Differences-in-Differences (using R)
(v. 1.0)

Oscar Torres-Reyna
otorres@princeton.edu

Difference in differences (DID)
Estimation step-by-step

 # Getting sample data.

```
library(foreign)
mydata = read.dta("https://www.princeton.edu/~otorres/Panel101.dta")
```

 # Create a dummy variable to indicate the time when the treatment started. Lets assume that treatment started in 1994. In this case, years before 1994 will have a value of 0 and 1994+ a 1. If you already have this skip this step.

```
mydata$time = ifelse(mydata$year >= 1994, 1, 0)
```

 # Create a dummy variable to identify the group exposed to the treatment. In this example lets assumed that countries with code 5,6, and 7 were treated (=1). Countries 1-4 were not treated (=0). If you already have this skip this step.

```
mydata$treated = ifelse(mydata$country == "E" |
                        mydata$country == "F" |
                        mydata$country == "G", 1, 0)
```

 # Create an interaction between time and treated. We will call this interaction 'did'.

```
mydata$did = mydata$time * mydata$treated
```

Difference in differences (DID)
Estimation step-by-step

 # Estimating the DID estimator

```
didreg = lm(y ~ treated + time + did, data = mydata)
summary(didreg)
```

```
Call:
lm(formula = y ~ treated + time + did, data = mydata)

Residuals:
        Min         1Q     Median         3Q        Max
 -9.768e+09 -1.623e+09  1.167e+08  1.393e+09  6.807e+09

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.581e+08  7.382e+08    0.485   0.6292
treated      1.776e+09  1.128e+09    1.575   0.1200
time         2.289e+09  9.530e+08    2.402   0.0191 *
did         -2.520e+09  1.456e+09   -1.731   0.0882 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.953e+09 on 66 degrees of freedom
Multiple R-squared:  0.08273, Adjusted R-squared:  0.04104
F-statistic: 1.984 on 3 and 66 DF,  p-value: 0.1249
```

**# The coefficient for 'did' is the differences-in-differences
estimator. The effect is significant at 10% with the treatment having
a negative effect.**

Difference in differences (DID)
Estimation step-by-step

 # Estimating the DID estimator (using the multiplication method, no need to generate the interaction)

```
didreg1 = lm(y ~ treated*time, data = mydata)
summary(didreg1)
```

```
Call:
lm(formula = y ~ treated * time, data = mydata)

Residuals:
        Min         1Q      Median         3Q         Max
-9.768e+09  -1.623e+09   1.167e+08   1.393e+09   6.807e+09

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.581e+08  7.382e+08    0.485   0.6292
treated       1.776e+09  1.128e+09    1.575   0.1200
time          2.289e+09  9.530e+08    2.402   0.0191 *
treated:time -2.520e+09  1.456e+09   -1.731   0.0882 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.953e+09 on 66 degrees of freedom
Multiple R-squared:  0.08273, Adjusted R-squared:  0.04104
F-statistic: 1.984 on 3 and 66 DF,  p-value: 0.1249
```

 # The coefficient for 'treated#time' is the differences-in-differences estimator ('did' in the previous example). The effect is significant at 10% with the treatment having a negative effect.

References

Introduction to econometrics, James H. Stock, Mark W. Watson. 2nd ed., Boston: Pearson Addison Wesley, 2007.

"Difference-in-Differences Estimation", Imbens/Wooldridge, Lecture Notes 10, summer 2007.

http://www.nber.org/WNE/lect_10_diffindiffs.pdf

"Lecture 3: Differences-in-Differences", Fabian Waldinger

http://www2.warwick.ac.uk/fac/soc/economics/staff/ffwaldinger/teaching/ec9a8/slides/lecture_3_-_did.pdf