EXERCISE III: PROBABILITY DISTRIBUTIONS

- **TOPICS COVERED:**
  1. Density Curves (mean and median)
  2. Normal Distribution
  3. Standardizing normal (z-score) and interpretation
  4. Assessing normality (histogram and QQ-plot)

- **REQUIRED DATASET: IPUMS-DHS**
- **REQUIRED VARIABLES:**
  1. WEALTHQ (household wealth index in quintiles)
  2. AGE (age)
  3. BIRTHWT_01 (birth weight in kilos of last birth)
  4. SAMPLE (country and year)
  5. PERWEIGHT (sample weight for persons)

[The only preselected variables that are needed in this exercise are SAMPLE and PERWEIGHT. Remove all of the other preselected variables by unchecking the purple boxes next to them. This will reduce the size of your data file and also make it easier to view the data in R.]

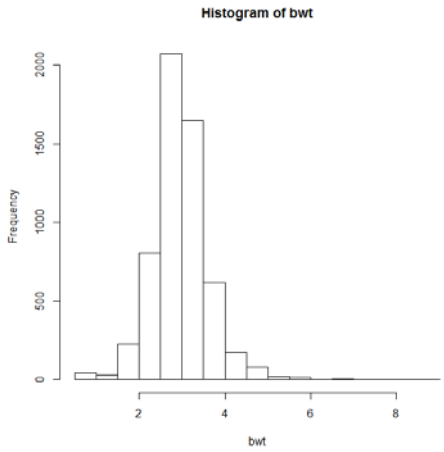
- **RECOMMENDED SAMPLES:**
  1. Benin [2011]
  2. Guinea [2012]
  3. Niger [2012]

- **UNIT OF ANALYSIS: Women**

***Make sure to eliminate the missing data for BIRTHWT_01. Also, change the BIRTHWT_01 variable from grams to kilograms by dividing by 1000.**

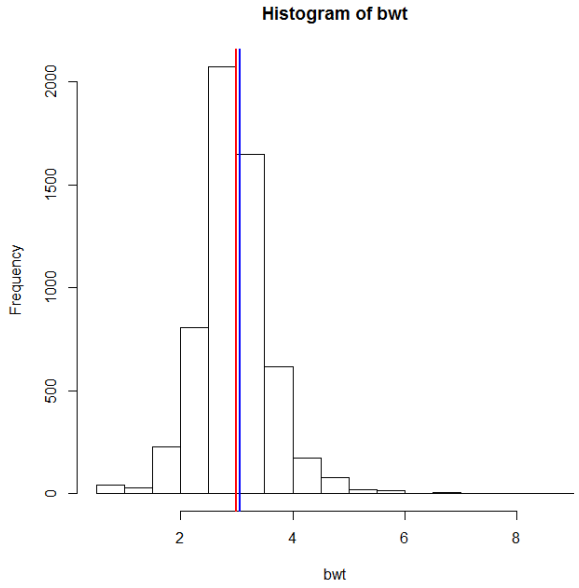
❖ Question 1

A. Create a histogram of the variable BIRTHWT_01 for Benin 2011 and describe its shape.



Shape: unimodal, pretty symmetric maybe very slightly right-skewed

B. Now, add vertical lines of the mean (blue) and median (red) to your graph, and write down the values of the mean __3.056669__ and the median __3__.



❖ Question 2

Using Benin 2011 as your sample, what is the z-score of a last-born child (var: BIRTHWT_01) who weighed (A) 1 kg? (B) 3 kg? (C) 6 kg? For each of these cases, explain in your own words what that number means.

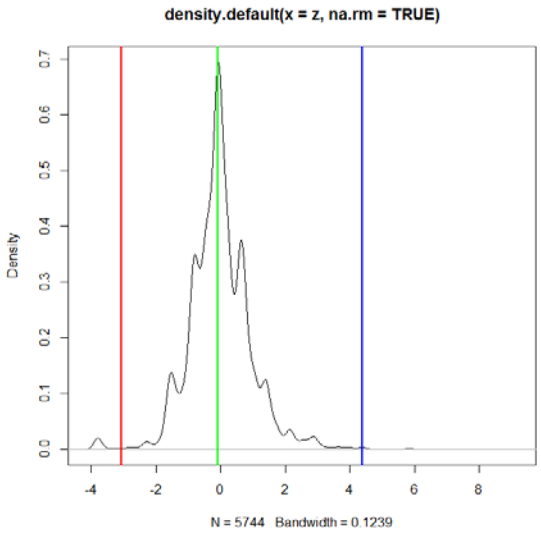
1 kg: z-score = -3.061096
Interpretation: A woman whose last-born child weighed 1 kg is -3.061096 standard deviations away from the mean.

3 kg: z-score = -0.08434408
Interpretation: A woman whose last-born child weighed 3 kg is -0.08434408 standard deviations away from the mean.

6 kg: z-score = 4.380785
Interpretation: A woman whose last-born child weighed 6 kg is 4.380785 standard deviations away from the mean.

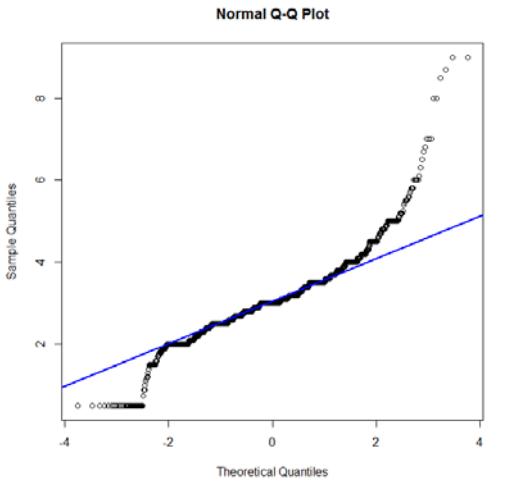
❖ Question 3

Create a density plot of the z-scores for variable BIRTHWT_01 (Benin 2011). Add vertical lines of the 3 z-scores to your graph.



❖ Question 4

Create a normal QQ-plot for variable BIRTHWT_01 (Benin 2011). Make sure to add the qqline to your graph. Does the variable BIRTHWT_01 follow a normal distribution? Explain.

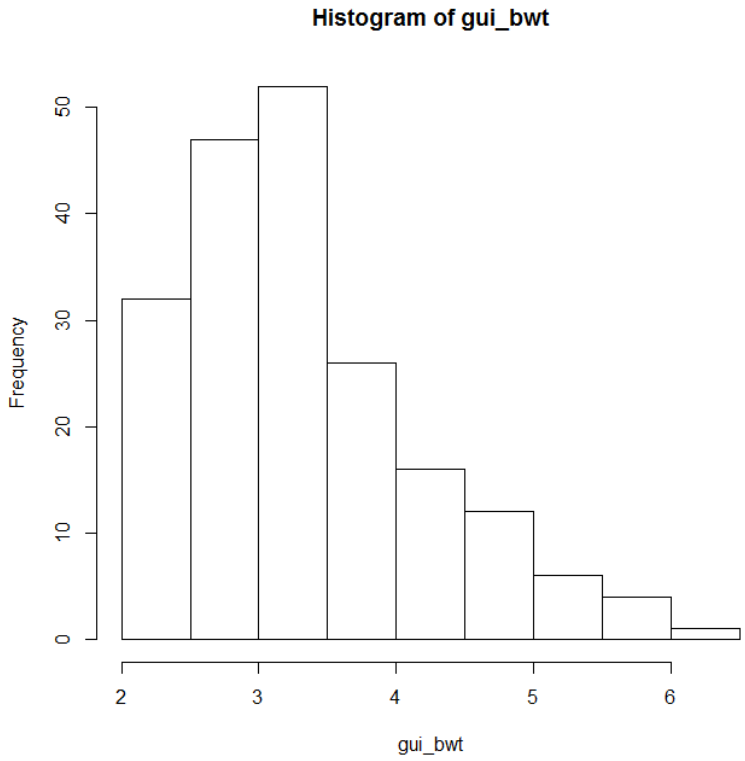


Commented [1]: Great tool for them to learn!

It does not appear to follow a normal distribution because most of the observations on the ends deviate away from the qqline on opposite sides suggesting that the data is too peaked in the middle.

❖ **Question 5**

Now create a histogram of the variable BIRTHWT_01 for Guinea 2012, limiting the sample to the poorest households. Describe the shape of this data.



Shape: right-skewed

❖ **Question 6**

A. **Using the data from the previous question, do you think that (A) the mean is greater than the median, (B) the mean is less than the median, or (C) that the mean and median are equal? Explain your reasoning.**

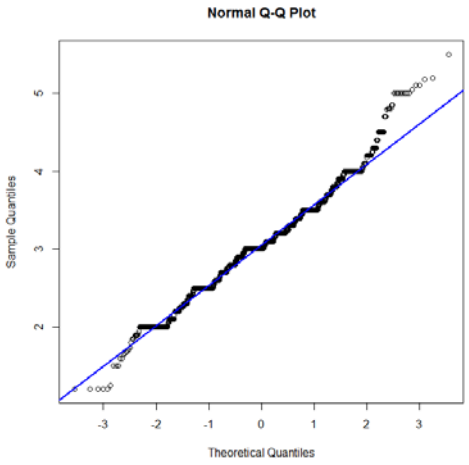
The mean is not resistant to outliers while the median is resistant to outliers. Since the data is slightly right-skewed, the mean is greater than the median.

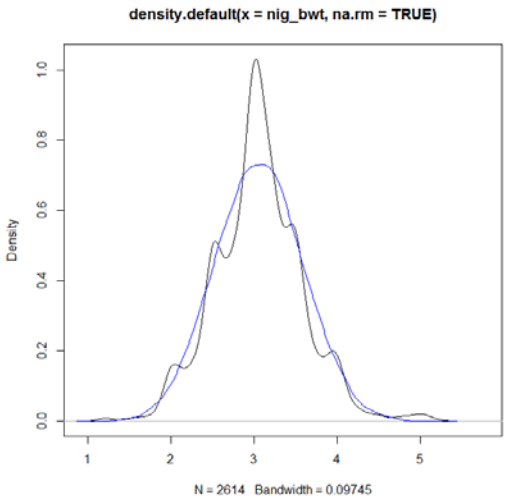
B. **Now, calculate the mean and the median values; did they match your original assumption?**

Mean: 3.462857
Median: 3.2

❖ Question 7

BIRTHWT_01 for Niger 2012 is approximately normal based on the qqplot below. Below the qqplot there is a density plot with the sample data in black and the normal data distribution in blue.





Use the normal data distribution to calculate the following: (Mean: 3.058/SD: 0.536)

A. **What is the probability that a woman's child from Niger 2012 weighs less than 2.25 kg?**

   0.0660031

B. **What is the probability that a woman's child from Niger 2012 weighs more than 3 kg?**

   0.543305

C. **What is the probability that a woman's child from Niger 2012 weighs between 2.5 kg and 3.5?**

   0.6456395

❖ Question 8

Using the data from the previous question, calculate:

A. The 25th percentile:

2.696385 kilograms

B. Calculate the 60th percentile

3.194338 kilograms

C. Calculate the 97th percentile

4.067761 kilograms

❖ Question 9

Suppose there is a normal distribution curve with a mean of 5 and a standard deviation of 3. A scientist realized there was a mistake in the calculations, and all of the values needed to be added by 3. How would the normal distribution curve change?

a) The curve would shift left.
b) The curve would shift right.
c) The curve would stretch out further and become flatter.
d) The curve would become narrower and become taller.

❖ Question 10

Suppose there is a normal distribution curve with a mean of 8 and a standard deviation of 2. A researcher realized he made a mistake, and the standard deviation was actually 4. How would the normal distribution curve change?

a) The curve would shift left.
b) The curve would shift right.
c) The curve would stretch out further and become flatter.
d) The curve would become narrower and become taller.