Pravděpodobnost a statistika – AR 2020/21

Katedra kvantitativních metod FVL UO Brno

Základní statistické pojmy, Popisná statistika

Marek Sedlačík

7. března 2024

Obsah

Pojem a úkoly statistiky

Statistika je věda, která se zabývá získáváním, zpracováním a analýzou dat pro potřeby rozhodování. Zkoumá stav a vývoj *hromadných jevů* a vztahů mezi nimi prostřednictvím *hromadných pozorování*.

Hromadná pozorování si představíme jako měření nebo zjišťování, kdy

- jev se může mnohokrát opakovat $\longrightarrow$ *opakované pokusy*
- jev pozorujeme na vybraném počtu objektů (jednotek) $\longrightarrow$ *výběry*

Pojem a úkoly statistiky

Statistika je věda, která se zabývá získáváním, zpracováním a analýzou dat pro potřeby rozhodování. Zkoumá stav a vývoj *hromadných jevů* a vztahů mezi nimi prostřednictvím *hromadných pozorování*.

Hromadná pozorování si představíme jako měření nebo zjišťování, kdy

- jev se může mnohokrát opakovat $\longrightarrow$ *opakované pokusy*
- jev pozorujeme na vybraném počtu objektů (jednotek) $\longrightarrow$ *výběry*

Pojem a úkoly statistiky

Statistika je věda, která se zabývá získáváním, zpracováním a analýzou dat pro potřeby rozhodování. Zkoumá stav a vývoj *hromadných jevů* a vztahů mezi nimi prostřednictvím *hromadných pozorování*.

Hromadná pozorování si představíme jako měření nebo zjišťování, kdy

- jev se může mnohokrát opakovat $\longrightarrow$ *opakované pokusy*
- jev pozorujeme na vybraném počtu objektů (jednotek) $\longrightarrow$ *výběry*

Pojem a úkoly statistiky

Etapy statistické práce:

1. statistické měření a zjišťování,
2. zpracování statistických údajů,
3. interpretace získaných výsledků.

Pojem a úkoly statistiky

Etapy statistické práce:

1. statistické měření a zjišťování,
2. zpracování statistických údajů,
3. interpretace získaných výsledků.

Pojem a úkoly statistiky

Etapy statistické práce:

1. statistické měření a zjišťování,
2. zpracování statistických údajů,
3. interpretace získaných výsledků.

Pojem a úkoly statistiky

Praktické užití statistiky se opírá o její 2 roviny:

- **popisnou statistiku, výběrový soubor** – zpracování naměřených dat a získání informací o těchto datech (zejména zobrazení dat pomocí tabulek, grafů a výpočet číselných charakteristik),
- **induktivní statistiku, základní soubor** – souhrn metod sloužících k odhadům sledovaných vlastností v základních souborech $\longrightarrow$ induktivní úvahy s využitím pravděpodobnosti, tedy zobecňování získaných informací z výběru na celý soubor, ze kterého byl výběr pořízen.

Pojem a úkoly statistiky

Praktické užití statistiky se opírá o její 2 roviny:

- **popisnou statistiku, výběrový soubor** – zpracování naměřených dat a získání informací o těchto datech (zejména zobrazení dat pomocí tabulek, grafů a výpočet číselných charakteristik),
- **induktivní statistiku, základní soubor** – souhrn metod sloužících k odhadům sledovaných vlastností v základních souborech $\longrightarrow$ induktivní úvahy s využitím pravděpodobnosti, tedy zobecňování získaných informací z výběru na celý soubor, ze kterého byl výběr pořízen.

Základní pojmy a prostředky

Definice

- **statistický soubor** – *množina zkoumaných objektů, které mají z daného hlediska společné vlastnosti (osoby, věci, rostliny, zvířata, podniky, události, . . .)*
- **statistická jednotka** – *prvek statistického souboru (1 člověk, 1 výrobek, 1 pokus, . . .)*

Základní pojmy a prostředky

Definice

- **statistický soubor** – *množina zkoumaných objektů, které mají z daného hlediska společné vlastnosti (osoby, věci, rostliny, zvířata, podniky, události, . . .)*
- **statistická jednotka** – *prvek statistického souboru (1 člověk, 1 výrobek, 1 pokus, . . .)*

Základní pojmy a prostředky

Definice

- **základní soubor** – *soubor, který je předmětem našeho zájmu, je předmětem statistického šetření a o jehož vlastnostech se mají dělat závěry (někdy se označuje jako populace)*

∼ **reálný** – všechny jednotky reálně existují (studenti VŠ, Felicie vyrobené v roce 1999, denní produkce rohlíků u pekaře, ...) ⟶ konečný

∼ **hypotetický** – obecně definován, ale při statistickém šetření statistcké jednotky reálně buď neexistují vůbec (výsledky laboratorních měření na 1 vzorku, sportovní výkony 1 sportovce, ...) nebo existují jen zčásti (pokračující výroba, přicházející zákazníci, ...) ⟶ nekonečný

Základní pojmy a prostředky

Definice

- **základní soubor** – *soubor, který je předmětem našeho zájmu, je předmětem statistického šetření a o jehož vlastnostech se mají dělat závěry (někdy se označuje jako populace)*

∼ **reálný** – všechny jednotky reálně existují (studenti VŠ, Felicie vyrobené v roce 1999, denní produkce rohlíků u pekaře, . . .) $\longrightarrow$ konečný

∼ **hypotetický** – obecně definován, ale při statistickém šetření statistcké jednotky reálně buď neexistují vůbec (výsledky laboratorních měření na 1 vzorku, sportovní výkony 1 sportovce, . . .) nebo existují jen zčásti (pokračující výroba, přicházející zákazníci, . . .) $\longrightarrow$ nekonečný

Základní pojmy a prostředky

Definice

- **základní soubor** – *soubor, který je předmětem našeho zájmu, je předmětem statistického šetření a o jehož vlastnostech se mají dělat závěry (někdy se označuje jako populace)*

$\sim$ **reálný** – všechny jednotky reálně existují (studenti VŠ, Felicie vyrobené v roce 1999, denní produkce rohlíků u pekaře, ...) $\longrightarrow$ konečný

$\sim$ **hypotetický** – obecně definován, ale při statistickém šetření statistcké jednotky reálně buď neexistují vůbec (výsledky laboratorních měření na 1 vzorku, sportovní výkony 1 sportovce, ...) nebo existují jen zčásti (pokračující výroba, přicházející zákazníci, ...) $\longrightarrow$ nekonečný

Základní pojmy a prostředky

Definice

- **výběrový soubor** – *podmnožina základního souboru vytvořená na základě tzv. výběrového (reprezentativního) šetření*

∼ **záměrný výběr** – výběr na základě známých vlastností základního souboru: jednotky vybíráme tak, aby výběrový soubor byl dobrým reprezentantem základního souboru

∼ **náhodný (pravděpodobnostní) výběr** – výběr na základě předem určené pravděpodobnosti zahrnutí jednotek do výběrového souboru, tedy vlastní výběr záleží na náhodě

Existují různé typy náhodných výběrů, nejdůležitější v našem kurzu je tzv. *prostý náhodný výběr*, tj. přímý výběr ze základního souboru, kde každá jednotka má stejnou pst výběru.

Základní pojmy a prostředky

Definice

- **výběrový soubor** – *podmnožina základního souboru vytvořená na základě tzv. výběrového (reprezentativního) šetření*

- ∼ **záměrný výběr** – výběr na základě známých vlastností základního souboru: jednotky vybíráme tak, aby výběrový soubor byl dobrým reprezentantem základního souboru
- ∼ **náhodný (pravděpodobnostní) výběr** – výběr na základě předem určené pravděpodobnosti zahrnutí jednotek do výběrového souboru, tedy vlastní výběr záleží na náhodě

Existují různé typy náhodných výběrů, nejdůležitější v našem kurzu je tzv. *prostý náhodný výběr*, tj. přímý výběr ze základního souboru, kde každá jednotka má stejnou pst výběru.

Základní pojmy a prostředky

Definice

- **výběrový soubor** – *podmnožina základního souboru vytvořená na základě tzv. výběrového (reprezentativního) šetření*

∼ **záměrný výběr** – výběr na základě známých vlastností základního souboru: jednotky vybíráme tak, aby výběrový soubor byl dobrým reprezentantem základního souboru

∼ **náhodný (pravděpodobnostní) výběr** – výběr na základě předem určené pravděpodobnosti zahrnutí jednotek do výběrového souboru, tedy vlastní výběr záleží na náhodě

Existují různé typy náhodných výběrů, nejdůležitější v našem kurzu je tzv. *prostý náhodný výběr*, tj. přímý výběr ze základního souboru, kde každá jednotka má stejnou pst výběru.

Základní pojmy a prostředky

Definice

- **rozsah výběrového souboru** – *počet jednotek tvořících vývěrový soubor; ozn.* n
- **statistický znak** – *vlastnost jednotek, která je předmětem našeho zájmu nebo na základě které byl vytvořen (definován) základní soubor (hmotnost rohlíku, rychlost auta, počet zákazníků, znalost cizího jazyka, pohlaví, známka u zkoušky ze Stat, . . .); ozn.* X
- **hodnota znaku** – *výsledek 1 zjištění - měření na 1 jednotce* $(X = x_i) \longrightarrow$ *zjištěné (naměřené) hodnoty představují tzv. data:* $x_1, x_2, \ldots, x_n$
- **obměny (varianty) znaku** – *různé hodnoty znaku v 1 souboru*

Základní pojmy a prostředky

Definice

- **rozsah výběrového souboru** – *počet jednotek tvořících vývěrový soubor; ozn.* n
- **statistický znak** – *vlastnost jednotek, která je předmětem našeho zájmu nebo na základě které byl vytvořen (definován) základní soubor (hmotnost rohlíku, rychlost auta, počet zákazníků, znalost cizího jazyka, pohlaví, známka u zkoušky ze Stat, ...); ozn.* X
- **hodnota znaku** – *výsledek 1 zjištění - měření na 1 jednotce* $(X = x_i) \longrightarrow$ *zjištěné (naměřené) hodnoty představují tzv. data:* $x_1, x_2, \ldots, x_n$
- **obměny (varianty) znaku** – *různé hodnoty znaku v 1 souboru*

Základní pojmy a prostředky

Definice

- **rozsah výběrového souboru** – *počet jednotek tvořících vývěrový soubor; ozn.* n
- **statistický znak** – *vlastnost jednotek, která je předmětem našeho zájmu nebo na základě které byl vytvořen (definován) základní soubor (hmotnost rohlíku, rychlost auta, počet zákazníků, znalost cizího jazyka, pohlaví, známka u zkoušky ze Stat, . . .); ozn.* X
- **hodnota znaku** – *výsledek 1 zjištění - měření na 1 jednotce* $(X = x_i) \longrightarrow$ *zjištěné (naměřené) hodnoty představují tzv. data:* $x_1, x_2, \ldots, x_n$
- **obměny (varianty) znaku** – *různé hodnoty znaku v 1 souboru*

Základní pojmy a prostředky

Definice

- **rozsah výběrového souboru** – *počet jednotek tvořících vývěrový soubor; ozn.* n
- **statistický znak** – *vlastnost jednotek, která je předmětem našeho zájmu nebo na základě které byl vytvořen (definován) základní soubor (hmotnost rohlíku, rychlost auta, počet zákazníků, znalost cizího jazyka, pohlaví, známka u zkoušky ze Stat, . . .); ozn.* X
- **hodnota znaku** – *výsledek 1 zjištění - měření na 1 jednotce* $(X = x_i) \longrightarrow$ *zjištěné (naměřené) hodnoty představují tzv. data:* $x_1, x_2, \ldots, x_n$
- **obměny (varianty) znaku** – *různé hodnoty znaku v 1 souboru*

Základní pojmy a prostředky

Klasifikace statistických znaků podle způsobu vyjádření hodnot:

- *číselné – měřitelné* (spojité, nespojité),
- *slovní – kategoriální*.

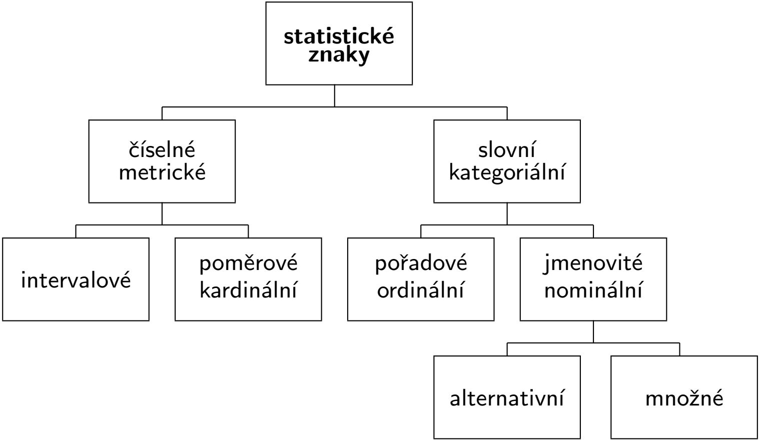
Podle typu vztahů mezi hodnotami a obměnami:

- *metrické – měřitelné* (kardinální, poměrové, intervalové),
- *ordinální – pořadové*,
- *nominální – jmenovité*.

Hledisko počtu obměn (pouze u slovní proměnné):

- *alternativní*,
- *množné*.

Základní pojmy a prostředky



Vyjadřovací prostředky statistiky

- tabulky $\longrightarrow$ tabulka rozdělení četností, korelační tabulka, různé typy výpočetních tabulek, ...
- grafy $\longrightarrow$ polygon četností, histogram, bodový graf, výsečový graf, krabicový graf, ...

Vyjadřovací prostředky statistiky

- tabulky $\longrightarrow$ tabulka rozdělení četností, korelační tabulka, různé typy výpočetních tabulek, ...
- grafy $\longrightarrow$ polygon četností, histogram, bodový graf, výsečový graf, krabicový graf, ...

Základní zpracování dat

Jedná se o práci s naměřenými daty, která směřuje k tomu **poznat nejdůležitější vlastnosti sledovaného znaku** prostřednictvím jednoduchých tabulek, grafů a numerických výpočtů.

Rozlišujeme zpracování dat

a) **ruční** $\longrightarrow$ provádí se na základě vzorců, zpravidla s využitím kalkulačky se statistickým režimem (SD-1, SD-2, STAT, REG, ...)

b) **počítačové** $\longrightarrow$ provádí se s využitím dostupného softwaru, např. STAT1, Matlab, Statistica, R, SAS, SPSS, Unistat, Statgraphics, QCExpert/Adstat, jednoduché procedury obsahuje také Excel

Základní zpracování dat

Jedná se o práci s naměřenými daty, která směřuje k tomu **poznat nejdůležitější vlastnosti sledovaného znaku** prostřednictvím jednoduchých tabulek, grafů a numerických výpočtů.

Rozlišujeme zpracování dat

a) **ruční** $\longrightarrow$ provádí se na základě vzorců, zpravidla s využitím kalkulačky se statistickým režimem (SD-1, SD-2, STAT, REG, ...)

b) **počítačové** $\longrightarrow$ provádí se s využitím dostupného softwaru, např. STAT1, Matlab, Statistica, R, SAS, SPSS, Unistat, Statgraphics, QCExpert/Adstat, jednoduché procedury obsahuje také Excel

Základní zpracování dat

Jedná se o práci s naměřenými daty, která směřuje k tomu **poznat nejdůležitější vlastnosti sledovaného znaku** prostřednictvím jednoduchých tabulek, grafů a numerických výpočtů.

Rozlišujeme zpracování dat

a) **ruční** $\longrightarrow$ provádí se na základě vzorců, zpravidla s využitím kalkulačky se statistickým režimem (SD-1, SD-2, STAT, REG, . . .)

b) **počítačové** $\longrightarrow$ provádí se s využitím dostupného softwaru, např. STAT1, Matlab, Statistica, R, SAS, SPSS, Unistat, Statgraphics, QCExpert/Adstat, jednoduché procedury obsahuje také Excel

Základní zpracování dat

Podle počtu a zejména charakteru měřených dat použijeme jednu ze 3 možností zpracování dat:

1. neroztříděná data
2. bodové rozdělení četností
3. intervalové rozdělení četností

Neroztříděná data

Neroztříděná data $\longrightarrow$ vhodné pro malý rozsah souboru ($n < 30$)

- uspořádání dat podle velikosti: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$
- grafické zobrazení dat – diagram rozptýlení
- výpočet charakteristik

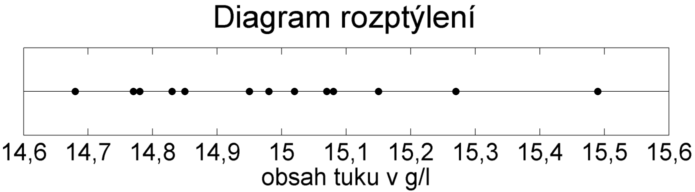
Neroztříděná data

Na 15 vzorcích mléka byl naměřen obsah tuku s těmito výsledky (v g/l):

14,85	14,68	15,27	14,77	14,83	14,95	15,08	15,02
15,07	14,98	15,15	15,49	14,83	14,95	14,78	

Sestavte diagram rozptýlení.

Neroztříděná data



Bodové rozdělení četností

Mějme uspořádaný datový soubor o rozsahu n prvků.

- *Absolutní četnost* n_j představuje počet výskytů varianty x_j v souboru. Pro absolutní četnosti platí $\sum_{j=1}^{k} n_j = n$, kde k je počet variant.
- *Relativní četnost* p_j je dána vztahem

$$p_j = \frac{n_j}{n}$$

  a představuje podíl výskytů varianty x_j v souboru. Pro relativní četnosti platí $\sum_{j=1}^{k} p_j = 1$.

Bodové rozdělení četností

- *Absolutní kumulativní četnost* N_j je dána vztahem

$$N_j = n_1 + \cdots + n_j$$

  a udává součet četností všech pozorování, která nepřekračují hodnotu x_j.
- *Relativní kumulativní četnost* F_j je určena vztahem

$$F_j = \frac{N_j}{n} = p_1 + \cdots + p_j$$

  a udává podíl četností všech pozorování, která nepřekračují hodnotu x_j.

Bodové rozdělení četností

Bodové rozdělení četností $\longrightarrow$ vhodné pro velký rozsah souboru, nespojitý znak a malý počet obměn ($k < 20$)

- tabulkové vyjádření rozdělení četností
  – $n_i,\ p_i,\ N_i,\ F_i,\ i = 1,\ 2,\ \ldots,\ k$, kde k udává počet obměn
- grafické zobrazení rozdělení četností
  – polygon četností, součtová křivka
- výpočet charakteristik

Bodové rozdělení četností

V rámci antropometrického průzkumu bylo podle metodiky lékařské komory provedeno měření tělesné výšky u 15měsíčních dětí. U 50 vybraných chlapců byly naměřeny tyto hodnoty (v cm):

83	85	81	82	84	82	79	84	80	81
82	82	80	82	80	82	83	84	82	79
83	82	83	82	82	82	81	80	82	82
83	80	82	85	81	83	81	81	83	82
81	85	83	79	81	81	81	84	81	82

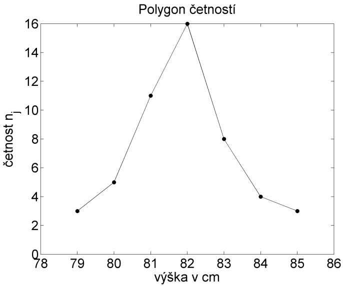
Sestavte tabulku rozdělení četností a graficky jej znázorněte.

Bodové rozdělení četností

x_i	n_i	N_i	p_i	F_i
79	3	3	0,06	0,06
80	5	8	0,1	0,16
81	11	19	0,22	0,38
82	16	35	0,32	0,7
83	8	43	0,16	0,86
84	4	47	0,08	0,94
85	3	50	0,06	1
Σ	50	x	1	x

Tabulka: Tabulka bodového rozdělení četností – výška 15-ti měsíčních dětí

Bodové rozdělení četností



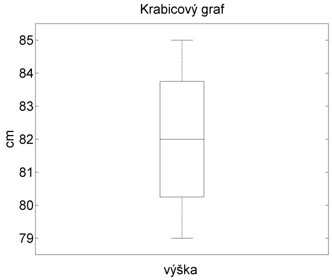
Bodové rozdělení četností

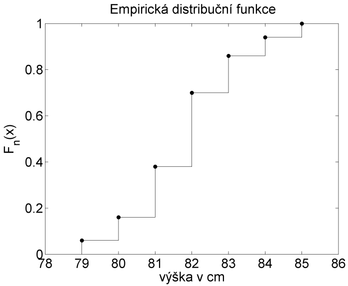
Rozdělení četností je také možné znázornit pomocí *empirické distribuční funkce*, kterou můžeme definovat vztahem

$$F_n(x) = \frac{N(x_i \leq x)}{n},$$

kde výraz v čitateli značí počet prvků výběru, jejichž hodnota je menší nebo rovna x.
Je možné sestrojit také *krabicový graf*, který zobrazuje nejmenší a největší hodnotu znaku, dále medián (případně aritmetický průměr), horní a dolní kvartil.

Bodové rozdělení četností





Intervalové rozdělení četností

Intervalové rozdělení četností $\longrightarrow$ vhodné pro velký rozsah souboru, spojitý znak nebo nespojitý znak s velkým počtem obměn

- konstrukce intervalů
  – počet, šířka a počátek intervalů
- tabulkové vyjádření rozdělení četností
  – n_j, p_j, N_j, F_j, $j = 1, 2, \ldots, k$, kde k udává počet intervalů
- grafické zobrazení rozdělení četností
  – histogram a součtový histogram
- výpočet charakteristik

Intervalové rozdělení četností

Postup při konstrukci intervalů (tříd):

1. zjistíme n, x_{min}, x_{max}
2. určíme *variační rozpětí* $R = x_{max} - x_{min}$
3. stanovení *počtu* intervalů k provedeme podle povahy a struktury dat pomocí:
   - Sturgesovo pravidlo: $k \approx 1 + 3,32 \log n$
   - Yuleovo pravidlo: $k \approx 2,5\sqrt[4]{n}$
   - jiná pravidla: $k \approx \sqrt{n}$; $k \leq 5 \log n$
4. stanovení *šířky* intervalů h: $h \approx \frac{R}{k}$

Intervalové rozdělení četností

Postup při konstrukci intervalů (tříd):

1. zjistíme n, x_{min}, x_{max}
2. určíme *variační rozpětí* $R = x_{max} - x_{min}$
3. stanovení *počtu* intervalů k provedeme podle povahy a struktury dat pomocí:
   - Sturgesovo pravidlo: $k \approx 1 + 3,32 \log n$
   - Yuleovo pravidlo: $k \approx 2,5 \sqrt[4]{n}$
   - jiná pravidla: $k \approx \sqrt{n}$; $k \leq 5 \log n$
4. stanovení *šířky* intervalů h: $h \approx \frac{R}{k}$

Intervalové rozdělení četností

Postup při konstrukci intervalů (tříd):

1. zjistíme n, x_{min}, x_{max}
2. určíme *variační rozpětí* $R = x_{max} - x_{min}$
3. stanovení *počtu* intervalů k provedeme podle povahy a struktury dat pomocí:
   - Sturgesovo pravidlo: $k \approx 1 + 3,32 \log n$
   - Yuleovo pravidlo: $k \approx 2,5\sqrt[4]{n}$
   - jiná pravidla: $k \approx \sqrt{n}$; $k \leq 5 \log n$
4. stanovení *šířky* intervalů h: $h \approx \frac{R}{k}$

Intervalové rozdělení četností

Postup při konstrukci intervalů (tříd):

1. zjistíme n, x_{min}, x_{max}
2. určíme *variační rozpětí* $R = x_{max} - x_{min}$
3. stanovení *počtu* intervalů k provedeme podle povahy a struktury dat pomocí:
   - Sturgesovo pravidlo: $k \approx 1 + 3,32 \log n$
   - Yuleovo pravidlo: $k \approx 2,5\sqrt[4]{n}$
   - jiná pravidla: $k \approx \sqrt{n}$; $k \leq 5 \log n$
4. stanovení *šířky* intervalů h: $h \approx \frac{R}{k}$

Intervalové rozdělení četností

Navíc

- počátek 1. intervalu, počet a šířku intervalů budeme volit tak, aby největší a nejmenší hodnota padly do prvního a posledního intervalu
- intervaly budeme volit polouzavřené zprava, tj. $(x_j - \frac{h}{2},\ x_j + \frac{h}{2}\rangle$
- hranice i středy intervalů by měly být vhodně zaokrouhlené
- způsob, jakým rozdělení provedeme, je individuální

Intervalové rozdělení četností

Navíc

- počátek 1. intervalu, počet a šířku intervalů budeme volit tak, aby největší a nejmenší hodnota padly do prvního a posledního intervalu
- intervaly budeme volit polouzavřené zprava, tj. $(x_j - \frac{h}{2},\ x_j + \frac{h}{2}\rangle$
- hranice i středy intervalů by měly být vhodně zaokrouhlené
- způsob, jakým rozdělení provedeme, je individuální

Intervalové rozdělení četností

Navíc

- počátek 1. intervalu, počet a šířku intervalů budeme volit tak, aby největší a nejmenší hodnota padly do prvního a posledního intervalu
- intervaly budeme volit polouzavřené zprava, tj. $(x_j - \frac{h}{2},\ x_j + \frac{h}{2}\rangle$
- hranice i středy intervalů by měly být vhodně zaokrouhlené
- způsob, jakým rozdělení provedeme, je individuální

Intervalové rozdělení četností

Navíc

- počátek 1. intervalu, počet a šířku intervalů budeme volit tak, aby největší a nejmenší hodnota padly do prvního a posledního intervalu
- intervaly budeme volit polouzavřené zprava, tj. $(x_j - \frac{h}{2},\ x_j + \frac{h}{2}\rangle$
- hranice i středy intervalů by měly být vhodně zaokrouhlené
- způsob, jakým rozdělení provedeme, je individuální

Intervalové rozdělení četností

Při kontrole dodržování hygienických norem v kuchyni se prováděl odběr vzduchu a pomocí filtru Pallflex se měřilo množství prachových částic. Ze 60 vzorků vzduchu jsme dostali následující výsledky (v $\mu g/m^3$):

1,23	1,10	1,54	1,34	1,06	1,09	1,41	1,48	1,52	1,37	1,37	1,63
1,51	1,53	1,31	1,23	1,31	1,27	1,17	1,27	1,34	1,27	1,09	1,01
1,41	1,22	1,27	1,37	1,14	1,22	1,43	1,40	1,41	1,51	1,51	1,47
1,14	1,34	1,16	1,51	1,58	1,33	1,31	1,04	1,58	1,12	1,19	1,17
1,47	1,24	1,45	1,29	1,17	1,63	1,39	1,02	1,38	1,39	1,43	1,28

Sestavte tabulku intervalového rozdělení četností a graficky jej znázorněte.

Intervalové rozdělení četností

Rozsah souboru je $n = 60$, nejmenší hodnota $x_{\min} = 1{,}01$, největší hodnota je $x_{\max} = 1{,}63$. Variační rozpětí je rovno $R = x_{\max} - x_{\min} = 0{,}62$. Určíme si optimální počet intervalů podle zmíněných pravidel:

- Sturgesovo pravidlo $k \approx 1 + 3{,}32 \log n \doteq 7$,
- Yuleovo pravidlo $k \approx 2{,}5\sqrt[4]{n} \doteq 7$,
- $k \approx \sqrt{n} \doteq 8$, $k \approx 5 \log n \doteq 9$.

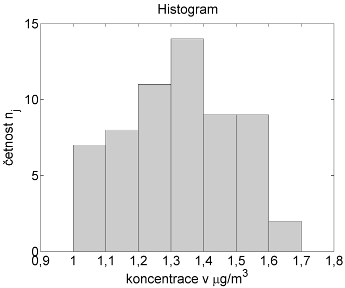
Na základě uvedených pravidel zvolíme např. počet intervalů $k = 7$, šířku intervalu $h = 0{,}1$ a počátek prvního intervalu $a = 1$.

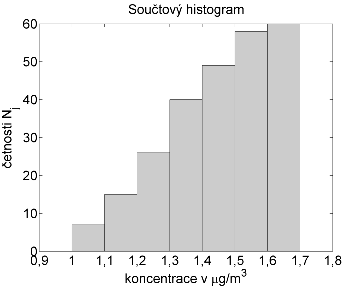
Intervalové rozdělení četností

	x_j	n_j	p_j	N_j	F_j
$(1{,}00; 1{,}10\rangle$	1,05	7	0,177	7	0,117
$(1{,}10; 1{,}20\rangle$	1,15	8	0,133	15	0,250
$(1{,}20; 1{,}30\rangle$	1,25	11	0,183	26	0,433
$(1{,}30; 1{,}40\rangle$	1,35	14	0,233	40	0,667
$(1{,}40; 1{,}50\rangle$	1,45	9	0,150	49	0,817
$(1{,}50; 1{,}60\rangle$	1,55	9	0,150	58	0,967
$(1{,}60; 1{,}70\rangle$	1,65	2	0,033	60	1,000
Σ	x	60	1	x	x

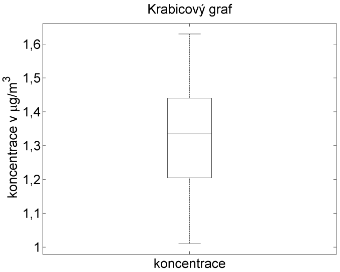
Tabulka: Tabulka intervalového rozdělení četností – množství prachových částic v $\mu g/m^3$

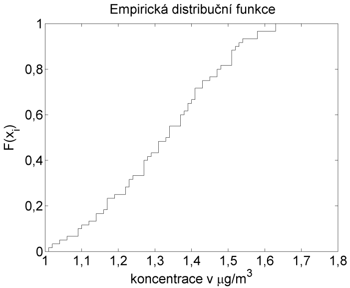
Intervalové rozdělení četností





Intervalové rozdělení četností





Charakteristiky polohy

Charakteristiky polohy (úrovně) měří obecnou velikost hodnot znaku v souboru a dělí se na průměry (počítané ze všech dat) a ostatní míry polohy (počítané z vybraných hodnot).

Aritmetický průměr

Definice

Aritmetický průměr je dán vztahem

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

kde $x_1, x_2 \ldots, x_n$ jsou naměřené hodnoty, n je celkový počet pozorování.

Aritmetický průměr nejčastěji užívaný druh průměru, který má uplatnění při řešení téměř všech úloh statistiky.

Aritmetický průměr

Jsou-li hodnoty statistického znaku uspořádány do tabulky rozdělení četností, určíme aritmetický průměr pomocí vztahu

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} n_i \cdot x_i,$$

kde $n_1, n_2, \ldots, n_k$ jsou četnosti jednotlivých variant znaku $x_1, x_2 \ldots, x_k$. Tyto četnosti udávají váhu jednotlivých variantám znaku x, proto mluvíme o *váženém aritmetickém průměru*.

Aritmetický průměr

Aritmetický průměr má tyto základní vlastnosti:

- součet jednotlivých odchylek od průměru je nulový, tj

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

- aritmetický průměr konstanty je opět roven konstantě, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c = c,$$

Aritmetický průměr

Aritmetický průměr má tyto základní vlastnosti:

- součet jednotlivých odchylek od průměru je nulový, tj

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

- aritmetický průměr konstanty je opět roven konstantě, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c = c,$$

Aritmetický průměr

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, zvýší se o tuto konstantu i aritmetický průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i + c) = c + \overline{x},$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je touto konstantou násoben i průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c \cdot x_i = c \cdot \overline{x}.$$

Aritmetický průměr

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, zvýší se o tuto konstantu i aritmetický průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i + c) = c + \overline{x},$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je touto konstantou násoben i průměr, tj.

$$\frac{1}{n}\sum_{i=1}^{n} c \cdot x_i = c \cdot \overline{x}.$$

Harmonický průměr

Aritmetický průměr však není jediným druhem průměru, existují i jiné, jenž se používají ve speciálních případech.

Definice

Harmonický průměr $\overline{x}_H$ je dán vztahem

$$\overline{x}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}.$$

Harmonický průměr

Harmonický průměr má specifické uplatnění v situacích, kdy má logický význam součet převrácených hodnot znaku. Bude tomu tak tehdy, kdy průměrovaná veličina má charakter části z celku, tedy průměrovat máme tzv. *poměrná čísla*. Např. průměrnou hustotu $\overline{h}$ obyvatelstva na km^2 v kraji, známe-li počet obyvatel p a hustotu h v okresech, určíme ze vztahu $\overline{h} = \frac{\sum p}{\sum r}$, kde rozloha $r = \frac{p}{h}$, nebo průměrnou rychlost $\overline{v}$ auta v km/hod., známe-li dráhu s a jí odpovídající rychlost v, určíme ze vztahu $\overline{v} = \frac{\sum s}{\sum t}$, kde čas $t = \frac{s}{t}$.

Geometrický průměr

Definice

Geometrický průměr $\overline{x}_G$ je dán vztahem

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

Geometrický průměr je např. využíván při jednoduché analýze časové řady pro určení tzv. průměrného tempa růstu nebo průměrného tempa poklesu. Např. pro tři meziroční indexy výroby 1,05; 1,06 a 1,02 je průměrné tempo růstu výroby rovno $\overline{x}_G = \sqrt[3]{1{,}05 \cdot 1{,}06 \cdot 1{,}02} \doteq 1{,}043$, což znamená, že průměrně za rok činil nárůst výroby 4,3 %.

Příklad

Určete aritmetický, harmonický, geometrický a kvadratický průměr z hodnot 1, 2, 5, 6, 7, 8, 8, 9.

- Aritmetický průměr

$$\overline{x} = \frac{1+2+5+6+7+8+8+9}{8} = 5{,}75.$$

- Harmonický průměr

$$\overline{x}_H = \frac{8}{\frac{1}{1}+\frac{1}{2}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{8}+\frac{1}{9}} \doteq 3{,}375.$$

- Geometrický průměr

$$\overline{x}_G = \sqrt[8]{1\cdot 2\cdot 5\cdot 6\cdot 7\cdot 8\cdot 8\cdot 9} \doteq 4{,}709.$$

Všimněte si, že pro naše průměry platí $\overline{x}_H \leq \overline{x}_G \leq \overline{x} \leq \overline{x}_K$, tento vztah mezi průměry platí obecně.

Kvantil

Definice

Kvantil x_p je hodnota znaku, pro kterou platí, že $100p\,\%$ jednotek uspořádaného souboru má hodnotu menší nebo rovnu x_p a $100(1-p)\,\%$ jednotek má hodnotu větší nebo rovnu x_p.

Takto definovaný kvantil není určen jednoznačně. Na jednoduchém příkladu ukážeme, jak počítají kvantily některé softwarové produkty.

Kvantil

Mějme následující datový soubor 2 5 7 10 12 13 18 21.
Možné výpočty kvantilů

- Uspořádejme data vzestupně od nejmenší hodnoty k největší. Určíme pořadový index i_p kvantilu x_p, který musí vyhovovat nerovnosti

$$np < i_p < np + 1.$$

Kvantil x_p je potom roven hodnotě znaku na pozici i_p, tedy $x_p = x_{(i_p)}$. Jsou-li hodnoty $np, np+1$ celočíselné, určíme kvantil jako aritmetický průměr hodnot $x_{(np)}$ a $x_{(np+1)}$, tj. $x_p = \frac{x_{(np)}+x_{(np+1)}}{2}$. Tímto způsobem určuje kvantily např. statistický software STATISTICA.

Kvantil

Mějme následující datový soubor 2 5 7 10 12 13 18 21.
Možné výpočty kvantilů

- Uspořádejme data vzestupně od nejmenší hodnoty k největší. Určíme pořadový index i_p kvantilu x_p, který musí vyhovovat nerovnosti

$$np < i_p < np + 1.$$

Kvantil x_p je potom roven hodnotě znaku na pozici i_p, tedy $x_p = x_{(i_p)}$. Jsou-li hodnoty $np, np + 1$ celočíselné, určíme kvantil jako aritmetický průměr hodnot $x_{(np)}$ a $x_{(np+1)}$, tj. $x_p = \frac{x_{(np)} + x_{(np+1)}}{2}$. Tímto způsobem určuje kvantily např. statistický software STATISTICA.

Kvantil

- Podle MATLABu
  Spočteme se číslo
  $$\bar{i}_p = \frac{np + np + 1}{2} = \frac{2np + 1}{2}$$
  určující polohu kvantilu. Hodnota kvantilu se určí lineární interpolací
  $$x_p = x_{([\bar{i}_p])} + (x_{([\bar{i}_p]+1)} - x_{([\bar{i}_p])})(\bar{i}_p - [\bar{i}_p]),$$
  kde $[\cdot]$ značí celou část čísla. Je-li $\bar{i}_p < 1$ položíme $x_p = x_{(1)}$, je-li $\bar{i}_p > n$ položíme $x_p = x_{(n)}$.

Kvantil

- Podle EXCELu
  Hodnotám uspořádaného souboru se přiřadí postupně hodnoty $0, \frac{1}{n-1}, \frac{2}{n-1}, \ldots, \frac{n-2}{n-1}, 1$. Pokud je hodnota P rovna násobku $\frac{1}{n-1}$, je kvantil x_p roven hodnotě znaku odpovídající danému násobku. Jestliže P není násobkem $\frac{1}{n-1}$, určí se hodnota kvantilu lineární interpolací.

Kvantil

x_p	0,10	0,25	0,50	0,75	0,90
STATISTICA	2	6	11	15,5	21
MATLAB	2,9	6	11	15,5	20,1
EXCEL	4,1	6,5	11	14,25	18,9

Příklad

Určete medián, dolní kvartil a horní decil z hodnot 1, 2, 5, 6, 7, 8, 8, 9.

Nejprve určíme medián, tedy prostřední hodnotu uspořádaného souboru. Rozsah souboru je $n = 8$, neexistuje tedy jedna prostřední hodnota, ale hodnoty dvě (6 a 7). Hodnotu mediánu učíme jako aritmetický průměr těchto hodnot

$$\tilde{x} = x_{0,50} = \frac{6+7}{2} = 6{,}5.$$

Tento výsledek budeme interpretovat takto: 50 % uspořádaných hodnot v souboru je menší nebo rovno 6,5, tedy nepřekročí hodnotu 6,5.

Příklad

Nyní určíme dolní kvartil $x_{0,25}$. Vyjdeme ze vztahu

$$np < i_p < np + 1$$

a dostáváme $8 \cdot 0{,}25 < i_p < 8 \cdot 0{,}25 + 1 \Leftrightarrow 2 < i_p < 3$. V případě, že žádné přirozené číslo nesplňuje danou nerovnici (i_p je pořadový index, tedy přirozené číslo), určíme hledaný kvartil jako aritmetický průměr hodnot, které jsou na pořadí np a $np + 1$, v našem případě průměr druhé a třetí hodnoty v uspořádaném souboru

$$x_{0,25} = \frac{x_{(2)} + x_{(3)}}{2} = \frac{2+5}{2} = 3{,}5.$$

Analogicky určíme horní decil $x_{0,90}$,
$8 \cdot 0{,}90 < i_p < 8 \cdot 0{,}90 + 1 \Leftrightarrow 7{,}2 < i_p < 8{,}2$, odkud $i_p = 8$ a

$$x_{0,90} = x_{(8)} = 9.$$

Řekneme, že 25 % uspořádaných hodnot v souboru je menší nejvýše rovno 3,5. Analogicky 90 % hodnot nepřekročí 9.

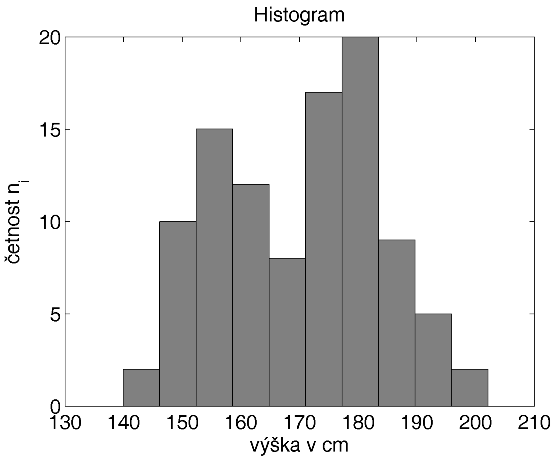
Modus

Definice

Modus $\hat{x}$ je hodnota znaku s největší četností.

V případě spojitého statistického znaku pojem nejčetnější hodnota obvykle nedává smysl, neboť četnosti jednotlivých hodnot znaku jsou buď jedničky, nebo velice malá čísla. Taková data se obvykle zpracovávají pomocí intervalového rozdělení četností a zobrazí pomocí histogramu. Ten interval, který má největší četnost, nazveme *modálním intervalem*.

Modus

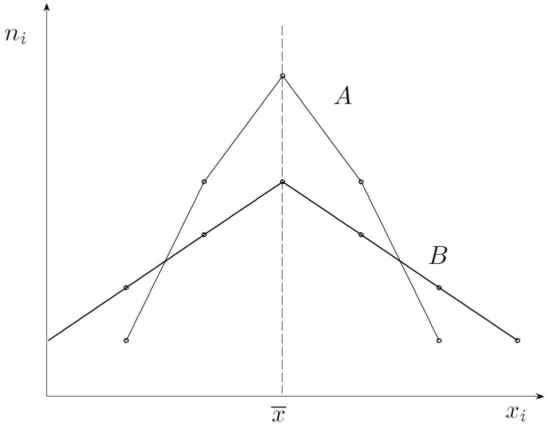


Obrázek: Dvoumodální rozdělení četností

Charakteristiky variability

Průměry, kvantily a modus, tedy charakteristiky o jež byly zmíněny v předchozím odstavci, v sobě shrnují informaci pouze o jedné vlastnosti rozdělení četností, o poloze. Při zpracování dat je možné se setkat s případem, kdy rozdělení četností budou mít shodnou polohu, ale přesto se od sebe budou lišit. Existuje řada měr variabily, zmíníme pouze ty nejdůležitější.

Charakteristiky variability



Obrázek: Rozdělení lišící se variabilitou

Variační rozpětí

Definice

Variační rozpětí R je definováno jako rozdíl největší a nejmenší hodnoty znaku

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

Je to nejjednodušší, ale i nejhrubší míra variabily. Udává šířku intervalu, v němž se nacházejí všechny hodnoty znaku.

Kvantilová rozpětí

Definice

- *kvartilové rozpětí*

$$R_Q = x_{0,75} - x_{0,25}$$

- *decilové rozpětí*

$$R_D = x_{0,90} - x_{0,10}$$

- *percentilové rozpětí*

$$R_C = x_{0,99} - x_{0,01}$$

Kvartilové (resp. decilové resp. percentilové) rozpětí udává šířku intervalu, ve kterém leží 50 % (resp. 80 % resp. 98 %) prostředních hodnot uspořádaného souboru.

Kvantilová rozpětí

Definice

- *kvartilové rozpětí*

$$R_Q = x_{0,75} - x_{0,25}$$

- *decilové rozpětí*

$$R_D = x_{0,90} - x_{0,10}$$

- *percentilové rozpětí*

$$R_C = x_{0,99} - x_{0,01}$$

Kvartilové (resp. decilové resp. percentilové) rozpětí udává šířku intervalu, ve kterém leží 50 % (resp. 80 % resp. 98 %) prostředních hodnot uspořádaného souboru.

Kvantilová rozpětí

Definice

- *kvartilové rozpětí*
$$R_Q = x_{0,75} - x_{0,25}$$
- *decilové rozpětí*
$$R_D = x_{0,90} - x_{0,10}$$
- *percentilové rozpětí*
$$R_C = x_{0,99} - x_{0,01}$$

Kvartilové (resp. decilové resp. percentilové) rozpětí udává šířku intervalu, ve kterém leží 50 % (resp. 80 % resp. 98 %) prostředních hodnot uspořádaného souboru.

Příklad

Data: 2, 5, 7, 10, 12, 13, 18 a 21. Spočteme kvantily (podle programu STATISTICA): $x_{0,10} = 2$, $x_{0,25} = 6$, $x_{0,50} = 11$, $x_{0,75} = 15{,}5$, $x_{0,90} = 21$).

- Variační rozpětí $R = x_{\max} - x_{\min} = 21 - 2 = 19$, všechny hodnoty se nacházejí v intervalu šířky 19.
- Kvartilové rozpětí $R_Q = x_{0,75} - x_{0,25} = 15{,}5 - 6 = 9{,}5$, tj. 50 % prostředních hodnot se nachází v intervalu šířky 9,5.
- Decilové rozpětí je rovno $R_D = x_{0,90} - x_{0,10} = 21 - 2 = 19$.

Kvantilové odchylky

Definice

- *kvartilová odchylka*
$$Q = R_Q/2$$
- *decilová odchylka*
$$D = R_D/8$$
- *percentilová odchylka*
$$C = R_C/98$$

Udává průměrnou vzdálenost mezi dvěma kvartily resp. decily, resp. percentily.

Kvantilové odchylky

Definice

- *kvartilová odchylka*

$$Q = R_Q/2$$

- *decilová odchylka*

$$D = R_D/8$$

- *percentilová odchylka*

$$C = R_C/98$$

Udává průměrnou vzdálenost mezi dvěma kvartily resp. decily, resp. percentily.

Kvantilové odchylky

Definice

- *kvartilová odchylka*

$$Q = R_Q/2$$

- *decilová odchylka*

$$D = R_D/8$$

- *percentilová odchylka*

$$C = R_C/98$$

Udává průměrnou vzdálenost mezi dvěma kvartily resp. decily, resp. percentily.

Příklad

Určete kvartilovou a decilovou odchylku z hodnot 2, 5, 7, 10, 12, 13, 18 a 21. Využijte dřívějších výsledků.

- Kvartilová odchylka $Q = R_Q/2 = 9{,}5/2 = 4{,}75$, tj. průměrná délka dvou prostředních kvartilových intervalů je 4,75,
- Decilová odchylka $D = R_D/8 = 19/8 = 2{,}375$, průměrná délka osmi prostředních decilových intervalů je 2,375.

Průměrná odchylka

Definice

Průměrná odchylka je definována jako aritmetický průměr absolutních odchylek jednotlivých hodnot od aritmetického průměru

$$\bar{d}_{\overline{x}} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|.$$

Příklad

Určete průměrnou odchylku z hodnot 1, 2, 5, 6, 7, 8, 8 a 9.

Hodnota aritmetického průměru je $\overline{x} = 5{,}75$. Dosazením do definičního vzorce dostáváme

$$\overline{d}_{\overline{x}} = \frac{|1-5{,}75| + |2-5{,}75| + |5-5{,}75| + |6-5{,}75|}{8} + \\ + \frac{|7-5{,}75| + |8-5{,}75| + |8-5{,}75| + |9-5{,}75|}{8} = 2{,}3125.$$

Rozptyl

Definice

Rozptyl s_n^2 je definován jako aritmetický průměr čtverců odchylek jednotlivých hodnot znaku od aritmetického průměru

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Patří k nejpoužívanějším mírám variability.

Rozptyl

Pro ruční výpočty rozptylu je možné odvodit jednodušší vzorec

$$
\begin{aligned}
s_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\overline{x}^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - 2n\overline{x}^2 - n\overline{x}^2\right) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2.
\end{aligned}
$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2 = 0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2 = s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i - c\cdot\overline{x})^2 = c^2\cdot s_n^2.$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2 = 0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2 = s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i - c\cdot\overline{x})^2 = c^2\cdot s_n^2.$$

Rozptyl

Rozptyl má tyto základní vlastnosti:

- rozptyl konstanty je roven nule, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c-c)^2 = 0,$$

- přičteme-li k jednotlivým hodnotám znaku x konstantu c, hodnota rozptylu se nezmění, tj.

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i+c)-(\overline{x}+c)]^2 = s_n^2,$$

- násobíme-li jednotlivé hodnoty znaku x konstantou c, je rozptyl násoben čtvercem této konstanty, tj.

$$\frac{1}{n}\sum_{i=1}^{n}(c\cdot x_i - c\cdot\overline{x})^2 = c^2\cdot s_n^2.$$

Směrodatná odchylka

Definice

Odmocnina z rozptylu se nazývá *směrodatná odchylka*

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

Směrodatná odchylka je, na rozdíl od rozptylu, vyjádřena ve stejných jednotkách jako sledovaný znak. Tvoří-li např. statistický soubor výsledky ve skoku vysokém vyjádřené v centimetrech, má i směrodatná odchylka jednotku *cm*, rozptyl je potom vyjádřen v jednotkách cm^2.

Výběrový rozptyl a směrodatná odchylka

Definice

Výběrový rozptyl s^2 je definovaný vztahem

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

odmocnina z výběrového rozptylu se nazývá *výběrová směrodatná odchylka*

$$s = \sqrt{s^2}.$$

Používá se v induktivní statistice. Jak plyne z definic rozptylu a výběrového rozptylu, platí mezi nimi vztah

$$s_n^2 = \frac{n-1}{n}s^2.$$

Příklad

Určete rozptyl, směrodatnou odchylku, výběrový rozptyl a výběrovou směrodatnou odchylku z hodnot 1, 2, 5, 6, 7, 8, 8 a 9.

Určili hodnotu aritmetického průměru $\overline{x} = 5{,}75$. Nejprve spočítáme hodnotu rozptylu z definičního vzorce

$$s_n^2 = \frac{(1-5{,}75)^2 + (2-5{,}75)^2 + (5-5{,}75)^2 + (6-5{,}75)^2}{8} + \\ + \frac{(7-5{,}75)^2 + (8-5{,}75)^2 + (8-5{,}75)^2 + (9-5{,}75)^2}{8} = 7{,}4375.$$

Příklad

Rozptyl je možné také určit pomocí vztahu $s_n^2 = \overline{x^2} - \overline{x}^2$. Určíme tedy hodnotu

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1^2 + 2^2 + 5^2 + 6^2 + 7^2 + 8^2 + 8^2 + 9^2}{8} = 40{,}5,$$

odtud potom dostáváme

$$s_n^2 = \overline{x^2} - \overline{x}^2 = 40{,}5 - 5{,}75^2 = 7{,}4375.$$

Směrodatná odchylka je

$$s_n = \sqrt{s_n^2} = \sqrt{7{,}4375} \doteq 2{,}72718.$$

Příklad

Výběrový rozptyl můžeme samozřejmě určit z definice, jednodušší bude ale využít vztahu

$$s^2 = \frac{n}{n-1} s_n^2 = \frac{8}{7} \cdot 7{,}4375 = 8{,}5.$$

Výběrová směrodatná odchylka má potom hodnotu

$$s = \sqrt{s^2} = \sqrt{8{,}5} \doteq 2{,}91548.$$

Obecný a centrální moment

Definice

r-tý obecný moment je definován vztahem

$$m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r,$$

r-tý centrální moment je definován vztahem

$$m_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^r.$$

Koeficient šikmosti

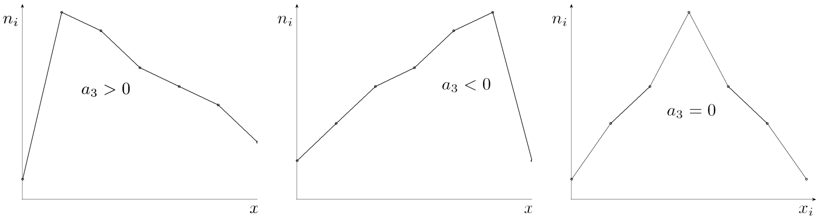
Definice

Koeficient šikmosti je dán vztahem

$$a_3 = \frac{m_3}{m_2^{3/2}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n s_n^3} = \frac{m_3}{s_n^3}$$

Je-li $a_3 = 0$, je stupeň hustoty malých a velkých hodnot stejný, což představuje souměrné rozdělení četností. Je-li $a_3 > 0$, je stupeň hustoty malých hodnot ve srovnání s hustotou velkých hodnot větší a rozdělení četností je proto zešikmené doleva. Analogicky je-li $a_3 < 0$, je rozdělení četností zešikmené doprava.

Koeficient šikmosti



Obrázek: Rozdělení lišící se šikmostí

Koeficient špičatosti

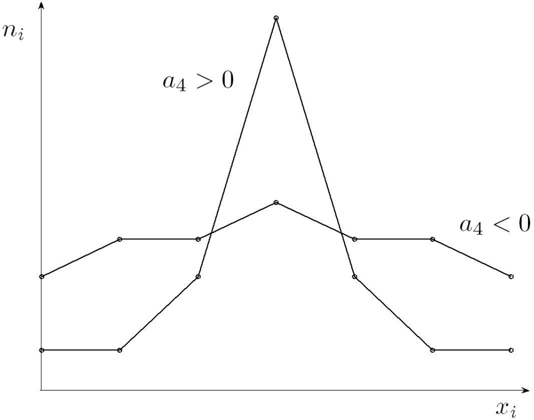
Definice

Koeficient špičatosti je dán vztahem

$$a_4 = \frac{m_4}{m_2^2} - 3 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{n s_n^4} - 3$$

Je-li $a_4 > 0$, je stupeň koncentrace prostředních hodnot ve srovnání s koncentrací všech hodnot větší a rozdělení četností se potom projeví špičatým tvarem. Analogicky je-li $a_4 < 0$, má rozdělení četností plochý tvar.

Koeficient špičatosti



Obrázek: Rozdělení lišící se špičatostí

Ukázka zpracování reálných dat v aplikaci STAT1

Příklad 1 Bodové rozdělení četností

- rozdělení četností – tabulka
- polygon četností
- součtová křivka
- číselné charakteristiky
- interpretace

Ukázka zpracování reálných dat v aplikaci STAT1

Příklad 2 Intervalové rozdělení četností

- rozdělení četností – tabulka
- histogram
- součtový histogram
- číselné charakteristiky
- interpretace

Základní literatura

- MANN, P.S. Introductory Statistics. 6th edition. Hoboken: Wiley, 2007. ISBN 978-0-471-75530-2.
- MOUČKA, J., RÁDL, P. Matematika pro studenty ekonomie. 1. vyd. Grada 2010. ISBN 978-80-247-3260-2.
- NEUBAUER, J., SEDLAČÍK, M., KŘÍŽ, O. Základy statistiky – Aplikace v technických a ekonomických oborech. Grada 2012.ISBN: 978-80-247-4273-1.
- ŘEZANKOVÁ, H. Analýza dat z dotazníkových šetření. 2. vydání, Professional Publishing, 2010. ISBN: 9788074310195.

Doporučená literatura

- AGRESTI, A. Categorical Data Analysis. Second Edition. Wiley 2002. ISBN: 0-471-36093-7.
- ANDĚL, J. Statistické metody. 3. vydání. Praha: Matfyzpress, 2003. ISBN 80-86732-08-8.
- ANDĚL, J. Základy matematické statistiky. 2. vyd. Praha: Matfyzpress, 2007, 358 s. ISBN 978-80-7378-001-2.
- VÁGNER, M. Integrální počet funkcí jedné proměnné. 1. vydání. Brno: UO, 2005,126 s. ISBN 80-7231-025-9.
- VÁGNER, M., KAŠTÁNKOVÁ, V. Posloupnosti a řady. 1. vydání. Brno: UO, 2006. ISBN 80-7231-131-X.