**GLMM**

This model seeks to find the relationship between amount of fact-checks performed and the variables taken from the Likert scale “concern for the war” : anxiety, importance, important for others.

Distribution of the response variable:

Chart, histogram

Description automatically generated

Because this is count data, we have to use a generalized linear model.

Additionally, the data was taken from many individuals in different countries. This means that observations of individuals within a country group could be correlated. Therefore, we have to use a mixed model to account for the hierarchical structure of the data that violates the independent observations assumption.

I am using the R package [glmmtmb](https://cran.r-project.org/web/packages/glmmTMB/glmmTMB.pdf) to run this model, and the package [DHARMa](https://cran.r-project.org/web/packages/DHARMa/DHARMa.pdf) for residual diagnostics.

Because of the nature of the count data, first I run a Poisson GLMM. The overdispersion test shows that the ratio is significantly higher than 1, meaning there is over-dispersion.

I move on to a negative binomial

Summary:

Text

Description automatically generatedGraphical user interface, text

Description automatically generated

Diagnostics:

* Test for overdispersion: in this case the ratio is close to one but insignificant, so it doesn’t indicate overdispersion.
* Test for zero inflation: the test indicates that the data has less zeros than expected. (I think a potential reason for this is that people choose to do one fact-check rather than zero "for-show").
* Mullticollinearity check is ok
* Test for categorical dependencies:

Chart, box and whisker chart

Description automatically generated

*(The function tests for two common problems: are residuals within each group distributed according to model assumptions, and is the variance between group heterogeneous.)*

Test for within-group uniformity : problematic groups are highlighted in red.

The test for homogeneity of variances is done with a Levene test. A significant p-value means that group variances are not constant. In this case, you should consider modelling variances.

* Residual Plots

Graphical user interface, chart, line chart

Description automatically generated

The plot function runs 4 tests i) KS test i) Dispersion test iii) Outlier test iv) quantile test

-KS test tests the overall uniformity of the simulated residuals in a DHARMa object. Result not ok. (Even if the residuals are not normally distributed, they should still be approximately uniformly distributed if the model is a good fit for the data)

-Dispersion is ok.

- Quantile test: The function fits quantile regressions on the residuals, and compares their location to the expected location (because of the uniform distribution, the expected location is 0.5 for the 0.5 quantile)

-Outlier tests is not ok

* Residual plots against predictors :they should reveal no patterns (only some as examples)

Anxiety

Chart, box and whisker chart

Description automatically generated

Importance Scale

Chart, line chart

Description automatically generated

Important for Others

Chart, box and whisker chart

Description automatically generated

**GLM by Country**

This model seeks to find the relationship between amount of fact-checks performed and the variables taken from the Likert scale “concern for the war” : anxiety, importance, important for others.

Example for country “Germany”

Distribution of the response variable:

Chart, histogram

Description automatically generated

Because this is count data, we have to use a generalized linear model.

I am using the R package MASS to run this model, and the package [DHARMa](https://cran.r-project.org/web/packages/DHARMa/DHARMa.pdf) for residual diagnostics.

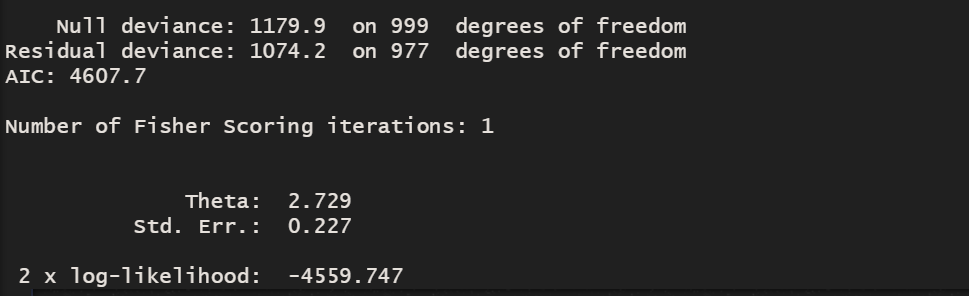
Using a negative binomial:

Summary

**Graphical user interface, text

Description automatically generated**

**Text

Description automatically generated with low confidence**

Diagnostics:

* Test for overdispersion: in this case the ratio is close to one but insignificant, so it doesn’t indicate overdispersion.
* Test for zero inflation: the test indicates that the data has less zeros than expected. (I think a potential reason for this is that people choose to do one fact-check rather than zero "for-show").
* Residual Plots

A picture containing chart

Description automatically generated

The plot function runs 4 tests i) KS test i) Dispersion test iii) Outlier test iv) quantile test

All the tests appear to be ok.

* Residual plots against predictors (they should reveal no patterns)

Anxiety

Chart, box and whisker chart

Description automatically generated

Importance Scale

Table

Description automatically generated with medium confidence

Important for Others

Chart, box and whisker chart

Description automatically generated

**Some thoughts:**

The model appears to work significantly better when it’s for one country alone. I think it may be because in the mixed model we would need to include other terms that capture variance by country. However, the more we include the more complicated the model becomes and probably won’t converge. Also, the interpretation becomes much harder. Would it make sense to have a model for each country?