

# How to run and interpret assumptions for initial models

Montserrat Valdivia

## 1. How to run the assumptions for initial models

### 1.1 Download the Assumptions for original models.R file

Make sure to download the file to the same directory where you run Victoria's code.

### 1.2. Run Victoria's code

You can run the seismic\_setup.R file, we will need the data and the models. Particularly we need the cleaned and filter data dat\_new to run the assumptions function.

### 1.3. Copy and paste the following code to run the assumptions

```
# Make sure you have the same working directory
getwd()
# Call in the function and the model specification
source("Assumptions for original models.R")

# Run the function specifying your data and your model
# model_1 uses ethnicity_cat
# model_2 uses urm
# the cleaned data from Victoria's code
assumptions(clean_data = dat_new, model = model_1)
assumptions(clean_data = dat_new, model = model_2)
```

The code should generate two folders, one for assumptions of the first model and the second for assumptions on the second model.

## 2. Interpretation

Inside the folders with assumptions you will find the following graphs, an excel file, and an RData file. The RData file contains all residuals (i.e., level 1 and level 2 residuals and their standardized residuals) in case you want to use them and explore more in depth.

### 2.1. Normality

Use the “Level1\_ResidualsvsFitted\_” graphs. There are as many graphs as courses were analyzed. Residuals should overlap the diagonal line in the qq-plot.

In the excel file, there is a Shapiro-Wilk test for the residuals to test the null hypothesis that residuals distribute normally. We want to retain the null hypothesis, therefore,  $p > .05$  to assume normality of the distribution of residuals.

### 2.2. Linearity

Scatterplot of standardized residuals vs fitted values OR scatterplots of residuals and predictors (i.e., female, ethnicity, etc.). We expect that the observations in these scatterplots do not show a pattern that deviates from a linear relationship, in other words, residuals are distributed randomly.

### 2.3. Homoscedasticity

Scatterplot of standardized residuals vs fitted values. We expect that for each fitted value in the X-axis, the residuals variability is similar. If the plot looks like a rombus or a funnel then equal variance cannot be assumed.