

Téma 5 – Příklad 2

Zadání příkladu

Následující tabulka zobrazuje pořadí skupiny dívek ve dvou různých soutěžích krásy. Můžeme říci, že hodnocení obou porot jsou lineárně nezávislá? Posuďte charakter případné závislosti.

Dívka	Umístění u poroty A	Umístění u poroty B
Esmeralda	2.	6.
Kassandra	5.	5.
Rosaneta	6.	2.
Julie	7.	1.
Betty	8.	3.
Gertruda	4.	4.
Žofie	1.	8.
Manuela	3.	7.

Vypracování příkladu

Z obou proměnných spočítáme v tomto případě Spearmanův korelační koeficient a otestujeme jeho významnost v základním souboru. Pokud test neprokáže jeho významnost, předpokládáme, že neexistuje statisticky významný vztah mezi oběma pořadími. Korelační analýza zkoumá závislost jako oboustrannou.

Korelační koeficient spočítáme dle vzorce:

$$r_s = 1 - \frac{6 \sum_{i=1}^n (i_x - i_y)^2}{n(n^2 - 1)}$$

Výběrový korelační koeficient v tomto případě vychází -0,881. Nyní otestujeme jeho významnost v základním souboru.

1. $H_0: \rho_S = 0$
2. $H_1: \text{non } H_0$
3. Testové kritérium

$$t = \frac{r_s \cdot \sqrt{n-2}}{\sqrt{1-r_s^2}} = -4,561$$

$$4. \quad W \equiv \left\{ t; t \leq t_{\frac{\alpha}{2}}(n-2) \cup t \geq t_{1-\frac{\alpha}{2}}(n-2) \right\}$$
$$W \equiv (t; t \leq -2,365 \cup t \geq 2,365)$$

5. Testové kritérium je prvkem kritického oboru. Nulovou hypotézu zamítáme, přijímáme hypotézu alternativní.

Sílu a směr závislosti posoudíme pomocí vypočteného korelačního koeficientu. V našem případě jde o silnou nepřímou lineární závislost. Definiční obor Spearmanova korelačního koeficientu se pohybuje od -1 v případě funkční nepřímé závislosti až po 1 v případě funkční přímé závislosti.

Řešení v SGP

Zadání do programu není v tomto případě vůbec problematické. Stačí přepsat obě proměnné do samostatných sloupců.

Procedura v SGP: Describe – Numeric Data – Multiple-Variable Analysis (Rank Correlations)
alternativně: **Describe – Multivariate Methods - Multiple-Variable Analysis (Rank Correlations)**

Při vstupním dialogu zadáme do položky *Data* jednotlivé proměnné (nezáleží na pořadí, jelikož jde o oboustrannou závislost). V druhém kroku (*Analysis Options*) se nás program zeptá, zda chceme do analýzy zahrnout pouze kompletní dvojice dat, nebo všechna data (dojde k aproximaci chybějících údajů). V našem případě jsou obě možnosti ekvivalentní.

V okně *Rank Correlations* vidíme výsledek testu. První údaj sděluje hodnotu Spearmanova korelačního koeficientu, druhý údaj rozsah výběru a třetí údaj je hodnotou *P-Value*.

Stejně jako v ostatních případech používá Statgraphics k vyhodnocení testu hypotézy ukazatel *P-Value* (v jiných programech např. *Significance Level* apod.), což je maximální možná hodnota hladiny významnosti, na které ještě nezamítáme nulovou hypotézu. Není tudíž nutné určovat kritický obor pro námi zvolenou hladinu významnosti.

Hodnotu testové statistiky tato procedura nezobrazuje. V tomto případě tedy na 5% hladině významnosti zamítáme nulovou hypotézu o nezávislosti Spearmanova korelačního koeficientu v základním souboru.

Interpretace

Na základě korelační analýzy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná lineární závislost** hodnocení obou porot. Závislost je silná a nepřímá. To znamená, že dívky, které se líbí první porotě, nejsou tak oblíbené u poroty druhé a naopak.

Řešení v MS Excel

Pro spočítání Spearmanova korelačního koeficientu je možné použít funkci *CORREL* v případě, že upravíme proměnnou na pořadí. V některých verzích existuje funkce *SCORREL*, která počítá hodnotu Spearmanova korelačního koeficientu. Výslednou hodnotu je pak možné podrobit výše popsanému testu.

V závislosti na verzi programu je také možné použít doplňkový analytický nástroj a po kliknutí na *Correlation – Spearman* lze zadat proměnné a následně formulovat, zda se jedná o

jednostrannou či oboustrannou hypotézu. Po kliknutí na OK lze z výstupu vyčíst hodnotu P-Value.