Lecture 11 | 12.05.2026

Regression models

beyond linearity

Linear regression models

Normal linear regression model

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, for $\varepsilon \sim N(0, \sigma^2)$, and $\sigma^2 > 0$
- random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, n\}$ from the distribution $F_{(Y, \boldsymbol{X})}$
- conditional distribution of $Y|\boldsymbol{X}$ is normal, i.e., $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top \beta, \sigma^2)$
- marginal distribution of $\boldsymbol{X} \sim f_{\boldsymbol{X}}$ does not depend on $\beta \in \mathbb{R}^p$, nor $\sigma^2 > 0$
- parameter estimates $\widehat{\beta}$ are BLUE and normaly distributed (LSE/MLE)

Linear regression model without normality

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2)$, $E\varepsilon^2 = \sigma^2 \in (0, \infty)$
- mean ($E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta$) and variance ($Var(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{X})$) specification
- conditional distribution of $Y|\boldsymbol{X}$ is, otherwise, left unspecified (LSE only)
- the marginal distribution of $f_{\boldsymbol{X}}$ does not depend on the unknown parameters
- parameter estimates (LSE only) are BLUE and asymptotically normal

Linear regression models

Normal linear regression model

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, for $\varepsilon \sim N(0, \sigma^2)$, and $\sigma^2 > 0$
- random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, n\}$ from the distribution $F_{(Y, \boldsymbol{X})}$
- conditional distribution of $Y|\boldsymbol{X}$ is normal, i.e., $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top \beta, \sigma^2)$
- marginal distribution of $\boldsymbol{X} \sim f_{\boldsymbol{X}}$ does not depend on $\beta \in \mathbb{R}^p$, nor $\sigma^2 > 0$
- parameter estimates $\widehat{\beta}$ are BLUE and normaly distributed (LSE/MLE)

Linear regression model without normality

- generic regression model $Y = \boldsymbol{X}^\top \beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2)$, $E\varepsilon^2 = \sigma^2 \in (0, \infty)$
- mean ($E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta$) and variance ($Var(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{X})$) specification
- conditional distribution of $Y|\boldsymbol{X}$ is, otherwise, left unspecified (LSE only)
- the marginal distribution of $f_{\boldsymbol{X}}$ does not depend on the unknown parameters
- parameter estimates (LSE only) are BLUE and asymptotically normal

Recall, that **linearity** + **normality** = **"lightness of being"** but linear regression models without the assumptions of normality introduce just a minor complication...

Linear regression models

Normal linear regression model

- generic regression model $Y = \boldsymbol{X}^\top\beta + \varepsilon$, for $\varepsilon \sim N(0, \sigma^2)$, and $\sigma^2 > 0$
- random sample $\{(Y_i, \boldsymbol{X}_i^\top)^\top;\ 1 = 1, \dots, n\}$ from the distribution $F_{(Y, \boldsymbol{X})}$
- conditional distribution of $Y|\boldsymbol{X}$ is normal, i.e., $Y|\boldsymbol{X} \sim N(\boldsymbol{X}^\top\beta, \sigma^2)$
- marginal distribution of $\boldsymbol{X} \sim f_{\boldsymbol{X}}$ does not depend on $\beta \in \mathbb{R}^p$, nor $\sigma^2 > 0$
- parameter estimates $\widehat{\beta}$ are BLUE and normaly distributed (LSE/MLE)

Linear regression model without normality

- generic regression model $Y = \boldsymbol{X}^\top\beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2)$, $E\varepsilon^2 = \sigma^2 \in (0, \infty)$
- mean $(E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta)$ and variance $(Var(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}))$ specification
- conditional distribution of $Y|\boldsymbol{X}$ is, otherwise, left unspecified (LSE only)
- the marginal distribution of $f_{\boldsymbol{X}}$ does not depend on the unknown parameters
- parameter estimates (LSE only) are BLUE and asymptotically normal

Recall, that **linearity** + **normality** = **"lightness of being"** but linear regression models without the assumptions of normality introduce just a minor complication...

Thus, the linearity property seems to be a way more crucial!
(linearity of the predictor, linearity of the least squares, linearity of the expectation, linearity of the normal distribution, ...)

Beyond linearity ...

- In practice, however, **the truth is (almost) never linear!**
  (on the other hand, the linearity assumption is a convenient approximation)
- **What to do, when the linearity assumption fails?**
  (the answer usually depends on specifc reasons why the linearity fails)
- **Note that there are a few levels of linearity in the model**
  (linearity of the predictor, linearity of the expectation, linearity of LS, ...)

Beyond linearity ...

- In practice, however, **the truth is (almost) never linear!**
  (on the other hand, the linearity assumption is a convenient approximation)
- **What to do, when the linearity assumption fails?**
  (the answer usually depends on specifc reasons why the linearity fails)
- **Note that there are a few levels of linearity in the model**
  (linearity of the predictor, linearity of the expectation, linearity of LS, ...)

  - the data are too volatile (robust estimation approaches)
  - the data are too flexibile (higher order approximations/splines)
  - the data are too irregular (piecewise approximation)
  - the data are too complex (additive models)
  - the nature of Y contradicts the linear model (GLM)
  - and many more reasons (and way more alternatives)

Two common generalizations beyond linearity

- **Linearity of the predictor**
  - linear predictor in terms of $\boldsymbol{X}^\top\beta$, where $\beta \in \mathbb{R}^p$ are unknown parameters
  - the linear predictor is directly associated with the (theoretical) quantity of interest – the (conditional) expectation of Y (i.e, $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$)
  - however, this direct association may not be realistic in some situations
  - $\Longrightarrow$ **generalized linear models (GLM)** & **non-linear models (NLS)**

Two common generalizations beyond linearity

- **Linearity of the predictor**
  - linear predictor in terms of $\boldsymbol{X}^\top\beta$, where $\beta \in \mathbb{R}^p$ are unknown parameters
  - the linear predictor is directly associated with the (theoretical) quantity of interest – the (conditional) expectation of Y (i.e, $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$)
  - however, this direct association may not be realistic in some situations
  - $\Longrightarrow$ **generalized linear models (GLM)** & **non-linear models (NLS)**

- **Linearity of the expectation**
  - the expectation $EY = \int_{\mathbb{R}} x \mathrm{d}F_Y(x)$ of some random variable $Y \sim F_Y$ is a linear functional
  - the expectation is also one of the most important characteristics of some unknown population (the distribution of the random variable)
  - on the other hand, the expectation offers only a very limited information about the behavior of $Y \sim F_Y$
  - $\Longrightarrow$ **quantile regression, expectile regression, or m-regression in general**

1. Generalized linear models (GLM)

So far, all regression models focussed on the response variable $Y \in \mathbb{R}$ that was apriori assumed to be continuous and the conditional distribution of $Y|\boldsymbol{X}$ was assumed to be normal or, at least, close to normal... with the expectation being $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$

In practical applications, however, the domain of Y can be somehow restricted...
(but the linear predictor $\boldsymbol{X}^\top\beta$ is generally unrestricted, i.e., $\boldsymbol{X}^\top\beta \in \mathbb{R}$)

- $Y \in \mathbb{N} \cup \{0\}$ (counts)
- $Y \in \{1, \dots, K\}$ for $K \in \mathbb{N}$ (categories/label)
- $Y \in \{0, 1\}$ (true/false)
- ...

1. Generalized linear models (GLM)

So far, all regression models focussed on the response variable $Y \in \mathbb{R}$ that was apriori assumed to be continuous and the conditional distribution of $Y|\boldsymbol{X}$ was assumed to be normal or, at least, close to normal... with the expectation being $E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top\beta$

In practical applications, however, the domain of Y can be somehow restricted...
(but the linear predictor $\boldsymbol{X}^\top\beta$ is generally unrestricted, i.e., $\boldsymbol{X}^\top\beta \in \mathbb{R}$)

- $Y \in \mathbb{N} \cup \{0\}$ (counts)
- $Y \in \{1, \dots, K\}$ for $K \in \mathbb{N}$ (categories/label)
- $Y \in \{0, 1\}$ (true/false)
- ...

Note that despite the fact that the domain of Y is restricted, the mean parameter of Y (the conditional mean of $Y|\boldsymbol{X}$ respectively) is still assumed to be from some compact subset, $\mathcal{M} \subset \mathbb{R}$... This becomes very useful in the following models...

Linear models with a flavour of nonlinearity

- In a standard linear model, the conditional mean is modelled as

$$E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \boldsymbol{\beta}, \quad \text{for } \boldsymbol{\beta} \in \mathbb{R}^p$$

while the variance structure $Var[Y|\boldsymbol{X}]$ is modeled separately from the mean structure (e.g., $Var[Y|\boldsymbol{X}] = \sigma^2 \mathbb{I}$)

Linear models with a flavour of nonlinearity

- In a standard linear model, the conditional mean is modelled as

$$E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

while the variance structure $Var[Y|\boldsymbol{X}]$ is modeled separately from the mean structure (e.g., $Var[Y|\boldsymbol{X}] = \sigma^2 \mathbb{I}$)

- In a generalized linear model, the conditional mean is modelled as

$$g(E[Y|\boldsymbol{X}]) = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

for some non-linear link function $g : \mathcal{M} \to \mathbb{R}$ (typically continuous, smooth, twice differentiable, but nonlinear)

Linear models with a flavour of nonlinearity

- In a standard linear model, the conditional mean is modelled as

$$E[Y|\boldsymbol{X}] = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

while the variance structure $Var[Y|\boldsymbol{X}]$ is modeled separately from the mean structure (e.g., $Var[Y|\boldsymbol{X}] = \sigma^2 \mathbb{I}$)

- In a generalized linear model, the conditional mean is modelled as

$$g(E[Y|\boldsymbol{X}]) = \boldsymbol{X}^\top \beta, \quad \text{for } \beta \in \mathbb{R}^p$$

for some non-linear link function $g : \mathcal{M} \to \mathbb{R}$ (typically continuous, smooth, twice differentiable, but nonlinear)

- Moreover, the variance structure typically depends on the mean structure

$$Var[Y|\boldsymbol{X}] = v(E[Y|\mathbb{X}])$$

where $v : \mathbb{M} \to (0, \infty)$ is some known (variance) function

Example 1: Logistic regression

Logistic regression

- the response variable $Y \in \mathbb{R}$ takes only two possible values, $Y \in \{0, 1\}$
- the conditional distribution of $Y|\boldsymbol{X}$ is alternative, with $p_{\boldsymbol{x}} = E[Y|\boldsymbol{X} = \boldsymbol{x}]$
- the conditional mean $\mu_{\boldsymbol{x}} = E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is modeled with the linear predictor $\boldsymbol{X}^\top\beta$ using the logit link function $g(x) = \log[x/(1-x)]$
- the model assumes the mean structure

$$logit(\mu_{\boldsymbol{x}}) = \log\frac{E[Y|\boldsymbol{X} = \boldsymbol{x}]}{1 - E[Y|\boldsymbol{X} = \boldsymbol{x}]} = \log\frac{P[Y = 1|\boldsymbol{X} = \boldsymbol{x}]}{1 - P[Y = 1|\boldsymbol{X} = \boldsymbol{x}]} = \boldsymbol{x}^\top\beta$$

- the model assumes the variance structure which depends on the mean $\mu_{\boldsymbol{x}}$

$$Var[Y|\boldsymbol{X} = \boldsymbol{x}] = v(\mu_{\boldsymbol{x}}) = \mu_{\boldsymbol{X}}(1 - \mu_{\boldsymbol{X}})$$

- the model is typically interpreted in terms of multiplicative comparisons, the odds ratios respectively (or logarithms of the odds ratios)
  (note the difference between a probability, an odd, and the odds ratio)

Example 2: Poisson regression

Poisson regression

- the response variable $Y \in \mathbb{N} \cup \{0\}$ represents integer counts (including 0)
- the conditional distribution of $Y|\boldsymbol{X}$ is Poisson, with $\lambda_{\boldsymbol{x}} = E[Y|\boldsymbol{X} = \boldsymbol{x}]$
- the conditional mean $\lambda_{\boldsymbol{X}} = E[Y|\boldsymbol{X} = \boldsymbol{x}]$ is modeled with the linear predictor $\boldsymbol{X}^\top \beta$ using the log link function $g(x) = \log x$
- the model assumes the mean structure

$$log(\lambda_{\boldsymbol{x}}) = \log E[Y|\boldsymbol{X} = \boldsymbol{x}] = \boldsymbol{x}^\top \beta$$

- the model assumes the variance structure which depends on the mean $\lambda_{\boldsymbol{X}} > 0$ and some additional dispersion parameter $\phi > 0$

$$Var[Y|\boldsymbol{X} = \boldsymbol{x}] = v(\lambda_{\boldsymbol{x}})\phi = \phi\lambda_{\boldsymbol{x}}$$

- the model is interpreted in terms of multiplicative comparisons and the parameters are interpreted in terms of the proportional changes of the conditional expectations of two sub-populations

Example 3: Special cases – back to normality

In general, the exponential family of distributions, expressed in terms of densities as

$$\mathcal{F} = \{f(y);\ f(y) = (\theta y - \psi(y))/\phi + c(y, \phi), \text{for } \theta \in \mathbb{R}, \text{ and } \phi > 0\},$$

contains also the normal distribution (beside many others: the alternative distribution, the binomial distribution, the Poisson distribution, exponential, etc.)

- **Classical linear regression model**
  - continuous response $Y \in \mathbb{R}$
  - identity link function $g(x) = x$
  - constant variance function $v(x) = 1$ and $\phi = \sigma^2$
- **Multinomial regression model**
- **Exponential data model**
- **...**

2. Nonlinear regression models

- In linear models and generalized linear models as well, the conditional mean is modeled (using a proper link function) as a linear combination of the response variables and the subset of unknown parameters...
- if the class of available models is not reach enough (and we still prefer a parametric model structure) then nonlinear regression models can serve a good alternative (sometimes computationally demanding)...
- the idea of nonlinear models is to use a general parametric (but nonlinear) regression function $f : \mathbb{R}^{p \times q} \to \mathbb{R}$, such that

$$E[Y|\boldsymbol{X}] = f(\boldsymbol{X}, \beta),$$

  where $\boldsymbol{X} \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ (generally for $p \neq q$)
- Note that the nonlinear element (the nonlinear function f) is now introduced on the right side of the regression model formula and typically it is not assumed that f should be regular (or continuous, etc.)

Nonlinear regression: Some examples

There are, of course, plenty of different models with various analytical structure and different regularity properties—smoothness, continuity. Typical nonlinear models are, for instance, various population models...

- **Exponential growth model**

$$f(x, \beta, \alpha) = \alpha \exp\{X\beta\}$$

$\rightarrow$ for some parameters $a > 0$ and $\beta > 0$;

- **Logistic growth model**

$$f(X, \beta, \alpha, K) = \frac{K}{1 + be^{-X\beta}};$$

$\rightarrow$ for some parameters $\alpha, \beta, K > 0$;

- **Gomertz growth model**

$$f(X, \beta, \alpha, K) = K \cdot exp\{-\beta e^{-\alpha t}\};$$

$\rightarrow$ for some parameters $\alpha, \beta, K > 0$;

Solutions for nonlinear regression models

- Note that all three nonlinear models above can not be solved by using classical method of the least squares...
  (i.e., no explicit solution can be directly derived)

- Thus, different computation strategies must be used to obtain the model solution—the estimates for the unkown parameters $\alpha, \beta, K > 0$

- Such computational methods may involve:
  - reparametrization into a linear model and applying least squares
  - model approximation and least squares
  - various iterative solutions

- Note, that as far as the unknown regression function is unspecified, the corresponding minimization problem may not even be convex!

Generalized nonlinear models

- **Advanced, but still possible....**

$$g(E[Y|\boldsymbol{X}]) = f(\boldsymbol{X}, \beta)$$

  where two additional sources of nonlinearity are introduced at the same time—the nonlinear link function g and the nonlinear predictor function f

- **Some challenges**
  - mostly, the interpretation of $\beta \in \mathbb{R}$ is not straightforward/possible
  - due to nonlinearity, various computational issues (solution instability)
  - difficult statistical inference typically perfomed by simulations

3. Regression models beyond expectation

Least squares

In an ordinary linear regression model (without the normality assumption) the likelihood can not be obtained and the estimates for $\beta' \in \mathbb{R}^p$ are obtained by minimizing least squares

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^p}{Argmin} \quad \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2$$

3. Regression models beyond expectation

- **Least squares**
  In an ordinary linear regression model (without the normality assumption) the likelihood can not be obtained and the estimates for $\beta' \in \mathbb{R}^p$ are obtained by minimizing least squares

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^p}{Argmin} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2$$

- **Maximum likelihood**
  In a normal linear regression model (under the normality assumption) the full likelihood for $\beta \in \mathbb{R}^p$ and $\sigma > 0$ can be formulated and the estimates are obtained by maximizing the likelihood function

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^p;\ \sigma^2 > 0}{Argmax} (2\pi\sigma^2)^{-N/2} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} (Y_i - \boldsymbol{X}_i^\top \beta)^2 \right\}$$

Theoretical and empirical expectation

- For some real random variable $X \sim F_X$ (and the density f with respect to the Lebesgue or count measure) and some measurable function $h : \mathbb{R} \to \mathbb{R}$ we can obtain the expectation (if the integral exists) as

$$Eh(X) = \int_{\mathbb{R}} h(x) \mathrm{d}F_X(x) = \int_{\mathbb{R}} h(x) f(x) \mathrm{d}x$$

Theoretical and empirical expectation

- For some real random variable $X \sim F_X$ (and the density f with respect to the Lebesgue or count measure) and some measurable function $h : \mathbb{R} \to \mathbb{R}$ we can obtain the expectation (if the integral exists) as

$$Eh(X) = \int_{\mathbb{R}} h(x) \mathrm{d}F_X(x) = \int_{\mathbb{R}} h(x) f(x) \mathrm{d}x$$

- For the random sample $X_1, \ldots, X_N$ drawn from the same distribution as the distribution of $X \sim F_X$ we can construct the empirical distribution function F_N and the empirical counterpart for $Eh(X)$ (i.e., the empirical estimate)

$$\widehat{Eh(X)} = \int_{\mathbb{R}} h(x) \mathrm{d}F_N(x) = \sum_{i=1}^{N} h(X_i)$$

Theoretical and empirical expectation

- For some real random variable $X \sim F_X$ (and the density f with respect to the Lebesgue or count measure) and some measurable function $h : \mathbb{R} \to \mathbb{R}$ we can obtain the expectation (if the integral exists) as

$$Eh(X) = \int_{\mathbb{R}} h(x) \mathrm{d}F_X(x) = \int_{\mathbb{R}} h(x) f(x) \mathrm{d}x$$

- For the random sample $X_1, \dots, X_N$ drawn from the same distribution as the distribution of $X \sim F_X$ we can construct the empirical distribution function F_N and the empirical counterpart for $Eh(X)$ (i.e., the empirical estimate)

$$\widehat{Eh(X)} = \int_{\mathbb{R}} h(x) \mathrm{d}F_N(x) = \sum_{i=1}^{N} h(X_i)$$

- The quantity (parameter) $\mu_h = Eh(X)$ is sometimes called the theoretical functional of the distribution F_X while the quantity $\widehat{\mu_h} = \widehat{Eh(X)}$ is called the (empirical) functional of the empirical distribution F_N *(of course, different functions can be used in place of h)*

Some common choices of the function *h*

- The expectation (theoretical quantity EX) and the average (empirical quantity $\overline{X}_N$) can be both obtained by a minimization problem with the choice of $h(x) = (x-a)^2$ where
  - $EX = Argmin_{a \in \mathbb{R}}\ E(X-a)^2 = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x-a)^2 \mathrm{d}F_X(x)$
  - $\overline{X}_N = \widehat{EX} = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x-a)^2 \mathrm{d}F_N(x) = Argmin_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$

Some common choices of the function h

- The expectation (theoretical quantity EX) and the average (empirical quantity $\overline{X}_N$) can be both obtained by a minimization problem with the choice of $h(x) = (x - a)^2$ where
  - $EX = Argmin_{a \in \mathbb{R}}\ E(X - a)^2 = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_X(x)$
  - $\overline{X}_N = \widehat{EX} = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_N(x) = Argmin_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$

- Note that in both cases we actually formulate the least squares problem (theoretical and empirical) and the solution is the theoretical mean and the empirical average (i.e., the estimate for the mean)
  This principle can be generalized even further—for the regression concepts and different forms of the function h

Some common choices of the function h

- The expectation (theoretical quantity EX) and the average (empirical quantity $\overline{X}_N$) can be both obtained by a minimization problem with the choice of $h(x) = (x - a)^2$ where
  - $EX = Argmin_{a \in \mathbb{R}}\ E(X - a)^2 = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_X(x)$
  - $\overline{X}_N = \widehat{EX} = Argmin_{a \in \mathbb{R}} \int_{\mathbb{R}} (x - a)^2 \mathrm{d}F_N(x) = Argmin_{a \in \mathbb{R}} \sum_{i=1}^{N} (X_i - a)^2$

- Note that in both cases we actually formulate the least squares problem (theoretical and empirical) and the solution is the theoretical mean and the empirical average (i.e., the estimate for the mean)
  This principle can be generalized even further—for the regression concepts and different forms of the function h

- Typical choices for h include: median regression for $h(x) = |x|$; quantile regression for $h_\tau(x) = \tau(x - \mathbb{I}_{\{x<0\}})$; expectile regression for $h_\tau(x) = |\tau - \mathbb{I}_{\{x<0\}}|x^2$; robust regression for $h(x) = \rho(x)$;

Basic properties of the regression variants

- **Median regression**
  - more robust than the standard least squares regression
  - for symmetric error distributions the median corresponds with the mean
  - easy and straightforward intepretation of the estimated parameters
- **Quantile regression**
  - generalization of the median regression (obtained as a special case)
  - provideds a complex insight about the conditional distribution of $Y|\boldsymbol{X}$
  - relatively easy interpretation but not that much popular in practice
- **Expectle regression**
  - generalization of the least squares (which are obtained for $\tau = 0.5$)
  - expectiles form elastic and coherent risk measures (unlike quantiles)
  - relatively difficult interpretation but very popular in risk theory
- **Robust regression**
  - generalization of the regression for outlyiers and heavy-talied distributions
  - least squares for $\rho(x) = X^2$; median regression for $\rho(x) = |x|$; maximum likelihood for $\rho(x) = -\log(x)$
  - other choices are common in practice as well (e.g., Huber function, Tukey function, Andrew's function, ...)

Exam terms – Summer Term 2026

- **Tuesday, 19.05.2026** (starting at 10:40 in K11)
- **Thursday, 21.05.2026** (starting at 12:20 in K11)
- **Tuesday, 26.05.2026** (starting at 10:40 in K11)
- **Thursday, 28.05.2026** (starting at 9:00 in K11)
- **Friday, 19.06.2026** (starting at 14:00 in K11)

- At least one other exam term in **September 202ž**