

REGRESNÁ ANALÝZA

Paulína Jašková

Regresná analýza

- Využívame v situácií, keď nás zaujíma závislosť určitej **kvantitatívnej (spojitej) premennej** na jednej alebo viacero ďalších **kvantitatívnych premenných = regresor**
- nezávislá premenná = vysvetľujúca
- závislá premenná = vysvetľovaná
- CIEĽ:
 - Popísať závislosť medzi premennými pomocou vhodného modelu
- ROZDELENIE:
 - Jednoduchá regresia
 - Viacnásobná regresia
 - lineárna (regresná funkcia je lineárna)
 - nelineárna

Jednoduchá lineární regresia

- Popisuje závislost vysvetľovanej premennej na jednom regresore

- MODEL:

$$Y = \beta_0 + \beta_1 X + e$$

- β_0, β_1 ... regresné koeficienty
- Y ... závislá premenná
- X ... nezávislá premenná
- e ... vektor náhodných chýb modelu
- $Y = \{y_1, y_2, \dots, y_n\}^T$
- $X = \{x_1, x_2, \dots, x_n\}^T$
- n ... počet pozorovaní

Viacnásobná lineárna regresia

- Závislá premenná závisí na viac ako jednom regresore

- MODEL:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$

- β_0, \dots, β_n ... regresné koeficienty
- Y ... závislá premenná
- X_1, \dots, X_n ... nezávislá premenná
- e ... vektor náhodných chýb modelu
- n ... počet pozorovaní

Regresný model

- TYPY:
 - Lineárny
 - Exponenciálny
 - Logaritmický
 - Polynomiálny
 - Kízavé priemery

Data

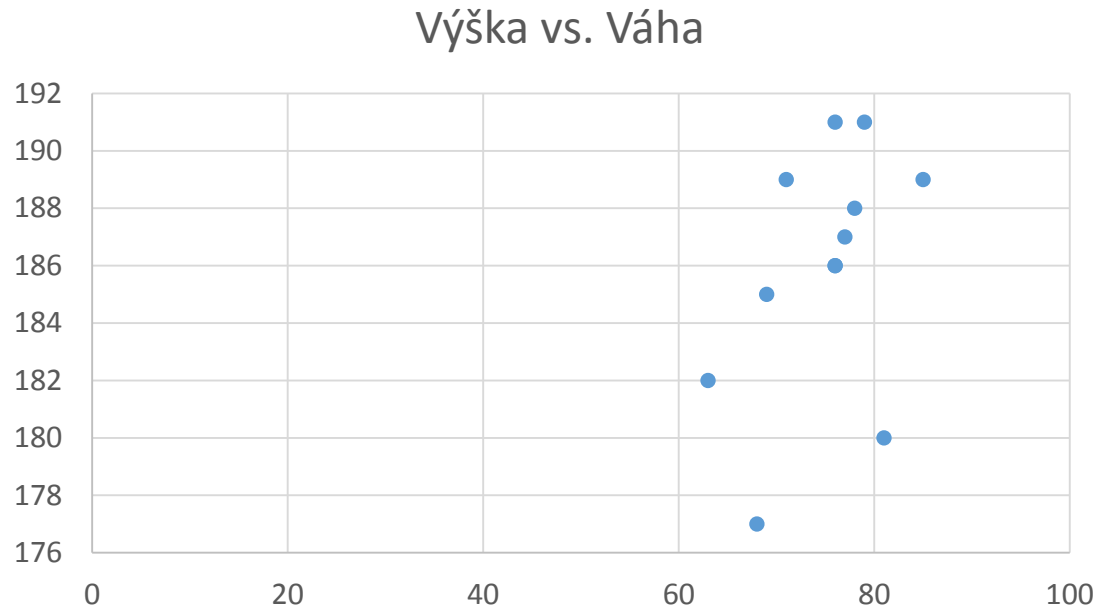
- Ex: Zisťovali sme výšku a hmotnosť náhodne vybraných športovcov. Údaje sú v tabuľke. Analyzujte závislosť výšky na váhe a zistite, ktorý model túto závislosť najlepšie popisujú ($\alpha = 0.05$).

Výška (v cm)	Hmotnosť (kg)
189	85
191	79
186	76
185	69
182	63
180	81
189	71
191	76
186	76
177	68
187	77
188	78

Ako začať s regresiou?

1. Bodový graf:

- predstava o vzťahu sledovaných spojitéch premenných
- vieme určiť rastovú tendenciu hmotnosti na výške človeka



2. Korelačný koeficient:

- kvantifikovaná miera závislosti
- Pearsonov korelačný koeficient
- =PEARSON(nezávislá premenná, závislá premenná)
- $r = \langle -1, 1 \rangle$
 - $|r| \leq 0,3$ malá závislosť
 - $0,3 < |r| \leq 0,8$ mierna závislosť
 - $|r| > 0,8$ silná závislosť

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

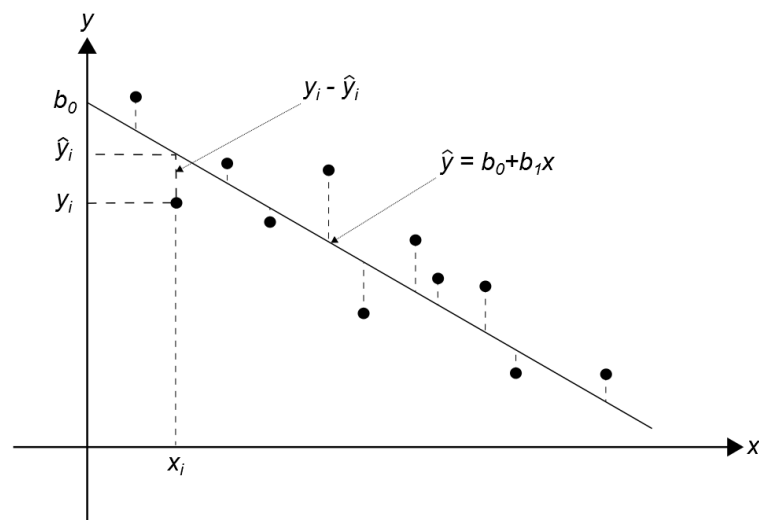
- $r=0,45$... Závislosť medzi výškou a váhou
 - Medzi výškou a váhou podľa Pearsnovho korelačného koeficientu je mierna závislosť

3. Typ vzťahu

- Z grafu odhadneme o aký typ závislosti pôjde
 - Lineárny model

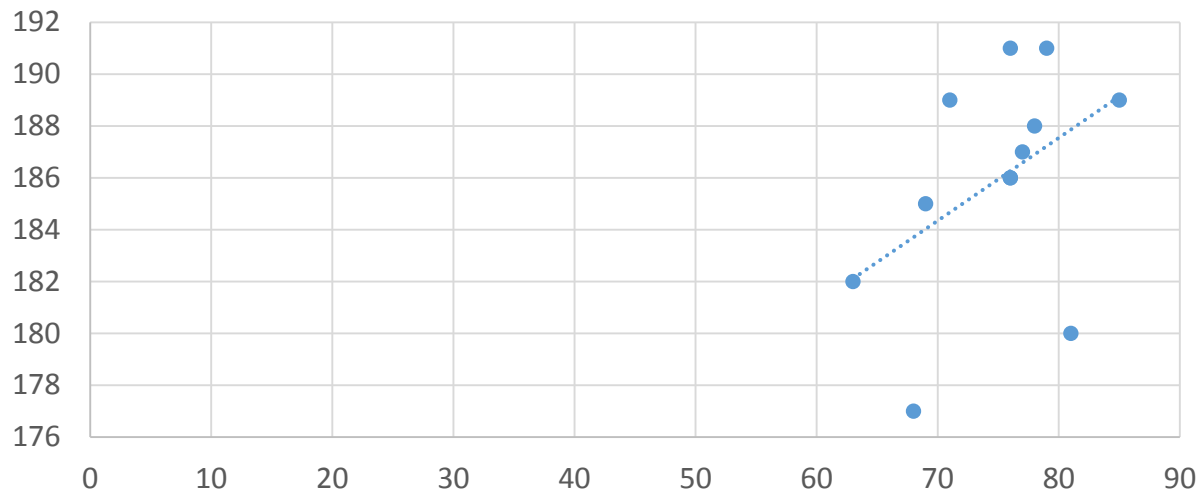
4. Odhad parametrov

- vzájomnú závislosť chceme vystihnúť pomocou priamky
- Ako najlepšie dáta priamkou preložiť?
 - ako určiť parametre β_0 a β_1
 - potrebujem mieru ktorá nám vypočíta ako moc je priamka blízko bodov
- **METÓDA NAJMENŠÍCH ŠTVORCOV**
 - Určí takú priamku, ktorá má najmenší súčet druhých mocnín z rozdielu predpovede (bod priamky príslušiaci hodnote X) a naozaj nameranej hodnote Y
 - =INTERCEPT(závislá,nezávislá) ... β_0
 - =SLOPE(závislá,nezávislá) ... β_1



- **5. Vykreslenie regresnej funkcie**
 - **do bodového grafu vložíme regresnú funkciu**

Závislosť



- **Lineárny regresný model: $Y = \beta_0 + \beta_1 X + e$**
- **Y – hmotnosť športovca (závislá premenná)**
- **X – výška športovca (nezávislá premenná)**

6. Interpretácia:

- regresných parametrov
- rastová tendencia
- miera tesnosti
- MODEL:

$$y = 161,98 + 0,3195 * x$$

- β_0 ... Hodnota pri $x=0$
- β_1 ... Nárast/pokles y pri jednotkovej x
- Interpretácia β_1 nemusí dávať vždy zmysel

Vhodnosť modelu

- Koeficient determinácie:
 - zn. $R^2 \in \langle 0,1 \rangle$
 - zhoda modelu s dátami
 - =RSQ(závislá, nezávislá)
 - $R^2 = 0,2037$ (nízky)
 - variabilita vysvetľovanej premennej (hmotnosť) je z 20% vysvetlená modelom s výškou športovca
- Upravený koeficient determinácie:
 - Adjusted R Squared
 - slúži k porovnávaní modelu

Záver

- Podarilo sa nám preukázať závislosť telesnej váhy na výške, pričom pri zmene váhy o 10 kg nám výška porastie o 3,2 cm
- Nízky podiel vysvetlenej variability naznačuje, že na výšku majú vplyv nejaké ďalšie, nami neuvažované, faktory (napr. vek)