Základy biostatistiky

(MD710P09)
ak. rok 2007/2008

Karel Zvára

karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 1. dubna 2008)

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \ldots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \ldots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme, ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \ldots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \ldots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme, ale **zda** jsou **co do polohy stejné**, nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \ldots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \ldots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme, ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \dots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \dots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme,
  ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \dots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \dots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme, ale **zda** jsou **co do polohy stejné**, nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \ldots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \ldots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme, ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \dots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \dots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme,
  ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párové testy
(převedou se na jednovýběrové)

- $(U_1, V_1), \ldots, (U_n, V_n)$ – párová pozorování
  nezávislé dvojice (možná) závislých náhodných veličin
- těsná závislost uvnitř dvojic je výhodná
- $X_i = U_i - V_i$ (označení rozdílů)
- předpokládáme **stejné rozdělení** $X_1, \ldots, X_n$
- U_i, V_i – dvojice měření na stejných jedincích, např. hodnota zjištěná před ošetřením a po něm
- např. věk otce a jeho syna nebo věk otce a věk matky
- **nezajímá nás** zda je mezi nimi **závislost**, tu připouštíme,
  ale **zda** jsou **co do polohy stejné**,
  nebo např. synové v (populačním) průměru vyšší, než otcové
- H_0 tvrdí, že např. mezi výškami otců a synů **není rozdíl**, tedy že rozdíly X_i **kolísají kolem nuly**

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \frac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X}}{S}\sqrt{n} = \frac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \dfrac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X}}{S}\sqrt{n} = \dfrac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \dfrac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X}}{S}\sqrt{n} = \dfrac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \dfrac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X}}{S}\sqrt{n} = \dfrac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \dfrac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X}}{S}\sqrt{n} = \dfrac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test
předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \dfrac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X}}{S}\sqrt{n} = \dfrac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test
předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ **nezávislé**
- odhad σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$
- $T = \frac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X}}{S}\sqrt{n} = \frac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

párový t-test

předpoklad normálního rozdělení rozdílů

- **normální** rozdělení: $X_i = U_i - V_i \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**
- odhad σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- $T = \frac{\bar{X}}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X}}{S}\sqrt{n} = \frac{\bar{U} - \bar{V}}{\mathrm{S.E.}(\bar{U} - \bar{V})}$
- $\mathrm{H}_0 : \mu = 0$ (pak je $\mu_U = \mu_V$)
- ve prospěch $\mathrm{H}_1 : \mu \neq 0$, když $|T| \geq t_{n-1}(\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu < 0$, když $T \leq -t_{n-1}(2\alpha)$
- ve prospěch $\mathrm{H}_1 : \mu > 0$, když $T \geq t_{n-1}(2\alpha)$
- vlastně jednovýběrový t-test pro $X_i = U_i - V_i$

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05$, $H_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99$, $\bar{u} = 179{,}267$, $\bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293$, $s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = P(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

příklad: výšky rodičů (párová pozorování!)

- U – výška otce, V – výška matky
- $\alpha = 0{,}05, \mathrm{H}_0 : \mu_U - 10 = \mu_V$ resp. $\mu_U - \mu_V = 10$
- $n = 99, \bar{u} = 179{,}267, \bar{v} = 166{,}970$
- $\bar{x} = \bar{u} - \bar{v} - 10 = 2{,}293, s_X = s_{U-10-V} = s_{U-V} = 8{,}144$
- $t = \frac{2{,}293}{8{,}144}\sqrt{99} = 2{,}801$, tedy $|t| > t_{98}(0{,}05) = 1{,}9845 \quad \Rightarrow$ zamítnout H_0
- $p = \mathrm{P}(|T| \geq t) = 0{,}0061 \qquad (0{,}61\ \%)$
- 95% interval spolehlivosti pro $\mu_U - \mu_V$:

$$\left(12{,}293 - \frac{8{,}144}{\sqrt{99}}1{,}9845\ ;\ 12{,}293 + \frac{8{,}144}{\sqrt{99}}1{,}9845\right) = (10{,}67; 13{,}92)$$

[shapiro.test(vyska.o-vyska.m)] ověření normality
[t.test(vyska.o,vyska.m, mu=10, paired=TRUE)]
[t.test(vyska.o-vyska.m, mu=10)]

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $\mathrm{P}(U_i > V_i) = \mathrm{P}(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $\mathrm{P}(U_i > V_i) = \mathrm{P}(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $P(U_i > V_i) = P(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $\mathrm{P}(U_i > V_i) = \mathrm{P}(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $\mathrm{P}(U_i > V_i) = \mathrm{P}(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

znaménkový test

bez předpokladu normálního rozdělení, stačí libovolné spojité

- stačí znát znaménka rozdílů $X_i = U_i - V_i$
- pozorování s $U_i = V_i$ (tj. $X_i = 0$) se vynechají, upraví se n
- Y – počet kladných znamének $X_i = U_i - V_i$
- H_0 : rozdělení U a V jsou stejná, pak je nutně $\mathrm{P}(U_i > V_i) = \mathrm{P}(X_i > 0) = 1/2$, tedy $Y \sim \mathrm{bi}(n, 1/2)$
- H_0 zamítáme pro velká nebo malá Y:

$$Z = \frac{Y - n/2}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

- pro malá n je bezpečnější použít Yatesovu korekci

$$Z = \frac{|Y - n/2| - 0{,}5}{\sqrt{n/4}}, \qquad |Z| \geq z(\alpha/2)$$

příklad: věk rodičů (párová pozorování!)

- celkem 99 dvojic (otec, matka), sledujeme jejich věk (U, V)
- $H_0 : E\,U = E\,V + 2$ (populační míra polohy věku otců je o 2 roky větší, než matek), H_1 oboustranná
- v jedenácti případech je vek.o – vek.m = 2, proto $n = 99 - 11 = 88$
- u 50 dvojic je vek.o – vek.m > 2, proto

$$z = \frac{50 - 88/2}{\sqrt{88/4}} = 1{,}279, \qquad p = 0{,}201\ (20{,}1\ \%)$$

- s Yatesovou korekcí: $z = 1{,}172$, $p = 0{,}241$ (24,1 %)

[n = sum(vek.o-vek.m != 2)] počet nenulových X_i
[y = sum(vek.o-vek.m > 2)] počet kladných X_i
[prop.test(y,n,correct=FALSE)] bez Yatesovy korekce
[prop.test(y,n,correct=TRUE)] s Yatesovou korekcí

příklad: věk rodičů (párová pozorování!)

- celkem 99 dvojic (otec, matka), sledujeme jejich věk (U, V)
- $H_0 : \mathrm{E}\, U = \mathrm{E}\, V + 2$ (populační míra polohy věku otců je o 2 roky větší, než matek), H_1 oboustranná
- v jedenácti případech je vek.o – vek.m $= 2$, proto $n = 99 - 11 = 88$
- u 50 dvojic je vek.o – vek.m > 2, proto

$$z = \frac{50 - 88/2}{\sqrt{88/4}} = 1{,}279, \qquad p = 0{,}201\ (20{,}1\ \%)$$

- s Yatesovou korekcí: $z = 1{,}172$, $p = 0{,}241$ (24,1 %)

[n = sum(vek.o-vek.m != 2)]	počet nenulových X_i
[y = sum(vek.o-vek.m > 2)]	počet kladných X_i
[prop.test(y,n,correct=FALSE)]	bez Yatesovy korekce
[prop.test(y,n,correct=TRUE)]	s Yatesovou korekcí

příklad: věk rodičů (párová pozorování!)

- celkem 99 dvojic (otec, matka), sledujeme jejich věk (U, V)
- $H_0 : \mathrm{E}\, U = \mathrm{E}\, V + 2$ (populační míra polohy věku otců je o 2 roky větší, než matek), H_1 oboustranná
- v jedenácti případech je vek.o – vek.m $= 2$, proto $n = 99 - 11 = 88$
- u 50 dvojic je vek.o – vek.m > 2, proto

$$z = \frac{50 - 88/2}{\sqrt{88/4}} = 1{,}279, \qquad p = 0{,}201\ (20{,}1\ \%)$$

- s Yatesovou korekcí: $z = 1{,}172$, $p = 0{,}241$ ($24{,}1$ %)

[n = sum(vek.o-vek.m != 2)]	počet nenulových X_i
[y = sum(vek.o-vek.m > 2)]	počet kladných X_i
[prop.test(y,n,correct=FALSE)]	bez Yatesovy korekce
[prop.test(y,n,correct=TRUE)]	s Yatesovou korekcí

příklad: věk rodičů (párová pozorování!)

- celkem 99 dvojic (otec, matka), sledujeme jejich věk (U, V)
- $H_0 : \mathrm{E}\,U = \mathrm{E}\,V + 2$ (populační míra polohy věku otců je o 2 roky větší, než matek), H_1 oboustranná
- v jedenácti případech je vek.o – vek.m $= 2$, proto $n = 99 - 11 = 88$
- u 50 dvojic je vek.o – vek.m > 2, proto

$$z = \frac{50 - 88/2}{\sqrt{88/4}} = 1{,}279, \qquad p = 0{,}201 \ (20{,}1\ \%)$$

- s Yatesovou korekcí: $z = 1{,}172$, $p = 0{,}241$ (24,1 %)

[n = sum(vek.o-vek.m != 2)] počet nenulových X_i
[y = sum(vek.o-vek.m > 2)] počet kladných X_i
[prop.test(y,n,correct=FALSE)] bez Yatesovy korekce
[prop.test(y,n,correct=TRUE)] s Yatesovou korekcí

příklad: věk rodičů (párová pozorování!)

- celkem 99 dvojic (otec, matka), sledujeme jejich věk (U, V)
- $H_0 : \mathrm{E}\, U = \mathrm{E}\, V + 2$ (populační míra polohy věku otců je o 2 roky větší, než matek), H_1 oboustranná
- v jedenácti případech je vek.o – vek.m $= 2$, proto $n = 99 - 11 = 88$
- u 50 dvojic je vek.o – vek.m > 2, proto

$$z = \frac{50 - 88/2}{\sqrt{88/4}} = 1{,}279, \qquad p = 0{,}201 \ (20{,}1\ \%)$$

- s Yatesovou korekcí: $z = 1{,}172$, $p = 0{,}241$ (24,1 %)

[n = sum(vek.o-vek.m != 2)]	počet nenulových X_i
[y = sum(vek.o-vek.m > 2)]	počet kladných X_i
[prop.test(y,n,correct=FALSE)]	bez Yatesovy korekce
[prop.test(y,n,correct=TRUE)]	s Yatesovou korekcí

párový Wilcoxonův test
(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

párový Wilcoxonův test
(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

párový Wilcoxonův test
(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

párový Wilcoxonův test

(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

párový Wilcoxonův test
(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

párový Wilcoxonův test

(silnější předpoklad, než u znaménkového testu)

- nutné **spojité a symetrické** rozdělení $X_i = U_i - V_i$
- opět vyloučíme případy $U_i = V_i$ (tj. $X_i = 0$)
- určíme pořadí R_i^+ hodnot $|X_i| = |U_i - V_i|$
- W součet těch pořadí, kde bylo $U_i > V_i$ (tj. $X_i > 0$)

$$Z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

- pod odmocninou bývá ještě oprava na výskyt shodných hodnot, která jmenovatele poněkud zmenší [wilcox.test(vyska.o,vyska.m,mu=10,paired=TRUE)]
- všimněte si zkrácených názvů parametrů (jednoznačnost!) [wilcox.test(vyska.o,vyska.m,m=10,p=TRUE,cor=FALSE)]

příklad: porovnání dvou metod učení nazpaměť

- H_0 : populační medián rozdílů $= 0$
- znaménkový test: $y = 5$; $n = 8$

$$z = \frac{|5 - 8/2| - 0{,}5}{\sqrt{8/4}} = 0{,}3536;\ p = 0{,}7237$$

- Wilcoxonův test (nově předpokládáme symetrii)

$u_i - v_i$	5	-1	2	3	-1	4	3	-3
r_i^+	8	1,5	3	5	1,5	7	5	5

$$w = 8 + 3 + 5 + 7 + 5 = 28$$

$$z = \frac{28 - 8 \cdot 9/4}{\sqrt{8 \cdot 9 \cdot 17/24}} = \frac{10}{\sqrt{51}} = 1{,}4$$

$$p = 0{,}1614 \qquad p = 16{,}1\ \%$$

- R dá $p = 15{,}9\ \%$, protože bere ohled na shodu

příklad: porovnání dvou metod učení nazpaměť

- H_0 : populační medián rozdílů $= 0$
- znaménkový test: $y = 5$; $n = 8$
  $$z = \frac{|5 - 8/2| - 0{,}5}{\sqrt{8/4}} = 0{,}3536;\ p = 0{,}7237$$
- Wilcoxonův test (nově předpokládáme symetrii)

$u_i - v_i$	5	-1	2	3	-1	4	3	-3
r_i^+	8	1,5	3	5	1,5	7	5	5

$$w = 8 + 3 + 5 + 7 + 5 = 28$$

$$z = \frac{28 - 8 \cdot 9/4}{\sqrt{8 \cdot 9 \cdot 17/24}} = \frac{10}{\sqrt{51}} = 1{,}4$$

$$p = 0{,}1614 \qquad p = 16{,}1\ \%$$

- R dá $p = 15{,}9\ \%$, protože bere ohled na shodu

příklad: porovnání dvou metod učení nazpaměť

- H_0 : populační medián rozdílů $= 0$
- znaménkový test: $y = 5$; $n = 8$
  $$z = \frac{|5 - 8/2| - 0{,}5}{\sqrt{8/4}} = 0{,}3536;\ p = 0{,}7237$$
- Wilcoxonův test (nově předpokládáme symetrii)

$u_i - v_i$	5	-1	2	3	-1	4	3	-3
r_i^+	8	1,5	3	5	1,5	7	5	5

$$w = 8 + 3 + 5 + 7 + 5 = 28$$

$$z = \frac{28 - 8 \cdot 9/4}{\sqrt{8 \cdot 9 \cdot 17/24}} = \frac{10}{\sqrt{51}} = 1{,}4$$

$$p = 0{,}1614 \qquad p = 16{,}1\ \%$$

- R dá $p = 15{,}9\ \%$, protože bere ohled na shodu

příklad: porovnání dvou metod učení nazpaměť

- H_0 : populační medián rozdílů $= 0$
- znaménkový test: $y = 5$; $n = 8$
  $$z = \frac{|5 - 8/2| - 0{,}5}{\sqrt{8/4}} = 0{,}3536;\ p = 0{,}7237$$
- Wilcoxonův test (nově předpokládáme symetrii)

$u_i - v_i$	5	-1	2	3	-1	4	3	-3
r_i^+	8	1,5	3	5	1,5	7	5	5

$$w = 8 + 3 + 5 + 7 + 5 = 28$$

$$z = \frac{28 - 8 \cdot 9/4}{\sqrt{8 \cdot 9 \cdot 17/24}} = \frac{10}{\sqrt{51}} = 1{,}4$$

$$p = 0{,}1614 \qquad p = 16{,}1\ \%$$

- R dá $p = 15{,}9\ \%$, protože bere ohled na shodu

dvouvýběrový t-test
(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé** (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\text{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test
(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé**
  (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\mathrm{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test
(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé**
  (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\mathrm{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test

(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé** (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\text{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test
(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé**
  (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\mathrm{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test
(předpoklad **normálního rozdělení**)

- n_X nezávislých pozorování X, n_Y nezávislých pozorování Y
- tyto výběry musí být **nezávislé**
  (musí vyplynout ze způsobu pořízení dat)
- rozptyly σ_X^2, σ_Y^2 shodné (odhady S_X^2, S_Y^2 podobné, lze ověřit)
- normální rozdělení v obou výběrech (lze ověřit, pro velká n_X, n_Y nenormalita nevadí)
- společný odhad rozptylu (vážený průměr odhadů z jednotlivých výběrů)

$$S^2 = \frac{n_X - 1}{n_X + n_Y - 2} S_X^2 + \frac{n_Y - 1}{n_X + n_Y - 2} S_Y^2$$

- statistika (pro test hypotézy, že rozdělení X a Y jsou stejná)

$$T = \frac{\bar{X} - \bar{Y}}{\mathrm{S.E.}(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_X n_Y}{n_X + n_Y}}$$

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska~Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska~Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska∼Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska∼Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska∼Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska∼Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska∼Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska∼Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska~Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska~Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska~Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska~Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska~Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska~Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

dvouvýběrový t-test

- $H_0 : \mu_X = \mu_Y$
  zamítnout ve prospěch alternativy
  - $H_1 : \mu_X \neq \mu_Y$ když $|T| \geq t_{n_X+n_Y-2}(\alpha)$
  - $H_1 : \mu_X > \mu_Y$ když $T \geq t_{n_X+n_Y-2}(2\alpha)$
  - $H_1 : \mu_X < \mu_Y$ když $T \leq -t_{n_X+n_Y-2}(2\alpha)$

  [t.test(hosi,divky,var.equal=TRUE)] nebo
  [t.test(vyska~Hoch,data=Vysky,var.equal=TRUE)]
- zamítáme-li H_0, říkáme, že rozdíl výběrových průměrů **je významný**
- pochyby o shodě rozptylů: Welchův test (modifikace t-testu)
  [t.test(hosi,divky,var.equal=FALSE)] (pro $\sigma_X \neq \sigma_Y$)
  [t.test(hosi,divky)] resp. [t.test(vyska~Hoch)] (pro $\sigma_X \neq \sigma_Y$)
- shodu rozptylů lze ověřit např. F-testem ($H_0 : \sigma_X = \sigma_Y$)
  [var.test(hosi,divky)]
- ověření normality nutně pro každý výběr zvlášť!

příklad: výšky dětí

	rozsah	průměr	výb. rozptyl
hoši	15	139,13	42,98
dívky	12	140,83	33,79

$$s^2 = \frac{15-1}{15+12-2} 42{,}98 + \frac{12-1}{15+12-2} 33{,}79 = 38{,}936$$

$$|t| = \frac{|139{,}13 - 140{,}83|}{\sqrt{38{,}936}} \sqrt{\frac{15 \cdot 12}{15+12}} = |-0{,}703| < 2{,}06 = t_{25}(0{,}05)$$

[shapiro.test(hosi)] $p = 80\ \%$
[shapiro.test(divky)] $p = 38\ \%$
[var.test(hosi,divky)] $p = 70\ \%$
[t.test(hosi,divky,var.equal=TRUE)]

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti $\Rightarrow$ významný rozdíl
  - nevýznamný rozdíl průměrů $\Rightarrow$ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový *t*-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti $\Rightarrow$ významný rozdíl
  - nevýznamný rozdíl průměrů $\Rightarrow$ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože *t*-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti ⇒ významný rozdíl
  - nevýznamný rozdíl průměrů ⇒ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový *t*-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti ⇒ významný rozdíl
  - nevýznamný rozdíl průměrů ⇒ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože *t*-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti $\Rightarrow$ významný rozdíl
  - nevýznamný rozdíl průměrů $\Rightarrow$ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti ⇒ významný rozdíl
  - nevýznamný rozdíl průměrů ⇒ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5\ \%$, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti $\Rightarrow$ významný rozdíl
  - nevýznamný rozdíl průměrů $\Rightarrow$ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5$ %, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti ⇒ významný rozdíl
  - nevýznamný rozdíl průměrů ⇒ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5\ \%$, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový t-test a intervaly spolehlivosti
(poznámka na okraj)

- zpravidla platí
  - disjunktní intervaly spolehlivosti $\Rightarrow$ významný rozdíl
  - nevýznamný rozdíl průměrů $\Rightarrow$ překryv intervalů
  - rozdíl průměrů může být významný a současně se intervaly mohou překrývat
  - pokud každý z intervalů spolehlivosti obsahuje výběrový průměr druhého výběru, rozdíl průměrů není významný (nemusí platit v případě, kdy oba rozsahy výběru jsou do čtyř)
- příklad: váha v 24. týdnu dětí matek maturantek
  - 95% interval spolehlivosti pro hochy [kg]: (7,51; 8,25)
  - 95% interval spolehlivosti pro dívky [kg]: (6,98; 7,59)
  - intervaly se poněkud překrývají, přestože t-test dal: $t = 2{,}52$, $p = 1{,}5\ \%$, tedy na odpovídající 5% hladině je rozdíl významný

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)
(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)
(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)

(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)
(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)

(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)
(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)

(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)

(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané"
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

dvouvýběrový Wilcoxonův test (Mannův-Whitneyův)
(stačí **spojité rozdělení**)

- dva nezávislé výběry rozsahu n_X, n_Y
- spojitá rozdělení
- H_0: rozdělení jsou stejná, tedy i **mediány** jsou stejné
- za H_0 jsou výběry „dobře promíchané“
- urči pořadí všech (promíchaných)
- kritický obor: průměrná pořadí se příliš liší
- W_X součet pořadí hodnot X

$$Z = \frac{W_X - n_X(n_X + n_Y + 1)/2}{\sqrt{n_X n_Y (n_X + n_Y + 1)/12}}$$

- shodu zamítni, pokud $|Z| \geq z(\alpha/2)$ (přibližný test)
- citlivý vůči posunutí, méně vůči nestejné variabilitě

hoši			dívky				poř.
127							1
130							2
			131				3
			132				4
133							5
			135				6
136	136						7,5
138							9
139	139	139					11
140							13
141			141	141	141	141	16
142			142				19,5
			143				21
			146	146			22,5
147							24
149							25
151			151				26,5

[wilcox.test(hosi,divky)]

$$
\begin{aligned}
w_X &= 1 + 2 + 5 + 2 \cdot 7{,}5 + 9 \\
&\quad + 3 \cdot 11 + 13 + 16 + 19{,}5 \\
&\quad + 24 + 25 + 26{,}5 = 189 \\
w_Y &= 3 + 4 + 6 + 4 \cdot 16 + 19{,}5 \\
&\quad + 21 + 2 \cdot 22{,}5 + 26{,}5 = 189
\end{aligned}
$$

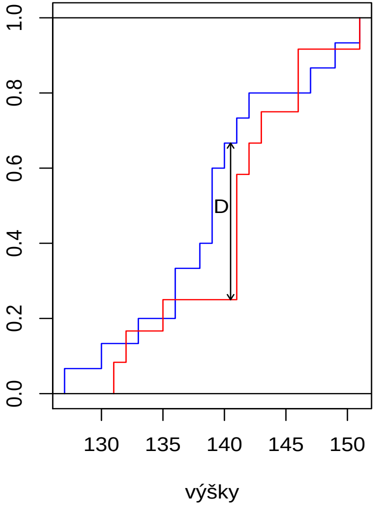
$$
z = \frac{189 - 15 \cdot (15 + 12 + 1)/2}{\sqrt{15 \cdot 12(15 + 12 + 1)/12}} = -1{,}025
$$

$$p = 0{,}3055$$

přesně: $p = 0{,}3149$

Kolmogorovův-Smirnovův test

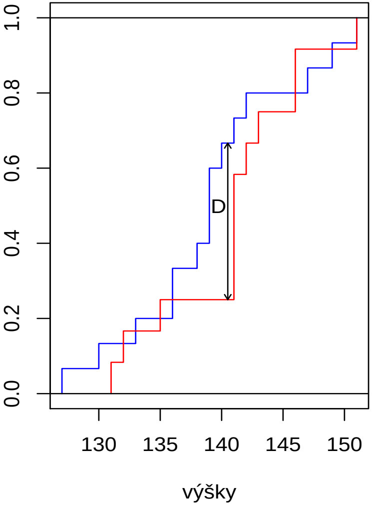
- porovná empirické distribuční funkce
- citlivý vůči všem neshodám (nejen co do populačního průměru či populačního mediánu)
- porovnání výšek hochů a dívek
- $D = \frac{10}{15} - \frac{3}{12} = 0{,}4167$
  $p = 19{,}7\ \%$



[ks.test(hosi,divky)]

Kolmogorovův-Smirnovův test

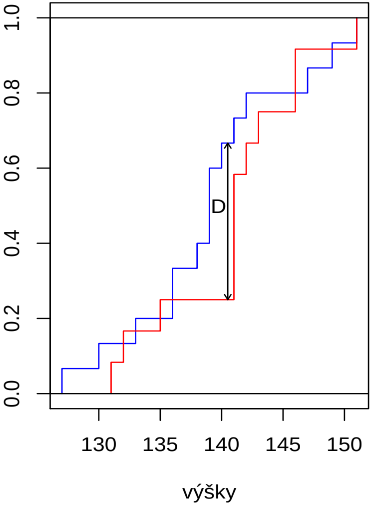
- porovná empirické distribuční funkce
- citlivý vůči všem neshodám (nejen co do populačního průměru či populačního mediánu)
- porovnání výšek hochů a dívek
- $D = \frac{10}{15} - \frac{3}{12} = 0{,}4167$
  $p = 19{,}7\ \%$



[ks.test(hosi,divky)]

Kolmogorovův-Smirnovův test

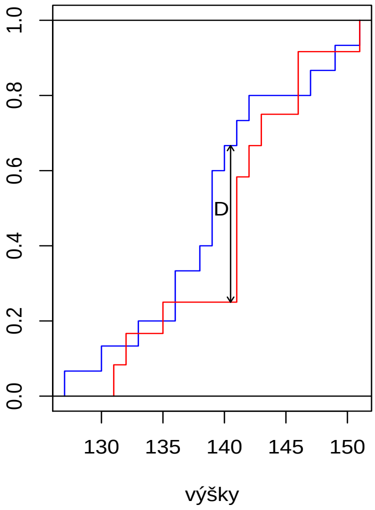
- porovná empirické distribuční funkce
- citlivý vůči všem neshodám (nejen co do populačního průměru či populačního mediánu)
- porovnání výšek hochů a dívek
- $D = \frac{10}{15} - \frac{3}{12} = 0{,}4167$
  $p = 19{,}7\ \%$



[ks.test(hosi,divky)]

Kolmogorovův-Smirnovův test

- porovná empirické distribuční funkce
- citlivý vůči všem neshodám (nejen co do populačního průměru či populačního mediánu)
- porovnání výšek hochů a dívek
- $D = \frac{10}{15} - \frac{3}{12} = 0{,}4167$
  $p = 19{,}7\ \%$



[ks.test(hosi,divky)]