Independent Samples ANOVA, Randomization Test with R

I obtained the code from Dave Howell. At the suggestion of Alex Schoemann, I deleted the row “par(mfrow = c(2,1)).” The data are simulated, based on the results of this research I did long ago. The first $p$ value is obtained from the usual $F$ distribution, and is not valid if the assumptions of normality and homogeneity of variance are seriously violated. The second $p$ value is based on the resampling distribution of $F$ under the null hypothesis.

**Descriptive Statistics**

<table>
<thead>
<tr>
<th>Latency</th>
<th>Father</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
</tr>
<tr>
<td>Mean</td>
<td>105.18</td>
</tr>
<tr>
<td>Median</td>
<td>63.50</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>99.758</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.245</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-.048</td>
</tr>
</tbody>
</table>

As shown in the table above, the scores in the Father-Present group are positively skewed and those in the other two groups are close to U-shaped (remember that a perfectly U-shaped distribution has kurtosis = -2).

The non-normal nature of the scores is illustrated by these histograms. The solid curve represents a normal distribution with the same mean and standard deviation.
The R Output:

The obtained value of F from the standard F test is  3.686965 
This has an associated probability of  0.03069425 
The calculated value of p from randomized samples is  0.0308
# The Code

# Randomization test for one way ANOVA using
# Data file has column1 = group, the Independent Variable and column2 = dv, the Dependent Variable

# Load the data

data <- read.table("http://core.ecu.edu/psyc/wuenschk/StatData/ANOVA-Resample.txt", header = TRUE)

# Rename variables

names(data)
data$group <- as.factor(data$group)

# Set number of repetitions

nreps <- 5000

# Calculate sample sizes

N <- length(data$dv)
n.i <- as.vector(table(data$group))  # Create vector of sample sizes

# Calculate original F statistic

k <- length(n.i)
model <- anova(lm(data$dv ~ data$group))
obt.F <- model$"F value"[1]  # Our obtained F statistic
obt.p <- model$"Pr(>F)"

# Print obtained F and p values

cat("The obtained value of F from the standard F test is ",obt.F, "\n")
cat("This has an associated probability of ", obt.p, "\n")

# Perform randomization test

samp.F <- numeric(nreps)
counter <- 0

set.seed(1086)

# Timing statistics

# time1 <- proc.time()

for (i in 1:nreps) {
    newdv <- sample(data$dv)
    newModel <- anova(lm(newdv ~ data$group))
    samp.F[i] <- newModel$"F value"[1]
    if (samp.F[i] > obt.F) counter = counter + 1
}

# time2 <- proc.time()

# The timing statistics are 
# (time2 - time1)\n"

# The computing time was approx. 10 sec.

pvalue <- counter/nreps

cat("The calculated value of p from randomized samples is ",round(pvalue, digits = 4), "\n")

hist(samp.F, breaks = 50, main = "Histogram of F on Randomized Samples", xlab = "F value", probability = TRUE, col = "green", border = 1,
     xlim = c(0,7), ylim = c(0,1))

legend("topright", paste("obtained.F = ",round(obt.F, digits = 4)), col=1, cex = 0.8)

legend("right",paste("p-value = ",round(pvalue, digits = 4)))

arrows( 5.5, 0.8,obt.F,0, length = .125)

f <- seq(0, 7,.01)
dens <- df(f,3,41)
par(new = T)
plot(f,dens, col = "red", type = "l", xlim = c(0,7), ylim = c(0,1), xlab = "", ylab = "")
#polygon(f,dens, col = "red")

The Output

It will be in the console, along with the code, with code in red and output in blue, like this:

```r
$ obtained value of F from the standard F test is ",obt.F, "\n"
$ed value of F from the standard F test is  3.686965
$s has an associated probability of ", obt.p, "\n"
$n associated probability of  0.03069425 NA
$- numeric(nreps)
$<- 0
$(1086)
$<- proc.time()
$n 1:nreps) {
$- sample(data$dv)
$s1 <- anova(lm(newdv~data$group))
$s[i] <- newModel$"F value"[1]
$p.F[i] > obt.F) counter = counter + 1

$<- proc.time()
$The timing statistics are ",(time2 - time1),"\n"
$mputing time was approx. 10 sec.
$- counter/nreps
$he calculated value of p from randomized samples is ",pvalue, "\n \n"

$ated value of p from randomized samples is  0.0308

$p.F, breaks = 50, main = "Histogram of F on Randomized Samples",
Karl L. Wuensch, February, 2016