For the entire course we assumed that our data were samples from some specific distribution:

- Gaussian responses $\rightarrow$ linear regression;
- binary responses $\rightarrow$ Bernoulli/binomial distributions $\rightarrow$ logistic regression;
- ... and other members from the exponential family $\rightarrow$ GLMs.

It is implicit that when we say “we assume a certain distribution” we should use diagnostic methods to verify that the assumption holds.

For example by using qq-plots, to compare sample quantiles with the exact quantiles from an hypothesized distribution.

In this half-lecture we consider a methodology which is “distribution free”.
Quantiles

General case: The \( \alpha \)-quantile \( y_\alpha \) is defined as \( P(Y \leq y_\alpha) = \alpha \).

Regression: The \( \alpha \)-quantile \( y_{i\alpha} \) is defined as \( P(Y_i \leq y_{i\alpha} | X_i) = \alpha \).

Linear regression

With \( Y_i = X_i \beta + \epsilon_i \) where \( \epsilon_i \sim N(0, \sigma^2) \) we have \( y_{i\alpha} = X_i \beta + \lambda_\alpha \cdot \sigma \).

Estimated by the prediction interval. Requires that \( \epsilon_i \) really are \( N(0, \sigma^2) \). [here \( \lambda_\alpha \) is the \( \alpha \)-quantile from \( N(0,1) \)]
Poisson regression

With $Y_i \sim Po(e^{X_i\beta})$ we can use the quantiles in the (estimated) Poisson distribution:

Other distributions

As long as we know (read “pretend to know”) the distribution type and can estimate all its parameters we can use its quantiles.

What if we don’t know the distribution type?
Engel (1857) analysed the relationship between household food expenditure and household income. Data (belgian francs) are from 235 Belgian working-class households.

We notice increasing dispersion for increasing household income. Difficult to explain variation without considering essential info (e.g. different levels of education; racial groups;...)
Quantile as solution to minimization problem

- Sample mean as a solution to minimization: $\hat{\mu} = \bar{y}$ solves
  $$\min_{\mu} \sum (y_i - \mu)^2$$

- Median (i.e. 50% quantile) $\hat{m} = \text{median}(y_1, \ldots, y_n)$ solves:
  $$\min_{m} \sum |y_i - m| \quad \text{(robust to outliers!)}$$

- Generic empirical quantile $y_\alpha$ corresponding to a probability $\alpha$ solves:
  $$\min_{y_\alpha} \left\{ (1 - \alpha) \sum_{y_i < y_\alpha} |y_i - y_\alpha| + \alpha \sum_{y_i \geq y_\alpha} |y_i - y_\alpha| \right\}$$
  and is robust to outliers.
Quantile regression

Set $y_{i\alpha} = \mathbf{X}_i \beta_\alpha$ where $P(Y_i \leq y_{i\alpha} | \mathbf{X}_i) = \alpha$.

Replace Least-squares $(Y_i - \mu_i)^2$ by

$$
\rho_\alpha(Y_i - y_{i\alpha}) = \begin{cases} 
(1 - \alpha) \cdot |Y_i - y_{i\alpha}| & \text{if } Y_i < y_{i\alpha}, \\
\alpha \cdot |Y_i - y_{i\alpha}| & \text{if } Y_i \geq y_{i\alpha}.
\end{cases}
$$

and minimize $\sum_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{X}_i \beta_\alpha)$ with respect to $\beta_\alpha = \begin{pmatrix} \beta_{0\alpha} \\
\beta_{1\alpha} \\
\vdots \\
\beta_{p\alpha} \end{pmatrix}$

Symbol $\beta_\alpha$ is meant to emphasize that it is not an estimate based on least squares or maximum likelihood (in the latter case it would not be possible as we do not specify the distribution for observed data).
the resulting $\alpha$-quantile regression line is such that, for given covariates $X_i$, a proportion of approximately $\alpha$ data points lies below the fitted value

$$y_{i\alpha} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$$

and a proportion $1 - \alpha$ lies above.

For example, when we estimate the coefficients for the .10th quantile regression line, 90% of the data points will lie above the fitted value leading to positive residuals, and 10% lie below the fitted value and thus have negative residuals.

Conversely, to estimate the coefficients for the .90th quantile regression, 90% of observations have negative residuals and the remaining 10% have positive residuals.
Engels data

Linear regression line (red); 50%-quantile (blue); 10,25,70,90% quantiles (gray).

Question: among families having income=2000, what food expenditure is exceeded by 10% of families at most?
Again Poisson regression:

10%−, 50%−, 90%− quantiles, given $X$

(a) True quantiles from the Poisson distribution

(b) Quantile regression, 10-50-90%
Advantages of QR: (i) does not require assumptions on the
distribution of $Y$. (ii) more robust to outliers than least squares
(quantiles do not change that much in presence of outliers).

Disadvantages of QR: because of (i) estimated asymptotic
variance of $\hat{\beta}_a$ does not attain minimal variance (Cramer-Rao
bound), unlike maximum likelihood estimates (MLE).

<table>
<thead>
<tr>
<th>Cramer-Rao</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under (mild) conditions an unbiased estimator $\hat{\beta}$ of a scalar parameter $\beta$ has variance</td>
</tr>
<tr>
<td>$Var(\hat{\beta}) \geq \frac{1}{I_{fish}(\beta)}$ (Cramer-Rao bound)</td>
</tr>
</tbody>
</table>

$I_{fish}(\beta)$ is the “Fisher information”, given by

$I_{fish}(\beta) = -E(\frac{\partial^2 \log L(y_1,..,y_n;\beta)}{\partial^2 \beta})$

As $n \to \infty$ a MLE reach the minimal variance $I_{fish}$.

Even though for most models $I_{fish}$ is unavailable analytically, as you
know by now, an approximate is given by $I_{fish} \approx H$, the Hessian
matrix at the MLE $\hat{\beta}$. 

10 / 10
Advantages of QR: (i) does not require assumptions on the distribution of $Y$. (ii) more robust to outliers than least squares (quantiles do not change that much in presence of outliers).

Disadvantages of QR: because of (i) estimated asymptotic variance of $\hat{\beta}_\alpha$ does not attain minimal variance (Cramer-Rao bound), unlike maximum likelihood estimates (MLE).

### Cramer-Rao

Under (mild) conditions an unbiased estimator $\hat{\beta}$ of a scalar parameter $\beta$ has variance

$$Var(\hat{\beta}) \geq \frac{1}{I_{fish}(\beta)} \quad \text{(Cramer-Rao bound)}$$

$I_{fish}(\beta)$ is the “Fisher information”, given by

$$I_{fish}(\beta) = -E\left(\frac{\partial^2 \log L(y_1,..,y_n;\beta)}{\partial^2 \beta}\right)$$

As $n \to \infty$ a MLE reach the minimal variance $I_{fish}$.

Even though for most models $I_{fish}$ is unavailable analytically, as you know by now, an approximate is given by $I_{fish} \approx H$, the Hessian matrix at the MLE $\hat{\beta}$. 