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 sixth-graders was actually 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 remains the same at 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 shape of a distribution does not change with a multiplicative linear transformation. 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.
20. The R command **summary(NELS$schattrt)** is used to obtain the mean, median, and range (by subtracting the minimum from the maximum).

../../../Desktop/Screen%20Shot%202019-06-19%20at%203.34.07%20P

The remaining statistics (*SD* = 4.69, variance = 22.02, IQR = 3, skewness = -6.21)are found using the following commands:

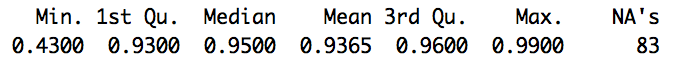
**sd(NELS$schattrt, na.rm = T)**

**var(NELS$schattrt, na.rm = T)**

**IQR(NELS$schattrt, na.rm = T)**

**skew(NELS$schattrt)**

1. **NELS$schattpp = NELS$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 R command **summary(NELS$schattpp)** is used to obtain the mean, median, and range (by subtracting the minimum from the maximum).

****

The remaining statistics (*SD* = .0469, variance = .0022, IQR = .03, skewness = -6.21)are found using the following commands:

**sd(NELS$schattpp, na.rm = T)**

**var(NELS$schattpp, na.rm = T)**

**IQR(NELS$schattpp, na.rm = T)**

**skew(NELS$schattpp)**

All of the outputs corroborate our predictions from part (c).

1. The R command **summary(NELS$expinc30)** is used to obtain the mean, median, and range (by subtracting the minimum from the maximum).

../../../Desktop/Screen%20Shot%202019-06-19%20at%203.52.26%20P

The remaining statistics (*SD* = 58,265.76; variance = 3,394,898,523; IQR = 25,000; skewness = 10.90)are found using the following commands:

**sd(NELS$expinc30, na.rm = T)**

**var(NELS$expinc30, na.rm = T)**

**IQR(NELS$expinc30, na.rm = T)**

**skew(NELS$expinc30)**

1. **NELS$expcents = NELS$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.
3. The R command **summary(NELS$expcents)** is used to obtain the mean, median, and range (by subtracting the minimum from the maximum).

../../../Desktop/Screen%20Shot%202019-06-19%20at%203.59.13%20P

The remaining statistics (*SD* = 5,826,576; variance = 3.394899e+13 or about 3.39\*1013; IQR = 2,500,000; skewness = 10.90)are found using the following commands:

**summary(NELS$expcents)**

**sd(NELS$expcents, na.rm = T)**

**var(NELS$expcents, na.rm = T)**

**IQR(NELS$expcents, na.rm = T)**

**skew(NELS$expcents)**

All of the outputs corroborate our predictions from part (c).

1. The mean is 1.47 and the standard deviation is .50. Note that we use the command **as.numeric** to access the numeric coding of the variable:

**mean(as.numeric(NELS$computer))**

**sd(as.numeric(NELS$computer))**

b) **NELS$comp1 = as.numeric(NELS$computer) - 1**

c) **NELS$comp2 = -1\*as.numeric(NELS$comp1) + 1**

d) For comp1, the mean is .47 and the standard deviation is .50.

For comp2, the mean is .53 and the standard deviation is .50.

We can verify the proportions of comp1 and comp2 as follows:

**percent.table(NELS$computer,NELS$comp1)**

**percent.table(NELS$computer,NELS$comp2)**

* 1. The new variable can be created by the command **Framingham$CIGTRANS = -2\*Framingham$CIGPDAY3 + 3**.

1. mean = 7.23, *SD* = 12.51, skewness = 2.09
2. (-2)\* 7.23+ 3 = -11.46, or **mean(Framingham$CIGTRANS, na.rm = T)**
3. (2)\* 12.51= 25.01, **sd(Framingham$CIGTRANS, na.rm = T)**
4. (-1)\*2.09 = -2.09, **skew(Framingham$CIGTRANS)**
   1. The mean of SEX is 1.5. If 1 were subtracted from every score for the variable sex, the mean would be 1.5 – 1 = .5 and the coding for the new variable would be 0 for Men and 1 for Women. Accordingly, the proportion of women in the Framingham dataset is 50 percent.
5. 3
6. .99
7. The R command to create the *z*-score of famsize is **NELS$zfamsize = scale(NELS$famsize)**. We then use the command **table(NELS$zfamsize[abs(NELS$zfamsize) > 2])** to obtain a frequency distribution of those *z*-scores less than -2 or greater than 2. The total number of outliers is 9 + 9 +13 = 31.

../../../Desktop/Screen%20Shot%202019-06-19%20at%204.59.54%20P

a) The mean is 21.06, and the standard deviation is 5.97.

1. **NELS$zslfcnc08ver1 = (NELS$slfcnc08 - 21.06) / 5.97**

OR

**NELS$zslfcnc08ver1 = (NELS$slfcnc08 - mean(NELS$slfcnc08)) / sd(NELS$slfcnc08)**

c) **NELS$zslfcnc08ver2 = scale(NELS$slfcnc08)**

d) The mean of each version of *z*-score distribution is 0 (within rounding), as it is for all *z*-score distributions. The standard deviation of each version of *z*-score distribution is 1 (within rounding), as it is for all *z*-score distributions.

1. Mean: 21.06 + 5 = 26.06.

Standard deviation = 5.97.

.

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% of the other students took more than 5.5 years of math, while 1.8% 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%; in tenth grade it is 67.2%; and in twelfth grade it is 20.6%.
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% of individuals have a self-concept score lower than or equal to 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 about 68.3.

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. Note that there are missing values of achsci08 for the males but not the females, so you need to include the argument **na.rm = T** when finding the mean and standard deviation for males only.

* 1. . The score of 89 is 1.5 standard deviations below the mean.
  2. .
  3. *Q25* ≤  < *z* = +1

1. mean = 2.54, *SD* = 1.20, and skewness ratio = 2.12 (.501/.236).
2. age = grade + 5
3. 2.54 + 5 = 7.54
4. 1.20
5. The skewness ratio is 2.12, the same as for grade. The shape is positively skewed because the shape of a distribution does not change under translation.
6. The skewness ratio for the self-contained classroom placement is -4.46, which indicates that the distribution is severely negatively skewed. The skewness ratio for the resource room placement is -1.42, which indicates that the distribution is negatively skewed, though not nearly as severely.
7. According to both the mean and the median, the students have a higher overall level of reading comprehension in the resource room placement (mean= 81.43, median = 83.00) than in the self-contained classroom placement (mean= 68.87, median = 69.00).
8. According to the interquartile range, the reading comprehension scores are more consistent for those in the self-contained classroom placement (IQR = 11.50) than for those in the resource room placement (IQR = 12.00). While the range and standard deviation lead to an opposite conclusion, they are not robust statistics and are influenced to varying degrees by the outliers present in these distributions.
9. 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.
10. 
11. 
12. ; . 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.
13. Yes. A percentile is a raw score in the distribution.
14. 17
15. 18.5%

c) Below the median since the median has 50% 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. **Framingham$BIRTHYR = 1956 - Framingham$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.
  7. The R commands used to check the three skewness variables are:

**skew(NELS$achmat10)**

**se.skew(NELS$achmat10)**

**skew.ratio(NELS$achmat10)**

**skew(NELS$achrdg10)**

**se.skew(NELS$achrdg10)**

**skew.ratio(NELS$achrdg10)**

These produce the following skewness results for math and reading, respectively:

achmat10: skewness = -0.33; standard error of skewness = .11; skewness ratio = -3.01

achrdg10: skewness = -0.56; standard error of skewness = .11; skewness ratio = -5.09

Given that both math and reading achievement distributions have skewness ratios greater than 2 in absolute terms, we seek non-linear transformations to reduce their skewness. Since both distributions are negatively skewed, we first must multiply each by negative one; we add a number greater than the original maximum value to ensure all values are greater than zero before applying each non-linear transformation.

**NELS$achmatlg = log(-1\*NELS$achmat10 + 72, 10)**

**NELS$achmatsq = sqrt(-1\*NELS$achmat10 + 72)**

**NELS$achrdglg = log(-1\*NELS$achrdg10 + 69, 10)**

**NELS$achrdgsq = sqrt(-1\*NELS$achrdg10 + 69)**

These produce the following skewness results, respectively:

achmatlg: skewness = -1.23; standard error of skewness = .11; skewness ratio = -11.31

achmathsq: skewness = -0.32; standard error of skewness = .11; skewness ratio = -2.91

achrdglg: skewness = -1.29; standard error of skewness = .11; skewness ratio = -11.82

achrdgsq: skewness = -0.08; standard error of skewness = .11; skewness ratio = -0.74

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.

**NELS$apofferlg = log(NELS$apoffer+1, 10)**

**NELS$apoffersq = sqrt(NELS$apoffer+1)**

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

apoffer: skewness = 4.08; standard error of skewness = .11; skewness ratio = 36.34

apofferlg: skewness = -.13; standard error of skewness = .11; skewness ratio = -1.16

apoffersq: skewness = .94; standard error of skewness = .11; skewness ratio = 8.34

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

**NELS$famsizelg = log(NELS$famsize, 10)**

**NELS$famsizesq = sqrt(NELS$famsize)**

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

famsize: skewness = 1.10; standard error of skewness = .11; skewness ratio = 10.05

famsizelg: skewness = .02; standard error of skewness = .11; skewness ratio = 0.19

famsizesq: skewness = .58; standard error of skewness = .11; skewness ratio = 5.33

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

**NELS$schattrtlg = log(-1\*NELS$schattrt+100, 10)**

**NELS$schattrtsq = sqrt(-1\*NELS$schattrt+100)**

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

schattrt: skewness = -6.21; standard error of skewness = .12; skewness ratio = -51.97

schattrtlg: skewness = .14; standard error of skewness = .12; skewness ratio = 1.20

schattrtsq: skewness = 2.27; standard error of skewness = .12; skewness ratio = 19.02

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

**NELS$cigarettlg = log(as.numeric(NELS$cigarett)+1, 10)**

**NELS$cigarettsq = sqrt(as.numeric(NELS$cigarett)+1)**

These transformations have no effect on the shape of the distribution as shown below.

cigarett: skewness = 2.05; standard error of skewness = .11; skewness ratio = 18.78

cigarettlg: skewness = 2.05; standard error of skewness = .11; skewness ratio = 18.78

cigarettsq: skewness = 2.05; standard error of skewness = .11; skewness ratio = 18.78

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

**Learndis$gradelg = log(Learndis$grade, 10)**

**Learndis$gradesq = sqrt(Learndis$grade)**

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

grade: skewness = .50; standard error of skewness = .24; skewness ratio = 2.12

gradelg: skewness = -.31; standard error of skewness = .24; skewness ratio = -1.31

gradesq: skewness = .10; standard error of skewness = .24; skewness ratio = .40

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

**Learndis$mathcomplg = log(Learndis$mathcomp, 10)**

**Learndis$mathcompsq = sqrt(Learndis$mathcomp)**

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

mathcomp: skewness = .56; standard error of skewness = .25; skewness ratio = 2.25

mathcomplg: skewness = .23; standard error of skewness = .25; skewness ratio = .92

mathcompsq: skewness = .39; standard error of skewness = .25; skewness ratio = 1.58

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

**Learndis$readcomplg = log(-1\*Learndis$readcomp+108, 10)**

**Learndis$readcompsq = sqrt(-1\*Learndis$readcomp+108)**

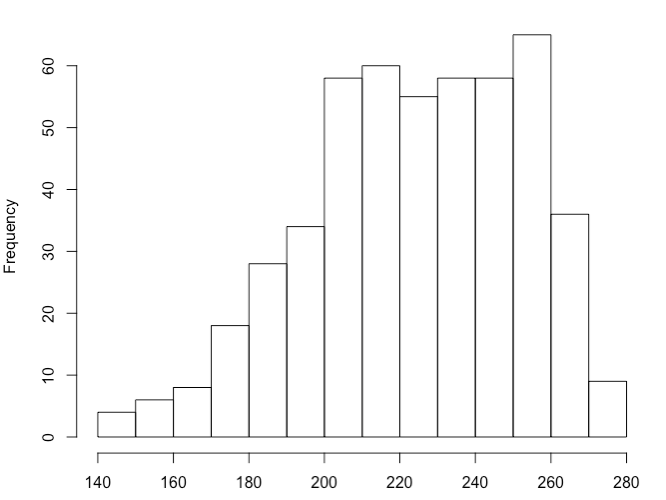
Based on the following skewness results, the square root transformation appears to be the more appropriate in this case for symmetrizing the variable.

readcomp: skewness = -.94; standard error of skewness = .28; skewness ratio = -3.40

readcomplg: skewness = -2.62; standard error of skewness = .28; skewness ratio = -9.49

readcompsq: skewness = -.41; standard error of skewness = .28; skewness ratio = -1.47

1. **NELS$unitmnc = NELS$unitmath - NELS$unitcalc**
2. One student took 6 units (years) of non-calculus math.
3. **NELS$achtot12 = NELS$achmat12 + NELS$achrdg12 + NELS$achsci12 + NELS$achsls12**
4. According to the histogram, obtained by the command **hist(NELS$achtot12)**, the distribution is negatively skewed. This conclusion is corroborated by the skewness ratio of -3.7.



1. Among the students in the NELS dataset, 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.
2. The first **ifelse** command is shown below, followed by the given command that ensures missing values are coded correctly. The second **ifelse** command checks if each value of apoffer is missing; if missing, apoffyn is assigned NA for “not available” and left as is otherwise.

**NELS$apoffyn = ifelse(NELS$apoffer > 0, "yes", "no")**

**NELS$apoffyn = ifelse(is.na(NELS$apoffer), NA, NELS$apoffyn)**

1. 128, obtained using **table(NELS$apoffyn)**
2. If apoffyn were a 0-1 coded dichotomous variable, the mean would be the proportion of 1’s in the variable, which would represent the proportion of schools that offer any AP courses, since 1 would equal “yes.”
3. **Statisticians$alive = Statisticians$Death - Statisticians$Birth**. At 92 years. Harald Cramer lived the longest.
4. **Statisticians$old = 2019 - Statisticians$Birth**. The R command assumes that the current year is 2019. If it’s not, adjust the number to the current year. If he were still alive today, John Tukey would be the youngest of the statisticians in the dataset at 104 years old.
5. 136 people (.34 \* 400) lost weight and 139 (.3475 \* 139) had reduced total cholesterol. Those with reduced BMI and cholesterol are taken as those with negative values on the DIFF variables.
6. 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) Both groups show an increase in total cholesterol, though the average increase is greater for those who experienced a CHD event than for those who did not. The mean change in total cholesterol for those who did not experience a CHD event is .95; for those who did experience a CHD event, the mean change in total cholesterol is 1.91.