Application of Quantile regression in clinical research: An overview with the help of R and SAS statistical package

Editor, International Journal of Statistics and Medical Informatics (ISJMI)

Abstract

Normally the relationship between two variables \( x \) and \( y \) is studied using the linear regression equation. Linear regression equation requires normality and homoscedasticity (equal variance) assumptions. When the normality and homoscedasticity assumptions are violated, linear regression estimates are not valid. Quantile regression method overcomes the drawbacks of linear regression and can be applied when the data is skewed and equal variance assumptions are violated. This paper provides an overview of application of quantile regression in clinical research using R and SAS statistical package.

Keywords: Quantile Regression; Linear Regression; SAS; R package

1. Introduction

Linear regression [1] is used to study the relationship between a dependent variable (response variable) and a set of independent variables (predictors). Linear regression provides the average (mean) value of a dependent variable for the specified values of the independent variables. Linear regression is useful when the data is normal and equal variance is assumed. If the data is having outliers and the distribution is skewed either outliers needs to be removed or data to be transformed to make it normal. Sometimes the removal of outliers is not possible as we may lose important information in the data. Hence an alternative methodology which can handle outliers needs to be adopted. Quantile regression [2, 3,4] is one such methodology which can overcome the problem of outliers. Quantile regression estimates quantiles (percentiles) of the dependent variable based on the quantiles (percentiles) of the independent variables. The quantile regression is used in different fields where the data is skewed such as finance and economics [2 , 3], environmental science and clinical research.

The following table-1 provides the difference between linear and quantile regression method [5]

<table>
<thead>
<tr>
<th>Linear Regression</th>
<th>Quantile Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses mean in estimating the response variable</td>
<td>Uses quantiles in estimating the response variable</td>
</tr>
<tr>
<td>Affected by outliers</td>
<td>Not affected by outliers</td>
</tr>
<tr>
<td>Equal variance assumption is needed</td>
<td>Equal variance assumption not necessary</td>
</tr>
<tr>
<td>Assumes normality</td>
<td>Distribution free</td>
</tr>
</tbody>
</table>

Table-1: Linear Regression vs. Quantile Regression

2. Quantile regression methodology

Quantile regression function [2] is defined as follows

For the \( p^{th} \) percentile the dependent variable \( y \) is defined as
i. \[ Q_p(Y) = \inf (y : F(y) \geq p) \] where \( p \) is the quantile (0.25, 0.5 and 0.75)

Conditional quantile function of \( y \) given \( x \) is

ii. \[ Q_p(Y/X=x) = \inf (y: F(y/x) \geq p) \]

The quantile regression equation is given by

iii. \[ Q_p(Y/x) = x^p b(p), 0 < p < 1 \]

Where \( b(p) \) is the quantile coefficient for the quantile values \( p \)

In quantile regression the absolute error function will be minimized to obtain the quantile regression coefficients \( b(p) \) and it can be done through linear programming methodology

iv. \[ \text{min} \sum (|y - x^p b|) \]

3. Examples of application of Quantile regression in clinical research [6, 7]
   a. Quantile regression to estimate probabilities of Low birth weight

When we carry out the analysis to find out the probability of mothers having low birth weight babies, linear regression model based on the average birth weight as a function of different predictors will leave the lower birth weight categories which is not correct as it losses the important information due to the skewness in the distribution. Quantile regression will be useful in these scenarios as it includes both the lower, middle and upper quantiles in predicting the probabilities of lower birth weight

The following example uses the dataset set from Hosmer book [8] and models the child birth weight based on the mother’s weight. SAS Procedure quantreg [9] is used to carried out the quantile regression analysis

**SAS Code**

```sas
proc quantreg data=lbw;
model bwt = lwt/quantile = 0.05 to 0.95 by 0.05
plot=quantplot;
```
The results provided the information related to different quantiles which clearly shows the effect of mother’s weight on the child birth weight and it captures the information related to extreme values which is not captured by the linear regression as shown below in figure-3 (Child birth weight with respect to mother’s weight using Linear regression model)
Linear Regression using SAS

```sas
proc reg data=lbw;
model bwt = lwt;
```

**Figure 3**: Fit for child birth weight with respect to mothers weight

Low birth weight example using R statistical package\[10\]: The above mentioned low birth weight dataset is used for carrying out the quantile regression and linear regression through R statistical package quantreg[11].

```r
library(quantreg)
library(xlsx)
data1 <- read.xlsx("new.xlsx",1,header=T)
summary(data1$LWT)
lw<- rq(data1$BWT~data1$LWT,tau=seq(0.2, 0.8, by=0.1))
lw1<- lm(data1$BWT~data1$LWT)
summary(lw)
summary(lw1)
plot(lw)
plot(lw1)
```

The following figure shows relationship between the birth weight of babies and mother weight with respect to quantiles ranging from 0.05 to 0.95 using quantile regression

**Figure 4**: Figure-3 : Fit for child birth weight with respect to mothers weight
The following figure-5 shows relationship between residual vs fitted values with respect to the birth weight of babies and mother weight using linear regression which is affected by the extreme values especially the lower birth weight.

Hence from the above results it can concluded that the quantile regression is capable of handling the extreme values than the linear regression.

b. Analysis of health care expenditures

Quantile regression is used to analyse the relationship between health care expending [12, 6] and other variables such as race, ethnicity, gender and other economic factors as normally income and expenditure data distributions are skewed and normal regression estimates may be affected by outliers.

- Quantile regression in survival analysis

Quantile regression can be used to obtain the quantile (percentile) estimates of the survival curve and assess the effects of covariates on the survival rate [13, 14, 15]. When the survival function is given as

\[ S(x) = 1 - F(x) = 1 - p \]

Where \( F(x) = F(X \leq x) = p \), \( x \) is the time to event variable.

Then Quantile estimate \( Q(p) \) is given by

\[ Q(p) = F^{-1} = x \]

Conclusion

The paper provided an overview of quantile regression in the clinical research filed using SAS and R statistical package and also discussed the difference between the quantile regression and the linear regression methods.

References


11. KOENKER, R. QUANTILE REGRESSION IN R: A VIGNETTE.


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.