

Nebojte se statistiky aneb Jak interpretovat výsledky studií

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

Psychiatrické centrum Praha, Klinika psychiatrie a lékařské psychologie 3. LF UK Praha

Článek je malým shrnutím deskriptivní statistiky.

Klíčová slova: deskriptivní statistika, výuka, přehled.

Do not be afraid of statistics. How to interpret study results.

This article contains mini-review about descriptive statistics.

Key words: descriptive statistics, education, review.

Psychiat. pro Praxi 2010; 11(2): 83–84

Milí čtenáři, přicházíme s novou rubrikou, která si klade za cíl srozumitelnou formou vysvětlit některé statistické pojmy a postupy, které jsou v odborné literatuře často používány, avšak jimž je věnována poměrně malá pozornost v průběhu studia lékařství. Jsme toho názoru, že mnohé statistické postupy a pojmy se dají vysvětlit i bez složitého matematického aparátu, který mnoho psychiatrů děsí. Znalost statistiky je však nezbytná, neboť jak dle Marka Twaina pravil bývalý britský předseda vlády a spisovatel Benjamin Disraeli či dle encyklopedie Wikipedie britský liberální kritik a redaktor Charles Wentworth Dilke. Jsou tři druhy lží. Obyčejné lži, proklaté lži a statistika. Winston Churchill prý věřil ze zásady jen té statistice, kterou si sám mohl zfalšovat. Abychom tolik nepodléhali zkreslenému dojmů, je dobré znát alespoň něco ze základů statistiky.

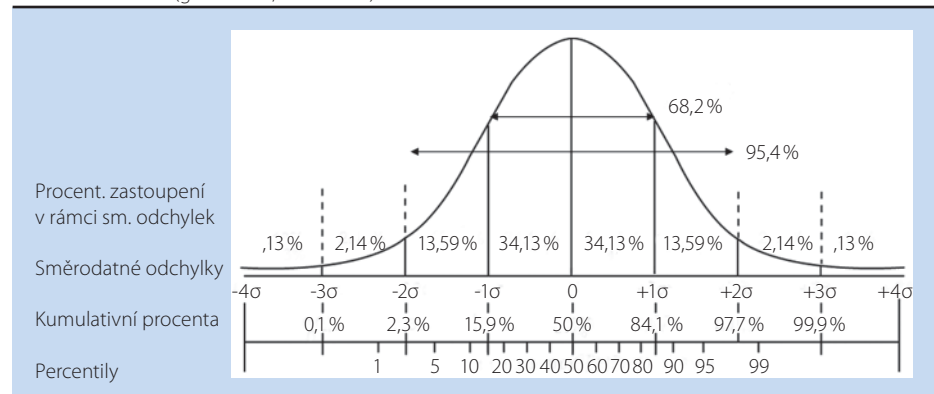
Ve statistické metodice se používají metody popisné (deskriptivní) a metody induktivní. Abychom hned na úvod nezastrašili a neodradili relativní začátečníky, začneme metodami deskriptivní statistiky. Postupy deskriptivní statistiky umožňují získat přehlednější a úspornější uspořádaný popis nasbíraných dat. Zde se používají pojmy průměr (mean), medián (median) a modus (mode) (viz slovníček). Pro lepší charakteristiku dat je nerozlučným bratrem průměru směrodatná odchylka (s. o.)/standard deviation (s. d.), pro medián pak bratři – 25. percentil (1. kvartil) a 75. percentil (3. kvartil), jež ohraničují tzv. mezikvartilové rozpětí (interquartile range). Při takto prezentovaných popisných datech můžeme mít přibližnou představu o distribuci dat, se kterou studie počítá. Průměr a s. o. lze použít tam, kde jsou vstupní data symetricky uspořádaná – mluví se často o tzv. normální (gausovské) distribuci (graf 1). Pokud ve článku objevíte údaj, že odečtení 2násobku s. o. od průměru nabývá záporných hodnot v případě, že jsou záporné hodnoty pro daný parametr ne-

smyslné (např. dny hospitalizace), snadno zjistíte, že byl nesprávně použit průměr.

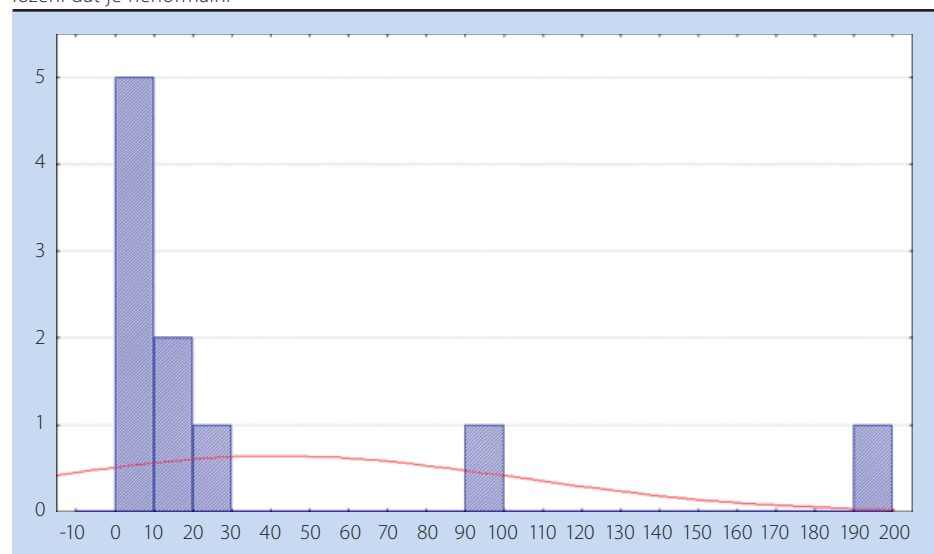
Příklad: Máme-li soubor o 10 jedincích s hodnotami (10, 10, 10, 10, 10, 20, 20, 30, 100, 200), bude průměr $42 \pm 61,9$. Zde po odečtení 2násobku (v tomto případě i 1násobku) s. o. od průměru dostáváme záporné číslo a průměr se s. o. je tak použit nesprávně. Správně použijeme v takovém případě medián, který nabývá

hodnoty 15 (polovina souboru je mezi hodnotou 10 a 20), 25. percentil je 10 (zde hodnota na 3. pozici), 75. percentil je 30 (zde hodnota na 8. pozici) či modus = 10. Kdyby uvedené hodnoty znamenaly distribuci příjmu na pracovišti, počet hodin pracovní neschopnosti či délku hospitalizace, snadno dovodíme, jak by pouhé udání průměru zkreslovalo data. Z grafického záznamu distribuce dat (graf 2) je vidět, že se nejedná

Graf 1. Normální (gausovská, zvonovitá) distribuce dat



Graf 2. Rozložení dat v našem příkladu a srovnání s normální distribucí (červená čára) ukazuje, že rozložení dat je nenormální



o normální distribuci (porovnej s obrázkem 1), a proto není správné používat průměr a s.o.

To, jaké použijeme parametry, určuje nejenom distribuce dat (normální vs. nenormální), ale také typ dat, který používáme. Průměr můžeme použít u dat kontinuálních (např. výška, váha), medián u dat ordinálních (např. posuzovací stupnice), modus u dat nominálních (např. barva očí, etnikum). Podrobněji o typu dat (proměnných) si přečtete v dalším čísle časopisu.

Pokud jsou data normálně rozložená, tak v rozsahu ± 1 s.o. je 68,2% souboru, v rozsahu 2 s.o. 95,4% a v rozsahu 3 s.o. 99,7% souboru (graf 1). Norma odvozovaná od statistického průměru říká, že normální hodnoty jsou 2 s.o. nalevo i napravo od průměru. Tím se např. řídí stanovené IQ, kde je průměr 100 bodů a s.o. 15 bodů.

Slovníček

- Průměr – suma všech hodnot dělená počtem hodnot
- Směrodatná odchylka – hodnota, která ukazuje, jak moc jsou hodnoty rozptýleny od středu

- Medián – střední hodnota, pokud seřadíme hodnoty od nejmenší k největší (také 50. percentil)
- 25. percentil – hodnota, která ukazuje, jakých hodnot dosahuje 25 % souboru, jinak též 1. kvartil
- 75. percentil – hodnota, která ukazuje, jakých hodnot dosahuje 75 % souboru, jinak též 3. kvartil
- Modus – nejčastěji se vyskytující hodnota
- Mezikvartilové rozpětí – vymezeno 25. a 75. percentilem (nebo 1. a 3. kvartilem)

Pamatuj

Je-li odečet 2násobku s.o. od průměru záporný v parametrech, které záporné být nemohou, distribuce dat není normální. Při nenormální distribuci dat použij medián, 25 % a 75 %.

Ukázky statistické negramotnosti v běžném životě

V médiích se zásadně dozvíte o průměrné mzdě (údaje o směrodatné odchylce či mediá-

nu chybí, takže nelze říci, jak je takový údaj zavadějící).

Výsledky známkování ve škole jsou hodnoceny průměrem, přičemž správnější je použití mediánu. Známkování jsou ordinální data.

V Severní Karolíně (USA) jsou praktické aplikace mediánu, modusu a průměru probírány na základní škole ve 4. ročníku a hodnocení žáků se odvíjí od národních percentilových norem.

Podporováno MZOPCP2005 (MZ ČR)

Literatura

1. Harris M, Taylor G. Medical Statistics Made it Easy, London, Martin Dunitz 2004.
2. Dušek L. Analýza dat v neurologii. Česká a slovenská neurologie a neurochirurgie 2007; 70(103): 99–100.

MUDr. Miloslav Kopeček, Ph.D.

Psychiatrické centrum Praha
Ústavní 91, 181 03 Praha 8-Bohnice
kopecek@pcp.lf3.cuni.cz
