Project 2c – Descriptive Statistics Part 2

Kelly Williams

 Name for the new variable: Mother's Education recoded into 4 categories

Number of categories: 4

Corresponding labels: High School or less, Post-secondary and some college, College graduate, Some graduate school and above

The rationale behind this particular grouping is to combine similar levels of schooling into smaller more manageable sets. I combined all the respondents that have either attended high school or less into one category. The next grouping is dedicated to respondents that have experienced some education beyond high school but do not hold degrees. I kept the label college graduate the same since it was the original median of the first report. Lastly, respondents whom have attended school beyond college at the graduate level and above were combined into the final group.

1. Frequencies for each of the 4 categories

HS or less	102982
Post-secondary and some college	84152
College graduate	109217
Some grad school and above	78170

2. Frequency and percent of missing cases
   Frequency: 9294   percent: 2.4%

3. Frequency and percent of the most frequent response to this item
   Item: College grad   Frequency: 109217   Percent: 29.2%

4. Mean: 2.4341
5. Median: 3.00
6. Minimum: 1
7. Maximum: 4

Brief description:

This data set also describes the responses to a question regarding the mother's level of education of college freshmen from the year 1998. However, the information is displayed differently because it has been recoded into four categories**: high school or less, post-secondary and some college, college graduate, and some graduate school and above.** The highest frequency of responses remains college graduate with a frequency of 109217 and a valid percent of 29.2%. Coincidently, by

combining all of the responses that were related to any schooling less than high school, the frequency grows significantly in comparison with the college graduates. There were a total of 102982 responses for mothers whom had grammar school or less, some high school and high school diplomas. The lowest frequency in the first analysis was "grammar or less" but in this recoding the lowest frequency changes to "some graduate school and above" based on the groupings of the new four categories. I think that this data set shows that the majority of freshmen entering college in 1998 either came from homes of college graduates and the next largest group were first generation college students.

Evidence that the variable was recoded correctly:

As Dr. Devita suggested there are a few indicators to help determine if the variable was recoded accurately. First, in analyzing the total number of frequencies between the two reports I could see that they were equivalent. Also, the frequency and percent of the missing cases are also identical between the two reports. When comparing the median for this variable, one can see that the median, college graduate, remains constant between the two reports even though different values were assigned. Lastly, I could compute figures and percentages in comparison to the first report and find matches there as well.

8. Skewness value: .019
9. Skewness standard error range (+/- 2 standard deviations): -0.008   to +0.008
10. Kurtosis value: -1.331
11. Kurtosis standard error range (+/- 2 standard deviations): -.016   to +.016

Histogram with a normal curve:
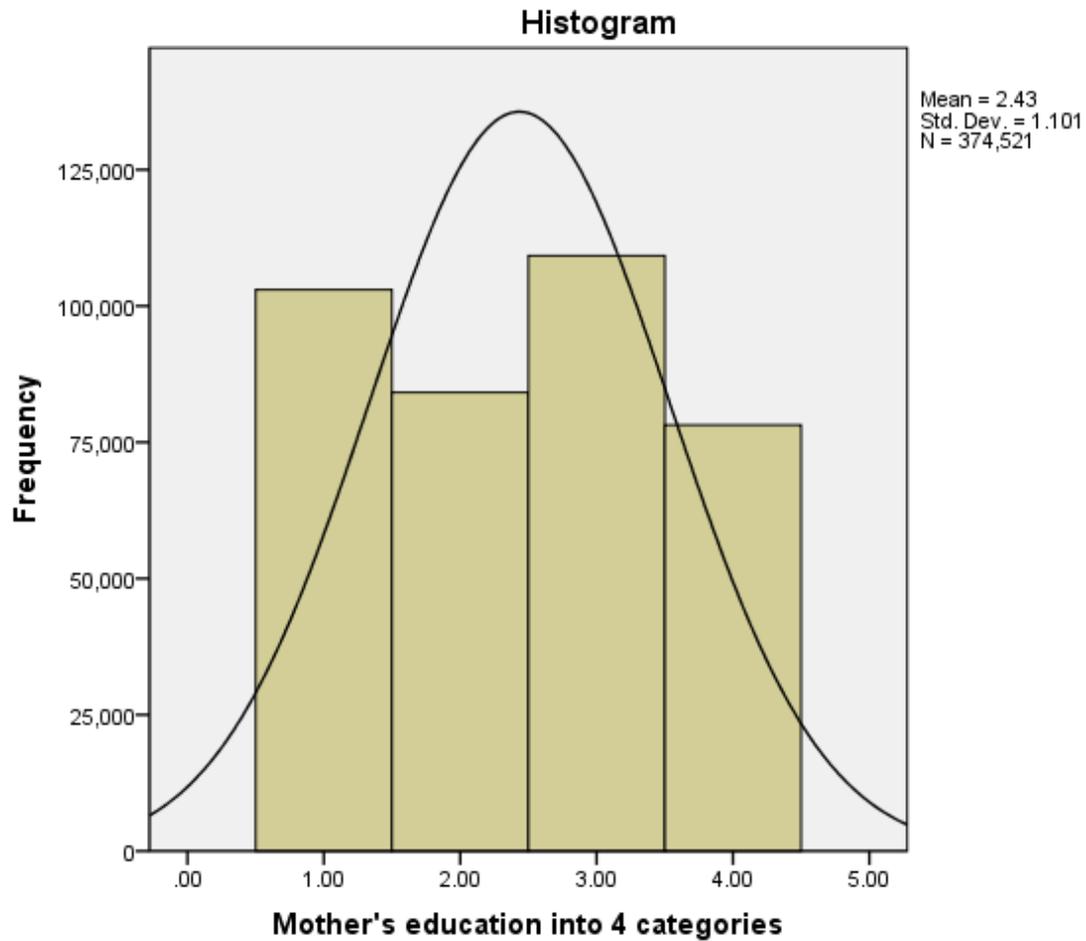
Based on the information provided from the analysis, are these data normally distributed? How can you tell? No, these data are not normally distributed. You can tell from two examples. First, there is a significant difference between the skewness value and the standard error of skewness. Since the skewness value does not fall within the range of the standard error of skewness it means that it is not normally distributed. Also, in analyzing the histogram you can visually see where the bars on the graph have points that extended outside of the normal curve.