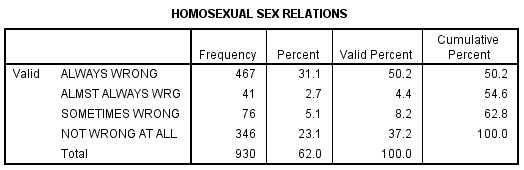
Chapter 3 SPSS Problems SOLUTIONS

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

1.

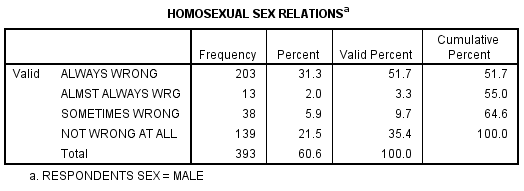
a. Looking at the following frequency distribution, the variable HOMOSEX is measured at the ordinal level. Therefore, the median is the appropriate measure of central tendency because the mean can only be used with interval-ratio variables. In this case, the median is “Always Wrong.”



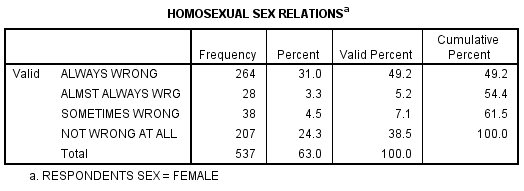
b. The following frequency distributions contains information on attitudes towards

homosexual relations for men and women:

Men:



Women:

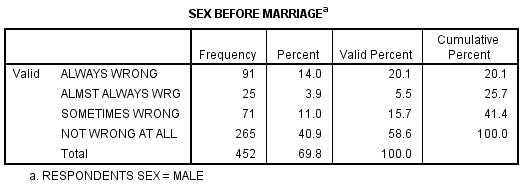


From the tables above, it appears that men and women have pretty similar attitudes on homosexual relations. That is, 51.7 percent of men and 49.2 percent of women say that homosexual relations are “Always Wrong.” Additionally, 35.4 percent of men and 38.5 percent of women say that homosexual relations are “Not Wrong at All.” And, about 13 percent of men and 12 percent of women are somewhere in the middle saying that homosexual relations are either “Almost Always Wrong” or “Sometimes Wrong.” What is clear is that both men and women seem to be on either extreme saying that homosexual relations are “Always Wrong” or “Not Wrong at All.”

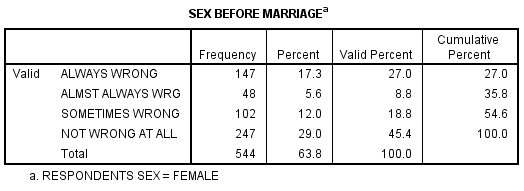
c. The following frequency distributions contain information on attitudes towards

premarital sex for men and women:

Men:



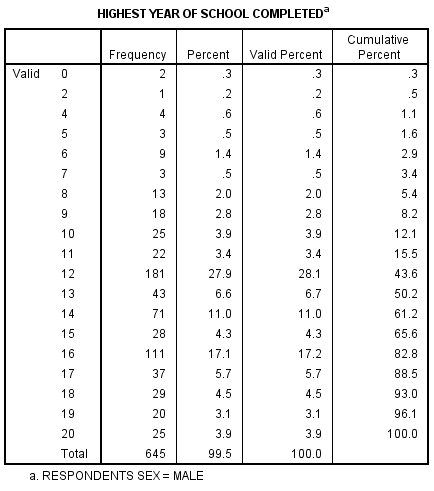
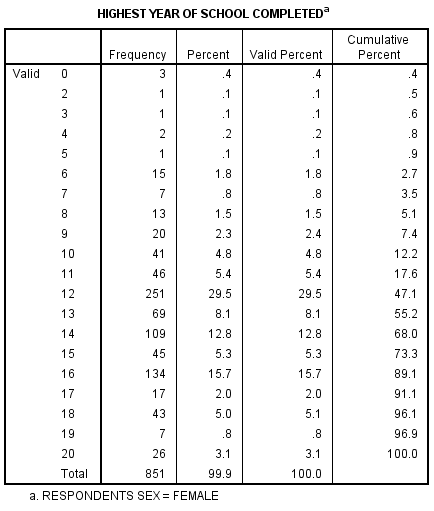
Women:

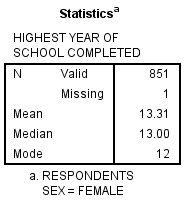
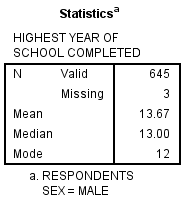


It is clear from the tables above that the majority of men view premarital sex as “Not Wrong at All” (58.6%). Women seem to be more evenly split: 27.0 percent of women view premarital sex as “Always Wrong” and 45.4 percent view premarital sex as “Not Wrong at All.” Also of interest is the category “Almost Always Wrong.” Only a small percentage of men and women fall into this category. Like attitudes on homosexual relations, it seems to some degree that men and women are prone toward the extreme response categories of “Always Wrong” and “Not Wrong at All.”

2.

a. The following frequency distributions have been adapted from the SPSS outputs and contain information on the educational levels of men and women:



For the most part, both distributions appear symmetrical; the mean, median, and mode for each group are very close. The mean number of years of education for men is 13.67. The mean number of years of education for women is 13.31. The median for men is 13, while the median for women is also 13. The modes for men and women are identical (12). The slight difference in mean number of years of education between men and women indicates that there are a small number of men who have a very high number of years of education; these individuals *pull* the mean upwards.

b. Practically speaking, the difference between male and female educational attainment is not that great and of little importance. This was not true in the past. Of course, years of education are a crude measure, and the type of school attended probably has an important effect on type of job and salary. Also, education is only one of the determinants of occupation and salary, which is why females can have educational levels close to males but still be making much less in salary.

3.

a. Theoretically, there is really no one best measure of central tendency. It is helpful to know the most common category, the mode, but also the mean and the median, since the distributions are not that highly skewed. The best measure is probably the mean, but since number of children (CHILDS) only comes in discrete values, the median is also valuable. For these particular distributions, the mean is definitely better, because the median makes it seem as if all three groups have the same number of children.

b. On average, upper class respondents have slightly more children per respondents (mean = 2.23, median = 2.0, mode = 2.0). The median (2.0) is the same for all classes. Working class and lower class have modes of zero, while middle class and upper class have modes of 2.00.

c.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Lower Class*** | ***Working Class*** | ***Middle Class*** | ***Upper Class*** |
| ***Mean*** | 3.55 | 3.09 | 3.19 | 3.30 |
| ***Median*** | 3.0 | 2.5 | 2.0 | 2.0 |
| ***Mode*** | 2.0 | 2.0 | 2.0 | 2.0 |

With the exception of working class respondents (median = 2.5 kids), the median is the same for all groups (2 kids). Additionally, all classes had the same mode of 2.0. The mean for lower class respondents (3.55) is the highest of all groups.

4.

a. TVHOURS: mode (2.0), median (2.0), Mean (3.01)

b. POLVIEWS: mode (moderate), median (moderate)

c. HRS1: mode (40.00), median (40.00), mean (40.62)

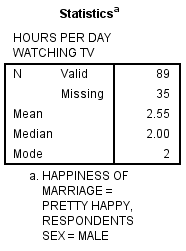
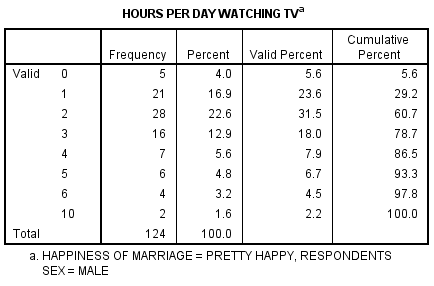
d. OWNGUN: Mode (no), median (no)

e. SIBS: mode (2.0), median (3.0), mean (3.68)

f. GRASS: mode (not legal), median (not legal)

5.

a. The frequency distribution, as well as statistics, for males that are pretty happy is shown below:



b. Males with “very happy” (2.84 hours) and “not too happy” (3.17 hours) marriages report watching more hours of TV than do males with “pretty happy” marriages (2.55 hours). For females, rates of hour spent watching TV have an inverse relationship with marital satisfaction. In other words, as marital satisfaction increases, hours spent watching TV decreases.

Chapter 3 Exercise SOLUTIONS

1.

a. The mode can be found two ways: looking for either the highest frequency (470) or the highest percentage (48.3). The mode is the category that corresponds to these values, “Exciting.”

b. The median can be found two ways: by using either the frequencies column of the cumulative percentages.

|  |  |
| --- | --- |
| ***Using Frequencies*** | ***Using Cumulative Percentages*** |
|  | Notice that 48.3% of the observations fall in the “Exciting” percentage category; 93.9% fall in or below the “Routine” category. |
| Starting with the frequency in the first category (470), add up the frequencies until you find where the 487th and 488th cases fall. Both of these cases correspond to the category “Routine,” which is the median. | The 50% mark, or the median, is located somewhere within the “Routine” category. So the median is “Routine.” |

c. The mode is simply the category with the highest frequency (or percentage) in the distribution. The median divides the distribution into two equal parts so that half the cases are below it and half above it.

d. Because this variable is an ordinal-level variable.

2.

a. Ordinal. The mode can be found two ways, either by looking for the highest frequency (467) or the highest percent (50.2%). The mode is the category that corresponds to these values, “Always Wrong.”

b. The median can be found two ways: by using either the frequencies column of the cumulative percentages.

|  |  |
| --- | --- |
| ***Using Frequencies*** | ***Using Cumulative Percentages*** |
|  | Notice that 50.2% of the observations fall in the “Almost always wrong” cumulative percentage category. |
| Starting with the frequency in the first category (472), add up the frequencies until you find where the 465th and 466th cases fall. Both of these cases correspond to the category “Almost always wrong,” which is the median. | The 50% mark, or the median, is located somewhere within the “Always almost wrong” category. So the median is “Always almost wrong.” |

A slight majority of people (50.2%) view homosexual relations as “Always wrong” or “Almost always wrong.” At the other extreme, 37.2% of people view homosexual relations as “Not wrong at all.”

3.

1. Interval-ratio. The mode can be found two ways: by looking either for the highest frequency (14) or the highest percentage (43.8%). The mode is the category that corresponds to the value “40 hours worked last week.” The median can be found two ways: by using either the frequencies column or the cumulative percentages.

|  |  |
| --- | --- |
| ***Using Frequencies*** | ***Using Cumulative Percentages*** |
|  | Notice that 34.4% of the observations fall in or below the “32 hours worked last week” category; 78.1% fall in or below the “40 hours worked last week” category. |
| Starting with the frequency in the first category (1), add up the frequencies until you find where the 16th and 17th cases fall. Both of these cases correspond to the category “40 hours worked last week,” which is the median. | The 50% mark, or the median, is located somewhere within the “40 hours worked last week” category. So the median is “40 hours worked last week.” |

1. Since the median is merely a synonym for the 50th percentile, we already know that its value is 40 hours worked last week.

25th percentile = (32 × 0.25) = 8th case = 30 hours worked last week

75th percentile = (32 × 0.75) = 24th case = 40 hours worked last week

4.

1. The median can be found two ways, either by using the frequencies column or by using the cumulative percentages. However, since the problem only gives the frequencies, we’ll use those to solve for the median.

|  |
| --- |
| *Using Frequencies* |
|  |
| Starting with the frequency in the first category (94), add up the frequencies until you find where the 240th and 241st cases fall. Both cases correspond to the category “Somewhat disagree,” which is the median. |

1. 20th percentile = (480 × 0.20) = 96th case = “Somewhat agree”  
   80th percentile = (480 × 0.80) = 384th case = “Somewhat disagree”

5.

1. The mode can be found by looking for the highest frequency in each column; the mode for each group is listed below:

18-29: Good

30-39: Good

40-49: Good

50-59: Good

The median can be found two ways: by using either the frequencies column or the cumulative percentages. However, since the problem only gives the frequencies, we’ll use those to solve for the median.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Age Group*** | | | |
| ***18–29*** | ***30–39*** | ***40–49*** | ***50–59*** |
|  |  |  |  |
| Starting with the frequency in the first category (56), add up the frequencies until you find where the 82nd and 83rd cases fall. Both cases correspond to “Good,” which is the median. | Starting with the frequency in the first category (55), add up the frequencies until you find where the 85th case falls. This case corresponds to “Good,” which is the median. | Starting with the frequency in the first category (41), add up the frequencies until you find where the 84th and 85th cases fall. Both cases correspond to “Good,” which is the median. | Starting with the frequency in the first category (38), add up the frequencies until you find where the 86th and 87th cases fall. Both of these cases correspond to “Good,” which is the median. |

1. Since the mode and median for all four age groups was “Good,” it has to do with how respondents interpreted the question. For instance, it is possible that one’s health status was assessed relative to his or her age. Neither the median nor the mode provides a better description of the data since they provide the same information.

6.

a. We begin by summing each of the columns.

|  |  |  |
| --- | --- | --- |
| ***State*** | ***2005*** | ***2009*** |
| Alabama | 740 | 828 |
| Delaware | 125 | 145 |
| Florida | 3,008 | 3,289 |
| Illinois | 1,674 | 1,806 |
| Minnesota | 691 | 767 |
| New Hampshire | 185 | 217 |
| New York | 2,758 | 2,937 |
| Washington | 807 | 938 |
|  | Σ = 9,988 | Σ = 10,927 |

Now we divide each sum by 8 (the total number of states). The mean in 2005 is

1,248.5 or 9,988/8. We divide by 8 because only 8 states are included in the table.

The mean in 2009 is 1,365.9 or 10,927/8.

Since the mean is higher in 2009, one could say that on average there are more Americans on Medicare in 2009 than in 2005. However, the mean is not the best measure of central tendency because three of the states (Florida, New York, and Illinois) have exponentially more cases than the other states, pulling the mean toward a higher value. You can verify this by computing the median for each year.

b. We begin by summing each of the columns after having removed Florida, Illinois, and New York.

|  |  |  |
| --- | --- | --- |
| ***State*** | ***2005*** | ***2009*** |
| Alabama | 740 | 828 |
| Delaware | 125 | 145 |
| Minnesota | 691 | 767 |
| New Hampshire | 185 | 217 |
| Washington | 807 | 938 |
|  | Σ = 2,548 | Σ = 2,895 |

The recalculated 2005 mean is 509.6; the mean for 2009 is 579. The mean now appears to be a more accurate measure of central tendency. Both means are better measures of central tendency now since they are closer to the values for most states.

7.

We begin by multiplying each household size by its frequency.

|  |  |  |
| --- | --- | --- |
| ***Household Size*** | ***Frequency*** | ***Frequency x Y (fY)*** |
| 1 | 381 | 381 |
| 2 | 526 | 1,052 |
| 3 | 227 | 681 |
| 4 | 200 | 800 |
| 5 | 96 | 480 |
| 6 | 42 | 252 |
| 7 | 19 | 133 |
| 8 | 5 | 40 |
| 9 | 2 | 18 |
| 10 | 2 | 20 |
| Total | *N* = 1,500 | ∑*fy* = 3,857 |

So, the mean number of people per household is 2.57.

8.

a. Students might become confused here and try to somehow relate the mode or median to the names of the states, which are simply labels for the cases. For the 2005 data, there is no mode. Since there are 8 cases, the median is the average of the 4th and 5th cases. Cases must be ordered, from low to high, so the median is (740 + 807)/2=773.5. Since the median (773.5) is lower than the mean (1,248.5) for 2005, the distributions are skewed in a positive direction.

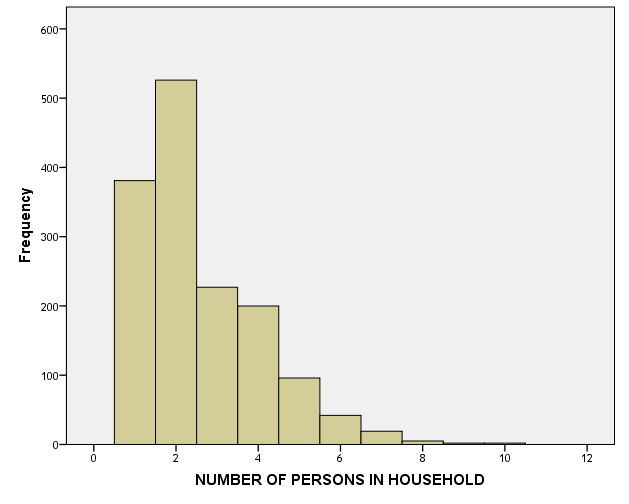
For the 2009 data, there is no mode. Since there are 8 cases, the median is the average of the 4th and 5th cases. Cases must be ordered, from low to high, so the median is (828 + 938)/2=883. Since the median (883) is lower than the mean (1,365.9) for 2009, the distributions are skewed in a positive direction.

b. Because of the very high values of Medicare enrollees for Illinois, New York, and Florida, the mean is inflated and does not accurately portray the number of enrollees in each state. Therefore, the median is a better estimation of 2005 and 2009 Medicare enrollees.

c. As mentioned before, the extreme number of Medicare enrollees for Florida, Illinois, and New York “pulls” the mean upward, thereby creating a positively skewed distribution.

9.

a. There appear to be a few outliers (i.e., extremely high values); this leads us to believe that the distribution is skewed in the positive direction.



b. The median can be found two ways: by using either the frequencies column or the cumulative percentages. The data are in frequencies; we’ll use those to solve the median. Since the median (2) is less than the mean (2.57), we can conclude that the distribution is skewed in a positive direction. Our answer to question 9a is further supported.

|  |
| --- |
| ***Using Frequencies*** |
|  |
| Starting with the frequency in the first category (381), add up the frequencies until you find where the 750th and 751st cases fall. Both of these cases correspond to the category “2,” which is the median. |

10.

a. We begin by multiplying each category by its frequency.

|  |  |  |
| --- | --- | --- |
| ***Hours Worked Last Week*** | ***Frequency*** | ***Frequency x Y (fY)*** |
| 20 | 3 | 60 |
| 25 | 2 | 50 |
| 28 | 1 | 28 |
| 29 | 1 | 29 |
| 30 | 3 | 90 |
| 32 | 1 | 32 |
| 40 | 14 | 560 |
| 50 | 2 | 100 |
| 52 | 1 | 52 |
| 55 | 1 | 55 |
| 60 | 1 | 60 |
| 64 | 1 | 64 |
| 70 | 1 | 70 |
| Total | *N* = 32 | *fY* = 1,250 |

b. Remember, the median was 40 hours worked last week. This distribution is skewed in a negative direction (i.e., the value of the mean is lower than the median).

11. Yes, both politicians can be correct, at least in a technical sense. One politician can be referring to the mean; the other could be using the median. It would be unusual if these two statistics were exactly equal. The average or mean income of Americans can be greater than the median if the distribution of income is positively skewed, which is certainly true.

12. First, let’s order each set of scores from lowest to highest.

|  |  |  |  |
| --- | --- | --- | --- |
| ***2008-2010 Male Murder Rate per 100,000*** | | | |
| ***Top 10 by Population*** | ***Murder Rate*** | ***Bottom 10 by Population*** | ***Murder Rate*** |
| Japan | 0.4 | Germany | 0.9 |
| China | 2.2 | Italy | 1.6 |
| India | 3.9 | United Kingdom | 1.7 |
| Pakistan | 4.3 | France | 1.9 |
| United States | 6.6 | Egypt | 2.2 |
| Indonesia | 13.9 | South Korea | 2.2 |
| Nigeria | 18.2 | Iran | 2.3 |
| Mexico | 23.0 | Vietnam | 2.6 |
| Russia | 29.1 | Turkey | 8.6 |
| Brazil | 54.7 | Democratic Republic of the Congo | 35.8 |
|  | ∑ = 156.3 |  | ∑ = 59.8 |

The median is simply the number in the middle. Since we have 10 countries in each group, we want the 5.5th case ([10 + 1]/2). For the countries in the top 10, the 5.5th case falls between the United States and Indonesia. So the median is 10.3 ([6.6+13.9]/2). For the countries in the bottom 10, the 5.5th case falls between Egypt and South Korea. So the median is 2.2. ([2.2+2.2]/2).

The mean for each group is calculated using the following formula:

For countries in the top 10 for murder rates, the mean is 15.63 (156.3/10). For countries in the bottom 10, the mean is 5.98 (59.8/10). Brazil has the highest murder rate (54.7) among these countries. On average, countries with higher populations (i.e., top 10) had higher murder rates. The distribution of murder rates for countries in the bottom 10 is skewed in a positive direction because only one country (i.e., Democratic Republic of the Congo) had a very high murder rate. The pattern is similar for the countries in the top 10 due to Brazil’s high murder rate, as well as the fairly high murder rates for Mexico and Russia.

13.

The mode can be found by looking for the highest frequency in each column; the mode for each group is listed below:

Males: Working

Females: Working

The median can be found two ways, either by using the frequencies column or by using the cumulative percentages. However, since the problem only gives the frequencies, we’ll use those to solve for the median.

|  |  |
| --- | --- |
| ***Males*** | ***Females*** |
|  |  |
| Starting with the frequency in the first category (390), add up the frequencies until you find where the 237th and 238th cases fall. We actually don’t need to do any adding, as both these cases correspond with the first category, “Working.” | Starting with the frequency in the first category (441), add up the frequencies until you find where the 323rd and 324th cases fall. We actually don’t need to do any adding, as both these cases correspond with the first category, “Working.” |

When using both the mode and median to estimate participation in the labor force, it appears there are no substantial differences between males and females. This might be due to the emphasis on being able to provide for oneself as well as the necessity for households to have two sources of income.

14.

a. We begin by ordering the countries from lowest to highest rate of infant mortality.

|  |  |
| --- | --- |
| ***Country*** | ***Infant Mortality Rates*** |
| Finland | 3.40 |
| Luxembourg | 4.39 |
| Canada | 4.85 |
| United States | 6.00 |
| Panama | 11.32 |
| Syria | 15.12 |
| Colombia | 15.92 |
| Turkey | 23.07 |
| Zimbabwe | 28.23 |
| Rwanda | 62.51 |
| Afghanistan | 121.63 |

Since we are working with an odd number of cases (11), the median is simply the frequency associated with the sixth case, Syria. Therefore, the median is 15.12. To calculate the mean, we total the number of infant mortality rates for the eleven countries.

|  |  |
| --- | --- |
| ***Country*** | ***Infant Mortality Rates*** |
| Finland | 3.40 |
| Luxembourg | 4.39 |
| Canada | 4.85 |
| United States | 6.00 |
| Panama | 11.32 |
| Syria | 15.12 |
| Colombia | 15.92 |
| Turkey | 23.07 |
| Zimbabwe | 28.23 |
| Rwanda | 62.51 |
| Afghanistan | 121.63 |
|  | ∑ = 296.44 |

b. Since the median value (15.12) is much lower than the mean value (26.95), we know that the data listed above is positively skewed.

c. Two reasons for the varying infant mortality rates are the types of health services available to women as well as the average level of education of women.

15. a. The data are measured at the ordinal level.

b. Whites: Mode = Somewhat agree and Somewhat disagree (bimodal), Median = Somewhat agree

Blacks: Mode = Somewhat disagree, Median = Somewhat disagree

c. The data suggest a difference between Whites and Blacks for whether racism is a thing of the past. The most common category for Blacks is “somewhat disagree”, while for Whites it is a tie between “somewhat agree” and “somewhat disagree”. The median – the most useful measure of central tendency for ordinal variables – is “somewhat disagree” for Blacks and “somewhat agree” for Whites. Whites are more likely to believe racism is in the past, while Blacks are more likely to see it as ongoing.

16. a. 4 children

b. 2 children

c. We begin by multiplying each number of children by its frequency.

|  |  |  |
| --- | --- | --- |
| Ideal number of children | Frequency | Frequency x Y (fY) |
| 0 | 8 | 0 |
| 1 | 29 | 29 |
| 2 | 477 | 954 |
| 3 | 238 | 714 |
| 4 | 83 | 332 |
| 5 | 18 | 90 |
| 6 | 9 | 54 |
| 7 | 3 | 21 |
| Total | N=865 | 2194 |

= 2.54

So, the mean number of children people would ideally want to have is 2.54.

d. The mean (2.54) is higher than the median (2), meaning the distribution of answers is positively skewed. This is due to the few extremely large responses of 6 or 7 children.