INSTRUCTOR'S SOLUTIONS MANUAL

Statistical Methods for Psychology

EIGHTH EDITION

David C. Howell
The University of Vermont

Table of Contents

General Notes

These solutions were checked using a variety of calculators and computer software. Answers often differ (sometimes a surprising amount) depending on how many decimal places the calculator or program carries. It is important not to be too concerned about differences, especially ones in the second or third decimal place, which may be attributable to rounding (or the lack thereof) in intermediate steps.

Although I do not provide detailed answers to all discussion questions, for reasons given elsewhere, I have provided pointers for what I am seeking for many (though not all) of them. I hope that these will facilitate using these items as a basis of classroom discussion.

Chapter 1 - Basic Concepts

1.1 The entire student body of your college or university would be considered a population under any circumstances in which you want to generalize **only** to the student body of your college or university and no further.

1.2 When you want to generalize or make inferences about a larger population of students (all U.S. students, for example), then the student body of your college or university would be considered a sample.

1.3 The students of your college or university are a nonrandom sample of U.S. students, for example, because all U.S. students do not have an equal chance of being included in the sample.

1.4 Not all residents are listed in the phone book and thus not all residents have an equal chance of being included in the sample. Transients, low income people, famous people with unlisted numbers, and especially women and children are underrepresented. Business or professional people with more than one phone are overrepresented. Many people now use only a cell phone, and those numbers are often omitted from phone books.

1.5 Independent variables: (a) First grade students who attended Kindergarten versus those who did not. (b) Seniors, Masters, Submasters, and Juniors as categories of marathon runners. Dependent variables: (a) Social-adjustment scores assigned by first-grade teachers. (b) Time to run 26 miles, 385 yards.

1.6 This experiment examined the difference in Length of hospital stay (dependent variable) among groups of new mothers using Traditional, Lamaze, or Leboyer delivery methods (independent variable).

1.7 Continuous variables: (a) Length of gestation. (b) Typing speed in words/minute. (c) Level of serotonin in a particular subcortical nucleus.

1.8 Discrete variables: (a) Number of siblings. (b) Political party affiliation. (c) Country of national origin.

1.9 The planners of a marathon race would like to know the average times of Senior, Master, Submaster, and Junior runners so as to facilitate planning for handling the finish line.

1.10 The promoters of a new diet plan want to know whether users of their plan lost more or less weight while following the plan than did another group of dieters following a different plan.

1

1.11 Categorical data: (a) The number of Brown University students in an October, 1984, referendum voting For and the number voting Against the university's stockpiling suicide pills in case of nuclear disaster. (b) The number of students in a small Midwestern college who are white, African-American, Hispanic-American, Asian, Native American, Alaskan Native, or Other. (c) One year after an experimental program to treat alcoholism, the number of participants who are "still on the wagon", "drinking without having sought additional treatment", or "again under treatment".

1.12 Measurement data: (a) Number of items endorsed on a 100-point scale of attitudes concerning capital punishment. (b) Weight (in grams) of newborn rat pups following maternal dietary restrictions. (c) Number of fatal traffic accidents in California during Labor Day weekend.

1.13 Children's scores in an inner-city elementary school could be reported numerically (a measurement variable), or the children could be categorized as Bluebirds ($\mathbf{X} > 90$), Robins ($\mathbf{X}$ = 70–90), or Cardinals ($\mathbf{X} < 70$).

1.14 Nominal: Brand of chocolate bars preferred. Ordinal: The finishing order of a group of runners. Interval: Prison inmates are rated from 1 to 5 on each of 20 behavioral items, producing a total score between 20 and 100. Ratio: The number of food pellets eaten by a hungry rat during the 24 hours after food deprivation ends.

1.15 For adults of a given height and sex, weight is a ratio scale of body weight, but it is **at best** an ordinal scale of physical health.

1.16 The implication is that speed is a poor measure of learning unless we assume that the animal that suddenly went to sleep had forgotten all he ever knew about the task.

1.17 Speed is probably a much better index of motivation than of learning.

1.18 (a) Vocational counselors would be interested in the relationship between high school students' scores on a vocational interest survey in 9th grade and their grades in vocational courses in 10th through 12th grades. (b) Marital counselors might investigate the relationship between quality of communication and rated marital satisfaction.

1.19 (a) The final grade point averages for low-achieving students taking courses that interested them could be compared with the averages of low-achieving students taking courses that don't interest them. (b) The quality of communication could be compared for happily versus unhappily married couples.

1.20 A good Web page for "internal validity" can be found at http://www.experiment-resources.com/internal-validity.html

1.21 For a synonym of "independent variable" see http://answers.yahoo.com/question /index?qid=20100919185852AAmMc57

2

1.22 Some discussion of this article can be found at http://www.dartmouth.edu/~chance/chance_news/recent_news/chance_news_4.10.html and http://www.dartmouth.edu/~chance/chance_news/recent_news/chance_news_4.12.html

1.23 An interesting study of the health effects of smoking in China can be found at http://www.berkeley.edu/news/media/releases/2005/09/04_smoking.shtml

3

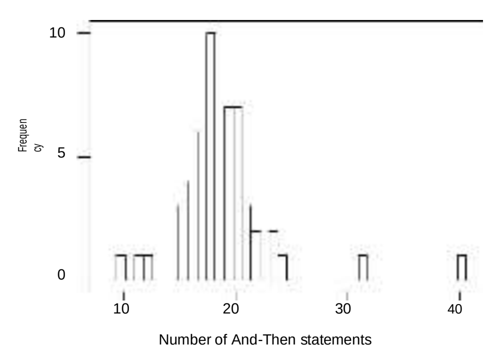
Chapter 2 - Describing and Exploring Data

2.1 Children's recall of stories:

a.

Children's "and then...s"	Frequency
10	1
11	1
12	1
15	3
16	4
17	6
18	10
19	7
20	7
21	3
22	2
23	2
24	1
31	1
40	1

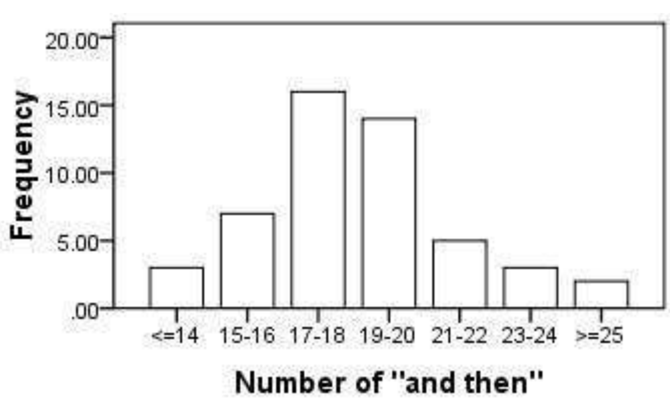
Distribution of And-Then statements for children



b. unimodal and positively skewed

2.2 For the data from Exercise 2.1—A second histogram:

Interval	Children's Midpoint	Frequency
≤ 14	13.5	3
15-16	15.5	7
17-18	17.5	16
19-20	19.5	14
21-22	21.5	5
23-24	23.5	3
≥ 25	25.5	2



4

2.3 The problem with making a stem-and-leaf display of the data in Exercise 2.1 is that almost all the values fall on only two leaves if we use the usual 10s' digits for stems.

Stem	Leaf
1	0125556666777777888888888889999999
2	000000011122334
3	1
4	0

And things aren't much better even if we double the number of stems.

Stem	Leaf
1*	012
1.	5556666777777888888888889999999
2*	000000011122334
2.	
3*	1
3.	
4*	0

Best might be to use the units digits for stems and add HI and LO for extreme values

Stem	Leaf
5	555
6	6666
7	7777777
8	8888888888
9	9999999
10	0000000
11	111
12	22
13	33
14	4
HI	31 40

2.4 Adults' recall of stories:

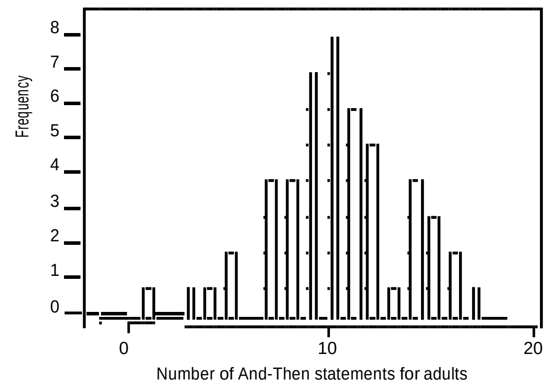
a. The scores for adults appear to be noticeably smaller. Adults seem to rely less strongly than do children on an "and then..." format for recalling stories.

5

b.Adults'

"and thens"	Freq.
1	1
3	1
4	1
5	2
7	4
8	4
9	7
10	8
11	6
12	5
13	1
14	4
15	3
16	2
17	1

Distribution of And-Then statements for adults



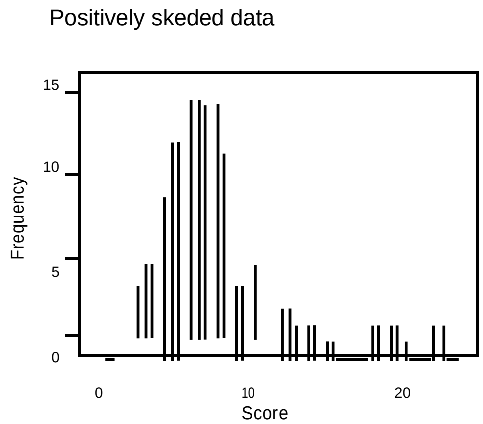
2.5 Stem-and-leaf diagram of the data in Exercises 2.1 and 2.4:

Children		**Adults**
	0*	1
	0t	34
	0f	55
	0s	7777
	0.	88889999999
10	1*	00000000111111
2	1t	222223
555	1f	4444555
7777776666	1s	667
77777778888888888	1.	
1110000000	2*	
3322	2t	
4	2f	
	2s	
	2.	
40 31	Hi	

6

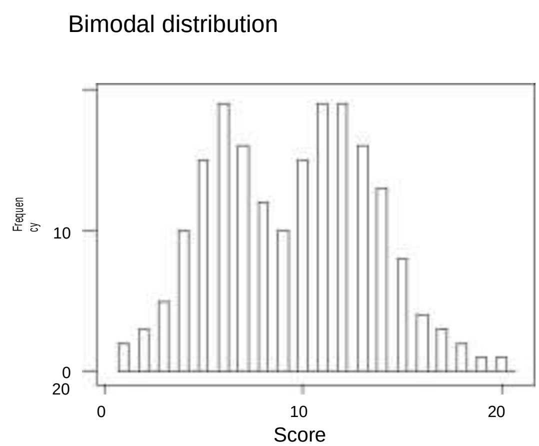
2.6 Invented positively skewed data:

Score	Freq.
1	2
2	3
3	5
4	10
5	13
6	15
7	14
8	11
9	5
10	5
11	4
12	3
13	2
14	2
15	1
18	2
19	2
20	1
22	2



2.7 Invented bimodal data:

Score	Freq
1	2
2	3
3	5
4	10
5	15
6	19
7	16
8	12
9	10
10	15
11	19
12	19
13	16
14	13
15	8
16	4
17	3
18	2
19	1
20	1



7

2.8 Although the median for males is slightly higher than the median for females, the center of the distribution has a smaller range for females. The overall range of the scores is roughly the same for both sexes.

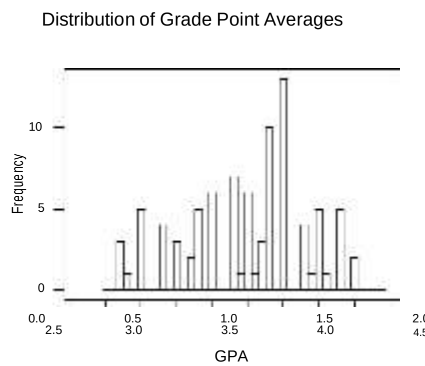
2.9 The first quartile for males is approximately 77, whereas for females it is about 80. The third quartiles are nearly equal for males and females, with a value of 87.

2.10 We can see that the data are more evenly distributed for males than for females. Females tend to be bunched more in the center of the distribution.

2.11 The shape of the distribution of number of movies attended per month for the next 200 people you met would be positively skewed with a peak at 0 movies per month and a sharp drop-off to essentially the baseline by about 5 movies per month.

2.12 Histogram for GPA:

GPA Interval	Midpoint	Frequency
.51–.75	.63	4
.76–1.00	.88	5
1.01–1.25	1.13	1
1.26–1.50	1.38	6
1.51–1.75	1.63	7
1.76–2.00	1.88	6
2.01–2.25	2.13	6
2.26–2.50	2.38	8
2.51–2.75	2.63	14
2.76–3.00	2.88	13
3.01–3.25	3.13	3
3.26–3.50	3.38	7
3.51–3.75	3.63	6
3.76–4.00	3.88	2



2.13 Stem-and-leaf for ADDSC

Stem	Leaf
2.	69
3*	0344
3.	56679
4*	00023344444
4.	5566677888899999
5*	00000000011223334
5.	55677889
6*	00012234
6.	55556899
7*	0024
7.	568
8*	
8.	55

8

2.14 a. X_3 9 X_5 10 X_8 8

b. X 1089...777

c. X_i $_{i\ 1}^{10}$

2.15 a. Y_1 9 Y_{10} 9

b. Y 99...257

2.16 a. X^2 10 8 ...

7 ^{2} 77 ^{2} 5929

X^2 10^2 8^2 ... 7^2 657

b. N X 10 77 7.7

c. The average, or the mean.

2.17 a. Y^2 (9 9 ... 2) ^{2} 3249

Y_2 9 9^2 ... 2^2 377

Y ^{2} Y_2 460 3249

b. N

N 1 10 5.789 9

c. $\sqrt{\text{answer to Exercise 17b}}$ $\sqrt{5.789}$ 2.406

d. The units of measurement were squared musicality scores in part (b) and musicality scores in part (c).

2.18 a. XY 10 9 8 9 ... 7 2 460

b. **X** **Y** = (77)(57) = 4389

XY X Y 460 4389

9

c. $\frac{10}{N-1} = 2.344 \cdot \frac{N}{9}$

10

2.19 a.

XY 109 89...72 134 X Y7757134

b.

XY 109 38 ... 37 460

X Y 77 57 4389

c.

CX 3X 310 38 ... 37 231

C X 377 231

d.

X^2 10^2 8^2 ... 7^2 657

X ^{2} 77^2 5929

2.20 Stem-and-leaf displays:

	1 Stimulus		3 Stimuli		5 Stimuli
3.	678899	3.		3.	9
4*	11122333444444	4*	23	4*	
4.	5555556666667777777889999	4.	666779	4.	6689
5*	111122222333444	5*	00111111122222223333333444	5*	13344
5.	566677778899	5.	556666667888889999999	5.	5555566778888899
6*	1124	6*	000000011111222222223333333444	6*	1111222222333444
6.	6777	6.	5566777799	6.	5555556666677777778899999
7*	112234	7*	22223344	7*	1122444
7.	69	7.	58	7.	566677889
8*		8*	3	8*	11233
8.		8.	6	8.	578
9*	4	9*		9*	4
9.		9.	5	9.	58
10*	44	10*		10*	
10.		10.		10.	
11*		11.		11*	
11.		11*		11.	
12*		12.		12*	
12.		12*		12.	5

As the number of digits in the comparison stimulus increases, the response time increases as well.

2.21 The results in Exercise 2.20 support the sequential processing hypothesis.

11

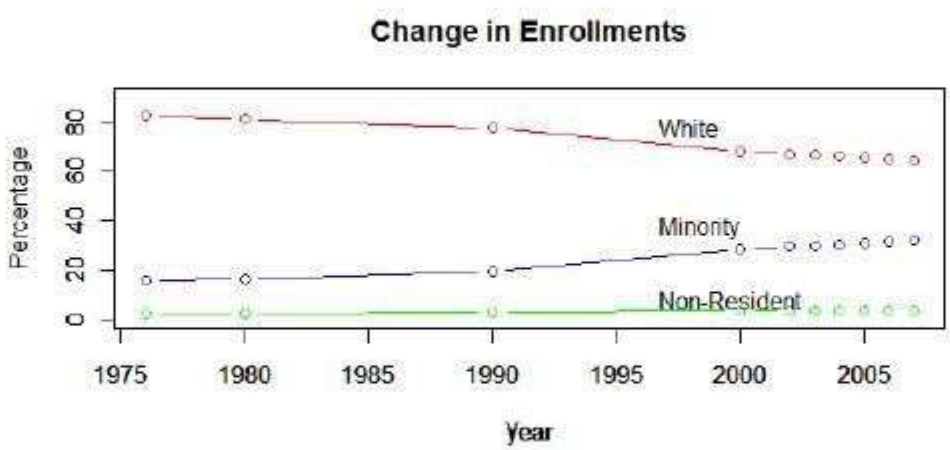
2.22 You could compare the reaction times for those cases in which the correct response was "Yes" and those cases in which it was "No." If we process information sequentially, the reaction times, on average, should be longer for the "No" condition than for the "Yes" condition because we would have to make comparisons against all stimuli in the comparison set. In the "Yes" condition we could stop as soon as we found a match.

2.23 The data are not likely to be independent observations because the subject is probably learning the task over the early trials, and later getting tired as the task progresses. Thus responses closer in time are more likely to be similar than responses further away in time.

2.24 For animals raised in a stable environment, there is little or no difference in immunity depending on Affiliation. However, for animals raised in an unstable environment, High Affiliation subjects showed much greater immunity than Low Affiliation subjects. Stability seems to protect against the negative effects of low affiliation.

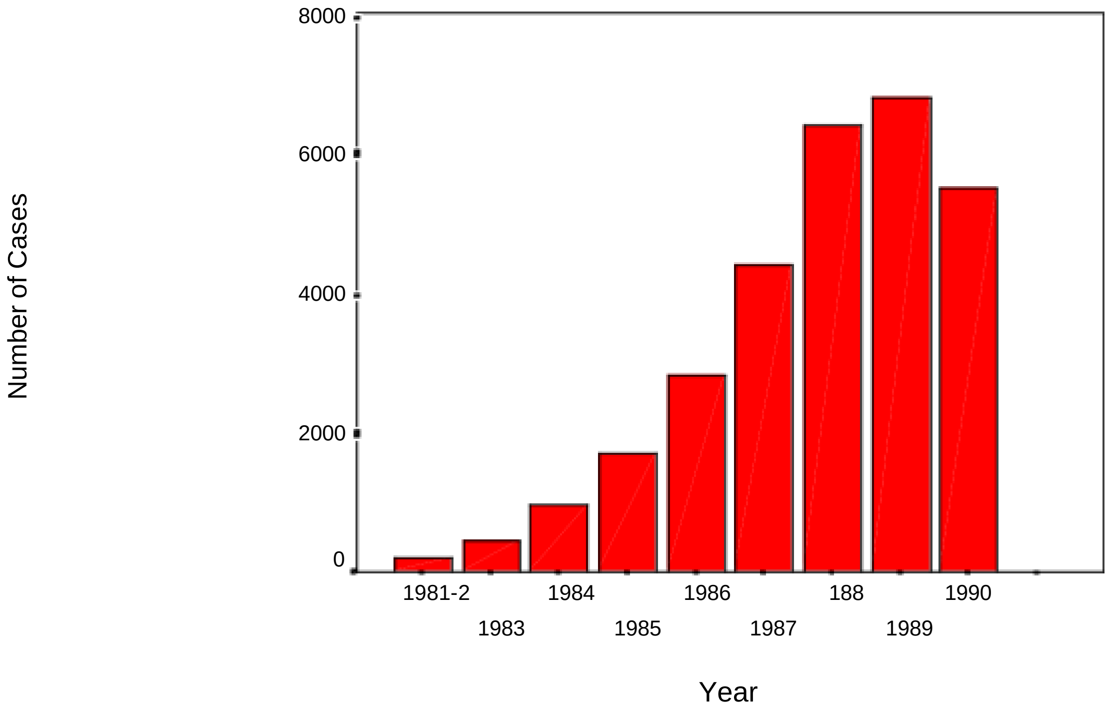
2.25 The amount of shock that a subject delivers to a white participant does not depend upon whether or not that subject has been insulted by the experimenter. On the other hand, black participants do suffer when the experimenter insults the subject.

2.26 There are any number of ways that these data could be plotted. Perhaps the simplest is to look at the change in the **percentages** of each major ethnic group's representation from 1976 to 2007.



12

2.27 AIDS cases among people aged 13–29 in U.S. population (in thousands):



2.28 Both the number of new cases and the number of people dying from AIDS is much lower in the U.S. than it is in the rest of the world except for Europe.

2.29 There is a strong increase in age at marriage, although the difference between males and females remains about the same. It is likely that what we are seeing is an increase in the percentage of couples living together without marrying. They finally get around to marrying about 5 years later than they used to.

2.30 Any positively skewed distribution will have a mean greater than the median.

2.31 The mean falls above the median.

2.32 Invented data: 1 9 10 15 15

Mean = Median = 10
Mode = 15

2.33 Rats running a straight alley maze:

$\bar{X} = \frac{\sum X}{N} = \frac{320}{15} = 21.33$; Median = 21

13

2.34 Subtracting a constant:

Original data (X):	8	7	12	14	3	7	$\bar{X}$ 8.5 Median = 7.5 Mode = 7
Transformed data (**Y** = **X** – 5):	3	2	7	9	-2	2	$\bar{Y}$ 3.5 Median = 2.5 Mode = 2

2.35 Multiplying by a constant:

Original data (X):	8	3	5	5	6	2	$\bar{Y}$ 4.83 Median = 5 Mode = 5
Transformed data (**Y** = 3**X**)	24	9	15	15	18	6	$\bar{Y}$ =14.5 Median = 15 Mode = 15

$3\bar{X} \quad \bar{Y}$	$3(Med_x) \quad Med_y$	$3(Mo_x) \quad Mo_y$
3(4.83) 14.5	3(5) 15	3(5) 15
14.5 14.5	15 15	15 15

2.36 Invented data with $\bar{X}$ 8.6 4 6 9 7 11 2 27 5 8 7

$$\bar{X} \quad \frac{86}{10} \quad 8.6$$

We had to be sure that the total came out to be 86. Therefore, one of the numbers was predetermined by the choice of the others.

2.37 They look just the way I would have expected.

2.38 The distribution of GPA is somewhat asymmetric, with Bs and Cs predominating. There were a few students with an A average, but none who failed all of their courses.

2.39 Computer exercise

2.40 For the data in Exercise 2.1:

range 40 10 30

variance = $s_X^2 X \quad \bar{X}^2 \quad 18 \quad 18.9^2 \quad 15 \quad 18.9^2 \quad ... \quad 16 \quad 18.9^2$
20.214

standard deviation $s_X \quad \sqrt{s_X^2} \quad \sqrt{20.214}$ 4.496

2.41 For the data in Exercise 2.4:

range 17 1 16

variance = $s_X^2 X \quad \bar{X}^2 \quad 10 \quad 10.2^2 \quad 12 \quad 10.2^2 \quad ... \quad 9 \quad 10.2^2$
11.592

standard deviation $s_X \quad \sqrt{s_X^2} \quad \sqrt{11.592}$ 3.405

2.42 The two standard deviations are roughly the same, although the range for the children is about twice the range for the adults.

2.43 For the data in Exercise 2.1:

The interval:
$X \; 2s_X \quad 18.9 \quad 2 \; 4.496 \quad 18.9 \; 8.992 \quad 9.908$ to 27.892

From the frequency distribution in Exercise 2.1 we can see that all but two scores (31 and 40) fall in this interval, therefore 48/50 = 96% of the scores fall in this interval.

2.44 For the data in Exercise 2.4:

The interval:
$X \; 2s_X \quad 10.2 \quad 2 \; 3.405 \quad 10.2 \quad 6.81 \quad 3.39$ to 17.01

From the frequency distribution in Exercise 2.4 we can see that all but two scores (1 and 3) fall in this interval, therefore 48/50 = 96% of the scores fall in this interval.

2.45 Original data: 5 8386997

$s_1 = 2.1$
If $\mathbf{X}_2 = \mathbf{c}\mathbf{X}_1$, then $s_2 = \mathbf{c}s_1$ and we want $s_2 = 1.00$

$s_2 = \mathbf{c}s_1$
$1 = \mathbf{c}(2.1)$
$\mathbf{c} = 1/(2.1)$

Therefore we want to divide the original scores by 2.1

$X_2 \quad \frac{X_1}{2.1}$ 2.3813.809 1.428 3.809 2.857 4.286 4.286 3.333

$s_2 = 1$

2.46 Boxplot for data in Exercise 2.4 [Refer to data in Exercise 2.4 and cumulative distribution in Exercise 2.7]

Median location = (**N** + 1)/2 = 51/2 = 25.5
Median = 10
Hinge location = (Median location + 1)/2 = (25 + 1)/2 = 26/2 = 13
Hinges = 8 and 12
H-spread = 12 - 8 = 4
Inner fences = Hinges $\pm$ 1.5*(H-spread)
=12+1.5(4)=12+6=18
and =8–1.5(4)=8–6=2
Adjacent values = 3 and 17

```
1    2  3    4   5   6    7    8     9  10   11   12   13   14    15  16   17
+---+  +   +   +   +   +   +   +   +   +   +   +   +   +   +  +-------------------------------------------
                                   ____________
*         ------------            |      |     |-----------
                                   ____________
```

2.47 Boxplot for ADDSC [Refer to stem-and-leaf in Exercise 2.15]:

Median location = (**N** + 1)/2 = (88 + 1)/2 = 89/2 = 44.5
Median = 50
Hinge location = (Median location + 1)/2 = (44 + 1)/2 = 45/2 = 22.5
Hinges = 44.5 and 60.5
H-spread = 60.5 - 44.5 = 16
Inner fences = Hinges $\pm$ 1.5*(H-spread)
= 60.5 + 1.5(16) = 60.5 + 24 = 84.5
and = 44.5 – 1.5(16) = 44.5 – 24 = 20.5
Adjacent values = 26 and 78

```
20          30          40          50         60           70          80          90
+---- +----+---- +----+----+---- +---- +----+---- +----+----+---- +----+----  +
                                   ____________
      ----------          |     |      |-------------------       **
                                   ____________
```


15

2.48 Coefficient of variation for Exercises 2.1 and 2.4: For Exercise 2.1 $cv = s / \bar{X} = 4.496 / 18.9 = 0.238$

For Exercise 2.4 $cv = s / \bar{X} = 3.405 / 10.2 = 0.334$

The adult sample shows somewhat greater variability when its smaller mean is taken into account.

2.49 Coefficient of variation for Appendix Data Set $s/\bar{X}$ = 0.8614/2.456 = 0.351

2.50 This is an Internet search question.

2.51 10% trimmed means of data in Table 2.6

3.13 3.17 3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34 3.36 3.38

Ten percent trimming would remove the two extreme observations at either end of the distribution, leaving

3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34

$$\bar{X} = \frac{52.28}{16} = 32.675$$

In this case the trimmed mean is very close to the untrimmed mean (3.266).

2.52 10% Winsorized standard deviation of data in Table 2.6

3.13 3.17 3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34 3.36 3.38

Ten percent Windsorizing would replace the two lowest observations with 3.19 and the two highest observations with 3.34. This leaves

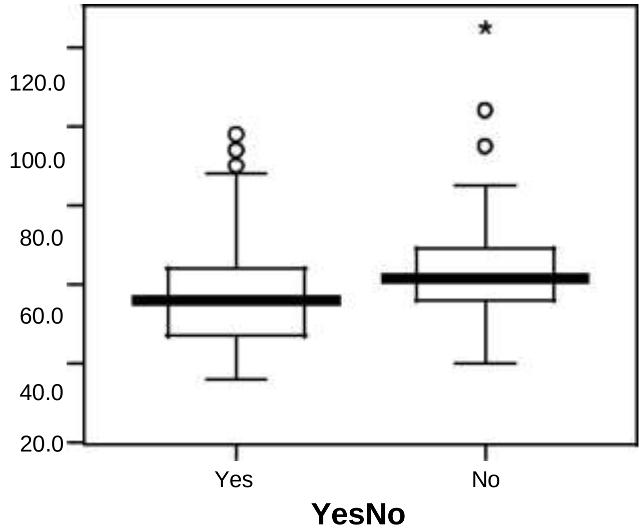
3.19 3.19 3.19 3.19 3.20 3.20 3.22 3.23 3.25 3.26

3.27 3.29 3.29 3.30 3.31 3.31 3.34 3.34 3.34 3.34

The standard deviation of the Winsorized sample is 0.058, whereas the standard deviation of the original sample was 0.069.

16

2.53 Reaction times when stimulus was present or absent.



2.54 A transformation will alter the shape of a distribution when it is a nonlinear transformation.

2.55 This is an Internet search that has no fixed answer.

17