# DVM Permutation Test - Work 1

Jeremy Williams October 11, 2017

```
#install.packages("combinat")
suppressMessages(suppressWarnings(library("combinat")))
```

# Correlation - Exact Test

#### **Pearson Correlation**

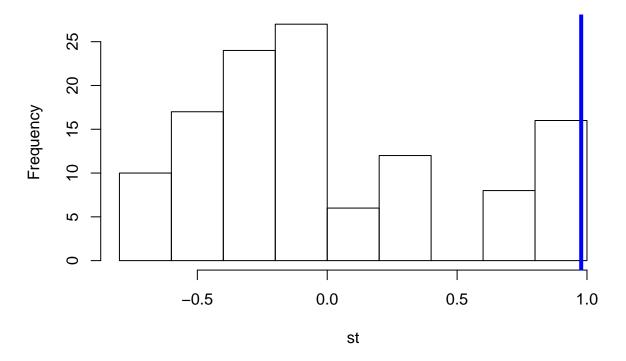
The first row is chest circumference (in inches) of five subjects. Let us call this X. The second row is the respective total volumes of air that can be breathed in and out in one minute (in liters) for the same five subjects. Let us call this Y.

The null hypothesis is, Ho: p = 0 "Chest circumference and volume of air are not correlated.".

The alternative hypothesis is, H1: p > 0 "Chest circumference and volume of air are positively correlated".

```
x= c(39,29,60,40,32)
y=c(11,5,20,8,6)
sttrue= cor(x,y,method = c("pearson"))
n=length(y)
nr=fact(n) #number of rearrangements to be examined
st=numeric(nr)
cnt=0
d=permn(y)
for (i in 1:nr){
   st[i]<-cor(d[[i]],x, method = c("pearson"))
   if(st[i] > sttrue)cnt=cnt+1 }
hist(st)
abline(v=sttrue, lwd=4, col="blue")
```

# Histogram of st



```
#True value using Pearson correlation
sttrue

## [1] 0.9777792

#p-value
cnt/nr
```

## [1] 0.008333333

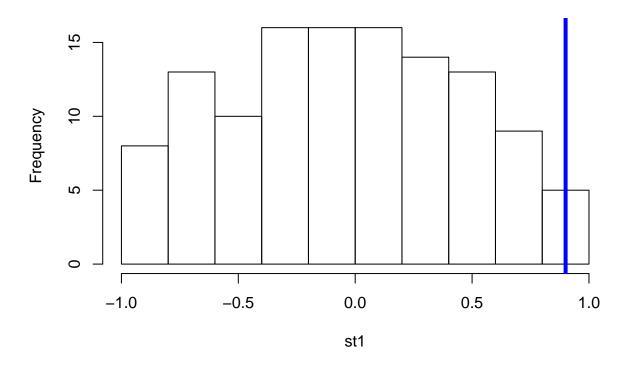
A small p-value (p "less than or equal to" 0.05) indicates strong evidence against the null hypothesis, so we reject the null hypothesis (Ho is rejected). Chest circumference and volume of air are positively correlated.

# **Spearman Correlation**

```
#---
x= c(39,29,60,40,32)
y=c(11,5,20,8,6)
sttrue1 = cor(x,y,method = c("spearman"))
n=length(y)
nr1=fact(n) #number of rearrangements to be examined
st1=numeric(nr1)
cnt=0
d=permn(y)
for (i in 1:nr1){
    st1[i]<-cor(d[[i]],x, method = c("spearman"))</pre>
```

```
if(st1[i] > sttrue1)cnt=cnt+1 }
hist(st1)
abline(v=sttrue1, lwd=4, col="blue")
```

# Histogram of st1



#True value using Spearman correlation sttrue1

## [1] 0.9

#p-value
cnt/nr1

## [1] 0.008333333

A small p-value (p "less than or equal to" 0.05) indicates strong evidence against the null hypothesis, so we also reject the null hypothesis (Ho is rejected). Chest circumference and volume of air are positively correlated.

# T-Test and Other Statistics

#### T-test with Permutation Test

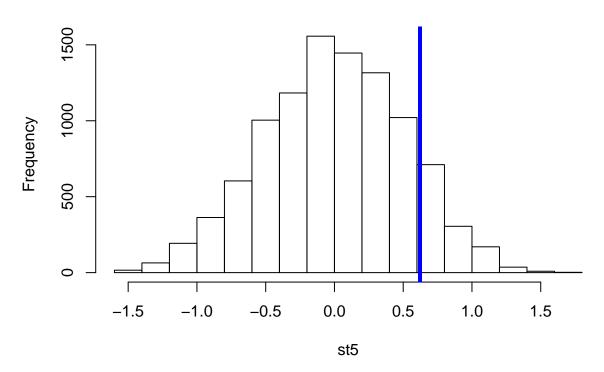
These are the increments of weight recorded in an experiment where a new additive has been added to a standard compound feed.

The null hypothesis is, Ho: ua - ub = 0 "The additive is not efficient".

The alternative hypothesis is, H1: ua - ub "not equal" 0 "The additive is efficient".

```
nr5=10000 #number of rearrangements to be examined
st5<-numeric(nr5)
stand5=c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
stand_ad5 = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8,
              3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
# Classical T-test
t.test(stand5,stand_ad5)
##
   Welch Two Sample t-test
##
## data: stand5 and stand_ad5
## t = -1.2936, df = 18.815, p-value = 0.2115
\#\# alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.629623 0.385179
## sample estimates:
## mean of x mean of y
## 3.177778 3.800000
n15=length(stand_ad5); n25=length(stand5)
total<-n15+n25
sttrue5= mean(stand_ad5)-mean(stand5)
cnt= 0 #zero the counter
#Put both sets of observations in a single vector
vect= c(stand_ad5, stand5)
for (i in 1:nr5){
 d= sample(vect,n15+n25)
 sta_ad<-d[1:n15]
 a<-n15+1;sta<-d[a:total]</pre>
 st5[i] <-mean(sta_ad)-mean(sta)</pre>
 if(st5[i] > sttrue5)cnt=cnt+1 }
hist(st5)
abline(v=sttrue5, lwd=4, col="blue")
```

# Histogram of st5



```
#True value using the difference of the means
sttrue5
```

## [1] 0.622222

#p-value
cnt/nr5

## [1] 0.109

A large p-value (p "greater than" 0.05) indicates weak evidence against the null hypothesis so, you fail to reject the null hypothesis (Ho is not rejected).

Therefore, the additive is efficient.

### Wilcoxon Rank Sum Test

Both t-Test and Wilcoxon rank test can be used to compare the mean of two (2) samples. The difference is t-Test assumes the samples being tests is drawn from a normal distribution, while, Wilcoxon's rank sum test does not.

## Warning in wilcox.test.default(stand, stand\_ad, alternative = "g"): cannot
## compute exact p-value with ties

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: stand and stand_ad
## W = 47.5, p-value = 0.9222
## alternative hypothesis: true location shift is greater than 0
```

A large p-value (p "greater than" 0.05) indicates weak evidence against the null hypothesis so, you fail to reject the null hypothesis (Ho is not rejected).

Therefore, the additive is efficient.

#### Kolmogorov And Smirnov Test

Kolmogorov-Smirnov test is used to check whether 2 samples follow the same distribution.

A large p-value (p "greater than" 0.05) indicates weak evidence against the null hypothesis so, we fail to reject the null hypothesis (Ho is not rejected).

Therefore, the additive is efficient.

#### Fisher's F-Test

Fisher's F test can be used to check if two samples have same variance.

```
stand=c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
stand_ad = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8,
              3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
var.test(stand, stand_ad)
##
## F test to compare two variances
##
## data: stand and stand_ad
## F = 0.75312, num df = 8, denom df = 15, p-value = 0.706
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.2354441 3.0887222
## sample estimates:
## ratio of variances
##
           0.7531241
```

A large p-value (p "greater than" 0.05) indicates weak evidence against the null hypothesis so, we fail to reject the null hypothesis (Ho is not rejected).

Therefore, the additive is efficient.

# **Summary Statistics**

```
suppressMessages(suppressWarnings(library(pastecs)))
stand=c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
stand_ad = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8,
              3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
#Standard only
stat.desc(stand)
##
        nbr.val
                     nbr.null
                                     nbr.na
                                                      min
                                                                    max
##
      9.0000000
                    0.0000000
                                  0.0000000
                                                1.8000000
                                                              5.3000000
##
                                     median
                                                                SE.mean
          range
                          sum
                                                     mean
                                                              0.3639207
##
      3.5000000
                   28.6000000
                                  3.3000000
                                                3.1777778
## CI.mean.0.95
                                    std.dev
                                                 coef.var
                          var
      0.8392026
##
                    1.1919444
                                  1.0917621
                                                0.3435615
#Standard + Additive
stat.desc(stand_ad)
##
        nbr.val
                     nbr.null
                                     nbr.na
                                                      min
                                                                    max
     16.0000000
                    0.0000000
                                  0.0000000
                                                1.8000000
                                                              6.3000000
##
##
                                                                SE.mean
          range
                          \operatorname{\mathtt{sum}}
                                     median
                                                     mean
##
      4.5000000
                   60.8000000
                                  3.6500000
                                                3.8000000
                                                              0.3145102
## CI.mean.0.95
                                    std.dev
                                                 coef.var
                          var
      0.6703626
                    1.5826667
                                  1.2580408
                                                0.3310634
```

# Conclusion

After exploring several statistics in order to perform the permutation test, we can conclude that the increments of weight recorded in an experiment, where a new additive has been added to a standard compound feed, is efficient.