

Sums of Squares 9 – 14

Today's activity will guide you to understand three quantities used to measure variability in data, SSE, SST, and SSA:

- The sums of squared errors (SSE) is a measure of the variability of responses within a group.

$$SSE = \sum_{j=1}^J \sum_{n=1}^{n_j} (y_{ij} - \bar{y}_j)^2 = \sum_{j=1}^J \sum_{n=1}^{n_j} e_{ij}^2$$

- The total sums of squares (SST) is a measure of the variability of responses around the overall average, \bar{y} .

$$SST = \sum_{j=1}^J \sum_{n=1}^{n_j} (y_{ij} - \bar{y})^2$$

- The sums of squares for the explanatory variable (SSA) is a measure of the variability of the group averages around the overall average.

$$SSA = \sum_{j=1}^J \sum_{n=1}^{n_j} (\bar{y}_j - \bar{y})^2 = SST - SSE$$

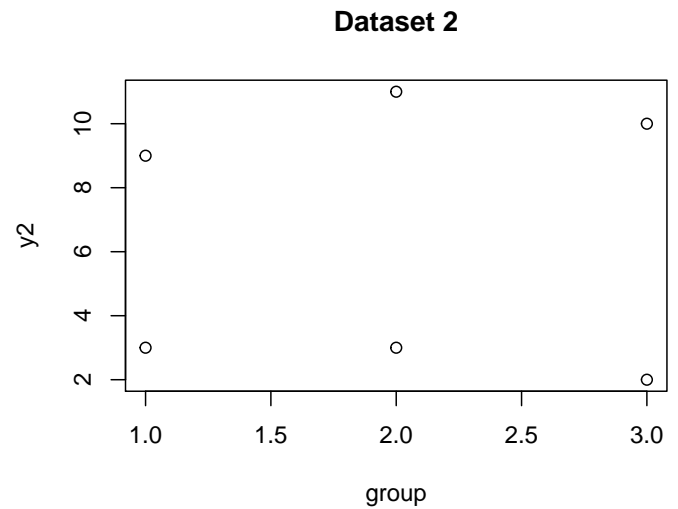
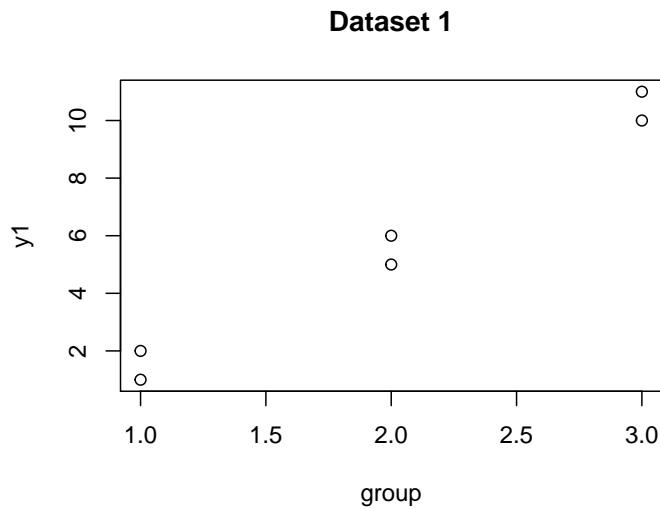
```
y1 <- c(1,2, 5,6, 10,11)
group <- c(rep(1,2), rep(2,2), rep(3,2))
one <- cbind.data.frame(y1,group)
one
```

```
##   y1 group
## 1  1     1
## 2  2     1
## 3  5     2
## 4  6     2
## 5 10     3
## 6 11     3
```

```
y2 <- c(3,9, 3,11, 2,10)
two <- cbind.data.frame(y2, group)
two
```

```
##   y2 group
## 1  3     1
## 2  9     1
## 3  3     2
## 4 11     2
## 5  2     3
## 6 10     3
```

```
par(mfrow=c(1,2))
plot(group, y1, main="Dataset 1")
plot(group, y2, main="Dataset 2")
```



1. For the first dataset, calculate $\bar{\bar{y}}$ and \bar{y}_j (for each of the groups).
2. For the first dataset, calculate SSE, SST, and SSA.
3. Find the ratio, SSA/SSE.
4. Repeat these steps for the second dataset.
5. For each dataset, describe whether you would choose the single mean model or the separate means model. In your answer, refer to the plot and the ratio SSA/SSE.