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 25. března 2008)

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí ...)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza“, co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu“
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí ...)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza", co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu"
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí ...)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza“, co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu“
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí …)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza", co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu"
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí …)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza“, co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu“
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí ...)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza“, co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu“
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

testování statistických hypotéz

[hypothesis testing, null hypothesis, alternative hypothesis, critical (rejection) region, Type I (II) error, significance level]

- **nulová hypotéza** H_0 tvrzení o populaci (parametru), o jehož platnosti chceme rozhodnout (**ne**ní rozdíl, **ne**závisí ...)
- **alternativní hypotéza** H_1 (alternativa) zbývající možnost (k H_0), často „vědecká hypotéza", co chceme dokázat
- volba mezi $\mathrm{H}_0, \mathrm{H}_1$ dána, volíme **o čem** budou hypotézy
- **kritický obor** možné výsledky pokusu, kdy H_0 zamítáme; zpravidla popsán pomocí statistiky (např. $|Z| \geq z(\alpha/2)$)
- **obor přijetí** možné výsledky pokusu, kdy H_0 nezamítáme
- **chyba prvního druhu** (náhodný jev) rozhodnutí zamítnout H_0, když platí H_0, tj. falešně prokázat „vědeckou hypotézu"
- **chyba druhého druhu** (náhodný jev) rozhodnutí nezamítnout H_0, když platí H_1, tj. nepoznat neplatnost H_0

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1-\beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu"
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p-**hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p-**hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1-\beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p-**hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu"
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p**-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p**-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu"
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p-**hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu"
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p-**hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p**-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p**-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = \mathrm{P}(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p **(p-hodnota)**
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = P(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1-\beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (**p-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = \mathrm{P}(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

statistické rozhodování

[significance level, power, p-value]

- **hladina testu** α (zpravidla 5 %, 1 %)
  - maximální dovolená pst chyby prvního druhu
  - volí se před pokusem, nezávisle na jeho výsledku
- **síla testu** $1 - \beta$
  - pravděpodobnost zamítnutí neplatné H_0
  - pst, s jakou prokážeme platnou „vědeckou hypotézu“
  - závisí na tom, co opravdu platí
- **dosažená hladina testu** p (p**-hodnota**)
  - za platnosti H_0 určená pst, že dostaneme statistiku, která stejně nebo ještě méně podporuje H_0
  - nejmenší hladina α, na které lze ještě H_0 zamítnout
  - např. $p = \mathrm{P}(|T| \geq t)$, kde t je skutečně realizovaná hodnota statistiky T
- H_0 se **zamítá**, právě když $\boxed{p \leq \alpha}$

testování statistických hypotéz

rozhodnutí	skutečnost	
	H_0 platí	H_0 neplatí
H_0 zamítnout (reject)	**chyba 1. druhu** ($\leq \alpha$)	správné rozhodnutí ($1 - \beta$)
H_0 nezamítnout (accept)	správné rozhodnutí ($\geq 1 - \alpha$)	chyba 2. druhu (β)

- zamítnutí $\Leftrightarrow$ výsledek pokusu v kritickém oboru
- přijetí $\Leftrightarrow$ výsledek pokusu v oboru přijetí
- nikdy spolehlivě nevíme, zda H_0 platí

testování statistických hypotéz

rozhodnutí	skutečnost: H_0 platí	skutečnost: H_0 neplatí
H_0 zamítnout (reject)	**chyba 1. druhu** ($\leq \alpha$)	správné rozhodnutí ($1-\beta$)
H_0 nezamítnout (accept)	správné rozhodnutí ($\geq 1-\alpha$)	chyba 2. druhu (β)

- zamítnutí $\Leftrightarrow$ výsledek pokusu v kritickém oboru
- přijetí $\Leftrightarrow$ výsledek pokusu v oboru přijetí
- nikdy spolehlivě nevíme, zda H_0 platí

testování statistických hypotéz

rozhodnutí	skutečnost: H_0 platí	skutečnost: H_0 neplatí
H_0 zamítnout (reject)	**chyba 1. druhu** ($\leq \alpha$)	správné rozhodnutí ($1 - \beta$)
H_0 nezamítnout (accept)	správné rozhodnutí ($\geq 1 - \alpha$)	chyba 2. druhu (β)

- zamítnutí $\Leftrightarrow$ výsledek pokusu v kritickém oboru
- přijetí $\Leftrightarrow$ výsledek pokusu v oboru přijetí
- nikdy spolehlivě nevíme, zda H_0 platí

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \ldots, X_n \sim \mathrm{N}\big(\mu, \sigma^2\big)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}\big(\mu, \sigma^2/n\big)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0, 1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0, 1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0, 1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

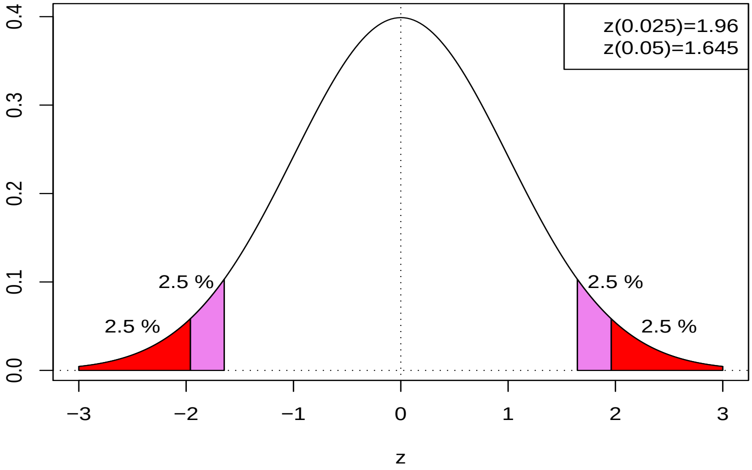
- $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

rozhodování o populačním průměru normálního rozdělení (σ známé)

- $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ **nezávislé**; $\sigma > 0$ známe
- $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, tedy $\mathrm{S.E.}(\bar{X}) = \sigma/\sqrt{n}$
- $\mathrm{H}_0 : \mu = \mu_0$ (dané číslo, jiný zápis $\mathrm{H}_0 : \mu - \mu_0 = 0$)
- platí-li H_0, pak $\boxed{Z = \dfrac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)}$
- $\mathrm{H}_1 : \mu \neq \mu_0 \Rightarrow$ kritický obor: $|Z|$ velké, tj. $\boxed{|Z| \geq z(\alpha/2)}$
- $\mathrm{H}_1 : \mu > \mu_0$: zamítnout pro $\boxed{Z \geq z(\alpha)}$
- $\mathrm{H}_1 : \mu < \mu_0$: zamítnout pro $\boxed{Z \leq -z(\alpha)}$
- volba jednostranné alternativy jen podle zadání úlohy, nikoliv podle výsledku pokusu

kritický obor pro Z

červeně na 5% hladině, červeně a fialově na 10% hladině



příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $\mathrm{H}_0 : \mu = \mu_0$ proti $\mathrm{H}_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $\mathrm{H}_0 : \mu = \mu_0$ proti $\mathrm{H}_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

příklad: výška desetiletých chlapců

- zvolíme klasickou hladinu $\alpha = 5\ \%$
- v roce 1951 velký výběr: $\mu_0 = 136{,}1$ cm, $\sigma = 6{,}4$ cm
- v roce 1961 změřeno $n = 15$ náhodně vybraných desetiletých hochů, $\bar{x} = 139{,}13$ cm
- stačí tento vzrůst k důkazu, že nová generace je vyšší?
- vzrostla výška desetiletých ? $H_0 : \mu = \mu_0$ proti $H_1 : \mu > \mu_0$

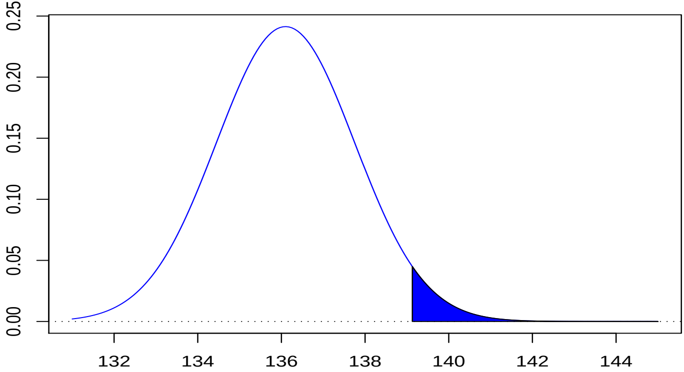
$$z = \frac{139{,}13 - 136{,}1}{6{,}4}\sqrt{15} = 1{,}836$$

- $z(0{,}05) = 1{,}645 < 1{,}836$, tedy H_0 na 5% hladině **zamítáme**
- na 5% hladině jsme prokázali, že nová generace je vyšší
- v případě, že nová generace není vyšší, riskovali jsme jen 5% pravděpodobnost, že budeme nesprávně tvrdit, že vyšší je

výška desetiletých hochů

hustota $\bar{X}$ za platnosti hypotézy

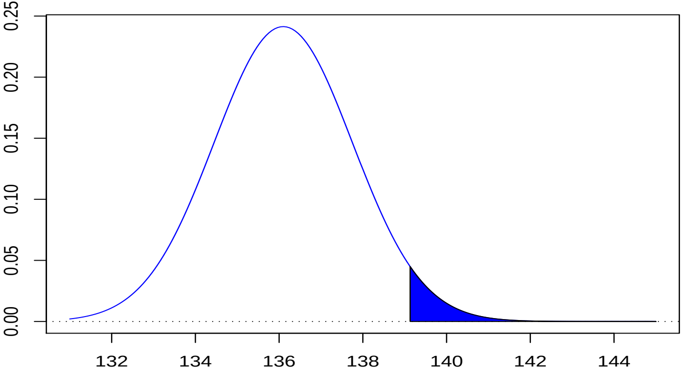
- p-hodnota – pst, že za H_0 vyjde $Z > 1{,}836$ [1-pnorm(1.836)]
- p-hodnota – modrá plocha napravo od zjištěného průměru, $p = 3{,}3\ \%$



výška desetiletých hochů

hustota $\bar{X}$ za platnosti hypotézy

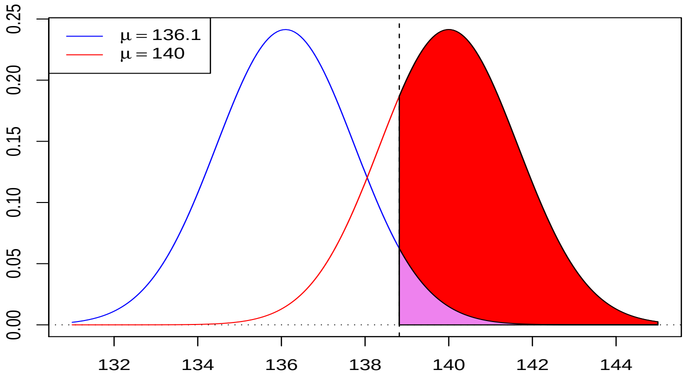
- p-hodnota – pst, že za H_0 vyjde $Z > 1{,}836$ [1-pnorm(1.836)]
- p-hodnota – modrá plocha napravo od zjištěného průměru, $p = 3{,}3\ \%$



výška desetiletých chlapců

hustota $\bar{X}$ za hypotézy a při $\mu = 140$

hladina testu – fialová plocha, síla testu – fialová + červená plocha



$$\mu_0 + 6{,}4/\sqrt{15} \cdot 1{,}96 = 138{,}8$$

volba rozsahu výběru

$H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$

- pro zvolenou hodnotu $\mu_1 \neq \mu_0$ požadujeme sílu $1 - \beta$
- $1 - \beta$ je pravděpodobnost, s jakou odhalíme neplatnost H_0, je-li skutečnost $\mu = \mu_1$

$$n \geq \left(\frac{z(\alpha/2) + z(\beta)}{\mu_1 - \mu_0} \right)^2 \sigma^2$$

- při jednostranné alternativě by bylo $z(\alpha)$ místo $z(\alpha/2)$
- aby pro $\mu_1 = 140$ byla síla 90 % (tj. $1 - \beta = 0{,}9$, $\beta = 0{,}1$, $z(0{,}1) = 1{,}282$), bude třeba aspoň

$$n \geq \left(\frac{1{,}96 + 1{,}282}{140 - 136{,}1} \right)^2 6{,}4^2 = 28{,}3$$

(místo 15 pozorování jich potřebujeme aspoň 29)

volba rozsahu výběru

$H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$

- pro zvolenou hodnotu $\mu_1 \neq \mu_0$ požadujeme sílu $1 - \beta$
- $1 - \beta$ je pravděpodobnost, s jakou odhalíme neplatnost H_0, je-li skutečnost $\mu = \mu_1$

$$n \geq \left(\frac{z(\alpha/2) + z(\beta)}{\mu_1 - \mu_0} \right)^2 \sigma^2$$

- při jednostranné alternativě by bylo $z(\alpha)$ místo $z(\alpha/2)$
- aby pro $\mu_1 = 140$ byla síla 90 % (tj. $1 - \beta = 0{,}9$, $\beta = 0{,}1$, $z(0{,}1) = 1{,}282$), bude třeba aspoň

$$n \geq \left(\frac{1{,}96 + 1{,}282}{140 - 136{,}1} \right)^2 6{,}4^2 = 28{,}3$$

(místo 15 pozorování jich potřebujeme aspoň 29)

volba rozsahu výběru

$H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$

- pro zvolenou hodnotu $\mu_1 \neq \mu_0$ požadujeme sílu $1 - \beta$
- $1 - \beta$ je pravděpodobnost, s jakou odhalíme neplatnost H_0, je-li skutečnost $\mu = \mu_1$

$$n \geq \left(\frac{z(\alpha/2) + z(\beta)}{\mu_1 - \mu_0} \right)^2 \sigma^2$$

- při jednostranné alternativě by bylo $z(\alpha)$ místo $z(\alpha/2)$
- aby pro $\mu_1 = 140$ byla síla 90 % (tj. $1 - \beta = 0{,}9$, $\beta = 0{,}1$, $z(0{,}1) = 1{,}282$), bude třeba aspoň

$$n \geq \left(\frac{1{,}96 + 1{,}282}{140 - 136{,}1} \right)^2 6{,}4^2 = 28{,}3$$

(místo 15 pozorování jich potřebujeme aspoň 29)

volba rozsahu výběru

$H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$

- pro zvolenou hodnotu $\mu_1 \neq \mu_0$ požadujeme sílu $1 - \beta$
- $1 - \beta$ je pravděpodobnost, s jakou odhalíme neplatnost H_0, je-li skutečnost $\mu = \mu_1$

$$n \geq \left(\frac{z(\alpha/2) + z(\beta)}{\mu_1 - \mu_0} \right)^2 \sigma^2$$

- při jednostranné alternativě by bylo $z(\alpha)$ místo $z(\alpha/2)$
- aby pro $\mu_1 = 140$ byla síla 90 % (tj. $1 - \beta = 0{,}9$, $\beta = 0{,}1$, $z(0{,}1) = 1{,}282$), bude třeba aspoň

$$n \geq \left(\frac{1{,}96 + 1{,}282}{140 - 136{,}1} \right)^2 6{,}4^2 = 28{,}3$$

(místo 15 pozorování jich potřebujeme aspoň 29)

jednovýběrový t-test

výběr z $\mathrm{N}(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x} \sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $\mathrm{N}(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x}\sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $\mathrm{N}\left(\mu, \sigma^2\right)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}\left(\mu, \sigma^2\right)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x}\sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $\mathrm{N}(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x} \sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $\mathrm{N}(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x}\sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $N(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $N(\mu, \sigma^2)$
- $H_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x} \sqrt{n}}$$

- $H_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $H_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $H_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

jednovýběrový t-test

výběr z $\mathrm{N}(\mu, \sigma^2)$, σ neznámé

- n nezávislých pozorování $X_1, \ldots, X_n$ z rozdělení $\mathrm{N}(\mu, \sigma^2)$
- $\mathrm{H}_0 : \mu = \mu_0$ (populační průměr roven dané konstantě)
- nutno odhadnout neznámý rozptyl σ^2

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- statistika (místo σ použijeme S_x)

$$\boxed{T = \frac{\bar{X} - \mu_0}{\mathrm{S.E.}(\bar{X})} = \frac{\bar{X} - \mu_0}{S_x} \sqrt{n}}$$

- $\mathrm{H}_1 : \mu \neq \mu_0$ zamítat při $|T| \geq t_{n-1}(\alpha)$
- $\mathrm{H}_1 : \mu > \mu_0$ zamítat při $T \geq t_{n-1}(2\alpha)$
- $\mathrm{H}_1 : \mu < \mu_0$ zamítat při $T \leq -t_{n-1}(2\alpha)$

výšky hochů pro případ neznámého σ

- $H_0 : \mu = 136{,}1$ proti $H_1 : \mu > 136{,}1$ ($\alpha = 5\ \%$)

$$\bar{x} = 139{,}133 \qquad s_x^2 = 6{,}556^2$$

$$t = \frac{139{,}133 - 136{,}1}{6{,}556}\sqrt{15} = 1{,}792 > 1{,}761 = t_{14}(0{,}10)$$

$$p = \mathrm{P}(T \geq 1{,}792) = 0{,}047 \qquad (\text{ tj. } 4{,}7\ \%)$$

- na 5% hladině jsme prokázali zvýšení populačního průměru (H_0 se na 5% hladině **zamítá**)
- [t.test(hosi,mu=136.1,alternative="greater")]

výšky hochů pro případ neznámého σ

- $\mathrm{H}_0 : \mu = 136{,}1$ proti $\mathrm{H}_1 : \mu > 136{,}1$ $(\alpha = 5\ \%)$

$$\bar{x} = 139{,}133 \qquad s_x^2 = 6{,}556^2$$

$$t = \frac{139{,}133 - 136{,}1}{6{,}556}\sqrt{15} = 1{,}792 > 1{,}761 = t_{14}(0{,}10)$$

$$p = \mathrm{P}(T \geq 1{,}792) = 0{,}047 \qquad (\text{ tj. } 4{,}7\ \%)$$

- na 5% hladině jsme prokázali zvýšení populačního průměru (H_0 se na 5% hladině **zamítá**)
- [t.test(hosi,mu=136.1,alternative="greater")]

výšky hochů pro případ neznámého σ

- $\mathrm{H}_0 : \mu = 136{,}1$ proti $\mathrm{H}_1 : \mu > 136{,}1$ ($\alpha = 5\ \%$)

$$\bar{x} = 139{,}133 \qquad s_x^2 = 6{,}556^2$$
$$t = \frac{139{,}133 - 136{,}1}{6{,}556}\sqrt{15} = 1{,}792 > 1{,}761 = t_{14}(0{,}10)$$
$$p = \mathrm{P}(T \geq 1{,}792) = 0{,}047 \qquad (\text{ tj. } 4{,}7\ \%)$$

- na 5% hladině jsme prokázali zvýšení populačního průměru (H_0 se na 5% hladině **zamítá**)
- [t.test(hosi,mu=136.1,alternative="greater")]

výšky hochů pro případ neznámého σ
(jiné zadání úlohy)

- **kdybychom** předem neměli určenu jednostrannou alternativu, ale zvolili $H_1 : \mu \neq 136{,}1$, pak

$$|t| = |1{,}792| < 2{,}145 = t_{14}(0{,}05)$$
$$p = P(|T| \geq 1{,}792) = 0{,}0948 \qquad \text{(tj. } 9{,}48\ \%\text{)}$$

- hypotézu na 5% hladině nezamítáme
- [t.test(hosi,mu=136.1,alternative="two.sided")], stačí ale [t.test(hosi,mu=136.1)]

výšky hochů pro případ neznámého σ

(jiné zadání úlohy)

- **kdybychom** předem neměli určenu jednostrannou alternativu, ale zvolili $H_1 : \mu \neq 136{,}1$, pak

$$|t| = |1{,}792| < 2{,}145 = t_{14}(0{,}05)$$
$$p = P(|T| \geq 1{,}792) = 0{,}0948 \qquad \text{(tj. 9,48 \%)}$$

- hypotézu na 5% hladině nezamítáme
- [t.test(hosi,mu=136.1,alternative="two.sided")], stačí ale [t.test(hosi,mu=136.1)]

výšky hochů pro případ neznámého σ

(jiné zadání úlohy)

- **kdybychom** předem neměli určenu jednostrannou alternativu, ale zvolili $\mathrm{H}_1 : \mu \neq 136{,}1$, pak

$$|t| = |1{,}792| < 2{,}145 = t_{14}(0{,}05)$$
$$p = \mathrm{P}(|T| \geq 1{,}792) = 0{,}0948 \qquad \text{(tj. } 9{,}48\ \%\text{)}$$

- hypotézu na 5% hladině nezamítáme
- [t.test(hosi,mu=136.1,alternative="two.sided")], stačí ale [t.test(hosi,mu=136.1)]

interval spolehlivosti pro μ

souvislost s testem o μ při oboustranné alternativě

- oboustranný interval spolehlivosti pro μ

$$\left(\bar{X} - \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha), \bar{X} + \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha)\right)$$

- μ_0 patří do intervalu spolehlivosti, právě když platí

$$|\bar{X} - \mu_0| < \frac{S_x}{\sqrt{n}} t_{n-1}(\alpha)$$

- tedy, právě když se nezamítne hypotéza $\mathrm{H}_0 : \mu = \mu_0$ při oboustranné alternativě $\mathrm{H}_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom **nezamítli** hypotézu $\mathrm{H}_0 : \mu = \mu_0$
- podobně u jednostranných intervalů spolehlivosti a jednostranných alternativ

interval spolehlivosti pro μ

souvislost s testem o μ při oboustranné alternativě

- oboustranný interval spolehlivosti pro μ

$$\left(\bar{X} - \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha), \bar{X} + \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha)\right)$$

- μ_0 patří do intervalu spolehlivosti, právě když platí

$$|\bar{X} - \mu_0| < \frac{S_x}{\sqrt{n}} t_{n-1}(\alpha)$$

- tedy, právě když se nezamítne hypotéza $\mathrm{H}_0 : \mu = \mu_0$ při oboustranné alternativě $\mathrm{H}_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom **nezamítli** hypotézu $\mathrm{H}_0 : \mu = \mu_0$
- podobně u jednostranných intervalů spolehlivosti a jednostranných alternativ

interval spolehlivosti pro μ

souvislost s testem o μ při oboustranné alternativě

- oboustranný interval spolehlivosti pro μ

$$\left(\bar{X} - \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha), \bar{X} + \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha)\right)$$

- μ_0 patří do intervalu spolehlivosti, právě když platí

$$|\bar{X} - \mu_0| < \frac{S_x}{\sqrt{n}} t_{n-1}(\alpha)$$

- tedy, právě když se nezamítne hypotéza $\mathrm{H}_0 : \mu = \mu_0$ při oboustranné alternativě $\mathrm{H}_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom **nezamítli** hypotézu $\mathrm{H}_0 : \mu = \mu_0$
- podobně u jednostranných intervalů spolehlivosti a jednostranných alternativ

interval spolehlivosti pro μ

souvislost s testem o μ při oboustranné alternativě

- oboustranný interval spolehlivosti pro μ

$$\left(\bar{X} - \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha), \bar{X} + \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha)\right)$$

- μ_0 patří do intervalu spolehlivosti, právě když platí

$$|\bar{X} - \mu_0| < \frac{S_x}{\sqrt{n}} t_{n-1}(\alpha)$$

- tedy, právě když se nezamítne hypotéza $H_0 : \mu = \mu_0$ při oboustranné alternativě $H_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom **nezamítli** hypotézu $H_0 : \mu = \mu_0$
- podobně u jednostranných intervalů spolehlivosti a jednostranných alternativ

interval spolehlivosti pro μ

souvislost s testem o μ při oboustranné alternativě

- oboustranný interval spolehlivosti pro μ

$$\left(\bar{X} - \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha), \bar{X} + \frac{S_x}{\sqrt{n}}\, t_{n-1}(\alpha)\right)$$

- μ_0 patří do intervalu spolehlivosti, právě když platí

$$|\bar{X} - \mu_0| < \frac{S_x}{\sqrt{n}} t_{n-1}(\alpha)$$

- tedy, právě když se nezamítne hypotéza $\mathrm{H}_0 : \mu = \mu_0$ při oboustranné alternativě $\mathrm{H}_1 : \mu \neq \mu_0$
- interval spolehlivosti obsahuje takové hodnoty μ_0, pro které bychom **nezamítli** hypotézu $\mathrm{H}_0 : \mu = \mu_0$
- podobně u jednostranných intervalů spolehlivosti a jednostranných alternativ

výšky hochů pro případ neznámého σ

- 95% interval spolehlivosti: $(135{,}5; 142{,}8)$
  s 95% pravděpodobností je skutečný populační průměr (střední hodnota μ) v uvedeném intervalu
- je jen 5% riziko, že leží mimo uvedený interval
- 99% interval spolehlivosti $(134{,}1; 144{,}2)$
  [t.test(hosi,mu=136.1,conf.level=0.99)] (vedlejší výsledek)
  [confint(lm(hosi~1),level=0.99)]
- aby byla zajištěna větší spolehlivost intervalu (větší pravděpodobnost, že zachytí skutečnou hodnotu), je nutně 99% interval spolehlivosti delší, než 95% interval spolehlivosti

výšky hochů pro případ neznámého σ

- 95% interval spolehlivosti: $(135{,}5; 142{,}8)$
  s 95% pravděpodobností je skutečný populační průměr (střední hodnota μ) v uvedeném intervalu
- je jen 5% riziko, že leží mimo uvedený interval
- 99% interval spolehlivosti $(134{,}1; 144{,}2)$
  [t.test(hosi,mu=136.1,conf.level=0.99)] (vedlejší výsledek)
  [confint(lm(hosi~1),level=0.99)]
- aby byla zajištěna větší spolehlivost intervalu (větší pravděpodobnost, že zachytí skutečnou hodnotu), je nutně 99% interval spolehlivosti delší, než 95% interval spolehlivosti

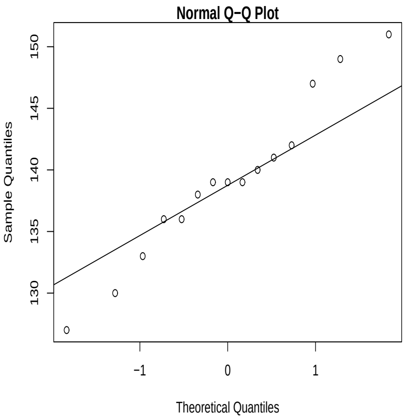
výšky hochů pro případ neznámého σ

- 95% interval spolehlivosti: $(135{,}5; 142{,}8)$
  s 95% pravděpodobností je skutečný populační průměr (střední hodnota μ) v uvedeném intervalu
- je jen 5% riziko, že leží mimo uvedený interval
- 99% interval spolehlivosti $(134{,}1; 144{,}2)$
  [t.test(hosi,mu=136.1,conf.level=0.99)] (vedlejší výsledek)
  [confint(lm(hosi~1),level=0.99)]
- aby byla zajištěna větší spolehlivost intervalu (větší pravděpodobnost, že zachytí skutečnou hodnotu), je nutně 99% interval spolehlivosti delší, než 95% interval spolehlivosti

výšky hochů pro případ neznámého σ

- 95% interval spolehlivosti: $(135{,}5; 142{,}8)$
  s 95% pravděpodobností je skutečný populační průměr (střední hodnota μ) v uvedeném intervalu
- je jen 5% riziko, že leží mimo uvedený interval
- 99% interval spolehlivosti $(134{,}1; 144{,}2)$
  [t.test(hosi,mu=136.1,conf.level=0.99)] (vedlejší výsledek)
  [confint(lm(hosi~1),level=0.99)]
- aby byla zajištěna větší spolehlivost intervalu (větší pravděpodobnost, že zachytí skutečnou hodnotu), je nutně 99% interval spolehlivosti delší, než 95% interval spolehlivosti

ověření předpokladu o normálním rozdělení



- **Shapirův-Wilkův test**
- H_0 : normální rozdělení s nějakými (neznámými) parametry
- [shapiro.test(hosi)]
- $W = 0{,}966$, $p = 80\ \%$
- hodnotí kvalitu přiblížení bodů k přímce na diagramu normality
- [qqnorm(hosi);qqline(hosi)]

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_o)/n}} \dot{\sim} \mathrm{N}(0,1)$$
- někdy s opravou na spojitost (Yates)
$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1-\pi_0)}} \operatorname{sign}(Y - n\pi_0) \dot{\sim} \mathrm{N}(0,1)$$
- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_o)/n}} \dot{\sim} \mathrm{N}(0,1)$$
- někdy s opravou na spojitost (Yates)

$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1-\pi_0)}} \operatorname{sign}(Y - n\pi_0) \dot{\sim} \mathrm{N}(0,1)$$

- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
  $$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_o)/n}} \mathrel{\dot{\sim}} \mathrm{N}(0,1)$$
- někdy s opravou na spojitost (Yates)

$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1-\pi_0)}} \,\mathrm{sign}(Y - n\pi_0) \mathrel{\dot{\sim}} \mathrm{N}(0,1)$$

- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_o)/n}} \dot{\sim} \mathrm{N}(0,1)$$
- někdy s opravou na spojitost (Yates)
$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1-\pi_0)}} \,\mathrm{sign}(Y - n\pi_0) \dot{\sim} \mathrm{N}(0,1)$$
- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1 - \pi_o)/n}} \dot{\sim} \mathrm{N}(0, 1)$$
- někdy s opravou na spojitost (Yates)
$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1 - \pi_0)}} \operatorname{sign}(Y - n\pi_0) \dot{\sim} \mathrm{N}(0, 1)$$
- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

pst výskytu jevu

test hypotézy o parametru π binomického rozdělení

- $Y \sim \mathrm{bi}(n, \pi) \quad \mathrm{H}_0 : \pi = \pi_0$:
$$Z = \frac{Y - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\mathrm{S.E.}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_o)/n}} \,\dot{\sim}\, \mathrm{N}(0,1)$$
- někdy s opravou na spojitost (Yates)

$$Z = \frac{|Y - n\pi_0| - 0{,}5}{\sqrt{n\pi_0(1-\pi_0)}} \,\mathrm{sign}(Y - n\pi_0) \,\dot{\sim}\, \mathrm{N}(0,1)$$

- $\mathrm{H}_1 : \pi \neq \pi_0$: zamítnout pokud $|Z| \geq z(\alpha/2)$
- $\mathrm{H}_1 : \pi > \pi_0$: zamítnout pokud $Z \geq z(\alpha)$
- $\mathrm{H}_1 : \pi < \pi_0$: zamítnout pokud $Z \leq -z(\alpha)$
- existuje přesný postup, bez použití aproximace

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů“, $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů“, $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů", $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů“, $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů“, $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů“, $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- pokusit se prokázat, že kalous dá přednost infikované myši před neinfikovanou
- Y – počet „zdarů", $n = 50$, π – pst, že zvolí infikovanou
- Y má **binomické rozdělení**
  za $\mathrm{H}_0 : \pi = 1/2 (= \pi_0)$ (myši se neliší) $Y \sim \mathrm{bi}(50, 1/2)$
- **alternativní hypotéza**: $\mathrm{H}_1 : \pi > 1/2$
- z 50 případů dal kalous ve 33 případech přednost infikované myši před neinfikovanou
- **kritický obor**: velká hodnota Y (tj. velké $\hat{\pi}$ resp. velké Z)

$$z = \frac{33 - 50 \cdot 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}263 \qquad p = \mathrm{P}(Z \geq 2{,}263) = 0{,}0118$$

- s opravou na spojitost:

$$z = \frac{33 - 50 \cdot 0{,}5 - 0{,}5}{\sqrt{50 \cdot 0{,}5 \cdot 0{,}5}} = 2{,}121 \qquad p = \mathrm{P}(Z \geq 2{,}121) = 0{,}0169$$

příklad kalous

- prop.test() počítá Z^2, která má za H_0 : rozdělení χ^2_1
  [prop.test(33,50,alternative="greater",correct=FALSE)]
  [prop.test(33,50,alternative="greater")]
  [binom.test(33,50,alternative="greater")]
- **dosažená hladina**: za H_0 počítaná pst, že dostaneme výsledek aspoň tolik odporující nulové hypotéze, jako ve skutečném pokusu:

$$\begin{aligned} p &= \mathrm{P}(Y \geq 33) = 1 - \mathrm{P}(Y \leq 32) \\ &= \sum_{k=33}^{50} \binom{50}{k} 0{,}5^k (1-0{,}5)^{50-k} \\ &= 0{,}0164 \end{aligned}$$

[1-pbinom(32,50,1/2)]

příklad kalous

- prop.test() počítá Z^2, která má za H_0 : rozdělení χ^2_1
  [prop.test(33,50,alternative="greater",correct=FALSE)]
  [prop.test(33,50,alternative="greater")]
  [binom.test(33,50,alternative="greater")]
- **dosažená hladina**: za H_0 počítaná pst, že dostaneme výsledek aspoň tolik odporující nulové hypotéze, jako ve skutečném pokusu:

$$\begin{aligned} p &= \mathrm{P}(Y \geq 33) = 1 - \mathrm{P}(Y \leq 32) \\ &= \sum_{k=33}^{50} \binom{50}{k} 0{,}5^k (1-0{,}5)^{50-k} \\ &= 0{,}0164 \end{aligned}$$

[1-pbinom(32,50,1/2)]