Lineární modely:

$$Y = X\beta + \sigma Z$$

$${}_n\left(\,\right)_1 = {}_n\left(\quad\right)_k \left(\,\right)_1^k + \left(\,\right)_1^n$$

Hledáme b, ŝt.

kde nejlepší odhad β.

= minimalizuje $\sum(\hat{Y}-Y)^2$

$\Longrightarrow$ přiřazuje průmět do podprostoru:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \beta_1 \begin{pmatrix} X_{11} \\ \vdots \\ X_{n1} \end{pmatrix} + \beta_2 \begin{pmatrix} X_{12} \\ \vdots \\ X_{n2} \end{pmatrix} + \dots + \beta_k \begin{pmatrix} X_{1k} \\ \vdots \\ X_{nk} \end{pmatrix} + \sigma Z$$

$M(X)$ = podprostor generovaný $\begin{pmatrix} X_{1j} \\ \vdots \\ X_{nj} \end{pmatrix}$

příklad:

předp. sloupce v X jsou ortonormální báze $M(X)$.

$\Rightarrow$ P matice s ortonormálními sloupci:

$$P^T P = E_k \qquad \begin{array}{c c} & \begin{array}{cc} P & R \end{array} \\ n & \left(\begin{array}{c|c} \cdots & \cdots \end{array} \right) \\ & \begin{array}{cc} \underbrace{\hphantom{xxxx}}_{k} & \underbrace{\hphantom{xxx}}_{n-k} \end{array} \end{array}$$

$$Q = (P, R)$$

$$R^T R = E_{n-k}$$

$$\underbrace{PP^T + RR^T}_{QQ^T} = E_n \qquad (P^T R = 0)$$

$$R^T X = 0 \quad (\text{sloupce v } X \text{ generují } M(X))$$

$$\boxed{y \in \mathbb{R}^n \qquad \hat{y} = PP^T y \text{ je kolmý průmět}}$$

$\hat{y}$ minimizuje $\|y-z\|^2$ pro $z \in M(X)$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad = Pt$

Dk: $\|y-z\|^2 = \|(y - PP^T y) + (PP^T y - Pt)\|^2$

$= \|RR^T y\|^2 + \|\hat{y} - Pt\|^2$

$\geq \|RR^T y\|^2$ ✓

$$\boxed{Y = X\beta + \sigma Z}$$

$$S(\beta) = \|Y - X\beta\|^2 = \sum_{i=1}^{n} \Big(Y_i - \sum_{j=1}^{k} x_{ij}\beta_j\Big)^2$$

$$\boxed{U = R^T Z}$$

$$\hat{Y} = PP^T Y = PP^T (X\beta + \sigma Z)$$

$$= X\beta + \sigma P \underbrace{P^T Z}_{V} = X\beta + \sigma P V$$

Lemma: $Z_i \sim N(0,1)$, Q ortogonální, pak $U = Q^T Z$ je vektor nezávislých veličin s rozdělením $N(0,1)$ pokud Z_i byly nezávislé.

$t = Q^T z$, inverzní k ní je $z = Qt$,
jakobiho matice je 1, $\sum z_i^2 = \sum t_i^2$:

$$F_U(u) = P(U < u) = \int\limits_{z:\, Q^T z < u} \cdots \int (2\pi)^{-n/2} e^{-\sum z_i^2/2} \, dz_1 \cdots dz_n$$

$$= \int\limits_{t:\, t < u} \cdots \int (2\pi)^{-n/2} e^{-\sum t_i^2/2} \, dt_1 \cdots dt_n$$

$$= \prod_{i=1}^{n} \left(\int_{-\infty}^{u_i} (2\pi)^{-1/2} e^{-t_i^2/2} \, dt_i \right)$$

$$= \prod_{i=1}^{n} F_{u_i}(u_i).$$

$$Y = \overbrace{X\beta + \sigma PV}^{\in M(X)} + \underbrace{\sigma PV + \sigma RU}_{Z}$$

$$\Rightarrow \hat{Y} = X\beta + \sigma PV$$

Lemma: (1) $\hat{Y} \sim N(X\beta, \sigma^2 PP^T)$

(2) $Y - \hat{Y} \sim N(0, \sigma^2 RR^T)$

(3) $\|Y - \hat{Y}\|^2 / \sigma^2 \sim \chi^2_{n-k}$

$$\hat{Y} = Xb \Rightarrow \underbrace{X^T X}_{\text{ex. } (\cdot/\cdot)^{-1}} b = X^T \hat{Y} \Rightarrow b = (X^T X)^{-1} X^T \hat{Y}$$

$$X = PT \quad , T \text{ regulární matice}$$

$$b = (T^T P^T P T)^{-1} T^T P^T Y$$

$$= T^{-1} (P^T P)^{-1} \cancel{(T^T)^{-1} T^T} P^T (PT\beta + \sigma Z)$$

$$= \cancel{T^{-1} T}\, \beta + \sigma T^{-1} V$$

<u>Lema</u> $\quad b \sim N(\beta, \sigma^2 (X^T X)^{-1})$

Příklad – jeden výběr: $x^T = (1, \dots, 1)$

$$Y_1, \dots, Y_n \sim N(\beta, \sigma^2)$$

$$b = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$T = \frac{\bar{Y} - \beta_0}{s} \sqrt{n}$$

Reliabilita:

A .. náhodná veličina, stanovení

$Y = A + e$ (výsledek testu)

$Ee = 0$, nezávislost realizací A

$$R_A = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2}$$

$$\boxed{\begin{array}{l} EY = \mu \\ \operatorname{var} Y = \sigma_A^2 + \sigma^2 \end{array}}$$

$$\rho_{Y_1Y_2} = \frac{\operatorname{cov}(Y_1, Y_2)}{\sqrt{\operatorname{var} Y_1 \operatorname{var} Y_2}} = \frac{\operatorname{cov}(A+e_1, A+e_2)}{\operatorname{var} Y}$$

$$= \frac{\operatorname{var} A}{\operatorname{var} Y} = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2}$$

$$X = M + e, \quad Y = B + f$$

$$\rho_{XY} = \frac{\operatorname{cov}(M+e, B+f)}{\sqrt{\operatorname{var}(M+e)\operatorname{var}(B+f)}} = \frac{\operatorname{cov}(M,B)}{\sigma_M \sigma_B}\sqrt{\frac{\sigma_M^2}{\sigma_M^2+\sigma^2} \cdot \frac{\sigma_B^2}{\sigma_B^2+\sigma^2}}$$

$$= \rho_{MB}\sqrt{R_M R_B}$$

$$R_\Sigma = \boxed{\frac{\operatorname{var} \sum B_{j\delta}}{\operatorname{var} \sum Y_j}}$$

Cronbach:

$$R_\Sigma \geq \frac{m}{m-1}\left(1 - \frac{\sum_j \operatorname{var} Y_j}{\operatorname{var} \sum_j Y_j}\right) = \alpha$$

je nestranný odhad

$\Leftarrow \quad B_j = A + \beta_j, \quad \sum \bar{\beta}_j \Rightarrow$ A je vydělané po míře.