

Procenta a procentní body

Kromě „normálních“ procent používáme ještě tzv. procentní body. Jaký je mezi procenty a procentními body rozdíl? Představte si, že prohlížeč Firefox před dvěma lety používalo 20 % lidí, ale tento rok ho používá již 30 % lidí. Jaký je nárůst mezi těmito dvěma lety? Chybou by bylo říci, že nárůst je desetiprocentní. Pokud Firefox používalo nejdříve 20 % lidí a poté 30 %, tak to znamená, že ho nyní používá o polovinu více lidí než předtím, tj. o 50 % více (základem, tj. 100 % byla pětina populace). Jenomže je trochu neobratné říkat, že nárůst je 50 %, když by pro všechny bylo srozumitelnější sdělení, že se počet procent zvedl o deset. A přesně k tomu máme procentní body. V tomto případě můžeme říci, že nárůst využívání prohlížeče Firefox činí deset procentních bodů. Procentní body tedy použijeme v případech, když chceme vyjádřit rozdíl mezi dvěma procentuálními údaji (procento je tedy vztaženo k absolutnímu základu, procentní body k relativnímu základu, tj. k procentům). Není překvapením, že běžně se procentní body používají v ekonomii (rozdíl v úrokových mírách, míře inflace), ale čím dál více pronikají i do ostatních věd, proto si na ně zvykněme. Dodejme, že zatímco procenta značíme známou značkou „%“, pro procentní body se buď užívá plného výrazu, nebo zkratky „p.b.“.

Rámeček 8.2

Literatura

- Linhart, Z., & Šafář, J. (1967). Programování třídění a statistických výpočtů pomocí samočinných počítačů. *Sociologický časopis*, 3(4), 435–443.
- Řehák, J., & Řeháková, B. (1978). Analýza kontingenčních tabulek: rozlišení dvou základních typů úloh a znaménkové schéma. *Sociologický časopis*, 14(6), 619–631.

Kapitola 9**Měření vztahů mezi dvěma proměnnými (korelační analýza)****9.1 Asociace a korelace**

V minulé kapitole jsme si ukázali, jak hledat asociaci mezi proměnnými v kontingenční tabulce na základě rozložení případů v jejich políčkách, a naučili jsme se také používat test chí-kvadrát pro potvrzení existence takové asociace v základním souboru. Řekli jsme si ale také, že chí-kvadrát nám sice dovolí zamítnout nulovou hypotézu o neexistenci asociace mezi sledovanými proměnnými, ale neposkytne nám žádnou informaci o síle této asociace. Vysvětlení, co asociace vlastně jsou a jaké jsou jejich charakteristiky, jsme v předchozí kapitole tak trochu odbyli, takže to nyní napravíme.

To, zdali je mezi dvěma proměnnými vztah (asociace, kontingence, korelace), je jedna ze základních otázek, kterou si při dvojrozměrné analýze dat klademe. Při hledání vztahu mezi dvěma proměnnými sledujeme, zda změny v jedné proměnné jsou doprovázeny i změnami v druhé proměnné.

Zajímá nás například, zdali existuje vztah mezi vzděláním a průměrným věkem v době prvního sňatku, zdali školní prospěch dětí souvisí s majetkovou úrovní jejich rodičů, zdali míra anomie souvisí s postojem k systému českého sociálního zabezpečení atd.

Hledání vztahu mezi proměnnými (hledání asociace) je z analytického hlediska operace, která má čtyři kroky. Při zkoumání párové asociace bychom si totiž, jak říkají Loether a McTavish (1988), měli klást čtyři následující otázky:

1. **Zdali asociace existuje, či nikoliv.**
2. **Jak je asociace silná (těsná)** – to je, jak silně rozložení variant jedné proměnné určují rozložení variant druhé proměnné.

3. **Jaký má asociace směr** – to je, zdali se jedná o asociaci kladnou, nebo zápornou.²⁰⁶

4. **Jakou má povahu** – zdali je monotónní (lineární) či jiného druhu.

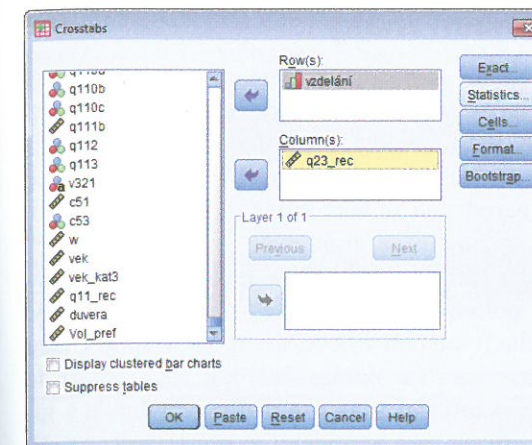
Pokud například zjistíme, že mezi vzděláním a průměrným věkem v době prvního sňatku existuje souvislost, zajímá nás, jak je tato souvislost silná/těsná, jaký má směr a zda tento věk se zvyšujícím se vzděláním (monotónně) roste, nebo klesá. Nebo zjistíme-li existenci asociace například mezi úhrnnou plodností a výší hrubého domácího produktu (GDP) na hlavu (jednotkami jsou v tomto případě jednotlivé státy světa), zajímá nás vedle síly této asociace a jejího směru (roste, nebo klesá úhrnná plodnost, když roste GDP?) také to, zda jde či nejde o lineární vztah. Blíže se k povaze lineárních vztahů a k jejich analýze dostaneme v kapitole 11, věnované lineární regresi. V této kapitole se budeme zabývat především otázkami měření síly/těsnosti asociací pomocí statistických koeficientů asociace. Pokud jste přeskočili první kapitoly této knihy, doporučujeme, abyste se k nim nyní raději vrátili. Konkrétní koeficienty asociace, o kterých bude řeč, byly totiž konstruovány pro jednotlivé úrovně měření (typy proměnných) a je velmi častou chybou nepoučených, že tento fakt ignorují. Použití nesprávného koeficientu je vlastně prohřešek stejného druhu, jako kdybychom tvrdošjně počítali aritmetické průměry u nominálních proměnných.

Jak jsme již řekli výše, pro zjištění síly či těsnosti vztahu počítáme tzv. koeficienty asociace (nebo korelace).²⁰⁷ **Koeficient asociace** je číslo, které zpravidla nabývá hodnot v intervalu od 0 do 1. Hodnota blízká nule indikuje nezávislost (neboli neexistenci vztahu), čím více se blíží jedné, tím silnější souvislost mezi proměnnými existuje. Hledáme-li souvislost mezi ordinálními či kardinálními znaky, míra korelace se zpravidla bude pohybovat v intervalu od -1 do $+1$.²⁰⁸ I zde platí, že čím blíže je hodnota $+1$ nebo -1 , tím silnější je mezi proměnnými vztah. Pro měření síly vztahu se používá řada nejrůznějších měř asociace a korelace. V SPSS je získáme následnou procedurou: *Analyze – Descriptive Statistics – Crosstabs – Statistics* – volba příslušných koeficientů (viz obr. 9.1a, 9.1b, datový soubor „EVS99-cvicny“).

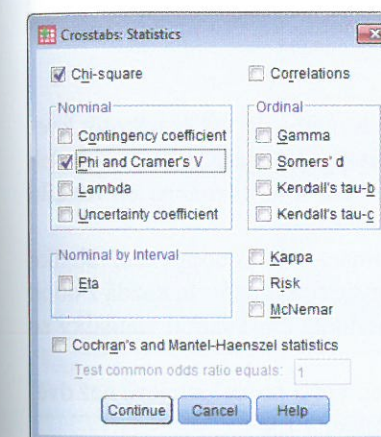
²⁰⁶ Dodejme, že toto ale můžeme dělat pouze u proměnných ordinálních a spojitých, u nominálních proměnných nemá smysl o pozitivní či negativní asociaci hovořit.

²⁰⁷ Existuje úzus, že při měření síly souvislosti mezi nominálními znaky hovoříme o asociaci nebo kontingenci, při měření síly souvislosti mezi ordinálními a kardinálními znaky hovoříme o korelaci. I příslušné dvourozměrné tabulky nazýváme kontingenční či korelační dle povahy proměnných.

²⁰⁸ Znaménko plus nebo minus u korelačních koeficientů logicky znamená pozitivní nebo negativní směr závislosti.



Obr. 9.1a Postup pro výpočet koeficientů asociace a korelace



Obr. 9.1b

Jaký druh koeficientu zaškrtnutím příslušného okénka zvolíme, závisí na několika okolnostech, z nichž ta nejdůležitější je povaha proměnných (tj. nominální, dichotomická, ordinální, kardinální), jejichž vztah hledáme. Ostatně i této logice odpovídá dialog v SPSS, který je členěn na jednotlivé části dle povahy proměnných.