

# Wackerly Section 1.3 Problems

John Nguyen

April 4, 2021

## Problem 1.9

- (a). By the Empirical Rule, about 0.68.
- (b). By the Empirical Rule, about 0.95.
- (c). According to the Empirical Rule, about 0.765.
- (d). According to the Empirical Rule, almost 0.0.

## Problem 1.10

- (a).  $14 - 17 = -3$
- (b). Assuming time spent online using the Internet is approximately normally distributed, about 0.16 of people spend less than -3 hours online per year.
- (c). No, because it is impossible to spend less than 0 hours online per year. Thus the normal distribution assumption is false.

## Problem 1.11

$$\begin{aligned}s^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2)\end{aligned}$$

Using (c),

$$\begin{aligned}\frac{1}{n-1} \sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2) &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \sum_{i=1}^n (-2y_i\bar{y} + \bar{y}^2) \right] \\ &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \sum_{i=1}^n (-2y_i\bar{y}) + \sum_{i=1}^n \bar{y}^2 \right]\end{aligned}$$

Using (a),

$$\frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \sum_{i=1}^n (-2y_i\bar{y}) + \sum_{i=1}^n \bar{y}^2 \right] = \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \sum_{i=1}^n (-2y_i\bar{y}) + n\bar{y}^2 \right]$$

Using (b),

$$\begin{aligned} \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \sum_{i=1}^n (-2y_i \bar{y}) + n\bar{y}^2 \right] &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \bar{y} \sum_{i=1}^n (-2y_i) + n\bar{y}^2 \right] \\ &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \left( \frac{1}{n} \sum_{j=1}^n y_j \right) \sum_{i=1}^n (-2y_i) + n \left( \frac{1}{n} \sum_{i=1}^n y_i \right)^2 \right] \end{aligned}$$

Consider the term  $\left( \frac{1}{n} \sum_{j=1}^n y_j \right) \sum_{i=1}^n (-2y_i)$ . Notice that when  $j = i$ :

$$\left( \frac{1}{n} \sum_{j=1}^n y_j \right) \left( \sum_{i=1}^n (-2y_i) \right) = \frac{1}{n} \sum_{i=1}^n -2y_i^2$$

Meanwhile, if  $i \neq j$ :

$$\left( \frac{1}{n} \sum_{j=1}^n y_j \right) \left( \sum_{i=1}^n (-2y_i) \right) = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (-4y_i y_j)$$

Therefore,

$$\begin{aligned} \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \left( \frac{1}{n} \sum_{j=1}^n y_j \right) \sum_{i=1}^n (-2y_i) + n \left( \frac{1}{n} \sum_{i=1}^n y_i \right)^2 \right] &= \\ \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{i=1}^n -2y_i^2 + \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (-4y_i y_j) + \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 \right] \end{aligned}$$

Notice that  $\frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n 2y_i y_j$ . Therefore,

$$\begin{aligned} \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{i=1}^n -2y_i^2 + \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (-4y_i y_j) + \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 \right] &= \\ \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{i=1}^n -2y_i^2 + \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (-4y_i y_j) + \frac{1}{n} \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n 2y_i y_j \right] \end{aligned}$$

So,

$$\begin{aligned} \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 - \frac{1}{n} \sum_{i=1}^n y_i^2 - \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n 2y_i y_j \right] &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i^2 + \sum_{j=1}^n \sum_{i=1}^n 2y_i y_j \right) \right] \\ &= \frac{1}{n-1} \left[ \sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2 \right] \end{aligned}$$

## Problem 1.12

(see code)

$$s = 1.4667$$

## Problem 1.13

(see code)

(a).  $\mu = 9.7911, s^2 = 17.1231, s = 4.1380$

(b). In the interval [5.65:, 13.93].  
According to the Empirical Rule, I expect to find 31  
I found 44 samples.

In the interval [1.52:, 18.07].  
According to the Empirical Rule, I expect to find 43  
I found 44 samples.

In the interval [-2.62:, 22.21].  
According to the Empirical Rule, I expect to find 45  
I found 44 samples.

The results do not follow the Empirical Rule because the Empirical Rule assumes my data obeys an approximately normal distribution. This is not the case because there is 1 extreme outlier sample in my data which causes my standard deviation to inflate. A more accurate analysis can be done by removing the outlier 35.1 value.

## Problem 1.14

(a).  $\mu = 3.2252, s = 3.1029$

(b). In the interval [0.12:, 6.33].  
According to the Empirical Rule, I expect to find 17  
I found 21 samples.

In the interval [-2.98:, 9.43].  
According to the Empirical Rule, I expect to find 24  
I found 23 samples.

In the interval [-6.08:, 12.53].  
According to the Empirical Rule, I expect to find 25  
I found 25 samples.

These results are what I expect.

## Problem 1.15

(a).  $\mu = 4.387, s = 1.8714$

(b). In the interval [2.52:, 6.26].  
According to the Empirical Rule, I expect to find 27  
I found 35 samples.

In the interval  $[0.64, 8.13]$ .

According to the Empirical Rule, I expect to find 38

I found 39 samples.

In the interval  $[-1.23, 10.00]$ .

According to the Empirical Rule, I expect to find 40

I found 39 samples.

For lower values of  $k$ , I found more samples than expected. This is likely due to the outlier value 11.88 in the data. Removing that sample will cause the data to better fit the Empirical Rule.

## Problem 1.16

$$\mu = 4.1949, s = 1.4232$$

In the interval  $[2.77, 5.62]$ .

According to the Empirical Rule, I expect to find 27

I found 25 samples.

In the interval  $[1.35, 7.04]$ .

According to the Empirical Rule, I expect to find 37

I found 36 samples.

In the interval  $[-0.07, 8.46]$ .

According to the Empirical Rule, I expect to find 39

I found 39 samples.

These results almost perfectly follow the Empirical Rule, as expected. By removing the outlier, the mean decreased by 0.19 and standard deviation went from 1.85 to 1.42.

## Problem 1.17

Analyzing Problem 1.2.

The standard deviation is 4.09.

The approximation is 7.35.

There is a large difference between the true standard deviation and the approximation because of the 35.1 outlier value. This is consistent with our analysis in Problem 1.13 where we found the Empirical Rule did not fit the data for similar reasons.

Analyzing Problem 1.3.

The standard deviation is 3.10.

The approximation is 3.04.

The true standard deviation and the approximation are close, which is consistent with our analysis in Problem 1.14.

Analyzing Problem 1.4.

The standard deviation is 1.85.

The approximation is 2.32.

There is a large difference between the standard deviation and approximation because of the outlier sample 11.88. This is consistent with our analysis in Problems 1.15 and 1.16 concerning the Empirical Rule.

### **Problem 1.18**

Using the approximation given in Problem 1.17, the standard deviation of the verbal test should be about 150.0.

### **Problem 1.19**

Chlorophorm amounts do not follow a normal distribution because if it did, the data should be approximated by the Empirical Rule. Therefore, about 2.5% of public water samples should have chlorophorm values between (-125.00, -72.00). However, it is physically impossible to record a chlorophorm sample below 0 micrograms per liter. Due to this physical constraint, the chlorophorm samples cannot follow a normal distribution with mean 34 and standard deviation 53.

### **Problem 1.20**

Assuming maintenance cost is normally distributed with mean 420.00 and standard deviation is 30.00, the Empirical Rule implies that the probability of running overbudget with a budget of 450.00 is 16%.

### **Problem 1.21**

Assuming cow weight fed this diet follows a normal distribution with mean 22 and standard deviation 2, the Empirical Rule implies the probability of weight gain exceeding 20 pounds is 82%. Thus the manufacturers claim is supported by these results.