

Téma 5 – Příklad 1

Zadání příkladu

Z korelační tabulky určete, zda existuje lineární závislost (korelace) proměnné y na proměnné x. U případné závislosti posuďte její sílu a směr.

	y _j	
x _i	6	9
2	2	-
5	4	1
9	-	3
11	1	2

Vypracování příkladu

Korelační tabulka zobrazuje dvě číselné proměnné. Z obou proměnných spočítáme výběrový Pearsonův korelační koeficient a otestujeme jeho významnost v základním souboru. Pokud test neprokáže jeho významnost, předpokládáme, že neexistuje statisticky významný vztah mezi oběma proměnnými. Korelační analýza zkoumá závislost jako oboustrannou.

Korelační koeficient spočítáme dle vzorce:

$$r_{yx} = r_{xy} = \sqrt{r_{yx}^2} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2) \cdot (\overline{y^2} - \bar{y}^2)}}$$

Výběrový korelační koeficient v tomto případě vychází 0,6313. Nyní otestujeme jeho významnost v základním souboru.

1. $H_0: \rho_{yx} = 0$

$H_1: \text{non } H_0$

2. Testové kritérium

$$t = \frac{r_{yx} \cdot \sqrt{n-2}}{\sqrt{1-r_{yx}^2}} = 6,395$$

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

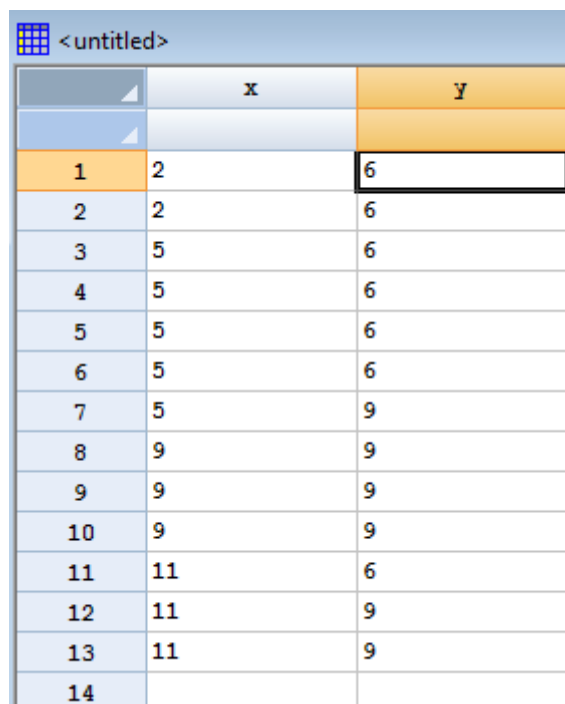
$W \equiv (t; t \leq -2,201 \cup t \geq 2,201)$

4. 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 středně silnou přímou lineární závislost. Definiční obor Pearsonova korelačního koeficientu se pohybuje od -1 v případě funkční nepřímé závislosti až po 1 v případě funkční přímé závislosti.

Řešení v SGP

Pro program Statgraphics je nejprve nutné vytvořit z korelační tabulky dva datové vektory odpovídající proměnným x a y . Pro jednu proměnnou lze využít operátoru REP a vytvořit vektor z okrajových absolutních četností, např. REP ($x_i; n_i$). Druhou proměnnou ovšem musím párově přiřadit tak, aby správně odpovídaly jednotlivé hodnoty x a y . Výsledné vektory pak vypadají takto.



	x	y
1	2	6
2	2	6
3	5	6
4	5	6
5	5	6
6	5	6
7	5	9
8	9	9
9	9	9
10	9	9
11	11	6
12	11	9
13	11	9
14		

Procedura v SGP: Describe – Numeric Data – Multiple-Variable Analysis (Correlations)

alternativně: Describe – Multivariate Methods - Multiple-Variable Analysis (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 *Correlations* vidíme výsledek testu. První údaj sděluje hodnotu Pearsonova 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.

	x	y
x		0,6313
		(13)
		0,0207
y	0,6313	
	(13)	
	0,0207	

Correlation
(Sample Size)
P-Value

Hodnotu testové statistiky tato procedura nezobrazuje, je ovšem možné použít například výstup z regresní analýzy. V tomto případě tedy na 5% hladině významnosti zamítáme nulovou hypotézu o nevýznamnosti korelačního koeficientu v populaci.

Interpretace

Na základě korelační analýzy můžeme říci, že na 5% hladině významnosti **existuje statisticky významná lineární závislost** proměnných x a y. Závislost je středně silná a přímá.

Řešení v MS Excel

Pro spočítání Pearsonova korelačního koeficientu je možné použít funkci CORREL. Výslednou hodnotu je pak možné podrobit výše popsanému testu.