# P-value in Psychology. Codebook

#### Distribution of p-values reported in eight major psychology journals in 1985-2013

The original file with the data was imported from [here](https://osf.io/gdr4q).

The procedure of data extraction is described by [Nuijten et al. (2016)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101263/), including the usage of the automated statcheck procedure: [Extract Statistics from Articles and Recompute P-Values](https://CRAN.R-project.org/package=statcheck).

Further analysis is conducted by [Hartgerin et al. (2016)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4830257/)

## 

The original dataset contains 19 variables:

## N Source Statistic df1 df2 Test\_Comparison Value Reported\_Comparison  
## 1 1 1 F 1 84 = 71,27 <  
## 2 2 1 F 2 84 = 12,5 <  
## 3 3 1 F 1 84 = 59,31 <  
## 4 4 1 F 1 78 = 634,8 <  
## 5 5 1 F 1 78 = 421,58 <  
## 6 6 1 F 2 78 = 13,38 <  
## Reported\_P\_Value Computed Raw Error DecisionError  
## 1 0,001 7,87E-13 F(1, 84) = 71.27, p < .001 FALSE FALSE  
## 2 0,001 1,77E-05 F(2, 84) = 12.50, p < .001 FALSE FALSE  
## 3 0,001 2,40E-11 F(1, 84) = 59.31, p < .001 FALSE FALSE  
## 4 0,001 3,20E-39 F(1, 78) = 634.80, p < .001 FALSE FALSE  
## 5 0,001 3,44E-33 F(1, 78) = 421.58, p < .001 FALSE FALSE  
## 6 0,001 1,01E-05 F(2, 78) = 13.38, p < .001 FALSE FALSE  
## OneTail OneTailedInTxt CopyPaste APAfactor journals\_jour\_ years\_y\_........  
## 1 FALSE FALSE FALSE 1 DP 1985,,,  
## 2 FALSE FALSE FALSE 1 DP 1985,,,  
## 3 FALSE FALSE FALSE 1 DP 1985,,,  
## 4 FALSE FALSE FALSE 1 DP 1985,,,  
## 5 FALSE FALSE FALSE 1 DP 1985,,,  
## 6 FALSE FALSE FALSE 1 DP 1985,,,

Out of them only 5 variables are used in the project [**new names used in the project are given in bold**]:

* N [**n**]- order number of a record
* Reported\_Comparison [**comparison**] - the way how p-value was reported (“=” means as an exact number, “>” or “<” means a range above or below Reported\_P\_Value)
* Reported\_P\_Value [**p\_report**] - p-value reported (or a threshold for it)
* Computed [**p\_compute**] - p-value recalculated based on the reported statistics value and degrees of freedom (recalculation is performed by [Nuijten et al. (2016)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101263/) using the automated procedure [statcheck](https://CRAN.R-project.org/package=statcheck)
* Years\_y\_ [**year**] - a year of a publication

*Other variables not used in the project:*

* Source - order number of a paper, from which data was extracted
* Statistic, Value, df1, df2 - type and value of statistics reported, relevant degrees of freedom
* Test\_Comparison - the way how statistics value was reported (“=” means as an exact number, “>” or “<” means a range above or below Value)
* Raw - raw data extracted from a paper
* Error, DecisionError, OneTail, OneTailedInTxt and CopyPaste - tests generated by [statcheck](https://CRAN.R-project.org/package=statcheck) but not used in this project
* APAfactor and journals\_jour\_ - the relevant journal’s characteristics

After data cleaning and formatting the dataset for the project contains 246476 rows and looks as follows:

## n comparison p\_report p\_compute year  
## 1 1 < 0.001 7.87e-13 1985  
## 2 2 < 0.001 1.77e-05 1985  
## 3 3 < 0.001 2.40e-11 1985  
## 4 4 < 0.001 3.20e-39 1985  
## 5 5 < 0.001 3.44e-33 1985  
## 6 6 < 0.001 1.01e-05 1985