3.2 Exercise

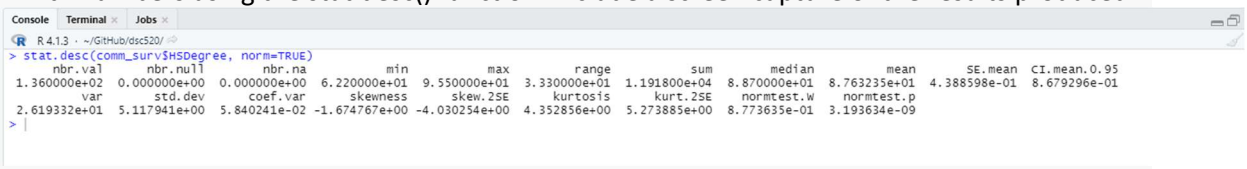
i. What are the elements in your data (including the categories and data types)?
   1. Data types: data frame, series, character, integer, numeric
   2. Categories: Id, Id2, Geography, PopGroupID, POPGROUP.display.label, RacesReported, HSDegree, BachDegree

ii. Answer the following questions based on the Histogram produced:
   1. Based on what you see in this histogram, is the data distribution unimodal?
      Unimodal
   2. Is it approximately symmetrical?
      . no
   3. Is it approximately bell-shaped?
      . no
   4. Is it approximately normal?
      . no
   5. If not normal, is the distribution skewed? If so, in which direction?
      . Negative skew (right)
   6. Include a normal curve to the Histogram that you plotted.
   7. Explain whether a normal distribution can accurately be used as a model for this data.
      . A normal distribution cannot accurately be used as a model for this data since the distribution is skewed.

iii. Create a Probability Plot of the HSDegree variable.

iv. Answer the following questions based on the Probability Plot:
   1. Based on what you see in this probability plot, is the distribution approximately normal? Explain how you know.
      . The distribution appears more normal than in the histogram because it looks fairly linear, with only a slight curve. However it is still not completely normal.
   2. If not normal, is the distribution skewed? If so, in which direction? Explain how you know.
      . The distribution is skewed to the left, since it curves to the left

v. Now that you have looked at this data visually for normality, you will now quantify normality with numbers using the stat.desc() function. Include a screen capture of the results produced.



vi. 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 results of the skew, kurtosis and z-scores can tell us a few things. The negative skewness means that the distribution is skewed to the right of the distribution. The positive kurtosis score tells us that the distribution is pointy and heavy-tailed. The absolute value of the z-scores for skew and kurtosis are both > 4, which mean there is a small chance of obtaining skew/kurtosis values that are more extreme than these ones by chance. This also means there is a significant skew/kurtosis in our sample. If we had other samples that measured things differently, we could also compare their z-scores to one another. A larger sample size would mean the z-scores would be less useful, since

the Standard Errors will be very small, so it would be better in that case to look at the data visually.