Assessing Potential Heteroscedasticity in Psychological Data: A GAMLSS approach

Juan C. Correa [a,f] ✉ ⓘ, Thomas Kneib [b] ⓘ, Raydonal Ospina [c] ⓘ, Julian Tejada [d] ⓘ & Fernando Marmolejo-Ramos [e] ⓘ

[a] Critical Centrality Institute; Monterrey; Mexico
[b] Statistics and Campus Institute Data Science; Georg-August-University Göttingen; Göttingen; Germany
[c] Department of Statistics; IME; Universidade Federal da Bahia; Salvador 40170-110; Brazil & Department of Statistics; CASTLab; Universidade Federal de Pernambuco; Recife 50670-901; Brazil
[d] Department of Psychology; Universidade Federal de Sergipe; Sergipe; Brazil
[e] University of South Australia Online & University of Adelaide; Adelaide; Australia
[f] Universidad de Monterrey; Department of Psychology; Monterrey; Mexico

Abstract ■ This paper provides a tutorial for analyzing psychological research data with `GAMLSS`, an `R` package that uses the family of generalized additive models for location, scale, and shape. These models extend the capacities of traditional parametric and non-parametric tools that primarily rely on the first moment of the statistical distribution. When psychological data fails the assumption of homoscedasticity, the GAMLSS approach might yield less biased estimates while offering more insights about the data when considering sources of heteroscedasticity. The supplemental material and data help newcomers understand the implementation of this approach in a straightforward step-by-step procedure.

Keywords ■ Heteroscedasticity, GAMLSS, Scientific Evidence.. **Tools** ■ R.

✉ jcc@criticalcentrality.com
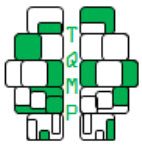
doi 10.20982/tqmp.19.4.p333

Introduction

Most statistical parametric inferential methods in behavioural sciences assume normality and homoscedasticity (Mair & Wilcox, 2020). Over-reliance on these assumptions, however, has been deemed problematic since the seminal paper of Micceri (1989), as he was among the first ones to pinpoint the non-normal shape of psychological data. Such data are characterized by heavy tails (e.g., the Log-Normal distribution), asymmetry, multi-modality, and modes outside the mean-median interval (e.g., Bono et al., 2017).

Empirical research in behavioural sciences has at least three interrelated strategies for data analysis. Mair and Wilcox (2020) identified them as 1) testing assumptions, 2) data transformation, and 3) employing robust statistical methods. Here, we add a fourth strategy that relies on applying the family of "*generalized additive models for location, scale, and shape*" (GAMLSS). GAMLSS is a powerful statistical modeling framework which extends the capabilities of generalized linear models introduced by Rigby and Stasinopoulos (2005). This family of models emerged under the assumption that the first moment of a statistical distribution (i.e., mean) is not necessarily the only helpful parameter for inferential purposes (Kneib, 2013). What GAMLSS offers in this regard is a wide variety of regression models, where all the parameters of the assumed distribution for the dependent variable can be modelled as additive functions of the predictor variables. Indeed, GAMLSS provides over 100 continuous, discrete, and mixed distributions for modelling several types of dependent variables. Furthermore, the analyst can use truncated, censored, log, and logit transformations and finite mixture versions of all these distributions.

GAMLSS is an encompassing set of techniques that contains, in this order of sophistication, linear models (LM;

also known as ordinary least squares or OLS), generalized linear models (GLM), and generalized additive models (GAM). Although GAMLSS offers an ample gamut of cutting-edge statistical algorithms, we believe GAMLSS has three main features that make it a powerful and flexible semi-parametric tool for data analysis. First, GAMLSS allows fitting the response variable via an ample range of continuous, discrete, and mixed distributions. For example, LM is limited to the Normal distribution, and GLM is limited to distributions from the exponential family (including the Normal distribution). Second, GAMLSS allows the usual linear fitting of the covariates (as LM and GLM do) and non-linear fitting via smoothing algorithms (i.e. GAM). Such smoothers include penalized smooth terms (e.g. P-splines), neural networks, and spatial and random effects. For the case of categorical variables, GAMLSS allows penalization to reduce the number of levels. Third, GAMLSS enables the modeling of the covariates' impact on all the distribution parameters used to describe the response variable. That is, while LM and GLM (including their mixed-model versions) can estimate the effects of covariates on the mean response only, GAMLSS enables to investigate if the covariates have effects on the location, scale (i.e. variance), and shape (e.g., skewness or kurtosis) of the response variable. Thus, GAMLSS stands as a broader applicable alternative to LM without suffering from the problems that emerge from its fundamental, often non-valid assumptions, such as homoscedasticity. Lastly, it is worth noting that present variations of GAMLSS render it appropriate for machine learning analyses, such as distributional regression trees and forests (see for example, the work of Klein, 2024 and Constable et al., 2023).

GAMLSS has been featured in several disciplines, including and not limited to scientometrics (Correa et al., 2022), nutrition sciences (Vasiljevic et al., 2019), or transportation (Hajmohammadi et al., 2019). In psychology, we have observed two types of usages: psychological test norming (Timmerman et al., 2021), neuroimaging norming (Dinga et al., 2021), or as a statistical method (Coupé, 2018; Campitelli et al., 2017). The current work provides an extension of this last type of usage. In the next section, we offer a minimum of technical details necessary for a basic understanding of the `gamlss` package (Stasinopoulos & Rigby, 2007). We revisited the data of Gignac and Zajenkowski (2020) to illustrate some of the most important capabilities of the `gamlss` package for psychological research, particularly about the modelling of heteroscedasticity. The work by Gignac and Zajenkowski (2020) aligns perfectly with our intentions as it illustrates a unique situation in which the results from the heteroscedasticity test can be interpreted as supportive evidence for the Dunning-Kruger hypothesis. The supplementary material accompanying this tuto-

rial provides a more fine-grained description of the syntaxes, calculations, and data interpretation from the results obtained as a reanalysis of the empirical findings by Gignac and Zajenkowski (2020).

Psychological Data Revisited

This tutorial examines a data set on the Dunning-Kruger Effect by Gignac and Zajenkowski (2020). This effect was originally studied in personality and social psychology (Kruger & Dunning, 1999) and refers to the simple idea that people's ignorance is often invisible to them. The meta-ignorance (i.e., ignorance of ignorance) emerges as the result of a lack of expertise and relies on disbelieves and background knowledge that is not appropriate for providing correct answers to a variety of questions (Dunning, 2011). Based on these considerations, studying the Dunning-Kruger Effect requires the analysis of self-assessed intelligence and comparing it with objective metrics. The raw data that was made available by Gignac and Zajenkowski (2020) on the Open Science Framework (osf.io/dg547), consist of a CSV file with 929 rows and 56 variables described as follows: 46 variables related to the Raven_IQ results and score transformations, 5 variables related to the SAIQ test, 2 variables with demographic information (sex and age), one variable (study) identifying each one of the three samples used by Gignac and Zajenkowski (2020) in their study and finally, two variables used to identify the participants. The "self-assessed intelligence questionnaire" (SAIQ) is on a discrete scale ranging from 1 to 25, while the intelligence objective metric (Raven_IQ) resulted from applying the Advanced Progressive Matrices (Raven et al., 1994). If the Dunning-Kruger Effect is correct, the results from the objective metric of IQ (Raven_IQ) and the self-assessment of intelligence (SAIQ) should show a particular resulting pattern in which the regression residuals should be larger at the lower-end of the measures range, and that had been tested using heteroscedasticity tests, such as the Glejser test (Machado & Santos Silva, 2000) used by Gignac and Zajenkowski (2020). In this context, we want to show how the GAMLSS could be used as an alternative test to evaluate the Dunning-Kruger Effect.

GAMLSS in practice

GAMLSS functionalities are available through two main R-based packages: the `gamlss` package and the `gamlss.dist` package. The GAMLSS Team provides both packages and are part of the packages needed to implement the framework described in the book "*Flexible Regression and Smoothing Using GAMLSS*" in R by Stasinopoulos et al. (2017). The `gamlss` package provides the functions for fitting the models, and the `gamlss.dist` package offers extensions to the `lm()`, and `glm()` functions from the `stats`
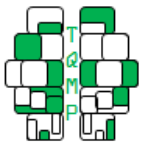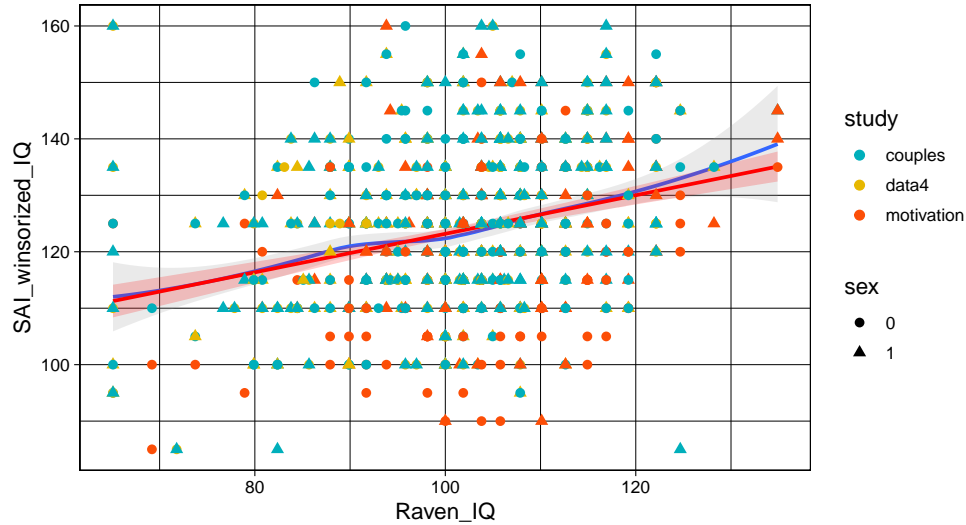
Figure 1 ■ Empirical relationships between an objective measurement of human intelligence and a subjective self-report of it, as functions of other two qualitative variables

package (R Core Team, 2023) and to the gam() function from the gam package (Hastie, 2020) for generalized additive models. The gamlss package needs the gamlss.dist package installed to work.

Installation setup and instructions for reproducibility

Installing both the gamlss and the gamlss.dist is straightforward through the Rtudio built-in "install" button. Sometimes the user might find interesting to work with a third related package called gamlss.util, which is not available through RStudio as it was archived on 2020-09-03. As it is available online (https://github.com/cran/gamlss.util), users can download it via devtools::install_github("cran/gamlss.util"). After these preliminary steps, the user can run GAMLSS functionalities like any other standard R package. To facilitate the reproduction of the results that will be shown below, interested readers can download and knit an RMarkdown file (GAMLSS Tutorial for Psychological Data.Rmd) from the following GitHub repository: https://github.com/jcorrean/GAMLSS4PsycData. We encourage interested readers to use RStudio to open this Rmd file to knit and reproduce the same results reported below.

First Step: Data Linearity Scrutiny

A recommended practice for data analysts is initializing data examination by scrutinizing data linearity assumptions. Here, we used a LOESS smoother for each variable in the model. A LOESS smoother selectively weights points in the regression. We can also overlay the standard regression line as the original paper of Gignac and Zajenkowski (2020) to determine how the LOESS smoothing (for detecting non-linearities) of data compares to linear regression. The following syntax yields a coloured scatterplot between the principal variables in the analysis: an objective measurement of human intelligence (Raven_IQ) and a subjective self-report of it (SAIQ) plotted by discriminating participant sex and the origin of the data (study variable which can take three values which represent the names of the three different samples used by Gignac and Zajenkowski, 2020).

```
# Loading libraries
library(ggplot2)
# Plotting data
ggplot(data, aes(x=Raven_IQ,y=SAI_winsorized_IQ)) +
 stat_smooth(method="LOESS", alpha=.2) +
 stat_smooth(method="lm", formula=y ~ x, color="red",
    fill="red", alpha=.2) +
 geom_point( aes(colour=study, shape=sex), size=2.3) +
 scale_color_manual(values = c("#00AFBB", "#E7B800", "#
    FC4E07"))+theme_linedraw()
```

The plot in Figure 1 shows the scatter plot of the Raven_IQ versus SAI_winsorized_IQ scores. Each point represents a participant, and as expected, the higher the Raven_IQ score, the higher the SAI_winsorized_IQ. Nevertheless, although the points follow a clustering trend, their dispersion on the y-axis is very large, indicating that for a specific Raven_IQ score, there are multiple values of SAIQ; conversely, for different participants with the same
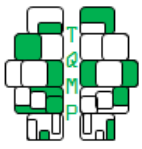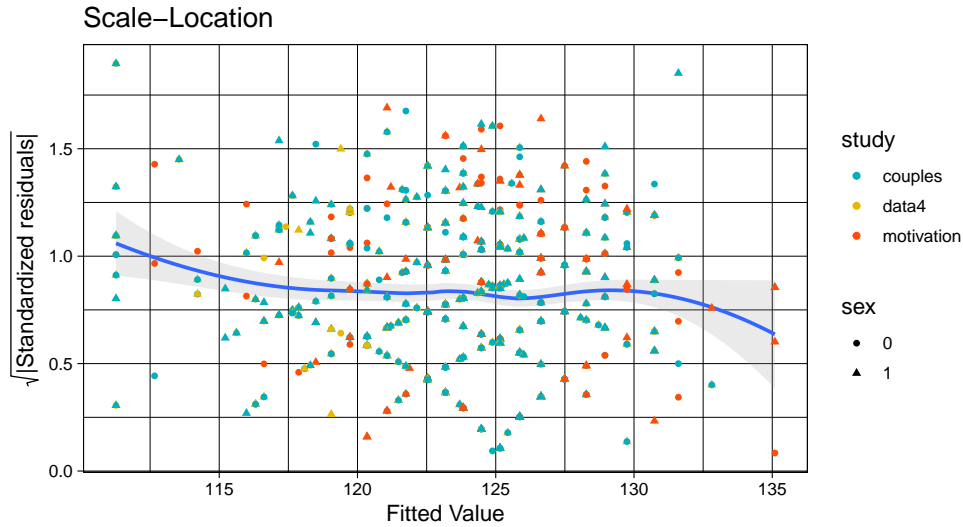
Figure 2 ■ Spread-Location Plot on the relationships between an objective measurement of human intelligence and a subjective self-report of it, as functions of other two qualitative variables

Raven_IQ there are different values of self-assessed intelligence.

The Figure also shows a blue curve with grey shading corresponding to the standard errors for the LOESS smoother fit, while the red straight line with red shading corresponds to the standard errors for the linear fit. The blue line is a LOESS smoothed line, and the red line is a linear regression line. As long as the LOESS smoother (blue line) roughly approximates the linear tendency (red line), the assumption of linearity has been apparently met. However, with a more careful, simple visual inspection, the linearity hypothesis seems not to be fully satisfied. The confidence bands are of short length (especially the confidence red band for the linear model), and LOESS, the blue line, shows local variations moving away from the red line, which may indicate some non-linearity in the data, plus, the square of the correlation coefficient $r^2 = 0.078$ is very low. The supplementary material offers the reader other visualization syntaxes that depict problems of linearity.

Second Step: Homoscedasticity Assumption

This assumption can be checked by examining the Scale-location plot, also known as the spread-location plot (see Figure 2), another scatter plot that examines the variance of the residuals versus the predicted values of the linear model. Ideally, in the plot, we check the presence of a pattern in the residuals, that will be consistent spread across the range of predicted values, indicating that the variability of the errors is consistent across all levels of the pre-

dictor. Conversely, if the blue line is flat and horizontal with equally and randomly spread data points, there is homoscedasticity. If the blue line has a positive or negative slope or the data points are not randomly scattered, this conveys the message of homoscedasticity violation. This plot shows that residuals are not spread equally across predictor ranges as expected to confirm the Dunning-Kruger Effect. In our example, this is not entirely the case; rigorously, we need to proceed with the variance homogeneity assumption. A second alternative to evaluate the homoscedasticity assumption is to use the absolute studentized residuals (also available in the supplementary material), which confirm the conclusions found with the spread-location plot.

```
# Building the linear model of SAIQ versus Raven_IQ
lm.temp <- lm(SAI_winsorized_IQ~Raven_IQ, data=db)
# Extracting the residuals from the model results
db$resid <- resid(lm.temp)
# Extracting the predictions from the
# model results
db$fitted <- fitted(lm.temp)
# Extracting the standardized residuals from the
# model results
db$stdresid <- rstandard(lm.temp)
# Estimating the square root of the
# standardized residuals
db$rstandard <-  sqrt(abs(db$stdresid ))
# Plotting the spread-location plot
ggplot(db, aes(x=fitted,y=rstandard)) +
 stat_smooth(method="LOESS", alpha=.2) +
 geom_point( aes(colour=study, shape=as.factor(sex)),
    size=2.3) +
```
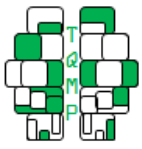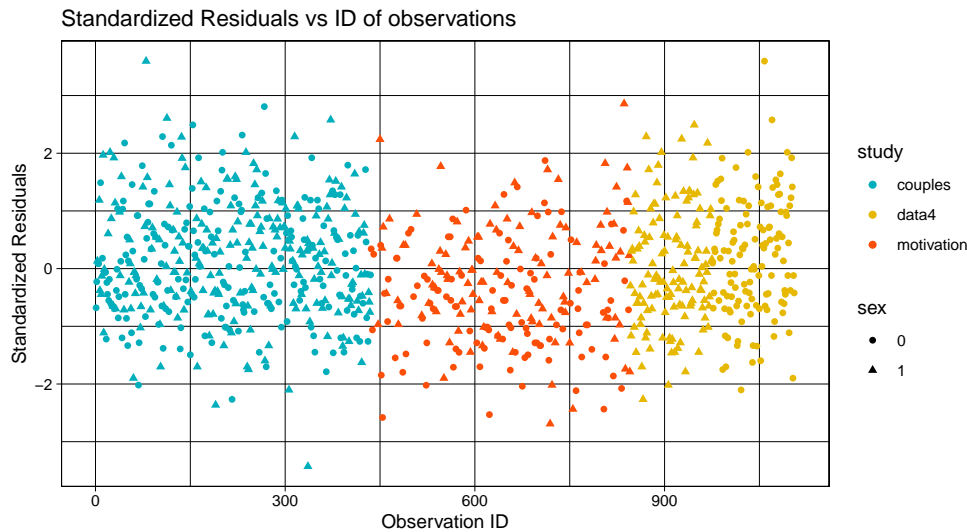
Figure 3 ∎ Presence of group patterns in the scatterplot of standardized residuals and observation ID

Standardized Residuals vs ID of observations

Third Step: Testing the homoscedasticity assumption

To increase our certainty in the interpretation of the previous step, we can perform homoscedasticity tests. The function `ncvTest()` and `bptest()` return a numerical output for the hypothesis of constant error variance against the alternative that the error variance changes with the level of the response (fitted values) or with a linear combination of predictors. On the one hand, the default `bptest()` from the `lmtest` package (Zeileis & Hothorn, 2002) uses the studentized Breusch-Pagan test proposed by Koenker (1981). On the other hand, `ncvTest()` from the `car` package (Fox & Weisberg, 2019), performs the original version of Breusch-Pagan test (Breusch & Pagan, 1979). The following syntaxes yield these results:

```
# Breusch & Pagan test
# load libraries
library(lmtest)
# Breusch & Pagan test using the results of the
# linear fit
bptest(lm.temp)


##
##  studentized Breusch-Pagan test
##
## data:  lm.temp
## BP = 4.8922, df = 1, p-value = 0.02698
```

```
# Koenker version of Breusch & Pagan test
# Load libraries
library(car)
# Koenker version of Breusch & Pagan test using
# the linear fit
ncvTest(lm.temp)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 4.942294, Df = 1, p = 0.026207
```
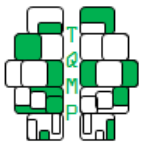
The results from both the Breusch-Pagan test (4.892; df = 1; p = 0.026), and the non-constant variance test (χ^2 = 4.942; df = 1; p = 0.026) led us to reject the hypothesis of homoscedasticity of errors at a .05 alpha level.

Fourth Step: Checking the Independence Assumption

In ideal conditions, it is expected the absence of data patterns when plotting standardized residuals against the index of observations. The presence of patterns conveys a clear message about the lack of independence or some uncontrolled endogeneity. After running the syntax below, we obtain Figure 3 with clear data patterns. The red dots make standardized residuals look a little more homogeneous than the other two samples, with fewer values greater than 1 and a few more values lower than -2.

```
# plotting the scatter plot of standardized residuals
# and observation ID's
p2 <- ggplot(db, aes(x=ID, y=stdresid)) +
 geom_point( aes(colour=study, shape=sex), size=2.3) +
 scale_color_manual(values = c("#00AFBB", "#E7B800", "#
    FC4E07")) +
 theme_linedraw()
 p2 + labs(x='Observation ID',
```

```
        y='Standardized Residuals',
        title='Standardized Residuals vs ID of
    observations')
\end{verbatim}
```

A Durbin-Watson (D-W) test allowed us to examine whether the errors are autocorrelated with themselves. The null hypothesis in this regard states that they are not autocorrelated. In other words, at this moment, it is useful to verify that we have not violated the independence assumption (presence of lack-of-fit).

```
# Durbin-Watson using the linear fit
durbinWatsonTest(lm.temp)
```

```
##  lag Autocorrelation D-W Statistic p-value
##   1       0.9278538    0.1342519        0
##  Alternative hypothesis: rho != 0
```

The results of the D-W Statistic and its p-value suggest a violation of the independence assumption. Thus, the data analysis requires the inclusion of a third grouping variable, in our case, 'sex' or s'tudy variables'.

Fifth Step: GAMLSS specification and assumptions checks

In the GAMLSS approach, the fit of relationships based on regression models does not have different assumptions between the response and the covariates than the linear models. We can use this property to find distributional evidence of lack of fit. Based on normality assumptions on errors and homoscedasticity, the linear model parametric fit based on OLS and likelihood are equivalent because if the residuals errors showed a normal distribution the least squares estimators should converge with the maximum likelihood estimators (Kutner et al., 2003). Also, the hypotheses of normality of residuals are useful for inferential purposes. Therefore, we can fit the model proposed by Gignac and Zajenkowski (2020) via GAMLSS using the family of normal distribution (NO). Thus, our following step requires a regression-like GAMLSS specification and employs it as an input in another chart.

```
# loading libraries
library(gamlss)
# Defining GAMLSS model
temp <- gamlss(SAI_winsorized_IQ ~ Raven_IQ, data=db,
    family=NO, trace=FALSE)
temp
```

```
## Family:  c("NO", "Normal")
## Fitting method: RS()

## Call:  gamlss(formula = SAI_winsorized_IQ ~ Raven_IQ,
    family = NO,      data = db, trace = FALSE)

## Mu Coefficients:
## (Intercept)     Raven_IQ
##     89.0364       0.3414
## Sigma Coefficients:
## (Intercept)
```

```
##       2.611

##  Degrees of Freedom for the fit: 3 Residual Deg. of
    Freedom   926
## Global Deviance:      7488.09
##          AIC:         7494.09
##          SBC:         7508.59
```

The residuals summary statistics from the resulting model can be obtained using the `summary` command:

```
> summary(temp)
*******************************************************
Family:  c("NO", "Normal")

Call:  gamlss(formula = SAI_winsorized_IQ ~ Raven_IQ,
    family = NO,      data = db, trace = FALSE)

Fitting method: RS()

-------------------------------------------------------
Mu link function:   identity
Mu Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) 89.03639    3.93286  22.639   <2e-16 ***
Raven_IQ     0.34144    0.03842   8.887   <2e-16 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1
    ` ' 1

-------------------------------------------------------
Sigma link function:  log
Sigma Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.6113     0.0232   112.6   <2e-16 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1
    ` ' 1

-------------------------------------------------------
No. of observations in the fit:  929
Degrees of Freedom for the fit:  3
    Residual Deg. of Freedom:  926
                at cycle:  2

Global Deviance:      7488.088
          AIC:         7494.088
          SBC:         7508.59
*******************************************************
```

This results showed that for the μ submodel the Raven_IQ is a significant predictor of SAIQ, but not for the σ submodel. As expected for the Dunning-Kruger Effect the scores for the IQ test does not necessarily predict the self-concept of intelligent scores. The result from the `gamlss` command could be also used to evaluated normality, following the next commands.

```
# Plotting the distribution of data
plotSimpleGamlss(x= Raven_IQ,
            y=SAI_winsorized_IQ,
            model=temp, data=db, pch=1,
            x.val=seq(65,160,6), val=200, N=500,
            ylim=c(80, 170),
            xlim=c(58, 135),
```
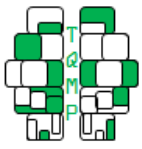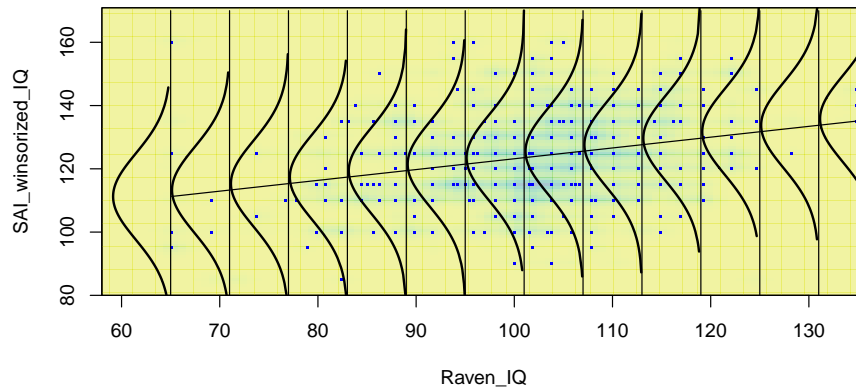
Figure 4 ■ Distribution of data in the Y axis at specific values of the data in the X axis. Note the shapes of the probability density function plots are based on the Normal distribution. Also, note that both location and scale can change according to X.


```
cols=viridis_pal(begin = 0.3, end = 0.95,
                 alpha=0.4, option ="D",
                 direction=1)(100))
```

The previous syntax leads to Figure 4. A simple visual inspection of this chart reveals that there is no strong evidence of normality deviation. Nonetheless, some density curves show different variability, which can be additional evidence of non-constant dispersion (i.e., heteroscedasticity) that is not so commonly checked in empirical psychological research.

Another way to visualize the normality assumption is through a worm-plot, a detrended Q-Q plot in which the deviation around the reference line can be seen better by having more space on the y-axis. Figure 5 depicts a worm-plot (Buuren & Fredriks, 2001) that identifies some characteristics of the data that the fitted model does not adequately capture.

The yellow dots indicate how far the residuals depart from their expected null value. A high (low) variance pattern is related to positive (negative) sloped dots. When the dots are U-shaped (inverted U-shaped), there is evidence of positive (negative) skewness. When the dots are S-shaped with left bent up (down), there is evidence of large (small) kurtosis (for more details as to the interpretation of the worm plot, see Table 12.1 in Stasinopoulos et al., 2017). Thus, when the model is correctly specified, all or nearly all of the dots are expected to lie inside the two dashed semicircles. A commonly used rule is to conclude that the model is correctly specified when at least 95% of

the dots do so. Figure 5 shows a U-shape and several points outside the confidence band, indicating that the hypothesis of normality and constant variance is not fulfilled, and at the same time, the wormplot U shape suggests a negative skew (Buuren & Fredriks, 2001), as was expected for the Dunning-Kruger effect for which the left tail should be longer, because the misestimation of cognitive ability should be larger at the lower end of the spectrum (left tail) (Gignac & Zajenkowski, 2020). In the supplementary material, other checks are also available.

Sixth Step: Evidence of Heteroscedasticity

Using regressograms, a smoother version of the histogram, we evaluate heteroscedasticity evidence by plotting the average of all points that fall into disjoint bins. We propose constructing piecewise functions in the intervals $\{B_k := [t_k, t_{k+1}) : t_k = t_0 + hk, k \in Z\}$ (uniform breaks) or $\{B_k := [t_k, t_{k+1}) : t_k - \text{empiric quantile}, k \in Z\}$ (quantile breaks) and calculate for y (output) a robust coefficient of variation based on the MAD across the midpoints of x (input) in B_k. This estimator, proposed by Ospina and Marmolejo-Ramos (2019), is defined as follows,

$$CV_{\text{MAD}} = \widehat{\theta}_{\text{MAD}} = \frac{\text{MAD}}{\text{Mdn}} \qquad (1)$$

where MAD is defined as,

$$\text{MAD} = 1.4826 \cdot \text{Mdn}\{|x - \widehat{Q}_2|\}, \qquad (2)$$

and in Eq. (1) $\text{Mdn} = \widehat{Q}_2 = \widehat{Q}(0.5) = F_n(0.5)^{-1}$ is an estimator of location that is robust as it has a high breakdown
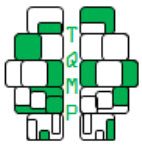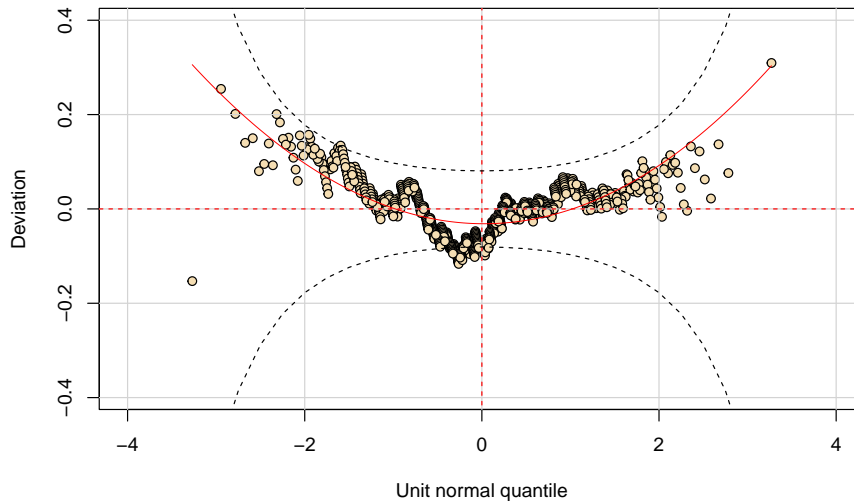
Figure 5 ▪ Worm-plot of the residuals of the gamlss model. Data were modelled with the Normal distribution.

based om Empirical function distribution. Note that the breakdown value is the smallest fraction of contamination that can cause the estimator to take on values far from its value on the uncontaminated data. Here, we choose the bandwidth h by the Sturges rule (Sturges, 1926) for histograms and we chose the percentiles for quantile breaks. The syntaxes included in Appendix facilitate the exploration of heteroscedasticity as defined above.

Figure 6 depicts the resulting regressogram plots in terms of the behaviour of the covariates against the response and in terms of the fitted values against the residuals. Regardless of how the breaks are constructed, the CV_{MAD} is not constant, giving further evidence of heteroscedasticity.

Final Validation

Finally, we can use a global test of model assumptions using the `gvlma` package in R (Pena & Slate, 2019). The package allows the user to check out (almost) all the ideas discussed so far. This package relies on the methodology proposed by Peña and Slate (2006). Appendix 2 at the end of this manuscript provides the syntaxes for this purpose.

Discussion

This article provided a GAMLSS-based tutorial for the analysis of potential heteroscedasticity in psychological data. As mentioned earlier, we revisited the data provided by Gignac and Zajenkowski (2020) about the Dunning-

Kruger effect, a phenomenon where heteroscedasticity is expected. However, Gignac and Zajenkowski (2020) reported that the Glejser test did not yield significant values, suggesting that their data were homoscedastic, contradicting the original hypothesis. Here we presented four different methods to reevaluate the heteroscedasticity hypothesis of the Gignac and Zajenkowski (2020) data, and our results revealed inconclusive evidence of heteroscedasticity as two statistical tests suggested its presence (i.e., Non-constant variance test and Breusch-Pagan test) while the other two tests did not (i.e., Glejser and the 'gvlma' tests). Our finding clearly extended the scrutiny of Gignac and Zajenkowski (2020) and illustrated that its examination should not be based on one test only. Substantive knowledge of the data-generating process, along with an ample gamut of statistical techniques, are, in our view, the best set of tools for analyzing psychological data with potential heteroscedasticity. Such an approach is in line with the multiverse analysis previously introduced by Steegen et al. (2016).

Although our approach has not found an unequivocal answer about the distribution of the Gignac and Zajenkowski (2020) data and the presence of the heteroscedasticity, we identify a set of specific characteristics of the distribution data that fit with the Dunning-Kruger effect. We found issues with the distribution of the residuals in that these were not equally spread along the predictor ranges, and the fact that the scores for the IQ test were not
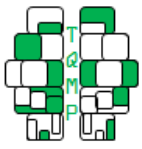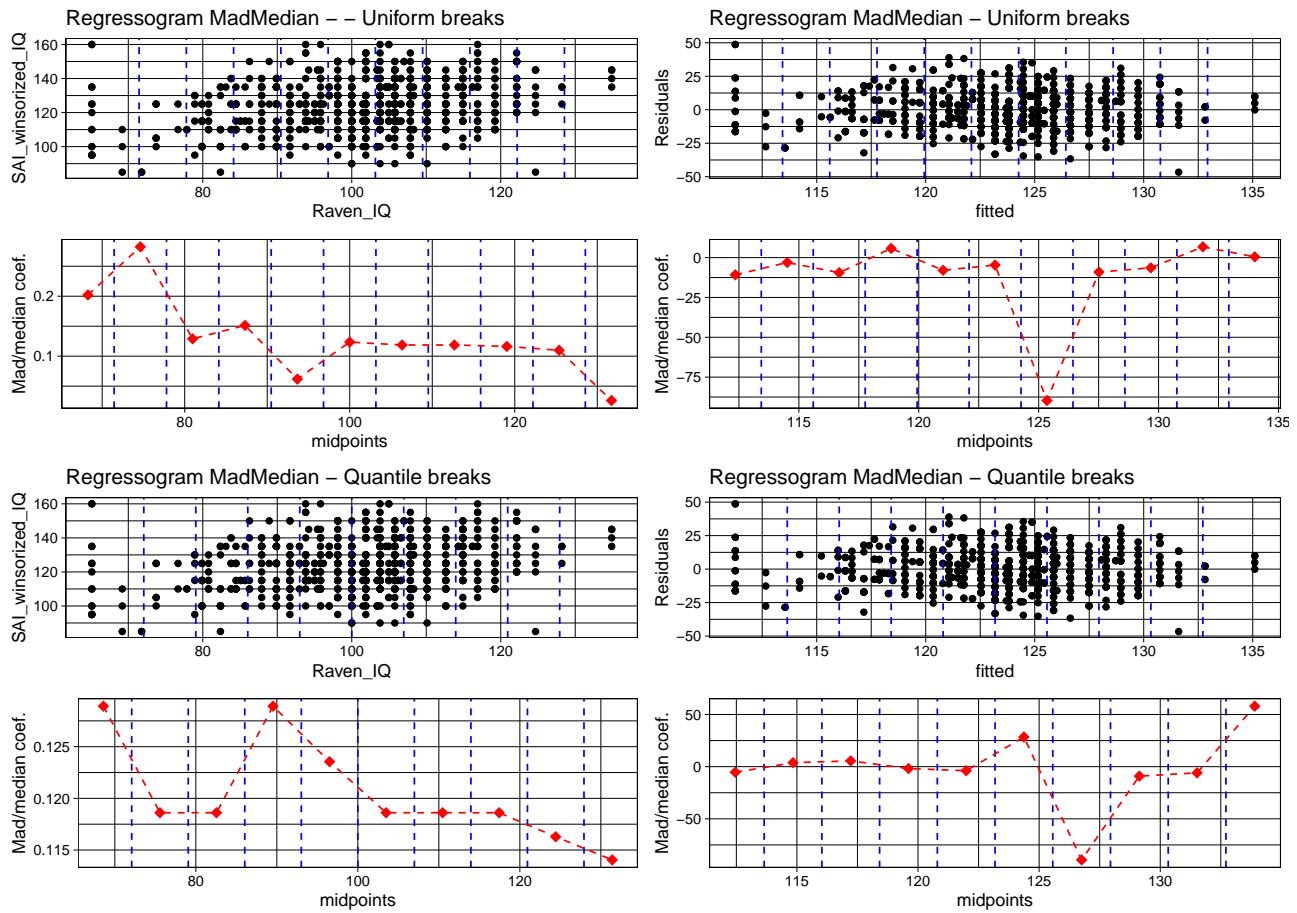
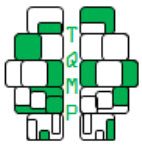**Figure 6 ∎** Regressograms with several uniform breaks (upper) and quantile breaks (bottom).

good predictors of the self-concept of intelligence scores for the gamlss σ submodel. These characteristics suggest that the data from Gignac and Zajenkowski (2020) could be used to support the Dunning-Kruger effect, demonstrating that the degree to which people mispredicted their measured intelligence is not equal for all levels of intelligence.

This reinterpretation of the Gignac and Zajenkowski (2020) data was made possible through the use of GAMLSS, and even though the use of the gamlss R package is not entirely novel in psychology, its employment has rarely gone beyond the disciplinary borders of psychometrics and methodological research (Timmerman et al., 2021; Voncken et al., 2019b, 2019a; Dinga et al., 2021; Coupé, 2018; Campitelli et al., 2017). This tutorial aimed to promote the use of the gamlss package for applied researchers working on different psychology areas (e.g., consumer psychology, industrial and organizational psychology, clinical psychology).
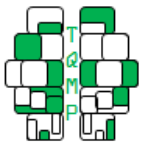
One aspect that is particularly relevant to enable more widespread usage of GAMLSS in psychology and beyond is the development of tools supporting the interpretation of the fitted models by the data analyst. In addition to appropriate summary statistics, graphical tools are particularly relevant for this purpose. Effect displays, scenario analyses (see for more details in Kneib et al., 2021) and bucketplots (de Bastiani et al., 2022) are potential ways of enabling a better understanding of how certain explanatory variables impact the fit of a model, and their construction can be facilitated by corresponding statistical software (see for example, Stadlmann and Kneib, 2021).

References

Bono, R., Blanca, M. J., Arnau, J., & Gómez-Benito, J. (2017). Non-normal distributions commonly used in health, education, and social sciences: A systematic review.

Frontiers in Psychology, 8, 1602. doi: 10.3389/fpsyg.2017.01602.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287–1294. doi: 10.2307/1911963.

Buuren, S. v., & Fredriks, M. (2001). Worm plot: A simple diagnostic device for modelling growth reference curves. *Statistics in Medicine, 20*(8), 1259–1277. doi: 10.1002/sim.746.

Campitelli, G., Macbeth, G., Ospina, R., & Marmolejo-Ramos, F. (2017). Three strategies for the critical use of statistical methods in psychological research. *Educational and Psychological Measurement, 77*(5), 881–895. doi: 10.1177/0013164416668234.

Constable, P. A., Loh, L., Prem-Senthil, M., & Marmolejo-Ramos, F. (2023). Visual search and childhood vision impairment: A GAMLSS-oriented multiverse analysis approach. *Attention, Perception, & Psychophysics, 85*, 968–977. doi: 10.3758/s13414-023-02670-z.

Correa, J. C., Laverde-Rojas, H., Tejada, J., & Marmolejo-Ramos, F. (2022). The sci-hub effect on papers' citations. *Scientometrics, 127*, 99–126. doi: 10.1007/s11192-020-03806-w.

Coupé, C. (2018). Modeling linguistic variables with regression models: Addressing non-gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in Psychology, 9*(APR). doi: 10.3389/fpsyg.2018.00513.

de Bastiani, F., Stasinopoulos, D., Rigby, R., Heller, G., & Silva, L. (2022). Bucket plot: A visual tool for skewness and kurtosis comparisons. *Brazilian Journal of Probability and Statistics, 36*(3), 421–440. doi: 10.1214/22-BJPS533.

Dinga, R., Fraza, C., Bayer, J., Kia, S., Beckmann, C., & Marquand, A. (2021). Normative modeling of neuroimaging data using generalized additive models of location scale and shape. doi: 10.1101/2021.06.14.448106.

Dunning, D. (2011). The Dunning-Kruger Effect. On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology, 44*, 247–296. doi: 10.1016/B978-0-12-385522-0.00005-6.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger Effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence, 80*. doi: 10.1016/j.intell.2020.101449.

Hajmohammadi, H., Marra, G., & Heydecker, B. (2019). Data-driven models for microscopic vehicle emissions. *Transportation Research Part D: Transport and Environment, 76*, 138–154. doi: 10.1016/j.trd.2019.09.013.

Hastie, T. (2020). *Gam: Generalized additive models* [R package version 1.20]. https://CRAN.R-project.org/package=gam

Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application, 11*(1). doi: 10.1146/annurev-statistics-040722-053607.

Kneib, T. (2013). Beyond mean regression. *Statistical Modelling, 13*(4), 275–303. doi: 10.1177/1471082X13494159.

Kneib, T., Silbersdorff, A., & Säfken, B. (2021). Rage against the mean - A review of distributional regression approaches. *Econometrics and Statistics, 26*, 99–123. doi: 10.1016/j.ecosta.2021.07.006.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics, 17*(1), 107–112. doi: 10.1016/0304-4076(81)90062-2.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121. doi: 10.1037/0022-3514.77.6.1121.

Kutner, M. H., Neter, J., & Nachtsheim, C. J. (2003). *Applied Linear Regression Models* (4th Revised edition). Irwin/McGraw-Hill.

Machado, J., & Santos Silva, J. (2000). Glejser's test revisited. *Journal of Econometrics, 97*, 189–202. doi: 10.1016/S0304-4076(00)00016-6.

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods, 52*, 464–488. doi: 10.3758/s13428-019-01246-w.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156. doi: 10.1037/0033-2909.105.1.156.

Ospina, R., & Marmolejo-Ramos, F. (2019). Performance of some estimators of relative variability. *Frontiers in Applied Mathematics and Statistics, 5*, 43. doi: 10.3389/fams.2019.00043.

Pena, E. A., & Slate, E. H. (2019). *Gvlma: Global validation of linear models assumptions* [R package version 1.0.0.3]. https://CRAN.R-project.org/package=gvlma

Peña, E. A., & Slate, E. H. (2006). Global validation of linear model assumptions. *Journal of the American Statistical Association, 101*(473), 341–354. doi: 10.1198/016214505000000637.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Raven, J. C., Court, J. H., & Raven, J. (1994). *Manual for raven's progressive matrices and mill hill vocabulary*

scales. advanced progressive matrices. Oxford Psychologists Press.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554. doi: 10.18637/jss.v023.i07.

Stadlmann, S., & Kneib, T. (2021). Interactively visualizing distributional regression models with distreg.vis. *Statistical Modelling*. doi: 10.1177/1471082X211007308.

Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, *23*(7). doi: 10.18637/jss.v023.i07.

Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible Regression and Smoothing Using GAMLSS in R.* CRC Press.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. doi: 10.1177/1745691616658637.

Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, *21*, 65–66. doi: doi:10.1080/01621459.1926.10502161..

Timmerman, M. E., Voncken, L., & Albers, C. J. (2021). A Tutorial on Regression-Based Norming of Psychological Tests With GAMLSS. *Psychological Methods*, *26*(3), 357–373. doi: 10.1037/met0000348.

Vasiljevic, M., Fuller, G., Pilling, M., Hollands, G. J., Pechey, R., Jebb, S. A., & Marteau, T. M. (2019). What is the impact of increasing the prominence of calorie labelling? a stepped wedge randomised controlled pilot trial in worksite cafeterias. *Appetite*, *141*, 104304. doi: 10.1016/j.appet.2019.05.035.

Voncken, L., Albers, C., & Timmerman, M. (2019a). Improving confidence intervals for normed test scores: Include uncertainty due to sampling variability. *Behavior Research Methods*, *51*(2), 826–839. doi: 10.3758/s13428-018-1122-8.

Voncken, L., Albers, C., & Timmerman, M. (2019b). Model Selection in Continuous Test Norming With GAMLSS. *Assessment*, *26*(7), 1329–1346. doi: 10.1177/1073191117715113.

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://CRAN.R-project.org/doc/Rnews/
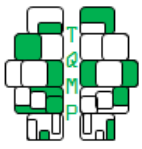
Appendix 1: Exploring heteroscedasticity in the data

The following syntaxes are helpful to explore the heteroscedasticity in the data.

```
## mad/median estimation
madmedian <- function(x){
  mdn <- median(x)
  MAD <- 1.4826 * median(abs(x - mdn))
  MAD/mdn
}


# classic coefficient of variation
oldcv <- function(x) sd(x)/mean(x)


# Defining a new function to estimate the regression mad/median
regres.madmedian <- function(x,y, x.lab="X",y.lab="Y",main="TITLE"){
  xy <- data.frame(x=x,y=y)
  xy <- xy[order(xy$x),]
  nbins <- nclass.Sturges(y)
  temp <- mean(y)
  z <- cut(xy$x,breaks=seq(min(xy$x),max(xy$x),length=nbins+1),
           labels=1:nbins,include.lowest=TRUE)
  xyz <- data.frame(xy,z=z)
  MEANS <- c(by(xyz$y,xyz$z,FUN=madmedian))
  x.seq <- seq(min(x), max(x), length = nbins + 1)
  midpts <- (x.seq[-1] + x.seq[-(nbins + 1)])/2
  d2 <- data.frame(midpts = midpts, MEANS = MEANS)
  p <- ggplot(xyz, aes(x, y)) + geom_point() + ggtitle(main) +
    xlab(x.lab) + ylab(y.lab) + theme(text = element_text(size = 20))
```

```
    p <- p + geom_vline(xintercept = x.seq[-c(1, nbins +
                           1)], linetype = "dashed", color = "blue")+theme_linedraw()

    p2 <-  ggplot()+ geom_point(data = d2, aes(x = midpts, y = MEANS),
                          color = "red", shape = 18, size = 3)+geom_vline(xintercept = x.seq[-c
    (1, nbins +
                           1)], linetype = "dashed", color = "blue")

    p2 <- p2 + geom_line(data = d2, aes(x = midpts, y = MEANS),
                          color = "red", linetype = "dashed") +
                          xlab("midpoints") +
                          ylab("Mad/median coef.") +
                          theme(text = element_text(size = 20)) +
                          theme_linedraw()

  output <- plot_grid(p, p2, ncol=1)
  output
}

# regression mad/median estimation for the Raven IQ and the SAI scores
regres.madmedian(x = db$Raven_IQ,
                 y = db$SAI_winsorized_IQ,
                 x.lab = "Raven_IQ",
                 y.lab = "SAI_winsorized_IQ",
                 main = "Regressogram MadMedian - - Uniform breaks")

# regression mad/median estimation for fitted values and residuals
regres.madmedian(x = db$fitted,
                 y = db$resid,
                 x.lab = "fitted",
                 y.lab = "Residuals",
                 main = "Regressogram MadMedian - Uniform breaks")

# Defining a new function to estimate R-square for the regression mad/median
regresquan.madmedian <- function(x,y, x.lab="X",y.lab="Y",main="TITLE"){
   xy <- data.frame(x=x,y=y)
   xy <- xy[order(xy$x),]
   brks <- as.numeric(with(xy, quantile(x, probs = seq(0,1, .1))))
   nbins <- length(brks) # nclass.Sturges(y)
   temp <- mean(y)
   z <- cut(xy$x,breaks= brks,
            # seq(min(xy$x),max(xy$x),length=nbins+1),
   labels=1:(nbins-1),include.lowest=TRUE)

   xyz <- data.frame(xy,z=z)
     MEANS <- c(by(xyz$y,xyz$z,FUN=madmedian))
     x.seq <- seq(min(x), max(x), length = nbins)
     midpts <- (x.seq[-1] + x.seq[-(nbins)])/2
     d2 <- data.frame(midpts = midpts, MEANS = MEANS)
     p <- ggplot(xyz, aes(x, y)) + geom_point() + ggtitle(main) +
       xlab(x.lab) + ylab(y.lab) + theme(text = element_text(size = 20))
```
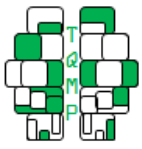
```
    p <- p + geom_vline(xintercept = x.seq[-c(1, nbins)],
                        linetype = "dashed", color = "blue")+theme_linedraw()

    p2 <-  ggplot()+ geom_point(data = d2, aes(x = midpts, y = MEANS),
                        color = "red", shape = 18, size = 3) +
                        geom_vline(xintercept = x.seq[-c(1, nbins )],
                        linetype = "dashed", color = "blue")

    p2 <- p2 + geom_line(data = d2, aes(x = midpts, y = MEANS),
                        color = "red", linetype = "dashed") +
                        xlab("midpoints") +
                        ylab("Mad/median coef.") +
                        theme(text = element_text(size = 20)) +
                        theme_linedraw()

    output <- plot_grid(p, p2, ncol=1)
    output
}

# Estimating the R-square of regression mad/median estimation for the Raven IQ and the SAI
    scores
regresquan.madmedian(x = db$Raven_IQ,
                     y = db$SAI_winsorized_IQ,
                     x.lab = "Raven_IQ",
                     y.lab = "SAI_winsorized_IQ",
                     main = "Regressogram MadMedian - Quantile breaks")

# Estimating the R-square for fitted vales and residuals
regresquan.madmedian(x = db$fitted,
                     y = db$resid,
                     x.lab = "fitted",
                     y.lab = "Residuals",
                     main = "Regressogram MadMedian - Quantile breaks")
```
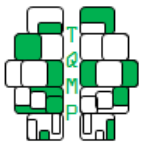
Appendix 2: Final Validation

The following syntaxes are helpful to conduct the final validation of results

```
library(gvlma)
gvmodel.gobble <- gvlma(lm.temp)
summary(gvmodel.gobble)
##
## Call:
## lm(formula = SAI_winsorized_IQ ~ Raven_IQ, data = db)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.605  -9.214  -0.984   9.134  48.734
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 89.03639    3.93709  22.615   <2e-16 ***
## Raven_IQ     0.34144    0.03846   8.878   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.63 on 927 degrees of freedom
## Multiple R-squared:  0.07836,    Adjusted R-squared:  0.07736
## F-statistic: 78.81 on 1 and 927 DF,  p-value: < 2.2e-16
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = lm.temp)
##
##                    Value p-value                   Decision
## Global Stat       6.24190 0.18180    Assumptions acceptable.
## Skewness          5.28391 0.02152 Assumptions NOT satisfied!
## Kurtosis          0.01622 0.89867    Assumptions acceptable.
## Link Function     0.36737 0.54444    Assumptions acceptable.
## Heteroscedasticity 0.57441 0.44851    Assumptions acceptable.
```

Citation