# Testy statistických hypotéz

Václav Adamec
**vadamec@mendelu.cz**

## Úvod

- Testování: kvalifikovaná procedura vedoucí v zamítnutí nebo nezamítnutí nulové hypotézy v podmínkách nejistoty
- Testy jsou vázány na rozdělení náhodných veličin
- Testy jsou podmíněny formulací vědeckého problému a experimentálním designem
- Hypotéza: předpoklad o statistických parametrech nebo jejich kombinacích. Typy: hypotéza nulová $H_0$ a alternativní $H_1$
- Testovací statistika: numerická charakteristika vypočtená z dat popř. z parametrů. Má známé rozdělení

## Statistické hypotézy

- Nulová hypotéza $H_0$: formulace negativním způsobem (efekt neexistuje) $H_0$: $\theta = \theta_0$
- $H_0$ vytváří předpoklad o rozdělení parametrů nebo testovací statistiky
- Alternativní hypotéza $H_1$: formulace mimo obor hodnot nulové hypotézy. Obsahuje alternativní předpoklad, který chceme dokázat
  - Dvoustranná $H_1$: $\theta \neq \theta_0$
  - Levostranná $H_1$ : $\theta < \theta_0$
  - Pravostranná $H_1$ : $\theta > \theta_0$
- $H_0$ zamítáme, je-li vypočtená testovací statistika v oblasti zamítnutí
- Oblast zamítnutí $H_0$ je určena typem $H_1$ a parametry rozdělení testovací statistiky

## Typy chyb při statistickém testování

| | | **Rozhodnutí o $H_0$** | |
|---|---|---|---|
| | | **$H_0$ zamítnuta** | **$H_0$ nezamítnuta** |
| **Realita o $H_0$** | **$H_0$ pravdivá** | **chyba I. typu ($\alpha$)** | **(1- $\alpha$)** |
| | **$H_0$ nepravdivá** | **síla testu (1-$\beta$)** | **chyba II. typu ($\beta$)** |

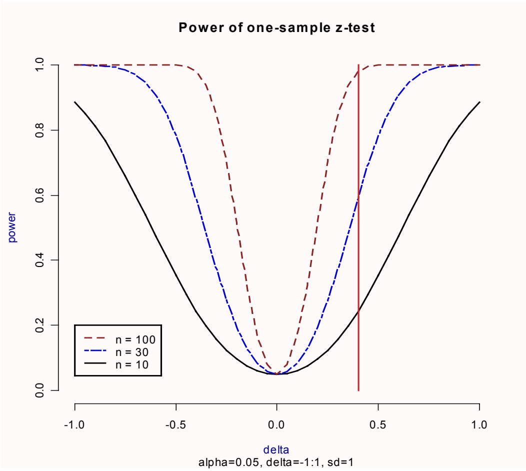
## Chyby při statistickém testování

- Chyba I. Typu ($\alpha$): $H_0$ zamítnuta a zároveň $H_0$ je pravdivá. Rozdíly prohlášeny za průkazné, avšak průkazné nejsou ($H_0$ je pravdivá). Test produkoval nesprávný pozitivní nález
- Chyba II. Typu ($\beta$): $H_0$ nezamítnuta a zároveň je $H_0$ nepravdivá. Rozdíly prohlášeny za neprůkazné, avšak průkazné jsou ($H_0$ je nepravdivá). Test produkoval nesprávný negativní nález
- Situace 1 - $\alpha$ a 1 - $\beta$ jsou korektní
  - V případě (1-$\alpha$) $H_0$ nezamítáme $H_0$ je pravdivá
  - V případě (1-$\beta$) $H_0$ zamítáme a $H_0$ je nepravdivá
- Oba typy chyb nemohou nastat současně
- Kontrolujeme pouze $\alpha$ s ohledem na míru přijatelného rizika důsledků chyby I. typu

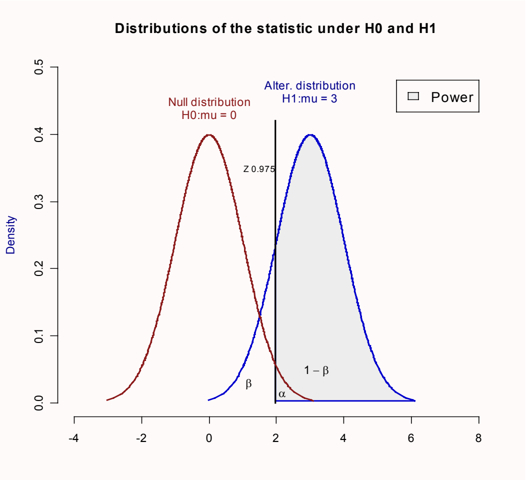
## Síla testu

- Pravděpodobnost 1-$\beta$ se nazývá síla statistického testu. Vysoká hodnota 1 - $\beta$ je žádoucí pro úspěch testování
- Síla testu závisí na:
  - Rozsahu souboru: vyšší n zvyšuje sílu
  - Variabilitě dat: kontrolovaná variabilita zvyšuje sílu
  - Zvolené hodnotě alfa: vysoká alfa zvyšuje sílu, (ale i chybu I. typu …)
  - Ploše překryvu rozdělení testovací statistiky za platnosti $H_0$ a $H_1$. Vysoký překryv snižuje sílu testu
- Žádoucí hodnota 1-$\beta$ u testů je > 80 %
- Testy s jednostrannou $H_1$ mají vyšší sílu než oboustranné $H_1$
- Testy jednovýběrové mají vyšší sílu než dvou-výběrové na shodných datech

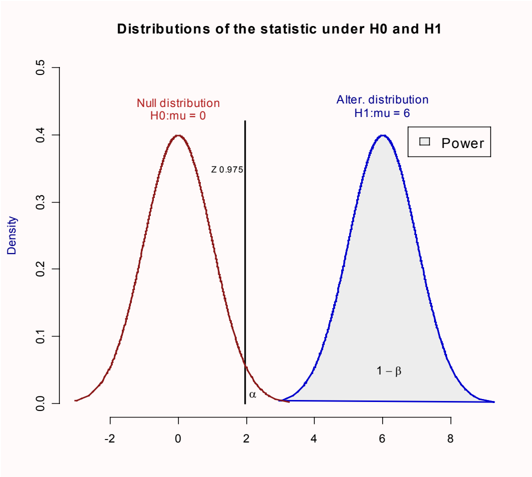
## Rozsah souboru a síla testu




## Vliv překrytí rozdělení za $H_0$ a $H_1$



# Malé překrytí rozdělení za $H_0$ a $H_1$




# Síla testu

- Negativně působí na sílu testu (obecně):
  - Heterogenita variancí
  - Sériové závislosti residuí (pod-specifikovaný model)
  - Odchylky od normality (extrémní šikmost)
  - Vysoký počet proměnných v testovací statistice

# P-value

- Je <u>pravděpodobnost</u> výskytu extrémnější hodnoty testovací statistiky, než té, která byla empiricky vypočtena
- Synonyma: vypočtená průkaznost nebo vypočtená alfa
- Účel: rovnocenné kritérium (ne)zamítnutí $H_0$
- Testovací kritérium: p-value < alfa → zamítáme $H_0$
- Hodnota p-value testu je určena typem $H_1$ a parametry rozdělení testovací statistiky, z kterého se zjišťuje.

# Úskalí statistických testů

- Zamítnutí $H_0$ závisí na:
  - rozsahu souboru
  - zvolené hodnotě alfa
  - variabilitě dat
  - parametrech rozdělení testovací statistiky za platnosti $H_0$ a $H_1$
- Manipulací těchto veličin lze dosáhnout <u>zamítnutí každé $H_0$</u> !
- Statisticky průkazný rozdíl neznamená prakticky významný rozdíl !

## Z-test nebo t-test ?

- Fakt: Lokační míry lze odhadovat relativně snadněji a spolehlivěji než dispersní míry
- Z - testy o parametrech se používají v situacích, je-li známa $\sigma^2$, tedy kvalitní odhad rozptylu
- Výběrová rozdělení se mohou lišit v lokační míře, avšak mají <u>stejný</u> a <u>známý rozptyl</u>
- T-testy o parametrech se používají v situacích, není-li známa $\sigma^2$, tedy když máme méně kvalitní odhad rozptylu $s^2$
- Výběrová rozdělení se mohou lišit v lokační míře, avšak mají <u>stejný</u> ale <u>neznámý rozptyl</u>
- Rozptyly z dvou nebo více výběrů mohou být <u>stejnorodé</u> nebo <u>různorodé</u>

## Testy na normalitu

- <u>Test dobré shody:</u> neparametrický aproximační test vycházející z $\chi^2$ rozdělení
- Použitelný pro více typů rozdělení, <u>především nespojitá</u>
- Využívá intervalového třídění výběru do p tříd ($p \approx n^{1/2}$)
- Kalkulace empirických (O) a očekávaných (E) četností (z funkce F(x))

$$\chi^2_{obs} = \sum_{i=1}^{p} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{p-1}$$

- <u>Shapiro – Wilks test</u>: Snad <u>nejlepší test normality</u>
- Kalkuluje statistiku W podobnou čtverci korelace mezi empirickými a teoretickými kvantilami
- Test je levostranný ! <u>Nízká hodnota W</u> (již okolo 0.9 !) zamítá $H_0$: $Y \sim N(\mu,\sigma^2)$.
- <u>Kolmogorov – Smirnov test</u>: test pro jeden (K) nebo dva soubory (S) a <u>spojitá</u> rozdělení.
- Porovnává $F_{(y)}$ empirického a rozdělení $H_0$:
- Vysoká hodnota D zamítá $H_0$. (pravostranný test)

$$D = \frac{\max\left|\hat{F}(y) - F_0(y)\right|}{n}$$

## Předpoklady testů

- Normální rozdělení výběrového souboru s daným průměrem a rozptylem
- Konstantnost a sériová nezávislost rozdělení hodnot (I.I.D.)
- Obecná formule Z – testu:

$$Z_{obs} = \frac{\theta - \theta_0}{se(\theta - \theta_0)} \sim N(0,1)$$

- Obecná formule T – testu:

$$t_{obs} = \frac{\theta - \theta_0}{se(\theta - \theta_0)} \sim t_v$$

- Pozn.: $\theta$ - $\theta_0$ rozdíl parametrů; se($\theta$ - $\theta_0$) = střední chyba rozdílu

## Jedno-výběrový z-test a t-test

- <u>Z - test</u>: Testujeme, zda-li se průměr výběru liší od specifikované hodnoty $\mu_0$, za dostupnosti <u>rozptylu populace</u> $\sigma^2$.

$$H_0 : \mu = \mu_0 \qquad Z_{obs} = \frac{\mu - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

- Oblasti zamítnutí $H_0$: $H_1 : \mu > \mu_0 \rightarrow z_{obs} > z_{1-\alpha}$ $\quad H_1 : \mu < \mu_0 \rightarrow z_{obs} < z_{\alpha}$

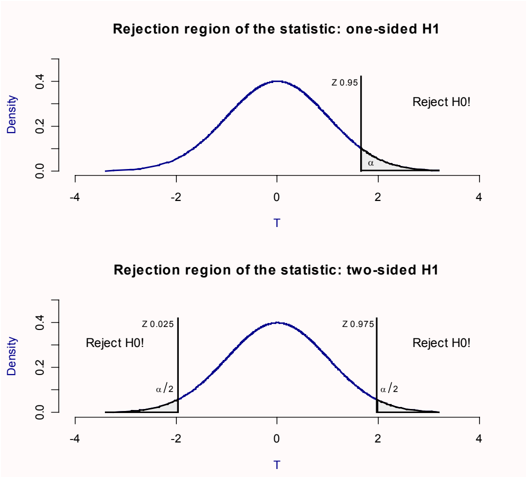
$H_1 : \mu \neq \mu_0 \rightarrow |z_{obs}| > z_{1-\alpha/2}$

- <u>T - test</u>: Testujeme, zda-li se průměr výběru liší od specifikované hodnoty $\mu_0$, za dostupnosti <u>výběrového rozptylu</u> $s^2$.

$$t_{obs} = \frac{\mu - \mu_0}{s / \sqrt{n}} \sim t_{n-1}$$

$H_1 : \mu \neq \mu_0 \rightarrow |t_{obs}| > t_{1-\alpha/2;n-1}$ $\quad H_1 : \mu > \mu_0 \rightarrow t_{obs} > t_{1-\alpha;n-1}$ $\quad H_1 : \mu < \mu_0 \rightarrow t_{obs} < t_{\alpha;n-1}$

## Oblast zamítnutí $H_0$




## F-test o dvou rozptylech

- F - test homogenity dvou nezávislých souborů

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \qquad H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \qquad H_1: \frac{\sigma_1^2}{\sigma_2^2} < 1 \qquad H_1: \frac{\sigma_1^2}{\sigma_2^2} > 1$$

- Testovací statistika: $$F_{obs} = \frac{s_1^2}{s_2^2} \sim F_{n_1-1;n_2-1}$$
- Oblasti zamítnutí $H_0$:
  - Oboustranná $H_1$: $F_{obs} < F_{\alpha/2;n_1-1;n_2-1}$ $\quad F_{obs} > F_{1-\alpha/2;n_1-1;n_2-1}$
  - Pravostranná $H_1$: $F_{obs} > F_{1-\alpha;n_1-1;n_2-1}$
  - Levostranná $H_1$: $F_{obs} < F_{\alpha;n_1-1;n_2-1}$

## Dvou-výběrový z-test

- Dva nezávislé soubory o průměrech a známých rozptylech $\sigma_1^2$ a $\sigma_2^2$.

$$\bar{y}_1 \sim N(\mu_1, \sigma_1^2/n_1) \qquad \bar{y}_2 \sim N(\mu_2, \sigma_2^2/n_2) \qquad \bar{y}_1 - \bar{y}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$H_0: \mu_1 = \mu_2 \qquad H_1: \mu_1 \neq \mu_2 \qquad H_1: \mu_1 < \mu_2 \qquad H_1: \mu_1 > \mu_2$$

- Testovací statistika: $$Z_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$
- Oblast zamítnutí $H_0$ je shodná s jedno-výběrovým z-testem

## Dvou-výběrový t-test (sdružený)

- Dva nezávislé soubory o průměrech a neznámých rozptylech $s_1^2$ a $s_2^2$. Rozptyly jsou si rovny!
- Sdružený odhad rozptylu $\sigma^2$: $$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$
- $S_p^2$ je odhadem společné variance obou souborů
- Za podmínky normality: $$\bar{y}_1 - \bar{y}_2 \sim N(\mu_1 - \mu_2, \sigma_p^2(\frac{1}{n_1} + \frac{1}{n_2}))$$
- Testovací statistika: $$t_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

# Dvou-výběrový t-test (Welschův)

- Dva nezávislé soubory o průměrech a neznámých rozptylech $s_1^2$ a $s_2^2$. Rozptyly si rovny nejsou !
- Welschova testovací statistika: $$w_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \sim t_\nu$$
- Stupně volnosti $\nu$ se aproximují podle Satterthwaita:

$$\nu = \left[ \frac{\left\{ \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right\}^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right]$$

- Welschův t - test vykazuje méně $\nu$ a proto má i nižší $1 - \beta$ než sdružený t-test.
- Použití: Je-li jedna z variancí je více než trojnásobek druhé

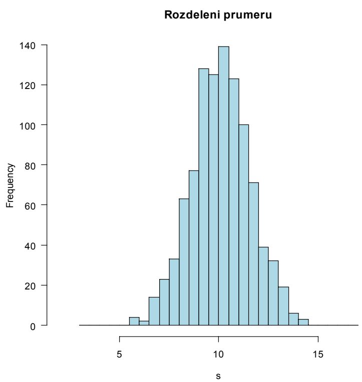
# Centrální limitní věta

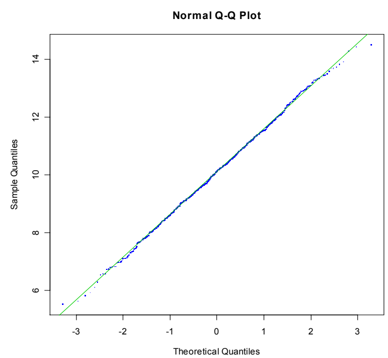
- Znění: Jestliže $y_i$ je náhodný výběrový soubor o rozsahu n, společné střední hodnotě $\mu$ a varianci $\sigma^2$, pak rozdělení výběrových průměrů aproximuje Gaussovo rozdělení o průměru $\mu$ a varianci $\sigma^2/n$ pokud se n přibližuje asymptotě ($\infty$).
- Důsledek: výběrové průměry souborů pocházející z kteréhokoliv rozdělení mají přibližně normální rozdělení, je-li n dostatečně velké

$$z = \frac{\bar{y} - E(y_i)}{\sqrt{Var(y_i)/n}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- Výhoda: možnost normální aproximace při dostatečném rozsahu
- Minimální rozsah n závisí na typu rozdělení (menší pro souměrná, větší pro nesouměrná)

# 1000 výběrů,$n_i$=10,a=0,b=20,uniformní densita




# Z-test o jednom výběrovém poměru

- Aplikace CLV:
- Testujeme, zda se výběrový poměr $\pi$ rovná populačnímu poměru $\pi_0$

$$H_0 : \pi = \pi_0 \qquad H_1 : \pi \neq \pi_0 \qquad H_1 : \pi > \pi_0 \qquad H_1 : \pi < \pi_0$$

- Testovací statistika: $$z_{obs} = \frac{\pi - \pi_0 \pm 0.5/n}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$
- n představuje rozsah výběrového souboru
- Korekce na kontinuitu v čitateli (Yates) když n*min($\pi$,1-$\pi$) < 5
  - Je záporná o hodnotě - 0.5/n když $\pi > \pi_0$
  - Je kladná o hodnotě + 0.5/n když $\pi < \pi_0$

## Z-test rovnosti dvou výběrových poměrů

- Testujeme, zda se výběrový poměr $\pi_1$ rovná výběrovému poměru $\pi_2$

$$H_0: \pi_1 - \pi_2 = 0 \qquad H_1: \pi_1 - \pi_2 \neq 0 \qquad H_1: \pi_1 - \pi_2 > 0 \qquad H_1: \pi_1 - \pi_2 < 0$$

- Testovací statistika:

$$z_{obs} = \frac{\pi_1 - \pi_2}{\sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

- $n_1$ a $n_2$ představují rozsahy výběrových souborů
- CLV aplikovatelná, když min($\pi$,1-$\pi$) * min($n_1$,$n_2$) ≥ 5
- Test je ekvivalentní testu rovnosti obou výběrových poměrů $\pi_1$ a $\pi_2$ sdruženému poměru $\pi$. Odhad $\pi$ se vypočítá:

$$\hat{\pi} = \frac{\sum_i n_{uspech}}{n}$$

- Ekvivalentní je Chi-kvadrát test 2 x 2 kontingenční tabulky

$$\chi^2_{obs} = \frac{n_{..}(n_{11}n_{00} - n_{10}n_{01})^2}{n_{1.}n_{0.}n_{.1}n_{.0}} \sim \chi^2_1$$

## Příklad testu dvou výběrových poměrů

Počty albínů ve 2 populacích: 8 a 15
Rozsahy populací: 67 a 78
Výběrové poměry populací: 8 / 67 = 0,1194 a 15 / 68 = 0,1923
Sdružený poměr populací: 23 / 145 = 0,1586
Testovací statistika : (0,1194 - 0,1923)/odm(0,1586*(1-0,1586)*(1/67 + 1/78))
= -1,198
P-value = 0,231 (dvoustranná $H_1$: rozdíly neprůkazné)

## $\chi^2$ test o rovnosti více poměrů

- Testujeme, zda se výběrové poměry $\pi_1$ $\pi_2$ ... $\pi_p$ rovnají

$$H_0: \pi_1 = \pi_2 = \ldots = \pi_p \qquad H_1: nerovnost$$

- Výpočet <u>sdružené hodnoty</u> $\pi$

$$\hat{\pi} = \frac{\sum_i n_{uspech}}{\sum_i n_i}$$

- Výpočet <u>dílčích statistik</u> $z_i$

$$z_i = \frac{\pi_i - \pi \pm 0.5/n_i}{\sqrt{\frac{\pi(1-\pi)}{n_i}}} \sim N(0,1)$$

- Souhrnná statistika $\chi^2$ je pak

$$\chi^2_{obs} = \sum_{i=1}^{p} z_i^2 \sim \chi^2_{p-1}$$

- Řešení možné i $\chi^2$ testem (2 x p) kontingenční tabulky

## Rozšíření na polychotomický případ

- Testujeme, zda se multinomické výběrové poměry $\pi_1$ $\pi_2$ ... $\pi_j$ rovnají mezi k skupinami
- Nejsnazší řešení přes rekonstrukci kontingenční tabulky a $\chi^2$ test nezávislosti
- Příklad:

Ve čtyřech výběrech z různých populací laboratorních myší ($n_1$ = 120; $n_2$ =120; $n_3$ = 96; $n_4$= 116; ) byly pozorovány relativní četnosti alel A1, A2 a A3. Testujte, zda se četnosti liší mezi populacemi.

## Potřebný rozsah náhodného výběru

- Minimální rozsah výběru (n) lze určit úpravou vzorců testovací statistiky nebo tzv. přípustné chyby $\Delta$ používané k sestavení (1-$\alpha$)*100 % intervalu spolehlivosti.
- Výsledné hodnoty n se obvykle zaokrouhlují nahoru
- jednovýběrový z - test jednostranné $H_1$: $$n = \frac{\left(z_{(1-\alpha)} + z_{(1-\beta)}\right)^2 \cdot \sigma^2}{\Delta^2}$$
- jednovýběrový z - test dvoustranné $H_1$: $$n = \frac{\left(z_{(1-\alpha/2)} + z_{(1-\beta)}\right)^2 \cdot \sigma^2}{\Delta^2}$$
- dvouvýběrový z - test jednostranné $H_1$: $$n = \frac{\left(z_{(1-\alpha)} + z_{(1-\beta)}\right)^2 \cdot (\sigma_1^2 + \sigma_2^2 / k)}{\Delta^2}$$
- dvouvýběrový z - test dvoustranné $H_1$: $$n = \frac{\left(z_{(1-\alpha/2)} + z_{(1-\beta)}\right)^2 \cdot (\sigma_1^2 + \sigma_2^2 / k)}{\Delta^2}$$
- k = koeficient $n_1$ / $n_2$
- V případě t-testu: n se násobí ($\nu$ + 3)/($\nu$ + 1), kde $\nu$ jsou stupně volnosti pro chybu (Cox and Cochran).

## Příklady na rozsah výběru

- Rozpětí křídel dospělých jedinců jistého ptačího druhu je v průměru $\mu_0$ = 48 cm. Biolog chce objevit nový poddruh, který se liší nejméně o 4 cm. Kolik jedinců musí být odchyceno a změřeno, jestliže $\sigma$ rozpětí je 6,5 cm? Předpokládáme 1-$\beta$ = 0,95, $\alpha$ = 0,05.

  jednostranná $H_1$: n = $(z_{0,95} + z_{0,95})^2 * 6,5^2 / 4^2$ = 28,5773 ≈ 29

  dvoustranná $H_1$: n = $(z_{0,975} + z_{0,95})^2 * 6,5^2 / 4^2$ = 34,3142 ≈ 35

- Chceme zjistit jestli se četnost homozygotů alely jistého genu u importované landrace průkazně liší od domácí populace, kde je známá frekvence $p_0$ = 30%. Kolik jedinců je potřeba genotypovat, chceme-li zjistit minimální rozdíl 3%. Předpokládáme 1-$\beta$ = 0,90, $\alpha$ = 0,05, maximální varianci.

  jednostranná $H_1$: n = $(z_{0,95} + z_{0,90})^2 * 0,5^2 / 0,03^2$ = 2378,85 ≈ 2379

  dvoustranná $H_1$: n = $(z_{0,975} + z_{0,90})^2 * 0,5^2 / 0,03^2$ = 2918.73 ≈ 2919

Statistiky jsou nejvyšším vývojovým stádiem lži...