FA23-Quantile Regression

2023-10-07

Table of contents

1

2 Introduction

Quantile regression (QR), like any regression model, illustrates the relationship between a response variable and one or more predictor variables. QR differs from traditional regression models, such as ordinary least squares (OLS) regression, in that it estimates the conditional *quantiles* of a response variable, given the predictors' values, as opposed to the conditional *mean* in OLS regression.

Due to its formula illustrated later, QR has several advantages over OLS, including relaxed assumptions, efficiency in non-Gaussian scenarios, and a broader perspective compared to traditional models. Unlike OLS, QR does not assume the normality of the conditional response variable distribution and is robust to heteroskedasticity. Furthermore, by considering the entire conditional distribution, QR offers a comprehensive understanding of distributions with higher moments—i.e., those with non-zero skewness, kurtosis, or even greater moments which may be significant in extreme distributions such as in financial data— enabling a detailed examination of their shape, asymmetry, and heavy-tailed characteristics. This makes QR a valuable tool for investigating extreme quantiles, which are of particular interest in fields such as epidemiology, and capturing the entire range of the distribution beyond the central tendency and variability, offering insights beyond traditional regression methods.

3 library(DHARMa)

```r
library(ggplot2)
#install.packages("lme4")
library(lme4)
#install.packages("DHARMa")
library(quantreg)
library(dplyr)
library(ggplot2)
library(tinytex)
```

4 Methods

4.1 Design Matrix

The design matrix is defined to be a matrix $\mathbf{X}$ such that $\mathbf{X}_{ij}$ (the j^{th}) column of the i^{th} row of $\mathbf{X}$) represents the value of the j^th variable associated with the i^{th} variable object.

A regression model may be represent via matrix multiplication as

$$y = \mathbf{X}\beta + e$$

where X is the design matrix, β is a vector of the model's coefficient (one for each variable), e is a vector of random errors with a mean zero, and y is the vector outputs for each object.

4.2 Ordinary least squares

Ordinary least squares model or OLS, works by creating a line through the data points. Then it calculates the difference between each prediction and observation (residual). And it tries to minimize the squared value of the residuals. The ordinary least squares is defined by:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

The least squares estimates in this case are given by simple formulas

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

4.3 How does the minimization of absolute deviations equal the media?

4.3.1 Definition of mean

Assume, without loss of generality, that Y is a continuous random variable. The expected value of the absolute sum of deviations from a given center c can be split into the following two terms:

$$E|Y-c| = \int_{y \in R} |y-c| f(y)dy = \int_{y<c} |y-c| f(y)dy + \int_{y>c} |y-c| f(y)dy$$

If y is less than c, then y-c will always be negative. Therefore, |y-c|=-(c-y). By a similar argument, |y-c| is just (y-c) when y > c.

$$= \int_{y<c} (c-y) f(y)dy + \int_{y>c} (y-c) f(y)dy$$

Since the absolute value is convex, differentiating E|y-c| with respect to c and setting the partial derivatives to zero will lead to the solution of the minimum.

$$\frac{\partial}{\partial c} E|y-c| = 0$$

$$\left\{ (c-y)f(y)\Big|_{-\infty}^{c} + \int_{y<c} \frac{\partial}{\partial c}(c-y)f(y)dy \right\} +$$
$$\left\{ (y-c)f(y)\Big|_{c}^{+\infty} + \int_{y>c} \frac{\partial}{\partial c}(y-c)f(y)dy \right\} = 0$$

The limit of any PDF approaching positive or negative infinity will equal 0, therefore the previous equation simplifies to:

$$\left\{ \int_{y<c} \frac{\partial}{\partial c}(c-y)f(y)dy \right\} +$$
$$\left\{ \int_{y>c} \frac{\partial}{\partial c}(y-c)f(y)dy \right\} = 0$$

Taking the partial, $\frac{\partial}{\partial c}(c-y)f(y) =$ f(y) and $\frac{\partial}{\partial c}(y-c)f(y) =$ -f(y).

$$\left\{ \int_{y<c} \theta f(y)dy \right\} +$$
$$\left\{ \int_{y>c} -\theta f(y)dy \right\} = 0$$

Using the CDF definition and the notion of reciprocals, the previous equation simplifies to: $F(c) - [1 - F(c)] = 0$ and thus $2F(c) - 1 = 0 \longrightarrow F(c) = \frac{1}{2} \longrightarrow$ c=Me.

Thus the minimization to a weighted least absolute deviation loss function is the value that gives the theta^{th} quantile.

5 Generalization least absolute deviations

The solution of the minimization problem formulated in Equation (1.2) is thus the median. The above solution does not change by multiplying the two components of $E|Y-c|$ by a constant θ and $(1-\theta)$, respectively. This allows us to formulate the same problem for the generic quantile θ. Namely, using the same strategy for Equation (1.5), we obtain:

$$\frac{\partial}{\partial c}E\left[\rho_\theta(Y-c)\right] = \frac{\partial}{\partial c}\left\{(1-\theta)\int_{-\infty}^{c}|y-c|f(y)dy + \theta\int_{c}^{+\infty}|y-c|f(y)dy\right\}.$$

Repeating the above argument, we easily obtain:

$$\frac{\partial}{\partial c}E\left[\rho_\theta(Y-c)\right] = (1-\theta)F(c) - \theta(1-F(c)) = 0$$

and then q_θ as the solution of the minimization problem:

$$F(c) - \theta F(c) - \theta + \theta F(c) = 0 \Longrightarrow F(c) = \theta \Longrightarrow c = q_\theta.$$

, interpreting Y as a response variable and $\mathbf{X}$ as a set of predictor variables, the idea of the unconditional median can be extended to the estimation of the conditional median function:

$$\hat{\mu}\left(\mathbf{x}_i, \beta\right) = \underset{\mu}{\operatorname{argmin}} E\left[\left|Y - \mu\left(\mathbf{x}_i, \beta\right)\right|\right],$$

In the case of a linear mean function, $\mu(x_i, \beta) = x_i^T\beta$ so the previous equation becomes:

$$\hat{\beta} = \underset{\beta}{argmin} E[|Y - x_i^T\beta|]$$

By the same argument,

$$q_\theta = \underset{c}{\operatorname{argmin}} E\left[\rho_\theta(Y-c)\right]$$

where $\rho_\theta(.)$ denotes the following loss function: