Randomization tests with R

```
> # Little Tubes Data: Don't throw the outlier away
> rm(list=ls())
> tubes =
read.table("http://www.utstat.toronto.edu/~brunner/appliedf14/code_n_data/l
ecture/littletubes.data")
> attach(tubes)
> mcg=factor(mcg)
>
> mod0 = lm(length10 ~ mcg) ; summary(mod0)

Call:
lm(formula = length10 ~ mcg)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7125 -0.3250  0.0125  0.2406  1.7875

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.7750     0.3018  92.030  < 2e-16 ***
mcg205       -2.4375     0.4268  -5.711 2.05e-05 ***
mcg213       -2.7125     0.4268  -6.355 5.49e-06 ***
mcg221       -4.8250     0.4268 -11.305 1.31e-09 ***
mcg223       -3.4250     0.4268  -8.025 2.35e-07 ***
mcg225       -3.6125     0.4268  -8.464 1.09e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6036 on 18 degrees of freedom
Multiple R-squared: 0.8889,   Adjusted R-squared: 0.8581
F-statistic: 28.81 on 5 and 18 DF,  p-value: 5.415e-08

> Fstat0 = summary(mod0)[[10]][1] # Just the F statistic
> Fstat0
  value
28.8147
>
```

There are $\binom{24}{4\ 4\ 4\ 4\ 4\ 4} = \frac{24!}{4!4!4!4!4!4!} = 3,246,670,537,110,000$ ways to scramble the data.

```
> set.seed(9999); nsim=1000; Fstat=numeric(nsim)
> for(j in 1:nsim)
+     {
+     y = sample(length10) # Sample with replacement = random permutation
+     Fstat[j] = summary(lm(y~mcg))[[10]][1]
+     } # Next simulation
> Fstat[1:10] # Look at the first 10
 [1] 0.4641829 0.7592105 0.3926211 0.6115303 0.7043749 1.2725364 1.5770560
3.1167085 0.4330515
[10] 0.3067711
> length(Fstat[Fstat>=Fstat0])/nsim
[1] 0
>
> # Try dropping mcg198 (highest) and 221 (lowest) as a test.
>
> dropit=numeric(24)
> dropit[mcg==198]=1; dropit[mcg==221]=1
> MCG = mcg[dropit==0]; Length10 = length10[dropit==0]
>
> mod0 = lm(Length10 ~ MCG) ; summary(mod0)

Call:
lm(formula = Length10 ~ MCG)

Residuals:
     Min       1Q   Median       3Q      Max
-0.71250 -0.36875 -0.01875  0.24062  1.78750

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.3375     0.3430  73.880   <2e-16 ***
MCG213       -0.2750     0.4850  -0.567   0.5812
MCG223       -0.9875     0.4850  -2.036   0.0644 .
MCG225       -1.1750     0.4850  -2.423   0.0322 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6859 on 12 degrees of freedom
Multiple R-squared: 0.4013,   Adjusted R-squared: 0.2516
F-statistic: 2.681 on 3 and 12 DF,  p-value: 0.09398
```

```
> Fstat0 = summary(mod0)[[10]][1] # Just the F statistic
> Fstat0
   value
2.681169
>
> # There are 3,246,670,537,110,000 ways to scramble the data.
> set.seed(9999); nsim=1000; Fstat=numeric(nsim)
> for(j in 1:nsim)
+     {
+     y = sample(Length10) # Sample with replacement = random permutation
+     Fstat[j] = summary(lm(y~MCG))[[10]][1]
+     } # Next simulation
> pval = length(Fstat[Fstat>=Fstat0])/nsim; pval
[1] 0.084
> # Compare p = 0.09398
```
