

Assignment #1

The Bureau of Labor Statistics administers surveys each March known as the Annual Demographic Supplement to the Current Population Survey. The data set WORKERS is drawn from one of these Annual Demographic Supplements and includes data about 71,096 respondents. The data includes the age, highest education level, sex, total income, and job class of respondents aged 25 to 65 who have worked, with the stipulation that the subjects' principal work experience is not in agriculture.

Figure 1 separates the respondents according to their highest education level in a bar graph. The subjects are divided into six categories, each with a corresponding numerical value: a 1 corresponds to not reaching high school; a 2, to having some high school but no high school diploma; a 3, to having a high school diploma; a 4, to having some college but no bachelor's degree (including respondents with associate degrees); a 5, to having a bachelor's degree; and a 6, to having a postgraduate degree (master's, professional, or doctorate). Of 55,899 responses, 2,165 fell under the "1" category, 3,682 under 2, 17,704 under 3, 15,756 under 4, 10,991 under 5, and 5,601 under 6. The distribution in Figure 1 is not symmetric, thus it is not normal. It is, however, unimodal: the "3" category applied to the greatest number of respondents.

According to the distribution in Figure 1, a high school diploma is the single most common level of education among the people surveyed. In fact, the number of subjects whose highest level of education is a high school diploma is three times the sum of respondents who either never reached high school or only had some high school (categories 1 and 2). Although a high school diploma is the most common level of education, more than half of the respondents (approximately 57.87%) had at least some college, possibly indicating the desirability of a college education. According to Figure 1, however, the proportions of respondents in categories

4, 5, and 6 become progressively smaller. This could be because for each successive level of education a person wishes to pursue, more time is required to complete the degree(s) desired. A separate category between 5 and 6 (for those with some graduate school but no graduate degree) may be helpful in further consideration of this possibility.

Figure 2 is a histogram of 91 intervals showing the earnings, or total income (measured in dollars) of the respondents. The possible values include all real positive and negative numbers, as well as the number zero. The minimum value in the distribution is -24,998; the first quartile is 17,000; the median is 29,717; the third quartile is 46,505; and the maximum is 425,510. The mean of the distribution is 37,865, the standard deviation is 36,158, and the interquartile range (IQR) is 29,505. The distribution is strongly skewed to the right, and not symmetric; therefore, it is not a normal distribution. The 1.5 x IQR rule indicates the presence of outliers on the right tail, further evidencing that the distribution is not normal. The distribution is, however, unimodal, with a peak between 17,500 and 22,500.

The distribution in Figure 2 has a strong skew to the right because the mean is not a resistant measure of center like the median is. Because of this, even a few high values (especially those such as the outliers) can pull the mean rightward. These high values may stem from a variety of factors. Certain jobs, especially those in professional fields such as medicine and law, can pay very high salaries. The subjects surveyed may also have supplemental sources of income, even if they only apply to the single year in question: lottery winnings, stocks sold for a profit, rented property, and inheritances are examples, though they by no means constitute an exhaustive list. The negative values, on the other hand, may be accounted for by debts accumulated over the course of the year that outweigh the total income. All in all, the data in Figure 1 and Figure 2 can reveal both patterns and abnormalities in the data they present.

Figure 1

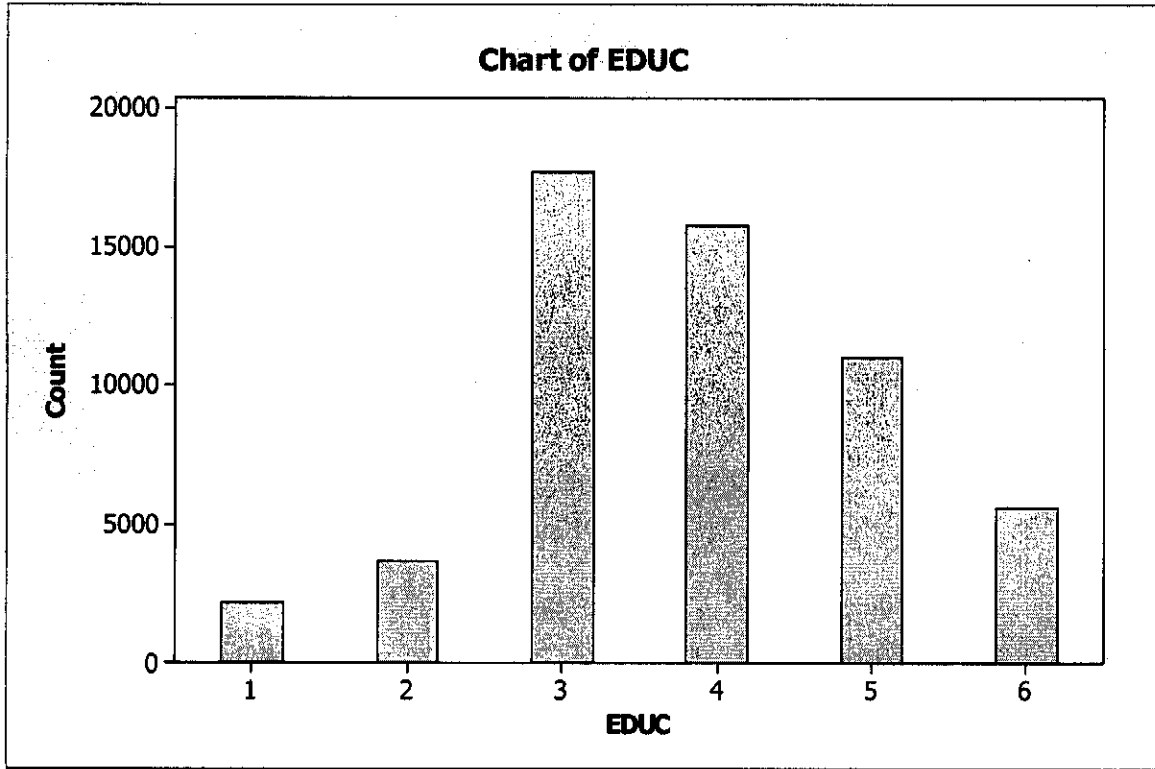


Figure 2

