Homework 1: Cardiovascular Disease

1. The mean is

$$\overline{x} = \frac{\sum x_i}{24} = \frac{469}{24} = 19.54 \text{ mg/dL}$$

2. We have that

$$s^2 = \frac{\sum_{i=1}^{24}(x_i - \overline{x})^2}{23} = \frac{(49-19.54)^2 + \cdots + (12-19.54)^2}{23} = \frac{6495.96}{23} = 282.43$$
$$s = \sqrt{282.43} = 16.81 \text{ mg/dL}$$

3. We provide two rows for each stem corresponding to leaves 5-9 and 0-4 respectively. We have

Stem-and-leaf plot		Cumulative frequency
+4	98	24
+4	1	22
+3	65	21
+3	21	19
+2	78	17
+2	13	15
+1	9699	13
+1	332	9
+0	88	6
+0	2	4
–0		
–0	8	3
–1	03	2

4. We wish to compute the average of the (24/2)th and (24/2 + 1)th largest values = average of the 12th and 13th largest points. We note from the stem-and-leaf plot that the 13th largest point counting from the bottom is the largest value in the upper +1 row = 19. The 12th largest point = the next largest value in this row = 19. Thus, the median $= \frac{19+19}{2} = 19$ mg/dL.

5. We first must compute the upper and lower quartiles. Because 24(75/100) =18 is an integer, the upper quartile = average of the 18th and 19th largest values $= \frac{32+31}{2} = 31.5$. Similarly, because 24(25/100) = 6 is an integer, the lower quartile average of the 6th and 7th smallest points $= \frac{8+12}{2} = 10$.
Second, we identify outlying values. An outlying value is identified as any value x such that

x > upper quartile + 1.5 × (upper quartile - lower quartile)
= 31.5 + 1.5 × (31.5 -10)
= 31.5 + 32.25
= 63.75

or

x < lower quartile – 1.5 ×(upper quartile - lower quartile)
= 10 – 1.5 × (31.5 – 10)
= 10 – 32.25
= -22.25

From the stem-and-leaf plot, we note that the range is from -13 to +49. Therefore, there are no outlying values. Thus, the box plot is as follows:

<table>
<tr><th colspan="2">Stem-and-leaf plot</th><th>Cumulative frequency</th><th>Box plot</th></tr>
<tr><td>+4</td><td>98</td><td>24</td><td>|</td></tr>
<tr><td>+4</td><td>1</td><td>22</td><td>|</td></tr>
<tr><td>+3</td><td>65</td><td>21</td><td>|</td></tr>
<tr><td>+3</td><td>21</td><td>19</td><td>+------+</td></tr>
<tr><td>+2</td><td>78</td><td>17</td><td>| |</td></tr>
<tr><td>+2</td><td>13</td><td>15</td><td>| |</td></tr>
<tr><td>+1</td><td>9699</td><td>13</td><td>*--+--*</td></tr>
<tr><td>+1</td><td>332</td><td>9</td><td>+------+</td></tr>
<tr><td>+0</td><td>88</td><td>6</td><td>|</td></tr>
<tr><td>+0</td><td>2</td><td>4</td><td>|</td></tr>
<tr><td>–0</td><td></td><td></td><td>|</td></tr>
<tr><td>–0</td><td>8</td><td>3</td><td>|</td></tr>
<tr><td>–1</td><td>03</td><td>2</td><td>|</td></tr>
</table>

Comments: The distribution is reasonably symmetric, since the mean = 19.54 mg/dL = 19 mg/dL = median. This is also manifested by the percentiles of the distribution since the **upper quartile - median** = 31.5 -19 = 12.5 = **median - lower quartile** = 19 -10 = 9 . The box plot looks deceptively asymmetric, since 19 is the highest value in the **upper + 1 row** and 10 is the lowest value in the **lower + 1 row**.

6. To compute the median cholesterol level, we construct a stem-and-leaf plot of the before-cholesterol measurements as follows.

Stem-and-leaf plot		Cumulative frequency
25	0	24
24	4	23
23	68	22
22	42	20
21		
20	5	18
19	5277	17
18	0	13
17	8	12
16	698871	11
15	981	5
14	5	2
13	7	1

Based on the cumulative frequency column, we see that the median = average of the 12th and 13th largest values = $\frac{178+180}{2} = 179$ mg/dL. Therefore, we look at the change scores among persons with baseline cholesterol ≥ 179 mg/dL and < 179 mg/dL, respectively. A stem-and-leaf plot of the change scores in these two groups is given as follows:

Baseline ≥ 179 mg/dL		Baseline < 179 mg/dL	
Stem-and-leaf plot		Stem-and-leaf plot	
+4	98	+4	
+4		+4	1
+3	65	+3	
+3	2	+3	1
+2	78	+2	
+2	1	+2	3
+1	699	+1	9
+1		+1	332
+0	8	+0	8
+0		+0	2
–0		–0	
–0		–0	8
–1		–1	03

Clearly, from the plot, the effect of diet on cholesterol is much greater among individuals who start with relatively high cholesterol levels (≥ 179 mg/dL) versus those who start with relatively low levels (< 179 mg/dL). This is also evidenced by the mean change in cholesterol levels in the two groups, which is 28.2 mg/dL in the ≥ 179 mg/dL group and 10.9 mg/dL in the < 179 mg/dL group.