

6. KORELACE

Korelace je statistická metoda, která ukazuje sílu souvislosti mezi dvěma proměnnými. To znamená, že pokud se změni jedna proměnná, změní se spolu s ní i druhá proměnná. Čím je hodnota tzv. korelačního koeficientu vyšší, tím je souvislost mezi oběma proměnnými silnější. Při změně jedné proměnné reaguje druhá proměnná změnou o to silněji, čím vyšší je korelační koeficient. Na druhou stranu síla korelace neříká nic o kauzalitě. Není nikde zaručeno, která proměnná způsobuje změny (je tak zvaně nezávisle nebo vysvětlující), a která těmto změnám podléhá (je takzvaně závisle nebo vysvětlovaná).

V sociálních vědách se obvykle posuzuje síla korelace následujícím způsobem [Rabušic, Mareš]:

0,00-0,09 – slabá či neexistující souvislost proměnných

0,10-0,29 – nízká až střední souvislost proměnných

0,30-0,49 – střední až podstatná souvislost proměnných

0,50-1,00 – podstatná až velmi silná souvislost proměnných

Při volbě vhodného korelačního koeficientu a tím i vhodného příkazu ve Statě se řídíme podle typu proměnných, které do korelace vstupují (pro připomenutí: nominální proměnné nelze seřadit, ordinální proměnné lze seřadit, ale nelze určit jejich vzdálenost či poměr, kardinální proměnné lze seřadit a současně lze určit jejich vzdálenost či poměr).

	Nominální	Ordinální	Kardinální
Nominální	Crammerovo V		
Ordinální	Crammerovo V	Spearmanovo rhó	
Kardinální	Crammerovo V	Spearmanovo rhó	Pearsonovo r

6.1 Jak vypočítat korelaci

Korelační koeficient vypíše i příkaz TAB, pokud použijeme parametr ALL. Kontingenční tabulky jsou ale vhodné jen pro výpis omezeného množství hodnot, tedy zejména pro krátké nominální a krátké ordinální proměnné (jednotky hodnot).

tab sex part, all

```
. tab sex part, all
```

H4. Pohlaví člena domácnosti	Má partnera/partnerku		Total
	Má partne	Nemá part	
Muž	2,439	537	2,976
Žena	2,979	1,124	4,103
Total	5,418	1,661	7,079


```

Pearson chi2(1) = 83.9736 Pr = 0.000
likelihood-ratio chi2(1) = 85.7184 Pr = 0.000
Cramér's V = 0.1089
gamma = 0.2630 ASE = 0.028
Kendall's tau-b = 0.1089 ASE = 0.011
    
```

Obrázek 1 Korelační koeficienty u kontingenční tabulky

Vhodnější je ale využít některý z následujících příkazů:

CORRELATE, PWCORR – vypočte korelační matici mezi všemi zadanými proměnnými.

```

correlate birthy age
correlate educ mstat [aweight=W_indi]
correlate educ mstat estat [aweight=W_indi]
pwcorr educ mstat estat [aweight=W_indi]
    
```

Rozdíl mezi oběma příkazy spočívá v tom, jak pracují s chybějícími hodnotami. Příkaz **PWCORR** počítá takzvané párové korelace. To znamená, že vezme vždy každý pár proměnných, vyřadí z něj chybějící hodnoty a spočítá

korelace. Každý pár proměnných má tak odlišný počet pozorování, který zjistíte zadáním parametru OBS. Oproti tomu příkaz CORRELATE nejprve vezme všechny proměnné, potom vyřadí všechny případy, v nichž má aspoň jedna proměnná chybějící hodnotu, a teprve poté spočte jednotlivé korelace. Všechny korelace tak mají stejný počet pozorování.

```
. corr educ mstat estat
(obs=10706)
```

	educ	mstat	estat
educ	1.0000		
mstat	-0.1018	1.0000	
estat	-0.1294	0.0084	1.0000

```
. pwcorr educ mstat estat, obs
```

	educ	mstat	estat
educ	1.0000 13047		
mstat	-0.1016 10710	1.0000 10712	
estat	-0.1369 10853	0.0086 10708	1.0000 10855

Obrázek 2 Ukázka výpočtu korelačních koeficientů

PCORR – vypočte parciální korelace (setkáte se i s názvem dílčí korelace) mezi první zadanou proměnnou a postupně všemi ostatními zadanými proměnnými, ale s tím, že každá uvedená korelace je očištěna od vlivu všech ostatních proměnných uvedených v seznamu.

Následující příkaz tak například vypočte postupně korelaci mezi proměnnou educ a mstat při očištění od vlivu proměnné estat, poté korelaci mezi proměnnou educ a estat při očištění od vlivu proměnné mstat.

pcorr educ mstat estat

```
. pcorr educ mstat estat
(obs=10706)
```

Partial and semipartial correlations of educ with

Variable	Partial Corr.	Semipartial Corr.	Partial Corr.^2	Semipartial Corr.^2	Significance Value
mstat	-0.1016	-0.1007	0.0103	0.0101	0.0000
estat	-0.1293	-0.1286	0.0167	0.0165	0.0000

Obrázek 3 Parciální korelace

SPEARMAN – vypočte korelační matici obsahující koeficienty Spearmanova rhó. Tento koeficient se používá pro ukázání souvislosti v případech, kdy je aspoň jedna proměnná ordinální (tedy ordinální × ordinální, nebo ordinální × kardinální).

spearman educ estat

KTAU – vypočte korelační matici obsahující koeficienty Kendaulova tau. Tento koeficient se používá pro ukázání souvislosti v případech, kdy je aspoň jedna proměnná ordinální (tedy ordinální × ordinální, nebo ordinální × kardinální).

ktau estat mstat