**6. Korelace**

Korelace je statistická metoda, která ukazuje sílu souvislosti mezi dvěma proměnnými. To znamená, že pokud se změní 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*

**

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.



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*

**

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*