Základy pravděpodobnosti a matematické statistiky

10. Testování statistických hypotéz

FAKULTA STROJNÍ ČVUT V PRAZE

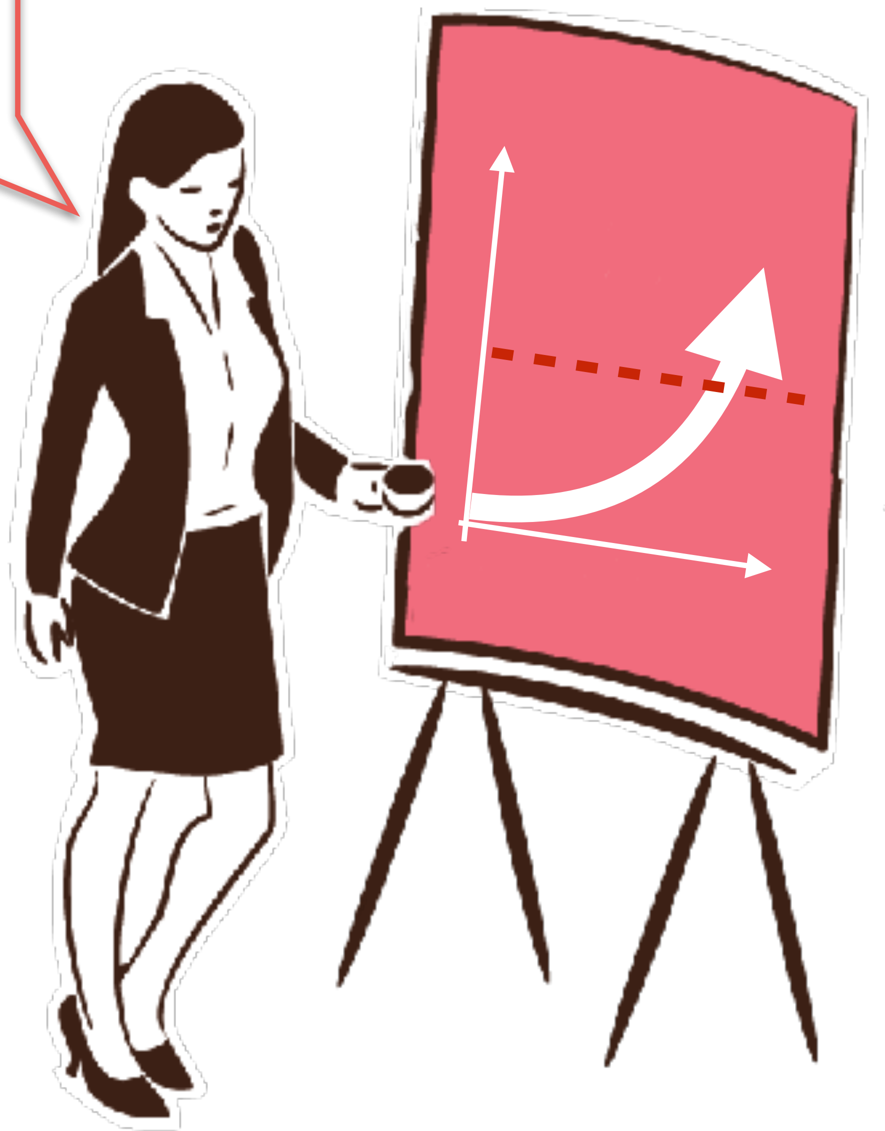
prof. RNDr. Gejza Dohnal, CSc.

ak. rok 2023/2024

10. Testování statistických hypotéz

Testování statistické hypotézy

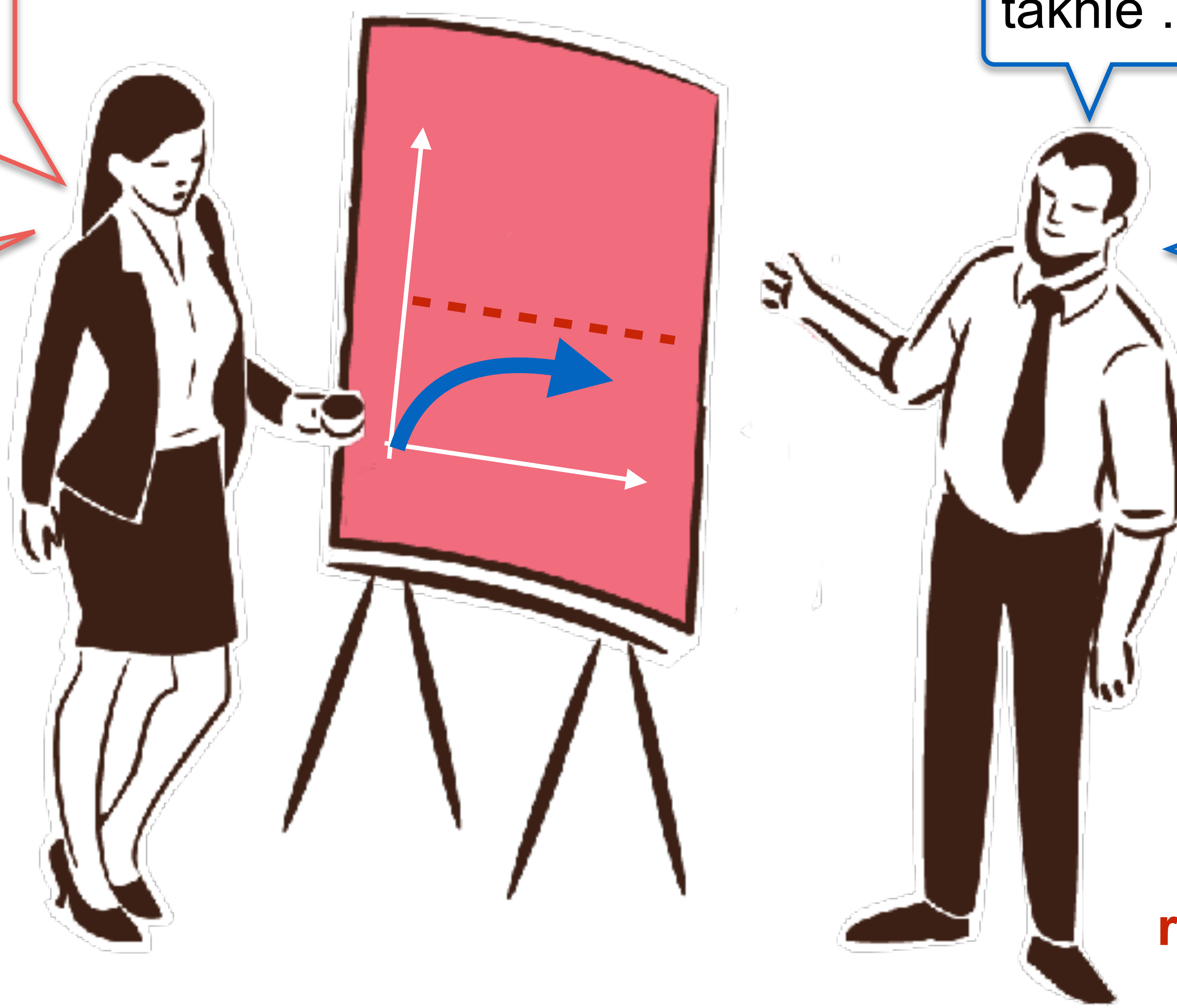
Na základě našich zkušeností jsme usoudili, že …



Testování statistické hypotézy

Na základě našich zkušeností jsme usoudili, že …

… tak si to ověřte …



A co kdyby to bylo takhle ….

Provedeme měření a uvidíme, jestli mám pravdu …

oba se ovšem musejí dohodnout na rozhodovacím kriteriu, aby mohli rozhodnout, kdo má pravdu

Provedli měření $X_1, X_2, \ldots, X_n$ a ukázalo se, že:

a) ON má pravdu => **Její hypotéza se tím vyvrací!**

b) ONA má pravdu => no ale to může být pouze náhoda! To ještě hypotézu nepotvrzuje => **takže ji tedy nelze vyvrátit.**

Testování statistické hypotézy

Příklad: Jednovýběrový t-test o střední hodnotě

1) Zformulujeme tzv. Nulovou hypotézu: $H_0: \quad \mu = m_0$

2) Zformulujeme tzv. Alternativní hypotézu: $H_A: \quad \mu \neq m_0$ (tzv. oboustrannou)

3) Provedeme n nezávislých měření $X_1, X_2, \ldots, X_n$ veličiny, které se hypotéza týká

4) Vytvoříme Testovou statistiku T:
(má t-rozdělení o (n-1) stupních volnosti)
$$T = \frac{\bar{X} - m_0}{s}\sqrt{n}$$

5) Zvolíme hladinu významnosti
a najdeme k ní kritickou hodnotu:
$$\alpha \quad \Rightarrow \quad t_\alpha(n-1)$$

6) Pomocí testové statistiky a kritické hodnoty
stanovíme rozhodovací pravidlo:
$$|T| \geq t_\alpha(n-1)$$

Nulová hypotéza se nepřijímá, pouze zamítá či nezamítá!

Statisticky nelze nic dokázat, pouze zamítnout!

$$\left|\frac{\bar{X} - m_0}{s}\sqrt{n}\right| \geq t_\alpha(n-1) \quad \Leftrightarrow \quad \frac{\bar{X} - m_0}{s}\sqrt{n} \notin \langle -t_\alpha(n-1), t_\alpha(n-1)\rangle$$

$$\bar{X} - m_0 \notin \langle -\frac{s}{\sqrt{n}}t_\alpha(n-1), \frac{s}{\sqrt{n}}t_\alpha(n-1)\rangle$$

$$\boxed{\bar{X} \notin \langle m_0 - \frac{s}{\sqrt{n}}t_\alpha(n-1), m_0 + \frac{s}{\sqrt{n}}t_\alpha(n-1)\rangle}$$

Hypotézu zamítáme, pokud naměřený průměr nepadne do intervalu spolehlivosti pro střední hodnotu m_0.

Testování statistické hypotézy

Příklad: Jednovýběrový t-test o střední hodnotě

1) Zformulujeme tzv. Nulovou hypotézu: $H_0: \quad \mu = m_0$

2) Zformulujeme tzv. Alternativní hypotézu: $H_A: \quad \mu \neq m_0$

3) Provedeme n nezávislých měření $X_1, X_2, \ldots, X_n$ veličiny, které se hypotéza týká

4) Vytvoříme Testovou statistiku T:
(má t-rozdělení o (n-1) stupních volnosti)

$$T = \frac{\bar{X} - m_0}{s}\sqrt{n}$$

5) Zvolíme hladinu významnosti
a najdeme k ní kritickou hodnotu:

$$\alpha \quad \Rightarrow \quad t_\alpha(n-1)$$

6) Pomocí testové statistiky a kritické hodnoty
stanovíme rozhodovací pravidlo:

$$|T| \geq t_\alpha(n-1)$$

Co je hladina významnosti α?

Chyba I. druhu: Zamítneme nulovou hypotézu i když ve skutečnosti platí.

Chyba II. druhu: Nezamítneme nulovou hypotézu i když ve skutečnosti neplatí.

α = P(chyba I. druhu)= pravděpodobnost, že učiníme chybné rozhodnutí, pokud nulová hypotéza platí (jakási míra „přísnosti“)

Testování statistické hypotézy

Příklad: Jednovýběrový t-test o střední hodnotě

1) Zformulujeme tzv. Nulovou hypotézu: $H_0: \quad \mu = m_0$

2) Zformulujeme tzv. Alternativní hypotézu: $H_A: \quad \mu \neq m_0$

3) Provedeme n nezávislých měření $X_1, X_2, \ldots, X_n$ veličiny, které se hypotéza týká

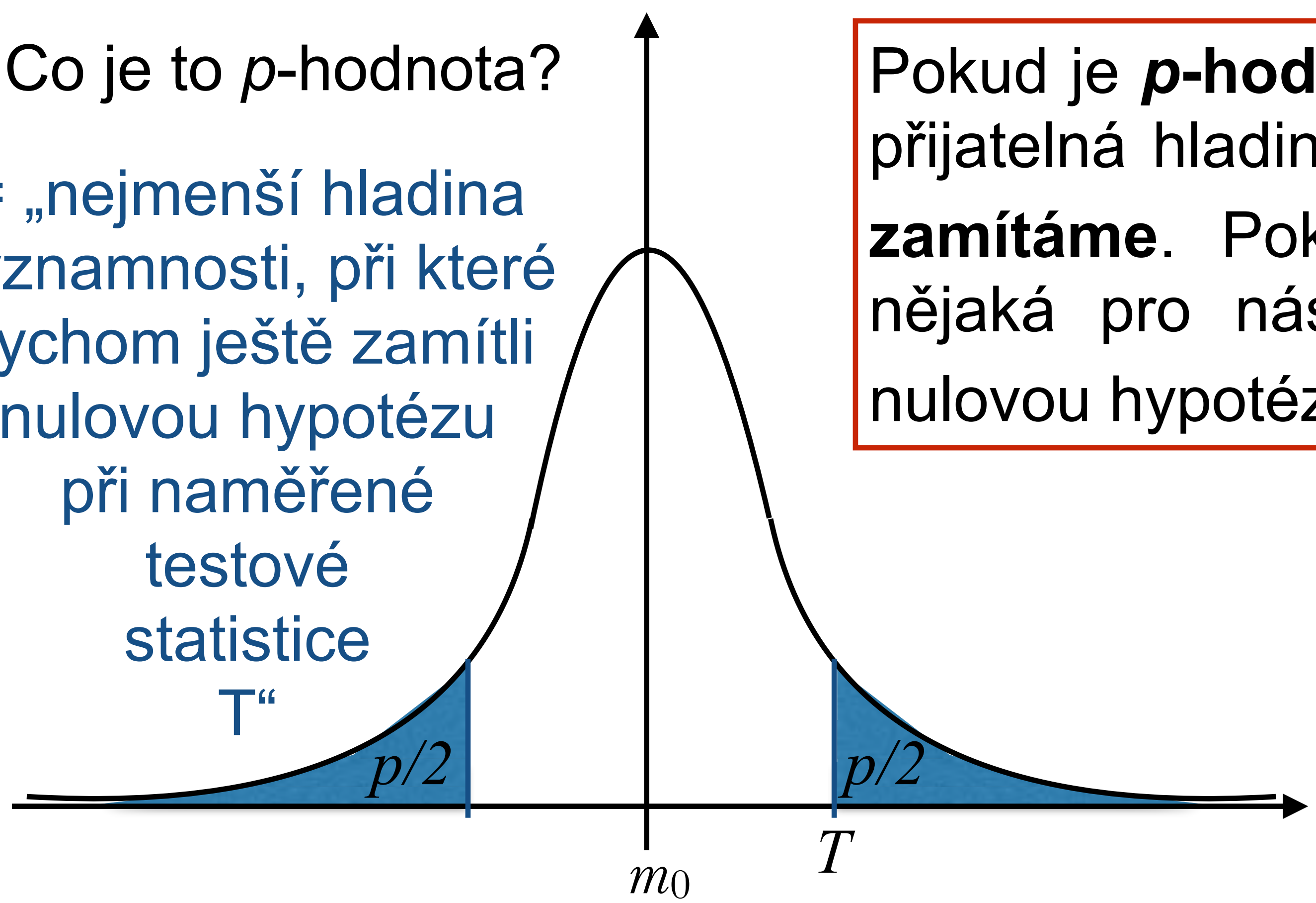
4) Vytvoříme Testovou statistiku T:
(má t-rozdělení o (n-1) stupních volnosti)

$$T = \frac{\bar{X} - m_0}{s}\sqrt{n}$$

5) Spočteme *p*-hodnotu a tu porovnáváme s naší volbou hladiny významnosti.

Co je to *p*-hodnota?

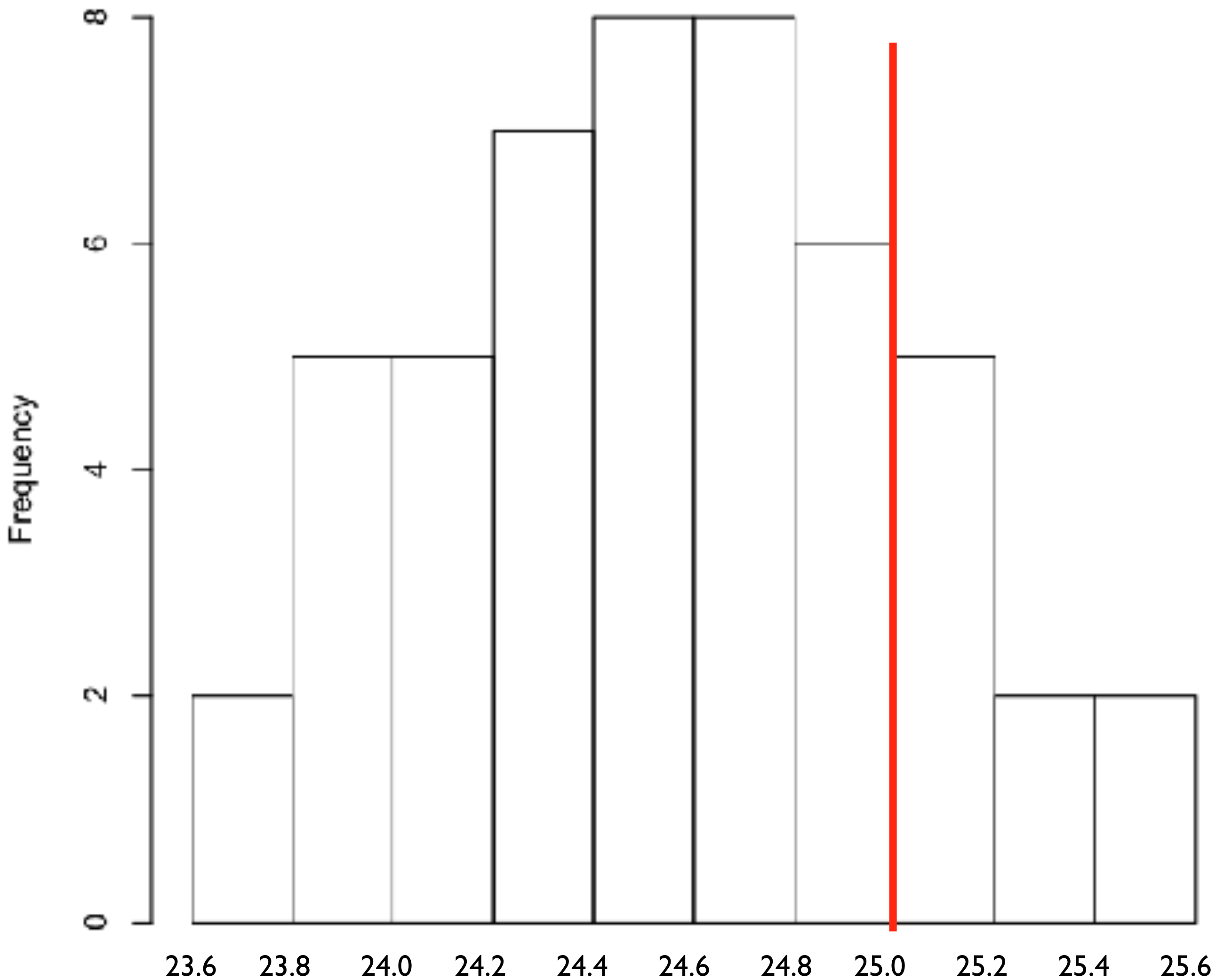
= „nejmenší hladina významnosti, při které bychom ještě zamítli nulovou hypotézu při naměřené testové statistice T“



Pokud je ***p*-hodnota malá** (menší než nějaká pro nás přijatelná hladina významnosti α), nulovou hypotézu **zamítáme**. Pokud je **p-hodnota velká** (větší než nějaká pro nás přijatelná hladina významnosti α), nulovou hypotézu **nezamítáme.**

Jednovýběrový t-test o střední hodnotě

Hmotnosti obsahu balení, dávkovaného automatem



24.52586	24.17119
24.54486	24.44240
23.93455	24.20389
24.19974	24.34851
23.94024	24.21022
24.87474	25.06155
25.48924	25.32572
23.71721	24.61622
25.06676	24.90055
24.36213	24.98580
24.80591	24.20853
24.72623	24.64437
24.70405	23.97645
25.29837	24.46910
24.99453	25.42994
24.66147	24.75773
25.03970	24.44901
25.13285	24.40205
24.78721	23.83656
24.17186	23.65390
24.48244	24.68550
24.22988	23.83956
24.09777	24.52098
24.89240	24.25332
24.14259	25.12906

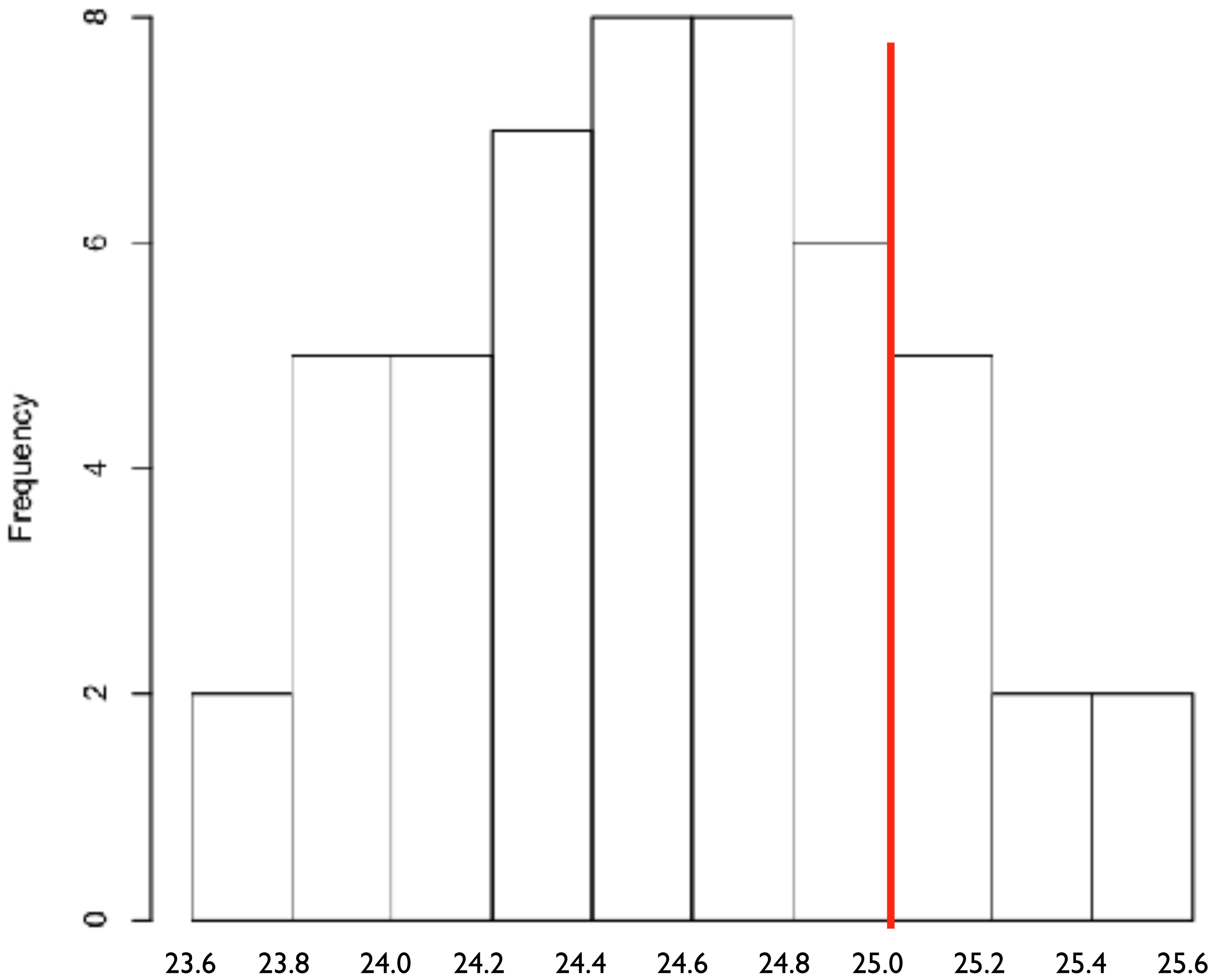
Lze považovat automat za správně seřízený (m_0=250 g)?

Pokusíme se ověřit, zda měření odpovídají rozdělení $N(m_0, s^2)$

… potom bychom mohli použít jednovýběrový t-test.

Jednovýběrový t-test o střední hodnotě

Hmotnosti obsahu balení, dávkovaného automatem

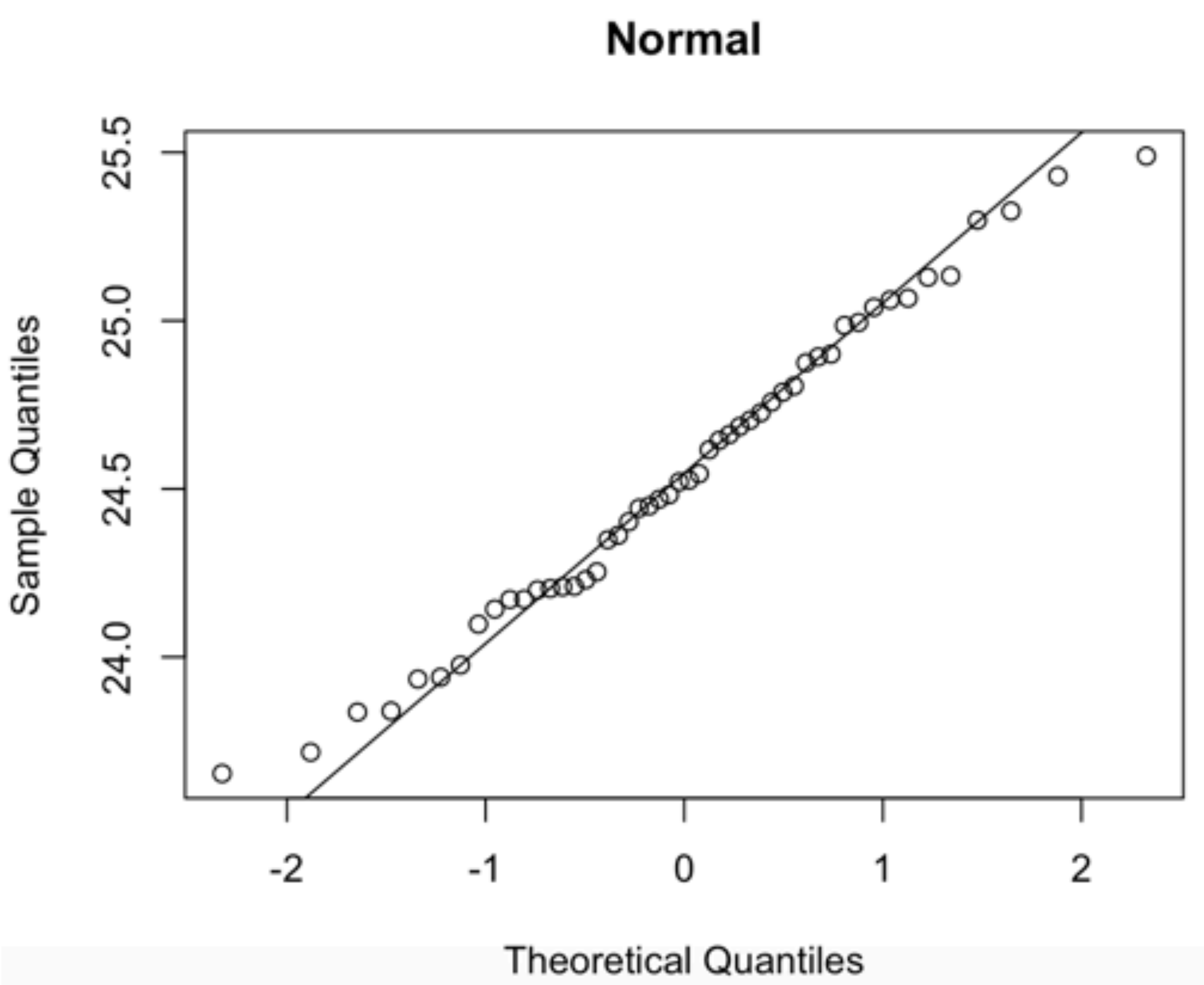


> x <- data.matrix(read.table("davkovac.txt"))

1) Nejprve ověříme normalitu rozdělení:

$$H_0 : X \sim N(\mu, \sigma^2)$$
$$H_A : X \nsim N(\mu, \sigma^2)$$

Jednovýběrový t-test o střední hodnotě



1) Nejprve ověříme normalitu rozdělení:

$$H_0 : X \sim N(\mu, \sigma^2)$$
$$H_A : X \nsim N(\mu, \sigma^2)$$

Z testů normality vybereme např. Shapiro-Wilkův test

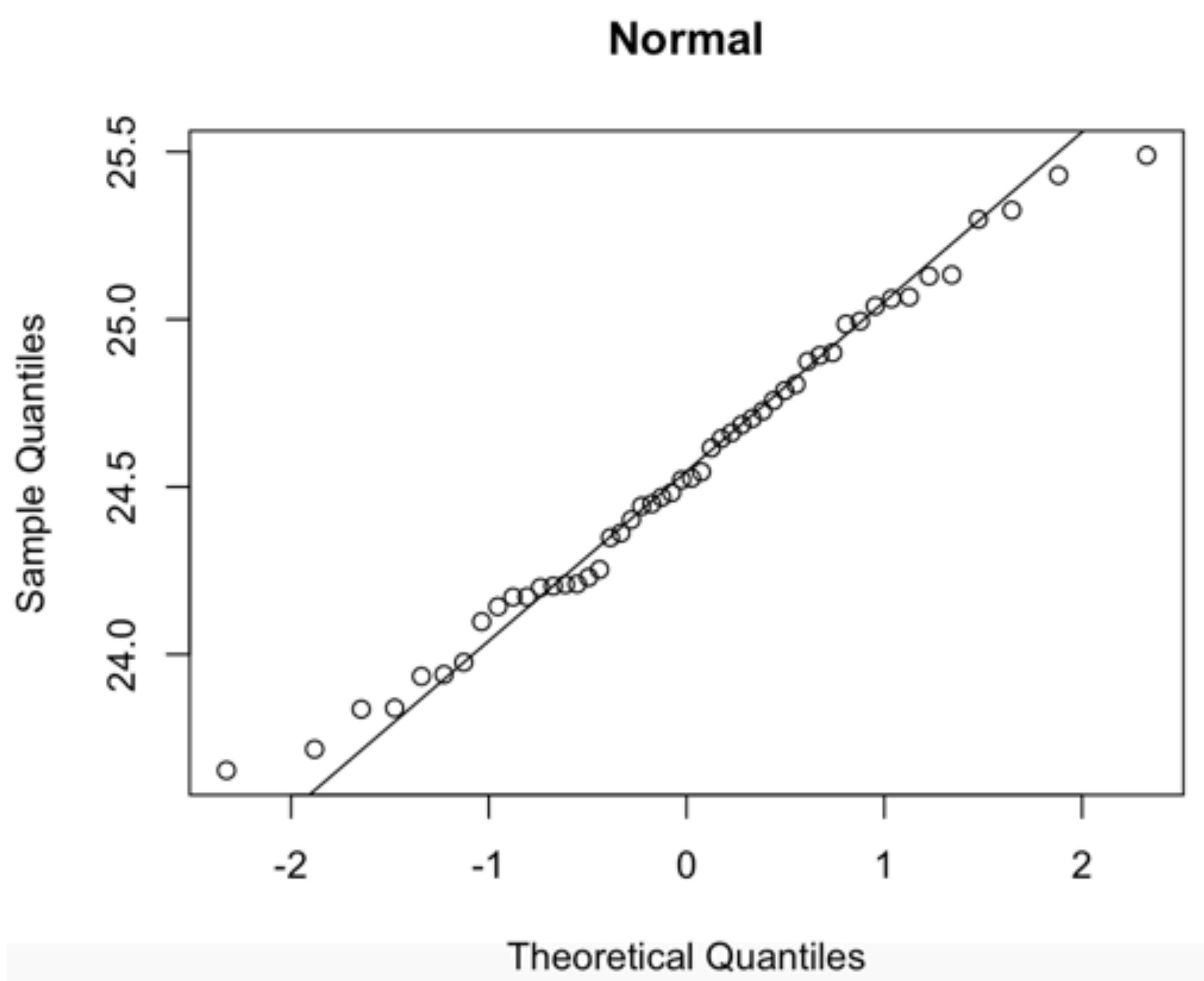
```
> shapiro.test(x)
```

Shapiro-Wilk normality test

data: x
W = 0.98406, p-value = 0.7307

```
> x <- data.matrix(read.table("davkovac.txt"))
> qqnorm(x, main='Normal')
> qqline(x)
```

Jednovýběrový t-test o střední hodnotě (při výběru z normálního rozdělení)



2) Zavoláme t-test:

$$H_0 : \mu = 25.0$$

$$H_A : \mu \neq 25.0$$

> t.test(x, mu=25.0)

One Sample t-test

data: x

t = -6.9875, df = 49,

p-value = 6.937e-09

alternative hypothesis:

true mean is not equal to 25

95 percent confidence interval:

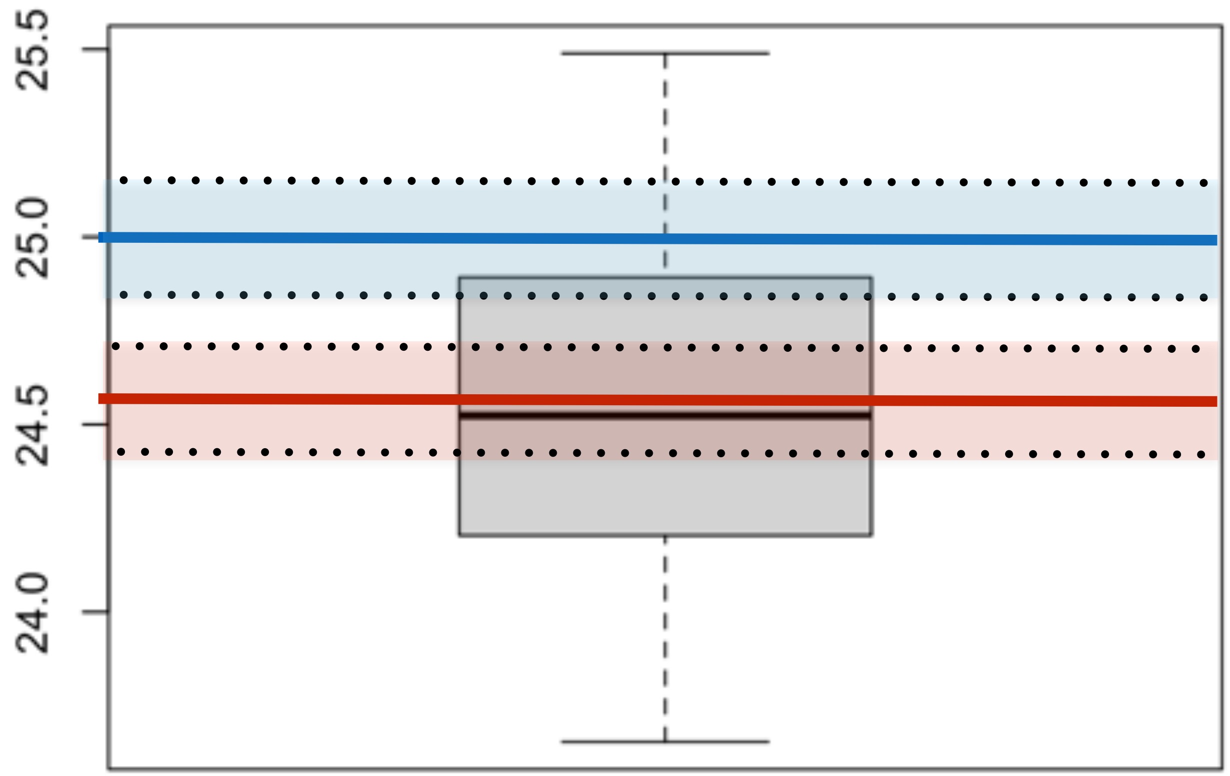
24.41658 24.67721

sample estimates:

mean of x

24.54689

Jednovýběrový t-test o střední hodnotě (při výběru z normálního rozdělení)



$$T = \frac{\bar{X} - m_0}{s}\sqrt{n} = \frac{24.5469 - 25.0}{0.4585}\sqrt{50} = -6.9875$$

$$\boxed{t_{0.975}(49) = 2.0096 < |T|}$$

$$SE = 0.1303$$

… tedy interval spolehlivosti pro průměr je

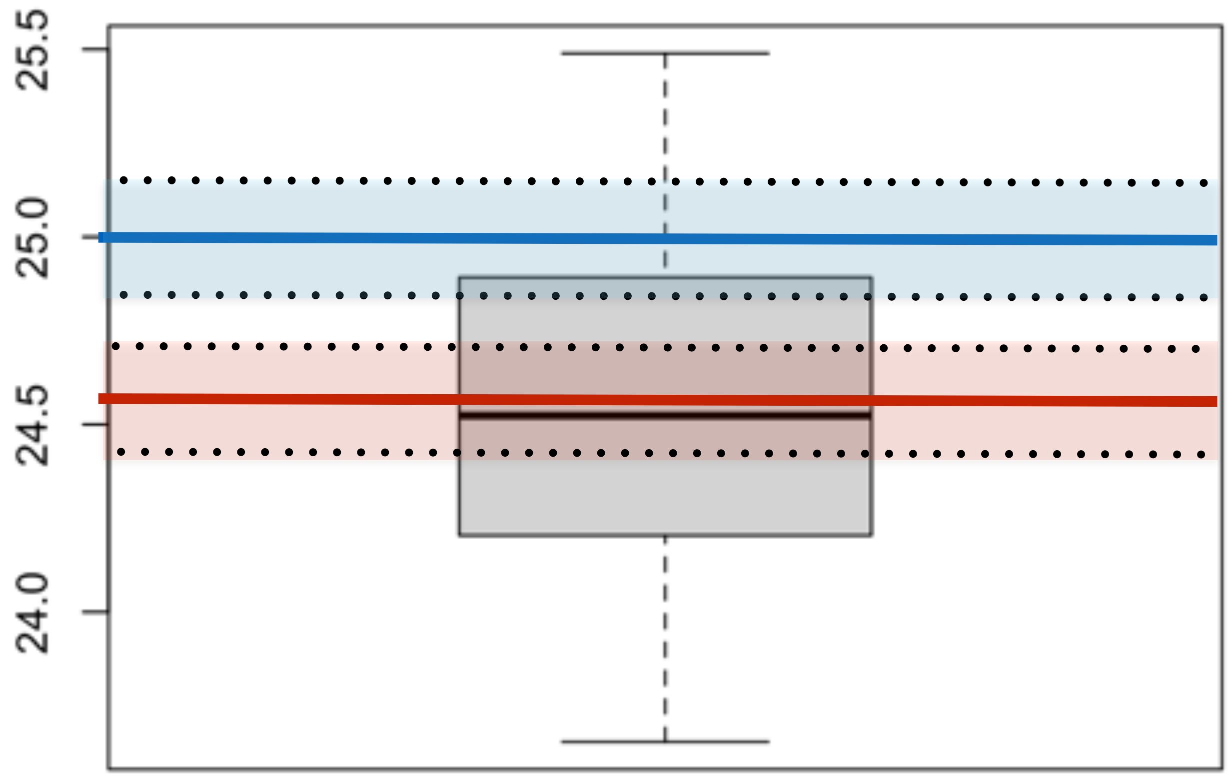
$$\bar{X} \notin (24.8697; 25.1303)$$

2) Zavoláme t-test:

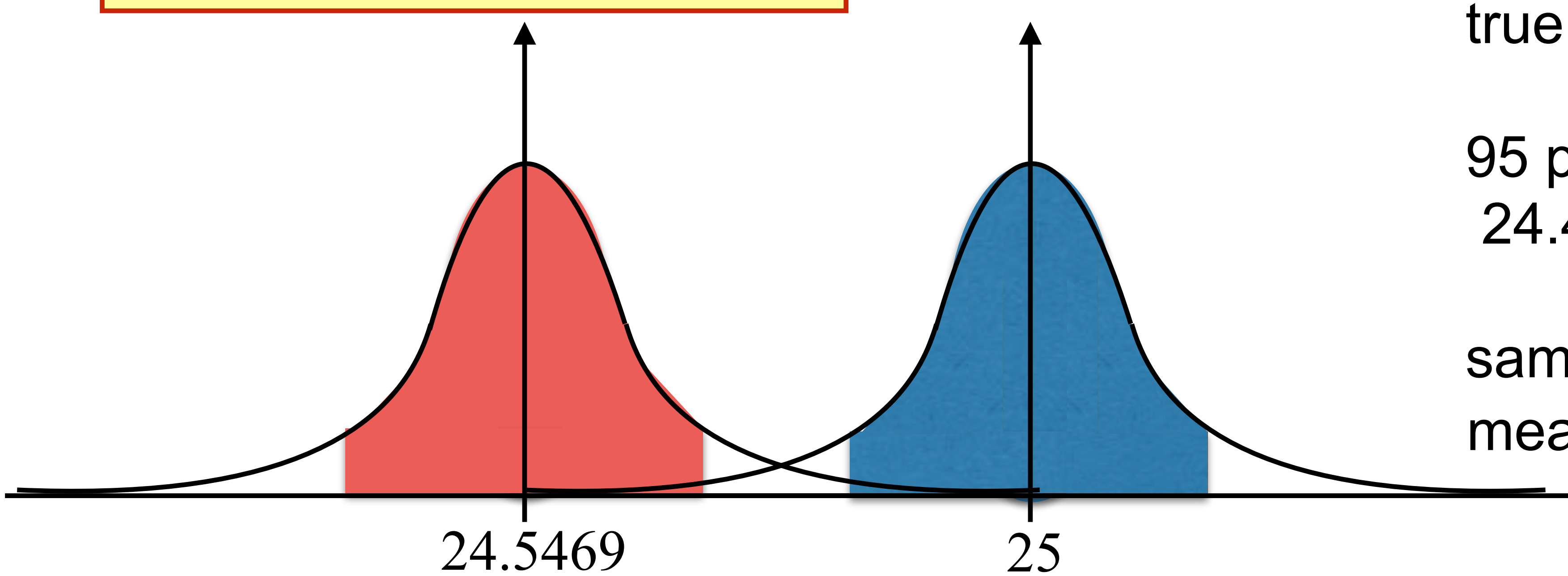
$$H_0 : \mu = 25.0$$
$$H_A : \mu \neq 25.0$$

```
> t.test(x, mu=25.0)

        One Sample t-test

data:  x
t = -6.9875, df = 49,
p-value = 6.937e-09

alternative hypothesis:
true mean is not equal to 25

95 percent confidence interval:
 24.41658 24.67721

sample estimates:
mean of x
 24.54689
```

Jednovýběrový t-test o střední hodnotě (při výběru z normálního rozdělení)



$$\boxed{|T| > t_{0,975}(49) = 2.0096}$$ = TINV(0,05; 49)



2) Zavoláme t-test:

$$H_0 : \mu = 25.0$$

$$H_A : \mu \neq 25.0$$

```
> t.test(x, mu=25,
  alternative = „two.sided")
```

One Sample t-test

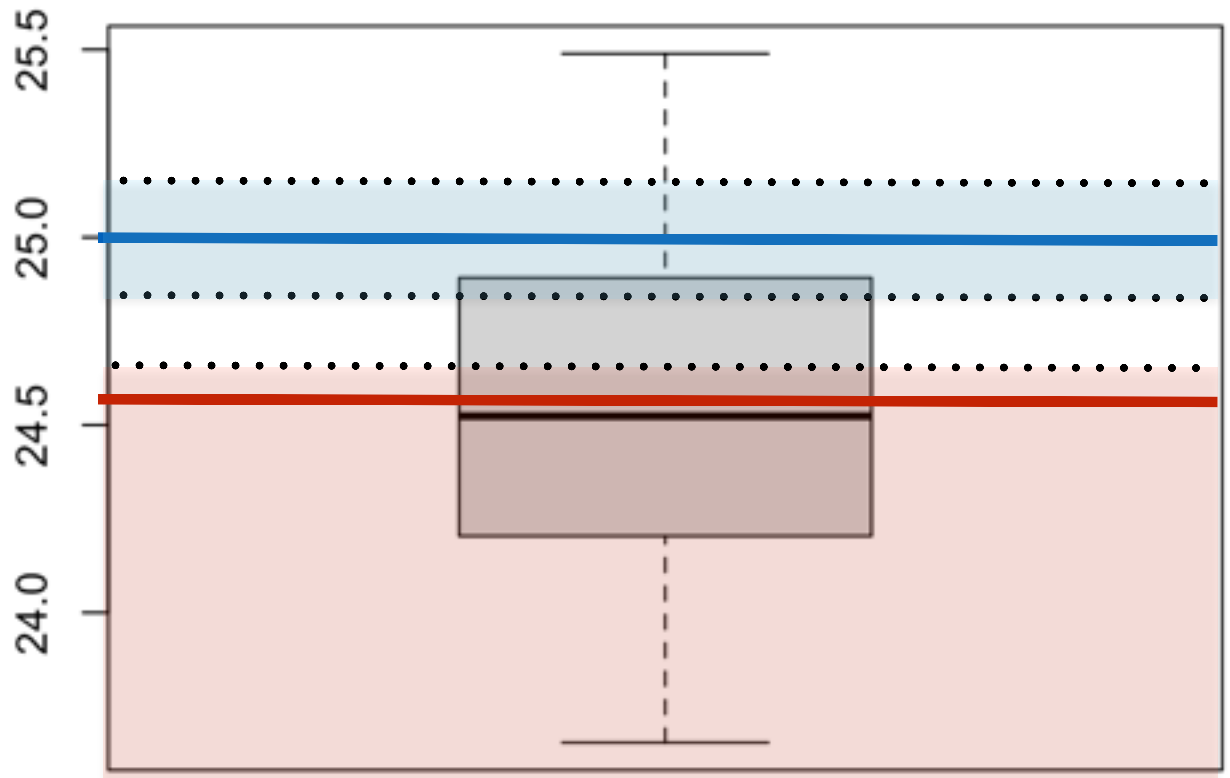
data: x
t = -6.9875, df = 49,
p-value = 6.937e-09

alternative hypothesis:
true mean is not equal to 25

95 percent confidence interval:
24.41658 24.67721

sample estimates:
mean of x = 24.54689

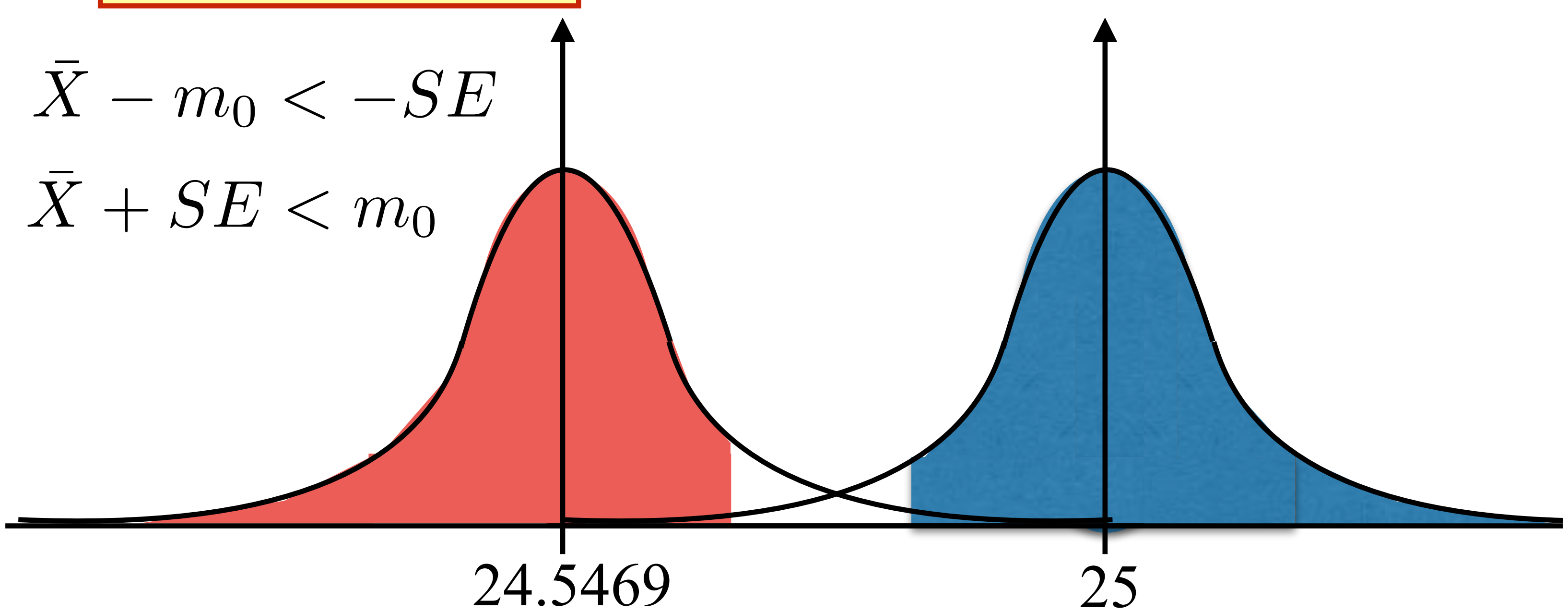
Jednovýběrový t-test o střední hodnotě (při výběru z normálního rozdělení)



$\boxed{T < -t_{0,95}(49)} = -1,6766$ = – TINV(0,1; 49)

$\bar{X} - m_0 < -SE$

$\bar{X} + SE < m_0$



2) Zavoláme t-test:

$$H_0 : \mu = 25.0$$

$$H_A : \mu < 25.0$$

> t.test(x, mu=25,
alternative = „less“)

One Sample t-test

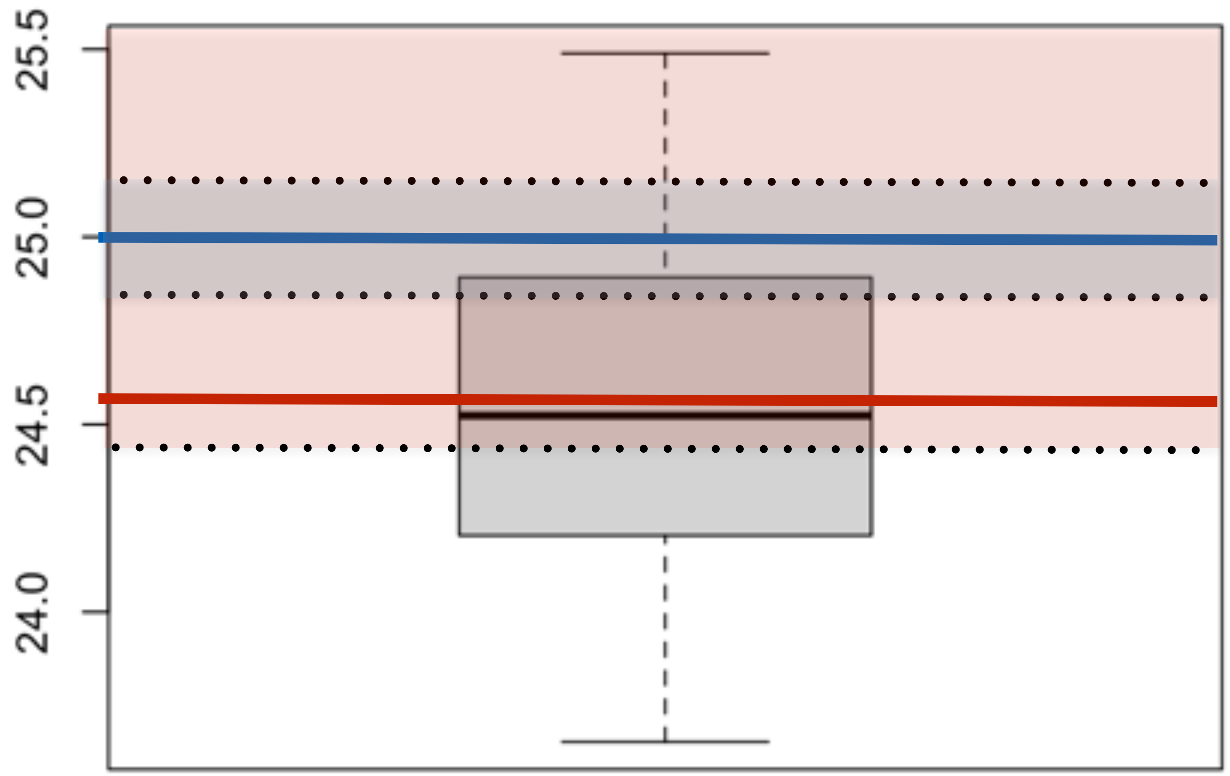
data: x
t = -6.9875, df = 49,
p-value = 3.469e-09

alternative hypothesis:
true mean is less than 25

95 percent confidence interval:
-Inf 24.65561

sample estimates:
mean of x = 24.54689

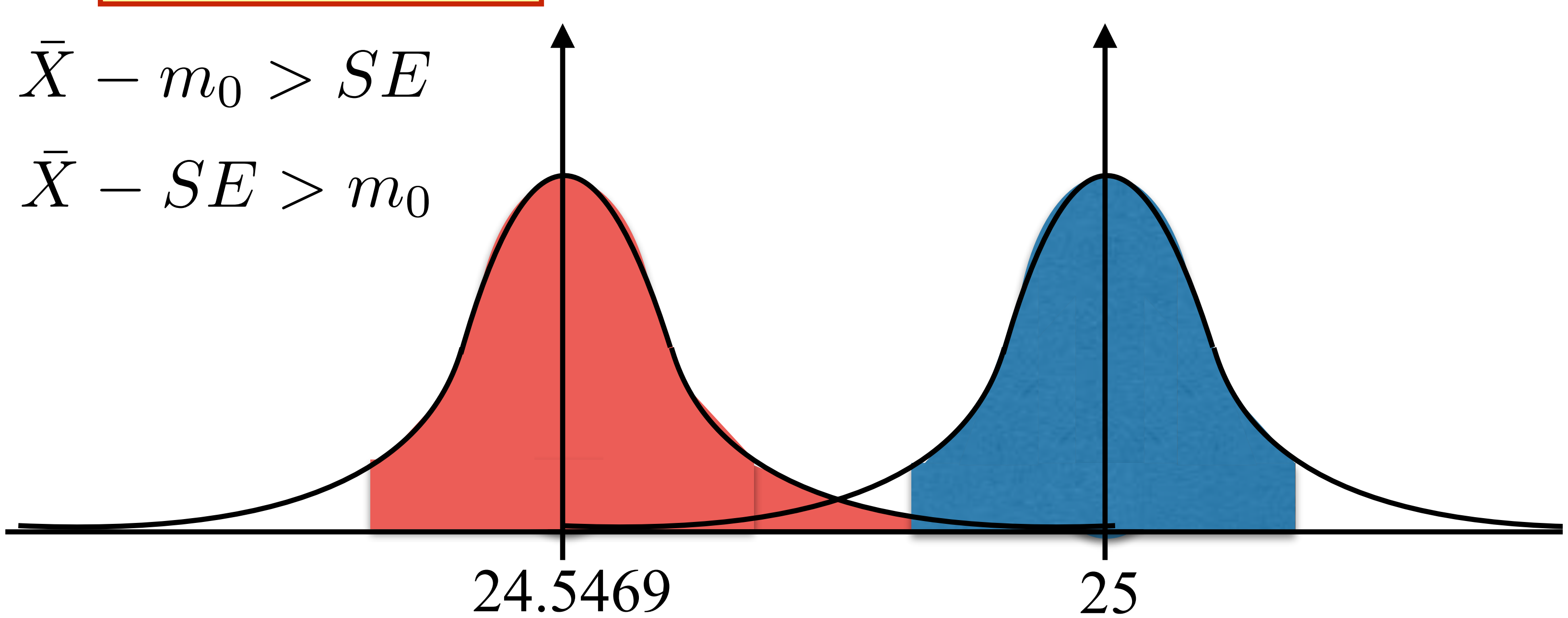
Jednovýběrový t-test o střední hodnotě (při výběru z normálního rozdělení)



$\boxed{T > t_{0,95}(49)} = 1,6766$

$\bar{X} - m_0 > SE$

$\bar{X} - SE > m_0$



2) Zavoláme t-test:

$H_0 : \mu = 25.0$

$H_A : \mu > 25.0$

> t.test(x, mu=25, alternative = „greater“)

One Sample t-test

data: x
t = -6.9875, df = 49,
p-value = 1

alternative hypothesis:
true mean is greater than 25

95 percent confidence interval:
24.43818 inf

sample estimates:
mean of x = 24.54689

Dvouvýběrové testy při výběru z normálního rozdělení

Chceme prokázat, že dvě náhodné veličiny s normálním rozdělením jsou statisticky srovnatelné na nějaké zvolené hladině významnosti

nebo

Snažíme se zjistit, zda normálně rozdělená náhodná veličina, kterou jsme měřili za dvou různých podmínek (pod vlivem dvou hodnot nějakého faktoru), se chovala statisticky významně různě nebo ne

nebo

Pokoušíme se porovnat dva soubory dat, které jsme dostali jako měření dvou normálně rozdělených náhodných veličin z hlediska jejich středních hodnot či rozptylů

Chceme zjistit, zda se dva soubory „normálních“ dat statisticky významně liší nebo jsou srovnatelné na určené hladině významnosti.

- K tomu je potřeba zjistit, zda jsou měření statisticky závislá (párovaná), či nikoli.
- Dále budeme předpokládat, že měřené veličiny mají normální rozdělení (je třeba ověřit)

Pozn. : Pokud nemají normální rozdělení, musíme použít jiné testy.

Dvouvýběrové testy při výběru z normálního rozdělení

1. Dvě nezávislá měření $X : X_1, X_2, \ldots, X_n$ $\qquad Y : Y_1, Y_2, \ldots, Y_m$

$$X \sim N(\mu_X, \sigma_X^2) \qquad Y \sim N(\mu_Y, \sigma_Y^2)$$

- **test shody rozptylů**
- **dvouvýběrový test shody středních hodnot při stejných rozptylech**
- **dvouvýběrový test shody středních hodnot při nestejných rozptylech**

2) Párová pozorování $X : X_1, X_2, \ldots, X_n$ $\qquad X \sim N(\mu_X, \sigma_X^2)$

$Y : Y_1, Y_2, \ldots, Y_n$ $\qquad Y \sim N(\mu_Y, \sigma_Y^2)$

- **párový test shody shody středních hodnot**

$$\Rightarrow \quad Z_1 = X_1 - Y_1,\ Z_2 = X_2 - Y_2, \ldots,\ Z_n = X_n - Y_n,$$

$$Z \sim N(\mu_X - \mu_Y, \sigma_Z^2)$$

Dvouvýběrové testy při výběru z normálního rozdělení

1) Test shody rozptylů

Liší se statisticky významně dvě nezávislá měření z hlediska velikosti rozptylu?

Lze považovat rozptyl dvou měření za shodný při dané hladině významnosti?

nulová hypotéza : $H_0 : \sigma_X^2 = \sigma_Y^2$

alternativní hypotéza: $H_A : \sigma_X^2 \neq \sigma_Y^2$

testová statistika : $F = \dfrac{s_X^2}{s_Y^2}$

hladina významnosti: α

F-test

Fisherovo-Snedecorovo rozdělení $F(n\text{-}1, m\text{-}1)$

H_0 nezamítneme, když pro dané α bude

$$F_{\alpha/2}(n-1, m-1) < F < F_{\alpha/2}(n-1, m-1)$$

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Liší se statisticky významně střední hodnoty dvou nezávislých měření?

Lze na dané hladině významnosti považovat střední hodnoty dvou měření za stejné?

Jsou dve měření statisticky rozlišitelná z hlediska střední hodnoty na dané hladině významnosti?

nulová hypotéza : $H_0 : \mu_X = \mu_Y$

alternativní hypotéza: $H_A : \mu_X \neq \mu_Y$ (oboustranná nebo jednostranná)

testová statistika : $T = \dfrac{\bar{X} - \bar{Y}}{s_{\bar{X}-\bar{Y}}}$

hladina významnosti: α

rozlišujeme dva případy (liší se odhadem $s_{\bar{X}-\bar{Y}}$ ve jmenovateli *T*):

- dvouvýběrový t-test se stejnými rozptyly, pokud $\sigma_X^2 = \sigma_Y^2$
- dvouvýběrový t-test s nestejnými rozptyly, je-li $\sigma_X^2 \neq \sigma_Y^2$

to zjistíme F-testem!

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

a) pokud $\sigma_X^2 = \sigma_Y^2$: $\quad = \sigma^2 \Rightarrow s_{\bar{X}-\bar{Y}}^2 = s_{\bar{X}}^2 + s_{\bar{Y}}^2 = \dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m} =$

$$= s^2\left(\frac{1}{n} + \frac{1}{m}\right) = s^2\left(\frac{m+n}{n.m}\right)$$

a odhadneme-li s^2 ze všech naměřených hodnot:

$$s^2 = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2\right) =$$

$$\frac{1}{n+m-2}\left((n-1)s_X^2 + (m-1)s_Y^2\right)$$

potom

$$s_{\bar{X}-\bar{Y}}^2 = \frac{n+m}{nm(n+m-2)}\left((n-1)s_X^2 + (m-1)s_Y^2\right)$$

a testová statistika bude mít tvar:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}}$$

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

b) pokud $\sigma_X^2 \neq \sigma_Y^2$: $\Rightarrow s^2_{\bar{X}-\bar{Y}} = s^2_{\bar{X}} + s^2_{\bar{Y}} = \frac{s_X^2}{n} + \frac{s_Y^2}{m} =$

a testová statistika bude mít jednodušší tvar:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n}s_X^2 + \frac{1}{m}s_Y^2}}$$

- V případě a) má testová statistika t-rozdělení (Studentovo rozdělení) pravděpodobnosti o $(n+m-2)$ stupních volnosti =>

H_0 nezamítneme, když pro dané α bude $|T| \leq t_\alpha(n+m-2)$, kde $t_\alpha(n+m-2)$ je (oboustranná) α-kritická hodnota t-rozdělení o $(n+m-2)$ stupních volnosti.

- V případě b) má testová statistika rozdělení, které je směsí t-rozdělení o $(n-1)$ a $(m-1)$ stupních volnosti. =>

H_0 nezamítneme, když pro dané α bude $|T| \leq At_\alpha(n-1) + Bt_\alpha(m-1)$, kde A a B jsou váhy

$$A = \frac{\frac{1}{n}s_X^2}{\frac{1}{n}s_X^2 + \frac{1}{m}s_Y^2}, \qquad B = \frac{\frac{1}{m}s_Y^2}{\frac{1}{n}s_X^2 + \frac{1}{m}s_Y^2}$$

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Příklad 1: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

Lze považovat délky tyčí od různých dodavatelů za shodné na hladině významnosti 5%?

Dodavatel X:

```
> x
  [1]  0.41379418  0.51040227  3.28722973  7.31995568  4.53994434 -1.07426821
  [7]  4.74575978  2.55201407  3.22058685 -1.17401554 -1.24119500  4.18294690
 [13]  0.65486399 -0.18908709 -0.73101186  1.27876451  1.26734875  2.78570344
 [19]  2.96834139  1.22145702  1.80851440 -0.80356569  2.57347292  3.42552806
 [25]  1.66904559 -2.21179295  4.17696270  2.15191523  3.62707736  0.06900211
 [31]  0.51371315  0.54983237  4.09554316  1.28465289  4.05350899  5.10504379
 [37]  4.25580572  0.79826235 -1.02042629  1.87299786  0.14051938  3.05622839
 [43]  4.74780021  4.54794140 -6.54132331  1.94429658  1.95488616  4.73267571
 [49]  4.83082378  2.95830720  2.99769818 -1.07337799  0.58403864  2.73050678
 [55]  0.28021230 10.49771713  2.36870296  0.60689702  8.42679434  1.29763889
 [61]  1.31289734  1.93230073  5.92597773  1.49746935  6.30721756  3.15585521
 [67]  5.38824907  3.27322441  3.41248356 -0.40437473  3.19350142 -4.06261001
 [73] -1.05763312 -0.39748962  0.86637433  2.02108109 -1.06445976  1.10375263
 [79]  4.51823259 -0.75725877 -0.87173075 -2.19932463  7.70167909  1.48655986
 [85]  4.90757730  5.51652338 -0.34615559  0.01031344  4.57582354  1.17516968
 [91] -0.21932558 -1.27848277  2.97655676  1.44863955  3.67881403  0.30868429
 [97] -2.52052309  0.05248743  0.07728483 -1.12975005  3.99585182  0.79045260
[103]  3.73159608  7.36490361  6.40646375 -1.54228149 -0.65100869  4.04305846
[109]  2.47766853 -3.48957597  6.20840771  0.40560482  0.49118447 -1.48277951
[115] -1.23675030  5.16138353  1.15383008  2.75286404  4.70183189 -2.29877355
```

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Příklad 1: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

Lze považovat délky tyčí od různých dodavatelů za shodné na hladině významnosti 5%?

Dodavatel Y:

```
> y
  [1]  6.65956934  2.78876119  0.33397602 -0.03763918  0.74993937  3.81490677
  [7]  1.70428804 -3.31291341 -0.22972370  4.02124752  5.93229834  3.30506070
 [13] -3.61277063  0.78809415  0.37976841  1.52357320  1.76230055  1.03078642
 [19] -2.74093726  2.77205578 -0.25596771 -0.79295335 -1.99567925  7.14183490
 [25]  6.56129569 -2.39785588 -2.30807391 -1.02088455 -2.26040839 -2.76088135
 [31]  1.81877126  0.14669279  4.21783231 -2.13184320  3.69196005 -2.69614367
 [37] -2.68014820  3.72209577  1.73709472 -0.70580812  0.07337669  2.17063230
 [43]  2.72495294  5.04390706  1.32219033  4.72349163 -0.67638087  2.64424944
 [49]  2.78769261 -2.10997705  4.26042721 -3.50266144  1.72564280 -2.07028305
 [55] -4.59779260 -1.71953774  2.90307934  1.38358058  3.42339203 -1.68000430
 [61]  7.55683608  6.32574310 -2.60318964  3.24511198  0.97390332  2.22611398
 [67]  0.83831831  0.07828888  2.29402602  2.68356827  0.07483911  3.38214384
 [73] -0.59180508  9.07209729 -1.27708114  4.77997853 -0.83918672  6.26383807
 [79]  1.50674691  3.25716693  5.70351834  5.80174051  3.61099316  2.19293272
 [85] -1.46102337 -0.97135778  1.54849399  4.34257358 -1.64886246  2.44942102
 [91]  2.68469434  1.64707956  5.49827517  1.01640668  4.43099277  2.23430799
 [97] -1.74337571  6.43458332  2.94137432 -1.01569579
```

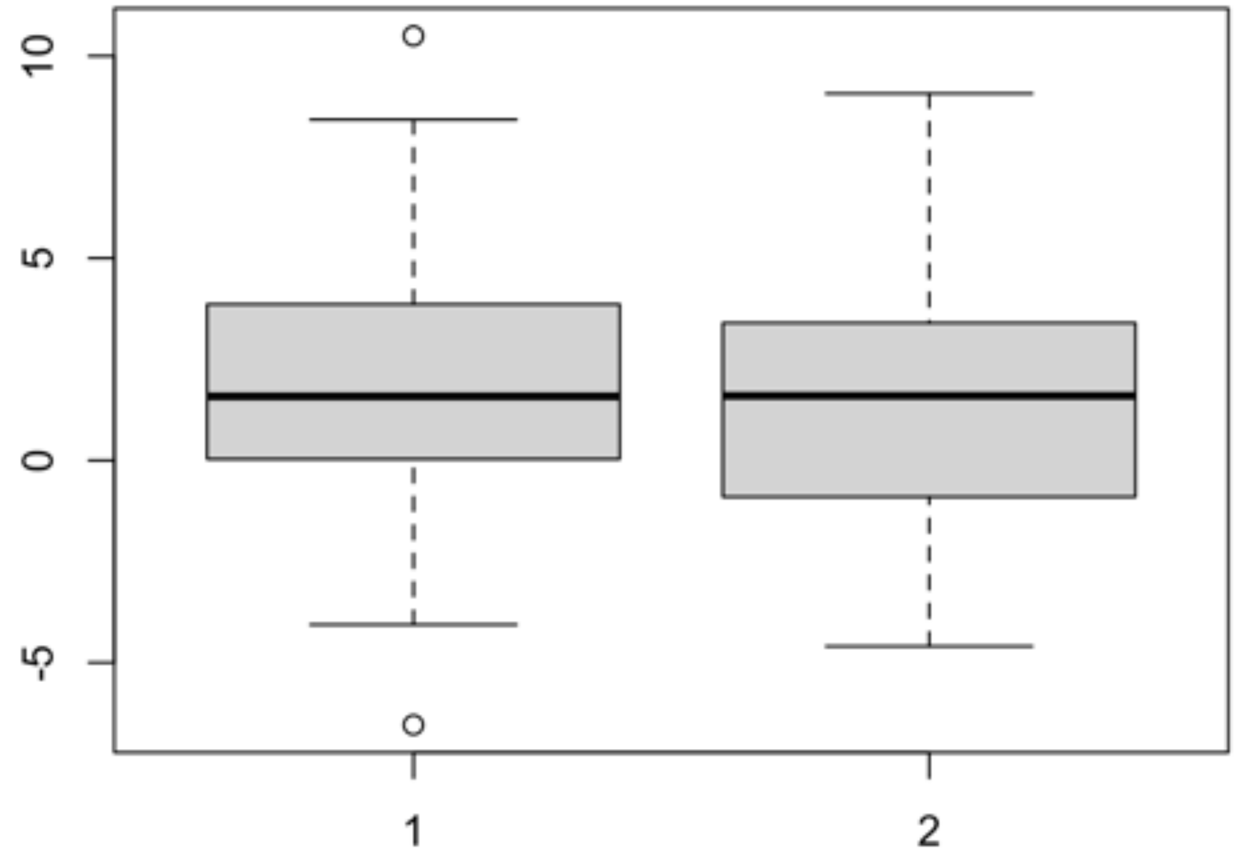
```
> x <- data.matrix(read.table(„dodavatelX.txt"))
> y <- data.matrix(read.table(„dodavatelY.txt"))
```

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Příklad 1: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

```
> boxplot(c(x), c(y))

> shapiro.test(x)
  Shapiro-Wilk normality test

  data: x
  W = 0.99023,  p-value = 0.5554

> shapiro.test(y)
  Shapiro-Wilk normality test

  data: y
  W = 0.98447,  p-value = 0.2907
```



Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Příklad 1: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

```
> var.test(x, y)

F test to compare two variances

data:  x and y
F = 0.87118, num df = 119, denom df = 99, p-value = 0.4701
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5943383 1.2684711
sample estimates:
ratio of variances
         0.8711758
```

=> nulovou hypotézu nezamítáme, rozptyly se statisticky významně neliší na hladině významnosti 5 %

Dvouvýběrové testy při výběru z normálního rozdělení

2) Test shody středních hodnot - dvouvýběrový t-test

Příklad 1: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

```
> t.test(x, y, var.equal = T)

   Two Sample t-test

data:  x   and y
t = 1.0375, df = 218, p-value = 0.3007
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3598731  1.1598308
sample estimates:
mean of x mean of y
 1.884360  1.484381
```

=> nulovou hypotézu nezamítáme, střední hodnoty se statisticky významně neliší na hladině významnosti 5 %

Délky tyčí dodávaných od dvou dodavatelů se statisticky významně neliší na hladině významnosti 5%.

Jak by to bylo s jinou hladinou významnosti?

Dvouvýběrové testy při výběru z normálního rozdělení

3) Párový test shody středních hodnot

$$X : X_1, X_2, \ldots, X_n$$
$$Y : Y_1, Y_2, \ldots, Y_n$$

$$\Rightarrow \quad Z_1 = X_1 - Y_1,\; Z_2 = X_2 - Y_2, \ldots,\; Z_n = X_n - Y_n,$$

$$Z \sim N(\mu_X - \mu_Y, \sigma_Z^2)$$

$$H_0 : \mu_Z = a$$
$$H_A : \mu_Z \neq a$$

$$T = \frac{\bar{Z} - a}{s_Z}\sqrt{n}$$

T má t-rozdělení (Studentovo rozdělení) pravděpodobnosti o $(n$-$1)$ stupních volnosti.

H_0 nezamítneme, když pro dané α bude $|T| \leq t_\alpha(n-1)$, kde $t_\alpha(n-1)$ je (oboustranná) α-kritická hodnota t-rozdělení o $(n$-$1)$ stupních volnosti.

Dvouvýběrové testy při výběru z normálního rozdělení

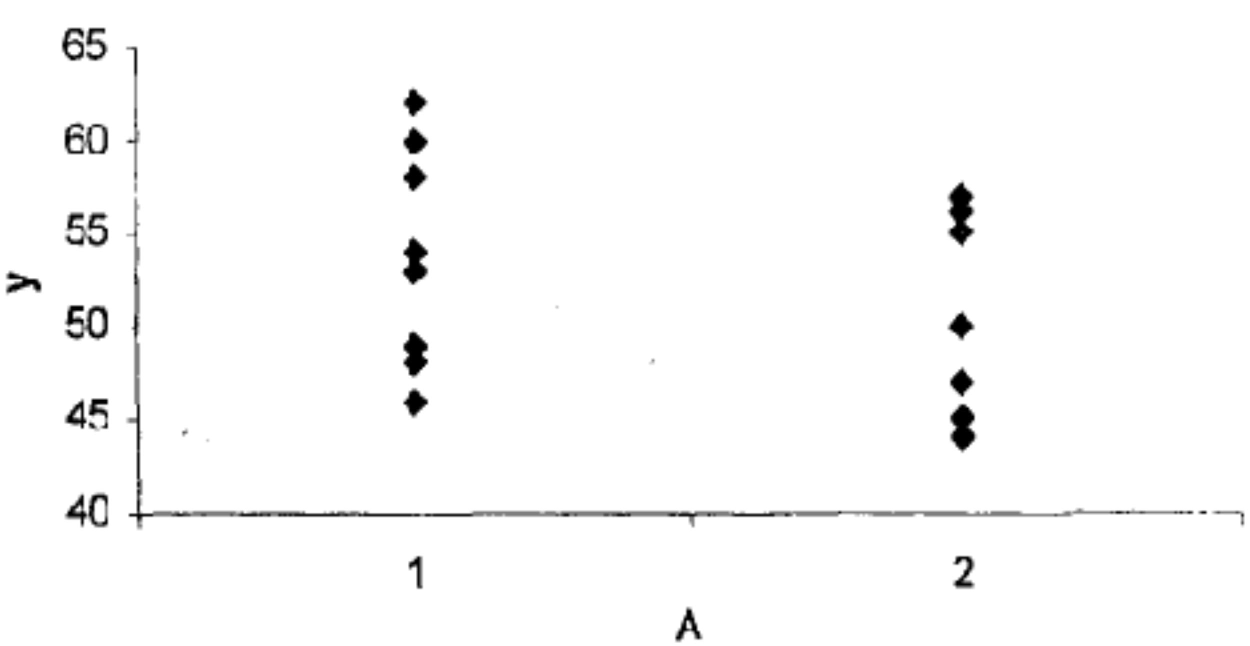
3) Párový test shody středních hodnot

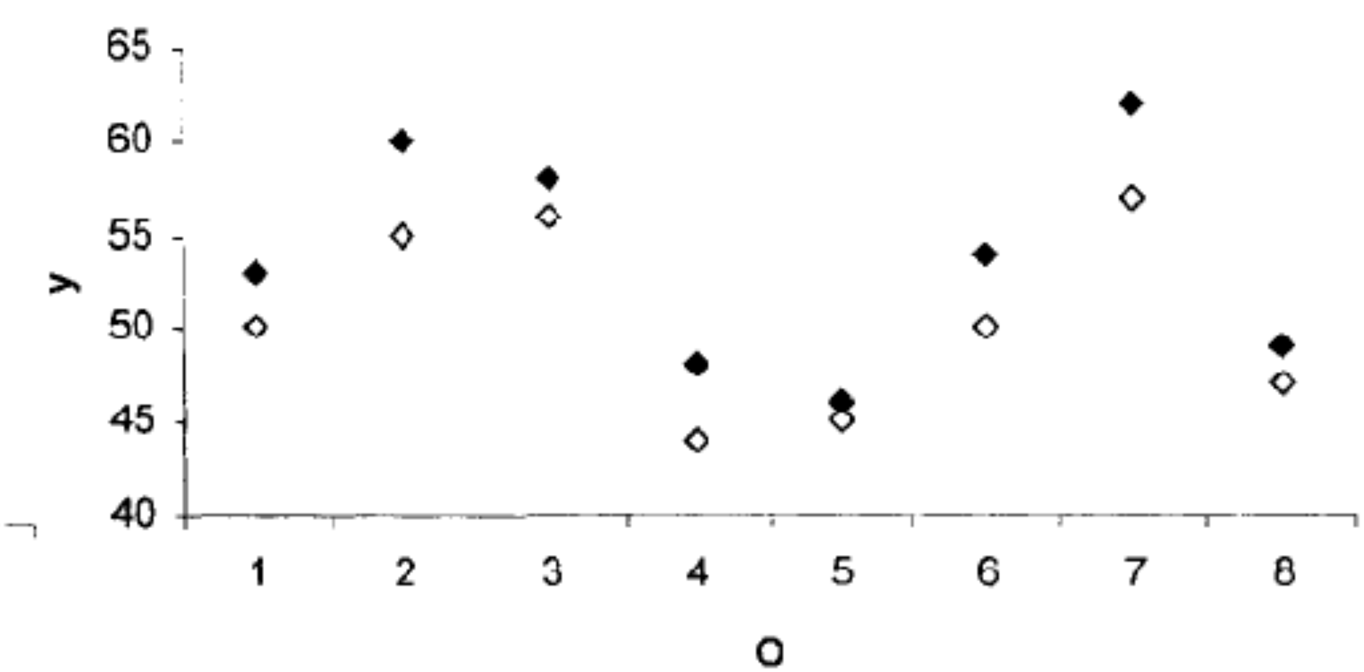
Příklad 2: Bylo zjišťováno, zda množství vyrobených kusů za hodinu (výkonnost) závisí na typu šicího stroje.

Data:

operátor	1	2	3	4	5	6	7	8
stroj A1	53	60	58	48	46	54	62	49
stroj A2	50	55	56	44	45	50	57	47
rozdíl d_j	3	5	2	4	1	4	5	2

Grafické znázornění dat:





Dvouvýběrové testy při výběru z normálního rozdělení

3) Párový test shody středních hodnot

Příklad 2: Bylo zjišťováno, zda množství vyrobených kusů za hodinu (výkonnost) závisí na typu šicího stroje.

$$H_0: \ \mu_{A1} = \mu_{A2}, \quad H_A: \ \mu_{A1} \neq \mu_{A2}$$

$$T = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{r}}}$$

kde

$$\bar{d} = \frac{1}{r}\sum_{j=1}^{r} d_j, \quad s_d^2 = \frac{1}{r-1}\sum_{j=1}^{r}(d_j - \bar{d})^2$$

	y1	*y2*
Stř. hodnota	53,75	50,5
Rozptyl	34,5	25,42857
Pozorování	8	8
Pears. korelace	0,974278	
Hyp. rozdíl stř. hodnot	0	
Stupně volnosti	7	
t stat	6,177483	
P(T<=t) (1)	0,000228	
t krit (1)	1,894578	
P(T<=t) (2)	0,000455	
t krit (2)	2,364623	

=> nulovou hypotézu zamítáme, typ stroje ovlivňuje výkonnost, která je statisticky významně vyšší u prvního typu stroje na jakékoli hladině významnosti, vyšší než $4{,}55.10^{-4}$.

Dvouvýběrové testy při výběru z normálního rozdělení

3) Párový test shody středních hodnot

Příklad 3: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

Data:

```
> pred_cvicenim
  [1] 12.666378
  [7]  9.999636
 [13] 18.019348
 [19] 14.238225
 [25] 21.246960
 [31] 18.258896
 [37] 13.502299
 [43]  8.282819
 [49] 11.488700
 [55] 12.904993
 [61] 15.208694
 [67] 14.647653
 [73]  9.548584
 [79]  9.854223
 [85] 15.605362
 [91] 13.554146
 [97] 12.485393
[103] 10.377856
[109] 14.195263
[115] 10.436798
```

```
> po_cviceni
  [1] 14.889379  8.627612  9.867455 13.141168  9.249122  8.490774
  [7] 10.217290 10.724403 14.669450 14.243944 10.826905 13.951521
 [13] 14.693401  9.449562 16.425888 16.392689 13.265474 14.704994
 [19] 12.718107 10.395385  8.756276  6.961521 12.688497 10.578342
 [25] 14.294064 13.763032  8.472324 15.605253 11.968936  9.897284
 [31] 14.788205 14.773378 11.723336 11.719464 11.824407 12.914485
 [37] 13.291805 13.272867 12.586791  9.202608 15.817188 11.197137
 [43]  8.974410 10.823942 12.289400 10.483861 11.119684  9.956822
 [49]  9.778551 12.062084 13.449972 15.481139  9.470557 11.143402
 [55] 10.793291  9.786869  8.547580  8.188947 12.532635 10.862473
 [61] 10.547040 13.774638 14.861969 11.180668  9.790466 12.469556
 [67] 11.837173 13.820717 11.476120  9.850563 10.440890 11.015557
 [73] 12.547672 12.041457  9.639740 11.368657 11.431948 15.449064
 [79]  9.110052 15.125478 13.433802 11.807514  9.632299 12.725762
 [85] 10.628523 10.824474 13.389953 10.077884  9.185360 13.697777
 [91] 10.116078 13.036067 14.412094 12.175099  7.835201 16.277825
 [97] 10.967441 10.892966 11.668289  9.340267 15.392018 13.323701
[103]  9.928631 14.378075 10.924935 11.448320 11.836161 13.397990
[109] 13.744963 14.083459 10.668370  9.139692 14.716621 15.173684
[115] 10.493444 14.308470 15.295041 13.748886 14.074436 12.261138
```

```
> cviceni <- data.frame(read.table("pred_cvicenim.txt"), read.table(„po_cviceni.txt"))
```

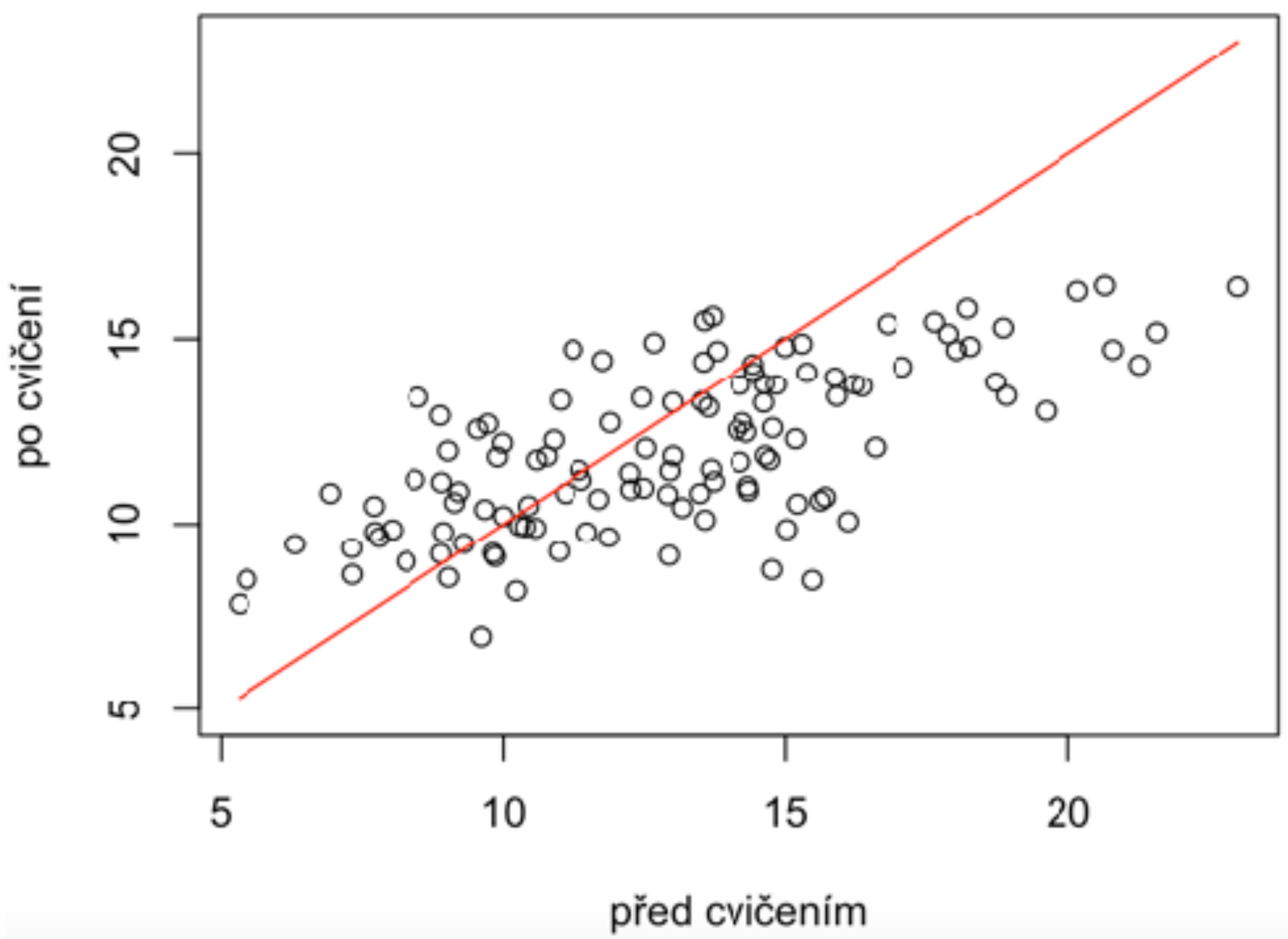
Dvouvýběrové testy při výběru z normálního rozdělení

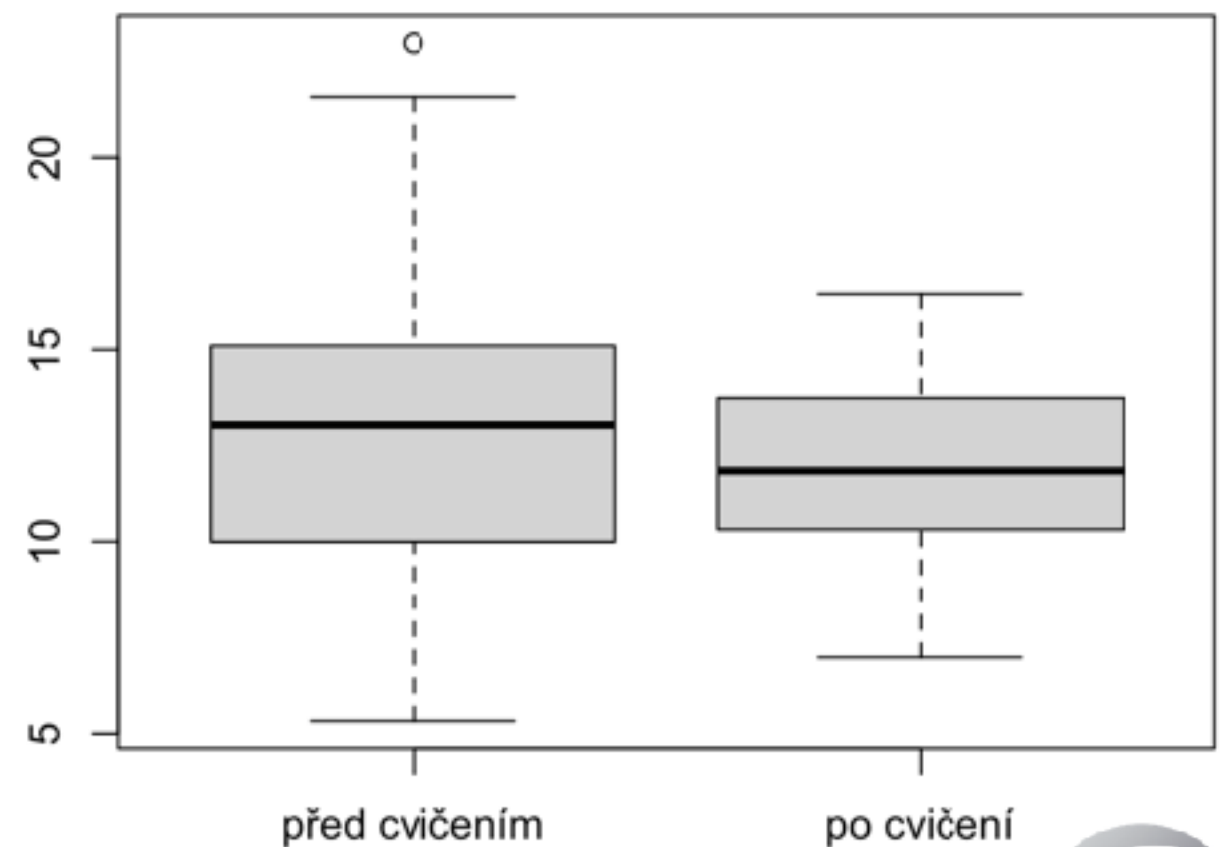
3) Párový test shody středních hodnot

Příklad 3: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

```
> plot(cviceni[,1], cviceni[,2], xlab="před cvičením", ylab="po cvičení", ylim= c(5,23))
> xfit <- seq(min(cviceni[,1]), max(cviceni[,1]), length=40)
> lines(xfit, xfit, col = „red“)
```

```
> boxplot(cviceni[,1], cviceni[,2], names = c("před cvičením","po cvičení"))
```





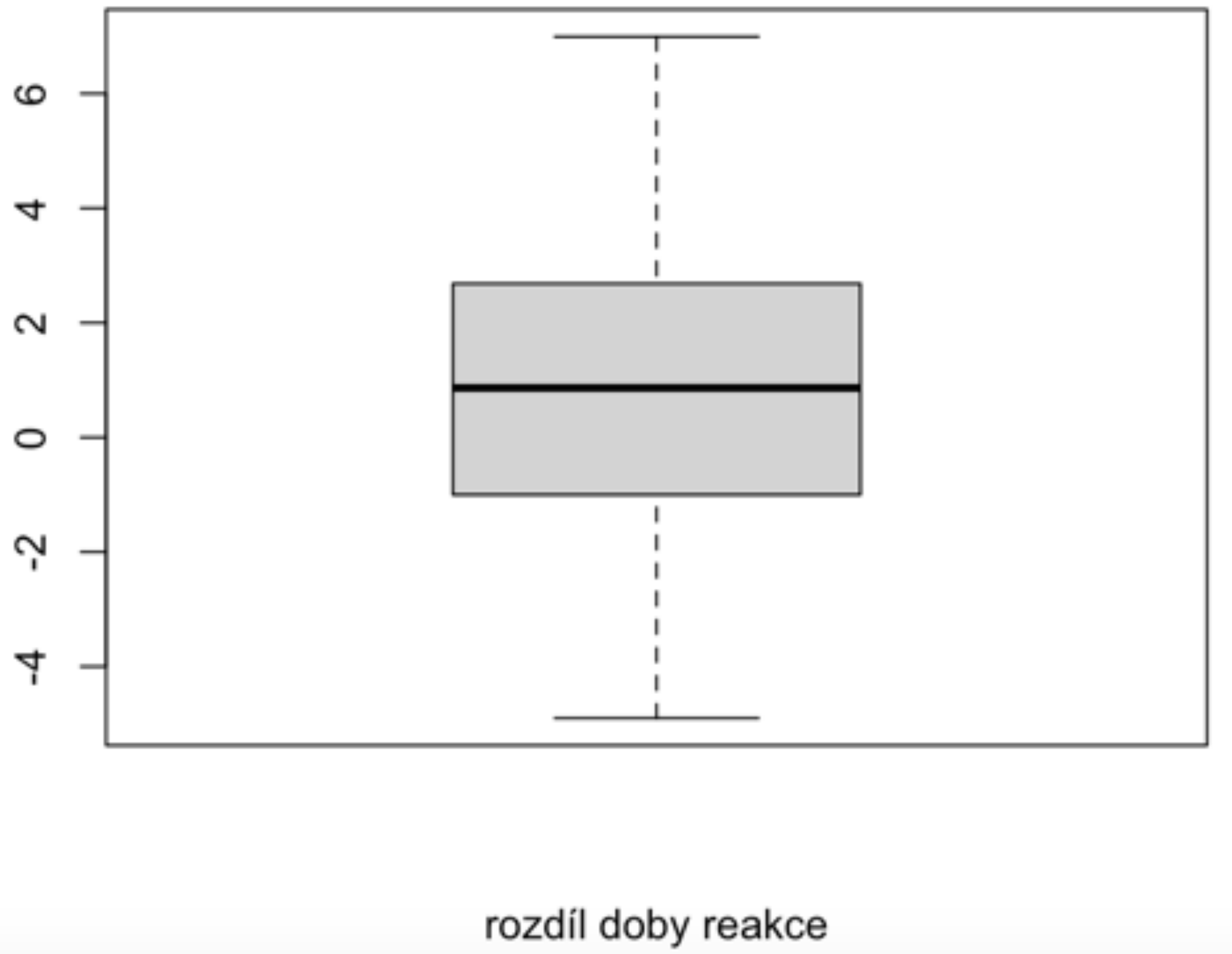
Dvouvýběrové testy při výběru z normálního rozdělení

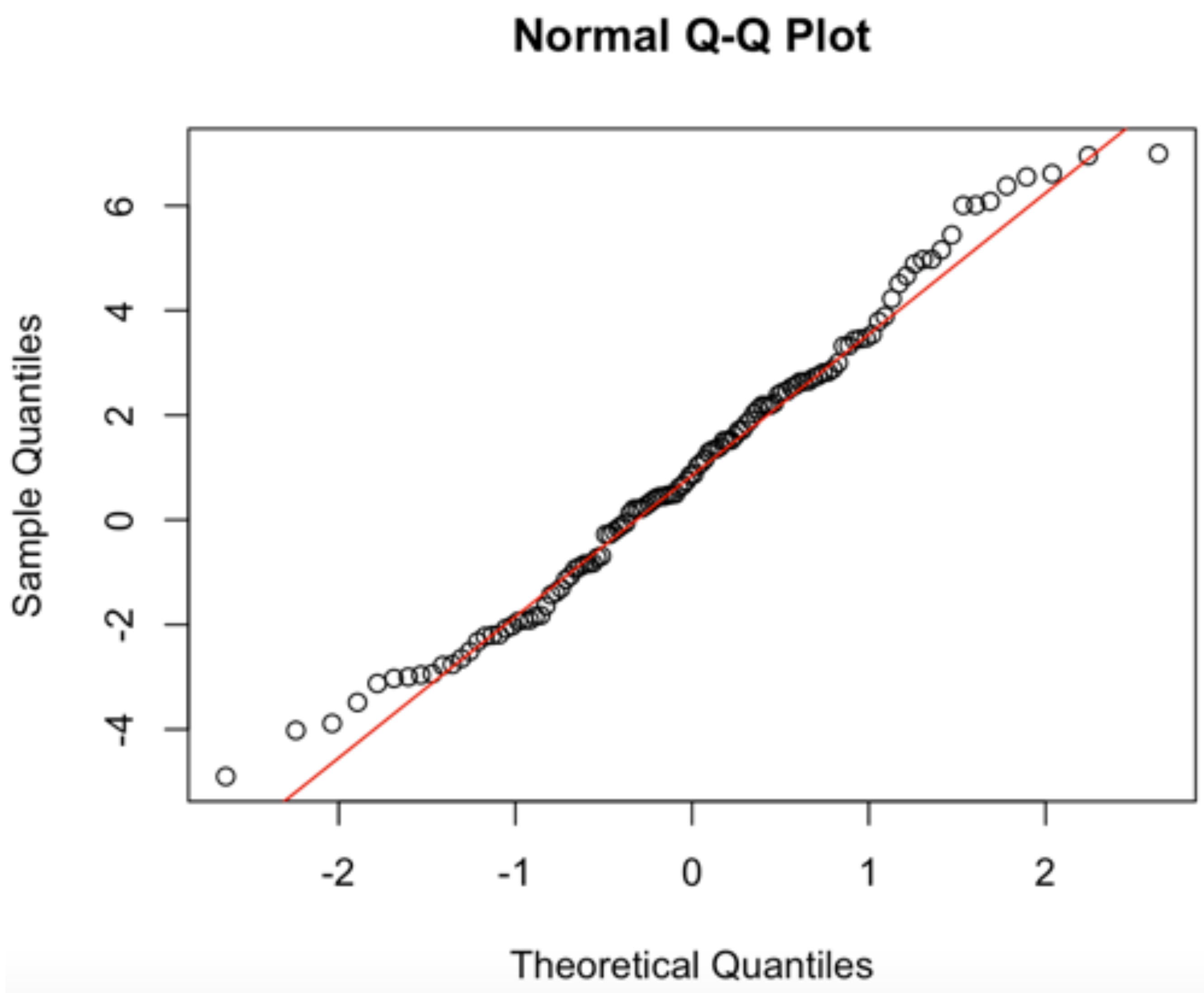
3) Párový test shody středních hodnot

Příklad 3: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

```
> rozdil  <-  cviceni[,1] - cviceni[,2]
> boxplot(rozdil, xlab = "rozdíl doby reakce")
```

```
> qqline(rozdil)
> qqline(rozdil, col = "red")
```





Dvouvýběrové testy při výběru z normálního rozdělení

3) Párový test shody středních hodnot

Příklad 3: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

```
> t.test(rozdil)

        One Sample t-test

data:  rozdil
t = 4.0391, df = 119, p-value = 9.541e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.5089508 1.4878397
sample estimates:
mean of x
0.9983952
```

=> nulovou hypotézu o tom, že rozdíl časů reakce je nulový, zamítáme ve prospěch oboustranné alternativy

Dvouvýběrové testy při výběru z normálního rozdělení

3) Párový test shody středních hodnot

Příklad 3: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

```
> t.test(cviceni[,1], cviceni[,2], paired = T)

	Paired t-test

data:  cviceni[, 1] and cviceni[, 2]
t = 4.0391, df = 119, p-value = 9.541e-05
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 0.5089508 1.4878397
sample estimates:
mean difference
      0.9983952
```

=> cviceni[,1] > cviceni[,2] , neboli
cas pred cvicenim > cas po cviceni
=> střední doba reakce se po cvičení zkrátila

Dvouvýběrové testy při výběru z normálního rozdělení

Jednostranné testy : “dolní” nebo “horní” jednostranná alternativa

$$H_0 : \mu_X = \mu_Y \qquad\qquad H_0 : \mu_X = \mu_Y$$
$$H_A : \mu_X < \mu_Y \qquad\qquad H_A : \mu_X > \mu_Y$$

H$_0$ nezamítneme, když pro dané α bude buď $T < -t_\alpha(n-1)$ nebo $T > t_\alpha(n-1)$ kde $t_\alpha(n-1)$ je (jednostranná) α-kritická hodnota t-rozdělení o $(n$-$1)$ stupních volnosti.

oboustranná α-kritická hodnota je $(1-\alpha/2)$-kvantil $t_{1-\alpha/2}(n)$

jednostranná α-kritická hodnota je $(1-\alpha)$-kvantil $t_{1-\alpha}(n)$

V EXCELu:

=TINV(α; n)
=T.INV.2T(α; n) — tyto funkce vracejí hodnotu $t_{1-\alpha/2}(n)$, tedy oboustrannou α-kritickou hodnotu t-rozdělení

=T.INV(1– α ; n) — vrací hodnotu $t_{1-\alpha}(n)$, tedy přímo (1-α)-kvantil t-rozdělení

Testy při výběru z jiných rozdělení

$$X_1, X_2, \ldots, X_n$$

$$X : X_1, X_2, \ldots, X_n \qquad Y : Y_1, Y_2, \ldots, Y_m$$

$$\cancel{X \sim N(\mu_X, \sigma_X^2) \qquad Y \sim N(\mu_Y, \sigma_Y^2)}$$

Neparametrické varianty t-testů:

1) Jednovýběrový Wilcoxonův test
2) Dvouvýběrový Wilcoxonův test
3) Znaménkový test

Neparametrické testy jsou založeny na **pořadových statistikách:** $R_1, R_2, \ldots, R_n$

Testy při výběru z jiných rozdělení

1. Jednovýběrový Wilcoxonův test

$$H_0 : \tilde{x}_{med} = m_0$$
$$H_A : \tilde{x}_{med} \neq m_0$$

- Provedeme n nezávislých měření $X_1, X_2, \ldots, X_n$ veličiny, které se hypotéza týká
- Spočítáme rozdíly $(X_1 - m_0)$, $(X_2 - m_0)$, ..., $(X_n - m_0)$ a vyloučíme nulové rozdíly
- Uspořádáme je podle velikosti absolutních hodnot a potom určíme:
  - R_k^+= pořadí k-tého kladného rozdílu, R_s^- = pořadí s-tého záporného rozdílu
  - R^+ = součet všech R_k^+ a R^-= součet všech R_s^-

Pro malá *n*, *m* (do 30) porovnáváme přímo jednu z hodnot buď R^+ nebo R^- s kritickými hodnotami pro jednovýběrový Wilcoxonův test (podrobnosti viz např. https://is.muni.cz/th/r5oe7/Bakalarska_prace.pdf).
Pro velké hodnoty *n* a *m* (nad 30) lze jako testovou statistiku použít

$$V = \frac{R^+ - E(R^+)}{\sqrt{var(R^+)}}$$

kde $E(R^+) = \dfrac{n^*(n^*+1)}{4}$, $\quad var(R^+) = \dfrac{n^*(n^*+1)(2n^*+1)}{24}$ a n^* je počet nenulových rozdílů. Tato statistika má přibližně N(0,1) rozdělení.

Testy při výběru z jiných rozdělení

2. Dvouvýběrový Wilcoxonův test

$$H_0 : \tilde{x}_{med} = \tilde{y}_{med}$$
$$H_A : \tilde{x}_{med} \neq \tilde{y}_{med}$$

- Provedeme n nezávislých měření $X_1, X_2, \ldots, X_n$ veličiny X a m měření $Y_1, Y_2, \ldots, Y_m$ veličiny Y.

- Sloučíme všechna měření v jeden soubor: $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$
- Uspořádáme jej podle velikosti a potom určíme:
  - R_k^X= pořadí k-tého pozorování X ve spojeném souboru
  - R_s^Y = pořadí s-tého pozorování Y ve spojeném souboru
  - R^X = součet všech R_k^X pro k=1, …, n a R^Y= součet všech R_s^Y pro s=1, …, m

Pro malá n, m (do 30) porovnáváme přímo jednu z hodnot buď R^X nebo R^Y s kritickými hodnotami pro dvouvýběrový Wilcoxonův test (podrobnosti viz např. https://is.muni.cz/th/r5oe7/Bakalarska_prace.pdf).
Pro velké hodnoty n a m (nad 30) lze jako testovou statistiku použít

$$W = \frac{R^X - E(R^X)}{\sqrt{var(R^X)}}$$

kde $E(R^X) = \dfrac{n(m+n+1)}{2}, \quad var(R^X) = \dfrac{mn(m+n+1)}{12}$. Tato statistika má přibližně N(0,1) rozdělení.

Testy při výběru z jiných rozdělení

3. Párový Wilcoxonův test

$$H_0 : \tilde{x}_{med} = \tilde{y}_{med}$$
$$H_A : \tilde{x}_{med} \neq \tilde{y}_{med}$$

- Spočítáme rozdíly $(X_1\text{-}Y_1)$, $(X_2\text{-}Y_2)$, ..., $(X_n\text{-}Y_n)$ a vyloučíme nulové rozdíly
- Uspořádáme je podle velikosti absolutních hodnot a potom určíme:
  - R_k^+= pořadí k-tého kladného rozdílu, R_l^- = pořadí l-tého záporného rozdílu
  - R^+ = součet všech R_k^+ a R^-= součet všech R_l^-

Pro malá n, m (do 30) porovnáváme přímo jednu z hodnot buď R^+ nebo R^- s kritickými hodnotami pro jednovýběrový Wilcoxonův test (viz např. https://is.muni.cz/th/r5oe7/Bakalarska_prace.pdf), pro velké hodnoty n a m (nad 30) lze jako testovou statistiku použít

$$V = \frac{R^+ - E(R^+)}{\sqrt{var(R^+)}}$$

kde $E(R^+) = \dfrac{n^*(n^*+1)}{4}$, $var(R^+) = \dfrac{n^*(n^*+1)(2n^*+1)}{24}$ a n^* je počet nenulových rozdílů. Tato statistika má přibližně N(0,1) rozdělení.

Testy při výběru z jiných rozdělení

4. Znaménkový (párový) test

$$H_0 : \tilde{x}_{med} = \tilde{y}_{med}$$
$$H_A : \tilde{x}_{med} \neq \tilde{y}_{med}$$

- Spočítáme rozdíly $(X_1\text{-}Y_1)$, $(X_2\text{-}Y_2)$, ..., $(X_n\text{-}Y_n)$ a vyloučíme nulové rozdíly
- Spočteme počet záporných rozdílů Z a počet nenulových rozdílů n*

Statistika Z má binomické rozdělení s parametry n* a 1/2, pro velké hodnoty n* ji lze aproximovat normáním rozdělením a tedy test založit na statistice

$$S = \frac{|2Z - n^*|}{\sqrt{n^*}}$$

která má přibližně N(0,1) rozdělení. Hypotézu o shodě zamítáme, když je

.$S \geq u(1 - \alpha/2)$

Testy při výběru z jiných rozdělení

```
> davkovac <- data.matrix(read.table(„davkovac.txt“))
> wilcox.test(davkovac, mu = 25, paired = F, correct = T)

	Wilcoxon signed rank test with continuity correction

data:  davkovac
V = 113, p-value = 4.23e-07
alternative hypothesis: true location is not equal to 25
```

```
> dodavatelX <- data.matrix(read.table("dodavatelX.txt"))
> dodavatelY <- data.matrix(read.table(„dodavatelY.txt"))
> wilcox.test(dodavatelX, dodavatelY, paired = F)

	Wilcoxon rank sum test with continuity correction

data:  dodavatelX and dodavatelY
W = 6488, p-value = 0.2997
alternative hypothesis: true location shift is not equal to 0
```

Testy při výběru z jiných rozdělení

```
> cviceni <- data.frame(read.table("pred_cvicenim.txt"), read.table("po_cviceni.txt"))
> wilcox.test(cviceni[,1], cviceni[,2], paired = T)

	Wilcoxon signed rank test with continuity correction

data:  cviceni[, 1] and cviceni[, 2]
V = 4973, p-value = 0.0004384
alternative hypothesis: true location shift is not equal to 0


> binom.test(sum(rozdil > 0), length(rozdil), p=0.5, alternative="two.sided")

	Exact binomial test

data:  sum(rozdil > 0) and length(rozdil)
number of successes = 77, number of trials = 120, p-value = 0.002447
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.5490390 0.7271215
sample estimates:
probability of success
             0.6416667
```