LonewolfModelConstruction

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 Lone wolf 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("./lone_wolf_metrics_hofstede.csv", sep = ";", header = TRUE, stringsAsFactors=FALSE)
# Reading input trompenaars data.
dataT <- read.csv("./lone_wolf_metrics_trompenaars.csv", sep = ";", header = TRUE, stringsAsFactors=FALS
# Reading input globe data.
dataG <- read.csv("./lone_wolf_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
lone <- lmer(working_data$lone~log(working_data$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)

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

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

```
## log(working_data$totalCommitters)    log(working_data$totalcommits)
##                           3.127397                           3.097453
##           working_data$projectAge             working_data$turnover
##                           1.416397                           1.395505
##           working_data$blauGender         working_data$tenureMedian
##                           2.506052                           1.114971
##      working_data$tenureDiversity        log(working_data$teamSize)
##                           1.066575                           3.153244
##         working_data$stCongruence          working_data$truckFactor
##                           3.170433                           1.089554
##               working_data$female            working_data$expertise
##                           1.072264                           1.101521
##           working_data$centrality                 working_data$CV_1
##                           1.223704                           4.740308
##                 working_data$CV_2                 working_data$CV_3
##                           6.153648                           3.225858
##                 working_data$CV_4                 working_data$CV_5
##                           8.420680                           3.878063
##                 working_data$CV_6
##                           7.464332
```

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

# print result
print(summary(lone))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$lone ~ log(working_data$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$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    994.5   1066.1   -475.2    950.5      170
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.59343 -0.65168  0.03541  0.81667  2.40098
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.05678  0.2383
##  Residual                             8.21281  2.8658
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                     -8.009780   4.344510  -1.844
## log(working_data$totalCommitters)   0.429837   0.314840   1.365
## log(working_data$totalcommits)      0.080659   0.244912   0.329
## working_data$projectAge             0.102810   0.051221   2.007
## working_data$turnover               0.164711   1.052333   0.157
## working_data$blauGender            -5.532499   3.583147  -1.544
## working_data$tenureMedian          -0.078096   0.121013  -0.645
## working_data$tenureDiversity        0.060009   0.084987   0.706
## log(working_data$teamSize)          4.525513   0.878066   5.154
## working_data$stCongruence          -8.261058   1.497358  -5.517
## working_data$truckFactor           -0.150398   0.157602  -0.954
## working_data$female                -0.002416   0.037965  -0.064
## working_data$expertise              0.607594   0.690377   0.880
## working_data$centrality             0.151354   0.515053   0.294
## working_data$CV_1                 -11.214253   4.466674  -2.511
## working_data$CV_2                 -14.548135   5.485614  -2.652
## working_data$CV_3                   0.641192   4.004666   0.160
## working_data$CV_4                  10.152959   7.263162   1.398
## working_data$CV_5                  -2.449517   3.363413  -0.728
## working_data$CV_6                  13.190426   6.088696   2.166
##
## Correlation matrix not shown by default, as p = 20 > 12.
## Use print(summary(lone), correlation=TRUE)  or
##     vcov(summary(lone))        if you need it
```

```
# Applying anova
Anova(lone)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$lone
##                                  Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  1.8639  1    0.17217
## log(working_data$totalcommits)     0.1085  1    0.74190
```

```
## working_data$projectAge            4.0288  1    0.04473 *
## working_data$turnover              0.0245  1    0.87562
## working_data$blauGender            2.3840  1    0.12258
## working_data$tenureMedian          0.4165  1    0.51870
## working_data$tenureDiversity       0.4986  1    0.48013
## log(working_data$teamSize)        26.5633  1  2.550e-07 ***
## working_data$stCongruence         30.4383  1  3.447e-08 ***
## working_data$truckFactor           0.9107  1    0.33994
## working_data$female                0.0041  1    0.94925
## working_data$expertise             0.7746  1    0.37881
## working_data$centrality            0.0864  1    0.76886
## working_data$CV_1                  6.3034  1    0.01205 *
## working_data$CV_2                  7.0334  1    0.00800 **
## working_data$CV_3                  0.0256  1    0.87279
## working_data$CV_4                  1.9540  1    0.16215
## working_data$CV_5                  0.5304  1    0.46644
## working_data$CV_6                  4.6932  1    0.03028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$lone ~ log(working_data$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$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    994.5   1066.1   -475.2    950.5      170
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.59343 -0.65168  0.03541  0.81667  2.40098
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.05678  0.2383
##  Residual                             8.21281  2.8658
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                   Estimate Std. Error t value
## (Intercept)                      -8.009780   4.344510  -1.844
## log(working_data$totalCommitters)   0.429837   0.314840   1.365
## log(working_data$totalcommits)      0.080659   0.244912   0.329
## working_data$projectAge             0.102810   0.051221   2.007
## working_data$turnover               0.164711   1.052333   0.157
```

```
## working_data$blauGender            -5.532499   3.583147  -1.544
## working_data$tenureMedian          -0.078096   0.121013  -0.645
## working_data$tenureDiversity        0.060009   0.084987   0.706
## log(working_data$teamSize)          4.525513   0.878066   5.154
## working_data$stCongruence          -8.261058   1.497358  -5.517
## working_data$truckFactor           -0.150398   0.157602  -0.954
## working_data$female                -0.002416   0.037965  -0.064
## working_data$expertise              0.607594   0.690377   0.880
## working_data$centrality             0.151354   0.515053   0.294
## working_data$CV_1                 -11.214253   4.466674  -2.511
## working_data$CV_2                 -14.548135   5.485614  -2.652
## working_data$CV_3                   0.641192   4.004666   0.160
## working_data$CV_4                  10.152959   7.263162   1.398
## working_data$CV_5                  -2.449517   3.363413  -0.728
## working_data$CV_6                  13.190426   6.088696   2.166
##
## Correlation matrix not shown by default, as p = 20 > 12.
## Use print(summary(lone), correlation=TRUE)  or
##     vcov(summary(lone))        if you need it
```

```
Anova(lone)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$lone
##                                Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  1.8639  1    0.17217
## log(working_data$totalcommits)     0.1085  1    0.74190
## working_data$projectAge            4.0288  1    0.04473 *
## working_data$turnover              0.0245  1    0.87562
## working_data$blauGender            2.3840  1    0.12258
## working_data$tenureMedian          0.4165  1    0.51870
## working_data$tenureDiversity       0.4986  1    0.48013
## log(working_data$teamSize)        26.5633  1  2.550e-07 ***
## working_data$stCongruence         30.4383  1  3.447e-08 ***
## working_data$truckFactor           0.9107  1    0.33994
## working_data$female                0.0041  1    0.94925
## working_data$expertise             0.7746  1    0.37881
## working_data$centrality            0.0864  1    0.76886
## working_data$CV_1                  6.3034  1    0.01205 *
## working_data$CV_2                  7.0334  1    0.00800 **
## working_data$CV_3                  0.0256  1    0.87279
## working_data$CV_4                  1.9540  1    0.16215
## working_data$CV_5                  0.5304  1    0.46644
## working_data$CV_6                  4.6932  1    0.03028 *
## ---
## 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
loneT <- lmer(working_dataT$lone~log(working_data$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(loneT))
```

```
## log(working_data$totalCommitters)   log(working_dataT$totalcommits)
##                          2.884017                          2.989136
##           working_data$projectAge           working_dataT$turnover
##                          1.398640                          1.495966
##          working_dataT$blauGender       working_dataT$tenureMedian
##                          2.258308                          1.099304
##     working_dataT$tenureDiversity      log(working_dataT$teamSize)
##                          1.085177                          3.205166
##        working_dataT$stCongruence        working_dataT$truckFactor
##                          3.238808                          1.091318
##              working_dataT$female          working_dataT$expertise
##                          1.089701                          1.132044
##          working_dataT$centrality               working_dataT$CV_1
##                          1.233525                         14.436426
##                 working_dataT$CV_2               working_dataT$CV_3
##                         11.764637                          6.260312
##                 working_dataT$CV_4               working_dataT$CV_5
##                         11.024564                         20.883748
##                 working_dataT$CV_6               working_dataT$CV_7
##                          2.946046                          6.169553
##                 working_dataT$CV_8
##                          4.445082
```

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

# print result
print(summary(loneT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$lone ~ log(working_data$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
##    993.9   1072.1   -473.0    945.9      168
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.55741 -0.63694  0.04516  0.78987  2.20695
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.1383   0.3719
##  Residual                             7.9459   2.8189
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                     -11.02558    4.36687  -2.525
## log(working_data$totalCommitters)   0.41449    0.30115   1.376
## log(working_dataT$totalcommits)     0.29011    0.24042   1.207
## working_data$projectAge             0.10274    0.05059   2.031
## working_dataT$turnover             -0.38377    1.08616  -0.353
## working_dataT$blauGender           -4.40217    3.37144  -1.306
## working_dataT$tenureMedian         -0.01189    0.11844  -0.100
## working_dataT$tenureDiversity       0.08668    0.08462   1.024
## log(working_dataT$teamSize)         4.79067    0.87435   5.479
## working_dataT$stCongruence         -7.60461    1.49209  -5.097
## working_dataT$truckFactor          -0.18626    0.15562  -1.197
## working_dataT$female                0.01488    0.03778   0.394
## working_dataT$expertise             0.66945    0.69033   0.970
## working_dataT$centrality            0.01405    0.51151   0.027
## working_dataT$CV_1                 -8.50224    7.44150  -1.143
## working_dataT$CV_2                  5.16374    5.49044   0.940
## working_dataT$CV_3                 -4.75285    4.48460  -1.060
## working_dataT$CV_4                  4.90803    5.93534   0.827
## working_dataT$CV_5                  0.46270    7.22820   0.064
## working_dataT$CV_6                 -2.08982    1.93993  -1.077
## working_dataT$CV_7                -13.15557    4.79748  -2.742
## working_dataT$CV_8                  9.96360    3.70005   2.693
##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(summary(loneT), correlation=TRUE)  or
##     vcov(summary(loneT))        if you need it
```

```
# Applying anova
Anova(loneT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  1.8943  1   0.168718
## log(working_dataT$totalcommits)    1.4562  1   0.227541
## working_data$projectAge            4.1246  1   0.042264 *
## working_dataT$turnover             0.1248  1   0.723842
## working_dataT$blauGender           1.7049  1   0.191646
## working_dataT$tenureMedian         0.0101  1   0.920023
## working_dataT$tenureDiversity      1.0493  1   0.305668
## log(working_dataT$teamSize)       30.0206  1  4.275e-08 ***
## working_dataT$stCongruence        25.9756  1  3.458e-07 ***
## working_dataT$truckFactor          1.4327  1   0.231329
## working_dataT$female               0.1551  1   0.693720
## working_dataT$expertise            0.9404  1   0.332173
## working_dataT$centrality           0.0008  1   0.978080
## working_dataT$CV_1                 1.3054  1   0.253228
## working_dataT$CV_2                 0.8845  1   0.346962
## working_dataT$CV_3                 1.1232  1   0.289229
## working_dataT$CV_4                 0.6838  1   0.408284
## working_dataT$CV_5                 0.0041  1   0.948960
## working_dataT$CV_6                 1.1605  1   0.281362
## working_dataT$CV_7                 7.5196  1   0.006103 **
## working_dataT$CV_8                 7.2513  1   0.007085 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$lone ~ log(working_data$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$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##    993.9   1072.1   -473.0    945.9      168
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.55741 -0.63694  0.04516  0.78987  2.20695
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.1383   0.3719
##  Residual                             7.9459   2.8189
```

```
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                     -11.02558    4.36687  -2.525
## log(working_data$totalCommitters)   0.41449    0.30115   1.376
## log(working_dataT$totalcommits)     0.29011    0.24042   1.207
## working_data$projectAge             0.10274    0.05059   2.031
## working_dataT$turnover             -0.38377    1.08616  -0.353
## working_dataT$blauGender           -4.40217    3.37144  -1.306
## working_dataT$tenureMedian         -0.01189    0.11844  -0.100
## working_dataT$tenureDiversity       0.08668    0.08462   1.024
## log(working_dataT$teamSize)         4.79067    0.87435   5.479
## working_dataT$stCongruence         -7.60461    1.49209  -5.097
## working_dataT$truckFactor          -0.18626    0.15562  -1.197
## working_dataT$female                0.01488    0.03778   0.394
## working_dataT$expertise             0.66945    0.69033   0.970
## working_dataT$centrality            0.01405    0.51151   0.027
## working_dataT$CV_1                 -8.50224    7.44150  -1.143
## working_dataT$CV_2                  5.16374    5.49044   0.940
## working_dataT$CV_3                 -4.75285    4.48460  -1.060
## working_dataT$CV_4                  4.90803    5.93534   0.827
## working_dataT$CV_5                  0.46270    7.22820   0.064
## working_dataT$CV_6                 -2.08982    1.93993  -1.077
## working_dataT$CV_7                -13.15557    4.79748  -2.742
## working_dataT$CV_8                  9.96360    3.70005   2.693
##
## Correlation matrix not shown by default, as p = 22 > 12.
## Use print(summary(loneT), correlation=TRUE)  or
##     vcov(summary(loneT))        if you need it
```

```
Anova(loneT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$lone
##                                   Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)  1.8943  1   0.168718
## log(working_dataT$totalcommits)    1.4562  1   0.227541
## working_data$projectAge            4.1246  1   0.042264 *
## working_dataT$turnover             0.1248  1   0.723842
## working_dataT$blauGender           1.7049  1   0.191646
## working_dataT$tenureMedian         0.0101  1   0.920023
## working_dataT$tenureDiversity      1.0493  1   0.305668
## log(working_dataT$teamSize)       30.0206  1  4.275e-08 ***
## working_dataT$stCongruence        25.9756  1  3.458e-07 ***
## working_dataT$truckFactor          1.4327  1   0.231329
## working_dataT$female               0.1551  1   0.693720
## working_dataT$expertise            0.9404  1   0.332173
## working_dataT$centrality           0.0008  1   0.978080
## working_dataT$CV_1                 1.3054  1   0.253228
## working_dataT$CV_2                 0.8845  1   0.346962
## working_dataT$CV_3                 1.1232  1   0.289229
## working_dataT$CV_4                 0.6838  1   0.408284
```

```
## working_dataT$CV_5                 0.0041  1   0.948960
## working_dataT$CV_6                 1.1605  1   0.281362
## working_dataT$CV_7                 7.5196  1   0.006103 **
## working_dataT$CV_8                 7.2513  1   0.007085 **
## ---
## 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
loneG <- lmer(working_dataG$lone~log(working_dataG$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)

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

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

```
## log(working_dataG$totalCommitters)    log(working_dataG$totalcommits)
##                           3.201664                           3.352180
##           working_dataG$projectAge            working_dataG$turnover
##                           1.393525                           1.579403
##           working_dataG$blauGender         working_dataG$tenureMedian
##                           2.351341                           1.112815
##      working_dataG$tenureDiversity        log(working_dataG$teamSize)
##                           1.086919                           3.292117
##         working_dataG$stCongruence          working_dataG$truckFactor
##                           3.390140                           1.119071
##               working_dataG$female            working_dataG$expertise
##                           1.105679                           1.190480
##           working_dataG$centrality                 working_dataG$CV_1
##                           1.226292                          18.679661
##                 working_dataG$CV_2                 working_dataG$CV_3
##                           7.308193                           5.529222
##                 working_dataG$CV_4                 working_dataG$CV_5
##                          10.360361                           5.173008
##                 working_dataG$CV_6                 working_dataG$CV_7
##                          15.018433                           5.947674
##                 working_dataG$CV_8                 working_dataG$CV_9
```

```
##                             16.882113                             8.048257
```

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

# print result
print(summary(loneG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$lone ~ log(working_dataG$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
##    990.2   1071.6   -470.1    940.2      167
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.63022 -0.60800  0.00714  0.82264  2.21879
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.09935  0.3152
##  Residual                             7.74283  2.7826
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                    -10.33562    4.33548  -2.384
## log(working_dataG$totalCommitters)   0.59835    0.31159   1.920
## log(working_dataG$totalcommits)      0.07653    0.24967   0.307
## working_dataG$projectAge             0.10014    0.04963   2.018
## working_dataG$turnover              -0.68879    1.09557  -0.629
## working_dataG$blauGender            -5.54141    3.38513  -1.637
## working_dataG$tenureMedian          -0.04912    0.11753  -0.418
## working_dataG$tenureDiversity        0.06810    0.08347   0.816
## log(working_dataG$teamSize)          4.76374    0.87323   5.455
## working_dataG$stCongruence          -7.67400    1.50545  -5.097
## working_dataG$truckFactor           -0.15647    0.15536  -1.007
## working_dataG$female                 0.01678    0.03751   0.447
## working_dataG$expertise              0.91410    0.69801   1.310
## working_dataG$centrality             0.09764    0.50227   0.194
## working_dataG$CV_1                 -34.33837   24.92454  -1.378
## working_dataG$CV_2                  -2.70069   18.79953  -0.144
## working_dataG$CV_3                  41.19147   22.75542   1.810
## working_dataG$CV_4                   8.64682   25.79679   0.335
## working_dataG$CV_5                  -9.43675   19.35032  -0.488
## working_dataG$CV_6                  34.55855   39.30303   0.879
## working_dataG$CV_7                 -58.41290   13.08040  -4.466
```

```
## working_dataG$CV_8                  76.52825   36.52777   2.095
## working_dataG$CV_9                 -19.13391   27.20753  -0.703
```

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

```
# Applying anova
Anova(loneG)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$lone
##                                      Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  3.6877  1    0.05482 .
## log(working_dataG$totalcommits)     0.0939  1    0.75922
## working_dataG$projectAge            4.0709  1    0.04363 *
## working_dataG$turnover              0.3953  1    0.52955
## working_dataG$blauGender            2.6797  1    0.10163
## working_dataG$tenureMedian          0.1747  1    0.67598
## working_dataG$tenureDiversity       0.6656  1    0.41459
## log(working_dataG$teamSize)        29.7605  1  4.888e-08 ***
## working_dataG$stCongruence         25.9841  1  3.442e-07 ***
## working_dataG$truckFactor           1.0143  1    0.31388
## working_dataG$female                0.2001  1    0.65464
## working_dataG$expertise             1.7150  1    0.19034
## working_dataG$centrality            0.0378  1    0.84587
## working_dataG$CV_1                  1.8980  1    0.16830
## working_dataG$CV_2                  0.0206  1    0.88577
## working_dataG$CV_3                  3.2768  1    0.07027 .
## working_dataG$CV_4                  0.1124  1    0.73748
## working_dataG$CV_5                  0.2378  1    0.62578
## working_dataG$CV_6                  0.7731  1    0.37925
## working_dataG$CV_7                 19.9423  1  7.981e-06 ***
## working_dataG$CV_8                  4.3893  1    0.03616 *
## working_dataG$CV_9                  0.4946  1    0.48189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$lone ~ log(working_dataG$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
##    990.2   1071.6   -470.1    940.2      167
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.63022 -0.60800  0.00714  0.82264  2.21879
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.09935  0.3152
##  Residual                             7.74283  2.7826
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                       -10.33562    4.33548  -2.384
## log(working_dataG$totalCommitters)   0.59835    0.31159   1.920
## log(working_dataG$totalcommits)      0.07653    0.24967   0.307
## working_dataG$projectAge             0.10014    0.04963   2.018
## working_dataG$turnover              -0.68879    1.09557  -0.629
## working_dataG$blauGender            -5.54141    3.38513  -1.637
## working_dataG$tenureMedian          -0.04912    0.11753  -0.418
## working_dataG$tenureDiversity        0.06810    0.08347   0.816
## log(working_dataG$teamSize)          4.76374    0.87323   5.455
## working_dataG$stCongruence          -7.67400    1.50545  -5.097
## working_dataG$truckFactor           -0.15647    0.15536  -1.007
## working_dataG$female                 0.01678    0.03751   0.447
## working_dataG$expertise              0.91410    0.69801   1.310
## working_dataG$centrality             0.09764    0.50227   0.194
## working_dataG$CV_1                 -34.33837   24.92454  -1.378
## working_dataG$CV_2                  -2.70069   18.79953  -0.144
## working_dataG$CV_3                  41.19147   22.75542   1.810
## working_dataG$CV_4                   8.64682   25.79679   0.335
## working_dataG$CV_5                  -9.43675   19.35032  -0.488
## working_dataG$CV_6                  34.55855   39.30303   0.879
## working_dataG$CV_7                 -58.41290   13.08040  -4.466
## working_dataG$CV_8                  76.52825   36.52777   2.095
## working_dataG$CV_9                 -19.13391   27.20753  -0.703
##
## Correlation matrix not shown by default, as p = 23 > 12.
## Use print(summary(loneG), correlation=TRUE)  or
##     vcov(summary(loneG))        if you need it
```

```
Anova(loneG)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  3.6877  1    0.05482 .
## log(working_dataG$totalcommits)     0.0939  1    0.75922
## working_dataG$projectAge            4.0709  1    0.04363 *
## working_dataG$turnover              0.3953  1    0.52955
## working_dataG$blauGender            2.6797  1    0.10163
```

```
## working_dataG$tenureMedian          0.1747  1    0.67598
## working_dataG$tenureDiversity       0.6656  1    0.41459
## log(working_dataG$teamSize)        29.7605  1  4.888e-08 ***
## working_dataG$stCongruence         25.9841  1  3.442e-07 ***
## working_dataG$truckFactor           1.0143  1    0.31388
## working_dataG$female                0.2001  1    0.65464
## working_dataG$expertise             1.7150  1    0.19034
## working_dataG$centrality            0.0378  1    0.84587
## working_dataG$CV_1                  1.8980  1    0.16830
## working_dataG$CV_2                  0.0206  1    0.88577
## working_dataG$CV_3                  3.2768  1    0.07027 .
## working_dataG$CV_4                  0.1124  1    0.73748
## working_dataG$CV_5                  0.2378  1    0.62578
## working_dataG$CV_6                  0.7731  1    0.37925
## working_dataG$CV_7                 19.9423  1  7.981e-06 ***
## working_dataG$CV_8                  4.3893  1    0.03616 *
## working_dataG$CV_9                  0.4946  1    0.48189
## ---
## 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
lone <- lmer(working_data$lone~log(working_data$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)
```

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

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

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

```
## log(working_data$totalCommitters)    log(working_data$totalcommits)
##                          2.539211                          2.361283
##           working_data$projectAge             working_data$turnover
##                          1.320130                          1.253354
##         working_data$tenureMedian      working_data$tenureDiversity
##                          1.063707                          1.046162
##        log(working_data$teamSize)         working_data$stCongruence
##                          3.077293                          3.078046
##           working_data$centrality          working_data$truckFactor
##                          1.156810                          1.067097
```

```
##            working_data$expertise                working_data$female
##                          1.069665                           1.036473
##           working_data$blauGender
##                          1.632092
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
lone <- lmer(working_data$lone~log(working_data$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(lone))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$lone ~ log(working_data$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                         8.813260   2.295216   3.840
## log(working_data$totalCommitters)   0.214510   0.308961   0.694
## log(working_data$totalcommits)      0.184696   0.231809   0.797
## working_data$projectAge             0.117693   0.053584   2.196
## working_data$turnover               0.869837   1.083180   0.803
## working_data$tenureMedian          -0.062020   0.129759  -0.478
## working_data$centrality             0.925333   0.533883   1.733
## working_data$tenureDiversity        0.076192   0.092256   0.826
## working_data$stCongruence         -14.353546   0.939173 -15.283
## working_data$truckFactor           -0.173819   0.171079  -1.016
## working_data$expertise              0.489398   0.745927   0.656
## working_data$female                 0.008265   0.040935   0.202
## working_data$blauGender            -2.227819   3.153427  -0.706
```

```
##
```

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

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

```
# Applying anova
Anova(lone)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)   0.4820  1    0.48750
## log(working_data$totalcommits)      0.6348  1    0.42559
## working_data$projectAge             4.8242  1    0.02806 *
## working_data$turnover               0.6449  1    0.42195
## working_data$tenureMedian           0.2285  1    0.63267
## working_data$centrality             3.0040  1    0.08306 .
## working_data$tenureDiversity        0.6821  1    0.40888
## working_data$stCongruence         233.5753  1    < 2e-16 ***
## working_data$truckFactor            1.0323  1    0.30962
## working_data$expertise              0.4305  1    0.51176
## working_data$female                 0.0408  1    0.84000
## working_data$blauGender             0.4991  1    0.47989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_data$lone ~ log(working_data$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##                                     Estimate Std. Error t value
```

```
## (Intercept)                          8.813260   2.295216   3.840
## log(working_data$totalCommitters)    0.214510   0.308961   0.694
## log(working_data$totalcommits)       0.184696   0.231809   0.797
## working_data$projectAge              0.117693   0.053584   2.196
## working_data$turnover                0.869837   1.083180   0.803
## working_data$tenureMedian           -0.062020   0.129759  -0.478
## working_data$centrality              0.925333   0.533883   1.733
## working_data$tenureDiversity         0.076192   0.092256   0.826
## working_data$stCongruence          -14.353546   0.939173 -15.283
## working_data$truckFactor            -0.173819   0.171079  -1.016
## working_data$expertise               0.489398   0.745927   0.656
## working_data$female                  0.008265   0.040935   0.202
## working_data$blauGender             -2.227819   3.153427  -0.706

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

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

```
Anova(lone)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_data$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_data$totalCommitters)   0.4820  1    0.48750
## log(working_data$totalcommits)      0.6348  1    0.42559
## working_data$projectAge             4.8242  1    0.02806 *
## working_data$turnover               0.6449  1    0.42195
## working_data$tenureMedian           0.2285  1    0.63267
## working_data$centrality             3.0040  1    0.08306 .
## working_data$tenureDiversity        0.6821  1    0.40888
## working_data$stCongruence         233.5753  1    < 2e-16 ***
## working_data$truckFactor            1.0323  1    0.30962
## working_data$expertise              0.4305  1    0.51176
## working_data$female                 0.0408  1    0.84000
## working_data$blauGender             0.4991  1    0.47989
## ---
## 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
loneT <- lmer(working_dataT$lone~log(working_dataT$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)
```

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

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

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

```
## log(working_dataT$totalCommitters)    log(working_dataT$totalcommits)
##                           2.539211                           2.361283
##          working_dataT$projectAge            working_dataT$turnover
##                           1.320130                           1.253354
##        working_dataT$tenureMedian      working_dataT$tenureDiversity
##                           1.063707                           1.046162
##       log(working_dataT$teamSize)         working_dataT$stCongruence
##                           3.077293                           3.078046
##          working_dataT$centrality          working_dataT$truckFactor
##                           1.156810                           1.067097
##           working_dataT$expertise               working_dataT$female
##                           1.069665                           1.036473
##          working_dataT$blauGender
##                           1.632092
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
loneT <- lmer(working_dataT$lone~log(working_dataT$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(loneT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataT$lone ~ log(working_dataT$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
```

```
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                        8.813260   2.295216   3.840
## log(working_dataT$totalCommitters)   0.214510   0.308961   0.694
## log(working_dataT$totalcommits)      0.184696   0.231809   0.797
## working_dataT$projectAge             0.117693   0.053584   2.196
## working_dataT$turnover               0.869837   1.083180   0.803
## working_dataT$tenureMedian          -0.062020   0.129759  -0.478
## working_dataT$centrality             0.925333   0.533883   1.733
## working_dataT$tenureDiversity        0.076192   0.092256   0.826
## working_dataT$stCongruence         -14.353546   0.939173 -15.283
## working_dataT$truckFactor           -0.173819   0.171079  -1.016
## working_dataT$expertise              0.489398   0.745927   0.656
## working_dataT$female                 0.008265   0.040935   0.202
## working_dataT$blauGender            -2.227819   3.153427  -0.706
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(loneT), correlation=TRUE)  or
##     vcov(summary(loneT))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
# Applying anova
Anova(loneT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$lone
##                                     Chisq Df Pr(>Chisq)
## log(working_dataT$totalCommitters)   0.4820  1    0.48750
## log(working_dataT$totalcommits)      0.6348  1    0.42559
## working_dataT$projectAge             4.8242  1    0.02806 *
## working_dataT$turnover               0.6449  1    0.42195
## working_dataT$tenureMedian           0.2285  1    0.63267
## working_dataT$centrality             3.0040  1    0.08306 .
## working_dataT$tenureDiversity        0.6821  1    0.40888
## working_dataT$stCongruence         233.5753  1    < 2e-16 ***
## working_dataT$truckFactor            1.0323  1    0.30962
## working_dataT$expertise              0.4305  1    0.51176
## working_dataT$female                 0.0408  1    0.84000
## working_dataT$blauGender             0.4991  1    0.47989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
```

```
## Formula:
## working_dataT$lone ~ log(working_dataT$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##                                     Estimate Std. Error t value
## (Intercept)                         8.813260   2.295216   3.840
## log(working_dataT$totalCommitters)  0.214510   0.308961   0.694
## log(working_dataT$totalcommits)     0.184696   0.231809   0.797
## working_dataT$projectAge            0.117693   0.053584   2.196
## working_dataT$turnover              0.869837   1.083180   0.803
## working_dataT$tenureMedian         -0.062020   0.129759  -0.478
## working_dataT$centrality            0.925333   0.533883   1.733
## working_dataT$tenureDiversity       0.076192   0.092256   0.826
## working_dataT$stCongruence        -14.353546   0.939173 -15.283
## working_dataT$truckFactor          -0.173819   0.171079  -1.016
## working_dataT$expertise             0.489398   0.745927   0.656
## working_dataT$female                0.008265   0.040935   0.202
## working_dataT$blauGender           -2.227819   3.153427  -0.706
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(loneT), correlation=TRUE)  or
##     vcov(summary(loneT))        if you need it
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
Anova(loneT)

## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataT$lone
##                                      Chisq Df Pr(>Chisq)
## log(working_dataT$totalCommitters)  0.4820  1    0.48750
## log(working_dataT$totalcommits)     0.6348  1    0.42559
## working_dataT$projectAge            4.8242  1    0.02806 *
## working_dataT$turnover              0.6449  1    0.42195
## working_dataT$tenureMedian          0.2285  1    0.63267
```

```
## working_dataT$centrality             3.0040  1    0.08306 .
## working_dataT$tenureDiversity        0.6821  1    0.40888
## working_dataT$stCongruence         233.5753  1    < 2e-16 ***
## working_dataT$truckFactor            1.0323  1    0.30962
## working_dataT$expertise              0.4305  1    0.51176
## working_dataT$female                 0.0408  1    0.84000
## working_dataT$blauGender             0.4991  1    0.47989
## ---
## 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
loneT <- lmer(working_dataG$lone~log(working_dataG$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)
```

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

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

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

```
## log(working_dataG$totalCommitters)    log(working_dataG$totalcommits)
##                           2.539211                           2.361283
##           working_dataG$projectAge             working_dataG$turnover
##                           1.320130                           1.253354
##         working_dataG$tenureMedian       working_dataG$tenureDiversity
##                           1.063707                           1.046162
##        log(working_dataG$teamSize)          working_dataG$stCongruence
##                           3.077293                           3.078046
##           working_dataG$centrality           working_dataG$truckFactor
##                           1.156810                           1.067097
##            working_dataG$expertise                working_dataG$female
##                           1.069665                           1.036473
##           working_dataG$blauGender
##                           1.632092
```

```
# Applying a Linear Mixed Model using the lmer function, after vif, NO REMOVALS
loneT <- lmer(working_dataG$lone~log(working_dataG$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(loneT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$lone ~ log(working_dataG$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                        8.813260   2.295216   3.840
## log(working_dataG$totalCommitters)   0.214510   0.308961   0.694
## log(working_dataG$totalcommits)      0.184696   0.231809   0.797
## working_dataG$projectAge             0.117693   0.053584   2.196
## working_dataG$turnover               0.869837   1.083180   0.803
## working_dataG$tenureMedian          -0.062020   0.129759  -0.478
## working_dataG$centrality             0.925333   0.533883   1.733
## working_dataG$tenureDiversity        0.076192   0.092256   0.826
## working_dataG$stCongruence         -14.353546   0.939173 -15.283
## working_dataG$truckFactor           -0.173819   0.171079  -1.016
## working_dataG$expertise              0.489398   0.745927   0.656
## working_dataG$female                 0.008265   0.040935   0.202
## working_dataG$blauGender            -2.227819   3.153427  -0.706
##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(summary(loneT), correlation=TRUE)  or
##     vcov(summary(loneT))        if you need it
```

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

```
# Applying anova
Anova(loneT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.4820  1    0.48750
## log(working_dataG$totalcommits)     0.6348  1    0.42559
## working_dataG$projectAge            4.8242  1    0.02806 *
## working_dataG$turnover              0.6449  1    0.42195
## working_dataG$tenureMedian          0.2285  1    0.63267
## working_dataG$centrality            3.0040  1    0.08306 .
## working_dataG$tenureDiversity       0.6821  1    0.40888
## working_dataG$stCongruence        233.5753  1    < 2e-16 ***
## working_dataG$truckFactor           1.0323  1    0.30962
## working_dataG$expertise             0.4305  1    0.51176
## working_dataG$female                0.0408  1    0.84000
## working_dataG$blauGender            0.4991  1    0.47989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula:
## working_dataG$lone ~ log(working_dataG$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
##   1015.6   1064.5   -492.8    985.6      177
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.44898 -0.65995  0.09741  0.78185  2.43141
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept) 0.000    0.000
##  Residual                             9.929    3.151
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##                                    Estimate Std. Error t value
## (Intercept)                        8.813260   2.295216   3.840
## log(working_dataG$totalCommitters) 0.214510   0.308961   0.694
## log(working_dataG$totalcommits)    0.184696   0.231809   0.797
## working_dataG$projectAge           0.117693   0.053584   2.196
## working_dataG$turnover             0.869837   1.083180   0.803
## working_dataG$tenureMedian        -0.062020   0.129759  -0.478
## working_dataG$centrality           0.925333   0.533883   1.733
## working_dataG$tenureDiversity      0.076192   0.092256   0.826
```

```
## working_dataG$stCongruence         -14.353546   0.939173 -15.283
## working_dataG$truckFactor           -0.173819   0.171079  -1.016
## working_dataG$expertise              0.489398   0.745927   0.656
## working_dataG$female                 0.008265   0.040935   0.202
## working_dataG$blauGender            -2.227819   3.153427  -0.706
```

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

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

```
Anova(loneT)
```

```
## Analysis of Deviance Table (Type II Wald chisquare tests)
##
## Response: working_dataG$lone
##                                    Chisq Df Pr(>Chisq)
## log(working_dataG$totalCommitters)  0.4820  1    0.48750
## log(working_dataG$totalcommits)     0.6348  1    0.42559
## working_dataG$projectAge            4.8242  1    0.02806 *
## working_dataG$turnover              0.6449  1    0.42195
## working_dataG$tenureMedian          0.2285  1    0.63267
## working_dataG$centrality            3.0040  1    0.08306 .
## working_dataG$tenureDiversity       0.6821  1    0.40888
## working_dataG$stCongruence        233.5753  1    < 2e-16 ***
## working_dataG$truckFactor           1.0323  1    0.30962
## working_dataG$expertise             0.4305  1    0.51176
## working_dataG$female                0.0408  1    0.84000
## working_dataG$blauGender            0.4991  1    0.47989
## ---
## 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
lone <- lmer(working_data$lone~(1 | working_data$window_idx ), REML=FALSE)
```

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

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

# print result
print(summary(lone))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_data$lone ~ (1 | working_data$window_idx)
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept)  0.00    0.000
##  Residual                             22.93    4.789
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_data$lone ~ (1 | working_data$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_data$window_idx (Intercept)  0.00    0.000
##  Residual                             22.93    4.789
## Number of obs: 192, groups:  working_data$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
sink()
```

Linear Mixed Model using lmer function on only random effect trompenaars

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

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

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

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

# print result
print(summary(loneT))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataT$lone ~ (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept)  0.00    0.000
##  Residual                              22.93    4.789
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataT$lone ~ (1 | working_dataT$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                    Name        Variance Std.Dev.
##  working_dataT$window_idx (Intercept)  0.00    0.000
##  Residual                              22.93    4.789
## Number of obs: 192, groups:  working_dataT$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
```

```
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
sink()
```

Linear Mixed Model using lmer function on only random effect globe

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

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

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

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

# print result
print(summary(loneG))
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataG$lone ~ (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept)  0.00    0.000
##  Residual                              22.93    4.789
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

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

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: working_dataG$lone ~ (1 | working_dataG$window_idx)
##
##      AIC      BIC   logLik deviance df.resid
##   1152.3   1162.1   -573.2   1146.3      189
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7097 -0.8744  0.1697  0.7961  1.6314
##
## Random effects:
##  Groups                   Name        Variance Std.Dev.
##  working_dataG$window_idx (Intercept)  0.00    0.000
##  Residual                              22.93    4.789
## Number of obs: 192, groups:  working_dataG$window_idx, 24
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   8.1875     0.3456   23.69
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
sink()
```