One Way Anova: Preparing Data

Sanskar Lalwani

1 Introduction

We Have so far learnt the theory of one way anova. In this section we will learn to apply that on a toy dataset of Agricultural yield using R.

2 Lets get started

We begin by loading up our data through a csv file. We also want to see the column headings and dimension of our data set.

```
agri <- read.csv("C:/Users/91776/Downloads/lec12data.csv")
dim(agri)
## [1] 8 2
names(agri)
## [1] "variety" "yield"
```

Now since our Dataset is small, we might as well print the whole dataset to see the actual cell values. We start by plotting to get a idea of whats happening in our dataset.

```
plot(yield~variety,agri)
```

variety	yield
1	210.3
2	245.0
2	248.9
3	212.3
3	230.4
2	250.1
1	213.5
1	212.4
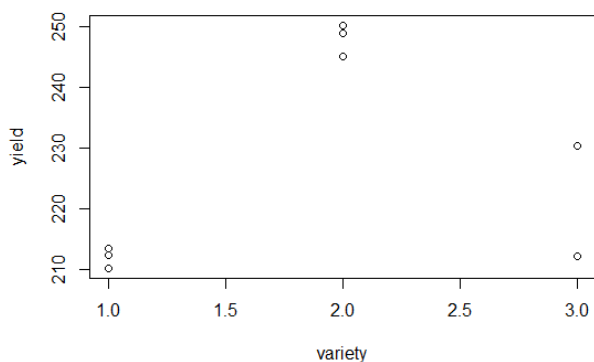
This is a scatter-plot. However our Variety is a nominal variable. So taking a values between 1 and 2, say 1.5 is meaningless here. Hence we need to change variable type to factor to inform R about the natural of the variable.
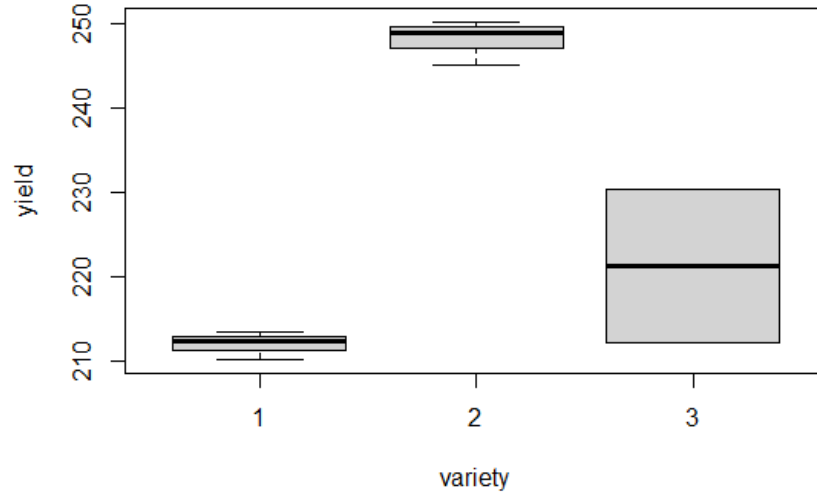
Figure 1: Box and Whisker plot

```
agri$variety=factor(agri$variety)
plot(yield~variety,agri)
```

Now we get a individual box and whisker plot for each of our variety. Here variety 3 has big boxplot, meaning that the scatter is huge. In any real life data,we should be skeptical about using the linear model for this type of plot.

In our toy dataset, we know we have just 2 observation for variety three, so we can pretend that it is okay and proceed forward.