

CHARAKTERISTIKA VIACROZMERNÝCH ŠTATISTICKÝCH SÚBOROV

VIACROZMERNÝ ŠTATISTICKÝ SÚBOR - pri výskume jeho prvkov je predmetom nášho záujmu vzťah viacerých štatistických znakov. Ak skúmame vzťah dvoch štatistických znakov, takýto súbor nazývame **dvojrozmerný**.

Hromadné javy sa nevyskytujú v realite izolovane, ale vždy v určitých vzťahoch s inými javmi. Jedným z cieľov štatistického výskumu je poznanie závislosti javov.

- ♦ Podľa stupňa závislosti jedného javu od druhého rozoznávajú sa dva základné druhy závislosti:
 - **PEVNÁ ZÁVISLOSŤ** – zmena jedného javu spôsobí zmenu javu druhého, a to v úplne presne zodpovedajúcej intenzite. Každéj hodnote jednej premennej veličiny zodpovedá vždy určitá hodnota druhej premennej veličiny. Priebeh tejto závislosti je možné presne charakterizovať určitou matematickou funkciou, preto sa označuje aj ako **FUNKČNÁ ZÁVISLOSŤ**. Vyskytuje sa väčšinou u prírodných javov (závislosť objemu telesa od teploty, závislosť bodu varu od atmosférického tlaku). Teplota bodu varu je **ZÁVISLÁ PREMENNÁ** a atmosférický tlak je **NEZÁVISLÁ PREMENNÁ**.
 - **VOLNÁ ZÁVISLOSŤ** – jeden jav podmieňuje druhý jav len s určitou pravdepodobnosťou a s rôznou intenzitou. Určitej hodnote jednej veličiny zodpovedajú rôzne hodnoty druhej veličiny. Pri tejto závislosti je možné charakterizovať len jej teoretický priebeh a jej pevnosť. Vyskytuje sa omnoho častejšie ako pevná závislosť a to najmä pri javoch spoločenských a biologických (jazda pod vplyvom alkoholu zvyšuje pravdepodobnosť dopravnej nehody, avšak vodič pod vplyvom alkoholu nemusí spôsobiť dopravnú nehodu vždy). Takáto závislosť je označovaná aj ako **STOCHASTICKÁ (PRAVDEPODOBNOŠTNÁ) ZÁVISLOSŤ**.

ZÁVISLOSŤ KVANTITATÍVNYCH ZNAKOV

KORELAČNÁ ZÁVISLOSŤ

- vyjadruje závislosť medzi kvantitatívnymi znakmi, t.j. vplyv zmeny nezávisle premenných na zmenu závisle premenných veličín

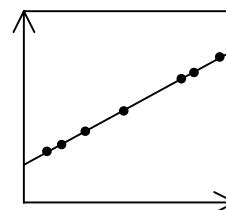
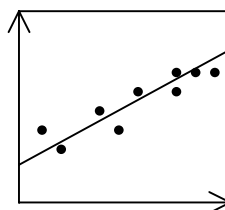
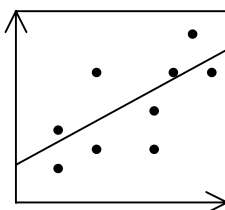
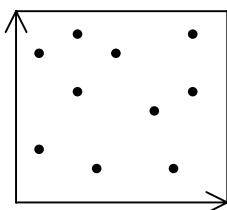
- ♦ pri skúmaní korelačnej závislosti charakterizujeme **regresiu a koreláciu**:

REGRESIA

- vo všeobecnosti charakterizuje priebeh závislosti a zmeny závisle premenného znaku na základe zmien nezávisle premenných znakov
- zvyčajne sa vyjadruje niektorou matematickou funkciou, z ktorej rovnice je možné robiť regresné odhady závislej premennej

KORELÁCIA

- charakterizuje tesnosť, resp. stupeň závislosti



Nižší stupeň

Vyšší stupeň

Štatistická nezávislosť

Volná závislosť

Pevná závislosť

Obr.: Schematické znázornenie rôznych stupňov závislosti

DRUHY KORELAČNEJ ZÁVISLOSTI

- korelačné závislosti sa členia z viacerých aspektov

- ♦ *podľa počtu kvantitatívnych znakov, závislosť ktorých skúmame*

korelačná závislosť jednoduchá (alebo **párová**) – skúma sa závislosť dvoch kvantitatívnych znakov (výskyt povodní – extrémne zrážky), teda uvažujeme iba s jednou nezávisle premennou

korelačná závislosť viacnásobná – skúma sa závislosť viac než dvoch kvantitatívnych znakov (výskyt povodní – extrémne zrážky, plocha areálov využitia zeme), teda výskyt nezávisle premenných je väčší ako jedna

♦ *podľa typu regresnej funkcie*

lineárna korelačná závislosť – zmeny jedného znaku sú zhruba lineárne závislé na zmenách druhého znaku. V prípade jednoduchej závislosti sa zobrazí pomocou regresnej priamky, v prípade viacnásobných závislostí pomocou roviny (pri dvoch nezávisle premenných), alebo nadroviny (pri viac ako dvoch nezávisle premenných).

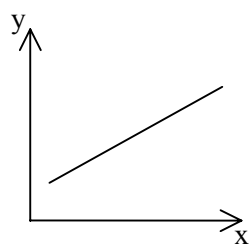
nelineárna korelačná závislosť – zmeny jedného a druhého znaku nie sú na sebe lineárne závislé. Závislosť je možné matematicky vyjadriť nelineárnou funkciou, ktorá sa v prípade jednoduchej závislosti graficky zobrazí ako regresná krivka (parabola, hyperbola, exponenciálna funkcia), alebo v prípade viacnásobnej závislosti prostredníctvom nelineárných funkcií rôznych typov a stupňov.

♦ *podľa smeru zmien kvantitatívnych znakov*

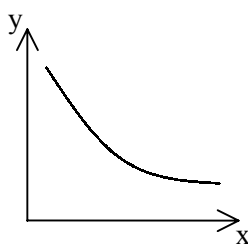
korelačná závislosť pozitívna (kladná, priama) – zmeny hodnôt skúmaných znakov prebiehajú v rovnakom smere; s rastom hodnôt jedného znaku narastajú aj hodnoty druhého znaku, a naopak s poklesom hodnôt jedného znaku klesajú aj hodnoty druhého znaku (veľkosť mesta – cena bytov)

korelačná závislosť negatívna (záporná, nepriama) – zmeny hodnôt skúmaných znakov prebiehajú v opačnom smere; s rastom hodnôt jedného znaku klesajú hodnoty druhého znaku (skeletnatosť pôdy – úrodnosť pôdy)

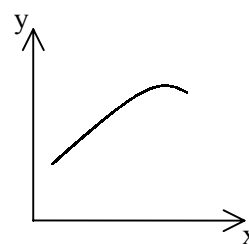
- jednoznačne charakterizovať smer závislosti je možné len pre lineárne závislosti. U nelineárných závislosti sa smer závislosti môže meniť



pozitívna lineárna



negatívna nelineárna



nelineárna so zmenou smeru

Obr. : Príklady jednoduchej korelačnej závislosti

♦ **KORELAČNÁ TABUĽKA**

Záhlavie tvoria hodnoty prvého znaku a legendu hodnoty druhého znaku. Do políčok sa píše početnosti štatistických jednotiek, ktoré nadobúdajú obidve hodnoty

Znak X \ Znak Y	y ₁	y ₂	...	y _l	Spolu „X“
x ₁	n ₁₁	n ₁₂	...	n _{1l}	n _{x1}
x ₂	n ₂₁	n ₂₂	...	n _{2l}	n _{x2}
⋮	⋮	⋮	⋮	⋮	⋮
x _k	n _{k1}	n _{k2}	...	n _{kl}	n _{xk}
Spolu „Y“	n _{y1}	n _{y2}	...	n _{yl}	n

Obr.: Korelačná tabuľka

ZÁVISLOSŤ KVALITATÍVNYCH ZNAKOV

- závislosť medzi kvalitatívnymi znakmi vyjadruje **asociácia a kontingencia**

ASOCIÁCIA

- závislosť medzi dvoma kvalitatívnymi znakmi alternatívnymi

◆ ASOCIAČNÁ TABUĽKA

- je základom pri výskume závislosti dvoch **kvalitatívnych alternatívnych** znakov

Znak A \ Znak B	b ₁	b ₂	Spolu „A“
a ₁	n ₁₁	n ₁₂	n _{a1}
a ₂	n ₂₁	n ₂₂	n _{a2}
Spolu „B“	n _{b1}	n _{b2}	n

Obr. : Schéma asociačnej tabuľky.

◆ DRUHY ASOCIÁCIE:

asociácia kladná – čím viac jednotiek má znak A, tým viac jednotiek má znak B

asociácia záporná – čím viac jednotiek má znak A, tým menej jednotiek má znak B

nezávislosť – znaky sú vzájomne nezávislé

Určuje sa to pomocou empirických kritérií, ktoré sa počítajú z relatívnych početností:

a) Prvé empirické kritérium:

$$\frac{n_{11}}{n_{a1}} > \frac{n_{21}}{n_{a2}} \text{ asociácia kladná} \quad \frac{n_{11}}{n_{a1}} < \frac{n_{21}}{n_{a2}} \text{ asociácia záporná} \quad \frac{n_{11}}{n_{a1}} \approx \frac{n_{21}}{n_{a2}} \text{ nezávislosť}$$

b) Druhé empirické kritérium:

$$\frac{n_{11}}{n_{a1}} > \frac{n_{b1}}{n} \text{ asociácia kladná} \quad \frac{n_{11}}{n_{a1}} < \frac{n_{b1}}{n} \text{ asociácia záporná} \quad \frac{n_{11}}{n_{a1}} \approx \frac{n_{b1}}{n} \text{ nezávislosť}$$

c) Tretie empirické kritérium:

$$\frac{n_{11}}{n} > \frac{n_{a1}}{n} \cdot \frac{n_{b1}}{n} \text{ asociácia kladná} \quad \frac{n_{11}}{n} < \frac{n_{a1}}{n} \cdot \frac{n_{b1}}{n} \text{ asociácia záporná} \quad \frac{n_{11}}{n} \approx \frac{n_{a1}}{n} \cdot \frac{n_{b1}}{n} \text{ nezávislosť}$$

V prípade nezávislosti približná rovnosť zahŕňa prípady, kedy sa hodnoty odlišujú do 5%.

◆ INTENZITA ASOCIÁCIE

- meria sa **koeficientom asociácie** alebo upraveným **koeficientom korelácie**, menej používanými koeficientmi sú Yuleov koeficient koligácie, Giniho koeficient podobnosti...

➤ KOEFICIENT ASOCIÁCIE

$$Q_{AB} = \frac{n_{11} \cdot n_{22} - n_{12} \cdot n_{21}}{n_{11} \cdot n_{22} + n_{12} \cdot n_{21}}$$

- Nadobúda hodnoty z intervalu (-1,1); pri nezávislosti sa blíži k 0; pri kladnej asociácii je kladný, pri zápornej je záporný; čím silnejšia je táto asociácia, tým viac sa jeho absolútna hodnota blíži k 1.

➤ KOEFICIENT KORELÁCIE KVALITATÍVNYCH ZNAKOV

$$R_{AB} = \frac{n \cdot n_{11} - n_{b1} \cdot n_{a1}}{\sqrt{n_{a1} \cdot n_{a2} \cdot n_{b1} \cdot n_{b2}}}$$

- nadobúda hodnoty z intervalu (-1,1) a interpretuje sa podobne ako koeficient asociácie

KONTINGENCIA

- závislosť medzi kvalitatívnymi znakmi množnými, teda tými ktoré nadobúdajú viacero obmien.

◆ **KONTINGENČNÁ TABUĽKA**

- je základom pri výskume závislosti dvoch **kvalitatívnych množných** znakov

Znak A \ Znak B	b₁	b₂	...	b_l	Spolu
a₁	n ₁₁	n ₁₂	...	n _{1l}	n _{a1}
a₂	n ₂₁	n ₂₂	...	n _{2l}	n _{a2}
⋮	⋮	⋮	⋮	⋮	⋮
a_k	n _{k1}	n _{k2}	...	n _{kl}	n _{ak}
Spolu	n _{b1}	n _{b2}	...	n _{bl}	n

Obr.: Schéma kontingenčnej tabuľky dvoch množných kvalitatívnych znakov, pričom znak *A* nadobúda obmeny a_i(i=1,2,...,k), znak *B* obmeny b_j(j=1,2,...,l) (napr. skúmame vzťah: vzdelanie – znalosť cudzieho jazyka)

◆ **INTENZITA KONTINGENCIE**

- meria sa pomocou **Pearsonovho koeficientu kontingencie**, **Čuprovovho koeficientu kontingencie**, ktorých základom je tzv. **štvorcová kontingencia**.

➤ **ČUPROVOV KOEFICIENT KONTINGENCIE:**

$$K = \frac{\chi^2}{n\sqrt{(k-1)(l-1)}}$$

n – rozsah súboru, *k* - počet obmien znaku *a*, *l* – počet obmien znaku *b* a výraz χ^2 (čítaj chí) –

ŠTVORCOVÁ KONTINGENCIA:

$$\chi^2 = n \cdot \left[\sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij})^2}{(n_{ai}) \cdot (n_{bj})} - 1 \right]$$

- koeficient nadobúda hodnotu 0 pri nezávislosti obidvoch znakov. Pri úplnej závislosti obidvoch znakov sa blíži k 1. Hodnotu 1 môže koeficient dosiahnuť pri úplnej závislosti obidvoch znakov len v prípade, že k=l. Ostatné hodnoty medzi týmito dvoma extrémami udávajú rôznu stupeň kontingencie.