Verzani Problem Set

Next are considered the problems from Verzani's book on page 92.

Problem 15.1

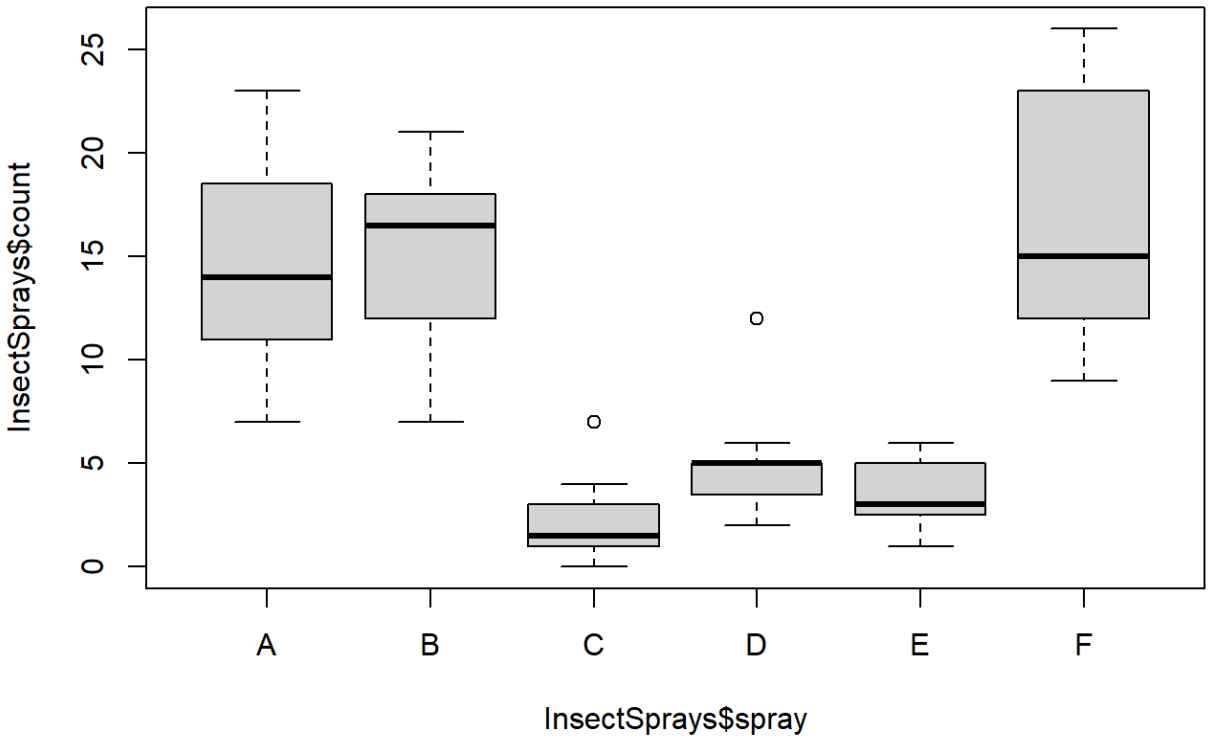
The dataset `InsectSpray` has data on the count of insects in areas treated with one of 6 different types of sprays. The dataset is already in the proper format for the one-way analysis of variance - a vector with the data `count`, and one with a factor describing the level `spray`. First make a side-by-side boxplot to see if the means are equal. Then perform a one-way ANOVA to check if they are. Do they agree?

Y is the count of insects.

X is the type of the spray. It has 6 levels: A, B, C, D and F.

```
> head(InsectSprays)
  count spray
1    10     A
2     7     A
3    20     A
4    14     A
5    14     A
6    12     A
```

```
> boxplot(InsectSprays$count ~ InsectSprays$spray)
```



Are the differences between these means statistically significant?

$$H_0 : \mathbb{E}(Y \mid X = A) = \mathbb{E}(Y \mid X = C) = \mathbb{E}(Y \mid X = D) = \\ = \mathbb{E}(Y \mid X = E) = \mathbb{E}(Y \mid X = F)$$

H_A : At least two of $\mathbb{E}(Y \mid X = A)$, $\mathbb{E}(Y \mid X = B)$, $\mathbb{E}(Y \mid X = C)$, $\mathbb{E}(Y \mid X = D)$, $\mathbb{E}(Y \mid X = E)$ and $\mathbb{E}(Y \mid X = F)$ are different.

or which is the same as

$$H_0 : \sigma_B^2 = \sigma_W^2$$
$$H_A : \sigma_B^2 \neq \sigma_W^2$$

```
> myanova <- aov(count ~ spray, data = InsectSprays)
> summary(myanova)
            Df Sum Sq Mean Sq F value Pr(>F)
spray        5   2669   533.8    34.7 <2e-16 ***
Residuals   66   1015    15.4
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
```

The first three numbers in the row `spray` are for the degrees of freedom, SS_B and MS_B.

The first three numbers in the row `Residuals` are for the degrees of freedom, SS_W and MS_W.

The number `F value` is for

$$F_{emp} = \frac{MS_B}{MS_W} = 34.702$$

The corresponding

$$p-value = \mathbb{P}(F_{stat} > F_{emp} \,|\, H_0) = \mathbb{P}(\eta > F_{emp}) < 2.2e-16 < 0.05 = \alpha,$$

so we reject H_0. According to the data the spray is statistically significant for the number of the insects in the area.

Problem 15.2

The `newcar` data set from here contains the interest rates in 6 different towns. Customer would like to know if there is statistically significant difference between the interest rates in these towns. Do a one-way analysis of variance to test the hypothesis that the interest rates are the same for all 6 towns. What do you conclude?

```
> library(readr)
> newcar <- read_table2("Data/newcar.dat", col_names =
c("Rate", "City"), skip = 24, comment = "--")

-- Column specification
--------------------------------------------------------
cols(
  Rate = col_double(),
  City = col_double()
)
> head(newcar)
# A tibble: 6 x 2
```

```
   Rate  City
  <dbl> <dbl>
1  13.8     1
2  13.8     1
3  13.5     1
4  13.5     1
5  13       1
6  13       1
> myanova <- aov(Rate ~ City, data = newcar)
> summary(myanova)
             Df Sum Sq Mean Sq F value Pr(>F)
City          1   0.75  0.7516   1.223  0.274
Residuals    52  31.95  0.6145
```

The first three numbers in the row `City` are for the degrees of freedom, SS_B and MS_B.

The first three numbers in the row `Residuals` are for the degrees of freedom, SS_W and MS_W.

The number `F value` is for

$$F_{emp} = \frac{MS_B}{MS_W} = 1.223$$

The corresponding

$$p-value = \mathbb{P}(F_{stat} > F_{emp} \,|\, H_0) = \mathbb{P}(\eta > F_{emp}) = 0.274 > 0.05 = \alpha,$$

so we have no evidence to reject H_0. According to the data the cities are not statistically significant for the interest rates.