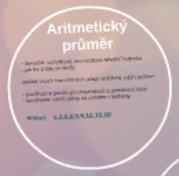
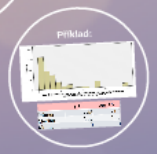
Popisné statistiky:

Míry centrální tendence a variability

Mgr. Zuzana Lenhartová

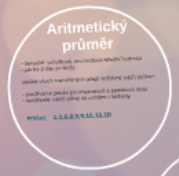


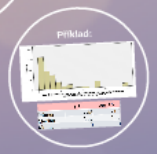


Popisné statistiky:

Míry centrální tendence a variability

Mgr. Zuzana Lenhartová





Popisné statistiky

- jednorozměrný popis a analýza proměnných

Kategoriální deskriptivy

- entropie
- modus

Pořadové deskriptivy

- medián
- kvartily
- percentily
- grafické zobrazení pomocí pořadových dekriptiv - boxplot

Odchylkové, momentové deskriptivy

- aritmetický průměr
- rozptyl, směrodatná odchylka
- zešikmení
- špičatost

Centrální tendence

Střední hodnoty, umístění

- nevýhoda tabulky četností i grafického zobrazení - neúspornost (hodně čísel -> špatná orientace)
- nemůžeme proměnnou popsat rychle

Jak zobrazení dat zredukovat?

- úsporně popsat rozložení proměnných skrze ukazatele centrální tendence a ukazatele variability
- najít hodnotu, která by všechny naměřené hodnoty dobře reprezentovavala

Ukazatel centrální tendence

= ukazatel středních hodnot; ukazatel míry polohy

- charakteristika typické hodnoty dat
- ukazuje, kde se na měřené škále (číselné ose) data nalézají
- popisuje rozložení četností jedné proměnné

Ukazatel variability

- udává, jak moc či málo jsou data na škále rozptýlená

Ukazatele centrální tendence

- popisná statistika (číselná charakteristika proměnné)
- ukazatel středních hodnot
- udávají průměrnou, typickou, reprezentativní, očekávanou hodnotu - jeden údaj
- jedno číslo - krásné a zrádné

- **modus**
- **medián**
- **aritmetický průměr**

Modus

$\hat{X}, Mo$

- kategoriální typická hodnota
- **nejčastější hodnota**
(hodnota s nejvyšší četností v datech)
- jediná možnost u nominálních dat, u vyšších úrovní často užitečnou volbou
- když známe všechny naměřené hodnoty, stanovíme modus tak, že zjistíme, která hodnota se v daném souboru vyskytuje nejčastěji

Příklad: 14, 3, 18, 4, 8, 18, 4, 6, 8, 10, 8

- v případě tabulky četnosti s intervaly lze modus určit přibližně jako střed intervalu s největší četností
- *nezávislý na extrémních hodnotách naměřené veličiny*

  - modus nemusí být určen jednoznačně - se stejnou nejvyšší frekvencí se může vyskytovat více hodnot
  - rozdělení s jedním modem (vrcholem) - unimodální
  - rozdělení pravděpodobnosti s dvěma vrcholy - dvouvrcholová (bimodální).

Medián

$\tilde{X}, Md$

- pořadová střední hodnota
- prostřední hodnota z řady hodnot seřazených podle velikosti
- 50. percentil - rozděluje soubor dat na dvě stejné části
- při sudém počtu prvků je mediánem průměr ze dvou prostředních hodnot/ kterékoli číslo z intervalu mezi nejbližší vyšší a nejbližší nižší hodnotou (konsensuálně střed intervalu)
- používáme pro (ordinální) pořadová data a výše
- nezávislý na extrémních hodnotách měřené veličiny

Příklad: Měření vědomostí žáků didaktickým testem, výsledky:

14, 3, 18, 4, 8, 18, 4, 6, 8, 10, 8

Aritmetický průměr

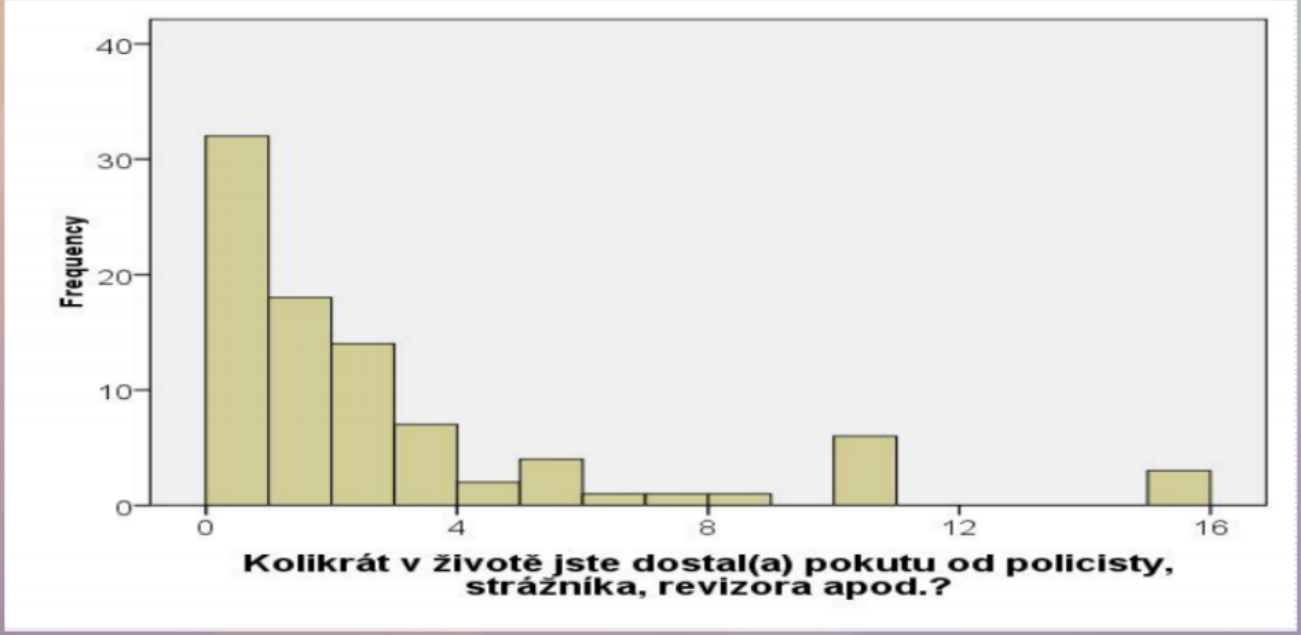
- deviační, ochylková, momentová střední hodnota
- jak ho znáte ze školy:

součet všech naměřených údajů vydělený jejich počtem

- používáme pouze pro intervalová a poměrová data
- nevýhoda: velmi citlivý na extrémní hodnoty

Příklad: **1,3,6,8,9,9,10,10,10**

Příklad:



	s 15	bez 15
Průměr	2,48	2,05
Medián	1,00	1,00
Modus	0	0

Příklad:

Určete průměr, medián a modus u těchto čtyř rozložení (sad dat):

a. 3, 3, 4, 5, 6, 8, 8, 8, 9
b. 2, 4, 4, 4, 6, 7, 7
c. 7, 7, 8, 9, 10, 10, 10
d. 1, 1, 3, 4, 5, 9

Míry variability

(rozptýlenosti)

- omezenost středních hodnot - udávají pouze to, kolem jaké hodnoty se data "centrují" - které jsou nejčastější
- data se stejnou střední hodnotou mohou mít různou rozptýlenost

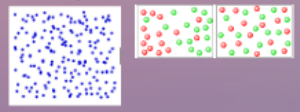
Variabilita - jak moc či málo jsou data na škále rozptýlená

- malá variabilita - většina hodnot v souboru je stejných nebo velmi blízkých
- vysoká variabilita - hodnoty jsou velmi rozmanité
- tři ukazatelé variability (podle škál)

Na nominální škále:

Entropie

- veličina udávající "míru neuspořádanosti" zkoumaného systému
- míra neurčitosti systému



- v sociálních vědách se moc nepoužívá

Pokud ano, tak:
- **variační poměr či nominální variance**

Na pořadové škále:

Variační rozpětí

- pokud můžeme seřadit hodnoty od nejmenší po největší a můžeme říct, co je minimum a co je maximum, máme rozpětí

- R = Xmax – Xmin

- extrémně roste s velikostí vzorku - čím větší soubor, tím větší hodnota rozpětí
- nevýhoda: Vysoká citlivost vůči outlierům

Příklad: 2, 8, 9, 10, 1, 0, 5

Interkvartilové rozpětí

- vzdálenost mezi dvěma body na škále, které jsou na nějakém místě, které můžeme snadno definovat - používá se 25. a 75. percentil

- Q = Q3 - Q1 (75. percentil minus 25. percentil)

- používáme spíše než jednoduché variační rozpětí

Na intervalové, poměrové škále

- charakteristiky založené na odchylkách od průměru
- měří rozptýlenost dat kolem aritmetického průměru

Rozptyl

- aritmetický průměr čtverců odchylek od aritmetického průměru
(průměrná kvadratická odchylka měření od aritmetického průměru, přičemž při průměrování této odchylky dělíme číslem (n-1))
= průměrná odchylka na druhou
- populační rozptyl: **(Σx2 / n)**
- výběrový rozptyl - vhodnější: **(Σx2 / (n – 1)** - při počítání pro všechny prvky populace
součet odchylek na druhou = **suma čtverců** (sečtu odchylky od průměru na druhou)
- používá se v inferenční statistice

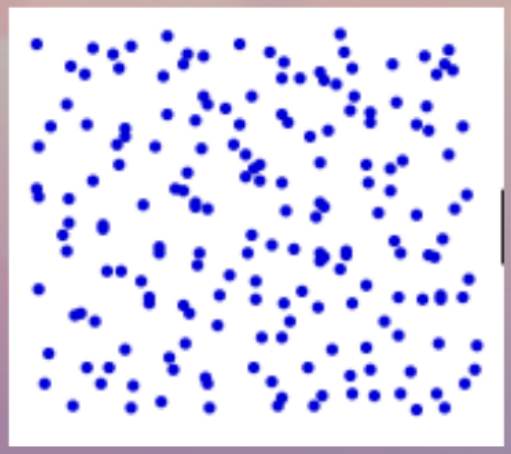
Směrodatná odchylka

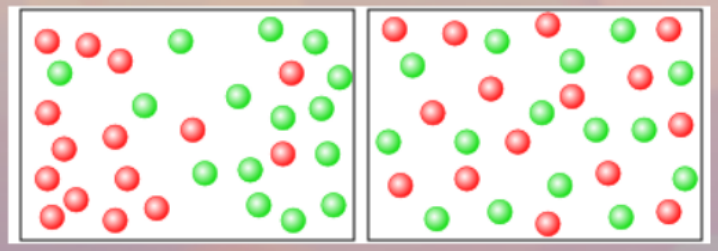
- standardní odchylka
- odmocnina rozptylu - návrat k původní jednotce, ve které měříme

Na nominální škále:

Entropie

- veličina udávající "míru neuspořádanosti" zkoumaného systému
- míra neurčitosti systému

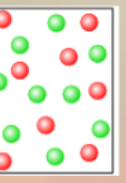




- v sociálních vědách se moc nepoužívá

Pokud ano, tak**:**

- variační poměr či nominální variance



Na pořadové škále:

Variační rozpětí

- pokud můžeme seřadit hodnoty od nejmenší po největší a můžeme říct, co je minimum a co je maximum, máme rozpětí

- R = Xmax – Xmin

- extrémně roste s velikostí vzorku - čím větší soubor, tím větší hodnota rozpětí
- nevýhoda: Vysoká citlivost vůči outlierům

Příklad: 2, 8, 9, 10, 1, 0, 5

Interkvartilové rozpětí

- vzdálenost mezi dvěma body na škále, které jsou na nějakém místě, které můžeme snadno definovat - používá se 25. a 75. percentil

- Q = Q3 - Q1 (75. percentil minus 25. percentil)

- používáme spíše než jednoduché variační rozpětí

Na intervalové, poměrové škále

- charakteristiky založené na odchylkách od průměru
- měří rozptýlenost dat kolem aritmetického průměru

Rozptyl

- aritmetický průměr čtverců odchylek od aritmetického průměru
(průměrná kvadratická odchylka měření od aritmetického průměru, přičemž při průměrování této odchylky dělíme číslem (n-1))
= průměrná odchylka na druhou
- populační rozptyl: **(Σx2 / n)**
- výběrový rozptyl - vhodnější: **(Σx2 / (n – 1)** - při počítání pro všechny prvky populace
součet odchylek na druhou = **suma čtverců** (sečtu odchylky od průměru na druhou)
- používá se v inferenční statistice

Směrodatná odchylka

- standardní odchylka
- odmocnina rozptylu - návrat k původní jednotce, ve které měříme

Popisné statistiky:

Míry centrální tendence a variability

Mgr. Zuzana Lenhartová