R Assignment 3: Simulating with distributions

Alec Brooks

2024-09-28

Question 1

a) Simulate 20 samples with normal dist. mean = 0, SD =1

```
set.seed(111)
rnorm(20,0,1)
```

```
##  [1]  0.23522071 -0.33073587 -0.31162382 -2.30234566 -0.17087604  0.14027823
##  [7] -1.49742666 -1.01018842 -0.94847560 -0.49396222 -0.17367413 -0.40659878
## [13]  1.84563626  0.39405411  0.79752850 -1.56666536 -0.08585101 -0.35913948
## [19] -1.19360897  0.36418674
```

b) Simulate 55 samples with normal dist. mean = 20.5, SD =10

```
set.seed(111)
rnorm(55,20.5,10)
```

```
##  [1]  22.852207  17.192641  17.383762  -2.523457  18.791240  21.902782
##  [7]   5.525733  10.398116  11.015244  15.560378  18.763259  16.434012
## [13]  38.956363  24.440541  28.475285   4.833346  19.641490  16.908605
## [19]   8.563910  24.141867  24.116625  23.969644  22.397365  18.904232
## [25]  23.765492  26.482542   2.084657  47.680556  22.412444   7.487039
## [31] -10.632173  11.086426  34.502588   4.295300  -2.159960  32.129936
## [37]  19.338450  23.842560  14.291419   7.401551   8.742740   9.287845
## [43]   6.880955  25.311246  27.919716  20.778246  23.813797  26.941141
## [49]  45.356616  40.099817  22.416634  36.025443  29.642423  24.086254
## [55]  22.250956
```

Question 2.

a) simulate 100 numbers with normal distribution, mean=0, SD=1

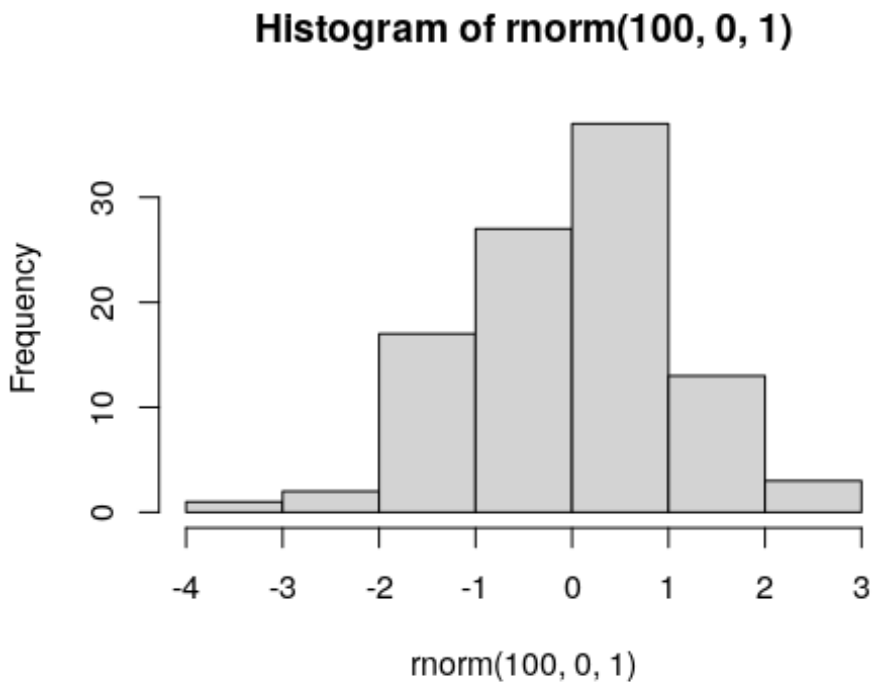
```
set.seed(111)
rnorm(100,0,1)
```

```
##   [1]  0.23522071 -0.33073587 -0.31162382 -2.30234566 -0.17087604  0.14027823
##   [7] -1.49742666 -1.01018842 -0.94847560 -0.49396222 -0.17367413 -0.40659878

...
##  [79]  1.03198498  0.38780843 -1.25612931 -0.78695273  0.42981155 -0.37641622
##  [85] -1.21622907  1.02927851  0.43039700 -1.24557402 -0.60272849  0.66006939
##  [91]  2.05074953  0.49080818 -1.73147942  0.71088366  0.01382291 -1.40104160
##  [97]  1.25912367 -0.12747752 -0.72938651 -1.21136136
```

b) Histogram

```
set.seed(111)
hist(rnorm(100,0,1))
```



c) What kind of shape do you see from the question b?

it looks like a bell shape with a slight negative skew and range of -4 to 3.

Question 3.

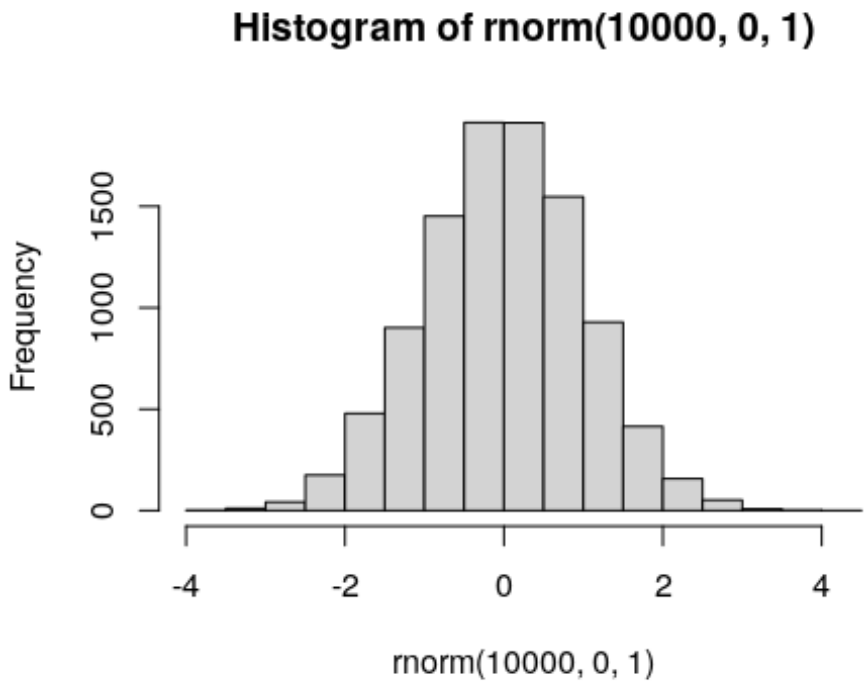
a) simulate 10000 numbers with the same condition

```
set.seed(111)
rnorm(10000,0,1)
```

```
##     [1]  0.2352207116 -0.3307358716 -0.3116238240 -2.3023456584 -0.1708760446
##     [6]  0.1402782251 -1.4974266556 -1.0101884191 -0.9484756050 -0.4939622172
##    [11] -0.1736741280 -0.4065987800  1.8456362638  0.3940541100  0.7975285013
##    [16] -1.5666653602 -0.0858510088 -0.3591394812 -1.1936089666  0.3641867370
##    [21]  0.3616624516  0.3469643701  0.1897365267 -0.1595768055  0.3265492384
##    [26]  0.5982542024 -1.8415343001  2.7180555995  0.1912443905 -1.3012960653
…
##  [9926]  1.9621309593 -1.3187787433 -0.7740403891  0.6767175693 -0.6651743653
##  [9996]  0.1923720500 -1.6935485373 -1.0078860862 -0.2308864252  0.0858787218
```

b) Histogram

```
set.seed(111)
hist(rnorm(10000,0,1))
```



c) Let's check its 5-number summary

```
set.seed(111)
summary(rnorm(10000,0,1))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -3.791832 -0.683419  0.004324 -0.002323  0.684187  4.058589
```

d) Let's check sd

```
set.seed(111)
sd(rnorm(10000,0,1))
```

```
## [1] 1.00197
```

e) What kind of shape do you see from the question b?

Like the previous chart, this one is bell-shaped but appears less skewed and more centered around the mean.

f) The law of large number says as a sample size grows, its mean gets closer to the average of the whole population. This is due to the sample being more representative of the population as the sample become larger. Do you see this comparing to two graphs with 100 samples and 10000 samples?

Very much so.

g) Between values -4 and -4, the frequency is close to 0. What does it mean?

I'm not sure I fully understand the question, but if we're referring to the values at -4 and 4, I would say that these values are low because they are the furthest from the mean.

h) Between values 0 and 0.5, the frequency is around 2000. What does it mean?

This is where most of our values are concentrated, as it represents the mean.

Question 4.

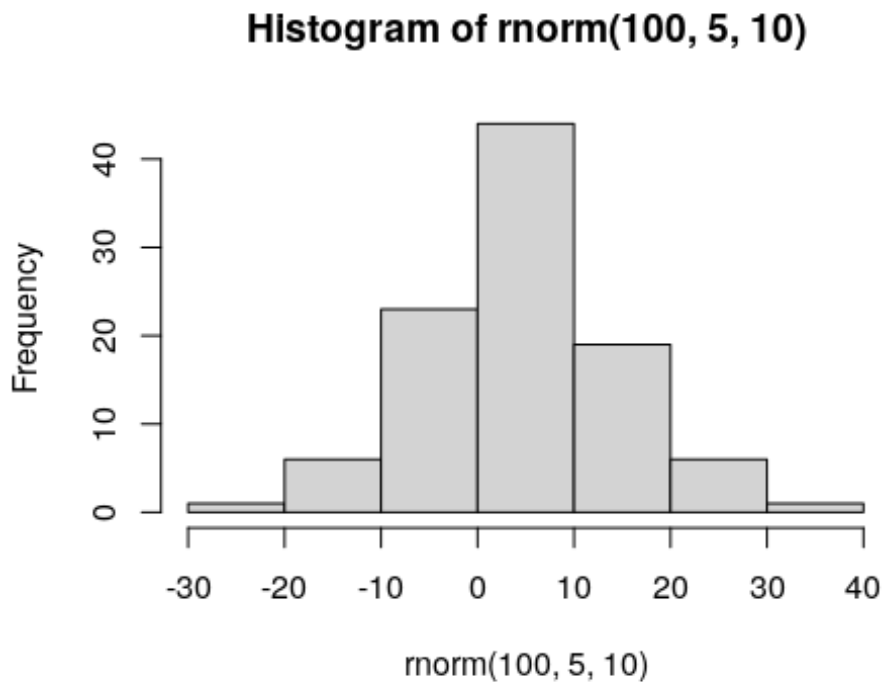
a) simulate 100 numbers with normal distribution, mean=5, SD=10

```
set.seed(111)
rnorm(100,5,10)
```

```
##   [1]   7.35220712   1.69264128   1.88376176 -18.02345658   3.29123955
##   [6]   6.40278225  -9.97426656  -5.10188419  -4.48475605   0.06037783
##  [11]   3.26325872   0.93401220  23.45636264   8.94054110  12.97528501
…
##  [86]  15.29278514   9.30396999  -7.45574020  -1.02728486  11.60069387
##  [91]  25.50749526   9.90808179 -12.31479419  12.10883659   5.13822911
##  [96]  -9.01041597  17.59123665   3.72522482  -2.29386514  -7.11361360
```

b) Histogram

```
set.seed(111)
hist(rnorm(100,5,10))
```



Question 5

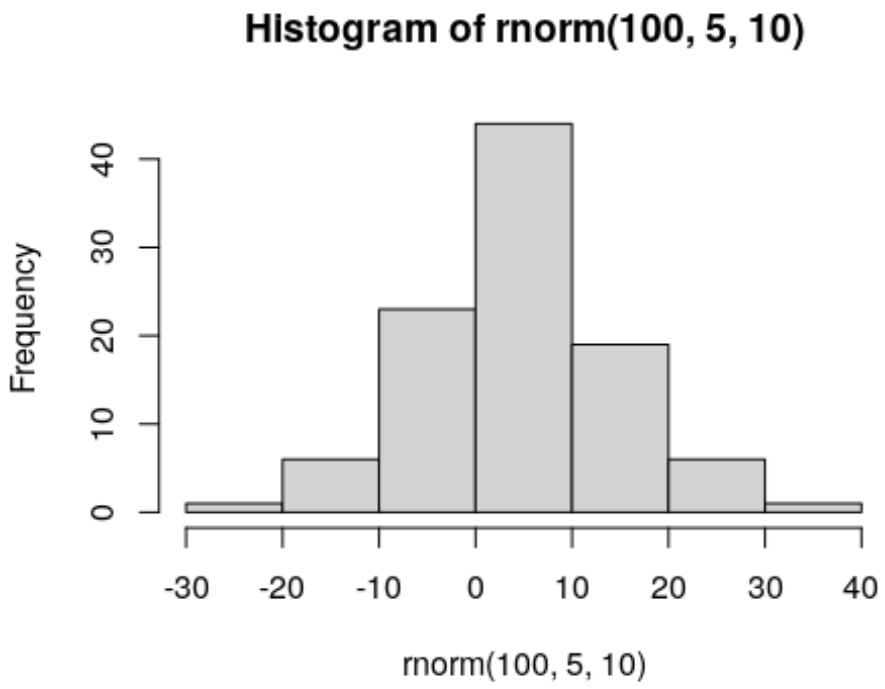
a) simulate 100 numbers with normal distribution, mean=5, SD=10

```
set.seed(111)
rnorm(100,5,10)
```

```
##   [1]   7.35220712   1.69264128   1.88376176 -18.02345658   3.29123955
##   [6]   6.40278225  -9.97426656  -5.10188419  -4.48475605   0.06037783
##  [11]   3.26325872   0.93401220  23.45636264   8.94054110  12.97528501
…
##  [91]  25.50749526   9.90808179 -12.31479419  12.10883659   5.13822911
##  [96]  -9.01041597  17.59123665   3.72522482  -2.29386514  -7.11361360
```

b) Histogram

```
set.seed(111)
hist(rnorm(100,5,10))
```



c) Let's check its 5-number summary

```
set.seed(111)
summary(rnorm(100,5,10))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -26.132  -2.438   6.316   4.872  11.097  32.181
```

d) Comparing two histograms from #3 and 5, #5 is shifted to the right. Why? Explian.

we increased the mean from 1 to 5, causing the mean to shift to the right.

e) The x-axis has different numbers. Why? Explain.

we increased the standard deviation, which significantly widened the range of the data, increasing both the minimum and maximum values.

Question 6

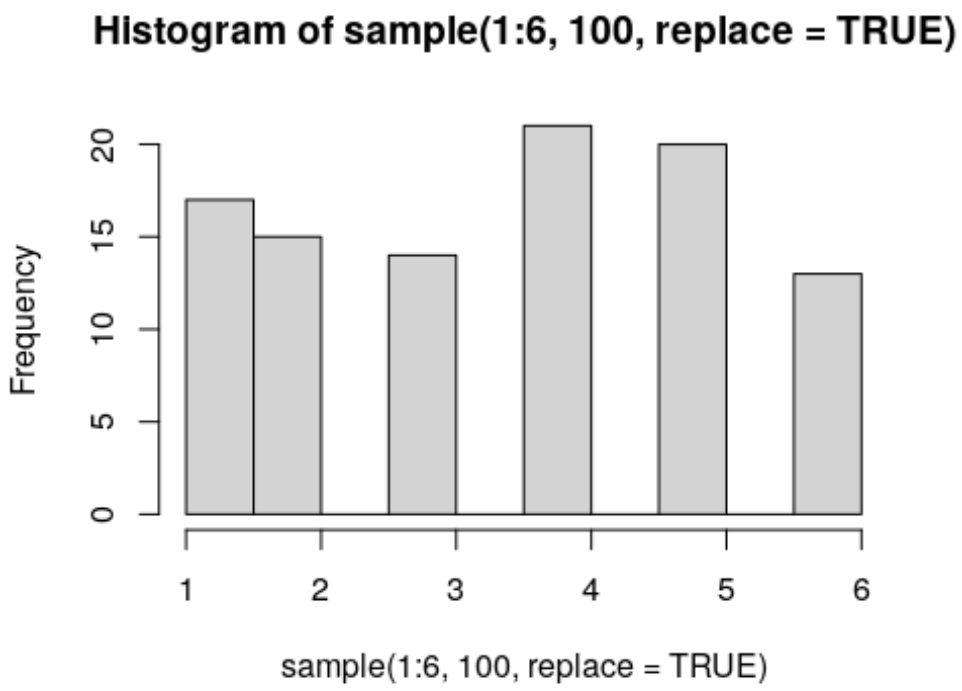
a) Simulate rolling a dice for 100 times using sample()

```
set.seed(111)
sample(1:6, 100,replace=TRUE)
```

```
##   [1] 6 3 4 3 1 3 5 3 4 2 1 5 5 2 4 6 2 1 3 1 3 5 1 1 1 3 4 4 4 6 1 5 4 4 6 6 5
##  [38] 5 5 6 4 1 4 2 2 2 5 5 3 5 5 5 1 4 2 2 4 4 1 1 3 6 5 5 3 3 5 2 6 4 2 1 5 1
##  [75] 2 2 2 5 3 1 1 3 4 4 3 4 6 4 6 2 1 4 2 6 5 4 4 5 6 6
```

b) Histogram

```
set.seed(111)
hist(sample(1:6, 100,replace=TRUE))
```



c) Let's check its 5-number summary

```
set.seed(111)
summary(sample(1:6, 100,replace=TRUE))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    2.00    4.00    3.51    5.00    6.00
```

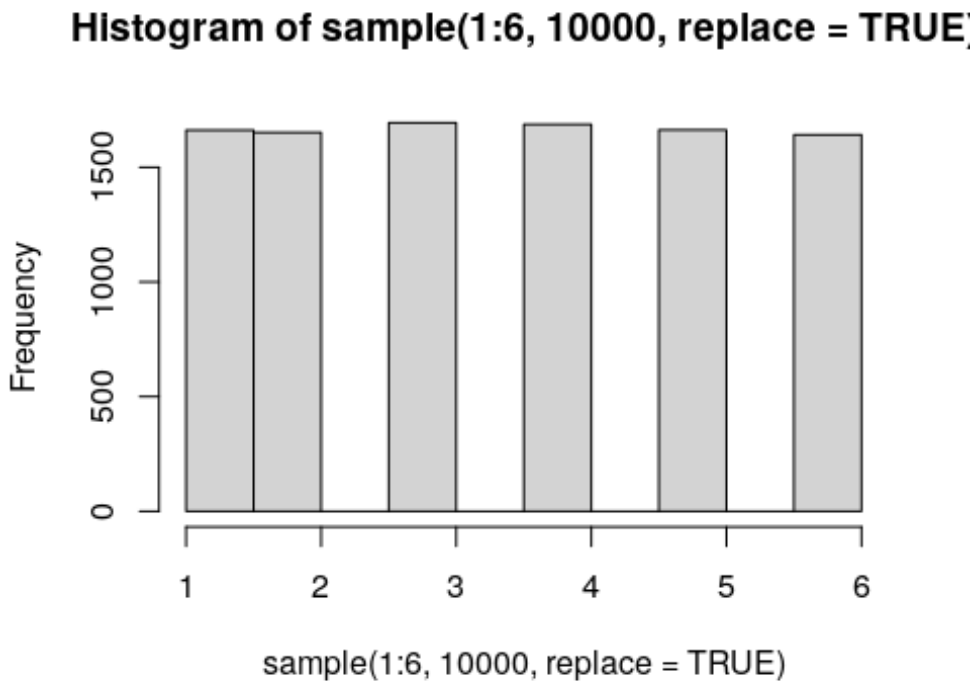
d) Simulate rolling a dice for 10000 times using sample()

```
set.seed(111)
sample(1:6, 10000,replace=TRUE)
```

```
##     [1] 6 3 4 3 1 3 5 3 4 2 1 5 5 2 4 6 2 1 3 1 3 5 1 1 1 3 4 4 4 6 1 5 4 4 6 6
##    [37] 5 5 5 6 4 1 4 2 2 2 5 5 3 5 5 5 1 4 2 2 4 4 1 1 3 6 5 5 3 3 5 2 6 4 2 1
…
##  [9937] 6 6 4 6 1 6 3 4 5 3 3 5 1 1 3 2 5 6 4 5 2 4 4 3 2 2 6 4 5 1 5 5 5 6 4 6
##  [9973] 3 4 2 3 3 3 3 3 3 2 4 5 1 4 6 6 3 3 2 2 3 5 2 2 5 3 1 4
```

e) Histogram

```
set.seed(111)
hist(sample(1:6, 10000,replace=TRUE))
```



f) 5-number summary

```
set.seed(111)
summary(sample(1:6, 10000,replace=TRUE))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   3.496   5.000   6.000
```

g) what kind of shape do you see?

All options from 1 to 6 are approximately the same height

h) What does it mean in term of the frequency?

All options from 1 to 6 show approximately the same frequency

i) What's this distribution?

this is a Uniform Distribution

Question 7

a) simulate 100 numbers from 5 to 20 with uniform distribution

```
set.seed(111)
runif(100,5,20)
```

```
##   [1] 13.894719 15.897217 10.556330 12.723857 10.664948 11.275060  5.159868
##   [8] 12.984429 11.482409  6.405223 13.336699 13.853427  6.007117  5.713218
…
##  [92] 17.794756 14.447317 13.685089 16.103739 10.807246 19.903017 10.971342
##  [99] 19.625016 17.367233
```

b) simulate 77 numbers from 0.5 to 7.77

```
set.seed(111)
runif(77,0.5,7.77)
```

```
##  [1] 4.8109739 5.7815178 3.1929680 4.2434962 3.2456116 3.5413124 0.5774825
##  [8] 4.3697864 3.6418077 1.1810646 4.5405200 4.7909611 0.9881161 0.8456729
…
## [64] 2.3845480 7.1831807 2.1861281 0.8821524 2.7129344 0.5852467 2.6861450
## [71] 6.8800353 5.3365764 3.7988703 0.8876534 5.0866925 3.7146863 2.4436090
```

Question 8

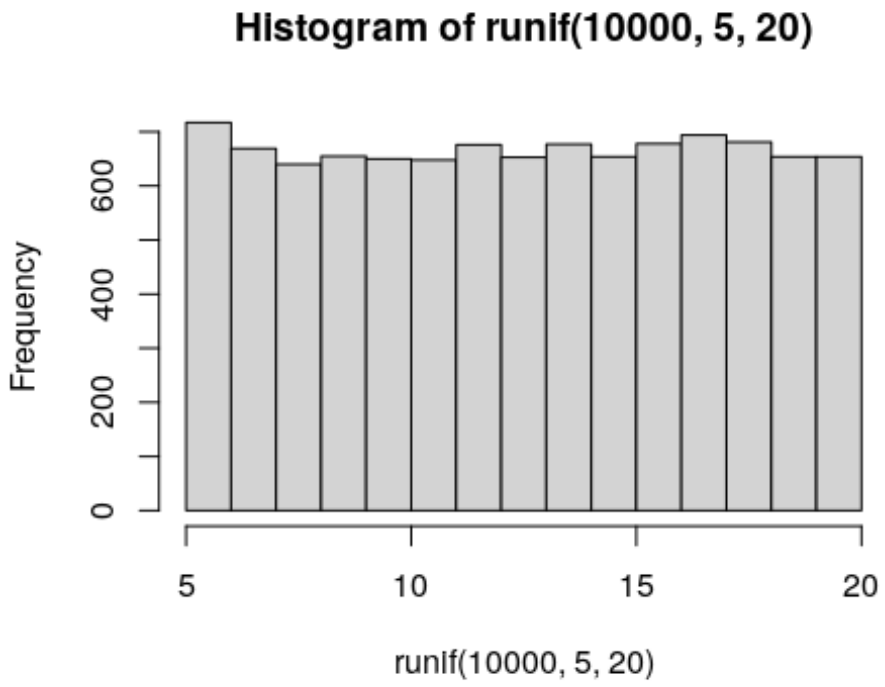
a) simulate 10000 numbers from 5 to 20 with uniform distribution

```
set.seed(111)
runif(10000,5,20)
```

```
##     [1] 13.894719 15.897217 10.556330 12.723857 10.664948 11.275060  5.159868
##     [8] 12.984429 11.482409  6.405223 13.336699 13.853427  6.007117  5.713218
##    [15]  7.343038 11.696416  7.571655 19.498014  9.659996 14.216996 11.465912
…
##  [9983] 15.826308 18.903183  5.621299  6.758795 16.415157 19.346479 17.165436
##  [9990] 13.721758  5.005755 10.548254 18.675993 18.352727  8.432662 13.831694
##  [9997]  9.173216 11.948201 13.451430 12.585942
```

b) Histogram

```
set.seed(111)
hist(runif(10000,5,20))
```



c) what kind of shape do you see?

It appears relatively even across the range.

d) What does it mean in term of the frequency?

Each bucket has an equal probability.

e) Does it belong to the uniform distribution?

I think so. However, this method differs from the sample() method because it outputs a continuous set of values. But, I believe a uniform distribution can consist of either discrete or continuous values.

f) Between values 10-11, looks like the frequency is about 600. What does it mean?

It means that values between 10 and 11 appear slightly less likely than other values. If we increase the number of tests, this discrepancy should disappear, as it's likely due to random variation.

Question 9 (prob 0.5)

a) Simulate flipping 2 coins for 10 times

```
set.seed(111)
rbinom(10,2,0.5)

##  [1] 1 1 1 1 1 1 0 1 1 0
```

b) Simulate flipping Flip 10 coins for 1 time

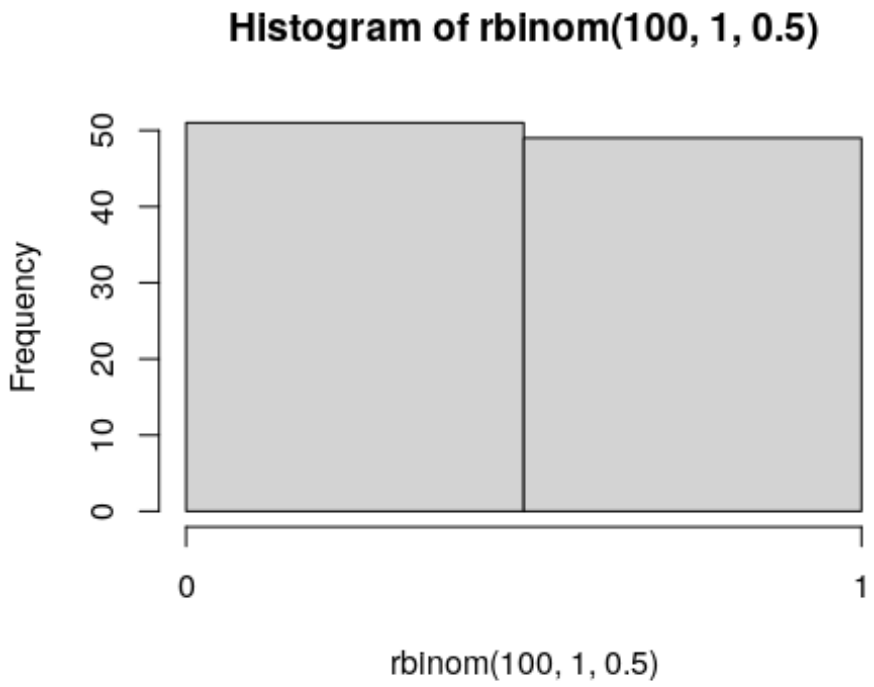
```
set.seed(111)
rbinom(1,10,0.5)

## [1] 5
```

Question 10 (prob 0.5)

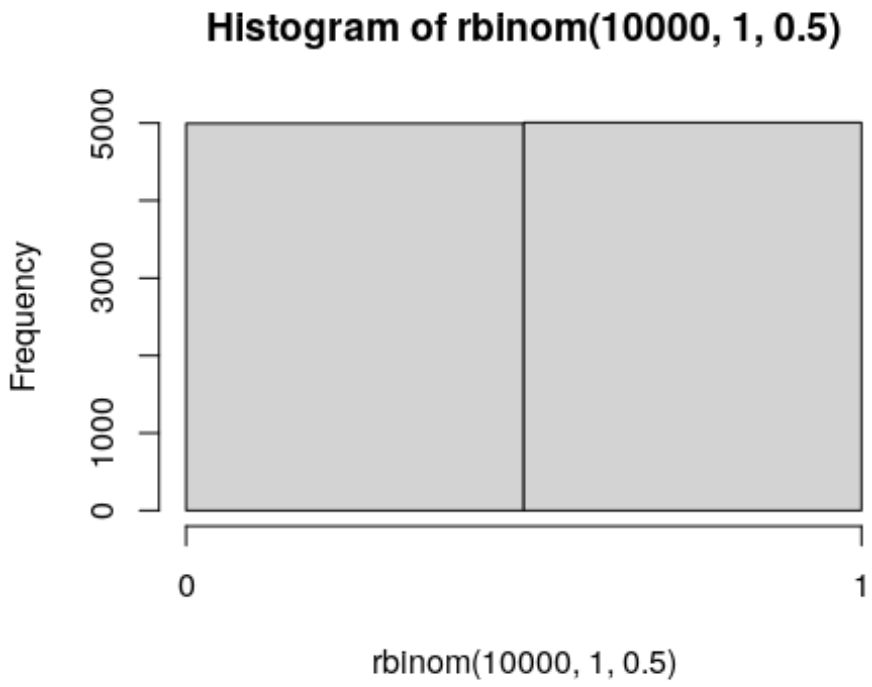
a) Simulate flipping 1 coin for 100 times and its histogram

```
set.seed(111)
hist(rbinom(100,1,0.5),breaks = 2, xaxt = "n")
axis(1, at = c(0, 1), labels = c(0, 1))
```



b) Simulate flipping 1 coin for 10000 times

```
set.seed(111)
hist(rbinom(10000,1,0.5),breaks = 2, xaxt = "n")
axis(1, at = c(0, 1), labels = c(0, 1))
```



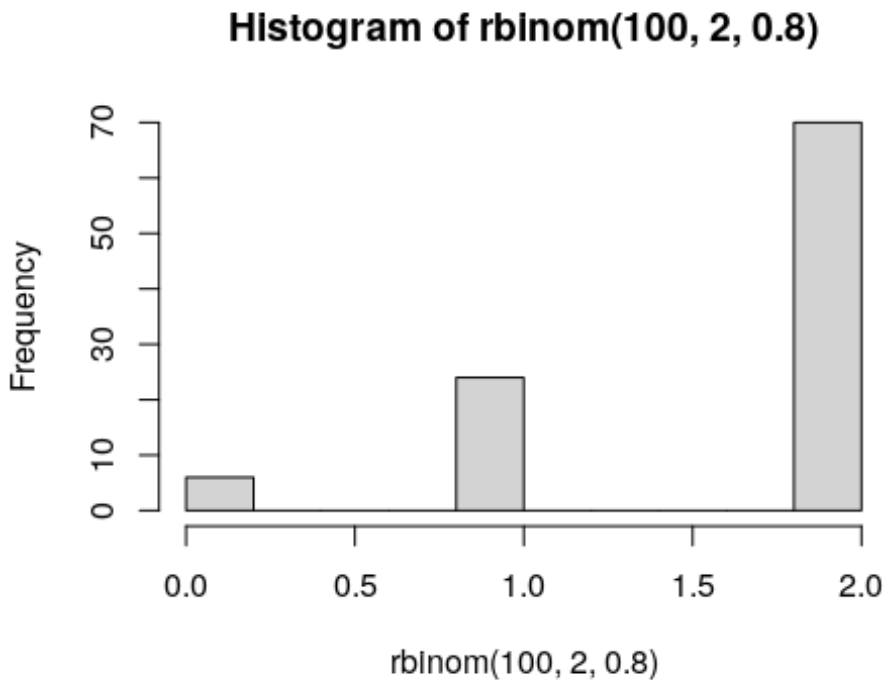
c) What is your observation?

Like our previous experiments, as we increase the number of tests, the gap in the frequency of heads and tails decreases.

Question 11 (prob 0.8)

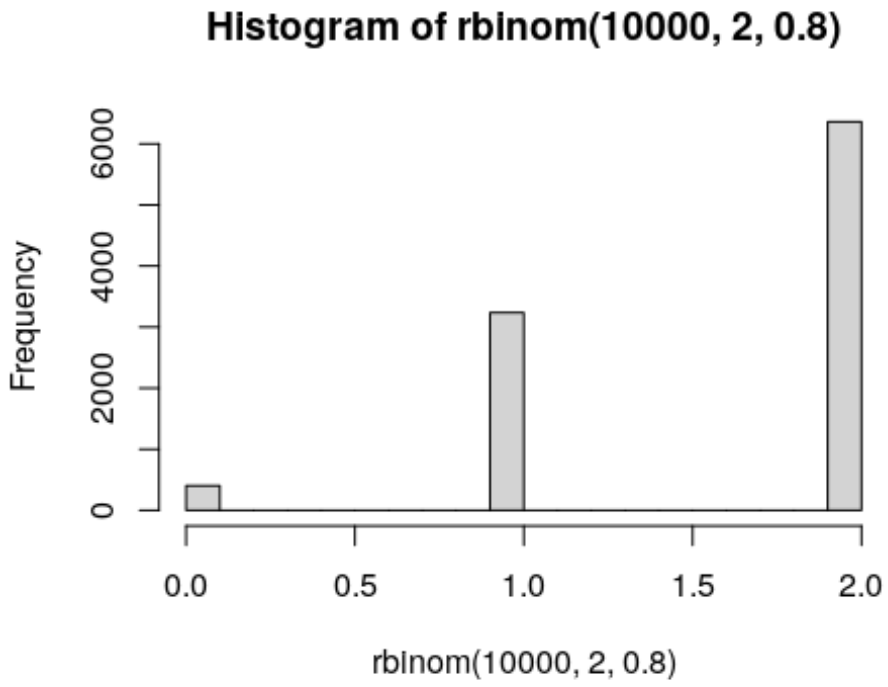
a) Simulate flipping 2 coins for 100 times and its histogram

```
set.seed(111)
hist(rbinom(100,2,0.8))
```



b) Simulate flipping 2 coins for 10000 times and its histogram

```
set.seed(111)
hist(rbinom(10000,2,0.8))
```



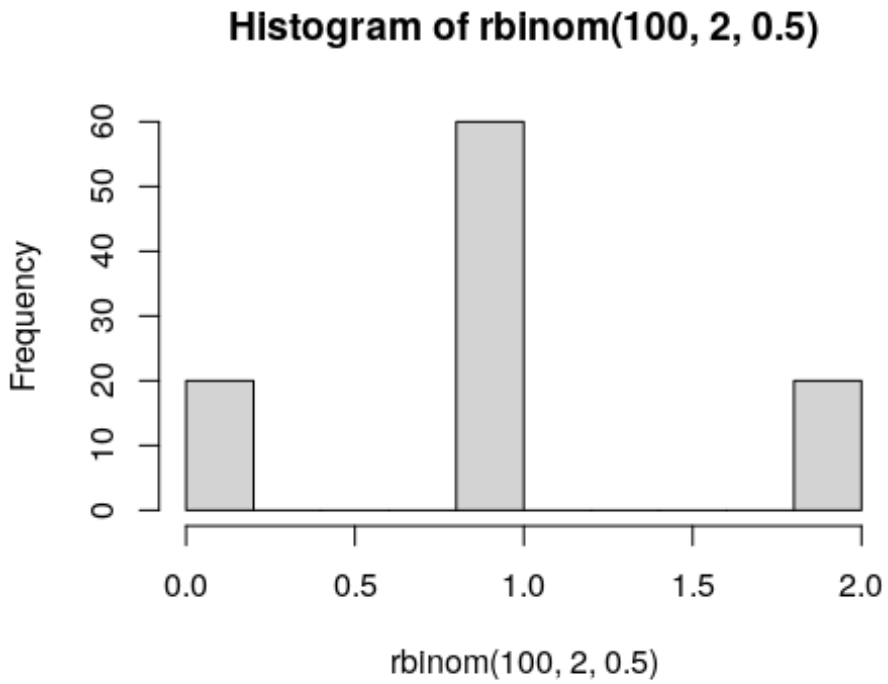
c) What is your observation?

By changing the probability from 0.5 to 0.8, we see a significant increase in tests with both coins landing heads.

Question 12. (prob 0.5)

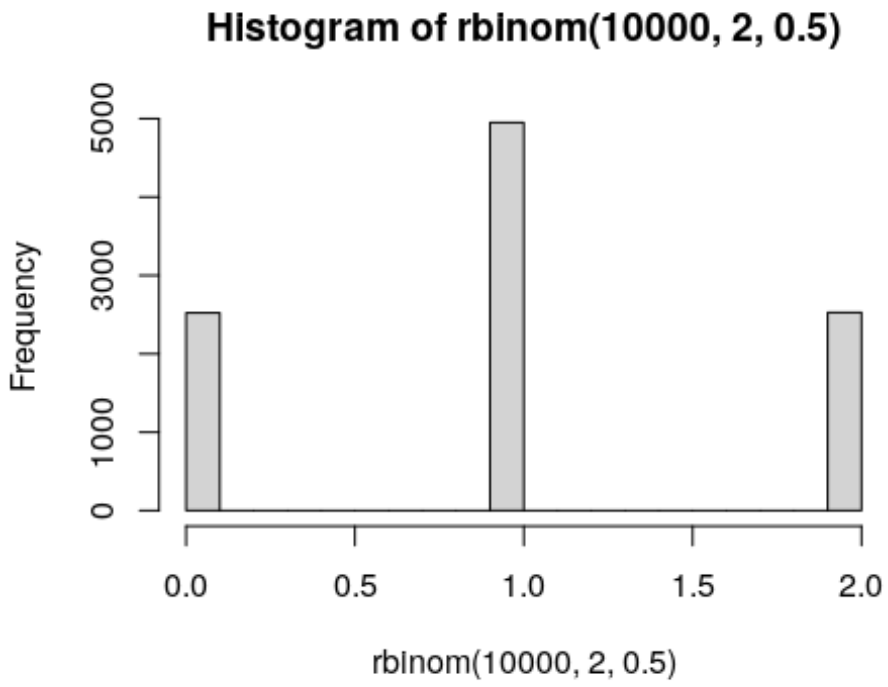
a) Simulate flipping 2 coins for 100 times and its histogram

```
set.seed(111)
hist(rbinom(100,2,0.5))
```



b) Simulate flipping 2 coins for 10000 times and its histogram

```
set.seed(111)
hist(rbinom(10000,2,0.5))
```



c) What is your observation?

By altering the probability back to 0.5, we observe the frequency returning closer to 1.

d) Why the value 5 has the highest bar?

I'm sorry, I'm not sure I fully understand the question. The bar for 1 is the highest in this graph because there are two combinations (H,T) and (T,H) that result in exactly one heads, while there is only one combination each for both heads (H,H) and no heads (T,T). Therefore, if you flip two coins, there's a 50% chance of getting **exactly** one heads and a 75% chance of getting **at least** one heads. I hope that covers the question.

Question 13

Suppose the probability a client will buy a product in a store is 0.35. What is the distribution of the probabilities of buying a product if in the store in one day arrive 15 clients?

a) Simulate it and predict how many customers will buy.

```
set.seed(111)
rbinom(15,1,0.35)
```

```
##  [1] 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
```

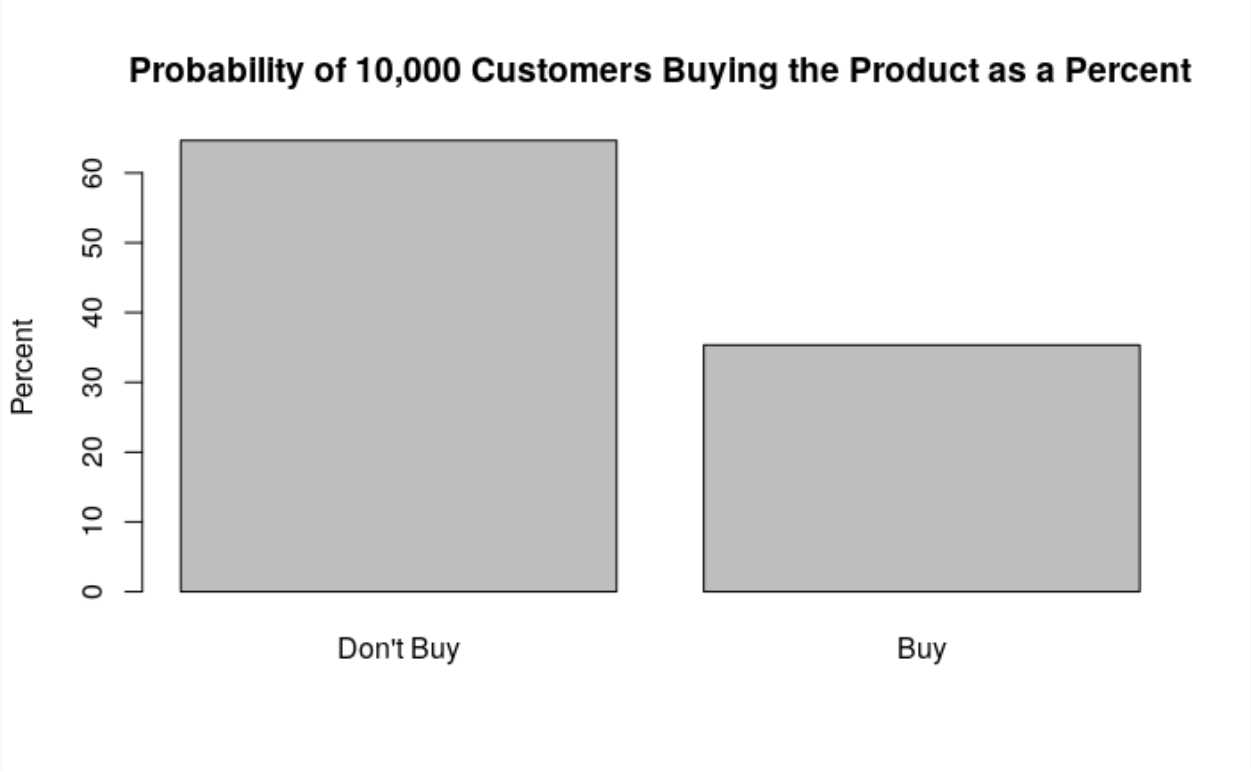
Using seed 111, only one customer would make a purchase. However, with a larger sample size, like 1,000 customers, approximately 350 should buy.

b) If the customer increases to 10,000, then what percent will buy? First predict it without simulating it

Prediction: 3,500 customers would buy or 35%

c) Plot the frequency table and see actual percentage if it is close to your prediction.

```
set.seed(111)
barplot(
  prop.table(table(rbinom(10000, 1, 0.35))) * 100,
  main = "Probability of 10,000 Customers Buying the Product as a Percent",
ylab = "Percent", names.arg = c("Don't Buy", "Buy"))
```



Full Code

```
set.seed(111)
rnorm(20,0,1)
set.seed(111)
rnorm(55,20.5,10)
set.seed(111)
rnorm(100,0,1)
set.seed(111)
hist(rnorm(100,0,1))

set.seed(111)
rnorm(10000,0,1)
```

```r
set.seed(111)
hist(rnorm(10000,0,1))

set.seed(111)
summary(rnorm(10000,0,1))
set.seed(111)
sd(rnorm(10000,0,1))
set.seed(111)
rnorm(100,5,10)
set.seed(111)
hist(rnorm(100,5,10))
set.seed(111)
rnorm(100,5,10)
set.seed(111)
hist(rnorm(100,5,10))

set.seed(111)
summary(rnorm(100,5,10))
set.seed(111)
hist(rnorm(100,5,10))
set.seed(111)
summary(rnorm(100,5,10))
set.seed(111)
sample(1:6, 100,replace=TRUE)
set.seed(111)
hist(sample(1:6, 100,replace=TRUE))

set.seed(111)
summary(sample(1:6, 100,replace=TRUE))
set.seed(111)
sample(1:6, 10000,replace=TRUE)
set.seed(111)
hist(sample(1:6, 10000,replace=TRUE))

set.seed(111)
summary(sample(1:6, 10000,replace=TRUE))
set.seed(111)
runif(100,5,20)
set.seed(111)
runif(77,0.5,7.77)
set.seed(111)
runif(10000,5,20)
set.seed(111)
hist(runif(10000,5,20))

set.seed(111)
rbinom(10,2,0.5)
set.seed(111)
rbinom(1,10,0.5)
set.seed(111)
```

```r
hist(rbinom(100,1,0.5),breaks = 2, xaxt = "n")
axis(1, at = c(0, 1), labels = c(0, 1))

set.seed(111)
hist(rbinom(10000,1,0.5),breaks = 2, xaxt = "n")
axis(1, at = c(0, 1), labels = c(0, 1))

set.seed(111)
hist(rbinom(100,2,0.8))

set.seed(111)
hist(rbinom(10000,2,0.8))

set.seed(111)
hist(rbinom(100,2,0.5))

set.seed(111)
hist(rbinom(10000,2,0.5))

set.seed(111)
rbinom(15,1,0.35)
set.seed(111)
barplot(
  prop.table(table(rbinom(10000, 1, 0.35))) * 100,
  main = "Probability of 10,000 Customers Buying the Product as a Percent",
ylab = "Percent", names.arg = c("Don't Buy", "Buy"))
```