ANOVA breakdown for quantile models

E-mail distributed on 25-08-2025

Dear all,

When fitting a regression model, a routine check should be made for influential cases, that is, observations which exert an undue influence on the model's parameter estimates or its fitted values.[1] When such cases are detected (e.g., with Cook's distance), there are two broad options to deal with them: **(1)** remove them and refit the model, or **(2)** keep them and use a "robust" alternative to ordinary least-squares (OLS) regression. A large literature has accumulated by now on robust regression methods, with the following a non-exhaustive list of examples:

- **Rank regression**: Regression where the outcome or predictors have been rank-transformed to reduce the impact of outliers and/or nonlinearity (e.g., Spearman correlation).
- **Least absolute regression (LAR)**: Regression which minimizes the sum of absolute residuals (`lad` in `L1pack`). Also sometimes called least absolute deviation (LAD) regression.
- **Least median of squares regression (LMS)**: Regression which minimizes the median squared error, rather than the mean squared error (`lqs` in `MASS`).
- **Iteratively reweighted least squares (IRLS)**: Regression in which case-weights are applied iteratively to reduce the impact of the most influential cases (various functions and methods in `robustbase` and `robust`).[2]
- **Quantile regression**: A generalized version of LAR, where a regression model can be estimated for one or more chosen quantiles of the outcome simultaneously (`rq` in `quantreg`). When the chosen quantile is 0.5 (=the median), the model is equivalent to LAR.

Quantile regression (Koenker, 2005) enables the analyst not just to reduce the impact of influential cases but even investigate what caused them by modelling the extreme quantiles of the outcome variable rather than the median (e.g., 0.80 or 0.90). Similarly, in outcome variables that have

[1] For an introduction to detecting influential cases, see Part 1 of my workshop on non-parametric data analysis.

[2] Note that the application of IRLS is not limited to the reweighting of influential cases, and more often used in the context of normalizing heteroscedastic residuals.

a long-tailed distribution, variables that predict the peak values may not predict the tail values (or vice-versa). Here as well quantile regression can provide insights that OLS regression could not.

A notable downside of the R implementations just cited is that typically they do not offer general ANOVA effects breakdowns, which limits their utility as a general-purpose alternative to OLS regression. Package `quantreg` has more flexibility in that it does provide a `summary` function, which optionally returns *t*-tests for individual parameters, and it also provides an `anova` function, although it requires the user to manually fit a full and a reduced model for the desired comparison. While the summary output is often sufficient, it cannot be used to return omnibus tests for multi-parameter effects (e.g., factor variables with more than 2 levels). For this reason, I decided to create a wrapper function (`AnovaRQ`) that computes a Type III ANOVA table for all effects automatically, and which is attached to this posting (see script hosted on the server).

1. Type III ANOVA for quantile models

The `AnovaRQ` function requires the packages `quantreg`, `insight`, `psych`, and `foreach` to be run. Once the script has been sourced, the function can be applied to a fitted `rq` model. The following example uses data from an experiment in which participants were exposed to a virtual height of 50 meters, while their fear was measured through various mental and bodily channels. Fear feeling itself (outcome variable) was measured with a pressure grip during exposure, with participants instructed to squeeze whenever they felt fear. The maximal force they applied during exposure was taken as their fear intensity. The data look as follows:

```
source("C:/myfilepath/.../AnovaRQ_ver3.r")

fear <- read.csv("https://drive.switch.ch/index.php/s/ NIUebpINqQptjmC/
                  download",header=TRUE,as.is=FALSE)
str(fear)

'data.frame':   81 obs. of  55 variables:
 $ num                : int  1 2 3 4 5 6 7 8 9 10 ...
 $ ID                 : Factor w/ 81 levels "APA","ARD","BEJ",..: 3 50 30 ...
 $ gender             : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 2 1 ...
 $ age                : int  26 26 29 28 23 27 31 32 30 34 ...
 $ mood_anxious       : int  0 1 1 0 2 0 1 1 2 0 ...
 $ phys_heart         : int  6 6 3 6 5 5 5 3 2 6 ...
 $ phys_respir        : int  0 4 3 5 6 5 1 4 1 4 ...
...
```

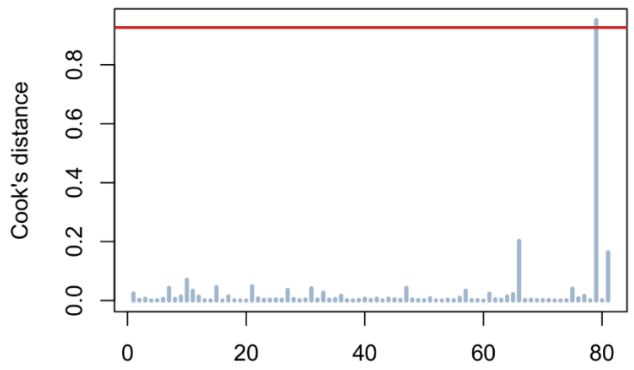
First, we fit a traditional OLS model to the data, predicting the amount of fear experienced (`force_max`) using different mental and bodily predictors, including appraised danger (`appr_danger`), motivation to stop (`mot_stop`), heart rate variability (`HR_SD`), maximal phasic skin conductance (`EDAP_max`), and spontaneous `vocalizations`. In addition, we control for `age` and `gender`. We obtain the traditional Type III ANOVA output using the `Anova` function from package `car`:

```
model <- lm(force_max~appr_danger+mot_stop+HR_SD+EDAP_max+
              vocalizations+gender+age, data=fear)
Anova(model,type=3)

> Anova Table (Type III tests)
> Response: force_max
>                  Sum Sq Df F value    Pr(>F)
> (Intercept)    10437995  1  4.9655 0.028938 *
> appr_danger     8579285  1  4.0813 0.047030 *
> mot_stop       16795741  1  7.9900 0.006065 **
> HR_SD             86339  1  0.0411 0.839961
> EDAP_max       16024383  1  7.6230 0.007284 **
> vocalizations     82331  1  0.0392 0.843670
> gender             1995  1  0.0009 0.975507
> age            16589334  1  7.8918 0.006369 **
> Residuals     153453187 73
> ---
> Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

While none of the effects reach significance below $\alpha = 0.005$, strong trend effects are suggested for motivation to stop, phasic skin conductance, and age. However, the outcome variable has a right-skewed (possibly multi-modal) distribution, and includes several extreme values (e.g., stronger participants were able to apply more force to the pressure grip), suggesting that OLS results may be distorted by influential cases and/or that different emotion variables predict different quantiles of fear intensity.

```
P <- length(coef(model))
N <- length(residuals(model))
threshold <- qf(0.5,P,N-P)
plot(cooks.distance(model),type="h",ylab="Cook's distance",lwd=3)
abline(h=threshold,col="firebrick3",lwd=2)
```



Influence diagnostics with Cook's distance turn up only one value that truly exceeds the threshold to be considered influential, suggesting that the OLS model may not be substantially affected

by this issue. Nevertheless, because we are interested in modelling the different quantiles of fear intensity, we fit a quantile regression model to gain more insight, using the median (50%) and the two quartiles (25%, 75%) as our target quantiles of interest (`tau` parameter in `rq`):

```
model <- rq(force_max~appr_danger+mot_stop+HR_SD+EDAP_max+
            vocalizations+gender+age, data=fear, tau=c(0.25,0.50,0.75))
summary(model, se="nid")
AnovaRQ(model, test="wald", type="within")
```

Because we do not have multi-parameter effects in these data, we could get effects tests through the summary function (note the `se="nid"` argument!). Even so, `AnovaRQ` will return a more compact effects breakdown, which eases quick inspection and comparison.

```
> Quantile regression Type III ANOVA breakdown
> Within-quantile Wald tests:
>
>     tau         effect DF1 DF2       F-value      p-value sig
> 1  0.25   appr_danger   1  73 1.775600e-01 0.6747167799
> 2  0.25      mot_stop   1  73 9.582674e+00 0.0027853040  **
> 3  0.25         HR_SD   1  73 7.942495e-02 0.7788756226
> 4  0.25      EDAP_max   1  73 3.266487e-01 0.5693947446
> 5  0.25 vocalizations   1  73 5.644033e-04 0.9811111225
> 6  0.25        gender   1  73 2.332030e-03 0.9616161270
> 7  0.25           age   1  73 2.458445e-02 0.8758398177
> 8  0.50   appr_danger   1  73 7.412232e+00 0.0080974053  **
> 9  0.50      mot_stop   1  73 8.518047e+00 0.0046721592  **
> 10 0.50         HR_SD   1  73 1.278841e-02 0.9102729229
> 11 0.50      EDAP_max   1  73 5.665385e+00 0.0199155057   *
> 12 0.50 vocalizations   1  73 1.113302e-02 0.9162580370
> 13 0.50        gender   1  73 6.463020e-01 0.4240481240
> 14 0.50           age   1  73 3.062801e+00 0.0843053434   .
> 15 0.75   appr_danger   1  73 9.579329e-02 0.7578192147
> 16 0.75      mot_stop   1  73 5.309643e+00 0.0240552241   *
> 17 0.75         HR_SD   1  73 6.240553e-02 0.8034351630
> 18 0.75      EDAP_max   1  73 1.451003e+01 0.0002882121 ***
> 19 0.75 vocalizations   1  73 5.229643e-02 0.8197538615
> 20 0.75        gender   1  73 3.455227e-02 0.8530525337
> 21 0.75           age   1  73 5.218222e+00 0.0252603423   *
> ------------------------
> Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> ------------------------
```

This time we get a slightly different picture. Whereas stop motivation remains predictive of fear intensity across all three quantiles, the effect of appraised danger is limited to the median, while the effect of phasic skin conductance becomes more important at higher quantiles. Age, finally, only affects fear intensity at the highest force values. By default the `test` argument of `AnovaRQ` uses a

Wald *F*-test,[3] but we can replace it with a non-parametric alternative by putting `test="rank"` instead, which typically produces more conservative *p*-values.

A notable feature of quantile regression is that we can test not only effects within quantiles, but also compare them *between* quantiles. This allows to explicitly address hypotheses that a variable affects different quantiles differently. For the fear data:

```
AnovaRQ(model, test="Wald", type="between")

> Quantile regression Type III ANOVA breakdown
> Between-quantile Wald tests:
>
>     tau.pair     parameter DF1 DF2      F-value      p-value sig
> 1   0.25-0.5   appr_danger   1 161  6.416968661 0.0122595541   *
> 2   0.25-0.5      mot_stop   1 161  1.422742755 0.2347073896
> 3   0.25-0.5         HR_SD   1 161  0.034369150 0.8531571371
> 4   0.25-0.5      EDAP_max   1 161  6.282703921 0.0131846186   *
> 5   0.25-0.5 vocalizations   1 161  0.017460440 0.8950401656
> 6   0.25-0.5       genderM   1 161  0.928849834 0.3366073121
> 7   0.25-0.5           age   1 161  3.910775937 0.0496844426   *
> 8  0.25-0.75   appr_danger   1 161  0.008337474 0.9273599636
> 9  0.25-0.75      mot_stop   1 161  0.100502578 0.7516380692
> 10 0.25-0.75         HR_SD   1 161  0.084932504 0.7710962191
> 11 0.25-0.75      EDAP_max   1 161 12.631001229 0.0004981599 ***
> 12 0.25-0.75 vocalizations   1 161  0.042682018 0.8365862167
> 13 0.25-0.75       genderM   1 161  0.033376556 0.8552691345
> 14 0.25-0.75           age   1 161  5.582128728 0.0193393044   *
> 15  0.5-0.75   appr_danger   1 161  2.865717842 0.0924194687   .
> 16  0.5-0.75      mot_stop   1 161  0.396548603 0.5297697386
> 17  0.5-0.75         HR_SD   1 161  0.049070568 0.8249692308
> 18  0.5-0.75      EDAP_max   1 161  1.916934142 0.1681108931
> 19  0.5-0.75 vocalizations   1 161  0.018035691 0.8933354393
> 20  0.5-0.75       genderM   1 161  0.271192495 0.6032484203
> 21  0.5-0.75           age   1 161  1.648173241 0.2010522951
> ------------------------
> Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> ------------------------
```

So only evidence that the effect of phasic skin conductance differs significantly between the the 25% and 75% quantiles of force values. Note that, for between-quantile comparisons, only Wald tests are available and omnibus tests are not possible![4] The function will always return single-parameter tests, which comes with the usual caution (see next section).

[3] In OLS regression the *F*-test always corresponds to a Wald test, whereas in GLMs with non-normal outcome distributions, test statistics can be derived from a Wald test, score test, or likelihood ratio test.

[4] The reason is that it was not implemented in quantreg's `anova` function.

2. Cautions

While applying `AnovaRQ` should be straightforward, the function comes with several cautions. Firstly, it produces a Type III ANOVA breakdown. As a consequence, in models that contain interaction effects, lower-order tests no longer reflect "marginal" effects but conditional contrasts (see my full explanation here). If you are fitting, say, `Y~A*B*C` and you want a quantile ANOVA breakdown of the marginal two-way interactions, then you should first remove the three-way interaction (e.g., by refitting the model as `Y~(A+B+C)^2`). The same logic applies to main effects which are involved in interactions. Misuse and misinterpretation of Type III ANOVA remains the most common and pervasive mistake I encounter in research! For this reason, I have long advocated the use of Type II ANOVA by default. However, for `AnovaRQ` I was not entirely confident that my Type II implementation was correct, and therefore I defaulted back to Type III out of caution.

Secondly, quantile models are more prone to instability issues than traditional OLS. Warnings with `rq` are fairly common, for example non-unique solutions and non-positive covariance matrices. These issues are more likely to occur for small samples, models with many predictors, models with interaction effects, models for extreme quantiles (e.g., 95%), and outcome variables on restricted integer scales (e.g., Likert scale data). My personal recommendation is that this type of model should primarily be used for median regression (tau=0.5), and only for other quantiles when it is required to address an explicit research hypothesis.

Thirdly, the `AnovaRQ` is currently somewhat inefficient in that it literally loops through the source model's effects and iteratively refits a full and reduced model to apply the `anova` function. For large data sets and/or large models this may take some time, especially when the number of quantiles being modelled is also large.

References

Koenker, R.W. (2005). *Quantile Regression*. Cambridge University Press.

--
Ben Meuleman, Ph.D.
Statistician
Swiss Center for Affective Sciences
University of Geneva | Campus Biotech
Chemin des Mines 9 | CH-1202 Genève
ben.meuleman@unige.ch | +41 (0)22 379 09 79