

Statistika

(MD360P03Z, MD360P03U)
ak. rok 2007/2008

Karel Zvára

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

2. října 2007



literatura

- ▶ K. Zvára: Biostatistika, Karolinum Praha, 1998, 2000, 2001, 2003, 2006
- ▶ Z. Pavlík, K. Kühnl: Úvod do kvantitativních metod pro geografy, SPN Praha, 1981
- ▶ T. H. Wonnacot, R. J. Wonnacot: Statistika pro obchod a hospodářství, Victoria Publishing Praha, 1992
- ▶ slajdy přednášky na adrese <http://www.karlin.mff.cuni.cz/~zvara> (celý semestr, může dojít k úpravám)
- ▶ postupně doplňované slajdy uskutečněných přednášek

literatura

- ▶ K. Zvára: Biostatistika, Karolinum Praha, 1998, 2000, 2001, 2003, 2006
- ▶ Z. Pavlík, K. Kühnl: Úvod do kvantitativních metod pro geografy, SPN Praha, 1981
- ▶ T. H. Wonnacot, R. J. Wonnacot: Statistika pro obchod a hospodářství, Victoria Publishing Praha, 1992
- ▶ slajdy přednášky na adrese <http://www.karlin.mff.cuni.cz/~zvara> (celý semestr, může dojít k úpravám)
- ▶ postupně doplňované slajdy uskutečněných přednášek

literatura

- ▶ K. Zvára: Biostatistika, Karolinum Praha, 1998, 2000, 2001, 2003, 2006
- ▶ Z. Pavlík, K. Kühnl: Úvod do kvantitativních metod pro geografy, SPN Praha, 1981
- ▶ T. H. Wonnacot, R. J. Wonnacot: Statistika pro obchod a hospodářství, Victoria Publishing Praha, 1992
- ▶ slajdy přednášky na adrese <http://www.karlin.mff.cuni.cz/~zvara> (celý semestr, může dojít k úpravám)
- ▶ postupně doplňované slajdy uskutečněných přednášek

literatura

- ▶ K. Zvára: Biostatistika, Karolinum Praha, 1998, 2000, 2001, 2003, 2006
- ▶ Z. Pavlík, K. Kühnl: Úvod do kvantitativních metod pro geografy, SPN Praha, 1981
- ▶ T. H. Wonnacot, R. J. Wonnacot: Statistika pro obchod a hospodářství, Victoria Publishing Praha, 1992
- ▶ slajdy přednášky na adrese <http://www.karlin.mff.cuni.cz/~zvara> (celý semestr, může dojít k úpravám)
- ▶ postupně doplňované slajdy uskutečněných přednášek

literatura

- ▶ K. Zvára: Biostatistika, Karolinum Praha, 1998, 2000, 2001, 2003, 2006
- ▶ Z. Pavlík, K. Kühnl: Úvod do kvantitativních metod pro geografy, SPN Praha, 1981
- ▶ T. H. Wonnacot, R. J. Wonnacot: Statistika pro obchod a hospodářství, Victoria Publishing Praha, 1992
- ▶ slajdy přednášky na adrese <http://www.karlin.mff.cuni.cz/~zvara> (celý semestr, může dojít k úpravám)
- ▶ postupně doplňované slajdy uskutečněných přednášek

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) \Rightarrow zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) ⇒ zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) \Rightarrow zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) \Rightarrow zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

cvičení, zápočet, zkouška

- ▶ cvičení v počítačových učebnách
 - ▶ PUA (suterén Albertov 6)
 - ▶ Z3 (Albertov 6, u schodů do suterénu)
 - ▶ B5 (Viničná 7, 1. patro)
- ▶ MS Excel
- ▶ volně šiřitelný program R (<http://cran.r-project.org/>)
- ▶ (aktivní účast na cvičení, maximálně dvě absence) & (napsání zápočtového testu) \Rightarrow zápočet
- ▶ obsah cvičení více přizpůsoben studovanému oboru
- ▶ přednášky formulovány obecněji
- ▶ zkouška nejspíš písemná, kombinovaná s ústní, zápočet **musí** zkoušce **předcházet**; přihlašování ke zkoušce přes SIS

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

přehled témat

- ▶ popisná statistika (měřítka, charakteristiky polohy, variability, souvislost znaků)
- ▶ statistika v geografických/demografických/sociálních vědách
- ▶ pravděpodobnost (základní kombinatorické pojmy, klasická definice, podmíněná pravděpodobnost, nezávislost)
- ▶ náhodná veličina (rozdělení, střední hodnota, rozptyl, hustota, distribuční funkce)
- ▶ důležitá rozdělení (normální, binomické, Poissonovo)
- ▶ statistické usuzování (populace a výběr, parametry a jejich odhady, interval spolehlivosti, volba rozsahu výběru)
- ▶ testování hypotéz (chyba 1. druhu, 2. druhu, hladina testu, síla testu, p -hodnota)
- ▶ testy (o populačním průměru, populačním podílu či podílech, nezávislosti, regresních koeficientech)
- ▶ regrese, kontingenční (čtyřpolní) tabulky

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, ..., 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, ..., 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, ..., 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování I

- ▶ zjišťování se týká 200 mužů středního věku
- ▶ v souboru je 80 kuřáků a 120 nekuřáků
- ▶ 85 mužů má oči modré, 25 hnědé, 90 jiné barvy
- ▶ 27 mužů má jen základní vzdělání, 44 neúplné střední, 65 maturitu, 64 vysokoškolské
- ▶ 22 se jich narodilo v roce 1942, 19 v roce 1943, 25 v roce 1944, . . . , 18 v roce 1951
- ▶ hmotnosti jednotlivých mužů jsou 83, 92, . . . , 63 kg
- ▶ výška jednotlivých mužů jsou 172, 176, . . . , 178 cm
- ▶ Co mají tyto údaje společného? Čím se údaje v jednotlivých podskupinách liší? Souvisí kouření a vzdělání? Souvisí příjem se vzděláním? Je tato souvislost stejná, jako v zemi XY?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

příklad statistického zjišťování II

- ▶ zjišťování se týká příjmů obyvatel
- ▶ hodnotíme hrubý příjem za rok
- ▶ přihlížíme k místu trvalého bydliště (velikost obce, který kraj)
- ▶ přihlížíme k vzdělání (druh, doba školní docházky)
- ▶ přihlížíme k věku a pohlaví
- ▶ Co mají tyto údaje společného? Čím se údaje liší?

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole . . .)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole ...)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole . . .)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole . . .)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole . . .)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

co a jak měříme (zjišťujeme)

- ▶ měříme na mnoha **statistických jednotkách** (osoba, domácnost, obec, okres, stát, pokusné pole . . .)
- ▶ měříme (zjišťujeme) hodnoty **znaků**
- ▶ zjištěnou hodnotu vyjadřujeme ve zvoleném **měřítku** (stupnici)
- ▶ na jedné jednotce můžeme měřit několik znaků (závislost)
- ▶ měříme na skupinách jednotek – **souborech**
- ▶ zajímají nás **hromadné** vlastnosti ve velkých souborech
- ▶ můžeme **porovnávat** vlastnosti znaku **mezi soubory**

měřítko

- ▶ **nula-jedničkové** (muž/žena, kuřák/nekuřák)
- ▶ **nominální** (země původu, barva očí) jednoznačně dané hodnoty
- ▶ **ordinální** (dosažené vzdělání, stupeň bolesti) jednoznačně dané hodnoty, možné hodnoty jsou *uspořádané*
- ▶ **intervalové** (teplota v Celsiově stupnici, rok narození) konstantní vzdálenosti mezi sousedními hodnotami, nula jen konvence; *o kolik* stupňů je je dnes tepleji, než bylo vloni?
- ▶ **poměrové** (hmotnost, výška, HDP, počet obyvatel, věk) násobek zvolené jednotky, nula = neexistence měřené vlastnosti *kolikrát* je A starší (vyšší ...) než B

měřítko

- ▶ **nula-jedničkové** (muž/žena, kuřák/nekuřák)
- ▶ **nominální** (země původu, barva očí) jednoznačně dané hodnoty
- ▶ **ordinální** (dosažené vzdělání, stupeň bolesti) jednoznačně dané hodnoty, možné hodnoty jsou *uspořádané*
- ▶ **intervalové** (teplota v Celsiově stupnici, rok narození) konstantní vzdálenosti mezi sousedními hodnotami, nula jen konvence; *o kolik* stupňů je je dnes tepleji, než bylo vloni?
- ▶ **poměrové** (hmotnost, výška, HDP, počet obyvatel, věk) násobek zvolené jednotky, nula = neexistence měřené vlastnosti *kolikrát* je A starší (vyšší ...) než B

měřítko

- ▶ **nula-jedničkové** (muž/žena, kuřák/nekuřák)
- ▶ **nominální** (země původu, barva očí) jednoznačně dané hodnoty
- ▶ **ordinální** (dosažené vzdělání, stupeň bolesti) jednoznačně dané hodnoty, možné hodnoty jsou *uspořádané*
- ▶ **intervalové** (teplota v Celsiově stupnici, rok narození) konstantní vzdálenosti mezi sousedními hodnotami, nula jen konvence; *o kolik* stupňů je je dnes tepleji, než bylo vloni?
- ▶ **poměrové** (hmotnost, výška, HDP, počet obyvatel, věk) násobek zvolené jednotky, nula = neexistence měřené vlastnosti *kolikrát* je A starší (vyšší ...) než B

měřítko

- ▶ **nula-jedničkové** (muž/žena, kuřák/nekuřák)
- ▶ **nominální** (země původu, barva očí) jednoznačně dané hodnoty
- ▶ **ordinální** (dosažené vzdělání, stupeň bolesti) jednoznačně dané hodnoty, možné hodnoty jsou *uspořádané*
- ▶ **intervalové** (teplota v Celsiově stupnici, rok narození) konstantní vzdálenosti mezi sousedními hodnotami, nula jen konvence; o *kolik* stupňů je je dnes tepleji, než bylo vloni?
- ▶ **poměrové** (hmotnost, výška, HDP, počet obyvatel, věk) násobek zvolené jednotky, nula = neexistence měřené vlastnosti *kolikrát* je A starší (vyšší ...) než B

měřítka

- ▶ **nula-jedničkové** (muž/žena, kuřák/nekuřák)
- ▶ **nominální** (země původu, barva očí) jednoznačně dané hodnoty
- ▶ **ordinální** (dosažené vzdělání, stupeň bolesti) jednoznačně dané hodnoty, možné hodnoty jsou *uspořádané*
- ▶ **intervalové** (teplota v Celsiově stupnici, rok narození) konstantní vzdálenosti mezi sousedními hodnotami, nula jen konvence; o *kolik* stupňů je je dnes tepleji, než bylo vloni?
- ▶ **poměrové** (hmotnost, výška, HDP, počet obyvatel, věk) násobek zvolené jednotky, nula = neexistence měřené vlastnosti *kolikrát* je A starší (vyšší . . .) než B

měřítko (stručnější dělení)

- ▶ **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 (není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ zařazení znaku k určitému měřítku může záviset na účelu šetření

měřítko (stručnější dělení)

- ▶ **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 (není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ zařazení znaku k určitému měřítku může záviset na účelu šetření

měřítko (stručnější dělení)

- ▶ **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 (není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ zařazení znaku k určitému měřítku může záviset na účelu šetření

měřítko (stručnější dělení)

- ▶ **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 (není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ zařazení znaku k určitému měřítku může záviset na účelu šetření

měřítko (stručnějšší dělení)

- ▶ **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 (není spojité)
- ▶ hodnoty kvantitativních – čísla
- ▶ zařazení znaku k určitému měřítku může záviset na účelu šetření

veličina

- ▶ číselně vyjádřený výsledek měření
- ▶ *hodnoty* znaků v intervalovém, poměrovém měřítku jsou husté – **spojitá veličina**
- ▶ *četnosti hodnot* znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ pro veličiny máme charakteristiky některých jejich hromadných vlastností (**charakteristiky polohy, variability, tvaru rozdělení**)
- ▶ popisné charakteristiky (statistiky) mají jedním číslem vyjádřit danou vlastnost

veličina

- ▶ číselně vyjádřený výsledek měření
- ▶ *hodnoty* znaků v intervalovém, poměrovém měřítku jsou husté – **spojitá veličina**
- ▶ *četnosti hodnot* znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ pro veličiny máme charakteristiky některých jejich hromadných vlastností (**charakteristiky polohy, variability, tvaru rozdělení**)
- ▶ popisné charakteristiky (statistiky) mají jedním číslem vyjádřit danou vlastnost

veličina

- ▶ číselně vyjádřený výsledek měření
- ▶ *hodnoty* znaků v intervalovém, poměrovém měřítku jsou husté – **spojitá veličina**
- ▶ *četnosti hodnot* znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ pro veličiny máme charakteristiky některých jejich hromadných vlastností (**charakteristiky polohy, variability, tvaru rozdělení**)
- ▶ popisné charakteristiky (statistiky) mají jedním číslem vyjádřit danou vlastnost

veličina

- ▶ číselně vyjádřený výsledek měření
- ▶ *hodnoty* znaků v intervalovém, poměrovém měřítku jsou husté – **spojitá veličina**
- ▶ *četnosti hodnot* znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ pro veličiny máme charakteristiky některých jejich hromadných vlastností (**charakteristiky polohy, variability, tvaru rozdělení**)
- ▶ popisné charakteristiky (statistiky) mají jedním číslem vyjádřit danou vlastnost

veličina

- ▶ číselně vyjádřený výsledek měření
- ▶ *hodnoty* znaků v intervalovém, poměrovém měřítku jsou husté – **spojitá veličina**
- ▶ *četnosti hodnot* znaků v nula-jedničkovém, nominálním (či ordinálním) měřítku – **diskrétní veličina**
- ▶ pro veličiny máme charakteristiky některých jejich hromadných vlastností (**charakteristiky polohy, variability, tvaru rozdělení**)
- ▶ popisné charakteristiky (statistiky) mají jedním číslem vyjádřit danou vlastnost

příklad: 100 hodů kostkou

počty puntíků coby různé obrázky – nominální znak

kostka A										kostka B									
4	2	5	6	3	1	1	2	2	2	1	4	6	2	3	2	6	1	5	2
2	4	5	3	1	1	3	5	5	5	5	6	5	5	6	4	2	4	5	6
4	3	2	5	5	5	2	2	5	2	3	6	3	6	5	6	1	3	5	1
2	6	5	5	2	3	6	6	4	6	6	6	2	1	1	2	6	3	2	3
5	4	1	4	2	2	4	5	2	5	4	4	1	6	6	2	6	3	2	6
5	5	3	3	5	3	6	6	6	5	2	6	1	2	6	1	5	5	6	5
3	5	4	5	1	1	4	3	2	4	6	6	5	1	6	6	6	1	2	6
1	2	4	6	6	3	4	6	1	2	6	2	5	6	2	6	6	5	6	4
6	6	1	2	6	2	4	3	2	3	6	1	2	6	2	1	6	6	6	6
1	1	6	5	2	6	4	4	6	3	6	5	1	5	6	6	1	6	6	6

příklad: 100 hodů kostkou

počty puntíků coby různé obrázky – nominální znak

kostka A										kostka B									
4	2	5	6	3	1	1	2	2	2	1	4	6	2	3	2	6	1	5	2
2	4	5	3	1	1	3	5	5	5	5	6	5	5	6	4	2	4	5	6
4	3	2	5	5	5	2	2	5	2	3	6	3	6	5	6	1	3	5	1
2	6	5	5	2	3	6	6	4	6	6	6	2	1	1	2	6	3	2	3
5	4	1	4	2	2	4	5	2	5	4	4	1	6	6	2	6	3	2	6
5	5	3	3	5	3	6	6	6	5	2	6	1	2	6	1	5	5	6	5
3	5	4	5	1	1	4	3	2	4	6	6	5	1	6	6	6	1	2	6
1	2	4	6	6	3	4	6	1	2	6	2	5	6	2	6	6	5	6	4
6	6	1	2	6	2	4	3	2	3	6	1	2	6	2	1	6	6	6	6
1	1	6	5	2	6	4	4	6	3	6	5	1	5	6	6	1	6	6	6

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

hody kostkou jako hromadný jev

- ▶ chceme 100 zjištěných hodnot (počtů puntíků) vyjádřit názorně, aby vypovídaly o vlastnostech kostky
- ▶ n_j (absolutní) **četnost** [frequency] hodnoty – kolikrát nastala
- ▶ $f_j = \frac{n_j}{n}$ **relativní četnost** hodnoty (lze vyjádřit v %)
– v jakém dílu měření nastala (nutně platí
$$n = n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j$$
)
- ▶ tabulka četností (absolutních, relativních)
- ▶ grafické vyjádření četností – **histogram** [histogram] (velikost plochy je úměrná četnosti)
- ▶ rozhodování o kvalitě kostky (zda je symetrická) je úlohou **statistické indukce** [inference] – později

zpracování četností (kostka A)

j	n_j	$f_j = n_j/n$
1	12	0,12
2	21	0,21
3	14	0,14
4	15	0,15
5	21	0,21
6	17	0,17
	<hr/>	
	$n = 100$	1,00

zpracování četností (kostka A)

j		n_j	$f_j = n_j/n$
1	### ###)	12	0,12
2	### ### ### ###)	21	0,21
3	### ###)	14	0,14
4	### ### ###	15	0,15
5	### ### ### ###)	21	0,21
6	### ### ###)	17	0,17
		<hr/>	
		$n = 100$	1,00

zpracování četností (kostka A)

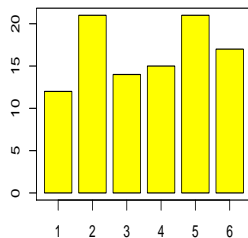
j		n_j	$f_j = n_j/n$
1	### ###)	12	0,12
2	### ### ### ###)	21	0,21
3	### ###)	14	0,14
4	### ### ###	15	0,15
5	### ### ### ###)	21	0,21
6	### ### ###)	17	0,17
		<hr/>	
		$n = 100$	1,00

zpracování četností (kostka A)

j		n_j	$f_j = n_j/n$
1	### ###)	12	0,12
2	### ### ### ###)	21	0,21
3	### ###)	14	0,14
4	### ### ###	15	0,15
5	### ### ### ###)	21	0,21
6	### ### ###)	17	0,17
		<hr/>	
		$n = 100$	1,00

zpracování četností (kostka A)

j		n_j	$f_j = n_j/n$
1	### ###)	12	0,12
2	### ### ### ###)	21	0,21
3	### ###	14	0,14
4	### ### ###	15	0,15
5	### ### ### ###)	21	0,21
6	### ### ###	17	0,17
		<hr/>	
		$n = 100$	1,00



zpracování četností (kostka B)

j	n_j	$f_j = n_j/n$
1	15	0,15
2	16	0,16
3	7	0,07
4	6	0,06
5	15	0,15
6	41	0,41
	<hr/>	
	$n = 100$	

zpracování četností (kostka B)

j		n_j	$f_j = n_j/n$
1	### ##	15	0,15
2	### ##	16	0,16
3	##	7	0,07
4	##	6	0,06
5	### ##	15	0,15
6	### ## ### ##	41	0,41
		<hr/>	
		$n = 100$	

zpracování četností (kostka B)

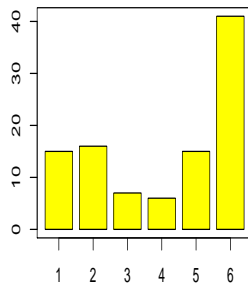
j		n_j	$f_j = n_j/n$
1	### ##	15	0,15
2	### ##	16	0,16
3	##	7	0,07
4	##	6	0,06
5	### ##	15	0,15
6	### ## ### ##	41	0,41
		<hr/>	
		$n = 100$	

zpracování četností (kostka B)

j		n_j	$f_j = n_j/n$
1	### ##	15	0,15
2	### ##	16	0,16
3	##	7	0,07
4	##	6	0,06
5	### ##	15	0,15
6	### ##	41	0,41
		<hr/>	
		$n = 100$	

zpracování četností (kostka B)

j		n_j	$f_j = n_j/n$
1	### ##	15	0,15
2	### ##	16	0,16
3	##	7	0,07
4	###	6	0,06
5	### ##	15	0,15
6	### ##	41	0,41
		<hr/>	
		$n = 100$	



příklad: věk 99 matek

99 zjištěných hodnot – soubor naměřených hodnot

26	35	21	25	27	24	24	30	23	18
35	21	25	26	26	19	29	22	21	27
26	30	28	28	27	29	27	26	21	23
24	21	28	25	34	24	21	28	25	28
22	26	32	22	32	25	21	25	24	32
24	22	31	33	23	30	26	27	25	24
24	23	25	23	26	28	24	25	25	26
28	28	22	23	20	20	21	31	24	21
29	28	26	38	20	23	25	37	33	23
27	23	21	25	21	33	22	29	21	

▶ Jdi k variační řadě

variační řada, pořadí

- ▶ x_1, x_2, \dots, x_n původní (neuspořádaná) data – hodnoty znaku v měřítku aspoň ordinálním uvedené v původním pořadí, bez ohledu na případná opakování
- ▶ **variační řada** $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ [sort(x)]
data uspořádána tak, aby hodnoty neklesaly
- ▶ proto **závorky u indexů**
- ▶ **pořadí** [rank] – umístění pozorování ve variační řadě; shodným hodnotám dáváme průměrné pořadí [rank(x)]

x_j	22	15	17	15	21	13	18
pořadí R_j	7	2,5	4	2,5	6	1	5

variační řada, pořadí

- ▶ x_1, x_2, \dots, x_n původní (neuspořádaná) data – hodnoty znaku v měřítku aspoň ordinálním uvedené v původním pořadí, bez ohledu na případná opakování
- ▶ **variační řada** $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ [sort(x)]
data uspořádána tak, aby hodnoty neklesaly
- ▶ proto **závorky u indexů**
- ▶ **pořadí** [rank] – umístění pozorování ve variační řadě; shodným hodnotám dáváme průměrné pořadí [rank(x)]

x_j	22	15	17	15	21	13	18
pořadí R_j	7	2,5	4	2,5	6	1	5

variační řada, pořadí

- ▶ x_1, x_2, \dots, x_n původní (neuspořádaná) data – hodnoty znaku v měřítku aspoň ordinálním uvedené v původním pořadí, bez ohledu na případná opakování
- ▶ **variační řada** $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ [sort(x)]
data uspořádána tak, aby hodnoty neklesaly
- ▶ proto **závorky u indexů**
- ▶ **pořadí** [rank] – umístění pozorování ve variační řadě; shodným hodnotám dáváme průměrné pořadí [rank(x)]

x_j	22	15	17	15	21	13	18
pořadí R_j	7	2,5	4	2,5	6	1	5

variační řada, pořadí

- ▶ x_1, x_2, \dots, x_n původní (neuspořádaná) data – hodnoty znaku v měřítku aspoň ordinálním uvedené v původním pořadí, bez ohledu na případná opakování
- ▶ **variační řada** $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ [sort(x)]
data uspořádána tak, aby hodnoty neklesaly
- ▶ proto **závorky u indexů**
- ▶ **pořadí** [rank] – umístění pozorování ve variační řadě; shodným hodnotám dáváme průměrné pořadí [rank(x)]

x_j	22	15	17	15	21	13	18
pořadí R_j	7	2,5	4	2,5	6	1	5

příklad: věk 99 matek – variační řada

uspořádaný soubor hodnot – variační řada

18	19	20	20	20	21	21	21	21	21
21	21	21	21	21	21	21	22	22	22
22	22	22	23	23	23	23	23	23	23
23	23	24	24	24	24	24	24	24	24
24	24	25	25	25	25	25	25	25	25
25	25	25	25	26	26	26	26	26	26
26	26	26	26	27	27	27	27	27	27
28	28	28	28	28	28	28	28	28	29
29	29	29	30	30	30	31	31	32	32
32	33	33	33	34	35	35	37	38	

▶ Jdi k původním pozorováním

třídění, třídí četnosti

- ▶ spojitá veličina s velkým počtem naměřených hodnot
- ▶ obor hodnot rozdělíme na nepřekrývající se třídy (intervaly), nejlépe stejné délky (ne vždy je to praktické či možné)
- ▶ všechna pozorování z daného intervalu nahradíme zástupnou hodnotou (zpravidla středem intervalu) x_j^*
- ▶ zjistíme (**absolutní**) četnosti n_1, \dots, n_k jednotlivých tříd
- ▶ **kumulativní četnosti** udávají počet hodnot v dané třídě a třídách předcházejících ($1 \leq j \leq k$) [cumsum()]

$$N_j = n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i$$

třídění, třídí četnosti

- ▶ spojitá veličina s velkým počtem naměřených hodnot
- ▶ obor hodnot rozdělíme na nepřekrývající se třídy (intervaly), nejlépe stejné délky (ne vždy je to praktické či možné)
- ▶ všechna pozorování z daného intervalu nahradíme zástupnou hodnotou (zpravidla středem intervalu) x_j^*
- ▶ zjistíme (**absolutní**) četnosti n_1, \dots, n_k jednotlivých tříd
- ▶ **kumulativní četnosti** udávají počet hodnot v dané třídě a třídách předcházejících ($1 \leq j \leq k$) [cumsum()]

$$N_j = n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i$$

třídění, třídí četnosti

- ▶ spojitá veličina s velkým počtem naměřených hodnot
- ▶ obor hodnot rozdělíme na nepřekrývající se třídy (intervaly), nejlépe stejné délky (ne vždy je to praktické či možné)
- ▶ všechna pozorování z daného intervalu nahradíme zástupnou hodnotou (zpravidla středem intervalu) x_j^*
- ▶ zjistíme (**absolutní**) četnosti n_1, \dots, n_k jednotlivých tříd
- ▶ **kumulativní četnosti** udávají počet hodnot v dané třídě a třídách předcházejících ($1 \leq j \leq k$) [cumsum()]

$$N_j = n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i$$

třídění, třídí četnosti

- ▶ spojitá veličina s velkým počtem naměřených hodnot
- ▶ obor hodnot rozdělíme na nepřekrývající se třídy (intervaly), nejlépe stejné délky (ne vždy je to praktické či možné)
- ▶ všechna pozorování z daného intervalu nahradíme zástupnou hodnotou (zpravidla středem intervalu) x_j^*
- ▶ zjistíme (**absolutní**) **četnosti** n_1, \dots, n_k jednotlivých tříd
- ▶ **kumulativní četnosti** udávají počet hodnot v dané třídě a třídách předcházejících ($1 \leq j \leq k$) [cumsum()]

$$N_j = n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i$$

třídění, třídí četnosti

- ▶ spojitá veličina s velkým počtem naměřených hodnot
- ▶ obor hodnot rozdělíme na nepřekrývající se třídy (intervaly), nejlépe stejné délky (ne vždy je to praktické či možné)
- ▶ všechna pozorování z daného intervalu nahradíme zástupnou hodnotou (zpravidla středem intervalu) x_j^*
- ▶ zjistíme (**absolutní**) **četnosti** n_1, \dots, n_k jednotlivých tříd
- ▶ **kumulativní četnosti** udávají počet hodnot v dané třídě a třídách předcházejících ($1 \leq j \leq k$) [cumsum()]

$$N_j = n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i$$

věk matek – třídní četnosti

 $k = 7$

interval	x_j^*	n_j	$f_j = n_j/n$	N_j	N_j/n
do 20	19	5	0,051	5	0,051
21 až 23	22	27	0,273	32	0,324
24 až 26	25	32	0,322	64	0,646
27 až 29	28	19	0,192	83	0,838
30 až 32	31	8	0,081	91	0,919
33 až 35	34	6	0,061	97	0,980
36 a více	37	2	0,020	99	1,000

▶ Jdi k histogramu věku matek

▶ Jdi k mírám polohy věku matek

grafické znázornění třídnicích četností

- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

grafické znázornění třídnicích četností

- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

grafické znázornění třídnicích četností

- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

grafické znázornění třídnicích četností

- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

grafické znázornění třídnicích četností

- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

grafické znázornění třídnicích četností

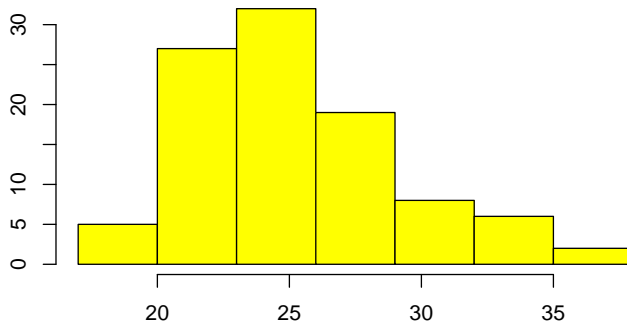
- ▶ **histogram** je založen na třídění do intervalů, výjimečně zobrazuje přímo četnosti jednotlivých hodnot (barplot) `[hist()]`
- ▶ každé třídě odpovídá obdélník o **ploše úměrné četnosti** (absolutní nebo relativní)
- ▶ při stejných šířkách intervalů h odpovídají četnostem výšky obdélníků (protože základny jsou stejně dlouhé)
- ▶ počet intervalů k : volí se 5–15 tak, aby středy byly okrouhlé
- ▶ pomůckou Sturgesovo pravidlo

$$k \approx 1 + 3,3 \cdot \log_{10} n = 1 + \log_2 n$$

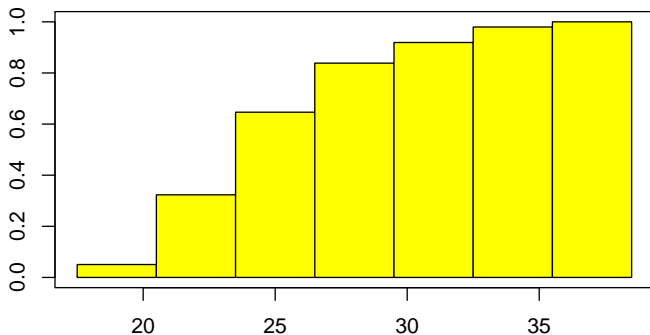
- ▶ příklad věk matek: $k \approx 1 + 3,3 \cdot \log_{10} 99 \approx 7,6$

příklad (věk matek): histogram, $h = 3$ ($k = 7$)

`[hist(vek.m,seq(17,38,by=3),col="yellow")]`



příklad (věk matek): kumulativní relativní četnosti



třídění při nestejně dlouhých intervalech

- ▶ někdy jsou data nepravidelně rozmístěna
- ▶ zpravidla jsou soustředěna u levého okraje intervalu hodnot (věkové či příjmové složení obyvatelstva)
- ▶ pak vhodné zvolit nestejně dlouhé intervaly
- ▶ je vhodné zvolit délky intervalů tak, aby delší byly násobkem kratších
- ▶ při nestejně dlouhých intervalech musí zjištěné četnosti odpovídat **plocha**, nikoliv výška; pak se na svislou osu nanáší **relativní** četnosti

třídění při nestejně dlouhých intervalech

- ▶ někdy jsou data nepravidelně rozmístěna
- ▶ zpravidla jsou soustředěna u levého okraje intervalu hodnot (věkové či příjmové složení obyvatelstva)
- ▶ pak vhodné zvolit nestejně dlouhé intervaly
- ▶ je vhodné zvolit délky intervalů tak, aby delší byly násobkem kratších
- ▶ při nestejně dlouhých intervalech musí zjištěné četnosti odpovídat **plocha**, nikoliv výška; pak se na svislou osu nanáší **relativní** četnosti

třídění při nestejně dlouhých intervalech

- ▶ někdy jsou data nepravidelně rozmístěna
- ▶ zpravidla jsou soustředěna u levého okraje intervalu hodnot (věkové či příjmové složení obyvatelstva)
- ▶ pak vhodné zvolit nestejně dlouhé intervaly
- ▶ je vhodné zvolit délky intervalů tak, aby delší byly násobkem kratších
- ▶ při nestejně dlouhých intervalech musí zjištěné četnosti odpovídat **plocha**, nikoliv výška; pak se na svislou osu nanáší **relativní** četnosti

třídění při nestejně dlouhých intervalech

- ▶ někdy jsou data nepravidelně rozmístěna
- ▶ zpravidla jsou soustředěna u levého okraje intervalu hodnot (věkové či příjmové složení obyvatelstva)
- ▶ pak vhodné zvolit nestejně dlouhé intervaly
- ▶ je vhodné zvolit délky intervalů tak, aby delší byly násobkem kratších
- ▶ při nestejně dlouhých intervalech musí zjištěné četnosti odpovídat **plocha**, nikoliv výška; pak se na svislou osu nanáší **relativní** četnosti

třídění při nestejně dlouhých intervalech

- ▶ někdy jsou data nepravidelně rozmístěna
- ▶ zpravidla jsou soustředěna u levého okraje intervalu hodnot (věkové či příjmové složení obyvatelstva)
- ▶ pak vhodné zvolit nestejně dlouhé intervaly
- ▶ je vhodné zvolit délky intervalů tak, aby delší byly násobkem kratších
- ▶ při nestejně dlouhých intervalech musí zjištěné četnosti odpovídat **plocha**, nikoliv výška; pak se na svislou osu nanáší **relativní** četnosti

příklad: tolary

měsíční příjmy 99 osob v tolarech

četnosti

x_j	10	11	12	13	14	15	16	17	18	19	20		
n_j	7	14	16	10	6	3	9	3	1	5	3		
x_j	21	22	24	26	27	28	32	35	36	40	43	45	47
n_j	4	3	3	1	2	1	1	1	2	1	1	1	1

třídní četnosti

třída	10	11	12	13–16	17–20	21–30	31–50	celkem
x_j^*	10	11	12	14,5	18,5	25,5	40,5	
n_j^*	7	14	16	28	12	14	8	99
hustota	7	14	16	7	3	1,4	0,4	

[▶ Jdi k hodnocení tolary](#)

příklad: tolary

měsíční příjmy 99 osob v tolarech

četnosti

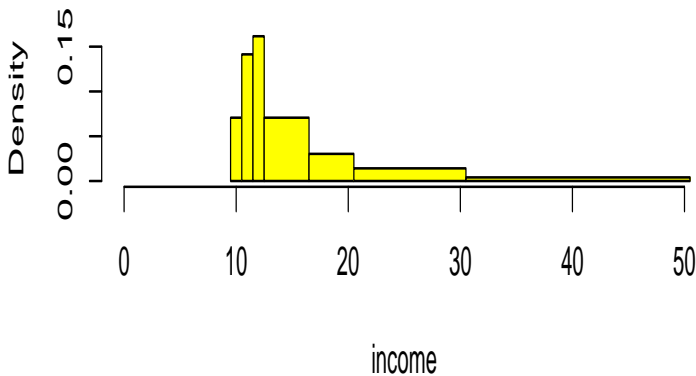
x_j	10	11	12	13	14	15	16	17	18	19	20		
n_j	7	14	16	10	6	3	9	3	1	5	3		
x_j	21	22	24	26	27	28	32	35	36	40	43	45	47
n_j	4	3	3	1	2	1	1	1	2	1	1	1	1

třídní četnosti

třída	10	11	12	13–16	17–20	21–30	31–50	celkem
x_j^*	10	11	12	14,5	18,5	25,5	40,5	
n_j^*	7	14	16	28	12	14	8	99
hustota	7	14	16	7	3	1,4	0,4	

► Jdi k hodnocení tolarů

příklad (tolary): histogram



výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7)

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7)

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7)

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7)

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7))

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7)

výběrové charakteristiky polohy: medián

snaha charakterizovat úroveň jediným číslem

- ▶ medián je číslo, které dělí data na dvě stejně velké části (větších hodnot a menších hodnot)

- ▶ **medián** [median] (prostřední hodnota) \tilde{x} [median(x)]

$$\tilde{x} = x_{(\frac{n+1}{2})} \quad \text{pro } n \text{ liché}$$

$$\tilde{x} = \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{pro } n \text{ sudé}$$

- ▶ závorky u indexů jsou nutné: znamenají, že hodnoty byly předem uspořádány do variační řady

- ▶ 5, 3, 4, 7, 6 $\tilde{x} = 5$ (3 < 4 < 5 < 6 < 7))

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[quantile(x,probs=c(1/4,3/4))]

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[quantile(x, probs=c(1/4, 3/4))]

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[quantile(x,probs=c(1/4,3/4))]

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[quantile(x,probs=c(1/4,3/4))]

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[quantile(x, probs=c(1/4,3/4))]

kvartily, percentily

- ▶ **dolní (horní) kvartil** Q_1 (Q_3) [lower (upper) quartile] vyděluje čtvrtinu nejmenších (největších) hodnot
- ▶ kvartil – speciální případ percentilu
- ▶ **percentil** [percentile] x_p vyděluje $100p$ % nejmenších hodnot od ostatních
- ▶ výpočet percentilů – mnoho vzorečků
- ▶ medián je také percentilem, totiž $x_{0,5}$
- ▶ podobně $Q_1 = x_{1/4} = x_{0,25}$, $Q_3 = x_{3/4} = x_{0,75}$
[`quantile(x, probs=c(1/4,3/4))`]

výpočet percentilů (jako v R), jen pro ilustraci

jedna z možných definic – Gumbel(1939)

- ▶ 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)
- ▶ 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

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

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

výpočet percentilů (jako v R), jen pro ilustraci jedna z možných definic – Gumbel(1939)

- ▶ 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)
- ▶ 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

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25$$
$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

výpočet percentilů (jako v R), jen pro ilustraci

jedna z možných definic – Gumbel(1939)

- ▶ 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)
- ▶ 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

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

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

výpočet percentilů (jako v R), jen pro ilustraci

jedna z možných definic – Gumbel(1939)

- ▶ 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)
- ▶ 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

$$k = \lfloor 1 + (99 - 1) \cdot 0,25 \rfloor = \lfloor 25,5 \rfloor = 25$$
$$Q_1 = x_{0,25} = 0,5 \cdot x_{(25)} + 0,5 \cdot x_{(26)}$$

příklad: věk 99 matek – variační řada

variační řada, medián $\tilde{x} = 25$

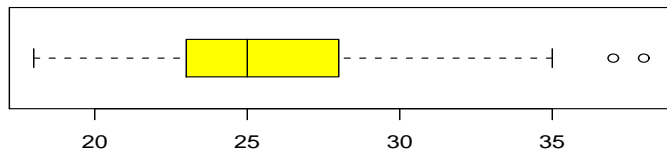
kvartily $Q_1 = (23+23)/2 = 23$, $Q_3 = (28+28)/2 = 28$

18	19	20	20	20	21	21	21	21	21
21	21	21	21	21	21	21	22	22	22
22	22	22	23	23	23	23	23	23	23
23	23	24	24	24	24	24	24	24	24
24	24	25	25	25	25	25	25	25	25
25	25	25	25	26	26	26	26	26	26
26	26	26	26	27	27	27	27	27	27
28	28	28	28	28	28	28	28	28	29
29	29	29	30	30	30	31	31	32	32
32	33	33	33	34	35	35	37	38	

▶ [Návrat míry var. věku matek](#)

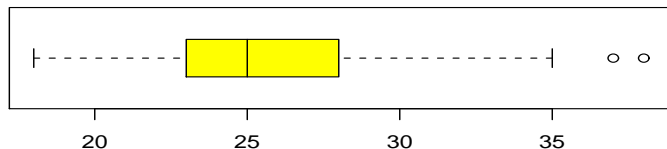
krabicový diagram

- ▶ **krabicový diagram** [box-plot] zobrazuje kvartily, medián, minimum, maximum, případně odlehlá pozorování: od bližšího kvartilu dál než $3/2 \cdot (Q_3 - Q_1)$ [boxplot(x)]
- ▶ příklad: věk matek ($Q_1 = 23$, $\bar{x} = 25$, $Q_3 = 28$, dvě odlehlá pozorování)



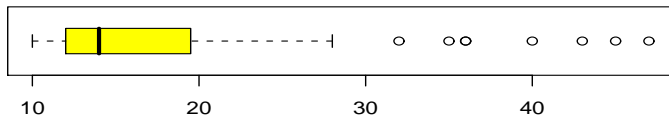
krabicový diagram

- ▶ **krabicový diagram** [box-plot] zobrazuje kvartily, medián, minimum, maximum, případně odlehlá pozorování: od bližšího kvartilu dál než $3/2 \cdot (Q_3 - Q_1)$ [boxplot(x)]
- ▶ příklad: věk matek ($Q_1 = 23$, $\bar{x} = 25$, $Q_3 = 28$, dvě odlehlá pozorování)



příklad: tolary ($\tilde{x} = 14$, $Q_1 = 12$, $Q_3 = 19,5$)

10	10	10	10	10	10	10	11	11	11
11	11	11	11	11	11	11	11	11	11
11	12	12	12	12	12	12	12	12	12
12	12	12	12	12	12	12	13	13	13
13	13	13	13	13	13	13	14	14	14
14	14	14	15	15	15	16	16	16	16
16	16	16	16	16	17	17	17	18	19
19	19	19	19	20	20	20	21	21	21
21	22	22	22	24	24	24	26	27	27
28	32	35	36	36	40	43	45	47	



průměr

- ▶ **průměr** [mean] (kdyby bylo všech n hodnot stejných)
[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**: [weighted mean] založen na četnostech

$$\bar{x} = \frac{1}{n} (n_1 x_1^* + \dots + n_k x_k^*) = \frac{1}{n} \sum_{j=1}^k n_j x_j^* = \sum_{j=1}^k \frac{n_j}{n} x_j^* = \frac{\sum_{j=1}^k n_j x_j^*}{\sum_{j=1}^k n_j}$$

- ▶ obecněji s vahami w_1, \dots, w_k hodnot x_1^*, \dots, x_k^*

$$\frac{\sum_{j=1}^k w_j x_j^*}{\sum_{j=1}^k w_j}$$

váhy musí být nezáporné ($w_j \geq 0$)

průměr

- ▶ **průměr** [mean] (kdyby bylo všech n hodnot stejných)
[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**: [weighted mean] založen na četnostech

$$\bar{x} = \frac{1}{n} (n_1 x_1^* + \dots + n_k x_k^*) = \frac{1}{n} \sum_{j=1}^k n_j x_j^* = \sum_{j=1}^k \frac{n_j}{n} x_j^* = \frac{\sum_{j=1}^k n_j x_j^*}{\sum_{j=1}^k n_j}$$

- ▶ obecněji s vahami w_1, \dots, w_k hodnot x_1^*, \dots, x_k^*

$$\frac{\sum_{j=1}^k w_j x_j^*}{\sum_{j=1}^k w_j}$$

váhy musí být nezáporné ($w_j \geq 0$)

průměr

- ▶ **průměr** [mean] (kdyby bylo všech n hodnot stejných)
[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**: [weighted mean] založen na četnostech

$$\bar{x} = \frac{1}{n} (n_1 x_1^* + \dots + n_k x_k^*) = \frac{1}{n} \sum_{j=1}^k n_j x_j^* = \sum_{j=1}^k \frac{n_j}{n} x_j^* = \frac{\sum_{j=1}^k n_j x_j^*}{\sum_{j=1}^k n_j}$$

- ▶ obecněji s vahami w_1, \dots, w_k hodnot x_1^*, \dots, x_k^*

$$\frac{\sum_{j=1}^k w_j x_j^*}{\sum_{j=1}^k w_j}$$

váhy musí být nezáporné ($w_j \geq 0$)

příklad: vážený průměr známek

předmět	známka	kredity	součin
A	1	6	6
B	1	6	6
C	2	4	8
D	3	4	12
<hr/>			
celkem	7	20	32

▶ průměr (nevážený): $\bar{x} = 7/4 = 1,75$

▶ vážený průměr (vahami kredity): $\bar{x} = 32/20 = 1,6$

příklad: vážený průměr známek

předmět	známka	kredity	součin
A	1	6	6
B	1	6	6
C	2	4	8
D	3	4	12
<hr/>			
celkem	7	20	32

- ▶ průměr (nevážený): $\bar{x} = 7/4 = 1,75$
- ▶ vážený průměr (vahami kredity): $\bar{x} = 32/20 = 1,6$

průměr pro nula-jedničkovou veličinu

- ▶ u nula-jedničkového měřítka: průměr = relativní četnost jedniček
- ▶ počet jedniček/počet všech hodnot (nul i jedniček)
- ▶ procento jedniček mezi všemi hodnotami (nulami a jedničkami)
- ▶ procento jedinců s danou vlastností
- ▶ pozor, nejde o pravděpodobnost, nanejvýš jde o její odhad!

průměr pro nula-jedničkovou veličinu

- ▶ u nula-jedničkového měřítka: průměr = relativní četnost jedniček
- ▶ počet jedniček/počet všech hodnot (nul i jedniček)
- ▶ procento jedniček mezi všemi hodnotami (nulami a jedničkami)
- ▶ procento jedinců s danou vlastností
- ▶ pozor, nejde o pravděpodobnost, nanejvýš jde o její odhad!

průměr pro nula-jedničkovou veličinu

- ▶ u nula-jedničkového měřítka: průměr = relativní četnost jedniček
- ▶ počet jedniček/počet všech hodnot (nul i jedniček)
- ▶ procento jedniček mezi všemi hodnotami (nulami a jedničkami)
- ▶ procento jedinců s danou vlastností
- ▶ pozor, nejde o pravděpodobnost, nanejvýš jde o její odhad!

průměr pro nula-jedničkovou veličinu

- ▶ u nula-jedničkového měřítka: průměr = relativní četnost jedniček
- ▶ počet jedniček/počet všech hodnot (nul i jedniček)
- ▶ procento jedniček mezi všemi hodnotami (nulami a jedničkami)
- ▶ procento jedinců s danou vlastností
- ▶ pozor, nejde o pravděpodobnost, nanejvýš jde o její odhad!

průměr pro nula-jedničkovou veličinu

- ▶ u nula-jedničkového měřítka: průměr = relativní četnost jedniček
- ▶ počet jedniček/počet všech hodnot (nul i jedniček)
- ▶ procento jedniček mezi všemi hodnotami (nulami a jedničkami)
- ▶ procento jedinců s danou vlastností
- ▶ pozor, nejde o pravděpodobnost, nanejvýš jde o její odhad!

modus

- ▶ **modus** \hat{x} [mode] nejčastější hodnota (lze počítat také pro nominální či ordinální měřítko)
- ▶ modus nemusí být určen jednoznačně, např. věk matek:

x_j^*	18	19	20	21	22	23	24	25	26	27
n_j	1	1	3	12	6	9	10	12	10	6
x_j^*	28	29	30	31	32	33	34	35	37	38
n_j	9	4	3	2	3	3	1	2	1	1

modus

- ▶ **modus** \hat{x} [mode] nejčastější hodnota (lze počítat také pro nominální či ordinální měřítko)
- ▶ modus nemusí být určen jednoznačně, např. věk matek:

x_j^*	18	19	20	21	22	23	24	25	26	27
n_j	1	1	3	12	6	9	10	12	10	6
x_j^*	28	29	30	31	32	33	34	35	37	38
n_j	9	4	3	2	3	3	1	2	1	1

příklad – věk matek

- ▶ již známe $\tilde{x} = 25$, $Q_1 = 23$, $Q_3 = 28$
- ▶ modus není určen jednoznačně: $\hat{x} = 21$, $\hat{x} = 25$
- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 35 + \dots + 21 + 23) = \frac{2544}{99} \doteq 25,7$$

- ▶ vážený průměr založený na třídění

$$\begin{aligned}\bar{x} &= \frac{5 \cdot 19 + 27 \cdot 22 + 32 \cdot 25 + 19 \cdot 28 + 8 \cdot 31 + 6 \cdot 34 + 2 \cdot 37}{5 + 27 + 32 + 19 + 8 + 6 + 2} \\ &= \frac{2547}{99} \doteq 25,7\end{aligned}$$

▶ Třídění: věk matek

příklad – věk matek

- ▶ již známe $\tilde{x} = 25$, $Q_1 = 23$, $Q_3 = 28$
- ▶ modus není určen jednoznačně: $\hat{x} = 21$, $\hat{x} = 25$
- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 35 + \dots + 21 + 23) = \frac{2544}{99} \doteq 25,7$$

- ▶ vážený průměr založený na třídění

$$\begin{aligned}\bar{x} &= \frac{5 \cdot 19 + 27 \cdot 22 + 32 \cdot 25 + 19 \cdot 28 + 8 \cdot 31 + 6 \cdot 34 + 2 \cdot 37}{5 + 27 + 32 + 19 + 8 + 6 + 2} \\ &= \frac{2547}{99} \doteq 25,7\end{aligned}$$

▶ Třídění: věk matek

příklad – věk matek

- ▶ již známe $\tilde{x} = 25$, $Q_1 = 23$, $Q_3 = 28$
- ▶ modus není určen jednoznačně: $\hat{x} = 21$, $\hat{x} = 25$
- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 35 + \dots + 21 + 23) = \frac{2544}{99} \doteq 25,7$$

- ▶ vážený průměr založený na třídění

$$\begin{aligned}\bar{x} &= \frac{5 \cdot 19 + 27 \cdot 22 + 32 \cdot 25 + 19 \cdot 28 + 8 \cdot 31 + 6 \cdot 34 + 2 \cdot 37}{5 + 27 + 32 + 19 + 8 + 6 + 2} \\ &= \frac{2547}{99} \doteq 25,7\end{aligned}$$

▶ Třídění: věk matek

příklad – věk matek

- ▶ již známe $\tilde{x} = 25$, $Q_1 = 23$, $Q_3 = 28$
- ▶ modus není určen jednoznačně: $\hat{x} = 21$, $\hat{x} = 25$
- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 35 + \dots + 21 + 23) = \frac{2544}{99} \doteq 25,7$$

- ▶ vážený průměr založený na třídění

$$\begin{aligned}\bar{x} &= \frac{5 \cdot 19 + 27 \cdot 22 + 32 \cdot 25 + 19 \cdot 28 + 8 \cdot 31 + 6 \cdot 34 + 2 \cdot 37}{5 + 27 + 32 + 19 + 8 + 6 + 2} \\ &= \frac{2547}{99} \doteq 25,7\end{aligned}$$

▶ Třídění: věk matek

příklad – tolary

- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 20 + \dots + 12 + 10) = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na četnostech jednotlivých hodnot

$$\bar{x} = \frac{7 \cdot 10 + 14 \cdot 11 + 16 \cdot 12 + \dots + 1 \cdot 47}{7 + 14 + 16 + \dots + 1} = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na třídnicích četnostech (obr. 24)

$$\begin{aligned}\bar{x} &= \frac{7 \cdot 10 + 14 \cdot 1 + 16 \cdot 12 + 28 \cdot 14,5 + \dots + 8 \cdot 40,5}{7 + 14 + 16 + 28 + 12 + 14 + 8} \\ &= \frac{1725}{99} \doteq 17,42\end{aligned}$$

- ▶ modus: $\hat{x} = 12$

příklad – tolary

- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 20 + \dots + 12 + 10) = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na četnostech jednotlivých hodnot

$$\bar{x} = \frac{7 \cdot 10 + 14 \cdot 11 + 16 \cdot 12 + \dots + 1 \cdot 47}{7 + 14 + 16 + \dots + 1} = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na třídnicích četnostech (obr. 24)

$$\begin{aligned} \bar{x} &= \frac{7 \cdot 10 + 14 \cdot 1 + 16 \cdot 12 + 28 \cdot 14,5 + \dots + 8 \cdot 40,5}{7 + 14 + 16 + 28 + 12 + 14 + 8} \\ &= \frac{1725}{99} \doteq 17,42 \end{aligned}$$

- ▶ modus: $\hat{x} = 12$

příklad – tolary

- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 20 + \dots + 12 + 10) = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na četnostech jednotlivých hodnot

$$\bar{x} = \frac{7 \cdot 10 + 14 \cdot 11 + 16 \cdot 12 + \dots + 1 \cdot 47}{7 + 14 + 16 + \dots + 1} = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na třídnicích četnostech (obr. 24)

$$\begin{aligned} \bar{x} &= \frac{7 \cdot 10 + 14 \cdot 1 + 16 \cdot 12 + 28 \cdot 14,5 + \dots + 8 \cdot 40,5}{7 + 14 + 16 + 28 + 12 + 14 + 8} \\ &= \frac{1725}{99} \doteq 17,42 \end{aligned}$$

- ▶ modus: $\hat{x} = 12$

příklad – tolary

- ▶ průměr

$$\bar{x} = \frac{1}{99} (26 + 20 + \dots + 12 + 10) = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na četnostech jednotlivých hodnot

$$\bar{x} = \frac{7 \cdot 10 + 14 \cdot 11 + 16 \cdot 12 + \dots + 1 \cdot 47}{7 + 14 + 16 + \dots + 1} = \frac{1687}{99} \doteq 17,04$$

- ▶ vážený průměr založený na třídnicích četnostech (obr. 24)

$$\begin{aligned}\bar{x} &= \frac{7 \cdot 10 + 14 \cdot 1 + 16 \cdot 12 + 28 \cdot 14,5 + \dots + 8 \cdot 40,5}{7 + 14 + 16 + 28 + 12 + 14 + 8} \\ &= \frac{1725}{99} \doteq 17,42\end{aligned}$$

- ▶ modus: $\hat{x} = 12$

useknutý průměr

také míra polohy

- ▶ **alfa-useknutý průměr** [trimmed mean]: nejprve se oddělí (usekne) 100α % nejmenších a 100α % největších hodnot, ze zbytku se spočítá průměr
- ▶ je robustní (necitlivý) vůči odlehlým hodnotám
- ▶ volí se zpravidla $\alpha = 0,1$ (0,15)
- ▶ příklad: věk matek [mean(vek.m,trim=0.1)]

$$\frac{1}{99 - 18} (x_{(10)} + x_{(11)} + \dots + x_{(89)} + x_{(90)}) = 25,3$$

useknutý průměr

také míra polohy

- ▶ **alfa-useknutý průměr** [trimmed mean]: nejprve se oddělí (usekne) 100α % nejmenších a 100α % největších hodnot, ze zbytku se spočítá průměr
- ▶ je robustní (necitlivý) vůči odlehlým hodnotám
- ▶ volí se zpravidla $\alpha = 0,1$ (0,15)
- ▶ příklad: věk matek [mean(vek.m,trim=0.1)]

$$\frac{1}{99 - 18} (x_{(10)} + x_{(11)} + \dots + x_{(89)} + x_{(90)}) = 25,3$$

useknutý průměr

také míra polohy

- ▶ **alfa-useknutý průměr** [trimmed mean]: nejprve se oddělí (usekne) 100α % nejmenších a 100α % největších hodnot, ze zbytku se spočítá průměr
- ▶ je robustní (necitlivý) vůči odlehlým hodnotám
- ▶ volí se zpravidla $\alpha = 0,1$ (0,15)
- ▶ příklad: věk matek [mean(vek.m,trim=0.1)]

$$\frac{1}{99 - 18} (x_{(10)} + x_{(11)} + \dots + x_{(89)} + x_{(90)}) = 25,3$$

useknutý průměr

také míra polohy

- ▶ **alfa-useknutý průměr** [trimmed mean]: nejprve se oddělí (usekne) 100α % nejmenších a 100α % největších hodnot, ze zbytku se spočítá průměr
- ▶ je robustní (necitlivý) vůči odlehlým hodnotám
- ▶ volí se zpravidla $\alpha = 0,1$ (0,15)
- ▶ příklad: věk matek [mean(vek.m,trim=0.1)]

$$\frac{1}{99 - 18} (x_{(10)} + x_{(11)} + \dots + x_{(89)} + x_{(90)}) = 25,3$$

příklad (věk matek): useknutý průměr (průměr počítán pouze z černých čísel)

vyloučí se $[0,1 \cdot 99] = [9,9] = 9$ ($[x]$ znamená celou část z x)
nejmenších a 9 největších hodnot

18	19	20	20	20	21	21	21	21	21
21	21	21	21	21	21	21	22	22	22
22	22	22	23	23	23	23	23	23	23
23	23	24	24	24	24	24	24	24	24
24	24	25	25	25	25	25	25	25	25
25	25	25	25	26	26	26	26	26	26
26	26	26	26	27	27	27	27	27	27
28	28	28	28	28	28	28	28	28	29
29	29	29	30	30	30	31	31	32	32
32	33	33	33	34	35	35	37	38	

vlastnosti charakteristik polohy

- ▶ změníme-li všechny hodnoty x_i tak, že přidáme ke každé stejnou konstantu a , změní se o tutéž konstantu také charakteristika polohy (posunutí)
- ▶ změníme-li všechny hodnoty x_i tak, že je vynásobíme kladnou konstantou b , toutéž konstantou musíme vynásobit původní charakteristiku polohy, abychom dostali charakteristiku polohy pro upravená data (změna měřítka)
- ▶ obecně pro míru polohy $m(x)$

$$m(a + x) = a + m(x),$$

$$m(b \cdot x) = b \cdot m(x), \quad b > 0$$

- ▶ v **obou** případech míra polohy **reaguje**

vlastnosti charakteristik polohy

- ▶ změníme-li všechny hodnoty x_i tak, že přidáme ke každé stejnou konstantu a , změní se o tutéž konstantu také charakteristika polohy (posunutí)
- ▶ změníme-li všechny hodnoty x_i tak, že je vynásobíme kladnou konstantou b , toutéž konstantou musíme vynásobit původní charakteristiku polohy, abychom dostali charakteristiku polohy pro upravená data (změna měřítka)
- ▶ obecně pro míru polohy $m(x)$

$$m(a + x) = a + m(x),$$

$$m(b \cdot x) = b \cdot m(x), \quad b > 0$$

- ▶ v **obou** případech míra polohy **reaguje**

vlastnosti charakteristik polohy

- ▶ změníme-li všechny hodnoty x_i tak, že přidáme ke každé stejnou konstantu a , změní se o tutéž konstantu také charakteristika polohy (posunutí)
- ▶ změníme-li všechny hodnoty x_i tak, že je vynásobíme kladnou konstantou b , toutéž konstantou musíme vynásobit původní charakteristiku polohy, abychom dostali charakteristiku polohy pro upravená data (změna měřítka)
- ▶ obecně pro míru polohy $m(x)$

$$m(a + x) = a + m(x),$$

$$m(b \cdot x) = b \cdot m(x), \quad b > 0$$

- ▶ v **obou** případech míra polohy **reaguje**

vlastnosti charakteristik polohy

- ▶ změníme-li všechny hodnoty x_i tak, že přidáme ke každé stejnou konstantu a , změní se o tutéž konstantu také charakteristika polohy (posunutí)
- ▶ změníme-li všechny hodnoty x_i tak, že je vynásobíme kladnou konstantou b , toutéž konstantou musíme vynásobit původní charakteristiku polohy, abychom dostali charakteristiku polohy pro upravená data (změna měřítka)
- ▶ obecně pro míru polohy $m(x)$

$$m(a + x) = a + m(x),$$

$$m(b \cdot x) = b \cdot m(x), \quad b > 0$$

- ▶ v **obou** případech míra polohy **reaguje**