Permutation_Test

Tingrui Huang

April 13, 2019

Snodgrass Problem

```
prop <- c(0.225, 0.262, 0.217, 0.240, 0.230, 0.229, 0.235, 0.217,
              0.209, 0.205, 0.196, 0.210, 0.202, 0.207, 0.224, 0.223, 0.220, 0.201)
index <- c(rep("Twin",8),rep("Snodgrass",10))
sno_df <- data.frame(cbind(prop,index))
sno_df$prop <- as.numeric(as.character(sno_df$prop))

library(FSA)
```

```
## Warning: package 'FSA' was built under R version 3.5.3
```

```
## ## FSA v0.8.22. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.
```

```
Summarize(prop~index, data = sno_df,digits = 3)
```

```
##       index  n  mean    sd   min    Q1 median    Q3   max
## 1 Snodgrass 10 0.210 0.010 0.196 0.203  0.208 0.218 0.224
## 2      Twin  8 0.232 0.015 0.217 0.223  0.230 0.236 0.262
```

```
boxplot(prop~index, data = sno_df, ylab="Proportion")

library(coin)
```

```
## Warning: package 'coin' was built under R version 3.5.3
```

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

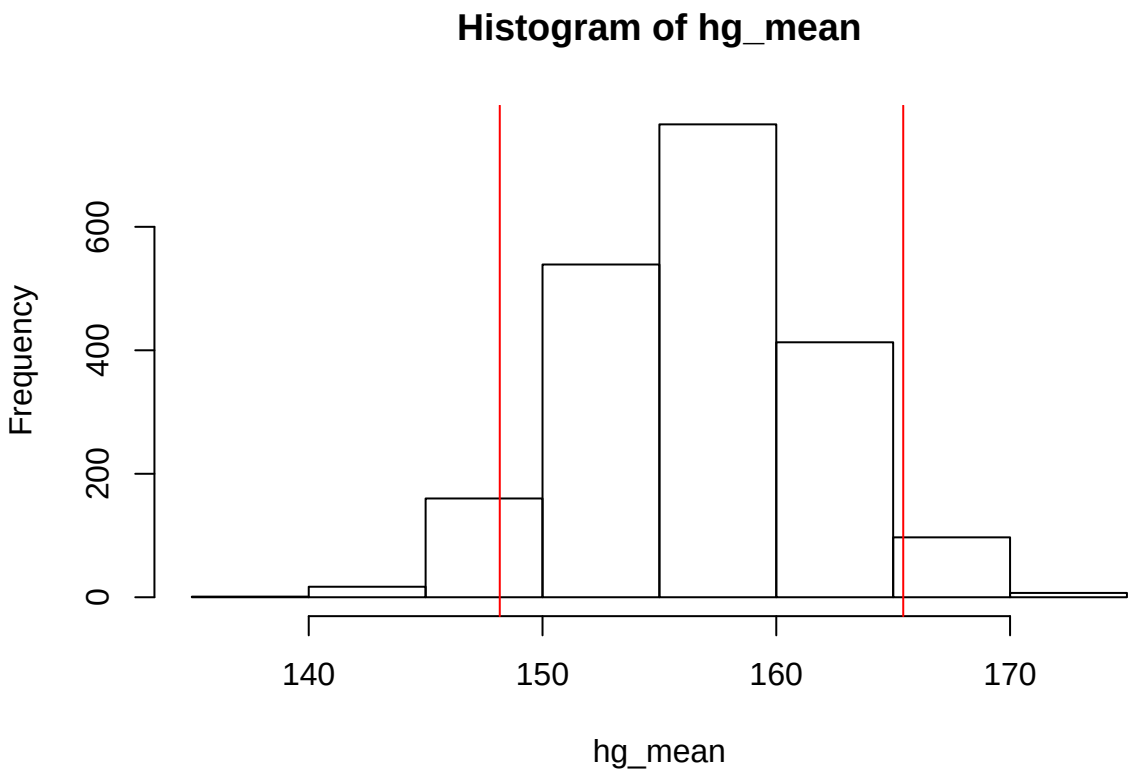

```
independence_test(prop~index, data = sno_df)
```

```
##
##  Asymptotic General Independence Test
##
## data:  prop by index (Snodgrass, Twin)
## Z = -2.87, p-value = 0.004104
## alternative hypothesis: two.sided
```

```
# Reject H0, the means are different
```

Hot Dog Problem

```
hotdog <- c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148,
            152, 111, 141, 153, 190, 157, 131, 149, 135, 132)
perm.sample <-  sample(hotdog,20,replace = TRUE)
hg_mean <- rep(NA,2000)
for (i in 1:2000) {
  perm.sample <-  sample(hotdog,20,replace = TRUE)
  hg_mean[i] <- mean(perm.sample)
}
hist(hg_mean)
abline(v=c(mean(hg_mean)+1.729*sd(hg_mean),mean(hg_mean)-1.729*sd(hg_mean)), col="red")
```



Reading Score Problem

```
# Data frame
score <- data.frame(c(1.23, 1.42, 1.41, 1.62, 1.55, 1.51, 1.60, 1.76,
          1.76, 1.41, 1.87, 1.49, 1.67, 1.81))
class <- data.frame(c(rep("Class1",8),rep("Class2",6)))
readingscore <- cbind.data.frame(class,score,stringsAsFactors = FALSE)
colnames(readingscore)[1] <- "class"
colnames(readingscore)[2] <- "score"

# Test Statistics
diff.means <- mean(readingscore$score[readingscore$class == "Class1"]) -
              mean(readingscore$score[readingscore$class == "Class2"])

# Create a function that randomly reassigns each observation to a different group and then takes the me
one.test <- function(grouping, variable) {
                resampled.group <- sample(grouping)
                mean(variable[resampled.group == "Class1"]) -
                mean(variable[resampled.group == "Class2"])
            }

# Repeat this permutation process 1,000 times to get a distribution of the mean difference of the permu
set.seed(1)
perm.means <- replicate(1000, one.test(readingscore$class, readingscore$score))
```

```r
# To check whether your test statistic is statistically different from 0
sig <- sum(perm.means > 0) # P-value is 0.481, mean of class 1 is not greater than mean of group 2

hist(perm.means)
abline(v=0, col="red")
```