High Diminsional Statistics-Sheet 2-Exercise 4

Hamed Vaheb

3/31/2022

The goal of an experimental research is to understand whether or not some hormonal treatments have effects on the hormonal concentrations of 12 adult female dogs. No treatment has been given to 4 dogs. Their hormonal concentrations are 117, 124, 40 and 88. 4 dogs have been treated with oestrogens. Their hormonal concentrations after the treatment are 440, 264, 221, 136. 4 dogs have been treated with progesterone. Their hormonal concentrations after the treatment are 605, 626, 385, 475.

Prepare

```
index <- c(1:4)
cnctr_before <- c(117, 124, 40, 88)
cnctr_after_oest <- c(440, 264, 221, 13)
cnctr_after_progest <- c(605, 626, 385, 475)
```

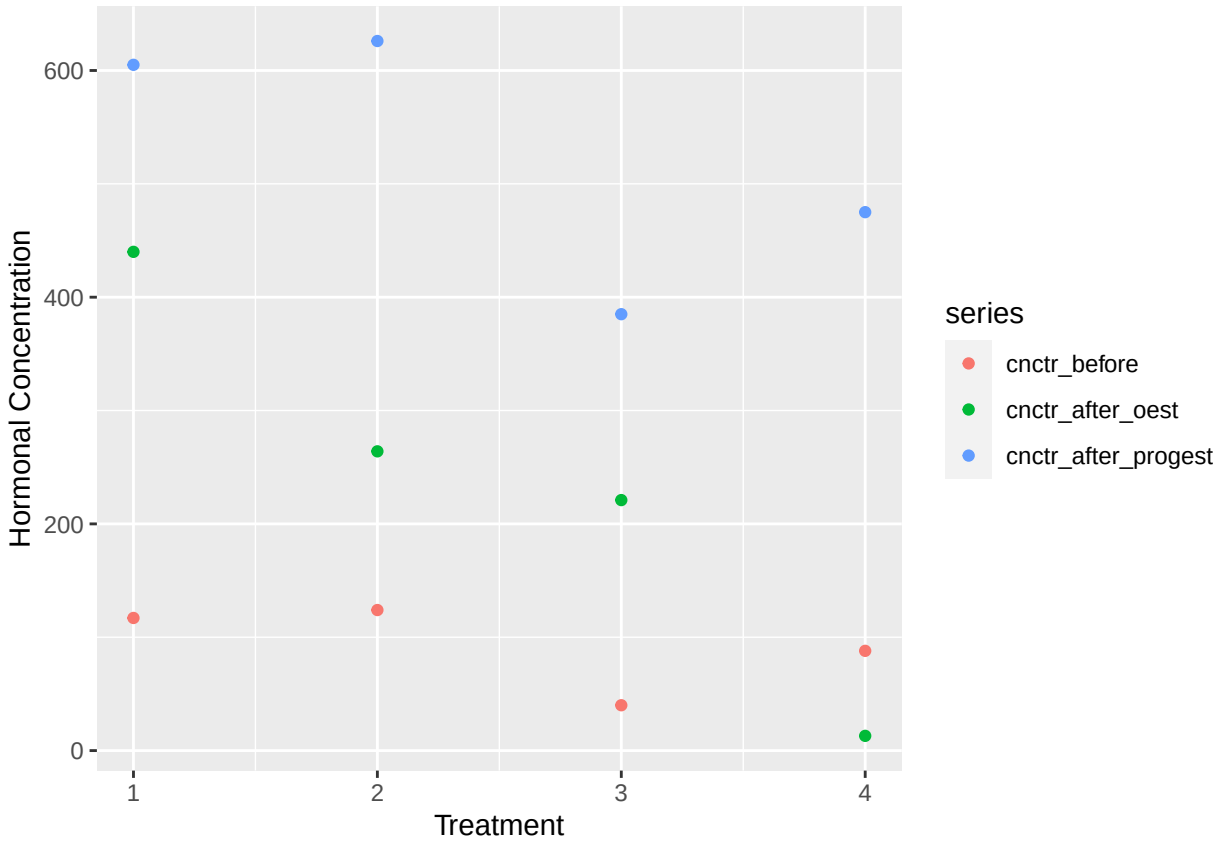
Question 1: Plotting Concentration by Treatment

Plot a multiple histogram showing hormonal concentrations by treatment.

```
# Plot with main and axis titles
# Change point shape (pch = 19) and remove frame.
# Add regression line
df <- data.frame(index, cnctr_before, cnctr_after_oest, cnctr_after_progest)

df_plot <- melt(df, id.vars = 'index', variable.name = 'series')


ggplot(data = df_plot, aes(index, value)) +
  geom_point(aes(colour = series)) +
  labs(x = "Treatment", y = "Hormonal Concentration")
```



Scatterplot

```
set.seed(1)
df
```

Histogram

```
##   index cnctr_before cnctr_after_oest cnctr_after_progest
## 1     1          117              440                 605
## 2     2          124              264                 626
## 3     3           40              221                 385
## 4     4           88               13                 475
```

```
df_plot <- df_plot[order(df_plot$index),]

df_plot
```

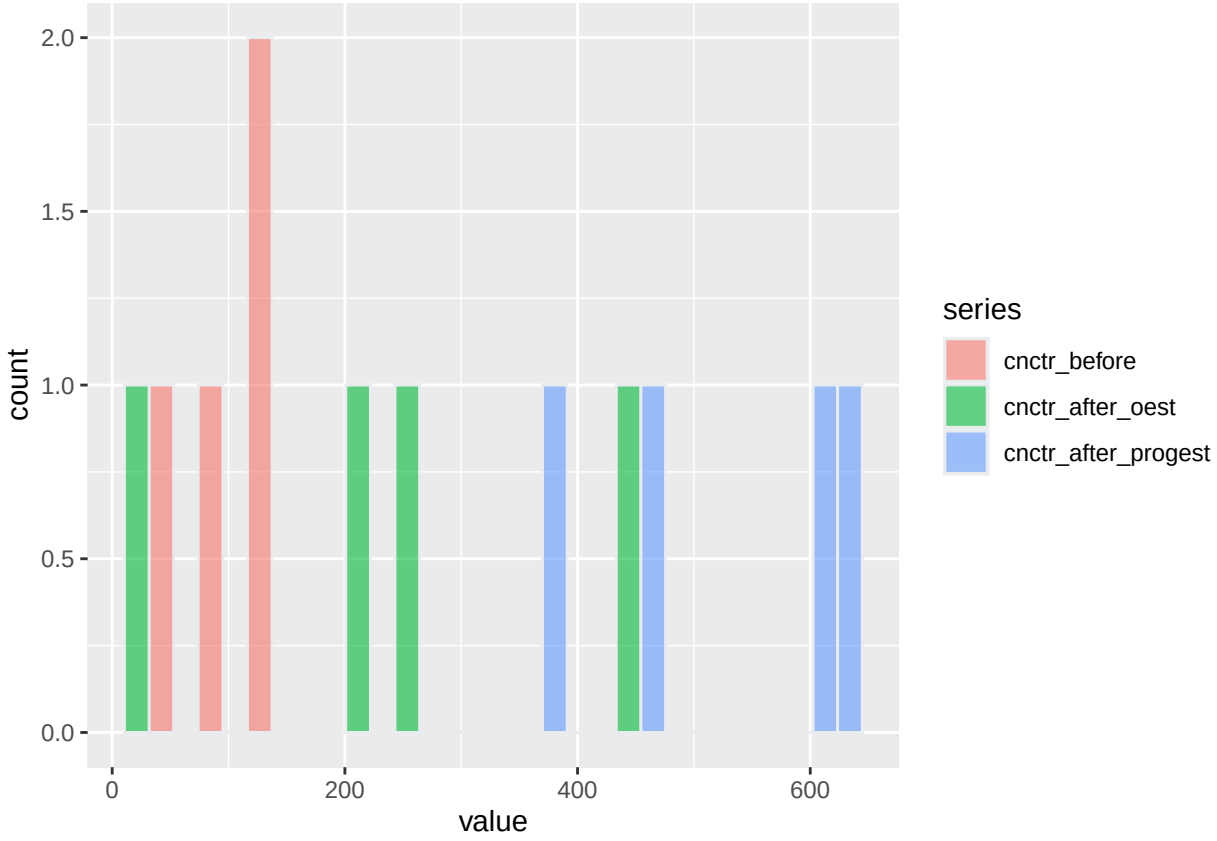
```
##    index              series value
## 1      1        cnctr_before   117
## 5      1    cnctr_after_oest   440
## 9      1 cnctr_after_progest   605
## 2      2        cnctr_before   124
## 6      2    cnctr_after_oest   264
## 10     2 cnctr_after_progest   626
## 3      3        cnctr_before    40
## 7      3    cnctr_after_oest   221
## 11     3 cnctr_after_progest   385
## 4      4        cnctr_before    88
## 8      4    cnctr_after_oest    13
```

```
## 12     4 cnctr_after_progest   475
```

```
df_plot$series <- as.factor(df_plot$series)


ggplot(df_plot, aes(x=value, fill=series)) +
  geom_histogram( color='#e9ecef', alpha=0.6, position='identity')
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Question 2: Table of man and stds

Calculate a table that shows the mean and standard deviation of hormonal concentrations by treatment.

```
 sd <- df %>% select(c('cnctr_before', 'cnctr_after_oest', 'cnctr_after_progest')) %>% apply(2,sd)

mean <- df %>% select(c('cnctr_before', 'cnctr_after_oest', 'cnctr_after_progest')) %>% apply(2,mean)

rbind(mean, sd)
```

```
##      cnctr_before cnctr_after_oest cnctr_after_progest
## mean     92.25000          234.500            522.7500
## sd       38.16084          175.447            113.5499
```

Question 3: ANOVA

Use ANOVA to test for a difference between treatments in the hormonal concentrations of the dogs, with a significance level $\alpha = 0.05$. What is your conclusion about the effect of the treatments?

One-way analysis of variance (ANOVA) is employed when there exists a categorical independent variable

(with two or more categories) and a normally distributed interval dependent variable and we aim at testing the difference between means of the dependent variable on broken down by the levels of the independent variable.

We assume that the categorical independent variable is hormonal treatment, which consists of two groups, namely oestrogens and progesterone. Mean of each group is denoted by μ_O and μ_P respectively. We also assume the dependent variable is hormonal concentrations. We formulate our hypothesis test as the following:

$H_0 : \mu_O - \mu_P = 0$ against $H_1 : \mu_O - \mu_P \neq 0$

```
aov.model <- aov(df_plot$value ~ df_plot$series)
summary(aov.model)
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## df_plot$series  2 384871  192436   12.79 0.00234 **
## Residuals       9 135395   15044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov.model$coefficients
```

```
##                         (Intercept)    df_plot$seriescnctr_after_oest
##                               92.25                             142.25
## df_plot$seriescnctr_after_progest
##                             430.50
```

The associated p-value helps to interpret the F-statistic. Given significance level $\alpha = 0.05$, since p_value = 0.00234 and hence p_value $<\alpha$, we reject null hypothesis and we increase our certainty in H_1, which states that there is significant difference between mean of concentration level of the groups that has been treated with oestrogen and the the one that has been treated with progesterone.