

- **Jednostupňové třídění statistických souborů**
- **Popis jednorozměrných statistických souborů**

2.

Zpracování a popis statistických souborů



Cíl kapitoly

V první kapitole jste se seznámili se základními pojmy, které jsou používány ve statistické vědě. Také jsme stručně zmínili konstrukci a požadavky na adekvátní datovou základnu popisující statistické znaky. Můžeme proto přistoupit k druhé a třetí etapě statistického zkoumání – jejich zpracování a vyhodnocení. Následující kapitola se Vás proto seznámí s nejčastěji používanými metodami třídění, shrnování a interpretace statistických údajů. Blíže Vás seznámí s běžně používanými pojmy jako je například četnost, či graf. Ve druhé části kapitoly se seznámíte se základními veličinami tzv. popisné statistiky, které vytvářejí kvantifikovanou představu o statistickém souboru. Patří sem zejména nejrůznější průměry, rozptyly a odchylky. V textu se zaměříme nejen na jejich definici a možnosti jejich výpočtu, ale i na vypovídací hodnotu těchto veličin a možnosti interpretace výsledků.



Časová zátěž

6 hodin (1. a 2. týden v říjnu)

Výsledkem statistického zjišťování je obvykle soubor údajů, který je poměrně rozsáhlý. Tento obrovský balík dat je velmi nepřehledný a pro vyslovení konkrétních závěrů o zkoumané skutečnosti nemůže sloužit jako adekvátní podklad. Je proto nutno zjištěná data nějakým způsobem zpřehlednit. Tak aby mohly vyniknout především charakteristické znaky souboru. Základním používaným nástrojem jsou metody třídění statistického souboru.

Třídění

Tříděním statistického souboru je nazýván postup, kdy jednotky statistického souboru rozdělíme do takových skupin, aby co nejlépe vynikly charakteristické vlastnosti zkoumaného jevu. Mimo uspořádání údajů v souboru tříděním také obvykle dosáhneme jejich zhuštění. Metody třídění dělíme do dvou základních skupin:

- **jednostupňové**, kdy uspořádáváme soubor do skupin dle jednoho zkoumaného statistického znaku
- **vícetupňové**, kdy třídíme soubor podle více statistických znaků najednou.



O jednostupňovém třídění například mluvíme, setřídíme-li zaměstnance vybrané organizace podle jejich hrubého měsíčního příjmu. Třídícím statistickým znakem zde zjevně bude výše měsíčního příjmu jednotlivých zaměstnanců. Setřídíme-li soubor zaměstnanců podle jejich průměrného příjmu, pracovního zařazení a věku, hovoříme již o třídění vícetupňovém.

2.1 Jednostupňové třídění statistických souborů

Základním typem statistického třídění je jednostupňová varianta, tedy jeho uspořádání podle jednoho znaku. V tomto případě rozdělujeme údaje v statistickém souboru do několika skupin, které předem stanovíme. Dostáváme tak údaje seřazené do posloupnosti, kde ke každé skupině (obvykle tato skupina

odráží jednotlivé varianty statistického znaku) je přiřazen příslušný počet jednotek, které do ní náleží (nabývají hodnotu rovnou dané variantě). Tento počet jednotek nazýváme **četností** statistického znaku. Pokud výsledky uvedeme ve formě tabulky dostáváme tabulku rozdělení četností.

Takto získané četnosti jsou někdy nazývány **absolutní četnost**. Vedle těchto absolutních četností je možno počítat i **relativní četnosti**, které odrážejí váhu jednotlivých skupin vztaženou k celému souboru. Relativních četností se používá například pro srovnávání více statistických souborů.

Relativní četnost je označována symbolem p_i a je definována jako

$$p_i = \frac{n_i}{\sum_{i=1}^k n_i} = \frac{n_i}{N}.$$

kde n_i je absolutní četnost i -té varianty znaku, N je počet prvků ve statistickém souboru a k je počet obměn statistického znaku v souboru,

$$N = \sum_{i=1}^k n_i.$$

Pro relativní četnosti daného souboru musí platit, že jejich součet pro všechny varianty zkoumaného znaku je roven jedné.

$$\sum_{i=1}^k p_i = 1.$$

Údaje o absolutních a relativních četnostech bývají obvykle doplňovány ještě o jejich kumulativní varianty (označíme je **nk** , resp. **pk**). **Kumulativní četnost** počítá, kolik (či jaké procento v případě relativních) prvků statistického souboru dosáhlo dané varianty znaku, či varianty nižší. Je zřejmé, že kumulativních četností je možno využít pouze u statistických souborů, kde je možno jednotlivé varianty statistického znaku ordinálně uspořádat od nejmenší po největší.

Kumulativní absolutní četnost i -té varianty znaku vypočítáme jako

$$nk_i = \sum_{j=1}^i n_j$$

a kumulativní relativní četnost jako

$$pk_i = \sum_{j=1}^i p_j = \sum_{j=1}^i \frac{n_j}{N}.$$

V souvislosti s pojmem četnost se někdy setkáváme s pojmem **modus**, který označuje nejčetnější obměnu statistické proměnné. Je to tedy ta varianta statistického znaku, která dosahuje nejvyšší hodnoty absolutní četnosti.

2. Zpracování a popis statistických souborů

Vztahy mezi výše uvedenými veličinami shrnuje následující tabulka:

Varianty znaku	Četnost		Kumulativní četnost	
	absolutní	relativní	absolutní	relativní
x_i	n_i	$p_i = \frac{n_i}{n}$	$N_i = \sum_{j=1}^i n_j$	$F_i = \frac{N_i}{n}$
x_1	n_1	$p_1 = \frac{n_1}{n}$	n_1	p_1
x_2	n_2	$p_2 = \frac{n_2}{n}$	$n_1 + n_2$	$p_1 + p_2$
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	$p_k = \frac{n_k}{n}$	$n_1 + \dots + n_k$	$p_1 + \dots + p_k$
Σ	n	1	Součet absolutních četností	Součet relativních četností



Příklad 2.1

Máme k dispozici statistický soubor, který obsahuje údaje charakterizující studijní výsledky posluchačů vybraného studijního oboru na vysoké škole. Studijní výsledky jsou popsány následujícími statistickými znaky známky z vykonaných zkoušek, počet vykonaných zkoušek, studijní průměr. Seříd'te tento soubor podle statistického znaku známka z předmětu Statistika. Víte, že 20 studentů daného oboru získalo následující známky (na stupnici 1–4): 1, 1, 1, 2, 1, 3, 4, 1, 2, 1, 3, 2, 1, 3, 4, 2, 2, 3, 1, 2.

Řešení

Máme-li seřadit uvedené studenty podle jednotlivých variant znaku, je nutno vytvořit čtyři skupiny studentů – studenty hodnocené *výborně* (1), studenty hodnocené stupněm *velmi dobrý* (2), studenty s hodnocením *dobrý* (3) a studenty, kteří danou zkoušku nesložili úspěšně – tedy hodnocené stupněm *nevyhověl* (4).

Tabulku rozdělení četností pak získáváme tak, že jednotlivým variantám hodnocení přiřadíme počet studentů s touto variantou. K vypočítaným absolutním četnostem poté doplníme relativní četnosti jako podíl absolutní četnosti na celku (tedy dělený počtem všech studentů). Pro variantu *výborně* dostáváme absolutní četnost rovnu 8. Relativní četnost je pak rovna

$$p_i = \frac{n_i}{\sum_{i=1}^k n_i} = \frac{8}{20} = 0,4.$$

Lze tedy říci, že u zkoušky z předmětu Statistika bylo stupněm *výborně* (1) hodnoceno 8 studentů, tj. 40% ze všech studentů daného studijního oboru.

Vypočítáme-li kumulativní absolutní četnost pro variantu *dobře* (3), obdržíme

$$nk_3 = \sum_{j=1}^i n_j = \sum_{j=1}^3 n_j = 8 + 6 + 4 = 18,$$

pro kumulativní relativní četnost dostáváme

$$pk_3 = \sum_{j=1}^i p_j = \sum_{j=1}^3 p_k = 0,4 + 0,3 + 0,2 = 0,9.$$

Z předmětu Statistika bylo tedy známkou *dobře* (3) a lepší hodnoceno 18 studentů, což je 90% ze všech studentů daného studijního oboru.

Kompletní výsledky přináší následující tabulka:

Varianta znaku	Absolutní četnost	Relativní četnost	Kumulativní absolutní četnost	Kumulativní relativní četnost
x_i	n_i	p_i	nk_i	pk_i
1 (výborně)	8	0,4	8	0,4
2 (velmi dobře)	6	0,3	14	0,7
3 (dobře)	4	0,2	18	0,9
4 (nevyhověl)	2	0,1	20	1,0
Σ	20	1,0	–	–

Z tabulky je tedy patrné, že drtivá většina – 18 (90%) studentů u zkoušky uspěla (hodnota kumulativních četností pro variantu 3), přičemž 14 (70%) studentů bylo hodnoceno výborně nebo velmi dobře (hodnota kumulativních četností pro variantu 2).

Modusem tohoto rozdělení by byla varianta výborně (1), jelikož této varianty dosáhlo nejvíce studentů.

2.1.1 Intervalové rozdělení četností

Uvedený postup konstrukce tabulky rozdělení četností má však několik omezení. Tato omezení jsou významná zejména při konstrukci tabulky četností pro spojitě statistické znaky. Výše uvedené úvahy (i příklad) uvažovaly pouze diskrétní statistické znaky, tedy znaky, které mohou nabývat pouze některých hodnot.

V případě, že chceme třídit statistický soubor, který obsahuje spojitě znaky (tedy znaky obecně nabývajících jakýchkoli hodnot), je vytvoření tabulky rozdělení četností pomocí uvedeného postupu podle jednotlivých variant nemožné, jelikož těchto variant může být nekonečně mnoho. Z těchto důvodů je nutno provést předstupu daného třídění a vytvořit několik intervalů, do kterých shrneme vždy výsek jednotlivých variant zkoumaného statistického znaku. Výslednou tabulku poté nazýváme **intervalovým rozdělením četností**. Tyto intervaly (někdy označované třídy) pak reprezentují určité varianty znaku, které tentokrát nejsou dány objektivně (tedy na základě vlastností zkoumaného znaku), ale subjektivně (na základě úsudku statistika).

Při konstrukci intervalového rozdělení četností čelíme třem podstatným problémům:

■ jak zvolit počet intervalů

Je nutno optimálně volit mezi počtem zvolených intervalů a jejich vypovídací hodnotou. Příliš vysoký počet intervalů má za následek nepřehlednost výsledků. Naopak, zvolíme-li příliš malý počet intervalů, vede to k velmi hrubé charakteristice daného statistického souboru a zjednodušenému pohledu na vlastnosti souboru. Neexistuje však žádné obecné pravidlo pro určení počtu intervalů, konkrétní rozhodnutí je vždy v rukou výzkumníka a mělo by se opírat o zkušenost a do jisté míry i o intuici. Počet intervalů je také úzce svázán s účelem statistického zkoumání.

■ jak určit hranice těchto intervalů

V případě spojitých statistických znaků může dojít k situaci, kdy stojíme před rozhodnutím, do kterého intervalu máme zařadit hodnotu přesně se shodující s hranicí intervalu.

Obvykle je postupováno podle následujícího pravidla: Prvek souboru, jehož hodnota je shodná s hranicí intervalu je přiřazen do toho intervalu, který má sudé pořadové číslo. Jinou možností – umožňující to hodnoty obsažené v statistickém souboru – je volit dichotomní (vzájemně se vylučující) hranice intervalu.

V případě, že ani jeden z těchto přístupů není možný, nebo vhodný, je možno započítat daný prvek do obou intervalů, vždy však pouze jednu polovinu (tedy absolutní četnost prvku v každém z obou intervalů je 0,5).

■ jaký prvek zvolit za reprezentanta daného intervalu (variantu znaku)

Jelikož je u spojitých veličin analyzována četnost výskytu v intervalu hodnot, nelze tento interval charakterizovat jedním údajem. V některých případech je však nutné nalézt reprezentanta tohoto intervalu. Za tohoto reprezentanta je obvykle považován střed daného intervalu (nebo průměrná hodnota jeho hranic).



Příklad 2.2

Budeme uvažovat stejný statistický soubor jako v prvním příkladu. Zkoumáme tedy soubor studijních výsledků studentů vybraného studijního oboru. Zkoumaným statistickým znakem tentokrát bude studijní průměr těchto studentů. K dispozici jsou následující hodnoty studijních průměrů: 1,1; 1,2; 1,5; 2,3; 2,0; 1,8; 2,1; 1,6; 1,4; 1,2; 1,0; 2,0; 2,8; 1,8; 1,1; 1,3; 1,7; 2,0; 2,1; 1,0. Roztřídte daný statistický soubor dle statistického znaku studijní průměr studenta.

Řešení

Jelikož studijní průměr je (pro větší počet zkoušek) možno považovat za spojitý statistický znak, konstruuje intervalové rozdělení četností. Musíme tedy nejprve rozhodnout, do kolika intervalů výsledky shrneme a rozdělíme. Účelným se v tomto případě jeví rozdělit studenty do pěti skupin – na vynikající (studijní průměr do 1,2), výborné (průměr do 1,5), studenty velmi dobré (průměr do 2,0), studenty průměrné (studijní průměr do 2,5) a studenty podprůměrné (průměr nad 2,5). Počet intervalů je v tomto případě dán

například požadavkem na jejich roztrídění pro účely poskytování stipendií, či ubytování na studentské koleji.

Prvky, které mají hodnoty shodné s hranicemi intervalů zařadíme (dle výše uvedeného pravidla) vždy do sudého intervalu. Tedy například student se studijním průměrem 1,2 bude zařazen mezi „výborné“, tedy do druhého intervalu (1,2 – 1,5), zatímco student se studijním průměrem 2,0 bude zařazen mezi „průměrné“, tedy do čtvrtého intervalu (2,0 – 2,5).

Výsledné intervalové rozdělení četností, včetně kumulativních variant těchto veličin, uvádí následující tabulka:

Interval (třída)	Absolutní četnost	Relativní četnost	Kumulativní absolutní četnost	Kumulativní relativní četnost
x_i	n_i	p_i	nk_i	pk_i
Vynikající (1,0 – 1,2)	4	0,20	4	0,20
Výborní (1,2 – 1,5)	4	0,20	4	0,40
Velmi dobří (1,5 – 2,0)	5	0,25	13	0,65
Průměrní (2,0 – 2,5)	6	0,30	19	0,95
Podprůměrní (nad 2,5)	1	0,05	20	1,00
Σ	20	1,0	–	–

Výsledky uvedené v tabulce tedy hovoří o tom, že plných 40% studentů je možno zařadit mezi vynikající, či výborné (hodnota kumulativní relativní četnosti pro druhý interval). Pod hranici 2,0 se se svým studijním průměrem dostalo 13 studentů, což činí 65% (hodnota kumulativních četností pro třetí interval – varianta „velmi dobří“).

2.1.2 Statistické grafy

Jinou možností, jak zpřehlednit údaje obsažené v rozsáhlém statistickém souboru, je možnost využití některého ze statistických grafů. Grafy jsou využívány také ke zobrazení již roztríděných souborů a mohou být tedy považovány za grafickou interpretaci tabulky rozdělení četností. V tomto případě hovoříme o **grafech rozdělení četností**. Mezi nejpoužívanější typy statistických grafů patří grafy:

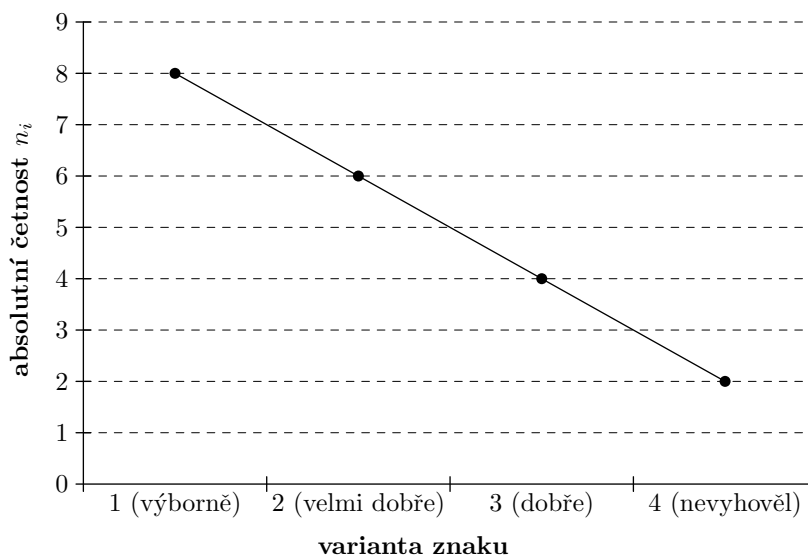
- spojnicové,
- sloupcové,
- výsečové,
- bodové.

Zobrazování rozdělení četností je obvyklé zejména pomocí spojnicových a sloupcových grafů. Grafy četností obvykle zachycují na vodorovné (horizontální) ose jednotlivé varianty znaků a na svislé (vertikální) ose jim příslušející hodnoty četností. Spojnicové grafy jsou nazývány **polygony četností** a jsou tvořeny úsečkami spojujícími body, jež charakterizují hodnotu četnosti (absolutní, či relativní) pro příslušnou variantu znaku. Z těchto důvodů jsou vhodnější pro zobrazení diskrétních rozdělení četností.

Grafy rozdělení četností

Polygon

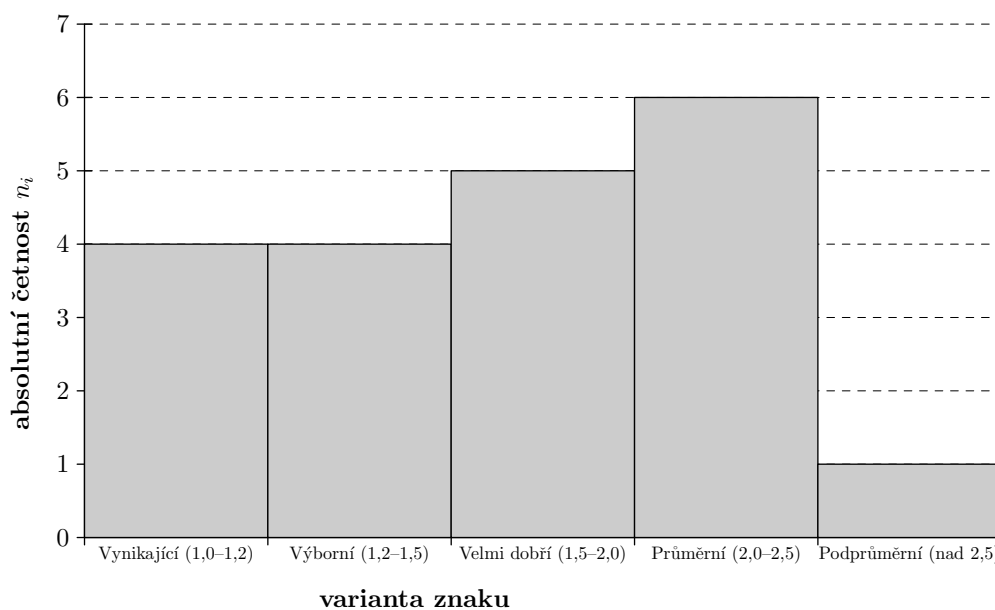
2. Zpracování a popis statistických souborů



Obrázek 2.1: Polygon rozdělení četností (pro výsledky příkladu 2.1)

Histogram

Pro zobrazení intervalového rozdělení četností je využíváno sloupcových grafů, nazývaných **histogram**. Hodnoty četnosti příslušející jednotlivým intervalům jsou v histogramu zachyceny pomocí sloupečků o výšce rovné dané četnosti.

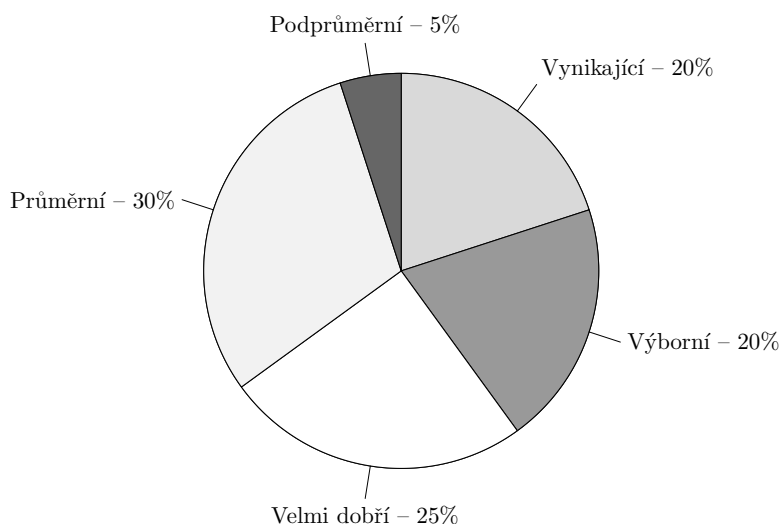


Obrázek 2.2: Histogram rozdělení četností (pro výsledky příkladu 2.2)

Výsečový graf

Výsečové (koláčové) **grafy** jsou využívány zejména pro zachycení struktury statistického souboru. Jednotlivé výseče tak zachycují relativní četnost jednotlivých variant ve statistickém souboru.

Bodové grafy jsou používány nejčastěji jako prostředek zobrazování závislostí mezi dvěma statistickými znaky. Znaky bývají zachyceny jako nezávisle proměnná (vynášená na vodorovné ose) a závisle proměnná (na vertikální)



Obrázek 2.3: Výšečový graf rozvrstvení studijních průměrů (pro výsledky příkladu 2.2)

ose. Bodové grafy jsou jednou z výchozích informací pro regresní a korelační analýzu.

Mezi další používané typy statistických grafů patří například krabicové diagramy nebo STEM-and-LEAF grafy. Jejich konstrukce i použití je podobné výše uvedeným typům. Význam těchto grafů však poněkud klesá v souvislosti s postupným nástupem výpočetní techniky a masovějšího rozšíření statistického software, který je schopen podat velmi rychlou informaci o skutečnostech, které tyto grafy obvykle zdůrazňují. Podobu uvedených typů grafů naleznete například v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na stranách 32–34.



2.2 Popis jednorozměrných statistických souborů

Tabulka či graf rozdělení četností přináší velmi užitečnou a konkrétní představu o základní struktuře statistického souboru. Nevýhodou těchto rozdělení je však skutečnost, že jsou stále poměrně rozsáhlou informací o tomto souboru. Pro potřeby srovnávání různých statistických souborů, případně pro jejich bližší charakterizaci, je výhodnější konstruovat jiné veličiny. Jejich posláním je převést informaci o statistickém souboru do jednoho či několika málo údajů. Požadavkem na tyto veličiny je, aby co možno nejpřesněji vystihly základní rysy zkoumaných dat.

Veličiny, které jsou pro tyto účely využívány se nazývají míry. Rozeznáváme míry čtyř typů dle vlastnosti statistického souboru, kterou popisují:

- míry polohy (úrovně),
- míry variability,
- míry šikmosti,
- míry špičatosti.

2.2.1 Míry polohy

Základní vlastností, která nás zajímá při popisu statistického souboru, je obvykle jeho úroveň. Hledáme tedy charakteristiky, které popisují, kde na číselné ose je umístěna podstatná část prvků statistického souboru. K tomu to účelu jsou využívány především **průměry** a **kvantily**.

Průměry

Průměry

Průměry jsou charakteristikami, které vystihují nejpravděpodobnější polohu hodnot obsažených ve statistickém souboru. Průměr se stává komplexní informací, která umožňuje nahradit všechny prvky statistického souboru jedním údajem. Rozlišujeme tři základní typy průměrů – aritmetický, geometrický a harmonický. Všechny tyto typy průměrů je možno doplnit o jejich vážené varianty, kdy přikládáme jednotlivým prvkům statistického souboru různou váhu v tomto souboru.

Aritmetický průměr \bar{x}_A

Aritmetický průměr je nejznámějším typem průměru. Je počítán jako součet všech hodnot statistického souboru dělený jejich počtem.

$$\bar{x}_A = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Vážený aritmetický průměr

$$\bar{x}_A = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_k \cdot w_k}{n} = \frac{\sum_{i=1}^k x_i \cdot w_i}{n}$$

kde w_1, w_2, \dots, w_k jsou váhy jednotlivých prvků.

Geometrický průměr \bar{x}_G

Geometrický průměr je obdobou aritmetického průměru. Pouze při výpočtu nahrazuje sčítání použité u aritmetického za násobení a dělení za odmocňování. Je tedy počítán jako n -tá odmocnina ze součinu všech prvků statistického souboru.

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

Vážený geometrický průměr

$$\bar{x}_G = \sqrt[n]{x_1^{w_1} \cdot x_2^{w_2} \cdot \dots \cdot x_k^{w_k}} = \sqrt[n]{\prod_{i=1}^k x_i^{w_i}}$$

Harmonický průměr \bar{x}_H

Harmonický průměr je výpočtově nejsložitějším typem průměru. Je definován jako převrácená hodnota aritmetického průměru převrácených hodnot

statistického souboru.

$$\bar{x}_H = \frac{1}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Vážený harmonický průměr

$$\bar{x}_H = \frac{1}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_k}{x_k}} = \frac{n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_k}{x_k}} = \frac{n}{\sum_{i=1}^k \frac{w_i}{x_i}}$$

Použití jednotlivých typů průměrů je velmi odlišné. Pro praktické příklady je nejobvyklejší výpočet aritmetického průměru. Pokud hodláme přikládat některým prvkům statistického souboru vyšší význam než jiným, je obvykle využíváno jeho vážené varianty.

Geometrický průměr je využíván tam, kde použití aritmetického průměru není možné, jelikož by mělo za následek zkreslený či nesprávný výsledek. Příkladem je například výpočet průměru relativních veličin. Geometrického průměru je proto využíváno například při výpočtu průměrného tempa růstu časové řady (tato veličina bude uvedena ve druhém díle Aplikované statistiky).

Harmonických průměrů, stejně jako vážených variant dvou výše uvedených typů průměrů, je využíváno například v oblasti cenové statistiky. Jedná se zejména o výpočet souhrnných cenových indexů, jejichž hlavním reprezentantem je Index spotřebitelských cen (CPI), který je výchozí informací pro výpočet míry inflace.

Užitím jednotlivých typů vážených průměrů se budeme zabývat ve druhé části Aplikované statistiky. Jejich výpočet je nastudujte v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na stranách 42–47.

Příklad 2.3

Vypočítejte průměrnou známku ze zkoušky ze statistiky u studentů uvedených v příkladu 2.1. Použijte všechny typy průměrů – aritmetický, geometrický a harmonický a jejich hodnoty porovnejte.

Řešení

Víme, že 20 studentů daného oboru získalo u zkoušky následující známky (na stupnici 1–4): 1, 1, 1, 2, 1, 3, 4, 1, 2, 1, 3, 2, 1, 3, 4, 2, 2, 3, 1, 2. Můžeme tedy označit jednotlivé prvky tohoto souboru symboly x_i , kde i nabývá hodnot od jedné do dvaceti.

Aritmetický průměr tedy vypočítáme jako

$$\bar{x}_A = \frac{\sum_{i=1}^n x_i}{n} = \frac{1 + 1 + \dots + 1 + 2}{20} = \frac{40}{20} = 2$$



2. Zpracování a popis statistických souborů

Geometrický průměr vypočítáme jako

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[20]{1 \cdot 1 \cdot \dots \cdot 1 \cdot 2} = \sqrt[20]{82944} = 1,76$$

Harmonický průměr je roven

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} = \frac{20}{\frac{1}{1} + \frac{1}{1} + \dots + \frac{1}{1} + \frac{1}{2}} = 1,56$$

Z výsledků je patrné, že průměrná známka u zkoušky z předmětu se pohybuje v rozmezí 1,56 – 2 dle použitého průměru. Nejvyšší hodnoty dosahuje průměr aritmetický, nejnižší hodnoty průměr harmonický.



Nerovnost, která byla výsledkem uvedeného příkladu, je obecným vztahem mezi třemi typy průměrů. Vypočítáme-li aritmetický, geometrický a harmonický průměr pro stejné hodnoty vždy platí

$$\bar{x}_A \geq \bar{x}_G \geq \bar{x}_H.$$

Rovnost nastává pouze v případě, že počítáme průměry souboru, který obsahuje identické hodnoty.

Kvantily

Kvantil

Kvantil je prvek statistického souboru, který jej rozděluje na dvě části. První část obsahuje prvky souboru, jejichž hodnota je menší nebo rovna hodnotě kvantilu. Druhá část obsahuje prvky s hodnotou vyšší než je hodnota kvantilu.

Kvantil se obvykle udává v procentním vyjádření. Symbol 15% kvantil tedy označuje prvek souboru, který jej rozdělí na 15% prvků souboru s hodnotami menšími než tento kvantil a 85% prvků, s hodnotami vyššími.



Uvažujeme-li následující sdělení: V České republice bylo v roce 2002 15% občanů s příjmy pod hranicí životního minima, které v tomto roce mělo hodnotu 5500 Kč.

Pak lze (ze statistického pohledu) říci, že hodnota 5500 Kč je 15% kvantil rozdělení příjmů v souboru obyvatel České republiky.

Medián

Z hlediska jejich dobré interpretace a využití jsou využívány především některé významné kvantily. Tyto kvantily také obvykle nesou speciální označení. O 10% kvantilu proto obvykle mluvíme jako o **decilu**, o 25% kvantilu jako o **(dolním) kvartilu** a nejnámějším je 50% kvantil označovaný jako **medián**. Medián je hodnota, která rozděluje statistický soubor na dvě identické poloviny. Je také někdy označován jako prostřední hodnota.

K výpočtu kvantilů je nutno seřadit prvky souboru do vzestupné posloupnosti. Příslušný kvantil v tom případě najdeme, jednoduchým úsudkem, kdy hledáme tolikátý prvek v souboru, který odpovídá příslušnému kvantilu.

Máme-li rozsáhlejší soubor, který je již uspořádán do tabulky (grafu) rozdělení četností, je určení kvantilu obtížnější. V tom případě je nutno použít některého z přibližných výpočtových vztahů. Jeden z těchto vztahů najdete v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na straně 40.



Příklad 2.4

Určete medián, první decil a horní kvartil souboru dvaceti známek ze zkoušky ze Statistiky, uvedené v příkladu 2.1.



Řešení

Uspořádáme-li získané známky těchto 20 studentů daného oboru obdržíme následující vzestupnou posloupnost: 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4.

Medián nalezneme jako prostřední hodnotu tohoto souboru. Jelikož máme soubor o sudém počtu členů, musíme medián určit jako hodnotu, která leží mezi dvěma prostředními prvky souboru (10. a 11. prvkem). Medián tedy určíme jako $(2 + 2)/2 = 2$. Prostředním prvkem souboru je tedy známka 2. Dostáváme tedy stejnou hodnotu jako je aritmetický průměr tohoto souboru. Tato skutečnost však není pravidlem. Obvykle se hodnota mediánu od hodnoty aritmetického průměru odlišuje, jelikož medián nezkoumá relativní rozdíly mezi hodnotami jednotlivých prvků souboru (jejich variabilitu).

První decil označuje hodnotu, která rozdělí statistický soubor na 10% nejmenších prvků a zbylých 90%, které jsou větší. V našem příkladu bude prvním decilem zjevně druhý prvek souboru, jehož hodnota je 1. Lze tedy konstatovat, že nejlepších 10% studentů nemá známku horší než 1.

Horní kvartil je hodnota, která rozděluje soubor na nejnižších 75% a nejvyšších 25% hodnot souboru. V uvedeném souboru bude horním kvartilem 15. prvek souboru, tedy známka 3. Můžeme tedy říci, že 25% nejslabších studentů má známku 3 a vyšší.

Všimněte si, že k výpočtu kvantilů je možno velmi dobře užít kumulativních četností. Dostaneme-li kumulativní četnost některého znaku, či varianty rovnou procentní hodnotě hledaného kvantilu, je tímto kvantilem právě tato hodnota.



2.2.2 Míry variability

Průměry, či kvantily mohou podávat naprosto dostatečnou informaci o zkoumaném statistickém souboru. Není však zaručeno, že takto podaná informace je naprosto správná. Lze si snadno představit dva soubory, které se shodují ve svém průměru, ale jsou naprosto odlišné. Jednu takovou situaci ilustruje následující příklad:

Příklad 2.5

Dva začínající střelci se zúčastnili soutěže v přesnosti na 3 zásahy. Jejich výsledky byly následující:



2. Zpracování a popis statistických souborů

1. střelec: 5, 5, 5
2. střelec: 10, 1, 4

Porota má rozhodnout o vítězi této soutěže a udělit cenu perspektivnějšímu střelci.

Spočítáme-li průměrné výsledky obou střelců (pomocí aritmetického průměru) dojdeme ke stejnému výsledku – oba střelci dosáhli v průmětu výsledku 5. Přesto se porota jednoznačně rozhodla pro udělení výhry prvnímu střelci. Jeho výsledky vykazovaly výrazně vyšší stabilitu než výsledky druhého střelce.

Je tedy vidět, že pouze průměr skutečně nemůže být jedinou veličinou charakterizující daný statistický soubor. Měl by být vždy doplněn údajem o struktuře tohoto souboru – o jeho **variabilitě**.

Ve statistické praxi je používáno celé řady veličin (měr), které slouží k zachycení míry variability statistického souboru. Těmi nejvýznamnějšími jsou zejména **rozptyl**, **směrodatná odchylka** a **variační rozpětí**.

Rozptyl a směrodatná odchylka

Rozptyl

Rozptyl statistického souboru patří mezi jeho nejvýznamnější charakteristiky. Zkoumá variabilitu všech hodnot obsažených v daném statistickém souboru. Tato variabilita je vztažena k aritmetickému průměru. Rozptyl je definován jako **průměr čtverců (druhých mocnin) odchylek jednotlivých hodnot statistického znaku od jejich aritmetického průměru**.

Vypočítáme jej tedy podle vztahu

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



Čím dostáváme vyšší hodnoty rozptylu, tím jsou hodnoty obsažené v souboru více rozvrstveny. Nulová hodnota rozptylu ukazuje na soubor s identickými hodnotami.

Uvedený vztah však není pro praktické výpočty příliš vhodný. Definiční vztah pro rozptyl je proto převáděn na tzv. výpočtový vztah. Jeho podoba je

$$s_x^2 = \overline{x^2} - \bar{x}^2$$



Odvození tohoto vztahu, stejně jako další varianty výpočtových tvarů rozptylu naleznete v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na stranách 49–50.

Nevýhodou rozptylu je však jeho poměrně špatná interpretace. Je počítán jako průměr odchylek, přičemž odchylky jsou brány ve své druhé mocnině. Je to z důvodu možnosti, kdy by se odchylky – bráné pouze v první mocnině – mohly vynulovat, aniž bychom mohli tvrdit, že daný soubor má nízkou variabilitu. Z důvodů špatné interpretace je v praktických příkladech rozptyl uváděn ve své odmocnině, která je nazývána **směrodatná odchylka**.

Směrodatná odchylka označuje jak se hodnoty ve statistickém souboru průměrně odchylují od svého aritmetického průměru. Spolu s hodnotou aritmetického průměru do jisté míry udává nejpravděpodobnější místo (pás hodnot) výskytu jednotlivých prvků souboru.

Směrodatnou odchylku tedy vypočítáme jako

$$s_x = \sqrt{s_x^2}$$

Variační rozpětí

Chceme-li získat pouze orientační informaci o variabilitě statistického souboru, je možno využít některé z výpočtově méně náročných veličin. Nejznámější z nich je **variační rozpětí**. Je definováno jako rozdíl maximální a minimální hodnoty statistického souboru, tedy jako

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

Jeho výhodou je velmi jednoduchý výpočet a velmi snadná interpretace. Oproti rozptylu, který je počítán ze všech hodnot daného souboru, je však variační rozpětí pouze výběrovou charakteristikou počítanou ze dvou prvků tohoto souboru. Hrozí proto riziko, že bude hodnota této veličiny ovlivněna netypickou (příliš odlehlou) hodnotou jedné z mezí, které jsou k výpočtu využívány. Může pak dojít k výraznému zkreslení pohledu na daný soubor, například ve smyslu příliš vysoké variability z důvodů nahodilé maximální hodnoty.

Variační koeficient

Pro srovnávání variability více statistických souborů je vhodné převést míry variability na stejné jednotky (či na relativní, bezrozměrné veličiny). Nejjednodušší možností je vydělit příslušnou charakteristiku průměrnou hodnotou souboru.

Podělíme-li směrodatnou odchylku aritmetickým průměrem hodnot daného statistického souboru, dostáváme nejznámější veličinu – **variační koeficient**. Vypočítáme jej tedy jako

$$V_x = \frac{s_x}{\bar{x}}$$

V některých případech bývají mezi míry variability zařazovány také míry vzájemného vztahu mezi dvěma statistickými soubory. Představitelem těchto veličin je především kovariance. Její definici a výpočet naleznete v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na stranách 52–62.



Příklad 2.6

Pro soubor známek studentů vybraného studijního oboru ze zkoušky ze statistiky (uvedeného v příkladu 2.1) posuďte jeho variabilitu.



2. Zpracování a popis statistických souborů

Řešení

Víme, že 20 studentů daného oboru získalo u zkoušky následující známky (na stupnici 1 – 4): 1, 1, 1, 2, 1, 3, 4, 1, 2, 1, 3, 2, 1, 3, 4, 2, 2, 3, 1, 2. Můžeme tedy označit jednotlivé prvky tohoto souboru symboly x_i , kde i nabývá hodnot od jedné do dvaceti.

Variabilitu souboru posoudíme všemi výše uvedenými mírami - rozptylem směrodatnou odchylkou, variačním rozpětím a variačním koeficientem.

K výpočtu rozptylu uijeme jeho výpočtový vztah. Postup výpočtu a dílčí výsledky jsou v následující tabulce.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Σ
x_i	1	1	1	2	1	3	4	1	2	1	3	2	1	3	4	2	2	3	1	2	40
x_i^2	1	1	1	4	1	9	16	1	4	1	9	4	1	9	16	4	4	9	1	4	100

Rozptyl určíme podle vztahu

$$s_x^2 = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \frac{1 + 1 + 1 + 4 + \dots + 1 + 4}{20} - 2^2 = 5 - 4 = 1.$$

Směrodatná odchylka je rovna druhé odmocnině rozptylu

$$s_x = \sqrt{s_x^2} = \sqrt{1} = 1.$$

Variační rozpětí je rozdíl mezi maximálním a minimálním prvkem souboru. Je tedy rovno

$$R = x_{\max} - x_{\min} = 4 - 1 = 3.$$

Variační koeficient je podíl aritmetického průměru a příslušné směrodatné odchylky.

$$V_x = \frac{s_x}{\bar{x}} = \frac{1}{2} = 0,5.$$

Na základě výše uvedených výsledků tedy lze konstatovat následující skutečnosti:

- Rozdíl mezi nejvyšší a nejnižší dosaženou výslednou známku studentů u zkoušky je roven 3.
- Rozptyl těchto známek je roven 1 a směrodatná odchylka, tedy průměrná odchylka od aritmetického průměru je rovna 1.
- Studenti u zkoušky tedy v průměru dosáhli známky v rozmezí 2 ± 1 .

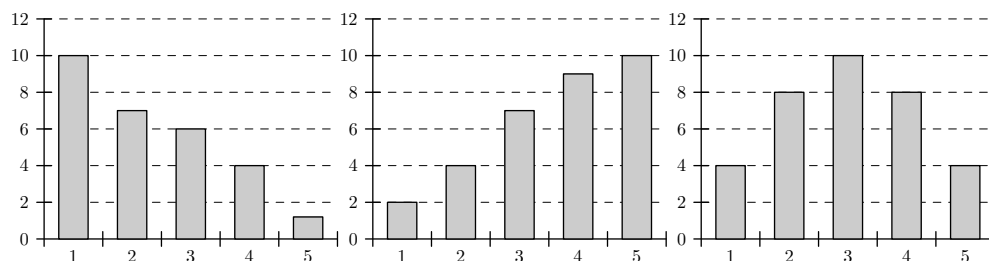
2.2.3 Míry šikmosti a špičatosti

Míry šikmosti a špičatosti statistického souboru jsou doplňující informací o hodnotách statistického souboru. Jsou určeny zejména ke srovnání míry nahuštění hodnot ve statistickém souboru.

Šikmost

Míry šikmosti srovnávají míru nahuštění malých hodnot sledovaného statistického znaku s mírou nahuštění velkých hodnot tohoto souboru. Míra šikmosti je úzce spjatá s tvarem grafu rozdělení četnosti prvků souboru.

Je-li míra koncentrace malých i velkých hodnot přibližně stejná, dostáváme obvykle velmi **symetrické** rozdělení četností. Je-li ve statistickém souboru vyšší stupeň koncentrace malých hodnot ve srovnání s koncentrací velkých hodnot, je tvar grafu rozdělení četností zešikmený. V tomto případě hovoříme o **kladné šikmosti**. V opačném případě – tedy kdy je míra koncentrace vyšších hodnot vyšší než koncentrace malých hodnot, hovoříme o **záporné šikmosti**.

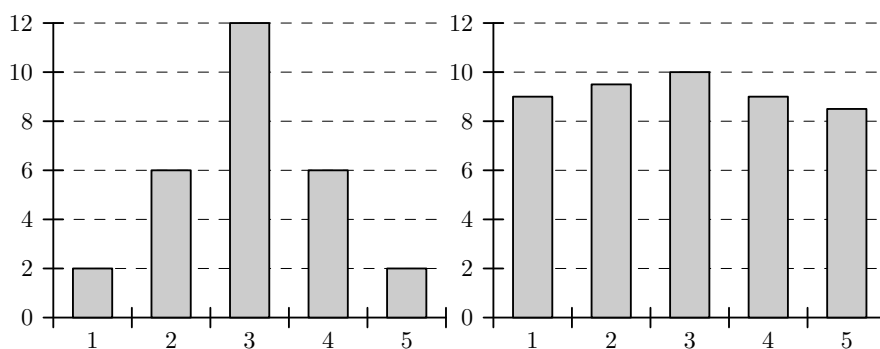


Obrázek 2.4: Kladná, záporná a nulová šikmost statistického souboru.

Míry špičatosti srovnávají míru nahuštění hodnot kolem prostřední hodnoty souboru s mírou nahuštění ostatních hodnot proměnné. Je-li koncentrace hodnot kolem prostředních hodnot přibližně stejná jako koncentrace v celém statistickém souboru, jsou hodnoty špičatosti poměrně nízké. Tato skutečnost se projeví přibližně plochým grafem rozdělení četností.

Špičatost

Naopak, je-li míra nahuštění kolem prostřední hodnoty výrazně vyšší než koncentrace ostatních hodnot ve statistickém souboru, dostáváme vysoké hodnoty špičatosti. V tomto případě se graf rozdělení četností projevuje špičatým tvarem.



Obrázek 2.5: Vysoká a nízká špičatost statistického souboru.

Konkrétní hodnoty šikmosti je možno určit na základě výpočtových vztahů, které jsou založeny na veličinách podobných rozptylu. Jejich výpočet je z těchto důvodů poměrně numericky náročný a zdlouhavý. Míry šikmosti i špičatosti jsou však obvyklou součástí statistického software a proto tyto vztahy neuvádíme. Můžete je nalézt například v učebnici SEGER, HINDLS, HRONOVÁ: *Statistika v hospodářství* na stranách 62–67.





Shrnutí kapitoly

Nejjednodušší nástroje popisu statistických souborů jsou zahrnovány pod oblast tzv. popisné statistiky. Kapitola shrnuje základní veličiny a postupy používané v popisné statistice. Základní informaci o rozsáhlém statistickém souboru lze získat jeho utříděním podle jednoho, či několika statistických znaků. Obvykle se zajímáme především o míru výskytu jednotlivých variant znaku v tomto souboru, které nazýváme četnosti. K zobrazení rozložení četností znaků ve statistickém souboru slouží tabulky rozdělení četností a především statistické grafy rozložení četností.

Druhá část kapitoly prezentuje definice, metody výpočtu a užití tzv. charakteristik (nebo měr) statistických souborů. Jsou užívány k doplnění informace o statistickém souboru, kterou získáme pomocí rozdělení četností. Do základních měr patří především průměry, které ukazují hodnotu, která nejlépe charakterizuje daný soubor. Průměr je nutno dále doplnit o jeho variabilitu, která udává jeho validitu (vypovídací hodnotu). Variabilita je měřena zejména rozptylem. Vyšší hodnoty rozptylu hovoří o poměrně vysoké variabilitě a tedy i malé vypovídací hodnotě aritmetického průměru.



Otázky k zamyšlení

1. K jakým účelům je možno využít kumulativních hodnot četností? Naleznete alespoň dva příklady z hospodářské praxe.
2. Vlekař si dělal statistiku přepravených lyžařů. Mezi 8–10 hod přepravil 65 lyžařů, 10–12 hod 103 lyžaře, 14–16 hod 85 lyžařů a 16–18 hod 30 lyžařů. Uspořádejte mu data do tabulky a znázorněte graficky.
3. V následující tabulce jsou obsaženy údaje o počtu žijících osob v pěti největších městech ČR. Doplň tabulku a interpretuj vlastními slovy čísla nacházející se v šedých buňkách.

	Absolutní četnost	Relativní četnost	Kumulativní absolutní četnost	Kumulativní relativní četnost
Praha	1 200 455			
Brno	385 866			
Ostrava	323 177			
Plzeň	169 391			
Olomouc	103 840			
Σ				

4. Pro 7-mi prvkový statistický soubor (2, 12, 14, 18, 10, 3, 15) dokažte vztah mezi aritmetickým, geometrickým a harmonickým průměrem uvedený na straně 30.

POT



Součástí studia předmětu je i vypracování a odevzdání dvou krátkých samostatných prací, které jsou označovány jako POT. Obě samostatné práce mají formu příkladu, který by Vám měl dát možnost otestovat vědomosti nabyté v předchozí části studijní opory. Výsledky obou POTů odevzdáte ve stanovených termínech tutorovi v elektronické podobě (soubor v MS EXCEL + případný doprovodný text).

Termíny odevzdání jednotlivých úkolů jsou následující:

- POT 1 – 4. týden v říjnu
- POT 2 – 2. týden v prosinci

Odevzdání POTů a jejich správné řešení je podmínkou připuštění ke zkoušce z předmětu.

Zadání POT 1

Máte k dispozici statistický soubor, který charakterizuje výsledky šetření mezi 20 osobami o jejich průměrných měsíčních výdajích na bydlení na osobu (v tisících).

x_i	3,5	7	12	2	14	5	6	4	10	3
x_i	5,5	10,5	20,5	3	11	7,5	9,5	3	11,5	1,5

Pro tento statistický soubor:

1. Vytvořte intervalové rozdělení četností.
2. Pomocí rozdělení četností odhadněte hodnotu mediánu a 20% kvantilu souboru.
3. Určete hodnotu aritmetického průměru a rozptylu tohoto souboru.
4. V prostředí programu EXCEL vytvořte histogram rozdělení četností a ověřte výsledky z bodů 1 – 3.

Příloha kapitoly 2

Výpočet četností a popisných statistických veličin v prostředí programu MS EXCEL

Pro analýzu statistických souborů je v současné době s úspěchem možno využít nejrůznějších specializovaných statistických programů. Z těch nejznámějších můžeme jmenovat např. program SPSS, STATISTICA, MATLAB, MAPLE. Mimo tyto programy je však možno využít i méně specializované programy. Patří sem zejména tzv. tabulkové procesory. Jak jsme již zmínili v úvodu, bude látka Aplikované statistiky doprovázena příklady v (patrně nejrozšířenějším) programu EXCEL od společnosti Microsoft. Příklady jsou vedeny v prostředí MS EXCEL verze 2000.

V následující kapitole naznačíme pouze některé z dostupných nástrojů pro popis a analýzu statistických souborů v EXCELU. Zmíníme se o některých funkcích, pomocí nichž můžeme určit **charakteristiky (míry)** statistických souborů zmíněné v této kapitole. Další oblastí, kde lze použít funkcí programu EXCEL, je výpočet četností a vytváření grafů **četností**, především **histogramu**.

Výpočet některých statistických veličin v EXCELU

a) pomocí vestavěných funkcí

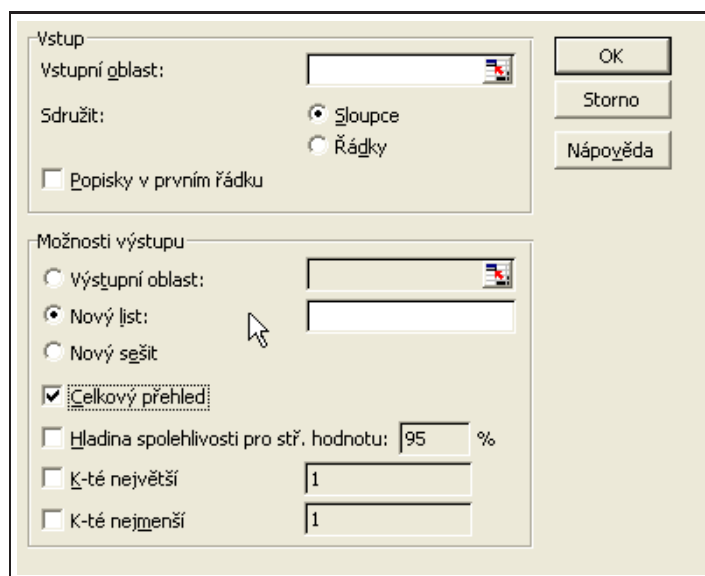
Pro výpočet charakteristik statistických souborů je možno užít vestavěných „excelovských“ funkcí. Jejich (nikoliv vyčerpávající přehled) uvádí následující tabulka. Syntaxe těchto příkazů není příliš složitá. Navíc lze s úspěchem použít tzv. Průvodce vložením funkce, proto uvádíme pouze názvy těchto funkcí.

COUNTIF	počet buněk v oblasti, splňujících kritérium
ČETNOSTI (data, hodnoty)	počet výskytů hodnot (vrátí vertikální matici dat)
GEOMEAN	geometrický průměr
HARMEAN	harmonický průměr
LARGE	k -tá nejvyšší hodnota množiny dat
MEDIÁN	medián
MODE	modus
PERCENTIL	k -tý percentil (kvantil)
QUARTIL	kvartil (25% kvantil)
RANK	pořadí čísla v seznamu
SKEW	šikmost
SMALL	k -tá nejmenší hodnota
SMODCH	směrodatná odchylka
VAR	rozptyl

b) pomocí Analytického nástroje „Popisná statistika“

Chceme-li získat komplexní přehled o statistickém souboru, je možno využít tzv. Analytického nástroje „Popisná statistika“. Lze jej aktivovat z nabídky Nástroje/Analýza dat/Popisná statistika.

Dostáváme se do následujícího okna:



Obrázek 2.6: Analýza dat – popisná statistika

Vyplníme „Vstupní oblast“, kam zadáme buňky, ve kterých jsou jednotlivé prvky statistického souboru. Dále je nutno zatrhnout nabídku „Celkový přehled“ a zvolit kam chceme výsledný přehled vypsat (je vhodné jej umístit na nový list). Výsledkem je např. následující přehled:

Sloupec 1	
Stř. hodnota	2
Chyba stř. hodnoty	0,229416
Medián	2
Modus	1
Směr. odchylka	1,025978
Rozptyl výběru	1,052632
Špičatost	-0,67059
Šikmost	0,649786
Rozdíl max–min	3
Minimum	1
Maximum	4
Součet	40
Počet	20

Výpočet četností a jejich zobrazení

Pro výpočet četností výskytu statistického znaku v souboru je možno využít několika funkcí. Nejjednodušším způsobem je využití „Filtrů“, které umožňují přímé zobrazení prvků odpovídajících danému kritériu. Je tak rychle možno

2. Zpracování a popis statistických souborů

třídít statistický soubor a určit příslušné absolutní četnosti (jak spojité tak i diskrétní veličiny).

a) Funkce ČETNOSTI

Pro intervalové rozdělení četností je v programu EXCEL možno využít také speciální vestavěné funkce ČETNOSTI. Její použití je možné dle následujícího postupu:

■ Určit hranice intervalů (tříd)

Do samostatného sloupce zadáme hodnoty, které budou tvořit hraniční hodnoty jednotlivých intervalů četnosti. Jejich volba je závislá především na našem úsudku, vycházejícím ze zkušenosti, či znalostí o daném statistickém souboru.

■ Označit místo, kam budou četnosti počítány

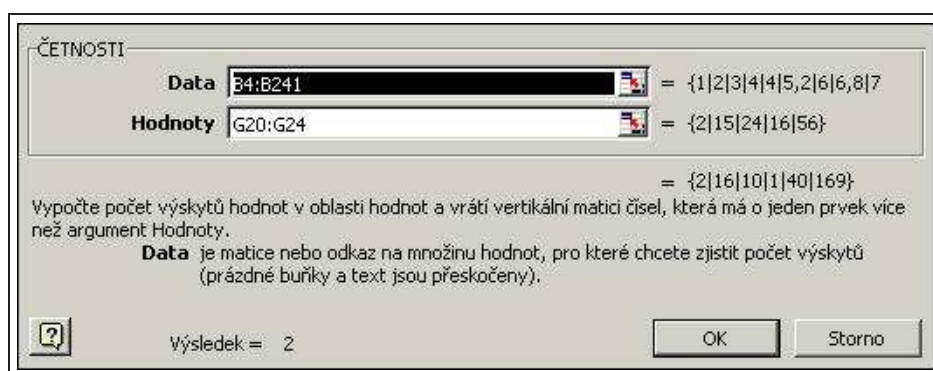
Druhým krokem je označení místa, kam bude směřovat výstup funkce ČETNOSTI. Je vhodné jej umístit vedle nadefinovaných hranic tříd. Vždy jenutno označit o jednu více buňku než je počet nadefinovaných tříd. Bude to interval, který odpovídá intervalu „vyšší než poslední hranice třídy“.

■ Zadání funkce ČETNOSTI

Po označení místak KAM chceme funkci počítat, zadáme z nabídky Vložit/Funkci/ funkci ČETNOSTI (v sekci Statistické funkce). Musíme vyplnit dva údaje :

Data – označují buňky, ve kterých jsou prvky zkoumaného statistického souboru

Hodnoty – označují buňky, ve kterých jsou nadefinované hranice tříd Jelikož je výstupem funkce ČETNOSTI pole (tedy několik hodnot naráz), musíme ji také jako pole zadat. Po potvrzení tlačítkem OK se ve výstupu vyplní pouze jedna buňka. Proto je nutno kliknout myší do prostoru definice funkcí a stitknout klávesy CTRL+SHIFT+ENTER. Vyplní se zbylé označené buňky a u definice funkce se objeví složené závorky.

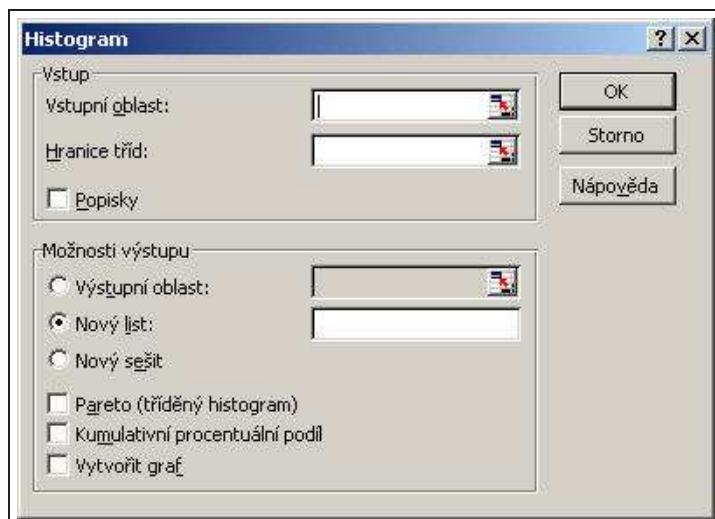


Obrázek 2.7: Zadání funkce ČETNOSTI v programu MS EXCEL 2000

Z takto vypočítaných hodnot je možno vytvořit graf rozdělení četností, ať už ve formě Polygonu četností (v EXCELU spojnicový graf) či ve formě Histogramu četností (sloupcový graf).

b) Analytický nástroj „Histogram“

Jinou možností jak získat hodnoty četností a histogram jejich rozdělení ve statistickém souboru je využití nástroje Histogram. Najdeme ho opět v nabídce Nástroje/Analytické nástroje. Zadání tohoto nástroje je velmi obdobné jako u výše zmíněné funkce ČETNOSTI. Opět nejprve musíme nadefinovat hranice tříd (pokud tak neučiníme a necháme příslušnou položku v nástroji nevyplněnou, nastaví EXCEL hranice automaticky).



Obrázek 2.8: Zadání analytického nástroje histogram v programu MS EXCEL 2000

Tento analytický nástroj nabízí výstup ve formě tabulky, kde jsou uvedeny hodnoty absolutních četností a kumulativních a relativních četností. Stejně tak je pomocí něj možno vytvořit histogram rozdělení četností a to i jako tříděný – tedy uspořádaný od nejčetnějšího znaku k nejméně četnému.

Příklad výstupu nástroje Histogram:

Třídy	Četnost	Kumul. %
2	2	1,63%
15	16	14,63%
16	1	15,45%
24	10	23,58%
56	40	56,10%
Další	54	100,00%

