

# Základy matematiky a statistiky pro humanitní obory

I

Pavel Rychlý Vojtěch Kovář

Fakulta informatiky, Masarykova univerzita  
Botanická 68a, 602 00 Brno, Czech Republic

{pary, xkovar3}@fi.muni.cz

část 9

# Obsah přednášky

- 1 Statistika
- 2 Statistický soubor
- 3 Jednorozměrný soubor
- 4 Dvourozměrný soubor

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech



# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistika

## ■ Motivace

- sumarizace informací o velkých souborech dat
- odhady informací o velkém souboru dat na základě menšího vzorku
- modelování různých souborů dat
- např. jazyka (prostřednictvím korpusů)

## ■ Cíl přednášky

- seznámit se se základními pojmy statistiky
- → využití v navazujících předmětech

# Statistický soubor

## ■ Základní soubor (populace)

- soubor údajů (statistických znaků) o objektech
- každý objekt souboru má **statistické znaky**
- jejich počet = **rozměr souboru**
- např. všichni sloni v Africe – výška a hmotnost

## ■ Statistický soubor

- výběr objektů ze základního souboru
- např. ti sloni, které se podařilo zvážit a změřit
- měl by být reprezentativní
- → můžeme vyvozovat znalosti o základním souboru
- často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr



# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
    - měl by být reprezentativní
    - → můžeme vyvozovat znalosti o základním souboru
    - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Statistický soubor

- Základní soubor (populace)
  - soubor údajů (statistických znaků) o objektech
  - každý objekt souboru má **statistické znaky**
  - jejich počet = **rozměr souboru**
  - např. všichni sloni v Africe – výška a hmotnost
- Statistický soubor
  - výběr objektů ze základního souboru
  - např. ti sloni, které se podařilo zvážit a změřit
  - měl by být reprezentativní
  - → můžeme vyvozovat znalosti o základním souboru
  - často náhodný výběr

# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3



# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3



# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3



# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3



# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3



# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3

# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnot
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3

# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3

# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3

# Jednorozměrný statistický soubor

- Např. hmotnost slonů v Africe
  - podařilo se nám zvážit 6 slonů
  - ti měli hmotnosti 2, 4, 4, 4, 5 a 11 tun
- Statistický soubor
  - šestice (2, 4, 4, 4, 5, 11)
- Rozsah statistického souboru
  - počet jeho prvků (6)
- Absolutní četnost hodnoty
  - počet jejích výskytů v souboru
  - např. absolutní četnost hodnoty 4 je 3

# Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd





## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

# Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnoty
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

# Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd



# Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

# Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

## Jednorozměrný statistický soubor (II)

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Relativní četnost hodnot
  - absolutní četnost / rozsah souboru
  - např. relativní četnost hodnoty 4 je 50 %
- Kumulativní četnost
  - četnost příslušné hodnoty + četnost všech menších hodnot
  - absolutní nebo relativní
  - např. kumulativní absolutní četnost hodnoty 4 je 4
- Hodnoty mohou být rozděleny do tříd
  - → absolutní/relativní/kumulativní četnost třídy
- Histogram
  - sloupcový graf znázorňující četnosti jednotlivých tříd

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4



# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky polohy

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Aritmetický průměr
  - „těžiště“ statistického souboru (někdy značíme *avg*)
  - např. 5
- Modus
  - hodnota (třída) s největší četností
  - např. 4
- Medián
  - „prostřední“ hodnota (nebo průměr ze dvou prostředních)
  - není citlivá na extrémní odchylky (jako průměr)
  - např. 4

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem



# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Charakteristiky variability

- Statistický soubor (2, 4, 4, 4, 5, 11)
- Rozptyl (disperze, variance)  $s^2$ 
  - aritmetický průměr druhých mocnin odchylek od průměrné hodnoty
  - $((x_1 - avg)^2 + (x_2 - avg)^2 + \dots + (x_n - avg)^2)/n$
  - např.  
 $((-3)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 0^2 + 6^2)/6 = 8$
  - větší rozptyl  $\equiv$  větší variabilita hodnot
- Směrodatná odchylka  $s$ 
  - odmocnina z rozptylu
  - vyjadřuje totéž, jen jiným číslem

# Dvouřozměrný statistický soubor

## ■ Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

## ■ Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost

■

$$\frac{\sum_{0 < i \leq n} (x_i - \text{avg}_x)(y_i - \text{avg}_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky



# Dvouřozměrný statistický soubor

- Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů

- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$

- lze chápat jako dva svázané jednorozměrné soubory

- Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$

- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“

- $0$  = žádná závislost;  $1$  = lineární závislost

- 

$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvourozměrný statistický soubor

- Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů

- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$

- lze chápat jako dva svázané jednorozměrné soubory

- Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$

- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“

- $0$  = žádná závislost;  $1$  = lineární závislost

- 

$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvouřozměrný statistický soubor

- Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

- Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvourozměrný statistický soubor

- Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

- Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvouřozměrný statistický soubor

- Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

- Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvouřozměrný statistický soubor

## ■ Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

## ■ Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou”
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvouřozměrný statistický soubor

## ■ Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

## ■ Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

# Dvourozměrný statistický soubor

## ■ Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

## ■ Koeficient korelace

- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky



# Dvouřozměrný statistický soubor

## ■ Dvě hodnoty pro každý objekt

- např. výška a hmotnost slonů
- $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$
- lze chápat jako dva svázané jednorozměrné soubory

## ■ Koeficient korelace

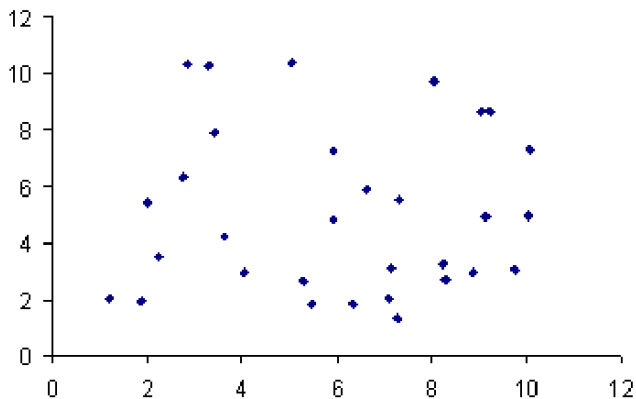
- určuje míru lineární závislosti znaků  $x$  a  $y$
- „jak dobře jde grafem závislosti  $x$  na  $y$  proložit přímkou“
- $0$  = žádná závislost;  $1$  = lineární závislost



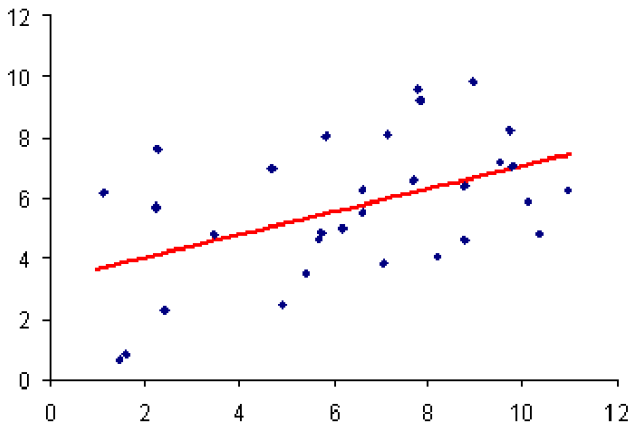
$$\frac{\sum_{0 < i \leq n} (x_i - avg_x)(y_i - avg_y)}{n * s(x) * s(y)}$$

- $(s(x), s(y))$  jsou směrodatné odchylky

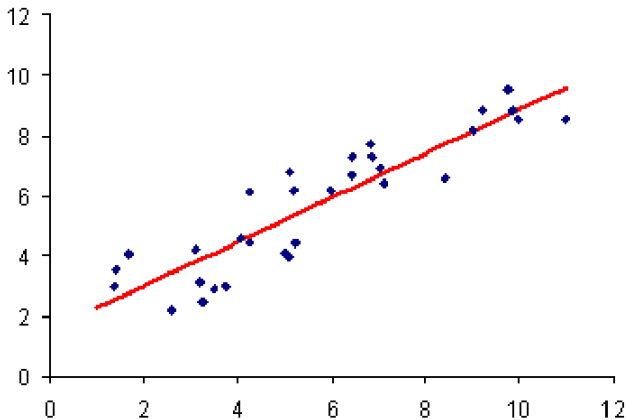
## Korelace 0



# Korelace 0,5



# Korelace 0,9



## Korelace -0,7

