Brett Sutow

DSC 520

9/15/20

1. **What are the elements in your data (including the categories and data types)?**
   1. This data is broken down into 8 different category groups which include: id, id2, Geography, PopGroupID, POPGROUP.display-label, Races Report, HS Degree, Bachelor Degree. When looking at these category groups they can be broken down as the following data types:

Numeric: id, id2, PopGroupID, Races Report, HS Degree, Bachelor Degree

String: Geography, POPGROUP.display-label

1. **Please provide the output from the following functions: str(); nrow(); ncol()**
   1. > str(ACS2014)

'data.frame': 136 obs. of 8 variables:

$ Id : chr "0500000US01073" "0500000US04013" "0500000US04019" "0500000US06001" ...

$ Id2 : int 1073 4013 4019 6001 6013 6019 6029 6037 6059 6065 ...

$ Geography : chr "Jefferson County, Alabama" "Maricopa County, Arizona" "Pima County, Arizona" "Alameda County, California" ...

$ PopGroupID : int 1 1 1 1 1 1 1 1 1 1 ...

$ POPGROUP.display.label: chr "Total population" "Total population" "Total population" "Total population" ...

$ RacesReported : int 660793 4087191 1004516 1610921 1111339 965974 874589 10116705 3145515 2329271 ...

$ HSDegree : num 89.1 86.8 88 86.9 88.8 73.6 74.5 77.5 84.6 80.6 ...

$ BachDegree : num 30.5 30.2 30.8 42.8 39.7 19.7 15.4 30.3 38 20.7 ...

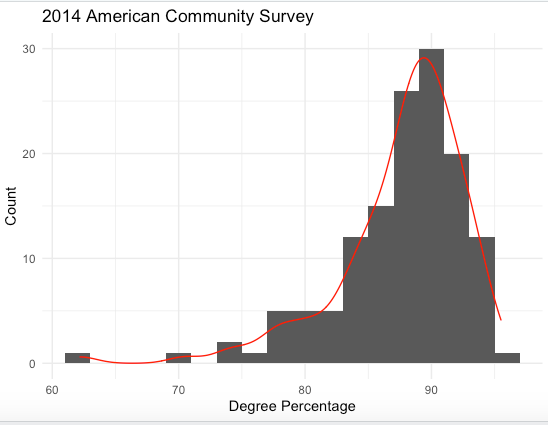
> nrow(ACS2014)

[1] 136

> ncol(ACS2014)

[1] 8

1. A picture containing clock

   Description automatically generated
2. **Answer the following questions based on the Histogram produced:**
   1. Yes, the data is unimodal because it has one clear peak in the graph between 80 and 90, peaking around 100.
   2. The graph is not approximately symmetric, it leans to the left with a tail and outliers on the left side of the graph.
   3. The graph is not approximately bell shaped, because it does not have a gradual increase/decrease to and from the mean of the histogram.
   4. No, it is not an approximately normal distribution, the graph leans to the left when looking at the scale.
   5. It is skewed to the left with a long tail.
   6. ****
   7. It cannot be used accurately for this model because of the skewed long tail of the left-hand side of the graph. This causes the normal distribution not to be accurate for this chart.
3. **A screenshot of a cell phone

   Description automatically generated**
4. **Answer the following questions based on the Probability Plot:**
   1. Yes, the plot is showing a close to normal probability plot. This is because it shows a semi straight line with most being centrally located in the middle of the distribution. With very few outliers on either side of the tail.
5. **A screenshot of a cell phone

   Description automatically generatedA screenshot of a cell phone

   Description automatically generated**
6. **In several sentences provide an explanation of the result produced for skew, kurtosis, and z-scores. In addition, explain how a change in the sample size may change your explanation?**
   1. The skew shows -1.67 which shows that tells that it is skewed to the left. The kurtosis of 4.35 tells us that it is not in the normal distribution, and not in the close to normally distributed range. Increasing the sample size may cause the distribution to even out creating a normal distribution curve.