

## Chapter 2 - Describing and Exploring Data

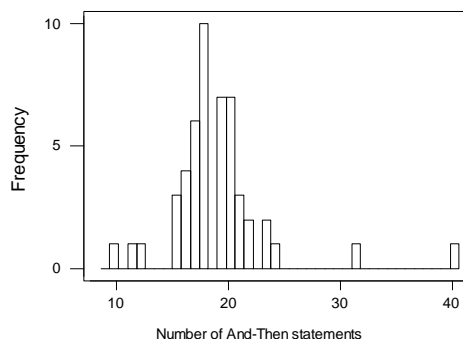
---

### 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

2.3 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.

<u>Stem</u>	<u>Leaf</u>
1	0125556666777777888888888889999999
2	000000011122334
3	1
4	0

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

<u>Stem</u>	<u>Leaf</u>
1*	012
1.	555666677777788888888889999999
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

<u>Stem</u>	<u>Leaf</u>
5	555

```

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

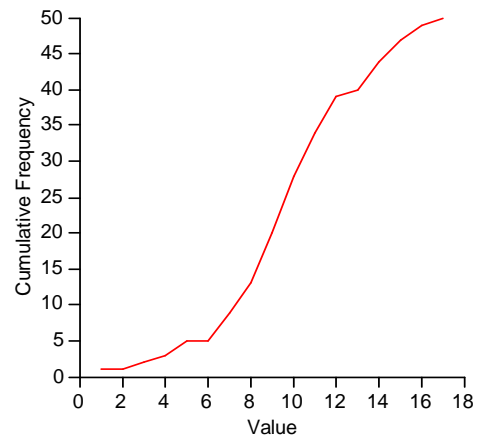
```

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

<u>Children</u>		<u>Adults</u>
	0*	1
	0t	34
	0f	55
	0s	7777
	0.	88889999999
10	1*	000000001111111
2	1t	222223
555	1f	4444555
7777776666	1s	667
777777788888888888	1.	
1110000000	2*	
3322	2t	
4	2f	
	2s	
	2.	
40 31	Hi	

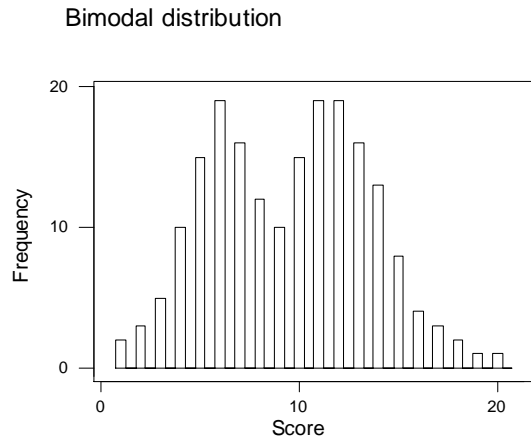
**2.7** Cumulative frequency distribution for data in Exercise 2-4:

Adult Scores	Frequ- ency	Cumul. Frequency
1	1	1
3	1	2
4	1	3
5	2	5
7	4	9
8	4	13
9	7	20
10	8	28
11	6	34
12	5	39
13	1	40
14	4	44
15	3	47
16	2	49
17	1	50



**2.9** Invented bimodal data:

<u>Score</u>	<u>Freq</u>
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



**2.11** 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.

**2.13** 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.

**2.15** Stem-and-leaf for ADDSC

<u>Stem</u>	<u>Leaf</u>
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

**2.17 a.**  $X_3 = 9$   $X_5 = 10$   $X_8 = 8$

$$\Sigma X = 10 + 8 + 9 + \dots + 7 = 77$$

b.

c.  $\sum_{i=1}^{10} X_i$

2.19 a.  $(\sum X)^2 = (10 + 8 + \dots + 7)^2 = (77)^2 = 5929$   
 $\sum X^2 = 10^2 + 8^2 + \dots + 7^2 = 657$

b.  $\frac{\sum X}{N} = \frac{77}{10} = 7.7$

c. The average, or the mean.

2.21 a.  $\sum XY = (10)(9) + (8)(9) + \dots + (7)(2) = 460$

b.  $\sum X \sum Y = (77)(57) = 4389$

c.  $\frac{\sum XY - \frac{\sum X \sum Y}{N}}{N - 1} = \frac{460 - \frac{4389}{10}}{9} = 2.344$

2.23 Stem-and-leaf displays:

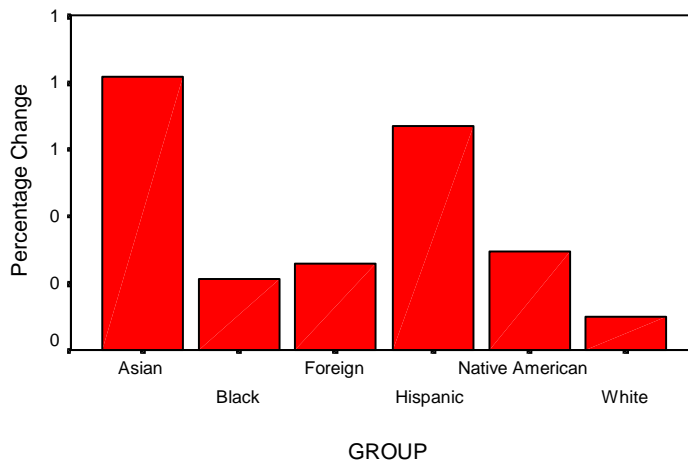
1 Stimulus		3 Stimuli		5 Stimuli	
3.	678899	3.	23	3.	9
4*	11122333344444	4*	666779	4*	6689
4.	55555666666777777889999	4.	001111111222222333333444	4.	13344
5*	111122222333444	5*	5566666788888999999	5*	555566778888899
5.	566677778899	5.	0000001111122222233333444	5.	111122222333444
6*	1124	6*	5566777799	6*	55555666667777778899999
6.	6777	6.	22223344	6.	1122444
7*	112234	7*	58	7*	566677889
7.	69	7.	3	7.	11233
8*		8*	6	8*	578
8.		8.	9*	9*	4
9*	4	9*	5	9.	58
9.		9.	10*	10*	
10*	44	10*	10.	10.	
10.		10.	11.	11*	
11*		11.	11*	11.	
11.		11*	12.	12*	
12*		12.	12*	12.	5
12.		12*			

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

- 2.25** 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.
- 2.27** 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.
- 2.29** 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 ethnic group’s representation from 1982 to 1991.

#### Change in Ethnic Distribution in U. S. Colleges

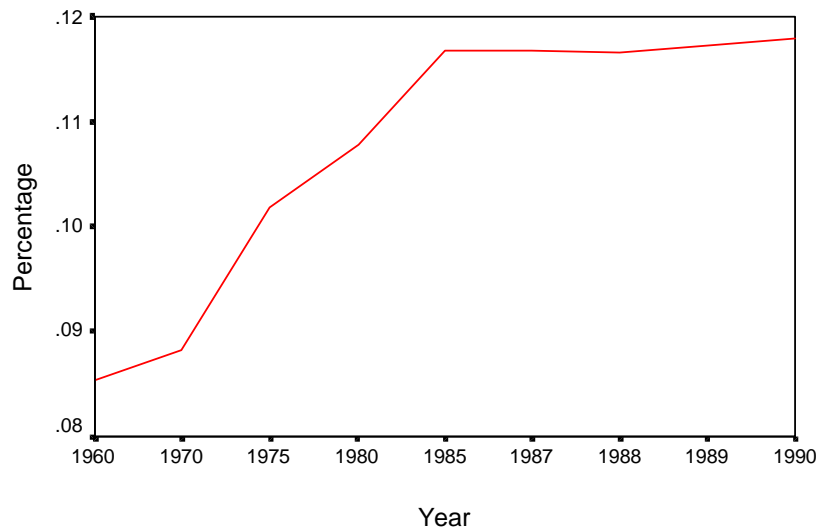
1982 to 1991



- 2.31** One way to look at these data is to plot the percentage of households headed by women and the family size separately against years. Notice that there is an uneven sampling of years.

## Percent households headed by women

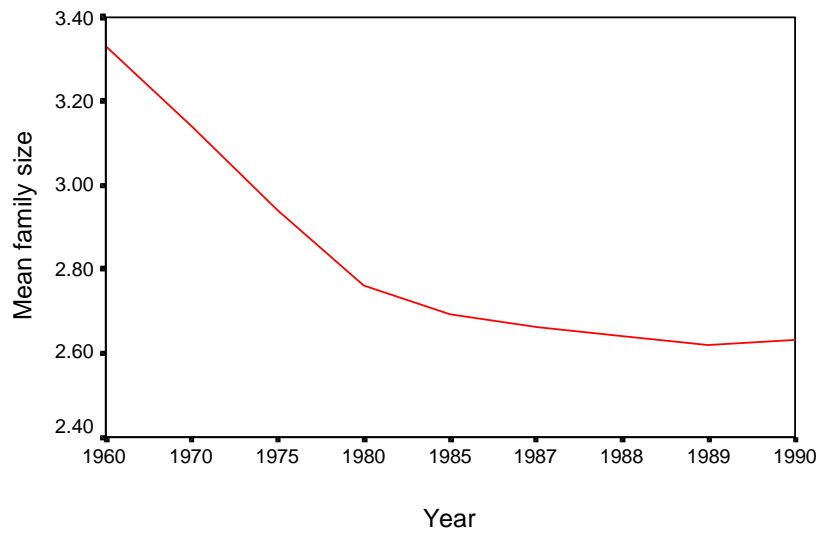
1960 - 1990



- There has been a dramatic increase in the percentage of households headed by women over the past 10 years.
- There has also been a corresponding decrease in family size, part of which is perhaps due to the increase in single-parent families.

## Family size

1960 - 1990



**2.33** The mean falls above the median.

**2.35** Rats running a straight alley maze:

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

**2.37** Multiplying by a constant:

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

$3\bar{X} = \bar{Y}$	$3(\text{Med}_x) = \text{Med}_y$	$3(\text{Mo}_x) = \text{Mo}_y$
$3(4.83) = 14.5$	$3(5) = 15$	$3(5) = 15$
$14.5 = 14.5$	$15 = 15$	$15 = 15$

**2.39** Computer printout

**2.41** For the data in Exercise 2.1:

$$\text{range} = 40 - 10 = 30$$

$$\begin{aligned} \text{variance} &= s_x^2 = \Sigma (X - \bar{X})^2 = (18 - 18.9)^2 + (15 - 18.9)^2 + \dots + (16 - 18.9)^2 \\ &= 20.214 \end{aligned}$$

$$\text{standard deviation} = s_x = \sqrt{s_x^2} = \sqrt{20.214} = 4.496$$

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

**2.45** For the data in Exercise 2.4:

$$\text{The interval: } \bar{X} \pm 2s_x = 10.2 \pm 2(3.405) = 10.2 \pm 6.81 = 3.39 \text{ to } 17.01$$

From the frequency distribution produced in Exercise 2.4 you 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.

**2.47** Original data: 2 3 4 4 5 5 9  $\bar{X}_1 = 4.57$   $s_1 = 2.23$   
(reordered)

$X_2 = 2X_1$ : 4 6 8 8 10 10 18  $\bar{X}_2 = 9.14$   $s_2 = 4.45$

$X_3 = X_1/2$ : 1 1.5 2 2 2.5 2.5 4.5  $\bar{X}_3 = 2.286$   $s_3 = 1.11$

Multiplying (or dividing) a distribution by a constant multiplies (or divides) the mean of that distribution by that constant. Here we find that the standard deviation of that distribution is multiplied (or divided) by that constant. (The variance is multiplied (or divided) by the square of that constant.)

Adding (or subtracting) a constant to (or from) a distribution adds (or subtracts) that constant from the mean of that distribution. Here we find that the standard deviation of that distribution is unchanged.

**2.49**  $X_2$  from Exercise 2.48: 2.381 3.809 1.428 3.809 2.857 4.286 4.286  
3.333

If  $X_3 = X_2 + c$ , then  $\bar{X}_3 = \bar{X}_2 + c$ , and we want  $\bar{X}_3 = 0$

$$\bar{X}_2 = \frac{26.19}{8} = 3.274$$

$$\bar{X}_3 = \bar{X}_2 + C$$

$$0 = 3.274 + C$$

$$-3.274 = C$$

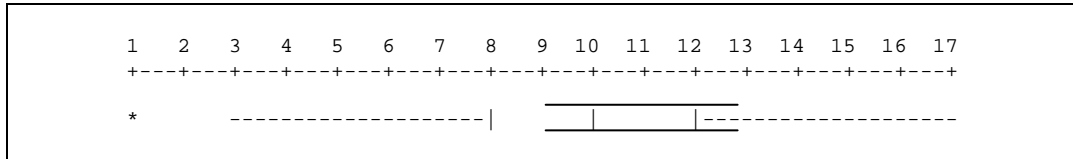
Therefore we want to subtract 3.274 from the  $X_2$  scores:

$X_3 = X_2 - 3.274$ : -0.893 0.535 -1.846 0.535 -0.417 1.012 1.012 0.059

$$\bar{X}_3 = 0$$

**2.51** 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 =  $(\text{Median location} + 1)/2 = (25 + 1)/2 = 26/2 = 13$   
 Hinges = 8 and 12  
 H-spread =  $12 - 8 = 4$   
 Inner fences = Hinges  $\pm 1.5*(\text{H-spread})$   
                   =  $12 + 1.5(4) = 12 + 6 = 18$   
                   and =  $8 - 1.5(4) = 8 - 6 = 2$   
 Adjacent values = 3 and 17



**2.53** 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.

**2.55** The answers are listed in a file at [www.uvm.edu/~dhowell/methods6/DataFiles/BadCancr.err](http://www.uvm.edu/~dhowell/methods6/DataFiles/BadCancr.err).

**2.57** 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 Winsorizing 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.

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

