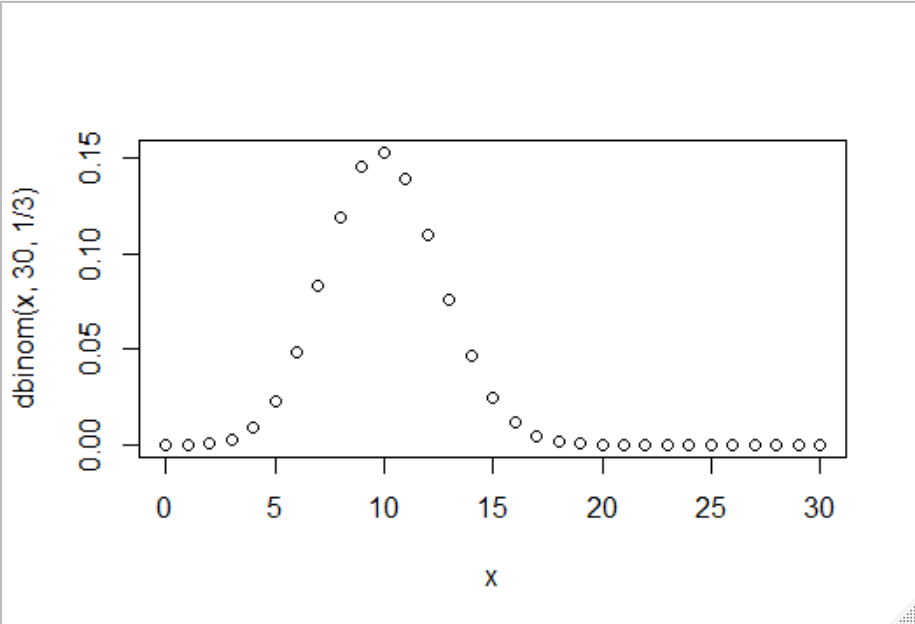
1. **In a population, there is 1/3 chance that a given person has a mutation in some gene.**

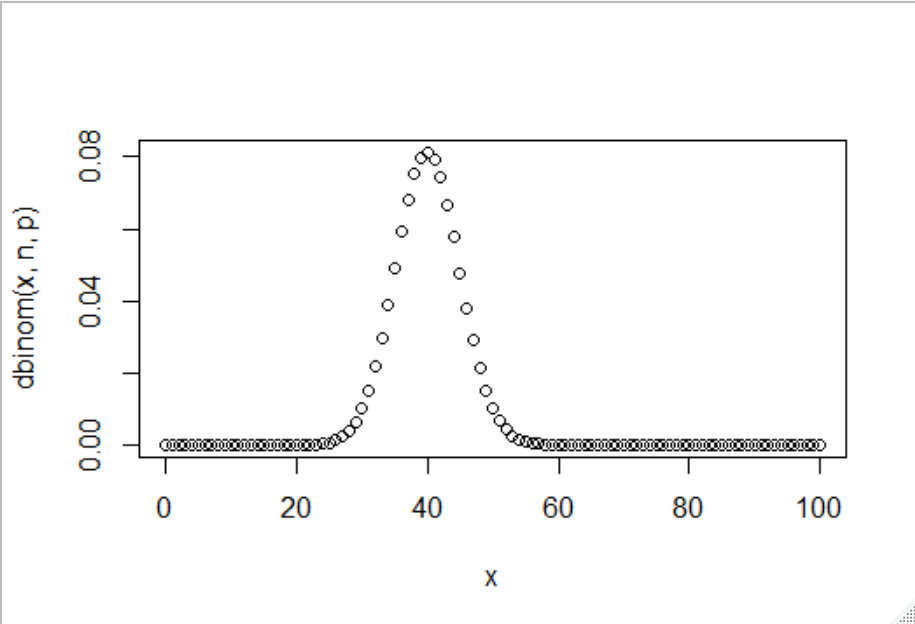
* **You sample 30 people; what are the odds that exactly 12 of the people have the mutation?**
  + > p <- 1/3; k <- 12; n <- 30  
    > p\_12 <- dbinom(k, n, p); p\_12  
    [1] 0.1101246
* **In R plot a probability density function (with dbinom) that shows the distribution for observing exactly (0,1,2,…30) people with the mutation.**  
  > x <- 0:30  
  > plot(x, dbinom(x, 30, 1/3))
* What is the mean and variance for the expected number of people with the mutation?
* > n <- 30; p <- 1/3  
  > b\_mean <- n\*p; b\_mean  
  [1] 10  
  > b\_variance <- n\*p\*(1-p); b\_variance  
  [1] 6.666667

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 bionomial test:**

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



> p <- 0.4; n <- 100; k <- 53; x <- 0:n  
> plot(x, dbinom(x, n,p))

**(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)**

> binom.test(53, 100, 0.4, alternative="two.sided")

p-value = 0.01036

**(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(….))**

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

p-value = 0.005761

**(3A)** **Use the rbiom 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, .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.**

|  |  |
| --- | --- |
| Expected Mean | 5000 |
| Actual Mean | 4999.051 |
| Expected Variance | 2500 |
| Actual Variance | 2555.884 |

> n <- 10000; p <- 0.5 ; expected\_mean <- n\*p; expected\_variance<- n\*p\*(1-p)

> expected\_mean

[1] 5000

> expected\_variance

[1] 2500

> myVals <- rbinom(1000, 10000, .5)

> actual\_mean <- mean(myVals); actual\_mean

[1] 4999.051

> actual\_variance <- var(myVals); actual\_variance

[1] 2555.884

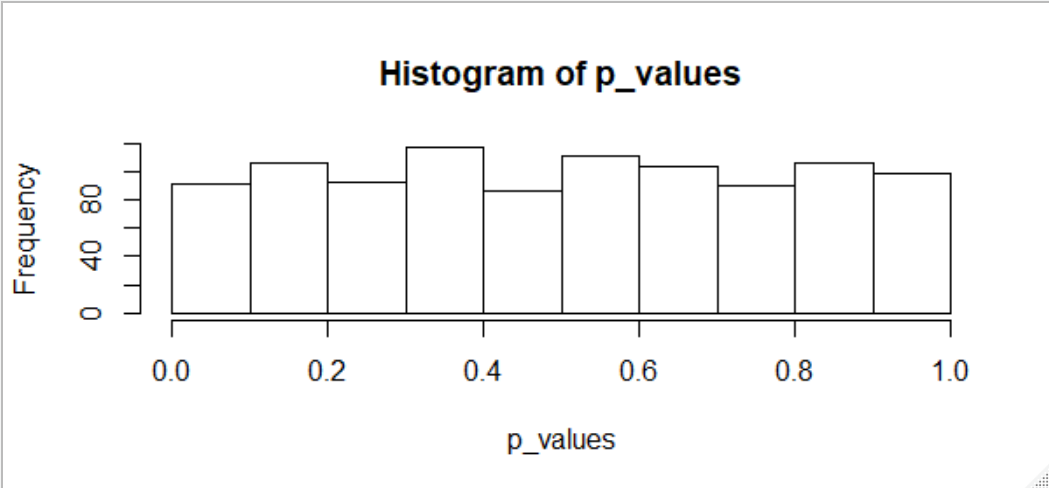
**(3C) Take the vector that results from (3A). For each element in that vector, calculate a p-value with binom.test(….)$p.value for the null hypothesis that the frequency of the allele in the population for that experiment is 1/2.**

**Graph the histogram of all of those p-values.**

* See figure 1.

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

* Since the function rbinom(1000, 10000, .5), with the underlying pdf illustrated in figure 2 below, will generate nearly all values closely clustered about the mean (5000), the multitude of corresponding p-values are greater than or equal to 0.05. Therefore, the null hypothesis will mostly not be able to be rejected. Consequently, the vector of p-values will approximate uniform distribution (per Fodor et al.).
* Yes, I observed an approximate normal distribution (figure 1).



**Figure 1: Histogram of p-values**

remove(list=objects())

prob <- 0.5

myVals <- rbinom(1000, 10000, prob)

p\_values <- vector(length = length(myVals))

cnt <- 1

for(value in myVals)

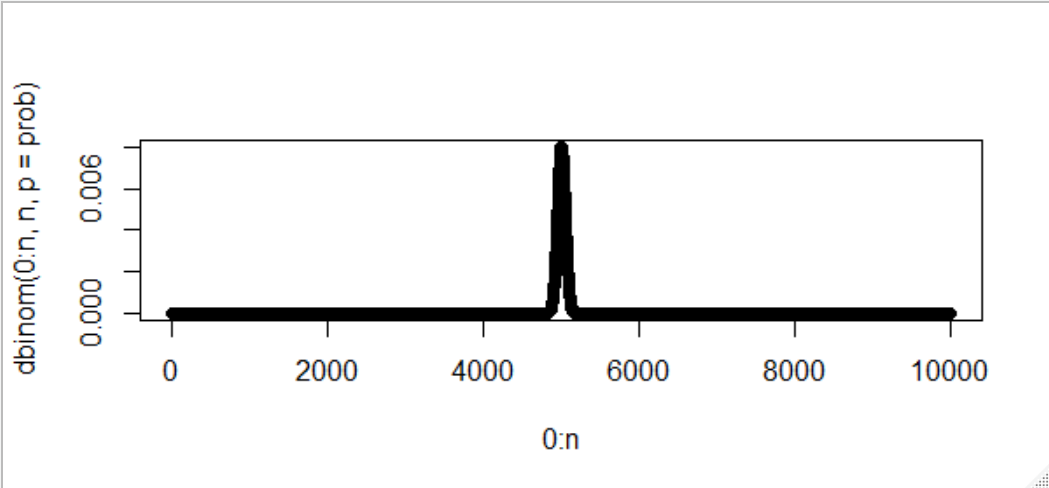
{

p\_values[cnt] <- binom.test(value, 10000, p=prob, alternative="two.sided")$p.value

cnt <- cnt + 1

}

hist(p\_values)



**Figure 2: Plot of dbinom(0:10000, 10000, p=0.5)**

remove(list=objects())

n <- 10000

prob <- 0.5

plot(0:n, dbinom(0:n, n, p=prob))

**(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?**