

Assignment 2

Aytijhya Saha
ROLL No. BS2002

INDIAN STATISTICAL INSTITUTE, KOLKATA

September 1, 2022

Two-factor model - R lab: Exploring data for additive model

Before trying to fit a linear model on a data, it is always good to explore the data. Here, we will explore a two-factor data.

```
1 dat=read.csv('data.csv')
2 dat
```

	variety	tilt	yield
1	IR8	plain	254.2
2	IR8	plain	253.9
3	IR8	plain	254.4
4	IR8	plain	254.2
5	IR8	plain	254.0
6	IR8	slope	261.0
7	IR8	slope	260.4
8	IR8	slope	261.1
9	IR8	slope	260.9
10	IR8	slope	261.1
11	Jaya	plain	264.5
12	Jaya	plain	264.7
13	Jaya	plain	264.3
14	Jaya	plain	264.2
15	Jaya	plain	264.3
16	Jaya	slope	270.7
17	Jaya	slope	271.3
18	Jaya	slope	270.6

```

19      Jaya slope 271.2
20      Jaya slope 270.7
21 Taichung plain 284.2
22 Taichung plain 284.6
23 Taichung plain 284.6
24 Taichung plain 285.0
25 Taichung plain 284.4
26 Taichung slope 291.1
27 Taichung slope 291.1
28 Taichung slope 291.2
29 Taichung slope 291.2
30 Taichung slope 291.5

```

=====

In our data, there are two factor variables- variety and tilt. So, we declare the first two columns of the data frame as factor variables.

```

1 dat$variety=factor(dat$variety)
2 dat$tilt=factor(dat$tilt)

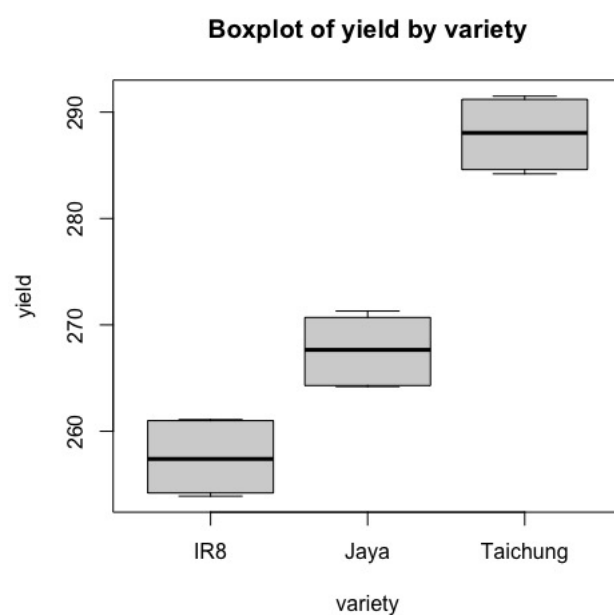
```

Now we plot the boxplots of yield vs tilt and yield vs variety.

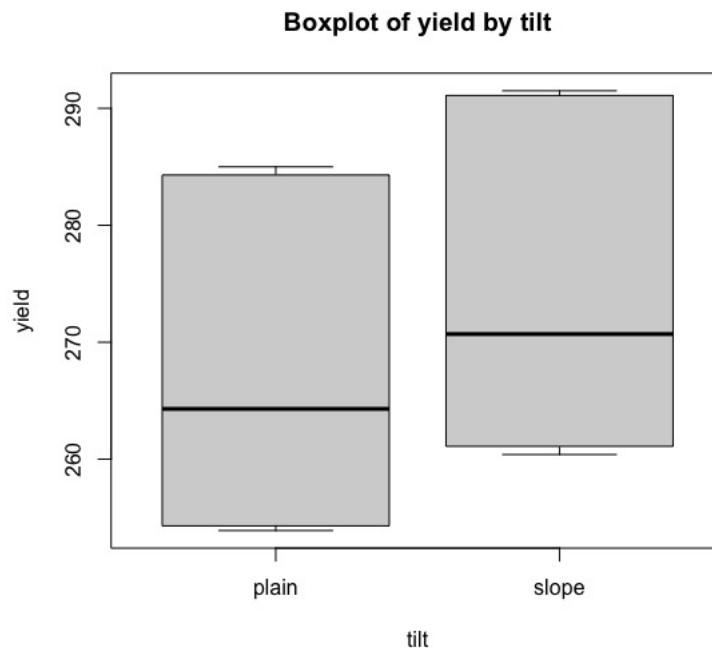
```

1 with(dat, plot(yield ~ variety,main="Boxplot of yield by variety"))
2 with(dat, plot(yield ~ tilt,main="Boxplot of yield by tilt"))

```



The central value of different varieties are different, but they are almost homoscedastic.



It seems again homoscedasticity is a pretty good assumption, but the central values are different.

Alternatively, to get the boxplots, we could also run the following line, which will give the same two plots one by one.

```
1 with(dat, plot(yield ~ variety*tilt))
```

Now, we shall create violin plots for the same set of variables as we did for boxplots. Violin plots are similar to box plots, except that they also show the kernel probability density of the data at different values. Typically, violin plots will include a marker for the median of the data and a box indicating the interquartile range, as in standard box plots.

```
1 library(ggplot2)
2
3 vp.var <- ggplot(dat, aes(x=variety, y=yield, fill=variety)) +
4   geom_violin(trim=FALSE)+
5   geom_boxplot(width=0.1, fill="white")+
6   xlab("Variety") + ylab("Yield")+
7   ggtitle("Violin plot of yield by variety")+
8   ggeasy::easy_center_title()
9
10 vp.var
11
12 vp.tilt <- ggplot(dat, aes(x=tilt, y=yield, fill=tilt)) +
```

```

13 geom_violin(trim=FALSE)+
14 geom_boxplot(width=0.1, fill="white")+
15 xlab("Tilt") + ylab("Yield")+
16 ggtitle("Violin plot of yield by tilt")+
17 ggeasy::easy_center_title()
18
19 vp.tilt

```

