

Two Way Classification Problem (ANOVA)

Anubhav Sarkar

2022-11-23

Abstract: Here the data given are plasma concentrations (in mg/100ml) of birds of both sexes, The birds of each sex being treated with a hormone and not treated with the hormone are equally distributed. Test for the significant difference between hormone treatment and between gender has been done with the help of 2-way ANOVA.

Introduction:

The analysis of variance is the systematic algebraic procedure of decomposing (i.e. partitioning) overall variation (i.e. total variation) in the responses observed in an experiment into different components of variations such as treatment variation and error variation. Each component is attributed identifiable cause or source of variation. Assumptions of ANOVA: For the validity of the F-test in ANOVA the following assumptions are made. 1. The effects of different factors (treatments and environmental effects) are additive in nature. 2. Observations of character under study follow normal distribution 3. Assumption of homogeneity of variance 4. Experimental errors are distributed independently and normally with mean zero and constant variation i.e. $\sim N(0, \sigma^2)$

Methodology:

The data is imported to R and the values are displayed.

```
Data <- read.csv("C:\\Users\\Desktop\\data10.csv")
```

```
Data$Treatment <- factor(Data$Treatment)
```

```
Data$Gender <- factor(Data$Gender)
```

```
attach(Data)
```

```
head(Data)
```

```
## Treatment Gender Plasma
```

```
## 1 Non Hormone Male 14.5
```

```
## 2 Non Hormone Male 11.0
```

```
## 3 Non Hormone Male 10.8
```

```
## 4 Non Hormone Male 14.3
```

```
## 5 Non Hormone Male 10.0
```

```
## 6 Non Hormone Female 16.5
```

```
str(Data)
```

```
## 'data.frame': 20 obs. of 3 variables:
```

```
## $ Treatment: Factor w/ 2 levels "Hormone","Non Hormone": 2 2  
2 2 2 2 2 2 2 ...
```

```
## $ Gender : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 1 1 1  
1 1 ...
```

```
## $ Plasma : num 14.5 11 10.8 14.3 10 16.5 18.4 12.7 14 12.8 ...
```

We will be conducting two tests. For the first one:

Null Hypothesis (H0): The average plasma concentrations are same among all the given treatments, (i.e. $\mu_1 = \mu_2$) where 1 is Non-Hormonal Treatment, 2 is Hormonal Treatment.

Alternative Hypothesis (H1): At least one of the group is different from the other. ($\mu_1 \neq \mu_2$) Level of significance (α): Let $\alpha = 0.05$

And for the second one: Null Hypothesis (H0): The average plasma concentrations are same among all the genders, (i.e. $\mu_1 = \mu_2$) where 1' is Male, 2' Female.

Alternative Hypothesis (H1): At least one of the group is different from the other. ($\mu_1 \neq \mu_2$) Level of significance (α): Let $\alpha = 0.05$

In ANOVA the model is like linear regression model, where all the component are additive in nature as given below. $y_{ij} = \mu + \tau_i + \beta_j + e_{ij}$ AIM: To check if the given dataset satisfies the assumptions required for ANOVA. (1) Observations of character under study follow a normal distribution. Testing for Normality: Shapiro-Wilks Test The Shapiro-Wilk test tests the null hypothesis that the samples come from a normal distribution against the alternative hypothesis that the samples do not come from a normal distribution.

```
shapiro.test(Data$Plasma)

##
## Shapiro-Wilk normality test
##
## data: Data$Plasma
## W = 0.91199, p-value = 0.06956
```

The conclusion above, is supported by the Shapiro-Wilk test ($W = 0.94235$, $p = 0.06956$) which indicates observations may follow normal distribution.

- (2) Assumption of homogeneity of variance Bartlett's test Bartlett's test tests the null hypothesis that the group variances are equal against the alternative hypothesis that the group variances are not equal. Bartlett's test should be used when the data is normal and Levene's test should be used when the data is non-normal

```
bartlett.test(Plasma ~ Treatment, data=Data)

##
## Bartlett test of homogeneity of variances
##
## data: Plasma by Treatment
```

```
## Bartlett's K-squared = 6.2299, df = 1, p-value = 0.01256

bartlett.test(Plasma ~ Gender, data=Data)
#Bartlett test of homogeneity of variances
##
## data: Plasma by Gender
## Bartlett's K-squared = 0.46751, df = 1, p-value = 0.4941
```

Since p value is 0.4941 which is greater than 0.05, we can conclude that the variance of the plasma across gender is homogeneously distributed.

- (3) Experimental errors are distributed independently and normally with mean zero and constant variation: The Shapiro-Wilk test tests the null hypothesis that experimental errors are distributed normally against the alternative hypothesis that the experimental errors do not distributed normally.

```
circum.anova.residual <- residuals(aov(Plasma ~ Gender+Treatment, data=Data))
shapiro.test(circum.anova.residual)

##
## Shapiro-Wilk normality test
##
## data: circum.anova.residual
## W = 0.97876, p-value = 0.9172
```

Since p values is greater than 0.05, we can conclude that experimental errors are normally distributed.

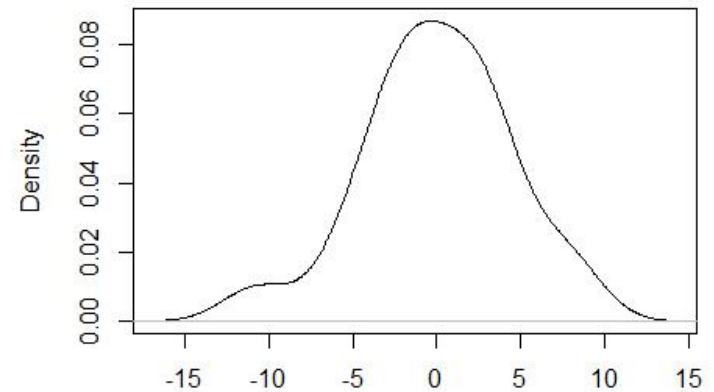
```
t.test(circum.anova.residual, mu=0, conf=0.95)

##
```

```
## One Sample t-test
##
## data: circum.anova.residual
## t = -3.9312e-17, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.068849 2.068849
## sample estimates:
## mean of x
## -3.885781e-17
```

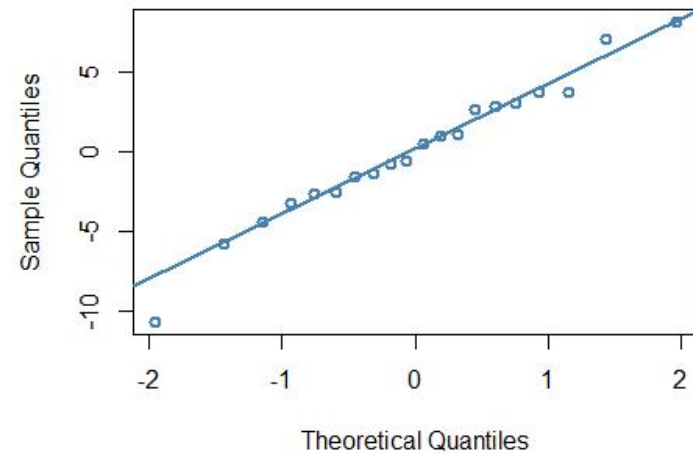
Since t-test's p-value is 1, we accept the fact that mean of the errors may be 0.
Checking constant variance (Note: if data follows a normal distribution, then data has constant variance)

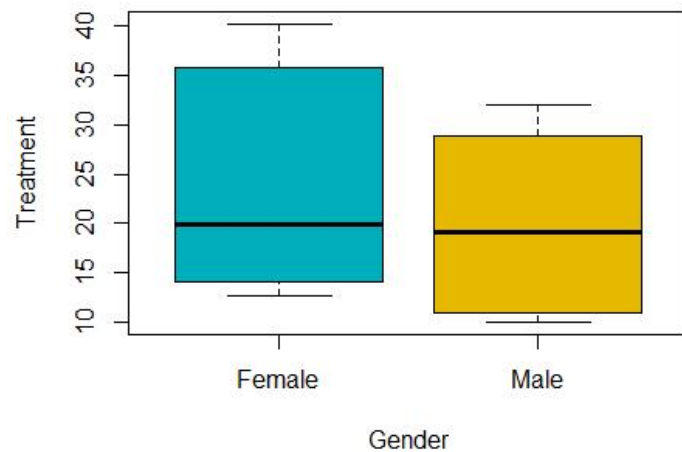
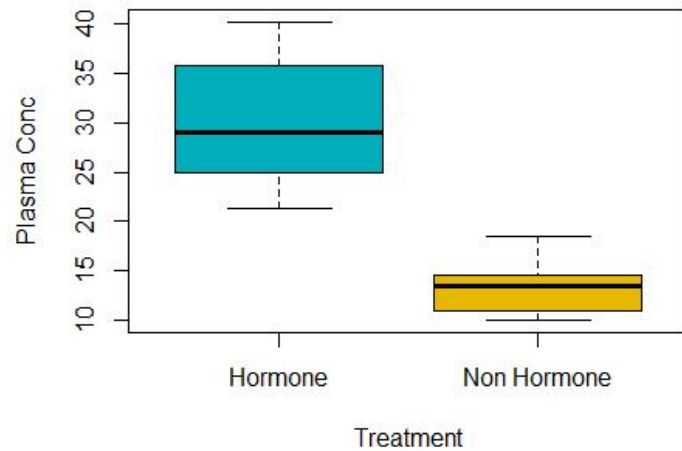
density.default(x = circum.anova.residual)



N = 20 Bandwidth = 2.034

Normal Q-Q Plot





```
model <- aov(Plasma ~ Gender+Treatment, data=Data)
summary(model)
```

```
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Gender      1   70.3    70.3    3.22  0.0906 .
## Treatment   1 1386.1  1386.1   63.47 3.86e-07 ***
## Residuals  17   371.3    21.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```