Homework Solutions 1

1 Exercise 1.6

a. The modal category in this case is 2 (quarts of milk). About 36% (9 people) of the 25 sampled fell into this category.

b. The proportion of people who purchased 3,4 and 5 quarts of milk are .2,.12, and .04 respectively. Therefore the answer is .2+.12+.04=.36

c. Note that 8% of the peolple purchased 0 while 4% purchased 5. Thus a total of $8\% + 4\% = 12\%$ purchased 0 or 5. Therefore 1-.12=.88 of the people purchased between 1 and 4 quarts of milk.

2 Exercise 1.7

a. Note that 9.7= 12-(1)2.3 and 14.3=12+(1)2.3. Therefore the interval (9.7,14.3) represents breathing rates within 1 standard deviation of the mean. According to the empirical rule approximately 68% of the college students should have breathing rates in this interval.

b. Note that 7.4=12-(2)2.3 and 16.6=12+(2)2.3 therefore we are now interested in the percentage of college students with breathing rates within 2 standard deviations of the mean. According to the empirical rule this percentage should be around 95%.

c. We know that 68% of students should have breathing rates between 9.7 and 14.3 (by part a). We also know 95% of students should have breathing rates between 7.4 and 16.6 (by part b). This leaves (95-68)=27 to lie between both 14.3 and 16.6 and 9.7 and 7.4. By symmetry then 13.5%=27 should lie between 14.3 and 16.6. Therefore, 68+13.5=81.5% of college students should have breathing rates between 9.7 and 16.6.

d. Note that 5.1=12-(3)2.3 and 18.9=12+(3)2.3 therefore we are interested in the proportion of college students that have breathing rates outside of 3 standard deviations of the mean. According to the empirical rule, this should be approximately 0.

3 Exercise 1.9

a. $\sum_{i=1}^{n} = c + c + ... + c$, where the sum involves n elements. Hence $\sum_{i=1}^{n} = nc$.

b. $\sum_{i=1}^{n} cy_i = cy_1 + ... + cy_n = c\sum_{i=1}^{n} y_i$.

c. $\sum_{i=1}^{n}(x_i + y_i) = (x_1 + ... + x_n) + (y_1 + ... + y_n) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$.
Consider the numerator of s^2, which is $\sum_{i=1}^{n}(y_i - \bar{y})^2$.

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) = \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} 2y_i\bar{y} + \sum_{i=1}^{n} \bar{y}^2$$

$\bar{y}$ and $\bar{y}^2$ are constant with respect to the variable of summation (i). Hence

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y}(n\bar{y}) + n\bar{y}^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

with the second equality following from the fact that $\sum_{i=1}^{n} y_i = n\bar{y}$.
Thus $s^2 = \frac{1}{n-1}(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)$ and we know $\bar{y}^2 = \frac{1}{n}(\sum_{i=1}^{n} y_i)^2$, thus we get the solution.

4 Exercise 1.15

For exercise 1.2 the approximation is:
$\frac{range}{4} = \frac{3168-565}{4} = 650.75$ while s=393.75.
Note the poor approximation due to the extreme values. For exercise 1.3, the approximation is: $\frac{range}{4} = \frac{12.48-0.32}{4} = 3.04$, while s=3.17.
For exercise 1.4, the approximation is $\frac{range}{4} = \frac{38.3-1.8}{4} = 9.125$, while s=7.48.

5 Exercise 1.20

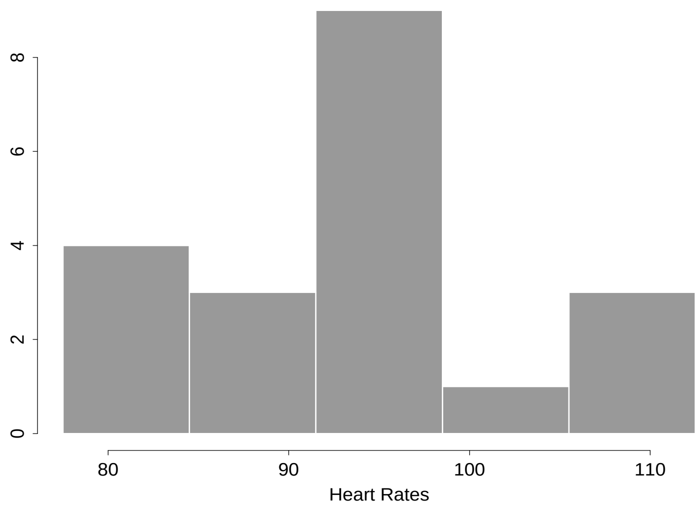
$\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} y_i - n\bar{y} = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i = 0$

6 Exercise 1.22

a. s is approximately equal to $\frac{range}{4}$ which equals $\frac{112-78}{4} = 8.5$.

b. Each student will obtain a slightly different frequency histogram. As an example chhose five intervals of length 7.

Class Boundaries	Frequency	Relative Frequency
77.5 - 84.5	4	.21
84.5 - 91.5	2	.11
91.5 - 98.5	9	.47
98.5 -105.5	1	.05
105.5 -112.5	3	.16



From the histogram, $\overline{y}$ appears to be about 95 and s appears to be about 10.

c. Calculate first, $\sum_{i=1}^{20} y_i = 1874.0$ and $\sum_{i=1}^{20} y_i^2 = 117328.0$. Then $\bar{y} = 93.7$ and s=9.55.

d.

k	$\overline{y} \pm ks$	Interval Boundaries	Freq	Expected Freq
1	93.7 ± 9.55	84.1 to 103.2	13	0.65
2	93.7 ± 19.11	74.6 to 112.8	20	1.00
3	93.7 ± 28.66	65.0 to 122.4	20	1.00

These results are reasonably consistant with the empirical rule.

7 Extra exercise

a. The interquartile range is 3.5-2=1.5

b. The median GPA for those watching less than 20 hourse per week is 3.6

c. The mean GPA for those watching less than 20 hourse per week is 3.377777

d. The data are left skewed.