6 Regression model

Let random variable X depends on random variable or vector Z (Z-values are given or observed).

Denote by $f(t)$ the function of dependence the average value of X on Z:

$$E(\mathrm{X}|\mathrm{Z}=t)=f(t),$$

$f(t)$ – is called regression line; and $x=f(t)$ – regression equation.

Let consider experiment: for n values Z: $t_1, t_2, \dots t_n$ obtain observed values $X_1, X_2, \dots X_n$; and introduce $\varepsilon_i \equiv X_i - E(X|z=t_i) = X_i - f(t_i)$ – difference between the observed random variables in *i*-th experiment and expectation of X provided that Z equal t_i .

About joint distribution ε_i usully assumed, that vector $\boldsymbol{\varepsilon}$ consists of independent, normally distributed random variables with zero mean:

$$E\varepsilon_i = EX_i - f(t_i) = E(X|Z=t_i) - E(X|X=t_i) = 0$$

Goal: To determine (estimate) $f(t)$, (t_i – are not random; ε_i , X_i – are random), use the most appropriate function class since the function is uniquely determined by the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ needed to estimate $\boldsymbol{\theta}$.

Likelihood method: *based on idea of maximization likelihood function of* $\vec{X}=(X_1, X_2, \dots X_n)$ *(depends on the sample).*

The basic assumption:

a) ε_i – independent, identically distributed
b) distribution $h(x)$ – symmetrical
c) family of distribution has zero mean and unknown variance (Normal, Students, …) X_i and ε_i have linear dependence, so X_i – have center in the point $f(t_i)$

So X_i have density function $h(x-f(t_i))$

$$f(\vec{X}, \theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^{n} h\big(X_i - f(t_i)\big) = h(\varepsilon_1)\cdot h(\varepsilon_2)\cdot \dots \cdot h(\varepsilon_k) \xrightarrow{\boldsymbol{\theta}} max$$

Case of normal distributed $\varepsilon_i (\varepsilon_i \in N(0, \sigma^2)$, independent)

Least square method

Maximum probability

$$f(\vec{X}, \boldsymbol{\theta}) = \prod \frac{1}{\sqrt{2\pi}\,\sigma} \, exp\left\{-\frac{(X_i - f(t_i))^2}{2\sigma^2}\right\} = \frac{1}{\sigma^n (2\pi)^{n/2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - f(t_i))^2\right\} \to max$$

achieves at a minimum of the sum squares: $\sum_{i=1}^{n}(X_i - f(t_i))^2 = \sum \varepsilon_i^2$.

$$\sum_{i=1}^{n}(X_i - f(t_i))^2 \to min$$

Example: linear regression:

$$X_i = \theta_1 + t_i\theta_2, \qquad i = 1, \dots, n$$

$$L \equiv \sum \varepsilon_i^2 = \sum (x_i - \theta_1 - t_i\theta_2)^2 \to min$$

Here, the points suspicious (suspicious [səsˈpɪʃəs] – подозрительный) of an extremum are determined by solving the linear system of two equations obtained as a derivative of L on each unknown parameter θ_1, θ_2, of the desired function F

$$\left\{\frac{\partial L}{\partial \theta_1} = 0\,, \qquad \frac{\partial L}{\partial \theta_2} = 0\right\}$$

$$\left\{\hat{\theta}_1 = \bar{x} - \bar{t}\hat{\theta}_2\,, \qquad \hat{\theta}_2 = \frac{\frac{1}{n}\sum X_i t_i - \bar{X}\,\bar{t}}{\frac{1}{n}\sum (t_i - \bar{t})^2}\right\}$$

Sample correlation coefficient:

$$\rho = \left(\frac{1}{n}\sum (X_i - \bar{X})(\,t_i - \bar{t})\right) / \sqrt{\frac{1}{n}\sum (t_i - \bar{t})^2 \;\frac{1}{n}(X_i - \bar{X})^2}$$

is the *measure* of linear dependence between $X_1, X_2, \dots X_n$ and $t_1, t_2, \dots t_n$.

General model of linear Regression

Let $\vec{Z} = (Z_1, Z_2, \dots, Z_k)^T$ – random regression factors (object properties or variables), $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ – unknown linear regression parameters and

$E(X|\vec{Z}) = f(\vec{z}) = \beta_1 Z_1 + \cdots + \beta_k Z_k$, here $Z^{(i)} = (Z_1^{(i)}, Z_2^{(i)}, \ldots, Z_k^{(i)})$ – i-th experiment. Results of n experiments, n$\gg k$, with responses $\vec{X} = (X_1, X_2, \ldots X_n)^T$ is represented by this system

$$\begin{cases} X_1 = \beta_1 Z_1^{(1)} + \cdots + \beta_k Z_k^{(1)} + \varepsilon_1 \\ X_2 = \beta_1 Z_1^{(2)} + \cdots + \beta_k Z_k^{(2)} + \varepsilon_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_n = \beta_1 Z_1^{(n)} + \cdots + \beta_k Z_k^{(n)} + \varepsilon_n \end{cases}$$

it is form of multivariate regression or by matrix notations, $\vec{X} = \vec{Z} * \vec{\beta} + \vec{\varepsilon}$ (*)

$$Z = \begin{pmatrix} Z_1^{(1)} & \ldots & Z_k^{(1)} \\ & \ldots & \\ & \ldots & \\ Z_1^{(n)} & \ldots & Z_k^{(n)} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ . \\ . \\ \beta_k \end{pmatrix} \equiv \begin{pmatrix} \vec{Z}^{(1)} \\ . \\ . \\ \vec{Z}^{(n)} \end{pmatrix}$$

Least square method and normal equations

Lemma 1. If Z has rank k, i.e. all column independent then A = $Z^T * Z$ – positive matrix (p.m.).

Proof

According to definition of (p.m.) if $\vec{t}^T \mathrm{A} t \geq 0$ for $\forall \vec{t} = (t_1, t_2, \ldots, t_n) \in R^n$ and $\vec{t}^T \mathrm{A} t = 0 \Leftrightarrow \vec{t} = \vec{0}$ [as *for* $u = (u_1, \ldots, u_n)^T$, $||\vec{u}||^2 = \vec{u}^T \vec{u} = \sum u_i^2 \geq 0$, $||\vec{u}||^2 = 0 \Leftrightarrow \vec{u} = \vec{0}$].

A – symmetrical, i.e. $A^T = A$: $\vec{t}^T A \vec{t} = \vec{t}^T Z^T Z \vec{t} = (Z\vec{t})^T Z\vec{t} = \|Z\vec{t}\|^2 \geq 0$ and $\|Z\vec{t}\|^2 = \vec{0}$ if $Z\vec{t} = \vec{0}$, but rank (Z)=k $\Rightarrow \vec{t} = \vec{0}$, this is contradiction because of $\vec{t}$ – an arbitrary vector (for $\forall \vec{t} = (t_1, t_2, \ldots, t_n) \in R^n$), so A – positive matrix.

Lemma 2 Positive definiteness and symmetry of matrix A imply the existence of $\sqrt{A}$ is a real symmetric matrix such that $\sqrt{A}\sqrt{A} = A$.

Proof For any positive and symmetrical matrix factorization A=$Q^T D Q$, where D – diagonal matrix (with eigenvalues on diagonal, eigenvalues > 0) and Q – orthogonal (consists of eigenvectors).

$A = Q^T \sqrt{D}\sqrt{D}\ Q = \left(\sqrt{D}\,Q\right)^T \sqrt{D}\,Q \Rightarrow \sqrt{A} = \sqrt{D}\,Q$, *proven.*

Let's find least squares estimation $\hat{\beta} = \vec{\beta}^*$, satisfying the following optimization problem (minimal solution exists):

$$S(\vec{\beta}) \rightarrow \min \qquad (1)$$

$$S(\vec{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = ||\vec{\varepsilon}||^2 = \left|\left|\vec{X} - Z\,\vec{\beta}\right|\right|^2 = \left(\vec{X} - Z\,\vec{\beta}\right)^T \left(\vec{X} - Z\,\vec{\beta}\right)$$

There are two approaches to find $\hat{\beta}$:

1. solving system $\left\{\frac{\partial S(\vec{\beta})}{\partial \beta_1} = 0, \ldots, \frac{\partial S(\vec{\beta})}{\partial \beta_k} = 0\right\}$ to find extremum points
2. $S(\vec{\beta})$ – square of distance between points $\vec{X} \in R^n$ and $Z\vec{\beta}$, $(\cdot)Z\vec{\beta} \in hyperplane$ where any vector of the form $Z\vec{t}, (\vec{t} \in R^k)$ lies. Minimal distance $S(\vec{\beta}^*)$ we obtain because of vector $\vec{X} - z\hat{\beta}$ is orthogonal to all vectors of hyper plane $Z\vec{t}\,(\forall \vec{t} \in R^k)$, so $\left(Z\vec{t}, \vec{X} - Z\hat{\beta}\right) = (Z\vec{t})^T\left(\vec{X} - Z\hat{\beta}\right) = \vec{t}^T\left(Z^T\vec{X} - Z^T Z\hat{\beta}\right) = 0$ (scalar product is equal to 0) and for any $\vec{t}^T \neq 0$, f.e. basis vector $\vec{t}^T = (00010\ldots0) \in R^k$ obtain

$$Z^T\vec{X} - Z^T Z\hat{\beta} = 0 \Rightarrow$$

$Z^T Z\hat{\beta} = Z^T\vec{X}$ or $A\hat{\beta} = Z^T\vec{X}$ (2)

so least squares estimation – is solution of (2). According to Lemma 1 they have the same single solution

$\hat{\beta} = \vec{\beta}^* = A^{-1}Z^T\vec{X}$ (3)

if rectangle matrix Z size (n,k) has rank = k, $k \leq n$, both equation (2) is called normal equations.

Conclusion:

If $\vec{\varepsilon}$ consists of independent r.v. $\forall \vec{\varepsilon}_i \in N(0, \sigma^2)$ ⇒least squares estimation is the same as likelihood estimation for $\vec{\sigma}^2 = \frac{1}{n}\sum \hat{\varepsilon}_i^2 = \frac{1}{n}||\vec{X} - Z\hat{\beta}||^2 = \frac{1}{n}S(\hat{\beta})$,

here S determined in (1).

Properties of least squares estimations:

p1 $\hat{\beta} - \vec{\beta} = A^{-1}Z^T\vec{\varepsilon}$ (4)

proof:

Substitute (3) in (4) and use (*) and equality $A = Z^T Z$, can obtain $A^{-1}Z^T\vec{X} - \vec{\beta} = A^{-1}Z^T(Z * \vec{\beta} + \vec{\varepsilon}) - \vec{\beta} = A^{-1}A * \vec{\beta} + A^{-1}Z^T\vec{\varepsilon} - \vec{\beta} = A^{-1}Z^T\varepsilon$, *proven.*

p2 If $E\vec{\varepsilon} = 0$ then $\hat{\beta}$ – unbiased estimation for $\vec{\beta}$.

proof: $E\hat{\beta} =^{(p.1.)} \vec{\beta} + A^{-1}Z^T E\vec{\varepsilon} = \vec{\beta}$. *proven.*

Let following assumptions are true:

Matrix Z has rank=k and all columns of Z linear independent.

I. Vector $\vec{\varepsilon}$ consists of independent random values$\in N(0, \sigma^2)$.

Recall for any $\vec{x}$: $D\vec{x} = E(\vec{x} - E\vec{x})(\vec{x} - E\vec{x})^T$ – covariance matrix;

II. $\operatorname{cov}(x_i, x_j) = E(x_i - Ex_i)(x_j - Ex_j)$ and $D\vec{\varepsilon} = \sigma^2 E_n$; $E - eye(n) \equiv E_n$ – identity matrix size *n*

p3 . Let I and II assumptions are true, then $\sqrt{A} * \hat{\beta}$ has covariance matrix of diagonal type and equal to $\sigma^2 E_k$, $(\sqrt{A} * \hat{\beta} = \sigma^2 E_k)$; it means that coordinates of $\sqrt{A}\hat{\beta}$ – uncorrelated (without proof)

Theorem. Let I and II assumptions are true, then

1) Vector $\frac{1}{\sigma}\sqrt{A}(\hat{\beta} - \vec{\beta})$ has k-dimensional normal standard distribution (consists of *k* independent random variables $\in N(0,1)$)
2) $n\hat{\sigma}^2/\sigma^2 = \left\|\vec{X} - Z\hat{\beta}\right\|^2/\sigma^2$ has χ^2_{n-k} distribution and doesn't depends on $\hat{\beta}$ (Interesting result!)
3) $(\sigma^2)^* = n\hat{\sigma}^2 = \frac{1}{n-k}||\vec{X} - Z\hat{\beta}||^2$ – unbiased estimation for σ^2 (without proof)