



## KORELÁCIA A KONTINGENCIA

Diana KOMENDÁTOVÁ - Jana MÜLLEROVÁ

## CORRELATION AND CONTINGENCY



ENVIRONMENTAL POLICY TOOLS '2018

### **ABSTRAKT**

*Príspevok pojednáva o aplikácii dvoch štatistických metód na konkrétny prípad kriminality. Korelácia a kontingencia sú teoreticky popísané a v ďalšej časti aplikované na štatistický súbor kriminality rôzneho charakteru. Korelačný koeficient a Čuprov koeficient sú vypočítané pre dané štatistické dáta. V závere sú popísané výsledky aplikácie.*

### **Kľúčové slová:**

*Kontingencia, korelácia, závislosť, asociácia.*

### **ABSTRACT**

*The paper deals with the application of two statistical methods to real criminality data. Correlation and contingency are theoretically described and in the following part also applied to a particular criminality statistic data of various character. Correlation coefficient and Cuprov coefficient are calculated for the statistic data. At the end the results of the application are described.*

### **Key words:**

*Contingency, correlation, dependence, association.*

## **1 ÚVOD**

Korelácia ako aj kontingencia sú nástroje štatistiky na meranie závislostí dvoch štatistických znakov. V závislosti od charakteru tohto znaku je možné popísať rozdiel medzi koreláciou a kontingenciou. Ak sa jedná o dva kvantitatívne znaky, závislosť bude korelačná. Ak sa jedná o dva kvalitatívne znaky, závislosť bude kontingenčná. Treťou závislosťou, ktorá skúma vzájomné závislosti dvoch štatistických znakov je asociácia, pri ktorej sú tieto znaky alternatívne.

## **2 MERANIE ZÁVISLOSTI V ŠTATISTIKE**

Meranie závislostí v štatistike má hlavne za cieľ skúmať vzájomné závislosti štatistických znakov viacrozmerových súborov. Zaradujú sa sem závislosti pevné, voľné, jednostranné, obojstranné, príčinné, zdanlivé a iné.

Pevná, inak nazývaná ja funkčná alebo deterministická závislosť, je závislosť, pri ktorej zmena jedného javu vyvoláva zmenu druhého javu. Pri spájaní hodnôt vzniká presná spojité krivka o známej rovnici. Odchýlenie sa od tejto priamky je spôsobené chybou merania. Závislosť je možné rozdeliť na



základe druhu skúmaných štatistických znakov na:

- korelačnú závislosť
- asociačnú závislosť
- kontingenčnú závislosť

## 2.1 KORELAČNÁ ZÁVISLOSŤ

Korelačná závislosť skúma závislosť medzi kvantitatívnymi (číselnými) znakmi. Pri skúmaní korelačnej závislosti rozoznávame dva pojmy, a to koreláciu a regresiu.

**Korelácia** je analytický nástroj, ktorý je využívaný k testovaniu závislosti dvoch kvantitatívnych (číselných) znakov. Závislosť v tomto prípade znamená, že hodnoty jedného znaku sú priamoúmerné znaku druhému, čo pojednáva o kladnej korelácii, alebo ak sú hodnoty nepriamoúmerné druhému znaku, pojednáva to o zápornej korelácii. V matematickom pojmání to znamená, že hodnoty korelačného koeficientu sú v intervale  $\langle -1, 1 \rangle$ . Ak sa jedná o dve nezávisle premenné, je hodnota korelačného koeficientu sa približuje nule. Opakom je vysoká závislosť oboch znakov, čo sa prejaví v korelačnom koeficiente ako hodnota približujúca sa -1 alebo 1 podľa jednotlivej úmery. Korelačný koeficient sa vypočíta [1] zo vzťahu

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n [(x_i - \bar{x})]^2 \sum_{i=1}^n [(y_i - \bar{y})]^2}}$$

**Regresia** má za úlohu popísať priebeh skúmaného vzťahu štatistických znakov a využitie výsledkov pri prognózach. Cieľom je vyjadrenie priebehu závislosti, čo znamená sledovanie zmien závislej premennej na zmenách nezávislej premennej. Grafickým riešením je korelačný bodový graf.

## 2.2 ASOCIAČNÁ ZÁVISLOSŤ

Asociačná závislosť sa chápe ako závislosť medzi dvoma alternatívnymi znakmi. Pre výpočet

asociačného koeficientu platí vzťah[1]:  $Q_{ab} = \frac{n_{ab}n_{a\beta} - n_{a\alpha}n_{ab}}{n_{ab}n_{a\beta} + n_{a\alpha}n_{ab}}$ . Znaky a,  $\alpha$  označujú premenné prvého znaku, zapísané v asociačnej tabuľke v stĺpci a znaky b,  $\beta$  označujú premenné druhého znaku, zapísané v asociačnej tabuľke v riadku. Pričom rozmedzie intervalu závislosti je rovnaké, ako pri korelácii.

## 2.2 KONTINGENČNÁ ZÁVISLOSŤ

Kontingenčná závislosť je závislosť medzi dvoma kvalitatívnymi (slovnými) znakmi. Závislosť sa

vypočíta podľa Čuprovho koeficientu, ktorý vyzerá nasledovne[1]:  $K = \frac{\sqrt{\chi^2}}{\sqrt{n\sqrt{(k-1)(t-1)}}$ .

V uvedenom vzťahu sú premennými k a l stupne voľnosti, čo predstavuje počet obmien jednotlivých štatistických znakov. Pre možnosť výpočtu Čuprovho koeficientu je potrebné vypočítať z empirickej početnosti teoretickú, čo znamená vynásobenie obmeny  $x_i$  obmenou  $y_i$  a to celé je vydelené celkovým počtom. Hodnoty testovacej početnosti sa vypočítajú podľa vzťahu[1]

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{((a_i b_j) - (a_i b_j)_o)^2}{(a_i b_j)_o}, \text{ čo v jednoduchšom znení znamená}$$

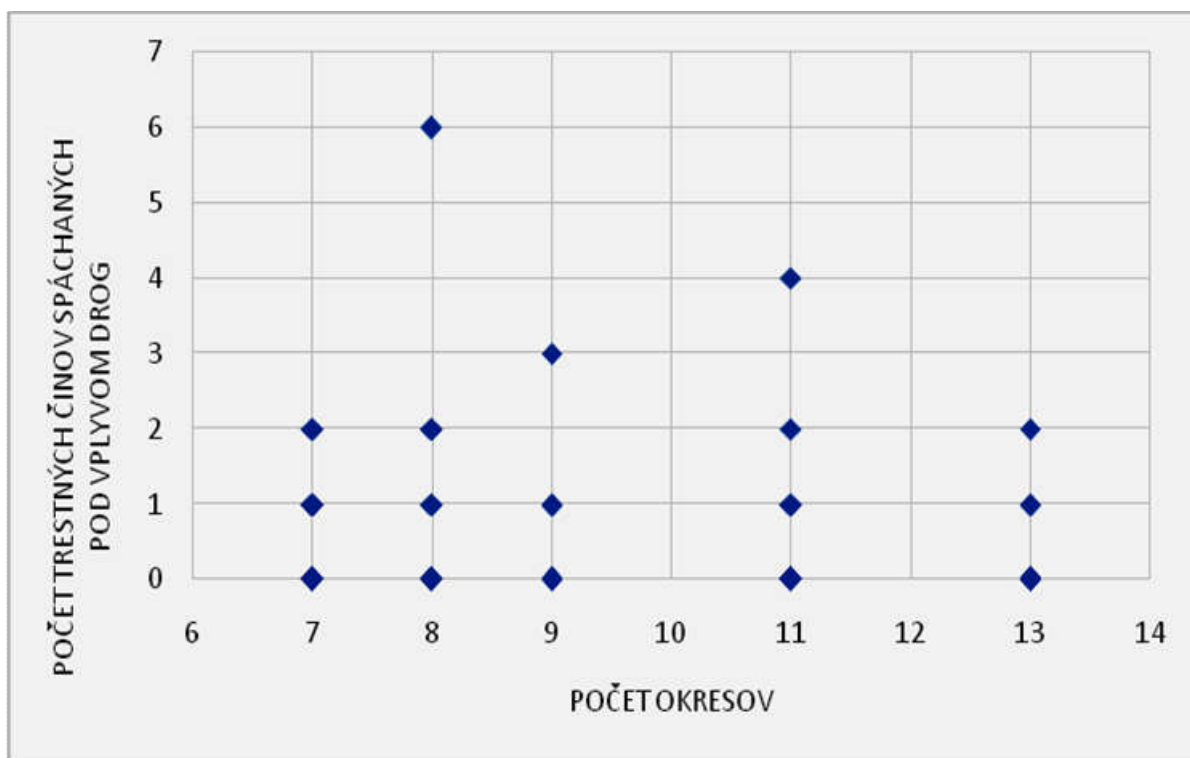


$$\chi^2 = \frac{(\text{empirická hodnota} - \text{teoretická hodnota})^2}{\text{teoretická hodnota}}$$

Intervalové rozdelenie závislosti je rovnaké ako pri predošlých dvoch závislostiach.

### 3 APLIKÁCIA KORELÁCIE

Na základe metódy korelácie boli využité štatistické znaky trestné činy spáchané pod vplyvom drog, ktorý je kvantitatívny, spojité znamenie, a počet okresov krajských riaditeľstiev policajného zboru, ktorý je kvantitatívny, nespojité znamenie. Pri riešení tejto metódy som si určila za nezávislú premennú počet okresov a za závislú premennú počet trestných činov pod vplyvom drog. Následne som určila korelačný koeficient a grafické zobrazenie danej závislosti. Na grafické vyjadrenie závislosti medzi znakmi som použila korelačný bodový graf.



Obrázok 1 Zobrazenie závislosti medzi počtom okresov a trestnými činmi spáchanými pod vplyvom drog

Tabuľka 1 Stupnica merania závislosti

Stupnica merania závislosti	
<0; (-)0,2)	bez závislosti
<(-)0,2; (-)0,4)	slabá závislosť
<(-)0,4; (-)0,6)	priemerná závislosť
<(-)0,6; (-)0,8)	silná závislosť
<(-)0,8; (-)1>	veľmi silná závislosť



#### 4 APLIKÁCIA KONTINGENCIE

Štatistické znaky v aplikácii metódy kontingencie boli použité údaje Krajských riaditeľstiev Policajného zboru, t.j. kvalitatívny, množný znak a škoda nad milión €, ktorý je kvalitatívny, alternatívny znak. Pre spracovanie údajov boli vytvorené 3 typy tabuliek, a to tabuľku empirickej početnosti (Tabuľka 2), tabuľku teoretickej početnosti (Tabuľka 3) a tabuľku testovacieho kritéria (Tabuľka 4)[2]

*Tabuľka 2 Empirická početnosť krajského riaditeľstva policajného zboru a škôd nad milión €*

Škoda nad milión €	Krajské riaditeľstvá policajného zboru								Spolu
	Banská Bystrica	Bratislava	Košice	Nitra	Prešov	Trenčín	Trnava	Žilina	
áno	5	12	5	8	3	8	11	8	60
nie	7	0	7	4	9	4	1	4	36
Spolu	12	12	12	12	12	12	12	12	96

*Tabuľka 3 Teoretická početnosť krajského riaditeľstva policajného zboru a škôd nad milión €*

Škoda nad milión €	Krajské riaditeľstvá policajného zboru								Spolu
	Banská Bystrica	Bratislava	Košice	Nitra	Prešov	Trenčín	Trnava	Žilina	
áno	8	8	8	8	8	8	8	8	60
nie	5	5	5	5	5	5	5	5	36
Spolu	12	12	12	12	12	12	12	12	96

*Tabuľka 4 Testovacie kritérium krajského riaditeľstva policajného zboru a škôd nad milión €*

Škoda nad milión €	Krajské riaditeľstvá policajného zboru								Spolu
	Banská Bystrica	Bratislava	Košice	Nitra	Prešov	Trenčín	Trnava	Žilina	
áno	1,13	2,00	1,13	0,00	3,13	0,00	1,13	0,00	8,50
nie	0,80	5,00	0,80	0,20	3,20	0,20	3,20	0,20	13,60
Spolu	1,93	7,00	1,93	0,20	6,33	0,20	4,33	0,20	22,10

#### ZÁVER

Podľa výpočtu korelačného koeficientu, ktorého hodnota je -0,22, je možné konštatovať, že závislosť medzi skúmanými znakmi je slabá, čo potvrdzuje aj stupnica merania závislosti (tabuľka 1). Nakoľko vyšla hodnota korelácie záporná, tak je možné tvrdiť, že rast hodnôt jedného znaku vyvoláva pokles hodnôt druhého znaku. Čo v danom prípade znamená, že s rastúcim počtom okresov krajských riaditeľstiev klesá hodnota trestnej činnosti spáchanej pod vplyvom drog. V jednoduchosti možno vysloviť hypotézu, že kriminalita by sa so stúpajúcim počtom krajských riaditeľstiev rozložila medzi tieto riaditeľstvá.

Z vypočítaných hodnôt testovacieho kritéria kontingencie sa dá odčítať hodnota chi-kvadrantu ( $\chi^2$ ), ktorá predstavuje hodnotu 22,10, na základe čoho je možné určiť štvorcovú kontingenciu, ktorej hodnota je 0,23. Daná hodnota je ďalej spracovaná v Čuprovom koeficiente ( $\tau^2$ ), ktorého hodnota je



0,78 pri stupňoch voľnosti 1 a 7, čo predstavuje silnú závislosť, čoho potvrdením je aj stupnica merania závislosti (tabuľka 1). Tabuľka empirickej početnosti krajského riaditeľstva policajného zboru a škôd nad milión € by sa dala využiť aj pri skúmaní pravdepodobnosti týchto dvoch znakov. Táto tabuľka by bola použiteľná aj pre určenie závislosti nezávislej premeny (znaku) - Krajské riaditeľstvá policajného zboru - podľa závislej premeny (znaku) - Škoda nad milión €.

## ZOZNAM BIBLIOGRAFICKÝCH ODKAZOV

- [1] HUŠEK, R. Základy ekonometrické analýzy II . Praha: VŠE, 1998
- [2] Ministerstvo vnútra SR: Štatistika kriminality v Slovenskej republike za rok 2018.  
[https://www.minv.sk/?statistika\\_kriminality\\_2018](https://www.minv.sk/?statistika_kriminality_2018)

## ADRESY AUTORIEK

### **Ing Diana Komendátová.,**

Žilinská univerzita v Žiline, Fakulta bezpečnostného inžinierstva, Ul. 1.mája 32, 010 08 Žilina, Slovenská republika.

E-mail: [diana.komendatova@fbi.uniza.sk](mailto:diana.komendatova@fbi.uniza.sk)

### **Prof. Ing Jana Müllerová., PhD.,**

Žilinská univerzita v Žiline, Fakulta bezpečnostného inžinierstva, Ul. 1.mája 32, 010 08 Žilina, Slovenská republika.

E-mail: [janka.mullerova.fsi@gmail.com](mailto:janka.mullerova.fsi@gmail.com)

### **RECENZIA TEXTOV V ZBORNÍKU**

*Recenzované dvomi recenzentmi, členmi vedeckej rady konferencie. Za textovú a jazykovú úpravu príspevku zodpovedajú autori.*

### **REVIEW TEXT IN THE CONFERENCE PROCEEDINGS**

*Contributions published in proceedings were reviewed by two members of scientific committee of the conference. For text editing and linguistic contribution corresponding authors.*