Quantile Regression
Winter/Spring 2021
Andy Poppick
(prerequisites: Statistical Inference and Applied Regression Analysis)

Modeling the behavior of a variable in the tails of its distribution is often even more important than modeling its mean. For example, climate change impacts depend disproportionately on the changing characteristics of extreme weather (e.g., heat events, floods, etc.); similar concerns arise in economics, epidemiology, ecology, and many other fields. In settings where the standard regression assumptions (i.i.d. normal noise) are violated, more flexible methods are therefore required for modeling the full distribution of a response variable as a function of covariates.

Quantile regression is a nonparametric method (i.e., does not depend on assuming a particular distribution) for modeling how different quantiles of the distribution of a response variable depend on covariates. Instead of estimating the mean by minimizing the sum of squared residuals (as in ordinary linear regression), it turns out that we can estimate quantiles by minimizing the sum of asymmetrically weighted absolute residuals. The picture to the right shows what these loss functions look like for three different quantile levels. For example, to estimate the 95th percentile, we want most of the residuals to be negative, so we penalize the positive residuals more. Quantile regression was proposed by Koenker and Bassett (1978), but has precursors dating to the 18th and 19th centuries and remains an active area of research in statistics.

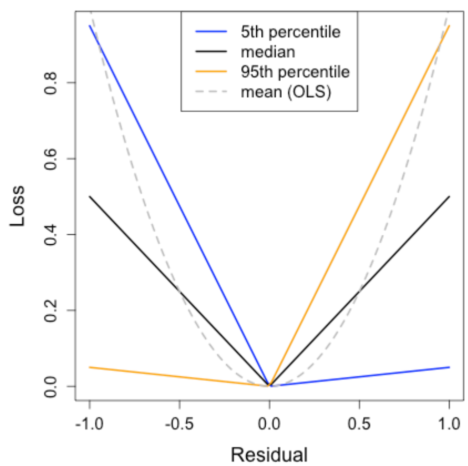

Figure 1: Examples of loss functions for estimating different quantiles. The squared loss function (used for ordinary least squares) is also given for reference.

On the next page, I give an example that illustrates just some of the potential insights provided by quantile regression methods. I am showing daily average temperatures at Minneapolis - Saint Paul International Airport during the winter months (December - February) over the last 40 years, along with a naive estimate of trends in the 5th, 50th, and 95th percentiles using quantile regression (Figure 2, left). The 5th percentile has apparently increased much more rapidly than the median or 95th percentile (the slope coefficients are about 0.87, 0.13, and 0.15°C per decade, respectively). That is, the coldest winter days have increased in temperature more than either the more typical or warmest days (and also, consequently, winter temperature variability has decreased). The general pattern holds at other quantiles levels, with lower quantiles warming more quickly than higher quantiles (Figure 2, right). Knowing that the coldest days are warming the most in Minnesota gives a much richer picture than just knowing that the mean is increasing.

In this comps, you'll learn all about quantile regression: how and why it works, its nice properties and its pitfalls, how to interpret results and do uncertainty quantification, and more. This will be an application-driven project, and I have some climate questions of potential interest that we could investigate, but I am also open to the group choosing a different application area depending on your own interests.

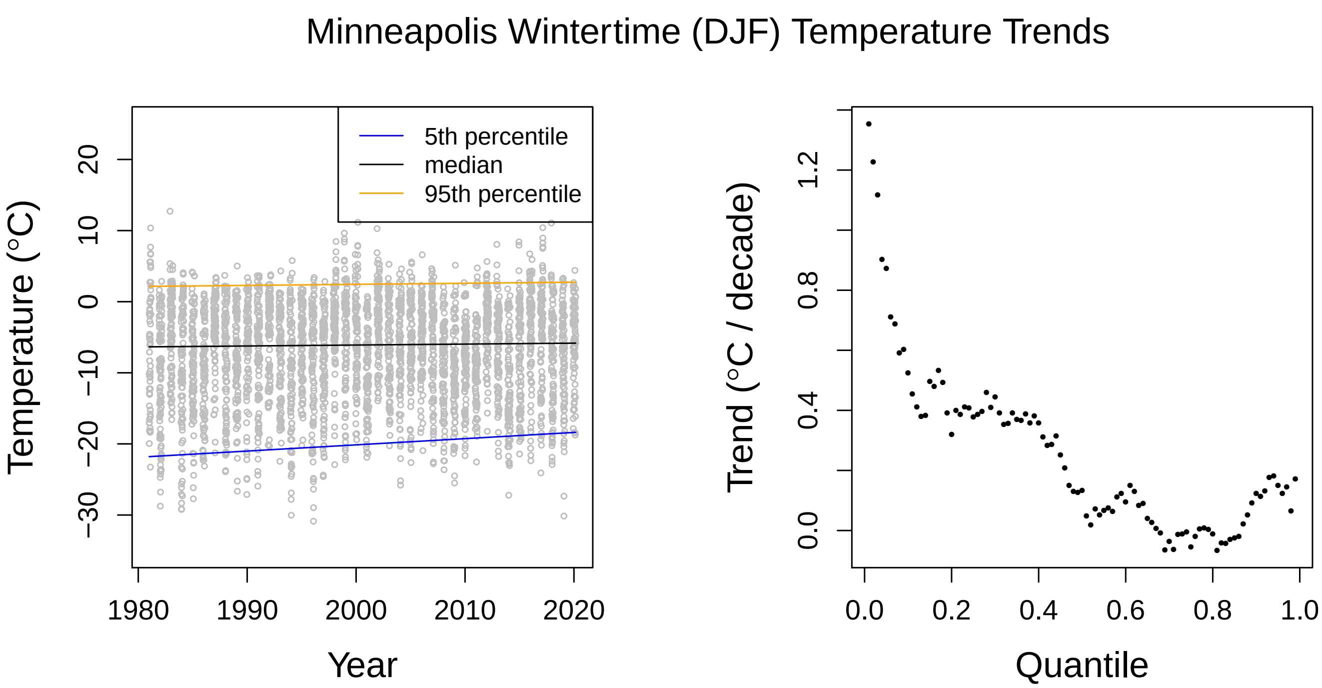

Figure 2: Left, daily winter (DJF) temperatures at Minneapolis - Saint Paul Airport with estimated 5th, 50th, and 95th percentile trends. Right, estimated trends (i.e., slope coefficients) by quantile level.