perché si dice
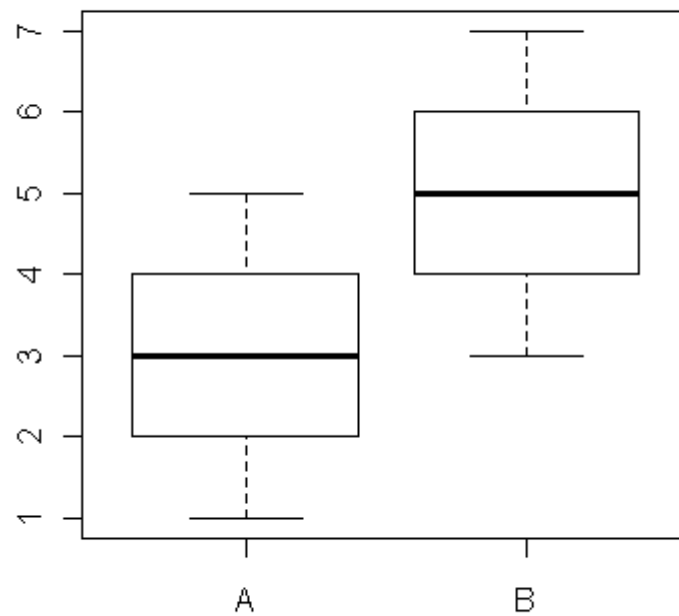# analisi della varianza?
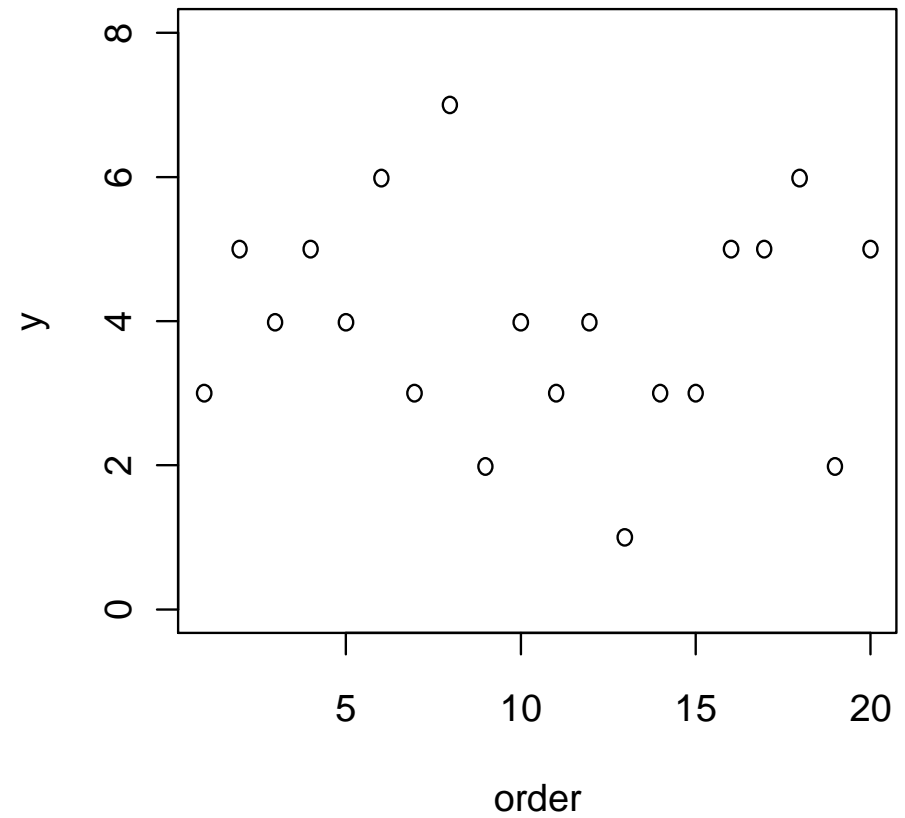
il dataset di Crawley

20 dati relativi all'ozono: 10 in A, 10 in B
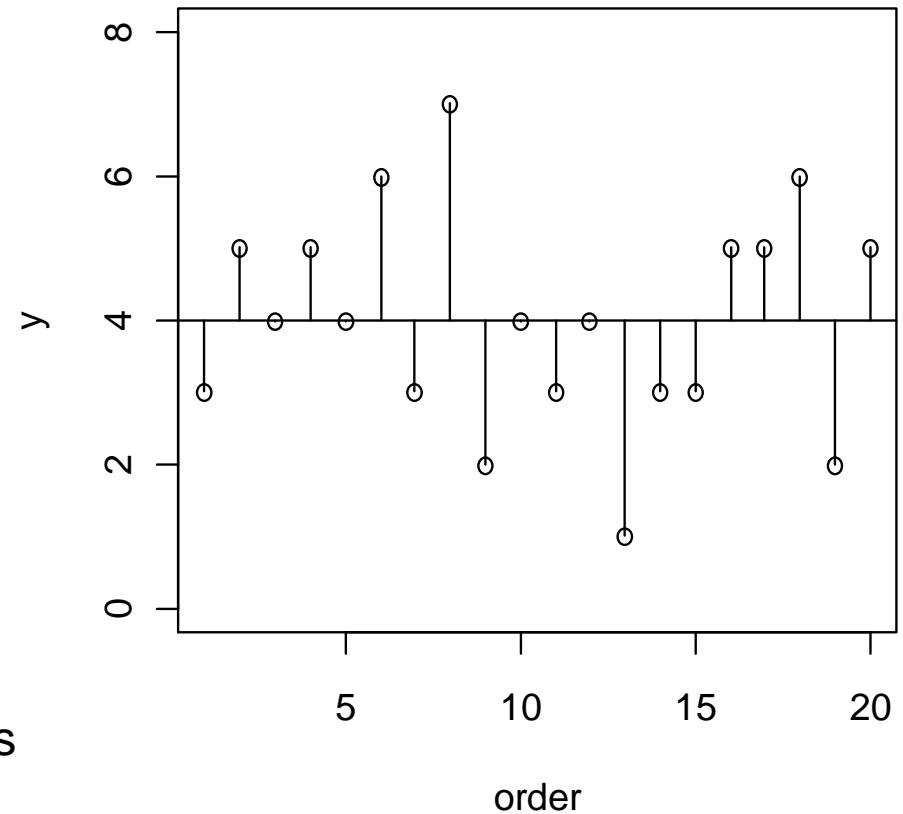
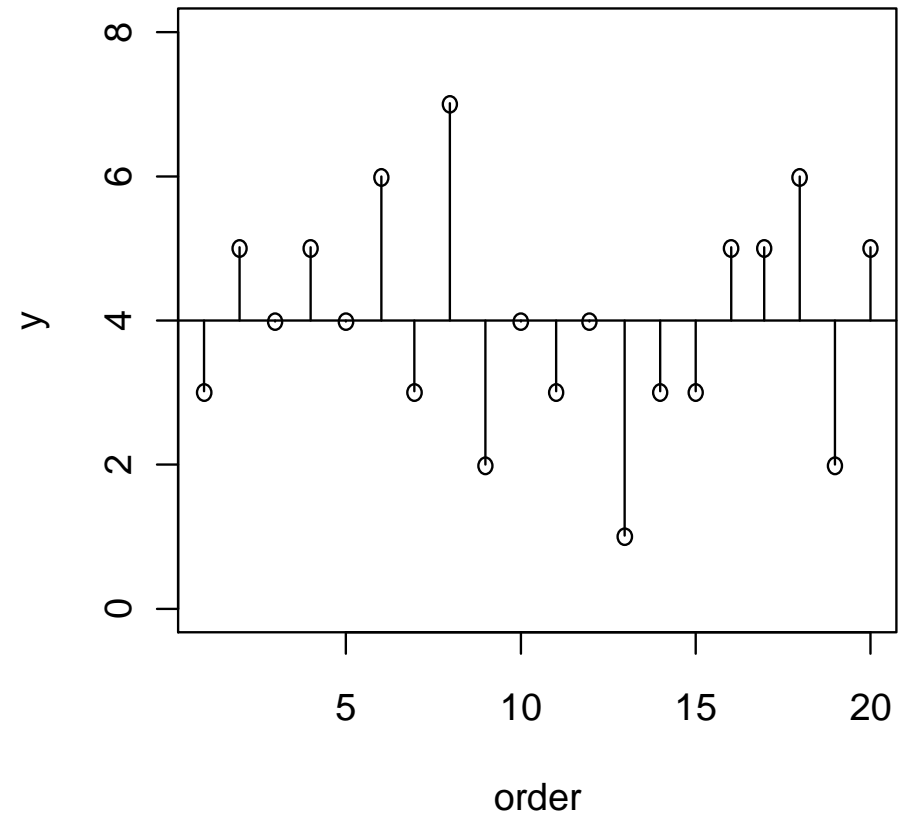| ozone	garden
3	A
5	B
4	A
5	B
4	A
6	B
3	A
7	B
2	A
4	B
3	A
4	B
1	A
3	B
3	A
5	B
5	A
6	B
2	A
5	B

il time-plot

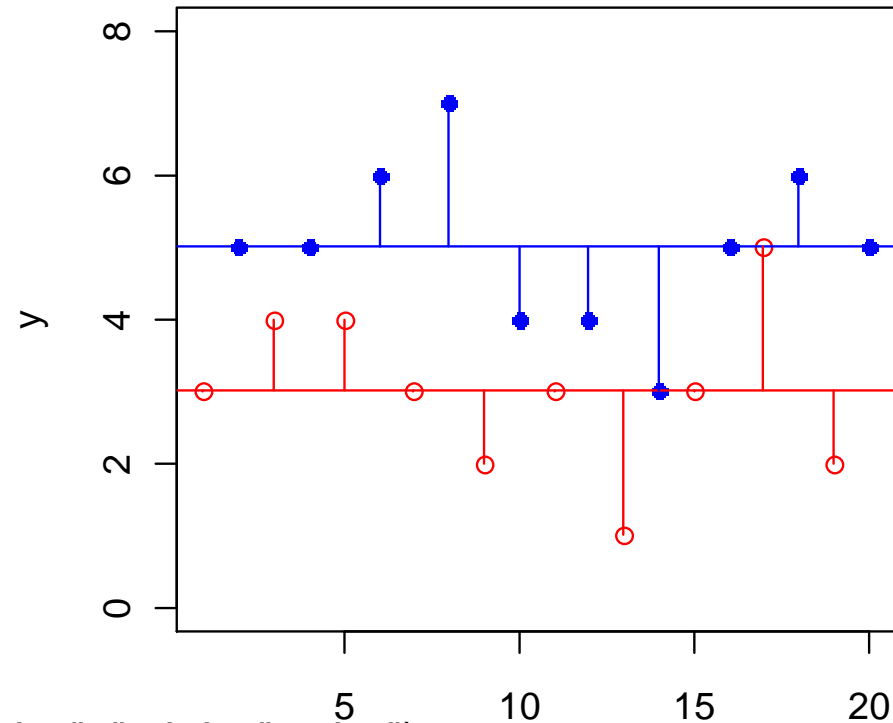plot(1:20,ozone,ylim=c(0,8),ylab="y",xlab="order")

c'è grande devianza


```
 # there is a lot of scatter
> # let's add the mean and residual lines
> abline(mean(ozone),0)
> for(i in 1:20) lines(c(i,i),c(mean(ozone),ozone[i]))
>
```

c'è grande devianza



```
> # Here is the total deviance
> SSY<-sum((ozone-mean(ozone))^2)
> SSY
[1] 44
```

garden A = red, garden B = blue


```
> plot(1:20,ozone,ylim=c(0,8),type="n",ylab="y",xlab="order")
> points(seq(1,19,2),ozone[garden=="A"],pch=1, col="red")
> points(seq(2,20,2),ozone[garden=="B"],pch=16, col="blue")
> abline(mean(ozone[garden=="A"]),0, col="red")
> abline(mean(ozone[garden=="B"]),0, col="blue")
> for (i in 1:11){  lines(c(2*i-1,2*i-1),
c(mean(ozone[garden=="A"]),ozone[garden=="A"][i]), col="red")
+ lines(c(2*i,2*i), c(mean(ozone[garden=="B"]),ozone[garden=="B"][i]),
col="blue")}
```
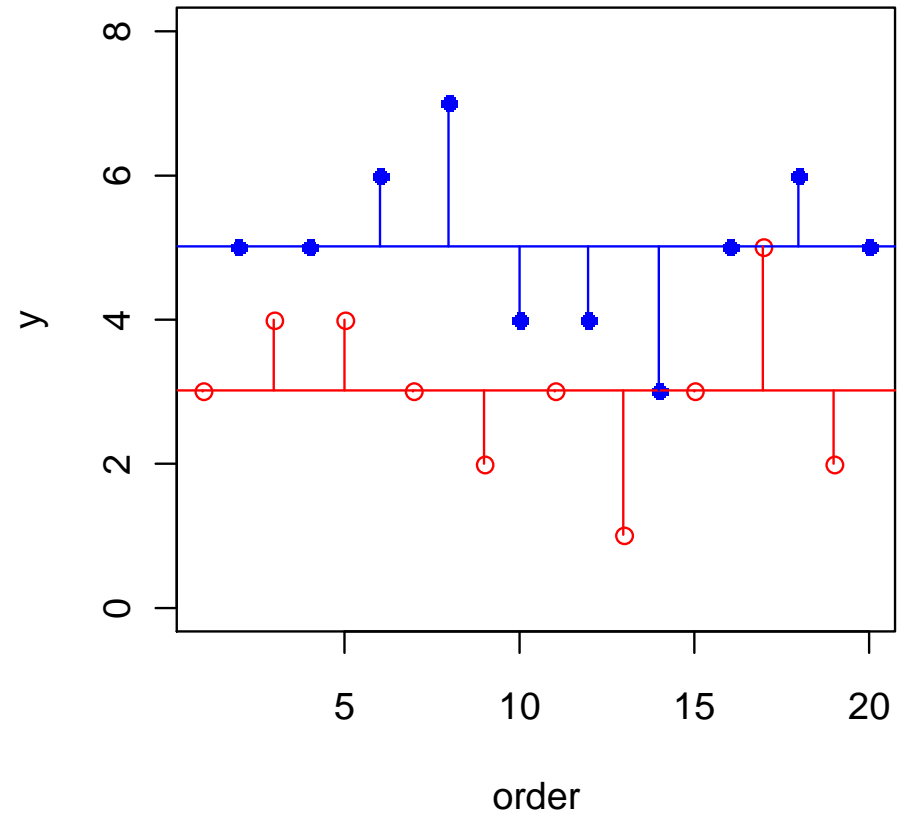
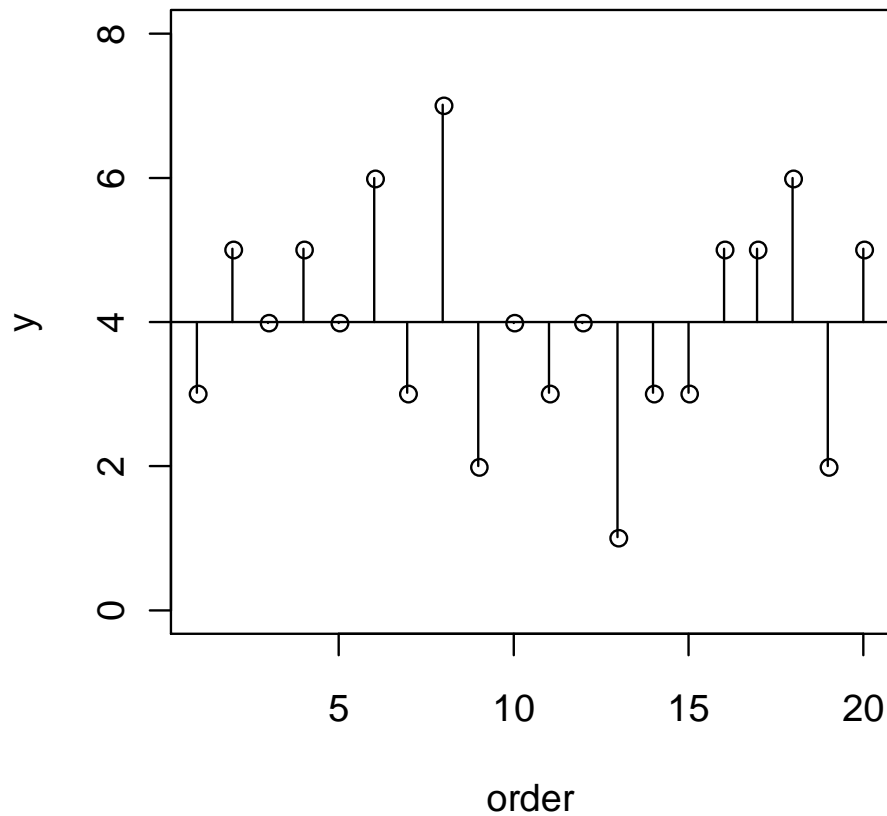# garden A = red,
garden B = blue
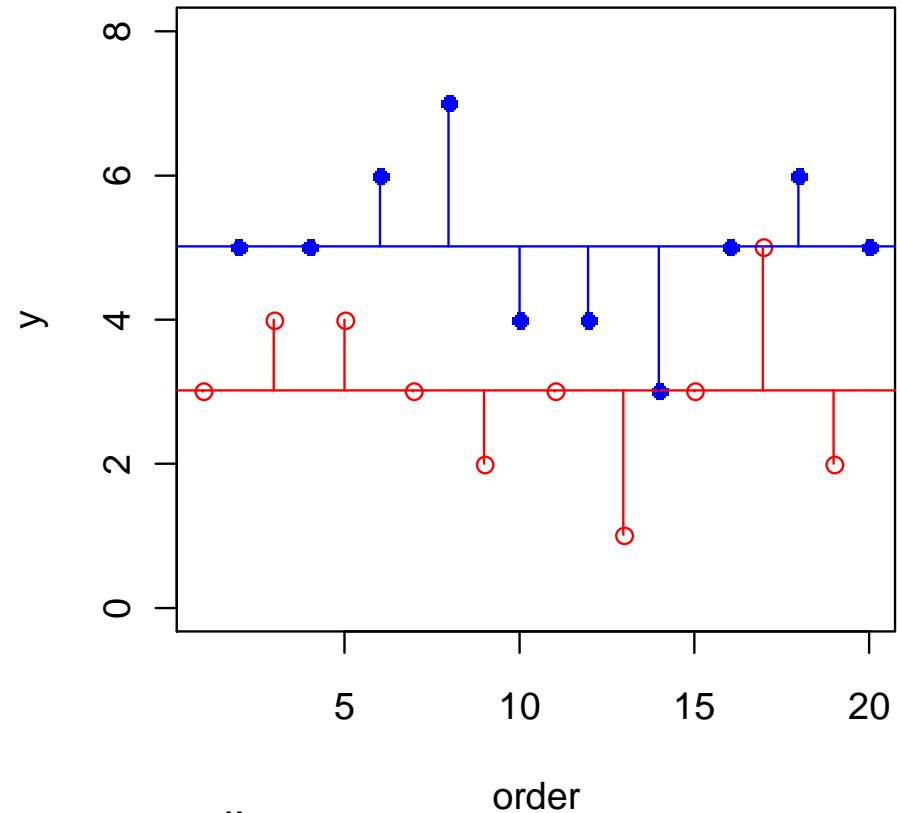

```
> # the deviances from individual means are smaller
> sum((ozone[garden=="A"]-mean(ozone[garden=="A"]))^2)
[1] 12
> sum((ozone[garden=="B"]-mean(ozone[garden=="B"]))^2)
[1] 12
>
```

garden A = red, garden B = blue
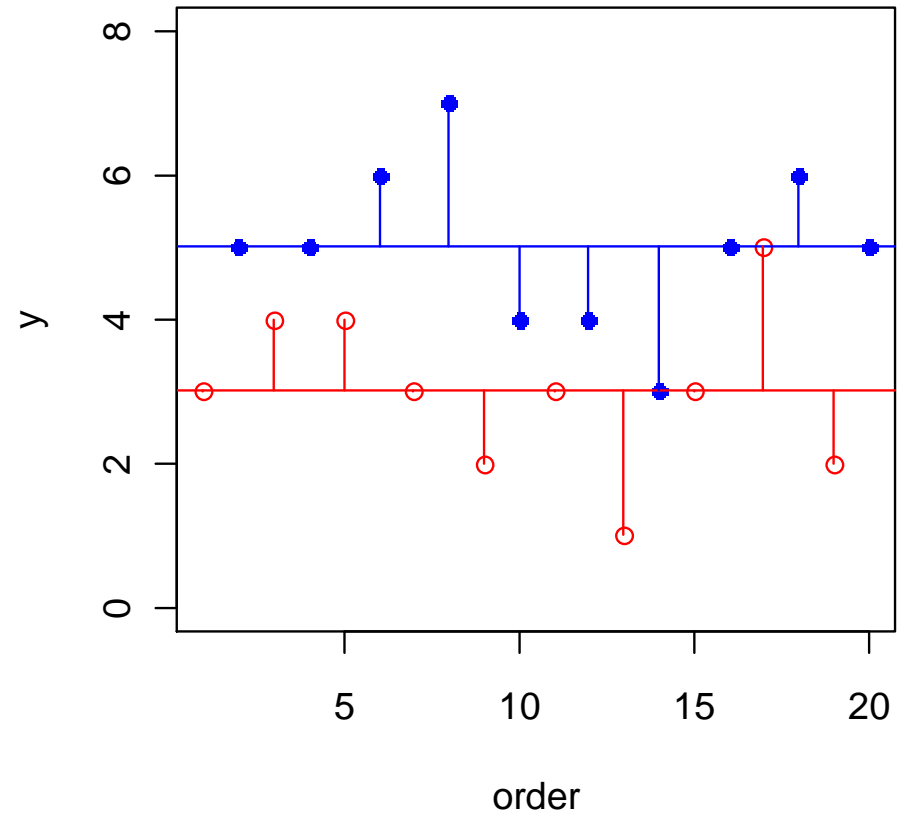

```
> # Here is the error sum of squares
> SSE <- sum((ozone[garden=="A"]-mean(ozone[garden=="A"]))^2) +
sum((ozone[garden=="B"]-mean(ozone[garden=="B"]))^2)
> SSE
[1] 24
>
```

```
> # conclusion: if the means were not different,
> # red and blue horizontal lines would be the same
> # But if the means are different, the residual are smaller,
> # and therefore the deviance form the individual treatment mean
> # is smaller than the deviance from the overall mean
```

La tavola Anova

> # Here it is, easily, the Anova table

```
> summary(aov(ozone~garden))

             Df  Sum Sq Mean Sq F value    Pr(>F)
garden        1 20.0000 20.0000      15 0.001115 **
Residuals    18 24.0000  1.3333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> 1-pf(15.0,1,18)
[1] 0.001114539
```