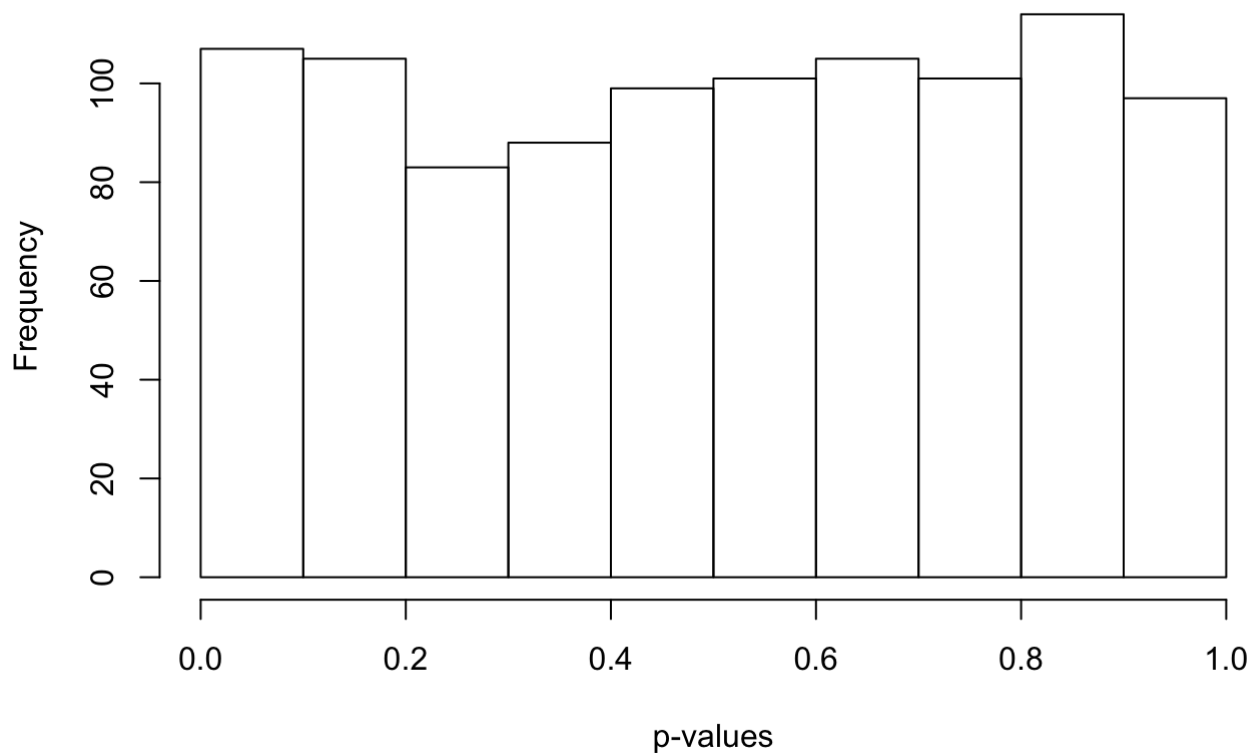
```
myPVals <- vector(length=length(myVals), mode = "double")  
  
for(i in 1:length(myVals))  
{  
  myPVals[i] <- binom.test(myVals[i], 10000, alternative = "two.sided")$p.value  
}
```

Graph the histogram of all of those p-values.

What distribution would you expect? Is that what you see?

```
hist(myPVals, xlab = "p-values", main = "Histograms of p-values")
```

Histograms of p-values



#the resulting histogram resembles a uniform distribution which is to be expected because if the null hypothesis of the experiment is true then the resulting p-values will resemble a uniform distribution

(3D) Change the expected value of 1/2 in (3C) to some other value. What happens to the p-values in the histogram. Would you expect the same shape of the p-value histogram with expected values of .49 as with .51? Why or why not?

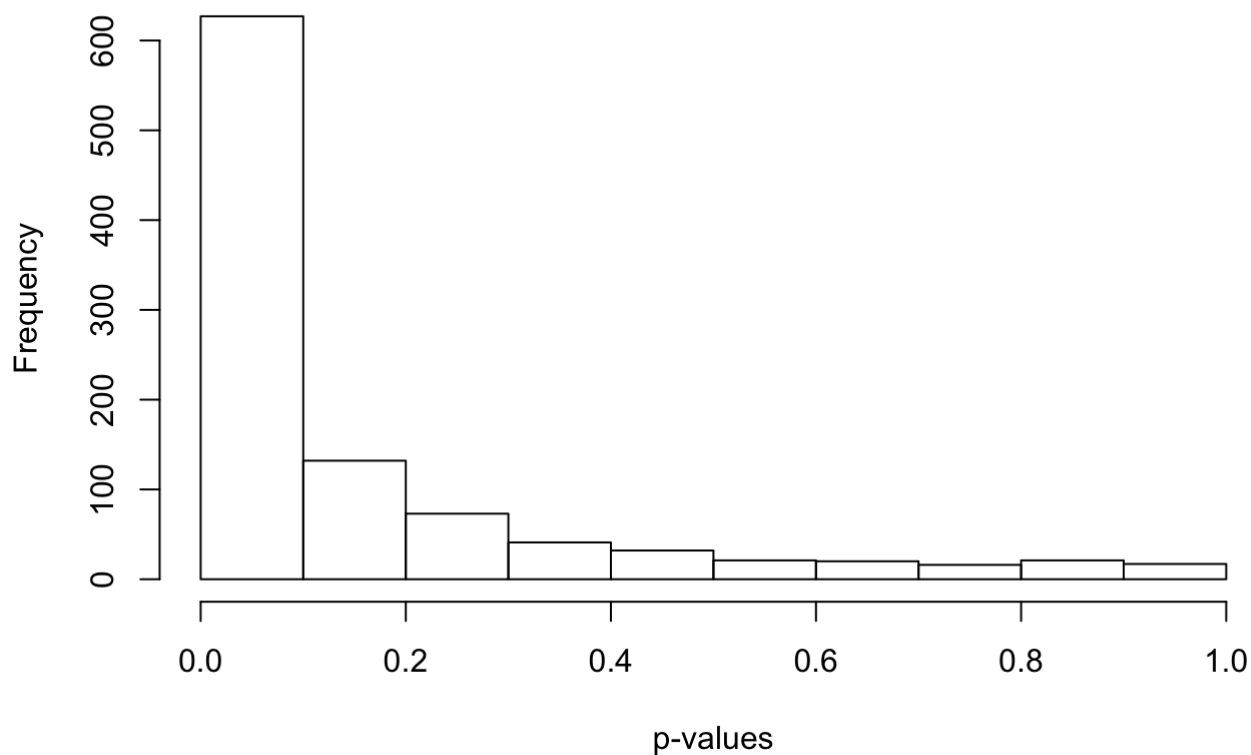
For an expected value of 0.49:

```
myPValsD_0.49 <- vector(length=length(myVals), mode = "double")

for(i in 1:length(myVals))
{
  myPValsD_0.49[i] <- binom.test(myVals[i], 10000, p = 0.49, alternative = "two.sided")
  $p.value
}
```

```
hist(myPValsD_0.49, xlab = "p-values", main = "Histograms of p-values with Mean = 0.49")
```

Histograms of p-values with Mean = 0.49



#we don't see a uniform distribution because the null hypothesis is not true, because we generated the data under a different expected value, therefore the resulting p-value are mainly small (same for 0.51)

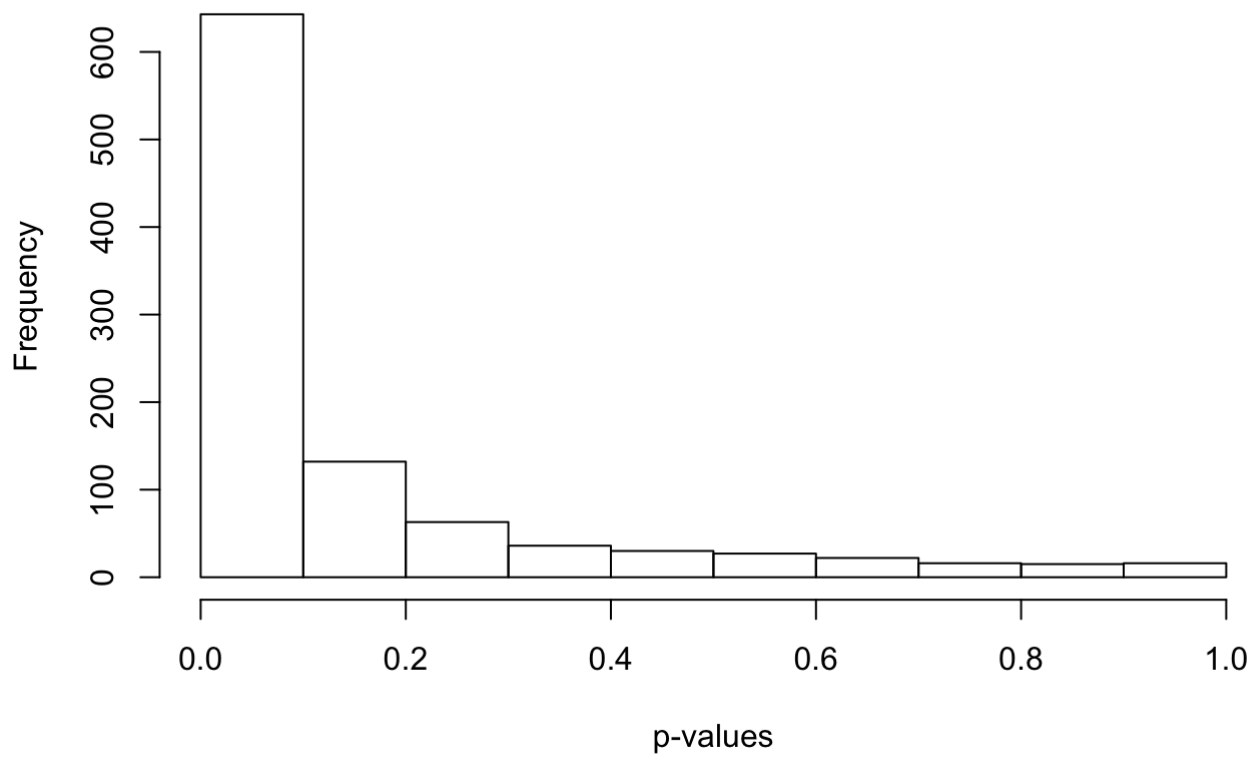
For and expected value of 0.51:

```
myPValsD_0.51 <- vector(length=length(myVals), mode = "double")

for(i in 1:length(myVals))
{
  myPValsD_0.51[i] <- binom.test(myVals[i], 10000, p = 0.51, alternative = "two.sided")
  $p.value
}
```

```
hist(myPValsD_0.51, xlab = "p-values", main = "Histograms of p-values with Mean = 0.51")
```

Histograms of p-values with Mean = 0.51



#same as above in the 0.49 expected value