DVM Permutation Test - Work 2

Jeremy Williams and Spyridoula Chrysikopoulou-Soldatou

October 18, 2017

Prediction of Multiple Regression Model

```
wc <- c(16,20,25,27,32,48,48)
t <- c(75,83,85,85,92,97,99)
tmg <- c(1.85,1.25,1.5,1.75,1.15,1.75,1.6)

# Fit a linear model and run a summary of its results.
mod1<-lm(wc ~ t + tmg)
summary(mod1)
```

```
##
## Call:
## lm(formula = wc ~ t + tmg)
##
## Residuals:
##       1       2       3       4       5       6       7
##  1.0441  0.4642 -0.6935 -1.8264  0.1061  1.0252 -0.1197
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -121.65500    6.54035 -18.601 4.92e-05 ***
## t              1.51236    0.06077  24.886 1.55e-05 ***
## tmg           12.53168    1.93302   6.483  0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.245 on 4 degrees of freedom
## Multiple R-squared:  0.9937, Adjusted R-squared:  0.9905
## F-statistic: 313.2 on 2 and 4 DF,  p-value: 4.027e-05
```

```
# Predicted values
fitted(mod1)
```

```
##        1        2        3        4        5        6        7
## 14.95591 19.53582 25.69347 28.82639 31.89393 46.97476 48.11973
```

```
# Model coefficients
coefficients(mod1)
```

```
## (Intercept)           t         tmg
## -121.654997    1.512364   12.531681
```

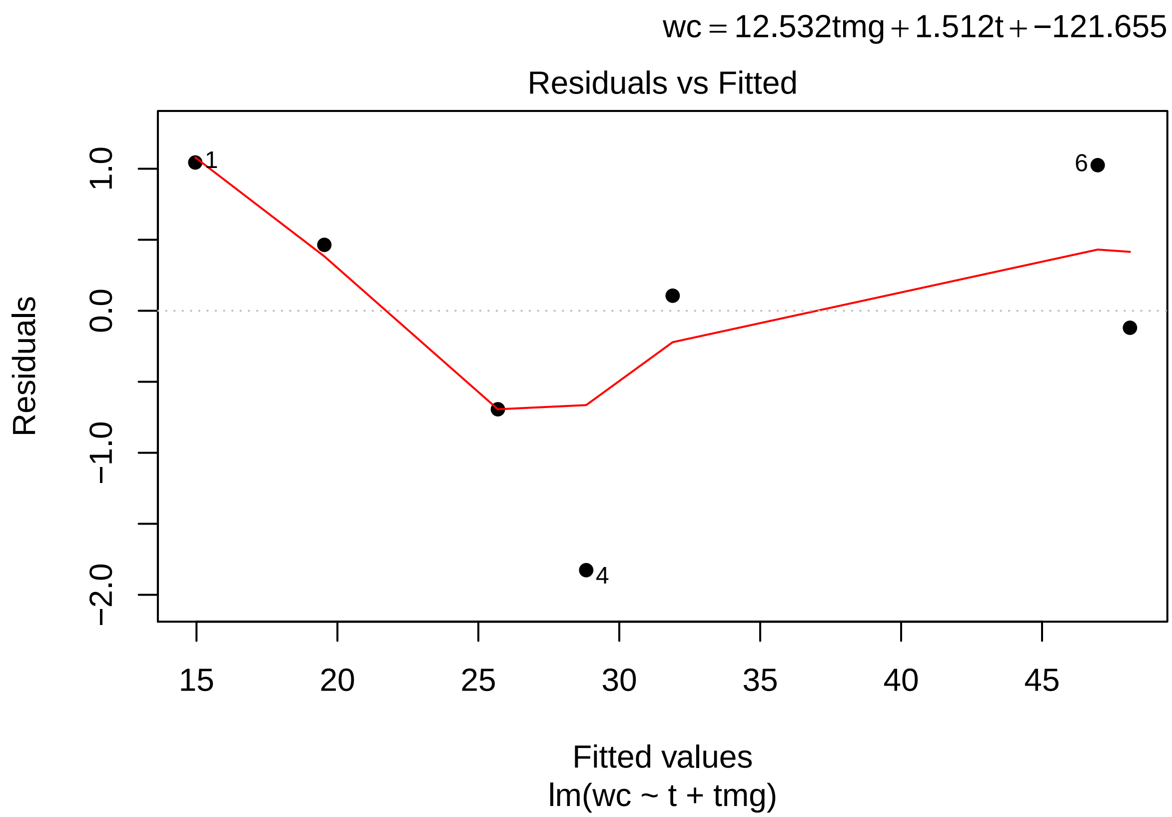
```
# CIs for model parameters
confint(mod1, level=0.95)
```

```
##                   2.5 %      97.5 %
## (Intercept) -139.813914 -103.496080
## t              1.343637    1.681091
## tmg            7.164757   17.898605
```

```r
# Plot model residuals.
plot(mod1, pch=16, which=1)

# Extract coefficients
mod1_coef <- round(coef(mod1), 3)

# Display equation
mtext(bquote(wc == .(mod1_coef[3])* tmg +.(mod1_coef[2])* t + .(mod1_coef[1])),
      adj=1, padj=-2)
```



Significance of the Variables T and TMG

We begin by testing whether a set of independent variables has no partial effect on the dependent variable, "Y".

Our model is:

wc = B0 + B1t + B2tmp + e

Null Hypothesis: The initial assumption is that there is no relation, which is expressed as:

Ho: B1 = B2 = 0.

Alternative Hypothesis: At least one of the independent variables IS useful in explaining/predicting Y, expressed as:

H1: At least one Bi is "not equal to" 0.

Exact Permuatition Test

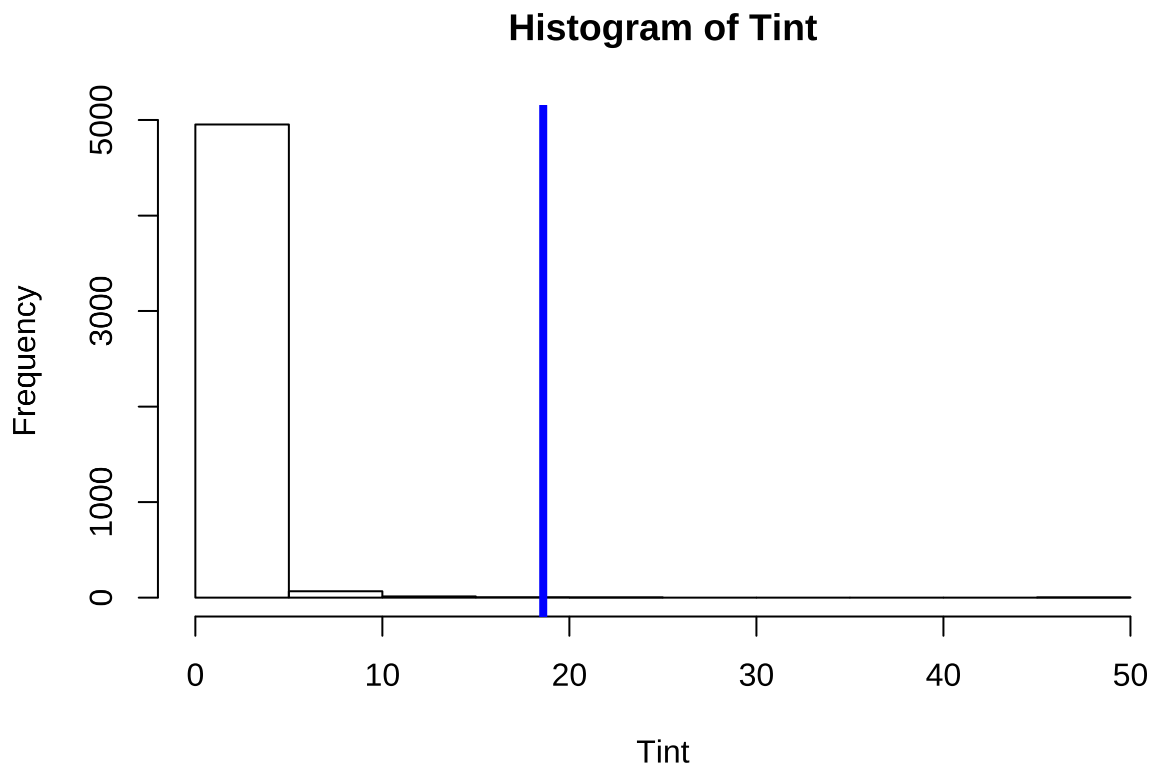
```
#install.packages("combinat")
suppressMessages(suppressWarnings(library("combinat")))
wc1 <- c(16,20,25,27,32,48,48)
t1 <- c(75,83,85,85,92,97,99)
tmg1 <- c(1.85,1.25,1.5,1.75,1.15,1.75,1.6)
mod2<-glm(wc1 ~ t1 + tmg1, family = gaussian)
a<-summary(mod2)
Tinttrue<-abs(a$coefficients[1,3])
Tttrue<-abs(a$coefficients[2,3])
Ttmgtrue<-abs(a$coefficients[3,3])

#number of rearrangements to be examined
n<-length(wc1)
nr<-fact(n)
nr
```

```
## [1] 5040
```

```
Tint=numeric(nr);Tt=numeric(nr);Ttmg=numeric(nr)
newy<-permn(wc1)
for (i in 1:nr){
  mod2<-glm(newy[[i]] ~ t1 + tmg1,family = gaussian)
  a<-summary(mod2)
  Tint[i]<-abs(a$coefficients[1,3])
  Tt[i]<-abs(a$coefficients[2,3])
  Ttmg[i]<-abs(a$coefficients[3,3])}

par(mfrow=c(1,1))
hist(Tint)
abline(v=Tinttrue, lwd=4, col="blue")
```



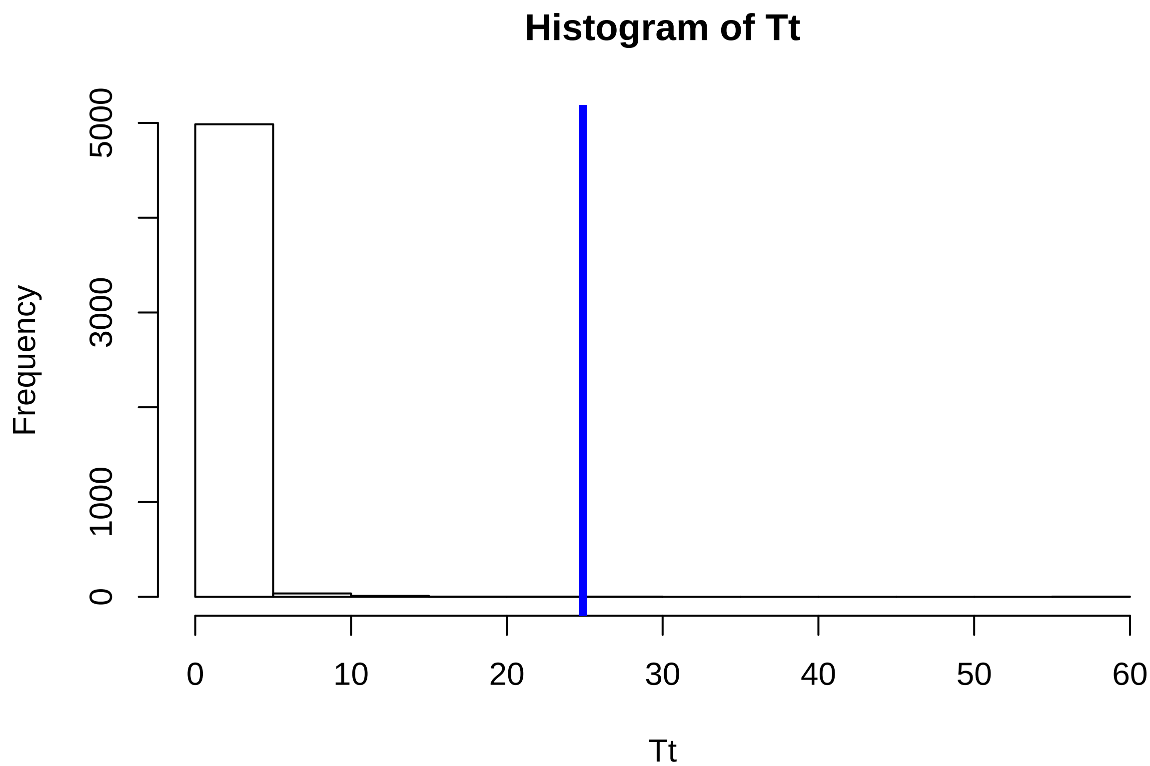

```
#True t-value of intercept "B0"
Tinttrue
```

```
## [1] 18.60069
```

```
#P-Value of intercept "B0"
length(Tint[Tint>= Tinttrue])/nr
```

```
## [1] 0.001190476
```

```
par(mfrow=c(1,1))
hist(Tt)
abline(v=Tttrue, lwd=4, col="blue")
```



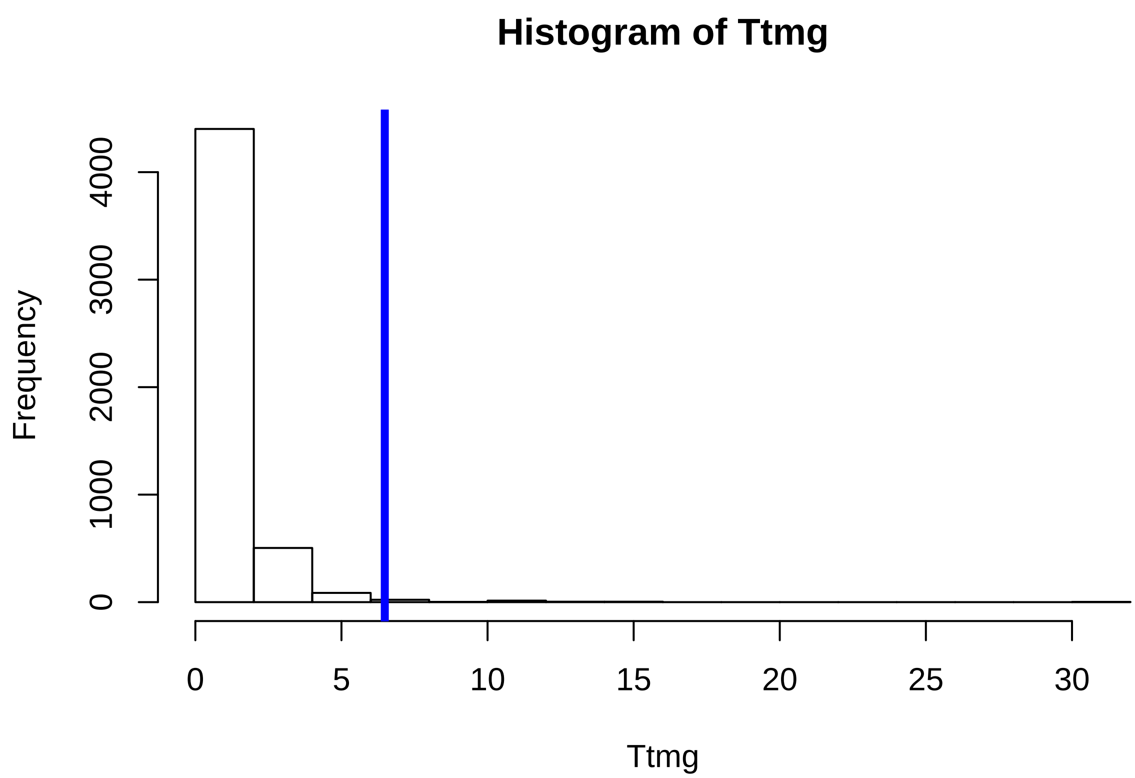

```
#True t-Value of Time (T) "B1"
Tttrue
```

```
## [1] 24.88626
```

```
#P-value of Time (T) "B1"
length(Tt[Tt>= Tttrue])/nr
```

```
## [1] 0.001190476
```

```
par(mfrow=c(1,1))
hist(Ttmg)
abline(v=Ttmgtrue, lwd=4, col="blue")
```




```
#True t-Value of Time moving the grass (TMG) "B2"
Ttmgtrue
```

```
## [1] 6.482954
```

```
#P-Value of Time moving the grass (TMG) "B2"
length(Ttmg[Ttmg>= Ttmgtrue])/nr
```

```
## [1] 0.009126984
```

All (intercept, T and TMG) small p-values (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).

Multiple Regression through the Origin

```
#install.packages("combinat")
suppressMessages(suppressWarnings(library("combinat")))
#install.packages("ape")
suppressMessages(suppressWarnings(library("ape")))
wc1 <- c(16,20,25,27,32,48,48)
t1 <- c(75,83,85,85,92,97,99)
tmg1 <- c(1.85,1.25,1.5,1.75,1.15,1.75,1.6)
a5<-data.frame(wc1,t1,tmg1)

#number of rearrangements to be examined
nr<-fact(length(wc1))
```

```r
nr
```

```
## [1] 5040
```

```r
#Permutation method using number of rearrangements
lmorigin(wc1 ~ t1 + tmg1, data =a5, nperm=nr)
```

```
## Regression through the origin
## Permutation method = raw data
## Computation time = 25.030000  sec
```

```
##
## Regression through the origin
##
## Call:
## lmorigin(formula = wc1 ~ t1 + tmg1, data = a5, nperm = nr)
##
## Coefficients and parametric test results
##
##      Coefficient Std_error t-value Pr(>|t|)
## t1       0.50764   0.23298  2.1789  0.08123 .
## tmg1    -8.50684  13.11494 -0.6486  0.54518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Two-tailed tests of regression coefficients
##
##      Coefficient p-param  p-perm
## t1       0.50764  0.0812 0.08669 .
## tmg1    -8.50684  0.5452 0.54930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## One-tailed tests of regression coefficients:
## test in the direction of the sign of the coefficient
##
##      Coefficient p-param  p-perm
## t1       0.50764  0.0406 0.04225 *
## tmg1    -8.50684  0.2726 0.28248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.41518 on 5 degrees of freedom
## Multiple R-square: 0.9290265   Adjusted R-square: 0.9006371
##
## F-statistic: 32.7244 on 2 and 5 DF:
##    parametric p-value   : 0.001341964
##    permutational p-value: 0.001586987
## after 5040 permutations of raw data
##
```

Both (parametric and permutational) small p-values (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).

LM and GLMs Comparison

```r
#install.packages("texreg")
suppressMessages(suppressWarnings(library("texreg")))
wc1 <- c(16,20,25,27,32,48,48)
t1 <- c(75,83,85,85,92,97,99)
tmg1 <- c(1.85,1.25,1.5,1.75,1.15,1.75,1.6)

# Estimate with OLS (Model 1):
reg1<-lm(wc1 ~ t1 + tmg1)

# Estimate with GLS (Model 2):
reg2<-glm(wc1 ~ t1 + tmg1, family = gaussian)

# Compare:
screenreg(l = list(reg1,reg2))
```

```
##
## ========================================
##                 Model 1      Model 2
## ----------------------------------------
## (Intercept)     -121.65 ***  -121.65 ***
##                   (6.54)       (6.54)
## t1                 1.51 ***     1.51 ***
##                   (0.06)       (0.06)
## tmg1              12.53 **     12.53 **
##                   (1.93)       (1.93)
## ----------------------------------------
## R^2                0.99
## Adj. R^2           0.99
## Num. obs.          7            7
## RMSE               1.24
## AIC                            27.01
## BIC                            26.80
## Log Likelihood                 -9.51
## Deviance                        6.20
## ========================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

Both (LM and GLM) small p-values (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).

Conclusion

After testing the significance of the variables T and TMG, we can conclude that at least one of the variables "T" and "TMG"" are significant.