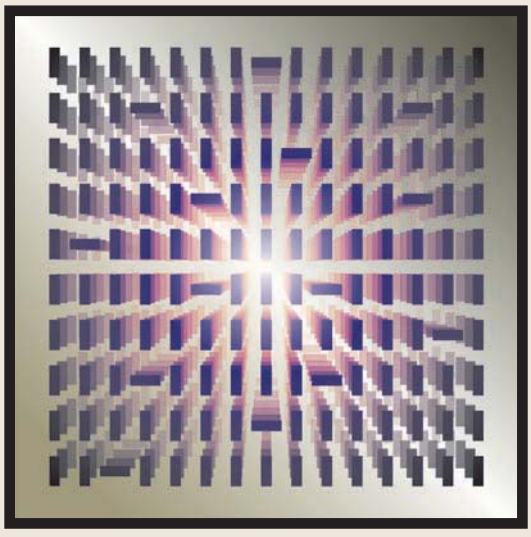
EDICE UČEBNÍCH TEXTŮ

KVANTITATIVNÍ METODY

STATISTIKA PRO EKONOMY

EDUARD SOUČEK



VYSOKÁ ŠKOLA EKONOMIE A MANAGEMENTU

Statistika pro ekonomy

Eduard Souček

Statistika pro ekonomy

VYSOKÁ ŠKOLA EKONOMIE A MANAGEMENTU
Praha 2007

Úvodem

Cílem této učební pomůcky je podat výklad základních statistických metod, s kterými ekonom přichází v praxi do styku a které nacházejí široké uplatnění při zpracování, prezentaci a analýze hospodářských a sociálních jevů. Výběr metod a způsob jejich objasnění je podřízen zájmu na zdůraznění postupů a aplikací, které jsou typické pro analytickou a rozhodovací činnost ekonomů a manažerů.

Obecně platí, že ideové zvládnutí statistického přístupu k hodnocení čísel zobrazujících reálný svět má dvojí význam. V prvé řadě je předpokladem pro kvalifikované využívání číselných informací, s kterými se v ekonomickém prostředí denně setkáváme. V druhé řadě je to nezbytný první krok pro racionální uplatnění výpočetní techniky v práci se statistickými daty. I v oblasti aplikace statistických metod existuje bohatá nabídka specializovaného statistického softwaru, jehož účelné využívání však vyžaduje dobrou znalost statistických procedur a zejména jejich cílů a podmínek jejich použití.

Skriptum je koncipováno tak, aby obsáhlo všechna základní témata standardního kurzu statistiky. Výklad jednotlivých partií není příliš zatížen popisem teorie a důkazy a akcentuje objasňování praktické stránky statistických metod, jejich použitelnosti při řešení typických statistických úloh a také při řešení problémů spojených s interpretací a hodnocením výsledků.

Doc. Ing. Eduard Souček, CSc.

Vysoká škola ekonomie a managementu

Statistika pro ekonomy

Eduard Souček

ISBN 978-80-86730-06-6

Vysoká škola ekonomie a managementu

www.vsem.cz

Obsah

Jak používat tuto učebnici

Tuto knihu můžete jednoduše přečíst od začátku do konce, ale mnohem užitečnější vám bude s perem a papírem. Nejefektivnější formou učení je aktivní učení, a proto jsme naplnili text cvičeními, abyste se přesvědčili, jak učivo zvládáte. Každá kapitola také obsahuje cíle, souhrn kapitoly a rychlý kviz. Následující body vám objasní, jak s knihou pracovat co nejefektivněji:

a) Vyberte si kapitolu, kterou budete studovat, přečtěte si úvod a cíle na začátku kapitoly.

b) Potom si přečtěte souhrn kapitoly na jejím konci (před rychlým kvizem a odpověďmi ke cvičením). Neočekávejte, že tento krátký závěr znamená v této fázi příliš mnoho, ale zkuste, zda můžete spojit některý z probraných bodů s některým z cílů.

c) Poté si přečtěte samotnou kapitolu. Vyřešte jednotlivá cvičení tak, jak jdou za sebou. Největší prospěch ze cvičení získáte, pokud si své odpovědi napíšete předem a poté je zkontrolujete s odpověďmi na konci kapitoly.

d) Při čtení používejte poznámkový sloupec a přidávejte vlastní komentáře, odkazy na další materiál atd. Pokuste se formulovat své vlastní názory. V ekonomii je mnoho věcí otázkou výkladu a často je zde prostor pro alternativní názory. Čím hlubší dialog s knihou provedete, tím více ze svého studia získáte.

e) Až dočtete kapitolu, znovu si přečtěte souhrn kapitoly. Poté se vraťte k cílům na začátku kapitoly a položte si otázku, zda jste jich dosáhli.

f) Nakonec upevněte své znalosti tím, že písemně vyřešíte příklady v závěru kapitoly. Své odpovědi si můžete zkontrolovat tak, že se podíváte zpět do textu. Návrat k textu a hledání významných detailů dále zlepší pochopení předmětu.

g) Nakonec si zkontrolujte svá řešení v přehledu správných odpovědí, který naleznete v závěru publikace.

Značky a symboly v učebním textu

Struktura distančních učebních textů je rozdílná již na první pohled, a to např. v zařazování grafických symbolů – značek.

Specifické grafické značky umístěné na okraji stránky upozorňují na definice, cvičení, příklady s postupem řešení, klíčová slova a shrnutí kapitol. Značky by měly studenta intuitivně vést tak, aby se již po krátkém seznámení s distanční učebnicí dokázal v textu rychle a snadno orientovat.

Poznámky

Označuje místo pro poznámky (vždy na začátku stránky v širším okraji).

Definice

Upozorňuje na definici nebo poučku pro dané téma.

Cvičení

Označuje jednotlivá číslovaná cvičení, jejichž řešení je uvedeno na konci kapitoly.

Příklad - případová studie

Označuje příklady s postupem řešení na konci kapitoly.

Klíčová slova

Upozorňuje na důležité výrazy či odborné termíny nezbytné pro orientaci v daném tématu.

Shrnutí kapitoly

Shrnutí kapitoly se zařazuje na konec dané kapitoly. Přehledně, ve strukturovaných bodech shrnuje to nejpodstatnější z předchozího textu.

kapitola 1

Popisná statistika

1. kapitola
Popisná statistika

Studium této kapitoly objasní

- **Cíle popisu statistického souboru popisnými charakteristikami.**
- **Způsoby prezentace dat v tabulkových a grafických formách.**
- **Výpočet a použití charakteristik úrovně.**
- **Výpočet a použití charakteristik variability.**
- **Výpočet a použití charakteristik symetrie rozdělení.**

Úvod

Statistický přístup ke zkoumání sociálně-ekonomické reality vychází z potřeby získání základních číselných popisných charakteristik statistického souboru, na základě kterých by bylo možno v přehledné podobě jednoznačně specifikovat vlastnosti hodnoceného souboru. K tomuto účelu slouží především dvě základní kategorie popisných měr: míry úrovně a míry variability hodnot. Znalost těchto měr je nejen výchozím bodem každé věcné analýzy, ale i podmínkou pro případné komparace více statistických souborů.

Vznik statistiky

Termín statistika je odvozen od latinského „status“, což v latině znamená „stav“ a ve slovním spojení „status rei republicae“ je to „stav věci veřejné“ neboli stát. Od tohoto významu vznikla v 16. a 17. století italská slova „statistica“ pro označení souhrnu znalostí o státních záležitostech. Tento termín se pak rozšířil v podobném významu i mezinárodně.

Činnosti blízké statistice však mají daleko starší historii. Známá jsou sčítání lidí před několika tisíciletími v Egyptě a v Číně. Běžná byla zjišťování pro účely vojenské a daňové ve starém Římě.

S prvními badatelskými aplikacemi statistiky se setkáváme v Anglii (John Graunt, 1620–1674, a William Petty, 1623–1687), kdy byla shromažďována data pro zkoumání pravidelností v úmrtnosti a porodnosti obyvatelstva. Graunt a Petty již usilovali o zobecnění významu jednotlivých případů tím, že zkoumali skutečnosti, které mají povahu hromadného jevu. Svůj postup zkoumání označil Petty jako „politickou aritmetiku“, aby tak vyjádřil fakt, že zkoumá skutečnosti důležité pro stát a současně, že jde o číselné charakterizování hodnocených jevů.

Významným vkladem pro teoretické zázemí statistických metod byl rozvoj počtu pravděpodobnosti. První kroky počtu pravděpodobnosti jsou spojeny s matematickými výpočty u hazardních her. Další vývoj teorie pravděpodobnosti je spojen se jmény slavných matematiků (B. Pascal, J. Bernoulli, T. Bayes, P. S. Laplace, K. F. Gauss, P. L. Čebyšev, A. A. Markov a další).

Pojetí statistiky

Pojem statistika se v současnosti používá ve třech významech:

a) pro vyjádření souhrnu dat o hromadných jevech,

b) pro činnost směřující k získávání statistických dat, jejich uspořádání a zpracování a následnou prezentaci,

c) pro metodologickou vědu, jejímž cílem je zkoumání zákonitostí hromadných jevů a kterou tvoří metodologie zjišťování, zpracování a analýzy dat.

Chápeme-li statistiku v uvedeném třetím významu, tedy jako metodologickou vědu, zjistíme, že jsou pro ni příznačné dvě skutečnosti:

1. Jejím předmětem jsou hromadné jevy, ne jevy jedinečné a neopakovatelné. Znamená to, že statistiku nezajímá konkrétní jedinec (předmět, objekt, událost) sám o sobě, ale jen jako součást souboru jedinců. Cílem statistiky je generalizace založená na zkoumání souborů případů.

2. Zkoumané poznatky o hromadných jevech vyjadřuje statistickými daty.

V tomto pojetí, jež chápe statistiku jako metodologickou disciplínu, která zkoumá svými specifickými metodami hromadné jevy, se bude statistikou zabývat tento učební text.

1.1 Základní statistické pojmy

1.1.1 Statistický soubor a statistická jednotka

Zkoumání hromadných jevů předpokládá definování – z hlediska účelu zkoumání – vymezené množiny objektů, prvků zkoumání neboli **statistického souboru** (soubor podniků, soubor obyvatelstva, soubor událostí apod.). Jednotlivé objekty, prvky statistického souboru, označujeme jako **statistické jednotky**. Jsou nositeli vlastností daného souboru. Počet jednotek statistického souboru se nazývá **rozsah souboru**.

Soubory, které jsou předmětem zkoumání, označujeme jako **základní soubor** (někdy se základní soubor označuje jako populace). V praxi často z různých důvodů nepracujeme s celým rozsahem statistického souboru, ale jen se vzorkem statistických jednotek neboli s **výběrovým souborem**. K tomu dochází buď proto, že zkoumání celého statistického souboru by bylo nákladné, časově zdlouhavé nebo z jiných praktických ohledů neuskutečnitelné, a dále proto, že zobecnění provedené z dat výběrového souboru považujeme pro daný účel zkoumání za dostatečně přesné a z hlediska poznání za reprezentativní.

1.1.2 Statistický znak

Zkoumané vlastnosti statistického souboru sleduje statistika prostřednictvím měřitelných vlastností statistických jednotek, které vyjadřuje tzv. statistickými znaky. **Statistický znak** nabývá slovních nebo číselných hodnot a je zjišťován u každé statistické jednotky statistického souboru. Jestliže ve statistickém souboru pracujeme jen s jedním znakem (s jednou proměnnou), říkáme, že se jedná o **jednorozměrný soubor**, máme-li současně více znaků, jde o dvou-, tří-, resp. obecně **vícerozměrný soubor**.

Základním tříděním statistických znaků je rozlišování znaků **číselných** (kvantitativních, numerických) a znaků **slovních** (kvalitativních, alfabetických, kategoriálních).

Číselné statistické znaky bezprostředně vyjadřují sledované vlastnosti čísly (např. při zkoumání souboru pracovníků podniku jsou to znaky jako mzda, věk, doba praxe). Rozlišujeme **znaky spojité** (kontinuální), které mohou teoreticky nabývat libovolných reálných číselných hodnot v určitém intervalu (průtok vody, hmotnost výrobku, výška, peněžní obrat apod.) a **znaky nespojité** (diskrétní), které mohou nabývat pouze určitých číselných hodnot v oboru reálných čísel (počet pracovníků, počet prodaných výrobků, počet členů domácnosti apod.).

Jsou-li hodnoty statistického znaku vyjádřeny slovně, nazývá se takový znak **slovní** (např. u osob je to vzdělání, odvětví činnosti, národnost, pohlaví). Zvláštní skupinou slovních statistických znaků jsou **ordinální (pořadové) znaky**. Ty jsou takové, že jejich obměny lze podle nějakého objektivního kritéria seřadit od nejmenší obměny do největší, např. na základě nějakého expertního ohodnocení. Taková situace vzniká kupř. při posuzování kvality výrobku, kdy výrobky jsou na základě hodnocení expertů seřazeny od nejlepšího k nejhoršímu. Namísto slovního popisu obměn pak u ordinálních znaků můžeme pracovat s pořadovými čísly jako s určitou formou kvantifikace těchto obměn.

1.2 Zjišťování a prezentace statistických dat

Statistické zkoumání prochází postupně několika pracovními etapami. Výchozí etapou je **statistické zjišťování (statistické šetření)**. Cílem je získávání neznámých statistických dat o hodnotách statistických znaků u jednotlivých statistických jednotek, které tvoří statistický soubor. Každé statistické zjišťování má určitý konkrétní účel, z kterého vyplývá, jaké proměnné statistické znaky budeme zjišťovat, co zvolíme za statistickou jednotku a jak vymezíme statistický soubor. Organizace statistického zjišťování musí obsahovat **prostorové, věcné a časové vymezení statistického souboru a statistických znaků**.

Např. při zjišťování ekonomických výsledků průmyslových podniků musí organizátor šetření stanovit, zda bude prostorově vymezen okruh průmyslových podniků územím České republiky nebo nějakým jiným regionem a zda o zařazení podniku do konkrétního území bude rozhodovat umístění sídla podniku nebo nějaké jiné hledisko. Věcné vymezení musí definovat, co považujeme za průmyslový podnik a jakými ukazateli budeme charakterizovat ekonomické výsledky každého podniku (objem produkce, rentabilita, produktivita práce, zisk apod.). Při časovém vymezení půjde o stanovení konkrétního časového intervalu nebo rozhodného časového okamžiku, ke kterému se budou jednotlivé zjišťované údaje vztahovat.

Elementární zpracování výsledků statistického zjišťování

Výsledky statistického zjišťování mají obvykle povahu velkého a nepřehledného množství číselných údajů, které je třeba pro analýzu vhodně uspořádat a utřídit. **Tříděním** rozumíme rozdělení jednotek souboru do skupin tak, aby vynikly charakteristické vlastnosti zkoumaných jevů. Provádíme-li třídění podle obměn jednoho statistického znaku, mluvíme o třídění jednostupňovém. Třídění podle více statistických znaků najednou označujeme jako třídění vícestupňové.

Je-li třídicím znakem číselný (kvantitativní) znak s malým počtem obměn, pak vhodným uspořádáním statistických dat je **tabulka rozdělení četností**, kdy napozorované hodnoty nejprve uspořádáme podle velikosti a ke každé variantě přiřadíme počty statistických jednotek, které udávají, s jakou četností se jednotlivé varianty hodnot vyskytují. Označíme-li obměny číselného statistického znaku x_i a četnosti n_i a předpokládáme-li, že tříděním vzniklo k obměn, pak tabulku rozdělení četností lze formálně vyjádřit takto:

TABULKA 1.1 Rozdělení četností

Obměna hodnoty znaku	Četnost
x_i	n_i
x_1	n_1
x_2	n_2
⋮	⋮
x_k	n_k
Celkem	n

Souhrn četností za k řádků $n_1 + n_2 + \ldots + n_k$ je roven rozsahu souboru n: $\sum_{i-1}^{k} n_i = n$.

Tímto způsobem lze především vyjadřovat rozdělení četností nespojitého statistického znaku. Např. při prezentaci velikostní struktury souboru domácností budou obměnami hodnot znaku jednotlivé vyskytující se varianty počtu členů domácností a četnostmi jsou údaje o počtu domácností u jednotlivých obměn.

Sledujeme-li nespojitý statistický znak s velkým počtem obměn nebo pracujeme-li se spojitým statistickým znakem, pak uvedený způsob prezentace výsledků statistického šetření by nepřinesl žádoucí zpřehlednění statistických dat. V takových případech namísto obměn jednotlivých číselných hodnot přecházíme na intervaly hodnot a přehlednost výsledků regulujeme počtem a šířkou zvolených intervalů. Výsledná tabulka je označována jako **intervalové rozdělení četností**.

Při sestavování intervalového rozdělení četností je třeba především vyřešit problém stanovení počtu a tím velikosti intervalů. Obvykle volíme řešení, které neohrožuje příliš informační hodnotu výsledků. Příliš široké intervaly snižují kvalitu prezentace, příliš úzké naopak zhoršují přehlednost a zvyšují rozsah tabulky.

Dalším problémem intervalového rozdělení četností je volba hranic intervalů, aby nedocházelo k nejasnostem, do kterého intervalu se mají jednotlivé jednotky zařadit. Nejčastěji se hranice intervalů volí tak, aby se intervaly nepřekrývaly. Např. při charakterizování věkové struktury obyvatelstva pětiletými věkovými skupinami se používají intervaly 0–4, 5–9, 10–14, 15–19 atd. V praxi se často neobejdeme bez tzv. otevřených intervalů, při jejich použití bychom však měli být opatrní a používat je jen pro intervaly s malou četností, kde nehrozí nebezpečí příliš velké informační ztráty. Např. u již zmíněné věkové struktury obyvatelstva to může být otevřený interval: 85 a více let.

Při výpočtech statistických charakteristik vzniká problém, jaká hodnota by ve výpočtu měla zastoupit (reprezentovat) jednotlivé intervaly. Za tuto zastupitelnou hodnotu se zpravidla volí střed intervalu.

Grafy rozdělení četností

Nejznámějším grafem rozdělení četností je tzv. **polygon** (řecky mnohoúhelník), který v pravoúhlém souřadnicovém systému používá osu x pro obměny znaku x a osu y pro četnosti n_1. Pro grafické vyjádření intervalového rozdělení četností se používá histogram. Velikost četností je vyjádřena sloupci, jejichž základna je rovna šířce intervalu.

A. Polygon četností

Příklad: „Rozdělení četností počtu žáků podle známky z matematiky“

OBRÁZEK 1.1 Polygon četnosti

Známka	Počet žáků
1	8
2	18
3	14
4	6
5	4
Celkem	50

B. Histogram četností

Příklad: „Intervalové rozdělení četností počtu škol podle průměrného počtu žáků na 1 třídu“

OBRÁZEK 1.2 Histogram četnosti

Průměrný počet žáků na třídu	Počet škol	Střed intervalu
16 – 17,99	6	17
18 – 19,99	10	19
20 – 21,99	22	21
22 – 23,99	16	23
24 – 25,99	10	25
26 – 27,99	4	27
28 – 29,99	2	29
Celkem	70	X

V případě, že jednotlivé intervaly zastoupíme středy intervalů, můžeme intervalové rozdělení četností graficky vyjádřit i polygonem.

Relativní a kumulativní četnosti

Abychom mohli vzájemně porovnávat různá rozdělení četností a jejich struktury v různě velkých statistických souborech, používáme namísto absolutních četností **relativní četnosti** p_i, které získáme jako poměr dílčích četností a rozsahu souboru:

$$p_i = \frac{n_i}{n} . \qquad (1.1)$$

U souboru většího rozsahu se relativní četnosti zpravidla vyjadřují v procentech.

Pro analýzy struktury souboru z hlediska určité vlastnosti může být také užitečné zjistit, jaký podíl jednotek má hodnotu menší nebo rovnou příslušné variantě. K tomu používáme tzv. **kumulativní četnosti** (absolutní nebo relativní). Získáme je postupným načítáním četností po sobě následujících tříd.

PŘÍKLAD 1.1

Za podnik máme k dispozici intervalové rozdělení četností hodinových mezd v členění na muže a ženy.

Interval hodinových mezd v Kč	Počet pracovníků		Relativní četnosti v %		Kumulativní relativní četnosti v %	
	Muži	**Ženy**	**Muži**	**Ženy**	**Muži**	**Ženy**
20 – 29,9	40	24	8	12	8	12
30 – 39,9	80	36	16	18	24	30
40 – 49,9	100	60	20	30	44	60
50 – 59,9	150	48	30	24	74	84
60 – 69,9	90	20	18	10	92	94
70 – 79,9	25	12	5	6	97	100
80 a více	15	–	3	–	100	100
Celkem	500	200	100	100	X	X

Příklad ilustruje, jak je možno řešit problém nepřekrývání intervalů. Interval v posledním řádku označujeme jako otevřený interval.

1.3 Kvantily

Kvantil je hodnota proměnné určená tak, že odděluje určitý podíl jednotek, které jsou menší než tato hodnota. Např. dvacetipětiprocentní kvantil $\tilde{x}_{25}$ odděluje 25 % malých hodnot a současně 75 % velkých hodnot. Tímto způsobem můžeme pak, kupř. při hodnocení úrovně mezd pracovníků v národním hospodářství, charakterizovat, jaká mzdová hranice odděluje 25 % pracovníků s nejnižšími mzdami.

V praxi se používají zejména tyto skupiny kvantilů:

Kvartily ($\tilde{x}_{25}$, $\tilde{x}_{50}$, $\tilde{x}_{75}$) patří mezi kvantily, které rozdělují uspořádanou řadu hodnot na 4 stejné části: první (dolní) kvartil $\tilde{x}_{25}$, který odděluje 25 % jednotek s nejnižšími hodnotami, druhý (prostřední) kvartil $\tilde{x}_{50}$, který odděluje 50 % jednotek s nízkými hodnotami a 50 % hodnot s vysokými hodnotami. Tento padesátiprocentní kvantil se také označuje jako **medián** (od latinského medius – prostřední). Třetí kvartil (horní) $\tilde{x}_{75}$ odděluje 75 % jednotek s nízkými hodnotami od 25 % jednotek s vyššími hodnotami.

Decily ($\tilde{x}_{10}$, $\tilde{x}_{20}$,..., $\tilde{x}_{90}$) rozdělují uspořádanou řadu na 10 stejných částí.

Centily, resp. **percentily** ($\tilde{x}_{1}$, $\tilde{x}_{2}$,..., $\tilde{x}_{99}$) rozdělují uspořádanou řadu hodnot na 100 stejně početných částí.

Nejužívanějším kvantilem je **medián**, který představuje prostřední hodnotu uspořádaného souboru, a je tedy svou vypovídací hodnotou blízký aritmetickému průměru. Je-li rozsah souboru udán sudým číslem, obsahuje soubor dvě prostřední hodnoty. V tomto případě bývá zvykem volit za medián průměr z těchto dvou prostředních hodnot a medián pak není konkrétní hodnotou původního souboru. Mediánu dáváme přednost před aritmetickým průměrem v těch situacích, kdy aritmetický průměr je výrazně ovlivněn existencí extrémních hodnot v souboru a poskytuje zkreslený obraz o úrovni hodnot, zatímco hodnota, která v daném souboru je co do velikosti prostřední, je vůči extrémům imunní.

Z povahy kvantilů je zřejmé, že prvním krokem při jejich výpočtu je uspořádání všech hodnot sledovaného znaku podle velikosti. Pak stanovíme pořadové číslo statistické jednotky, jejíž hodnota je hledaným kvantilem. Označíme-li toto pořadové číslo z_p, pak platí:

$$z_p = np + 0{,}5, \tag{1.2}$$

kde n je rozsah souboru a p je relativní četnost nejnižších hodnot. Např. pořadové číslo z_p pro 1. kvartil ($\tilde{x}_{25}$) v souboru n = 80 zjistíme takto: $z_{25} = 80 \cdot 0{,}25 + 0{,}5 = 20{,}5$.
Při odvozování pořadového čísla z_p z četností vyjádřených v procentech se hodnota 0,5 ve vzorci obvykle zanedbává.
Poněkud složitější je výpočet kvantilů z intervalového rozdělení četností. Pokud se spokojíme pouze s určením intervalu, v němž hledaný kvantil leží, je postup stejný jako v předchozím případě. Chceme-li kvantil odhadnout jedním konkrétním číslem, je třeba použít při výpočtu lineární interpolaci založenou na předpokladu, že ve stejných proporcích, v jakých rozděluje pořadové číslo hledaného kvantilu interval četností, rozděluje kvantil interval hodnot. Tento postup hypoteticky předpokládá, že v intervalu, kde leží hledaný kvantil, jsou hodnoty rozděleny rovnoměrně.

PŘÍKLAD 1.2

Hledáme hodnotu všech tří kvartilů ($\tilde{x}_{25}$, $\tilde{x}_{50}$, $\tilde{x}_{75}$) v rozdělení četností hodinových mezd v návaznosti na údaje z příkladu 1.1. Výpočet provedeme zvlášť za muže a ženy. Využijeme k tomu poslední dva sloupce obsahující v procentech vyjádřené kumulativní četnosti:

Interval hodinových mezd v Kč	Relativní četnosti v %		Kumulativní relativní četnosti v %	
	Muži	**Ženy**	**Muži**	**Ženy**
20 – 29,9	8	12	8	12
30 – 39,9	16	18	24	30
40 – 49,9	20	30	44	60
50 – 59,9	30	24	74	84
60 – 69,9	18	10	92	94
70 – 79,9	5	6	97	100
80 a více	3	–	100	100
Celkem	100	100	X	X

Pro stanovení jednotlivých kvartilů potřebujeme zjistit k pořadovým číslům z_{25}, z_{50} a z_{75} odpovídající hodnoty mezd:

Hodinové mzdy mužů

Ze sloupce kumulativních četností zjistíme, že pořadové číslo 25 patří do třetího intervalu s hodnotami 40 až 49,9 Kč, chápané vždy zaokrouhleně jako 50 Kč. Z těchto podkladů můžeme pro přibližný výpočet prvního kvartilu použít lineární interpolaci, při které bude jeho hodnota rozdělovat tento interval ve stejném poměru, jako pořadové číslo 25 rozděluje odpovídající interval četností:

$$\frac{\tilde{x}_{25} - 40}{50 - 40} = \frac{25 - 24}{44 - 24}.$$

Z toho pak snadno odvodíme, že: $\tilde{x}_{25} = 40 + \frac{1}{20} 10 = 40{,}5.$

Podobně zjistíme, že: $\tilde{x}_{50} = 50 + \frac{6}{30}$ a $\tilde{x}_{75} = 60 + \frac{1}{18} 10 = 60{,}6.$

Hodinové mzdy žen

$$\tilde{x}_{25} = 37{,}2 \qquad \tilde{x}_{50} = 46{,}7 \qquad \tilde{x}_{75} = 56{,}2.$$

1.4 Statistické charakteristiky

1.4.1 Charakteristiky úrovně

Úroveň jevů vyjadřovaných kvantitativními znaky vyjadřují střední hodnoty. Ty v koncentrované podobě shrnují informaci obsaženou v údajích o statistickém znaku. Hlavní skupinu středních hodnot tvoří **průměry** (aritmetický průměr, geometrický průměr, harmonický průměr), jejichž společnou vlastností je, že jsou určovány ze všech naměřených hodnot znaku. Druhou skupinu středních hodnot tvoří tzv. **poziční střední hodnoty** (medián a modus), které jsou určeny pozicí některých jednotek souboru. Medián $\tilde{x}$ je určen hodnotou znaku, kterou má jednotka statistického souboru s hodnotou co do velikosti prostřední. Modus $\hat{x}$ je určen hodnotou znaku u jednotek, které jsou v souboru nejčastěji zastoupeny, jinak řečeno, tou hodnotou souboru, která má největší četnost.

A. Průměry

Aritmetický průměr $\bar{x}$

Je nejznámějším a nejužívanějším typem průměru. Ze zjištěných hodnot x_1, x_2, ... x_n za n-členný statistický soubor jej lze vypočítat takto:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i . \tag{1.3}$$

Tuto formu aritmetického průměru nazýváme **prostý aritmetický průměr**. Výpočet nepředpokládá žádné předběžné uspořádání hodnot. Aritmetický průměr je použitelný všude tam, kde má nějaký informační smysl součet hodnot.

Pokud jsou hodnoty statistického souboru **uspořádány do rozdělení četností**, což je zejména případ velkých souborů a souborů, kde stejné obměny hodnot statistického znaku má vždy více statistických jednotek, předchozí vzorec upravujeme do tvaru, který se označuje jako **vážený aritmetický průměr**. Při jeho použití využíváme skutečnost, že k úhrnu všech hodnot můžeme dospět přes stanovení pomocných součinů $x_i n_i$ pro k obměn znaku. Vzorec váženého aritmetického průměru pak zapisujeme takto:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}, \text{ resp. jako } \bar{x} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i . \tag{1.4}$$

Četnosti $n_1, n_2,..., n_k$ zde vystupují jako váhy k jednotlivým obměnám hodnot.

Máme-li k dispozici intervalové rozdělení četností, bereme při výpočtu aritmetického průměru za hodnoty znaku středy odpovídajících intervalů.

Chceme porovnat aritmetický průměr hodinových mezd mužů a žen v návaznosti na údaje z příkladu 1.2:

PŘÍKLAD 1.3

Interval hodinových mezd v Kč	Relativní četnosti v % n_i		Středy intervalů x_i	$x_i\, n_i$	
	Muži	Ženy		Muži	Ženy
20 – 29,9	8	12	25	200	300
30 – 39,9	16	18	35	560	630
40 – 49	20	30	45	900	1 530
50 – 59	30	24	55	1 650	1 100
60 – 69,9	18	10	65	1 170	650
70 – 79,9	5	6	75	375	450
80 a více	3	–	85	255	–
Celkem	100	100	X	5 110	4 460

Pro výpočet aritmetického průměru z intervalového rozdělení četností použijeme vážený aritmetický průměr, v kterém jsou hodnoty znaku zastoupeny středy intervalů:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i\, n_i}{\sum} \Rightarrow \text{muži} = \frac{5\,110}{100} = 51{,}10 \quad \text{ženy} = \frac{4\,660}{100} = 46{,}60.$$

Použití váženého aritmetického průměru přichází v úvahu i tam, kde váhy nejsou odvozeny z četností, ale z relativního významu (důležitosti) jednotlivých hodnot. Např. při hodnocení likvidity podniku musíme počítat s tím, že jednotlivá aktiva podniku mají různou schopnost využití pro splácení krátkodobých závazků. Proto se v této oblasti setkáváme s tím, že k jednotlivým aktivům jsou na základě expertního ocenění přiřazovány váhy, určující důležitost dané skupiny aktiv z hlediska likvidity podniku. Celkový (průměrný) ukazatel likvidity je pak váženým aritmetickým průměrem z objemů peněžních prostředků, vázaných v jednotlivých skupinách aktiv, kdy jako váhy vystupují nějaké koeficienty kvality aktiv z hlediska stupně likvidity.

PŘÍKLAD 1.4

Při souhrnném hodnocení studijních výsledků z určitého předmětu chceme použít bodových výsledků ze tří testů, dvou průběžných a jednoho závěrečného. Bodům z průběžných testů dáváme stejnou 25% váhu a závěrečnému testu 50% váhu.

Předpokládejme, že student získal v průběžných testech 60 a 80 bodů a v závěrečném 52 bodů.

Celkový průměr $\bar{x} = 1/100\,(60 \cdot 25 + 80 \cdot 25 + 52 \cdot 50) = 61$ bodů.

K důležitým vlastnostem aritmetického průměru patří:

1. Součet odchylek jednotlivých hodnot od jejich aritmetického průměru je nulový.
2. Součet čtverců odchylek jednotlivých hodnot od průměru je minimální.
3. Transformace jednotlivých hodnot přičtením (nebo odečtením) konstanty zvýší (nebo sníží) aritmetický průměr o tuto konstantu.
4. Při transformaci jednotlivých hodnot násobením (nebo dělením) nenulovou konstantou je i aritmetický průměr znásoben (nebo vydělen) touto konstantou.

Geometrický průměr

Je definován pro kladné hodnoty x jako n-tá odmocnina ze součinu těchto hodnot:

$$\bar{x}_G = \sqrt[n]{x_1, x_2, \ldots x_n}. \tag{1.5}$$

Má uplatnění tam, kde má informační smysl součin hodnot. K použití geometrického průměru při výpočtu průměrného koeficientu růstu se vrátíme v kapitole věnované časovým řadám.

Harmonický průměr

Je definován jako poměr mezi rozsahem souboru a součtem převratných hodnot:

$$\bar{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}. \tag{1.6}$$

Má uplatnění tam, kde má informační smysl součet převratných hodnot.

B. Ostatní střední hodnoty

Do této skupiny řadíme medián a modus jako tzv. poziční střední hodnoty.

Medián $\tilde{x}$

Je padesátiprocentním kvantilem, který charakterizuje hodnotu souboru co do velikosti prostřední. Odděluje polovinu hodnot menších od poloviny hodnot větších.

Medián je na rozdíl od aritmetického průměru necitlivý k extrémním hodnotám, protože závisí pouze na jedné, nejvýše dvou prostředních hodnotách souboru. Nemůže být tedy zkreslen ani přítomností nějaké chybné extrémní hodnoty. Výhodou mediánu je i to, že jej můžeme stanovit i u intervalových rozdělení četností s otevřenými intervaly u minimálních a maximálních hodnot.

Modus $\hat{x}$

Představuje hodnotu, která je v rámci šetřeného souboru nejtypičtější. Jinak řečeno, jde o nejčetnější hodnotu znaku. Také modus není ovlivněn extrémními hodnotami.

V případě intervalového rozdělení četností se při stanovení modu spokojujeme buď s určením modálního (nejčetnějšího) intervalu, nebo v rámci tohoto intervalu modus odhadujeme, např. středem intervalu. Existují však i přesnější postupy, které vycházejí z rekonstrukce vrcholu souboru podle rozdělení četností v okolí modálního intervalu. Pokud se spokojíme jen s určením modálního intervalu, pak je třeba si uvědomit, že má smysl jej určovat pouze tehdy, jsou-li všechny intervaly stejně velké.

Modus považujeme za důležitou doplňkovou charakteristiku k aritmetickému průměru. Pokud se obě míry úrovně významněji liší, pak to znamená, že aritmetický průměr nevyjadřuje dobře typickou úroveň hodnot souboru, např. pro existenci extrémních hodnot nebo pro asymetrické rozložení četností.

1.4.2 Charakteristiky variability

Variabilitou (měnlivostí) kvantitativního statistického znaku rozumíme kolísání hodnot této veličiny. Pokud soubor obsahuje všechny hodnoty stejné (x_i = konstanta), mluvíme o nulové variabilitě. Kolísání hodnot v souboru můžeme posuzovat buď jako vzájemnou rozdílnost jednotlivých hodnot sledované veličiny, nebo jako rozdílnost jednotlivých hodnot od aritmetického průměru. Tento druhý princip měření variability převažuje.

Měření variability lze využít k hodnocení stejnorodosti (homogenity) souboru a také k posuzování kvality informace, kterou o úrovni hodnot v souboru poskytla některá ze středních hodnot.Vycházíme přitom z úvahy, že čím je soubor stejnorodější, s menší variabilitou, tím je např. aritmetický průměr výstižnější z hlediska hodnocení úrovně hodnot souboru. V ekonomické praxi mají míry variability uplatnění např. při hodnocení rovnoměrnosti dodávek, prodeje nebo výroby, při hodnocení stability ukazatele v časové řadě. Hlavně však se s mírami variability setkáme při zkoumání závislosti mezi jevy.

K základním charakteristikám variability patří **variační rozpětí, rozptyl (a jeho odmocnina – směrodatná odchylka) a variační koeficient.**

Variační rozpětí R

Variační rozpětí je rychlou, jednoduchou, ale jen orientační charakteristikou variability založenou na informaci o maximální a minimální hodnotě v souboru:

$$\mathbf{R} = x_{\max} - x_{\min}. \tag{1.7}$$

Při použití variačního rozpětí si musíme vždy být vědomi toho, že hodnoty minima a maxima v souboru mohou mít charakter nahodilých extrémů a tím nepřiměřeně zvětší naši představu o míře variability ve zkoumaném souboru.

Rozptyl a směrodatná odchylka

Rozptyl je nejznámější a nejužívanější mírou variability. Je definován jako aritmetický průměr ze čtverců odchylek jednotlivých hodnot od průměru:

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}. \tag{1.8}$$

Tento vzorec používáme při počítání rozptylu z neuspořádaného souboru všech hodnot souboru, kdy u každé jednotlivé hodnoty souboru zjišťujeme její odchylku od průměru a čtverec této odchylky. Mluvíme pak o výpočtu tzv. **prostého rozptylu**.

Při výpočtu z rozdělení četností, kdy přihlížíme k četnostem jednotlivých obměn, používáme **vážený rozptyl**:

$$s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i}, \text{ resp. } s_x^2 = \frac{1}{n} \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i. \tag{1.9}$$

Pro praktické výpočty se někdy oba vzorce rozptylu upravují do formy tzv. výpočtových tvarů. Způsob této úpravy si ukážeme na vzorci prostého rozptylu.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2\right)=\frac{1}{n}\sum_{i=1}^{n}x_i^2-2\bar{x}\,\frac{1}{n}\sum_{i=1}^{n}x_i+\bar{x}^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2. \tag{1.10}$$

Podobnou úpravou je možno odvodit různé podoby výpočtových tvarů i pro vážený rozptyl, nejpoužívanější je tato úprava:

$$s_x^2=\frac{1}{n}\sum_{i=1}^{k}x_i^2\,n_i-\left(\frac{1}{n}\sum_{i=1}^{k}x_i\,n_i\right)^2. \tag{1.11}$$

Rozptyl sám o sobě není interpretovatelnou veličinou, protože výsledek je dán ve čtvercích měrných jednotek. Proto se při hodnocení variability dává přednost druhé odmocnině rozptylu, tzv. **směrodatné odchylce** s_x (brané s kladným znaménkem).

PŘÍKLAD 1.5

Z výsledků přijímacích zkoušek jsme u 12 studentů z určitého gymnázia zjišťovali dosažené bodové výsledky z testu z matematiky (znak x) a angličtiny (znak y). Chceme porovnat úroveň a variabilitu bodových výsledků u obou předmětů:

Student	x_i	y_i	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
1	60	50	100	25
2	40	30	100	625
3	20	60	900	25
4	40	60	100	25
5	55	55	25	–
6	50	55	–	–
7	80	55	900	–
8	40	55	100	–
9	80	50	900	25
10	10	60	1 600	25
11	100	80	2 500	625
12	25	50	625	25
Celkem	600	660	7 850	1 400

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{600}{12}=50,\qquad \bar{y}=\frac{660}{12}=55,$$

$$s_x^2=\frac{\sum_{i=1}(x_i-\bar{x})^2}{n}=\frac{7\,850}{12}=654{,}2\qquad s_y^2=\frac{1\,400}{12}=116{,}7.$$

Z výsledků jednoznačně vyplývá, že matematika vykazuje podstatně vyšší míru nestejnorodosti bodových výsledků než angličtina.

Variační koeficient

Při srovnávání variability více souborů narážíme na problém rozdílných měrných jednotek a rozdílné úrovně hodnot v souborech. V takových případech je pro potřeby srovnání nejvhodnější charakteristikou variability **variační koeficient** V_x**:**

$$V_x = \frac{s_x}{\bar{x}}. \tag{1.12}$$

Patří mezi relativní míry variability, protože nevyjadřuje variabilitu v původních měrných jednotkách, ale jako poměr směrodatné odchylky a průměru. Obvykle tento poměr prezentujeme v procentech. Pak udává, z kolika procent se v průměru odchylují jednotlivé hodnoty od aritmetického průměru.

Snadná interpretace hodnot variačního koeficientu jej řadí mezi nejpoužívanější charakteristiky variability.

Z následujících dat za odvětví chceme porovnat variabilitu hodinových mezd mužů a žen pomocí variačního koeficientu. Vzhledem k tomu, že výchozí data jsou k dispozici ve formě intervalového

PŘÍKLAD 1.6

rozdělení četností, bude třeba pro výpočet průměru a rozptylu pracovat se středy intervalů:

Interval hodinových mezd v Kč	Relativní četnosti v %		Středy intervalů x_i	muži	ženy	muži	ženy
	muži	ženy		$x_i n_i$		$x_i^2 n_i$	
	n_i						
20 – 29,9	8	12	25	200	300	5 000	7 500
30 – 39,9	16	18	35	560	630	19 600	22 050
40 – 49,9	20	30	45	900	1 530	40 500	68 850
50 – 59,9	30	24	55	1 650	1 100	90 750	60 500
60 – 69,9	18	10	65	1 170	650	76 050	42 250
70 – 79,9	5	6	75	375	450	28 125	33 750
80 a více	3	–	85	255	–	21 675	–
Celkem	100	100	x	5 110	4 660	281 700	234 900

aritmetický průměr $\bar{x} = \dfrac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \Rightarrow$ muži $= \dfrac{5\,110}{100} = 51{,}1,$ ženy $= \dfrac{4\,660}{100} = 46{,}5.$

Pro výpočet použijeme vzorec váženého rozptylu: $s_x^2 = \dfrac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left\{\dfrac{1}{n}\sum_{i=1}^{k} x_i n_i\right\}^2,$

$$s^2_{x\ \text{muži}} = \frac{281\ 700}{100} - 51{,}10^2 = 205{,}8 \quad \Rightarrow \quad V_x = \frac{\sqrt{205{,}8}}{51{,}10} = 0{,}281,$$

$$s^2_{x\ \text{ženy}} = \frac{234\ 900}{100} - 46{,}6^2 = 177{,}44 \quad \Rightarrow \quad V_x = \frac{\sqrt{177{,}44}}{46{,}6} = 0{,}286.$$

I když z číselných hodnot variačních koeficientů vyplývá, že větší stejnorodost hodinových mezd (větší koncentraci kolem průměru) mají muži, nelze považovat zjištěný malý rozdíl v diferenciaci mezd za příliš významný.

K důležitým vlastnostem rozptylu patří:

1. Rozptyl lze vyjádřit jako průměr čtverců hodnot zmenšený o čtverec průměru ($s_x^2 = \overline{x^2} - \overline{x}^2$).
2. Přičte-li se ke všem hodnotám konstanta a, pak se rozptyl nezmění ($s^2_{x+a} = s_x^2$).
3. Násobí-li se všechny hodnoty souboru konstantou k, pak rozptyl je znásoben čtvercem této konstanty ($s^2_{kx} = k^2 s_x^2$).

1.4.3 Charakteristiky tvaru rozdělení

Znázorníme-li jednorozměrná rozdělení četností pomocí polygonu, získáme možnost posoudit tvar rozdělení, např. polohu vrcholu, symetrii rozdělení, míru koncentrace hodnot v určité části variačního rozpětí apod. Z těchto aspektů má největší praktický význam zjištění míry symetrie (souměrnosti) rozdělení četností, protože tím lze významně obohatit hodnocení vypovídací ceny všech popisných charakteristik souboru. Souměrná symetrická rozdělení jsou v ekonomické praxi spíše vzácností. Zřetelným projevem asymetrie rozdělení je především odlišnost hodnot aritmetického průměru od mediánu a modu. Pro zcela symetrické rozdělení je naopak charakteristické, že všechny hlavní charakteristiky úrovně jsou totožné:

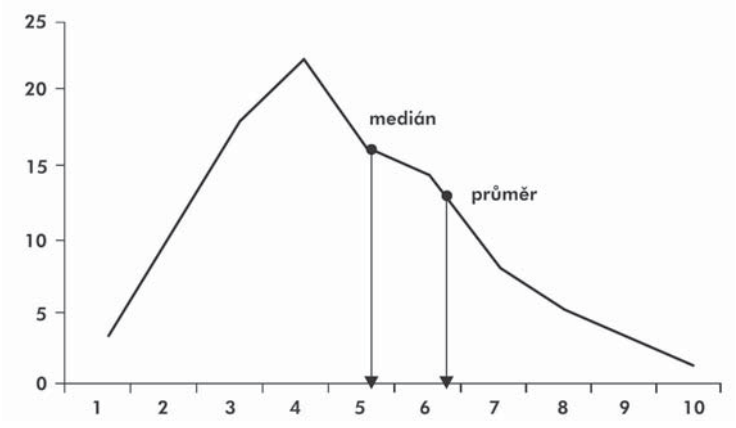
$$\bar{x} = \tilde{x} = \hat{x}.$$

U nesymetrických rozdělení tato identita neplatí. Graf A charakterizuje kladně zešikmené rozdělení, pro které je obvyklé, že aritmetický průměr je menší než medián a modus:

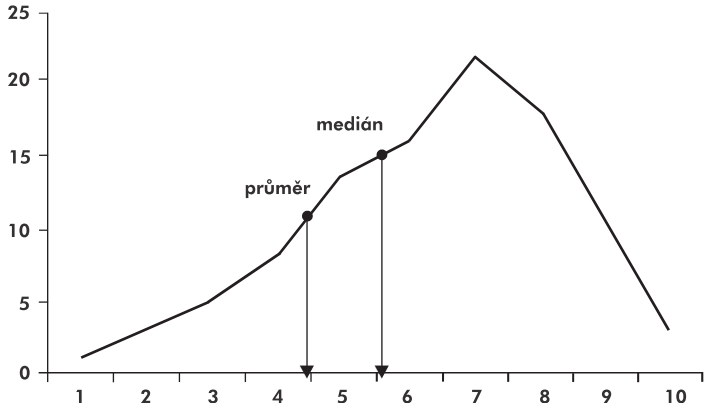
$$\bar{x} > \tilde{x} > \hat{x}.$$

Je to rozdělení s velkým nakupením hodnot menších než průměr. Tento typ rozdělení je v praxi typický např. pro rozdělení mezd.

GRAF A — Rozdělení s kladnou šikmostí



GRAF B Záporně zešikmené rozdělení, kde platí $x > \tilde{x} > \hat{x}$



Jednoduchou charakteristikou šikmosti je Pearsonův koeficient α, který využívá k hodnocení stupně šikmosti vztah mezi velikostí aritmetického průměru a mediánu:

$$\alpha = \frac{3\,(\bar{x} - \tilde{x})}{s_x}\,. \tag{1.13}$$

Pro symetrická rozdělení má nulovou hodnotu. Velikost koeficientu a jeho znaménko pak ukazuje stupeň a charakter zešikmení.

Jiný přístup k měření šikmosti je založen na aplikaci tzv. momentových charakteristik. Při práci s daty uspořádanými do rozdělení četností je vhodná tzv. momentová míra šikmosti (označovaná také jako třetí moment směrodatné proměnné) se vzorcem:

$$\frac{1}{n}\sum_{i=1}^{k}\left(\frac{x_i - \bar{x}}{s_x}\right)^3 n_i\,. \tag{1.14}$$

Opět platí, že nulová hodnota charakterizuje symetrická rozdělení a kladné a záporné hodnoty vyjadřují různý stupeň tzv. kladné a záporné šikmosti.

Shrnutí

- Tato kapitola byla věnována praktickým problémům zpracování, prezentace a vstupní analýzy dat získaných statistickým zjišťováním, kde je třeba vymezit **statistickou jednotku**, **statistický znak** (proměnnou, ukazatel) a **statistický soubor**.
- Pokud pracujeme s proměnnou, jejíž hodnoty se ve statistickém souboru vyskytují opakovaně, je výhodné pro další analýzu uspořádat hodnoty zkoumaného souboru ve formě **rozdělení četností**. To má za následek, že je třeba upravit i způsob výpočtu charakteristik, kterými popisujeme vlastnosti statistického souboru. Rozlišujeme pak např. prostý a vážený aritmetický průměr, prostý a vážený rozptyl.
- Má-li zkoumaný kvantitativní statistický znak (proměnná) charakter spojité veličiny nebo příliš mnoho obměn, prezentujeme statistický soubor ve formě **intervalového rozdělení četností**.
- Grafickým vyjádřením rozdělení četností je **polygon**.
- Grafickým vyjádřením intervalového rozdělení četností je **histogram**.
- O rozložení hodnot zkoumané proměnné ve statistickém souboru nás informují **kvantily**. Typy kvantilů jsou rozlišeny stupněm podrobnosti, v kterém rozdělují soubor do stejně obsazených částí. V praxi se nejčastěji setkáme **s mediánem**, kterým je soubor rozdělen do dvou částí, a je tedy určen hodnotou, která rozděluje soubor na 50 % prvků menších a 50 % prvků větších.
- Pro základní deskripci statistického souboru kvantitativního znaku používáme systém popisných charakteristik, který tvoří:
  - **míry úrovně** hodnot souboru (míry polohy rozdělení četností),
  - **míry variability** hodnot,
  - **míry šikmosti** (asymetrie) rozdělení.
- K nejužívanějším mírám úrovně patří **aritmetický průměr**, **medián** a **modus**.
- V situacích, kdy hodnota aritmetického průměru reprezentujícího statistický soubor je výrazně ovlivněna existencí extrémních hodnot, je vhodné jako charakteristiku úrovně použít medián.
- Způsob výpočtu popisných charakteristik je odlišný, pracujeme-li s netříděnými hodnotami a s hodnotami uspořádanými do rozdělení četností. V případě, kdy údaje statistického souboru máme k dispozici ve formě rozdělení četností, používáme vzorce váženého aritmetického průměru a váženého rozptylu, v nichž jako váhy vystupují četnosti jednotlivých obměn statistického znaku. Při uspořádání hodnot statistického souboru ve formě intervalového rozdělení četností je třeba počítat se ztrátou možnosti získat přesnou hodnotu popisných charakteristik.
- K nejužívanějším mírám variability patří **variační rozpětí**, **rozptyl**, **směrodatná odchylka** a **variační koeficient**.
- Pro porovnávání variability několika souborů dáváme přednost variačnímu koeficientu jako **relativní míře variability**.
- O souborech, kde úroveň všech hodnot souboru je stejná, říkáme, že mají nulovou variabilitu.
- Pro hodnocení **stupně asymetrie (šikmosti) rozdělení** nás může informovat jednak vzájemná poloha aritmetického průměru, mediánu a modu, jednak tzv. **momentová míra šikmosti**. V souborech zcela symetrických mají aritmetický průměr, medián i modus totožnou hodnotu.
- U souborů, které jsou výrazně asymetrické, je třeba počítat s tím, že aritmetický průměr nevyjadřuje typickou úroveň hodnot souboru a při hodnocení dáváme přednost informaci získané z mediánu.

Klíčová slova

aritmetický průměr
modus
medián
rozdělení četností
polygon
intervalové rozdělení četností
histogram
kvantily
variabilita
rozptyl
variační koeficient
směrodatná odchylka
variační rozpětí
statistická jednotka
statistický soubor
statistický znak

Řešené příklady

Příklad 1

Máme k dispozici údaje o hodinových mzdách 10 pracovníků jednoho oddělení firmy:

51, 58, 70, 64, 60, 50, 58, 55, 66 a 138.

Chceme charakterizovat vhodnou charakteristikou úroveň mezd v daném oddělení.

Řešení:

Nabízí se především zjištění aritmetického průměru hodinové mzdy:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{670}{10} = 67 \text{ Kč.}$$

Z konfrontace získané hodnoty aritmetického průměru s výchozími daty vyplývá, že prakticky všichni pracovníci – až na jednoho – mají podprůměrný plat. Přitom je zřejmé, že na výši průměru se výrazně podepsala nevyšší hodinová mzda 138 Kč, která je však v daném souboru netypická (říkáme také odlehlá, extrémní).

Charakteristikou, která by nás v daném případě lépe informovala o typické úrovni mezd v souboru, je medián, protože ten obecně není ovlivněn extrémními hodnotami v souboru.

Pro jeho výpočet v prvním kroku seřadíme jednotlivé hodnoty souboru podle velikosti a ke každé přiřadíme pořadové číslo:

Pořadí	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
x_i	50	51	55	58	58	60	64	66	70	138

Pořadové číslo prostřední hodnoty stanovíme jako $\frac{n+1}{2} = \frac{11}{2} = 5{,}5$. Z toho vyplývá, že hodnota mediánu nebude dána některou konkrétní hodnotou zkoumaného souboru, ale odhadneme ji jako průměr ze dvou prostředních hodnot (z 5. a 6. hodnoty):

$$\tilde{x} = \frac{58 + 60}{2} = 59.$$

Zjištěná hodnota mediánu 59 Kč nám v našem případě daleko výstižněji charakterizuje typickou úroveň platů v oddělení.