Mosteller and Tukey (1977) 
What the regression curve does is give a grand summary for the averages of the distributions 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 distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so 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.
A fist example with simulated data

**Homogeneous model**

\[ y_i = 1 + 2x + e \]
\[ x \sim N(10; 1) \quad e \sim N(0; 1) \]

**Location model**
- homoskedasticity assumption
- change in central tendency
- any deviation among the regression estimates is simply due to sampling variation

---

A simple example

**A sample of 60 families**

**Location model**
- homoskedasticity assumption
- change in central tendency
- any deviation among the regression estimates is simply due to sampling variation

---

A simple example

**A sample of 60 families**

**Location model**
- homoskedasticity assumption
- change in central tendency
- any deviation among the regression estimates is simply due to sampling variation

---

A simple example

**A sample of 60 families**

**Location-scale model**
- heteroskedasticity assumption
- heterogeneous variances
Two examples with simulated data

- **homogeneous model**
  \[
  y_1 = 1 + 2x + e \\
  x \sim N(10; 1) \quad e \sim N(0; 1)
  \]

- **heterogeneous model**
  \[
  y_2 = 1 + 2x + (1 + x)\epsilon \\
  x \sim N(10; 1) \quad \epsilon \sim N(0; 1)
  \]

### OLS results

#### homogeneous model

#### heterogeneous model

### A preliminary analysis

- **homogeneous model**
  
- **heterogeneous model**

<table>
<thead>
<tr>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Max</th>
<th>sd</th>
<th>asym</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1</td>
<td>19.5</td>
<td>20.9</td>
<td>28.9</td>
<td>2.2</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Max</th>
<th>sd</th>
<th>asym</th>
</tr>
</thead>
<tbody>
<tr>
<td>-14.6</td>
<td>-13.0</td>
<td>20.6</td>
<td>70.5</td>
<td>11.2</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### Classical vs quantile linear regression

#### Classical linear regression (conditional expected value)

- estimation of the conditional mean of a response variable \( y \) as a function of a set \( X \) of predictor variables

  \[
  E(y \mid X) = X\beta
  \]

#### Quantile regression (conditional quantiles)

- estimation of the conditional quantiles of a response variable \( y \) as a function of a set \( X \) of predictor variables

  \[
  Q_\alpha(y \mid X) = X\beta(\theta)
  \]

  where: \( 0 < \theta < 1 \)

(Koenker and Basset, 1978) (Koenker, 2005) (Davino et al., 2013)
A quick introduction
An in-depth analysis
Inference
Properties
Assessment

OLS and QR results

**homogeneous model**

<table>
<thead>
<tr>
<th>OLS</th>
<th>$\theta = 0.1$</th>
<th>$\theta = 0.25$</th>
<th>$\theta = 0.5$</th>
<th>$\theta = 0.75$</th>
<th>$\theta = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.52</td>
<td>-0.49</td>
<td>-0.73</td>
<td>0.40</td>
<td>1.63</td>
</tr>
<tr>
<td>$x$</td>
<td>2.04</td>
<td>2.02</td>
<td>2.10</td>
<td>2.06</td>
<td>2.00</td>
</tr>
</tbody>
</table>

**heterogeneous model**

<table>
<thead>
<tr>
<th>OLS</th>
<th>$\theta = 0.1$</th>
<th>$\theta = 0.25$</th>
<th>$\theta = 0.5$</th>
<th>$\theta = 0.75$</th>
<th>$\theta = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.05</td>
<td>-2.24</td>
<td>-9.38</td>
<td>-4.61</td>
<td>1.55</td>
</tr>
<tr>
<td>$x$</td>
<td>2.47</td>
<td>0.91</td>
<td>2.28</td>
<td>2.53</td>
<td>2.64</td>
</tr>
</tbody>
</table>

A real data example (Sole24Ore dataset)

**Descriptive statistics of the rate of suicide dependent variable**

<table>
<thead>
<tr>
<th>Min</th>
<th>$\theta = 0.25$</th>
<th>Median</th>
<th>Mean</th>
<th>$\theta = 0.75$</th>
<th>Max</th>
<th>Variance</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.77</td>
<td>11.22</td>
<td>15.56</td>
<td>16.85</td>
<td>19.73</td>
<td>48.20</td>
<td>64.44</td>
<td>22.53</td>
</tr>
</tbody>
</table>

**Scatter plot of the rate of suicide and mean per capita income**

- Heteroscedastic relationships
- Skewed dependent variable
A real data example (Sole24Ore dataset)

Quantile regression lines and coefficients

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>3.11</td>
<td>0.38</td>
</tr>
<tr>
<td>$\theta = 0.1$</td>
<td>1.32</td>
<td>0.22</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>0.57</td>
<td>0.31</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>1.51</td>
<td>0.38</td>
</tr>
<tr>
<td>$\theta = 0.75$</td>
<td>5.45</td>
<td>0.40</td>
</tr>
<tr>
<td>$\theta = 0.9$</td>
<td>-1.26</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Scale shift effect

“A reconstruction” of the conditional and unconditional response distribution

1. estimate the QR model using a grid of quantiles
2. combine the different estimates for the different values of $\theta$
3. select the “best” model for each unit according to the deviation between the observed value and the estimated values
4. “reconstruct” the response variable using the “best” models determined at the previous step

Another real data application (unimc dataset)

The evaluation of University educational processes

- random sample of 685 students graduated at University of Macerata (Italy)
- dependent variable: degree mark
- 7 regressors related to the student profile:
  - gender
  - place of residence during University (Macerata and its province, Marche region, outside Marche)
  - course attendance (no attendance, regular)
  - foreign experience (yes, no)
  - working condition (full time student, working student)
  - number of years to get a degree
  - diploma mark

The response variable

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>0.25</th>
<th>Median</th>
<th>0.75</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>all sample</td>
<td>77</td>
<td>102.0</td>
<td>110.0</td>
<td>106.4</td>
<td>110</td>
</tr>
<tr>
<td>regular</td>
<td>85</td>
<td>107</td>
<td>110</td>
<td>107.8</td>
<td>110</td>
</tr>
<tr>
<td>1 year out-of-course</td>
<td>85</td>
<td>106</td>
<td>110</td>
<td>107.2</td>
<td>110</td>
</tr>
<tr>
<td>2 year out-of-course</td>
<td>83</td>
<td>101.2</td>
<td>106</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>3 year out-of-course</td>
<td>77</td>
<td>98</td>
<td>104</td>
<td>102.5</td>
<td>110</td>
</tr>
<tr>
<td>4 year out-of-course</td>
<td>85</td>
<td>97</td>
<td>104</td>
<td>102.4</td>
<td>110</td>
</tr>
<tr>
<td>≥ 5 years out-of-course</td>
<td>82</td>
<td>96.2</td>
<td>103</td>
<td>101.3</td>
<td>110</td>
</tr>
</tbody>
</table>

The QR estimates

- Intercept
- gender
- placeOfResidence
- coursesAttendance
- foreignExperience
- workingCondition
- yearsToGetDegree
- diplomaMark
Quantile Regression

QR model for a given conditional quantile $\theta$ (linear regression):

$$Q_\theta(y|X) = X\hat{\beta}(\theta)$$

where

- $0 < \theta < 1$
- $Q_\theta(.)$ denotes the conditional quantile function for $\theta^{th}$ quantile

**Interpretation**

$$\hat{\beta}(\theta) = \frac{\partial Q_\theta(y|X)}{\partial x_i}$$

Rate of change of the $\theta^{th}$ quantile of the dependent variable distribution per unit change in the value of the $i^{th}$ quantile

- Classical regression focuses on $E(y|X)$
- QR extends this approach to study the conditional distribution of a response variable
  - $\theta$ regression lines
  - $100(1-\theta)\%$ of points above the QR line and $100\theta\%$ below
  - QR generalizes univariates quantiles for conditional distributions

**QR pros:**

- Regressor effects on the whole dependent variable distribution
- Heteroscedastic relationships
- Presence of outliers
- Skewed dependent variable
On the objective functions

Unconditional mean, median and quantiles
Let \( Y \) be a generic random variable:
- Mean (and its objective function): \( \mu = E(Y - c)^2 \)
- Median (and its objective function): \( Me = E|Y - c| \)
- Generic quantile \( \theta \) (and its objective function):
\[
q_\theta = E[\rho_\theta(Y - c)]
\]
- \( \hat{\mu} \) is \( Me \) denotes the sample estimators for such centers
- \( \rho(.) \) denotes the following location functions:
\[
\rho_\theta(y) = [\theta - I(y < 0)]y
= [(1 - \theta)(y \leq 0) + \theta I(y > 0)]|y|
\]
- \( \rho_\theta(.) \) is an asymmetric absolute loss function; that is a weighted sum of absolute deviations, where a \((1 - \theta)\) weight is assigned to the negative deviations and a \(\theta\) weight is used for the positive deviations.

On the objective functions

- Denoting with \( Y \) a discrete random variable with probability distribution \( f(y) = P(Y = y) \):
\[
q_\theta = E[\rho_\theta(Y - c)]
= E \left\{ (1 - \theta) \sum_{y < c} |y - c|f(y) + \theta \sum_{y > c} |y - c|f(y) \right\}
\]
- If \( Y \) is a continuous r.v. with density function \( f(y) \):
\[
q_\theta = E[\rho_\theta(Y - c)]
= \int_{-\infty}^{c} (1 - \theta) |y - c|f(y)dy + \theta \int_{c}^{+\infty} |y - c|f(y)dy
\]

Note 1: \( \hat{q}_\theta \), with \( \theta \in [0, 1] \), denotes the sample estimator
Note 2: when \( \theta = 0.5 \) we have the solution for the median

Conditional mean and conditional quantiles: the regression setting

Least squares linear regression estimator
\[
\hat{\beta} = \beta E[Y - X^T \beta]^2
\]

The conditional quantile estimator
\[
\hat{\beta}(\theta) = \beta E[\rho_\theta(Y - X\beta)]
\]

Note: the \((\theta)\)–notation denotes that the parameters and the corresponding estimators are for a specific quantile \( \theta \)
\[
\rho_\theta = \begin{cases} 
\theta u & \text{if } u > 0 \\
(\theta - 1) u & \text{if } u \leq 0
\end{cases}
\]
On the objective functions

\( \theta = 0.25 \)

- 75% of points above the QR line and 25% below
- unbalanced weighting system: 0.75 (0.25) for sum of negative (positive) deviations
- \( m = 2 \) points lies exactly on the line (\( m \)-number of model parameters)

\( \theta = 0.75 \)

- 25% of points above the QR line and 75% below
- unbalanced weighting system: 0.25 (0.75) for sum of negative (positive) deviations
- \( m = 2 \) points lies exactly on the line (\( m \)-number of model parameters)

The linear programming formulation of the QR problem

- Wagner (1959) proved that the least absolute deviation criterion can be formulated as a linear programming technique and then solved efficiently exploiting proper methods and algorithms
- Koenker and Bassett (1978) pointed out how conditional quantiles could be estimated by an optimization function minimizing a sum of weighted absolute deviations, using weights as asymmetric functions of the quantiles
- The linear programming formulation of the problem was therefore natural, offering researchers and practitioners a tool for looking inside the whole conditional distribution apart from its center

Methods for solving the linear programming problem

- The simplex method (Dantzig, 1947) is the widespread solution for the linear programming problem
- It is an iterative process, starting from a solution that satisfies the imposed constraints and looking for new and better solution
- The process iterates until a solution that cannot be further improved is reached, moving along the edges of the simplex corresponding to the feasible set
- For the QR problem, the efficient version of the simplex algorithm, proposed by Barrodale and Roberts (1974) and adapted by Koenker e D'Orey (1987) to compute conditional quantiles, is typically used with a moderate size problem
- The simplex method is the default option in most of the QR software
- A completely different method approaches the solution from the interior of the feasible set rather than on its boundary, that is starting in the zone where all the inequalities are strictly satisfied
- Such methods, called interior–point methods, have their roots in the seminal paper of Karmakar (1984) and are usually superior on very large problems
- The QR solution using interior–point methods has been proposed by Portnoy e Koenker (1997)
- A heuristic approach (finite smoothing algorithm) has been proposed by Chen (2004, 2007): it is faster and more accurate in the presence of a large number of covariates
Main approaches to inference in QR

- Small sample theory
  (Koenker and Basset, 1978)
  “The practical of this theory would entail a host of hazardous assumptions and an exhausting computational effort” (Koenker, 2005)

- Asymptotic theory
  (Koenker and Basset, 1978, 1982a,b)

- Rank–based theory
  (Gutenbrunner and Jureckova, 1992) (Gutenbrunner, 1993)

- Resampling methods

Asymptotic theory

\[ Q_\theta (\hat{y} | x) = \hat{\beta}_0 (\theta) + \hat{\beta}_1 (\theta) x \]

“under mild regularity conditions”

Asymptotic distribution of the estimator:

1. case of i.i.d. errors
   \[ \sqrt{n} \left[ \beta (\theta) - \beta (\theta) \right] \to N \left( 0, \varpi^2 (\theta) J^{-1} \right) \]

2. case of i.i.d. errors
   \[ \sqrt{n} \left[ \beta (\theta) - \beta (\theta) \right] \to N \left( 0, \theta (1 - \theta) H (\theta)^{-1} J H (\theta)^{-1} \right) \]

The error distribution affects the variance–covariance matrix of the QR estimator

The variance of the estimator - case i.i.d.

\[ \sqrt{n} \left[ \beta (\theta) - \beta (\theta) \right] \to N \left( 0, \varpi^2 (\theta) J^{-1} \right) \] (1)

- \( \theta \): selected quantile
- \( J = \lim_{n \to \infty} \frac{1}{n} \sum_i x_i x_i^\top = X^\top X \)

\[ \varpi^2 (\theta) = \frac{\theta (1 - \theta)}{F^{-1} (\theta)^2} \] scale parameter

\[ s (\theta) = \frac{1}{F^{-1} (\theta)} \] sparsity (Tukey, 1965),
quantile-density function (Parzen, 1979)

Assumption 1 (density):
F has a continuous and positive density, \( f \), for each \( z \) such that
\[ 0 < F(z) < 1 \] for all the quantiles of interest

Assumption 2 (design):
\( J \) is a positive definite matrix
The variance of the estimator - case i.i.d.

\[
\sqrt{n} \left[ \hat{\beta}(\theta) - \beta(\theta) \right] \rightarrow N \left( 0, \theta (1 - \theta) H(\theta)^{-1} \right)
\]

- \( \theta \): quantile of interest
- \( J = \lim_{n \to \infty} \frac{1}{n} \sum_i x_i^T x_i = X^T X \)
- \( H(\theta) = \lim_{n \to \infty} \frac{1}{n} \sum_i f_i [F^{-1}(\theta)] x_i^T x_i = X^T S^{-1} X \)
- \( S = \text{diag} \{ s_i(\theta) \} \)
- \( s_i(\theta) = \frac{1}{i [F^{-1}(\theta)]} \)

Problems:
- Estimation of \( F^{-1}(\theta) \) | bandwidth computation

xy-pair method: a single quantile \( \theta \)

Simple quantile regression model

\[
Q_{\theta}(\hat{y}|x) = \hat{\beta}_0 + \hat{\beta}_1(\theta)x
\]

Bootstrap estimate: \( \overline{\beta}(\theta) = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_b(\theta) \)

Bootstrap standard error: \( \text{se} \left( \overline{\beta}_j(\theta_q) \right) \)

- **xy-pair or design matrix bootstrap method** (Kocherginsky, 2003)
- method based on pivotal estimation functions (Parzen, 1979)
**xy-pair method: k quantiles**

Estimated variance covariance matrix:

\[
V_{q,j} = \frac{1}{B} \sum_{b=1}^{B} \left( \beta_{q,j}(\theta_q) - \overline{\beta}_j(\theta_q) \right) \left( \beta_{q,j}(\theta_q) - \overline{\beta}_j(\theta_q) \right)^T
\]

(3)

where \( j = 1, \ldots, p; \ q = 1, \ldots, k \) bootstrap estimates:

\[
\overline{\beta}_j(\theta_q) = \frac{1}{B} \sum_{b=1}^{B} \beta_{q,j}(\theta_q)
\]

**OLS and QR results**

**homogeneous model**

<table>
<thead>
<tr>
<th>OLS</th>
<th>( \theta = 0.1 )</th>
<th>( \theta = 0.25 )</th>
<th>( \theta = 0.5 )</th>
<th>( \theta = 0.75 )</th>
<th>( \theta = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.52</td>
<td>-0.49</td>
<td>-0.73</td>
<td>0.40</td>
<td>1.83</td>
</tr>
<tr>
<td>( x )</td>
<td>2.04</td>
<td>2.02</td>
<td>2.10</td>
<td>2.06</td>
<td>2.00</td>
</tr>
</tbody>
</table>

**heterogeneous model**

<table>
<thead>
<tr>
<th>OLS</th>
<th>( \theta = 0.1 )</th>
<th>( \theta = 0.25 )</th>
<th>( \theta = 0.5 )</th>
<th>( \theta = 0.75 )</th>
<th>( \theta = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-4.05</td>
<td>-2.24</td>
<td>-9.38</td>
<td>-4.61</td>
<td>1.55</td>
</tr>
<tr>
<td>( x )</td>
<td>2.47</td>
<td>0.91</td>
<td>2.28</td>
<td>2.53</td>
<td>2.64</td>
</tr>
</tbody>
</table>

**Two examples with simulated data**

**homogeneous model**

\[
y_1 = 1 + 2x + e \quad x \sim N(10; 1) \quad e \sim N(0; 1)
\]

**heterogeneous model**

\[
y_2 = 1 + 2x + (1 + x)e \quad x \sim N(10; 1) \quad e \sim N(0; 1)
\]

**Empirical distributions of the estimates**

\[ \theta = 0.1 \]

**homogeneous model**

\( \hat{\beta}_1 = 2.02 \)

**heterogeneous model**

\( \hat{\beta}_1 = 0.91 \)
Empirical distributions of the estimates

\[ \theta = 0.5 \]

**homogeneous model**
\[ \hat{\beta}_1 = 2.06 \]

**heterogeneous model**
\[ \hat{\beta}_1 = 2.53 \]

Inference in QR

Hypothesis tests

\[ H_0 : \beta = 0 \]

\[ t(\hat{\beta}(\theta)) = \frac{\hat{\beta}(\theta)}{se(\hat{\beta}(\theta))} \]

Confidence intervals

\[ \beta(\theta) \pm se(\hat{\beta}(\theta)) \]

**homogeneous model**

| coeff    | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| intercept| 0.40515    | 0.43001 | 0.94220  | 0.34632  |
| x        | 2.05600    | 0.04272 | 48.13042 | 0.00000  |

**heterogeneous model**

| coeff    | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| intercept| -4.60710   | 4.49361 | -1.02526 | 0.30549  |
| x        | 2.53023    | 0.44492 | 5.68698  | 0.00000  |
Coefficient QQ-plot according to B ($\theta = 0.5$)

**heterogeneous model**

<table>
<thead>
<tr>
<th>B = 50</th>
<th>B = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Plot" /></td>
<td><img src="image2.png" alt="Plot" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B = 500</th>
<th>B = 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.png" alt="Plot" /></td>
<td><img src="image4.png" alt="Plot" /></td>
</tr>
</tbody>
</table>

**Bootstrap results**

<table>
<thead>
<tr>
<th>$\theta = 0.5$</th>
<th>B = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nid</strong></td>
<td><strong>bootstrap</strong></td>
</tr>
<tr>
<td><strong>coef</strong></td>
<td><strong>Std. Error</strong></td>
</tr>
<tr>
<td>intercept</td>
<td>-4.61</td>
</tr>
<tr>
<td>x</td>
<td>2.53</td>
</tr>
</tbody>
</table>

**Equivariance properties**

- scale equivariance
- shift or regression equivariance
- equivariance to reparametrization of design
- equivariance to monotone transformations
**Scale equivariance**

\[
Q_\theta(\hat{y} | x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x
\]

\[
Q_\theta(c\hat{y} | x) = c\hat{\beta}_0(\theta) + c\hat{\beta}_1(\theta)x
\]

- if the dependent variable is multiplied by a positive constant \(c\), the coefficients of the new model changes accordingly
- useful when it is necessary to modify the unit of measurement of the dependent variable (for example to reduce its variability or to allow comparisons with other models) without altering the interpretation of the results.

\[
Q_\theta(d\hat{y} | x) = d\hat{\beta}_0(1 - \theta) + d\hat{\beta}_1(1 - \theta)x
\]

\(d\) is a negative constant

For \(\theta = 0.5\), the scale equivariance does not depend on the constant sign

**Equivariance to monotone transformations**

\[
Q_\theta(\hat{y} | x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x
\]

\[
Q_\theta[h(\hat{y}) | x] = h[\hat{\beta}_0(\theta)] + h[\hat{\beta}_1(\theta)]x
\]

where \(h(.)\) is a non decreasing function in \(\mathbb{R}\)

- The quantiles of the transformed \(y\) variable are the transformed quantiles of the original ones
- appropriate selection of \(h(.)\) corrects different kinds of skewness

\[
\log[E(\hat{y} | x)] \neq E[\log(\hat{y}) | x]
\]

\[
\log[Q_\theta(\hat{y} | x)] = Q_\theta[\log(\hat{y}) | x]
\]

---

**The Sole24Ore dataset**

- Dependent variable: *rate of suicide* (per 100,000 hinab.)
- Regressors: *income*, *geographic area* (North-west, North-East, Center, South and Islands)
- Units: 103 italian provinces

<table>
<thead>
<tr>
<th></th>
<th>suicide</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>2.77</td>
<td>19.1</td>
</tr>
<tr>
<td>Q1</td>
<td>11.22</td>
<td>26.4</td>
</tr>
<tr>
<td>Median</td>
<td>15.56</td>
<td>37</td>
</tr>
<tr>
<td>Mean</td>
<td>16.65</td>
<td>35.45</td>
</tr>
<tr>
<td>Q3</td>
<td>19.73</td>
<td>42.35</td>
</tr>
<tr>
<td>Max</td>
<td>48.2</td>
<td>57.1</td>
</tr>
<tr>
<td>st.dev</td>
<td>8.03</td>
<td>9.35</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.26</td>
<td>-0.052</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.31</td>
<td>2.035</td>
</tr>
</tbody>
</table>

---

**Data description**

![Data description graph](image)
Data description

The logarithmic transformation

Quantile Regression ($\theta=[0.1:0.9, 0.1]$)

The logarithmic transformation: effect on OLS results

White Test  | Shapiro-Wilk normality test
---|---
(p-value) | (p-value)

| suicide | 0.02345 | 1.875e-05 |
| log(suicide) | 0.2836 | 0.7268 |
The logarithmic transformation: effect on OLS and QR results

\[
E(y|x) = E(\log(y)|x) = Q_{0.1}(y|x) = Q_{0.1}(\log(y)|x)
\]

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>3.111</td>
<td>0.382</td>
</tr>
<tr>
<td>slope</td>
<td>1.817</td>
<td>0.025</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
E(\text{suicide}|\text{income} = 28) &= 3.111 + 28 \times 0.382 = 13.807 \\
E(\log(\text{suicide})|\text{income} = 28) &= 1.817 + 28 \times 0.025 = 2.517. \\
Q_{0.1}(\text{suicide}|\text{income} = 28) &= 1.319 + 28 \times 0.225 = 7.61 \\
Q_{0.1}(\log(\text{suicide})|\text{income} = 28) &= 1.358 + 28 \times 0.024 = 2.03.
\end{align*}
\]

Temperature = high

**OLS results**

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{suicide}</td>
<td>intercept</td>
<td>13.934</td>
<td>1.512</td>
</tr>
<tr>
<td></td>
<td>income</td>
<td>4.215</td>
<td>2.138</td>
</tr>
<tr>
<td>\log(\text{suicide})</td>
<td>intercept</td>
<td>2.475</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>income</td>
<td>0.359</td>
<td>0.121</td>
</tr>
</tbody>
</table>

**QR results**

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{suicide}</td>
<td>intercept</td>
<td>11.600</td>
<td>1.082</td>
</tr>
<tr>
<td></td>
<td>income</td>
<td>5.320</td>
<td>1.437</td>
</tr>
<tr>
<td>\log(\text{suicide})</td>
<td>intercept</td>
<td>2.451</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>income</td>
<td>0.377</td>
<td>0.097</td>
</tr>
</tbody>
</table>
QR assessment

Model: $Q_0(\hat{y} | x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x$

Residual absolute sum of weighted differences:

$RASW_\theta = \sum_{y_i \geq \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x_i} y_i - \hat{\beta}_0(\theta) - \hat{\beta}_1(\theta)x_i$

$\sum_{y_i < \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x_i} (1 - \theta) |y_i - \hat{\beta}_0(\theta) - \hat{\beta}_1(\theta)x_i|

Model: $Q_0(\hat{y}) = \hat{\beta}_0(\theta)$

Total absolute sum of weighted differences:

$TASW_\theta = \sum_{y_i \geq \hat{\theta}} |y_i - \hat{\theta}| + \sum_{y_i < \hat{\theta}} (1 - \theta) |y_i - \hat{\theta}|

pseudoR^2_\theta = 1 - \frac{RASW}{TASW}$

(Koenker and Machado, 1999)

Main references


Concluding remarks: motivation and caution

Motivation (Mosteller and Tukey, 1977)

What the regression curve does is give a grand summary for the averages of the distributions 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 distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so 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.

Caution

QR offers information on the whole conditional distribution of the response variable, allowing us to discern effects that would otherwise be judged equivalent using only conditional expectation. Nonetheless, the QR ability to statistically detect more effects can not be considered a panacea for investigating relationships between variables: in fact, the improved ability to detect a multitude of effects forces the investigator to clearly articulate what is important to the process being studied and why.

Main references

Main references