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?

In R plot a probability density function (with dbinom) that shows

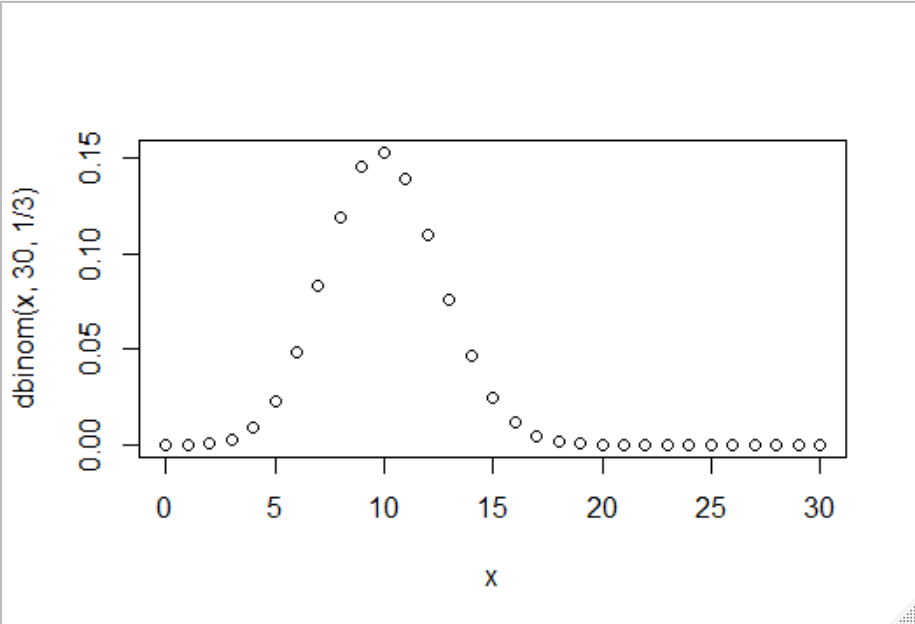
the distribution for observing exactly (0,1,2,…30) people with the mutation .

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

> p <- 1/3; k <- 12; n <- 30

> p\_12 <- dbinom(k, n, p); p\_12

[1] 0.1101246



> n <- 30; p <- 1/3

> b\_mean <- n\*p; b\_mean

[1] 10

> b\_variance <- n\*p\*(1-p); b\_variance

[1] 6.666667

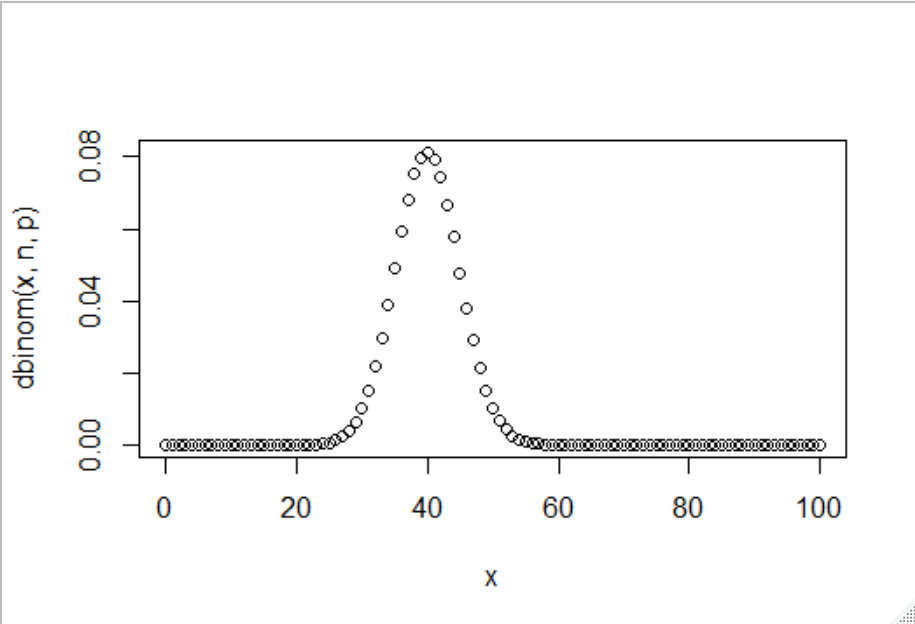
2. The background expected survival rate for a disease is 0.4. You are running a clinical trial.

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 <- 1000; k <- 53

> p <- 0.4; n <- 100; k <- 53; x <- 0:n

(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(k,n,p, alternative="two.sided")

Exact binomial test

data: k and n

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(….))

> binom.test(k,n,p, alternative="greater")

Exact binomial test

data: k and n

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

(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.

(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.

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

(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?