One way ANOVA

Let us consider the PlantGrowth data set in R. We first load it into R, then copy it a data frame called `df`.

```
data(PlantGrowth)
df <- data.frame(PlantGrowth)
```

What have we got?

```
str(df)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
##  $ group : Factor w/ 3 levels "ctrl","trt1",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
overall_mean <- mean(df$weight)
overall_mean
```

```
## [1] 5.073
```

```
gpcounts<-sapply(split(df$weight,df$group),length)
gpcounts
```

```
## ctrl trt1 trt2
##   10   10   10
```

```
gpmeans<-sapply(split(df$weight,df$group),mean)
gpmeans
```

```
##  ctrl  trt1  trt2
## 5.032 4.661 5.526
```

```
vars<-sapply(split(df$weight,df$group),var)
vars
```

```
##      ctrl      trt1      trt2
## 0.3399956 0.6299211 0.1958711
```

Between-group variability

```
bgv <- sum((gpcounts*(gpmeans-overall_mean)^2)/(3-1))
bgv
```

```
## [1] 1.88317
```

Within-group variability

```
wgv <- sum(vars/(30-3))
wgv
```

```
## [1] 0.04317733
```

F-statistic

```
F <- bgv/wgv
F
```

```
## [1] 43.61479
```