STAT 327

<u>Solutions to Assignment #1</u>

(2.8) (a) quantitative

(b) discrete

(c) proportions: 0.270, 0.158, 0.255, 0.164, 0.086, 0.031, 0.018, 0.010, .008

percentages: 27%, 15.8%, 25.5%, 16.4%, 8.6%, 3.1%, 1.8%, 1.0%, 0.8%

(d) mode = 0

<u>Note</u>: The <u>mode</u> is the value with highest frequency.

(2.15) (a)

```
17 | 8
18 |
19 | 9
20 | 0
21 | 0 0
22 | 5 5 8 5 5
23 | 5 2
24 | 0 9 0 6 6 0 5 6
25 | 0 0 0 5 5
```

(b) Prices fall between $178 and $255, with more falling at the upper end of the range.

(2.16) (a) no. of students = 33 (just count the "leaves")

min = 65, max = 98

(c) Use the intervals 60-69, 70-79, 80-89, 90-99.
The resulting histogram will look like the stem-and-leaf plot ~~321~~ rotated 90°, with "leaves" replaced by rectangles.

2.17 (a)
```
1| 3 3 3 3 4 4 5 6 7 7 7 7 8 8 9 9
2| 0 4
```

The plot is too compact, making it difficult to see where the data fall.

(c)
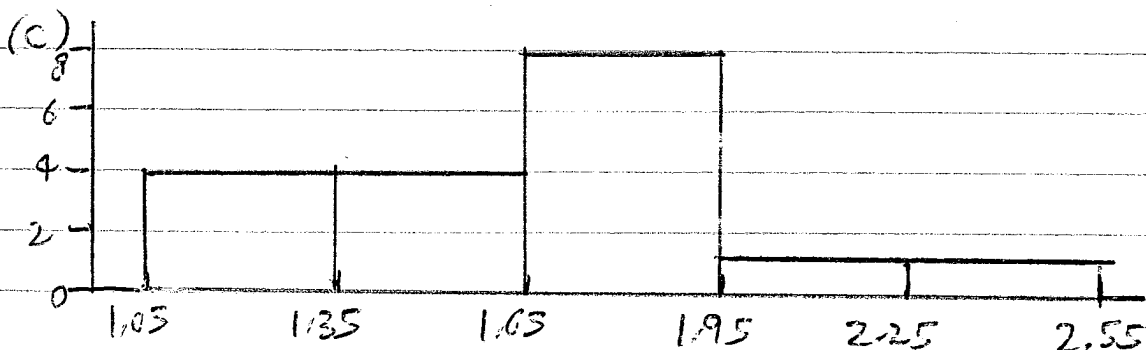
2.32 (a) The median for all three data sets is 10. (b) Set 1: $\bar{x} = 10$; Set 2: $\bar{x} = 27.6$; Set 3: $\bar{x} = 207.6$ (c) The median is unaffected by the outlier, the mean gets higher and higher as the outlier becomes more extreme.

2.37 (a) Mean: 2; median: 0; and mode: 0 (b) Mean: 10; median: 0. Outlier affects mean but not median.

2.45 (a) 0.17; (b) 0; (c) Median: 0; mean: 1.95; (d) The median is the same for both because the median ignores much of the data.

2.47 (a) Range = 6; six days separate the fewest and most sick days taken. (b) $s = 2.38$; the typical distance of an observation from the mean is 2.38. (c) Range = 60; standard deviation = 21.06. Both increase when an outlier is added

2.48 (a) Africa because the spread of the data is greater. (b) Western Europe: $s = 1.0$; Africa: $s = 7.1$.

2.64 (a) 6.7. (b) Q1 = 4.5; Q3 = 7.8. (c) 6.28.

2.70 (a) Q1 = 76; Q3 = 101, (b) IQR = 25,

(2.108) mean = 30.4, std dev = 16.7

(2.118) ordered observations are:

2, 6, 14, 19, 23, 26, 28, 37, 47, 80

(a) stem-and-leaf:

```
0 | 2 6
1 | 4 9
2 | 3 8 8
3 | 7
4 | 7
5 |
6 |
7 |
8 | 0
```

(b) median = $\frac{23+26}{2}$ = 24.5

1st quartile = median of $\{2, 6, 14, 19, 23\}$ = 14

3rd quartile = median of $\{26, 28, 37, 47, 80\}$ = 37

(c) IQR = 37 − 14 = 23, 1.5 IQR = 34.5   37 + 34.5 = 71.5

min   Q1   Med   Q3   Q3 + 1.5 IQR   Vmax
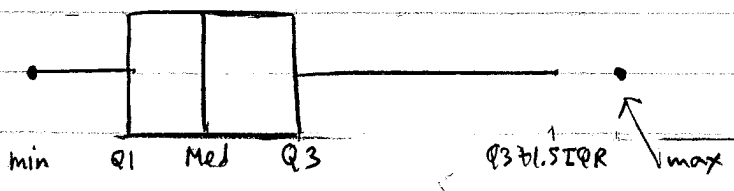
The formula for r presented in class was:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} .$$

Starting with this formula, and using the algebraic identity $n - 1 = \sqrt{n-1} \cdot \sqrt{n-1}$ and the definitions

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{and} \quad s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} , \quad \text{we obtain}$$
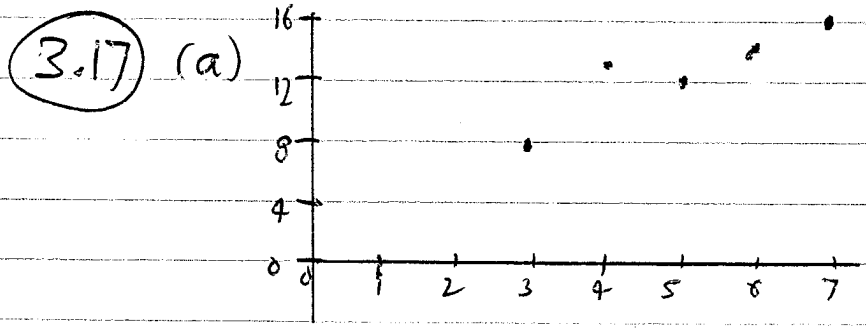
$$r = \frac{1}{n-1} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}} = \frac{1}{n-1} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \, s_y}$$

$$= \frac{1}{n-1} \sum \frac{(x_i - \bar{x})}{s_x} \frac{(y_i - \bar{y})}{s_y} , \quad \text{the formula in the text.}$$

Starting with the formula for b given in the notes, we obtain

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{s_x^2}$$

$$= \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{s_x \, s_y} \cdot \frac{s_y}{s_x} = r \frac{s_y}{s_x} \quad \longleftarrow \text{the formula in the text.}$$

(3.16) 1(c); 2(a); 3(d); 4(b)

(3.17) (a)

(b) (4, 13)

(c) Change the value 13 to 10.

(3.21) Solution omitted (too time-consuming!)

(3.24) (a) $a = 7, b = 0.5$
(b) $a = 7, b = 1$
(c) $a = 7, b = -1$
(d) $a = 7, b = 0.$

(3.25)

(3.25)(a) (i) 6.47;
(ii) 5.75. Connect the points ($x = 10$, $y = 6.47$) and ($x = 40$, $y = 5.75$). (b) The y-intercept indicates that when a person cannot do any sit-ups, she/he would be predicted to run the 40-yard dash in 6.71 seconds. The slope indicates that every increase of one sit-up leads to a decrease in running time of 0.024 seconds. (c) The slope indicates a negative correlation (the slope and correlation have the same sign.)