Answers to the survey question, “How much do you like coffee?” are given on a 1, 2, 3,

4, and 5 scale. Suppose the following distribution, in list form, is a reasonable model

for producing the data that you might see:

***y p*(*y*)**

1 0.25

2 0.15

3 0.20

4 0.10

5 0.30

Total 1.00

A. Show how to write Pr(*Y* > 3) using the summation and set notation formula; then

show how to calculate that probability using that formula.

reals  = {1,2,3,4,5}

{k member of reals | k > 3}

: - p (4) + p (5)

Thus, Pr (Y member of k): - 0.1 + 0.3

Hence, Pr (Y member of k): - 0.4

B. Show how to write Pr (Y ≥ 3) using the summation and set notation formula; then

show how to calculate that probability using that formula.

reals  = {1,2,3,4,5}

{k member of reals | k > = 3}

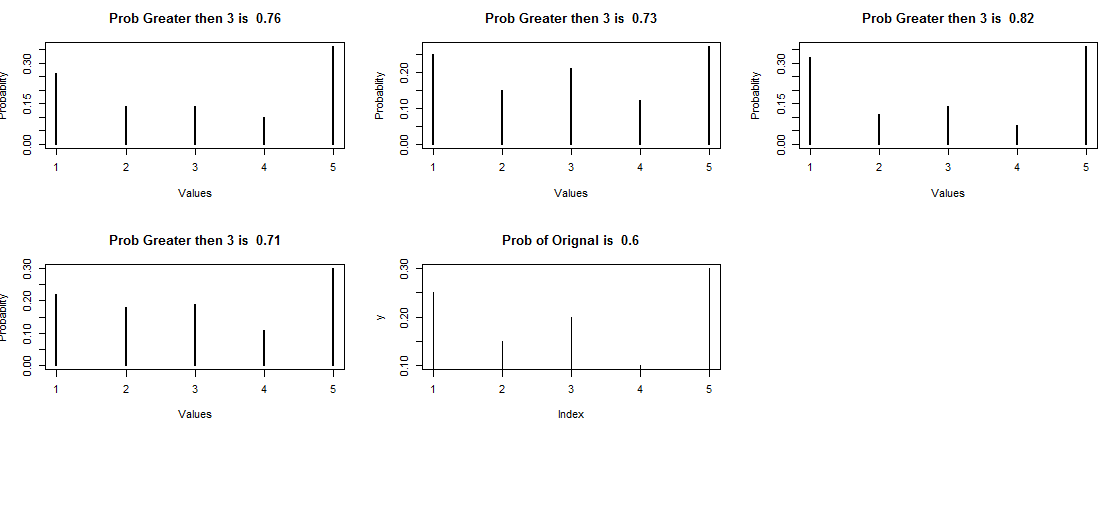
: - p (3) + p (4) + p (5)

Thus, Pr (Y member of k): - 0.2 + 0.1 + 0.3

Hence, Pr (Y member of k): - 0.6

C. Use simulation to estimate the probability in Exercise 3.6B. Compare and contrast

your simulation-based result with your answer to Exercise 3.6B.



Comparison of Different probabilities which are randomly generated and Original Data.

Consider the following function *p*(*y*):

***y p*(*y*)**

0.00 0.50

0.50 0.00

1.20 0.10

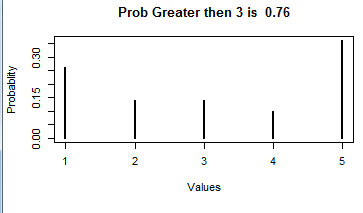
1.70 0.20

1.90 0.15

1.95 0.05

Total 1.00

1. Draw a graph of this function using the computer.



1. Show that this is a valid pdf.

Submission of all the probabilities equal to 1. And individual probabilities between 0 and 1.

1. Find Pr(Y < 1.70).

Y = p(y)

A = {Y < 1.7}

A = {0.0, 0.5, 1.2}

1. Find Pr(Y = 1.95).
2. Find Pr(0.50 < Y < 1.90).
3. Find Pr(0.50 ≤ Y ≤ 1.90).
4. Find Pr(|Y − 1.2| < 0.5).

Here Y ∊ A means. (Y<1.7 and Y>0.7)

H. Find Pr(|Y − 1.2| > 0.5).

Here Y ∊ A means. (Y<1.7 and Y>0.7)

3.8 Hans is investigating effectiveness of banner ads. He asks the participants in the study to mark their intention to click on the banner ad by moving their mouse cursor to a position between 0 and 100 on a specially designed web page. Suppose that the distribution of the continuous responses is p(y) = 0.0002y, for 0 ≤ y ≤ 100, as in Example

2.10.A. Show how to write Pr(Y > 50) using the integral formula; then show how to calculate that probability using the integral formula.

=





=0.75

B. Show how to calculate Pr(Y > 50) as the difference between the areas of two triangles.



=1

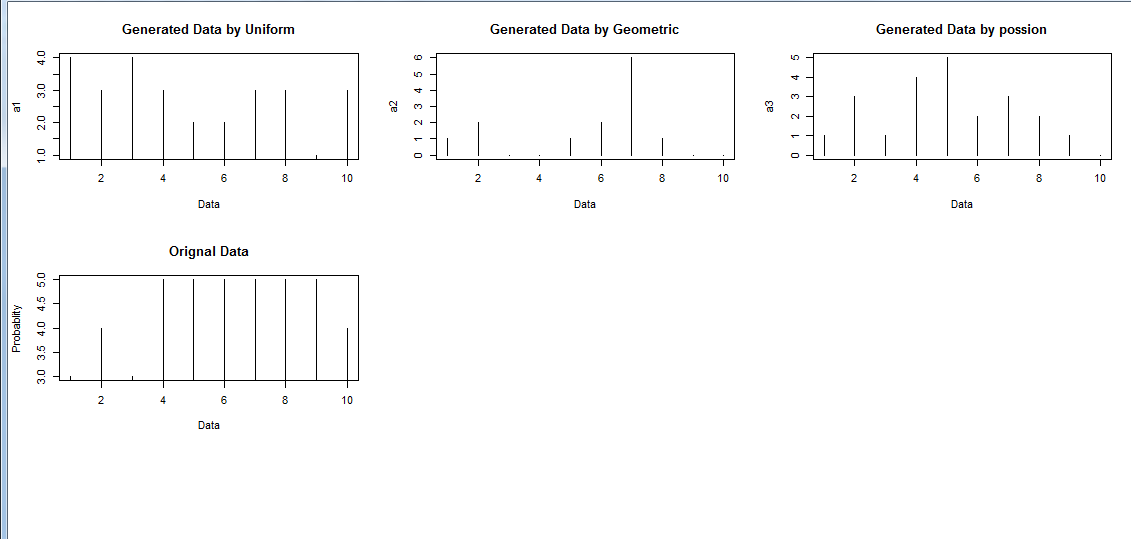
Difference between the areas of the Triangle: - 1-0.25 = 0.75

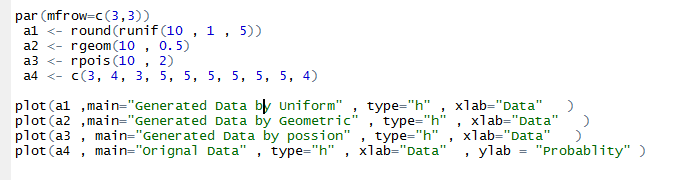
4.1 Data from surveys are often recorded on a 1, 2, 3, 4, 5 scale. Here are some responses from n = 10 people surveyed by an automobile retailer regarding their level of customer satisfaction: 3, 4, 3, 5, 5, 5, 5, 5, 5, and 4.

A. Using the computer, draw three distinct distributions that could have produced these data. Try not to make the distributions too similar to one another. Be realistic:

It is possible that there are dissatisfied customers. You just didn’t

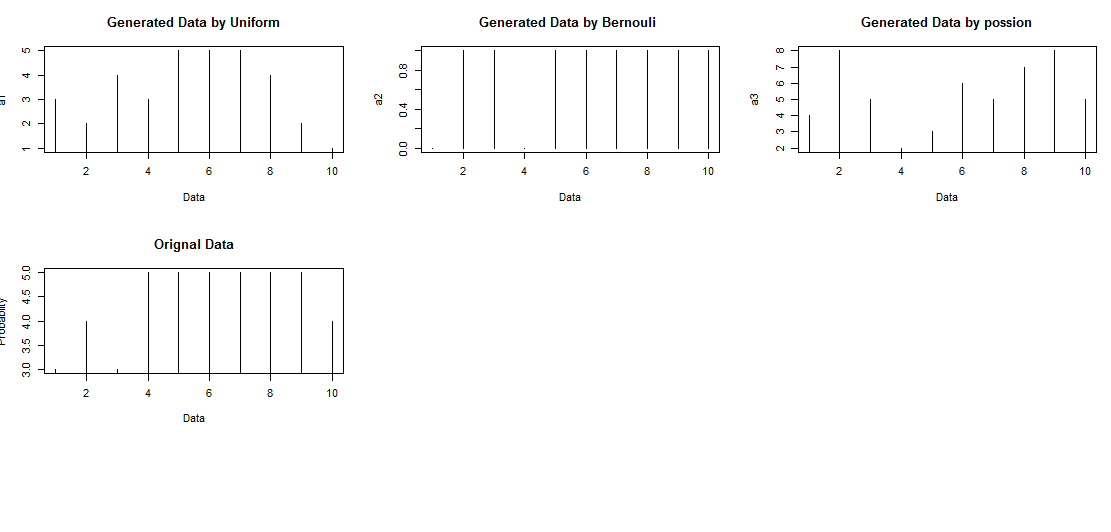
happen to see any in this particular data set.





Generated the Data by Different models.

B. Using the computer, draw three distinct distributions that most likely could not have produced these data.



None of the Data Generated by the Models are likely to be similar by Original Data.

C. Write down the generic distribution p(y) for these data in table form, in terms of unknown parameters, and explain why this distribution is more believable than any of the distributions you drew in Exercise 4.1A or B.

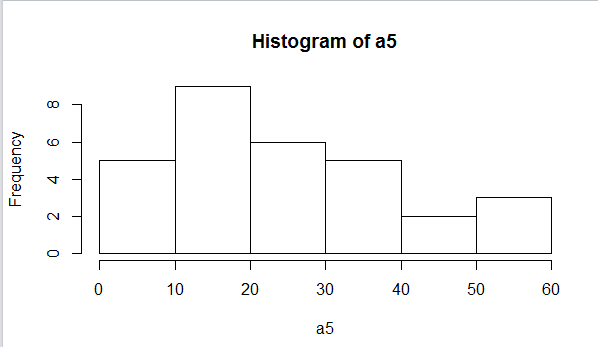
Original Data - 3, 4, 3, 5, 5, 5, 5, 5, 5, and 4.

|  |  |  |
| --- | --- | --- |
| Y | count | p(y) |
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 2 | 0.2 |
| 4 | 2 | 0.2 |
| 5 | 6 | 0.6 |

The Distribution produces data which has the same frequencies as the original data. The Earlier model created data which has not the same frequency as Original data.

4.7) The following data were obtained from the Internet data and story library, or DASL, with URL http://lib.stat.cmu.edu/DASL/ at the time of writing this book. The data are scores given by taste testers on cheese, as follows: 12.3, 20.9, 39, 47.9, 5.6, 25.9, 37.3,21.9, 18.1, 21, 34.9, 57.2, 0.7, 25.9, 54.9, 40.9, 15.9, 6.4, 18, 38.9, 14, 15.2, 32, 56.7, 16.8, 11.6, 26.5,0.7, 13.4, and 5.5.

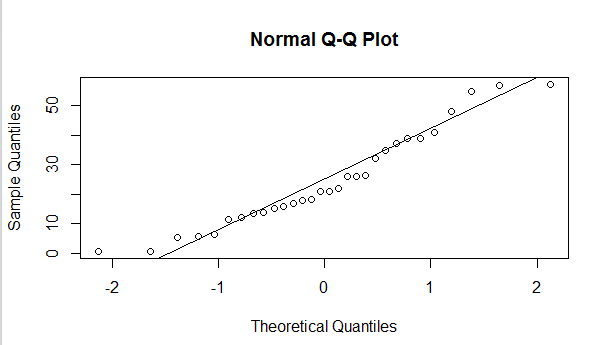
A) Construct and interpret the histogram of the data.



B. Create a table like Table 4.3 for these data.

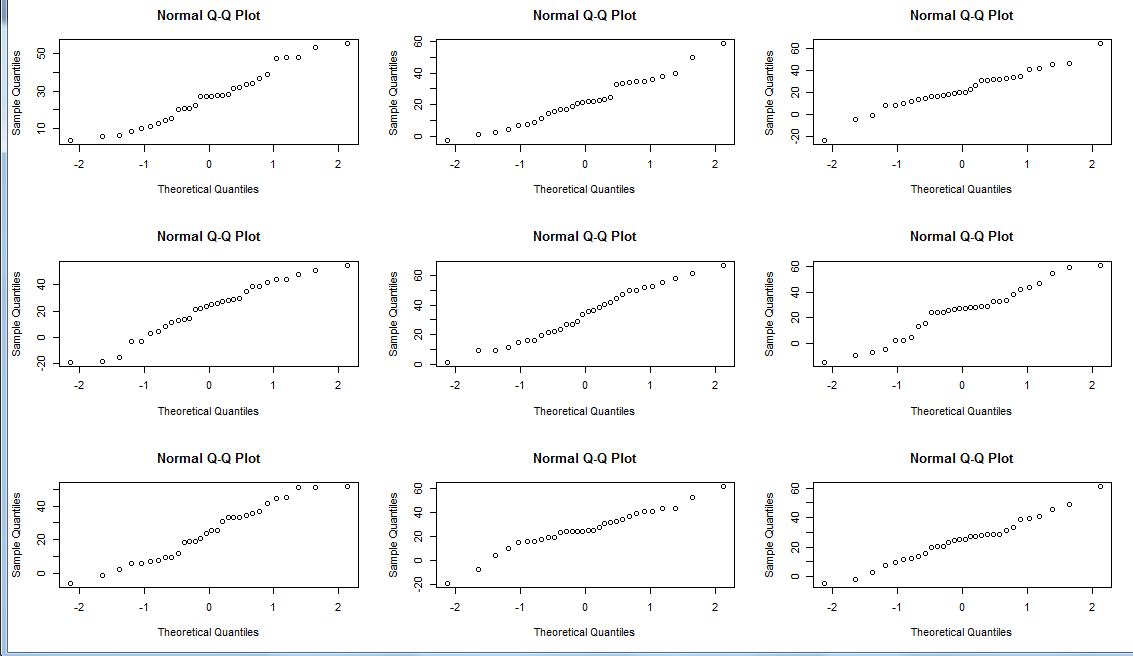
|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **i** | **y** | **p= i-0.5/n** | **norm.inv** |
| 1 | 12.3 | 0.016667 | -9.47616 |
| 2 | 20.9 | 0.05 | -1.75476 |
| 3 | 39 | 0.083333 | 2.429754 |
| 4 | 47.9 | 0.116667 | 5.484778 |
| 5 | 5.6 | 0.15 | 7.967794 |
| 6 | 25.9 | 0.183333 | 10.1043 |
| 7 | 37.3 | 0.216667 | 12.00966 |
| 8 | 21.9 | 0.25 | 13.75165 |
| 9 | 18.1 | 0.283333 | 15.37398 |
| 10 | 21 | 0.316667 | 16.90689 |
| 11 | 34.9 | 0.35 | 18.37258 |
| 12 | 57.2 | 0.383333 | 19.78813 |
| 13 | 0.7 | 0.416667 | 21.16735 |
| 14 | 25.9 | 0.45 | 22.52193 |
| 15 | 54.9 | 0.483333 | 23.86221 |
| 16 | 40.9 | 0.516667 | 25.19779 |
| 17 | 15.9 | 0.55 | 26.53807 |
| 18 | 6.4 | 0.583333 | 27.89265 |
| 19 | 18 | 0.616667 | 29.27187 |
| 20 | 38.9 | 0.65 | 30.68742 |
| 21 | 14 | 0.683333 | 32.15311 |
| 22 | 15.2 | 0.716667 | 33.68602 |
| 23 | 32 | 0.75 | 35.30835 |
| 24 | 56.7 | 0.783333 | 37.05034 |
| 25 | 16.8 | 0.816667 | 38.9557 |
| 26 | 11.6 | 0.85 | 41.09221 |
| 27 | 26.5 | 0.883333 | 43.57522 |
| 28 | 0.7 | 0.916667 | 46.63025 |
| 29 | 13.4 | 0.95 | 50.81476 |
| 30 | 5.5 | 0.983333 | 58.53616 |

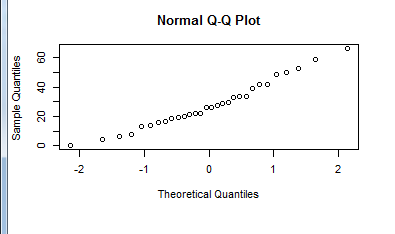
C. Construct and interpret the normal q–q plot of the data.



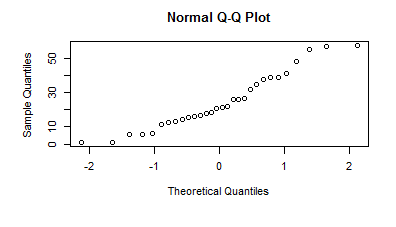
D. Simulate a data\* set from a normal distribution having the same mean, standard deviation, and sample size as for the taste variable. Construct the histogram and q–q plot of the simulated data\*. Repeat nine times, getting a total of 10 histograms and 10 q–q plots. You now have an idea of the effect of chance variation on the histograms and q–q plots. Without doing any formal test, do the histogram and q–q plot of the original data differs from normal by amounts that are explainable by chance alone? Discuss.

Q-Q Plot



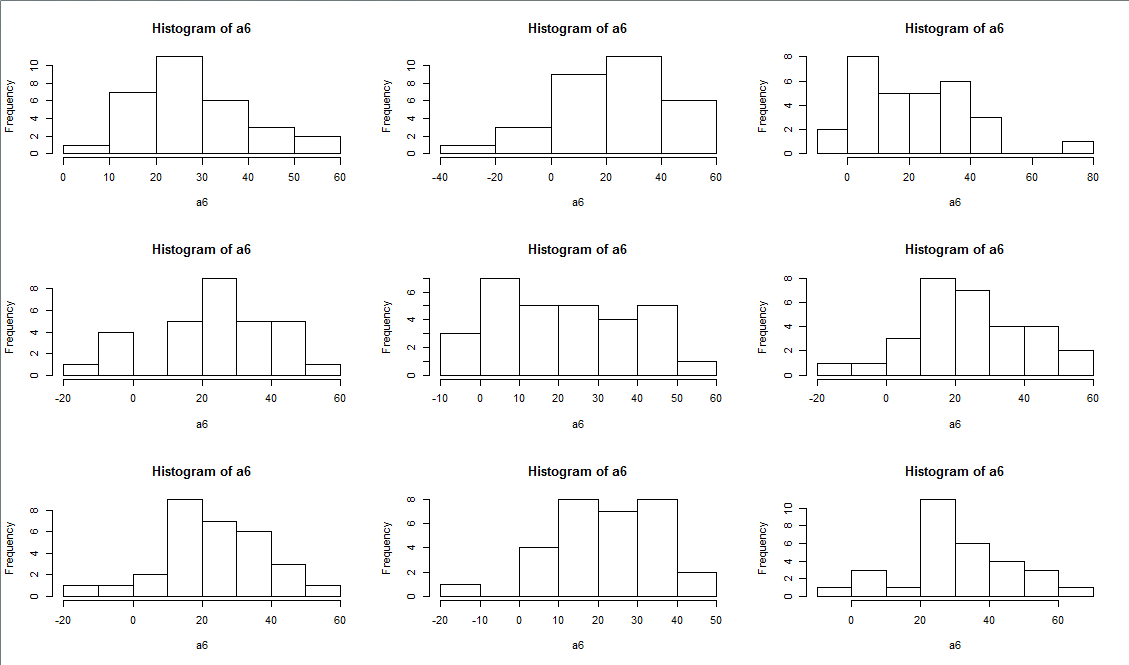


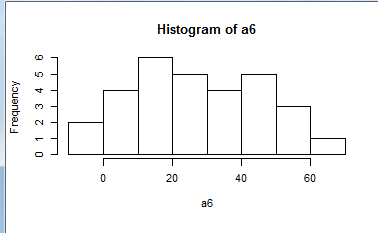
Original Data



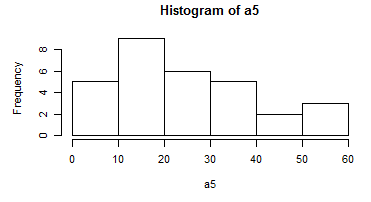
The values in the Random generated plot differs from the Original data in Norm QQ plot.

Histogram Plot





Original



The values in the Random generated plot differs from the Original data in Histogram plot.

4.10 Hans spins a top and lets it fall and settle. He had marked a location on the top in

indelible ink; if the top settles with the mark perfectly at the peak, then Hans records

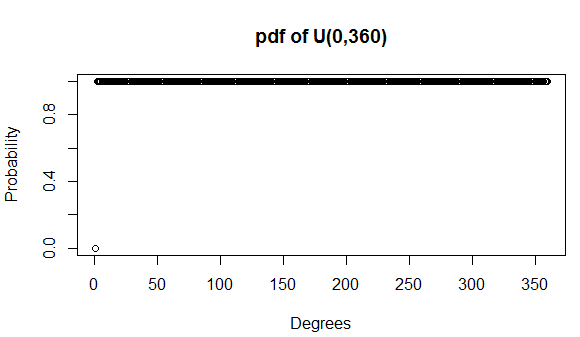
Y = 0. If the top settles with the mark perfectly at the bottom, then he records Y = 180.

Otherwise, he records Y, the angle in degrees from the peak, where the mark appears.

The range of Y is 0 ≤ Y < 360°.

A. Explain why the pdf p(y) that produces Y might be the U (0, 360) distribution, and

draw a graph of p(y). Label and number all axes.



The values Equally lies between 0 and 360, that’s why it is Uniform Distribution.

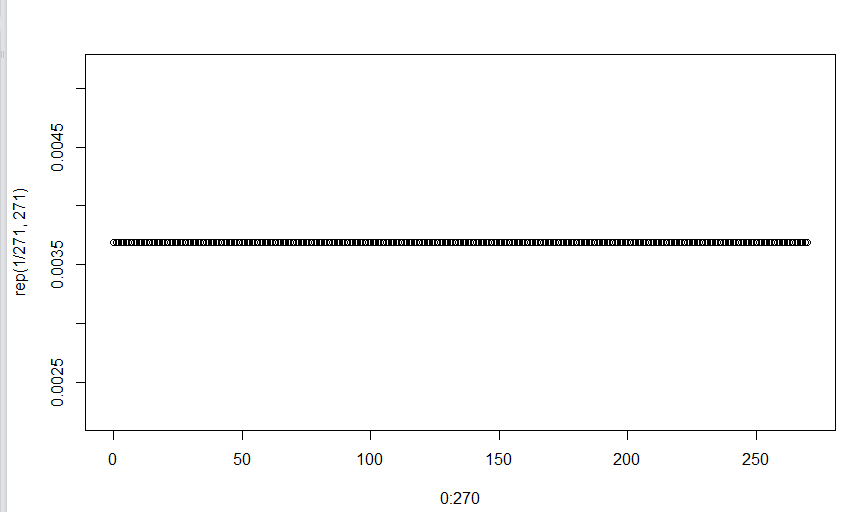
B. Explain a circumstance where the U (0, 360) pdf might be wrong, and draw a

graph of a possible p(y) that might be more correct in this circumstance. Label

and number all axes.

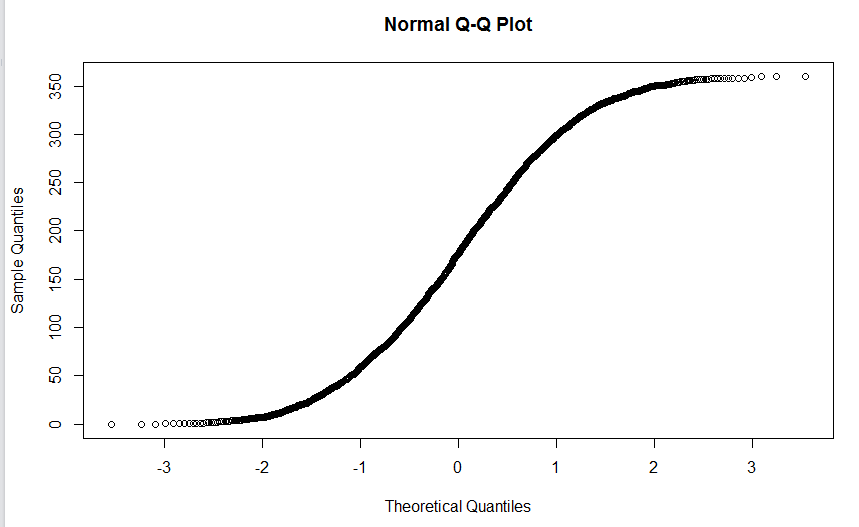
In the above Scenario where the top is not fair, that is one side is slightly heavier than that of the other one, so that the unmarked side won’t come at all.





C. Suppose you decide to c]heck whether the U (0, 360) model is reasonable using a

q–q plot. Describe Nature, design, measurement, and DATA for this study.



**Nature**: - Settlement and Spins of the Top is the nature of the U (0, 360) Model.

**Design: -** It is the plan to collect the data, In the above case the Collection of data is Determined by the “runif (2500,0,360)”, In the Above Scenario the Spinning and the measuring the angle of the of the Top is considered as Design

**Measurement: -** Measurement in the above scenario is the units in which we plan to collect the Data that is Degree in which we collect the Data

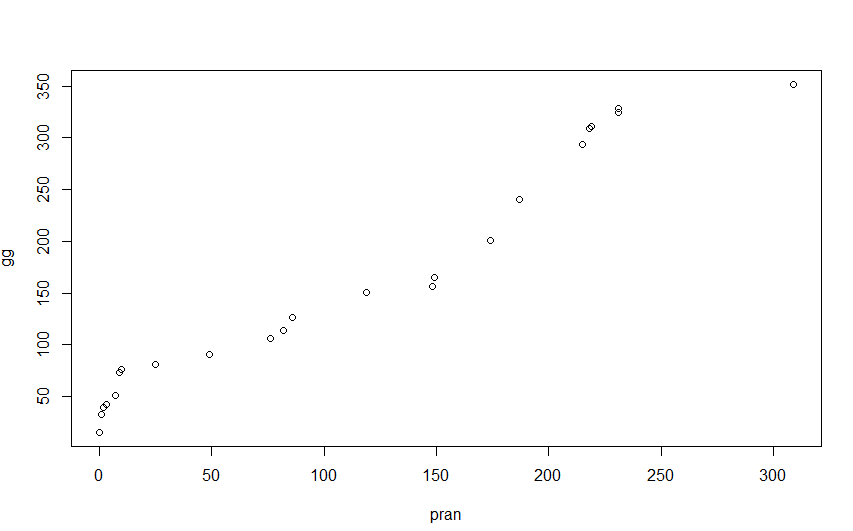
**DATA: -** The Output generated by the Q-Q plot that is the Data, any angle between 0 to 360 is known as the DATA

D. Suppose Hans’ data set from Exercise 4.10C is 149, 174, 309, 1, 82, 9, 218, 231, 49,

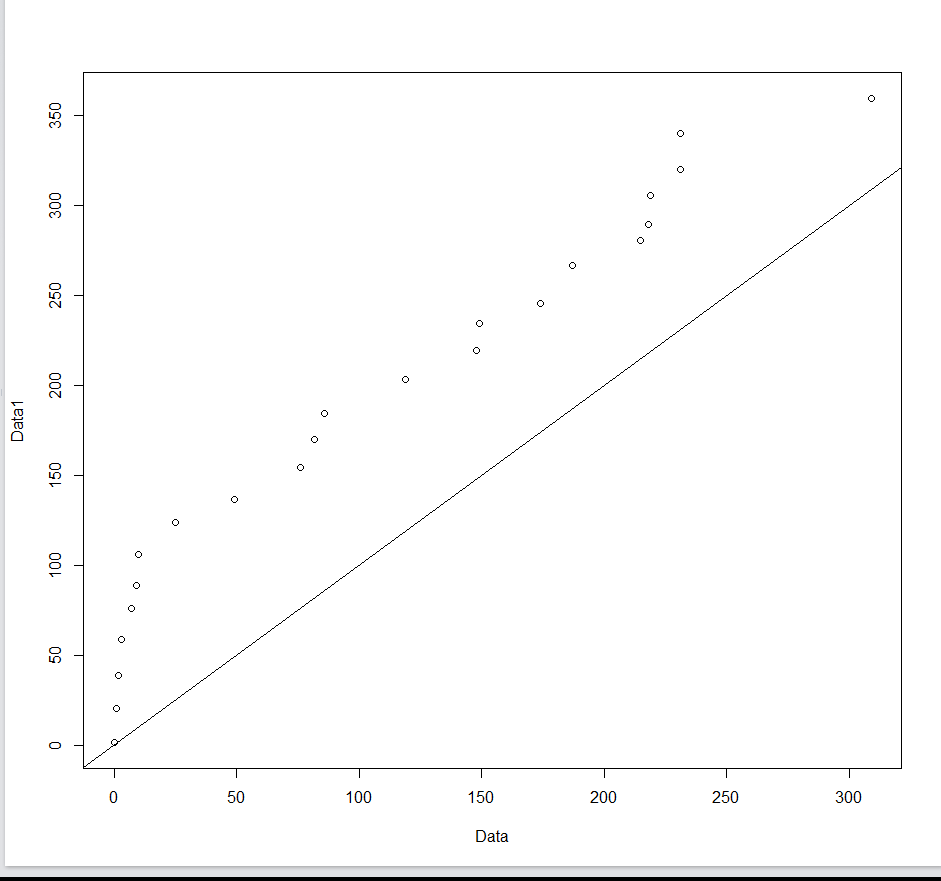
76, 0, 219, 215, 119, 148, 187, 231, 7, 2, 3, 10, 25, and 86, all degrees from the peak.

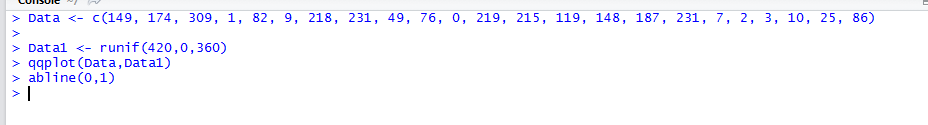
Construct the q–q plot for checking whether the U (0, 360) model might have produced

Hans’ data and discuss.



Constructed the data using the





Generate the Data by Hans Data set by putting the value in the vector.

Also Generate the Random data by runif Command and taken the sample of 420, with the data between 0 to 360

Plotted the simulated data on Y axis and the given data on x axis.

Plotted the abline 0 and 1 where 0 is intercept and 1 is slope of the above Data.

Plot does not produce the straight line, hence model does not produce Hanes data