

# Základy biostatistiky

(MD710P09)

ak. rok 2008/2009

Karel Zvára

karel.zvara@mff.cuni.cz

<http://www.karlin.mff.cuni.cz/~zvara>

katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 9. března 2009)



## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)



## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, ..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvára>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998, . . . , 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

- ▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998,..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998,..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

▶ **konzultace** pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)

## ▶ cvičení na počítačích v B5

- ▶ od úterka 24. února ve Viničné 7, 1. patro B5
- ▶ nutno **zapsat se do paralelky** prostřednictvím SIS
- ▶ zápočet za aktivní účast (+ odevzdávání souborů/písemky)
- ▶ nutno mít aktivní účet v učebnách, znát svoje heslo
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)

## ▶ zkouška v B5

- ▶ jen se zápočtem, přihlašování prostřednictvím SIS
- ▶ kombinace písemného a ústního zkoušení
- ▶ řešení úloh na počítači
- ▶ základy teorie (pojmy, metody a jejich volba, interpretace)

## ▶ literatura

- ▶ K. Zvára: Biostatistika. Karolinum 1998,..., 2008
- ▶ internetová stránka <http://www.karlin.mff.cuni.cz/~zvara>

## ▶ konzultace pondělí 13:15-14:00 v pracovně, II. patro K234, Sokolovská 83, Karlín (případně po předchozí dohodě jindy, možno také v ÚAMVT, Albertov 6)



# tři části přednášky

## ▶ popisná statistika

- ▶ několika čísly vystihnout důležitou vlastnost
- ▶ jednoduchým grafem vyjádřit důležitou vlastnost
- ▶ porovnat soubory dat

## ▶ abstraktní pohled (teorie)

- ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
- ▶ populace a výběr, populační parametr a jeho odhad
- ▶ interval spolehlivosti pro parametr
- ▶ test statistické hypotézy

## ▶ některé metody (modely)

- ▶ testy o jednom, dvou či několika výběrech
- ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ **popisná statistika**
  - ▶ **několika čísla vystihnout důležitou vlastnost**
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ **abstraktní pohled (teorie)**
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ **některé metody (modely)**
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ **popisná statistika**
  - ▶ **několika čísla vystihnout důležitou vlastnost**
  - ▶ **jednoduchým grafem vyjádřit důležitou vlastnost**
  - ▶ porovnat soubory dat
- ▶ **abstraktní pohled (teorie)**
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ **některé metody (modely)**
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků



# tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

## tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

## tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

## tři části přednášky

- ▶ popisná statistika
  - ▶ několika čísly vystihnout důležitou vlastnost
  - ▶ jednoduchým grafem vyjádřit důležitou vlastnost
  - ▶ porovnat soubory dat
- ▶ abstraktní pohled (teorie)
  - ▶ pravděpodobnost, Bayesův vzorec, náhodná veličina, distribuční funkce, nezávislost
  - ▶ populace a výběr, populační parametr a jeho odhad
  - ▶ interval spolehlivosti pro parametr
  - ▶ test statistické hypotézy
- ▶ některé metody (modely)
  - ▶ testy o jednom, dvou či několika výběrech
  - ▶ rozhodování o závislosti kvantitativních či kvalitativních znaků

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace
- ▶ **příklady dat:**
  - ▶ **výšky:** výška desetiletých chlapců/dívek
  - ▶ **děti:** pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
  - ▶ **kojení:** hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace

## ▶ příklady dat:

- ▶ **výšky**: výška desetiletých chlapců/dívek
- ▶ **děti**: pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
- ▶ **kojení**: hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace

## ▶ příklady dat:

- ▶ **výšky**: výška desetiletých chlapců/dívek
- ▶ **děti**: pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
- ▶ **kojení**: hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace

## ▶ příklady dat:

- ▶ **výšky**: výška desetiletých chlapců/dívek
- ▶ **děti**: pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
- ▶ **kojení**: hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice



# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace

## ▶ příklady dat:

- ▶ **výšky**: výška desetiletých chlapců/dívek
- ▶ **děti**: pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
- ▶ **kojení**: hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace
- ▶ **příklady dat:**
  - ▶ **výšky:** výška desetiletých chlapců/dívěk
  - ▶ **děti:** pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
  - ▶ **kojení:** hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

# statistika

## ▶ statistika

- ▶ **popisná** (deskriptivní):  
data stručně popsat, něco z dat „vydolovat“  
tvrdit něco o daných datech, nezobecňovat
- ▶ **induktivní** (konfirmatorní):  
tvrdit něco nového, zobecnit na větší soubor,  
důležitá je interpretace

## ▶ příklady dat:

- ▶ **výšky**: výška desetiletých chlapců/dívek
- ▶ **děti**: pohlaví, porodní hmotnost a délka, hmotnost a délka v jednom roce, věk otce a matky, počet onemocnění otitidou v prvním roce věku
- ▶ **kojení**: hmotnost a délka porodní a ve 24. týdnu, věk a výška obou rodičů, zda těhotenství plánováno, zda dudlík, porodnice

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára



## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

## co měříme (zjišťujeme) a kde

- ▶ měříme na **statistických jednotkách** (osoba, obec, stát, pokusné pole, rostlinka pšenice, třetí list rostlinky pšenice, ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ **znak** - vlastnost měřená na objektu (statistické jednotce)
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (vyšetřování závislosti)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti, které charakterizují celou velkou skupinu (**populaci**)
- ▶ hodnoty znaků zjišťujeme u jedinců, nechceme vypovídat pouze o jednotlivcích
- ▶ kolik procent mužů kouří, ne, zda kouří Karel Zvára

# měřítko

- ▶ **nula-jedničkové**  
pouze dvě možné hodnoty (muž/žena, kouří/nekouří)
- ▶ **nominální**  
seznam všech rozlišitelných hodnot, **faktor**  
(porodnice, pohlaví, odrůda)
- ▶ **ordinální**  
hodnoty nominálního měřítka jsou uspořádány,  
**uspořádaný faktor** (vzdělání matky, stupeň bolesti)
- ▶ **intervalové**  
stejně vzdálenosti sousedních hodnot (rok narození)  
„o kolik je x menší než y?“ (nikoliv „kolikrát“)
- ▶ **poměrové**  
srovnání se zvolenou jednotkou (hmotnost, výška, věk)  
„kolikrát je x větší, než y?“

# měřítko

- ▶ **nula-jedničkové**  
pouze dvě možné hodnoty (muž/žena, kouří/nekouří)
- ▶ **nominální**  
seznam všech rozlišitelných hodnot, **faktor**  
(porodnice, pohlaví, odrůda)
- ▶ **ordinální**  
hodnoty nominálního měřítka jsou uspořádány,  
**uspořádaný faktor** (vzdělání matky, stupeň bolesti)
- ▶ **intervalové**  
stejně vzdálenosti sousedních hodnot (rok narození)  
„o kolik je x menší než y?“ (nikoliv „kolikrát“)
- ▶ **poměrové**  
srovnání se zvolenou jednotkou (hmotnost, výška, věk)  
„kolikrát je x větší, než y?“

# měřítko

- ▶ **nula-jedničkové**  
pouze dvě možné hodnoty (muž/žena, kouří/nekouří)
- ▶ **nominální**  
seznam všech rozlišitelných hodnot, **faktor**  
(porodnice, pohlaví, odrůda)
- ▶ **ordinální**  
hodnoty nominálního měřítka jsou uspořádány,  
**uspořádaný faktor** (vzdělání matky, stupeň bolesti)
- ▶ **intervalové**  
stejně vzdálenosti sousedních hodnot (rok narození)  
„o kolik je x menší než y?“ (nikoliv „kolikrát“)
- ▶ **poměrové**  
srovnání se zvolenou jednotkou (hmotnost, výška, věk)  
„kolikrát je x větší, než y?“

# měřítko

- ▶ **nula-jedničkové**  
pouze dvě možné hodnoty (muž/žena, kouří/nekouří)
- ▶ **nominální**  
seznam všech rozlišitelných hodnot, **faktor**  
(porodnice, pohlaví, odrůda)
- ▶ **ordinální**  
hodnoty nominálního měřítka jsou uspořádány,  
**uspořádaný faktor** (vzdělání matky, stupeň bolesti)
- ▶ **intervalové**  
stejně vzdálenosti sousedních hodnot (rok narození)  
„**o kolik** je  $x$  menší než  $y$ ?“ (nikoliv „kolikrát“)
- ▶ **poměrové**  
srovnání se zvolenou jednotkou (hmotnost, výška, věk)  
„**kolikrát** je  $x$  větší, než  $y$ ?“



# měřítko

- ▶ **nula-jedničkové**  
pouze dvě možné hodnoty (muž/žena, kouří/nekouří)
- ▶ **nominální**  
seznam všech rozlišitelných hodnot, **faktor**  
(porodnice, pohlaví, odrůda)
- ▶ **ordinální**  
hodnoty nominálního měřítka jsou uspořádány,  
**uspořádaný faktor** (vzdělání matky, stupeň bolesti)
- ▶ **intervalové**  
stejně vzdálenosti sousedních hodnot (rok narození)  
„o kolik je  $x$  menší než  $y$ ?“ (nikoliv „kolikrát“)
- ▶ **poměrové**  
srovnání se zvolenou jednotkou (hmotnost, výška, věk)  
„kolikrát je  $x$  větší, než  $y$ ?“

# hrubší dělení měřitek

(důležitější, bezprostředně ovlivní volbu metod)

## ▶ **kvalitativní**

nula-jedničkové, nominální, často i ordinální

▶ u kvalitativních se zpravidla udávají **četnosti** jednotlivých hodnot (kolikrát která hodnota nastala)

## ▶ **kvantitativní** (spojité)

intervalové, poměrové, někdy ordinální (ale není spojitě)

▶ hodnoty kvantitativních – čísla

▶ pro četnosti hodnot v kvalitativním měřítku se používají zpravidla jiné charakteristiky a metody, než pro hodnoty v kvantitativním měřítku

# hrubší dělení měřítek

(důležitější, bezprostředně ovlivní volbu metod)

- ▶ **kvalitativní**

nula-jedničkové, nominální, často i ordinální

- ▶ u kvalitativních se zpravidla udávají **četnosti** jednotlivých hodnot (kolikrát která hodnota nastala)

- ▶ **kvantitativní** (spojité)

intervalové, poměrové, někdy ordinální (ale není spojité)

- ▶ hodnoty kvantitativních – čísla

- ▶ pro četnosti hodnot v kvalitativním měřítku se používají zpravidla jiné charakteristiky a metody, než pro hodnoty v kvantitativním měřítku

# hrubší dělení měřítek

(důležitější, bezprostředně ovlivní volbu metod)

- ▶ **kvalitativní**  
nula-jedničkové, nominální, často i ordinální
- ▶ u kvalitativních se zpravidla udávají **četnosti** jednotlivých hodnot (kolikrát která hodnota nastala)
- ▶ **kvantitativní** (spojité)  
intervalové, poměrové, někdy ordinální (ale není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ pro četnosti hodnot v kvalitativním měřítku se používají zpravidla jiné charakteristiky a metody, než pro hodnoty v kvantitativním měřítku

# hrubší dělení měřitek

(důležitější, bezprostředně ovlivní volbu metod)

- ▶ **kvalitativní**  
nula-jedničkové, nominální, často i ordinální
- ▶ u kvalitativních se zpravidla udávají **četnosti** jednotlivých hodnot (kolikrát která hodnota nastala)
- ▶ **kvantitativní** (spojité)  
intervalové, poměrové, někdy ordinální (ale není spojitě)
- ▶ hodnoty kvantitativních – čísla
- ▶ pro četnosti hodnot v kvalitativním měřítku se používají zpravidla jiné charakteristiky a metody, než pro hodnoty v kvantitativním měřítku

# hrubší dělení měřitek

(důležitější, bezprostředně ovlivní volbu metod)

- ▶ **kvalitativní**  
nula-jedničkové, nominální, často i ordinální
- ▶ u kvalitativních se zpravidla udávají **četnosti** jednotlivých hodnot (kolikrát která hodnota nastala)
- ▶ **kvantitativní** (spojité)  
intervalové, poměrové, někdy ordinální (ale není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ pro četnosti hodnot v kvalitativním měřítku se používají zpravidla jiné charakteristiky a metody, než pro hodnoty v kvantitativním měřítku

# veličina, statistika

- ▶ číselně vyjádřený výsledek měření, pokusu
- ▶ možné hodnoty znaků v intervalovém nebo poměrovém měřítku jsou hustě rozmístěné – **spojitá veličina**
- ▶ četnosti hodnot znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ u veličin používáme číselné charakteristiky některých hromadných vlastností (**charakteristiky polohy**, **charakteristiky variability**, charakteristiky tvaru)
- ▶ **statistika** (další význam) – funkce pozorovaných hodnot např. průměrná teplota nebo nejvyšší teplota v roce; číselně charakterizuje důležitou vlastnost veličiny (veličin)

# veličina, statistika

- ▶ číselně vyjádřený výsledek měření, pokusu
- ▶ možné hodnoty znaků v intervalovém nebo poměrovém měřítku jsou hustě rozmístěné – **spojitá veličina**
- ▶ četnosti hodnot znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ u veličin používáme číselné charakteristiky některých hromadných vlastností (**charakteristiky polohy**, **charakteristiky variability**, charakteristiky tvaru)
- ▶ **statistika** (další význam) – funkce pozorovaných hodnot např. průměrná teplota nebo nejvyšší teplota v roce; číselně charakterizuje důležitou vlastnost veličiny (veličin)



# veličina, statistika

- ▶ číselně vyjádřený výsledek měření, pokusu
- ▶ možné hodnoty znaků v intervalovém nebo poměrovém měřítku jsou hustě rozmístěné – **spojitá veličina**
- ▶ četnosti hodnot znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ u veličin používáme číselné charakteristiky některých hromadných vlastností (**charakteristiky polohy**, **charakteristiky variability**, charakteristiky tvaru)
- ▶ **statistika** (další význam) – funkce pozorovaných hodnot např. průměrná teplota nebo nejvyšší teplota v roce; číselně charakterizuje důležitou vlastnost veličiny (veličin)

# veličina, statistika

- ▶ číselně vyjádřený výsledek měření, pokusu
- ▶ možné hodnoty znaků v intervalovém nebo poměrovém měřítku jsou hustě rozmístěné – **spojitá veličina**
- ▶ četnosti hodnot znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ u veličin používáme číselné charakteristiky některých hromadných vlastností (**charakteristiky polohy**, **charakteristiky variability**, charakteristiky tvaru)
- ▶ **statistika** (další význam) – funkce pozorovaných hodnot např. průměrná teplota nebo nejvyšší teplota v roce; číselně charakterizuje důležitou vlastnost veličiny (veličin)

# veličina, statistika

- ▶ číselně vyjádřený výsledek měření, pokusu
- ▶ možné hodnoty znaků v intervalovém nebo poměrovém měřítku jsou hustě rozmístěné – **spojitá veličina**
- ▶ četnosti hodnot znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ u veličin používáme číselné charakteristiky některých hromadných vlastností (**charakteristiky polohy**, **charakteristiky variability**, charakteristiky tvaru)
- ▶ **statistika** (další význam) – funkce pozorovaných hodnot např. průměrná teplota nebo nejvyšší teplota v roce; číselně charakterizuje důležitou vlastnost veličiny (veličin)

## označení

rozlišujte  $n$ ,  $n_i$ ,  $m$ ,  $x_i$ ,  $x_i^*$ 

$x_1,$	$x_2,$	$\dots,$	$x_n$	zjištěné hodnoty
$x_1^*,$	$x_2^*,$	$\dots,$	$x_m^*$	možné hodnoty (různé)
$n_1,$	$n_2,$	$\dots,$	$n_m$	<b>četnosti</b> hodnot

$$n_1 + n_2 + \dots + n_m = \sum_{j=1}^m n_j = n$$

$$\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_m}{n} \quad - \text{relativní četnosti}$$

$$N_j = \sum_{i=1}^j n_i \quad \text{kumulativní četnosti}$$

pro kumulativní četnosti nutno aspoň ordinální měřítko

# příklad kojení (vzdělání 99 matek)

ordinální měřítko se třemi hodnotami

vzděl.	zákl.	maturita	VŠ	celkem	pozn.
$x_j^*$	1	2	3		možné hodnoty
$n_j$	34	47	18	99	absolutní čet.
$n_j/n$	0,343	0,475	0,182	1,000	relativní čet.
$n_j/n$	34,3 %	47,5 %	18,2 %	100 %	relativní čet.
$N_j$	34	81	99		kumulativní čet.

# histogram (barplot u kvalitativní veličiny)

- ▶ **histogram**  
grafické znázornění intervalových četností spojité veličiny
- ▶ **barplot**  
grafické znázornění četností (počtů hodnot) kvalitativního znaku
- ▶ plocha (výška) obdélníku úměrná četnosti
- ▶ relativní četnosti mají jen jiné měřítko svislé osy
- ▶ **výsečový diagram** pro relativní četnosti kvalitativního znaku (podíly nějakého celku)

# histogram (barplot u kvalitativní veličiny)

- ▶ **histogram**  
grafické znázornění intervalových četností spojité veličiny
- ▶ **barplot**  
grafické znázornění četností (počtů hodnot) kvalitativního znaku
  - ▶ plocha (výška) obdélníku úměrná četnosti
  - ▶ relativní četnosti mají jen jiné měřítko svislé osy
  - ▶ **výsečový diagram** pro relativní četnosti kvalitativního znaku (podíly nějakého celku)

# histogram (barplot u kvalitativní veličiny)

- ▶ **histogram**  
grafické znázornění intervalových četností spojité veličiny
- ▶ **barplot**  
grafické znázornění četností (počtů hodnot) kvalitativního znaku
- ▶ plocha (výška) obdélníku úměrná četnosti
- ▶ relativní četnosti mají jen jiné měřítko svislé osy
- ▶ **výsečový diagram** pro relativní četnosti kvalitativního znaku (podíly nějakého celku)

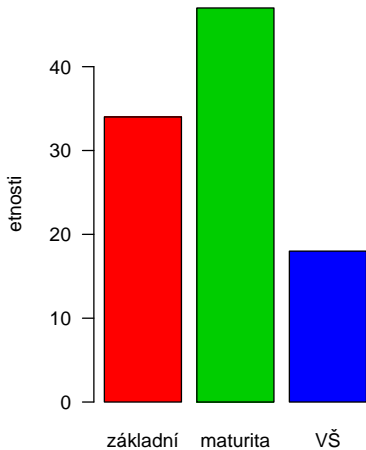


# histogram (barplot u kvalitativní veličiny)

- ▶ **histogram**  
grafické znázornění intervalových četností spojité veličiny
- ▶ **barplot**  
grafické znázornění četností (počtů hodnot) kvalitativního znaku
- ▶ plocha (výška) obdélníku úměrná četnosti
- ▶ relativní četnosti mají jen jiné měřítko svislé osy
- ▶ **výsečový diagram** pro relativní četnosti kvalitativního znaku (podíly nějakého celku)

## histogram (barplot u kvalitativní veličiny)

- ▶ **histogram**  
grafické znázornění intervalových četností spojité veličiny
- ▶ **barplot**  
grafické znázornění četností (počtů hodnot) kvalitativního znaku
- ▶ plocha (výška) obdélníku úměrná četnosti
- ▶ relativní četnosti mají jen jiné měřítko svislé osy
- ▶ **výsečový diagram** pro relativní četnosti kvalitativního znaku (podíly nějakého celku)



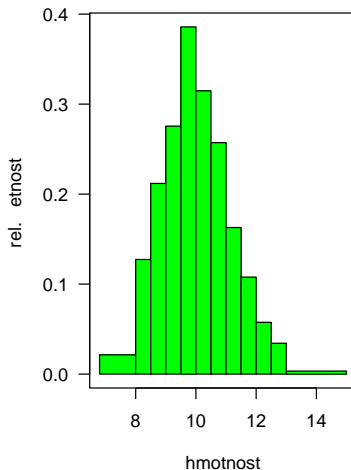
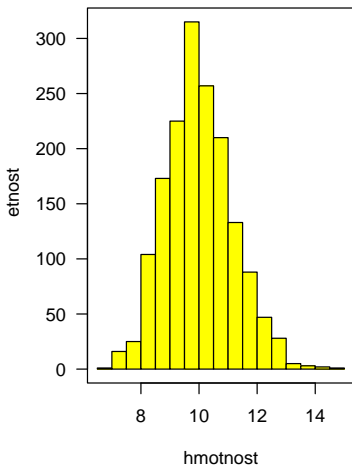
## histogram u **spojité** veličiny

**třídění:** všechny hodnoty z daného intervalu  $(t_{j-1}, t_j)$  nahradíme prostřední hodnotou  $x_j^* = (t_{j-1} + t_j)/2$   
hmotnost dětí ve 12. měsíci (příklad **děti**)

$j$	$x_j^*$	$t_j$	$n_j$	$n_j/n$	$N_j$	$N_j/n$
1	7750	8000	42	0,026	42	0,026
2	8250	8500	104	0,063	146	0,089
3	8750	9000	173	0,106	319	0,195
4	9250	9500	225	0,138	544	0,333
5	9750	10000	315	0,193	859	0,526
6	10250	10500	257	0,157	1116	0,683
7	10750	11000	210	0,129	1326	0,812
8	11250	11500	133	0,081	1459	0,893
9	11750	12000	88	0,054	1547	0,947
10	12250	12500	47	0,029	1594	0,976
11	12750	13000	28	0,017	1622	0,992
12	13250	$\infty$	11	0,007	1633	1,000

# histogram pro hmotnost v jednom roce

histogram napravo podle třídnicích četností udává relativní četnosti



## variační řada, pořadí

nutno rozlišovat  $x_i$  a  $x_{(i)}$

- ▶ původní hodnoty spojité veličiny (kvantitativní znak)

$$x_1, x_2, \dots, x_n \quad \text{např. } 7, 4, 5, 4, 2$$

- ▶ variační řada  $[\text{sort}(x)]$

$$x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad \text{např. } 2, 4, 4, 5, 7$$

- ▶ pořadí:  $[\text{rank}(x)]$

na které místo ve variační řadě se dostane daná hodnota nejmenší dostane pořadí 1, druhá nejmenší dostane 2, ...

- ▶ je-li několik hodnot stejných, dostanou průměr z odpovídajících pořadí
- ▶ pořadí hodnot 7, 4, 5, 4, 2 jsou po řadě 5, 2,5, 4, 2,5, 1

## variační řada, pořadí

nutno rozlišovat  $x_i$  a  $x_{(i)}$

- ▶ původní hodnoty spojité veličiny (kvantitativní znak)

$$x_1, x_2, \dots, x_n \quad \text{např. } 7, 4, 5, 4, 2$$

- ▶ **variační řada**  $[\text{sort}(x)]$

$$x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad \text{např. } 2, 4, 4, 5, 7$$

- ▶ **pořadí**:  $[\text{rank}(x)]$

na které místo ve variační řadě se dostane daná hodnota nejmenší dostane pořadí 1, druhá nejmenší dostane 2, ...

- ▶ je-li několik hodnot stejných, dostanou průměr z odpovídajících pořadí
- ▶ pořadí hodnot 7, 4, 5, 4, 2 jsou po řadě 5, 2,5, 4, 2,5, 1

## variační řada, pořadí

nutno rozlišovat  $x_i$  a  $x_{(i)}$

- ▶ původní hodnoty spojité veličiny (kvantitativní znak)

$$x_1, x_2, \dots, x_n \quad \text{např. } 7, 4, 5, 4, 2$$

- ▶ **variační řada**  $[\text{sort}(x)]$

$$x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad \text{např. } 2, 4, 4, 5, 7$$

- ▶ **pořadí**:  $[\text{rank}(x)]$

na které místo ve variační řadě se dostane daná hodnota nejmenší dostane pořadí 1, druhá nejmenší dostane 2, ...

- ▶ je-li několik hodnot stejných, dostanou průměr z odpovídajících pořadí
- ▶ pořadí hodnot 7, 4, 5, 4, 2 jsou po řadě 5, 2,5, 4, 2,5, 1



## variační řada, pořadí

nutno rozlišovat  $x_i$  a  $x_{(i)}$

- ▶ původní hodnoty spojité veličiny (kvantitativní znak)

$$x_1, x_2, \dots, x_n \quad \text{např. } 7, 4, 5, 4, 2$$

- ▶ **variační řada**  $[\text{sort}(x)]$

$$x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad \text{např. } 2, 4, 4, 5, 7$$

- ▶ **pořadí**:  $[\text{rank}(x)]$

na které místo ve variační řadě se dostane daná hodnota nejmenší dostane pořadí 1, druhá nejmenší dostane 2, ...

- ▶ je-li několik hodnot stejných, dostanou průměr z odpovídajících pořadí
- ▶ pořadí hodnot 7, 4, 5, 4, 2 jsou po řadě 5, 2,5, 4, 2,5, 1

## variační řada, pořadí

nutno rozlišovat  $x_i$  a  $x_{(i)}$

- ▶ původní hodnoty spojité veličiny (kvantitativní znak)

$$x_1, x_2, \dots, x_n \quad \text{např. } 7, 4, 5, 4, 2$$

- ▶ **variační řada**  $[\text{sort}(x)]$

$$x_{(1)} \leq x_{(2)} \dots \leq x_{(n)} \quad \text{např. } 2, 4, 4, 5, 7$$

- ▶ **pořadí**:  $[\text{rank}(x)]$

na které místo ve variační řadě se dostane daná hodnota nejmenší dostane pořadí 1, druhá nejmenší dostane 2, ...

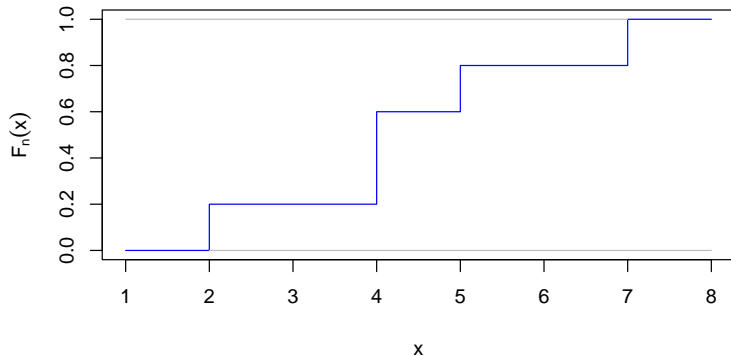
- ▶ je-li několik hodnot stejných, dostanou průměr z odpovídajících pořadí
- ▶ pořadí hodnot 7, 4, 5, 4, 2 jsou po řadě 5, 2,5, 4, 2,5, 1

# empirická distribuční funkce

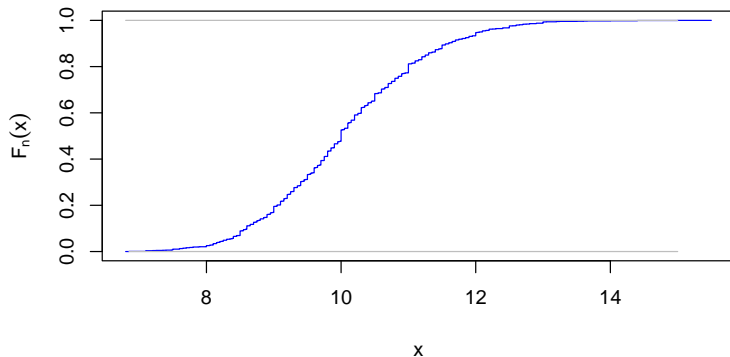
[empirical distribution function]

relativní četnost hodnot, které jsou nejvýše  $x$   
naše variační řada: 2, 4, 4, 5, 7

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

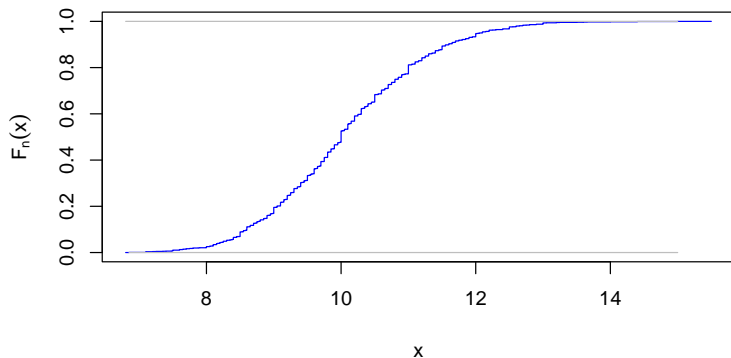


# empirická distribuční funkce



- ▶ příklad: váha dětí v jednom roce
- ▶ připomíná hladkou neklesající funkci

# empirická distribuční funkce



- ▶ příklad: váha dětí v jednom roce
- ▶ připomíná hladkou neklesající funkci

# průměry

- ▶ **průměr** [mean(x)]

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

- ▶ **vážený průměr** s využitím četností ( $n = \sum_j n_j$ )

$$\bar{x} = \frac{1}{n}(n_1 x_1^* + n_2 x_2^* + \dots + n_m x_m^*) = \frac{1}{n} \sum_{j=1}^m n_j x_j^*$$

- ▶ obecněji s nezápornými vahami  $w_j$  hodnot  $x_j^*$

$$\bar{x} = \frac{\sum_j w_j x_j^*}{\sum_j w_j}$$

[weighted.mean(x, w)]

# průměry

- ▶ **průměr** [mean(x)]

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

- ▶ **vážený průměr** s využitím četností ( $n = \sum_j n_j$ )

$$\bar{x} = \frac{1}{n}(n_1 x_1^* + n_2 x_2^* + \dots + n_m x_m^*) = \frac{1}{n} \sum_{j=1}^m n_j x_j^*$$

- ▶ obecněji s nezápornými vahami  $w_j$  hodnot  $x_j^*$

$$\bar{x} = \frac{\sum_j w_j x_j^*}{\sum_j w_j}$$

[weighted.mean(x, w)]

# průměry

- ▶ **průměr** [mean(x)]

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

- ▶ **vážený průměr** s využitím četností ( $n = \sum_j n_j$ )

$$\bar{x} = \frac{1}{n}(n_1 x_1^* + n_2 x_2^* + \dots + n_m x_m^*) = \frac{1}{n} \sum_{j=1}^m n_j x_j^*$$

- ▶ obecněji s nezápornými vahami  $w_j$  hodnot  $x_j^*$

$$\bar{x} = \frac{\sum_j w_j x_j^*}{\sum_j w_j}$$

[weighted.mean(x, w)]



## příklad: vážený průměr známek vážený kredity

jaký je nevážený průměr?

známka	kreditů	součin
$x_j^*$	$w_j$	$x_j^* \cdot w_j$
1	6	6
2	4	8
2	2	4
3	4	12
celkem	16	30

$$\bar{x} = \frac{6 \cdot 1 + 4 \cdot 2 + 2 \cdot 2 + 4 \cdot 3}{6 + 4 + 2 + 4} = \frac{30}{16} = 1,875$$

[weighted.mean(x=c(1,2,2,3),w=c(6,4,2,4))]

## další míry polohy

opět jsou důležité závorky kolem indexů

- ▶ **medián** (prostřední hodnota, NIKOLIV *střední hodnota*)

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ liché} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & n \text{ sudé} \end{cases} \quad [\text{median}(x)]$$

- ▶ **minimum, maximum**

$$x_{\min} = x_{(1)} \quad [\min(x)]$$

$$x_{\max} = x_{(n)} \quad [\max(x)]$$

$[\text{range}(x)]$  spočítá dvojici  $(x_{\min}, x_{\max})$

- ▶ **variační průměr**  $[\text{mean}(\text{range}(x))]$

$$\frac{1}{2} (x_{(1)} + x_{(n)}) = \frac{1}{2} (x_{\min} + x_{\max})$$

## další míry polohy

opět jsou důležité závorky kolem indexů

- ▶ **medián** (prostřední hodnota, NIKOLIV *střední hodnota*)

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ liché} \\ \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & n \text{ sudé} \end{cases} \quad [\text{median}(x)]$$

- ▶ **minimum, maximum**

$$x_{\min} = x_{(1)} \quad [\min(x)]$$

$$x_{\max} = x_{(n)} \quad [\max(x)]$$

$[\text{range}(x)]$  spočítá dvojici  $(x_{\min}, x_{\max})$

- ▶ **variační průměr**  $[\text{mean}(\text{range}(x))]$

$$\frac{1}{2} (x_{(1)} + x_{(n)}) = \frac{1}{2} (x_{\min} + x_{\max})$$

## další míry polohy

opět jsou důležité závorky kolem indexů

- ▶ **medián** (prostřední hodnota, NIKOLIV *střední hodnota*)

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ liché} \\ \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & n \text{ sudé} \end{cases} \quad [\text{median}(x)]$$

- ▶ **minimum, maximum**

$$x_{\min} = x_{(1)} \quad [\min(x)]$$

$$x_{\max} = x_{(n)} \quad [\max(x)]$$

$[\text{range}(x)]$  spočítá dvojici  $(x_{\min}, x_{\max})$

- ▶ **variační průměr**  $[\text{mean}(\text{range}(x))]$

$$\frac{1}{2} (x_{(1)} + x_{(n)}) = \frac{1}{2} (x_{\min} + x_{\max})$$

## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$

## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$

## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$

## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$



## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$

## kvartily, decily

- ▶ **medián**  $\tilde{x}$  je číslo, které dělí data na dvě poloviny: hodnot menších nebo stejných jako medián – hodnot větších nebo stejných jako medián  $[\text{median}(x)]$   $[\text{quantile}(x, \text{probs}=1/2)]$
- ▶ **dolní kvartil**  $Q_1$  je číslo, které oddělí čtvrtinu hodnot (menších či stejných jako  $Q_1$ ) od tří čtvrtin hodnot (větších či stejných jako  $Q_1$ )  $[\text{quantile}(x, \text{probs}=1/4)]$
- ▶ **horní kvartil**  $Q_3$  je číslo, které oddělí tři čtvrtiny hodnot (menších či stejných jako  $Q_3$ ) od čtvrtiny hodnot (větších či stejných jako  $Q_3$ )  $[\text{quantile}(x, \text{probs}=3/4)]$
- ▶ **první decil** je číslo, které oddělí desetinu nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=1/10)]$
- ▶ **percentil**  $x_p$  je číslo, které oddělí  $100p$  % nejmenších hodnot od ostatních hodnot  $[\text{quantile}(x, \text{probs}=p)]$
- ▶ několik percentilů současně  $[\text{quantile}(x, \text{probs}=(0:4)/4)]$

## výpočet percentilu $x_p$

jeden z nejčastěji užívaných postupů, též v R

- ▶ najde se celé číslo  $k$  splňující

$$\frac{k-1}{n-1} \leq p < \frac{k}{n-1}$$

- ▶ tedy  $k = \lfloor 1 + (n-1) \cdot p \rfloor$

( $\lfloor x \rfloor$  znamená celou část z  $x$ , zaokrouhlí dolů)

- ▶ provede se lineární interpolace mezi  $x_{(k)}$  a  $x_{(k+1)}$

( $\{x\}$  znamená zlomkovou část  $x$ , o kolik přesahuje celé číslo)

$$q = \{1 + (n-1) \cdot p\} = (1 + (n-1) \cdot p) - k$$

$$x_p = (1 - q) \cdot x_{(k)} + q \cdot x_{(k+1)}$$

- ▶ např. pro  $n = 99$ ,  $p = 0,25$  bude

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25, \quad q = 25,5 - 25 = 0,5$$

$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

## výpočet percentilu $x_p$

jeden z nejčastěji užívaných postupů, též v R

- ▶ najde se celé číslo  $k$  splňující

$$\frac{k-1}{n-1} \leq p < \frac{k}{n-1}$$

- ▶ tedy  $k = \lfloor 1 + (n-1) \cdot p \rfloor$

( $\lfloor x \rfloor$  znamená celou část z  $x$ , zaokrouhlí dolů)

- ▶ provede se lineární interpolace mezi  $x_{(k)}$  a  $x_{(k+1)}$

( $\{x\}$  znamená zlomkovou část  $x$ , o kolik přesahuje celé číslo)

$$q = \{1 + (n-1) \cdot p\} = (1 + (n-1) \cdot p) - k$$

$$x_p = (1 - q) \cdot x_{(k)} + q \cdot x_{(k+1)}$$

- ▶ např. pro  $n = 99$ ,  $p = 0,25$  bude

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25, \quad q = 25,5 - 25 = 0,5$$

$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

## výpočet percentilu $x_p$

jeden z nejčastěji užívaných postupů, též v R

- ▶ najde se celé číslo  $k$  splňující

$$\frac{k-1}{n-1} \leq p < \frac{k}{n-1}$$

- ▶ tedy  $k = \lfloor 1 + (n-1) \cdot p \rfloor$

( $\lfloor x \rfloor$  znamená celou část z  $x$ , zaokrouhlí dolů)

- ▶ provede se lineární interpolace mezi  $x_{(k)}$  a  $x_{(k+1)}$

( $\{x\}$  znamená zlomkovou část  $x$ , o kolik přesahuje celé číslo)

$$q = \{1 + (n-1) \cdot p\} = (1 + (n-1) \cdot p) - k$$

$$x_p = (1 - q) \cdot x_{(k)} + q \cdot x_{(k+1)}$$

- ▶ např. pro  $n = 99, p = 0,25$  bude

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25, \quad q = 25,5 - 25 = 0,5$$

$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

## výpočet percentilu $x_p$

jeden z nejčastěji užívaných postupů, též v R

- ▶ najde se celé číslo  $k$  splňující

$$\frac{k-1}{n-1} \leq p < \frac{k}{n-1}$$

- ▶ tedy  $k = \lfloor 1 + (n-1) \cdot p \rfloor$

( $\lfloor x \rfloor$  znamená celou část z  $x$ , zaokrouhlí dolů)

- ▶ provede se lineární interpolace mezi  $x_{(k)}$  a  $x_{(k+1)}$

( $\{x\}$  znamená zlomkovou část  $x$ , o kolik přesahuje celé číslo)

$$q = \{1 + (n-1) \cdot p\} = (1 + (n-1) \cdot p) - k$$

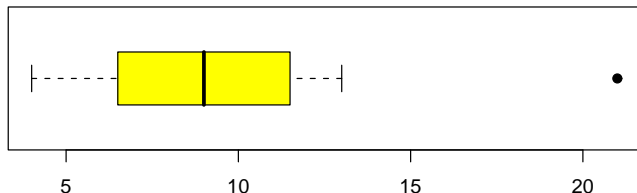
$$x_p = (1 - q) \cdot x_{(k)} + q \cdot x_{(k+1)}$$

- ▶ např. pro  $n = 99$ ,  $p = 0,25$  bude

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25, \quad q = 25,5 - 25 = 0,5$$

$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

# krabicový diagram

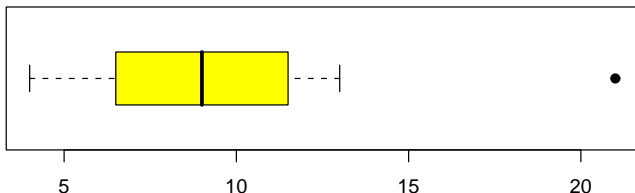


`[boxplot(c(4,5,8,9,10,13,21),horizontal=TRUE,col=7,pch=16)]`

znázorěna řada statistik pro data: 4, 5, 8, 9, 10, 13, 21

- ▶ medián ( $\tilde{x} = 9$ ) – příčka obdélníka
- ▶ kvartily ( $Q_1 = 6,5$ ,  $Q_3 = 11,5$ ) – kratší strany obdélníka
- ▶ tykadla od kvartilu k minimu (maximu), pokud není odlehlé
- ▶ odlehlé pozorování – je dál, než  $3/2 \cdot (Q_3 - Q_1)$  ( $= 7,5$ ) od bližšího kvartilu

## krabicový diagram



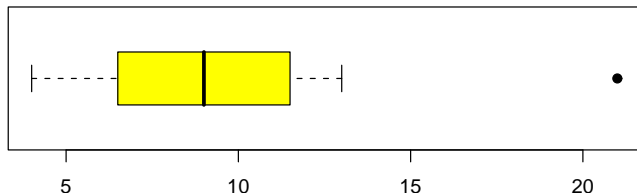
`[boxplot(c(4,5,8,9,10,13,21),horizontal=TRUE,col=7,pch=16)]`

znázorěna řada statistik pro data: 4, 5, 8, 9, 10, 13, 21

- ▶ medián ( $\tilde{x} = 9$ ) – příčka obdélníka
- ▶ kvartily ( $Q_1 = 6,5$ ,  $Q_3 = 11,5$ ) – kratší strany obdélníka
- ▶ tykadla od kvartilu k minimu (maximu), pokud není odlehlé
- ▶ odlehlé pozorování – je dál, než  $3/2 \cdot (Q_3 - Q_1)$  ( $= 7,5$ ) od bližšího kvartilu



## krabicový diagram

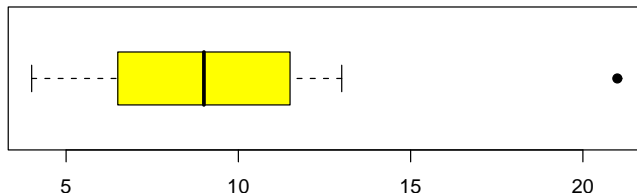


`[boxplot(c(4,5,8,9,10,13,21),horizontal=TRUE,col=7,pch=16)]`

znázorěna řada statistik pro data: 4, 5, 8, 9, 10, 13, 21

- ▶ medián ( $\bar{x} = 9$ ) – příčka obdélníka
- ▶ kvartily ( $Q_1 = 6,5$ ,  $Q_3 = 11,5$ ) – kratší strany obdélníka
- ▶ tykadla od kvartilu k minimu (maximu), pokud není odlehlé
- ▶ odlehlé pozorování – je dál, než  $3/2 \cdot (Q_3 - Q_1)$  ( $= 7,5$ ) od bližšího kvartilu

## krabicový diagram



`[boxplot(c(4,5,8,9,10,13,21),horizontal=TRUE,col=7,pch=16)]`

znázorěna řada statistik pro data: 4, 5, 8, 9, 10, 13, 21

- ▶ medián ( $\bar{x} = 9$ ) – příčka obdélníka
- ▶ kvartily ( $Q_1 = 6,5$ ,  $Q_3 = 11,5$ ) – kratší strany obdélníka
- ▶ tykadla od kvartilu k minimu (maximu), pokud není odlehlé
- ▶ odlehlé pozorování – je dál, než  $3/2 \cdot (Q_3 - Q_1)$  ( $= 7,5$ ) od bližšího kvartilu

## vlastnosti míry polohy

- ▶ přičteme-li ke každé hodnotě  $x$  stejnou konstantu  $a$ , musíme tutéž konstantu  $a$  přičíst k průměru (mediánu, kvartilu, ...)
- ▶ vynásobíme-li každou hodnotu  $x$  stejnou kladnou konstantou  $b$ , musíme průměr (medián, kvartil, ...) vynásobit totéž konstantou  $b$
- ▶ pro dobrou míru polohy  $\mu(X)$  platí:

$$\mu(a + X) = a + \mu(X)$$

$$\mu(b \cdot X) = b \cdot \mu(X) \quad (b > 0)$$

- ▶ dobrá míra polohy je citlivá vůči posunutí i vůči změně měřítka, zachová je

## vlastnosti míry polohy

- ▶ přičteme-li ke každé hodnotě  $x$  stejnou konstantu  $a$ , musíme tutéž konstantu  $a$  přičíst k průměru (mediánu, kvartilu, ...)
- ▶ vynásobíme-li každou hodnotu  $x$  stejnou kladnou konstantou  $b$ , musíme průměr (medián, kvartil, ...) vynásobit totéž konstantou  $b$
- ▶ pro dobrou míru polohy  $\mu(X)$  platí:

$$\mu(a + X) = a + \mu(X)$$

$$\mu(b \cdot X) = b \cdot \mu(X) \quad (b > 0)$$

- ▶ dobrá míra polohy je citlivá vůči posunutí i vůči změně měřítka, zachová je

## vlastnosti míry polohy

- ▶ přičteme-li ke každé hodnotě  $x$  stejnou konstantu  $a$ , musíme tutéž konstantu  $a$  přičíst k průměru (mediánu, kvartilu, ...)
- ▶ vynásobíme-li každou hodnotu  $x$  stejnou kladnou konstantou  $b$ , musíme průměr (medián, kvartil, ...) vynásobit totéž konstantou  $b$
- ▶ pro dobrou míru polohy  $\mu(X)$  platí:

$$\mu(a + X) = a + \mu(X)$$

$$\mu(b \cdot X) = b \cdot \mu(X) \quad (b > 0)$$

- ▶ dobrá míra polohy je citlivá vůči posunutí i vůči změně měřítka, zachová je

## vlastnosti míry polohy

- ▶ přičteme-li ke každé hodnotě  $x$  stejnou konstantu  $a$ , musíme tutéž konstantu  $a$  přičíst k průměru (mediánu, kvartilu, ...)
- ▶ vynásobíme-li každou hodnotu  $x$  stejnou kladnou konstantou  $b$ , musíme průměr (medián, kvartil, ...) vynásobit totéž konstantou  $b$
- ▶ pro dobrou míru polohy  $\mu(X)$  platí:

$$\mu(a + X) = a + \mu(X)$$

$$\mu(b \cdot X) = b \cdot \mu(X) \quad (b > 0)$$

- ▶ dobrá míra polohy je citlivá vůči posunutí i vůči změně měřítka, zachová je

# míry variability

- ▶ míra variability  $\sigma(x)$  číselně charakterizuje jinou vlastnost, než míry polohy, proto na poloze nesmí záviset
- ▶ ukazuje nakolik jsou zjištěné hodnoty nestejně, velikost jejich kolísání, jejich **variabilitu**
- ▶ pro dobrou míru variability  $\sigma(X)$  platí:

$$\sigma(a + X) = \sigma(X)$$

$$\sigma(b \cdot X) = b \cdot \sigma(X) \quad b > 0$$

- ▶ přičtení konstanty  $a$  míru variability nezmění, na vynásobení kladnou konstantou  $b$  reaguje

# míry variability

- ▶ míra variability  $\sigma(x)$  číselně charakterizuje jinou vlastnost, než míry polohy, proto na poloze nesmí záviset
- ▶ ukazuje nakolik jsou zjištěné hodnoty nestejně, velikost jejich kolísání, jejich **variabilitu**
- ▶ pro dobrou míru variability  $\sigma(X)$  platí:

$$\sigma(a + X) = \sigma(X)$$

$$\sigma(b \cdot X) = b \cdot \sigma(X) \quad b > 0$$

- ▶ přičtení konstanty  $a$  míru variability nezmění, na vynásobení kladnou konstantou  $b$  reaguje



## míry variability

- ▶ míra variability  $\sigma(x)$  číselně charakterizuje jinou vlastnost, než míry polohy, proto na poloze nesmí záviset
- ▶ ukazuje nakolik jsou zjištěné hodnoty nestejně, velikost jejich kolísání, jejich **variabilitu**
- ▶ pro dobrou míru variability  $\sigma(X)$  platí:

$$\sigma(a + X) = \sigma(X)$$

$$\sigma(b \cdot X) = b \cdot \sigma(X) \quad b > 0$$

- ▶ přičtení konstanty  $a$  míru variability nezmění, na vynásobení kladnou konstantou  $b$  reaguje

## míry variability

- ▶ míra variability  $\sigma(x)$  číselně charakterizuje jinou vlastnost, než míry polohy, proto na poloze nesmí záviset
- ▶ ukazuje nakolik jsou zjištěné hodnoty nestejně, velikost jejich kolísání, jejich **variabilitu**
- ▶ pro dobrou míru variability  $\sigma(X)$  platí:

$$\sigma(a + X) = \sigma(X)$$

$$\sigma(b \cdot X) = b \cdot \sigma(X) \quad b > 0$$

- ▶ přičtení konstanty  $a$  míru variability nezmění, na vynásobení kladnou konstantou  $b$  reaguje

## směrodatná odchylka, rozptyl

- **rozptyl** (variance,  $s_{b.x}^2 = b^2 s_x^2$ )

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

[var(x)]

- např. pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme  $\bar{x} = 10$ , tedy

$$s_x^2 = \frac{1}{7-1} ((4-10)^2 + (5-10)^2 + \dots + (21-10)^2) = \frac{196}{6}$$

- **směrodatná odchylka** [standard deviation]

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

[sd(x)]

## směrodatná odchylka, rozptyl

- ▶ **rozptyl** (variance,  $s_{b.x}^2 = b^2 s_x^2$ )

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

[var(x)]

- ▶ např. pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme  $\bar{x} = 10$ , tedy

$$s_x^2 = \frac{1}{7-1} ((4-10)^2 + (5-10)^2 + \dots + (21-10)^2) = \frac{196}{6}$$

- ▶ **směrodatná odchylka** [standard deviation]

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

[sd(x)]

## směrodatná odchylka, rozptyl

- **rozptyl** (variance,  $s_{b \cdot x}^2 = b^2 s_x^2$ )

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

[var(x)]

- např. pro data: 4, 5, 8, 9, 10, 13, 21 dostaneme  $\bar{x} = 10$ , tedy

$$s_x^2 = \frac{1}{7-1} ((4-10)^2 + (5-10)^2 + \dots + (21-10)^2) = \frac{196}{6}$$

- **směrodatná odchylka** [standard deviation]

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

[sd(x)]