UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH

FILOZOFICKÁ FAKULTA

Katedra sociálnej práce

Základy štatistiky v sociálnych vedách I.

Teoretické východiská a deskriptívna analýza

Vladimír LICHNER

Košice 2020

Základy štatistiky v sociálnych vedách I.

Teoretické východiská a deskriptívna analýzy

Vysokoškolská učebnica

Mgr. Vladimír Lichner, PhD.

Univerzita Pavla Jozefa Šafárika v Košiciach, Filozofická fakulta, Katedra sociálnej práce

Recenzenti:

doc. Ing. Mgr. Jozef Bavoľár, PhD.

Univerzita Pavla Jozefa Šafárika v Košiciach, Filozofická fakulta, Katedra psychológie

Mgr. Ing. Zuzana Poklembová, PhD.

Prešovská univerzita v Prešove, Filozofická fakulta, Inštitút edukológie a sociálnej práce

Za odbornú stránku tejto publikácie zodpovedajú autori. Rukopis prešiel redakčnou a jazykovou úpravou.

Umiestnenie: www.unibook.upjs.sk

Dostupné od: 9.11.2020

ISBN 978-80-8152-925-2 (e-publikácia)

Obsah

Úvod

Napriek tomu, že sa to na prvý pohľad javiť nemusí, štatistika, alebo jej oblasti každodenne obklopujú každého z nás. Ak sa zamyslíme nad tým, že každodenná realita je do značnej miery ovplyvnená procesom zberu, vyhodnocovania a reprodukovania rôznych informácií, údajov, či skúseností, vystihneme realitu, ktorú nám ponúka práve štatistika.

Zjednodušene povedané, štatistika je nástroj, ktorý nás učí spôsobom, ako prevádzať číselné údaje (číselné pozorovania) na informácie, ktoré umožňujú utvoriť si presný obraz o javoch, z ktorých boli tieto údaje získané a vyvodzovať správne a efektívne postupy a rozhodnutia.

V každej oblasti štúdia, ale aj v neakademickom prostredí sa stretávame s používaním štatistickej praxe. Médiá nám na pravidelnej báze ponúkajú rôzne prieskumy verejnej mienky, kde prezentujú často percentuálny prehľad rôznych názorov. Rovnako ak počúvame o príčinách, či dôsledkoch javov, ktoré predkladajú rôzne štúdie (napr. že spôsob rodičovskej kontroly ovplyvňuje zapojenie sa do rizikových, či nerizikových prejavov správania detí a pod.), hovoríme o skutočnostiach založených na štatistickom dokazovaní. Rovnako je štatistika dôležitou súčasťou i v štatistickej praxi študentov rôznych odborov. V sociálnych a behaviorálnych vedách sa stretávame s uchopením štatistiky v rôznych smeroch. Jednak je to pre pochopenie empirických dôkazov o existencii, resp. neexistencii rôznych sociálnych javov či ich príčin alebo dôsledkov. Na druhej strane, je štatistika dôležitou súčasťou vlastných empirických výstupov študentov v podobe tvorby a spracovania záverečných prác.

Predložená vysokoškolská učebnica predstavuje základný prehľad štatistickej teórie a praxe určený pre študentov a študentky sociálnej práce vrátane príkladov ich použitia v konkrétnych štúdiách. Tento text sa zaoberá prvou časťou štatistickej praxe, a to deskriptívnou analýzou, ktorú vnímame ako základnú podmienku spracovania akéhokoľvek empirického výskumu a je dôležitou i pri uchopení typických štatistických princípov a pojmov.

Pri spracovaní textu tejto vysokoškolskej učebnice sme vychádzali z relevantnej literatúry doplnenej o konkrétne príklady, ktorých vypracovanie bolo navrhnuté i s ohľadom na dostupný matematický a štatistický software v rozsahu, ktorý je pre študentov sociálnej práce relevantný. Jednotlivé časti sú doplnené o praktické príklady výpočtov v najpoužívanejších štatistických a matematických programov (Excel a SPSS). Uvedomujeme si, že na výpočet jednotlivých konštruktov je možné použiť i iné spôsoby, preto sme tieto štatistické postupy uviedli iba ako tipy pre čitateľov.

I. Teoretické východiská štatistiky

1 Štatistika – predmet a jeho chápanie

Definične je štatistika vnímaná ako veda, ktorá sa zaoberá zberom, komparáciou a systematickou klasifikáciou kvantitatívnych údajov, najmä ako základ pre štatistické usudzovanie a indukciu (Donnelly 2007).

Štatistika nás informuje o tom, akým spôsobom môžeme získavať relevantné informácie z pôvodne numerických informácií – dát, pričom:

- je určujúcou pri príprave a realizácii výskumu a procesu vyhodnotenia výsledkov,
- poskytuje nástroje pre spracovanie a vyhodnotenie výsledkov tak, aby bolo možné problému porozumieť (Hendl 2012).

V súčasnom chápaní štatistikou označujeme najčastejšie:

1. **kvantitatívny údaj** o hromadných javoch v podobe konkrétnych numerických údajov, ktoré poskytujú najmä prehľadové informácie (napr. štatistiky o počte (ne)zamestnaných, poskytovateľov sociálnych služieb, klientoch, vo všeobecnejšej rovine o počte obyvateľov, novonarodených,...);
2. **štatistickú charakteristiku** v podobe údajov vypočítaných na základe príslušných vzorcov (napr. údaj o priemernom veku klientov,...);
3. **praktickú činnosť** orientovanú na proces zberu, grafickým a textovým spracovaním, vyhodnocovaním a zobrazovaním štatistických údajov;
4. **vednú disciplínu** zaoberajúcu sa hromadnými javmi. V tomto kontexte je súhrnom metód určených na zhromažďovanie, triedenie, analýzu, vysvetlenie a komparáciu hromadných javov s cieľom vyvodzovať všeobecne platné závery (Sodomová 2013; Kanderová, Úradníček 2007; Ferjenčík 2006).

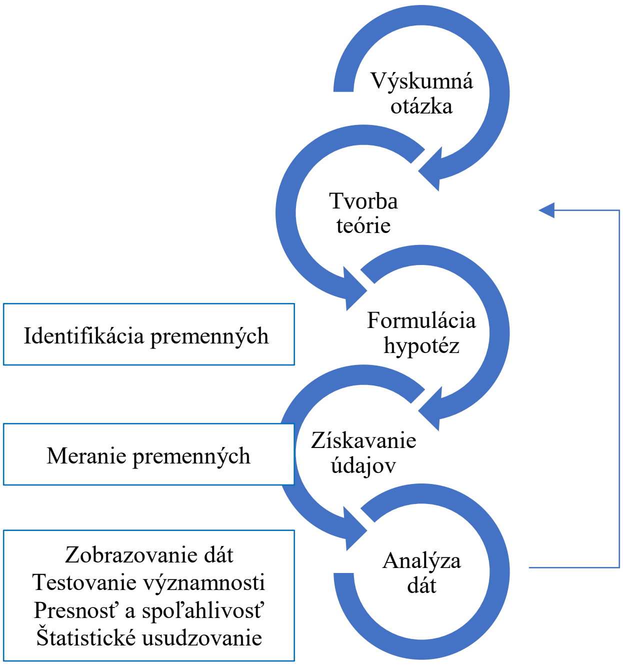
Štatistika je označovaná aj pojmom „veda vied". Tento názov vysvetľuje integrujúci princíp, ktorým je štatistika charakteristická – štatistika totiž ponúka spoločné metodologické nástroje, napomáha pomenovávať a riešiť problémy vo všetkých oblastiach vedeckého poznania (Rimarčík 2007). Opäť zjednodušene, napriek odlišnostiam rôznych vedných disciplín je štatistika nástrojom, ktorý umožňuje získavať a pomenovávať vedecké dôkazy v každej z nich. Je nástrojom komunikácie vedeckých poznatkov naprieč disciplínami. V tomto kontexte je možné o štatistike hovoriť ako o efektívnom **nástroji** pre vedecké poznanie, konkrétne pre dokázanie platnosti vedeckých predikcií, názorov, či teórií.

Keďže štatistika je chápaná aj ako veda, je nutné formulovať **vedeckú metódu**, ktorú štatistika používa. Tá spočíva v troch krokoch:

1. formulácia hypotézy (na základe teórie),
2. určenie dôsledkov (teórie alebo hypotézy),
3. komparácia dôsledkov s faktami získanými prostredníctvom výskumných zistení.

Ak sa v tejto fáze získajú potvrdzujúce informácie, teória (hypotéza) môže byť potvrdená, ak však fakty odporujú teórii, tá by mala byť modifikovaná a predpokladané dôsledky prehodnotené (Rimarčík 2007). Štatistika je tak nástrojom pre napredovanie vedeckého bádania.

Úlohy a postup, ktorý využíva štatistika je možné pozorovať i na základe grafického vyjadrenia štádií výskumného procesu. Konkrétne štatistické úlohy v jej jednotlivých štádiách zobrazujú obdĺžniky v ľavej časti schémy.



Obr. 1 - štádiá výskumného procesu

Zdroj: Field (2009), preklad a úprava: autor

V súvislosti s uvedeným, Cobb a Moore (1997) rozlišujú **štatistickú prax** v troch úrovniach:

1. ***získavanie údajov***, ktoré zahŕňa postupy pre zber dát, návrhy metodologických postupov, výberov, výskumných nástrojov a pod. V oblasti sociálnych a behaviorálnych vied túto úlohu zaobstaráva **metodológia**;
2. ***exploračná analýza údajov***, kedy hovoríme o prvom stupni štatistického spracovania. Ide o základný popis, triedenie, číselné a grafické vyhodnocovanie, prezentovanie a vysvetlenie zozbieraných údajov. Táto prax štatistiky je veľmi dôležitá, je tiež

označovaná termínom **deskriptívna** (popisná) štatistika, čo indikuje jej základné využitie. Rovnako má do značnej miery exploračnú povahu, ide o dôsledné vysvetlenie dát, ktoré je v podstate základom štatistickej praxe;

3. ***štatistické usudzovanie***, ktoré okrem popisu skúma i povahu získaných dát. Často hovoríme, že ide „za dáta" a snaží sa údaje popísať v širšom kontexte. Konkrétne sem patria spôsoby vyvodenia napr. v deskriptívnej štatistike nepozorovateľných súvislostí, rozdielov a pod. Okrem takéhoto vysvetlenia skúma aj spoľahlivosť získaných záverov prostredníctvom pravdepodobnosti. Takúto prax označujeme aj pomenovaním **induktívna štatistika,** alebo tiež inferenčná štatistika, ktorá umožňuje formuláciu záverov platných pre celú populáciu.

V praxi sa často stretávame so snahou študentov ponúknuť čo najviac štatistických výpočtov, overovaní platnosti svojich zistení a snahou o zovšeobecnenie tzv. „za každú cenu". V snahe ponúknuť to najlepšie induktívne zisťovanie však zabudnú na deskripciu toho, čo vlastne zistili. To však nie je správne, pretože základný popis, pomenovanie a utriedenie zistení musí predchádzať akékoľvek snahy o zložitejšie štatistické výpočty. Napr. ak sa výskumník orientuje na popis príčin nejakého sociálneho javu prostredníctvom štatistických testov a nepopíše, či vôbec tento sociálny jav identifikoval, môže identifikovať výsledky nesprávne.

Oblasť štatistiky, ktorá sa zaoberá teoretickými základmi štatistiky nazývame matematická štatistika. Oblasť štatistiky, ktorá sa zaoberá deskriptívnou a induktívnou štatistikou podľa predchádzajúceho vymedzenia nazývame **aplikovanou štatistikou.** Práve tá je oblasťou, ktorá je v záujme sociálnych a behaviorálnych vied vrátane sociálnej práce.

Aplikovaná štatistika je tá časť štatistiky, ktorá je založená na aplikovanom pozorovaní. Matematická štatistika je založená na matematických vzorcoch a výpočtoch a odpovede sú na základe týchto postupov jasne predvídateľné. Aplikovaná štatistika je založená na rozsiahlych pozorovaniach a prvky matematickej štatistiky využíva na overovanie predpokladov a predpovedí na základe pozorovaného.

Otázky a úlohy:

1. Čo je to štatistika?
2. Definujte 4 chápania štatistiky.
3. Z čoho pozostáva vedecká metóda vo výskumnom procese? Čo z toho zabezpečuje prax štatistiky?

1.1 Niečo z histórie štatistiky

Ak by sme chceli byť dôslední, počiatky štatistických zisťovaní by sme dokázali nájsť už pred viac než 3000 rokmi a opierali by sme sa o biblické texty. Úlohou tejto časti je však zhrnúť vývoj chápania štatistickej praxe v niekoľkých bodoch. Štatistika vo svojom pôvodnom význame vychádza z latinského pomenovania štátu. V tomto chápaní ponúka informácie pre tých, ktorí spravovali štát (najmä zisťovanie počtu žijúcich ľudí, ich pôdy, dobytka,...). Takto získavané informácie slúžili rôznym, či už vojenským alebo civilným účelom (Donnelly 2007). V súčasnosti by sme dokázali nájsť podobnosť v rozsiahlych štatistických zisťovaniach príslušných inštitúcií (najmä štatistických úradov) v podobe napr. sčítania obyvateľov. Špecifikom takéhoto zisťovania je, že sa realizuje na **celej populácii** (kedy hovoríme o tzv. **cenzovom** zisťovaní ako štatistickom zisťovaní a vyhodnocovaní komplexných demografických údajov o každej jednotke v určitom priestore) (Rimarčík 2007).

Medzi ďalšie bežne zobrazované štatistické údaje patria napr. počty narodení, úmrtí, svadieb, rozvodov, ale i informácie o geografických podmienkach jednotlivých štátov (rozlohách, nadmorských výškach,...). Pravidelné uverejňovanie takýchto informácií však nebolo štandardom. To sa mení v podstate až začiatkom 16.storočia, kedy sa vo Veľkej Británii (Londýn) na týždennej báze uverejňovala štatistika úmrtnosti a v Benátkach vyšla publikácia, ktorá zhŕňala štatistické, hospodárske a geografické podmienky v rôznych krajinách (Donnelly 2007).

Rozvoj štatistiky v ďalšom období súvisel najmä s rozvojom jej praktického používania a rovnako s rozvojom matematiky. Za významný míľnik sa považuje vznik teórie pravdepodobnosti v 17. storočí. Skutočný a prudký rozvoj štatistiky nastal koncom 19. a začiatkom 20. storočia, kedy bola vyvinutá v podstate väčšina štatistických metód.

Zmenila sa i podstata štatistických zisťovaní, kedy vyčerpávajúce zisťovania nahradila tzv. **moderná** štatistika, charakteristická metódami umožňujúcimi realizovať závery na základe výberov, bez nutnosti osloviť celú populáciu. Táto moderná štatistika je charakterizovaná i ako matematická štatistika (Sodomová a kol. 2013).

Ďalší obrovský pokrok vzniká s príchodom moderných informačných technológií, kedy zaznamenávame opäť výrazné napredovanie teórie i praxe.

Rozvoj modernej štatistiky je možné charakterizovať v troch fázach:

1. vývoj nových metód štatistickej analýzy (do 50.rokov 20.st.),
2. vysvetľovanie matematických aspektov metód (do 80. rokov 20.st.),

3. rozvoj exploračných techník štatistiky – používaných na hĺbkové skúmanie, tzv. „data mining“ (od 90. rokov 20.st.).

Po tomto období prichádza tzv. **integrovaný prístup**, ktorý spája uvedené prístupy do jedného celku a je charakteristický rozvojom používania technológií.

1.2 A čo sociálna práca?

Sociálna práca ako profesionálna prax i ako teoretický súbor vedeckých poznatkov sa masívne rozvíjala na prelome 19. a 20. storočia. Vznik sociálnej práce bol spojený s pozorovaním individuálnej činnosti prvých profesionálnych sociálnych pracovníkov a posudzovaním efektívnosti ich práce. V tomto kontexte môžeme hovoriť o praxi založenej na kvalitatívnej analýze. Táto prax je charakteristická najmä pre tradičné, diagnostické smerovanie sociálnej práce. S rozvojom sociálnej práce najmä v období okolo 2. svetovej vojny sa rapídne rozširoval okruh sociálnych problémov, pri ktorých sociálna práca našla svoje profesionálne pole pôsobnosti. S týmto rozvojom súvisí i rozvoj poznatkovej bázy, kedy sa rozširoval okruh teórie ako aj vzdelávania. Prax založená na individuálnych dôkazoch úspešnosti však pre takto rapídne rozvíjajúcu sa vednú disciplínu nepostačovala. Preto bolo potrebné začleniť do teórie i do vzdelávania platné vedecké dôkazy o úspešnosti/neúspešnosti jednotlivých postupov a intervencií. Začalo sa tak hovoriť o **praxi sociálnej práce založenej na dôkazoch**, ktorými sú jasné vedecké štúdie s využitím štatistických metód vylučujúcich náhodné efekty činnosti. Máme za to, že práve tým sa štatistika stala neoceniteľnou súčasťou sociálnej práce ako vednej disciplíny i ako oblasti vzdelávania.

Štatistika v sociálnej práci sa tak zameriava na identifikovanie príčin, dôsledkov či podmienok, na pozadí ktorých sa dejú určité sociálne javy v konkrétnom sociálnom prostredí.

Otázky a úlohy:

1. Z akého základu pochádza pojem štatistika?
2. Čo znamená slovo cenzus?
3. Na aké účely bola používaná štatistika v tradičnom chápaní?
4. Čo znamená moderná štatistika?
5. Akú má históriu používanie štatistiky v sociálnej práci?

2 Pojmy špecifické pre štatistiku

Sociálna práca, podobne ako i iné vedné disciplíny, vo svojich počiatkoch reflektovala na spoločenské potreby a priniesla v praxi novú oblasť intervencie a teórie, ktorá v pomáhajúcich profesiách absentovala. Intervencie, ktoré boli využívané v praxi, reagovali na konkrétny prípad, konkrétneho klienta, ktorému sa sociálni pracovníci snažili pomôcť. Nedokázali sa tak oprieť o overené spôsoby. V tomto zmysle hovoríme o práci s individuálnymi prípadmi. Výsledky takejto práce nie je možné považovať za všeobecné, nakoľko chýba dôkaz o ich efektivite či účinkoch aj v iných prípadoch. I štatistika pozná pojem **individuálne javy**, ako také javy, udalosti, či pozorovania, ktoré sú založené na zisťovaní, pozorovaní jedného konkrétneho prípadu, objektu, či jednej skúsenosti. Z hľadiska štatistiky je však takéto pozorovanie nepostačujúce, nakoľko nedokáže pozorovať prípadné pravidlá, nie je dostatočným dôkazom toho, že to, čo sa deje, sa nedeje na princípoch náhody a pod. (Ferjenčík 2006).

Ako sme uviedli pri definovaní základného predmetu štatistiky, základom pre štatistickú prax je skúmanie hromadných javov, najmä ich kvalitatívnej a kvantitatívnej stránky (Kanderová, Úradníček 2007). **Hromadný jav** je udalosť, ktorá sa za jasne vymedzených vecných, časových a priestorových podmienok opakuje hromadne (mnohokrát). Je výsledkom neobmedzene opakovaných pozorovaní takýchto udalostí. V prípade hromadného opakovania takýchto javov je možné pozorovať určité pravidelnosti, ktoré sú v prípade jednorazových udalostí pred očami výskumníkov ukryté (Sodomová 2013; Kanderová, Úradníček 2007). V štatistike teda ide o skúmanie viacerých procesov, udalostí, javov, či objektov a hľadanie spoločných zákonitostí, ktoré by dokázali tieto procesy, udalosti, javy a objekty vysvetliť a poskytnúť reprezentatívny náhľad na skúmanú skutočnosť (Rimarčík 2007).

Pri skúmaní hromadných javov teda predpokladáme, že sa v spoločnosti vyskytujú v takom počte, že ich rôzne hodnoty môžu vytvárať také pravidlá, ktoré by sme dokázali v štatistike skúmať a odhaliť. Koho/čo však pre tieto účely skúmať? Akým spôsobom? A ako je možné výsledky zovšeobecniť? Na koho budú konkrétne platiť? Aj na tieto otázky odpovedá prax štatistiky. V tomto prípade je však potrebné pracovať s presnými štatistickými vymedzeniami a definíciami ďalších štatistických pojmov.

V prvom rade, ide o to, odkiaľ naše jednotlivé pozorovania pochádzajú. Pochádzajú iba tak zo spoločnosti, alebo sa musia vyznačovať nejakými spoločnými vlastnosťami? Základné (no veľmi zjednodušené) pravidlo, ktoré v tomto prípade platí je, že výsledky štatistických zisťovaní je možné zovšeobecňovať (ak sú splnené všetky podmienky, ktoré ešte v tejto učebnici vymedzíme) na tú časť spoločnosti, z ktorej pochádza časť, ktorú skúmame. Pracujeme s pojmami ako populácia (alebo základný štatistický súbor), vzorka (štatistický súbor), štatistická jednotka a štatistický znak.

Základný štatistický súbor (populácia) je najširší z týchto pojmov. Predstavuje súhrn všetkých prvkov, ktoré by mohli byť teoreticky zaradené do výskumu. Všetky tieto prvky vykazujú požadované vlastnosti. Základným štatistickým cieľom je, aby výsledky mohli byť zovšeobecnené práve na základný štatistický súbor (čím získajú v danej skupine/časti spoločnosti všeobecnú platnosť). V praxi sa často stáva, že populácia, na ktorej chceme realizovať výskum je veľmi veľká (hovoríme o desiatkach až stovkách tisícoch), preto nie je efektívne a často ani možné osloviť všetkých členov. Tento rozpor však rieši štatistika prostredníctvom možnej redukcie na základe **štatistického výberu.** Proces výberu je relatívne zložitý a zohľadňuje rôzne špecifiká a spôsoby, ktoré musí výskumník zobrať do úvahy. V praxi nám na to, aký spôsob výberu je najvhodnejší spolu so štatistikou ukazuje metodológia, preto sa o procesu výberu nebudeme v tejto učebnici zmieňovať.
Štatistický súbor (vzorka, výber, alebo tiež označovaný ako výberový súbor) je definovaný ako podmnožina v rámci základného štatistického súboru identifikovaná prostredníctvom výberových kritérií (Kanderová, Úradníček 2007). Ide o množinu prvkov, ktoré majú určité spoločné vlastnosti – tie sú podmienkou pre príslušnosť k štatistickému súboru (Sodomová a kol. 2013). Pri štatistickom súbore definujeme jeho vlastnosti, ktorými sú **rozsah** a **obsah**. Rozsah predstavuje počet prvkov v súbore (Hendl 2012) a obsah vymedzujú tie konkrétne vlastnosti, ktoré majú prvky súboru spoločné. Čím je takýchto prvkov viac, tým je štatistický súbor homogénnejší (no na druhej strane menší). Na základe počtu skúmaných vlastností/znakov zase rozlišujeme **jednorozmerné** (obsahujú jednu zisťovanú vlastnosť/znak) a **viacrozmerné** (zisťujeme dva a viac znakov/vlastností (Sodomová a kol. 2013).
Štatistická jednotka je základným prvkom štatistiky na ktorom skúmame vlastnosti, znaky, alebo prejavy hromadného javu. Je zároveň najmenšou jednotkou štatistického súboru (Sodomová a kol. 2013; Kanderová, Úradníček 2007). Jednoduchšie by sme mohli v oblasti sociálnych vied hovoriť o jednom pozorovaní konkrétneho objektu (jednotlivca; respondenta; účastníka; ale i rodiny, skupiny, komunity v prípade, že ich nedelíme na menšie časti – teda jednotlivých členov). Aby sme konkrétny prvok mohli považovať za štatistickú jednotku, musí spĺňať základné kritériá. Tými sú tzv. *priestorová, časová a vecná príslušnosť*. ***Priestorová príslušnosť*** vyjadruje územie (priestor), z ktorého štatistické jednotky pochádzajú. Všetky

štatistické jednotky prislúchajúce do tohto územia tak tvoria štatistický súbor. ***Časová príslušnosť*** spočíva v časovom rámci, pre ktorý je štatistický výsledok platný. Je samozrejmé, že nie je štatisticky možné zachytiť minulé ani budúce ukazovatele. Preto sú štatistické jednotky zaradené v súbore v konkrétnom čase zberu údajov. ***Vecná príslušnosť*** sa orientuje na nájdenie spoločných vlastností – znakov, ktorými sa vyznačuje každá štatistická jednotka v štatistickom súbore. Môžu tak byť predmetom štatistického skúmania, pretože sú komparovateľné medzi štatistickými jednotkami (Sodomová a kol. 2013; Kanderová, Úradníček 2007).

Štatistický znak je vlastnosť štatistickej jednotky, ktorá je jasne identifikovateľná. Štatistický znak musí spĺňať podmienku merateľnosti – musí byť identifikovaná prostredníctvom procesu merania – kvantifikácie. ***Kvantifikáciou*** tak rozumieme snahu charakterizovať štatistický znak prislúchajúci každej štatistickej jednotke prostredníctvom číselnej charakteristiky (miery, veľkosti, intenzity). Zjednodušene, ide o proces merania charakterizovaný ako snaha o identifikovanie skúmanej vlastnosti prostredníctvom konkrétneho čísla (Ferjenčík 2006). ***Hodnota štatistického znaku*** je následne konkrétne číslo, hodnota, typ, označenie toho ktorého štatistického znaku.

Štatistické znaky samozrejme môžu mať rôzny charakter. Podľa ich povahy a teda aj povahy zisťovanej vlastnosti rozlišujeme základné **typy štatistických znakov**. ***Kvalitatívne*** štatistické znaky sa označujú taktiež ako slovné, alebo kategoriálne. Ide o štatistické znaky, ktoré odpovedajú najčastejšie na otázky „aký typ“, „aká kategória“. Sú to slovné štatistické znaky prostredníctvom ktorých je možné zaradiť štatistické jednotky do niekoľkých kategórií. Patrí sem napr. rodové, vzdelanostné, národnostné rozdelenie, rozdelenie podľa miesta bydliska, výkonu zamestnania a pod. ***Kvantitatívne***, alebo taktiež číselné štatistické znaky sú také, ktoré je možné vyjadriť konkrétnym číslom (podľa miery, veľkosti, intenzity,...). Najčastejšia otázka je v tomto prípade otázka „koľko“, pričom odpoveďou je stále konkrétne číslo. Môže ísť o diskrétne alebo spojité štatistické znaky. Diskrétne môžu nadobúdať iba konečné a celočíselné hodnoty. Spojité nadobúdajú ľubovoľné hodnoty z vymedzeného intervalu.

Rozdelenie základných štatistických pojmov a ich identifikovanie vo výskume sú v praxi pomerne jednoduché. Predstavme si, že bol realizovaný výskum, ktorý slúžil pre zistenie úrovne rizikového správania a jeho sociálnych súvislostí u slovenských adolescentov. V tomto prípade základným štatistickým súborom, teda populáciou budú slovenskí adolescenti. V tomto prípade ide o všetkých adolescentov. Keďže nie je prakticky realizovateľné osloviť všetkých z populácie, na základe štatistického výberu je identifikovaný štatistický súbor (vzorka, výber) na ktorom sa konkrétny výskum zrealizuje. Každý jeden respondent je považovaný za štatistickú jednotku. Keďže je výskum orientovaný *na zisťovanie rizikového správania, samotné rizikové správanie je štatistickým znakom – teda tým, čo je v danom prípade zisťované. Konkrétny typ štatistického znaku závisí od spôsobu položenia otázok, ich typu a neskoršieho vyhodnotenia (ak je rizikové správanie identifikované ako číslo – počet takýchto rizikových prejavov v sledovanom období, bude to kvantitatívny znak, ak sa budeme zaujímať o konkrétne typy rizikového správania, bude to kvantitatívny znak).*

V teórii i praxi sociálnej práce sa často stretávame s potrebou preskúmať určitý, často negatívny, sociálny jav, jeho príčiny či následky. Je to najmä preto, aby sme dokázali nastaviť účinné spôsoby a postupy intervencie. Rovnako i tie by mali byť verifikované ako funkčné a pre klientov zmysluplné. Jedinou účinnou cestou ako poskytnúť empirické dôkazy pre tieto účely sú štatistické postupy. Základným štatistickým súborom sú tak všetci klienti vykazujúci daný sociálny jav (napr. nezamestnanosť a pod.). Z tejto populácie je na základe štatistického výberu selektovaný štatistický súbor, jednotliví členovia (osoby, respondenti) sú tak štatistickými jednotkami (prvkami štatistického súboru). U týchto štatistických jednotiek sú pozorované štatistické znaky a súhrnné štatistické výsledky sú následne (zasa procesom štatistiky, tentoraz induktívnej, resp. inferenčnej) zovšeobecňované na pôvodnú populáciu. Tento nikdy nekončiaci cyklus je základom pre každú vednú disciplínu a poskytuje dôkazy o efektivite činností a opatrení v rámci teórie a praxe.

Cvičenie:

V nasledujúcom zhrnutí realizovaného výskumu identifikujte **základný štatistický súbor, štatistický súbor, štatistickú jednotku** (pokúste sa ich vymedziť prostredníctvom základných kritérií)**, štatistické znaky** (vrátane typov) **a ich hodnoty.** Tieto údaje doplňte do prehľadu pod zhrnutie výskumu. V prípade potreby doplňte riadky.

V rámci výskumu realizovaného v mesiaci máj 2020, ktorý sa zaoberal starostlivosťou o seba u pracovníkov v zariadeniach sociálnych služieb na Slovensku, oslovili výskumníci 850 pracovníkov pracujúcich v týchto zariadeniach. Priemerný vek týchto pracovníkov bol 42 rokov, priemerná dĺžka praxe bola 12,6 roka. Okrem iných ukazovateľov bola zisťovaná frekvencia preventívnych prehliadok u lekára (pričom pracovníci vyberali z možností: každý rok, každé dva roky, raz za 5 rokov a viac, alebo vôbec); ich športová aktivita (denne, aspoň 3x týždenne, aspoň raz za týždeň, príležitostne, nikdy).
Z výsledkov vyplýva, že až 40% respondentov nevyvíja žiadnu športovú aktivitu a 52% respondentov chodí na preventívne prehliadky raz za dva roky.

Základný štatistický súbor: ....

Štatistický súbor: ....

Štatistická jednotka: ....

priestorová príslušnosť: ....

časová príslušnosť: ....

vecná príslušnosť: ....

Počet štatistických znakov: ....

konkrétne – názov	typ	konkrétna hodnota
....	....	....

Otázky a úlohy:

1. Ako by ste vymedzili rozdiel medzi individuálnym a hromadným javom?
2. Čo je to základný štatistický súbor a čo štatistický súbor?
3. Ako je vnímaný pojem štatistická jednotka? Aké vymedzenie štatistických jednotiek je potrebné brať do úvahy?
4. Čo je to štatistický znak? Aké štatistické znaky rozoznávame?
5. Ako by ste charakterizovali proces štatistiky prostredníctvom pojmov príznačných pre štatistiku?

3 Premenné a ich úlohy v štatistike

Ako sme v predchádzajúcej časti, pri definovaní základných pojmov naznačovali, štatistické jednotky sa vyznačujú rôznymi vlastnosťami. Tieto vlastnosti sú predmetom výskumov a označujú sa pojmami charakteristika, alebo štatistický znak**. Konkrétnu číselnú reprezentáciu konkrétneho štatistického znaku nazývame pojmom premenná**. Aký je teda rozdiel medzi charakteristikou a premennou? Keďže štatistika pracuje s číslami, štatistický znak sa stáva premennou vtedy, **ak ho dokážeme číselne vyjadriť**. V tomto kontexte Ferjenčík (2006, s. 13) uvádza podmienku ktorú je v prípade konkrétnej charakteristiky splniť. Tou je formulácia „pravidiel, podľa ktorých môžeme jej jednotlivé podoby (hodnoty-úrovne) vyjadriť číslami“. Každá premenná teda medzi štatistickými jednotkami nadobúda konkrétne číselné hodnoty (tie sa u jednotlivých štatistických jednotiek líšia – variujú, čo je i podstatou anglického označenia „Variable“). Odlišnosť hodnôt premennej však nespočíva iba v odlišnostiach medzi jednotlivými štatistickými jednotkami, ale aj v možnej zmene hodnoty v čase u konkrétnej štatistickej jednotky (ak napríklad meriame konkrétnu premennú, napr. pracovnú spokojnosť dnes a následne o niekoľko mesiacov, je predpoklad, že tieto hodnoty u rovnakých štatistických jednotiek nebudú rovnaké).

Premenné môžu nadobúdať niekoľko, alebo mnoho hodnôt z určitej, často presne definovanej množiny údajov. Ako sme uviedli, premenné pracujú s číselnými údajmi. Čo ale v prípade, že zisťujeme charakteristiku, ktorá nemá číselnú odpoveď? Sú tieto charakteristiky ako premenné štatisticky nepoužiteľné? Toto tvrdenie by bolo príliš prehnané. Takéto premenné v štatistike slúžia najmä na identifikáciu rôznych štatistických aspektov, najmä rozdelenie štatistického súboru podľa rôznych kritérií na skupiny a pod. Aby sme však takýto štatistický znak použili ako premennú, stále platí podmienka uvedená vyššie – nájsť kritérium, na základe ktorého ho dokážeme číselne vyjadriť.

Jednoduchý príklad na uvedené je charakteristika rod. V štatistickom použití je potrebné nájsť spôsob číselného vyjadrenia, teda napríklad všetky ženy číselne označíme ako 1, všetkých mužov ako 2. Rod je v tomto prípade premenná, ktorá nadobudne 2 hodnoty. O niečo zložitejšie je to napríklad v premennej miesto pôsobenia, kde musíme jednotlivým typom organizácií prideliť číselné označenie, s ktorým môžeme štatisticky ďalej pracovať.

Dôležitou vlastnosťou, ktorú musí premenná spĺňať, je **nezameniteľnosť jej hodnôt**. V štatistickej praxi je neprípustné, aby premenná mohla nadobúdať v jednej chvíli dve, alebo viac hodnôt. Ak by to tak bolo, musíme hovoriť o viacerých premenných, nie o jednej. Premennou tak môžu byť tie otázky, na ktoré je možné pre každého respondenta nájsť iba jednu odpoveď (Rimarčík 2007).

Na vysvetlenie tohto javu použijeme síce neštandardný príklad, ale pre ilustráciu takéhoto stavu ho považujeme za vhodný. Predstavme si, že je úlohou výskumníka urobiť si predstavu o farbe očí skupiny študentov. Tento výskum by bol veľmi jednoduchý, výskumník by si zaznamenával farbu očí jednotlivých študentov. Následne by jednotlivým farbám pridelil konkrétne číslo, čím by sa zo štatistického znaku stala premenná. Vysvetľovanou podmienkou je však nezameniteľnosť hodnôt. Tu však výskumník narazí na problém. Jeden zo študentov má každé oko odlišnej farby. Akú farbu výskumník označí v prípade takéhoto študenta? Problém je však niekde inde – už na začiatku výskumu mal výskumník myslieť na to, že takýto jav môže nastať a zisťovanie tejto premennej malo prebiehať na úrovni nie jednej, ale dvoch premenných. Teda samostatná hodnota pre farbu pravého a samostatná pre farbu ľavého oka.

Číselná povaha premenných je pre štatistiku kľúčová. Umožňuje realizovať matematické operácie a viacero spôsobov manipulácie, napr. triedenie, komparácia, a pod. Napomáha rovnako k získaniu komplexného náhľadu na skúmaný jav. Samozrejme, premenné môžu nadobúdať viacero podôb, čo je determinované najmä charakterom skúmaného štatistického znaku. To, aký má premenná charakter, následne ovplyvní ďalší štatistický postup. Preto v nasledujúcej časti priblížime niekoľko základných kategorizácií premenných, ktoré v štatistike rozlišujeme.

3.1 Kategorizácia premenných a ich použitie v štatistike

Ucelený prehľad v prípade kategorizácie môže poslúžiť ku korektnému rozhodovaniu o použití konkrétnych štatistických metód. Nie v každom prípade totiž môžeme použiť všetky metódy, ktoré štatistika rozoznáva. Do výraznej miery to závisí od povahy údajov, ktoré tvoria premenné a tým pádom i od spôsobu ich zisťovania meraním.

I. číselné vs. kategoriálne premenné

Triedenie premenných na číselné a kategoriálne je základnou kategorizáciou, ktorá sa uvádza v rôznych štatistických publikáciách v oblasti sociálnych vied. Vyjadruje povahu dát, ktoré je možné vo forme premenných získavať.

Číselné premenné rovnako nazývame i kvantitatívnymi. Odpovedajú na základnú otázku „koľko?“ čím je vyjadrená ich číselná povaha. Patria sem všetky údaje, ktoré sa dajú vyjadriť konkrétnym číslom (a samozrejme konkrétnou jednotkou merania).

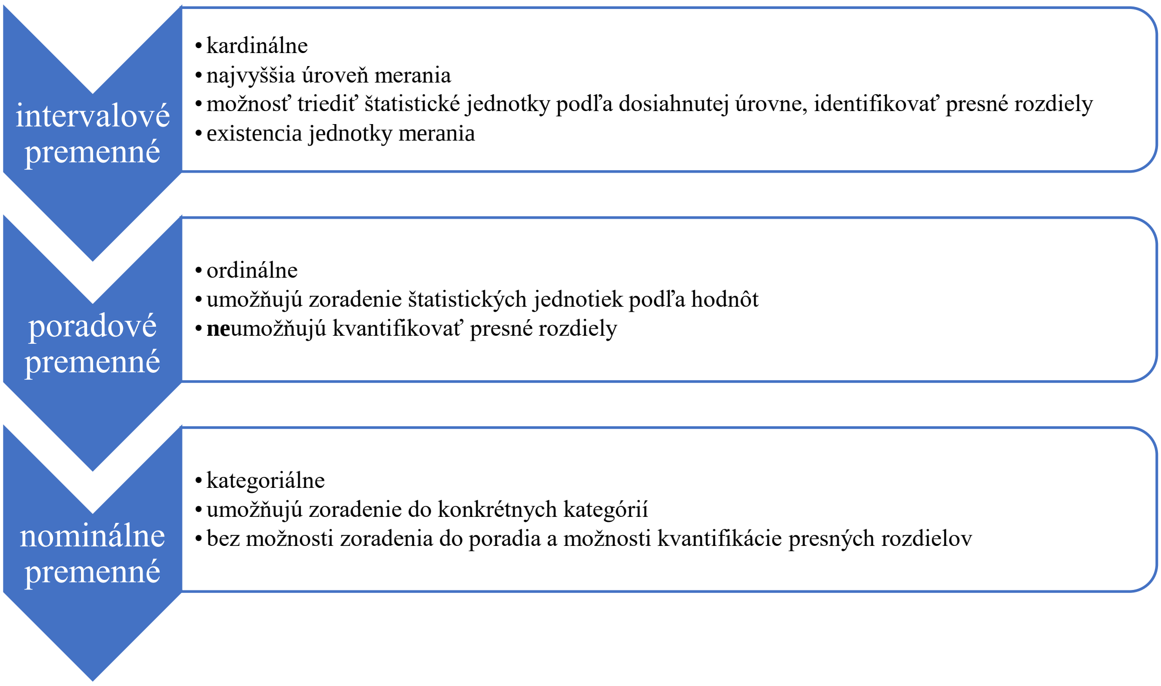
Patria sem napríklad premenné ako výška (meraná v centimetroch); hmotnosť (v kilogramoch); vek (v rokoch); príjem (v eurách); ale aj premenné, ktoré sú výsledkom merania v sociálnych vedách (číselné skóre dotazníka) či premenné, ktoré sú výsledkom štatistického testu (priemerná hodnota v populácii).

Kategoriálne premenné tiež označované ako kvalitatívne, resp. slovné. Odpovedajú na otázku „aký typ?“. Patria sem údaje, ktoré sa síce nedajú vyjadriť číslom, ale slovne označujú určitú kategóriu, s ktorou neskôr vieme v štatistike pracovať.

Patria sem premenné ako napr. pohlavie; národnosť; rodinný stav; bydlisko; vzdelanie a pod. Základnou podmienkou ich použitia v štatistike je ich transformácia do číselného vyjadrenia (napr. vyjadriť jednotlivé kategórie konkrétnym číselným označením, ktoré budeme v štatistickom spracovaní používať – v prípade pohlavia je to napríklad označenie žien číslom 1, mužov číslom 2 a pod.).

II. intervalové, poradové a nominálne premenné

Kategorizácia premenných na intervalové, poradové a nominálne je najpoužívanejšou a jednou z najvýznamnejších kategorizácií. Do značnej miery ovplyvňuje budúce štatistické spracovanie a pre výskumníkov je toto triedenie kľúčovým. Každá kategória predstavuje inú úroveň merania. Toto triedenie je do výraznej miery spojené s metodologickými zásadami merania premenných a dôležité je zohľadňovať tento fakt ešte pred samotným štatistickým spracovaním – a teda už pri návrhu výskumnej stratégie. Premenné podľa tejto kategorizácie nadobúdajú rôznu úroveň a faktom je, že poskytujú i určitú hierarchiu tak ako je to uvedené nižšie.



Obr. 2 - hierarchia kategórií premenných

Intervalové premenné, inak označované aj ako kardinálne. Matematicky aj štatisticky umožňujú najrozsiahlejšie operácie a poskytujú najhodnovernejšie výsledky. Dôvodom je samotná povaha týchto premenných. Každá štatistická jednotka tu dostáva hodnotu pochádzajúcu zo stupnice, na ktorej vieme presne identifikovať rovnaké rozdiely medzi jednotlivými jednotkami (úrovňami/dielikmi). Na základe toho je tak možné štatistické jednotky zoraďovať podľa dosiahnutej úrovne (konkrétneho nameraného čísla) ako aj hovoriť o presných rozdieloch medzi jednotlivými hodnotami. Pri vzájomnom porovnávaní je možné napríklad presne určiť, o koľko je nameraná hodnota štatistickej jednotky vyššia/nižšia od inej štatistickej jednotky, alebo aj vykonávať matematické operácie (napr. sčítavanie, odčítavanie a pod.). Presné výsledky zaručuje štandardná jednotka merania, ktorá ostáva rovnaká po celej dĺžke škály rovnaká.

Kde je na intervalovej premennej počiatok? Diskusie ohľadne intervalových premenných sa viedli o začiatku intervalu, od ktorého je táto premenná identifikovaná. Znamená bod nula úplný začiatok stupnice, pod ktorú nemôže žiadna hodnota klesnúť? Intervalové premenné rátajú s oboma možnosťami. V prípade, že nula na intervalovej stupnici je dohodnutá (tzv. „arbitrálna“) a hodnota môže klesnúť aj do záporných hodnôt (typickým príkladom je teplota meraná v stupňoch Celzia), hovoríme o „klasickej“ intervalovej úrovni. V prípade, že bod nula je absolútnym začiatkom stupnice a pod tento bod nemôže hodnota nikdy klesnúť (napr. teplota meraná v stupňoch Kelvina), hovoríme o **intervalovej pomerovej premennej**. Tá je vnímaná ako osobitný typ intervalovej premennej, spĺňa všetky jej parametre, no navyše umožňuje aj určovanie tzv. pomerov (o aký pomer je hodnota v prípade jednej štatistickej jednotky vyššia/nižšia ako u druhej štatistickej jednotky). Intervalová pomerová premenná je tak považovaná v pomyselnej hierarchii za absolútne najvyššiu možnú úroveň merania.

Jedným z príkladov na intervalové meranie je dĺžka praxe sociálneho pracovníka v konkrétnom zariadení. Keďže prax sa meria vo vopred určených a prijatých jednotkách (dni/týždne/mesiace/roky – čo je na rozhodnutí výskumníka), spĺňa táto premenná parametre pre intervalovú premennú (existencia jednotky, rovnaké rozostupy medzi dielikmi). Výskumník tak v rámci vyhodnotenia dokáže povedať, aké sú rozdiely medzi jednotlivými respondentmi v prípade dĺžky praxe, o koľko má niekto vyššiu/nižšiu prax a pod. Keďže však v tomto prípade ide aj o pomerovú premennú (prax = 0 je absolútnym začiatkom stupnice a jej hodnota pod 0 nikdy neklesne), vie výskumník presne identifikovať pomery medzi jednotlivými respondentmi.
Napr.: respondent A má prax 10 rokov, respondent B má prax 20 rokov. Respondent B má 2x dlhšiu prax ako respondent B.

Poradové premenné označované ako ordinálne. Na hierarchii sú o stupeň nižšie než intervalové premenné. Ako už zo samotného názvu vyplýva, umožňujú zoradiť na základe ich hodnôt štatistické jednotky do konkrétneho poradia vzhľadom na ich veľkosť (intenzitu, mieru,...). Na základe týchto hodnôt vie výskumník spoľahlivo konštatovať, ktorá hodnota je väčšia, menšia, rovná,... . Keďže na rozdiel od intervalovej premennej nepoznáme presnú hodnotu (poznáme iba poradie), neumožňujú určiť presné rozdiely medzi hodnotami štatistických jednotiek, čo neumožňuje realizovať niektoré matematické a štatistické operácie, ako napríklad sčítanie, odčítanie, výpočet celkového skóre či výpočet aritmetického priemeru (ktorý v tomto prípade nemá zmysel). Rovnako oproti intervalovým premenným nemusia byť jednotlivé dieliky v poradí od seba rovnako vzdialené.

Príkladom sa opäť vrátime k dĺžke praxe sociálneho pracovníka v konkrétnom zariadení. Predstavme si, že respondenti nebudú určovať dĺžku svojej praxe konkrétnym číslom, ale iba prostredníctvom niekoľkých možností (budú mať na výber prax do 5 rokov, 5-10 rokov a nad 10 rokov). Z takto položenej otázky je možné určiť, ktorá skupina respondentov má prax najkratšiu, ktorá najdlhšiu, prípadne ktorí respondenti patria do rovnakej skupiny podľa dĺžky praxe. Takýmto spôsobom je tak možné určiť poradie. Keďže však nepoznáme presnú dĺžku praxe, nedokážeme rozdiely kvantifikovať úplne presne.

Ďalším príkladom by mohlo byť poradie klientov konkrétneho zariadenia podľa miery ich sociálnej odkázanosti. Ak nepoznáme presné percentuálne ohodnotenie ich sociálnej odkázanosti a poznáme iba kategórie, môžeme presné hodnoty iba odhadnúť na základe pridelených kategórií. Tu však narážame na problém presného určenia, keďže hranice jednotlivých kategórií sú určené ako rozsah. Preto sa môžu v jednej skupine ocitnúť klienti viac a menej odkázaní, ktorí síce stále spĺňajú kritérium rozsahu zaradenia do rovnakej skupiny, no ich konkrétne odkázanie je iné.

V oblasti sociálnych a behaviorálnych vied, kde patrí i sociálna práca, sa pre zisťovanie väčšiny premenných používajú otázky vyjadrujúce odpovede respondentov na odpoveďovej škále (napr. od 1 do 5, pričom 1 znamená úplne nesúhlasím, 5 úplne súhlasím, alebo iné typy odpoveďových škál). Ak by sme mali byť matematicky a štatisticky úplne presní, konštatovali by sme, že identifikujeme poradové premenné. To by však znamenalo, že by sme nemohli používať množstvo štatistických metód a vylúčená by bola i štatistická indukcia. Preto v oblasti štatistiky sociálnych a behaviorálnych vied platí úzus o používaní štatistických metód určených aj pre intervalové premenné v prípadoch, že poradová premenná má aspoň 5 rôznych hodnôt (teda minimálne pri škálach 1-5) a nie je dôvod predpokladať, že medzi jednotlivými hodnotami sú zásadné rozdiely (čo pri štandardne používaných škálach nie je) (Rimarčík 2007). Tým je zabezpečená možnosť štatistického spracovania pre väčšinu používaných škál, čo umožňuje vedecké bádanie tak, ako v iných vedných oblastiach.

Nominálne premenné, inak označované ako kategoriálne sú v hierarchii najnižšie a ich názov predpovedá, že slúžia najmä pre roztriedenie štatistických jednotiek do rôznych kategórií podľa určitého slovného kritéria (aj názov nominálne pochádza z výrazu „nomen“, teda názov, meno). Keďže jednotlivé skupiny sa zakladajú na iných než na matematických princípoch (neoznačujú číselné vyjadrenie), sú si z matematického hľadiska rovnocenné. Preto nedokážeme určiť ani poradie (nevieme povedať, čo je viac, alebo menej), a rovnako nie je možné realizovať žiadne matematické operácie. V štatistike sa však často využívajú na roztriedenie výskumného súboru do kategórií a následné hľadanie rôznych štatistických významov (napr. rozdielností a pod.) v iných, číselných premenných no podľa nominálne stanovených kategórií. V štatistickej praxi (ktorá pracuje s číselnými vyjadreniami) sa aj takéto premenné kvantifikujú – teda premieňajú na čísla. To sa realizuje najčastejšie jednoduchým priradením konkrétneho čísla ku konkrétnej kategórii (no tieto čísla nemajú žiaden matematický význam a je na rozhodnutí výskumníka, ktorej kategórii aké číslo priradí).

Príkladom z oblasti sociálnej praxe môže byť pracovná pozícia pracovníkov konkrétneho zariadenia. Prax sociálnej práce je špecifická tým, že v zariadeniach sociálnych služieb pracuje mnoho zamestnancov vykonávajúcich rôzne pracovné činnosti podľa ich vzdelania, pracovného zaradenia, skúseností a pod. (v jednom zariadení sa tak môžeme stretnúť so sociálnymi pracovníkmi, ošetrovateľmi, opatrovateľmi, zdravotnými sestrami, lekármi, psychológmi, terapeutmi a ďalšími pracovníkmi). Pri realizovaní rôznych zisťovaní (napr. pracovnej spokojnosti pri výkone zamestnania) slúžia pracovné pozície ako kategórie podľa ktorých sa hľadajú medzi nimi rozdiely (v číselnej premennej, napr. pracovná spokojnosť). Samotné pozície však vystupujú ako nominálna premenná, keďže žiadne matematické operácie ani určovanie poradia nemá matematický ani štatistický význam.

Osobitným druhom nominálnej premennej sú **nominálne binárne /dichotomické/ premenné**. Sú to tie premenné, ktoré v rámci možností odpovede nadobúdajú dve možné kategórie (napr. premenná rod – muž, žena a pod.). Takáto premenná môže byť z hľadiska vzdialeností medzi jednotlivými bodmi považovaná za intervalovú premennú, zo svojej samotnej povahy sa správa aj ako najjednoduchšia intervalová premenná (vzdialenosť medzi jednotlivými dielikmi /keďže sú iba 2/ je rovnaká z každej strany). V určitých prípadoch je tak možné použiť aj v tu niektoré výpočty určené pre intervalové premenné. Z hľadiska ďalších kritérií a realizovaných matematických a štatistických operácií sa za takúto považovať nemôže.

Vhodnou pomôckou pre určovanie typu premennej je uvedomenie si jej prípadného ďalšieho použitia vo forme štatistickej operácie – napr. výpočtu aritmetického priemeru (ktorý bude podrobnejšie opísaný v ďalšej časti). Aritmetický priemer sa dá síce vypočítať z akejkoľvek číselnej premennej, no jeho interpretácia má význam iba v prípade intervalovej premennej. Výpočet aritmetického priemeru z poradovej ani nominálnej premennej nemá logický význam.

Transformácia premenných, inak označovaná ako prekategorizovanie premenných je jednoduchá operácia, kedy môže výskumník v rámci predstavenej hierarchie pracovať s premennou vyššej kategórie ako s premennou nižšej kategórie (teda s intervalovou ako s poradovou, alebo nominálnou; s poradovou ako nominálnou). Je dôležité si uvedomiť, že transformácia funguje stále z vyššej kategórie na nižšiu, v opačnom prípade by ju nebolo možné realizovať.

Ak napríklad výskumník zisťuje vek na intervalovej úrovni (jednoducho sa pýta na vek v určených jednotkách, napr. v rokoch), jednoduchým roztriedením výskumného súboru do poradia podľa veku *získa poradovú premennú alebo roztriedením do 2 kategórií (starší ako 15 rokov?: áno-nie), nominálnu premennú.*

Aj vzhľadom na možnosť neskoršej transformácie je stále z pozície výskumníka nutné uvažovať nad tým, akým spôsobom bude potrebné so získanými premennými ďalej narábať. To si musí výskumník premyslieť ešte pred samotnou realizáciou výskumu vo fáze metodologickej prípravy. Z praktického hľadiska je vhodné premenné vždy kvantifikovať na najvyššej možnej úrovni, ktorá je z hľadiska danej premennej a neskoršieho použitia zmysluplná a následne umožňuje transformáciu na nižšiu úroveň.

III. premenné z hľadiska kontinuity

Kontinuitu premenných vnímame ako vlastnosť, ktorá určuje povahu údajov, ktoré môžu jednotlivé premenné nadobúdať. Z tohto hľadiska vnímame premenné **diskrétne** a **spojité**. Diskrétne (alebo nespojité) premenné môžu nadobúdať stanovený počet hodnôt, pričom medzery medzi jednotlivými hodnotami nemusia byť medzi jednotlivými bodmi rovnaké. Rovnako nemusí existovať jednotná jednotka merania týchto premenných. Do tejto kategórie patria všetky premenné nominálnej a poradovej úrovne, no môžu sem patriť i premenné intervalovej úrovne (čo záleží od rozhodnutia výskumníka a použitého spôsobu merania). Spojité (tiež kontinuitné) premenné sú typické pre intervalové premenné a charakterizujú sa ako tie škály, ktoré môžu nadobúdať nekonečné množstvo hodnôt zo stupnice bez medzier medzi jednotlivými bodmi (Ferjenčík 2006), resp. z určitého intervalu reálnych čísel (Hendl 2012). Je dôležité upozorniť, že aj s diskrétnymi premennými je možné štatisticky pracovať ako so spojitými, čo potvrdzuje úvahy o štatistickom spracovaní poradových premenných v sociálnych a behaviorálnych vedách uvedené v predchádzajúcej časti.

Rozdelenie premenných na diskrétne a spojité je relatívne zložito pochopiteľné bez ilustrácie. Jednoduchšie vyjadrené, ak je možné identifikovať medzi jednotlivými bodmi nekonečné množstvo menších hodnôt ako je to napríklad pri premennej výška, vzdialenosť a pod. (kde dokážeme určiť stále menšie časti na základe štandardnej stupnice), ide o spojitú premennú. Ak je meranie obmedzené na presne určený počet hodnôt, kde medzi nimi neexistujú (alebo ich nevnímame a nepracujeme s nimi), pôjde o diskrétnu premennú (ako príklad môžu slúžiť odpoveďové škály alebo stupnice). Diskrétnosť v tomto prípade neznamená, že sa nebude dať s premennou ďalej pracovať – je predsa možné identifikovať rôzne matematické a štatistické operácie (napr. aritmetický priemer) aj pri odpoveďových škálach, teda diskrétnych premenných.

IV. premenné z hľadiska ich (ne)závislosti

Kategorizácia premenných z hľadiska ich (ne)závislosti je významná najmä pri štatistickom spracovaní identifikovaných premenných testami induktívnej štatistiky. Z tohto pohľadu rozlišujeme **závislé** a **nezávislé** premenné. Veľmi zjednodušene povedané, v rôznych výskumných zisťovaniach sa výskumníci snažia identifikovať, do akej miery niektoré premenné vysvetľujú (predikujú) iné premenné. Z druhého uhla pohľadu, do akej miery sú niektoré premenné závislé na iných (vysvetľujúcich) premenných. Ešte jednoduchšie, ako sú závislé premenné závislé na nezávislých. Takýmto spôsobom teda rozoznávame závislé (odpoveďové, kriteriálne) a nezávislé (explanačné, prediktorové) premenné (Hendl 2012).

Určenie týchto premenných nemá vplyv na ich kvantifikáciu, je to teda na rozhodnutí výskumníka a záleží na tom, akým spôsobom výskumník pripravuje projekt výskumu a ako položí výskumné otázky a výskumné hypotézy. Ak sa napríklad snažíme identifikovať vplyv poskytnutého poradenstva na kvalitu života klientov sociálnej práce, premennú kvalita života musíme postaviť do pozície závislej premennej a rozsah poskytnutého poradenstva bude nezávislá premenná. Ak by sme však na druhej strane identifikovali vplyv vzdelania a skúseností poradcov na poskytované poradenstvo, vzdelanie a skúsenosti by boli nezávislou a poskytované poradenstvo závislou premennou. Je dôležité uvedomiť si, že v jednom výskumnom spracovaní môže byť premenná považovaná za závislú, v inom za nezávislú. Toto (odlišný prístup k premennej, raz ako závislej, inokedy ako nezávislej) sa môže stať aj v rámci jedného realizovaného výskumu, kde sa práve prostredníctvom zmeny prístupu môžu preukázať rôzne kauzálne vzťahy, ktoré môžu byť pre interpretáciu dôležité. Takýmto spôsobom je možné preukazovať (okrem iného) účinnosť poskytnutých intervencií, čo môže posunúť teóriu i prax na kvalitatívne významne vyššiu úroveň tým, že budú poskytované také intervencie, ktoré prinášajú kvalitatívne lepšie výsledky. Tento príklad preukazuje, že aj v oblasti sociálnych vied má používanie štatistiky ako i rozdeľovania na závislé a nezávislé zmysel.

V. nepriame a mätúce premenné

Ako sme v predchádzajúcej časti uviedli, základnou snahou vo vedeckom výskume je prepojiť nezávislé a závislé premenné. To sa deje prostredníctvom rôznych metód. V klasickom psychologickom výskume sa toto realizuje prostredníctvom experimentálnej manipulácie, pričom výskumníci sa snažia, aby ostatné relevantné premenné ostali konštantné. V rigidných učebniciach najmä z oblasti psychologického výskumu sa uvádza, že delenie na závislé a nezávislé premenné má zmysel iba pri experimentálnych typoch výskumov (iba tak je možné mať pod kontrolou všetky potenciálne ovplyvňujúce premenné). Iné zdroje však uvádzajú, že pojmy závislá a nezávislá premenná pôvodne pochádzajú z matematických výskumov a týkajú sa akýchkoľvek súvisiacich premenných (Coolican 2014). K týmto zdrojom sa prikláňame i my keď konštatujeme, že v oblasti sociálnych a behaviorálnych vied, kde radíme i sociálnu prácu, sa vykonáva veľké množstvo výskumov, ktoré nie sú experimentálneho charakteru a napriek tomu v nich pracujeme so závislými a nezávislými premennými. Pri týchto výskumoch priamo pracujeme s predpokladom, že nie všetky premenné máme ako výskumníci pod kontrolou. Tieto premenné nazývame **nepriamymi** (vedľajšími) premennými. Do tejto kategórie radíme tie premenné, ktoré nie sú sledovanými nezávislými premennými, no môžu ovplyvniť výsledky výskumu (stav a výsledok závislej premennej). Nepriame premenné môžu na závislé premenné pôsobiť rôznym spôsobom. Niekedy sa stáva, že pôsobia ako **mätúce** premenné. Tie sú charakterizované ako premenné, ktoré ovplyvnia výsledky, no pokiaľ si ich výskumník neuvedomí, je možné, že prijme neexistujúci (alebo omnoho menej/viac výrazný) vplyv nezávislej premennej.

Príkladom by mohlo byť napr. nezohľadnenie sezónneho aspektu nezamestnanosti v skúmaní efektivity rôznych programov pre prácu s mladými nezamestnanými absolventmi. Ak by boli realizované výskumy preukazujúce efektivitu konkrétnych programov uplatniteľnosti mladých absolventov škôl a tieto výskumy by boli realizované pred koncom školského roka, kedy je nezamestnaných absolventov najmenej (keďže študenti posledných ročníkov ešte neukončili vzdelávanie), premenná obdobie (sezóna) skúmania by do výsledkov vstúpila ako mätúca premenná. To by sa mohlo odstrániť niekoľkými nezávislými testovaniami v priebehu celého roka a pod. Rovnako zmätočne môžu v sociálnych a behaviorálnych výskumoch pôsobiť rôzne demografické premenné, ale napríklad aj spoločenská, politická, či sociálna situácia v danom regióne (ktorá sa môže odlišovať od situácie v inom regióne).

Cvičenie

1. Nasledujúca tabuľka vyjadruje zhrnutie výskumných zistení. Identifikujte typy premenných podľa troch základných kategorizácií.

Tab. 1- identifikácia premenných

Číslo dotazníka	Rod	Vek	Pracovná pozícia	Počet rokov v praxi	Skóre dotazníka Únava z pomáhania
1	Muž	36	Sociálny pracovník	15	19
2	Žena	42	Terapeut	18	15
3	Žena	28	Sociálny pracovník	5	14
4	Žena	23	Sociálny pracovník	1	18
5	Muž	42	Inštruktor	20	21
6	Muž	39	Ošetrovateľ	19	26
7	Žena	41	Ošetrovateľ	15	11
8	Žena	56	Riaditeľ	30	14
9	Muž	30	Sociálny pracovník	5	19
10	muž	29	Terapeut	3	17

Tab. 2 - identifikácia premenných (vyhodnotenie)

	Číslo dotazníka	Rod	Vek	Pracovná pozícia	Počet rokov v praxi	Skóre dotazníka Únava z pomáhania
I. číselné/ kategoriálne						
II. intervalové, poradové, nominálne						
III. z hľadiska kontinuity						

2. Ak by bolo zámerom výskumníka identifikovať, či počet rokov v praxi ovplyvňuje to, do akej miery je respondent „unavený z pomáhania“, ktorá premenná by bola v úlohe závislej a ktorá v úlohe nezávislej?

závislá ....
nezávislá ....

3. Skúste sa zamyslieť nad príkladom predchádzajúcej úlohy. Aké iné premenné by mohli v danom prípade pôsobiť na výsledok z pozície nepriamych a mätúcich premenných?

nepriame: ....

mätúce: ....

4. Vráťme sa k cvičeniu 1. Pozorne si prezrite premennú „Počet rokov v praxi“. Pokúste sa túto premennú transformovať na poradovú a nominálnu úroveň. Ako by vyzerali premenné po transformácii?

Otázky a úlohy:

1. Definujte pojem premenná.
2. Kedy sa stáva charakteristika premennou?
3. Ako by ste charakterizovali kvantifikáciu?
4. Aké kategorizácie premenných rozoznávame?
5. Čo je to transformácia premenných a akým smerom je možná?
6. Je možné pracovať so závislými a nezávislými premennými i vo výskumoch týkajúcich sa sociálnej práce?
7. Aké sú to nepriame a mätúce premenné?

II. Deskriptívna analýza dát

4 Deskriptívna – popisná štatistika

Ako sme uvádzali v predchádzajúcej časti, deskriptívna štatistika je druhou časťou štatistickej praxe hneď po získavaní údajov a ide o tzv. exploračnú analýzu údajov. Samotný názov napovedá, že ide najmä o popis údajov. Definične však pojem popis nepostačuje, preto by sme mohli povedať, že ide o súhrn štatistických techník grafickej a numerickej povahy, ktoré sa používajú na usporiadanie, prezentáciu a analýzu údajov získaných prostredníctvom výskumu (Fisher, Marshall 2008). Dôležité je, že tieto techniky sú výlučne exploračné/deskriptívne, čo znamená, že nezahŕňajú žiadne zovšeobecnenia, ktoré by zachádzali za rámec nastavený prostredníctvom získania údajov meraním. Zhŕňajú, analyzujú, triedia a charakterizujú stav a povahu premenných. Prostredníctvom týchto techník je tak možné získať úplný obraz o údajoch, s ktorými výskumník pracuje.

Prakticky sa ukazuje, že techniky deskriptívnej štatistiky sa využívajú na začiatku štatistického spracovania (ako sme uvádzali v príklade v predchádzajúcej časti tejto učebnice). Pri štatistickom spracovaní výskumných projektov a rôznych vedeckých štúdií (či v prípade študentov záverečných prác) sa techniky deskriptívnej štatistiky môžu prakticky použiť na:

- *opis charakteristík výskumného súboru (napr. rodové, vekové rozdelenie, rozloženie podľa vzdelania, a pod.);*
- *kontrolu premenných, či sú vhodné pre ďalšie štatistické spracovanie a použitie štatistických techník (či sú dáta kompletné, či údaje nie sú skreslené, akú majú povahu, či sa v rámci distribúcie nenachádzajú rôzne extrémy,...);*
- *riešenie konkrétnych výskumných otázok (ak je cieľ výskumu orientovaný výlučne na deskripciu zistení – napríklad pri mapovacom výskume) (Pallant 2007).*

Deskriptívne štatistické metódy poskytujú usporiadanie údajov z výskumu do zmysluplného celku, následne sa ich snažia opísať a charakterizovať, na čo sa používajú:

- **tabuľkové deskriptívne metódy** – ako už zo samotného názvu vyplýva predstavujú tabuľkové zobrazenie jednotlivých hodnôt nameranej premennej. Ich použitie predstavuje pre štatistické spracovanie základ, pretože umožňujú zobrazovať základný prehľad nameraných hodnôt;
- **číselné deskriptívne metódy** – zahŕňajú numerické ukazovatele, ktoré popisujú konkrétne hodnoty vypočítané z údajov. Sú presnejšie a objektívnejšie. Na druhej strane je pri ich interpretácii potrebné vysvetľovanie a pochopenie;

- **grafické deskriptívne metódy** – zahŕňajú grafy a oproti číselným sú ich využitie a interpretácia rýchlejšie. Prostredníctvom nich je možné identifikovať v údajoch rôzne vzory a to na základe grafického zobrazenia, ktoré je pri ilustrovaní údajov veľmi praktické. Nie sú však tak presné ako číselné deskriptívne metódy.

V štatistickej praxi sa často stretávame s otázkou, akú metódu deskriptívnej štatistiky je najefektívnejšie použiť. Jednoduchý princíp tu však neexistuje. Vhodné je jednotlivé metódy kombinovať a to najmä s ohľadom na čo najefektívnejší popis údajov. Tam, kde výskumník potrebuje viac ilustrovať použije grafické zobrazenie, kde je potrebné byť taxatívnejší, vhodnejšie je použiť číselné alebo tabuľkové zobrazenie. *V praxi študentských záverečných prác sa stretávame s tým, že v snahe docieliť čo najväčší rozsah empirickej časti študenti kombinujú tabuľkové, číselné a grafické metódy. Jednoducho povedané, to, čo zobrazia prostredníctvom tabuliek, duplicitne zobrazujú prostredníctvom grafu. To vyvoláva dojem nadbytočnosti a v rámci záverečných prác to neodporúčame. Ak si vezmeme príklad z odborných* textov *v rôznych vedeckých zdrojoch, tam takúto prax nenachádzame a údaje sú zobrazené čo najefektívnejšie s cieľom popísať údaje čo najpresnejšie, najobjektívnejšie a najmä čo najúspornejšie (niekedy sa číselné údaje vpisujú priamo do textu v snahe nepridávať ani len tabuľky). Aj spôsob prezentácie tak patrí k štatistickým „danostiam“ a je potrebné dôsledne ho zvážiť vzhľadom na povahu informácií či cieľ prezentovania.*

Okrem toho, akej povahy sú deskriptívne metódy sa táto oblasť štatistiky delí podľa toho, koľko premenných je predmetom zobrazenia, resp. koľko premenných má výskumník za cieľ prezentovať. Počet premenných sa v tomto kontexte taktiež označuje pojmom „rozmer“ a rozlišujeme tak:

- **jednorozmerné deskriptívne metódy** *(tie ktoré zobrazujú iba jednu premennú, napr. frekvenčná tabuľka, aritmetický priemer, koláčový graf a* pod.),
- **dvojrozmerné deskriptívne metódy** *(zobrazujúce dve premenné, napr. kontingenčná tabuľka),*
- **viacrozmerné deskriptívne metódy** *(zobrazujúce viac ako dve premenné, napr. viaccestná tabuľka).*

V princípe však ide o veľmi podobné metódy, najmä v prípade tabuľkového a grafického zobrazenia. V prípade metód určených pre zobrazenie dvoch a viacerých premenných ide o rovnaké metódy s tým, že sú doplnené o potrebný počet „rozmerov“, teda konkrétnych premenných, ktoré zobrazujú.

4.1 Tabuľkové deskriptívne metódy

Tabuľkové deskritpívne metódy sú časťou popisnej štatistiky, ktorá umožňuje systematizovať, kategorizovať a zobrazovať inokedy ťažko čitateľné informácie v kompaktnej forme prostredníctvom rôznych foriem tabuľkového spracovania. Základnou metódou zobrazenia sú, ako už z názvu vyplýva, tabuľky. V nasledujúcej časti budeme hovoriť o najčastejšie používaných tabuľkových metódach – frekvenčnej, kontingenčnej tabuľke a viaccestnej tabuľke.

Frekvenčná tabuľka – základná metóda deskriptívnej štatistiky

Frekvenčnú tabuľku považujeme za základnú metódu deskriptívnej štatistiky radenú do jednorozmerných deskriptívnych metód (najčastejšie zobrazuje jednu premennú). V prípade zobrazenia viacerých premenných hovoríme o kontingenčnej tabuľke. V prípade štatistickej analýzy by malo ísť o prvý krok, keďže táto metóda dokáže odhaliť chyby v súbore údajov, ktoré by mohli nasledujúce štatistické analýzy ovplyvniť, príp. znehodnotiť.

Frekvenčná tabuľka je metóda organizovania nespracovaných (tzv. „surových“) údajov v kompaktnej podobe zobrazením série hodnôt vo vzostupnom alebo zostupnom poradí spolu s ich frekvenciami (počtami výskytov konkrétneho údaju) v príslušnej množine údajov (Salkind 2010).

Frekvenčná tabuľka ponúka zobrazenie prehľadu početností, relatívnych početností a kumulatívnych početností číselných premenných. Je vhodná pre všetky typy číselných premenných a poskytuje úplný prehľad o nameraných hodnotách. To umožňuje zistiť, ako sú vo výskumnom súbore rozložené namerané hodnoty premennej (predstavuje zistenie charakteru – distribúcie premennej). Pre ilustráciu zobrazenia sa niekedy tabuľky používajú ako východisko pre iné grafické spôsoby prezentácie. Spôsob prípravy frekvenčnej tabuľky je veľmi jednoduchý. V prvom kroku je potrebné jednotlivé hodnoty kvantitatívnej premennej utriediť (v kľúči zvolenom výskumníkom, najčastejšie od najnižšej hodnoty po najvyššiu), následne sa zisťuje pre každú hodnotu jej absolútna početnosť, relatívna početnosť, absolútna kumulatívna početnosť a relatívna kumulatívna početnosť (Hendl 2012). Tieto ukazovatele tvoria súčasť frekvenčnej tabuľky najmä v učebnicových príkladoch. V praxi sa takmer vôbec nestretávame, že by sa v rámci frekvenčnej tabuľky sa zobrazovali všetky 4 ukazovatele početností. Najčastejšie sa zobrazuje absolútna a relatívna početnosť (ako východisko pri výpočte percentuálneho podielu zastúpenia jednotlivých hodnôt). Keďže však účelom tejto vysokoškolskej učebnice je predstaviť základné štatistické princípy a techniky, v nasledujúcej časti si povieme viac o všetkých štyroch ukazovateľoch početností a v príklade si ukážeme aj najčastejšie používanú formu zobrazenia.

Ako príklad opäť použijeme premennú počet rokov v praxi, ktorá bola zisťovaná vo výskume u pracovníčok a pracovníkov v sociálnej oblasti. Do výskumu sa zapojilo 30 respondentov, ktorí boli v praxi od 1 roka do 10 rokov (čo vyjadruje prvý stĺpec). Celkovo prezentuje frekvenčná tabuľka všetky ukazovatele uvedené v predchádzajúcom texte.

Tab. 3 - frekvenčná tabuľka – príklad

Počet rokov v praxi	f	cum f	rel f	cum rel f
1	2	2	0,07	0,07
2	4	6	0,13	0,20
3	3	9	0,10	0,30
4	1	10	0,03	0,33
5	5	15	0,17	0,50
6	1	16	0,03	0,53
7	2	18	0,07	0,60
8	6	24	0,20	0,80
9	3	27	0,10	0,90
10	3	30	0,10	1,00
	absolútna početnosť	**absolútna kumulatívna početnosť**	**relatívna početnosť**	**relatívna kumulatívna početnosť**

Absolútna početnosť predstavuje jednoduché zobrazenie absolútneho počtu respondentov, ktorí v rámci výskumu nadobudli konkrétnu hodnotu (v našom prípade počet rokov praxe). Táto početnosť sa získa jednoduchým sčítaním respondentov (ich odpovedí, dotazníkov,...) v jednotlivých kategóriách. Samotný počet rokov v praxi je v tabuľke zobrazený v vzostupnom poradí, k jednotlivým rokom je potom priradený celkový počet respondentov rozdelených podľa rokov ich praxe.

Prakticky je prvý stĺpec výsledkom jednoduchého „triedenia“. V prípade použitia dotazníkov si môžu študenti vytvoriť príslušný počet „kôpok“ (v tomto prípade 10 - podľa počtu najvyššieho počtu rokov), na ktoré budú triediť jednotlivé dotazníky podľa vpísaného čísla vyjadrujúceho počet rokov v praxi. Absolútna početnosť následne bude počet dotazníkov v jednotlivých kôpkach.

V súčasnej štatistickej praxi sa však so samotným „triedením“ papierových dotazníkov pracuje pomerne zriedkavo. To je možné a efektívne pri malých výskumných súboroch. Pri rozsiahlejších štatistických zisťovaniach sa údaje z jednotlivých dotazníkov prenášajú do tabuľkových procesorov (PC programov, napr. Excel; či iných špecializovaných štatistických programov, napr. SPSS Statistics). Tým vznikne dátový súbor (ide o verný prepis jednotlivých dotazníkov), s ktorým sa pracuje veľmi podobne ako s papierovými dotazníkmi. Výhodou však je, že prostredníctvom týchto programov sa dajú jednotlivé hodnoty veľmi efektívne filtrovať, čím sa minimalizuje ručné spočítavanie odpovedí a tým aj možnosť numerickej chyby zo strany študentov, či výskumníkov.

Absolútna kumulatívna početnosť predstavuje kumulované vyjadrenie početností, jednoduchšie povedané ide o postupný súčet všetkých doposiaľ vyjadrených absolútnych početností. Jej výpočet je možné zrealizovať jednoduchým súčtom absolútnych početností predchádzajúcich a príslušnej hodnoty premennej. Prakticky vyjadruje, ako sú hodnoty premennej rozložené vo výskumnom súbore, keďže pri každej hodnote máme prehľad o doterajšej početnosti (ak vieme, aká je celková početnosť – teda koľko respondentov odpovedalo, vieme povedať to, kde sa v rámci komplexného vyjadrenia pri ktorej hodnote nachádzame).
Zároveň pri absolútnej kumulatívnej početnosti vieme odčítať i celkový rozsah štatistického súboru (veľkosť vzorky), keďže posledná hodnota (po sčítaní všetkých predchádzajúcich početností) vyjadruje práve celkovú početnosť štatistického súboru.

Najjednoduchší spôsob, ako si predstaviť zmysel využitia a spôsob výpočtu absolútnej kumulatívnej početnosti, je jednoduchá stavebnica s kockami. Predstavme si, že máme toľko kociek, aký je početný výskumný súbor. Na základe početností jednotlivých hodnôt budeme stále pridávať na seba príslušný počet kociek. Na konci tejto snahy budeme mať postavenú vežu, kde využijeme všetky pripravené kocky. Využitie takéhoto postupu je najmä pre predstavu toho, aká vysoká je veža vzhľadom na celkový počet kociek. Takto vieme napríklad zistiť, či sa na koncoch hodnôt nenachádza vysoké množstvo respondentov, resp. ako sú jednotlivé hodnoty z hľadiska početností rozložené.

Výpočet z predchádzajúceho príkladu premennej „Počet rokov v praxi“ by vyzeral nasledovne. Pri prvej hodnote, teda 1 roku je absolútna početnosť na úrovni „2“. Keďže ide o prvú hodnotu, nemá sa v tomto prípade s čím sčítať, preto sa do stĺpca absolútnej kumulatívnej početnosti prepíše hodnota „2“. Pri pokračovaní v ďalšom riadku pri hodnote 2 roky je absolútna početnosť „4“ keďže v tu už môžeme sčítať hodnoty s predchádzajúcim riadkom, realizujeme výpočet (2+4=6) a do príslušného stĺpca vpíšeme hodnotu „6“. Prakticky to znamená, že doposiaľ šiesti respondenti mali prax do 4 rokov (2 a 4 roky). V ďalších riadkoch je to rovnaké (6+3=9;...). Na konci celej frekvenčnej tabuľky získame celkovú početnosť (celkový počet respondentov, v našom prípade 30).

Tab. 4 - absolútna kumulatívna početnosť - výpočet

Počet rokov v praxi	f	cum f	rel f	cum rel f
1	2	2		
2	4	6		
3	3	9		

...

10	3	30		

Relatívna početnosť je ďalšie z vyjadrení početnosti, ktoré môže obsahovať frekvenčná tabuľka. V praxi je tento ukazovateľ východiskovým pri prezentovaní percentuálneho podielu počtu odpovedí na jednotlivé možnosti. Taktiež je proporcionálnym ukazovateľom a číselne vyjadruje časť údajov reprezentujúcu pomer konkrétnej hodnoty k celkovému počtu údajov (buď k vzorke, alebo základnému súboru) (Ferjenčík 2006).

Výpočet v tomto prípade (s použitím príkladu „Počet rokov v praxi“ prebieha ako podiel analyzovanej početnosti k celkovej početnosti v každom riadku. Tu vieme, že celková vzorka dosiahla 30 respondentov, preto sa bude i relatívna početnosť počítať práve s týmto číslom. V prvom riadku pri 1 ročnej praxi je absolútna početnosť „2“. Vytvoríme teda podiel (2/30=0,07, čo je v podstate podiel zisťovanej početnosti, t.j. 2 k celkovej početnosti, t.j.30), ktorý zapíšeme do tabuľky, rovnako postupujeme v ďalších riadkoch (4/30=0,13;...).

Tab. 5 - relatívna početnosť – výpočet

Počet rokov v praxi	f	cum f	rel f	cum rel f
1	2		0,07	
2	4		0,13	

Ako sme uviedli, relatívna početnosť je východisko pri zobrazení percentuálneho podielu jednotlivých odpovedí. Percentuálny podiel je pritom zrejme najčastejším ukazovateľom, ktorý sa v prípade bežných frekvenčných tabuliek používa. Je to aj preto, že väčšina populácie sa s nejakým vyjadrením percent stretla aj v bežnom živote a pre ilustráciu a porozumenie je tento ukazovateľ vhodný. V praxi prezentovania štatistických údajov je tak najjednoduchšia frekvenčná tabuľka zostavená výlučne z početností a percentuálneho podielu. Tento percentuálny podiel sa jednoducho získa z relatívnej početnosti a to tak, že každá hodnota relatívnej početnosti sa vynásobí hodnotou „100".

Tab. 6 - výpočet percentuálneho podielu odpovedí

Počet rokov v praxi	f	rel f		%
1	2	0,07	→	7
2	4	0,13	**x 100**	13
3	3	0,10		10
4	1	0,03		3
5	5	0,17		17
6	1	0,03		3
7	2	0,07		7
8	6	0,20		20
9	3	0,10		10
10	3	0,10		10

Relatívna kumulatívna početnosť je ukazovateľ, ktorý kumulatívne vyjadruje relatívne početnosti na rovnakom princípe ako absolútna kumulatívna početnosť. Jej výpočet sa realizuje ako postupný súčet relatívnych početností, ktoré predchádzali príslušnej hodnote. Takýmto spôsobom je možné zistiť proporcionálne rozloženie jednotlivých hodnôt. Na konci frekvenčnej tabuľky by mala relatívna kumulatívna početnosť dosiahnuť hodnotu 1. Rovnako je možné tento ukazovateľ vyjadriť percentuálne tak, ako sme uviedli v prípade relatívnej početnosti.

Výpočet relatívnej kumulatívnej početnosti prebieha prostredníctvom sčítania relatívnych početností. Pri prvej hodnote – 1 rok sa opäť uvedie hodnota relatívnej početnosti (keďže nie je žiadna predchádzajúca hodnota, s ktorou by sa mohla táto hodnota sčítať). Pri ďalšej hodnote (2 roky) sa však už hodnota relatívnej početnosti sčítava s predchádzajúcou hodnotou kumulatívnej relatívnej početnosti (0,07+0,13=0,20) čím dostaneme príslušnú hodnotu relatívnej kumulatívnej početnosti. Takýmto spôsobom postupujeme pri ďalších hodnotách (0,20+0,10=0,30; ...) až po poslednú hodnotu, ktorá by mala dosiahnuť „1“ (v prípade percentuálneho vyjadrenia 100%).

Tab. 7 - relatívna kumulatívna početnosť – výpočet

Počet rokov v praxi	f	cum f	rel f	cum rel f
1			0,07	0,07
2			0,13	0,20
3			0,10	0,30
…				
10			0,10	1,00

<u>*Vytváranie tried vo frekvenčných tabuľkách*</u>

V prípade rozsiahlych štatistických zisťovaní, ale aj v prípade, že výskumník má za cieľ prehľadne prezentovať zistené údaje sa ukazuje, že takéto vyčerpávajúce frekvenčné tabuľky (jednak na počet zobrazovaných frekvencií ale aj na počet jednotlivých hodnôt – všetkých nameraných hodnôt) sú veľmi neprehľadné a zaťažujúce. Navyše, pri samostatnom prezentovaní všetkých jednotlivých hodnôt (ako v príklade v premennej „Počet rokov v praxi“ boli výsledky prezentované samostatne po rokoch) sa stáva, že množstvo hodnôt premennej je príliš veľké a sú sýtené veľmi malými početnosťami, v niektorých hodnotách je početnosť nulová, v niektorých sa nachádza iba malý počet respondentov. Pre sprehľadnenie sa preto pristupuje k vytváraniu tzv. **tried**. Tie vnímame ako kategórie hodnôt premennej s jasne vymedzenými hranicami (od-do), čo sa taktiež označuje pojmom **šírka** triedy (Ferjenčík 2006). V prípade usporiadania do tried je prezentácia hodnôt jednoduchšia, keďže sa počet kategórií zredukuje, čím je zobrazenie prehľadnejšie a omnoho jednoduchšie. Nevýhodou, najmä pri veľkej šírke triedy môže byť strata dôležitých údajov (napr. rozdielov), preto je dôležité voliť šírku tried s ohľadom na účel prezentovaných výsledkov.

Ak by sme použili príklad s premennou „Počet rokov v praxi“, mohli by sme triedy vytvoriť viacerými spôsobmi s ohľadom na účel prezentovania údajov. Najjednoduchšie by bolo vytvoriť 2 triedy – v rovnakej šírke 5 rokov. Úplná frekvenčná tabuľka by v takomto prípade vyzerala nasledovne:

Tab. 8 - vytváranie tried – príklad I.

Počet rokov v praxi	f	cum f	rel f	cum rel f
1-5	15	15	0,5	0,5
6-10	15	30	0,5	1,0

Pre ilustratívne prezentovanie dlhšej a kratšej praxe je takáto tabuľka postačujúca. Jednotlivé triedy sú dokonca i vyrovnané početnosťou. Ak sa však pozrieme na pôvodnú frekvenčnú tabuľku, zistíme, že tabuľka s dvoma triedami nedokáže zobraziť nerovnomerné vyváženie dĺžky praxe, ktoré sa v triedach nachádza (napr. až tretina respondentov v prvej zobrazenej triede má dĺžku praxe 5 rokov a podobné ukazovatele), ktoré by mohlo prezentované výsledky do istej miery negatívne ovplyvniť. Ak by sme chceli prezentovať výsledky detailnejšie, môžeme zvoliť 5 tried so šírkou 2 roky. V tom prípade by frekvenčná tabuľka vyzerala takto:

Tab. 9 - vytváranie tried – príklad II.

Počet rokov v praxi	f	cum f	rel f	cum rel f
1-2	6	6	0,20	0,20
3-4	4	10	0,13	0,33
5-6	6	16	0,20	0,53
7-8	8	24	0,27	0,80
9-10	6	30	0,20	1,00

Samozrejme, to, ktorú alternatívu výskumník zvolí závisí od cieľa prezentovania výsledkov i charaktereu výskumných zistení a rozhodnutí samotného výskumníka. Najčastejšie sa v prácach študentov ale i vo výskumných štúdiách stretávame s kombinovanými tabuľkami s absolútnou početnosťou, percentuálnym podielom a konečnou absolútnou kumulatívnou početnosťou (vyjadrenú ako sumu hodnôt). Príklady tabuliek z dvoch a piatich tried sú zobrazené nižšie.

Tab. 10 - vytváranie tried s vyjadrením percentuálneho podielu – príklad 2 tried

Počet rokov v praxi	f	%
1-5	15	50
6-10	15	50
∑	30	100

Tab. 11 - vytváranie tried s vyjadrením percentuálneho podielu – príklad 5 tried

Počet rokov v praxi	f	%
1-2	6	20
3-4	4	13
5-6	6	20
7-8	8	27
9-10	6	20
∑	30	100

Samozrejme, vytváranie tried je efektívne u premenných číselnej povahy (teda intervalových a poradových). V prípade nominálnych premenných je vytváranie tried síce čiastočne možné, no vo väčšine prípadov neefektívne (a mnohokrát i nemožné). Je to najmä preto, že z nečíselnej povahy premenných nie je možné stanoviť šírku tried s presne vymedzenými hranicami.

Príkladom pre frekvenčné zobrazenie nominálnych premenných by mohlo byť rozdelenie respondentov podľa pracovnej pozície, ktorú v práci zastávajú. Premenná pracovná pozícia spĺňa všetky kritériá pre nominálnu premennú, jej základná frekvenčná tabuľka môže byť rovnaká, ako frekvenčná tabuľka číselných premenných. Pre tento príklad použijeme výskum, ktorého sa zúčastnilo 50 respondentov pracujúcich v rôznych zariadeniach sociálnych služieb.

Napriek tomu, že frekvenčná tabuľka môže byť aj v tomto prípade rovnaká, praktický nemá žiaden praktický význam radiť do takéhoto prehľadu relatívnu kumulatívnu početnosť a to preto, že poradie jednotlivých hodnôt nezohráva žiaden význam – je na výskumníkovi, ktorú hodnotu bude prezentovať ako prvú, druhú a pod. V praxi sa najefektívnejšie ukazuje, keď sa poradie možností odvíja od početnosti – napr. od najpočetnejšej po najmenej početnejšiu.

Tab. 12 - frekvenčná tabuľka pre nominálne premenné

Pracovná pozícia	f	cum f	rel f
opatrovateľ	18	18	0,36
ošetrovateľ	15	33	0,30
sociálny pracovník	7	40	0,14
terapeut	7	47	0,14
psychológ	3	50	0,06

Aj z prezentovaného prehľadu vyplýva, že triedy ako také sa z nominálnej premennej nebudú dať vytvoriť. Za určitých okolností by bolo možné spojiť niektoré z prezentovaných profesií podľa ich charakteru a spojiť ich do spoločných kategórií. To, ktoré by sa takto dali spojiť je na rozhodnutí výskumníka, no musí to zároveň aj teoreticky odôvodniť. Nemôže to byť iba svojvoľné spojenie za účelom „aby bolo dosť respondentov v každej skupine" (čo je častý jav u študentov). V prípade prezentovanej premennej „Pracovná pozícia" by mohol výskumník postupovať nasledovnou úvahou: ak si analyzujeme podmienky a požiadavky vykonávania jednotlivých pracovných pozícií v zariadeniach sociálnych služieb, z povahy samotných profesií, najviac kontaktu s klientom vr. pomoci so sebaobslužnými činnosťami a pod. majú profesie ošetrovateľ a opatrovateľ, ktorých by sme mohli nazvať pracovníkmi denného kontaktu. Charakter výkonu profesie sociálny pracovník, psychológ a terapeut by v zariadeniach mohol byť podobný (samozrejme, nemáme na mysli, že títo pracovníci robia to isté, každý z nich vykonáva pridelené pracovné činnosti vo svojej odbornosti, máme na mysli charakter kontaktu s klientmi), preto by sme ich teoreticky mohli spojiť do spoločnej kategórie, ktorú nazveme „odborní pracovníci". Ak si takýmto spôsobom dokáže výskumník odôvodniť spojenie jednotlivých hodnôt (upozorňujeme, že toto je iba učebnicový príklad a v žiadnom prípade netvrdíme, že uvedené je univerzálne použiteľná kategória), môže frekvenčná tabuľka vyzerať napríklad takto:

Tab. 13 - frekvenčná tabuľka pre nominálne premenné – 2 kategórie

Pracovná pozícia	f	cum f	rel f
Pracovníci denného kontaktu	33	33	0,66
Odborní pracovníci	17	50	0,34

S takýmito kategóriami je následne možné v štatistike ďalej pracovať podľa toho, aký je účel prezentácie štatistických výsledkov. Rovnako je opätovne dôležité poukázať na to, že typ použitej frekvenčnej tabuľky, či údaje v nej zobrazené je na rozhodnutí výskumníka, ktorý zvažuje všetky spôsoby prezentácie. Základnou zásadou je pritom, aby boli prezentované výsledky čo najjasnejšie a najjednoduchšie komunikované. Aj z tohto pohľadu sú v rámci prezentovania preferované jednoduchšie tabuľky spojené do zmysluplných tried so zobrazením absolútnej početnosti a percentuálneho podielu.

Kontingenčná tabuľka

Kontingenčná tabuľka je tabuľkový spôsob zobrazenia numerických dát, ktorý je príznačný pre dvojrozmernú deskriptívnu štatistiku, teda prezentuje hodnoty dvoch premenných. Je to tzv. „**krížová tabuľka**“. Každá bunka tejto tabuľky najčastejšie prezentuje početnosť príslušnej kombinácie premenných v riadkoch a stĺpcoch (Rimarčík 2007). Okrem početnosti môže byť doplnená o ďalšie deskriptívne ukazovatele (relatívna početnosť, percentuálny podiel,...).

Praktický príklad využitia takejto tabuľky by mohol byť pri prezentovaní dvoch doposiaľ používaných premenných, počet rokov v praxi a pracovná pozícia. V tomto prípade využijeme premenné po úprave – spojení do tried a kategórií tak, ako boli prezentované v predchádzajúcej časti. Pracujeme teda s premennou počet rokov v praxi, ktorá je rozdelená do 5 tried a premennou pracovná pozícia rozdelená do dvoch kategórií. Vytvorenie kontingenčnej tabuľky je tak otázkou pospájania a prepočítania jednotlivých kombinácií. V tomto prípade budeme pracovať s výskumom, ktorý zahŕňa 30 respondentov. Samotná tabuľka môže vyzerať rôzne, je na výskumníkovi, ktorú premennú bude prezentovať v stĺpcoch a ktorú v riadkoch. Jeden z príkladov spracovania by mohol vyzerať nasledovne:

Tab. 14 - kontingenčná tabuľka – príklad použitia

		Počet rokov praxe				
		1-2	3-4	5-6	7-8	9-10
Pracovná pozícia	Odborní pracovníci	4	1	3	5	1
	Pracovníci denného kontaktu	2	3	3	3	5

TIP: Zostavenie kontingenčnej tabuľky sa realizuje buď ručným usporiadaním údajov a vytvorením tabuľky podľa zobrazených zásad, alebo s využitím technológie – prostredníctvom počítačových programov.

V programe Excel je možné využiť automatické vloženie kontingenčnej tabuľky: Vložiť → Kontingenčná tabuľka, kde je potrebné vybrať rozsah hodnôt, ktoré chceme prostredníctvom kontingenčnej tabuľky analyzovať.

V štatistickom programe SPSS je vloženie tabuliek možné po zadaní Analyze → Tables → Custom Tables. Tu v novom okne vyberieme, ktoré premenné chceme použiť ako riadkové a ktoré ako stĺpcové. Po odsúhlasení (OK) sa tabuľka zobrazí v novom okne (Output). Rovnako je možné použiť funkciu „Crosstabs“ v Analyze → Descriptive Statistics.

Viaccestná tabuľka

Viaccestná tabuľka je ďalšou tabuľkovou deskriptívnou metódou využívanou pri prezentovaní viac ako dvoch premenných. V tomto prípade ide buď o rozšírenie kontingenčnej tabuľky (pridanie ďalších kontingenčných tabuliek) alebo vytvorenie samostatnej tabuľky, ktorá by obsahovala viacero premenných a ich vzájomné početnosti, alebo iné ukazovatele (percentuálny podiel, aritmetické priemery,...).

Ako príklad použijeme opäť premenné „Pracovná pozícia“ a „Počet rokov v praxi“, ku ktorým pridáme ďalšiu premennú (nominálnej povahy) a to „Rod“ respondentov. Po prispôsobení a prepočítaní by mohla viaccestná tabuľka vyzerať napríklad takto:

Tab. 15 - viaccestná tabuľka – príklad

		Rod	Počet rokov v praxi				
			1-2	3-4	5-6	7-8	9-10
Pracovná pozícia	Odborní pracovníci	muži	1	0	2	1	1
		ženy	3	1	1	4	0
	Pracovníci denného kontaktu	muži	1	1	2	0	2
		ženy	1	2	1	3	3

Cvičenie

V tabuľke nižšie je k dispozícii dátový súbor (odpovede respondentov) 30 klientov v seniorskom veku, ktorí odpovedali na dotazník kvality života. Sú to klienti, ktorí boli oslovení v rôznych zariadeniach sociálnych služieb. Okrem toho, v dotazníku sa výskumníci zaujímali o ich rod, vek, stupeň odkázanosti.

Preštudujte si dátový súbor a odpovedzte na otázky pod tabuľkou.

Tab. 16 - dátový súbor pre vytvorenie frekvenčných tabuliek

Poradové číslo respondenta	Rod	Vek	Stupeň odkázanosti klienta	Typ inštitúcie, v ktorej je klient posudzovaný	Skóre dotazníka kvality života
1	muž	67	IV	ADOR	14
2	muž	62	IV	komunitné centrum	28
3	žena	71	IV	denný stacionár	19
4	muž	69	III	denný stacionár	11
5	žena	83	II	komunitné centrum	15
6	žena	74	III	komunitné centrum	19
7	žena	68	IV	ADOR	14
8	žena	90	V	DSS	20
9	muž	76	V	DSS	22
10	muž	70	V	DSS	19
11	žena	71	IV	ADOR	20
12	muž	65	IV	ADOR	15
13	žena	92	V	DSS	11
14	žena	84	VI	DSS	21
15	žena	75	VI	DSS	29
16	muž	69	VI	DSS	25
17	muž	75	VI	DSS	14
18	muž	69	V	DSS	23
19	žena	74	V	DSS	18
20	žena	69	III	denný stacionár	19
21	žena	83	III	denný stacionár	21
22	žena	76	V	DSS	25
23	žena	80	VI	DSS	18
24	muž	69	VI	DSS	20
25	muž	73	V	DSS	21
26	žena	86	V	DSS	20
27	žena	77	V	DSS	14
28	muž	80	IV	ADOR	11
29	žena	83	IV	ADOR	21
30	muž	79	IV	komunitné centrum	13

Otázky a úlohy:

1. Identifikujte typy premenných v dátovom súbore.
2. Zostrojte samostatné frekvenčné tabuľky pre všetky nominálne a poradové premenné, ktoré sa v dátovom súbore nachádzajú.
3. Navrhnite spôsob vytvorenia tried v intervalovej premennej nachádzajúcej sa v dátovom súbore.
4. Zostrojte frekvenčnú tabuľku pre intervalovú premennú po vytvorení tried.
5. Vytvorte krížovú tabuľku pre premenné „Skóre dotazníka kvality života“ a „Typ inštitúcie, v ktorej je klient posudzovaný“ /môžete vytvoriť krížové tabuľky i pre iné kombinácie premenných/. Uvažujte, či krížová tabuľka môže vysvetliť rôzne vzťahy medzi premennými, ak áno, aké?
6. Zostrojte viaccestnú tabuľku z ľubovoľných premenných v dátovom súbore.

4.2 Číselné deskriptívne metódy

Číselné deskriptívne metódy definujeme ako súbor ukazovateľov charakterizujúcich výskumné dáta prostredníctvom konkrétnych čísel. Tie numericky charakterizujú štatistický súbor a sú vypočítané z konkrétnych hodnôt skúmanej premennej. Rovnako sú označované aj pojmom opisné /deskriptívne/ charakteristiky (Rimarčík 2007). Číselné deskriptívne metódy sa počítajú prostredníctvom vzorcov z konkrétnej premennej, preto sú súčasťou jednorozmernej deskriptívnej štatistiky.

Z hľadiska použitia v deskriptívnej štatistike rozlišujeme:

- ukazovatele centrálnej tendencie (rovnako označované ako miery polohy, alebo miery strednej hodnoty);
- ukazovatele rozptýlenosti (variability);
- ukazovatele symetrie (tiež ukazovatele tvaru distribúcie, konkrétne šikmosti a špicatosti) (napr. Hendl 2012; Rimarčík 2007; Kanderová, Úradníček 2007; Sodomová 2013).

V nasledujúcej kapitole postupne charakterizujeme jednotlivé ukazovatele, spôsob ich použitia a interpretácie v deskriptívnej štatistike.

Ukazovatele centrálnej tendencie

Čísla, ktoré reprezentujú centrálnu tendenciu, sa v dátach snažia nájsť tzv. „**typickú hodnotu**" – teda pomyselnú stredovú hodnotu na číselnej osi, okolo ktorej sú údaje skúmanej premennej vo výskumnom súbore rozložené. Stredových ukazovateľov poznáme viacero, pričom každý z nich reprezentuje iné chápanie stredovej hodnoty. Pri každom z nich si teda definujeme, ktorú stredovú hodnotu daný ukazovateľ reprezentuje.

Aritmerický priemer

Aritmetický priemer ako najznámejší ukazovateľ stredu je hodnota, ktorá predstavuje geometrický stred – ťažisko všetkých údajov premennej. Ťažiskom sa nazýva preto, že ide o rovnosť súčtov hodnôt premennej na oboch stranách aritmetického priemeru. V praxi to znamená, že súčet hodnôt premennej pod priemernou hodnotou je totožný so súčtom hodnôt premennej nad priemernou hodnotou (Ferjenčík 2006; Hendl 2012).

Ak by sme si predstavili pomyselné váhy, na ktoré sa kladú závažia, aritmetický priemer je hodnota, kedy by boli váhy presne v rovnováhe. Každý jeden údaj sa podieľa na konštrukcii aritmetického priemeru svojou hodnotou.

Čisto matematicky, ak by sme urobili súčet všetkých rozdielov (odchýlok) jednotlivých nameraných hodnôt od hodnoty aritmetického priemeru, dostali by sme hodnotu 0 (Hendl 2012; Coolican 2014)

Výpočet:
Aritmerický priemer sa vypočíta ako súčet všetkých nameraných hodnôt vydelený počtom týchto hodnôt.

Matematicky sa tento výpočet dá vyjadriť vzorcom:
$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ alebo v zjednodušenom tvare $\overline{x} = \frac{x_1+x_2+x_3+\cdots+x_n}{n}$

Označenie:
Ako vyplýva z uvedeného, aritmetický priemer sa najčastejšie označuje symbolom $\overline{x}$ pokiaľ ide o aritmetický priemer výberu (konkrétnej skupiny). Pokiaľ ide o aritmetický priemer populácie, označuje sa symbolom μ, pričom princíp výpočtu sa nemení (s ohľadom na celú populáciu sa iba zvyšuje počet opakovaní).
V rôznych zdrojoch sa však môžeme stretnúť i s iným označením, a to M (z angl. „mean“).

Aritmetický priemer ako najpoužívanejší ukazovateľ stredu má zo svojej povahy množstvo výhod, no na druhej strane sa nájdu i nevýhody, pre ktoré je potrebné neustále dôsledne zvažovať pravý zmysel jeho hodnoty. Ten sa totiž v určitých prípadoch stráca (ak je napr. ovplyvnený extrémnymi hodnotami premennej), čo môže viesť k dezinterpretáciám informácií o samotnom výskumnom súbore.

Medzi *výhody aritmetického priemeru* najčastejšie radíme jeho citlivosť, pretože je najcitlivejším ukazovateľom centrálnej tendencie. Stredová hodnota sa nachádza presne v matematickom strede súboru údajov, čo spôsobuje to, že sa kladné a záporné rozdiely smerom k priemeru zrušia (výsledok sčítania kladných a záporných hodnôt smerom k priemeru je nulový). Aritmetický priemer sa najčastejšie používa pri tzv. odhade parametrov, ďalej i pri testovaní rozdielov medzi priemermi niekoľkých populačných skupín. Na druhej strane, hodnota aritmetického priemeru je častokrát úplne odlišná ako ktorákoľvek hodnota vo výskumnom súbore, čo môže spôsobiť praktické problémy pri jeho interpretácii. To je výrazné najmä pri existencii extrémnych (nízkych alebo vysokých) hodnôt premennej. Práve to sa prakticky ukazuje ako *nevýhoda aritmetického priemeru*, keďže jedna extrémna hodnota v jednom smere dokáže negatívne ovplyvniť hodnotu aritmetického priemeru, ktorý už adekvátne nereprezentuje stredovú hodnotu. Tu je nutné povedať, že pokiaľ sa takéto hodnoty objavujú v oboch smeroch, hodnoty sa eliminujú a táto nevýhoda sa stráca.

V odbornej literatúre (napr. Hendl 2012; Coolican 2014) sa môžeme stretnúť i s konkrétne vyšpecifikovanými podmienkami, kedy je možné použiť aritmetický priemer (ako i ďalšie ukazovatele centrálnej tendencie). Tie sa týkajú charakteru premennej (minimálnou podmienkou je intervalová premenná, pri poradových a nominálnych nemá aritmetický priemer praktický význam) a konkrétnych hodnôt premennej (rozdelenie hodnôt by malo byť symetrické – bez existencie extrémnych hodnôt – viac o symetrickosti v ďalšej časti).

Príklad: *Úlohou výskumníka je identifikovať priemerný čas poskytovania sociálneho poradenstva klientom v konkrétnej organizácii poskytujúcej túto sociálnu službu. Keďže zadanie tejto výskumnej úlohy bolo identifikovať čas poskytovania poradenstva všetkými poradcami bez rozdielu vykonávaného poradenstva, zaradil výskumník do priemeru všetkých pracovníkov (teda tých, ktorí vykonávajú prvý kontakt s klientom, základné sociálne poradenstvo a rovnako i špecializované sociálne poradenstvo). Konkrétny čas poradenstva poradcov v sledovanom dni bol nasledovný.*

Tab. 17 - aritmetický priemer – čas poskytovania poradenstva

Pracovníci	**Čas v minútach**
Pracovník 1	10
Pracovník 2	10
Pracovník 3	65
Pracovník 4	70
Pracovník 5	85
Pracovník 6	80
Pracovník 7	90

Výpočet aritmetického priemeru je v tomto prípade po dosadení do vzorca jednoduchý, konkrétne:

$$\bar{x} = \frac{10+10+65+70+85+80+90}{7} = 59$$

Odpoveď výskumníka na zadanú úlohu by teda bola, že priemerný čas poskytovania sociálneho poradenstva pracovníkmi organizácie je 59 minút (po zaokrúhlení na celé minúty).

Pri praktickej interpretácii tejto hodnoty by sme však mohli konštatovať, že dvaja pracovníci poskytujú poradenstvo výrazne kratšie, zatiaľ čo ďalší výrazne dlhšie ako je priemer. Preto je na mieste zamyslieť sa, či hodnota aritmetického priemeru zodpovedá tomu, čo bolo vlastne potrebné identifikovať.

Výskumník pre tento účel doplnil k pracovníkom typ poskytovaného poradenstva, keďže sa domnieva, že by mohlo ísť o nepriamu mätúcu premennú.

Tab. 18 - aritmetický priemer – čas a typ poskytovaného poradenstva

Pracovníci	Typ poradenstva	Čas v minútach
Pracovník 1	Prvý kontakt	10
Pracovník 2	Prvý kontakt	10
Pracovník 3	Základné	65
Pracovník 4	Základné	70
Pracovník 5	Špecializované	85
Pracovník 6	Špecializované	80
Pracovník 7	Špecializované	90

Po identifikovaní ďalšej premennej, a to typu poskytovaného poradenstva, by mohol výskumník dospieť k relevantnejším záverom. V tomto prípade by lepší prehľad poskytoval výpočet nie jedného, ale troch aritmetických priemerov, pre každý typ poskytovaného poradenstva (prvý kontakt = 10 minút; základné = 68 minút; špecializované = 85 minút). Tieto výsledky poskytujú pre účely interpretácie adekvátnejší prehľad o časoch poskytovaného poradenstva.

Nie vo všetkých prípadoch je však možné zmeniť náhľad prostredníctvom zohľadnenia nepriamych premenných, keďže údaje o nich nemusia byť k dispozícii, alebo jednoducho neexistuje premenná, ktorá by mohla takto pôsobiť. Napriek tomu sa však vo výskumnom súbore vyskytujú hodnoty na jednej strane hodnôt premennej (príliš nízka alebo príliš vysoká hodnota), ktoré môžu negatívne ovplyvniť hodnotu premennej. V takom prípade môžeme siahnuť po niektorej z ***iných foriem výpočtu priemeru****, ktoré uvádzame v ďalšej časti.*

TIP: so spôsobom výpočtu aritmetického priemeru sa stretávame na dennej báze. Pri menších počtoch hodnôt nie je problém tento výpočet realizovať aj ručne. Pri vyšších počtoch hodnôt je efektívnejšie využiť niektorý z matematických, štatistických, alebo tabuľkových programov.

V programe Excel – predpokladáme, že údaje sú v tabuľkovom procesore Excel pripravené na použitie, čo znamená, že hodnoty premennej uvádzané jednotlivými respondentmi sú v stĺpci pod sebou (teda napr. hodnoty premennej „čas v minútach“ sa nachádzajú v stĺpci označenom „B“ a konkrétny hodnoty sa nachádzajú v riadkoch č. 1-7. V takomto prípade je možné využiť priamy zápis vzorca:

=(B1+B2+B3+B4+B5+B6+B7)/7,

alebo vopred pripravenú funkciu:

=AVERAGE(B1:B7),

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých aritmetický priemer počítame.

V štatistickom programe SPSS je aritmetický priemer súčasťou viacerých analýz. Opätovne predpokladajme, že dátový súbor je pripravený – hodnoty jednotlivých respondentov sú uvedené v jednom stĺpci. Výpočet aritmetického priemeru ponúkajú analýzy s názvami „Frequencies“, „Descriptives“, či „Explore“ v záložke Analyze → Descriptive Statistics. Aritmetický priemer je tu označený ako „Mean“.

Redukovaný priemer

Redukovaný priemer označovaný taktiež pojmami ***orezaný aritmetický priemer*** alebo ***trimmed mean*** čiastočne rieši problém extrémnych hodnôt na oboch koncoch premennej. Hodnota redukovaného priemeru sa vypočíta z hodnôt premennej po redukcii (odstránení) extrémnych (najvyšších a najnižších) hodnôt v súbore. Počet odstránených hodnôt musí byť z každej strany vyvážený, pričom odstraňuje sa najviac 25% z každej strany. Po tejto redukcii sa vypočíta aritmetický priemer klasickým spôsobom zo zostávajúcich hodnôt premennej. Predpokladom pre výpočet redukovaného priemeru je utriedenie hodnôt buď vzostupným alebo zostupným spôsobom.

Výpočet redukovaného priemeru rieši problém s extrémnymi hodnotami, avšak iba čiastočne. Tým, že sa odstráni 25% hodnôt premennej z každej strany hodnôt sa samotný výpočet redukuje obvykle na polovicu hodnôt. Tu sa strácajú niektoré výhody klasického aritmetického priemeru, najmä pokiaľ ide o to, že ten zohľadňuje každú jednu hodnotu premennej. Pri malých výskumných súboroch, ako je napríklad ten z predchádzajúceho príkladu je 25% redukovanie problematické (v tomto príklade je možné odstrániť najviac jednu hodnotu v každom smere) a po ich odstránení (konkrétne by išlo o hodnoty 10 a 90) získame hodnotu redukovaného priemeru opätovne skreslenú (hodnota 62 je iba *nepatrne vyššia než bola hodnota aritmetického priemeru) a tá taktiež nemusí interpretačne zodpovedať cieľu. Pri väčších výskumných súboroch to však môže byť efektívnejšie.*

<u>*TIP:*</u> aj v prípade redukovaného priemeru je možné postupovať manuálne, teda utriediť jednotlivé hodnoty premennej zostupne alebo vzostupne a vypočítať aritmetický priemer zvyšných hodnôt po odstránení príslušného počtu hodnôt. Aj v tomto prípade však môžeme počítačové programy.

<u>Excel</u> – rovnako ako v prípade aritmetického priemeru môžeme využiť priamu formu vzorca, do ktorého neuvádzame redukované hodnoty

=(B2+B3+B4+B5+B6)/5,

alebo vopred pripravenú funkciu:

=TRIMMEAN(B1:B7;25%)

pričom v zátvorke sa uvádza pôvodný rozsah hodnôt, a za bodkočiarkou percentuálne vyjadrenie počtu orezaných hodnôt.

<u>*Geometrický priemer*</u>

Je ďalším ukazovateľom centrálnej tendencie, ktorý sa snaží identifikovať stredovú – priemernú hodnotu premennej. Často sa používa ako alternatíva k aritmetickému priemeru, ktorá eliminuje vplyv extrémnych hodnôt premennej na samotný priemer. Je možné ho použiť pri tzv. pozitívne zošikmených distribúciách premennej (kedy sa vpravo od priemeru nachádzajú viac vzdialenejšie hodnoty a väčšina hodnôt sa nachádza vľavo od priemeru – viac o ukazovateľoch šikmosti a špicatosti v ďalšej časti). Základným princípom pre výpočet geometrického priemeru je úvaha, že priemerná hodnota by mohla byť skôr súčinom ako súčtom jednotlivých hodnôt. Všetko uvedené by sme mohli zaradiť k *výhodám,* pre ktoré sa tento priemer používa. Zároveň je dôležité zvážiť najmä typ premennej a jej hodnoty. Matematicky je výpočet tohto priemeru obmedzený na kladné a nenulové hodnoty (ak sa niektorá z hodnôt premennej rovná nule aj geometrický priemer bude nulový, ak je niektorá z hodnôt záporná, nedá sa odmocňovať).

Výpočet:

Geometrický priemer sa vypočíta ako n-tá odmocnina súčinu všetkých hodnôt, matematicky:

$$\overline{x}_G = \sqrt[n]{x_1 . x_2 . x_3 . \dots . x_n}$$

Označenie:

Geometrický priemer označujeme podobne ako aritmetický priemer s pridaním indexu označujúceho geometrický priemer: $\overline{x}_G$

Príklad I. *Ak využijeme zadanie z predchádzajúcich strán, teda výskum priemerného času poskytovania poradenských služieb, môžeme uvažovať o tom, ako by priemerná hodnota poradenského času vyzerala s použitím geometrického priemeru. V tomto prípade by konkrétny výpočet vyzeral nasledovne:* $\bar{x}_G = \sqrt[7]{10.10.65.70.85.80.90} = 43$. *V tomto prípade je na výskumníkovi, ako by zvládol túto situáciu interpretačne – na hodnotu priemeru majú nízke hodnoty vplyv, preto tento priemer čiastočne zohľadnil. Je však geometrický priemer 43 skutočnou priemernou hodnotou? V súvislosti s konkrétnymi hodnotami premennej je táto hodnota skutočne reálnejšia. Interpretácia by však aj v tomto prípade bola zložitá a to najmä s ohľadom na to, že ide o rôzne typy vykonávanej činnosti. Preto by aj v tomto prípade bola vhodnejšia samostatná interpretácia prostredníctvom jednotlivých typov poradenstva (v tomto prípade by boli hodnoty geometrického priemeru 10; 67 a 85 čím sa značne približujú k hodnotám aritmetického priemeru).*

Príklad II. *Iný príklad použitia geometrického priemeru, kde nám prispieva k lepšiemu porozumeniu stredovej hodnoty, by mohol byť pri hodnotách premennej, ktoré vyjadrujú čas riešenia konkrétnej úlohy klientmi. Predstavme si, že konkrétne hodnoty tejto premennej by boli 1, 2, 3, 10. Ako vidíme, v súbore sa nachádza jedna hodnota, ktorá je omnoho vyššia ako ostatné (v tomto prípade extrémna hodnota), čo ovplyvní i výpočet priemeru. Aritmetický priemer by v tomto prípade mal hodnotu* $\bar{x} = 4$. *Táto hodnota je iná ako ktorákoľvek hodnota v súbore, zároveň ju extrémna (najvyššia) hodnota ovplyvnila a pri interpretácii by mohlo dôjsť k problémom. Preto je vhodné zamyslieť sa nad výpočtom alternatívneho priemeru. Geometrický priemer by v tomto prípade vyzeral* $\bar{x}_G = \sqrt[4]{1.2.3.10} = 2{,}8$. *Z hľadiska ďalšej interpretácie ako priemerného času riešenia konkrétnej úlohy klientmi považujeme túto hodnotu za prijateľnejšiu. Samozrejme, to, akú hodnotu výskumník použije ostáva na zvážení aj z hľadiska účelu prezentácie.*

TIP: počítačové programy pre analýzu údajov je možné využiť aj v prípade výpočtu geometrického priemeru, pričom aj tu je možné postupovať viacerými spôsobmi. Uvedené vysvetlíme na pôvodnom príklade s časom poskytovaného poradenstva.

Excel – prvou možnosťou je postupovať ako pri manuálnom výpočte, kde zohľadníme všetky matematické špecifiká výpočtu tohto priemeru. S ohľadom na to, že v programe Excel pristupujeme k výpočtu odmocním ako zlomku mocnín, musíme uviesť nasledovný vzorec:

=POWER((B1*B2*B3*B4*B5*B6*B7);/1/7),

kde funkcia POWER znamená mocnina a hodnota za bodkočiarkou (1/7) hovorí o mocnine 1/7 čo v podstate znamená 7 odmocninu.

Rovnako je aj v tomto prípade možné použiť vopred nastavenú funkciu, ktorej zápis je nasledovný:

=GEOMEAN(B1:B7)

pričom v zátvorke sa uvádza pôvodný rozsah hodnôt.

V predchádzajúcom prehľade sme uviedli 3 typy priemerov, ktoré považujeme za najčastejšie použiteľné aj v oblasti sociálnej práce. Samozrejme, uvedomujeme si existenciu ďalších spôsobov výpočtu priemernej hodnoty, kde patrí *harmonický priemer*, ktorý sa najčastejšie používa v situáciách kedy sa výskumník domnieva, že existuje nepriama súvislosť medzi hodnotami štatistického znaku a výsledným identifikovaným javom a je v podstate reverznou hodnotou aritmetického priemeru z reverzných hodnôt štatistického znaku (v programe Excel funkcia HARMMEAN) (Kulčár 2007); *kubický priemer*; *kvadratický priemer* a ďalšie. Z hľadiska praxe sú však tieto stredové hodnoty pomerne špecifické pre konkrétne oblasti vedy a v oblasti sociálnych vied sa štandardne nepoužívajú.

Osobitnou kategóriou je *vážený priemer*. Ten sa najčastejšie počíta v prípadoch, kedy sú jednotlivé hodnoty usporiadané do radov s tým, že viacnásobný výskyt je priamo označený. V tom prípade sa viacnásobný výskyt eviduje ako násobok hodnoty a jej výskytu. Vážené formy priemerov sa dajú počítať pri všetkých uvedených typoch priemerov so zohľadnením ich matematickej povahy.

Ďalším príkladom pre výpočet váženého priemeru je situácia, kedy sú rôzne hodnoty, z ktorých sa priemer počíta, rôzne dôležité. V tom prípade je každej hodnote priradená váha podľa jej dôležitosti a priemer sa počíta po zohľadnení týchto váh (vynásobením samotnej hodnoty jej váhou).

Medián

Medián je jeden z ukazovateľov centrálnej tendencie, ktorý identifikuje presnú stredovú hodnotu, ktorá nie je ovplyvnená extrémnymi hodnotami premennej. Je to v podstate miesto v strede radu hodnôt premennej, ktoré ich delí na dve rovnako početné polovice. Identifikuje tak presný stred hodnôt, ktoré sú predtým zoradené vzostupne, alebo zostupne.

Ak si tento číselný rad predstavíme, 50% hodnôt premennej sa bude nachádzať pred hodnotou mediánu, ďalších 50% sa bude následne nachádzať nad hodnotou mediánu. Zdôrazňujeme však, že ide o počet jednotlivých hodnôt (teda konkrétnych čísel). Dôležité je upozorniť, že v prípade opakujúcich sa hodnôt sa táto hodnota do číselného radu uvádza viackrát podľa počtu opakovaní. To sa stáva častým problémom pri výpočte mediánu z dát uvedených vo frekvenčnej tabuľke (kde sa k hodnotám udáva početnosť). Častou chybou je to, že výskumník pri interpretácii nezohľadní početnosť hodnôt.

Výpočet:
Výpočet mediánu je veľmi jednoduchý, pred samotným výpočtom je však nutné utriediť konkrétne hodnoty premennej, a to vzostupne, alebo zostupne. Samotný výpočet mediánu sa potom vzťahuje na určenie pozície konkrétnej hodnoty v súbore, ktorá reprezentuje medián.

V prípade nepárneho počtu hodnôt sa medián nachádza na pozícii strednej hodnoty, ktorá sa vypočíta: $r = \frac{n+1}{2}$
V prípade párneho počtu hodnôt sa pozícia mediánu vypočíta ako aritmetický priemer hodnôt na pozíciách $\frac{n}{2}$ a $\frac{n}{2} + 1$, matematicky: $r = \frac{1}{2}(x\frac{n}{2} + (x\frac{n}{2} + 1))$
Označenie:
Medián sa najčastejšie označuje dvomi spôsobmi, a to buď ako „$\tilde{x}$“, alebo označením „Me$_{(x)}$“, alebo „Med$_{(x)}$“ čo vyplýva z jeho anglického označenia „median“.
Pozícia, na ktorej sa medián v súbore nachádza, sa označuje ako „r“.

Príklad I. *Výskumník sa zameriava na poskytovanú sociálnu prácu konkrétnymi sociálnymi pracovníkmi v organizácii. Ako jednu z premenných, ktorá môže ovplyvniť to, ako je poskytovaná sociálna práca pracovníkmi, zvolil si množstvo pracovných skúseností. Keďže samotné skúsenosti by bolo veľmi komplikované (ak nie nemožné) merať, výskumník pre identifikáciu tejto premennej zvolí počet rokov, ktoré respondenti strávili v praxi sociálnej práce. Samotné hodnoty premennej počet rokov v praxi by mohli vyzerať nasledovne: 1, 3, 1, 5, 4, 2, 1, 25. Ak by chcel výskumník identifikovať strednú hodnotu, má samozrejme na výber viacero spôsobov. V prípade použitia klasického aritmetického priemeru, výsledkom by bola hodnota* $\overline{x} = 5{,}4$. *Po zohľadnení jednotlivých respondentov a ich praxe môžeme konštatovať, že priemerný počet rokov 5,4 nereflektuje celý kolektív, keďže iba jedna hodnota je nad priemerom a všetci ostatní pracovníci vykazujú menší počet rokov v praxe. Zameriame sa preto na výpočet stredovej hodnoty nachádzajúcej sa v presnom strede, teda mediánu.*
V prvom rade je potrebné utriediť hodnoty premennej. Dostaneme nasledovný číselný rad:

1	*1*	*1*	*2*	*3*	*4*	*5*	*25*

V tomto prípade ide o párny počet hodnôt, keďže respondentov bolo 8. Pre výpočet pozície mediánu postupujeme podľa spôsobu pre párny počet hodnôt, čiže ide o aritmetický priemer hodnôt na pozíciách $\frac{8}{2} = 4$ *a* $\frac{8}{2} + 1 = 5$. *Konkrétne pôjde o hodnoty „2“ a „3“ a medián* $\tilde{x} = 2{,}5$. *V tomto prípade môžeme povedať, že hodnota mediánu lepšie reprezentuje stredovú hodnotu v danom výskumnom súbore než hodnota aritmetického priemeru.*

V prípade, ak by bol počet hodnôt nepárny (predstavme si, že vo výskumnej vzorke je ešte jeden respondent s počtom rokov v praxi „7“, číselný rad by vyzeral:

1	*1*	*1*	*2*	*3*	*4*	*5*	*7*	*25*

Pozícia mediánu by sa počítala podľa vzorca pre nepárny počet hodnôt $r = \frac{n+1}{2} = \frac{9+1}{2} = 5$. *Medián sa teda nachádza na 5. mieste, v tomto prípade pôjde o* hodnotu *Me(x) = 3.*

Výhodou mediánu je to, že okrem určovania pozície prostredníctvom vzorcov sa dá odhadnúť i vizuálne, keďže podľa definícií je známe, že počet hodnôt pred a za mediánom má byť rovnaký.

(2+3)/2 = 2,5

1	*1*	*1*	*2*	*3*	*4*	*5*	*25*

1	*1*	*1*	*2*	*3*	*4*	*5*	*7*	*25*

Medián ako hodnota, ktorá ukáže presný stred hodnôt údajov, môže byť v určitých prípadoch nepresná. Hovoríme o prípadoch, kedy sa medián nachádza v intervale rovnakých hodnôt. V takýchto výpočtoch sa uplatňuje výpočet **mediánového intervalu**, a následne sa vypočíta hodnota mediánu pomocou vzorca.

Príklad II. *Opäť použijeme príklad s dĺžkou praxe sociálnych pracovníkov. V tomto prípade však výskumník zistí nasledovné dĺžky praxe v rokoch: 1, 1, 2, 3, 3, 3, 3, 4, 5, 5, 8, 25. Keďže hodnôt je v* tomto *prípade 12, pre výpočet pozície mediánu sa použije postup pre párny počet hodnôt. V tomto prípade sa medián bude nachádzať medzi 6. a 7. hodnotou (v tomto príklade medzi hodnotami „3“ a „3“). Princíp výpočtu aritmetického priemeru by následne hovoril, že hodnota mediánu bude „3“. Mnoho výskumníkov by tento záver prijalo s tým, že nemá v tomto prípade praktický zmysel zvažovať, či presný stred nie je o málo vyšší ako hodnota 3. Matematicky sa však zohľadňuje práve fakt, že medián sa nachádza v intervale, ktorý ohraničuje všetky hodnoty „3“ v dátovom súbore. Pre ilustráciu číselný* rad *vyjadríme tabuľkou:*

1	*1*	*2*	*3*	*3*	*3*	*3*	*4*	*5*	*5*	*8*	*25*

V tabuľke sme modrou vyznačili stredovú pozíciu a červenou interval obsahujúci všetky hodnoty „3“, kde sa medián bude nachádzať. Vzhľadom na ostatné hodnoty sa medián nachádza v intervale 2,5 – 3,5. Konkrétne sa nachádza na tretej pozícii zo štyroch rovnakých hodnôt. Ak teda zohľadníme tento fakt, medián sa vypočíta ako súčet spodnej hranice mediánového intervalu a pozície mediánu v intervale. V našom prípade to bude $\tilde{x} = 3 + \frac{3}{4}$ *= 3,75. Hodnota skutočného stredu, teda mediánu je 3,75.*

Výpočet:
Výpočet mediánu prostredníctvom mediánového intervalu je možný i prostredníctvom vzorca:

$$\tilde{x} = L + \frac{\frac{N}{2} - F}{f_m} . h$$

pričom:
L predstavuje označenie dolnej hranice mediánového intervalu
N predstavuje celkový počet hodnôt v dátovom súbore
F predstavuje celkový počet hodnôt nižších ako mediánový interval
f_m predstavuje počet hodnôt v mediánovom intervale
h predstavuje šírku mediánového intervalu (rozsah hodnôt ktoré ho tvoria) (Coolican 2014).

Ak použijeme tento vzorec na predstavený príklad, dosadíme do vzorca tieto čísla:

L = 3; N = 12; F = 3; f_m *= 4; h = 1 (v tomto prípade je šírka 1, keďže v intervale sa nachádzajú iba hodnoty 3, v ďalšom príklade však vysvetlíme i širší mediánový interval).*

Po takomto dosadení bude vzorec a samotný výpočet vyzerať nasledovne: $\tilde{x} = 3 + \frac{\frac{12}{2} - 3}{4} . 1 = 3{,}75.$

Príklad III. *Medián sa okrem uvedených prípadov, kedy sú známe všetky konkrétne hodnoty premennej dá identifikovať i pri premenných, kedy poznáme iba intervaly a ich početnosti. Predstavme si, že opäť pracujeme s výskumnými dátami o počte rokov v praxi sociálnych pracovníkov. V tomto prípade však výskumník nemá k dispozícii prehľad o všetkých konkrétnych hodnotách a pracuje s prehľadom v tabuľke. Pri takýchto prípadoch je medián jedinou cestou určenia stredovej hodnoty, keďže výpočet iných hodnôt (napr. priemeru) je ovplyvnený nedostatkom konkrétnych dát.*

Tab. 19 - príklad mediánový interval – počet rokov v praxi

Prax	*f*
1-5	*13*
6-10	*25*
11-15	*10*
16 →	*7*

Z uvedeného príkladu vyplýva, že výskumu sa zúčastnilo 55 respondentov. Ak je úlohou výskumníka nájsť stredovú hodnotu v podobe mediánu, prvým krokom je určenie, v ktorom z intervalov sa bude nachádzať. Keďže stále platí pravidlo, že medián je presným stredom, bude sa nachádzať v intervale 6-10 (mal by sa nachádzať na pozícii r = $\frac{55}{2}$ *= 27,5). Je teda potrebné nájsť pozíciu 27,5 vo výskumnej vzorke. Keďže v prvom intervale sa nachádza 13 hodnôt, nemôže to byť v tomto intervale. Keďže sa v druhom nachádza ďalších 25 hodnôt, celkový počet hodnôt v prvom a druhom intervale je 38. Pozícia 27,5 sa teda logicky nachádza v druhom intervale. V tomto prípade máme dostatok údajov na to, aby sme medián vypočítali podľa vyššie uvedeného vzorca, pričom dosadíme nasledovné hodnoty: L = 6 (dolná hranica intervalu, v ktorom sa medián nachádza); N = 55 (celkový počet hodnôt); F = 13 (počet hodnôt predchádzajúcich mediánový interval);* f_m *= 25 (počet hodnôt v mediánovom intervale); h = 5 (šírka intervalu v ktorom sa nachádza medián, teda „vzdialenosť" medzi hodnotami 6-10, resp. počet rôznych hodnôt, ktoré sa v tomto intervale nachádzajú). Konkrétny výpočet bude teda vyzerať nasledovne:*

$$\tilde{x} = 6 + \frac{\frac{55}{2} - 13}{25} . 5 = 8{,}9$$

Stred tohto súboru údajov teda predstavuje hodnota 8,9, s čím môže výskumník pri interpretácii ďalej pracovať.

Ako sme naznačili vyššie, medián má v štatistike široké využitie. Je vnímaný ako jeden z ukazovateľov stredu hodnôt a zobrazuje presnú stredovú pozíciu. Je vhodný pre interpretáciu v prípadoch, že v súbore hodnôt sa nachádzajú extrémne hodnoty, teda pre distribúcie, ktoré sú zošikmené. Vo veľkej časti prípadov sa počíta jednoduchšie ako priemer a rovnako sa používa i v prípadoch poradových premenných, a tam, kde výpočet aritmetického priemeru nemá význam, alebo nie je možný. Toto všetko radíme k *výhodám* použitia mediánu.
Okrem toho, má medián i svoje *nevýhody.* Tu patrí najmä to, že jeho výpočet sa obmedzuje iba na veľmi malý počet hodnôt v súbore (pričom ostatné hodnoty sa žiadnym spôsobom nezohľadňujú); nemožno ho použiť pre žiadne ďalšie štatistické odhady (napr. pre odhad parametrov populácie) ako aj to, že má komplikovanú, niekedy až zmätočnú interpretáciu pri malom počte hodnôt (ak je veľmi malý počet hodnôt /napr. 1,2,5,100,130/, hodnota mediánu „5" nereprezentuje tento súbor dát) (Coolican 2014).

TIP: hodnota mediánu sa dá pomerne jednoducho určiť manuálne, alebo pomocou uvedených vzorcov. Aj v tomto prípade je však možné využiť dostupné matematické a štatistické nástroje.

V programe Excel – medián má v programe Excel vlastnú funkciu, ktorá sa zapisuje v tvare:

=MEDIAN(hodnota1:hodnota2)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má medián určiť (napr. =MEDIAN(B1:B7).

V štatistickom programe SPSS je medián súčasťou viacerých analýz. Výpočet mediánu ponúkajú analýzy s názvami „Frequencies“, „Descriptives“, či „Explore“ v záložke Analyze → Descriptive Statistics. Medián je označený ako „Median“.

Modus

Modus je posledným z ukazovateľov centrálnej tendencie. Tento ukazovateľ identifikuje tzv. **typickú hodnotu** premennej v súbore. Je to teda najčastejšie vyskytujúca sa hodnota premennej. Ako z uvedeného vyplýva, ide o hodnotu, ktorá sa opakuje najčastejšie a teda má najvyššiu početnosť. Často sa používa pri premenných, kde nie je možné určiť priemery ani medián, resp. kde ich zisťovanie nemá zmysel (napríklad ak pracujeme s nominálnymi premennými).

V prípade modusu je nájdenie veľmi jednoduché, stačí poznať absolútne početnosti jednotlivých hodnôt a na ich základe určiť, ktorá z hodnôt je modusom.

Príklad*. V prípade premennej počet rokov v praxi, s ktorou sme pracovali v časti pojednávajúcej o hodnote mediánu by určenie hodnoty modusu prebiehalo pomerne jednoducho. Využijeme hodnoty premennej: 1, 3, 1, 5, 4, 2, 1, 25. Po ich usporiadaní je možné konštatovať, že hodnota modusu je* $\hat{x} = 1$. *Táto hodnota sa v súbore vyskytuje najčastejšie (3x), pričom ostatné hodnoty sa v tomto prípade vyskytujú iba raz.*

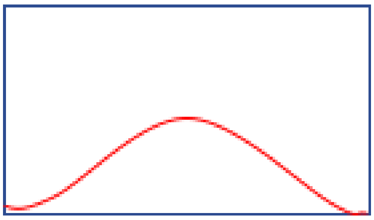
Označenie:
Modus sa najčastejšie označuje ako $\hat{x}$, v niektorých zdrojoch sa stretneme s označením „$Mod_{(x)}$“, resp. „$Mo_{(x)}$“.

Modus ako typická hodnota premennej tvorí tzv. „vrchol distribúcie“, teda na pomyselnom grafe rozdelenia distribúcie hodnôt je hodnota modusu na najvyššej pozícii. Prakticky sa však pomerne často stáva, že vrcholov distribúcie je viacero – je teda možné nájsť viac modusov. Môžeme tak vidieť distribúciu, ktorá je dvojvrcholová, resp. viacvrcholová. To však v praxi znamená, že štatistický súbor v prípade skúmanej premennej je nekonzistentný. Pre výskumníka to značí, že by sa mal zamyslieť nad spôsobom interpretácie tejto premennej – keďže ostatné stredové hodnoty by mohli byť do značnej miery ovplyvnené práve nekonzistenciou premennej. Na druhej strane, výskumník si tak vytvára obraz o samotnej premennej a rozdelení hodnôt, čo môže napovedať o samotnom rozdelení – distribúcii.

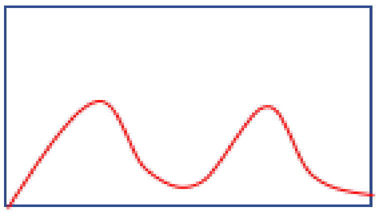
V prípade, že distribúcia premennej je dvojvrcholová, označuje sa taktiež pojmom bi-modálna, v niektorých zdrojoch sa stretávame s označovaním distribúcie v tvare U (podľa dvoch vrcholov). Určovanie modusu má v tomto prípade svoje pravidlá:

- v prípade, že sa v rade hodnôt vyskytujú 2 rovnako veľké početnosti za sebou, modusom je štandardne ich aritmetický priemer (pri hodnotách 1,1,2,2,2,3,3,3,4,5 je modusom hodnota 2,5);
- pokiaľ sa v súbore hodnôt vyskytujú 2 rovnako početné hodnoty, ktoré nie sú susediace, uvádzajú sa 2 hodnoty modusu (pri hodnotách 1,1,1,2,3,3,3,4,5,5 sú modusmi hodnoty $Mo_{(x1)} = 1$ a $Mo_{(x2)} = 3$);
- v prípade, že sa v súbore nachádzajú viac ako 2 hodnoty s rovnakou početnosťou, ide o multimodálnu premennú a takéto modusy sa štandardne nezobrazujú.

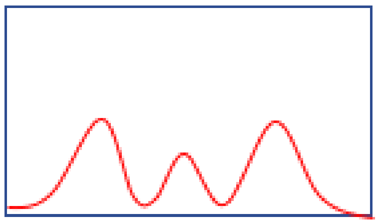
V prípade modusov môžeme hovoriť i o tzv. vrcholoch premennej (keďže v grafickej podobe tvorí modus najvyššie miesto v distribúcii.



Obr. 3 - modus jednomodálne rozloženie



Obr. 4 - modus bimodálne rozloženie



Obr. 5 – modus viacmodálne rozloženie

Postupy pre určenie modusu sú platné pre číselné rady, kde poznáme každú hodnotu premennej. Modálna hodnota sa však dá vypočítať i pre premenné, ktorých početnosť je intervalovo rozdelená. V tomto prípade sa modus počíta podľa vzorca a pre výpočet sa použije tzv. **modálny interval.**

Výpočet:
Výpočet modusu prostredníctvom modálneho intervalu je možný prostredníctvom vzorca:

$$\hat{x} = a_{\hat{x}} + \frac{d_0}{d_0 + d_1} . h_{\hat{x}}$$

pričom:
$a_{\hat{x}}$ je spodná hranica modálneho intervalu
d_0 je rozdiel medzi početnosťou (počtom hodnôt) modálneho intervalu a intervalu, ktorý mu predchádza
d_1 je rozdiel medzi početnosťou modálneho intervalu a intervalu, ktorý nasleduje
$h_{\hat{x}}$ je šírka modálneho intervalu (Sodomová 2013).

***Príklad**. V tomto prípade opäť použijeme príklad s premennou počet rokov v praxi. Vychádzame z intervalového rozdelenia a absolútnych početností:*

Tab. 20 - príklad modálneho intervalu – počet rokov v praxi

Prax	*f*
1-5	*13*
6-10	*25*
11-15	*10*
16 →	*7*

Z uvedeného je zrejmé, že modálnym intervalom (teda intervalom s najväčšou početnosťou) je interval 6-10, máme všetky potrebné informácie pre výpočet modusu. Tieto môžeme dosadiť do vzorca, kde $a_{\hat{x}} = 6$; $d_0 = 25 - 13 = 12$; $d_1 = 25 - 10 = 15$; $h_{\hat{x}} = 5$. *Výpočet modusu tak bude vyzerať* $\hat{x} = 6 + \frac{12}{12+15} . 5 = 8{,}8$. *V tomto prípade môžeme konštatovať, že modálna hodnota sa nachádza v intervale 6-10, konkrétne je to hodnota 8,8.*

Použitie modusu v deskriptívnej štatistike má svoj význam najmä tam, kde nie je možné použiť hodnotu aritmetického priemeru alebo mediánu. Najčastejšie sa teda používa pri nominálnych typoch premenných. Určuje najpodstatnejšiu hodnotu v rámci distribúcie, ktorou je početnosť a z povahy určenia nie je zaťažený extrémnymi hodnotami. Pokiaľ je rozdelenie hodnôt viacvrcholové, modus dáva výskumníkom dôležité informácie, ktoré musia pri interpretácii zvážiť – niekedy je toto dôležitým krokom pri rozhodnutí, či nie je vhodnejšie zmeniť prístup k interpretácii. Toto všetko sú *výhody* používania modusu.

Rovnako i tu vieme hovoriť o *nevýhodách*, ktorými sú najmä hodnoty, z ktorých sa identifikuje. Modus berie do úvahy iba veľmi málo hodnôt premennej, nemôže sa používať na zložitejšie štatistické procedúry, pri malých vzorkách môže spôsobovať problémy kedy sa stáva nereprezentatívnym. Rovnako v mnohých prípadoch nie je možné modus určiť (v súbore sa nenachádza opakovaná početnosť pri žiadnej hodnote), alebo na druhej strane je ich veľké množstvo (Coolican 2014).

TIP: napriek tomu, že modus sa dá pomerne presne a spoľahlivo určiť i prostredníctvom vizuálnych pomôcok a frekvenčnej tabuľky, aj v tomto prípade sa dajú využiť softwarové riešenia.

V programe Excel – modus má v programe Excel vlastnú funkciu, ktorá sa zapisuje v tvare:

=MODE.SNGL(hodnota1:hodnota2)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má modus určiť (napr. =MODE.SNGL(B1:B7).

V štatistickom programe SPSS je modus súčasťou viacerých analýz. Výpočet ponúkajú analýzy s názvami „Frequencies“, „Descriptives“, či „Explore“ v záložke Analyze → Descriptive Statistics. Medián je označený ako „Mode“.

Kvantily

Kvantily sú v štatistike rovnako ukazovateľmi polohy, ktoré však nezobrazujú stredovú pozíciu. Slúžia však na rozdelenie štatistického súboru na vopred definovaný počet rovnako početných častí. Základnou podmienkou v tomto prípade je zoradenie súboru na jednotlivé hodnoty od najnižšej po najvyššiu).

Princíp výpočtu a použitia jednotlivých kvantilov je obdobný ako v prípade **mediánu**. Samotný medián sa taktiež medzi kvantily zaraďuje a je z nich najpoužívanejší.

Okrem mediánu, ktorý identifikuje v súbore dve rovnako početné časti v štatistickej praxi rozoznávame:

- **kvartily** (3 hodnoty deliace súbor na **4** rovnako početné časti – každý kvartil tak obsahuje 25% hodnôt – ak by sme si predstavili, že prvý kvartil bude mať hodnotu 5, 25% respondentov malo hodnoty do 5),
- **kvintily** (4 hodnoty deliace súbor na **5** rovnako početných častí – každý kvintil obsahuje 20% hodnôt),
- **decily** (9 hodnôt deliacich súbor na **10** rovnako početných častí – každý decil obsahuje 10% hodnôt),
- **percentily** (99 hodnôt deliacich súbor na **100** rovnako početných častí – každý percentil obsahuje 1% hodnôt).

Samotný medián, keďže stojí v presnom strede súboru údajov je 2. kvartilom, 5. decilom a rovnako 50. percentilom.

Príklad. *V prípade, že je zámerom výskumníka identifikovať hodnoty kvartilov z hodnôt 1, 1, 2, 3, 3, 3, 3, 4, 5, 5, 8, 25, bude postupovať analogicky k určeniu mediánu, so zohľadnením pozície jednotlivých kvartilov. Prvý kvartil sa teda nachádza na pozícii* $r = \frac{n+1}{4}$, *tretí kvartil na pozícii* $r = 3.\frac{n}{4}$. *V prípade výpočtu jednotlivých kvartilov postupujeme rovnako ako pri výpočte mediánu a rovnako zohľadňujeme pozíciu jednotlivých kvartilov.*

1	*1*	*2*	*3*	*3*	*3*	*3*	*4*	*5*	*5*	*8*	*25*

V uvedenom príklade sme vyznačili miesta, kde by sa jednotlivé kvartily mali nachádzať. Spôsobom výpočtu aritmetického priemeru môžeme určiť, že prvý kvartil má hodnotu $\tilde{x}_{25\%} = 2{,}5$. *Druhý kvartil je zároveň mediánom a má hodnotu* $\tilde{x} = 3{,}25$ *(postup pri určení mediánu v tomto príklade je popísaný v časti o Mediáne). Tretí kvartil sa v tomto prípade vypočíta podľa vzorca (keďže ide o viacnásobný výskyt hodnoty „5“). Vzorec pre výpočet mediánu sa pre tento účel upraví:* $\tilde{x} = 5 + \frac{3.\frac{12}{4} - 8}{2}.1$ *V prípade, ak by sa nachádzal medzi rôznymi hodnotami, bol by tretím kvartilom ich aritmetický priemer.* V *tomto prípade je hodnota tretieho kvartilu* $\tilde{x}_{75\%} = 5{,}5$.

V prípade výpočtu ďalších kvantilov postupujeme podľa uvedeného postupu, zohľadňujeme pozíciu týchto hodnôt (v prípade kvintilov sú to pozície $r = \frac{n}{5}$, $2.\frac{n}{5}$, $3.\frac{n}{5}$, $4.\frac{n}{5}$, percentuálne vyjadrené na pozíciách 20%, 40%, 60%, 80%.

TIP: aj v prípade kvartilov je možné využiť technológie. Tu je potrebné zohľadniť, ktorý z kvartilov má výskumník záujem vypočítať.

V programe Excel – kvartily majú v programe Excel vlastnú funkciu v tvare:

=QUARTILE(hodnota1:hodnota2;kvartil)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má kvartil určiť a za bodkočiarkou sa uvádza poradové číslo kvartilu (napr. =QUARTILE(B1:B7;1).

V štatistickom programe SPSS sa kvartily ako i iné kvantily dajú zobraziť prostredníctvom analýzy „Frequencies“ v záložke Analyze → Descriptive Statistics. Tu v karte „Statistics“ je potrebné označiť možnosť „Percentile(s)“ a percentuálne doplniť, aké kvantily je potrebné zobraziť (v príp. kvartilov 25%, 50%, 75%; kvintilov 20%, 40%, 60%, 80% a pod.).

Ukazovatele rozptýlenosti (variability)

Na rozdiel od predchádzajúcich ukazovateľov centrálnej tendencie sú ukazovatele rozptýlenosti zamerané na popis variability hodnôt, zjednodušene povedané, do akej miery sa jednotlivé hodnoty od seba líšia, alebo sa podobajú. Pod pojmom variabilita teda rozumieme premenlivosť hodnôt.
Ukazovatele rozptýlenosti majú za cieľ hlbšie popísať štatistický súbor a hodnoty premennej a poskytujú širší obraz o premennej než iba ukazovatele centrálnej tendencie.

V štatistickej praxi je dôležité, aby sa jednotlivé ukazovatele vzájomne kombinovali. Ukazuje sa ako nedostatočné, ak by boli súbory popísané iba prostredníctvom ich stredových hodnôt, keďže samotná premenlivosť môže do značnej miery ovplyvniť aký je skutočný charakter premennej a jej hodnôt.
V nasledujúcom príklade sa opätovne orientujeme na premennú počet rokov v praxi sociálnej práce. Úlohou výskumníka je v tomto prípade popísať túto premennú v dvoch na seba nezávislých kolektívoch. Prehľad konkrétnych hodnôt prezentujú nasledujúce tabuľky.

6	6	6	6	6	6	6

1	3	6	6	7	8	11

Ak by sme sa orientovali na ukazovatele centrálnej tendencie, identifikovali by sme aritmetický priemer, medián a modus v hodnote „6" a to v oboch týchto skupinách. Po zameraní sa na konkrétne hodnoty môžeme konštatovať, že minimálne v druhom prípade tieto centrálne hodnoty nedostatočne reprezentujú, ako v skutočnosti hodnoty premennej vyzerajú. Kým v prvom prípade nie je žiadna premenlivosť hodnôt premennej, v druhom prípade je premenlivosť značná a hodnoty premennej sa pohybujú od 1 roka do 11 rokov. Preto konštatovať, že tieto súbory sú identické by nebolo správne. Popis rozptýlenosti (variability) by mal byť teda ďalším krokom, ktorý je súčasťou analýzy.

V rôznych štatistických publikáciách sa stretávame s mnohými ukazovateľmi, ktoré sa orientujú na variabilitu súborov, pričom najpoužívanejšie sú: variačné rozpätie, kvantilové rozpätie, rozptyl a štandardná odchýlka. Tieto ukazovatele taktiež označujeme ako **absolútne miery variability.** Okrem toho rozoznávame tzv. **relatívne miery variability,** medzi ktoré radíme variačný koeficient a pomernú priemernú odchýlku.

Variačné rozpätie

Variačné rozpätie sa používa na identifikáciu rozsahu premenlivosti hodnôt premennej a vo svojej podstate je to identifikácia rozsahu premenlivosti – teda toho, či a ako sa hodnoty v súbore menia.

Vo výpočte ide o rozdiel medzi najvyššou a najnižšou hodnotou premennej v súbore.

Výpočet: výpočet variačného rozpätia je pomerne jednoduchý, konkrétna hodnota sa vypočíta aplikáciou vzorca:

$$R = x_{max} - x_{min}$$

teda od najvyššej hodnoty sa odčíta najnižšia hodnota premennej.

Označenie: ako z uvedeného vyplýva, variačné rozpätie sa označuje ako R, čo pochádza z anglického označenia „Range".

V prípade príkladov s dĺžkou praxe sociálnych pracovníkov použitých na predchádzajúcej strane, variačné rozpätie v prvom prípade má hodnotu R = 6 – 6 = 0; v druhom prípade R = 11 – 1 = 10. Tieto hodnoty vypovedajú o celkovom obraze oboch súborov a pre ich komparáciu predstavujú vhodné východisko.

Výhodou variačného rozpätia je jeho veľmi rýchly výpočet, rovnako poskytuje základný prehľad o premenlivosti hodnôt. Prakticky je však menej uplatňované a to najmä preto, že je silne ovplyvnené extrémnymi hodnotami, ktoré sa nachádzajú na oboch koncoch hodnôt premennej (prakticky sa počíta práve z nich) a jeho výpočet ignoruje ostatné hodnoty premennej čo sú často uvádzané *nevýhody* použitia.

TIP:

V štatistickom programe SPSS sa je možné variačné rozpätie vypočítať prostredníctvom analýz s názvami „Frequencies", „Descriptives" a „Explore" v záložke Analyze → Descriptive Statistics. Variačné rozpätie je v tomto prípade označené ako „Range"

Kvantilové rozpätie

Kvantilové rozpätie, resp. kvantilové rozpätia (keďže poznáme viacero kvantilov ako bolo uvádzané v predchádzajúcich častiach) sú používané ako alternatívy k variačnému rozpätiu. Ich zmyslom je eliminovanie negatívnych vplyvov extrémnych hodnôt, ktoré môžu byť prítomné pri variačnom rozpätí.

Podstatou kvantilových rozpätí je vypočítať prehľad premenlivosti hodnôt premennej medzi prvou a poslednou hodnotou príslušného kvantilu. Keďže extrémne hodnoty premennej sa nachádzajú na jej jednotlivých koncoch, vynechaním prvého a posledného kvintilu by mali byť práve extrémy vynechané a takéto rozpätie by mohlo viac zodpovedať súboru.

V tomto prípade môže byť kvantilových rozpätí niekoľko typov, podľa toho, aký kvantil je použitý. Medzi najpoužívanejšie kvantilové rozpätie patria: **medzikvartilové**, **kvintilové** a **decilové** rozpätie, konkrétne podľa toho, aký počet hodnôt má výskumník záujem eliminovať. Princíp výpočtu jednotlivých typov kvantilového rozpätia závisí na použitom kvantile, v ostatnom je analogický k výpočtu variačného rozpätia.

Výpočet: výpočet kvantilového rozpätia závisí od použitého kvantilu, pričom určenie pozície príslušných kvantilov je prvou podmienkou pre výpočet, v prípade najpoužívanejších kvantilových rozpätí:

medzikvartilové (kvartilové) rozpätie: $R_Q^4 = x_{75\%} - x_{25\%}$
kvintilové rozpätie: $R_Q^5 = x_{80\%} - x_{20\%}$
decilové rozpätie: $R_Q^{10} = x_{90\%} - x_{10\%}$

Označenie: kvantilové rozpätia sa označujú podľa použitého kvantilu horným indexom, ako je uvedené v časti výpočet. Okrem toho sa v prípade medzikvartilového rozpätia používa označenie „IQR“ (z angl. „interquartile range“).

Rozptyl

Rozptyl je ukazovateľ rozptýlenosti, ktorý sa v štatistickej praxi používa veľmi často. Identifikuje, do akej miery sú jednotlivé hodnoty premennej vzdialené od stredovej hodnoty, ktorú reprezentuje aritmetický priemer. Jednou hodnotou je tak možné identifikovať celkový stav rozloženia a variability premennej. Čím je hodnota rozptylu vyššia, tým sú konkrétne hodnoty premennej rozdielne a variabilita je tým pádom vyššia.

Rozptyl je definovaný ako priemer druhých mocnín odchýlok jednotlivých hodnôt od spoločného aritmetického priemeru.

Výpočet: výpočet rozptylu je podmienený poznaním každej hodnoty premennej v súbore a rovnako aritmetického priemeru všetkých hodnôt:

$$s^2 = \frac{\sum d^2}{(n-1)}$$

pričom „d“ je odchýlka konkrétnej hodnoty od hodnoty aritmetického priemeru a môže sa zapísať ako: $d = x_i - \overline{x}$

v prípade, že je záujmom výskumníka vyjadriť rozptyl všetkých prvkov populácie, nepoužije sa na delenie hodnota n-1, ale počet všetkých prvkov N.

Označenie: rozptyl sa štandardne označuje ako s^2, v niektorých zdrojoch sa stretávame s označením „Variance“.

Poznámka: nutnosť výpočtu prostredníctvom druhej mocniny rozdielov vychádza z algebraického princípu aritmetického priemeru, ktorý sme uvádzali v predchádzajúcich častiach. Rozdiel v kladných a záporných rozdielov všetkých hodnôt od aritmetického priemeru sa rovnajú nule. Tým pádom by sa tieto rozdiely vzájomne eliminovali, čím by sa v čitateli zlomku nachádzala 0 a došlo by k matematickej chybe. Toto je riešené práve druhými mocninami.

Príklad. *Príkladom pre výpočet rozptylu môže byť opätovne premenná počet rokov v praxi sociálnej práce. Budeme pracovať s hodnotami premennej známymi z predchádzajúcich ukazovateľov.*

1	*3*	*6*	*6*	*7*	*8*	*11*

Vieme, že aritmetický priemer má hodnotu $\overline{x} = 6$. *Pre výpočet rozptylu teda použijeme vzorec:* $s^2 = \frac{(1-6)^2+(3-6)^2+(6-6)^2+(6-6)^2+(7-6)^2+(8-6)^2+(11-6)^2}{6} = \frac{25+9+0+0+1+4+25}{6} = 10{,}7$. *V prípade výpočtu všetkých prvkov populácie bude menovateľ v zlomku 7 a rozptyl bude mať hodnotu* $s^2 = 9{,}14$.

Výhodou rozptylu je najmä fakt, že informuje výskumníka o celkovej variabilite súboru ako i to, že berie do úvahy všetky hodnoty premennej, čím je pre štatistiku prínosnejší než variačné rozpätie, rovnako je vnímaný ako najpresnejší ukazovateľ variability premennej. Je tiež použiteľný i v inferenčnej štatistike pri odhadoch parametrov populácie.

Praktickou *nevýhodou*, i keď vynútenou algebraickým princípom je na druhej strane fakt, že nie je tvorený samotnými rozdielmi hodnôt od priemeru, ale ich druhými mocninami. Tak sa stáva zle interpretovateľným, keďže druhá mocnina nie je reprezentantom pôvodných hodnôt. To sa však prakticky rieši prostredníctvom výpočtu **štandardnej odchýlky** (Coolican 2014).

<u>TIP:</u> matematické a štatistické programy môžu byť užitočné i v prípade výpočtu rozptylu

V programe <u>Excel</u> – rozptyl vieme vypočítať prostredníctvom pripravenej funkcie:

=VAR.P(hodnota1:hodnota2)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má rozptyl určiť (napr. =VAR.P(B1:B7).

rovnako je možné využiť zápis prostredníctvom vzorca, ktorý je ale v tomto prípade komplikovanejší, keďže musíme využiť funkcie pre mocniny:

=((POWER((hodnota1-priemer);2))+(POWER((hodnota2-priemer);2))+ ... + (POWER((hodnota$_n$-priemer);2)))/počet hodnôt

V štatistickom programe <u>SPSS</u> sa je možné rozptyl vypočítať prostredníctvom analýz s názvami „Frequencies“, „Descriptives“ a „Explore“ v záložke Analyze → Descriptive Statistics. Rozptyl je v tomto prípade označený ako „Variance“

<u>Štandardná odchýlka</u>

Nevýhodu rozptylu, ktorou je jeho umocnenie, kompenzuje štandardná odchýlka. Pôvodnú hodnotu rozptylu udávanú v štvorcoch (v druhej mocnine) je možné vrátiť do pôvodných jednotiek, v ktorých je premenná zisťovania prostredníctvom jeho odmocniny. Touto procedúrou získame tzv. **štandardnú odchýlku**, ktorá je definovaná ako miera variability identifikujúca ako veľmi sa jednotlivé hodnoty odlišujú od typickej hodnoty a to v pôvodných jednotkách. Štandardná odchýlka sa rovnako označuje i ako **smerodajná odchýlka**.

<u>TIP:</u> i výpočet štandardnej odchýlky je možný prostredníctvom dostupných programov

V programe <u>Excel</u> – štandardnú odchýlku vieme vypočítať prostredníctvom pripravenej funkcie:

=STDEV.P(hodnota1:hodnota2)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má rozptyl určiť (napr. =STDEV.P(B1:B7).

v prípade, že poznáme rozptyl, ktorý je uvedený v programe Excel, môžeme využiť i priamy vzorec, ktorý využije matematický princíp štadnardnej odchýlky (odmocnenie rozptylu):

=POWER(rozptyl;1/2)

V štatistickom programe <u>SPSS</u> sa je možné štandardnú odchýlku vypočítať prostredníctvom analýz s názvami „Frequencies“, „Descriptives“ a „Explore“ v záložke Analyze → Descriptive Statistics. Rozptyl je v tomto prípade označený ako „Std. Deviation“

Štandardná odchýlka sa používa v interpretácii dát s pomerne veľkou frekvenciou. Jej *výhodou* je to, že identifikuje mieru variability v pôvodných jednotkách. Na druhej strane, *nevýhodou* je jej značné ovplyvnenie samotným aritmetickým priemerom. Tu je pre výskumníka neustále potrebné zvažovať, či je vhodné aritmetický priemer používať ako ukazovateľ centrálnej tendencie. Pokiaľ nie (v prípade ovplyvnenia extrémnymi hodnotami, zošikmenej distribúcie), štandardná odchýlka a tým i posúdenie variability je tým do značnej miery ovplyvnené (Hendl 2012).

Relatívne miery variability

Relatívne miery variability kompenzujú nedostatok predchádzajúcich ukazovateľov, ktorým je ich absolútnosť. Nedajú sa teda použiť na vzájomnú komparáciu viacerých premenných, najmä vyjadrených v rozdielnych jednotkách merania alebo rozdielnych hodnôt znaku. Relatívne miery variability komparujú vypočítané absolútne miery so samotným aritmetickým priemerom a určujú rovnorodosť, resp. rôznorodosť súboru, čím umožňujú komparáciu viacerých súborov aj s odlišnými jednotkami merania či hodnotami znaku.

Najčastejšie sú používané:

- variačný koeficient – vyjadruje do akej miery je štandardná odchýlka vyjadrením priemeru, udáva sa v percentách pričom vysoká hodnota indikuje nerovnorodosť súboru. Počíta sa ako vzťah: $V_x = \frac{s}{\bar{x}}.100$;
- pomerná priemerná odchýlka – vyjadruje do akej miery je priemerná odchýlka vyjadrením priemeru a počíta sa ako podiel priemernej odchýlky $\bar{d}$ a aritmetického priemeru vyjadrený v percentách. I v tomto prípade platí, že vyššia hodnota svedčí o nerovnorodosti súboru.

Ukazovatele symetrie (tvaru distribúcie)

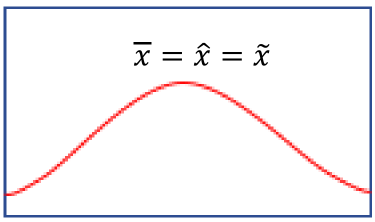
Kým sme sa v predchádzajúcich častiach zaoberali najmä polohou konkrétnych hodnôt v súbore či mierou variability hodnôt v súbore, ukazovatele symetrie umožňujú odhadnúť tvar rozdelenia hodnôt. Na viacerých miestach v tejto učebnici sme uvádzali nemožnosť použitia konkrétnych metód v prípade šikmej, alebo špicatej distribúcie. Práve ukazovatele symetrie umožňujú identifikovať tieto fenomény jasným, matematickým, alebo grafickým spôsobom.
V praxi štatistiky sociálnych a behaviorálnych vied nám práve tieto ukazovatele napovedajú, ako sú hodnoty rozdelené, na ktorých miestach sa nachádzajú jednotlivé ukazovatele centrálnej tendencie, čo priamo predikuje, aké metódy ďalšej analýzy je možné ďalej použiť.

Konkrétne ukazovatele, ktoré je v tejto oblasti používame sú:

- ukazovatele šikmosti,
- ukazovatele špicatosti.

Východiskom pri týchto ukazovateľoch je tzv. symetrická distribúcia, ktorá vychádza z normálneho rozdelenia reprezentovaného Gaussovým rozdelením (Gaussovou krivkou).
Ak by sme rozloženie hodnôt premennej graficky zobrazili, teoreticky je možné už na základe tohto grafu odčítať niektoré charakteristiky tvaru distribúcie. Nasledujúca časť tejto učebnice pojednáva o ďalších, negrafických spôsoboch určenia tvaru distribúcie s ich využitím pri deskriptívnej interpretácii výskumných dát.

Normálne rozdelenie sa rovnako označuje pojmom symetrické, je charakteristické tým, že grafickú reprezentáciu hodnôt reprezentuje Gaussova krivka. Ak by sme identifikovali jednotlivé doposiaľ prezentované ukazovatele (aritmetický priemer, modus, medián), nachádzali by sa v strede, na najvyššom bode tohto grafu. Pri takto rozdelených hodnotách sa v súbore nenachádzajú extrémne hodnoty, pričom všetky hodnoty sú v distribúcii vyvážené.



Obr. 6 - symetrické – normálne rozdelenie

Takéto rozdelenie je v štatistickej praxi základom pre rôzne metódy tzv. parametrickej štatistiky, teda takých štatistických procedúr patriacich do oblasti inferenčnej štatistiky, ktoré používajú ukazovatele centrálnej tendencie, vzájomne ich komparujú, resp. s nimi inak pracujú. Pre lepšie porozumenie povahy dát v štatistickom súbore sa komparuje skutočný tvar distribúcie s uvedeným symetrickým rozdelením, na čo sa používajú práve ukazovatele šikmosti a špicatosti.

Šikmosť

Šikmosť dát identifikuje asymetriu, resp. nerovnomernosť hodnôt premennej sprava, alebo zľava. Šikmosť dát sa identifikuje graficky, alebo matematicky, pričom pre tento účel sa počíta tzv. koeficient šikmosti.

Výpočet: koeficient šikmosti vieme určiť podľa vzorca:

$$b_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{s^3}$$

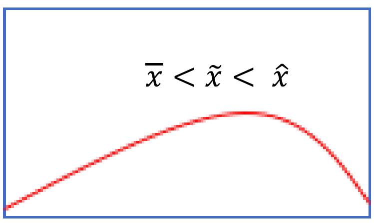
Označenie: v prípade koeficientu šikmosti sa stretávame s viacerými označeniami, pričom najčastejšími sú b_1, γ_1 alebo S_1.

Interpretácia: výsledkom výpočtu je číslo, ktoré má bezrozmerný charakter. Interpretácia šikmosti sa vzťahuje k tomuto číslu nasledovne:

- $b_1 = 0$ – rozdelenie približne symetrické,
- $b_1 < 0$ – rozdelenie zošikmené zľava,
- $b_1 > 0$ – rozdelenie zošikmené sprava.

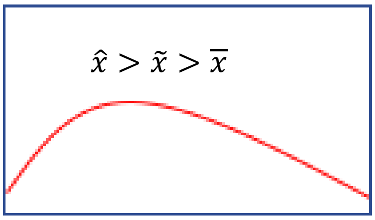
Zošikmenie zľava, označované ako negatívne, alebo záporné je charakteristické predĺženou ľavou stranou distribúcie (v nižších hodnotách premennej). To prakticky znamená, že hodnota aritmetického priemeru je posunutá smerom doľava, nasleduje medián a napokon modus.

Prakticky je toto zošikmenie spôsobené tým, že naľavo od stredu (priemeru) sa nachádzajú odľahlejšie – vzdialenejšie hodnoty, pričom napravo od priemeru sa nachádza väčšina hodnôt premennej.



Obr. 7 - nesymetrická distribúcia – záporné zošikmenie

Zošikmenie sprava sa označuje ako pozitívne či kladné a charakterizuje ho opačne predĺžená, teda pravá strana distribúcie. V tomto prípade je prvou hodnotou v distribúcii modus, následne medián a napokon aritmetický priemer, a to vzhľadom na početnosť hodnôt v súbore a ich rozloženie.



Obr. 8 - nesymetrická distribúcia – kladné zošikmenie

V zošikmení sprava (alebo tiež označovanom ako pozitívnom zošikmení) sú oproti ľavostrannému odľahlejšie hodnoty na pravej strane od priemeru a väčšina hodnôt sa nachádza pred hodnotou aritmetického priemeru.

<u>*TIP:*</u> výpočet šikmosti značne uľahčuje použitie štatistických a matematických programov

V programe <u>Excel</u> – šikmosť má nastavenú funkciu

=SKEW(hodnota1:hodnota2)

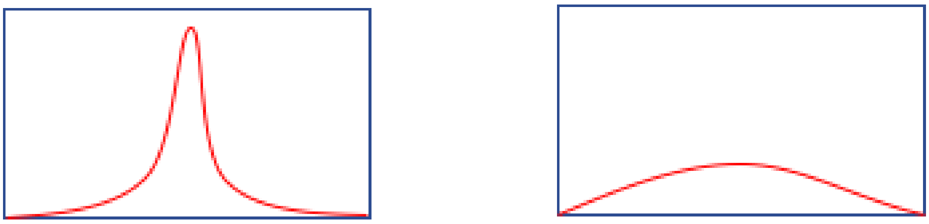
pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má šikmosť určiť (napr. =SKEW(B1:B7).
Pre určenie šikmosti pre celú populáciu sa používa funkcia SKEW.P(hodnota1:hodnota2).

V štatistickom programe <u>SPSS</u> sa je možné štandardnú odchýlku vypočítať prostredníctvom analýz s názvami „Frequencies", „Descriptives" a „Explore" v záložke Analyze → Descriptive Statistics. Šikmosť je označená ako „Skewness".

Okrem koeficientu zošikmenia sa pre výpočet šikmosti používa i tzv. **Pearsonova miera šikmosti**, ktorá sa označuje ako S_p. Interpretácia tejto miery je identická s interpretáciou koeficientu šikmosti. Rozdiel je v spôsobe výpočtu, ktorý sa realizuje prostredníctvom vzorca: $S_p = \frac{\bar{x}-\hat{x}}{s}$.

<u>*Špicatosť*</u>

Podobne ako šikmosť je špicatosť ukazovateľom distribúcie hodnôt premennej, pričom v tomto prípade ide o mieru zoskupenia hodnôt na mieste, ktoré je stredom distribúcie. V prípade, ak je väčšina hodnôt zoskupená v oblasti stredu, sám stred je extrémnou hodnotou a hovoríme o tzv. špicatej (alebo tiež leptokurtickej) distribúcii. Ak sú si všetky hodnoty veľmi podobné, distribúcia je plochá (tiež platykurtická). Konkrétne sa špicatosť identifikuje prostredníctvom **koeficientu špicatosti.**



Obr. 9 - nesymetrická distribúcia – špicaté a ploché rozdelenie

Výpočet: koeficient špicatosti vieme určiť podľa vzorca:

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{s^4} - 3$$

Označenie: v prípade koeficientu špicatosti sa stretávame s viacerými označeniami, pričom najčastejšími sú b_2, γ_2 alebo S_2.

Interpretácia: výsledkom výpočtu je číslo, ktoré má bezrozmerný charakter. Interpretácia špicatosti sa vzťahuje k tomuto číslu nasledovne:

- $b_2 = 0$ – rozdelenie približne normálne,
- $b_2 < 0$ – rozdelenie špicatejšie,
- $b_2 > 0$ – rozdelenie plochšie.

TIP: výpočet špicatosti sa realizuje aj prostredníctvom počítačových programov

V programe Excel – špicatosť vypočíta funckia:

=KURT(hodnota1:hodnota2)

pričom v zátvorke sa stále uvádza príslušný rozsah hodnôt, z ktorých sa má určiť (napr. =KURT(B1:B7).

V štatistickom programe SPSS sa je možné štandardnú odchýlku vypočítať prostredníctvom analýz s názvami „Frequencies“, „Descriptives“ a „Explore“ v záložke Analyze → Descriptive Statistics. Šikmosť je označená ako „Kurtosis“.

Cvičenie

V tabuľke nižšie sa nachádza dátový súbor respondentov – pracovníkov zariadenia poskytujúceho sociálne služby, ktorí odpovedali na dotazník týkajúci sa niektorých okolností ich práce. Zároveň tento dotazník obsahoval metodiku identifikujúcu úroveň ich vyhorenia ako negatívneho javu spôsobeného výkonom pomáhajúcej profesie.

Preštudujte si prosím dátový súbor a odpovedzte na otázky pod tabuľkou. Použite pritom jednak počítanie prostredníctvom vzorcov a prostredníctvom počítačových programov.

Tab. 21 - cvičenie – číselné deskriptívne metódy

Rod	vek	Vyhorenie - skóre	Prax v rokoch	Funkčný plat
muž	25	10	3	455
žena	27	12	2	460
muž	26	10	5	470
muž	25	15	8	470
žena	20	11	3	480
muž	22	15	5	480
žena	20	12	7	480
žena	23	14	9	490
žena	23	15	2	490
muž	30	11	1	490
žena	35	10	1	490
muž	31	12	8	500
muž	34	13	6	500
muž	32	12	5	500
muž	30	10	7	500
žena	29	15	8	500
žena	27	12	3	510
žena	28	13	5	510
muž	19	15	6	510
muž	38	13	2	510
žena	33	12	6	510
žena	26	10	7	510
muž	23	14	2	520
muž	23	12	4	520
žena	24	15	6	520

žena	27	13	3	520
žena	28	15	1	520
žena	31	12	8	530
muž	31	14	6	530
muž	33	15	7	530
muž	40	10	3	530
muž	37	10	1	540
žena	38	13	4	540
žena	32	12	7	540
žena	37	15	8	550
žena	25	13	6	550
žena	20	15	3	560
muž	24	14	5	565

Otázky a úlohy:

1. Transformujte tieto hodnoty do počítačového programu podľa vlastného výberu (Excel, SPSS,...).
2. Identifikujte aritmetický priemer z premenných, kde sa dá určiť. Zdôvodnite, pre ktoré premenné to nie je vhodné a prečo.
3. Rozhodnite, akým spôsobom by bolo možné vypočítať redukovaný priemer a vypočítajte ho.
4. Vypočítajte geometrický priemer.
5. Určte medián, modus a kvartily.
6. Vypočítajte miery variability hodnôt a interpretujte ich.
7. Vypočítajte koeficient šikmosti a špicatosti, pokúste sa identifikovať distribúciu graficky.

4.3 Grafické deskriptívne metódy

Grafické metódy deskriptívnej štatistiky sú súčasťou deskriptívnej analýzy, ktorá má za cieľ pomerne rýchlo a prehľadne prezentovať výskumné dáta prostredníctvom zobrazenia vo forme grafov. Používajú sa najmä na ilustráciu rozdelenia hodnôt jednej, alebo viacerých premenných. *Výhodami* použitia grafických deskriptívnych metód sú ich intuitívna interpretácia a relatívne jednoduché spracovanie a použitie. Umožňujú identifikovať v dátach rôzne pravidlá (vzory), ktoré nemusia byť na prvý pohľad viditeľné pri použití iných metód deskriptívnej analýzy. Rovnako je výhodné, že tieto metódy dokáže z hľadiska ich grafickej povahy interpretovať i laik, preto sú vhodné aj pre prezentáciu dát smerom k verejnosti. Grafické metódy sú vhodnými i pre prezentáciu nominálnych premenných, čo je opäť oproti číselným charakteristikám výhodou. Na druhej strane, ich ilustratívnosť je zároveň *nevýhodou*, keďže nie sú tak presné ako číselné deskriptívne metódy. Preto je ich použitie obmedzené a v štatistickej praxi je vhodné ich kombinovať.

Prakticky sa pri interpretácii výskumných dát často stretávame s otázkou, akú štatistickú metódu je vhodné pri tých-ktorých premenných použiť. Grafické metódy poskytujú rýchly spôsob prehľadu o rozložení hodnôt premenných, pričom sú ľahko interpretovateľné aj bez použitia dodatočných tabuľkových, alebo číselných metód. To je však možné najmä pri nižšom počte hodnôt, keďže v opačnom prípade sa stávajú neprehľadnými. Výskumník by preto mal vždy zvoliť taký spôsob zobrazenia, ktorý bude čo najjednoduchšie a najprehľadnejšie ilustrovať zmysel prezentovania výskumných dát.

Z praktického hľadiska sa grafické deskriptívne metódy uplatňujú najmä ako **jednorozmerné**, no je možné sa stretnúť i so zobrazením **dvoch** a **viacerých rozmerov.** V nasledujúcej časti budeme prezentovať niektoré najpoužívanejšie metódy grafickej reprezentácie výskumných dát.

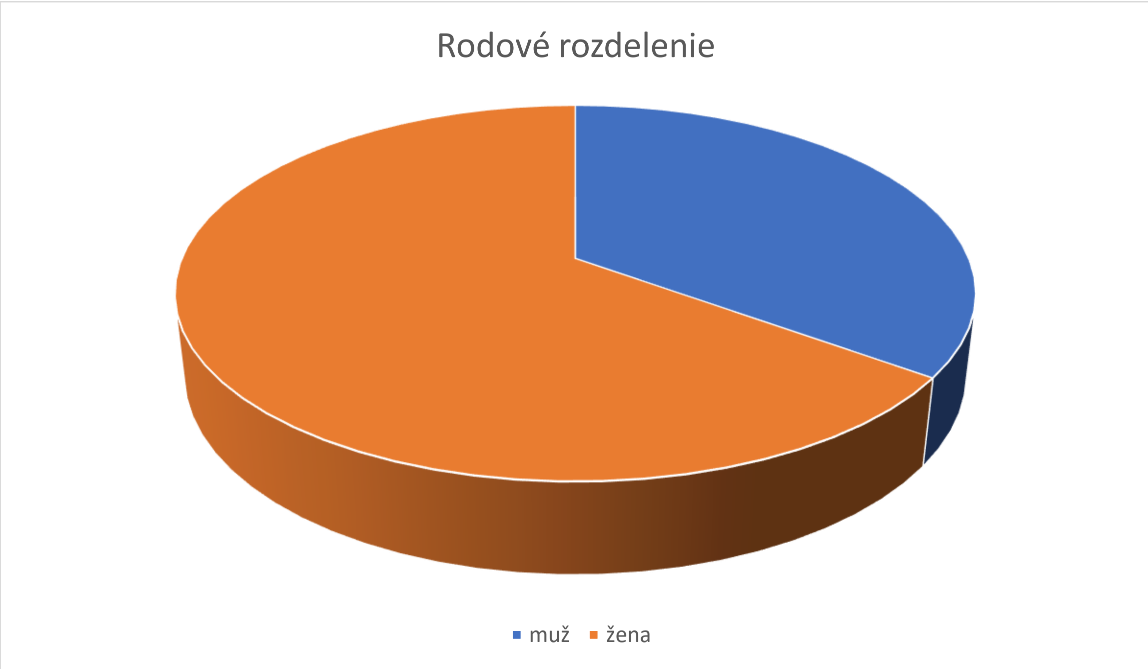
Jednorozmerné grafické metódy

<u>*Koláčový graf*</u>

Je jedna zo základných metód grafického zobrazenia dát. Zostrojenie je pomerne jednoduché, stačí poznať relatívne početnosti hodnôt premennej, ktoré sa nachádzajú vo frekvenčnej tabuľke. Zjednodušene povedané, koláčový graf zobrazuje relatívne, či percentuálne rozdelenie hodnôt – odpovedí respondentov na konkrétnu premennú.

Kruh (koláč) v tomto prípade predstavuje celý štatistický súbor, jednotlivé výseky (často farebne označené) zase podiely výskytu konkrétnych hodnôt.

Prakticky sa koláčový graf používa najmä pri prezentácii menšieho počtu hodnôt. Ako príklad použijeme rodové rozdelenie výskumného súboru. Predstavme si, že z frekvenčnej tabuľky premennej rod vieme, že výskumnú vzorku tvorilo 100 respondentov, pričom podiel mužov a žien je 35 k 65. Po dosadení hodnôt bude koláčový graf vyzerať nasledovne:



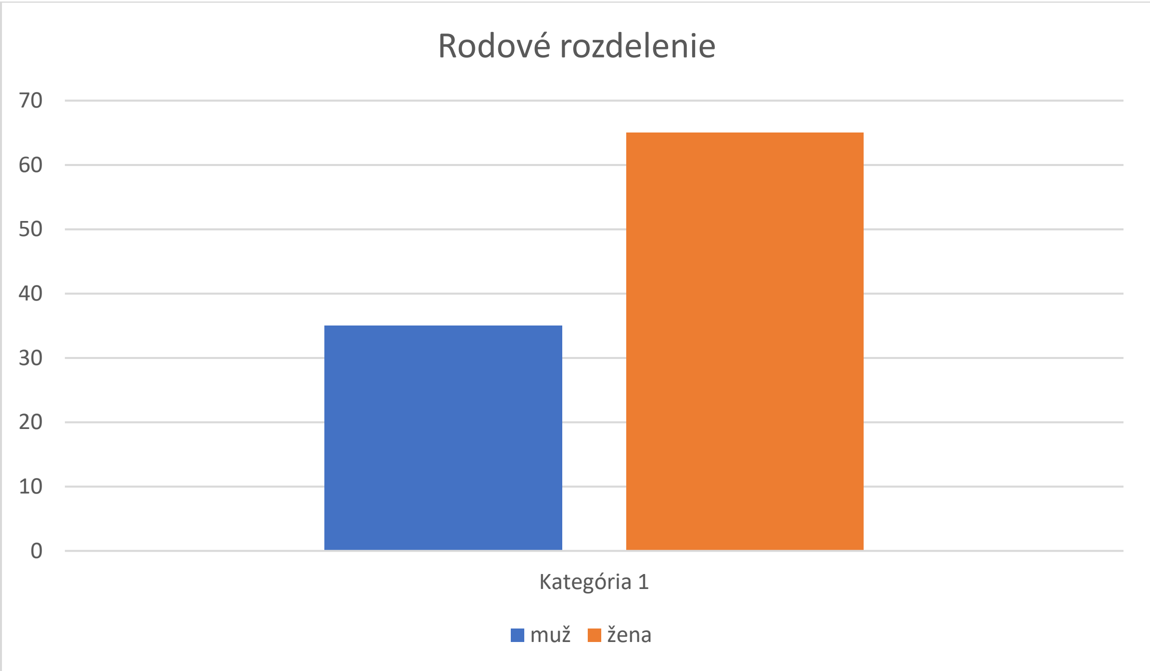
Obr. 10 - koláčový graf – rodové rozdelenie

Pri snahe o prezentáciu väčšieho počtu hodnôt nemusí byť koláčový graf dobrou voľbou. Čím viac rôznych hodnôt máme za cieľ takýmto spôsobom interpretovať, koláčový graf sa stáva menej prehľadným, pričom i rozdiely medzi jednotlivými výsekmi sa pri malých rozdieloch v relatívnych početnostiach interpretujú veľmi ťažko (keďže často sú to iba nepatrné rozdiely jednotlivých výsekov).

<u>Stĺpcový graf</u>

Stĺpcový graf zobrazuje absolútne početnosti premennej, ktoré sú vyjadrené frekvenčnou tabuľkou. Často sa používa aj v kombinácii s vyjadrením ďalších ukazovateľov zobrazujúcich centrálnu tendenciu – napr. vyznačenie aritmetického priemeru či mediánu. Princíp stĺpcového grafu je vyjadrenie absolútnych početností v jednotlivých stĺpcoch, ktoré samotnú početnosť graficky znázorňujú.

Pre ilustráciu využijeme rovnaký príklad, ako v prípade koláčového grafu. Absolútne početnosti sa vyjadria prostredníctvom stĺpcov:



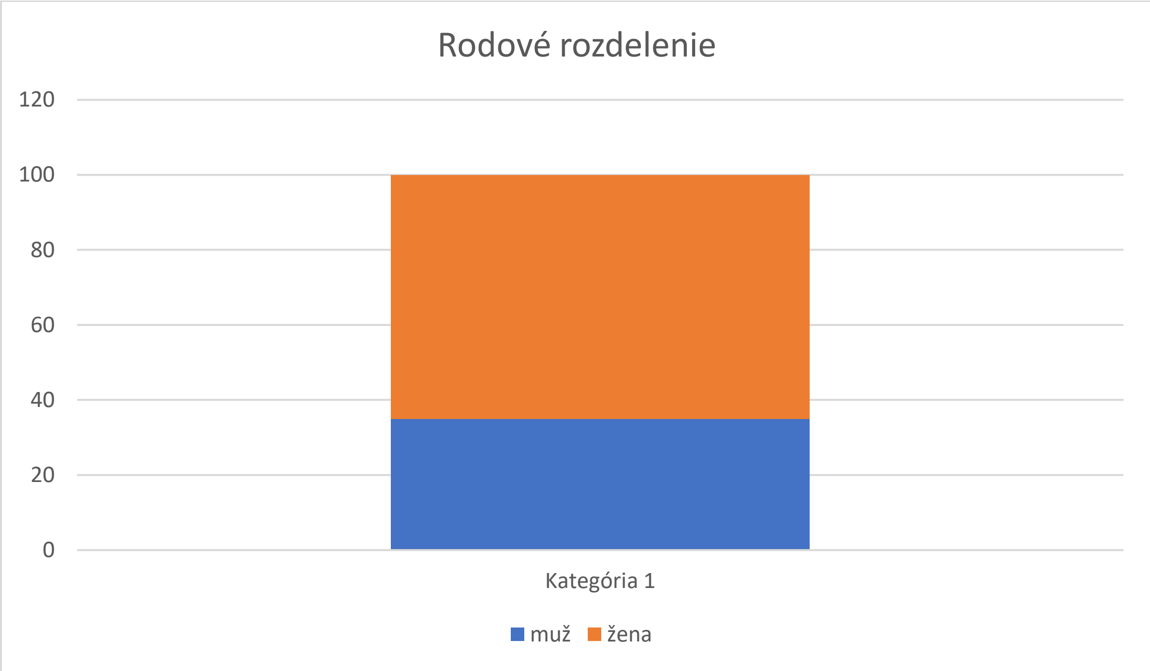
Obr. 11 - stĺpcový graf – rodové rozdelenie

Tento grafický spôsob je na rozdiel od koláčového grafu vhodný i pre prezentáciu viacerých hodnôt. Jednotlivé stĺpce sú ľahšie komparovateľné ako výseky koláčového grafu.

Kumulatívny stĺpcový graf

Používa sa ako alternatíva ku koláčovému grafu s rozdielom v spôsobe zobrazenia, keďže ide o stĺpcové zobrazenie. V tomto prípade je zobrazením relatívnej početnosti, pričom samotný stĺpec predstavuje všetky hodnoty štatistického súboru, početnosť jednotlivých hodnôt je najčastejšie farebne odlíšená.

Pri prezentácii rodového rozdelenia štatistického súboru môžeme použiť i kumulatívny stĺpcový graf. Pre ilustráciu je vhodné si predstaviť, že tento typ grafu môže predstavovať akúsi pomyselnú vežu, ktorá pozostáva z viacerých farebných kociek, ktorých pomer závisí od početnosti jednotlivých hodnôt. Týmto spôsobom je graf vhodný i pre väčší počet rôznych hodnôt premennej.



Obr. 12 - kumulatívny stĺpcový graf – rodové rozdelenie

Na ilustráciu hodnôt je tento graf používaný ako alternatíva ku koláčovému grafu, avšak jednotlivé časti sú v tomto type grafu lepšie čitateľné než v koláčovom. Nevýhodou môže byť jeho následný rozsah – pri väčších počtoch hodnôt je veľmi vysoký, čím sa prakticky vytráca ilustratívna funkcia.

TIP: tieto grafické charakteristiky sa manuálne zostroja relatívne ťažko, keďže je potrebné využívať geometrické pomôcky a rôzne farby. Zároveň je ich zostrojenie pomerne pracné. Preto sa na realizáciu používajú najmä matematické a štatistické programy, ktoré túto prácu značne zjednodušujú

Excel – jednotlivé grafické funkcie predpokladajú, že dáta sú v programe uvedené v stĺpci, resp. stĺpcoch. Vloženie grafov sa realizuje prostredníctvom funkcie Vložiť → Graf → ...v tejto ponuke je potrebné vybrať si príslušný typ grafu...a následne určiť rozsah dát, z ktorých sa má graf vytvoriť.

SPSS v programe SPSS je možné využiť 2 „cesty" zostrojenia grafov. Prvou je využitie funkcie „Chart Builder" v záložke „Graphs". Tu je potrebné vybrať typ grafu a premennú, z ktorej sa má zostrojiť.
Druhým spôsobom je využiť funkciu „Frequencies" v záložke Analyze → Descriptive Statistics. Pričom typ grafu sa zvolí v časti „Charts". Možnosti v tomto prípade sú koláčový graf (Pie charts), stĺpcový graf (Bar charts) a histogram (Histograms).

Histogram

Kým predchádzajúce grafické metódy sú efektívne používané pri zobrazovaní najmä nominálnych typov premenných, histogram je považovaný za ideálny prostriedok pre ilustráciu rozdelenia hodnôt číselných premenných (vrátane intervalových). Histogram identifikuje hodnoty premennej v podobe ich intervalov a početnosti výskytu (absolútne, alebo relatívne) v týchto intervaloch. Histogram sa podobá stĺpcovému grafu, no na osi y sa oproti stpcovému grafu zobrazujú práve intervaly premennej. To spôsobuje, že jednotlivé stĺpce histogramu sú tesne pri sebe.

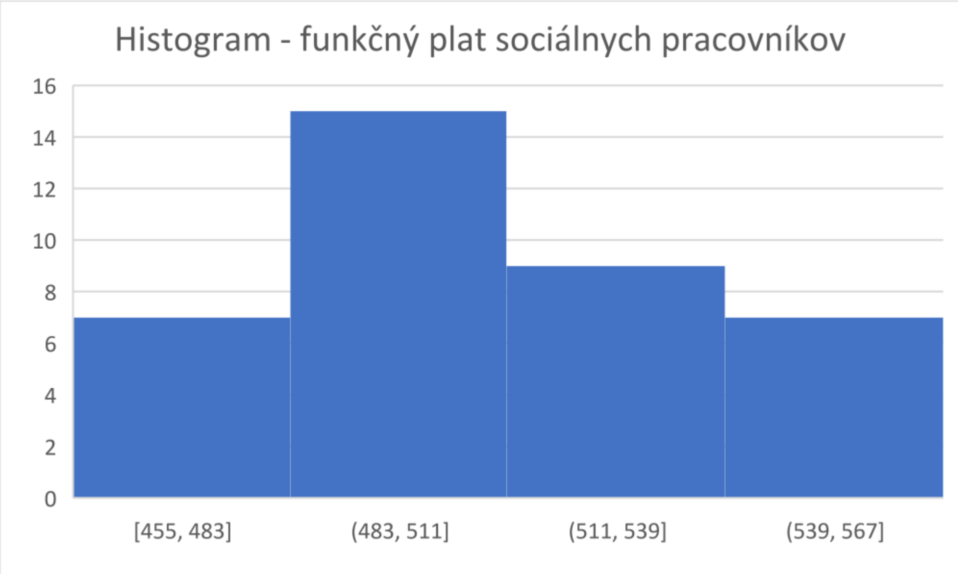
Histogram sa v praxi používa najmä na zachytenie distribúcie premennej. V predchádzajúcej časti tejto učebnice sme diskutovali okrem iného o ukazovateľoch distribúcie. To sme relizovali komparovaním distribúcie hodnôt so symetrickým rozdelením, pričom sme pre ilustráciu používali krivky distribúcie. V reálnej štatistickej praxi sa však krivka distribúcie hodnotí práve na základe histogramu, kedy sa vrcholy jednotlivých stĺpcov pretnú krivkou. Z dôvodu, že sa histogram používa i takýmto spôsobom je jednou zo základných štatistických metód, ktoré sa používajú na identifikáciu povahy premennej a posúdenie vhodnosti ďalších štatistických analýz.

Prakticky sa zobrazenie histogramu realizuje v niekoľkých na seba nadväzujúcich krokoch. V prvom rade je potrebné identifikovať premennú, z ktorej sa má zobraziť, posúdiť jej číselnúpovahu. V druhom kroku je potrebné realizovať usporiadanie hodnôt premennej do tried, o čom pojednáva táto učebnica v jej prvej časti. Triedy by mali byť rovnako široké, pričom sa odporúča, ak by mali šírku aspoň 5 hodnôt.. Po tomto utriedení je možné zostrojiť histogram v podobe grafu, pričom os y zobrazuje identifikované triedy (alebo intervaly) a os x ich absolútnu alebo relatívnu početnosť v súbore.

TIP: rovnako ako v prípade ostatných grafov, použitie software značne uľahčuje i zostrojenie histogramu. Pričom výhodou je, že dáta sa nemusia manuálne utriediť ako by to bolo pri manuálnom zostrojení

Excel – histogram je v programe Excel dostupný rovnako ako predchádzajúce grafy vo funkcii Vložiť → Graf.

SPSS obdobne ako pri iných typoch grafov je histogram dostupný v záložkách „Graphs“ → „Chart Builder“, ako i v záložke Analyze → Descriptive Statistics → „Frequencies“, kde sa v ponuke „Charts“ zvolí Histogram.



Obr. 13 - histogram premennej - funkčný plat

Pre ilustráciu sme v predchádzajúcom obrázku použili histogram vypracovaý v programe Excel. Ako je možné vidieť, pre výpočet histogramu bola použitá premenná Funkčný plat z cvičenia v predchádzajúcej časti pojednávajúcej o číselných deskriptívnych metódach. Šírka jednotlivých tried bola stanovená na „28“, pričom takýmto spôsobom boli prijaté 4 triedy.

Na základe uvedeného histogramu je možné identifikovať typ distribúcie, rovnako je možné identifikovať mediánový i modusový interval (ako interval s najyššou početnosťou). Tento typ grafu tak značne pomáha pri ďalšom štatistickom spracovaní.

Škatuľový graf

Škatuľový graf označovaný ako „Box Plot“ je prostriedkom grafických deskriptívnych metód, ktorý sa používa na zobrazenie viacerých charakteristík premennej (minimálna, maximálna hodnota, kvartily, medián a aritmetický priemer). Je to vhodný nástroj pre grafické zobrazenie ukazovateľov centrálnej tendencie a rovnako sa dá použiť i ako ukazovateľ tvaru distribúcie vrátane smeru zošikmenia distribúcie.

Pre zostrojenie škatuľového grafu je potrebné poznať charakteristiky, ktoré sa doň priamo vyznačujú, konkrétne minimálna hodnota, maximálna hodnota, kvartily, medián, aritmetický priemer.

Škatuľový graf, ako jeho názov napovedá sa zostrojí ako „škatuľa“, ktorá identifikuje centrálne hodnoty premennej. Pracuje teda s hodnotami, nie s početnosťou. Na jednu z osí grafu (podľa použitej osi je *spracovaný ako vertikálny, alebo horizontálny) sa uvádza rozsah hodnôt premennej. Samotná škatuľa začína na hodnote prvého kvartilu a končí na hodnote tretieho kvartilu. Okrem týchto hodnôt sa do grafu vyznačujú minimálna hodnota a maximálna hodnota, ktoré sa so škatuľou spájajú čiarou. Do škatule sa môže vyznačiť hodnota mediánu a rovnako aritmetického priemeru.*

Obr. 14 - škatuľový graf premennej - prax v rokoch

Predchádzajúci obrázok identifikuje škatuľový graf premennej Prax v rokoch z cvičenia na konci predchádzajúcej kapitoly. Z tohoto grafu je možné zistiť hodnoty nasledujúcich ukazovateľov: Min = 1; Max = 9; $\tilde{x}_{25\%} = 3$; $\tilde{x}_{75\%} = 7$; $\tilde{x} = 5$; $\overline{x} = 4{,}81$ *(vyznačený oranžovou). Na tomto princípe je rovnako možné identifikovať tvar distribúcie.*

Dvoj a viacrozmerné grafické metódy

I grafické metódy je možné realizovať pre zobrazenie viacerých premenných. Pre tento účel je možné použiť niektoré doposiaľ uvedené grafy, alebo použiť niektoré, ktoré sú pre takéto účel priamo navrhnuté. Prvá možnosť ako zobraziť viacero premenných je použitie stĺpcového grafu, ktorý by mohol byť rozšírený o ďalšiu premennú. Prakticky sa tento princíp používa s cieľom komparovať početnosť rovnakých odpovedí vo viacerých premenných. Rovnako je možné použiť kumulatívny stĺpcový graf.

Z hľadiska samostatných grafických metód, ktoré by boli navrhnuté ako dvoj, alebo viacrozmerné sú najznámejšie rozptylový diagram a bivariačný histogram.

Rozptylový diagram

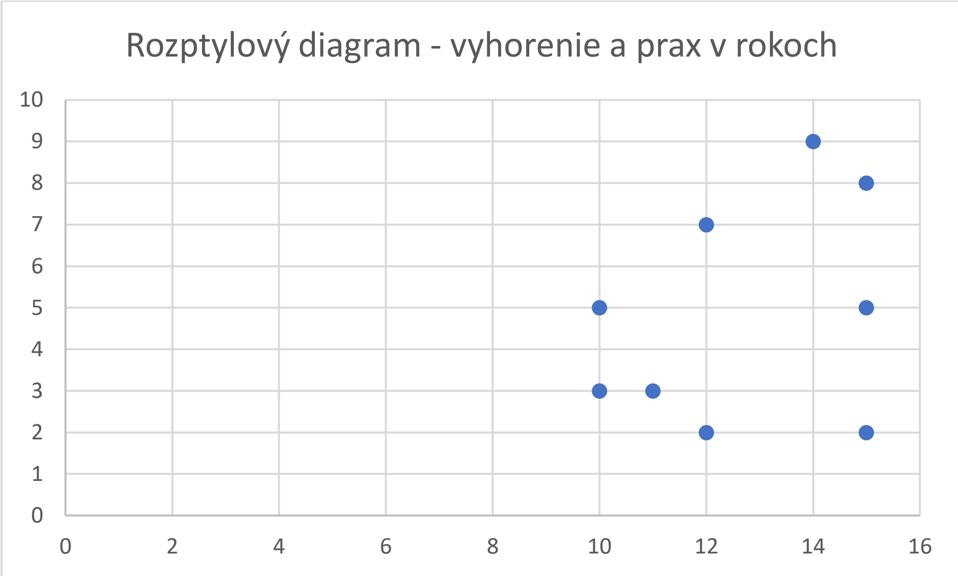
Rozptylový diagram, tiež označovaný ako X-Y graf, anglicky „Scatter Plot" je v štatistike významným vo vzťahu k ďalšiemu štatistickému spracovaniu. Graficky znázorňuje vzájomný vzťah medzi dvoma číselnými premennými prostredníctvom ich kombinácie. Pracuje priamo s hodnotami týchto premenných u jednotlivých štatistických jednotiek v súbore. Na základe rozptylového diagramu je možné prakticky veľmi jednoducho odhadnúť, či hodnoty dvoch premenných spoločne rastú/klesajú, alebo či spolu zdanlivo nesúvisia.

Princíp zostrojenia rozptylového diagramu je veľmi jednoduchý. Základom je dátový súbor s minimálne dvoma premennými u každej štatistickej jednotky. Ako príklad môžeme použiť výskum, ktorý pri pracovníkoch v sociálnej oblasti identifikoval mieru vyhorenia. Okrem tejto premennej sa zaujímal o dĺžku praxe. Zámerom výskumníka je identifikovať vzťah medzi týmito premennými grafickou metódou. Dátový súbor je uvedený v nasledujúcej tabuľke.

Tab. 22 - dátový súbor k rozptylovému diagramu – vyhorenie a prax v rokoch

Vyhorenie - skóre	Prax v rokoch
10	3
12	2
10	5
15	8
11	3
15	5
12	7
14	9
15	2

V ďalšom kroku môžeme pristúpiť k spracovaniu samotného rozptylového diagramu. Základom je prideľovanie konkrétnych pozícií (konkrétneho bodu) každej štatistickej jednotke na základe kombinácie dvoch premenných. Na os X sa nanesú hodnoty jednej premennej (v našom prípade vyhorenie), na os Y hodnoty druhej premennej (v našom prípade prax v rokoch). Príslušné kombinácie sa vyznačia v priestore grafu tak ako je to uvedené na obrázku nižšie.



Obr. 15 - rozptylový diagram – vyhorenie a prax v rokoch

Na základe rozptylového diagramu a umiestnenia jednotlivých bodov je možné identifikovať vzťahy medzi týmito dvoma premennými. Ak hodnoty premenných spoločne stúpajú, priestorom grafu by bolo možné preložiť priamku, ktorá by mala stúpajúci smer (spoločný rast), ak naopak, jedna premenná bola v poklese, mala by priamka klesajúci smer. Ak by boli body v priestore grafu umiestnené nesúvisiaco v priestore, medzi premennými nie je zdanlivá interakcia, resp. súvislosť. To je taktiež základom pre ďalšie metódy patriace do oblasti inferenčnej štatistiky.

Táto metóda sa dá použiť i ako metóda viacrozmernej štatistiky, kde sa môžu pridať ďalšie osi, čím sa priestor grafu stáva trojrozmerným.

Bivariačný histogram

Bivariačný histogram je metóda deskriptívnej štatistiky, ktorá vychádza z klasického jednorozmerného histogramu. Ku klasickému histogramu sa pridávajú ďalšie premenné v podobe ďalších rozmerov, čím sa stáva trojrozmerným.

Cvičenie

Nasledujúca tabuľka reprezentuje výsledky výskumu realizovaného na pracovníkoch v sociálnej oblasti, ktorý sa zaoberal viacerými charakteristikami. Preštudujte si prosím odpovede respondentov a odpovedajte na otázky uvedené pod tabuľkou.

Tab. 23 - grafické deskriptívne metódy - cvičenie

Rod	Vek	Pracovná pozícia	Počet rokov v praxi	Skóre dotazníka Únava z pomáhania
Muž	36	Sociálny pracovník	15	19
Žena	42	Terapeut	18	15
Žena	28	Sociálny pracovník	5	14
Žena	23	Sociálny pracovník	1	18
Muž	42	Inštruktor	20	21
Muž	39	Ošetrovateľ	19	26
Žena	41	Ošetrovateľ	15	11
Žena	56	Riaditeľ	30	14
Muž	30	Sociálny pracovník	5	19
muž	29	Terapeut	3	17

Otázky a úlohy:

1. Navrhnite grafické metódy, ktoré by ste použili pri prezentovaní jednotlivých premenných. Svoje rozhodnutie odôvodnite.
2. Graficky vyjadrite jednotlivé premenné prostredníctvom navrhnutých grafických deskriptívnych metód.
3. Vypracujte škatuľový graf a histogram z číselných premenných.
4. Použite matematický program Excel pre vypracovanie jednotlivých grafických metód.
5. Vypracujte rozptylový diagram z premenných počet rokov v praxi a skóre dotazníka Únava z pomáhania.
6. Je možné medzi premennými z predchádzajúcej úlohy identifikovať vzájomné vzťahy?

Zoznam použitej literatúry

COBB., G. W. & D. S. MOORE. 1997. Mathematisc, Statistics, and Teaching. *The American Mathematical Monthly,* 104(9), 801-823. ISSN 0002-9890.

COOLICAN, H. 2014. *Research Methods and Statistics in Psychology.* London: Psychology Press. ISBN 978-144-417-0115.

DONNELLY, R. A. 2007. The Complete Idiot's Guide to Statistics. New York: Alpha Books. ISBN 1-4295-1390-X.

FERJENČÍK, J. 2006. *Základy štatistických metód v sociálnych vedách*. Košice: UPJŠ. ISBN 80-7097-639-X.

FIELD, A. 2009. *Discovering Statistics Using SPSS.* London: SAGE. ISBN 978-1-84787-906-6.

FISHER, M. & A.P. MARSHALL. 2009. Understanding descriptive statistics. *Australian Critical Care Nurses,* 22(2), 93-97. ISSN 1036-7314.

HENDL, J. 2012. Přehled statistických metod: analýza a metaanalýza dat. Praha: Portál. ISBN 978-80-262-0200-4.

KANDEROVÁ, M. a V. ÚRADNÍČEK. 2007. *Štatistika a pravdepodobnosť pre ekonómov.* Banská Bystrica: oz FINANC. ISBN 978-80-969535-5-4.

KULČÁR, L. 2007. Harmonický priemer a jeho praktická aplikácia. In *Matematika v škole dnes a zajtra. Zborník príspevkov* [online]. Ružomberok: Katedra Matematiky PF KU. Dostupné na: http://math.ku.sk/data/portal/data/zbornik2007/Articles/Kulcar_Ladislav.pdf

PALLANT, J. 2007. *SPSS Survival Manual.* Berkshire: Open University Press. ISBN 978-0-335-22366-4.

RIMARČÍK, M. 2007. *Štatistika pre prax.* Košice: Marián Rimarčík. ISBN 978-80-969813-1-1.

SALKIND, N.J. (ed.) 2010. *Encyclopedia of Research Design.* Thousand Oaks: SAGE Publications. ISBN 978-1-4129-6127-1.

SODOMOVÁ, E. a kol. 2013. *Štatistika pre bakalárov.* Bratislava: Ekonóm. ISBN 978-80-225-4283-8.

ZLACKÁ, A. 2004. *Základy kvantitatívnych metód. Skriptá.* [online] Prešov: FHPV PU. Dostupné na: http://www.fhpv.unipo.sk/cvt/statistika/zlacka/geoinfo1.pdf

Základy štatistiky v sociálnych vedách I.
Teoretické východiská a deskriptívna analýza

Vysokoškolská učebnica

Autor:	Mgr. Vladimír Lichner, PhD.
Vydavateľ:	Univerzita Pavla Jozefa Šafárika v Košiciach Vydavateľstvo ŠafárikPress
Rok vydania:	2020
Počet strán:	83
Rozsah:	4,2 AH
Vydanie:	prvé

ISBN 978-80-8152-925-2 (e-publikácia)