Exercise 1.2

According to the journal Chemical Engineering, an important property of a fiber is its water absorbency. A random sample of 20 pieces of cotton fiber was taken and the absorbency on each piece was measured. The following are the absorbency values:

18.71 21.41 20.72 21.81 19.29 22.43 20.17
23.71 19.44 20.50 18.92 20.33 23.00 22.85
19.25 21.77 22.11 19.77 18.04 21.12

(a) Calculate the sample mean and median for the above sample values.

(b) Compute the 10% trimmed mean.

Solution

Part (a) - Calculation of Sample Mean

We'll first calculate the sample mean. Let $x_1, x_2, \ldots, x_n$ denote n observations. The formula for the sample mean, $\bar{x}$, is given by the formula

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{1}$$

For this data set we have $n = 20$ observations as shown above: $x_1 = 18.71$, $x_2 = 21.41$, ..., $x_{20} = 21.12$. We compute $\bar{x}$ using equation (1).

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{n} x_i$$

Using the software **R** we compute the mean as follows:

```
> x <- c(18.71,21.41,20.72,21.81,19.29,22.43,20.17,23.71,19.44,20.50,18.92,
20.33,23.00,22.85,19.25,21.77,22.11,19.77,18.04,21.12)
> mean(x)
[1] 20.7675
```

The sample mean absorbancy of the 20 pieces of cotton fiber is 20.77 (rounded).

Part (a) - Calculation of Sample Median

Next we'll show the calculation of the sample median. The median is defined as the 50^{th} percentile of the n observations. To find the p^{th} percentile we first need to order the data from smallest to largest, then find the observation at position $(p/100)(n+1)$. Since $p = 50$ and $n = 20$ we are looking for position

$$(p/100)(n+1) = (50/100) \times (20+1) = 10.5$$

Since this number is not an integer, we must take the average of the observation at position 10 (the floor of 10.5) and the observation a position 11 (the ceiling of 10.5). To find those two observations, we order the data from smallest to largest.

18.04 18.71 18.92 19.25 19.29 19.44 19.77 20.17 20.33 20.50
20.72 21.12 21.41 21.77 21.81 22.11 22.43 22.85 23.00 23.71

The observation at position 10 is 20.50 and the observation at position 11 is 20.72. Thus, the sample median is the average of these two values

$$m = \frac{20.50 + 20.72}{2} = 20.61.$$

We verify this calculation using **R**.

```
> x <- c(18.71,21.41,20.72,21.81,19.29,22.43,20.17,23.71,19.44,20.50,18.92,
20.33,23.00,22.85,19.25,21.77,22.11,19.77,18.04,21.12)
> median(x)
[1] 20.61
```

The sample median absorbancy of the 20 pieces of cotton fiber is 20.61.

Part (b) - Calculation of the 10% trimmed mean

The α% trimmed mean is the mean of the data with the smallest $\alpha\% \times n$ observations and the largest $\alpha\% \times n$ observations truncated from the data. For this example we have $\alpha = 10$ and $n = 20$. Then we are to remove the smallest $10\% \times 20 = 2$ observations and the largest $10\% \times 20 = 2$ observations. In part (a) when we calculated the sample median we ordered the data as shown above. The two smallest observations are 18.04 and 18.71 and the two largest observations are 23.00 and 23.71. Removing those 4 observations gives us the following data:

18.92 19.25 19.29 19.44 19.77 20.17 20.33 20.50
20.72 21.12 21.41 21.77 21.81 22.11 22.43 22.85

The sample size of this new data set is $n = 16$ and take $y_1 = 18.92$, $y_2 = 19.25$, ..., $y_{16} = 22.85$. The sample mean of this data set, calculated using equation (1), is the 10% trimmed mean.

$$\bar{y}_{tr(10)} = \frac{1}{16}\sum_{i=1}^{16} y_i$$

Using the software **R** we compute the mean as follows:

```
> y <- c(18.92,19.25,19.29,19.44,19.77,20.17,20.33,20.50,20.72,
21.12,21.41,21.77,21.81,22.11,22.43,22.85)
> mean(y)
[1] 20.74312
```

We can alternatively calculate the trimmed mean in **R** directly from the original data of 20 observations, $x_1, \ldots, x_{20}$, using the `mean()` function with additional input `trim=0.10`.

```
> x <- c(18.71,21.41,20.72,21.81,19.29,22.43,20.17,23.71,19.44,20.50,18.92,
20.33,23.00,22.85,19.25,21.77,22.11,19.77,18.04,21.12)
> mean(x, trim = 0.10)
[1] 20.74312
```

The 10% trimmed mean of the absorbancy of the 20 pieces of cotton fiber is 20.74 (rounded).

Summary

In this exercise we calculated the sample mean, median, and 10% trimmed mean for the mean absorbancy of the 20 pieces of cotton fiber. We found the sample mean to be 20.77, median to be 20.61, and 10% trimmed mean to be 20.74. All three of these values represent measures of location, specifically where the center of the data lies. All three are very close to one another in value.