Quantile regression: a classical and Bayesian approach

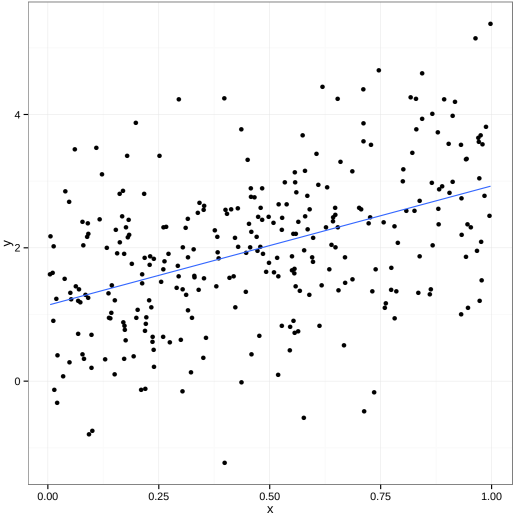
Bruno Santos

Instituto de Matemática e Estatística
Universidade Federal da Bahia

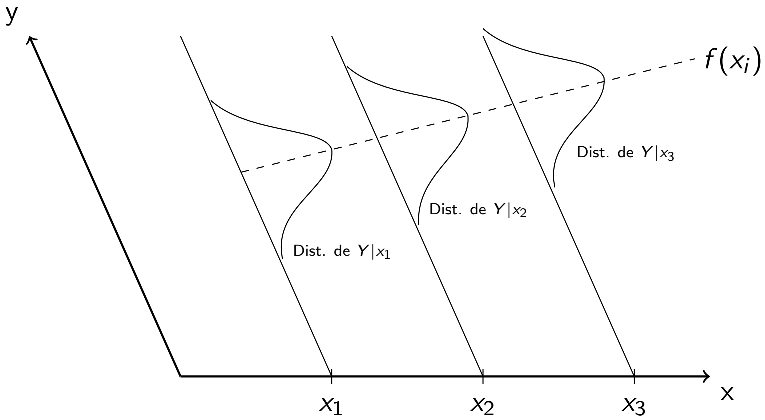
Universidad de Antofagasta
Antofagasta, Chile

Outline

Regression model



Context



$$
\begin{aligned}
E[Y|x_i] &= f(x_i) \\
&= \beta_0 + \beta_1 x_i
\end{aligned}
$$

On regression models

Mosteller and Tukey (1977), pg. 266

What the regression curve does is give a grand summary for the averages of the distribution corresponding to the set of x's. We could go further and compute several different regression curves corresponding to the various percentage points of the distribution and thus get a more complete picture of the set. Ordinarily this is not done, and so the regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions.

Assumptions and estimation methods

Linear regression model

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Least squares estimator

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

Least absolute deviations estimator

$$\min_{\alpha,\beta} \sum_{i=1}^{n} |Y_i - \alpha - \beta X_i|$$

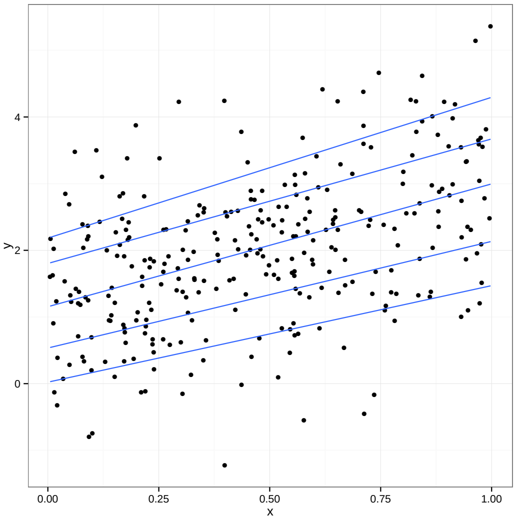
Historical context

Stigler (1986) - The History of Statistics

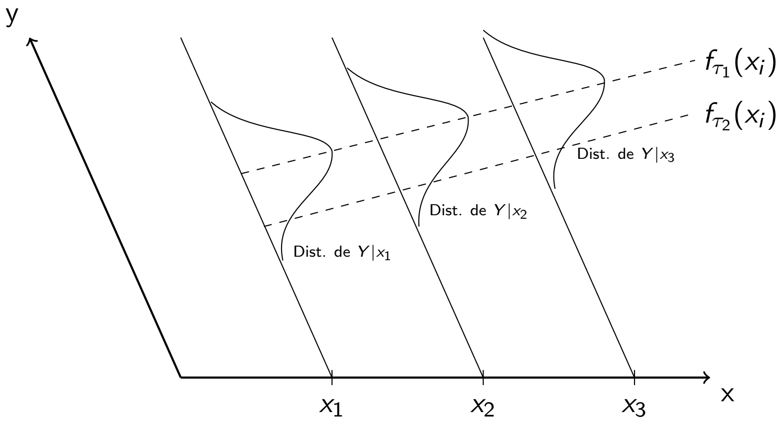
- 1805: least squares estimator
  - Adrien Marie Legendre.
  - Study on orbits of comets.
  - Conditional mean.

- 1757: L_1 estimator
  - Roger Boscovich.
  - Study on the shape of Earth.
  - Conditional median.

Quantile regression models



Quantile regression



$$
\begin{aligned}
Q_Y[\tau_j|x_i] &= f_{\tau_j}(x_i) \\
&= \beta_0(\tau_j) + \beta_1(\tau_j)x_i
\end{aligned}
$$

Usual setting

Consider the usual setting of a regression model

$$Y \sim N(X\beta, \sigma^2).$$

Remember the definition of the τth quantile, which is $m(\tau)$ that

$$P\left(Y < m(\tau)\right) = \tau$$

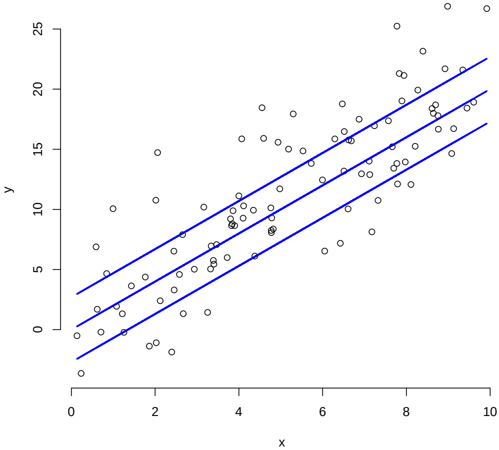
Basic probability theory shows that $m(\tau)$ can be defined as

$$m(\tau) = \sigma\Phi^{-1}(\tau) + X\beta,$$

where $\Phi(z) = P(Z < z), Z \sim N(0, 1)$.

Example

$$Y_i = 1 + 2 \times X_i + \epsilon_i, \epsilon_i \sim N(0, 4^2), i = 1, \cdots, 100.$$



Why quantile regression models?

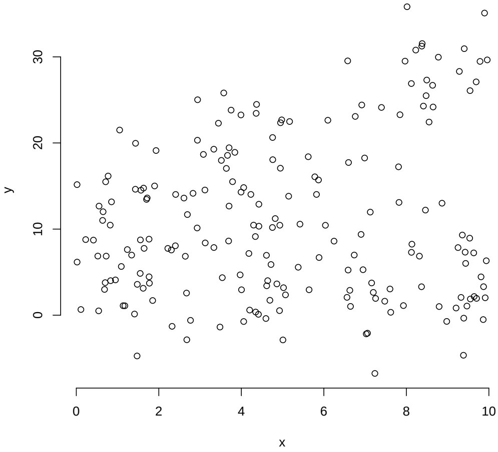
Why do we need quantile regression models then?

Suppose the following scenario,

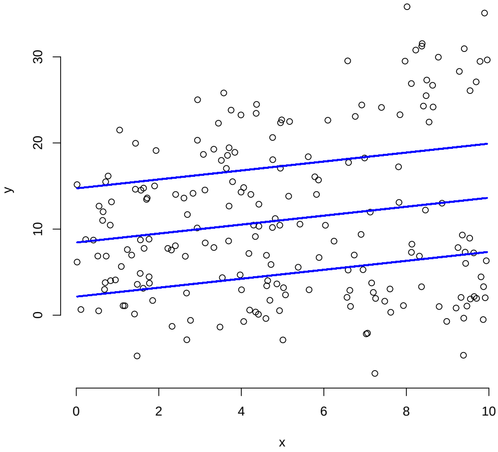
$$Y_i = \begin{cases} \alpha + \epsilon_i, & \text{if } Y_i \leqslant m(\tau) \\ \alpha + \beta \cdot x_i + \epsilon_i, & \text{otherwise} \end{cases}$$

Intuition: the regression coefficient should be significant only for quantiles greater than τ.

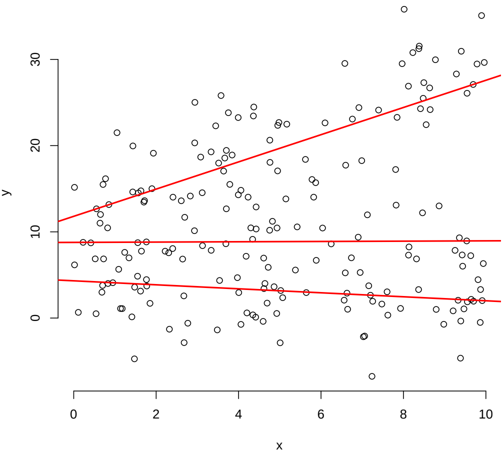
Scenario



Scenario with regular estimates



Scenario with quantile regression estimates



Motivation

nature

Vol 455|4 September 2008|doi:10.1038/nature07234

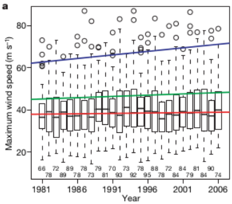
LETTERS

The increasing intensity of the strongest tropical cyclones

James B. Elsner[1], James P. Kossin[2] & Thomas H. Jagger[1]

Atlantic tropical cyclones are getting stronger on average, with a 30-year trend that has been related to an increase in ocean temperatures over the Atlantic Ocean and elsewhere[1–4]. Over the rest of the tropics, however, possible trends in tropical cyclone intensity are less obvious, owing to the unreliability and incompleteness of the observational record and to a restricted focus, in previous trend analyses, on changes in average intensity. Here we overcome these two limitations by examining trends in the upper quantiles of per-cyclone maximum wind speeds (that is, the maximum intensities that cyclones achieve during their lifetimes), estimated from homogeneous data derived from an archive of satellite records. We find significant upward trends for wind speed quantiles above the 70th percentile, with trends as high as $0.3 \pm 0.09\ \mathrm{m\,s^{-1}\,yr^{-1}}$ (s.e.) for the strongest cyclones. We note separate upward trends in the estimated lifetime-maximum wind speeds of the very strongest tropical cyclones (99th percentile) over each ocean basin, with the largest increase at this quantile occurring over the North Atlantic, although not all basins show statistically significant increases. Our results are qualitatively consistent with the hypothesis that as the seas warm, the ocean has more energy to convert to tropical cyclone wind.

To quantify and determine the significance of these trends, we use quantile regression. Quantile regression as employed here is a method to estimate the change (trend) in lifetime-maximum wind speed quantile as a function of year. A quantile is a point taken from



Quantile definition

Definition

Let X be a random variable with cumulative distribution function (cdf)

$$F(x) = P(X \leqslant x).$$

Then using the inverse of the cdf at τ, we can define

$$F^{-1}(\tau) = \inf\{x : F(x) \geqslant \tau\}$$

as the τth quantile of X.

About loss functions

If we define the squared loss function, which is

$$L(X, d) = (X - d)^2.$$

It is well known that, considering the expected loss

$$E\left[(X - \hat{x})^2\right],$$

the predictor $\hat{x}$ which minimizes this value is

$$\hat{x} = E(X)$$

Considering a sample of values

If we consider a sample $X_1, \cdots, X_n$, then the solution of

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \mu)^2.$$

is given by the sample mean, $\bar{X}$.

If we consider a regression setting, where $\mu = x_i^t \beta$, then the solution of

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i^t \beta)^2.$$

is said to be the conditional mean.

Loss function for the quantile regression model

Using loss function

$$\rho_\tau(u) = u(\tau - \mathbb{I}(u < 0)), \quad 0 < \tau < 1,$$

where $\mathbb{I}$ is the indicator function.

Let $\hat{x}$, a predictor of X, which minimizes the expected loss

$$E\big[\rho_\tau(X - \hat{x})\big].$$

One can check that $\hat{x} = F^{-1}(\tau)$ minimizes this expected loss.

Seminal paper

Econometrica, Vol. 46, No. 1 (January, 1978)

REGRESSION QUANTILES[1]

BY ROGER KOENKER AND GILBERT BASSETT, JR.

A simple minimization problem yielding the ordinary sample quantiles in the location model is shown to generalize naturally to the linear model generating a new class of statistics we term "regression quantiles." The estimator which minimizes the sum of absolute residuals is an important special case. Some equivariance properties and the joint asymptotic distribution of regression quantiles are established. These results permit a natural generalization to the linear model of certain well-known robust estimators of location.

Estimators are suggested, which have comparable efficiency to least squares for Gaussian linear models while substantially out-performing the least-squares estimator over a wide class of non-Gaussian error distributions.

Definition of the quantile regression estimator

Koenker and Bassett (1978):

- To obtain $\boldsymbol{\beta(\tau)}$ as solution of the minimization problem.

Weighted least absolute deviation

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x_i}'\beta).$$

Inferential aspects

Estimation methods:

- Linear programming algorithms:
  - simplex method;
  - interior point method.

Observations:

- simplex algorithm is more stable, as it always finds a solution.
- interior point method can have difficulties if there are extreme observations in the explanatory variables.
- the interior point algorithm is faster when there are too many observations.

Example using R

R package: **quantreg**.

Developed by: Roger Koenker.

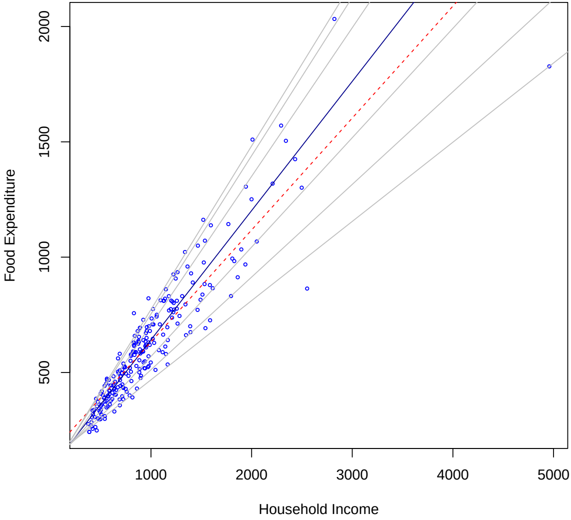
Observations:

- a lot of major contributions to quantile regression were added to this package, such as:
  - quantile regression for censored observations;
  - additive models;
  - dynamic models;
  - penalized regression (LASSO, SCAD);
- it is very fast compared to Bayesian estimation methods, for instance.
- it is very stable.

Example using R

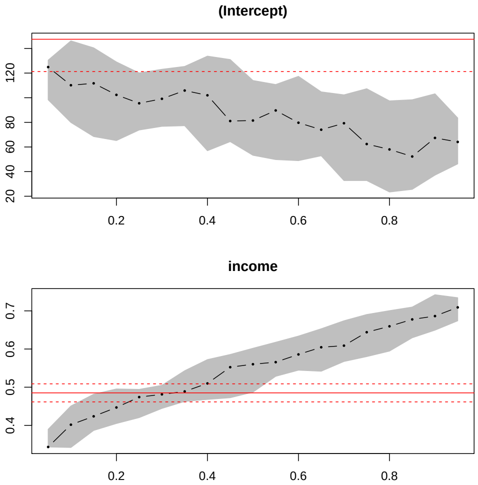
```
data(engel)

plot(foodexp   income, data = engel, cex= .5,
   col = "blue", xlab = "Household Income",
   ylab = "Food Expenditure")

z <- rq(foodexp ~ income, tau= .50,
   data = engel, method = ’br’)

abline(z, col = "dark blue")

abline(lm(foodexp ~ income, data = engel),
   lty=2, col="red")

taus <- c(.05,.1,.25,.75,.90,.95)

for( i in 1:length(taus)) {
   abline(rq(foodexp ~ income, tau=taus[i],
      data = engel), col="gray")
}
```

Example plot



Summary of the results

```
plot(summary(rq(foodexp~income, tau=1:19/20,
    data = engel)))
```



Confidence intervals

Important observation:

- There is no need to assume a probability distribution for the response distribution.

Confidence intervals for quantile regression models:

- Asymptotic results;
- Bootstrap;
- Rank score tests.

Asymptotic results

Considering a few assumptions for the distribution of Y (see Koenker and Bassett, 1978), one can show that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{D} N_p(0, V(\tau)),$$

where

$$V(\tau) = \frac{\tau(1-\tau)}{f^2(0)}(X'X)^{-1}.$$

An estimate of $1/f(0)$ can be obtained using the empirical cdf of the residuals with

$$\frac{\hat{F}^{-1}(\tau + h_n) - \hat{F}^{-1}(\tau - h_n)}{2h_n}$$

where $\lim_{n\to\infty} h_n = 0$

Bootstrap

- Efron e Tibshirani (1993) - Bootstrap for regression models:
  - estimate the covariance matrix of the estimators.
- He e Hu (2002) - Markov Chain Marginal Bootstrap (MCMB), adapted to quantile regression models by Kocherginsky et al. (2005).

With the estimate of the standard errors of $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau})$ one can build confidence intervals with

$$\hat{\beta}_i(\tau) \pm z_{\alpha/2}\widehat{\mathrm{SE}(\hat{\beta}_i(\tau))}.$$

Hypothesis testing - Wald test

Considering the following model

$$Q_{Y_i}(\tau|X) = \beta_1(\tau) + \beta_2(\tau)x_2 + \cdots + \beta_p(\tau)x_p,$$

we could be interested in the hypothesis

$$\beta_2(\tau) = \beta_3(\tau) = \cdots = \beta_p(\tau) = 0.$$

The test statistic for this hypothesis is given by

$$T_n = n\frac{f^2(0)}{\tau(1-\tau)}\sum_{i=2}^{p}\frac{\hat{\beta}_i^2(\tau)}{v_{ii}},$$

where v_{ii} is ith element of the diagonal of the matrix $(X'X)^{-1}$.

Hypothesis testing - Wald test

The test statistic T_n has the following limiting distribution

$$T_n \xrightarrow{D} \chi^2_{p-1}$$

In the R package `quantreg`, one can use

```
data(barro)
fit0 <- rq(y.net ~ lgdp2 + fse2 + gedy2 ,
   data = barro)
fit1 <- rq(y.net ~ lgdp2 + fse2 + gedy2 +
   Iy2 + gcony2, data = barro)
anova.rq(fit1, fit0, test = 'Wald')
```

Hypothesis testing - Analysis of weighted absolute residuals

Chen et al. (2008) propose the following test statistic

$$M_n = \min_{\boldsymbol{\beta} \in \Omega_0} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x_i'}\beta) - \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x_i'}\beta).$$

The authors show that

$$M_n \xrightarrow{\quad D \quad} \frac{\chi^2_q}{4f(0)},$$

where q is the number of restrictions imposed by the hypothesis $H_0 : C\beta\boldsymbol{(\tau)} = \mathbf{c}$, $f(.)$ is the density function of the errors and Ω_0 is the parametric space of the null hypothesis.

However, to avoid estimating $f(0)$, propose the following transformation

$$\begin{aligned} M_n^* &= \min_{\beta \in \Omega_0} \sum_{i=1}^{n} w_i \rho_\tau(y_i - \mathbf{x_i'}\boldsymbol{\beta}) - \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} w_i \rho_\tau(y_i - \mathbf{x_i'}\boldsymbol{\beta}) \\ & \quad - \left(\sum_{i=1}^{n} w_i \rho_\tau(y_i - \mathbf{x_i'}\hat{\boldsymbol{\beta}}_{\mathbf{r}}) - \sum_{i=1}^{n} w_i \rho_\tau(y_i - \mathbf{x_i'}\hat{\boldsymbol{\beta}}_{\mathbf{c}}) \right), \end{aligned}$$

where $w_1, \ldots, w_n$ is a sequence of non-negative r.v. iid with mean and variance equals to 1.

It is possible to show that

$$M_n^* \xrightarrow{D} \frac{\chi_q^2}{4f(0)}.$$

Using probability distributions to study quantile regression

Important observations:

- There is no need for any probability distribution for the estimation of the parameters, using linear programming algorithms.
- Although, when constructing confidence intervals and making hypothesis testing, we need to obtain some estimate of $f(0)$.
- But if we use Bootstrap methods we can avoid these specific difficulties.

A recent proposal in the literature brings the attention to the possibility of using some probability distributions to study these quantile regression models.

Paper - Galarza et al. (2017)

Stat

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.140

Robust quantile regression using a generalized class of skewed distributions

Christian Galarza Morales[a], Victor Lachos Davila[b]*, Celso Barbosa Cabral[c] and Luis Castro Cepero[d]

Received 1 February 2017; Accepted 18 February 2017

It is well known that the widely popular mean regression model could be inadequate if the probability distribution of the observed responses do not follow a symmetric distribution. To deal with this situation, the quantile regression turns to be a more robust alternative for accommodating outliers and the misspecification of the error distribution because it characterizes the entire conditional distribution of the outcome variable. This paper presents a likelihood-based approach for the estimation of the regression quantiles based on a new family of skewed distributions. This family includes the skewed version of normal, Student-*t*, Laplace, contaminated normal and slash distribution, all with the zero quantile property for the error term and with a convenient and novel stochastic representation that facilitates the implementation of the expectation–maximization algorithm for maximum likelihood estimation of the pth quantile regression parameters. We evaluate the performance of the proposed expectation–maximization algorithm and the asymptotic properties of the maximum likelihood estimates through empirical experiments and application to a real-life dataset. The algorithm is implemented in the R package `lqr`, providing full estimation and inference for the parameters as well as simulation envelope plots useful for assessing the goodness of fit.

Keywords: EM algorithm; quantile regression model; scale mixtures of normal distributions

Generalized class of skew-densities (SKD)

Preliminaries:

Skew-normal distribution ($SKN(\mu, \sigma, \tau)$):

$$f(x; \mu, \sigma, \tau) = 2\left[\tau\phi\left(x\middle|\mu, \frac{\sigma^2}{4(1-\tau)^2}\right) I(x < \mu)\right.$$
$$\left. +(1-\tau)\phi\left(x\middle|\mu, \frac{\sigma^2}{4\tau^2}\right) I(x \geqslant \mu)\right]$$

- μ location parameter, $\mu \in \mathbb{R}$.
- σ scale parameter, $\sigma > 0$.
- τ skewness parameter, $\tau \in [0, 1]$.

An important feature of this distribution is

$$P(X \leqslant \mu) = \tau$$

Skew-normal distribution

This density can also be written as

$$f(x;\mu,\sigma,\tau) = \frac{4\tau(1-\tau)}{\sqrt{2\pi\sigma^2}} \exp\left\{-2\rho_\tau^2\left(\frac{x-\mu}{\sigma}\right)\right\},$$

where $\rho_\tau(u) = u(\tau - I(u<0))$.

This family is closed under location and scale transformations:

$$Z \sim SKN(0,1,\tau) \Rightarrow X = \mu + \sigma Z \sim SKN(\mu,\sigma,\tau)$$

Stochastic representation of $SKN(\mu, \sigma, \tau)$

Let $T_0 \sim N(0,1)$ and I with probability function

$$P\left(I = -\frac{1}{2(1-\tau)}\right) = \tau \quad P\left(I = \frac{1}{2\tau}\right) = 1 - \tau$$

be independent. Then, the random variable with representation

$$X = \mu + \sigma I |T_0|$$

follows a $SKN(\mu, \sigma, \tau)$.

Scale mixture of normal distributions (*SMN*)

This class can be conveniently represented as

$$Y = \mu + \sigma\kappa(U)^{1/2} T_0,$$

where μ is a location parameter, $\kappa(.)$ is the weight function, U is a positive random variable with pdf $h(u|\nu)$ and cdf $H(u|\nu)$, ν is the parameter indexing U, and $T_0 \sim N(0,1)$, with T_0 independent of U.

The marginal pdf of W is given by

$$f(w|\mu,\sigma,\nu) = \int_0^\infty \phi(w|\mu,\sigma^2\kappa(u),\nu)dH(u|\nu).$$

Scale mixture of *SKN* distributions

Y follows a skewed distribution if it can represented stochastically as

$$Y = \mu + \sigma\kappa(U)^{1/2}Z,$$

where $Z \sim SKN(0, 1, \tau)$.

A consequence of using $SKN(0, 1, \tau)$ is

$$P(Y \leqslant \mu) = \tau, \quad P(Y > \mu) = 1 - \tau$$

The marginal pdf of Y is

$$f(y|\mu, \sigma, \tau, \nu) = \int_0^\infty \frac{4\tau(1-\tau)}{\sqrt{2\pi\kappa(u)\sigma^2}} \exp\left\{-2\rho_\tau^2\left(\frac{x-\mu}{\kappa^{1/2}(u)\sigma}\right)\right\} dH(u|\nu).$$

Examples of this class

Table from Galarza et al. (2017)

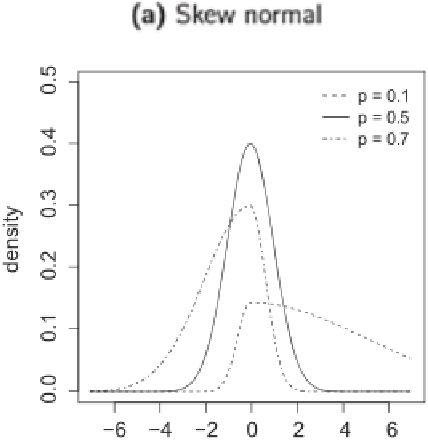
Table I. $\kappa(\cdot), h(u|\nu)$ and *pdf* for some members of the SKD family.

Distribution	$\kappa(u)$	$h(u \mid \nu)$	$f(y \mid \mu, \sigma, \nu)$
Skewed Student-t	u^{-1}	$G(\frac{\nu}{2}, \frac{\nu}{2})$	$\frac{4p(1-p)\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{2\pi\sigma^2}} \left\{\frac{4}{\nu}\rho_p^2\left(\frac{y-\mu}{\sigma}\right) + 1\right\}^{-\frac{\nu+1}{2}}$
Skewed Laplace	u	Exp(2)	$\frac{2p(1-p)}{\sigma} \exp\left\{-2\rho_p\left(\frac{y-\mu}{\sigma}\right)\right\}$
Skewed slash	u^{-1}	Beta$(\nu, 1)$	$\nu \int_0^1 u^{\nu-1} \phi_{skd}(y \mid \mu, u^{-1/2}\sigma, p) du$
Skewed contaminated normal	u^{-1}	$\nu\mathbb{I}\{u = \gamma\} + (1-\nu)\mathbb{I}\{u = 1\}$	$\nu\phi_{skd}(y \mid \mu, \gamma^{-1/2}\sigma, p) + (1-\nu)\phi_{skd}(y \mid \mu, \sigma, p)$

$G(\alpha, \beta)$ denotes the Gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$; Exp(β) denotes the exponential distribution with mean β; Beta(α, β) denotes the beta distribution; and $\phi_{skd}(y|\mu, \sigma, p)$ denotes the *pdf* of the SKN distribution defined in (2) where $0 \le \nu, \gamma \le 1$.

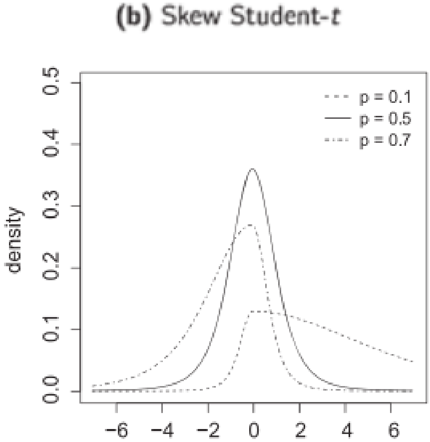
Examples of the densities

Figure from Galarza et al.(2017)



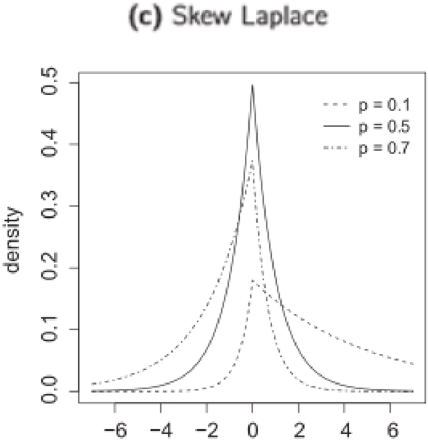
Examples of the densities

Figure from Galarza et al.(2017)



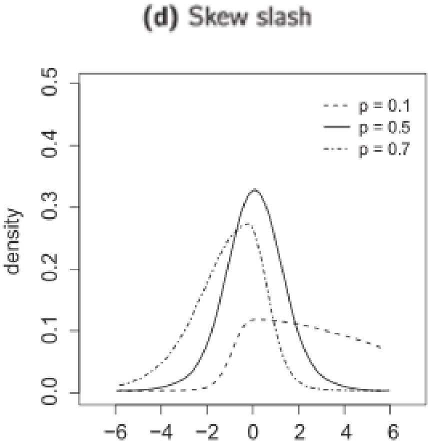
Examples of the densities

Figure from Galarza et al.(2017)



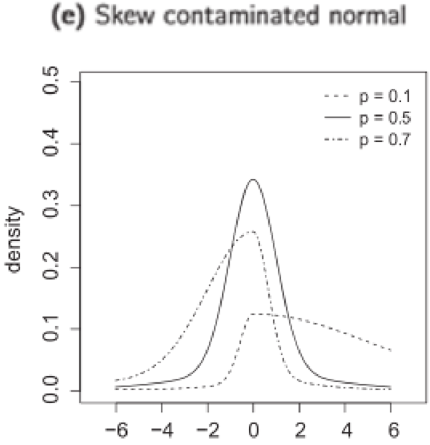
Examples of the densities

Figure from Galarza et al.(2017)



Examples of the densities

Figure from Galarza et al.(2017)



Quantile regression using SKD family

Consider the following linear model

$$Y_i = x_i^t \beta(\tau) + \epsilon_i.$$

If we would like to have

$$Q_{Y_i}(\tau|X) = x_i^t \beta(\tau),$$

then we need $Q_{\epsilon_i}(\tau) = 0$. We could achieve that by considering

$$\epsilon_i \sim SKD(0, \sigma, \tau, \nu)$$

Estimation via EM algorithm

We need to consider the hierarchical representation of the SKD family, then

$$
\begin{aligned}
Y_i|U_i = u_i &\sim SKN(x_i^t\beta(\tau), \sqrt{\kappa(u_i)}\sigma, \tau \\
U_i &\sim h(u_i|\nu)
\end{aligned}
$$

where $h(u_i|\nu)$ represents the mixture density.

Consider $y = (y_1, \cdots, y_n)$ and $u = (u_1, \cdots, u_n)$ the complete and the missing data, respectively. Then, the complete data log-likelihood of $\theta = (\beta(\tau), \sigma, \nu)$ given (y, u) is given by

$$
l_c(\theta|y, u) = \sum_{i=1}^{n} l_c(\theta|y_i, u_i),
$$

where

$$
l_c(\theta|y_i, u_i) = \log \phi\left(y_i|x_i^t\beta(\tau), \frac{\kappa(u_i)\sigma^2}{4\xi_i^2}\right) + \log h(u_i|\nu)
$$

Estimation via EM algorithm (cont.)

$$l_c(\theta|y_i, u_i) = \log \phi\left(y_i|x_i^t\beta(\tau), \frac{\kappa(u_i)\sigma^2}{4\xi_i^2}\right) + \log h(u_i|\nu),$$

where

$$\xi_i = (1-\tau)I(y_i \leqslant x_i^t\beta(\tau) + \tau I(y_i > x_i^t\beta(\tau).$$

In the E step, we need to evaluate the $Q(\theta|\theta^{(k)})$ function, where

$$\begin{aligned} Q(\theta|\theta^{(k)}) &= E[l_c(\theta|y, u)|y, \theta^{(k)}] \\ &\propto -n\log\sigma - 2\sum_{i=1}^{n}\{\widehat{\kappa^{-1}(u_i)}\xi_i^2 z_i^2\} + \sum_{i=1}^{n} E[\log h(u_i|\nu)|y_i, \theta^{(k)}] \end{aligned}$$

where

$$\widehat{\kappa^{-1}(u_i)} = E[\kappa^{(-1)}(U_i)|y_i, \theta^{(k)}], \quad z_i = (y_i - x_i^t\beta(\tau))/\sigma$$

Results for the EM algorithm

Table from Galarza et al. (2017)

Table III. Conditional distribution of U given Y for specific SKD distributions.

Distribution	Distribution of U	Conditional distribution of $U\|Y$	$\widehat{\kappa^{-1}(u_i)}$
Skewed Student-t	$\mathrm{G}(\frac{\nu}{2}, \frac{\nu}{2})$	$\mathrm{G}\left(\frac{\nu+1}{2}, \frac{\nu+4\xi_i^2 z_i^2}{2}\right)$	$\frac{\nu+1}{\nu+4\xi_i^2 z_i^2}$
Skewed Laplace	$\mathrm{Exp}(2)$	$\mathrm{GIG}\left(\frac{1}{2}, 2\xi_i^2 z_i^2, \frac{1}{2}\right)$	$\frac{1}{2\xi_i \|z_i\|}$
Skewed slash	$\mathrm{Beta}(\nu, 1)$	$\mathrm{TG}\left(\nu+\frac{1}{2}, 2\xi_i^2 z_i^2, 1\right)$	$\left[\frac{\nu+\frac{1}{2}}{2\xi_i^2 z_i^2}\right]\frac{\mathcal{F}(1\mid\nu+\frac{3}{2}, 2\xi_i^2 z_i^2)}{\mathcal{F}(1\mid\nu+\frac{1}{2}, 2\xi_i^2 z_i^2)}$
Skewed contaminated normal	$\nu\mathbb{I}\{u=\gamma\}+(1-\nu)\mathbb{I}\{u=1\}$ $0 \le \nu, \gamma \le 1$	$\frac{a\,\mathbb{I}\{u=\gamma\}+b\,\mathbb{I}\{u=1\}}{a+b}$	$\frac{a\gamma+b}{a+b}$

EM algorithm

Summary of the algorithm, at the kth step:

1. **E step:** Given $\theta^{(k)}$, compute $\widehat{\kappa^{-1}(u_i)}$.

2. **M step:** Update $\theta^{(k)}$ by maximizing $Q(\theta|\theta^{(k)})$ over θ, with these expressions

$$\hat{\beta}(\tau)^{(k+1)} = (X^t\Omega^{(k)}X)^{-1}X^t\Omega^{(k)}y,$$

$$(\hat{\sigma}^2)^{(k+1)} = \frac{4}{n}\left(y - X\hat{\beta}(\tau)^{(k+1)}\right)^t \Omega^{(k)}\left(y - X\hat{\beta}(\tau)^{(k+1)}\right),$$

where Ω is a diagonal matrix with entries $\xi_i^2\widehat{\kappa^{-1}(u_i)}$. After the M step, ν is updated with

$$\hat{\nu}^{(k+1)} = \arg\max_{\nu} \log f(y_i|\hat{\beta}(\tau)^{(k+1)}, \hat{\sigma}^{(k+1)}, \nu)$$

Standard error approximation

Using Louis (1982) missing information principle, one can obtain the score function based on the log-likelihood of the complete data,

$$
\begin{aligned}
\frac{\partial l_{c_i}}{\partial \beta(\tau)} &= \frac{4}{\sigma}[\Omega_1 X]_i \\
\frac{\partial l_{c_i}}{\partial \sigma} &= \frac{4}{\sigma^3}\left(y - X\hat{\beta}(\tau)\right)^t \Omega \left(y - X\hat{\beta}(\tau)\right) - \frac{1}{\sigma}
\end{aligned}
$$

where Ω_1 is a diagonal matrix with entries

$$\xi_i^2 z_i \widehat{\kappa^{-1}(u_i)}$$

and $[.]_i$ denotes the ith element of the vector.

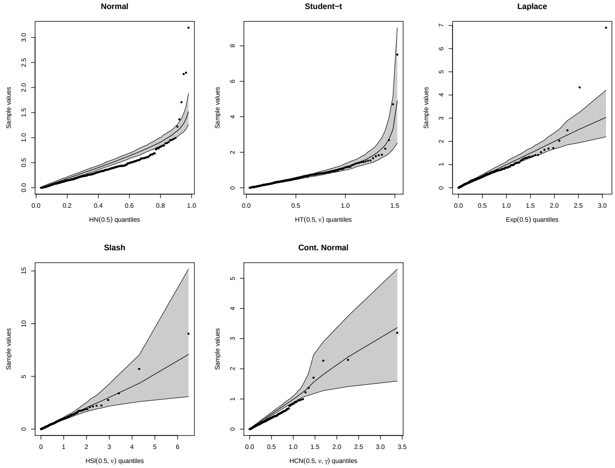
R package `lqr`

For these models, one can use `lqr`.

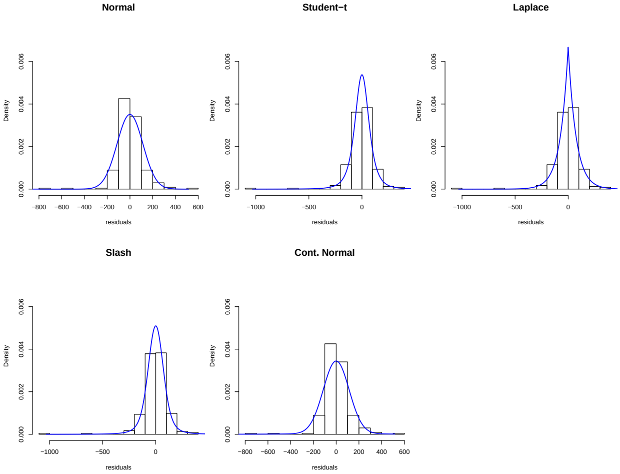
For instance, the following code is able to compare the fit for a few distributions of the family, in the median case:

```
library(lqr)
data(engel)
y <- engel$foodexp
X <- cbind(1, engel$income)
res = best.lqr(y, X, p = 0.55, criterion = "AIC")
```

Plots produced



Plots produced



Paper - Santos and Elian (2014)

Communications in Statistics—Theory and Methods, 44: 1842–1853, 2015

ISSN: 0361-0926 print / 1532-415X online
DOI: 10.1080/03610926.2013.799699

Taylor & Francis
Taylor & Francis Group

Influence Measures in Quantile Regression Models

BRUNO R. SANTOS AND SILVIA N. ELIAN

Departmento de Estatística, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

In this article, we use the asymmetric Laplace distribution to define a new method to determine the influence of a certain observation in the fit of quantile regression models. Our measure is based on the likelihood displacement function and we propose two types of measures in order to determine influential observations in a set of conditional quantiles conjointly or in each conditional quantile of interest. We verify the validity of our average measure in a simulated data set as well in an illustrative example with data about air pollution.

Likelihood displacement definition

The likelihood displacement function (LD) suggested by Cook and Weisberg (1982) could be defined as

$$LD_i(\theta) = 2[L(\hat{\theta}, y) - L(\hat{\theta}_{(i)}, y)], \quad i = 1, \ldots, n,$$

where $\hat{\theta}$ is the MLE of θ based on all observations and $\hat{\theta}_{(i)}$ is the MLE based on all observations except the observation i.

If the value for the ith observation is large, then this observation is influential, because its deletion may cause changes in important conclusions of the model fitted.

Likelihood displacement in quantile regression models

In quantile regression models, $\theta = (\beta_\tau, \sigma)$.

Then, we have that the LD for the ith observation is

$$LD_i(\beta(\tau), \sigma) = 2\left[n \log\left(\frac{\hat{\sigma}_{(i)}}{\hat{\sigma}}\right) + \frac{\rho_\tau(y_i - x_i'\hat{\beta}(\tau(i)))}{\hat{\sigma}_{(i)}} - 1\right],$$

where $\hat{\beta}_{\tau(i)}$ is the QR estimator of $\beta(\tau)$ based on all observations except observation i and $\hat{\sigma}_{(i)}$ is given by

$$\hat{\sigma}_{(i)} = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} \rho_\tau(y_j - x_j'\hat{\beta}(\tau(i))).$$

About the likelihood displacement in quantile regression models

Remarks about the measured proposed:

- It will be large
  - ◇ if the ratio between the estimates of the scale parameter with and without the observation i is too big;
  - ◇ if $\rho_\tau(y_i - x_i'\hat{\beta}_{\tau(i)})/\hat{\sigma}_{(i)}$ is large
- the likelihood displacement function will be close to zero when the observation i is not influential;
- large values could denote observations that might be influential.

Conditional likelihood displacement in quantile regression models

- Possible influence of the observation i only in the parameter β_τ.
- Using a conditional likelihood function on σ.

For this, we must consider the following measure

$$LD_i(\beta(\tau)|\sigma) = 2[L(\hat{\beta}(\tau), \hat{\sigma}) - \max_{\sigma} L(\hat{\beta}(\tau(i)), \sigma)].$$

We should note that

$$L(\hat{\beta}(\tau(i)), \sigma) = n \log \tau(1-\tau) - n \log \sigma - \frac{\sum_{j=1}^{n} \rho_\tau(y_j - x_j' \hat{\beta}(\tau(i)))}{\sigma}$$

is maximized when $\hat{\sigma}_m = \sum_{j=1}^{n} \rho_\tau(y_j - x_j' \hat{\beta}_{\tau(i)})/n$.

Conditional likelihood displacement in quantile regression models

The conditional LD function on σ is given by

$$LD_i(\beta_\tau|\sigma) = 2n\log\left(\frac{\sum_{j=1}^{n}\rho_\tau(y_j - x_j'\hat{\beta}_{\tau(i)})}{\sum_{j=1}^{n}\rho_\tau(y_j - x_j'\hat{\beta}_\tau)}\right).$$

- If the ith observation is not influential, then this measure $LD_i(\beta_\tau|\sigma)$ will be close to zero.
- Because we compare the sum of the weighted absolute residuals with and without the observation i and we expect that these values are close when the deletion of this observation does not affect much the fitted model.

Air pollution in US cities

- Data about the air pollution in 41 US cities.
- Response: annual mean concentration of sulphur dioxide ($\mu g/m^3$).
- For the explanatory variables, we have:
  - TEMP: Average annual temperature in degrees Fahrenheit,
  - MAN: Number of manufacturing enterprises employing 20 or more workers,
  - POP: Population size (1970 census), in thousands,
  - WIND: Average annual wind speed in miles per hour,
  - RAIN: Average annual precipitation in inches.

We consider the following linear model

$$Q_{SO2}(\tau|X) = \beta_0 + \beta_1 \text{TEMP} + \beta_2 \text{MAN} + \beta_3 \text{POP} + \beta_4 \text{WIND} + \beta_5 \text{RAIN},$$

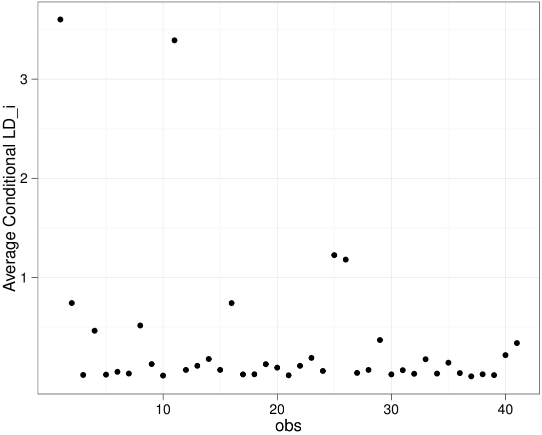
for $\tau = 0.25, 0.50, 0.75$.

Estimates and confidence interval for the parameters

	Conditional quantiles		
Variables	0.25	0.50	0.75
(Intercept)	81.863	96.464	121.028
	[37.036;139.126]	[81.954;144.567]	[16.777;157.411]
TEMP	-0.792	-0.859	-1.118
	[−1.686;−0.168]	[−1.865;−0.723]	[−1.868;−0.343]
MAN	0.042	0.055	0.063
	[−0.025;0.110]	[0.0364;0.082]	[0.033;0.072]
POP	-0.010	-0.028	-0.035
	[−0.081;0.018]	[−0.056;−0.005]	[−0.046;−0.017]
WIND	-4.446	-3.789	-4.688
	[−7.100;−0.192]	[−6.841;−1.883]	[−6.481;2.349]
RAIN	0.277	0.174	0.380
	[0.084;1.379]	[0.084;0.840]	[0.047;0.705]

Influential observations

We consider the average conditional likelihood displacement function.



About the influential observations

- The observations #1 and #11 as the ones that stand out from the others.
- City number one is the city of Phoenix and city number 11 is the city of Chicago.

By removing these cities to fit the quantile regression models:

- we obtain different point estimates;
- but the conclusions that we reach with the confidence intervals only change with the deletion of Chicago.

Estimates and confidence interval for the parameters without Chicago.

	Conditional quantiles		
Variables	0.25	0.50	0.75
(Intercept)	80.508	94.889	118.954
	[30.927;146.903]	[76.748;154.469]	[20.328;162.528]
TEMP	-0.867	-0.799	-1.086
	[−1.962;−0.252]	[−1.839;−0.579]	[−2.402;−0.823]
MAN	0.014	0.075	0.065
	[−0.034;0.105]	[0.004;0.090]	[0.053;0.073]
POP	0.002	-0.040	-0.036
	[−0.085;0.033]	[−0.054;0.006]	[−0.049;−0.019]
WIND	-2.579	-3.726	-4.660
	[−6.987;−0.786]	[−6.712;−2.616]	[−6.352;3.193]
RAIN	0.104	0.113	0.381
	[0.081;1.324]	[0.061;0.876]	[0.082;0.691]

Explanation of the influential observation

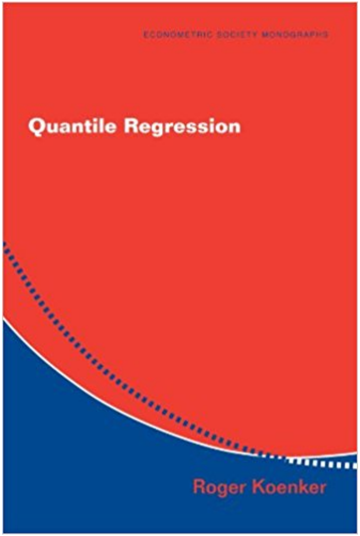
Remarks about the influential observation:

- Instead of having an impact in the conditional median and in the 0.75 quantile, the variable POP should not be considered significant in the conditional median.
- One possible explanation for this change is that the city of Chicago is the city with the biggest population among the cities of the sample.

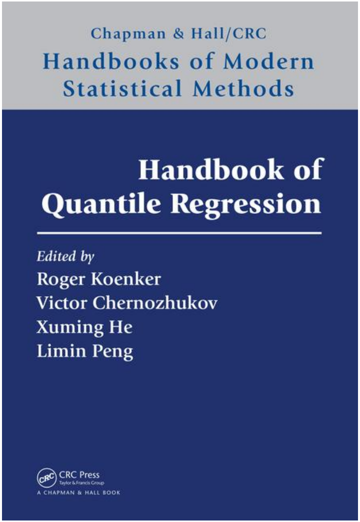
Discussion

- Quantile regression models are still a great source for applications studies.
  - these models are able to show a new perspective for the data.
- Further developments considering other probability distributions could be studied as well.
- Specific areas could be targeted to propose new models, where quantile estimates could prove to be helpful.
  - Survival analysis;
  - Dynamic modeling;

Book - Koenker (2005)



Handbook of Quantile Regression



¡Gracias!

brunorsantos@ufba.br