

# PSY117/454 Statistická analýza dat v psychologii – přednáška 6

---

## Vztahy mezi dvěma proměnnými II

Statistická predikce

Lineární regrese

# Statistická predikce

---

- Předpovídání nejpravděpodobnější hodnoty proměnné ze statistických údajů, které již známe.
  
- Jaký výsledek v inteligenčním testu lze nejspíše očekávat od náhodně přišedšího, víme-li, že test má přibližně normální rozložení s průměrem  $m$  a směrodatnou odchylkou  $s$ ?
  
- Jaká informace by nám pomohla zpřesnit náš odhad?
  - jeho/její pohlaví
  - vzdělání
  - výsledek v testu paměti
  - výsledek v jiném inteligenčním testu
  
- Korelace nám přímo neumožňuje určit predikovanou hodnotu (je třeba učinit přepočít přes  $z$ -skóry), protože udává pouze těsnost vztahu (těsnost nakupení bodů kolem nějaké přímky) nikoliv však sklon té přímky. Korelace není funkce.

# K predikci je třeba funkce

---

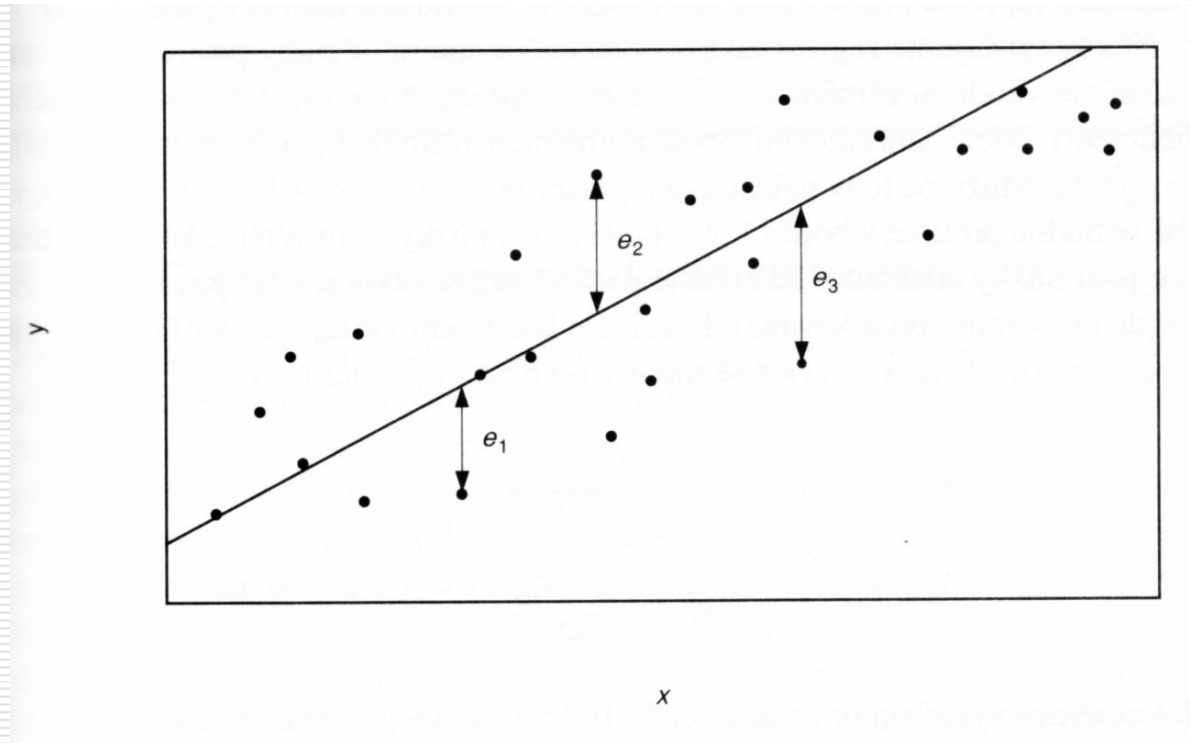
- $Y = f(X)$
- funkce je „návod“, jak ze známé hodnoty ( $X$ ) vypočítat tu neznámou ( $Y$ )
  - jsou různé funkce...
    - stanovené výčtem
    - trigonometrické, exponenciální a logaritmické ...
    - polynomické
      - lineární:  $Y = bX + a$  (rovná čára)
      - kvadratické:  $Y = cX^2 + bX + a$  (jedna zatáčka)
- ve statistice...
  - tuto funkci odhadujeme (modelujeme)
    - Jak dobře dokážeme vyjádřit (=predikovat) proměnnou  $Y$ , pomocí proměnné  $X$  a funkce  $f$ ?
  - říkáme výsledku výpočtu **odhad** ( $Y'$ ) a stanovení té funkce říkáme **regrese**
  - regrese  $Y$  na  $X$ :  $Y' = f(X) + e$  ,kde  $e = Y - Y'$  (1)
    - $e$  je reziduální hodnota (reziduum),  $Y$  je závislá p.,  $X$  je prediktor (nezáv.)
    - $e$  představuje všechny ostatní zdroje variability vyjma  $X$

---

AJ: function, polynomial, linear, quadratic, estimation, modelling, estimate  $n$ , regression, residual  $n$ , predictor, sources of variability(variance), dependent and independent variable

# Lineární regrese I. - odhad

- $Y' = a + bX + e$
- $b$  – směrnice
- $a$  – průsečík
  
- metoda nejmenších čtverců
- $b = r_{xy}(s_y/s_x)$
- $a = m_y - bm_x$



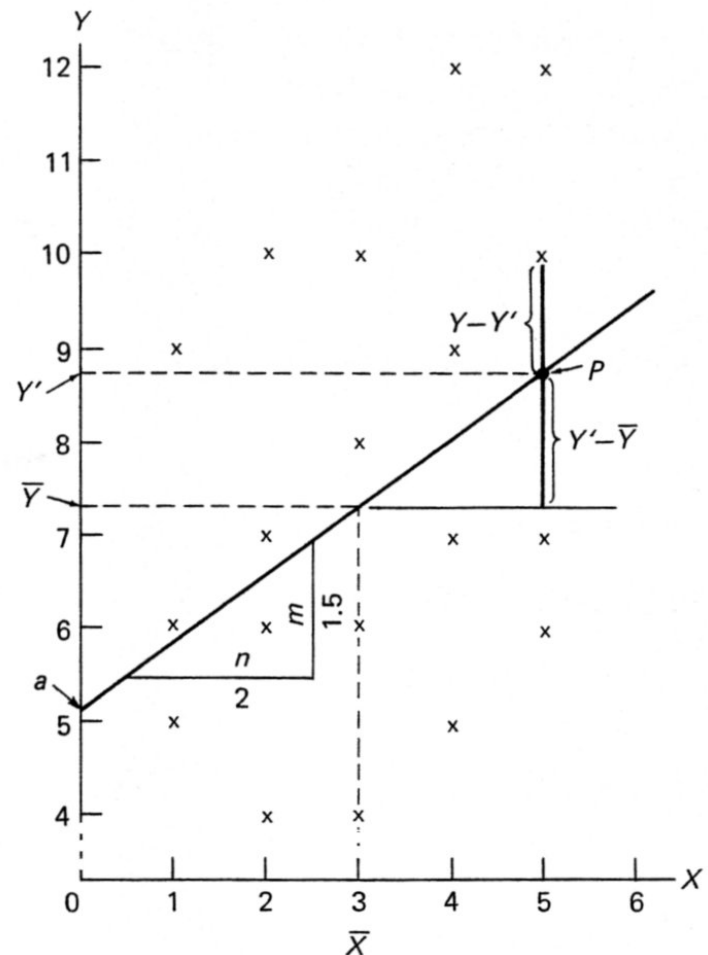
- jsou-li X a Y vyjádřeny v z-skórech, pak  $b = r_{xy}$

# Lineární regrese II. – úspěšnost predikce

$$s_{reg}^2 = \frac{\sum (\bar{Y} - Y')^2}{n-1} \quad s_{res}^2 = \frac{\sum (Y - Y')^2}{n-1}$$

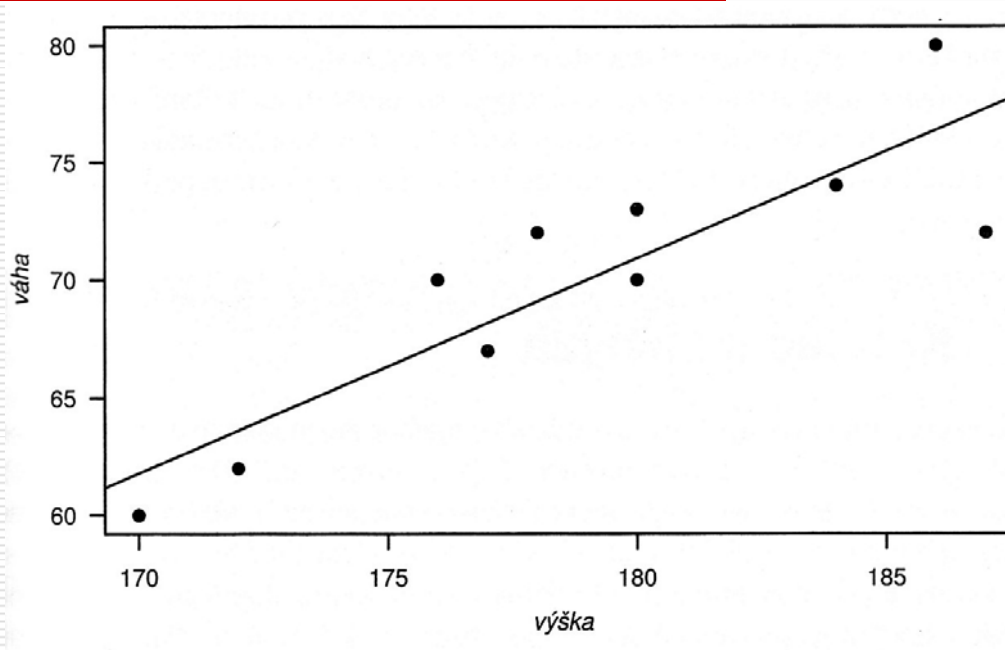
$$s_y^2 = \frac{\sum (Y - \bar{Y})^2}{n-1}$$

- $s_y^2 = s_{reg}^2 + s_{res}^2$  ( $SS_y = SS_{res} + SS_{reg}$ )
- $R^2 = s_{reg}^2 / s_y^2$
- Koeficient determinace ( $R^2$ )
  - Podíl vysvětleného rozptylu
  - Je ukazatelem kvality, úspěšnosti regrese
  - Vyjadřuje shodu modelu s daty
- Pro jednoduchou lin. regr. platí  $R^2 = r^2$



AJ: regression and residual variance (sum of squares), explained variance, model fit with the data, coefficient of determination (R square)

# Lineární regrese III. – příklad z Hendla

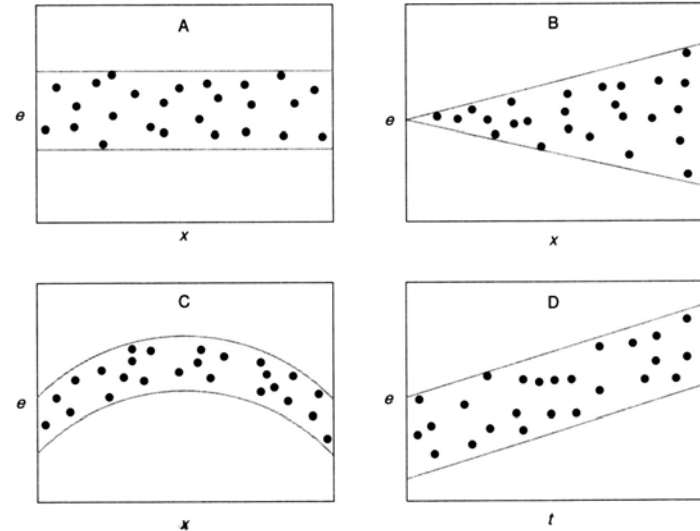


- Regrese váhy na výšku (s. 239 a 268)
  - $\text{váha (kg)} = 0,912 * \text{výška (cm)} - 93,24$
  - $r = 0,878 \quad R^2 = 0,77$

# Lineární regrese IV. – předpoklady, platnost

Předpoklady oprávněnosti použití lineárně-regresního modelu

- konceptuální předpoklad: vztah je ve skutečnosti lineární
- rezidua mají normální rozložení s průměrem 0
- homoskedascita
  - =rozptyl reziduí (chyb odhadu) se s rostoucím  $X$  nemění



- Platnost modelu je omezena daty, z nichž byl získán, a teorií.
  - Extrapolace, neoprávněná extrapolace (≈jako generalizace nad rámec empirických dat)
  - Pozor na odlehlé hodnoty – jako u všech ostatních momentových statistik

# Další druhy regrese

---

Zde je prezentovaná pouze jednoduchá lineární regrese, tj. s jednou závislou a jednou nezávislou proměnnou. Potřeb a možností je více.

- mnohočetná (mnohonásobná) lineární regrese
  - $Y = a + b_1X_1 + b_2X_2 + \dots + b_mX_m$
  - komplikují ji vztahy mezi nezávislými proměnnými
- logistická regrese
  - pokud je závislá dichotomie, nominální proměnná
  - predikuje se tak pravděpodobnost jednotlivých hodnot závislé
- Není-li vztah lineární, snažíme se transformovat proměnné tak, aby byl lineární.
  - nelineární regrese je velmi obtížná



# Hendl

---

kapitoly 7.3 – 7.3.2, 7.3.6  
pro absolventy metodologie i 7.4

---