

COVIDiSTRESS Global Survey Data

The file *COVIDiSTRESS_April_27_clean.csv* is the cleaned version of the COVIDiSTRESS data from the April 27th 2020. At the present moment the cleaned data contains $N = 115\,213$ participants. In 50 countries, more than 100 participants took part in the survey. The data is still being collected worldwide (see <https://osf.io/z39us/>).

The raw data was cleaned in the following way:

1. All participants who did not give their consent and who were younger than 18 were filtered out.
2. There were some language-specific issues:
 - a. The variable “Country” was mixed up for participants who took part in the survey in Bulgarian language between March 28th 2020 and April 8th 2020. The data was recoded to account for this error.
 - b. The variable “Country” was mixed up for participants who took part in the survey in Afrikaans language before April 7th. The data was recoded to account for this error.
3. Cases from all countries recorded before the official launch date (March 30th 2020) were excluded, except for Denmark and Kosovo, where the study started earlier.
4. The levels of the variable “Dem_maritalstatus” were mixed up in all languages apart from English. They values were recoded to fix this error.
5. There were some uninformative responses in variables “Dem_maritalstatus” (response “5”), “Dem_edu” and “Dem_edu_mom” (response “1”). These responses were recoded to “Uninformative response”.
6. The text values in variables “AD_gain” and “AD_loss” were shortened.

7. The text values in variables “Scale_PSS_10_UCLA_X”, “OECD_X”, “Corona_concerns_X”, “Compliance_X”, “BFF_X”, “SPS_X”, “Expl_Coping_X”, “Expl_media_X” were recoded to numeric values for further statistical analyses.
8. The text values in variables “Expl_Distress_X” were recoded to numeric values. Please, note that 99 means “Does not apply to my current situation”.
9. The text values in variable “Trust_countrymeasure” were recoded to numeric values. Please, note that values 0 and 10 signify that measures are perceived as inappropriate (too little or too much), whereas 5 signifies that the measures are perceived as appropriate
10. Three variables “Scale_PSS_10_UCLA_11”, “Scale_PSS_10_UCLA_12” and “Scale_PSS_10_UCLA_13” were renamed to “Scale_Lon_1”, “Scale_Lon_2” and “Scale_Lon_3”, because they belong to a separate scale.
11. Composite scores were created for:
 - a. PSS
 - b. Perceived loneliness scale
 - c. BFF
 - d. SPS10
12. All new line characters in “Expl_Distress_txt”, “Expl_coping_txt” and “Final_open” were replaced by “ “. Also, all “;” characters in “Expl_Distress_txt”, “Expl_coping_txt” and “Final_open” were replaced by “|”. This was done in order to preserve the formatting of the .csv file.

For detailed cleaning procedures, please take a look at the R cleaning notebook.