Tutorial 3

Xiaonan Da

2020-01-27

As part of a study, the experimenters obtained test scores from a small group of students who read a certain passage as follow: 66, 75, 72, 71, 55, 56, 72, 93, 73, 72, 72, 73, 91, 66, 71, 56, 59. They also collected data on students who did not read the passage: 54, 52, 51, 50, 36, 55, 44, 46, 57, 44, 43, 52, 38, 46, 55, 34, 44, 39, 43, 36, 55, 57, 36, 46, 49, 46, 49, 47. Answer the following questions.

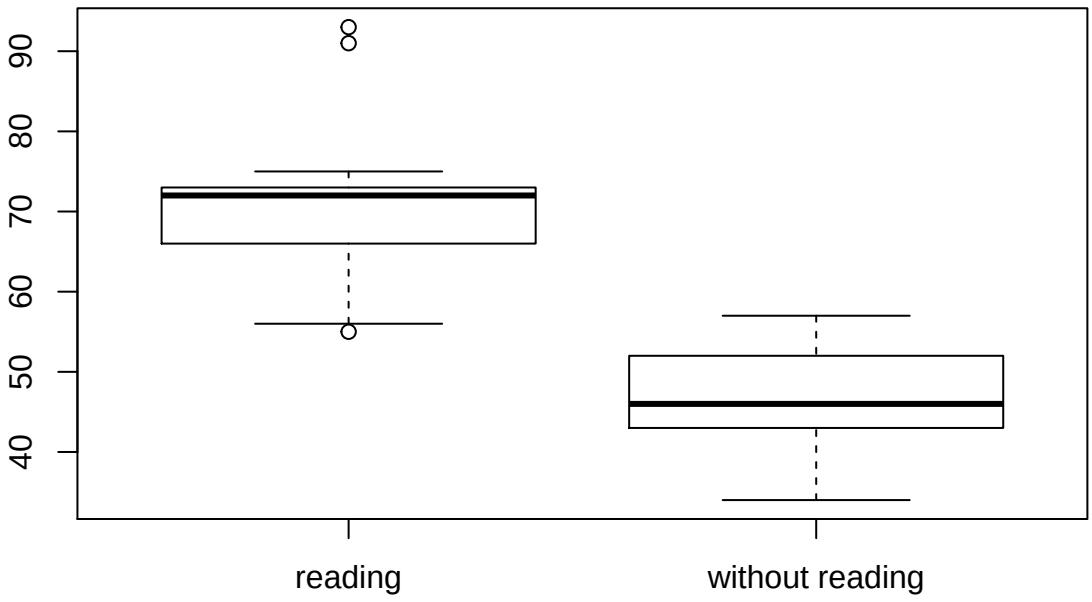
(a)

Produce approriate graphs to demonstrate the above two sets of data.

Solution:

Boxplot of test scores of students who read the passage

```
x <- c(66, 75, 72, 71, 55, 56, 72, 93, 73, 72, 72, 73, 91, 66, 71, 56, 59)
y <- c(54, 52, 51, 50, 36, 55, 44, 46, 57, 44, 43, 52, 38, 46,
       55, 34, 44, 39, 43, 36, 55, 57, 36, 46, 49, 46, 49, 47)
boxplot(x,y,names=c("reading","without reading"))
```



(b)

Are there any outliers in the data sets?

Solution: Yes in the first set, and no in the second.

(c)

Calculate the range, mode, median, and the mean for both data sets.

Solution:

For the first group, range is 38, mode is 72, median is 72, and mean is 70.176

```
# For the first set of data
range(x)
```

```
## [1] 55 93
```

```
range(x)[2]-range(x)[1]
```

```
## [1] 38
```

```
# the range is 38
sort(x)
```

```
##  [1] 55 56 56 59 66 66 71 71 72 72 72 72 73 73 75 91 93
```

```
# 72 is the mode since it appeared 4 times
median(x)
```

```
## [1] 72
```

```
# 72 is the median
mean(x)
```

```
## [1] 70.17647
```

```
# 70.17647 is the mean
sum(x)/length(x)
```

```
## [1] 70.17647
```

For the second group, range is 23, mode is 46, median is 46, and mean is 46.571

```
# For the second set
range(y)
```

```
## [1] 34 57
```

```
range(y)[2]-range(y)[1]
```

```
## [1] 23
```

```
# the range is 23
sort(y)
```

```
##  [1] 34 36 36 36 38 39 43 43 44 44 44 46 46 46 46 47 49 49 50 51 52 52 54
## [24] 55 55 55 57 57
```

```
# 46 is the mode since it appeared 4 times
median(y)
```

```
## [1] 46
```

```
# 46 is the median
mean(y)
```

```
## [1] 46.57143
```

```
# 46.57143 is the mean
sum(y)/length(y)
```

```
## [1] 46.57143
```

(d)

Are the data skewed? Explain.

Solution: The first set of data is skewed to the left, since the median is greater than the mean. The second set of data is skewed to the right, since the median is less than the mean.

(e)

Calculate the sample variance and standard deviation of both data sets.

Solution:

The sample mean is defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

The sample variance is defined as

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

```
var(x)
```

```
## [1] 112.5294
```

```
sum((x-mean(x))^2)/(length(x)-1)
```

```
## [1] 112.5294
```

```
# the variance is 112.5294
sqrt(var(x))
```

```
## [1] 10.60799
```

```
sd(x)
```

```
## [1] 10.60799
```

```
sqrt(sum((x-mean(x))^2)/(length(x)-1))
```

```
## [1] 10.60799
```

```
# the standard deviation is 10.60799
```

```
# Second set
var(y)
```

```
## [1] 46.62434
```

```
sum((y-mean(y))^2)/(length(y)-1)
```

```
## [1] 46.62434
```

```
# the variance is 46.62434
sd(y)
```

```
## [1] 6.828202
```

```
sqrt(sum((y-mean(y))^2)/(length(y)-1))
```

```
## [1] 6.828202
```

```
# the standard deviation is 6.828202
```