SOM: Analysis of model building blocks usefulness

2023-02-27

Table of Contents

# Edits

# Introduction

Now let us read data in!

# Clearing all  
rm(list=ls())  
  
# Packages:  
library(tidyverse)  
library(readr)  
library(dplyr)  
library(tibble)  
library(forcats)  
library(ggplot2)  
library(rstatix)  
library(stringr)  
library(knitr)  
# library(kableExtra)  
library(stargazer)  
  
#### Reading data in:  
# source("SOM\_loading.R")  
load("SOM.RData")

# Step 1 VS. Step 2

Firstly, data filtering:

df = tc %>% filter((Step == "No Identity, Constant Boundary" | Step == "No Identity, Normal Boundary"),   
 Boundary\_STD == 0, Conformity\_STD == 0, HK\_distribution == "FALSE")  
df %>% count(Step)

## # A tibble: 2 × 2  
## Step n  
## <fct> <int>  
## 1 No Identity, Constant Boundary 5040  
## 2 No Identity, Normal Boundary 5040

Comparison:

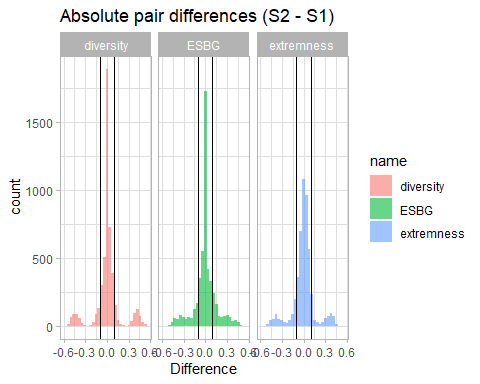
Computing sim-by-sim difference:

# We split data into pairs according all relevant varibles, except STEP:  
dp = df %>% #group\_by(RS, N, Boundary, Conformity) %>%   
 pivot\_longer(cols = diversity:ESBG) %>%   
 pivot\_wider(id\_cols = c(HK\_distribution:Conformity\_STD, name), names\_from = Step, values\_from = value) %>%   
 rename("S1" = 14, "S2" = 15) %>%   
 mutate(  
 Difference = S2 - S1,  
 highDifference = abs(Difference) > 0.1) %>%   
 # Then we compute difference:  
 # summarise(across(diversity:ESBG, ~sd(.x) / sqrt(2), .names = "{.col}\_sd")) %>% ungroup() %>%   
 # And now, also pivoting data for the future graph:  
 # select(ends\_with("\_sd")) %>%   
 mutate(across(N:name, ~factor(.x)))  
  
# We split data into groups according all relevant varibles, except STEP:  
# dg = df %>% group\_by(N, Boundary, Conformity) %>%   
# # Then we compute difference:  
# summarise(across(diversity:ESBG, ~sd(.x), .names = "{.col}\_sd")) %>% ungroup() %>%   
# # And now, also pivoting data for the future graph:  
# # select(ends\_with("\_sd")) %>%   
# pivot\_longer(cols = ends\_with("\_sd")) %>%   
# mutate(across(N:name, ~factor(.x)))  
dg = dp %>% group\_by(N, Boundary, Conformity, name) %>%   
 summarise(Difference\_sd = sd(Difference)) %>%   
 mutate(highDifference = Difference\_sd > 0.1)

## `summarise()` has grouped output by 'N', 'Boundary', 'Conformity'. You can  
## override using the `.groups` argument.

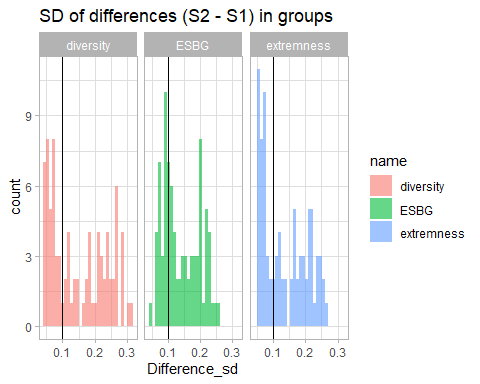
dp %>%   
 ggplot() +  
 aes(x = Difference, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1, -0.1)) +  
 labs(title = "Absolute pair differences (S2 - S1)") +   
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



dg %>%   
 ggplot() +  
 aes(x = Difference\_sd, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1)) +  
 labs(title = "SD of differences (S2 - S1) in groups") +  
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



mp = lm(Difference ~ name+Conformity+Boundary+N, data = dp)  
mpa = lm(abs(Difference) ~ name+Conformity+Boundary+N, data = dp)  
mg = lm(Difference\_sd ~ name+Conformity+Boundary+N, data = dg)  
stargazer(mp, mpa, mg, type = "text")

##   
## ===================================================================================================  
## Dependent variable:   
## -------------------------------------------------------------------------------  
## Difference abs(Difference) Difference\_sd   
## (1) (2) (3)   
## ---------------------------------------------------------------------------------------------------  
## nameESBG -0.004 -0.0002 -0.009\*\*\*   
## (0.003) (0.002) (0.003)   
##   
## nameextremness 0.001 0.001 -0.013\*\*\*   
## (0.003) (0.002) (0.003)   
##   
## Conformity0.8 0.0005 -0.007\*\*\* -0.010\*\*\*   
## (0.003) (0.002) (0.003)   
##   
## Boundary0.11 -0.0004 0.004 0.005   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.12 -0.001 0.006 0.009   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.13 0.002 0.011\* 0.014\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.14 -0.002 0.014\*\* 0.020\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.15 0.001 0.019\*\*\* 0.027\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.16 0.002 0.023\*\*\* 0.033\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.17 0.001 0.026\*\*\* 0.036\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.18 0.009 0.037\*\*\* 0.057\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.19 0.016\* 0.055\*\*\* 0.084\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.2 0.013 0.073\*\*\* 0.116\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.21 0.012 0.093\*\*\* 0.140\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.22 0.015\* 0.110\*\*\* 0.161\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.23 0.023\*\*\* 0.133\*\*\* 0.183\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.24 0.005 0.146\*\*\* 0.200\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.25 -0.005 0.126\*\*\* 0.188\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.26 -0.004 0.106\*\*\* 0.178\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.27 -0.002 0.076\*\*\* 0.154\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.28 -0.021\*\* 0.040\*\*\* 0.114\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.29 -0.034\*\*\* 0.016\*\*\* 0.081\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## Boundary0.3 -0.015\* -0.006 0.044\*\*\*   
## (0.008) (0.006) (0.008)   
##   
## N101 -0.015\*\*\* -0.005\*\* -0.004   
## (0.003) (0.002) (0.003)   
##   
## Constant 0.010 0.049\*\*\* 0.069\*\*\*   
## (0.006) (0.005) (0.006)   
##   
## ---------------------------------------------------------------------------------------------------  
## Observations 15,120 15,120 252   
## R2 0.009 0.141 0.929   
## Adjusted R2 0.007 0.140 0.922   
## Residual Std. Error 0.159 (df = 15095) 0.118 (df = 15095) 0.020 (df = 227)   
## F Statistic 5.485\*\*\* (df = 24; 15095) 103.604\*\*\* (df = 24; 15095) 124.636\*\*\* (df = 24; 227)  
## ===================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Tables with extremes

dp %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme absolute differences in pairs comparison")

Table with fraction of extreme absolute differences in pairs comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.2333333 |
| ESBG | 0.3162698 |
| extremness | 0.2549603 |

dg %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme SD of within-groups comparison")

Table with fraction of extreme SD of within-groups comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.5952381 |
| ESBG | 0.6785714 |
| extremness | 0.5833333 |

# Step 2 VS. Step 3

Firstly, data filtering:

df = tc %>% filter((Step == "Constant SPIRO, Normal Boundary" | Step == "No Identity, Normal Boundary"), SPIRO\_Mean == 0.25)  
df %>% count(Step)

## # A tibble: 2 × 2  
## Step n  
## <fct> <int>  
## 1 No Identity, Normal Boundary 80640  
## 2 Constant SPIRO, Normal Boundary 80640

Comparison:

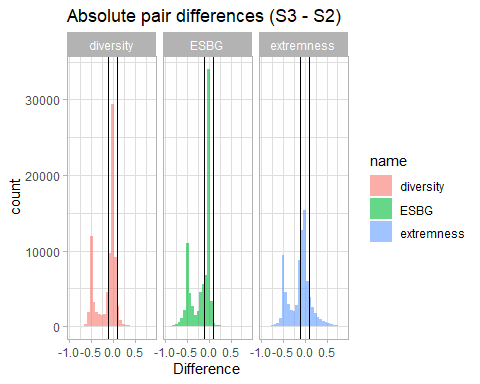
Computing sim-by-sim difference:

# We split data into pairs according all relevant varibles, except STEP:  
# dp = df %>% group\_by(RS, HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD) %>%   
# # Then we compute difference:  
# summarise(across(diversity:ESBG, ~sd(.x) / sqrt(2), .names = "{.col}\_sd")) %>% ungroup() %>%   
# # And now, also pivoting data for the future graph:  
# # select(ends\_with("\_sd")) %>%   
# pivot\_longer(cols = ends\_with("\_sd")) %>%   
# mutate(across(N:name, ~factor(.x)))  
dp = df %>%   
 pivot\_longer(cols = diversity:ESBG, values\_drop\_na = T) %>%   
 pivot\_wider(id\_cols = c(RS, HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, name),   
 names\_from = Step, values\_from = value) %>%   
 rename("S1" = 9, "S2" = 10) %>%   
 mutate(  
 Difference = S2 - S1,  
 highDifference = abs(Difference) > 0.1,  
 asymetryDifference = Difference < -0.25) %>%   
 mutate(across(N:name, ~factor(.x)))  
  
# We split data into groups according all relevant varibles, except STEP:  
# dg = df %>% group\_by(HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD) %>%   
# # Then we compute difference:  
# summarise(across(diversity:ESBG, ~sd(.x), .names = "{.col}\_sd")) %>% ungroup() %>%   
# # And now, also pivoting data for the future graph:  
# # select(ends\_with("\_sd")) %>%   
# pivot\_longer(cols = ends\_with("\_sd")) %>%   
# mutate(across(N:name, ~factor(.x)))  
dg = dp %>% group\_by(HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, name) %>%   
 summarise(Difference\_sd = sd(Difference)) %>%   
 mutate(highDifference = Difference\_sd > 0.1)

## `summarise()` has grouped output by 'HK\_distribution', 'N', 'Boundary',  
## 'Boundary\_STD', 'Conformity', 'Conformity\_STD'. You can override using the  
## `.groups` argument.

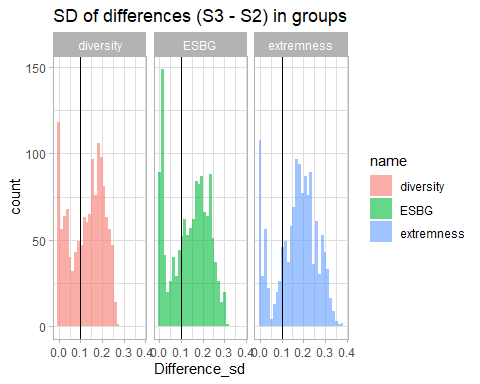
dp %>%   
 ggplot() +  
 aes(x = Difference, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1, -0.1)) +  
 labs(title = "Absolute pair differences (S3 - S2)") +   
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



dg %>%   
 ggplot() +  
 aes(x = Difference\_sd, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1)) +  
 labs(title = "SD of differences (S3 - S2) in groups") +  
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



mp = lm(Difference ~ name+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dp)  
mpa = lm(abs(Difference) ~ name+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dp)  
mg = lm(Difference\_sd ~ name+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dg)  
stargazer(mp, mpa, mg, type = "text")

##   
## ============================================================================================================  
## Dependent variable:   
## ----------------------------------------------------------------------------------------  
## Difference abs(Difference) Difference\_sd   
## (1) (2) (3)   
## ------------------------------------------------------------------------------------------------------------  
## nameESBG -0.036\*\*\* 0.016\*\*\* 0.010\*\*\*   
## (0.001) (0.001) (0.003)   
##   
## nameextremness 0.020\*\*\* 0.037\*\*\* 0.042\*\*\*   
## (0.001) (0.001) (0.003)   
##   
## Conformity0.8 -0.028\*\*\* 0.042\*\*\* 0.018\*\*\*   
## (0.001) (0.001) (0.002)   
##   
## Conformity\_STD0.1 0.012\*\*\* -0.012\*\*\* 0.006\*\*\*   
## (0.001) (0.001) (0.002)   
##   
## Boundary0.11 -0.004 0.007\*\*\* 0.009   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.12 -0.020\*\*\* 0.022\*\*\* 0.013\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.13 -0.028\*\*\* 0.023\*\*\* 0.012\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.14 -0.031\*\*\* 0.027\*\*\* 0.013\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.15 -0.029\*\*\* 0.027\*\*\* 0.015\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.16 -0.038\*\*\* 0.033\*\*\* 0.017\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.17 -0.033\*\*\* 0.029\*\*\* 0.019\*\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.18 -0.035\*\*\* 0.030\*\*\* 0.020\*\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.19 -0.043\*\*\* 0.034\*\*\* 0.021\*\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.2 -0.055\*\*\* 0.049\*\*\* 0.020\*\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.21 -0.072\*\*\* 0.065\*\*\* 0.015\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.22 -0.084\*\*\* 0.075\*\*\* 0.011   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.23 -0.102\*\*\* 0.089\*\*\* 0.005   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.24 -0.112\*\*\* 0.098\*\*\* 0.007   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.25 -0.117\*\*\* 0.104\*\*\* -0.001   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.26 -0.113\*\*\* 0.099\*\*\* -0.001   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.27 -0.103\*\*\* 0.086\*\*\* -0.003   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.28 -0.091\*\*\* 0.072\*\*\* -0.006   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.29 -0.073\*\*\* 0.053\*\*\* -0.009   
## (0.003) (0.002) (0.007)   
##   
## Boundary0.3 -0.060\*\*\* 0.036\*\*\* -0.018\*\*\*   
## (0.003) (0.002) (0.007)   
##   
## Boundary\_STD0.05 -0.039\*\*\* 0.047\*\*\* 0.108\*\*\*   
## (0.001) (0.001) (0.003)   
##   
## Boundary\_STD0.1 0.128\*\*\* -0.062\*\*\* 0.133\*\*\*   
## (0.001) (0.001) (0.003)   
##   
## Boundary\_STD0.15 0.209\*\*\* -0.109\*\*\* 0.087\*\*\*   
## (0.001) (0.001) (0.003)   
##   
## N101 -0.008\*\*\* 0.007\*\*\* -0.003   
## (0.001) (0.001) (0.002)   
##   
## HK\_distribution -0.055\*\*\* 0.049\*\*\* -0.026\*\*\*   
## (0.001) (0.001) (0.002)   
##   
## Constant -0.103\*\*\* 0.098\*\*\* 0.042\*\*\*   
## (0.002) (0.002) (0.006)   
##   
## ------------------------------------------------------------------------------------------------------------  
## Observations 241,920 241,920 4,032   
## R2 0.244 0.152 0.421   
## Adjusted R2 0.244 0.152 0.417   
## Residual Std. Error 0.198 (df = 241890) 0.180 (df = 241890) 0.066 (df = 4002)   
## F Statistic 2,696.208\*\*\* (df = 29; 241890) 1,497.269\*\*\* (df = 29; 241890) 100.310\*\*\* (df = 29; 4002)  
## ============================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Tables with extremes

dp %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(asymetryDifference)) %>%   
 kable(caption = "Table with fraction of extreme negative (asymetric) absolute differences in pairs comparison")

Table with fraction of extreme negative (asymetric) absolute differences in pairs comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.2651290 |
| ESBG | 0.3010541 |
| extremness | 0.2675099 |

dp %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme absolute differences in pairs comparison")

Table with fraction of extreme absolute differences in pairs comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.3848586 |
| ESBG | 0.4413442 |
| extremness | 0.5521577 |

dg %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme SD of within-groups comparison")

Table with fraction of extreme SD of within-groups comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.6450893 |
| ESBG | 0.6688988 |
| extremness | 0.7879464 |

# Step 4 VS. Step 3

Firstly, data filtering:

df = tc %>% filter((Step == "Constant SPIRO, Normal Boundary" | Step == "Normal SPIRO, Normal Boundary"), SPIRO\_STD == 0)  
df %>% count(Step)

## # A tibble: 2 × 2  
## Step n  
## <fct> <int>  
## 1 Constant SPIRO, Normal Boundary 483840  
## 2 Normal SPIRO, Normal Boundary 483840

Comparison:

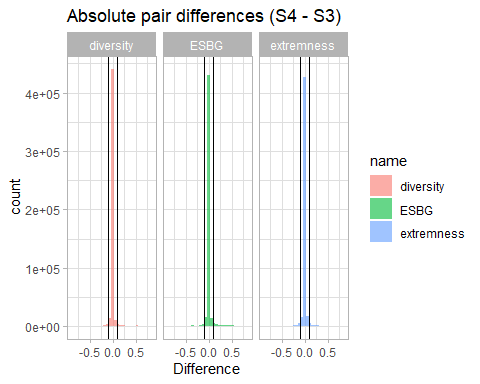
Computing sim-by-sim difference:

# We split data into pairs according all relevant varibles, except STEP:  
# dp = df %>% group\_by(RS, HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, SPIRO\_Mean) %>%   
# # Then we compute difference:  
# summarise(across(diversity:ESBG, ~sd(.x) / sqrt(2), .names = "{.col}\_sd")) %>% ungroup() %>%   
# # And now, also pivoting data for the future graph:  
# # select(ends\_with("\_sd")) %>%   
# pivot\_longer(cols = ends\_with("\_sd")) %>%   
# mutate(across(N:name, ~factor(.x)))  
#   
# # We split data into groups according all relevant varibles, except STEP:  
# dg = df %>% group\_by(HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, SPIRO\_Mean) %>%   
# # Then we compute difference:  
# summarise(across(diversity:ESBG, ~sd(.x), .names = "{.col}\_sd")) %>% ungroup() %>%   
# # And now, also pivoting data for the future graph:  
# # select(ends\_with("\_sd")) %>%   
# pivot\_longer(cols = ends\_with("\_sd")) %>%   
# mutate(across(N:name, ~factor(.x)))  
dp = df %>%   
 pivot\_longer(cols = diversity:ESBG, values\_drop\_na = T) %>%   
 pivot\_wider(id\_cols = c(RS, HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, SPIRO\_Mean, name),   
 names\_from = Step, values\_from = value) %>%   
 rename("S1" = 10, "S2" = 11) %>%   
 mutate(  
 Difference = S2 - S1,  
 highDifference = abs(Difference) > 0.1) %>%   
 mutate(across(N:name, ~factor(.x)))  
  
# We split data into groups according all relevant varibles, except STEP:  
dg = dp %>% group\_by(HK\_distribution, N, Boundary, Boundary\_STD, Conformity, Conformity\_STD, SPIRO\_Mean, name) %>%   
 summarise(Difference\_sd = sd(Difference)) %>%   
 mutate(highDifference = Difference\_sd > 0.1) %>% ungroup()

## `summarise()` has grouped output by 'HK\_distribution', 'N', 'Boundary',  
## 'Boundary\_STD', 'Conformity', 'Conformity\_STD', 'SPIRO\_Mean'. You can override  
## using the `.groups` argument.

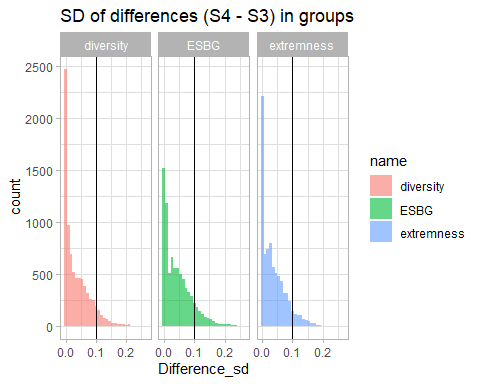
dp %>%   
 ggplot() +  
 aes(x = Difference, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1, -0.1)) +  
 labs(title = "Absolute pair differences (S4 - S3)") +   
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



dg %>%   
 ggplot() +  
 aes(x = Difference\_sd, fill = name) +  
 facet\_wrap(~name) +  
 geom\_histogram(alpha = 0.6) +  
 geom\_vline(xintercept = c( 0.1)) +  
 labs(title = "SD of differences (S4 - S3) in groups") +  
 theme\_light()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



mp = lm(Difference ~ name+SPIRO\_Mean+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dp)  
mpa = lm(abs(Difference) ~ name+SPIRO\_Mean+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dp)  
mg = lm(Difference\_sd ~ name+SPIRO\_Mean+Conformity+Conformity\_STD+Boundary+Boundary\_STD+N+HK\_distribution, data = dg)  
stargazer(mp, mpa, mg, type = "text" )

##   
## =============================================================================================================  
## Dependent variable:   
## -----------------------------------------------------------------------------------------  
## Difference abs(Difference) Difference\_sd   
## (1) (2) (3)   
## -------------------------------------------------------------------------------------------------------------  
## nameESBG 0.003\*\*\* 0.004\*\*\* 0.009\*\*\*   
## (0.0001) (0.0001) (0.0005)   
##   
## nameextremness 0.0004\*\*\* 0.002\*\*\* 0.001\*\*\*   
## (0.0001) (0.0001) (0.0005)   
##   
## SPIRO\_Mean0.37 0.011\*\*\* 0.015\*\*\* 0.043\*\*\*   
## (0.0002) (0.0002) (0.001)   
##   
## SPIRO\_Mean0.49 0.013\*\*\* 0.017\*\*\* 0.044\*\*\*   
## (0.0002) (0.0002) (0.001)   
##   
## SPIRO\_Mean0.61 -0.004\*\*\* 0.009\*\*\* 0.024\*\*\*   
## (0.0002) (0.0002) (0.001)   
##   
## SPIRO\_Mean0.73 0.001\*\*\* 0.019\*\*\* 0.034\*\*\*   
## (0.0002) (0.0002) (0.001)   
##   
## SPIRO\_Mean0.85 -0.001\*\*\* 0.011\*\*\* 0.016\*\*\*   
## (0.0002) (0.0002) (0.001)   
##   
## Conformity0.8 -0.001\*\*\* 0.002\*\*\* 0.005\*\*\*   
## (0.0001) (0.0001) (0.0004)   
##   
## Conformity\_STD0.1 -0.0002\*\* -0.0003\*\*\* 0.0001   
## (0.0001) (0.0001) (0.0004)   
##   
## Boundary0.11 -0.0002 0.0001 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.12 0.001\*\* -0.0005 -0.00001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.13 -0.0001 -0.001\*\* -0.00000   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.14 0.001\*\* 0.0001 0.002   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.15 0.0004 -0.0005 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.16 0.0005 -0.001\* 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.17 0.001\*\*\* -0.001\*\* 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.18 0.001\* -0.001\*\*\* 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.19 0.001\*\*\* -0.001\*\*\* 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.2 0.001\*\*\* -0.001\*\*\* 0.001   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.21 0.002\*\*\* -0.001\* 0.003\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.22 0.001\*\*\* -0.001\*\*\* 0.002\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.23 0.001\*\*\* -0.001\*\*\* 0.002   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.24 0.002\*\*\* -0.0004 0.005\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.25 0.003\*\*\* 0.001\*\*\* 0.006\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.26 0.004\*\*\* 0.002\*\*\* 0.009\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.27 0.004\*\*\* 0.002\*\*\* 0.011\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.28 0.004\*\*\* 0.002\*\*\* 0.010\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.29 0.005\*\*\* 0.002\*\*\* 0.012\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary0.3 0.006\*\*\* 0.003\*\*\* 0.013\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## Boundary\_STD0.05 0.0004\*\*\* 0.004\*\*\* 0.014\*\*\*   
## (0.0001) (0.0001) (0.001)   
##   
## Boundary\_STD0.1 0.005\*\*\* 0.020\*\*\* 0.054\*\*\*   
## (0.0001) (0.0001) (0.001)   
##   
## Boundary\_STD0.15 0.006\*\*\* 0.022\*\*\* 0.057\*\*\*   
## (0.0001) (0.0001) (0.001)   
##   
## N101 -0.0004\*\*\* 0.004\*\*\* 0.009\*\*\*   
## (0.0001) (0.0001) (0.0004)   
##   
## HK\_distribution -0.0005\*\*\* 0.003\*\*\* 0.004\*\*\*   
## (0.0001) (0.0001) (0.0004)   
##   
## Constant -0.005\*\*\* -0.014\*\*\* -0.034\*\*\*   
## (0.0003) (0.0003) (0.001)   
##   
## -------------------------------------------------------------------------------------------------------------  
## Observations 1,451,520 1,451,520 24,192   
## R2 0.015 0.044 0.510   
## Adjusted R2 0.015 0.044 0.509   
## Residual Std. Error 0.059 (df = 1451485) 0.056 (df = 1451485) 0.030 (df = 24157)   
## F Statistic 637.678\*\*\* (df = 34; 1451485) 1,947.352\*\*\* (df = 34; 1451485) 738.023\*\*\* (df = 34; 24157)  
## =============================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Tables with extremes

dp %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme absolute differences in pairs comparison")

Table with fraction of extreme absolute differences in pairs comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.0345672 |
| ESBG | 0.0451306 |
| extremness | 0.0403191 |

dg %>% group\_by(name) %>% summarise(fractionOfHighDifferences = mean(highDifference)) %>%   
 kable(caption = "Table with fraction of extreme SD of within-groups comparison")

Table with fraction of extreme SD of within-groups comparison

| name | fractionOfHighDifferences |
| --- | --- |
| diversity | 0.0879216 |
| ESBG | 0.1256200 |
| extremness | 0.0825893 |