Case study: leukemia data

From Efron and Hastie (2016), pages 8-10.

Gene expression measurements on 72 leukemia patients, 47 ALL (acute lymphoblastic leukemia), 25 AML (acute myeloid leukemia, a worse prognosis). These data arise from the landmark Golub et al (1999) Science paper.

1. Download the data

```
leukemia <- read.csv("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
```

2. Compare the genetic expressions in the two groups for gene 136 by using a two-sample t-test

```
##
##  Two Sample t-test
##
## data:  leukemia[136, group == "ALL"] and leukemia[136, group == "AML"]
## t = -3.014, df = 70, p-value = 0.003589
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.32817995 -0.06680742
## sample estimates:
## mean of x mean of y
## 0.7524794 0.9499731
```

3. Install the R package `globaltest`

```
source("https://bioconductor.org/biocLite.R")
biocLite("globaltest")
```

perform the global test for the leukemia data

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     leukemia
```

```
##    p-value Statistic Expected Std.dev #Cov
## 1 1.48e-19      6.91     1.41   0.243 7128
```