Cornell University
Cornell Statistical Consulting Unit

StatNews #70: Quantile Regression

November 2007

Updated 2012

Linear regression is a statistical tool used to model the relation between a set of predictor variables and a response variable. It estimates the mean value of the response variable for given levels of the predictor variables. Suppose we are interested in investigating the relationship between infants' birth weight (in grams) and a set of predictors, such as: the gender of the infant, marital status of the mother, prenatal care, and smoking status of the mother during pregnancy. The data set used for this example is a subset of 50,000 observations from a study on birth weight that was carried out by Koenker and Hallock in 2001. The linear regression model for this example is:

Y= 3224 + 115. 9*Boy + 161.1 * Married - 227*Prenatal_Care - 200.9*Smoke.

This model estimates how, on average, these mothers' characteristics affect the birth weights of infants. The prenatal care predictor variable compares the effect of prenatal care for babies born to mothers who received no prenatal with babies born to mothers who had a prenatal visit in the first trimester.

While this model can address the question "is prenatal care important?" it cannot answer an important question: "does prenatal care influence birth weight differently for infants with low birth weight than for those with average birth weights? ".

A more comprehensive picture of the effect of the predictors on the response variable can be obtained by using Quantile regression. Quantile regression models the relation between a set of predictor variables and specific percentiles (or quantiles) of the response variable. It specifies changes in the quantiles of the response. For example, a median regression (median is the 50^{th} percentile) of infant birth weight on mothers' characteristics specifies the changes in the median birth weight as a function of the predictors. The effect of prenatal care on median infant birth weight can be compared to its effect on other quantiles of infant birth weight.

In linear regression, the regression coefficient represents the increase in the response variable produced by a one unit increase in the predictor variable associated with that coefficient. The quantile regression parameter estimates the change in a specified quantile of the response variable produced by a one unit change in the predictor variable. This allows comparing how some percentiles of the birth weight may be more affected by certain mother characteristics than other percentiles. This is reflected in the change in the size of the regression coefficient.

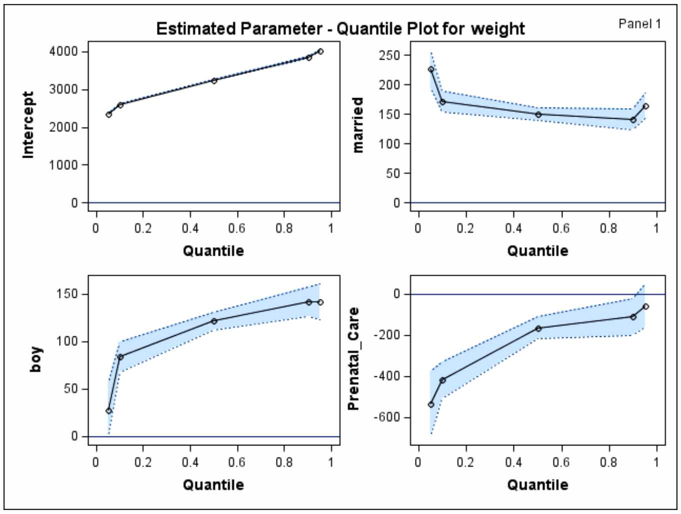
Coefficient estimates for the 5^{th}, 10^{th}, 50^{th}, 90^{th}, 95^{th} quantile regression and the linear regression coefficient estimates for the birth weight example are presented in the following table (all predictors are dummy variables):

Characteristic	**Linear Regression**	**Quantile Regression**				
		5^{th}	**10^{th}**	**50^{th}**	**90^{th}**	**95^{th}**
Intercept	3224	2353	2608	3252	3856	4031
Married	161.1	227	171	149	141	165
Boy	115.9	28	84	121	142	142
Prenatal care	-227.0	-536	-418	-164	-111	-57
Smoke	-200.9	-255	-226	-190	-177	-199

According to the linear regression model, the mean weight of babies born to mothers with no prenatal care is -227 grams lower than that of babies born to mothers who had a prenatal visit in the first trimester. The quantile regression results indicate that the effect of no prenatal care has a larger negative impact on the lower quantiles of birth weight. The 5th quantile of birth weight for infants born to mothers who had no prenatal care is 536 grams lower than for infants born to mothers had a prenatal visit in the first trimester. The linear regression model underestimates this effect at the 5th quantile.

Standard errors and confidence limits for the quantile regression coefficient estimates can be obtained with asymptotic and bootstrapping methods. Both methods provide robust results (Koenecker and Hallock 2001), with the bootstrap method preferred as more practical (Hao and Naiman, 2007).

Quantile regression can be implemented in various statistical software packages including SAS, using Proc Quantreg, and STATA, using Qreg. The quantile plots for the intercept and predictor variables shown below were obtained using SAS. In each plot, the regression coefficient at a given quantile indicates the effect on birth weight of a unit change in that variable, assuming that the other variables are fixed, with 95% confidence interval bands. The intercept can be interpreted as the estimated conditional quantile function of the birth weight distribution of a girl born to an unmarried mother who didn't smoke, and had her first prenatal visit in the first trimester of the pregnancy (Koenker and Hallock, 2001).



These plots help us understand how variable these effects can be. They also highlight that a linear regression might not be an optimal solution to assess this relationship.

If you feel that Quantile Regression might be a useful methodology for your research, contact a staff statistician at the Cornell Statistical Consulting Unit for further assistance.

References:
1. Koenker, R. and Hallock, K. (2001), Quantile Regression: An Introduction, *Journal of Economic Perspectives*, 15, 143–156.
2. Hao L. and Naiman D. Q.(2007), Quantile Regression, Sage Publications, Thousand Oaks.
3. SAS 9.1 Proc Quantreg Documentation.

Author: Simona Despa

(This newsletter was distributed to faculty and graduate students in the Division of Nutritional Sciences, the College of Human Ecology, the College of Agriculture and Life Sciences, the College of Arts and Sciences, the School of Industrial and Labor Relations, the Department of Statistical Sciences, and the Department of City and Regional Planning by the Cornell Statistical Consulting Unit. Please forward it to any interested colleagues, students, and research staff. Anyone not receiving this newsletter who would like to be added to the mailing list for future newsletters should contact us at cscu@cornell.edu. Information about the Cornell Statistical Consulting Unit and copies of previous newsletters can be obtained at http://www.cscu.cornell.edu)