BlackCloudModelConstruction

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

This notebook has been used to create a Linear Mixed Effect Model for the Black Cloud analysis.

```
# Package names
packages <- c("carData", "car", "Matrix", "lme4", "LMERConvenienceFunctions")

# Install packages not yet installed
installed_packages <- packages %in% rownames(installed.packages())
if (any(installed_packages == FALSE)) {
  install.packages(packages[!installed_packages])
}

# Packages loading
invisible(lapply(packages, library, character.only = TRUE))
```

```
## Warning: il pacchetto 'carData' è stato creato con R versione 4.2.3

## Warning: il pacchetto 'car' è stato creato con R versione 4.2.3

## Warning: il pacchetto 'Matrix' è stato creato con R versione 4.2.3

## Warning: il pacchetto 'lme4' è stato creato con R versione 4.2.3

## Warning: il pacchetto 'LMERConvenienceFunctions' è stato creato con R versione
## 4.2.3
```

Datasets load

```
# Reading input hofstede data.
data <- read.csv("./black_cloud_metrics_hofstede.csv", sep = ";", header = TRUE, stringsAsFactors=FALSE)
# Reading input trompenaars data.
dataT <- read.csv("./black_cloud_metrics_trompenaars.csv", sep = ";", header = TRUE, stringsAsFactors=FA
# Reading input globe data.
dataG <- read.csv("./black_cloud_metrics_globe.csv", sep = ";", header = TRUE, stringsAsFactors=FALSE)

# Excluding some columns from hofstede data
working_data <- na.omit(data)
# Excluding some columns from trompenaars data
working_dataT <- na.omit(dataT)
# Excluding some columns from globe data
working_dataG <- na.omit(dataG)
```

Linear Mixed Model using lmer function on all the variables for Hofstede

```
#----------------------------
#ALL THE VARIABLES

# Applying a Linear Mixed Model using the lmer function
black <- lmer(working_data$black~log(working_data$totalCommitters)+log(working_data$totalcommits)
              +working_data$projectAge+working_data$turnover+working_data$blauGender
              +working_data$tenureMedian+working_data$tenureDiversity+log(working_data$teamSize)
              +working_data$stCongruence+working_data$truckFactor+working_data$female
              +working_data$expertise+working_data$centrality+working_data$CV_1
              +working_data$CV_2+working_data$CV_3
              +working_data$CV_4+working_data$CV_5+working_data$CV_6
              +(1 | working_data$window_idx ), REML=FALSE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# Remove outlier
#romr.fnc(black, working_data, trim = 2.5)

# Applying vif <5
print(vif(black))
```

```
## log(working_data$totalCommitters)    log(working_data$totalcommits)
##                          3.315400                          3.173282
##           working_data$projectAge            working_data$turnover
##                          1.441401                          1.421853
##           working_data$blauGender        working_data$tenureMedian
##                          2.740519                          1.125818
##      working_data$tenureDiversity       log(working_data$teamSize)
##                          1.069988                          2.684843
##         working_data$stCongruence         working_data$truckFactor
##                          1.066486                          1.086931
##               working_data$female           working_data$expertise
##                          1.142353                          1.099845
##           working_data$centrality                working_data$CV_1
##                          1.177195                          4.994493
##                 working_data$CV_2                working_data$CV_3
##                          6.339214                          3.294542
##                 working_data$CV_4                working_data$CV_5
##                          8.515881                          4.333004
##                 working_data$CV_6
##                          7.708871
```

```
# Applying a Linear Mixed Model using the lmer function, after vif - NO REMOVAL

# print result
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$black ~ log(working_data$totalCommitters) + log(working_data$totalcommits) +
##     working_data$projectAge + working_data$turnover + working_data$blauGender +
##     working_data$tenureMedian + working_data$tenureDiversity +
```

```
##     log(working_data$teamSize) + working_data$stCongruence +
##     working_data$truckFactor + working_data$female + working_data$expertise +
##     working_data$centrality + working_data$CV_1 + working_data$CV_2 +
##     working_data$CV_3 + working_data$CV_4 + working_data$CV_5 +
##     working_data$CV_6 + (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    870.2    941.9   -413.1    826.2      170
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.88030 -0.87700 -0.03409  0.84003  1.99275
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.000    0.00
##  Residual                             4.328    2.08
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                         5.049027   2.067613   2.442
## log(working_data$totalCommitters) -0.114491   0.233109  -0.491
## log(working_data$totalcommits)     0.023803   0.177852   0.134
## working_data$projectAge           -0.020489   0.037219  -0.551
## working_data$turnover              0.683087   0.763294   0.895
## working_data$blauGender           -8.573778   2.704472  -3.170
## working_data$tenureMedian          0.063706   0.088142   0.723
## working_data$tenureDiversity       0.015988   0.061633   0.259
## log(working_data$teamSize)         0.269396   0.232283   1.160
## working_data$stCongruence         -0.212388   0.456211  -0.466
## working_data$truckFactor           0.009162   0.114015   0.080
## working_data$female                0.037219   0.028296   1.315
## working_data$expertise            -0.207151   0.499771  -0.414
## working_data$centrality           -0.590572   0.365201  -1.617
## working_data$CV_1                 -0.934249   3.300639  -0.283
## working_data$CV_2                  0.896752   4.017696   0.223
## working_data$CV_3                  1.367557   2.923935   0.468
## working_data$CV_4                 -1.104878   5.257865  -0.210
## working_data$CV_5                 -1.257845   2.563971  -0.491
## working_data$CV_6                  2.155360   4.461876   0.483
##
## Correlation matrix not shown by default, as p = 20 > 12.
## Use print(summary(black), correlation=TRUE)  or
##     vcov(summary(black))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```r
# Applying anova
Anova(black)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$black
```

```
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.2412  1   0.623322
## log(working_data$totalcommits)     0.0179  1   0.893531
## working_data$projectAge            0.3031  1   0.581968
## working_data$turnover              0.8009  1   0.370830
## working_data$blauGender           10.0503  1   0.001523 **
## working_data$tenureMedian          0.5224  1   0.469824
## working_data$tenureDiversity       0.0673  1   0.795321
## log(working_data$teamSize)         1.3451  1   0.246140
## working_data$stCongruence          0.2167  1   0.641540
## working_data$truckFactor           0.0065  1   0.935952
## working_data$female                1.7301  1   0.188395
## working_data$expertise             0.1718  1   0.678514
## working_data$centrality            2.6151  1   0.105854
## working_data$CV_1                  0.0801  1   0.777138
## working_data$CV_2                  0.0498  1   0.823379
## working_data$CV_3                  0.2188  1   0.639991
## working_data$CV_4                  0.0442  1   0.833560
## working_data$CV_5                  0.2407  1   0.623720
## working_data$CV_6                  0.2333  1   0.629052
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Save in a txt file
sink("hofstede/output_black_hofstede_all_variables.txt")
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$black ~ log(working_data$totalCommitters) + log(working_data$totalcommits) +
##     working_data$projectAge + working_data$turnover + working_data$blauGender +
##     working_data$tenureMedian + working_data$tenureDiversity +
##     log(working_data$teamSize) + working_data$stCongruence +
##     working_data$truckFactor + working_data$female + working_data$expertise +
##     working_data$centrality + working_data$CV_1 + working_data$CV_2 +
##     working_data$CV_3 + working_data$CV_4 + working_data$CV_5 +
##     working_data$CV_6 + (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    870.2    941.9   -413.1    826.2      170
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.88030 -0.87700 -0.03409  0.84003  1.99275
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.000    0.00
##  Residual                             4.328    2.08
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                       5.049027   2.067613   2.442
## log(working_data$totalCommitters) -0.114491   0.233109  -0.491
```

```
## log(working_data$totalcommits)     0.023803   0.177852   0.134
## working_data$projectAge           -0.020489   0.037219  -0.551
## working_data$turnover              0.683087   0.763294   0.895
## working_data$blauGender           -8.573778   2.704472  -3.170
## working_data$tenureMedian          0.063706   0.088142   0.723
## working_data$tenureDiversity       0.015988   0.061633   0.259
## log(working_data$teamSize)         0.269396   0.232283   1.160
## working_data$stCongruence         -0.212388   0.456211  -0.466
## working_data$truckFactor           0.009162   0.114015   0.080
## working_data$female                0.037219   0.028296   1.315
## working_data$expertise            -0.207151   0.499771  -0.414
## working_data$centrality           -0.590572   0.365201  -1.617
## working_data$CV_1                 -0.934249   3.300639  -0.283
## working_data$CV_2                  0.896752   4.017696   0.223
## working_data$CV_3                  1.367557   2.923935   0.468
## working_data$CV_4                 -1.104878   5.257865  -0.210
## working_data$CV_5                 -1.257845   2.563971  -0.491
## working_data$CV_6                  2.155360   4.461876   0.483

##
## Correlation matrix not shown by default, as p = 20 > 12.
## Use print(summary(black), correlation=TRUE)  or
##     vcov(summary(black))        if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
Anova(black)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$black
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.2412  1   0.623322
## log(working_data$totalcommits)     0.0179  1   0.893531
## working_data$projectAge            0.3031  1   0.581968
## working_data$turnover              0.8009  1   0.370830
## working_data$blauGender           10.0503  1   0.001523 **
## working_data$tenureMedian          0.5224  1   0.469824
## working_data$tenureDiversity       0.0673  1   0.795321
## log(working_data$teamSize)         1.3451  1   0.246140
## working_data$stCongruence          0.2167  1   0.641540
## working_data$truckFactor           0.0065  1   0.935952
## working_data$female                1.7301  1   0.188395
## working_data$expertise             0.1718  1   0.678514
## working_data$centrality            2.6151  1   0.105854
## working_data$CV_1                  0.0801  1   0.777138
## working_data$CV_2                  0.0498  1   0.823379
## working_data$CV_3                  0.2188  1   0.639991
## working_data$CV_4                  0.0442  1   0.833560
## working_data$CV_5                  0.2407  1   0.623720
## working_data$CV_6                  0.2333  1   0.629052
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on all the variables for trompenaars

```
#----------------------------
#ALL THE VARIABLES

# Applying a Linear Mixed Model using the lmer function
blackT <- lmer(working_dataT$black~log(working_data$totalCommitters)+log(working_dataT$totalcommits)
             +working_data$projectAge+working_dataT$turnover+working_dataT$blauGender
             +working_dataT$tenureMedian+working_dataT$tenureDiversity+log(working_dataT$teamSize)
             +working_dataT$stCongruence+working_dataT$truckFactor+working_dataT$female
             +working_dataT$expertise+working_dataT$centrality+working_dataT$CV_1
             +working_dataT$CV_2+working_dataT$CV_3
             +working_dataT$CV_4+working_dataT$CV_5+working_dataT$CV_6
             +working_dataT$CV_7+working_dataT$CV_8
             +(1 | working_dataT$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(blackT, working_dataT, trim = 2.5)

# Applying vif <5
print(vif(blackT))
```

```
## log(working_data$totalCommitters)   log(working_dataT$totalcommits)
##                          3.044192                          3.081570
##           working_data$projectAge           working_dataT$turnover
##                          1.462583                          1.488128
##          working_dataT$blauGender       working_dataT$tenureMedian
##                          2.580912                          1.099754
##     working_dataT$tenureDiversity      log(working_dataT$teamSize)
##                          1.087902                          2.313761
##        working_dataT$stCongruence        working_dataT$truckFactor
##                          1.059188                          1.087671
##              working_dataT$female          working_dataT$expertise
##                          1.151478                          1.149538
##          working_dataT$centrality               working_dataT$CV_1
##                          1.189510                         15.114620
##                working_dataT$CV_2               working_dataT$CV_3
##                         11.620736                          6.366079
##                working_dataT$CV_4               working_dataT$CV_5
##                         11.210657                         21.568299
##                working_dataT$CV_6               working_dataT$CV_7
##                          3.015897                          6.744459
##                working_dataT$CV_8
##                          4.426151
```

```
# Applying a Linear Mixed Model using the lmer function, after vif - NO REMOVAL

# print result
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$black ~ log(working_data$totalCommitters) + log(working_dataT$totalcommits) +
##     working_data$projectAge + working_dataT$turnover + working_dataT$blauGender +
##     working_dataT$tenureMedian + working_dataT$tenureDiversity +
##     log(working_dataT$teamSize) + working_dataT$stCongruence +
##     working_dataT$truckFactor + working_dataT$female + working_dataT$expertise +
##     working_dataT$centrality + working_dataT$CV_1 + working_dataT$CV_2 +
##     working_dataT$CV_3 + working_dataT$CV_4 + working_dataT$CV_5 +
##     working_dataT$CV_6 + working_dataT$CV_7 + working_dataT$CV_8 +
##     (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    860.1    938.3   -406.0    812.1      168
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.10641 -0.89713 -0.07154  0.82073  2.18141
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.03831  0.1957
##  Residual                             3.98472  1.9962
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                      5.121611   1.973793   2.595
## log(working_data$totalCommitters) -0.006787   0.217109  -0.031
## log(working_dataT$totalcommits)  -0.045591   0.170887  -0.267
## working_data$projectAge          -0.020487   0.036357  -0.563
## working_dataT$turnover            0.022607   0.759723   0.030
## working_dataT$blauGender         -8.487706   2.538158  -3.344
## working_dataT$tenureMedian        0.067029   0.083761   0.800
## working_dataT$tenureDiversity     0.014801   0.059844   0.247
## log(working_dataT$teamSize)       0.419001   0.209524   2.000
## working_dataT$stCongruence       -0.158755   0.436897  -0.363
## working_dataT$truckFactor         0.038243   0.109774   0.348
## working_dataT$female              0.036555   0.027372   1.335
## working_dataT$expertise          -0.170993   0.491624  -0.348
## working_dataT$centrality         -0.628640   0.354252  -1.775
## working_dataT$CV_1                4.319710   5.350418   0.807
## working_dataT$CV_2                5.404610   3.846560   1.405
## working_dataT$CV_3                2.886457   3.185623   0.906
## working_dataT$CV_4                7.408224   4.214292   1.758
## working_dataT$CV_5               -7.396776   5.175349  -1.429
## working_dataT$CV_6               -3.473067   1.379752  -2.517
## working_dataT$CV_7              -10.465609   3.530592  -2.964
## working_dataT$CV_8                1.316173   2.602450   0.506
##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
```

```r
# Applying anova
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$black
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.0010  1  0.9750599
## log(working_dataT$totalcommits)    0.0712  1  0.7896319
## working_data$projectAge            0.3175  1  0.5730979
## working_dataT$turnover             0.0009  1  0.9762609
## working_dataT$blauGender          11.1826  1  0.0008257 ***
## working_dataT$tenureMedian         0.6404  1  0.4235717
## working_dataT$tenureDiversity      0.0612  1  0.8046605
## log(working_dataT$teamSize)        3.9991  1  0.0455238 *
## working_dataT$stCongruence         0.1320  1  0.7163289
## working_dataT$truckFactor          0.1214  1  0.7275511
## working_dataT$female               1.7835  1  0.1817191
## working_dataT$expertise            0.1210  1  0.7279803
## working_dataT$centrality           3.1490  1  0.0759715 .
## working_dataT$CV_1                 0.6518  1  0.4194595
## working_dataT$CV_2                 1.9742  1  0.1600064
## working_dataT$CV_3                 0.8210  1  0.3648889
## working_dataT$CV_4                 3.0901  1  0.0787678 .
## working_dataT$CV_5                 2.0427  1  0.1529375
## working_dataT$CV_6                 6.3361  1  0.0118303 *
## working_dataT$CV_7                 8.7869  1  0.0030341 **
## working_dataT$CV_8                 0.2558  1  0.6130364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Save in a txt file
sink("trompe/output_black_trompenaars_all_variables.txt")
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$black ~ log(working_data$totalCommitters) + log(working_dataT$totalcommits) +
##     working_data$projectAge + working_dataT$turnover + working_dataT$blauGender +
##     working_dataT$tenureMedian + working_dataT$tenureDiversity +
##     log(working_dataT$teamSize) + working_dataT$stCongruence +
##     working_dataT$truckFactor + working_dataT$female + working_dataT$expertise +
##     working_dataT$centrality + working_dataT$CV_1 + working_dataT$CV_2 +
##     working_dataT$CV_3 + working_dataT$CV_4 + working_dataT$CV_5 +
##     working_dataT$CV_6 + working_dataT$CV_7 + working_dataT$CV_8 +
##     (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    860.1    938.3   -406.0    812.1      168
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.10641 -0.89713 -0.07154  0.82073  2.18141
##
## Random effects:
```

```
##  Groups                   Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.03831  0.1957
##  Residual                             3.98472  1.9962
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                         5.121611   1.973793   2.595
## log(working_data$totalCommitters)  -0.006787   0.217109  -0.031
## log(working_dataT$totalcommits)    -0.045591   0.170887  -0.267
## working_data$projectAge            -0.020487   0.036357  -0.563
## working_dataT$turnover              0.022607   0.759723   0.030
## working_dataT$blauGender           -8.487706   2.538158  -3.344
## working_dataT$tenureMedian          0.067029   0.083761   0.800
## working_dataT$tenureDiversity       0.014801   0.059844   0.247
## log(working_dataT$teamSize)         0.419001   0.209524   2.000
## working_dataT$stCongruence         -0.158755   0.436897  -0.363
## working_dataT$truckFactor           0.038243   0.109774   0.348
## working_dataT$female                0.036555   0.027372   1.335
## working_dataT$expertise            -0.170993   0.491624  -0.348
## working_dataT$centrality           -0.628640   0.354252  -1.775
## working_dataT$CV_1                  4.319710   5.350418   0.807
## working_dataT$CV_2                  5.404610   3.846560   1.405
## working_dataT$CV_3                  2.886457   3.185623   0.906
## working_dataT$CV_4                  7.408224   4.214292   1.758
## working_dataT$CV_5                 -7.396776   5.175349  -1.429
## working_dataT$CV_6                 -3.473067   1.379752  -2.517
## working_dataT$CV_7                -10.465609   3.530592  -2.964
## working_dataT$CV_8                  1.316173   2.602450   0.506
##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
```

```
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$black
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.0010  1  0.9750599
## log(working_dataT$totalcommits)    0.0712  1  0.7896319
## working_data$projectAge            0.3175  1  0.5730979
## working_dataT$turnover             0.0009  1  0.9762609
## working_dataT$blauGender          11.1826  1  0.0008257 ***
## working_dataT$tenureMedian         0.6404  1  0.4235717
## working_dataT$tenureDiversity      0.0612  1  0.8046605
## log(working_dataT$teamSize)        3.9991  1  0.0455238 *
## working_dataT$stCongruence         0.1320  1  0.7163289
## working_dataT$truckFactor          0.1214  1  0.7275511
## working_dataT$female               1.7835  1  0.1817191
## working_dataT$expertise            0.1210  1  0.7279803
## working_dataT$centrality           3.1490  1  0.0759715 .
## working_dataT$CV_1                 0.6518  1  0.4194595
```

```
## working_dataT$CV_2                 1.9742  1  0.1600064
## working_dataT$CV_3                 0.8210  1  0.3648889
## working_dataT$CV_4                 3.0901  1  0.0787678 .
## working_dataT$CV_5                 2.0427  1  0.1529375
## working_dataT$CV_6                 6.3361  1  0.0118303 *
## working_dataT$CV_7                 8.7869  1  0.0030341 **
## working_dataT$CV_8                 0.2558  1  0.6130364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on all the variables for Globe

```
#----------------------------
#ALL THE VARIABLES

# Applying a Linear Mixed Model using the lmer function
blackG <- lmer(working_dataG$black~log(working_dataG$totalCommitters)+log(working_dataG$totalcommits)
            +working_dataG$projectAge+working_dataG$turnover+working_dataG$blauGender
            +working_dataG$tenureMedian+working_dataG$tenureDiversity+log(working_dataG$teamSize)
            +working_dataG$stCongruence+working_dataG$truckFactor+working_dataG$female
            +working_dataG$expertise+working_dataG$centrality+working_dataG$CV_1
            +working_dataG$CV_2+working_dataG$CV_3
            +working_dataG$CV_4+working_dataG$CV_5+working_dataG$CV_6+working_dataG$CV_7
            +working_dataG$CV_8+working_dataG$CV_9
            +(1 | working_dataG$window_idx ), REML=FALSE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# Remove outlier
#romr.fnc(blackG, working_dataG, trim = 2.5)

# Applying vif <5
print(vif(blackG))
```

```
## log(working_dataG$totalCommitters)    log(working_dataG$totalcommits)
##                           3.366585                           3.468358
##           working_dataG$projectAge           working_dataG$turnover
##                           1.460067                           1.584891
##           working_dataG$blauGender        working_dataG$tenureMedian
##                           2.604123                           1.122257
##      working_dataG$tenureDiversity       log(working_dataG$teamSize)
##                           1.092801                           2.538959
##         working_dataG$stCongruence         working_dataG$truckFactor
##                           1.054311                           1.113453
##               working_dataG$female           working_dataG$expertise
##                           1.134056                           1.203001
##           working_dataG$centrality                working_dataG$CV_1
##                           1.173959                          18.888681
##                working_dataG$CV_2                working_dataG$CV_3
##                           7.516169                           5.884378
##                working_dataG$CV_4                working_dataG$CV_5
```

```
##                          10.432447                          5.294252
##                 working_dataG$CV_6                 working_dataG$CV_7
##                          15.184290                          6.141369
##                 working_dataG$CV_8                 working_dataG$CV_9
##                          17.375580                          8.349998
```

```
# Applying a Linear Mixed Model using the lmer function, after vif - NO REMOVAL

# print result
print(summary(blackG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$black ~ log(working_dataG$totalCommitters) + log(working_dataG$totalcommits) +
##     working_dataG$projectAge + working_dataG$turnover + working_dataG$blauGender +
##     working_dataG$tenureMedian + working_dataG$tenureDiversity +
##     log(working_dataG$teamSize) + working_dataG$stCongruence +
##     working_dataG$truckFactor + working_dataG$female + working_dataG$expertise +
##     working_dataG$centrality + working_dataG$CV_1 + working_dataG$CV_2 +
##     working_dataG$CV_3 + working_dataG$CV_4 + working_dataG$CV_5 +
##     working_dataG$CV_6 + working_dataG$CV_7 + working_dataG$CV_8 +
##     working_dataG$CV_9 + (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    874.9    956.4   -412.5    824.9      167
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.78640 -0.86922 -0.01937  0.81916  1.84091
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.0      0.000
##  Residual                             4.3      2.074
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                         5.767405   1.999907   2.884
## log(working_dataG$totalCommitters) -0.176431   0.234137  -0.754
## log(working_dataG$totalcommits)     0.029119   0.185331   0.157
## working_dataG$projectAge           -0.031341   0.037337  -0.839
## working_dataG$turnover              0.666161   0.803244   0.829
## working_dataG$blauGender           -8.783163   2.627727  -3.342
## working_dataG$tenureMedian          0.029774   0.087715   0.339
## working_dataG$tenureDiversity       0.004652   0.062083   0.075
## log(working_dataG$teamSize)         0.301574   0.225148   1.339
## working_dataG$stCongruence         -0.253891   0.452122  -0.562
## working_dataG$truckFactor          -0.020100   0.115022  -0.175
## working_dataG$female                0.031106   0.028101   1.107
## working_dataG$expertise            -0.239756   0.520981  -0.460
## working_dataG$centrality           -0.557379   0.363511  -1.533
## working_dataG$CV_1                  0.739376  18.468735   0.040
## working_dataG$CV_2                 -6.445341  14.032364  -0.459
## working_dataG$CV_3                -19.779981  17.238554  -1.147
```

```
## working_dataG$CV_4                  12.427726  19.099964   0.651
## working_dataG$CV_5                  -1.812215  14.483755  -0.125
## working_dataG$CV_6                   3.928708  29.068627   0.135
## working_dataG$CV_7                   7.288826   9.718622   0.750
## working_dataG$CV_8                   6.368313  27.350137   0.233
## working_dataG$CV_9                 -11.317567  20.410769  -0.554

##
## Correlation matrix not shown by default, as p = 23 > 12.
## Use print(summary(blackG), correlation=TRUE)  or
##     vcov(summary(blackG))        if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
# Applying anova
Anova(blackG)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$black
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.5678  1  0.4511266
## log(working_dataG$totalcommits)     0.0247  1  0.8751510
## working_dataG$projectAge            0.7046  1  0.4012447
## working_dataG$turnover              0.6878  1  0.4069127
## working_dataG$blauGender           11.1723  1  0.0008303 ***
## working_dataG$tenureMedian          0.1152  1  0.7342760
## working_dataG$tenureDiversity       0.0056  1  0.9402652
## log(working_dataG$teamSize)         1.7941  1  0.1804256
## working_dataG$stCongruence          0.3153  1  0.5744204
## working_dataG$truckFactor           0.0305  1  0.8612791
## working_dataG$female                1.2253  1  0.2683289
## working_dataG$expertise             0.2118  1  0.6453720
## working_dataG$centrality            2.3511  1  0.1251969
## working_dataG$CV_1                  0.0016  1  0.9680661
## working_dataG$CV_2                  0.2110  1  0.6460046
## working_dataG$CV_3                  1.3166  1  0.2512052
## working_dataG$CV_4                  0.4234  1  0.5152611
## working_dataG$CV_5                  0.0157  1  0.9004281
## working_dataG$CV_6                  0.0183  1  0.8924910
## working_dataG$CV_7                  0.5625  1  0.4532634
## working_dataG$CV_8                  0.0542  1  0.8158826
## working_dataG$CV_9                  0.3075  1  0.5792436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Save in a txt file
sink("globe/output_black_globe_all_variables.txt")
print(summary(blackG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$black ~ log(working_dataG$totalCommitters) + log(working_dataG$totalcommits) +
##     working_dataG$projectAge + working_dataG$turnover + working_dataG$blauGender +
##     working_dataG$tenureMedian + working_dataG$tenureDiversity +
##     log(working_dataG$teamSize) + working_dataG$stCongruence +
```

```
##     working_dataG$truckFactor + working_dataG$female + working_dataG$expertise +
##     working_dataG$centrality + working_dataG$CV_1 + working_dataG$CV_2 +
##     working_dataG$CV_3 + working_dataG$CV_4 + working_dataG$CV_5 +
##     working_dataG$CV_6 + working_dataG$CV_7 + working_dataG$CV_8 +
##     working_dataG$CV_9 + (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    874.9    956.4   -412.5    824.9      167
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.78640 -0.86922 -0.01937  0.81916  1.84091
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.0      0.000
##  Residual                             4.3      2.074
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                        5.767405   1.999907   2.884
## log(working_dataG$totalCommitters) -0.176431   0.234137  -0.754
## log(working_dataG$totalcommits)     0.029119   0.185331   0.157
## working_dataG$projectAge           -0.031341   0.037337  -0.839
## working_dataG$turnover              0.666161   0.803244   0.829
## working_dataG$blauGender           -8.783163   2.627727  -3.342
## working_dataG$tenureMedian          0.029774   0.087715   0.339
## working_dataG$tenureDiversity       0.004652   0.062083   0.075
## log(working_dataG$teamSize)         0.301574   0.225148   1.339
## working_dataG$stCongruence         -0.253891   0.452122  -0.562
## working_dataG$truckFactor          -0.020100   0.115022  -0.175
## working_dataG$female                0.031106   0.028101   1.107
## working_dataG$expertise            -0.239756   0.520981  -0.460
## working_dataG$centrality           -0.557379   0.363511  -1.533
## working_dataG$CV_1                  0.739376  18.468735   0.040
## working_dataG$CV_2                 -6.445341  14.032364  -0.459
## working_dataG$CV_3                -19.779981  17.238554  -1.147
## working_dataG$CV_4                 12.427726  19.099964   0.651
## working_dataG$CV_5                 -1.812215  14.483755  -0.125
## working_dataG$CV_6                  3.928708  29.068627   0.135
## working_dataG$CV_7                  7.288826   9.718622   0.750
## working_dataG$CV_8                  6.368313  27.350137   0.233
## working_dataG$CV_9                -11.317567  20.410769  -0.554
##
## Correlation matrix not shown by default, as p = 23 > 12.
## Use print(summary(blackG), correlation=TRUE)  or
##     vcov(summary(blackG))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
Anova(blackG)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
```

```
##
## Response: working_dataG$black
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.5678  1  0.4511266
## log(working_dataG$totalcommits)     0.0247  1  0.8751510
## working_dataG$projectAge            0.7046  1  0.4012447
## working_dataG$turnover              0.6878  1  0.4069127
## working_dataG$blauGender           11.1723  1  0.0008303 ***
## working_dataG$tenureMedian          0.1152  1  0.7342760
## working_dataG$tenureDiversity       0.0056  1  0.9402652
## log(working_dataG$teamSize)         1.7941  1  0.1804256
## working_dataG$stCongruence          0.3153  1  0.5744204
## working_dataG$truckFactor           0.0305  1  0.8612791
## working_dataG$female                1.2253  1  0.2683289
## working_dataG$expertise             0.2118  1  0.6453720
## working_dataG$centrality            2.3511  1  0.1251969
## working_dataG$CV_1                  0.0016  1  0.9680661
## working_dataG$CV_2                  0.2110  1  0.6460046
## working_dataG$CV_3                  1.3166  1  0.2512052
## working_dataG$CV_4                  0.4234  1  0.5152611
## working_dataG$CV_5                  0.0157  1  0.9004281
## working_dataG$CV_6                  0.0183  1  0.8924910
## working_dataG$CV_7                  0.5625  1  0.4532634
## working_dataG$CV_8                  0.0542  1  0.8158826
## working_dataG$CV_9                  0.3075  1  0.5792436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on all the confounding variables hofstede

```
#-----------------------------
#ALL THE CONFOUNDING VARIABLES

# Applying a Linear Mixed Model using the lmer function
black <- lmer(working_data$black~log(working_data$totalCommitters)+log(working_data$totalcommits)
              +working_data$projectAge+working_data$turnover
              +working_data$tenureMedian+working_data$tenureDiversity+log(working_data$teamSize)
              +working_data$stCongruence+working_data$centrality+working_data$truckFactor
              +working_data$expertise+working_data$female+working_data$blauGender
              +(1 | working_data$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(black, working_data, trim = 2.5)

# Applying vif <5
print(vif(black))
```

```
## log(working_data$totalCommitters)    log(working_data$totalcommits)
##                          2.546307                          2.307622
##           working_data$projectAge             working_data$turnover
```

```
##                           1.319851                           1.245872
##         working_data$tenureMedian      working_data$tenureDiversity
##                           1.068053                           1.044572
##        log(working_data$teamSize)         working_data$stCongruence
##                           1.819064                           1.035011
##           working_data$centrality          working_data$truckFactor
##                           1.123582                           1.064854
##            working_data$expertise               working_data$female
##                           1.073942                           1.068560
##           working_data$blauGender
##                           2.040802
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
black <- lmer(working_data$black~log(working_data$totalCommitters)+log(working_data$totalcommits)
              +working_data$projectAge+working_data$turnover
              +working_data$tenureMedian+working_data$centrality+working_data$tenureDiversity
              +working_data$stCongruence+working_data$truckFactor
              +working_data$expertise+working_data$female+working_data$blauGender
              +(1 | working_data$window_idx ), REML=FALSE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# print result
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$black ~ log(working_data$totalCommitters) + log(working_data$totalcommits) +
##     working_data$projectAge + working_data$turnover + working_data$tenureMedian +
##     working_data$centrality + working_data$tenureDiversity +
##     working_data$stCongruence + working_data$truckFactor + working_data$expertise +
##     working_data$female + working_data$blauGender + (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                        5.822913   1.561796   3.728
## log(working_data$totalCommitters) -0.135296   0.206874  -0.654
## log(working_data$totalcommits)     0.102238   0.155227   0.659
## working_data$projectAge           -0.024549   0.036136  -0.679
## working_data$turnover              0.782085   0.724888   1.079
## working_data$tenureMedian          0.049157   0.087089   0.564
## working_data$centrality           -0.622232   0.361357  -1.722
## working_data$tenureDiversity       0.012062   0.061803   0.195
```

```
## working_data$stCongruence          -0.222344   0.456140  -0.487
## working_data$truckFactor            0.006548   0.114495   0.057
## working_data$expertise             -0.159103   0.498861  -0.319
## working_data$female                 0.031600   0.027610   1.145
## working_data$blauGender           -10.475619   2.120067  -4.941
```

```
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(black), correlation=TRUE)  or
##     vcov(summary(black))        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
# Applying anova
Anova(black)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$black
##                                  Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.4277  1    0.51311
## log(working_data$totalcommits)     0.4338  1    0.51013
## working_data$projectAge            0.4615  1    0.49692
## working_data$turnover              1.1640  1    0.28063
## working_data$tenureMedian          0.3186  1    0.57245
## working_data$centrality            2.9651  1    0.08508 .
## working_data$tenureDiversity       0.0381  1    0.84525
## working_data$stCongruence          0.2376  1    0.62594
## working_data$truckFactor           0.0033  1    0.95439
## working_data$expertise             0.1017  1    0.74978
## working_data$female                1.3099  1    0.25241
## working_data$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Save in a txt file
sink("hofstede/output_black_hofstede_confounding_variables.txt")
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$black ~ log(working_data$totalCommitters) + log(working_data$totalcommits) +
##     working_data$projectAge + working_data$turnover + working_data$tenureMedian +
##     working_data$centrality + working_data$tenureDiversity +
##     working_data$stCongruence + working_data$truckFactor + working_data$expertise +
##     working_data$female + working_data$blauGender + (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                  Name        Variance Std.Dev.
```

```
##  working_data$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                        5.822913   1.561796   3.728
## log(working_data$totalCommitters)  -0.135296   0.206874  -0.654
## log(working_data$totalcommits)      0.102238   0.155227   0.659
## working_data$projectAge            -0.024549   0.036136  -0.679
## working_data$turnover               0.782085   0.724888   1.079
## working_data$tenureMedian           0.049157   0.087089   0.564
## working_data$centrality            -0.622232   0.361357  -1.722
## working_data$tenureDiversity        0.012062   0.061803   0.195
## working_data$stCongruence          -0.222344   0.456140  -0.487
## working_data$truckFactor            0.006548   0.114495   0.057
## working_data$expertise             -0.159103   0.498861  -0.319
## working_data$female                 0.031600   0.027610   1.145
## working_data$blauGender           -10.475619   2.120067  -4.941
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(black), correlation=TRUE)  or
##     vcov(summary(black))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
Anova(black)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$black
##                                   Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  0.4277  1    0.51311
## log(working_data$totalcommits)     0.4338  1    0.51013
## working_data$projectAge            0.4615  1    0.49692
## working_data$turnover              1.1640  1    0.28063
## working_data$tenureMedian          0.3186  1    0.57245
## working_data$centrality            2.9651  1    0.08508 .
## working_data$tenureDiversity       0.0381  1    0.84525
## working_data$stCongruence          0.2376  1    0.62594
## working_data$truckFactor           0.0033  1    0.95439
## working_data$expertise             0.1017  1    0.74978
## working_data$female                1.3099  1    0.25241
## working_data$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on all the confounding variables trompenaars

```
#----------------------------
#ALL THE CONFOUNDING VARIABLES

# Applying a Linear Mixed Model using the lmer function
blackT <- lmer(working_dataT$black~log(working_dataT$totalCommitters)+log(working_dataT$totalcommits)
             +working_dataT$projectAge+working_dataT$turnover
             +working_dataT$tenureMedian+working_dataT$tenureDiversity+log(working_dataT$teamSize)
             +working_dataT$stCongruence+working_dataT$centrality+working_dataT$truckFactor
             +working_dataT$expertise+working_dataT$female+working_dataT$blauGender
             +(1 | working_dataT$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(blackT, working_dataT, trim = 2.5)

# Applying vif <5
print(vif(blackT))
```

```
## log(working_dataT$totalCommitters)    log(working_dataT$totalcommits)
##                           2.546307                           2.307622
##           working_dataT$projectAge            working_dataT$turnover
##                           1.319851                           1.245872
##         working_dataT$tenureMedian      working_dataT$tenureDiversity
##                           1.068053                           1.044572
##        log(working_dataT$teamSize)         working_dataT$stCongruence
##                           1.819064                           1.035011
##           working_dataT$centrality          working_dataT$truckFactor
##                           1.123582                           1.064854
##            working_dataT$expertise               working_dataT$female
##                           1.073942                           1.068560
##           working_dataT$blauGender
##                           2.040802
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
blackT <- lmer(working_dataT$black~log(working_dataT$totalCommitters)+log(working_dataT$totalcommits)
             +working_dataT$projectAge+working_dataT$turnover
             +working_dataT$tenureMedian+working_dataT$centrality+working_dataT$tenureDiversity
             +working_dataT$stCongruence+working_dataT$truckFactor
             +working_dataT$expertise+working_dataT$female+working_dataT$blauGender
             +(1 | working_dataT$window_idx ), REML=FALSE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# print result
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$black ~ log(working_dataT$totalCommitters) + log(working_dataT$totalcommits) +
##     working_dataT$projectAge + working_dataT$turnover + working_dataT$tenureMedian +
##     working_dataT$centrality + working_dataT$tenureDiversity +
##     working_dataT$stCongruence + working_dataT$truckFactor +
##     working_dataT$expertise + working_dataT$female + working_dataT$blauGender +
```

```
##     (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                         5.822913   1.561796   3.728
## log(working_dataT$totalCommitters) -0.135296   0.206874  -0.654
## log(working_dataT$totalcommits)     0.102238   0.155227   0.659
## working_dataT$projectAge           -0.024549   0.036136  -0.679
## working_dataT$turnover              0.782085   0.724888   1.079
## working_dataT$tenureMedian          0.049157   0.087089   0.564
## working_dataT$centrality           -0.622232   0.361357  -1.722
## working_dataT$tenureDiversity       0.012062   0.061803   0.195
## working_dataT$stCongruence         -0.222344   0.456140  -0.487
## working_dataT$truckFactor           0.006548   0.114495   0.057
## working_dataT$expertise            -0.159103   0.498861  -0.319
## working_dataT$female                0.031600   0.027610   1.145
## working_dataT$blauGender          -10.475619   2.120067  -4.941
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
# Applying anova
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$black
##                                    Chisq Df Pr(>Chisq)
## log(working_dataT$totalCommitters) 0.4277  1    0.51311
## log(working_dataT$totalcommits)    0.4338  1    0.51013
## working_dataT$projectAge           0.4615  1    0.49692
## working_dataT$turnover             1.1640  1    0.28063
## working_dataT$tenureMedian         0.3186  1    0.57245
## working_dataT$centrality           2.9651  1    0.08508 .
## working_dataT$tenureDiversity      0.0381  1    0.84525
## working_dataT$stCongruence         0.2376  1    0.62594
## working_dataT$truckFactor          0.0033  1    0.95439
## working_dataT$expertise            0.1017  1    0.74978
```

```
## working_dataT$female                1.3099  1    0.25241
## working_dataT$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Save in a txt file
sink("trompe/output_black_trompenaars_confounding_variables.txt")
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$black ~ log(working_dataT$totalCommitters) + log(working_dataT$totalcommits) +
##     working_dataT$projectAge + working_dataT$turnover + working_dataT$tenureMedian +
##     working_dataT$centrality + working_dataT$tenureDiversity +
##     working_dataT$stCongruence + working_dataT$truckFactor +
##     working_dataT$expertise + working_dataT$female + working_dataT$blauGender +
##     (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                        5.822913   1.561796   3.728
## log(working_dataT$totalCommitters) -0.135296   0.206874  -0.654
## log(working_dataT$totalcommits)     0.102238   0.155227   0.659
## working_dataT$projectAge           -0.024549   0.036136  -0.679
## working_dataT$turnover              0.782085   0.724888   1.079
## working_dataT$tenureMedian          0.049157   0.087089   0.564
## working_dataT$centrality           -0.622232   0.361357  -1.722
## working_dataT$tenureDiversity       0.012062   0.061803   0.195
## working_dataT$stCongruence         -0.222344   0.456140  -0.487
## working_dataT$truckFactor           0.006548   0.114495   0.057
## working_dataT$expertise            -0.159103   0.498861  -0.319
## working_dataT$female                0.031600   0.027610   1.145
## working_dataT$blauGender          -10.475619   2.120067  -4.941
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
##
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$black
##                                    Chisq Df Pr(>Chisq)
## log(working_dataT$totalCommitters)  0.4277  1    0.51311
## log(working_dataT$totalcommits)     0.4338  1    0.51013
## working_dataT$projectAge            0.4615  1    0.49692
## working_dataT$turnover              1.1640  1    0.28063
## working_dataT$tenureMedian          0.3186  1    0.57245
## working_dataT$centrality            2.9651  1    0.08508 .
## working_dataT$tenureDiversity       0.0381  1    0.84525
## working_dataT$stCongruence          0.2376  1    0.62594
## working_dataT$truckFactor           0.0033  1    0.95439
## working_dataT$expertise             0.1017  1    0.74978
## working_dataT$female                1.3099  1    0.25241
## working_dataT$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on all the confounding variables globe

```
#-----------------------------
#ALL THE CONFOUNDING VARIABLES

# Applying a Linear Mixed Model using the lmer function
blackT <- lmer(working_dataG$black~log(working_dataG$totalCommitters)+log(working_dataG$totalcommits)
             +working_dataG$projectAge+working_dataG$turnover
             +working_dataG$tenureMedian+working_dataG$tenureDiversity+log(working_dataG$teamSize)
             +working_dataG$stCongruence+working_dataG$centrality+working_dataG$truckFactor
             +working_dataG$expertise+working_dataG$female+working_dataG$blauGender
             +(1 | working_dataG$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(blackT, working_dataG, trim = 2.5)

# Applying vif <5
print(vif(blackT))
```

```
## log(working_dataG$totalCommitters)    log(working_dataG$totalcommits)
##                           2.546307                           2.307622
##           working_dataG$projectAge            working_dataG$turnover
##                           1.319851                           1.245872
##         working_dataG$tenureMedian      working_dataG$tenureDiversity
##                           1.068053                           1.044572
##        log(working_dataG$teamSize)         working_dataG$stCongruence
##                           1.819064                           1.035011
##           working_dataG$centrality          working_dataG$truckFactor
##                           1.123582                           1.064854
```

```
##            working_dataG$expertise                working_dataG$female
##                           1.073942                            1.068560
##           working_dataG$blauGender
##                           2.040802
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
blackT <- lmer(working_dataG$black~log(working_dataG$totalCommitters)+log(working_dataG$totalcommits)
            +working_dataG$projectAge+working_dataG$turnover
            +working_dataG$tenureMedian+working_dataG$centrality+working_dataG$tenureDiversity
            +working_dataG$stCongruence+working_dataG$truckFactor
            +working_dataG$expertise+working_dataG$female+working_dataG$blauGender
            +(1 | working_dataG$window_idx ), REML=FALSE)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
# print result
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$black ~ log(working_dataG$totalCommitters) + log(working_dataG$totalcommits) +
##     working_dataG$projectAge + working_dataG$turnover + working_dataG$tenureMedian +
##     working_dataG$centrality + working_dataG$tenureDiversity +
##     working_dataG$stCongruence + working_dataG$truckFactor +
##     working_dataG$expertise + working_dataG$female + working_dataG$blauGender +
##     (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                     Estimate Std. Error t value
## (Intercept)                         5.822913   1.561796   3.728
## log(working_dataG$totalCommitters) -0.135296   0.206874  -0.654
## log(working_dataG$totalcommits)     0.102238   0.155227   0.659
## working_dataG$projectAge           -0.024549   0.036136  -0.679
## working_dataG$turnover              0.782085   0.724888   1.079
## working_dataG$tenureMedian          0.049157   0.087089   0.564
## working_dataG$centrality           -0.622232   0.361357  -1.722
## working_dataG$tenureDiversity       0.012062   0.061803   0.195
## working_dataG$stCongruence         -0.222344   0.456140  -0.487
## working_dataG$truckFactor           0.006548   0.114495   0.057
## working_dataG$expertise            -0.159103   0.498861  -0.319
## working_dataG$female                0.031600   0.027610   1.145
## working_dataG$blauGender          -10.475619   2.120067  -4.941
```

```
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
# Applying anova
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$black
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.4277  1    0.51311
## log(working_dataG$totalcommits)     0.4338  1    0.51013
## working_dataG$projectAge            0.4615  1    0.49692
## working_dataG$turnover              1.1640  1    0.28063
## working_dataG$tenureMedian          0.3186  1    0.57245
## working_dataG$centrality            2.9651  1    0.08508 .
## working_dataG$tenureDiversity       0.0381  1    0.84525
## working_dataG$stCongruence          0.2376  1    0.62594
## working_dataG$truckFactor           0.0033  1    0.95439
## working_dataG$expertise             0.1017  1    0.74978
## working_dataG$female                1.3099  1    0.25241
## working_dataG$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Save in a txt file
sink("globe/output_black_globe_confounding_variables.txt")
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$black ~ log(working_dataG$totalCommitters) + log(working_dataG$totalcommits) +
##     working_dataG$projectAge + working_dataG$turnover + working_dataG$tenureMedian +
##     working_dataG$centrality + working_dataG$tenureDiversity +
##     working_dataG$stCongruence + working_dataG$truckFactor +
##     working_dataG$expertise + working_dataG$female + working_dataG$blauGender +
##     (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    861.9    910.7   -415.9    831.9      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.82099 -0.89953  0.01655  0.89299  1.87441
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.000    0.000
##  Residual                             4.459    2.112
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
```

```
## Fixed effects:
##                                     Estimate Std. Error t value
## (Intercept)                         5.822913   1.561796   3.728
## log(working_dataG$totalCommitters) -0.135296   0.206874  -0.654
## log(working_dataG$totalcommits)     0.102238   0.155227   0.659
## working_dataG$projectAge           -0.024549   0.036136  -0.679
## working_dataG$turnover              0.782085   0.724888   1.079
## working_dataG$tenureMedian          0.049157   0.087089   0.564
## working_dataG$centrality           -0.622232   0.361357  -1.722
## working_dataG$tenureDiversity       0.012062   0.061803   0.195
## working_dataG$stCongruence         -0.222344   0.456140  -0.487
## working_dataG$truckFactor           0.006548   0.114495   0.057
## working_dataG$expertise            -0.159103   0.498861  -0.319
## working_dataG$female                0.031600   0.027610   1.145
## working_dataG$blauGender          -10.475619   2.120067  -4.941
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(blackT), correlation=TRUE)  or
##     vcov(summary(blackT))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
Anova(blackT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$black
##                                     Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.4277  1    0.51311
## log(working_dataG$totalcommits)     0.4338  1    0.51013
## working_dataG$projectAge            0.4615  1    0.49692
## working_dataG$turnover              1.1640  1    0.28063
## working_dataG$tenureMedian          0.3186  1    0.57245
## working_dataG$centrality            2.9651  1    0.08508 .
## working_dataG$tenureDiversity       0.0381  1    0.84525
## working_dataG$stCongruence          0.2376  1    0.62594
## working_dataG$truckFactor           0.0033  1    0.95439
## working_dataG$expertise             0.1017  1    0.74978
## working_dataG$female                1.3099  1    0.25241
## working_dataG$blauGender           24.4152  1  7.765e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sink()
```

Linear Mixed Model using lmer function on only random effect hofsetde

```
#-----------------------------
#ONLY RANDOM EFFECT

# Applying a Linear Mixed Model using the lmer function
black <- lmer(working_data$black~(1 | working_data$window_idx ), REML=FALSE)
```

```r
# Remove outlier
#romr.fnc(black, working_data, trim = 2.5)

# print result
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_data$black ~ (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```r
# Save in a txt file
sink("hofstede/output_black_hofstede_random.txt")
print(summary(black))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_data$black ~ (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```r
sink()
```

Linear Mixed Model using lmer function on only random effect trompenaars

```
#----------------------------
#ONLY RANDOM EFFECT

# Applying a Linear Mixed Model using the lmer function
blackT <- lmer(working_dataT$black~(1 | working_dataT$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(blackT, working_dataT, trim = 2.5)

# print result
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataT$black ~ (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```
# Save in a txt file
sink("trompe/output_black_trompenaars_random.txt")
print(summary(blackT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataT$black ~ (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_dataT$window_idx, 24
```

```
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```
sink()
```

Linear Mixed Model using lmer function on only random effect globe

```
#----------------------------
#ONLY RANDOM EFFECT

# Applying a Linear Mixed Model using the lmer function
blackG <- lmer(working_dataG$black~(1 | working_dataG$window_idx ), REML=FALSE)

# Remove outlier
#romr.fnc(blackG, working_dataG, trim = 2.5)

# print result
print(summary(blackG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataG$black ~ (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```
# Save in a txt file
sink("globe/output_black_globe_random.txt")
print(summary(blackG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataG$black ~ (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    878.8    888.6   -436.4    872.8      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -1.9295 -1.0075 -0.1742  0.7660  1.5643
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.05738  0.2395
##  Residual                             5.46349  2.3374
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   4.5060     0.1766   25.51
```

```
sink()
```