Analýza rozptylu

- jednofaktorová,
- vícefaktorová.

Jednofaktorová analýza rozptylu

- Zkoumá, zda lze změny hodnot měřitelné proměnné *y* vysvětlovat faktorem *x*, který může být slovní i číselná proměnná.
- **Použití :** AR slouží k ověření významnosti rozdílu mezi výběrovými průměry většího počtu náhodných výběrů.

- **Předpoklady AR :**

  I. Náhodné výběry jsou nezávislé.
  II. Každý z výběrů má normální rozdělení s neznámou střední hodnotou $\mu_1, \mu_2, ..., \mu_k$ a s neznámým rozptylem $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$.
  III. Rozptyly všech skupin jsou stejné**,** tj. $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$.
  IV. Počet pozorování musí být větší než počet skupin, tj. $n > k$.
  V. Sledovaný statistický znak *y* je číselný, faktor *x* může být proměnná číselná i slovní.

- **Hlavní myšlenka AR :** Celkový rozptyl hodnot proměnné *y* (s_y^2) rozložíme na 2 části :

  1. *meziskupinový rozptyl* s_{ym}^2 *(rozptyl podmíněných průměrů)*
  - Odráží variabilitu mezi skupinami.
  - Jde o rozptyl, který se vztahuje k vlivu, podle něhož bylo provedeno třídění hodnot *y*.
  - Meziskupinová variabilita je vysvětlitelná daným faktorem *x*.

  2. *vnitroskupinový (reziduální) rozptyl* s_{yv}^2 *(průměr podmíněných rozptylů)*
  - Odráží variabilitu uvnitř skupin.
  - Kolísání je důsledkem závislosti *y* na jiných faktorech než na *x*.
  - Vnitroskupinová variabilita není vysvětlitelná daným faktorem.

- Platí tedy : $s_y^2 = s_{ym}^2 + s_{yv}^2$,

  kde $s_y^2 = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2}{n} = \dfrac{S_y}{n}$, kde

  y_{ij} .... jednotlivé hodnoty sledované proměnné,
  $\bar{y}$ .... celkový průměr,
  n .... rozsah výběru,
  S_y .... celkový součet čtverců (= součet čtverců odchylek hodnot od celkového průměru).

$$s_{ym}^2 = \frac{\sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2 n_i}{n} = \frac{S_{ym}}{n}$$, kde

$\bar{y}_i$ ...... podmíněný průměr,
S_{ym} .... meziskupinový součet čtverců (= součet čtverců odchylek hodnot podmíněných průměrů od celkového průměru).

$$s_{yv}^2 = \frac{\sum_{i=1}^{k} s_i^2 n_i}{n} = \frac{S_{yv}}{n}$$, kde

s_i^2 ..... podmíněný rozptyl,
S_{yv} .... vnitroskupinový (reziduální) součet čtverců.

$$s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i} \qquad \bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{n} \qquad \bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

- Jestliže platí : $s_y^2 = s_{ym}^2 + s_{yv}^2$, platí také $S_y = S_{ym} + S_{yv}$ (Za účelem zjednodušení lze tudíž používat pouze čitatele vzorců, tzv. součty čtverců.).

- **Test hypotézy o nezávislosti :**
  1) $H_0 : \mu_1 = \mu_2 = ... = \mu_k$
     (NEBO H$_0$: *y* nezávisí na *x*)
     $H_1 : non\ H_0$

  2) **Testovým kritériem** je statistika :

  $$F = \frac{\frac{S_{ym}}{k-1}}{\frac{S_{yv}}{n-k}}$$

  , která má při platnosti H$_0$ rozdělení F s $\nu_1 = k-1$ a $\nu_2 = n-k$ stupni volnosti.

  3) **Kritický obor:**
     $W \equiv \{F, F \geq F_{1-\alpha}(k-1, n-k)\}$.

- Měření síly závislosti proměnné *y* na faktoru *x*:
  Poměr determinace P² : $P^2 = \frac{S_{ym}}{S_y}$; $P^2 \in \langle 0,1 \rangle$

  Poměr korelace P: $P = \sqrt{P^2}$; $P \in \langle 0,1 \rangle$