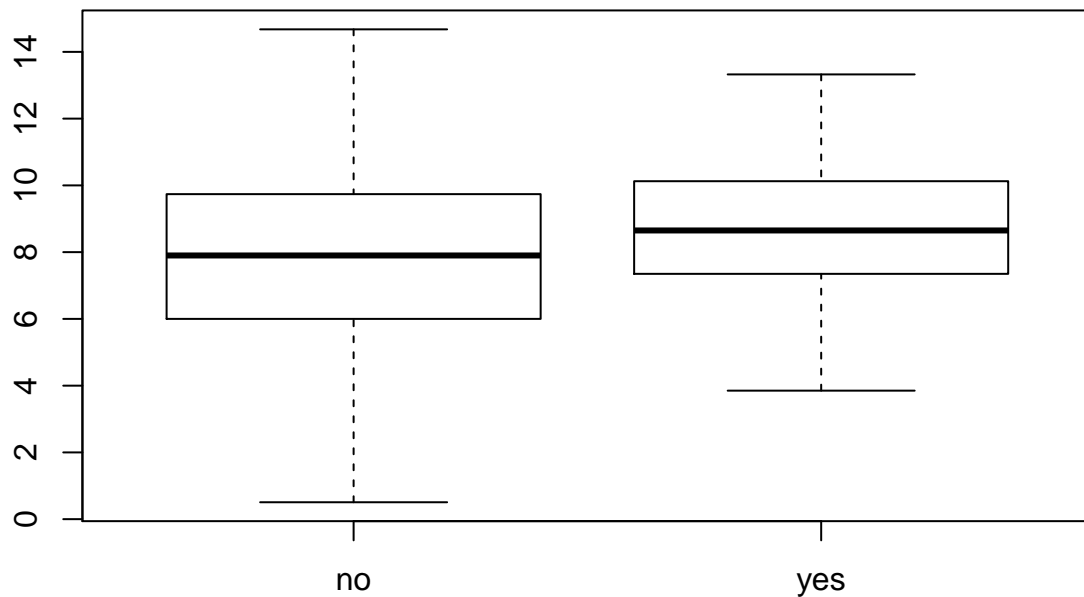


Two_Sample_T-test.R

matiasberretta

Sun May 5 13:35:26 2019

```
#####  
# Effect of Smoking on Lung Capacity  
#####  
# Load Data  
setwd("~/Documents/Career/Data_Science/R/Two_Sample_Test")  
LungCapData <- read.csv("LungCapData.txt", header=T, sep = "\t")  
attach(LungCapData)  
names(LungCapData)  
  
## [1] "LungCap" "Age" "Height" "Smoke" "Gender" "Caesarean"  
  
class(LungCap)  
  
## [1] "numeric"  
  
levels(Smoke)  
  
## [1] "no" "yes"  
  
# Visualize Smokers an Non-Smokers Lung Capacity  
boxplot(LungCap ~ Smoke)
```



```
# We run Levene's Test to determine whether population variances are equal
#
# Ho: population variances are equal
library(car)
```

```
## Loading required package: carData
```

```
leveneTest(LungCap~Smoke)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value    Pr(>F)
## group  1  12.955 0.0003408 ***
##      723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We run a Two-Sided T-test to determine whether lung capacity differs between smokers and non-smokers
#
# Ho: mean lung cap of smokes == mean lung cap of non smokers
# Two-Sided Test
# Assume non-equal variance
t.test(LungCap~Smoke, mu=0, alt="two.sided", conf=0.95, var.eq=F, paired=F)
```

```
##
```

```
## Welch Two Sample t-test
##
## data: LungCap by Smoke
## t = -3.6498, df = 117.72, p-value = 0.0003927
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.3501778 -0.4003548
## sample estimates:
## mean in group no mean in group yes
## 7.770188 8.645455
```

Given the p-value we can conclude that the alternative hypothesis is true
There is a significant difference in lung capacity
between smokers and non-smokers