Popisné ukazatele

Analyzujeme-li nějaká data, může být užitečné rozložit je na intervaly, různými způsoby je zakreslit do grafu nebo popsat jejich vlastnosti na základě čísel odvozených z určitých vzorců. Tyto postupy nazýváme *popisné (deskriptivní) ukazatele*.

Percentily

Možná, že někteří z vás už zažili národní srovnávací zkoušky, které probíhají na českých středních školách od roku 1996 a kterým se obecně říká „SCIO testy". Jejich cílem je objektivní srovnání účastníků několika různě obtížných verzí testů a určení jejich celkového pořadí, což vyžaduje přepočítání na *percentily*.

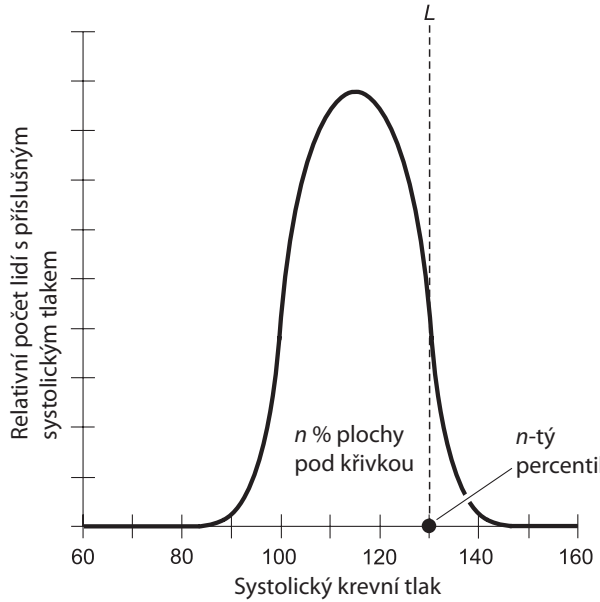
Percentil tedy vyjadřuje, jak se jednotlivý účastník umístil v rámci všech ostatních účastníků, neboli kolik procent ostatních účastníků dosáhlo horšího výsledku, popř. absolutní pořadí účastníka.

Percentily v normálním rozložení

Percentily dělí velkou množinu čísel na 100 intervalů, kde každý z nich obsahuje 1 % prvků v množině. Existuje 99 možných percentilů, ne 100, protože percentily představují hranice, kde se těchto 100 intervalů setkává.

Představte si experiment, ve kterém se obrovskému počtu lidí odečte systolický krevní tlak. Jedná se o vyšší ze dvou čísel, které se vám zobrazí na měřicím přístroji. Pokud tedy máte tlak 110/70, čteme „110 na 70", pak má váš systolický tlak hodnotu 110. Dejme tomu, že výsledky tohoto testu dostaneme v grafické podobě a že křivka vypadá jako spojité rozložení, protože je v tomto případě populace tak velká. Jedná se o normální rozložení, které má zvonovitý tvar a je symetrické (obrázek 4.1).

Vybereme jakoukoli hodnotu tlaku na vodorovné ose a nahoru od ní promítneme čáru. Abychom mohli určit percentil odpovídající této hodnotě, musíme najít takové číslo n, aby se nejméně n % plochy pod křivkou nacházelo nalevo od čáry L. Poté n zaokrouhlíme na nejbližší celé číslo od 1 do 99 včetně a získáme percentil p. Například oblast nalevo od čáry L představuje 93,3 % plochy pod křivkou, což znamená, že $n = 93{,}3$. Z toho vyplývá, že se krevní tlak odpovídající bodu, kde čára L protíná vodorovnou osu, rovná 93. percentilu.



Obrázek 4.1: Určení percentilů v normálním rozložení

Polohu jakéhokoli určitého bodu percentilu (hranice), řekněme *p*-tý percentil, najdeme tak, že promítneme svislou čáru tím způsobem, aby *n* procent plochy pod křivkou přesně odpovídalo *p*, a pak si poznamenáme bod, kde tato čára protnula vodorovnou osu. Představte si u obrázku 4.1, že bychom čárou *L* mohli pohybovat tam a zpět jako posuvnými dveřmi. Rovná-li se číslo *n*, které představuje procentuální podíl plochy pod křivkou nalevo od *L*, přesně 93, pak čára protíná vodorovnou osu v hraničním bodu 93. percentilu. Přestože nás to svádí myslet si, že by mohl existovat „nultý percentil" ($n = 0$) a „stý percentil" ($n = 100$), nepředstavuje žádný z těchto percentilů hranici, kde by se setkaly dva intervaly.

Povšimněte si rozdílu mezi tím, když řekneme, že se nějaký určitý tlak „nachází v" *p*-tém percentilu anebo že se nějaký tlak „nachází na" *p*-tém percentilu. V prvním případě popisujeme interval údajů, ve druhém mluvíme o hraničním bodu mezi dvěma intervaly.

Percentily vyjádřené pomocí tabulky

Představte si 1 000 žáků, kteří píší test se 40 otázkami, přičemž dosáhnou všech z 41 možných výsledků. Existují lidé, kteří napíší dokonalý test, ale i nešťastníci, kteří nemají ani jedinou odpověď správně. Tabulka 4.1 obsahuje v prvním sloupci výsledky testů uvedené vzestupně od 0 do 40. Druhý sloupec obsahuje počet žáků, kteří dosáhli odpovídajícího výsledku (absolutní četnost), a třetí sloupec udává kumulativní absolutní četnost vyjádřenou od nejnižšího do nejvyššího výsledku.

Kam umístíme 99 bodů percentilu (hranic) v této množině dat? Jak můžeme do jedné množiny s 41 možnými výsledky umístit 99 čar? Odpověď očividně zní, že nemůžeme. A co tedy rozdělit žáky do skupin? Testu se podrobilo tisíc lidí. Proč je tedy nerozdělíme na 100 různých skupin s 99 rozdílnými hranicemi a nenazýváme je „body percentilu" podle následujících znaků?

- 10 „nejhorších“ testů a první bod percentilu na vrcholu této skupiny.
- 10 „druhých nejhorších“ testů a druhý bod percentilu na vrcholu této skupiny.
- 10 „třetích nejhorších“ testů a třetí bod percentilu na vrcholu této skupiny.

  ↓
- 10 „*p*-tých nejhorších“ testů a *p*-tý bod percentilu na vrcholu této skupiny.

Tabulka 4.1: Výsledky hypotetického testu s 40 otázkami, který psalo 1 000 žáků

Výsledek testu	Absolutní četnost	Kumulativní absolutní četnost
0	5	5
1	5	10
2	10	20
3	14	34
4	16	50
5	16	66
6	18	84
7	16	100
8	12	112
9	17	129
10	16	145
11	16	161
12	17	178
13	22	200
14	13	213
15	19	232
16	18	250
17	25	275
18	25	300
19	27	327
20	33	360
21	40	400
22	35	435
23	30	465
24	35	500
25	31	531
26	34	565
27	35	600
28	34	634
29	33	667
30	33	800
31	50	750
32	50	800
33	45	845

Výsledek testu	Absolutní četnost	Kumulativní absolutní četnost
34	27	872
35	28	900
36	30	930
37	28	958
38	20	978
39	12	990
40	10	1 000

↓

- 10 „*q*-tých nejlepších“ testů a *p*-tý bod percentilu na vrcholu této skupiny.

↓

- 10 „třetích nejlepších“ testů a 97. bod percentilu na konci této skupiny.
- 10 „druhých nejlepších“ testů a 98. bod percentilu na konci této skupiny.
- 10 „nejlepších“ testů a 99. bod percentilu na konci této skupiny.

Na první pohled vypadá toto rozřazení výborně, ale má to háček. Pokud zkontrolujeme tabulku 4.1, uvidíme, že 50 lidí získalo v testu 31 bodů, což odpovídá pěti skupinám po deseti lidech se stejným výsledkem. Tyto výsledky jsou všechny „stejně dobré“ (nebo „stejně špatné“). Pokud řekneme, že se kterýkoli z těchto testů nachází „v *p*-tém percentilu“, pak musíme samozřejmě také říct, že se v tomto „*p*-tém percentilu“ nacházejí i všechny ostatní testy s tímto výsledkem. Nemůžeme si svévolně vybrat 10 testů s výsledkem 31 a dát je do *p*-tého percentilu, pak si vzít dalších 10 testů s výsledkem 31 a dát je do *p*-tého + 1 percentilu, pak si vzít dalších 10 testů s výsledkem 31 a dát je do *p*-tého + 2 percentilu, pak si vzít dalších 10 testů s výsledkem 31 a dát je do *p*-tého + 3 percentilu a nakonec si vzít dalších 10 testů s výsledkem 31 a dát je do *p*-tého + 4 percentilu. To by nebylo spravedlivé.

Situace s percentily se nám pomalu komplikuje a stává se nepřehlednou, že? Možná vás napada, kdo tento návrh vůbec vymyslel. Na tom nezáleží. Toto schéma se běžně používá a my se ho budeme držet. Co tedy můžeme udělat, abychom to všechno vysvětlili a našli vzorec, který by dával smysl pro všechny možné scénáře?

Body percentilu

Výše uvedený hlavolam můžeme vyřešit tak, že definujeme schéma na vypočítání pozic bodů percentilu v množině *uspořádaných datových prvků.* Jedná se o množinu uspořádanou v tabulce od „nejlepšího po nejhorší“, jako například v tabulce 4.1. Jakmile jednou určíme schéma pozic percentilu, přijmeme ho jako pravidlo a navždy tak ukončíme všechen chaos.

Tak tedy: Máme za úkol najít pozici *p*-tého percentilu v množině *n* uspořádaných datových prvků. Nejdříve vynásobíme *p* a *n* a následně jejich součin vydělíme 100. Vyjde nám číslo *i*, které nazýváme *index*:

$$i = pn/100$$

Zde jsou uvedena pravidla:

- Není-li *i* celé číslo, pak je pozice *p*-tého bodu percentilu rovna *i* + 1.
- Je-li *i* celé číslo, pak je pozice *p*-tého bodu percentilu rovna *i* + 0,5.

Pořadí percentilu

Chceme-li najít pořadí percentilu p pro daný prvek nebo pozici s v množině uspořádaných datových prvků, můžeme použít jinou definici. Počet prvků menších než s (říkejme tomuto číslu t) vydělíme celkovým počtem prvků n a tuto hodnotu vynásobíme 100, čímž získáme předběžný percentil p^*:

$$p^* = 100\ t/n$$

Hodnotu p^* potom zaokrouhlíme na nejbližší celé číslo od 1 do 99 včetně, čímž získáme pořadí percentilu p pro daný prvek nebo jeho pozici v množině.

Pořadí percentilů určená tímto způsobem odpovídají intervalům, jejichž středy se nacházejí na hranicích percentilů tak, jak jsme to popsali výše. Podle tohoto schématu jsou pořadí 1. a 99. percentilu často příliš velká, a to zejména, jedná-li se velkou populaci. Je to proto, že 1. a 99. pořadí percentilu obsahují prvky, které se nacházejí na nejzazších místech nějaké množiny nebo nějakého rozložení.

Inverze percentilu

Čas od času možná uslyšíte, jak lidé používají výraz „percentil" v opačném smyslu slova. Budou mluvit o „prvním percentilu", i když ve skutečnosti budou myslet 99., o „2. percentilu", i když ve skutečnosti budou myslet 98., a tak dále. Vyvarujte se toho! Pokud budete skládat nějaký test, budete z něj mít dobrý pocit, a nakonec vám bude řečeno, že patříte do „4. percentilu", nepanikařte. Zeptejte se vyučujícího nebo správce testů, co to doopravdy znamená. Nejlepší 4 %? Nejlepší 3 %? Nejlepších 3,5 %? Nebo co tedy? Nebuďte překvapení, pokud si učitel nebo správce nebudou jistí.

Problém 4.1

Kde se nachází 56. bod percentilu v množině dat, kterou znázorňuje tabulka 4.1?

Řešení 4.1

Testu se účastnilo 1 000 žáků (datových prvků), proto $n = 1000$. Chceme najít 56. bod percentilu, tedy $p = 56$. Nejdříve spočítáme index:

$$i = (56 \times 1\,000)/100$$

$$= 56\,000/100$$

$$= 560$$

Vyjde nám celé číslo, takže k němu musíme přičíst 0,5, čímž dostaneme $i + 0{,}5 = 560{,}5$. 56. percentil je tedy hranice mezi „560. nejhorším" a „561. nejhorším" výsledkem testu. Abychom zjistili, jaký konkrétní výsledek je s tímto percentilem spojen, musíme ověřit kumulativní absolutní četnosti v tabulce 4.1. Kumulativní četnost odpovídající výsledku 25 správných odpovědí je 531 (což je méně než 560,5), kumulativní četnost odpovídající výsledku 26 správných odpovědí je 565 (což je více než 560,5). Proto se 56. percentil nachází mezi výsledky 25 a 26 správných odpovědí.

Problém 4.2

Pokud byste patřili k žákům, kteří se testu účastnili a dosáhli 33 správných odpovědí, jaké by bylo pořadí vašeho percentilu?

Řešení 4.2

Prohlédněte si opět tabulku a všimněte si, že 800 žáků má nižší výsledek nežli vy. (Nejedná se o hodnotu menší nebo rovnou, než je váš výsledek, ale pouze o menší než váš výsledek!) V souladu s druhou definicí uvedenou výše to znamená, že $t = 800$. Předběžný percentil p^* je tedy:

$$p^* = 100 \times 800/1000$$

$$= 100 \times 0{,}8$$

$$= 80$$

Vyšlo nám celé číslo, takže zaokrouhlení na nejbližší celé číslo od 1 do 99 má za výsledek $p = 80$. Jste tedy v 80. percentilu.

Problém 4.3

Pokud byste patřili k žákům, kteří se testu účastnili a dosáhli 0 správných odpovědí, jaké by bylo vaše pořadí percentilu?

Řešení 4.3

V tomto případě nemá nikdo nižší výsledek než vy. V souladu s druhou definicí uvedenou výše to znamená, že $t = 0$. Předběžný percentil p^* je tedy:

$$p^* = 100 \times 0/1000$$

$$= 100 \times 0$$

$$= 0$$

Vzpomeňte si na pravidlo, že abychom získali skutečnou hodnotu percentilu, musíme výsledek zaokrouhlit na nejbližší celé číslo od 1 do 99. Proto $p = 1$ a vy se řadíte do prvního percentilu.

Kvartily a decily

Množiny dat můžeme rozdělit i jinými způsoby, nežli pomocí percentilů: Je docela běžné určovat body nebo hranice, které dělí data na čtvrtiny nebo desetiny.

Kvartily v normálním rozložení

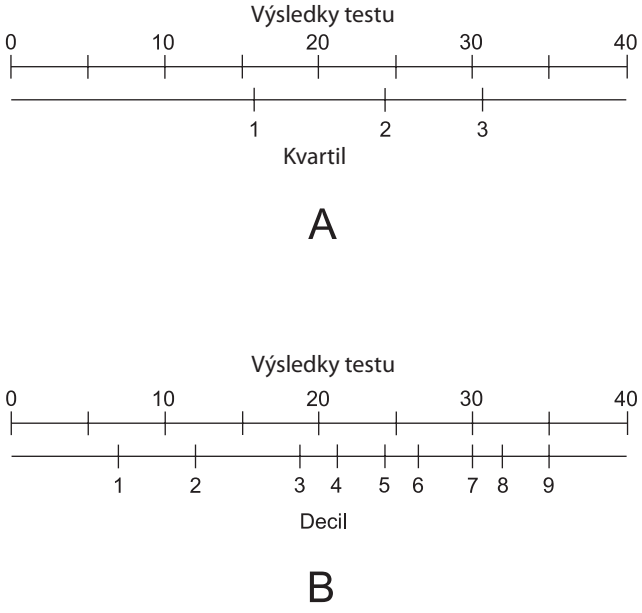
Kvartil je číslo, které dělí množinu dat na čtyři intervaly, kde každý z nich obsahuje 1/4 neboli 25 % prvků v množině. Existují tři kvartily, ne čtyři, protože kvartily představují hranice, kde se tyto čtyři intervaly setkávají. Kvartilům jsou proto přiřazovány hodnoty 1, 2 nebo 3 a někdy je také nazýváme jako první, druhý nebo třetí kvartil.

Prohlédněte si opět obrázek 4.1. Polohu q-tého bodu kvartilu najdeme tak, že promítneme svislou čáru L tím způsobem, aby n procent plochy pod křivkou přesně odpovídalo $25q$, a pak si poznamenáme bod, kde čára L vodorovnou osu protnula. Představte si u obrázku 4.1, že bychom čárou L mohli pohybovat sem a tam. Hodnota n představuje procentuální podíl plochy pod křivkou, která se nachází nalevo od L. Je-li n = 25 %, pak čára L protíná vodorovnou osu v prvním bodu kvartilu. Je-li n = 50 %, pak čára L protíná vodorovnou osu v druhém bodu kvartilu, a je-li n = 75 %, pak čára L protíná vodorovnou osu ve třetím bodu kvartilu.

Kvartily vyjádřené pomocí tabulky

Vraťme se nyní k testu se 40 otázkami, který je popsán výše a který je znázorněn v tabulce 4.1. Existuje 41 možných výsledků a 1 000 skutečných datových prvků. Těchto tisíc výsledků rozdělíme na čtyři skupiny se třemi různými hraničními body dle následujících znaků:

- Nejvyšší možný hraniční bod představuje „nejhorších" 250 nebo méně testů a první bod kvartilu na vrcholu této skupiny.
- Nejvyšší možný hraniční bod představuje „nejhorších" 500 nebo méně testů a druhý bod kvartilu na vrcholu této skupiny.
- Nejvyšší možný hraniční bod představuje „nejhorších" 750 nebo méně testů a třetí bod kvartilu na vrcholu této skupiny.



Obrázek 4.2: Obrázek A znázorňuje pozice kvartilů pro výsledky popsaného testu, obrázek B pozice decilů

Nomogram v obrázku 4.2A znázorňuje pozice bodů kvartilu pro výsledky testu uvedených v tabulce 4.1. Údaje v tabulce jsou neobvyklé: Jedná se totiž o shodu okolností, protože kvartily jsou jasně definované. Mezi „nejhoršími" a „druhými nejhoršími" 250 testy, mezi „druhými a třetími nejhoršími" 250 testy, ale i mezi „třetími nejhoršími a nejlepšími" 250 testy se vyskytují zřetelné

hranice. Ty se objevují přesně mezi testy s 16 a 17 správnými odpověďmi pro první kvartil, 24 a 25 správnými odpověďmi pro 2. kvartil a 31 a 32 správnými odpověďmi pro 3. kvartil. Je celkem jisté, že pokud by stejných 1 000 žáků dostalo jiný test se 40 otázkami nebo pokud by tento test se 40 otázkami dělalo jiných 1 000 žáků, pak by kvartily nebyly tak jasně zřetelné.

Problém 4.4

V tabulce 4.2 je uvedena část výsledků stejných testů se 40 otázkami, ale s trochu odlišnými výsledky, než jaké obsahovala tabulka 4.1, takže první bod kvartilu není „jasně" definován. Kde se v tomto případě nachází třetí bod kvartilu?

Řešení 4.4

Definici si vyložíme doslovně: První kvartil je „nejvyšší možný" hraniční bod na vrcholu množiny „nejhorších" 250 *nebo méně* testů. V tabulce 4.2 to odpovídá testům s 16 a 17 správnými odpověďmi.

Tabulka 4.2: Tabulka k problému 4.4

Výsledek testu	Absolutní četnost	Kumulativní absolutní četnost
↑	↑	↑
↑	↑	↑
↑	↑	↑
13	22	200
14	13	213
15	19	232
16	16	248
17	30	278
18	22	300
19	27	327
↓	↓	↓
↓	↓	↓
↓	↓	↓

Decily v normálním rozložení

Decil je číslo, které dělí množinu dat na 10 intervalů, kde každý z nich obsahuje 1/10 neboli 10 % prvků v množině. Existuje devět decilů, které představují body, kde se těchto 10 množin setkává. Decilům jsou proto přiřazovány hodnoty celých čísel od 1 do 9 včetně a někdy je také nazýváme první, druhý nebo třetí decil a tak dále až po devátý decil.

Vraťte se opět k obrázku 4.1. Polohu d-tého bodu decilu najdeme tak, že promítneme svislou čáru L tím způsobem, aby n procent plochy pod křivkou přesně odpovídalo $10d$, a pak si poznamenáme bod, kde čára L vodorovnou osu protnula. Opět si představte, že bychom čárou L mohli pohybovat sem a tam podle libosti. Hodnota n představuje procentuální podíl plochy pod křivkou, která se nachází nalevo od L. Je-li n = 10 %, pak čára L protíná vodorovnou osu v prvním bodu decilu. Je-li n = 20 %, pak čára L protíná vodorovnou osu v druhém bodu decilu, a je-li n = 30 %, pak čára L protíná vodorovnou osu ve třetím bodu decilu. Takto bychom mohli pokračovat dále až do n = 90 %, kde čára L vodorovnou osu protíná v devátém bodu decilu.

Decily vyjádřené pomocí tabulky

Ještě jednou se vraťme k testu se 40 otázkami, jehož výsledky jsou uvedeny v tabulce 4.1. Kam umístíme body decilu? 1 000 testů rozdělíme na 10 různých skupin s devíti různými hraničními body dle následujících znaků:

- Nejvyšší možný hraniční bod představuje „nejhorších" 100 nebo méně testů a první bod decilu na vrcholu této skupiny.
- Nejvyšší možný hraniční bod představuje „nejhorších" 200 nebo méně testů a druhý bod decilu na vrcholu této skupiny.
- Nejvyšší možný hraniční bod představuje „nejhorších" 300 nebo méně testů a třetí bod decilu na vrcholu této skupiny.

  ↓
- Nejvyšší možný hraniční bod představuje „nejhorších" 900 nebo méně testů a devátý bod decilu na vrcholu této skupiny.

Nomogram v obrázku 4.2B znázorňuje pozice bodů decilu pro výsledky testu uvedených v tabulce 4.1. Podobně jako v případě s kvartily jsou údaje v tabulce pouze shoda okolností, protože decily jsou jasně zřetelné. Mezi „nejhoršími" a „druhými nejhoršími" 100 testy, mezi „druhými a třetími nejhoršími" 100 testy, ale i mezi „třetími a čtvrtými nejhoršími" 100 testy a tak dále se vyskytují patrné hranice. Je celkem jisté, že pokud by stejných 1 000 žáků dostalo jiný test se 40 otázkami nebo pokud by tento test se 40 otázkami dělalo jiných 1 000 žáků, pak by decily byly méně zřejmé. (Teď už byste měli být schopní rozpoznat, že data v tabulce byla upravena tak, aby vycházely přehledné výsledky.)

Problém 4.5

Tabulka 4.3 obsahuje část výsledků stejného testu se 40 otázkami, ale s trochu odlišnými výsledky, než jaké obsahovala tabulka 4.1. V tomto případě není 6. bod decilu „jasně" definován. Kde se tedy nachází tento bod?

Tabulka 4.3: Tabulka k problému 4.5

Výsledek testu	Absolutní četnost	Kumulativní absolutní četnost
↑	↑	↑
↑	↑	↑
↑	↑	↑
24	35	500
25	31	531
26	34	565
27	37	602
28	32	634
29	33	667
30	33	700
↓	↓	↓
↓	↓	↓
↓	↓	↓

Řešení 4.5

Definici si opět vyložíme doslovně: 6. decil je *nejvyšší možný* hraniční bod na vrcholu množiny „nejhorších" 600 *nebo méně* testů. V tabulce 4.3 to odpovídá testům s 26 a 27 správnými odpověďmi.

Intervaly podle množství prvků

Percentily, kvartily a decily mohou být poněkud matoucí, například jako v případě, když vám někdo tvrdí: „Patříte do 99. percentilu této maturující třídy, což je nejvyšší možné pořadí." Poté, co by žáci této třídy takový výrok slyšeli, by se bezpochyby nejméně jeden z nich zeptal: „Nechtěl jste říci, že patřím do 100. percentilu?" Konec konců, výraz „percentil" vyjadřuje, že by mělo existovat 100 skupin, a ne 99.

Je v pořádku uvažovat o intervalech v tom smyslu, že se nacházejí mezi hranicemi percentilu, kvartilu nebo decilu, nežli že se soustředí kolem těchto hranic. Z čistě matematického hlediska vlastně tento postoj dává větší smysl. 99 bodů percentilu v uspořádané množině dat dělí tuto množinu na 100 intervalů, přičemž každý z nich obsahuje stejný počet (nebo co možná nejvíce podobný počet) prvků. Podobně dělí i tři body kvartilu nějakou uspořádanou množinu na čtyři intervaly, které jsou pokud možno stejně velké, a devět bodů decilu dělí uspořádanou množinu dat na 10 pokud možno stejně velkých intervalů.

25% intervaly

Vraťte se ještě jednou k tabulce 4.1 a představte si, že chcete výsledky vyjádřit jako nejnižších 25 %, druhých nejnižších 25 %, druhých nejvyšších 25 % a nejvyšších 25 %. Tabulka 4.4 ukazuje výsledky testu v 25% intervalech, kterým také můžeme říkat nejnižší čtvrtina, druhá nejnižší čtvrtina, druhá nejvyšší čtvrtina a nejvyšší čtvrtina. Tato určitá množina výsledků je i v tomto případě zvláštní, protože intervaly jsou „jasně" definované. Pokud by situace nebyla tak přehledná, nutilo by nás to vypočítat body kvartilu a následně určit 25% intervaly jako množiny výsledků mezi těmito hranicemi.

Tabulka 4.4: Výsledky hypotetického testu se 40 otázkami, kterého se účastnilo 1 000 žáků, s uvedenými 25% intervaly

Pořadí výsledků	Absolutní četnost	Kumulativní absolutní četnost	25% intervaly
0–16	250	250	nejnižších 25 %
17–24	250	500	2. nejnižších 25 %
25–31	250	750	2. nejvyšších 25 %
32–40	250	1 000	nejvyšších 25 %

10% intervaly

Znovu se podívejme na test, jehož výsledky jsou uvedeny v tabulce 4.1. Dejme tomu, že namísto toho, abychom mysleli na percentily, chceme výsledky vyjádřit jako nejnižších 10 %, druhých nejnižších 10 %, třetích nejnižších 10 % a tak dále až po nejvyšších 10 %. Tabulka 4.5 znázorňuje výsledky testu s těmito intervaly, jejichž rozpětí také můžeme nazývat jako první, druhá nebo

třetí desetina a tak dále až po nejvyšší desetinu (nebo chcete-li zůstat u prvního způsobu, desátá desetina).

Tabulka 4.5: Výsledky hypotetického testu se 40 otázkami, kterého se účastnilo 1 000 žáků, s uvedenými 10% intervaly

Pořadí výsledků	Absolutní četnost	Kumulativní absolutní četnost	10% intervaly
0–7	100	100	nejnižších 10 %
8–13	100	200	2. nejnižších 10 %
14–18	100	300	3. nejnižších 10 %
19–21	100	400	4. nejnižších 10 %
22–24	100	500	5. nejnižších 10 %
25–27	100	600	5. nejvyšších 10 %
28–30	100	700	4. nejvyšších 10 %
31–32	100	800	3. nejvyšších 10 %
33–35	100	900	2. nejvyšších 10 %
36–40	100	1 000	nejvyšších 10 %

Tato určitá množina výsledků je zvláštní, protože nejzazší body intervalů jsou „zřejmé". Pokud by tato tabulka nebyla upravena tak, aby byla diskuse co možná nejsnazší, museli bychom vypočítat body decilu a následně určit 10% intervaly jako množiny výsledků mezi těmito hranicemi.

Problém 4.6

V tabulce 4.6 je uvedena část výsledků testu se 40 otázkami, který řešilo 1 000 žáků, ale s trochu odlišnými výsledky, než jaké obsahovala tabulka 4.1. Jaké pořadí výsledků představuje v tomto případě druhých nejvyšších 10 %?

Tabulka 4.6: Tabulka k problému 4.6.

Test	Absolutní četnost	Kumulativní absolutní četnost
↑	↑	↑
↑	↑	↑
↑	↑	↑
30	35	702
31	51	753
32	50	803
33	40	843
34	27	870
35	31	901
36	30	930
↓	↓	↓
↓	↓	↓
↓	↓	↓

Řešení 4.6

Druhých nejvyšších 10 % rovněž odpovídá devátým nejnižším 10 %. Jedná se o výsledky, které jsou ohraničeny odspodu 8. a odshora 9. decilem. Nejvyššímu možnému hraničnímu bodu na vrcholu množiny „nejhorších" 800 *nebo méně* testů odpovídá 8. decil. V souladu s tabulkou 4.6 to odpovídá výsledkům s 31 a 32 správnými odpověďmi. 9. decil je nejvyšší možný hraniční bod na vrcholu množiny „nejhorších" 900 *nebo méně* testů, což se podle tabulky 4.6 shoduje s testy s 34 a 35 body. Devátých nejnižších (nebo druhých nejvyšších) 10 % výsledků se tedy nachází v rozmezí výsledků od 32 do 34 správných odpovědí včetně.

Problém 4.7

Tabulka 4.7 ukazuje část výsledků testu se 40 otázkami, kterého se účastnilo 1 000 žáků, ale s trochu odlišnými výsledky, než jaké obsahovala tabulka 4.1. Jaké pořadí výsledků představuje v tomto případě nejnižších 25 %?

Tabulka 4.7: Tabulka k problému 4.7

Test	Absolutní četnost	Kumulativní absolutní četnost
↑	↑	↑
↑	↑	↑
↑	↑	↑
14	12	212
15	18	230
16	19	249
17	27	276
18	26	302
↓	↓	↓
↓	↓	↓
↓	↓	↓

Řešení 4.7

U nejnižších 25 % se jedná o výsledky, které jsou odspodu ohraničeny nejnižším možným výsledkem a odshora prvním kvartilem. Nejnižší možný výsledek je roven 0. První kvartil je *nejvyšší možný* hraniční bod na vrcholu množiny „nejhorších" 250 *nebo méně* testů, což se podle tabulky 4.7 shoduje s výsledky s 16 a 17 správnými odpověďmi. Nejnižších 25 % výsledků se tedy nachází v rozmezí od 0 do 16 včetně.

Pevně stanovené intervaly

V této kapitole jsme až doposud množiny dat dělili na podmnožiny se stejným (nebo co možná nejvíce stejným) počtem prvků a pak jsme sledovali rozsah hodnot v každé podmnožině. Můžeme ale postupovat jinak: Můžeme definovat pevně stanovený rozsah hodnot nezávislých proměnných a následně si všímat počtu prvků v každém z nich.

Návrat k testu

Podívejme se opět na test, jehož výsledky jsou uvedeny v tabulce 4.1, tentokrát se ale zaměřme na rozsah výsledků. Existuje mnoho způsobů, jak to můžeme udělat, a tři z nich jsou zobrazeny v tabulkách 4.8, 4.9 a 4.10.

V tabulce 4.8 jsou výsledky testů rozvrženy podle počtu testů s výsledky v rámci čtyř následujících rozsahů: 0–10, 11–20, 21–30 a 31–40. Vidíme, že největší počet žáků dosáhl výsledku v rozsahu 21–30, který následují rozsahy 31–40, 11–20 a 0–10.

Tabulka 4.8: Výsledky hypotetického testu se 40 otázkami, kterého se účastnilo 1 000 žáků, rozděleného na čtyři stejně velké rozsahy výsledků.

Rozsah výsledků	Absolutní četnost	Procentuální podíl výsledků
0–10	145	14,5 %
11–20	215	21,5 %
21–30	340	34,0 %
31–40	300	30,0 %

Tabulka 4.9: Výsledky hypotetického testu se 40 otázkami, kterého se účastnilo 1 000 žáků, rozděleného na deset stejně velkých rozsahů výsledků.

Rozsah výsledků	Absolutní četnost	Procentuální podíl výsledků
0–4	50	5,0 %
5–8	62	6,2 %
9–12	66	6,6 %
13–16	72	7,2 %
17–20	110	11,0 %
21–24	140	14,0 %
25–28	164	13,4 %
29–32	166	16,6 %
33–36	130	13,0 %
37–40	70	7,0 %

Tabulka 4.10: Výsledky hypotetického testu se 40 otázkami, kterého se účastnilo 1 000 žáků, rozděleného na rozsahy podle subjektivních známek

Známka	Rozsah výsledků	Absolutní četnost	Procentuální podíl výsledků
F	0–18	300	30,0 %
D	19–24	200	20,0 %
C	25–31	250	25,0 %
B	32–37	208	20,8 %
A	38–40	42	4,2 %

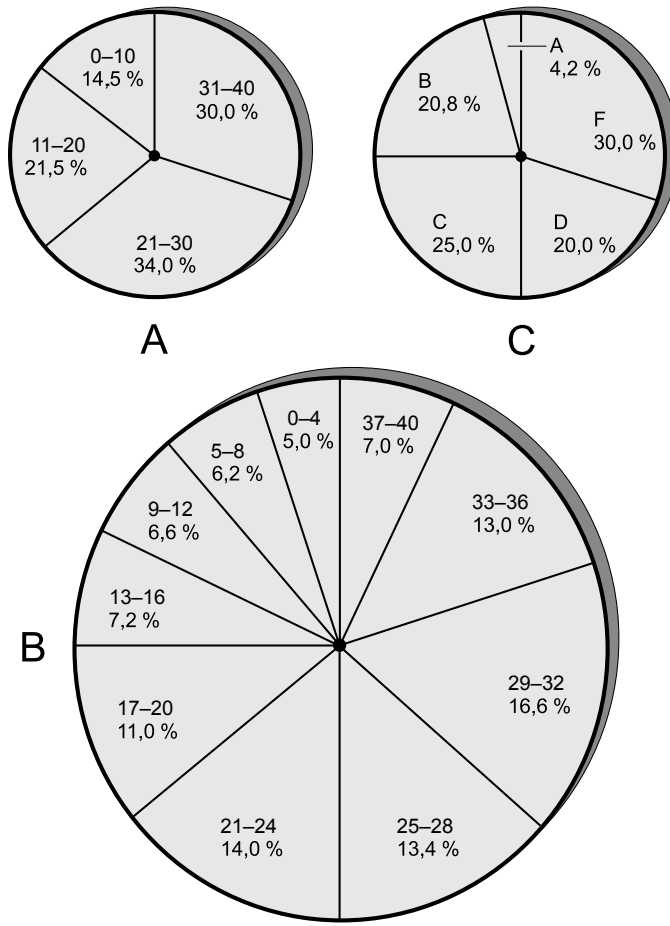
V tabulce 4.9 jsou výsledky testů znázorněny podle počtu testů s výsledky v rámci 10 různých rozsahů. V tomto případě byl „nejčastěji se vyskytující“ rozsah v rozmezí 29–32 správných odpovědí,

následující „nejčastěji se vyskytující“ rozsah odpovídal 21–24 správným odpovědím a „nejméně“ se vyskytoval rozsah 0–4.

Obě tabulky 4.8 i 4.9 dělí výsledky testů na stejně velké rozsahy (kromě nejnižšího rozsahu, který zahrnuje jeden výsledek navíc, neboli 0), ovšem tabulka 4.10 se od nich liší. Místo toho, aby obsahovala data rozdělená do stejných skupin, vyjadřuje výsledky v souladu se známkou v podobě písmen A, B, C, D a F. Přidělování známek je často subjektivní a závisí na výkonnosti třídy celkově, na obtížnosti testu a na povaze učitele. (Pomyslný učitel, který hodnotil tento test, musel být velmi přísný.)

Výsečový graf

Údaje v tabulkách 4.8, 4.9 a 4.10 můžeme zobrazit v grafické formě, použijeme-li kruhové grafy rozdělené do částí. Takové ilustrace nazýváme *výsečové grafy* nebo *výsečové diagramy*. Jsou rozděleny do částí ve formě klínů podobně jako u koláče. Čím se zvyšuje počet množiny dat, tím se přímo úměrně zvyšuje i úhel jednotlivé části.



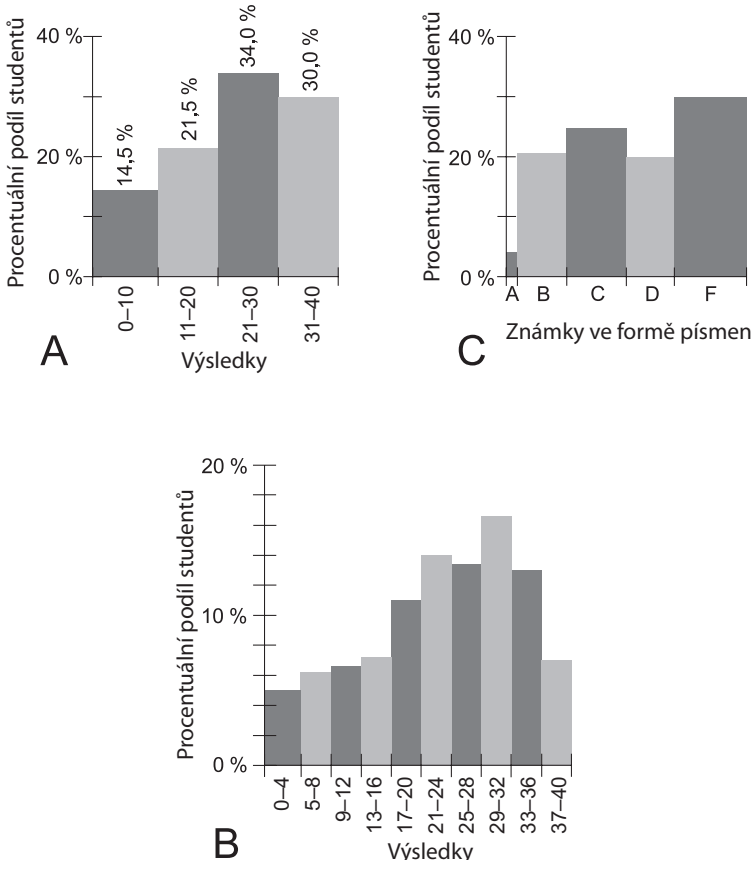
Obrázek 4.3: Ilustrace A je graf znázorňující údaje z tabulky 4.8, graf B obsahuje údaje z tabulky 4.9 a obrázek C z tabulky 4.10.

Na obrázku 4.3 znázorňuje graf A údaje vyplývající z tabulky 4.8, graf B výsledky z tabulky 4.9 a graf C výsledky z tabulky 4.10. Úhel (ve stupních) na špičce nebo vrcholu každého klínku je úměrný k procentuálnímu podílu datových prvků v podmnožině. Pokud tedy nějaký klínek vyobrazuje 10 % žáků, představuje jeho vrcholový úhel 10 % z 360° neboli 36°. Podobně platí, že znázorňuje-li nějaký klínek 25 % žáků, představuje jeho vrcholový úhel 25 % z 360°, tedy 90°. Obecně řečeno: Pokud nějaký klínek znázorňuje *x* % prvků populace, je vrcholový úhel (ve stupních) jeho klínku ve výsečovém grafu roven 3,6*x*.

Velikosti těchto klínů můžeme také vyjádřit na základě procentuální plochy. Všechny klínky mají stejný poloměr, který odpovídá poloměru kruhu. To znamená, že jejich plochy jsou úměrné k procentuálním podílům datových prvků v podmnožině, kterou znázorňují. Například na obrázku 4.3A představuje rozsah výsledků 31–40 oblast obsahující „30 % nebo 3/10 koláče", zatímco u obrázku 4.3C si můžeme všimnout, že žáci, kteří dostali známku C, představují „25 % nebo 1/4 koláče".

Histogram s proměnlivou šířkou

Histogramy jsme probírali v 1. kapitole a příklad, který tam byl uveden, bychom mohli označit za přílišně zjednodušený, protože se jednalo o *histogram s pevnou šířkou*. Existuje ale i přizpůsobivější druh, který nazýváme *histogram s proměnlivou šířkou*.



Obrázek 4.4: Ilustrace A je histogram s údaji z tabulky 4.8, histogram B obsahuje údaje z tabulky 4.9 a histogram C z tabulky 4.10

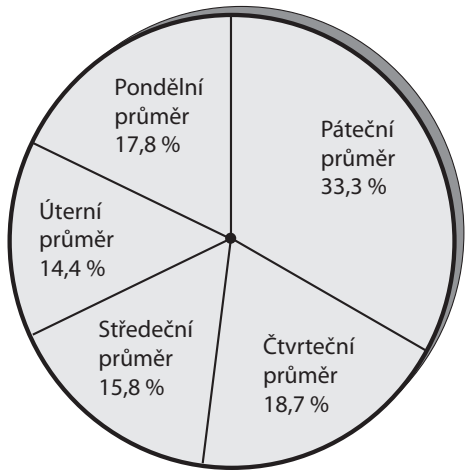
Tento typ grafu je ideální pro znázornění výsledků našeho hypotetického testu se 40 otázkami, kterého se různými způsoby účastnilo 1 000 žáků.

Obrázek 4.4 je histogram s proměnlivou šířkou, který vyjadřuje ty samé údaje jako v tabulkách a výsečových grafech. Graf A vyobrazuje výsledky z tabulky 4.8, graf B z tabulky 4.9 a graf C z tabulky 4.10. Šířka každého svislého sloupce je přímo úměrná k rozsahům výsledků a jejich výška je přímo úměrná k procentům žáků, kteří dosáhli výsledku v uvedeném rozsahu.

Na obrázku 4.4A jsou uvedeny údaje v procentech, protože tam pro ně bylo dostatek místa, aniž by graf vypadal chaoticky nebo přeplněně. Histogramy 4.4B a C údaje v procentech neobsahují. Jedná se pouze o otázku upřednostnění, protože v tomto případě by graf B pro některé lidi mohl vypadat přeplněně a v grafu C by se naskytl problém, kdybychom chtěli uvést procentuální hodnotu známky A, což by vypadalo velmi chaoticky. Neobsahují-li histogramy procentuální údaje nad sloupci, je velmi výhodné doplnit je o tabulky obsahující příslušné údaje.

Problém 4.8

Představte si hodně velkou společnost, která pracuje na bázi pětidenního pracovního týdne (od pondělí do pátku). Dejme tomu, že po dlouhé časové období zjišťujeme průměrný počet zaměstnanců, kteří se na každý den v průběhu jednoho týdne telefonicky omluvili kvůli nemoci, a zároveň průměrný počet jednotlivých dnů v týdnu, kdy nějaký zaměstnanec zůstal doma nemocen. Průměrný počet lidí, kteří se v určitý den ohlásí jako nemocní, vydělíme pro každý z pěti dnů jednoho pracovního týdne průměrným počtem jednotlivých dnů tohoto týdne, kdy někdo zůstal doma z důvodu nemoci, a výsledné údaje zapíšeme do tabulky jako procento pro daný pracovní den. Výsledky jsou vyobrazeny jako výsečový graf na obrázku 4.5. Pojmenujte dvě skutečnosti, které tento obrázek vypovídá o pátcích. Dále pojmenujte jednu skutečnost, u které se zpočátku zdálo, že bude vypovídat něco o pátcích, ale nakonec o nich vlastně nic nevypovídá.



Obrázek 4.5: Ilustrace k problémům 4.8 až 4.9

Řešení 4.8

Výsečový graf vyjadřuje, že se průměrně kvůli nemoci omluví více lidí v pátek než v jakýkoli jiný den pracovního týdne a že z celkového počtu dnů, kdy někdo zůstane nemocen doma, připadá

v průměru 33 % na pátky. Na první pohled se také mohlo zdát, ale to ve skutečnosti není pravda, že se průměrně 33 % zaměstnanců společnosti omluví pro nemoc právě v pátek.

Problém 4.9

Dejme tomu, že ve výše popsané společnosti a během pozorovaného období zobrazeného ve výsečovém grafu 4.5 se průměrně vyskytne 1 000 dnů, kdy nějaký zaměstnanec zůstane doma kvůli nemoci. Jaký průměrný počet takových dnů připadá na pondělí? Jaký je průměrný počet lidí, kteří se omluví kvůli nemoci právě v pondělí?

Řešení 4.9

Převedeme-li situaci na jeden jediný den, pak den, kdy nějaký zaměstnanec zůstane doma, odpovídá jedné osobě, která se telefonicky omluví pro nemoc. To ovšem nemusí nutně platit pro období delší než jeden den. V tomto případě, kdy se jedná pouze o pondělky, můžeme 1 000 vynásobit 17,8 %, čímž dostaneme 178 a odpověď na obě otázky. Průměrně připadá na pondělky 178 dnů, kdy někdo zůstane doma, a průměrně se v pondělí omluví 178 zaměstnanců.

Problém 4.10

Zůstaneme-li ještě u scénáře popsaného v předchozích dvou problémech, jaký je průměrný počet dnů, kdy nějaký zaměstnanec zůstane doma nemocen, připadající na pondělí a úterý? Jaký je průměrný počet osob, které se v pondělí a v úterý telefonicky omluví pro nemoc?

Řešení 4.10

Při řešení minulého problému jsme vypočítali, že na pondělky připadá průměrně 178 dnů, kdy někdo zůstane doma. Abychom našli průměrný počet těchto dnů i pro úterky, musíme 1 000 vynásobit 14,4 %, čímž dostaneme 144. Průměrný počet takových dnů připadajících jak na pondělky, tak i na úterky je tedy 178 + 144, neboli 322. Není možné vypočítat průměrný počet osob, které se v pondělí i v úterý telefonicky omluví, protože nevíme, kolik pondělků a úterků, kdy někdo zůstal doma, odpovídá jedné jediné nemocné osobě po oba dva dny (dva dny, kdy někdo zůstal doma, ale pouze jedna nemocná osoba).

Další upřesnění

Vlastnosti dat můžeme popsat i pomocí doplňkových deskriptivních ukazatelů, z nichž jsou některé blíže popsány v následující části.

Rozpětí

Pro množinu dat nebo pro jakýkoli sousedící interval v této množině můžeme pojem *rozpětí* definovat jako rozdíl mezi nejmenší a nejvyšší hodnotou této množiny nebo tohoto intervalu.

V grafu znázorňujícím hypotetický systolický krevní tlak (obrázek 4.1) se nejnižší hodnota systolického tlaku v množině dat rovnala 60 a nejvyšší hodnota 160. Rozpětí mezi těmito hodnotami je tedy 100. Je samozřejmě možné, že někteří lidé měli tlak nižší než 60 nebo popřípadě vyšší než 160, ale jejich hodnoty byly z této množiny dat ve skutečnosti vyškrtnuty.

V případě testu se 40 otázkami, kterému jsme se v této kapitole už tak často věnovali, byl nejnižší výsledek 0 a nejvyšší 40 správných odpovědí, proto je rozpětí rovno 40. Možná, že bychom se

chtěli soustředit pouze na rozpětí určité části všech výsledků, například na druhých nejnižších 25 % z nich. Zjistíme ho v tabulce 4.4: Odpovídají mu výsledky s 24–17 správnými odpověďmi, což znamená, že se rozpětí rovná 7.

Variační koeficient

Vzpomínáte si ještě na definici průměru (μ) a směrodatné odchylky (σ) z 2. kapitoly? Nyní si je stručně zopakujeme, protože z nich můžeme odvodit jedno důležité upřesnění.

U normálního rozložení, jako například u toho, které znázorňuje výsledky našeho hypotetického testu s krevním tlakem, je průměr taková hodnota (v tomto případě krevního tlaku), že plocha pod křivkou je po obou stranách svislé čáry, která průměr představuje, naprosto stejná.

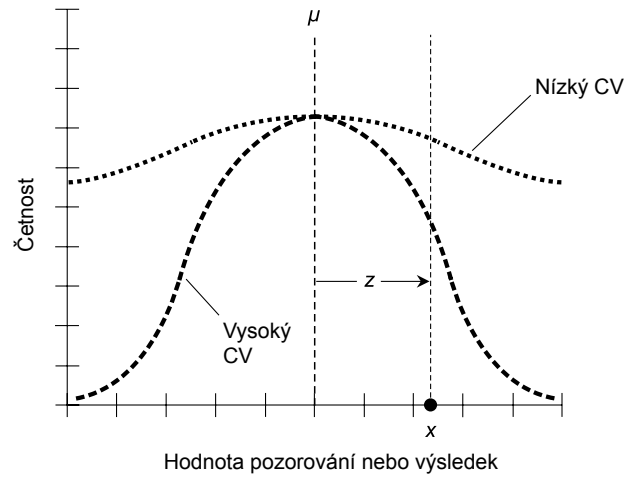
Jsou-li údaje spojitých prvků uvedeny v tabulce, pak se průměr rovná aritmetickému průměru těchto výsledků. Máme-li výsledky $\{x_1, x_2, x_3, \ldots, x_n\}$, jejichž průměrem je μ, pak je směrodatná odchylka:

$$\sigma = \sqrt{\{(1/n)[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_n - \mu)^2]\}}$$

Průměr je míra střední polohy nebo také „vycentrovanost". Směrodatná odchylka je míra rozptylu nebo také „rozptýlení po prostoru". Dejme tomu, že chceme vědět, jak jsou údaje vzhledem k průměru rozptýlené. To zjistíme, když směrodatnou odchylku vydělíme průměrem. Vyjde nám hodnota, kterou nazýváme *variační koeficient* a kterou značíme velkými písmeny CV. Matematicky CV odpovídá:

$$\mathrm{CV} = \sigma/\mu$$

Směrodatná odchylka i průměr jsou vyjádřeny ve stejných jednotkách, například jako systolický krevní tlak nebo výsledky testu. Vydělíme-li tyto jednotky samy sebou, vzájemně se vykrátí, takže CV nemá žádné jednotky. Takové číslo bez specifické jednotky je známé jako *bezrozměrná veličina*.



Obrázek 4.6: Dvě rozložení znázorněná v grafické formě, jedno s nízkým variačním koeficientem (CV) a jedno s vyšším CV. Odvození Z-skóre (z) pro výsledky x je vysvětleno v textu.

Protože je variační koeficient bezrozměrný, můžeme ho použít za účelem porovnání „rozptýlení po prostoru" množin dat, které popisují vysoce rozdílné věci, jako například krevní tlak nebo výsledky nějakého testu. Vysoký variační koeficient znamená, že jsou data kolem průměru rozptýlená pouze relativně, malý koeficient znamená, že se soustředí v jeho blízkosti. V krajním případě, je-li CV = 0, jsou všechny hodnoty dat stejné a nacházejí se přesně na průměru. Obrázek 4.6 představuje dvě rozložení v grafické formě, jedno s docela malým variačním koeficientem a jedno s vyšším.

Věřili byste tomu, že výše uvedený problém skrývá jeden potenciální problém? Pokud se divíte tomu, co se stane v rozložení, kde mohou údaje nabýt buď kladné nebo záporné hodnoty – například u teplot ve stupních Celsia – pak je vaše starost oprávněná. Je-li $\mu = 0$ (bod tuhnutí vody na Celsiově stupnici), pak máme problém. Můžeme tomu zabránit tím, že změníme jednotky, ve kterých jsou údaje vyčísleny, tak, že se 0 nebude vyskytovat uvnitř množiny možných hodnot. Pokud například vyjadřujeme teploty, můžeme raději použít Kelvinovu nežli Celsiovu stupnici, kde se všechny teplotní údaje nacházejí nad 0.

V situaci, kde se všechny prvky nějaké množiny dat rovnají 0, což by například nastalo, pokud by celá třída žáků odevzdala nevyplněné testy, by variační koeficient nebyl určený, protože se průměr ve skutečnosti rovná 0.

Z-skóre

Někdy možná uslyšíte některé lidi říkat, že se takové a takové pozorování nebo takový a takový výsledek nachází „2,2 směrodatné odchylky pod průměrem" nebo „1,6 směrodatné odchylky nad průměrem". *Z-skóre*, které zapisujeme malým písmenem *z*, je kvantitativní určení pozice určitého prvku s ohledem na průměr. Z-skóre nějakého prvku je rovno počtu směrodatných odchylek vyjadřujících, jak se tento prvek liší od průměru, ať už kladně nebo záporně.

Pro určitý prvek x v množině dat závisí hodnota z jak na průměru (μ), tak i na směrodatné odchylce (σ) a můžeme ji vypočítat pomocí následujícího vzorce:

$$z = (x - \mu)/\sigma$$

Nachází-li se x pod průměrem, pak je z záporné číslo. Je-li x nad průměrem, pak je z kladné. Rovná-li se x průměru, pak $z = 0$.

U grafických rozložení na obrázku 4.6 je $z > 0$ pro znázorněný bod x. To platí pro obě křivky. Jen na základě grafu nemůžeme říct, jaké je Z-skóre pro x s ohledem na jednu z křivek, ale alespoň vidíme, že je v obou případech kladný.

Mezikvartilové rozpětí

Někdy může být užitečné, známe-li „střední polovinu" dat v množině – neboli *mezikvartilové rozpětí*, které zkracujeme na IQR. IQR se rovná hodnotě 3. bodu kvartilu, od něhož odečteme hodnotu prvního bodu kvartilu. Pokud se bod kvartilu vyskytne mezi dvěma celými čísly, můžeme ho považovat za průměrnou hodnotu těchto dvou čísel (menší číslo plus 0,5).

Vzpomeňte si opět na test se 40 otázkami, který dělalo 1 000 žáků a jehož body kvartilu jsou znázorněny v obrázku 4.2A. První kvartil se nachází mezi výsledky 16 a 17, třetí mezi výsledky 31 a 32. Proto:

$$\text{IQR} = 31 - 16$$

$$= 15$$

Problém 4.11

Dejme tomu, že žákům zadáme rozdílný test se 40 otázkami a že výsledky jsou nyní mnohem více soustředěny nežli u testu, který je znázorněn na obrázku 4.2A. Jaké by bylo IQR tohoto testu ve srovnání s IQR předchozí zkoušky?

Řešení 4.11

IQR by bylo menší, protože první a třetí kvartil by si byly vzájemně blíže.

Problém 4.12

Vzpomeňte si na empirické pravidlo, které jsme probírali v předcházející kapitole a které říká, že všechna normální rozložení mají následující tři vlastnosti:

- Přibližně 68 % všech hodnot se nachází v rozmezí ±1 směrodatné odchylky σ od průměru μ.
- Přibližně 95 % všech hodnot se nachází v rozmezí ±2 směrodatných odchylek σ od průměru μ.
- Přibližně 99,7 % všech hodnot se nachází v rozmezí ±3 směrodatných odchylek σ od průměru μ.

Přeformulujte toto pravidlo na základě Z-skóre.

Řešení 4.12

Jak jsme uvedli výše, Z-skóre nějakého prvku je počet směrodatných odchylek vyjadřujících, jak moc se tento prvek odchyluje od průměru, ať už kladně nebo záporně. Všechna normální rozložení mají následující tři vlastnosti:

- Přibližně 68 % všech hodnot má Z-skóre mezi -1 a +1.
- Přibližně 95 % všech hodnot se nachází v rozmezí -2 a +2.
- Přibližně 99,7 % všech hodnot se nachází v rozmezí -3 a +3.

Test

Pokud to bude nutné, vraťte se k jednotlivým odstavcům této kapitoly. Dobrý výsledek je 8 správných odpovědí. Odpovědi najdete na konci knihy.

1. Dejme tomu, že se nějakého testu zúčastní velký počet lidí. Třetí bod decilu určíme tak, že:
   (a) najdeme nejvyšší výsledek představující „nejhorších“ 20 % nebo méně testů, přičemž třetí decil je na vrcholu této množiny,
   (b) najdeme nejvyšší výsledek představující „nejhorších“ 30 % nebo méně testů, přičemž třetí decil je na vrcholu této množiny,
   (c) najdeme nejnižší výsledek představující „nejlepších“ 20 % nebo méně testů, přičemž třetí decil je na dně této množiny,
   (d) najdeme nejnižší výsledek představující „nejlepších“ 30 % nebo méně testů, přičemž třetí decil je na dně této množiny.

2. Představte si, že se velký počet žáků podrobí testu s 10 otázkami a že průměrný výsledek bude 7,22 správných odpovědí a směrodatná odchylka se bude rovnat 0,722. Jakou hodnotu bude mít variační koeficient?
   (a) Tuto otázku není možné zodpovědět, pokud neznáme přesný počet žáků, kteří test píší.
   (b) 0,1
   (c) 10
   (d) 100

3. Test s 10 otázkami píše několik žáků. Jejich nejhorší výsledek jsou 3 správné odpovědi, jejich nejlepší výsledek je 10 správných odpovědí. Jaký je rozsah výsledků?
   (a) Tuto otázku není možné zodpovědět, pokud neznáme přesný počet žáků, kteří test píší.
   (b) 3/7
   (c) 7
   (d) 7/3

4. Několik žáků píše test s 10 otázkami, přičemž jejich nejhorší výsledek jsou 3 správné odpovědi a jejich nejlepší výsledek je 10 správných odpovědí. Jakou hodnotu má 50. percentil?
   (a) 7, což je rovno 10 – 3.
   (b) 6,5, což se rovná (3 + 10)/2.
   (c) 301/2, což se rovná druhé odmocnině ze (3 × 10).
   (d) Tuto otázku není možné zodpovědět, pokud nevíme, kolik žáků dosáhlo každého výsledku.

5. Představte si, že jste psali normovaný test a že vám na konci sdělili, že patříte do 91. percentilu. To znamená, že:
   (a) 90 žáků, kteří se podrobili tomu samému testu, dosáhlo vyšších výsledků než vy,
   (b) 90 % všech žáků, kteří se podrobili tomu samému testu, dosáhlo vyšších výsledků než vy,
   (c) 90 žáků, kteří se podrobili tomu samému testu, dosáhlo nižších výsledků než vy.
   (d) Nic to neznamená, protože nic jako 91. percentil neexistuje.

6. Tabulka 4.11 ukazuje výsledky hypotetického testu s 10 otázkami, který dostala skupina žáků. Kde se nachází první bod percentilu?
   (a) Mezi výsledky 1 a 2.
   (b) Mezi výsledky 4 a 5.
   (c) Mezi výsledky 6 a 7.
   (d) Není možné to určit, nebudeme-li mít více informací.

7. Jaký je rozsah výsledků, kterých dosáhli žáci ve scénáři v tabulce 4.11?
   (a) 5
   (b) 8
   (c) 10
   (d) Není možné to určit, protože to není jednoznačné.

Tabulka 4.11: Ilustrace ke kvizovým otázkám 6 až 9. Výsledky hypotetického testu s 10 otázkami, kterého se účastnilo 128 žáků

Výsledek testu	Absolutní četnost	Kumulativní absolutní četnost
0	0	0
1	0	0
2	3	3
3	8	11
4	12	23
5	15	38
6	20	58
7	24	82
8	21	103
9	14	117
10	11	128

8. Jaké je mezikvartilové rozpětí výsledků v tabulce 4.11?
   (a) 3
   (b) 4
   (c) 6
   (d) Není možné to určit, protože to není jednoznačné.

9. Ve scénáři znázorněném v tabulce 4.11 je jedním z žáků, kteří test psali, i Jan N., který dosáhl 6 správných odpovědí. V jakém intervalu se Jan nachází s ohledem na celou třídu?
   (a) Ve spodních 25 %.
   (b) V druhých nejnižších 25 %.
   (c) V druhých nejvyšších 25 %.
   (d) Není možné to určit, protože to není jednoznačné.

10. Dejme tomu, že se do vašich rukou dostane kruhový graf zobrazující výsledky nějakého průzkumu, jehož cílem je zjistit počet a poměr rodin v určitém městě, které vydělávají peníze v rozdílných rozsazích. Skutečné údaje z nějakého důvodu v tomto grafu nejsou uvedeny. Vy ale vidíte, že jedno z rozpětí se nachází v části s vrcholovým úhlem 90°. Na základě toho můžete předpokládat, že tato část odpovídá:
    (a) rodinám, jejichž výdělky spadají do středních 25 % vzhledem k příjmům,
    (b) 25 % sledovaných rodin,
    (c) $1/\pi$ sledovaných rodin (π je podíl obvodu kruhu a jeho průměru),
    (d) rodinám, jejichž výdělky spadají do mezikvartilového rozpětí.