# *M*-quantiles for binary and categorical data

James Dawber

University of Wollongong, Australia

15 July, 2016

Can quantile-like regression be applied to categorical response data?

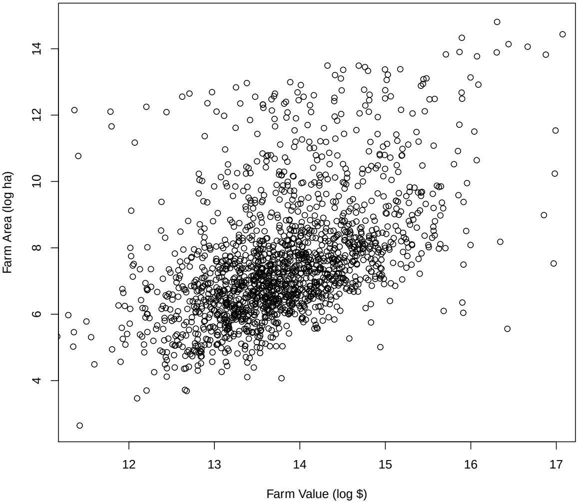
# Outline

1. *M*-quantiles for continuous response
2. Uses in small area estimation
3. Current methods for *M*-quantiles for binary response
4. An alternative method for binary responses
5. Extending to categorical responses
6. Small area estimation with binary response
7. Small area estimation with categorical response
8. An example - unemployment in the UK

# Quantiles, expectiles and something in between...

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon \le 0)} + q\mathbb{I}_{(\epsilon > 0)}\right] \rho(\epsilon)$$

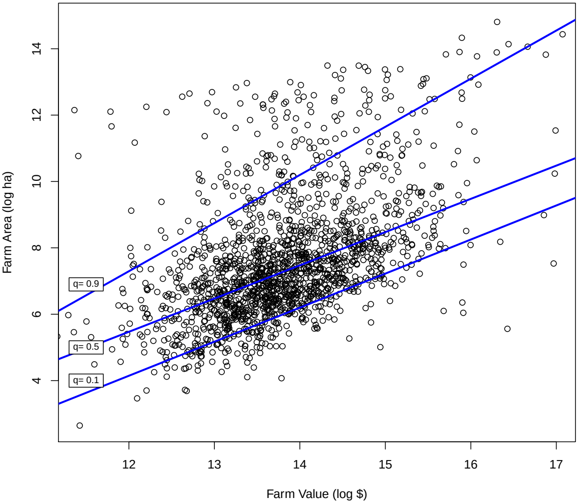
$$\epsilon = \frac{\boldsymbol{y} - \boldsymbol{X}\beta_{\boldsymbol{q}}}{\sigma_q}$$



# Quantiles, expectiles and something in between...

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon \le 0)} + q\mathbb{I}_{(\epsilon > 0)}\right] \rho(\epsilon)$$

$$\epsilon = \frac{\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta_q}}{\sigma_q}$$



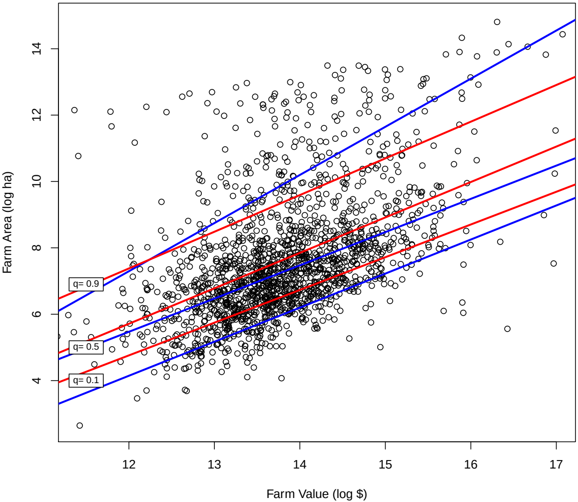

**Quantile**

Minimise $\sum \rho_q(\epsilon)$ where $\rho(\epsilon) \propto |\epsilon|$

# Quantiles, expectiles and something in between...

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon \le 0)} + q\mathbb{I}_{(\epsilon > 0)}\right] \rho(\epsilon)$$

$$\epsilon = \frac{\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta_q}}{\sigma_q}$$




**Quantile**

Minimise $\sum \rho_q(\epsilon)$ where $\rho(\epsilon) \propto |\epsilon|$

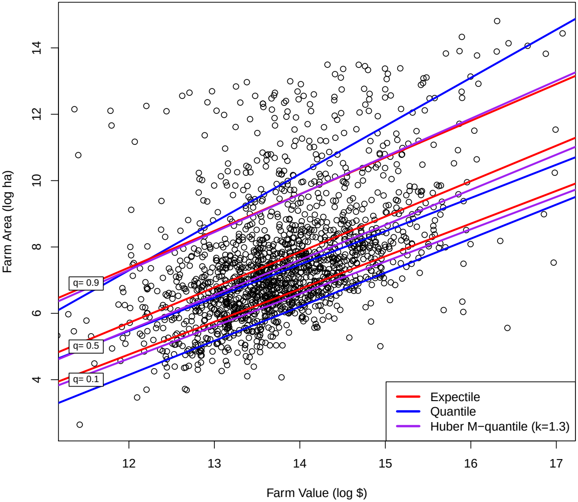
**Expectile**

Minimise $\sum \rho_q(\epsilon)$ where $\rho(\epsilon) \propto \epsilon^2$

# Quantiles, expectiles and something in between...

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon\leq 0)} + q\mathbb{I}_{(\epsilon>0)}\right] \rho(\epsilon)$$

$$\boldsymbol{\epsilon} = \frac{\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta_q}}{\sigma_q}$$




## Quantile

Minimise $\sum \rho_q(\epsilon)$ where $\rho(\epsilon) \propto |\epsilon|$

## Expectile

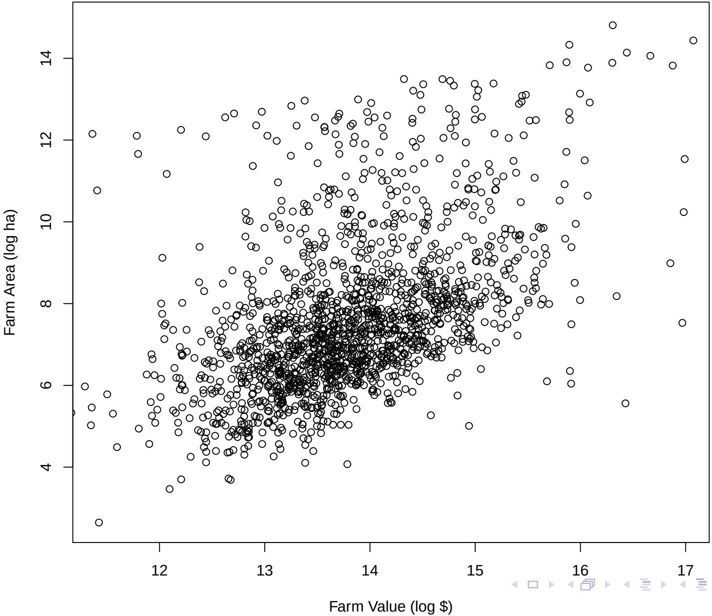
Minimise $\sum \rho_q(\epsilon)$ where $\rho(\epsilon) \propto \epsilon^2$

## Huber *M*-quantile

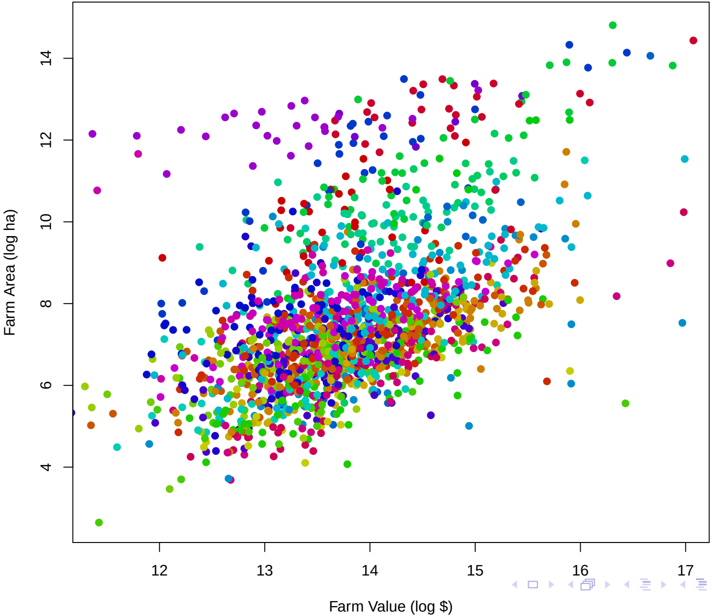
Minimise $\sum \rho_{q,k}(\epsilon)$ where

$$\rho(\epsilon, k) \propto \begin{cases} \epsilon^2, & \text{if } |\epsilon| < k \\ |\epsilon|, & \text{if } |\epsilon| \geq k \end{cases}$$

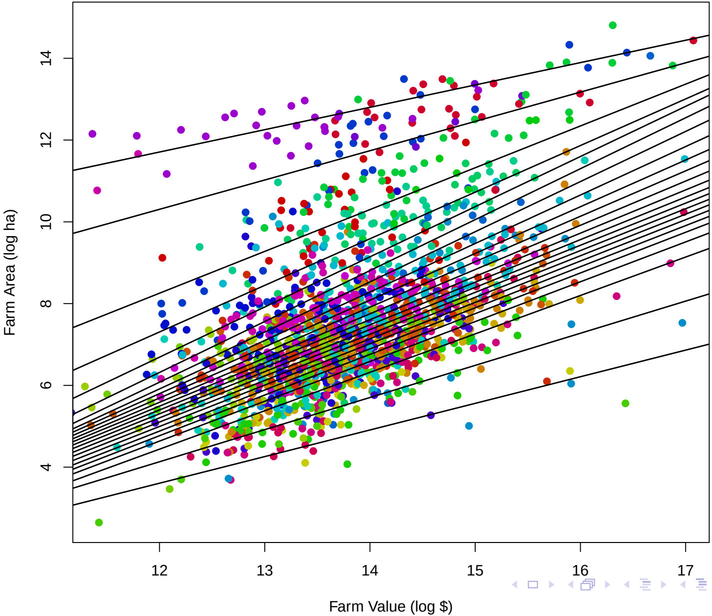
# *M*-quantiles - uses in Small Area Estimation



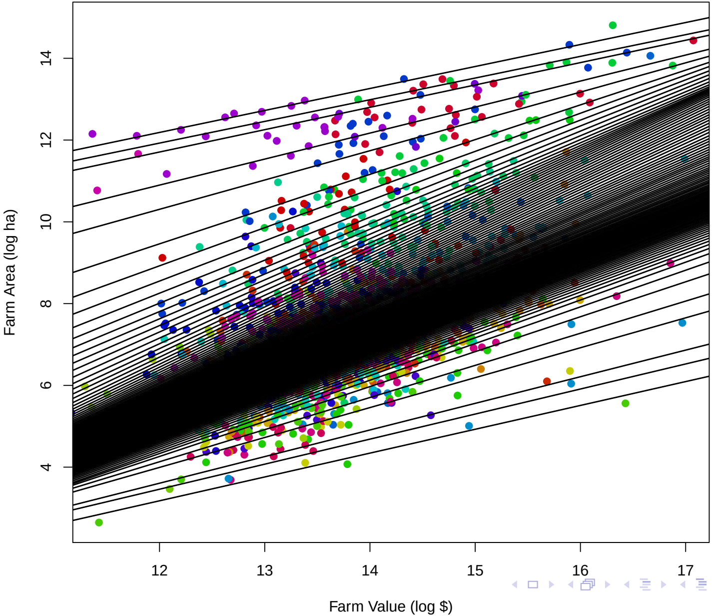
# $M$-quantiles - uses in Small Area Estimation



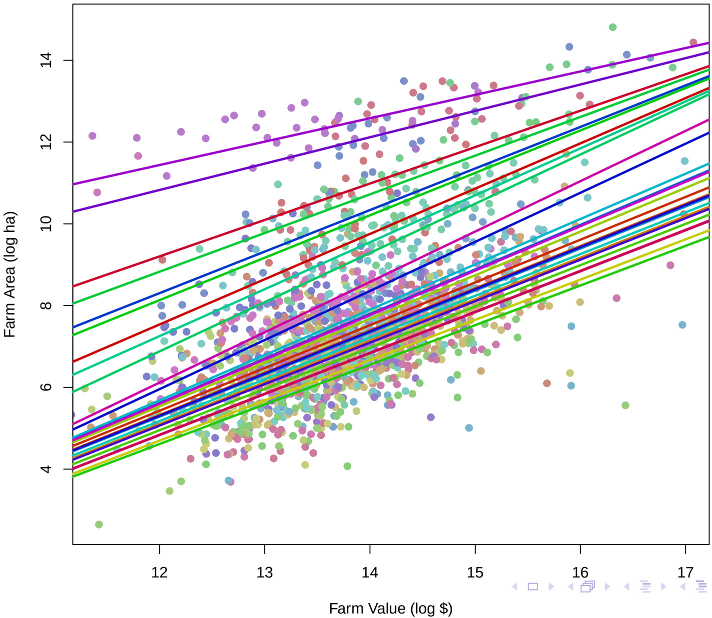
# $M$-quantiles - uses in Small Area Estimation



# *M*-quantiles - uses in Small Area Estimation



# $M$-quantiles - uses in Small Area Estimation



# $M$-quantiles with continuous data

1. Huber $M$-quantiles are $M$-estimators with influence function:

$$\psi_{q,k}(x) = \begin{cases} -2(1-q)k, & \text{if } x < -k \\ 2(1-q)x, & \text{if } -k \leq x \leq 0 \\ 2qx, & \text{if } 0 < x \leq k \\ 2qk, & \text{if } x > k \end{cases}$$

2. As $k \to 0$ we get the quantile and as $k \to \infty$ we get the expectile influence function.
3. So with continuous $Y$ we obtain the $M$-quantile estimate $M_{q,k}$ by solving:

$$E\left[\psi_{q,k}\left(\frac{Y - M_{q,k}}{\sigma_q}\right)\right] = 0.$$

4. Where $\sigma_q$ is a scale estimator which ensures scale equivariance of $M_{q,k}$.

# $M$-quantiles with binary data

1. Quantiles for binary data are very difficult, but $M$-quantiles and expectiles are nice.
2. With binary $Y$ we obtain $M$-quantile estimate according to Chambers et al (2015):

$$E\left[\psi_{q,k}\left(\frac{Y - M_{q,k}}{\sigma(M_{q,k})}\right)\sigma(M_{q,k}) - \alpha\right] = 0.$$

   Where $\sigma(M_{q,k}) = \sqrt{M_{q,k}(1 - M_{q,k})}$ and $\alpha$ is a correction term for consistency.
3. Just extend this to categorical data with more than two groups?
4. Is there another way?

# Binary expectile calculation

Suppose we have $Y \sim Bernoulli(\pi)$, then a simplified binary expectile $\mu_q$ is found by solving:

$$E\left[\psi_{q,k=\infty}(Y - \mu_q)\right] = 0$$

---

$$\begin{aligned}
0 &= E\left[2(1-q)(Y-\mu_q)I_{Y\leq\mu_q} + 2q(Y-\mu_q)I_{Y>\mu_q}\right] \\
&= \sum_{y=0}^{1}\left[2(1-q)(y-\mu_q)I_{y\leq\mu_q} + 2q(y-\mu_q)I_{y>\mu_q}\right] Pr(Y=y) \\
&= 2(1-q)(0-\mu_q)(1-\pi) + 2q(1-\mu_q)\pi \\
&= -2(1-q)\mu_q(1-\pi) + 2q(1-\mu_q)\pi
\end{aligned}$$

which rearranges to:

$$\mu_q = \frac{\pi q}{(1-\pi)(1-q) + \pi q}.$$

# Binary expectile properties

$$\mu_q = \frac{\pi q}{(1-\pi)(1-q)+\pi q}.$$

Some of the properties of $\mu_q$ include:

1. $\mu_q \to 0$ as $q \to 0$ or $\pi \to 0$.
2. $\mu_q \to 1$ as $q \to 1$ or $\pi \to 1$.
3. $\mu_q = \pi$ when $q = 0.5$.
4. $1 - \mu_q(Y) = \mu_{1-q}(1 - Y)$.
5. When $\pi + q = 1$, $\mu_q = 0.5$.

# Binary expectile regression

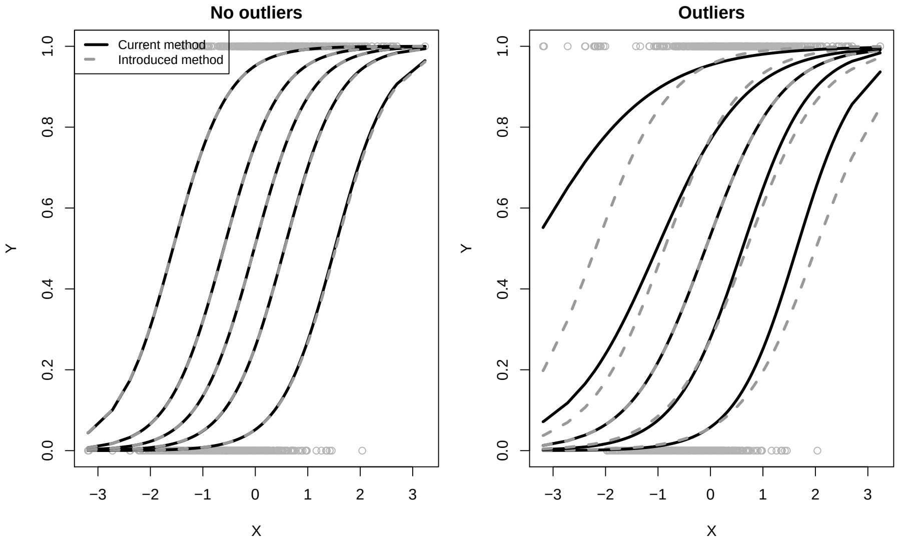
We can use estimates of $\pi$ to get $\mu_q$.

$$\pi = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$$

which gives estimates of $\mu_q$:

$$\mu_q = \frac{\exp(X\beta + \log \frac{q}{1-q}))}{1 + \exp(X\beta + \log \frac{q}{1-q})}$$

# Binary expectile regression example



# Categorical expectile regression

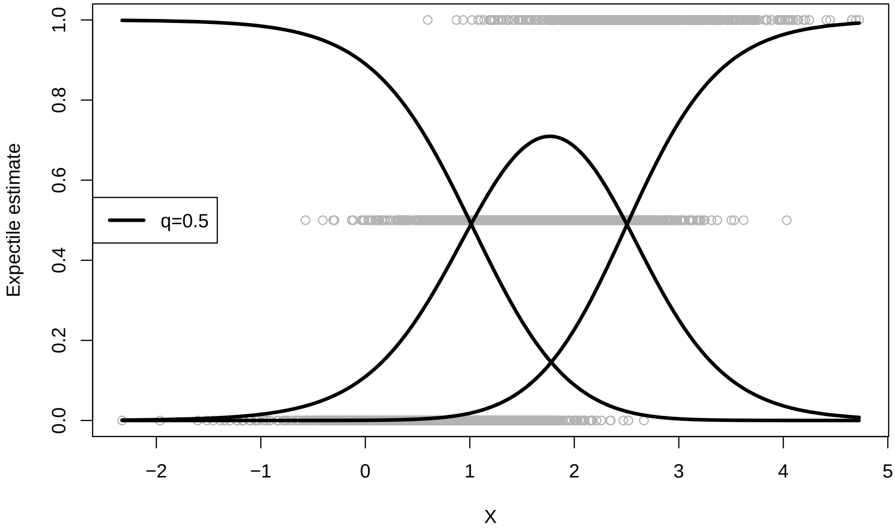
If instead of two categorical groups we have $J$ groups then we can use estimates of $\pi_j$ to get $\mu_{qj}$.

$$\pi_j = \frac{\exp(X\beta_j)}{1 + \sum_{j=1}^{J-1} \exp(X\beta_j)}$$
$$\pi_J = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(X\beta_j)}$$

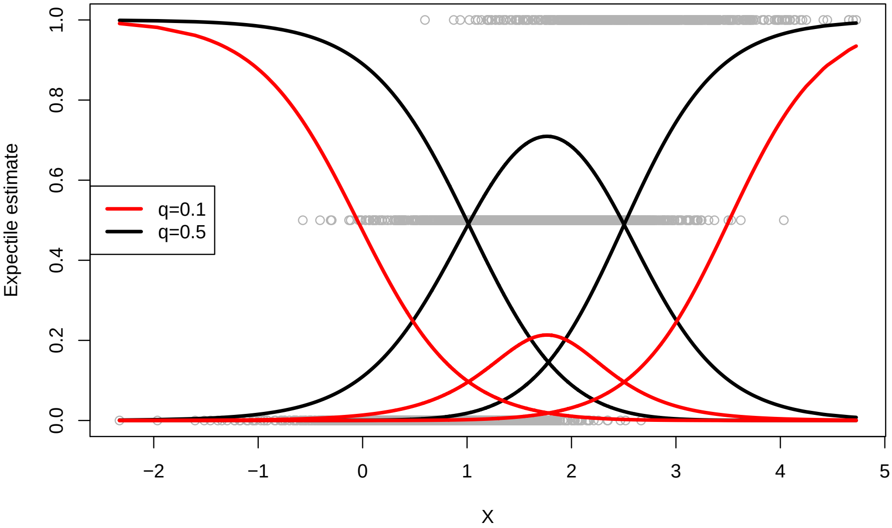
which gives estimates of $\mu_{qj}$:

$$\mu_{qj} = \frac{\exp(X\beta_j + \log \frac{q}{1-q})}{1 + \sum_{j=1}^{J-1} \exp(X\beta_j) - \exp(X\beta_j) + \exp(X\beta_j + \log \frac{q}{1-q})}$$
$$\mu_{qJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(X\beta_j + \log \frac{1-q}{q})}$$

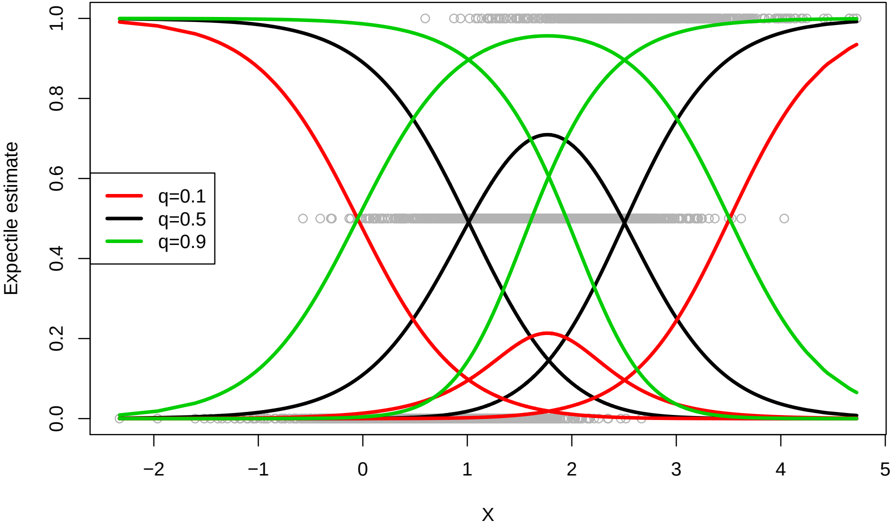
# Categorical expectile regression example



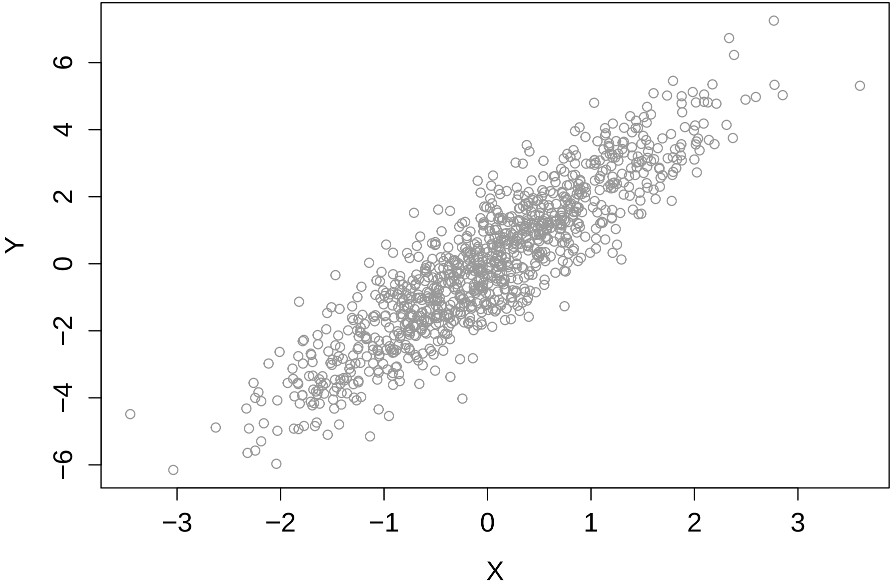
# Categorical expectile regression example



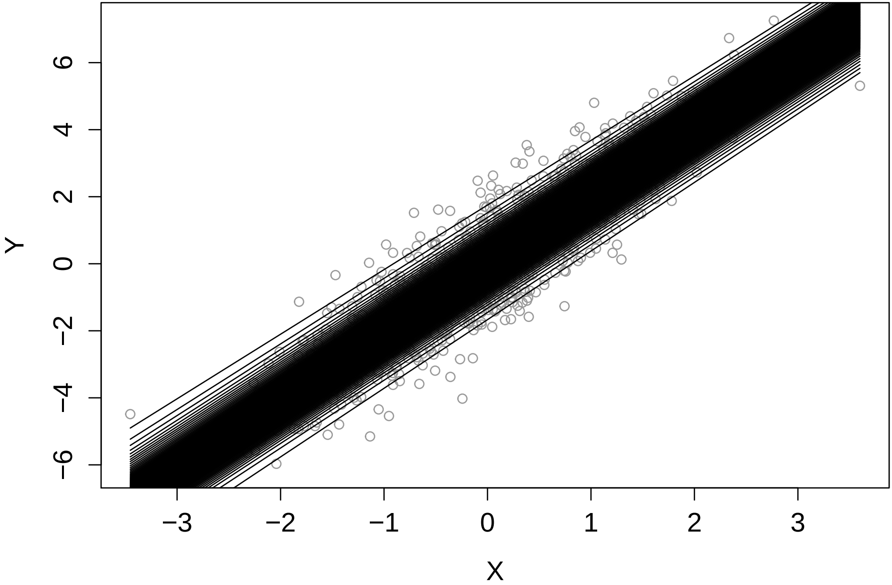
# Categorical expectile regression example



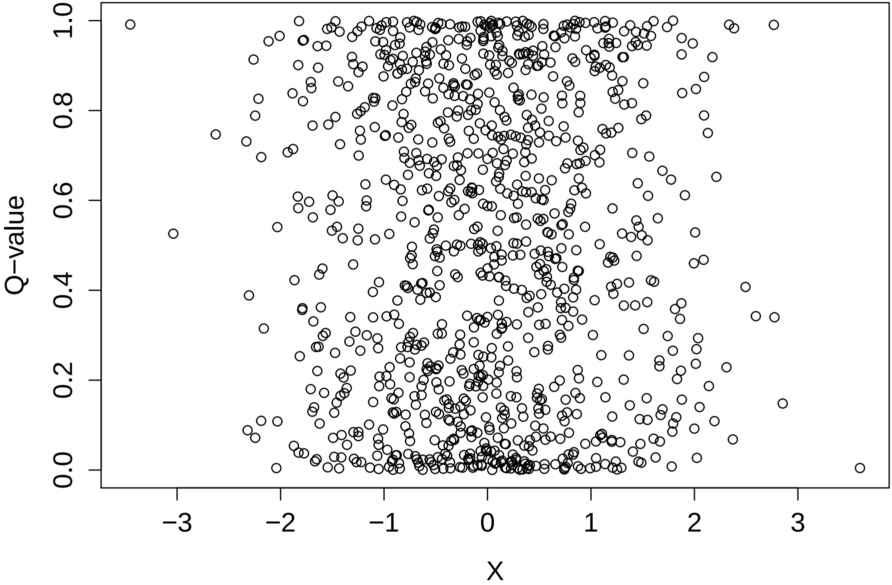
# Continuous expectile q-values



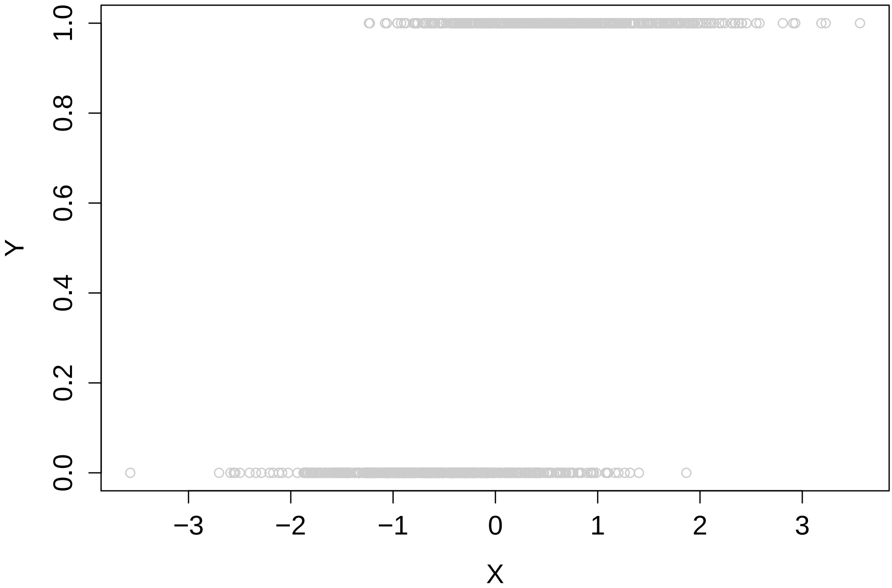
# Continuous expectile q-values



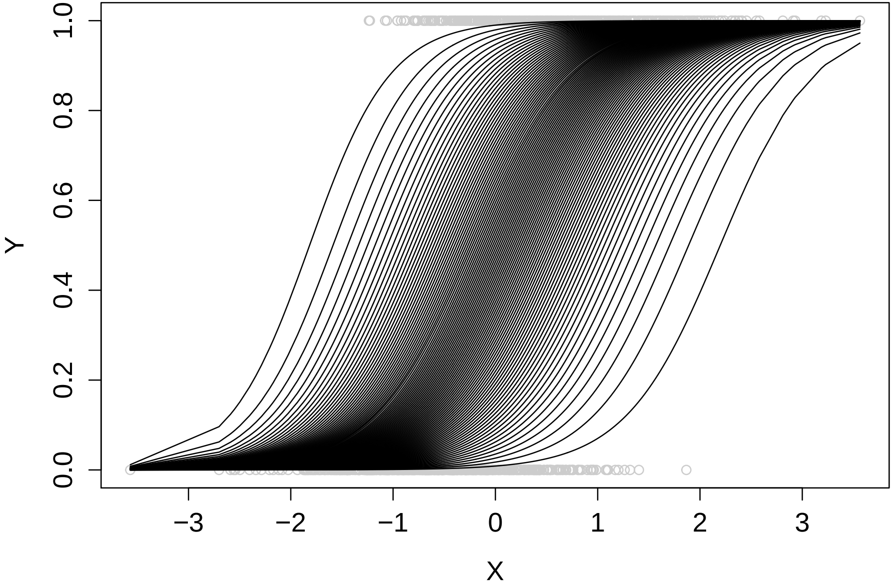
# Continuous expectile q-values



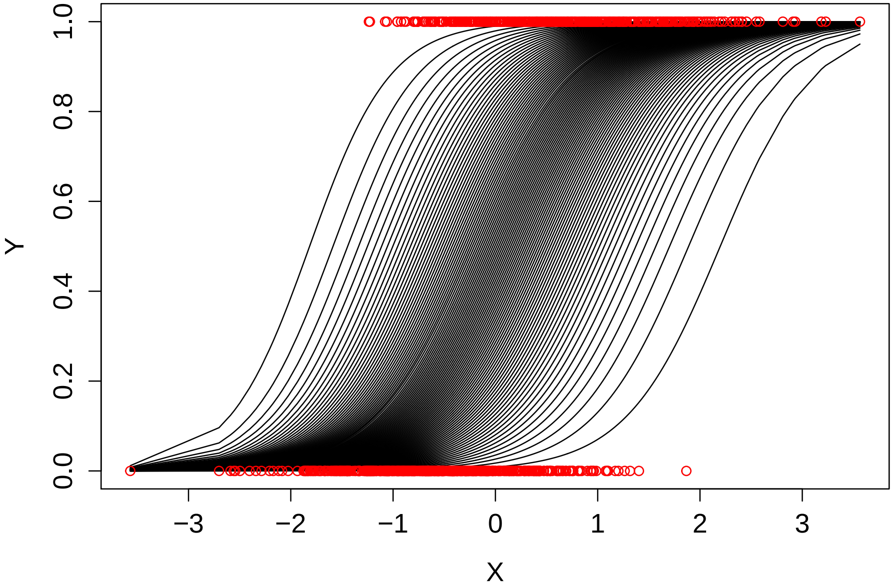
# Binary expectile q-values



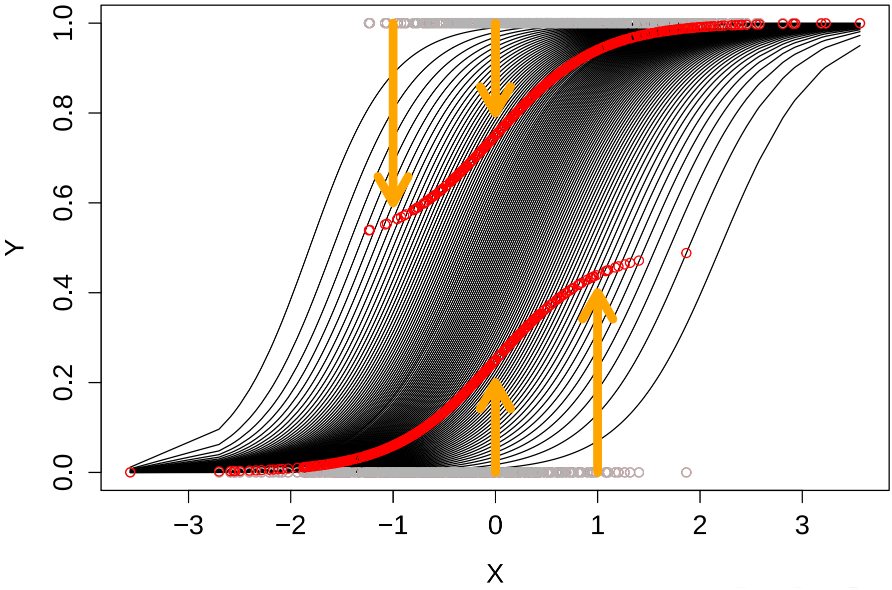
# Binary expectile q-values



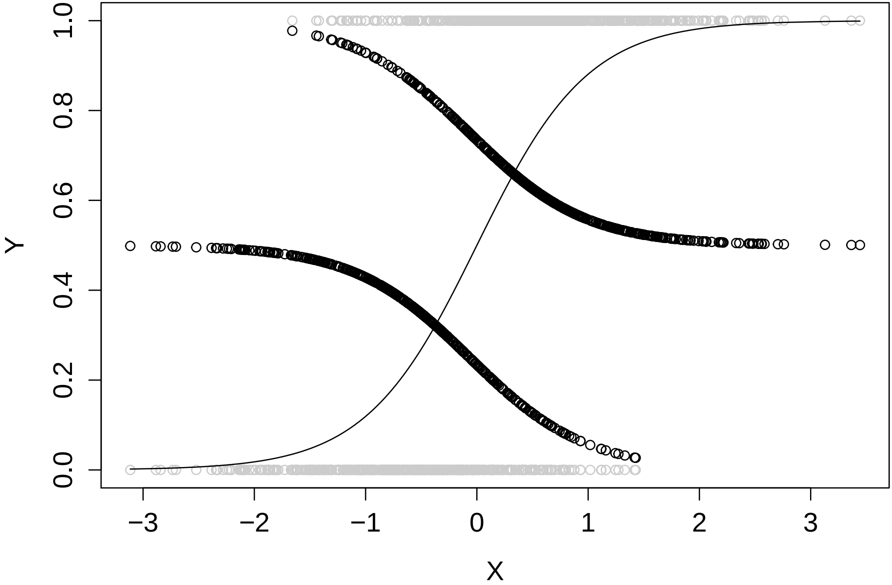
# Binary expectile q-values



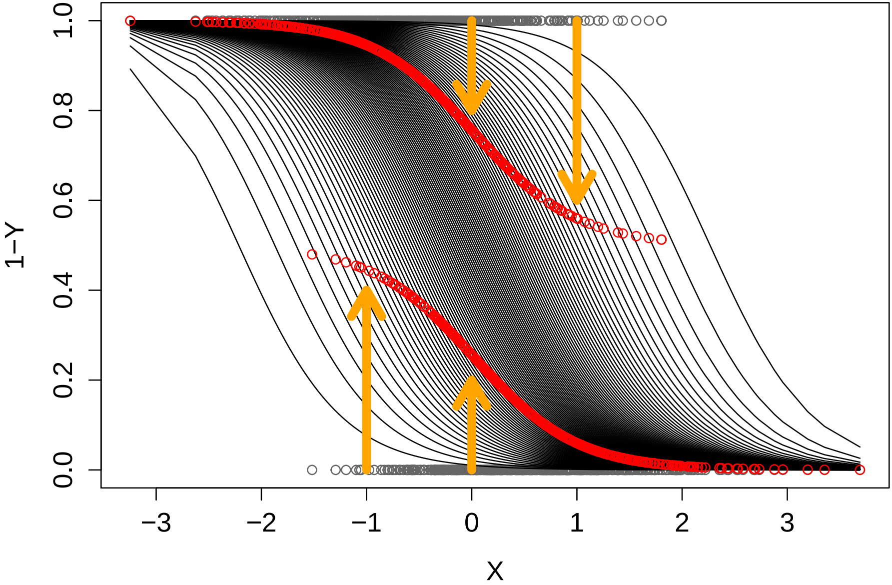
# Binary expectile q-values



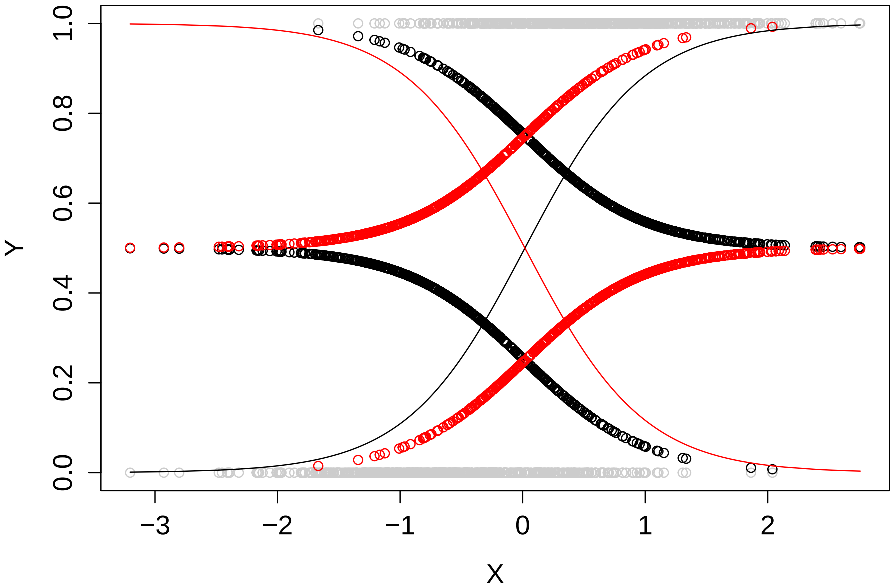
# Binary expectile q-values



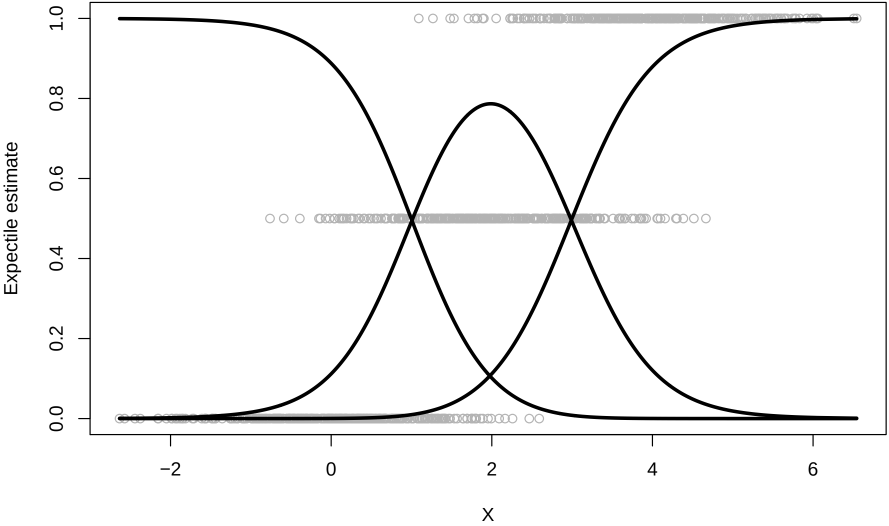
# Binary expectile q-values - swap groups



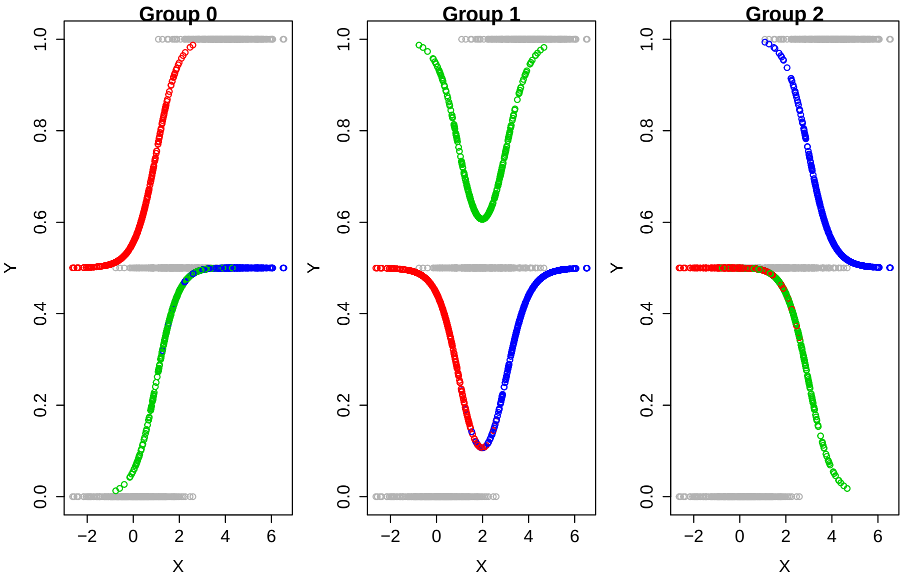
# Binary expectile q-values



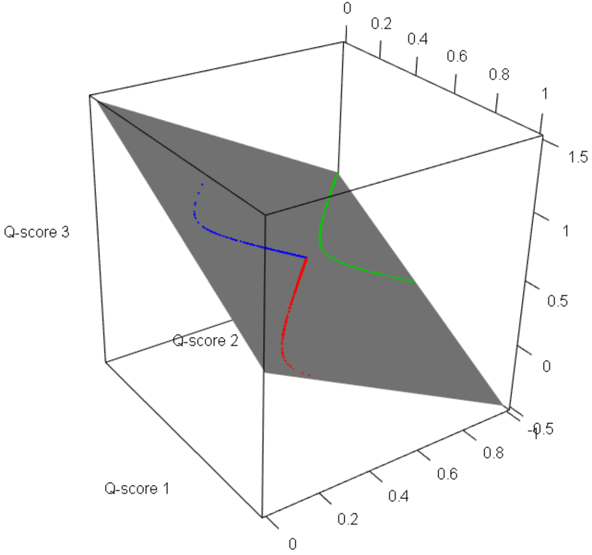
# Categorical expectile q-values



# Categorical expectile q-values



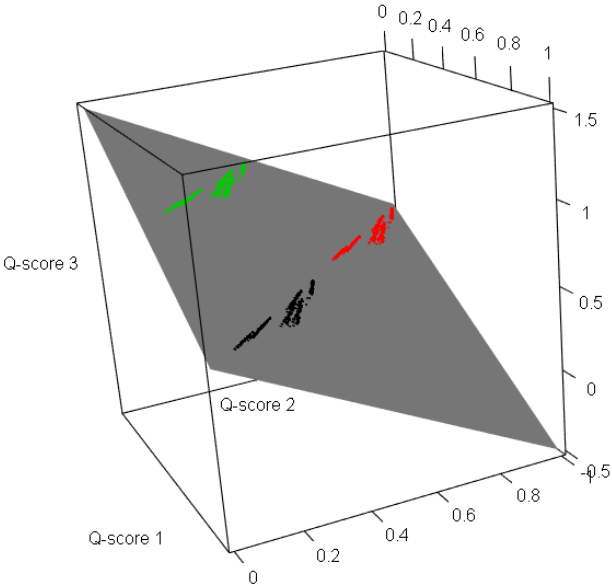
# Categorical expectile q-values



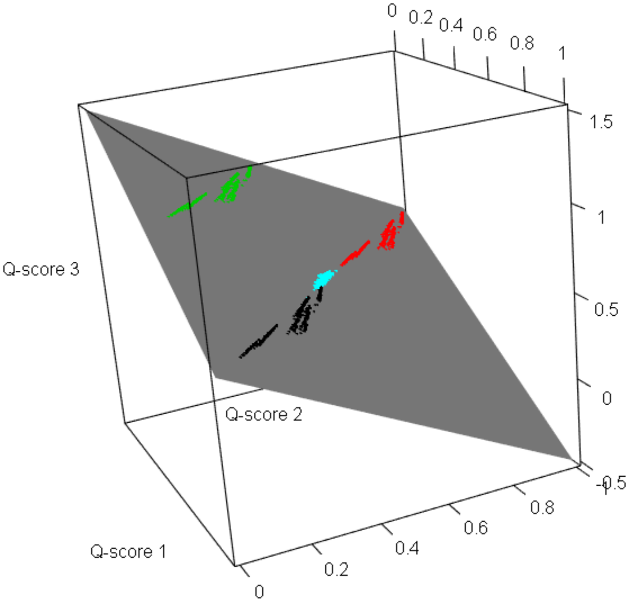
# Example - unemployment in UK

| UALAD | GOR | CLUSTER | SEXAGE | UNEMPLOYED | EMPLOYED | INACTIVE | CELLPOP | Registered Unemployed |
|---|---|---|---|---|---|---|---|---|
| 379 | 7 | 6 | 1 | 5 | 35 | 16 | 14944 | 753 |
| 379 | 7 | 6 | 2 | 2 | 35 | 11 | 13035 | 285 |
| 379 | 7 | 6 | 3 | 8 | 143 | 8 | 41217 | 1992 |
| 379 | 7 | 6 | 4 | 9 | 153 | 26 | 42817 | 471 |
| 379 | 7 | 6 | 5 | 4 | 51 | 69 | 27791 | 438 |
| 379 | 7 | 6 | 6 | 3 | 41 | 104 | 33139 | 112 |
| 380 | 7 | 1 | 1 | 6 | 30 | 12 | 12641 | 465 |
| 380 | 7 | 1 | 2 | 3 | 32 | 12 | 12509 | 197 |
| 380 | 7 | 1 | 3 | 4 | 138 | 8 | 38006 | 1111 |
| 380 | 7 | 1 | 4 | 6 | 121 | 51 | 40196 | 339 |
| 380 | 7 | 1 | 5 | 0 | 87 | 68 | 34613 | 411 |

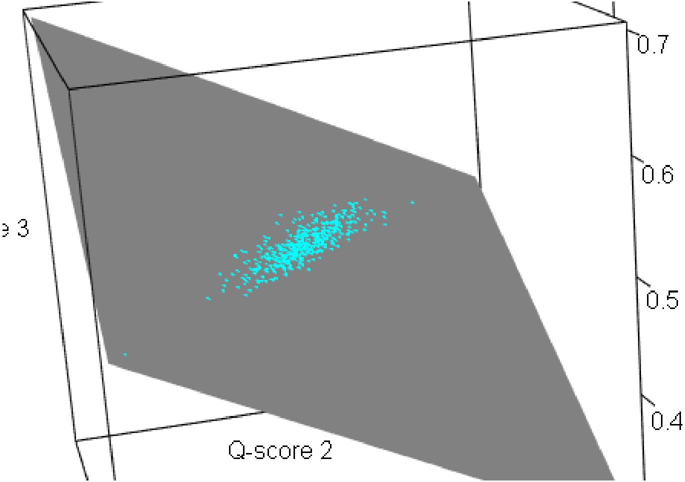
# Example - unemployment in UK



# Example - unemployment in UK



# Example - unemployment in UK



# Binary $M$-quantile

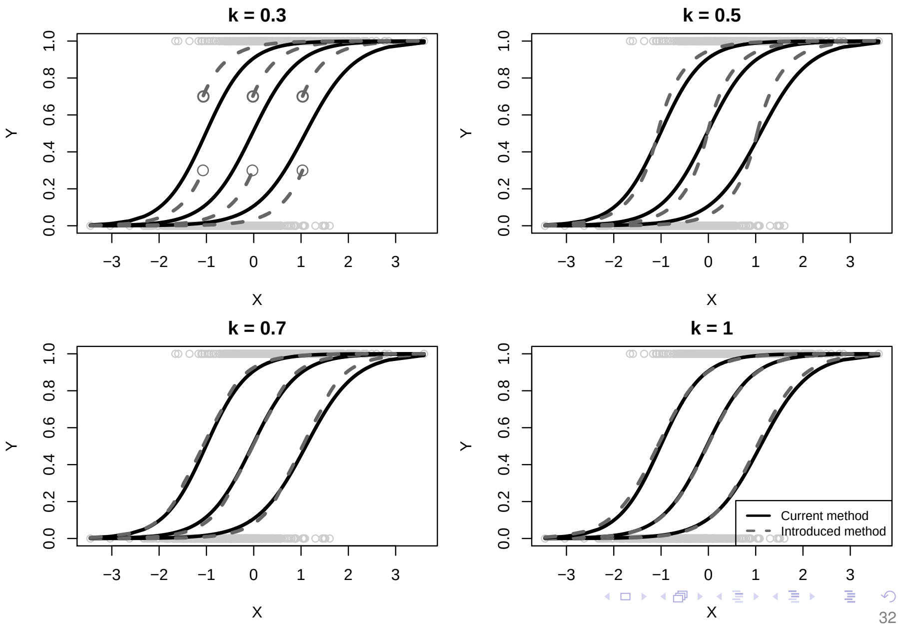
Suppose we have $Y \sim Bernoulli(\pi)$, then the binary $M$-quantile is:

$$E\left[\psi_{q,k}(Y - M_{q,k})\right] = 0$$

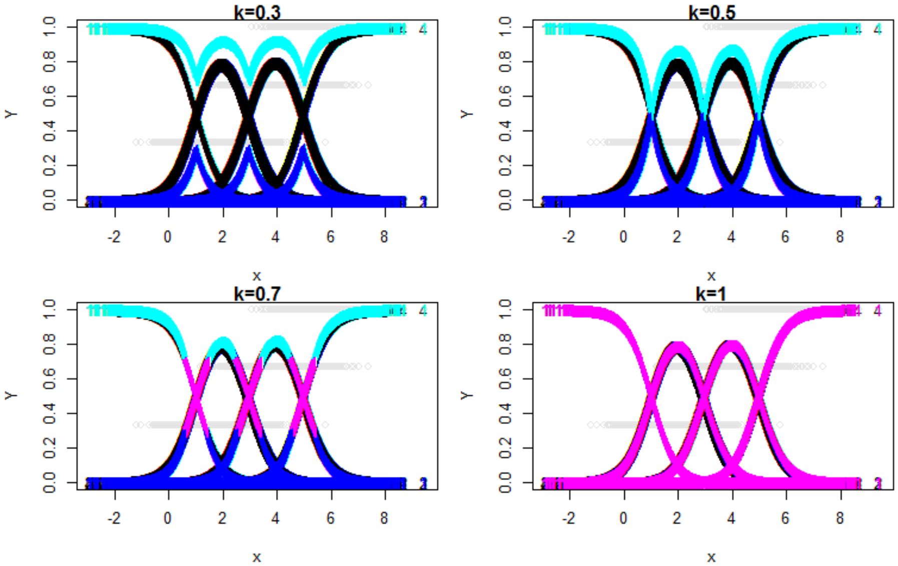
which can be solved and expressed as:

$$\mu_{q,k} = \begin{cases} \dfrac{q\pi}{(1-q)(1-\pi)+q\pi}, & \text{if } \mu_{q,k} < k \cap \mu_{q,k} > 1-k \\ \dfrac{kq\pi}{(1-q)(1-\pi)}, & \text{if } \mu_{q,k} < k \cap \mu_{q,k} < 1-k \\ 1 - \dfrac{k(1-q)(1-\pi)}{q\pi}, & \text{if } \mu_{q,k} > k \cap \mu_{q,k} > 1-k \end{cases}$$

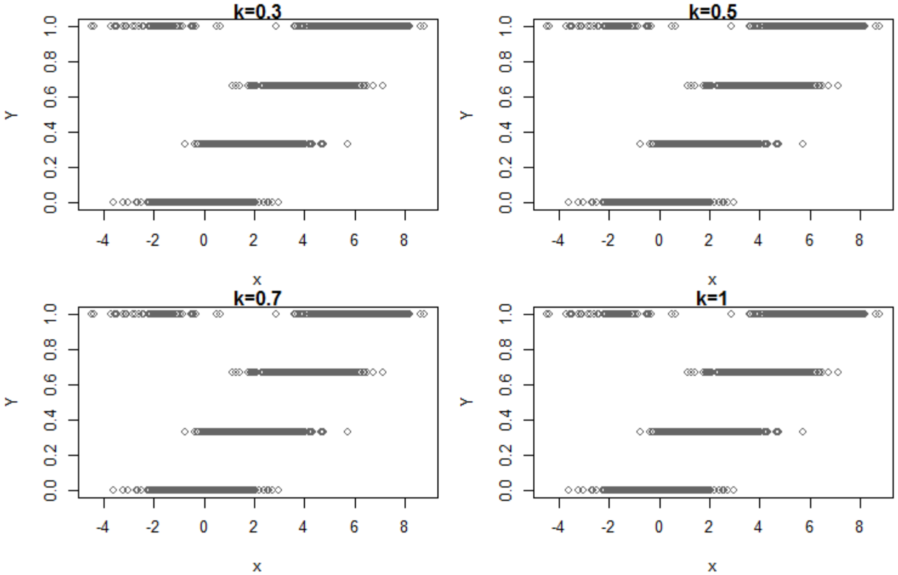
# Binary $M$-quantile



# Categorical $M$-quantile



# Categorical *M*-quantile



# Categorical $M$-quantile