

KENDALLOVO TAU A JEHO VYUŽITIE V AKTUÁRSTVE

Michal Páleš, Ján Gogola

Abstrakt

Kendallovo τ (tau) je neparametrický štatistický ukazovateľ, ktorý meria silu a smer monotónnej závislosti medzi dvoma poradovými (ordinalnými) premennými. Je obzvlášť užitočný v prípadoch, keď sú dáta ovplyvnené odľahlými hodnotami alebo keď nie je splnený predpoklad normality, ktorý je potrebný pre Pearsonovu koreláciu. V aktuárskej vede, ktorá sa zaoberá analýzou a riadením rizika poisťovní, je schopnosť presne modelovať závislosti medzi rôznymi rizikovými faktormi kľúčová. Kendallovo τ poskytuje robustný nástroj na analýzu týchto závislostí.

Kľúčové slová

koeficient korelácie, miery závislosti, Kendallovo tau, jazyk R

1 ÚVOD

Modelovanie závislostí medzi faktormi opisujúcimi riziko (t. j. medzi príslušnými náhodnými premennými) je veľmi dôležitou oblasťou dnešného prístupu riadenia rizík, a to obzvlášť v rámci regulačných pravidiel typu Basel III a Solvency II. Základným nástrojom, ktorý sa dnes v tomto kontexte používa sú kopula funkcie, pozri napríklad [1].

Miery závislosti (*dependence measures*) zohrávajú v teórii rizika veľmi dôležitú úlohu, pretože napríklad v rámci tzv. ERM (*Enterprise Risk Management*) sa riziká často agregujú a ich prípadná závislosť práve do procesu agregácie významne zasahuje. V prípade ak uvažujeme, že riziká sú pozitívne korelované, potom riziko ich súčasného nastatia môže byť vysoké (napríklad nemocenské riziko a riziko úmrtnosti); v prípade ak riziká sú negatívne korelované, potom dochádza nie len k ich diverzifikácii, ale dokonca sa môžu vzájomne vykompenzovať (napríklad riziko úmrtnosti a riziko dlhovekosti). Miery závislosti sú dôležité nie len pre výpočet celkovej rizikovej expozície poisťovne, ale poisťovňa by k prípadnej závislosti mala tiež prihliadať v situáciách, keď stanovuje rozsah svojich aktivít v rôznych oblastiach akceptácie rizika.

Podľa [1] môžeme rozlišovať prinajmenšom tri typy mier závislosti:

- A. **lineárne korelácie,**
- B. **poradové korelácie,**
- C. **chvostové závislosti.**

Koreláciám typu A. (Pearsonovo ρ) sa venujú napríklad autori v [6], [3], [2], [1] a i. Tieto a ani chvostové závislosti (záležitosť kopúl) nebudú predmetom záujmu v týchto analýzach. My sa ďalej zameriame len na poradové korelácie s aplikáciami v jazyku R. Pre jazyk R pozri napríklad [4], [5].

Poradové korelácie sú miery závislosti, ktoré na rozdiel od lineárnej korelácie nezávisia na marginálnom rozdelení príslušných náhodných premenných, ale iba na ich kopule. Ich empirické odhady môžu byť skonštruované len na základe poradia pozorovaných hodnôt, čo tiež súvisí s ich názvom. Poradové korelácie sú jednoducho empiricky odhadnuteľné a preto sa v praxi používajú na kalibráciu kopúl. Príklady a teoretický rámec dvoch najobvyklejších poradových korelácií uvádza napríklad [1].

Kendallov koeficient poradovej korelácie τ možno chápať ako mieru založenú na výskyte konkordantných a diskordantných párov náhodných premenných X a Y , pričom dva páry pozorovaní (x, y) a (x^*, y^*) týchto náhodných premenných sú konkordantné, ak ich rozdiely $x - x^*$ a $y - y^*$ majú rovnaké znamienka, v opačnom prípade sú diskordantné. Ak (X^*, Y^*) je nejaký náhodný vektor s rovnakým rozdelením ako (X, Y) , ale nezávislý od (X, Y) , potom teoretický Kendallov koeficient poradovej korelácie τ má tvar:

$$\rho_\tau = \rho_\tau(X, Y) = P((X - X^*)(Y - Y^*) > 0) - P((X - X^*)(Y - Y^*) < 0)$$

tzn. v tomto prípade ako pravdepodobnosť konkordancie mínus pravdepodobnosť diskordancie. Kendallov koeficient je symetrická miera závislosti, pričom platí $\rho_\tau \in (-1; 1)$. [1] V odbornej literatúre možno nájsť aj iné tvary vyjadrenia Kendallovho τ .

2 PRÍKLADY ANALÝZ V JAZYKU R

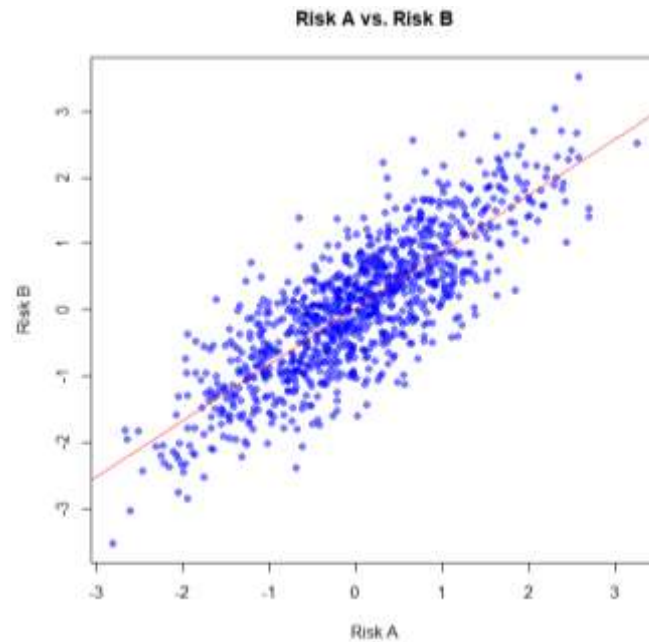
Ako príklad uvažujme dve náhodné premenné, ktoré sú medzi sebou korelované. Predstavme si, že tieto hodnoty reprezentujú riziko A a riziko B. Simulujeme dva vektory hodnôt, pričom druhý vektor bude závislý na prvom (lineárne závislý s koreláciou 0,8). Vypočítame Kendallove τ ($\tau = 0,5978058$) a keďže p -hodnota pre testovanie hypotézy, že medzi premennými nie je žiadna monotónna závislosť, je menšia ako hladina významnosti ($\alpha = 0,05$), môžeme odmietnuť nulovú hypotézu a potvrdiť, že medzi premennými existuje štatisticky významná závislosť. Na lepšiu vizualizáciu závislosti medzi dvoma rizikami môžeme vytvoriť rozptylový diagram (obrázok 1). Kód v jazyku R (podľa [7] s úpravami) je nasledovný:

```
set.seed(123)
n <- 1000
x <- rnorm(n) # Riziko A
y <- 0.8 * x + sqrt(1 - 0.8^2) * rnorm(n) # Riziko B, závislé na x

head(data.frame(x, y))
(kendall_test <- cor.test(x, y, method = "kendall"))

plot(x, y, main = "Risk A vs. Risk B",
     xlab = "Risk A", ylab = "Risk B",
     pch = 19, col = rgb(0, 0, 1, 0.5))
abline(lm(y ~ x), col = "red")
```

Pre pokročilejšie analýzy okrem základného rozhrania jazyka R možno využiť aj doplnkové knižnice ako napríklad `DescTools` alebo `rankr`.



Obr. 1, zdroj: [1]

Môžeme uviesť ešte ďalšie dva jednoduché príklady z oblasti aktuárstva, pričom využijeme nasimulované hodnoty v jazyku R (zdroj: [7]).

Závislosť medzi frekvenciou poistných udalostí a vekom poistenca

Predstavme si, že aktuár analyzuje databázu poistných zmlúv, kde má k dispozícii nasledovné premenné: vek poistenca a počet nahlásených poistných udalostí za posledné dva roky, pričom cieľom je zistiť, či medzi vekom a frekvenciou poistných udalostí existuje monotónna závislosť.

```
set.seed(123)
vek <- sample(18:80, 100, replace = TRUE)
frekvencia <- rpois(100, lambda = 0.03 * vek)
cor.test(vek, frekvencia, method = "kendall")
```

Potom, ak je $\tau > 0$, možno usudzovať, že so zvyšujúcim sa vekom sa zvyšuje aj frekvencia poistných udalostí a p -hodnota testuje nulovú hypotézu, že medzi premennými nie je žiadna monotónna závislosť.

Závislosť medzi výškou poistného a počtom nahlásených poistných udalostí

Aktuári často potrebujú identifikovať rizikové produkty, ktoré pri nízkom poistnom vykazujú vysokú škodovosť.

```
set.seed(123)
poistne <- runif(100, 100, 500)
udalosti <- rpois(100, lambda = 0.01 * poistne)
cor.test(poistne, udalosti, method = "kendall")
```

Tieto výsledky opäť môžu pomôcť aktuárovi pri preceňovaní produktov, analýze profitability, resp. pri rizikovom manažmente portfólia.

3 ZHRNUTIE A ZÁVER

Spearmanov koeficient poradovej korelácie ρ_s je možné opäť vyjadriť pomocou výskytu konkordancie a diskonkordancie medzi náhodnými premennými X a Y , pričom pre túto definíciu na rozdiel od ρ_r potrebujeme ešte tretí pár (X^{**}, Y^{**}) s identickým rozdelením ako (X, Y) , ale nezávislý od (X, Y) a (X^*, Y^*) . Potom, podľa [1]:

$$\rho_s = \rho_s(X, Y) = 3[P((X - X^*)(Y - Y^{**}) > 0) - P((X - X^*)(Y - Y^{**}) < 0)]$$

Spearmanov koeficient má rovnaké vlastnosti ako Kendallov koeficient. Výhodou je, že je menej citlivý na odľahlé hodnoty ako Pearsonov koeficient a zachytí aj nelineárne monotónne vzťahy.

Zatiaľ čo Pearsonov koeficient je výhodný pri dobre rozložených lineárnych dátach a v kombinácii s regresnými modelmi, Spearmanov a najmä Kendallov koeficient sú vhodnejšie pri nelineárnych monotónnych vzťahoch a robustnej analýze dát. V aktuárskej vede, kde dáta často nie sú normálne rozložené a môžu byť zaťažené odľahlými hodnotami, majú neparametrické prístupy významnú výhodu. Kendallovo τ je preto často preferované najmä pri modelovaní závislostí medzi rizikovými premennými alebo pri výpočtoch s kopulami. Porovnanie vlastností týchto koeficientov zobrazuje aj tabuľka 1 nižšie.

Kritérium/koeficient	Pearsonov	Spearmanov	Kendallov
Typ vzťahu	Lineárny	Monotónny	Monotónny
Odolnosť voči extrémom	Nízka	Stredná	Vysoká
Dátové predpoklady	Normalita, metrická škála	Poradie postačuje	Poradie postačuje
Výpočtová náročnosť	Nízka	Stredná	Vyššia
Príklady použitia	Lineárna regresia, PCA	Nelineárne modely, ordinálne dáta	Modely závislosti, kopuly

Tab. 1, zdroj: [7]

Kendallovo τ môže byť mimoriadne užitočný nástroj v „arzenáli“ aktuára. Vďaka svojej robustnosti a neparametrickej povahe je vhodný pre reálne dátové súbory s neštandardnými rozdeleniami, rovnakými či odľahlými hodnotami. Jeho aplikácia siaha od základnej analýzy závislostí až po zložité modely s kopulami, čím napomáha presnejšiemu oceňovaniu rizík a efektívnejšiemu spravovaniu poisťných portfólií.

ZDROJE

- [1] CIPRA, T. Riziko ve financích a pojišťovnictví: Basel III a Solvency II. Praha : Ekopress, 2015.
- [2] HINDLS, R. – ARLTOVÁ, M. – HRONOVÁ, S. – MALÁ, I. – MAREK, L. – PECÁKOVÁ, I. – ŘEZANKOVÁ, H. Statistika v ekonomii. Průhonice : Professional Publishing, 2018.
- [3] PACÁKOVÁ, V. a kol. Štatistické metódy pre ekonómov. Bratislava : Wolters Kluwer, 2009.
- [4] PÁLEŠ, M. Aktuárstvo. Bratislava : Letra Edu, 2021.
- [5] PÁLEŠ, M. Jazyk R pre aktuárov. Bratislava : Letra Edu, 2019.

- [6] ŠOLTÉS, E. Regresná a korelačná analýza s aplikáciami v softvéri SAS. Bratislava : Letra Edu, 2019.
- [7] OpenAI. ChatGPT. Verzia 4. Dostupné na: chat.openai.com [zobrazené: 2025-3-26; 2025-5-6].

PROJEKT

VEGA č. 1/0377/25 – Inovatívne metódy Enterprise risk managementu a ich využitie v riadení a modelovaní rizík; VEGA č. 1/0497/25 – Implementácia inovatívnych prístupov v oblasti riadenia a modelovania rizík v rámci interných modelov poisťovní

KONTAKTNÉ ÚDAJE

Páleš, Michal, doc. Ing., PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1, 852 35 Bratislava, e-mail: michal.pales@euba.sk

Gogola, Ján, RNDr., PhD., Katedra matematiky a aktuárstva, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave, Dolnozemska cesta 1, 852 35 Bratislava, e-mail: jan.gogola@euba.sk