

Typy proměnných

MUDr. Tomáš Novák, MUDr. Miloslav Kopeček, Ph.D.

Psychiatrické centrum, 3. LF UK v Praze

Data mohou nabývat různé formy. Je důležité vědět, v jaké formě se získávaná data vyskytují, abychom mohli zvolit nejvhodnější statistické metody pro jejich zpracování. Jedno z možných dělení rozlišuje data spojitá, diskrétní, ordinální a nominální.

Klíčová slova: spojitá data, diskrétní data, ordinální data, nominální data.

Types of variables

Data may take many different forms. It is important to know what forms every variable takes before we can make a decision regarding the most appropriate statistical methods to use. One possible division distinguishes continuous, discrete, ordinal and nominal data.

Key words: continuous data, discrete data, ordinal data, nominal data.

Psychiat. pro Praxi 2010; 11(4): 176–177

I v druhé části našeho seriálu o statistice se věnujeme popisné statistice, která je prvním krokem před tzv. **induktivní statistikou (testováním hypotéz)**.

Od pacientů, kteří souhlasí s účastí v některém z výzkumných projektů, které probíhají na našem pracovišti, získáváme řadu dat (proměnných; pozn.¹), u kterých očekáváme, že by nám mohla pomoci odpovědět na naše výzkumné otázky (hypotézy). Jde o různorodá data, která vyžadují odlišné „zacházení“.

Příkladem sbíraných dat jsou pohlaví, věk, rodinný stav, body mass index (BMI), délka současné depresivní epizody, počet předchozích depresivních epizod, úroveň psychopatologie hodnocená posuzovacími stupnicemi, hladina cholesterolu apod.

U pohlaví mohou nastat jen dvě možnosti (muž – žena), v případě rodinného stavu je možností několik (svobodný/á, ženatý/vdaná, rozvedený/á, vdovec/vdova), hodnoty věku či délky deprese se mohou pohybovat v určitém rozpětí a vyjadřuje se je v celých jednotkách času (obvykle roky, resp. měsíce), hodnoty hladin cholesterolu v závislosti na přesnosti měření lze uvádět s různým počtem čísel za desetinnou čárkou (5,1 vs. 5,14), BMI lze vyjádřit jako více či méně přesnou hodnotu (např. 31,4) nebo ji rozdělit do několika kategorií (normální váha, nadváha, obezita).

Pracujeme tedy s řadou různých proměnných, které můžeme podle toho, jakých mohou nabývat hodnot (odborněji dle použitého měřítka), rozdělit do několika skupin. Nejde jen o samoučelné dělení, typ proměnné totiž rozho-

duje, jakou statistickou metodu můžeme použít při jejich zpracování. Následující dělení je jen jedno z možných a stačí srovnat několik učebnic statistiky, abychom došli ke zjištění, že plná shoda neexistuje (např. 1, 2).

Kontinuální (spojité) proměnné mohou nabývat (teoreticky) všech hodnot v daném intervalu a jejich konkrétní hodnota závisí jen na přesnosti měření. Sem patří tělesná výška, hladina cholesterolu či BMI (byť jej získáme až ze dvou kontinuálních proměnných – výšky a váhy). Další vlastností kontinuální proměnné je možnost určit rozdíly mezi dvěma měřeními (osoba A je o 7 cm vyšší než osoba B nebo že má 1,5krát vyšší hladinu cholesterolu). Tento typ dat obvykle vykazuje normální rozložení a pro popis souboru lze použít průměr a směrodatnou odchylku (skupina A má průměrnou hladinu cholesterolu 4,98 se směrodatnou odchylkou 1,56).

Diskrétní proměnné nabývají pouze celočíselných hodnot a mohou být považovány za variantu dat spojitých. Příkladem je věk uvedený v letech, délka depresivní epizody v týdnech, počet dnů hospitalizace či počet sebevražedných pokusů v anamnéze. Věk by samozřejmě šlo vyjádřit i v minutách, ale není to příliš praktické. S hodnotami diskrétních proměnných lze pracovat podobně jako s kontinuálními (lze je porovnávat), pro popis souboru však je třeba nejprve zjistit, zda jsou data normálně rozložena. K popisu souboru se v závislosti na rozložení dat používá průměr a směrodatná odchylka (normální rozložení) nebo medián a mezikvartilové rozpětí (není normální rozložení) (3).

Ordinální proměnné jsou patrně nejméně povědomé a často se zaměňují s předchozími typy. Hlavní vlastností ordinální proměnné je, že její hodnoty mohou být dle určité charakteristiky pouze seřazeny do pořadí (např. od nej-

lepšího k nejhoršímu), ale nelze je přímo porovnávat („o kolik“, „kolikrát“). Příkladem tohoto typu proměnných je závažnost deprese (mírná, středně těžká, těžká), stupeň rezistence na léčbu (0–5 stupňů), dosažené vzdělání (základní, středoškolské, vysokoškolské) či kategorie dle BMI (podváha, norma, nadváha, obezita). Podobně sem patří závažnost psychopatologie dle posuzovacích škál.

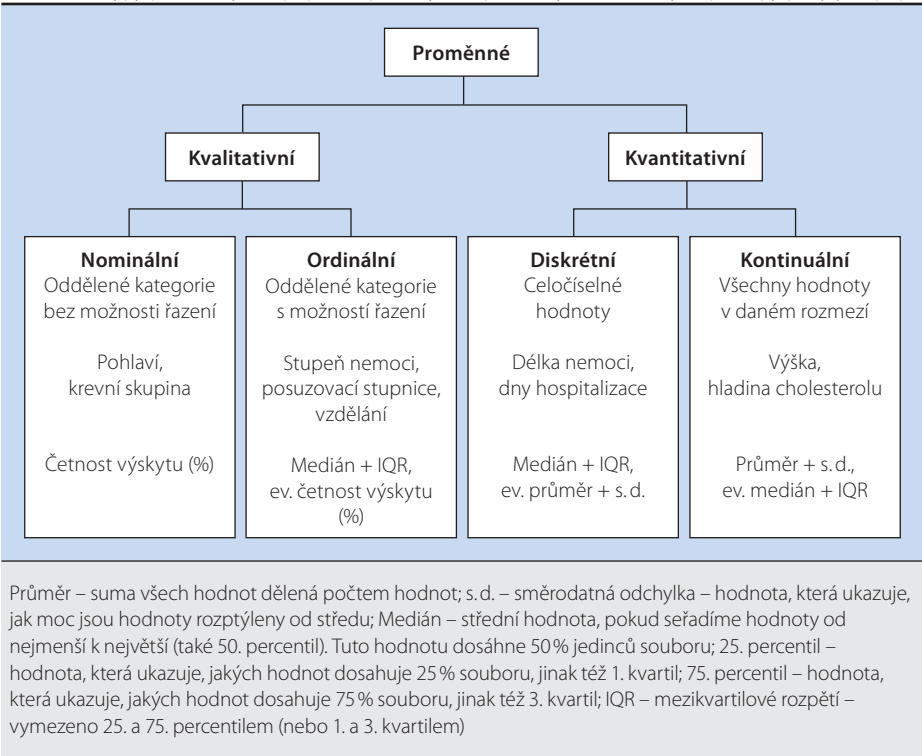
Nejzřetelněji to lze ukázat na stupnici celkového klinického dojmu (CGI, sedmibodová stupnice). Pokud je stav osoby A dle této stupnice hodnocen 3 body „mírné příznaky duševní nemoci“ a stav osoby B 6 body „silně vyjádřené známky nemoci“, můžeme pouze říct, u koho je stav závažnější, ale konstatovat, že je stav osoby B o 3 body či 2krát závažnější než u osoby A, nelze. U stupnic s větším rozptylem možností (v MADRS lze teoreticky dosáhnout 0–60 bodů) není situace tolik výrazná a snadno svádí s hodnotami zacházet, jako by šlo o kontinuální proměnné. I zde však platí, že jde o ordinální (pořadovou) proměnnou.

Ordinální data by se měla vyjadřovat pomocí pořadové statistiky (medián, mezikvartilové rozpětí), v situaci malého počtu možností pak jako absolutní či relativní (%) četnost výskytu (základní vzdělání 21 %, střední vzdělání 59 % a vysokoškolské vzdělání 20 %).

Nominální (kategoriální) proměnné jsou takové, kde můžeme u zjištěné hodnoty určit, do které kategorie patří, aniž bychom je mohli seřadit či srovnávat (pokud opustíme předsudky). Tyto proměnné mohou nabývat dvou (např. pohlaví, odpověď na léčbu, pacienti s komorbidní/bez komorbidní poruchy osobnosti) nebo více možností (etnikum, krevní skupina, typ léčebné intervence). Tato data lze obvykle prezentovat jako absolutní či relativní četnost výskytu v souboru (55 % žen a 45 % mužů). Některá

1. Pojmy proměnné a data používáme v textu promiskue, proměnnými (znaky) jsou nicméně myšleny určité charakteristiky subjektů v našem souboru (pohlaví, BMI) a daty jsou míněny konkrétní, aktuální hodnoty proměnné u daného subjektu (žena, 25,6).

Schéma 1. Typy proměnných s popisem, příklady a doporučenými statistickými postupy pro jejich popis



nominální data jsou jasně daná (pohlaví, krevní skupina), některá vznikají našim zásahem uměle, kdy v původně kontinuální, spojitě či ordinální proměnné stanovíme hranice, které vymezí nové kategorie.

Příklad: Změnu psychopatologie lze hodnotit jako bodový rozdíl či procentuální změnu dle posuzovací stupnice nebo v kategoriích – remise, odpověď, částečná odpověď, bez odpovědi. Tato změna dat s sebou nese určitou ztrátu původní informace

a vyžaduje jiné statistické postupy. Proto by mělo mít převádění dat na jiný typ jasné zdůvodnění a je vhodné uvádět výsledky jak v původních datech (změna psychopatologie v bodech), tak v kategoriích (procento respondérů).
Kontinuální a diskrétní proměnné se souhrnně nazývají **kvantitativní (metrické) proměnné** a nominální a ordinální se obvykle společně uvádějí jako **kvalitativní proměnné** (schéma 1).

Podporováno MZOPCP2005 (MZ ČR).

Literatura

1. Hendl J. Přehled statistických metod zpracování dat. Praha: Portál 2004: 584.
2. Petrie A, Sabin C. Medical Statistics at a Glance. 2nd ed. Oxford: Blackwell Publishing 2005: 157.
3. Kopeček M, Novák T. Nebojte se statistiky aneb Jak interpretovat výsledky studií. Psychiat. pro Praxi 2010; 11(2): 83–84.

MUDr. Tomáš Novák
Psychiatrické centrum, 3. LF UK
Ústavní 91, 181 03 Praha 8
novak@pcp.lf3.cuni.cz