Section 5.3 Fitting the ANOVA Model

Loaded needed packages

```
library(Stat2Data)
library(mosaic)
```

EXAMPLE 5.5 Seeing the leafhopper data as a sum of overlays

Create a dataframe for **Leafhoppers** and look at the structure of the data.

```
data("Leafhoppers")
str(Leafhoppers)
```

```
## 'data.frame':    8 obs. of  3 variables:
##  $ Dish: int  1 2 3 4 5 6 7 8
##  $ Diet: Factor w/ 4 levels "Control","Fructose",..: 1 1 4 4 3 3 2 2
##  $ Days: num  2.3 1.7 3.6 4 2.9 2.7 2.1 2.3
```

Question 0: What is the overall or "grand"" average?

```
ybar=mean(Leafhoppers$Days)
ybar
```

```
## [1] 2.7
```

Question 1: How far is each group average from the grand average?

```
GroupMeans=mean(Days~Diet,data=Leafhoppers)
GroupMeans
```

```
##  Control Fructose  Glucose  Sucrose
##      2.0      2.2      2.8      3.8
```

```
GroupMeansDiff=GroupMeans-ybar
GroupMeansDiff
```

```
##  Control Fructose  Glucose  Sucrose
##     -0.7     -0.5      0.1      1.1
```

Question 2: How far is each response from its group average?

```
Residuals=Leafhoppers$Days-GroupMeans[Leafhoppers$Diet]
Residuals
```

```
##  Control  Control  Sucrose  Sucrose  Glucose  Glucose Fructose Fructose
##      0.3     -0.3     -0.2      0.2      0.1     -0.1     -0.1      0.1
```

FIGURE 5.9 Leafhopper data as a sum of overlays

Put the pieces of the decomposition into a dataframe

```
Overlays=data.frame(Response=Leafhoppers$Days)
Overlays$GrandAverage=rep(ybar,8)
Overlays$TreatmentEffects=GroupMeansDiff[Leafhoppers$Diet]
Overlays$Residuals=Residuals
Overlays
```

```
##   Response GrandAverage TreatmentEffects Residuals
## 1      2.3          2.7             -0.7       0.3
## 2      1.7          2.7             -0.7      -0.3
## 3      3.6          2.7              1.1      -0.2
## 4      4.0          2.7              1.1       0.2
## 5      2.9          2.7              0.1       0.1
## 6      2.7          2.7              0.1      -0.1
## 7      2.1          2.7             -0.5      -0.1
## 8      2.3          2.7             -0.5       0.1
```

EXAMPLE 5.6 Evaluating psychotherapy

Create a dataframe for **Undoing** and look at the structure of the data.

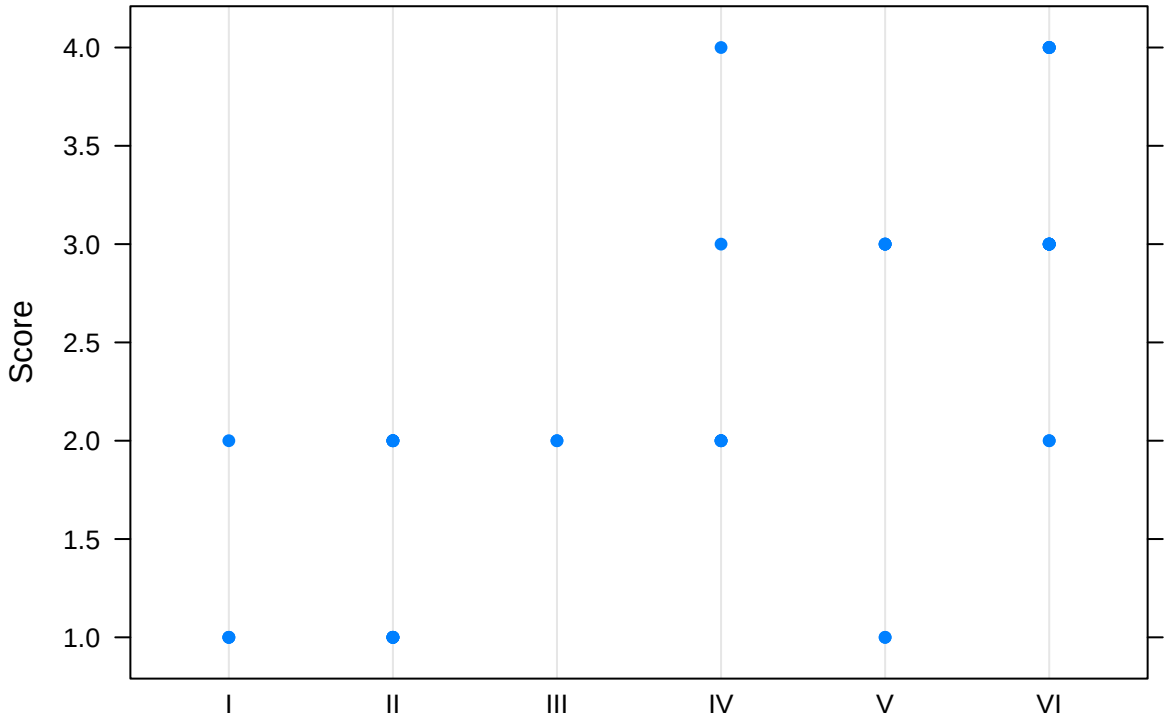
```
data(Undoing)
str(Undoing)
```

```
## 'data.frame':    44 obs. of  3 variables:
##  $ Group : Factor w/ 6 levels "I","II","III",..: 1 1 1 1 2 2 2 2 2 2 ...
##  $ Score : int  1 1 1 2 1 1 1 1 1 1 ...
##  $ Symbol: int  1 1 1 1 0 0 0 0 0 0 ...
```

FIGURE 5.10 The Undoing data

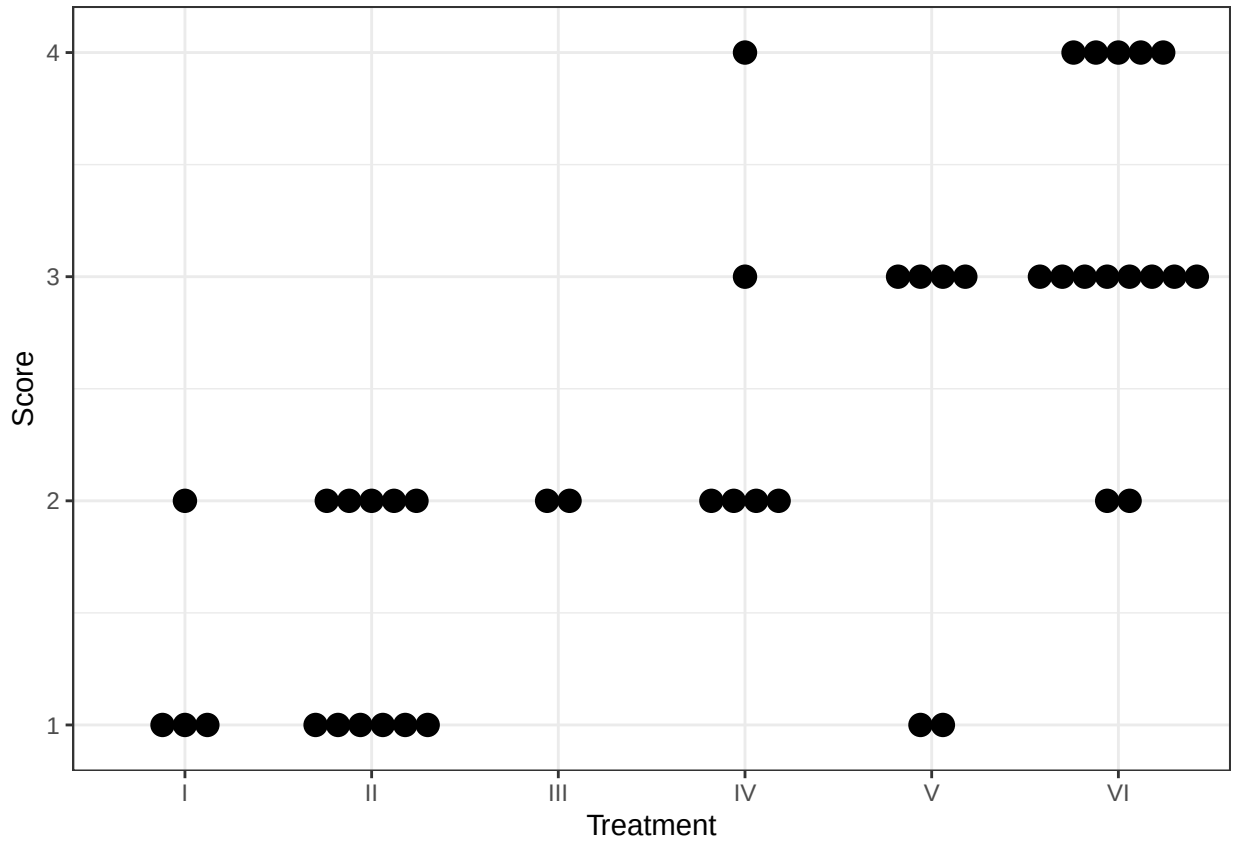
Note: There's some trouble if we use the generic dotplot( ) function, since multiple values are at the same point.

```
dotplot(Score~Group,data=Undoing)
```



A much better dotplot can be created with ggplot().

```
ggplot(Undoing, aes(x = Group, y = Score)) +
  geom_dotplot(binaxis = "y", stackdir = "center",binwidth=0.1) +
  labs(x="Treatment") +theme_bw()
```



Create a subset with just Groups I, III, and IV.

```
Undoing134=subset(Undoing,Group %in% c("I","III","IV"))
Undoing134$Group=factor(Undoing134$Group,levels=c("I","III","IV"))
Undoing134
```

```
##    Group Score Symbol
## 1      I     1      1
## 2      I     1      1
## 3      I     1      1
## 4      I     2      1
## 16   III     2      1
## 17   III     2      1
## 18    IV     2      1
## 19    IV     2      1
## 20    IV     2      1
## 21    IV     2      1
## 22    IV     3      1
## 23    IV     4      1
```

Question 0: What is the overall or "grand" average?

```
ScoreMean=mean(Undoing134$Score)
ScoreMean
```

```
## [1] 2
```

Question 1: How far is each group average from the grand average?

```
ScoreMeans=mean(Score~Group,data=Undoing134)
ScoreMeans
```

```
##    I  III   IV
## 1.25 2.00 2.50
```

```
GroupEffects=ScoreMeans-ScoreMean
GroupEffects
```

```
##     I   III    IV
## -0.75  0.00  0.50
```

Question 2: How far is each response from its group average?

```
UndoingResid=Undoing134$Score-ScoreMeans[Undoing134$Group]
UndoingResid
```

```
##     I     I     I     I   III   III    IV    IV    IV    IV    IV    IV
## -0.25 -0.25 -0.25  0.75  0.00  0.00 -0.50 -0.50 -0.50 -0.50  0.50  1.50
```

FIGURE 5.11 Decomposition of the psychotherapy data

Put the pieces of the decomposition into a dataframe

```
UndoingDecomposition=data.frame(Response=Undoing134$Score)
UndoingDecomposition$GrandMean=rep(ScoreMean,12)
UndoingDecomposition$TreatmentEffects=GroupEffects[Undoing134$Group]
UndoingDecomposition$Residuals=UndoingResid
UndoingDecomposition
```

```
##    Response GrandMean TreatmentEffects Residuals
## 1         1         2            -0.75     -0.25
## 2         1         2            -0.75     -0.25
## 3         1         2            -0.75     -0.25
## 4         2         2            -0.75      0.75
## 5         2         2             0.00      0.00
## 6         2         2             0.00      0.00
## 7         2         2             0.50     -0.50
## 8         2         2             0.50     -0.50
## 9         2         2             0.50     -0.50
## 10        2         2             0.50     -0.50
## 11        3         2             0.50      0.50
## 12        4         2             0.50      1.50
```

EXAMPLE 5.7 ANOVA table for the leafhoppers

Note: R's ANOVA table includes a P-value that is introduced in Section 5.4, but does not include a row for Total.

```
anovamodel=aov(Days~Diet,data=Leafhoppers)
summary(anovamodel)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Diet         3   3.92   1.307   17.42 0.00925 **
## Residuals    4   0.30   0.075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

EXAMPLE 5.8 ANOVA table for the psychotherapy data

```
summary(aov(Score~Group,data=Undoing134))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Group        2   3.75  1.8750   3.971 0.0581 .
## Residuals    9   4.25  0.4722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compute sums of squares for Leafhoppers

```
SSGroups=sum(Overlays$TreatmentEffects^2)
SSE=sum(Overlays$Residuals^2)
SSTotal=SSGroups+SSE
paste("SSGroups=",SSGroups)
```

```
## [1] "SSGroups= 3.92"
```

```
paste("SSE = ",SSE)
```

```
## [1] "SSE =  0.3"
```

```
paste("SSTotal=",SSTotal)
```

```
## [1] "SSTotal= 4.22"
```