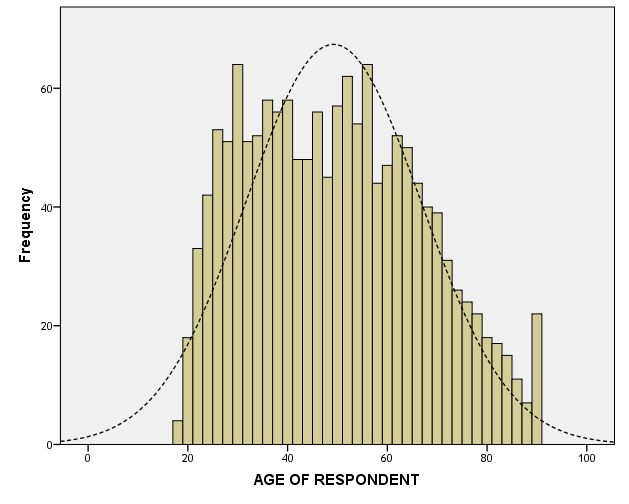
**CHAPTER 5 SPSS Problems SOLUTIONS**

\* Selected, but not all, output is shown below with some graphic modification.

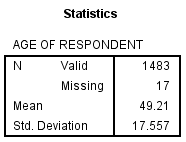
1.

a.

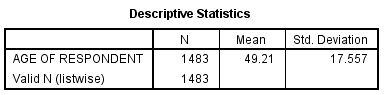


Looking at the histogram for AGE, we see that the distribution for age is close to a normal distribution when compared to the normal curve. However, because of several outliers at the upper end of the distribution, the distribution itself is slightly positively skewed.

b. Below is the output for the mean and standard deviation values using the Frequencies procedure.



Below is the output for the mean and standard deviation values using the Descriptives procedure.

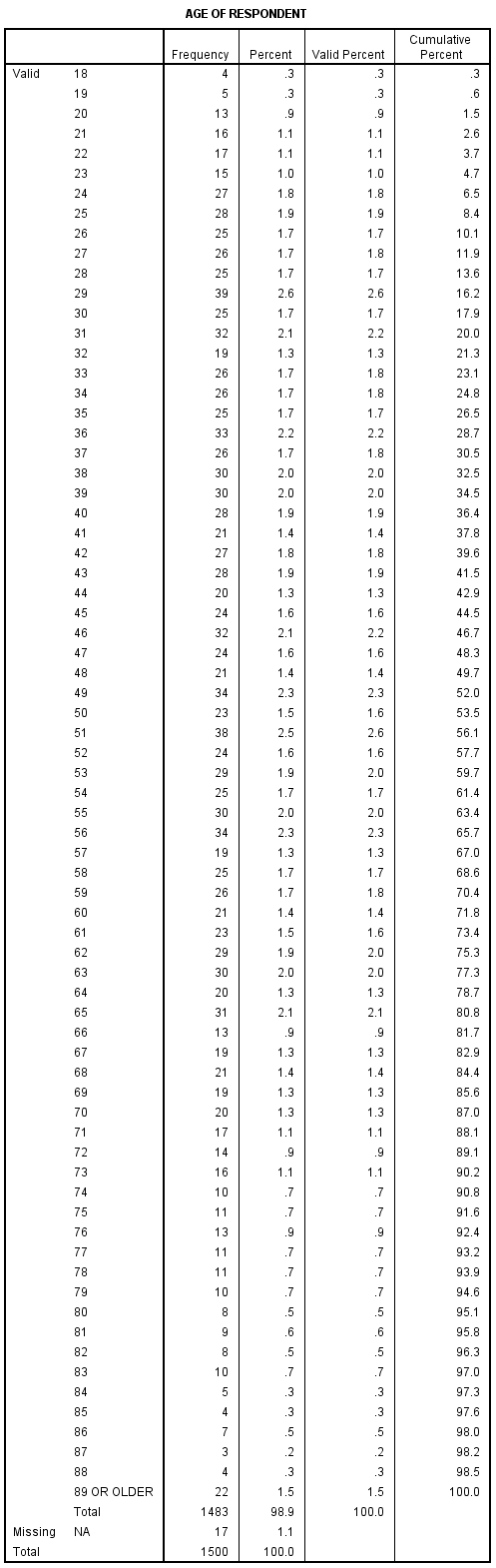


As depicted in the output, either procedure produces the same values for the mean (49.21) and standard deviation (17.557).

c. The *Z* score for the age of 25 is

From Appendix A, we see that the area beyond a *Z* score of -1.38 is 0.0838, or

8.38%. So, 8.38% of the distribution should be at or below age 25.

d. Below is a frequency distribution for AGE:

Looking at the output, we see that 8.4% of respondents are 25 years of age or younger. This is nearly identical to our theoretical calculation of 8.38%, even though the distribution is not precisely normal.

2.

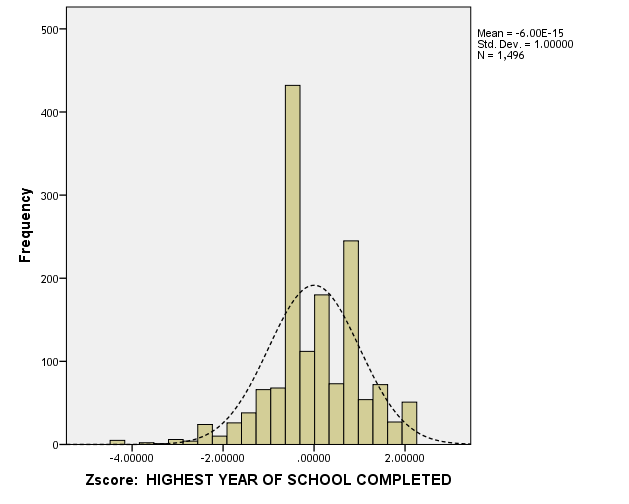
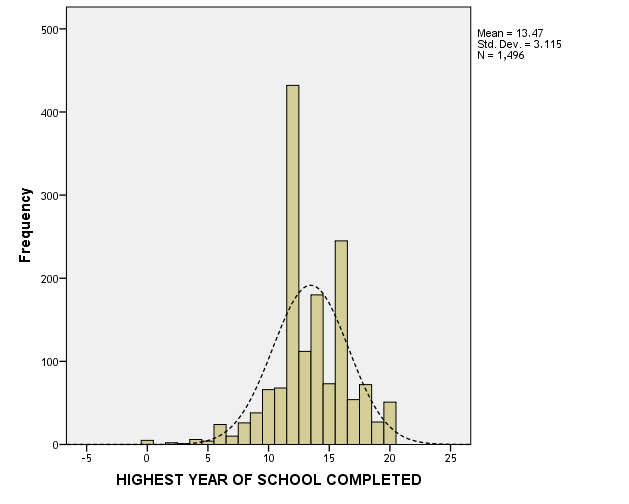
a. SPSS creates and saves *Z* scores using the Descriptives procedure and by selecting the “Save standardized values as variables” option.

b. The easiest way to find the equivalent *Z* score for 18 years of education is to first switch the SPSS Data Editor Window to “Data View” (bottom left portion of the screen). Next, we locate EDUC in the columns, right click the EDUC as the column header, and select the “Sort Ascending” option. After the scores have been stratified, we scroll down the EDUC column until we find where 18 years of education begins. Finally, once we locate these values, we simply scroll over to ZEDUC variable column and note that 1.45547 is the corresponding *Z* score with 18 years of education.

c. The Frequencies table shows that the cumulative percent of respondents with 18 years or less of education is 94.8 percent. Thus, the percentile rank is the same, 94.8.

d. The percentile rank of 94.8 is associated with a *Z* score of about 1.63 because we look for the *Z* score that has 1.00 – .948 = .052 of the distribution above it. SPSS calculated a *Z* score of 1.45547 for 18 years, a calculation based on the assumption that the distribution of hours worked per week is normal (even though that is untrue).

e. The distributions are identical. Because SPSS chose to use a different number of intervals for each variable, the histograms don’t appear identical, but they are. Transforming raw scores into Z scores changes the scale on which a variable is measured but doesn’t change its distribution (mathematically, subtracting and dividing by constants preserves the order and relative magnitude of the scores).



3. For this exercise, the instructor (or students) should select an appropriate EDUC value to calculate equivalent Z scores or to determine whether the distribution is normal for men/women and blacks/whites. Follow the same procedures as in 2a-e.

**CHAPTER 5 EXERCISE SOLUTIONS**

1.

a. The *Z* score for a person who watches more than 8 hr/day:



b. We first need to calculate the *Z* score for a person who watches 5 hr/day:



The area between *Z* and the mean is 0.2734. We then need to add 0.50 to 0.2734 to find the proportion of people who watch television less than 5 hrs/day. Thus, we conclude that the proportion of people who watch television less than 5 hrs/day is 0.7734. This corresponds to 783.45 (0.7734 × 1,013).

c. 5.66 television hours per day corresponds to a *Z* score of +1.



d. The *Z* score for a person who watches 1 hr of television per day is



The area between the mean and *Z* is 0.2764.

The *Z* score for a person who watches 6 hr or television per day is



The area between the mean and *Z* is 0.3708.

Therefore, the percentage of people who watch between 1 and 6 hr of television per day is 64.72% (0.2764 + 0.3708 = 0.6472 × 100).

2.

The statement is partly true. The use of *Z* scores is predicated on the assumption that the distribution being studied is normal in shape. *Z* scores can be calculated for any distribution, so a *Z* score can always be calculated for any raw score. However, a *Z* score will only correspond to a particular, known amount of area under the normal curve. We would need to multiply this area by the sample size in order to find out how many participants this area under the curve corresponds with.

3.

a. For an individual with 13.47 years of education, his or her *Z* score would be



b. Since our friend’s number of years of education completed is associated with the 60th percentile, we need to solve for *Y*. However, we must first use the logic of the normal distribution to find *Z*. For any normal distribution, 50% of all cases will fall above the mean. Since our friend is in the 60th percentile, we know that the area between the mean and our friend’s score is 0.10. Similarly, the area beyond our friend’s score is 0.40. We can now look in Appendix A column “B” for 0.10 or in column “C” for 0.40. We find that the *Z* associated with these values is 0.25. Now, we can solve for *Y*:



4.

a. Among working-class respondents:

The *Z* score for a value of 12 is



The *Z* score for a value of 16 is



The area between a *Z* of -0.28 and the mean is 0.1103. The area between a *Z* of 1.12 and the mean is 0.3686, so the total area between the scores is



So the proportion of working-class respondents with 12 to 16 years of education is 0.4789.

Among upper-class respondents:

The *Z* score for a value of 12 is



The *Z* score for a value of 16 is



The area between a *Z* of -1.16 and the mean is 0.3770. The area between a *Z* of 0.18 and the mean is 0.0714, so the total area between the scores is



So the proportion of upper-class respondents with 12 to 16 years of education is 0.4484.

b. Among working-class respondents:

The *Z* score for a value of 16 is



The area between a *Z* of 1.12 and the tail of the distribution (Column C) is 0.1314. So the probability of a working-class respondent having more than 16 years of education is 0.1314.

Among middle-class respondents:

The *Z* score for a value of 16 is



The area between a *Z* of 0.50 and the tail of the distribution (Column C) is 0.3085. So the probability of a middle-class respondent having more than 16 years of education is 0.3085.

c. Among lower-class respondents:

The *Z* score for a value of 12 is



The area between a *Z* of 0.15 and the mean (Column B) is 0.0596. To this, we must add 0.50 (the lower half of the distribution) to 0.0596. So the probability of a lower-class respondent having less than 12 years of education is 0.5596 (0.0596 + 0.50).

Among upper-class respondents:

The *Z* score for a value of 12 is



The area between a *Z* of -1.16 and the tail of the distribution (Column C) is 0.1230.

Remember, in this case, the fact that the *Z* score is a negative value tells us that we are working on the lower half of the distribution. So unlike our previous answer, we do not need to add 0.50. So the probability of an upper-class respondent having less than 12 years of education is 0.1230.

d. First, we find the *Z* score that has 25%, or 0.25, of the area between it and the mean. This is a *Z* score of about 0.68. The lower limit is



And the upper limit is



So the middle 50% of working-class respondents falls between 10.86 and 14.74.

e. If years of education is positively skewed, then the proportion of cases with high levels of education will be less than for a normal distribution. This means, for example, that the probabilities associated with high levels of education will be smaller.

5.

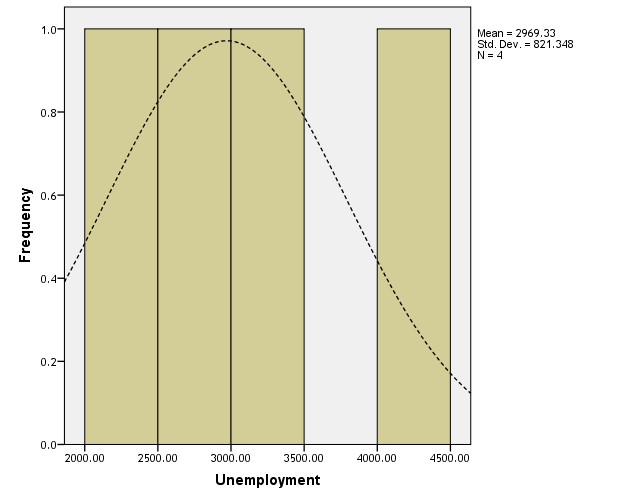
a. The mean and standard deviation are

b. One standard deviation above the mean is 2,969.33 + 1.0(821.35) = 3,790.68. One

region has unemployment above this value. We would expect, from Appendix A, about 15.9% of the distribution to lie above a *Z* score of 1.0. Then 15.9% of 4 regions is less than one region, which means that the number of regions with scores greater than 1 standard deviation above the mean is more than what we would expect from a normal distribution.

c. Because we only have four regions and each has a different unemployment number, the distribution is obviously not normal. Because the South has such a high number of unemployed persons, the distribution is technically positively skewed.



6.

a. An occupational prestige score of 60 corresponds to a *Z* score of



The area between a *Z* of 1.07 and the tail of the distribution is 0.1423. So about 14% of Whites should have occupational prestige scores above 60. This corresponds to approximately 157 Whites (0.1423 × 1,100) in our sample who should have occupational prestige scores above 60.

b. An occupational prestige score of 60 corresponds to a *Z* score of



The area between a *Z* of 1.47 and the tail of the distribution is 0.0708. So about 7% of Blacks should have occupational prestige scores above 60. This corresponds to approximately 14 Blacks (0.0708 × 195) in our sample who should have occupational prestige scores above 60.

c. An occupational prestige score of 30 corresponds to a *Z* of score of



The area between a *Z* of -1.08 and the mean is 0.3599.

An occupational prestige score of 70 corresponds to a *Z* score of



The area between a *Z* of 1.79 and the mean is 0.4633. So the proportion of Whites with occupational prestige scores between 30 and 70 is 0.8232 (0.3599 + 0.4633). Thus, approximately 906 Whites (0.8232 × 1,100) in the sample should have occupational prestige scores between 30 and 70.

d. An occupational prestige score of 30 corresponds to a *Z* of score of



The area between a *Z* of -0.83 and the mean is 0.2967.

An occupational prestige score of 60 corresponds to a *Z* score of



The area between a *Z* of 1.47 and the mean is 0.4292. So the proportion of Blacks with occupational prestige scores between 30 and 60 is 0.7259 (0.2967 + 0.4292). Thus, approximately 142 Blacks (0.7259 × 195) in the sample should have occupational prestige scores between 30 and 60.

7.

a. The area beyond the *Z* is about 0.1056, so 10.56% of students should score above 625.



b. The area between this score and the mean is 0.3413.





The area between this score and the mean is also 0.3413, so 68.26% of all students should score between 400 and 600 (or ±1 standard deviation from the mean of 500).

8.

a. About 0.1894 of the distribution falls above the *Z* score, so that is the proportion of crime incidents with more than 2 victims.



b. The area between the mean and the *Z* score is about 0.1331, so the total area above 1 victim is 0.50 + 0.1331 = 0.6331, or 63.31%.



c. The area between the mean and the *Z* score is about 0.4995, so the total area below 4 victims is 0.50 + 0.4995 = 0.9995.



9.

a. The area between the value and the upper tail of the distribution is 0.1020. So, the probability that someone will work more than 60 hours per week is 0.1020. This translates into approximately 85 (838 x 0.1020) respondents in the sample.



b. The area between the value and the lower tail of the distribution is 0.2420. So, the probability that someone will work less than 30 hours per week is 0.2420. This translates into approximately 203 (838 x 0.2420) respondents in the sample.



10.

a. For a team with an APR score of 975



From Appendix A, the area beyond 0.81 is 0.2090, or about the 79th percentile. The team is at the upper quartile because it is above the 75th percentile.

b. The *Z* value which corresponds to a cutoff score with an area of about 0.25 toward the tail of the distribution is 0.67. This is translated into a cutoff score by

Students should round off to the whole number that has a value closest to a *Z* of 0.67.

c. The *Z* value is 0.67.

11.

a. For the eligibility criterion, the team A has a *Z* score of



For the retention criterion,



Team A has a higher *Z* score on the eligibility criterion and therefore is better on that.

b. For the eligibility criterion, the team B has a *Z* score of



For the retention criterion,



Team B is better on the retention criterion.

c. Team B’s retention *Z* score was -0.21, below the mean. We simply look to “Column C” and find that the proportion of teams that did worse than Team B on the retention criterion is 0.4168. This is the area between Team B’s retention *Z* score and the tail end of the distribution.

12. For any *Z* distribution, the value of the mean is 0. The standard deviation of a *Z* distribution is 1. *Z* distributions are based on the mean of a variable and are centered on that value, so they have a mean of 0 by definition. A *Z* score of 1 or -1 is equivalent to a score in the original distribution that is 1 standard deviation above or below the mean, respectively. This direct mapping from the original distribution to a *Z* score means that the standard deviation of a *Z* distribution must be equal to 1.

13.

a. The 95th percentile corresponds to a *Z* score of about 1.65. Translating this into a raw score for the number of women needing shelter yields

women

Unfortunately, a capacity of 350 is below this value, so there will not be enough space for all abused women on 95% of all nights. Obviously, the city needs at least 374 beds.

b. The area below this value is 0.3446, so the area exceeding this *Z* is 1 – .3446 = 0.6554. Or 65.54% of all nights the number of women seeking shelter will exceed the capacity of 220.



14. The Z score for sociologists is

z = = = 1.19

The area above this Z score is .1170. Therefore, .1170 of people have higher occupational prestige scores than sociologists.

15. a. The Z score with only .01 of the population above it (in column C) is 2.33.

65,628.80 + 2.33(59,373.92) = 203,970.03

So, a household would have to have made $203,970.03 in the past year to have had more income than 99% of other houses.

b. z = = = -1.11

The area below this Z score is .1335. Our classmate has a point: it is problematic to assume the distribution is normal, when that would entail that 13.35% of people come from households making less than $0 in the past year.