* [The data](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html" \l "the-data" \t "_blank)
* [Standardize](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#standardize)
* [Effect of Standardization](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#effect-of-standardization)
  + [At a general level](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#at-a-general-level)
  + [At a participant level](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#at-a-participant-level)
  + [Distribution](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#distribution)
* [Correlation](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#correlation)
* [Test](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#test)
* [Conclusion](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#conclusion)
* [Credits](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#credits)
* [Previous blogposts](https://neuropsychology.github.io/psycho.R/2018/07/14/standardize_grouped_df.html#previous-blogposts)

To make sense of their data and effects, psychologists often standardize (Z-score) their variables. However, in repeated-measures designs, there are three ways of standardizing data:

* **Variable-wise**: The most common method. A simple scaling and reducing of each variable by their mean and SD.
* **Participant-wise**: Variables are standardized “within” each participant, *i.e.*, for each participant, by the participant’s mean and SD.
* **Full**: Participant-wise first and then re-standardizing variable-wise.

Unfortunately, the method used is often not explicitly stated. This is an issue as these methods can generate important discrepancies that contribute to the reproducibility crisis of psychological science.

In the following, we will see how to perform those methods and look for differences.

**The data**

We will take a dataset in which participants were exposed to negative pictures and had to rate their emotions (**valence**) and the amount of memories associated with the picture (**autobiographical link**). One could make the hypothesis that for young participants with no context of war or violence, the most negative pictures (mutilations) are less related to memories than less negative pictures (involving for example car crashes or sick people). In other words, **we expect a positive relationship between valence** (with high values corresponding to less negativity) **and autobiographical link**.

Let’s have a look at the data, averaged by participants:

library(tidyverse)

library(psycho)

df <- psycho::emotion %>% # Load the dataset from the psycho package

filter(Emotion\_Condition == "Negative") # Discard neutral pictures

df %>%

group\_by(Participant\_ID) %>%

summarise(n\_Trials = n(),

Valence\_Mean = mean(Subjective\_Valence, na.rm=TRUE),

Valence\_SD = sd(Subjective\_Valence, na.rm=TRUE),

Autobiographical\_Link\_Mean = mean(Autobiographical\_Link, na.rm=TRUE),

Autobiographical\_Link\_SD = sd(Autobiographical\_Link, na.rm=TRUE)) %>%

mutate(Valence = paste(Valence\_Mean, Valence\_SD, sep=" +- "),

Autobiographical\_Link = paste(Autobiographical\_Link\_Mean, Autobiographical\_Link\_SD, sep=" +- ")) %>%

select(-ends\_with("SD"), -ends\_with("Mean"))

| **Participant\_ID** | **n\_Trials** | **Valence** | **Autobiographical\_Link** |
| --- | --- | --- | --- |
| 10S | 24 | -58.07 +- 42.59 | 49.88 +- 29.60 |
| 11S | 24 | -73.22 +- 37.01 | 0.94 +- 3.46 |
| 12S | 24 | -57.53 +- 26.56 | 30.24 +- 27.87 |
| 13S | 24 | -63.22 +- 23.72 | 27.86 +- 35.81 |
| 14S | 24 | -56.60 +- 26.47 | 3.31 +- 11.20 |
| 15S | 24 | -60.59 +- 33.71 | 17.64 +- 17.36 |
| 16S | 24 | -46.12 +- 24.88 | 15.33 +- 16.57 |
| 17S | 24 | -1.54 +- 4.98 | 0.13 +- 0.64 |
| 18S | 24 | -67.23 +- 34.98 | 22.71 +- 20.07 |
| 19S | 24 | -59.61 +- 33.22 | 28.37 +- 12.55 |

As we can see from the means and SDs, there is a lot of variability **between and within** participants.

**Standardize**

We will create three dataframes standardized with each of the three techniques.

Z\_VarWise <- df %>%

standardize()

Z\_ParWise <- df %>%

group\_by(Participant\_ID) %>%

standardize()

Z\_Full <- df %>%

group\_by(Participant\_ID) %>%

standardize() %>%

ungroup() %>%

standardize()

**Effect of Standardization**

Let’s see how these three standardization techniques affected the **Valence** variable.

**At a general level**

# Create convenient function

print\_summary <- function(data){

paste(deparse(substitute(data)), ":",

format\_digit(mean(data[["Subjective\_Valence"]])),

"+-",

format\_digit(sd(data[["Subjective\_Valence"]])),

"[", format\_digit(min(data[["Subjective\_Valence"]])),

",", format\_digit(max(data[["Subjective\_Valence"]])),

"]")

}

# Check the results

print\_summary(Z\_VarWise)

[1] "Z\_VarWise : 0 +- 1.00 [ -1.18 , 3.10 ]"

print\_summary(Z\_ParWise)

[1] "Z\_ParWise : 0 +- 0.98 [ -2.93 , 3.29 ]"

print\_summary(Z\_Full)

[1] "Z\_Full : 0 +- 1.00 [ -2.99 , 3.36 ]"

**At a participant level**

# Create convenient function

print\_participants <- function(data){

data %>%

group\_by(Participant\_ID) %>%

summarise(Mean = mean(Subjective\_Valence),

SD = sd(Subjective\_Valence)) %>%

mutate\_if(is.numeric, round, 2) %>%

head(5)

}

# Check the results

print\_participants(Z\_VarWise)

# A tibble: 5 x 3

Participant\_ID Mean SD

1 10S -0.0500 1.15

2 11S -0.460 1.00

3 12S -0.0300 0.720

4 13S -0.190 0.640

5 14S -0.0100 0.710

print\_participants(Z\_ParWise)

# A tibble: 5 x 3

Participant\_ID Mean SD

1 10S 0. 1.

2 11S 0. 1.

3 12S 0. 1.

4 13S 0. 1.

5 14S 0. 1.

print\_participants(Z\_Full)

# A tibble: 5 x 3

Participant\_ID Mean SD

1 10S 0. 1.02

2 11S 0. 1.02

3 12S 0. 1.02

4 13S 0. 1.02

5 14S 0. 1.02

**Distribution**

data.frame(VarWise = Z\_VarWise$Subjective\_Valence,

ParWise = Z\_ParWise$Subjective\_Valence,

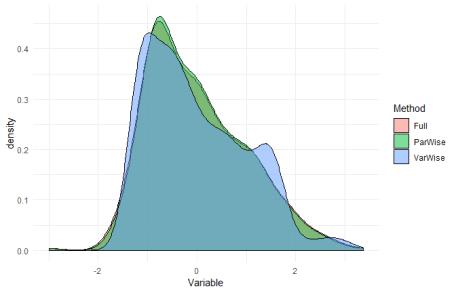
Full = Z\_Full$Subjective\_Valence) %>%

gather(Method, Variable) %>%

ggplot(aes(x=Variable, fill=Method)) +

geom\_density(alpha=0.5) +

theme\_minimal()



The distributions appear to be similar…

**Correlation**

Let’s do a **correlation** between the **variable-wise and participant-wise methods**.

psycho::bayes\_cor.test(Z\_VarWise$Subjective\_Valence, Z\_ParWise$Subjective\_Valence)

Results of the Bayesian correlation indicate moderate evidence (BF = 8.65) in favour of an absence of a negative association between Z\_VarWise$Subjective\_Valence and Z\_ParWise$Subjective\_Valence (r = -0.015, MAD = 0.047, 90% CI [-0.096, 0.057]). The correlation can be considered as small or very small with respective probabilities of 3.46% and 59.43%.

data.frame(Original = df$Subjective\_Valence,

VarWise = Z\_VarWise$Subjective\_Valence,

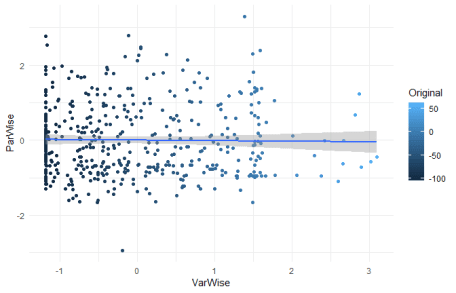
ParWise = Z\_ParWise$Subjective\_Valence) %>%

ggplot(aes(x=VarWise, y=ParWise, colour=Original)) +

geom\_point() +

geom\_smooth(method="lm") +

theme\_minimal()



**While the three standardization methods roughly present the same characteristics at a general level (mean 0 and SD 1) and a similar distribution, their values are very different and completely uncorrelated!**

**Test**

Let’s now answer to the original question by investigating the **linear relationship between valence and autobiographical link**. We can do this by running a mixed model with participants entered as random effects.

# Convenient function

print\_model <- function(data){

type\_name <- deparse(substitute(data))

lmerTest::lmer(Subjective\_Valence ~ Autobiographical\_Link + (1|Participant\_ID), data=data) %>%

psycho::analyze(CI=NULL) %>%

summary() %>%

filter(Variable == "Autobiographical\_Link") %>%

mutate(Type = type\_name,

Coef = round(Coef, 2),

p = format\_p(p)) %>%

select(Type, Coef, p)

}

# Run the model on all datasets

rbind(print\_model(df),

print\_model(Z\_VarWise),

print\_model(Z\_ParWise),

print\_model(Z\_Full))

Type Coef p

1 df 0.09 > .1

2 Z\_VarWise 0.07 > .1

3 Z\_ParWise 0.08 = 0.08°

4 Z\_Full 0.08 = 0.08°

As we can see, in our case, using **participant-wise standardization resulted in a significant (at *p* = .1) effect**! But keep in mind that *this is not always the case*. In can be the contrary, or generate very similar results. **No method is better or more justified, and its choice depends on the specific case, context, data and goal**.

**Conclusion**

1. **Standardization can be useful in *some* cases and should be justified**
2. **Variable and Participant-wise standardization methods produce “in appearance” similar data**
3. **Variable and Participant-wise standardization can lead to different and uncorrelated results**
4. **The choice of the method can strongly influence the results and thus, should be explicitly stated**

We showed here yet another way of **sneakily tweaking the data** that can change the results. **To prevent its use for *p*-hacking, we can only support the generalization of open-data, open-analysis and preregistration**.