ANALÝZA ZÁVISLOSTÍ – 1. ČÁST

- zkoumání závislosti dvou event. více proměnných, měření síly této závislosti, atd.
- cílem je hlubší vniknutí do podstaty sledovaných jevů a procesů, přiblížení k tzv. příčinným souvislostem.

Dvourozměrná tabulka rozdělení četností

- je elementární metodou popisu závislosti
- rozlišujeme různé typy tabulek.

Korelační tabulka: obě proměnné jsou numerické.

Kontingenční tabulka: alespoň jedna proměnná je slovní.

Asociační tabulka: obě proměnné jsou alternativní.

Čtyřpolní tabulka: obě proměnné nabývají pouze dvou obměn.

Dvourozměrná tabulka rozdělení četností

x_i	y_j				*Součty četností* $n_{i\bullet}$
	y_1	y_2	$\cdots$	y_s	
x_1	n_{11}	n_{12}	$\cdots$	n_{1s}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2s}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{r1}	n_{r2}	$\cdots$	n_{rs}	$n_{r\bullet}$
Součty četností $n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\cdots$	$n_{\bullet s}$	n

Symbolika:

n_{ij} … sdružené (simultánní) absolutní četnosti

$n_{i\bullet}$, $n_{\bullet j}$ … okrajové (marginální) absolutní četnosti

p_{ij} … sdružené relativní četnosti

$p_{i\bullet}$, $p_{\bullet j}$ … marginální relativní četnosti.

$$n_{i\bullet} = \sum_{j=1}^{s} n_{ij}\ ; \quad n_{\bullet j} = \sum_{i=1}^{r} n_{ij}\ ; \quad p_{i\bullet} = \frac{n_{i\bullet}}{n}\ ; \quad p_{\bullet j} = \frac{n_{\bullet j}}{n}\ ; \quad p_{ij} = \frac{n_{ij}}{n}$$

$$\sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{s} n_{\bullet j} = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n$$

Podmíněné rozdělení četností:

Rozdělení četností jedné proměnné, které odpovídá určité obměně druhé proměnné (tj. za podmínky, že druhá proměnná nabyla určité obměny).

Podmíněné relativní četnosti: $p_{j/i} = \frac{n_{ij}}{n_{i\bullet}}$; $p_{i/j} = \frac{n_{ij}}{n_{\bullet j}}$

Podmíněný průměr: $$\overline{y}_i = \frac{\sum_{j=1}^{s} y_j n_{ij}}{n_{i\bullet}} = \frac{\sum_{j=1}^{s} y_{ij}}{n_{i\bullet}}$$

Podmíněný rozptyl: $$s_{yi}^2 = \frac{\sum_{j=1}^{s} \left(y_j - \overline{y}_i\right)^2 n_{ij}}{n_{i\bullet}} = \frac{\sum_{j=1}^{s} \left(y_{ij} - \overline{y}_i\right)^2}{n_{i\bullet}}$$

Pro výpočty je často používána také jiná forma tabulky, která umožňuje třídění hodnot proměnné *y* podle proměnné *x* do *k* skupin.

Tabulka třídění proměnné y podle proměnné x

$i = 1, 2, \ldots, k$ x_i	$j = 1, 2, \ldots, n_i$ y_{ij}	n_i	$\overline{y}_i$	s_{yi}^2
x_1	y_{11}, y_{12}, y_{13}	n_1	$\overline{y}_1$	s_{y1}^2
x_2	$y_{21}, y_{22}, y_{23}, y_{24}, y_{25}$	n_2	$\overline{y}_2$	s_{y2}^2
. . .	. . .	. . .	. . .	. . .
x_k	$y_{k1}, y_{k2}, y_{k3}, y_{k4}$	n_k	$\overline{y}_k$	s_{yk}^2
Celkem		n	$\overline{y}$	s_y^2

Podmíněný průměr: $$\overline{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

Podmíněný rozptyl: $$s_{yi}^2 = \frac{\sum_{j=1}^{n_i} \left(y_{ij} - \overline{y}_i\right)^2}{n_i}$$

Grafické znázornění dvourozměrného rozdělení četností

- je další formou popisu závislosti
- lze použít různé typy grafů.

❖ ***čára podmíněných průměrů***

❖ ***čára podmíněných rozptylů***

❖ ***bodový graf (diagram)***.

Rozklad rozptylu

Princip: celkový rozptyl proměnné y (s_y^2) lze vyjádřit jako součet rozptylu podmíněných průměrů ($s_{y.m}^2$) a průměru podmíněných rozptylů ($s_{y.v}^2$).

Vzorec pro rozklad rozptylu:

$$s_y^2 = s_{y.m}^2 + s_{y.v}^2$$

Celkový rozptyl (s_y^2)

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

y_{ij} … jednotlivé hodnoty sledované proměnné
$\bar{y}$ … celkový průměr
n … rozsah výběru
S_y … celkový součet čtverců.

1. Rozptyl podmíněných průměrů ($s_{y.m}^2$)

- meziskupinový rozptyl
- odráží variabilitu mezi skupinami
- kolísání podmíněných průměrů je důsledkem závislosti y na x
- meziskupinová variabilita je vysvětlitelná faktorem x.

$$s_{y.m}^2 = \frac{\sum_{i=1}^{k}\left(\bar{y}_i - \bar{y}\right)^2 n_i}{n} = \frac{S_{y.m}}{n}$$

$\bar{y}_i$ … podmíněný průměr
$S_{y.m}$ … meziskupinový součet čtverců.

2. *Průměr podmíněných rozptylů* ($s_{y.v}^2$)

- vnitroskupinový rozptyl
- odráží variabilitu uvnitř skupin
- kolísání je důsledkem závislosti y na jiných faktorech než na x.

$$s_{y.v}^2 = \frac{\sum_{i=1}^{k} s_i^2 n_i}{n} = \frac{S_{y.v}}{n}$$

s_i^2 … podmíněný rozptyl

$S_{y.v}$ … vnitroskupinový (reziduální) součet čtverců.

Poznámka: jestliže platí $s_y^2 = s_{y.m}^2 + s_{y.v}^2$, pak také platí $S_y = S_{y.m} + S_{y.v}$. Za účelem zjednodušení vzorců lze tudíž používat pouze čitatele vzorců, tzv. součty čtverců.

Analýza rozptylu

- jednofaktorová analýza, faktorem je proměnná x (číselná nebo slovní)
- je to test, který zkoumá, zda změny hodnot numerické proměnné y lze vysvětlit změnami faktoru x
- slouží k ověření významnosti rozdílu výběrových průměrů více náhodných výběrů.

Předpoklady testu:

- ze základního souboru s normálním rozdělením $N(\mu;\sigma^2)$ je pořízeno k nezávislých náhodných výběrů
- každý z výběrů má normální rozdělení s neznámou střední hodnotou $\mu_1, \mu_2, ..., \mu_k$ a s neznámým rozptylem $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$
- rozptyly všech skupin jsou stejné, tj. $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ (tzv. homoskedasticita)
- shodu rozptylů je třeba ověřit vhodným testem, např. Bartlettovým
- počet pozorování musí být větší než počet skupin, tj. $n > k$.

Testovací postup:

1) $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ (tj. y nezávisí na x; rozdělení proměnné y mají na různých úrovních faktoru x stejné střední hodnoty)

$H_1 : non\ H_0$

2) *Testové kritérium:*

$$F = \frac{\frac{S_{y.m}}{k-1}}{\frac{S_{y.v}}{n-k}};$$

Statistika F má při platnosti H_0 rozdělení $F(k-1; n-k)$

3) ***Kritický obor:***

$W \equiv \{F;\, F \geq F_{1-\alpha}(k-1; n-k)\}$

4) Závěr testu:

Pokud leží hodnota testového kritéria v kritickém oboru, zamítáme H_0 a přijímáme H_1, tedy prokázali jsme hypotézu H_1 o závislosti proměnné *y* na faktoru *x*.

Měření síly (intenzity, těsnosti) závislosti proměnné *y* na faktoru *x*:

Poměr determinace: $P^2 = \dfrac{S_{y.m}}{S_y}$; $\quad P^2 \in \langle 0;1 \rangle$

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