**CHAPTER 4 SOLUTIONS**

1. 15.
2. 15.
3. 1.14
4. 11.
5. 5.
6. 1.14.
7. -10.
8. 10.
9. -1.14.
10. -1.
11. 5.
12. 1.14.
13. The average time for the fourth graders was 20 minutes. The average time for the sixth graders was 15 minutes. Thus, the sixth grade students were faster. The standard deviation for the fourth graders was 7 minutes. The standard deviation for the sixth graders was 5.4 minutes. Thus, the sixth grade students were more similar to each other in the amount of time taken to complete the task.
14. Adjusting the sixth grade scores so that the time spent reading the instructions was not included in the calculation of the time to complete the task, the average time for the fourth graders was 20 minutes. The average time for the sixth grade teacher was 18 minutes. Thus, the sixth grade students were faster. The standard deviation for the fourth graders was 7 minutes. The standard deviation for the sixth graders was 5 minutes. Thus the sixth grade students were more similar to each other in the amount of time it took to complete the task.
15. CLASSES = CREDITS/3.
16. 16.26/3 = 5.42
17. 2.40/3 = .8
18. Skewness ratio = -12.05. The distribution of CLASSES is severely negatively skewed.
19. Because of the presence of part-time students in the sample who could be taking only one credit during the semester. Full-time students typically take at least 15 credits during a semester.

****

1. SCHATTPP = SCHATTRT/100.
2. In this case, every data value in the distribution of SCHATTRT has been divided by 100, or multiplied by .01. Thus, we expect the mean, median, standard deviation, range, and interquartile range for SCHATTPP to be equal to the corresponding statistic for SCHATTRT divided by 100. We expect the variance for SCHATTPP to be equal to the variance for SCHATTRT divided by 1002 or 10,000. Because the transformation is linear, we expect that it will have no effect on the shape of the distribution so that the skew will remain the same.
3. The SPSS output below corroborates our expectations. Note that the range is given as 1 because SPSS rounds .56 to 1.

****

****

1. Compute EXPCENTS = EXPINC30 \*100
2. The mean and median for EXPCENTS are equal, respectively, to the mean and median for EXPINC30 multiplied by 100. The standard deviation and interquartile range for EXPCENTS are equal, respectively, to the standard deviation and interquartile range for EXPINC30 multiplied by 100. The skew for EXPCENTS is equal to the skew for EXPINC30.

****

a) .47

b) The standard deviation is .50.

c) COMP1 = COMPUTER + 1.

d) COMP2 = (-1)\*COMPUTER + 1.

e) For COMP1, the mean is 1.47 and the standard deviation is .50.

For COMP2, the mean is .53 and the standard deviation is .50.

1. *M* = 9.44, *SD* = 12.37, skewness ratio = 1.15/.122 = 9.42.
2. (-2)\*9.44 + 3 = -15.88.
3. (2)\*12.37 = 24.74.
4. (-1)\*9.42 = -9.42.
	1. The mean of SEX is 1.5. If 1 were subtracted from every score for the variable sex, the mean would be 1.5 – 1 = .5 and the coding for the new variable would be 0 for Men and 1 for Women. Accordingly, the proportion of women in the *Framingham* data set is 50 percent.
5. 3.
6. .99.
7. 31.

a)

****

b) ZSCORE = (SLFCNC08-21.06)/5.971

c)

****

d) The mean of this z-score distribution is 0, as it is for all z-score distributions. The standard deviation of this z-score distribution is 1, as it is for all z-score distributions. This is verified by using SPSS to calculate the mean and standard deviation of ZSCORE and ZSLFCNC0.

1. Mean: 21.06 + 5 = 26.06.

Standard deviation = 5.971.

.

1. X = (.6605)(-1.71) + 4.132 = 3.00.
2. 25. There are 7 z-scores below –2 and 18 above 2.
3. The cumulative percentage of the value 5.5 in English is 98.2. The cumulative percentage of the value 5.5 in Math is 99.0. Accordingly, 5.5 represents a slightly more unusual number in math than in English because only 1 percent of the other students took more than 5.5 years of math, while 1.8 percent of the other students took more than 5.5 years of English.
4. The *z*-score of the value 5.5 in English is 2.07. The *z*-score of the value 5.5 in Math is 2.31. Accordingly, 5.5 is slightly more unusual in math than in English because its z-score is higher.
5. According to cumulative percentages, a score of 25 is relatively highest in eighth grade because the cumulative percentage in eighth grade is 74.8 percent, in tenth grade it is 67.2 percent, and in twelfth grade it is 20.6 percent.
6. According to *z*-scores, a score of 25 is highest in 8th grade because the *z*-score in eighth grade is .66, in tenth grade it is .35, and in twelfth grade it is -.90.

a) According to a cumulative percentage criterion, a self-concept score of 25 in eighth grade represents a higher level of self-concept for females than for males. For males, 70.9 percent of individuals have lower than or equal to a self-concept score of 25. That percentage for females is 78.0.

b) According to a z-score criterion, a self-concept score of 25 in eighth grade represents a higher level of self-concept for females than males. The z-score for males is .52 and for females it is .78.

a) The female is slightly higher based on the cumulative percentage criterion. The percentage of females who scored at or below 58 is 70.0, while the percentage of males who scored at or below 63 is 68.6.

b) The female is only slightly higher based on a z-score criterion. The z-score computed relative to females is .51 , while the z-score computed relative to males is .50.

* 1. . The score of 89 is 1.5 standard deviations below the mean.
	2. .
	3. Q1 ≤  < z = +1
1. *M* = 2.54, *SD* = 1.20, and Skewness ratio = 2.15 (.508/.236).
2. AGE = GRADE + 5
3. 2.54 + 5 = 7.54
4. 1.20
5. The skewness ratio is 2.15, the same as for GRADE. The shape is positively skewed because the shape of a distribution does not change under translation.
	1. Jkj



1. The skewness ratio for the resource room placement is , which indicates that the skewed negatively, although not severely so. The skewness ratio for the self-contained classroom placement is , which indicates that the distribution is severely negatively skewed.

1. According to both the mean and the median, the students have a higher overall level of reading comprehension in the resource room placement (*M* = 81.43, *Median* = 83.00) than in the self-contained classroom placement (*M* = 68.87, *Median* = 69.00).
2. According to the interquartile range, the reading comprehension scores are more consistent for those in the self-contained classroom placement (*IQR* = 12) than for those in the resource room placement (*IQR* = 12.50). While the range and standard deviation lead to an opposite conclusion, they are not robust statistics and are influenced by the outliers present to varying degrees in these distributions.
3. The maximum reading comprehension score for students in the resource room is 107, while for those in a self-contained classroom it is only 86.

1. ; . A student with a reading comprehension score of 75 would be above the mean in the self-contained classroom and below the mean in the resource room classroom. Furthermore, the student would be relatively farther away from the mean in the resource room than in the self-contained classroom.

1. Yes. A percentile is a raw score in the distribution.
2. 17
3. 18.5 percent.

c) Below the median since the median has 50 percent of the distribution below it. Also, the median age is 48.

* 1. Below the mean by 1.07 standard deviations because 
	2. 10 (all positive outliers)
	3. There are five outliers among the men and five among the women.
	4. BIRTHYR = 1956 – AGE1
	5. Given that the median for AGE1 is 48, the median for BIRTHYR must be 1956 - 48 = 1908.
	6. The same as for AGE1, 8.425.

****

Given that both math and reading achievement distributions have skewness ratios greater than 2 in absolute terms (for math it is -3.02 and for reading it is -5.12), we seek non-linear transformations to reduce their skewness. The following transformations are explored:

ACHMATLG = LG10((-1)\*ACHMAT10 + 72)

ACHMATSQ = SQRT((-1)\*ACHMAT10 + 72)

ACHRDGLG = LG10((-1)\*ACHRDG10 + 69)

ACHRDGSQ = SQRT((-1)\*ACHRDG10 + 69)

These produce the following skewness results:

****

Based on these results, the square root transformations appear to be most appropriate.

1. The following square root and log transformations were applied to APOFFER.

COMPUTE APOFFERSQ = SQRT(APOFFER+1) .

COMPUTE APOFFERlg = LG10(APOFFER+1) .

Based on the following skewness results, the log transformation appears to be the more appropriate in this case for symmetrizing the variable.

****

Assuming the *NELS* data set was already open, the Syntax file used to generate these results should look like:

DATASET ACTIVATE DataSet1.

COMPUTE apoffersq=sqrt(apoffer+1).

EXECUTE.

COMPUTE apofferlg=lg10(apoffer+1).

EXECUTE.

DESCRIPTIVES VARIABLES=apoffer apoffersq apofferlg

 /STATISTICS=MEAN STDDEV MIN MAX SKEWNESS.

1. The following square root and log transformations were applied to FAMSIZE:

COMPUTE FAMSIZEsq = SQRT(FAMSIZE) .

COMPUTE FAMSIZElg = LG10(FAMSIZE) .

Based on the following skewness results, the log transformation appears to be the more appropriate in this case for symmetrizing the variable ****

1. The following square root and log transformations were applied to SCHATTRT:

COMPUTE SCHATTsq = SQRT(-1\*SCHATTRT+100) .

COMPUTE SCHATTlg = LG10(-1\*SCHATTRT+100) .

Based on the following skewness results, the log transformation appears to be the more appropriate in this case for symmetrizing the variable ****

1. No effect as shown below.

****

1. The following square root and log transformations were applied to GRADE:

COMPUTE GRADEsq = SQRT(GRADE) .

COMPUTE GRADElg = LG10(GRADE) .

Based on the following skewness results, the square root transformation appears to be the more appropriate in this case for symmetrizing the variable ****

1. The following square root and log transformations were applied to MATHCOMP:

COMPUTE MATHCOMPsq = SQRT(MATHCOMP) .

COMPUTE MATHCOMPlg = LG10(MATHCOMP) .

Based on the following skewness results, the log transformation appears to be the more appropriate in this case for symmetrizing the variable ****

1. The following square root and log transformations were applied to READCOMP:

COMPUTE READCOMPsq = SQRT(-1\*READCOMP+108) .

COMPUTE READCOMPlg = LG10(-1\*READCOMP+108) .

Based on the following skewness results, the square root transformation appears to be the more appropriate in this case for symmetrizing the variable ****

1. Construct a compute statement, as follows:

UNITMNC = UNITMATH – UNITCALC.

Assuming the *NELS* data set was already open, the Syntax file used to generate these results should look like:

DATASET ACTIVATE DataSet1.

COMPUTE sumunits=unitmath + unitengl.

EXECUTE.

1. One student took 6 units (years) of non-calculus math.
2. Construct a compute statement, as follows:

 ACHTOT12 = ACHMATH12 + ACHRDG12 + ACHSCI12 + ACHSLS12.

1. According to the histogram, the distribution is negatively skewed. This conclusion is corroborated by the skewness ratio of -3.7.



1. Among the students in the *NELS* data set, boys score higher than girls on the total achievement measure, as they have a higher mean as well as a higher median. The mean total achievement for boys is 230.14, while for girls it is 218.77. The median total achievement for boys is 234.19, while for girls it is 220.04.



1. One way to create the variable is with the following sequence of old and new

 Values in the Recode procedure: 0 → 0, miss→ sysmis, and else → 1.

1. 128.
2. The mean of APOFFYN is .729, indicating that, not including those schools with a missing value on this variable, approximately 73% offer at least one AP course.
3. COMPUTE ALIVE = Death - Birth. At 92 years. Harald Cramer lived the longest.
4. COMPUTE OLD = 2014 - Birth. If he were still alive today, John Tukey, at 99 would be the youngest of the statisticians in the data set.
5. The Syntax file should include:

DATASET ACTIVATE DataSet4.

COMPUTE Alive=Death - Birth.

EXECUTE.

DESCRIPTIVES VARIABLES=Alive

 /STATISTICS=MEAN STDDEV MIN MAX.

COMPUTE Old=2014-Birth.

EXECUTE.

DESCRIPTIVES VARIABLES=Old

 /STATISTICS=MEAN STDDEV MIN MAX.

1. 136 people (.447 x 304) lost weight and 74 (.243 x 304) reduced their cigarette smoking (those with negative values on the diff variables).
2. For those who did not experience a CHD event, the mean change in BMI is .34, indicating an increase in BMI over this time period. For those who did experience a CHD event, the mean change in BMI is -.28, indicating a decrease in BMI over this time period. In short, those without a CHD event gained weight and those with such an event lost weight.

c) Because the distribution of the difference in cigarettes smoked per day is severely positively skewed, we employ the median, a robust statistic, for the analysis. For both those who did and did not experience a CHD event, the median change in cigarettes smoked per day is 0, indicating no change in cigarette use over time for wither group. However, the means lead to a different conclusion. The mean change in cigarettes per day of those who did not experience a CHD event is -1.21, indicating a decrease in the average number of cigarettes smoked per day over this time period. For those who did experience a CHD event, the mean change in CIGDAY is -3.01, indicating a decrease in the average number of cigarettes smoked per day over this time period. In short, although both groups smoked less, the decrease is greater by about two cigarettes per day for those who experienced a CHD event than those who did not.