R Assignment 1

Albert Gator (agator)

January 13, 2025

Exercise 14.1 (modified)

Recall the built-in **InsectSprays** data frame, containing counts of insects on various agricultural units treated with one of six sprays.

a) Find the five number summary for the counts of insects for the total data set and also according to each spray type.

b) Obtain Use base R graphics to produce a boxplot of the insects in aggregate, and then a second plot showing the number of insects by spray type, labeling the plots appropriately. Provide a summary of the two plots.

Solution

Part (a) - Five number summary of insect counts

The **InsectSprays** data frame is a built-in data frame that can be accessed directly in **R**. The data set contains data on the number of insects found on various agricultural units, as well as the type of insect spray that was used on each unit.

We will use **R** to calculate the five number summary of the counts of insects, which includes the minimum, 25th percentile, median, 75th percentile and maximum. This is done using the **quantile()** function, specifying the quantiles as 0, 0.25, 0.5, 0.75, and 1.

```
quantile(InsectSprays$count, probs = c(0, 0.25, 0.5, 0.75, 1))
```

```
##    0%   25%   50%   75%  100%
##  0.00  3.00  7.00 14.25 26.00
```

From this summary we see that the minimum number of insects is 0, the maximum number of insects is 26, the median is 7, 25th percentile is 3, and 75th percentile is 14.25.

Part (a) - Five number summary of insect counts by spray type

Next we determine the five number summary of insect counts by the type of spray, which is the other information given in the **InsectSprays** data frame. There are six spray types: A, B, C, D, E, and F. To determine this, we subset the vector of number of insects by each spray type.

Five number summary for Spray Type A

```r
quantile(InsectSprays$count[InsectSprays$spray == "A"])
```

```
##    0%   25%   50%   75%  100%
##  7.00 11.50 14.00 17.75 23.00
```

Five number summary for Spray Type B

```r
quantile(InsectSprays$count[InsectSprays$spray == "B"])
```

```
##   0%  25%  50%  75% 100%
##  7.0 12.5 16.5 17.5 21.0
```

Five number summary for Spray Type C

```r
quantile(InsectSprays$count[InsectSprays$spray == "C"])
```

```
##   0%  25%  50%  75% 100%
##  0.0  1.0  1.5  3.0  7.0
```

Five number summary for Spray Type D

```r
quantile(InsectSprays$count[InsectSprays$spray == "D"])
```

```
##    0%   25%   50%   75%  100%
##  2.00  3.75  5.00  5.00 12.00
```

Five number summary for Spray Type E

```r
quantile(InsectSprays$count[InsectSprays$spray == "E"])
```

```
##   0%  25%  50%  75% 100%
## 1.00 2.75 3.00 5.00 6.00
```

Five number summary for Spray Type F

```r
quantile(InsectSprays$count[InsectSprays$spray == "F"])
```

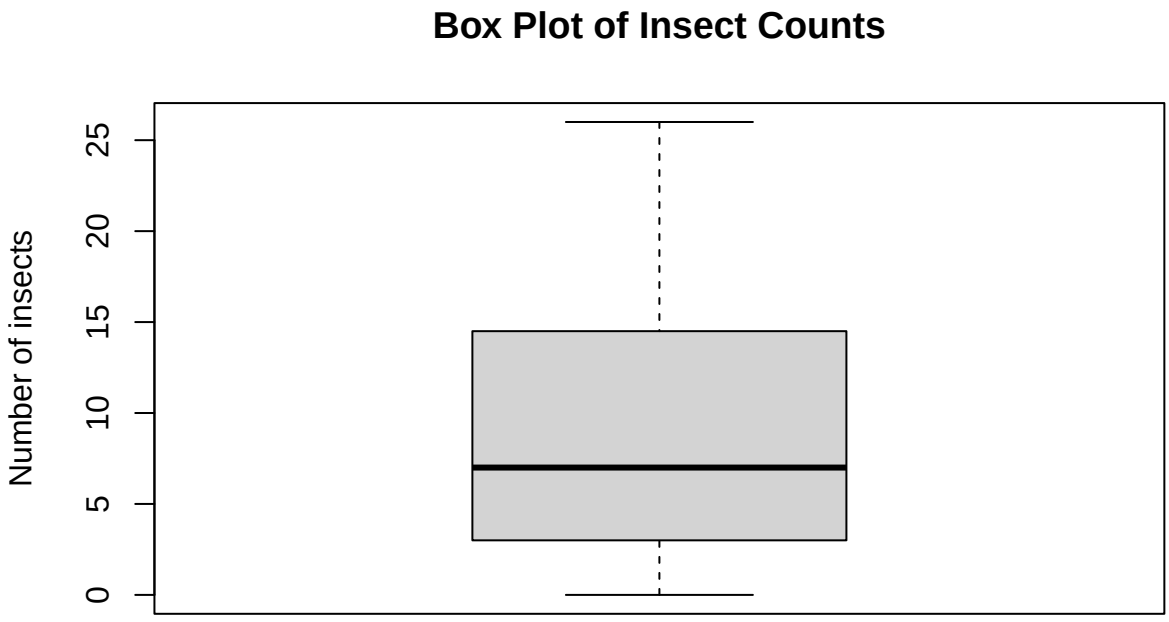
```
##   0%  25%  50%  75% 100%
##  9.0 12.5 15.0 22.5 26.0
```

As seen in the five number summaries of the number of insects, there is a lot of differences between each spray type in terms of the central tendency and spread. For example, with sprays A, B, and F there seem to be more insects compared to sprays C, D, and E. Also the spread for the number of insects seems to be much smaller for sprays C and E compared to the others.

Part (b) - Box plot of insect counts for all spray types

We'll use the base **R** boxplot as a visualization of the five number summary shown in part (a). The function to produce a boxplot is **boxplot()**.

```r
boxplot(InsectSprays$count, main = "Box Plot of Insect Counts",
    ylab = "Number of insects")
```

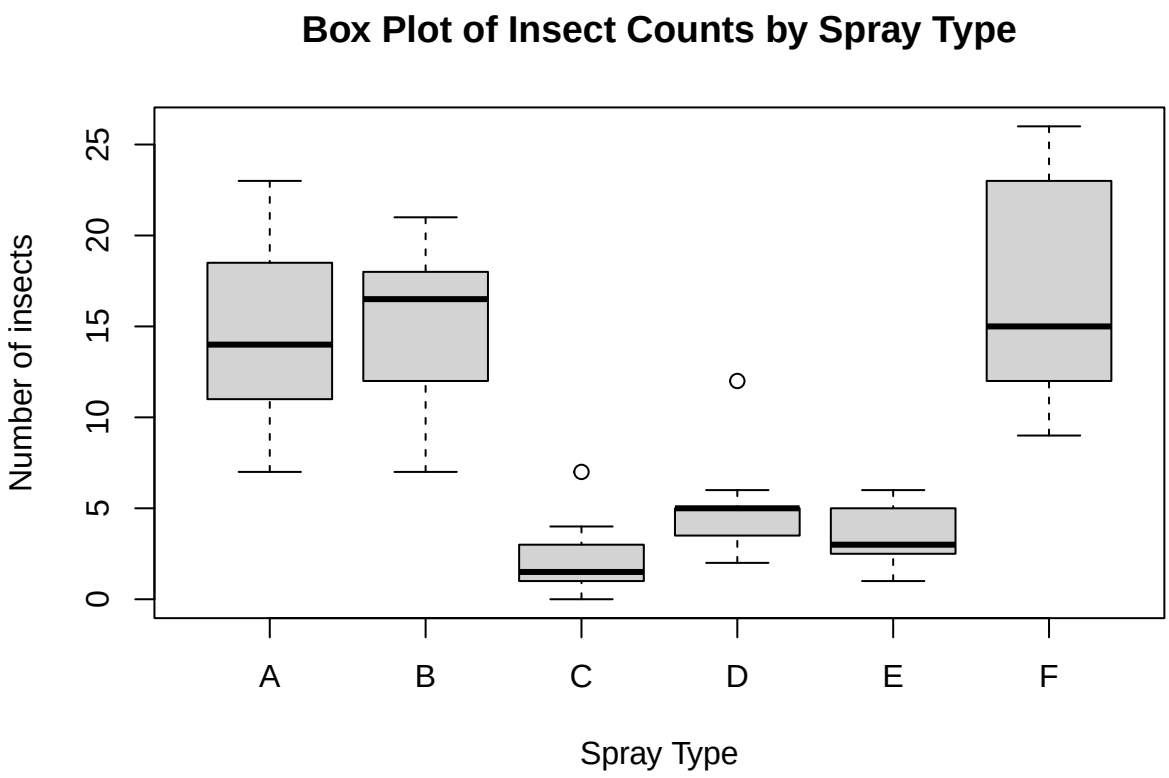


As seen in the box plot, the median (central black line of the box) is at 7, the 25th percentile (lower edge of box) is around 3 and 75th percentile (upper edge of box), is around 15. The whiskers extend out to the minimum at 0 and maximum at 27. The distribution does seem to be slightly skewed right as the top whisker is longer than the bottom whisker.

Part (b) - Box plot of insect counts for all spray types

Next we'll produce box plots for insect counts by spray type using the **boxplot()** function.

```r
boxplot(InsectSprays$count ~ InsectSprays$spray, main = "Box Plot of Insect Counts by Spray Type",
    xlab = "Spray Type", ylab = "Number of insects")
```



The distribution of number of insects for sprays A, B, and F are all very similar in terms of central tendency and spread, while the distributions for sprays C, D, and E are all very similar in terms of central tendency and spread. Sprays A, B, and F all consistently have more insects than sprays C, D, and E. The spreads are also higher for sprays A, B, and F.

Summary

In this assignment we investigated the number of insects for different spray types. We calculated a five number summary and produced box plots for each spray type. We can conclude that there are fewer insects overall for sprays C, D, and E compared to sprays A, B, and F.