Pravděpodobnostní metody ve strojírenství

10. Testování hypotéz

FAKULTA STROJNÍ ČVUT V PRAZE

prof. RNDr. Gejza Dohnal, CSc.

ak. rok 2021/2022

10. Testování hypotéz

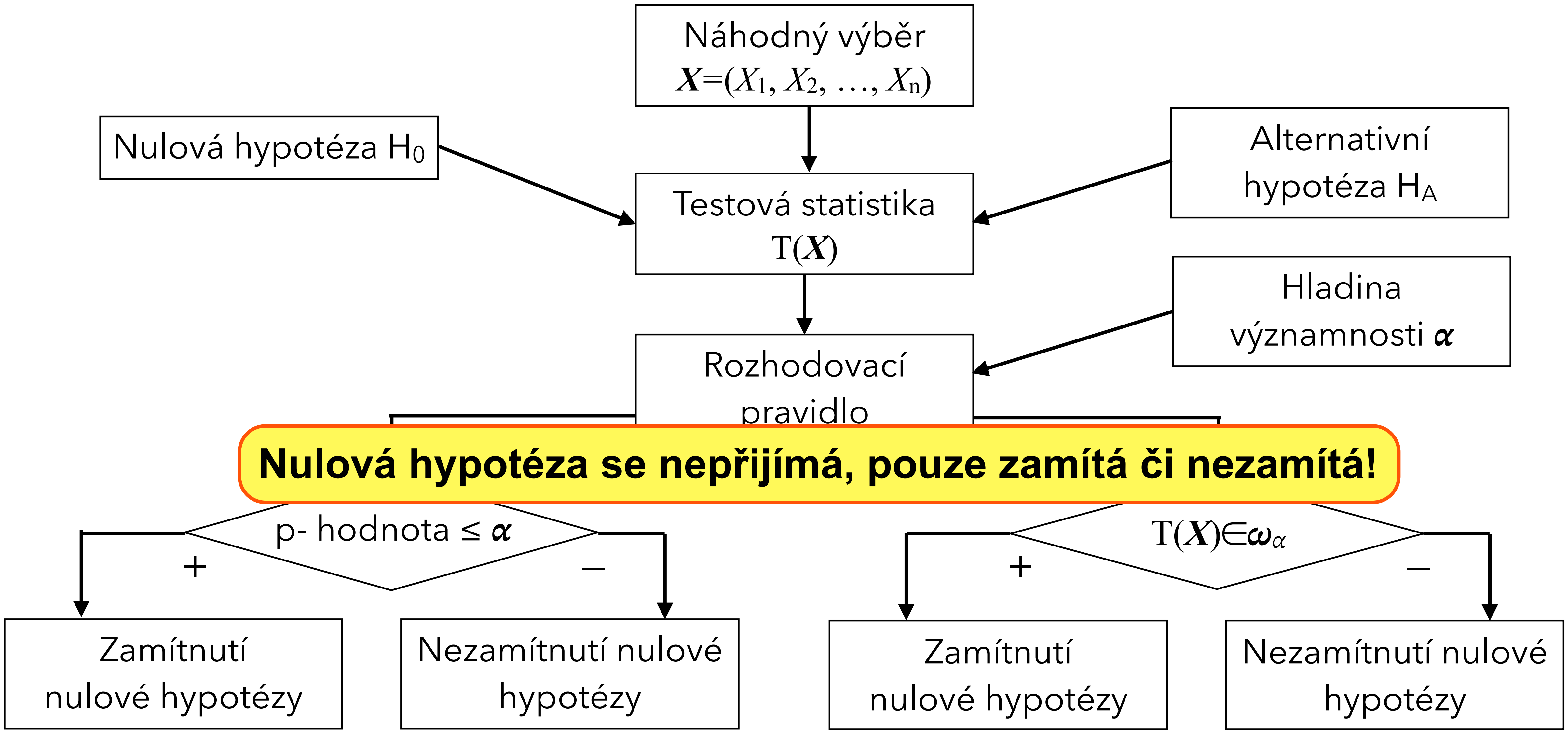
Klíčové pojmy:

- Nulová a alternativní hypotéza
- Hladina významnosti testu
- Chyby 1. a 2. druhu
- p-hodnota

Klíčové vztahy:

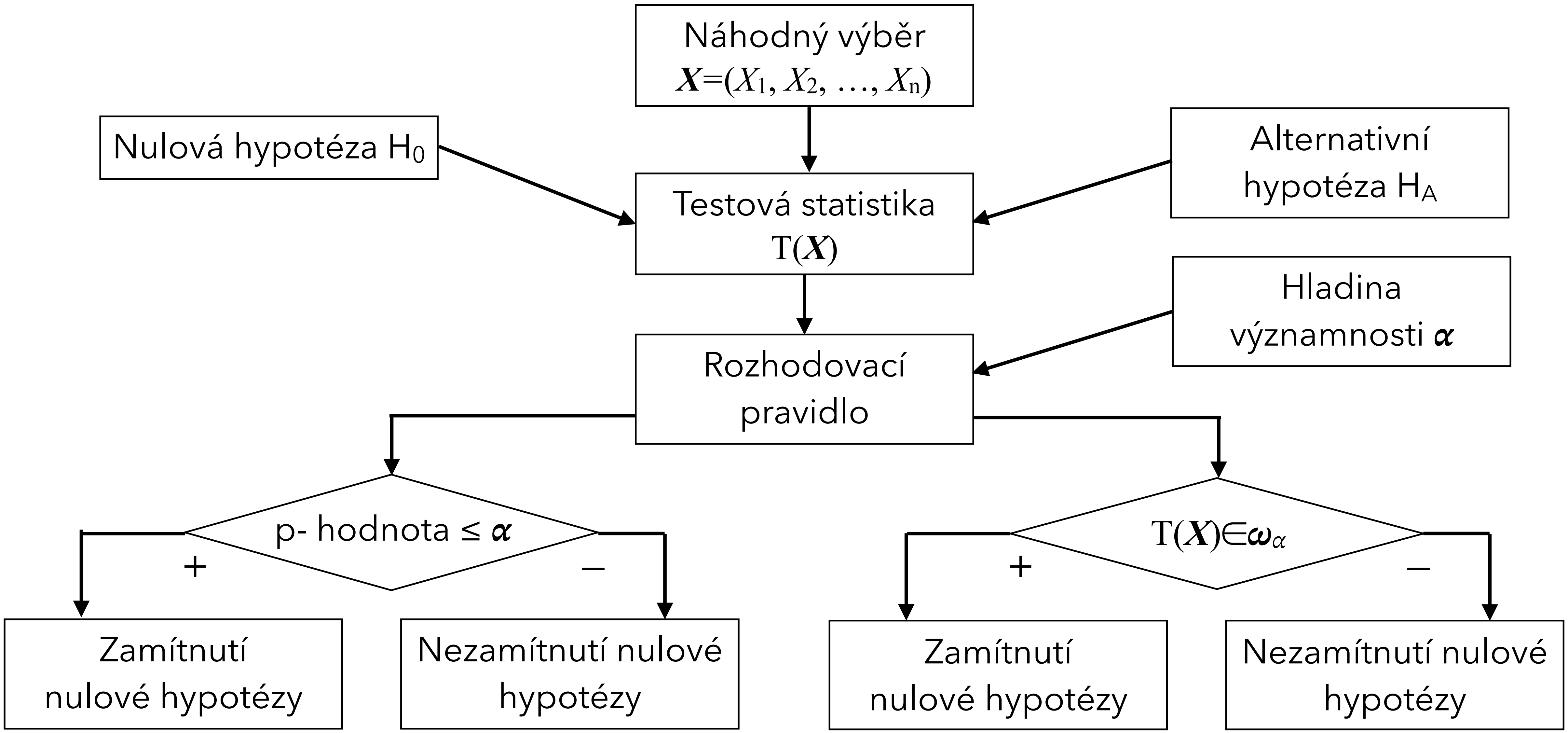
- Jednovýběrové testy (parametrické i neparametrické
- Dvouvýběrové testy při nezávislých výběrech
- Párové testy
- Test ANOVA, podmínky
- Kruskal-Wallis test

Základní algoritmus



Statisticky nelze nic dokázat, pouze zamítnout!

Základní algoritmus



- **Chyba 1. druhu:** Zamítneme nulovou hypotézu, která ve skutečnosti platí.
  P(nastane chyba 1. druhu) = α … **hladina významnosti testu**
- **Chyba 2. druhu:** Nezamítneme nulovou hypotézu, i když ve skutečnosti neplatí.
  P(nastane chyba 2. druhu) = β … **síla testu**

Testování (statistických) hypotéz

Různé statistické testy se liší v testových statistikách a jejich kritických hodnodách:

- testy o parametrech rozdělení pravděpodobnosti (jednovýběrové testy)
- srovnání dvou či více souborů z hlediska jejich charakteristik (dvouvýběrové testy, analýza rozptylu)
- neparametrické testy
- testy o rozdělení pravděpodobnosti (testy dobré shody)
- testy nezávislosti (nekorelovanosti)
- …..

Jednovýběrový test o střední hodnotě

$$X_1, X_2, \ldots, X_n \quad X \sim N(\mu, \sigma^2)$$

$H_0: \quad \mu = m_0$ - Nulová hypotéza

$H_A: \quad \mu \neq m_0$ - Alternativní hypotéza

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

- má t-rozdělení o (n-1) stupních volnosti

$\alpha \quad \Rightarrow \quad t_\alpha(n-1)$ - hladina významnosti, kritická hodnota

---

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

$p = 2\min\{\, F(T),\, 1 - F(T)\,\}$ - p-hodnota

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

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

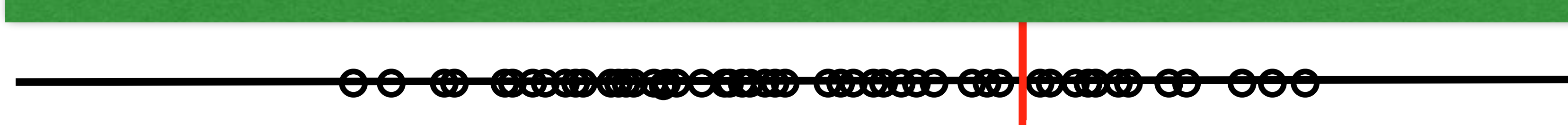
$$\boxed{\bar{X} \notin \langle m_0 - \frac{s}{\sqrt{n}}t_\alpha(n-1), m_0 + \frac{s}{\sqrt{n}}t_\alpha(n-1)\rangle}$$

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?

	A	B	C	D	E	F
1	24,52586					
2	24,17119		Minimum:	23,65390	=MIN(A1:A50)	
3	24,54486		Dolníkvartil:	24,20505	=QUARTIL(A1:A50;1)	
4	24,44240		Median:	24,52342	=MEDIAN(A1:A50)	
5	23,93455		Horníkvartil:	24,887985	=QUARTIL(A1:A50;1)	
6	24,20389		Maximum:	25,48924	=MAX(A1:A50)	
7	24,19974		Průměr:	24,54689	=PRŮMĚR(A1:A50)	
8	24,34851		Směr.odch.:	0,45852795	=SMODCH.VÝBĚR(A1:A50)	
9	23,94024		Šikmost:	0,09593675	=SKEW(A1:A50)	
10	24,21022		Špičatost	-0,6732486	=KURT(A1:A50)	
11	24,87474					
12	25,06155		t(n-1,0.95)	2,00957524	=T.INV.2T(0,05;49)	
13	25,48924		SE:	0,13	=D8*D1/SQRT(50)	
14	25,32572					
15	23,71721		CI(mean):	24,87		
16	24,61622			25,13		
17	25,06676					

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



$\bar{X} = 24.55, \;\; s^2(X) = 0.21, \;\; s(X) = 0.46, \;\; n = 50$

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

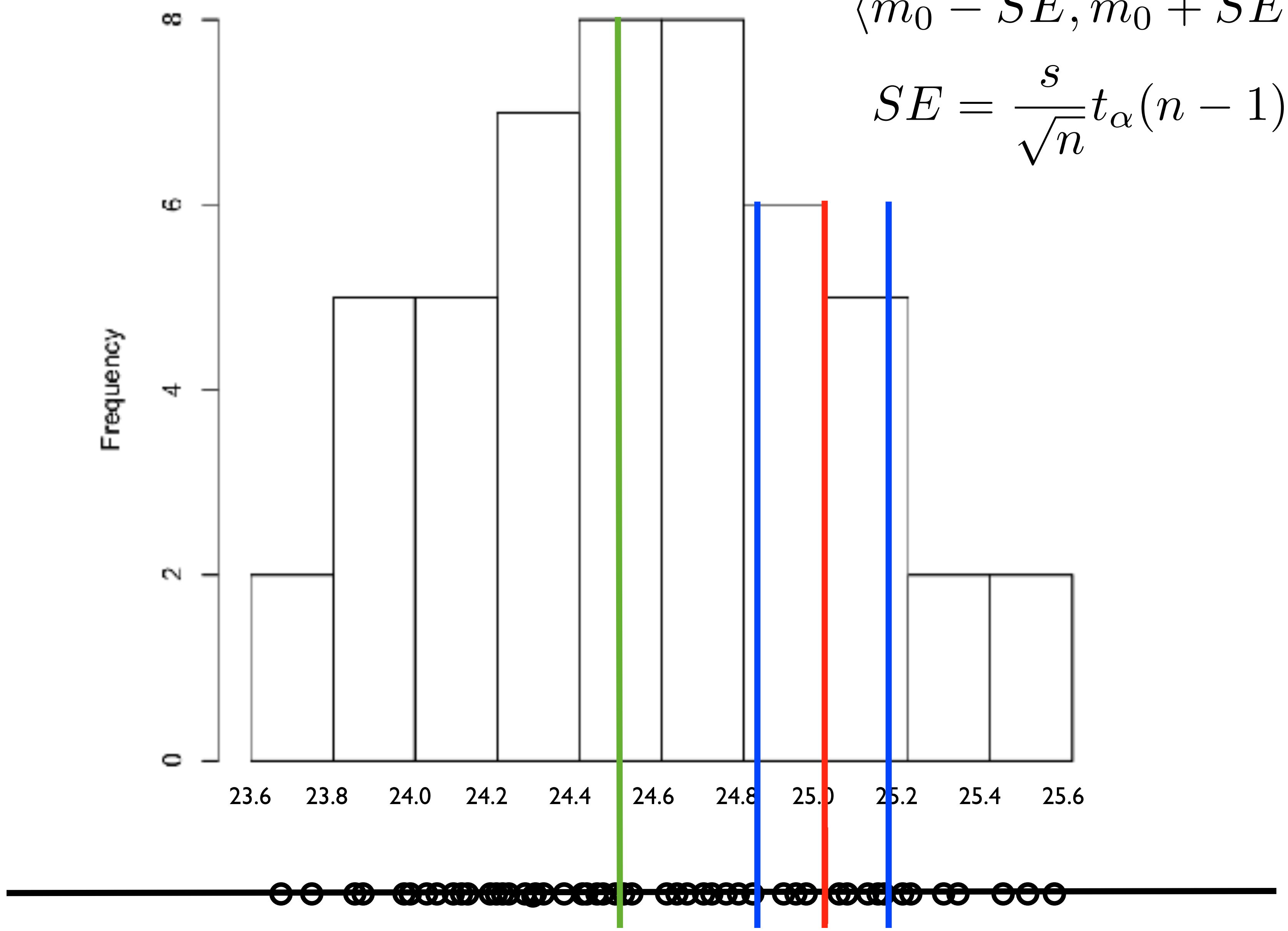
$\alpha = 0.05 \;\; \Rightarrow t_{0.05}(49) = 2.01 \qquad SE = 0.132$

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?

$$\langle m_0 - SE, m_0 + SE \rangle$$

$$SE = \frac{s}{\sqrt{n}} t_\alpha(n-1)$$



$\bar{X} = 24.55,\ s^2(X) = 0.21,\ s(X) = 0.46,\ n = 50$

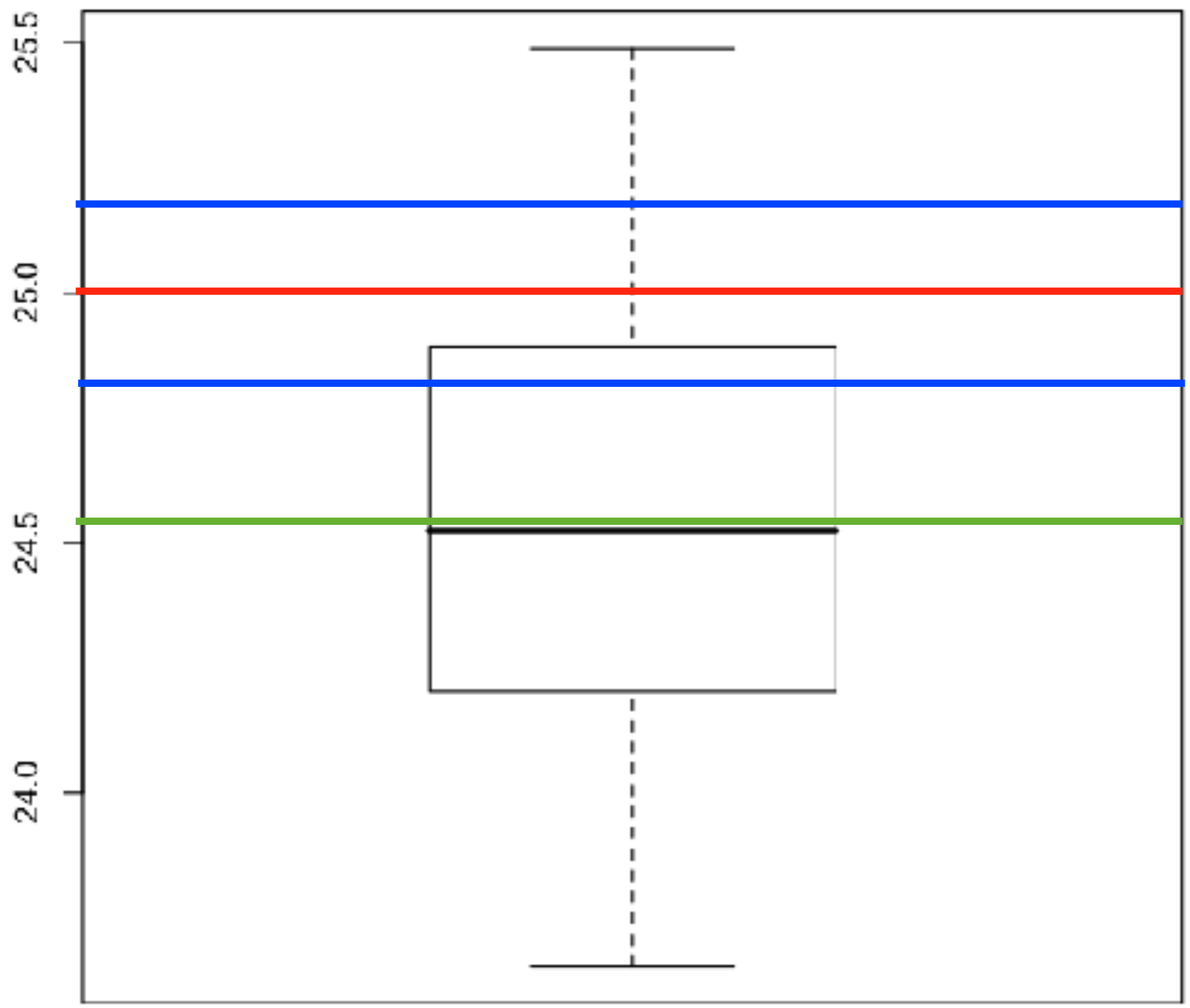
Chceme ověřit, zda měření odpovídají rozdělení $N(m_0, s^2)$

$\alpha = 0.05 \ \Rightarrow t_{0.05}(49) = 2.01$ $\quad SE = 0.132$ $\quad \bar{X} \notin \langle 24.868, 25.132 \rangle$

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

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?



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

$\bar{X} = 24.55,\ s^2(X) = 0.21,\ s(X) = 0.46,\ n = 50$

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

$\alpha = 0.05 \ \Rightarrow t_{0.05}(49) = 2.01$ $\quad SE = 0.132$ $\quad \bar{X} \notin \langle 24.868, 25.132 \rangle$

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?

```
x <- data.matrix(read.table("davkovac.txt"))
t.test(x, mu=25.0, alternative = "two.sided", conf.level=0.95)

        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
```

$\bar{X} \pm SE$

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
24.2298[illegible]	[illegible]
24.0977[illegible]	[illegible]
24.892[illegible]	[illegible]
24.14259	25.12906

$m_0 \pm SE$

$\bar{X} = 24.55, \ s^2(X) = 0.21, \ s(X) = 0.46, \ n = 50$

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

$\alpha = 0.05 \ \Rightarrow t_{0.05}(49) = 2.01$ $\quad SE = 0.132$ $\quad \bar{X} \notin \langle 24.868, 25.132 \rangle$

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?

```
x <- data.matrix(read.table("davkovac.txt"))
t.test(x, mu=25.0, 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
```

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

$\bar{X} = 24.55,\ s^2(X) = 0.21,\ s(X) = 0.46,\ n = 50$

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

$\alpha = 0.05 \ \Rightarrow t_{0.05}(49) = 2.01$ $\quad SE = 0.132$ $\quad \bar{X} \notin \langle 24.868, 25.132 \rangle$

Jednovýběrový test o střední hodnotě - příklad

Lze považovat automat za správně seřízený (m_0=25,0g)?

```
x <- data.matrix(read.table("davkovac.txt"))
t.test(x, mu=25.0, 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
```

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

$\bar{X} = 24.55, \;\; s^2(X) = 0.21, \;\; s(X) = 0.46, \;\; n = 50$

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

$\alpha = 0.05 \;\; \Rightarrow t_{0.05}(49) = 2.01$ $\qquad SE = 0.132$ $\qquad \bar{X} \notin \langle 24.868, 25.132 \rangle$

Srovnání dvou výběrů

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

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

Mohou nastat tři situace:

- oba parametry v obou případech známe
- známe střední hodnoty a neznáme rozptyly
- známe rozptyly a neznáme střední hodnoty
- žádný z parametrů neznáme

Odhady středních hodnot: $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \bar{Y} = \frac{1}{m}\sum_{i=1}^{n} Y_i$$

Odhady rozptylů: $$s_X^2 = \frac{1}{n-1}\sum (X - \bar{X})^2, \quad s_Y^2 = \frac{1}{n-1}\sum (Y - \bar{Y})^2$$

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

Dvě závislá měření

$X: X_1, X_2, \ldots, X_n$

$Y: Y_1, Y_2, \ldots, Y_n$

párová pozorování

- párový test shody středních hodnot

Srovnání dvou náhodných veličin

2) Srovnání středních hodnot dvou nezávislých měření

- Dvouvýběrový t-test

Liší se statisticky významně dvě nezávislá měření z hlediska jejich střední hodnoty?

Lze považovat střední hodnoty dvou nezávislých měření za shodné při dané hladině významnosti?

Lze od sebe statisticky významně odlišit dvě nezávislá měření podle jejich jejich střední hodnoty?

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

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

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

hladina významnosti: α

Srovnání dvou náhodných veličin

2) Srovnání středních hodnot dvou nezávislých měření

- Dvouvýběrový t-test

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

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

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

hladina významnosti: α

pokud $\sigma_X^2 = \sigma_Y^2$ → dvouvýběrový t-test se stejnými rozptyly

pokud $\sigma_X^2 \neq \sigma_Y^2$ → dvouvýběrový t-test s nestejnými rozptyly

Srovnání dvou náhodných veličin

2) Srovnání rozptylů dvou nezávislých měření

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

Lze považovat rozptyl dvou nezávislých 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

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

Srovnání dvou náhodných veličin

2) Srovnání středních hodnot dvou nezávislých měření

- Dvouvýběrový t-test se stejnými rozptyly

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

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

dále odhadneme s^2 ze
všech naměřených hodnot: $s^2 = \dfrac{1}{n+m-2}\left(\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{m}(Y_i - \bar{Y})^2\right) =$

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

tedy: $s_{\bar{X}-\bar{Y}}^2 = \dfrac{n+m}{nm(n+m-2)}\left((n-1)s_X^2 + (m-1)s_Y^2\right)$ a testová statistika

bude mít tvar: $T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}}\sqrt{\dfrac{nm(n+m-2)}{n+m}}$

T má t-rozdělení (Studentovo) 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.

Srovnání dvou náhodných veličin

2) Srovnání středních hodnot dvou nezávislých měření

- Dvouvýběrový t-test s nestejnými rozptyly

pokud $\sigma_X^2 \neq \sigma_Y^2$, testová statistika bude mít tvar:

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

a má rozdělení, které je směsí t-rozdělení o $(n\text{-}1)$ a $(m\text{-}1)$ stupních volnosti.

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

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

Srovnání dvou náhodných veličin

3) Párový test shody středních hodnot dvou závislých měření

- pozorování stejné veličiny před a po nějakém zásahu
- měření stejných objektů za různých podmínek
- měření stejné veličiny ve dvou různých časech
- .....

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

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

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

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

$$H_0 : \mu_X = \mu_Y \qquad H_A : \mu_X \neq \mu_Y$$

$$\Rightarrow$$

$$H_0 : \mu_Z = 0 \qquad H_A : \mu_Z \neq 0$$

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

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

Srovnání dvou náhodných veličin

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 $t_{1-\alpha/2}(n-1)$

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

Srovnání dvou náhodných veličin

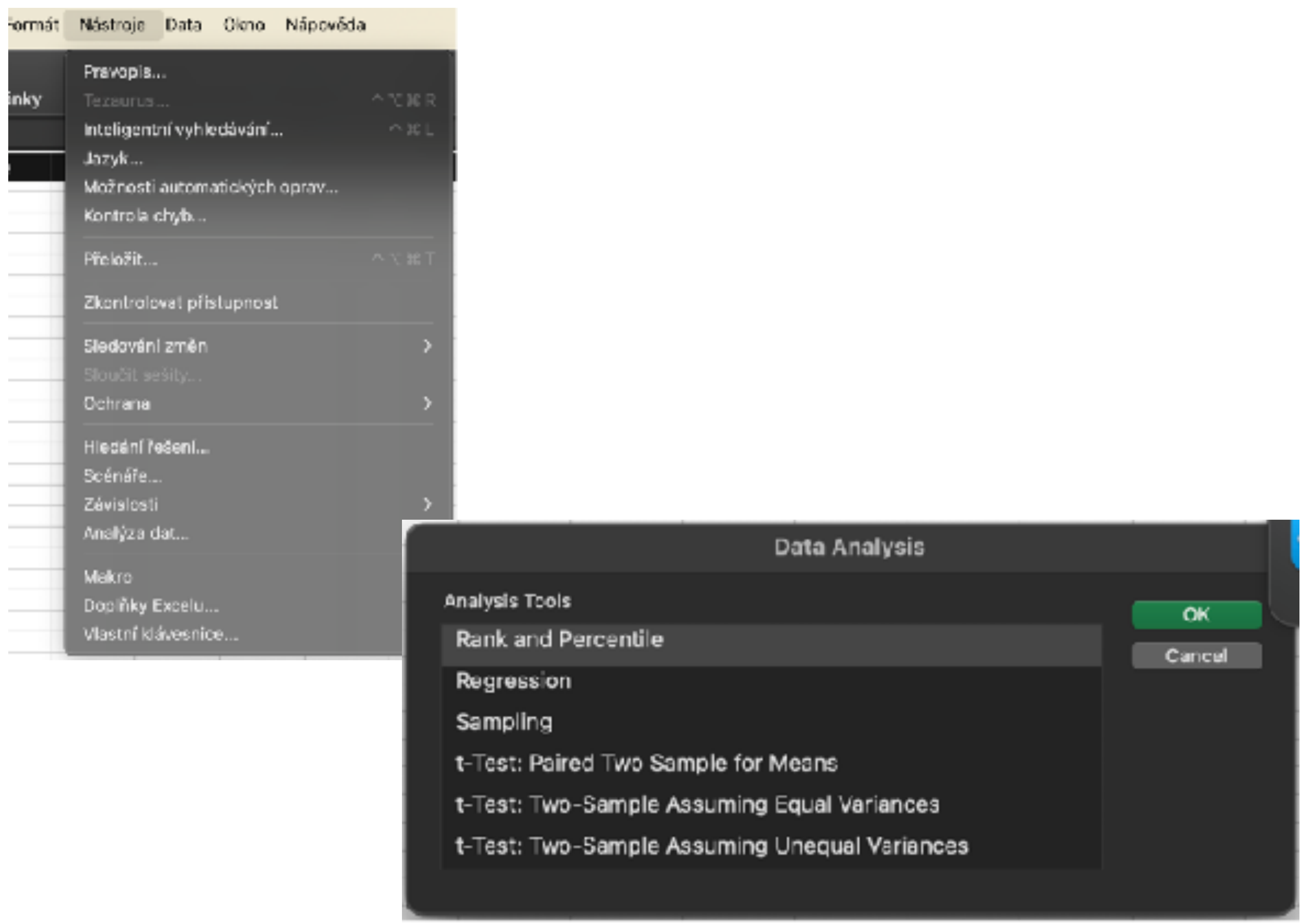
Příklad: Byly měřeny odchylky od požadované délky 4m ocelových tyčí od dvou dodavatelů. Odchylky jsou uvedeny v cm.

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

Řešení v MS Excel:

K19

	A	B
1	Dodavetel_X	Dodavatel_Y
2	0,41379418	6,65956934
3	0,51040227	2,78876119
4	3,28722973	0,33397602
5	7,31995568	-0,0376392
6	4,53994434	0,74993937
7	-1,07426821	3,81490677
8	4,74575978	1,70428804
9	2,55201407	-3,3129134
10	3,22058685	-0,2297237
11	-1,17401554	4,02124752
12	-1,241195	5,93229834
13	4,1829469	3,3050607
14	0,65486399	-3,6127706
15	-0,18908709	0,78809415
16	-0,73101186	0,37976841
17	1,27876451	1,5235732
18	1,26734875	1,76230055
19	2,78570344	1,03078642
20	2,96834139	-2,7409373
21	1,22145702	2,77205578
22	1,8085144	-0,2559677
23	-0,80355569	-0,7929534
24	2,57347292	-1,9956793
25	3,42552806	7,1418349
26	1,66904559	6,56129569
27	-2,21179295	-2,3978559
28	4,1769627	-2,3080739
29	2,15191523	-1,0208846
30	3,62707736	-2,2604084
31	0,06900211	-2,7608814
32	0,51371315	1,81877126
33	0,54983237	0,14609279
34	4,09554316	4,21763231



Srovnání dvou náhodných veličin

Příklad: 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%?

Řešení v MS Excel:

K19

	A	B
1	Dodavetel_X	Dodavatel_Y
2	0,41379418	6,65956934
3	0,51040227	2,78876119
4	3,28722973	0,33397602
5	7,31995568	-0,0376392
6	4,53994434	0,74993937
7	-1,07426821	3,31490677
8	4,74575978	1,70428804
9	2,55201407	-3,3129134
10	3,22058685	-0,2297237
11	-1,17401554	4,02124752
12	-1,241195	5,93229834
13	4,1829469	3,3050607
14	0,65486399	-3,6127706
15	-0,18908709	0,78809415
16	-0,73101186	0,37976841
17	1,27876451	1,5235732
18	1,26734875	1,76230055
19	2,78570344	1,03078642
20	2,96834139	-2,7409373
21	1,22145702	2,77205578
22	1,8085144	-0,2559677
23	-0,80355569	-0,7929534
24	2,57347292	-1,9956793
25	3,42552806	7,1418349
26	1,66904559	6,56129569
27	-2,21179295	-2,3978559
28	4,1769627	-2,3080739
29	2,15191523	-1,0208846
30	3,62707736	-2,2604084
31	0,06900211	-2,7608814
32	0,51371315	1,81877126
33	0,54983237	0,14609279
34	4,09554316	4,21763231

F-Test Two-Sample for Variances

	Variable 1	*Variable 2*
Mean	1,88435964	1,4843808
Variance	7,59727603	8,72071529
Observations	120	100
df	119	99
F	0,87117579	
P(F<=f) one-	0,23503862	
F Critical one	0,72967687	

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	1,88435964	1,4843808
Variance	7,59727603	8,72071529
Observations	120	100
Pooled Varia	8,10746175	
Hypothesized	0	
df	218	
t Stat	1,03746593	
P(T<=t) one-	0,15033407	
t Critical one	1,65187337	
P(T<=t) two-	0,30066814	
t Critical two	1,9709056	

Srovnání dvou náhodných veličin

Příklad: 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%?

Řešení v R:

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

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
```

Srovnání dvou náhodných veličin

Příklad: 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%?

\> y <- data.matrix(read.table(„dodavatelY.txt"))

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
```

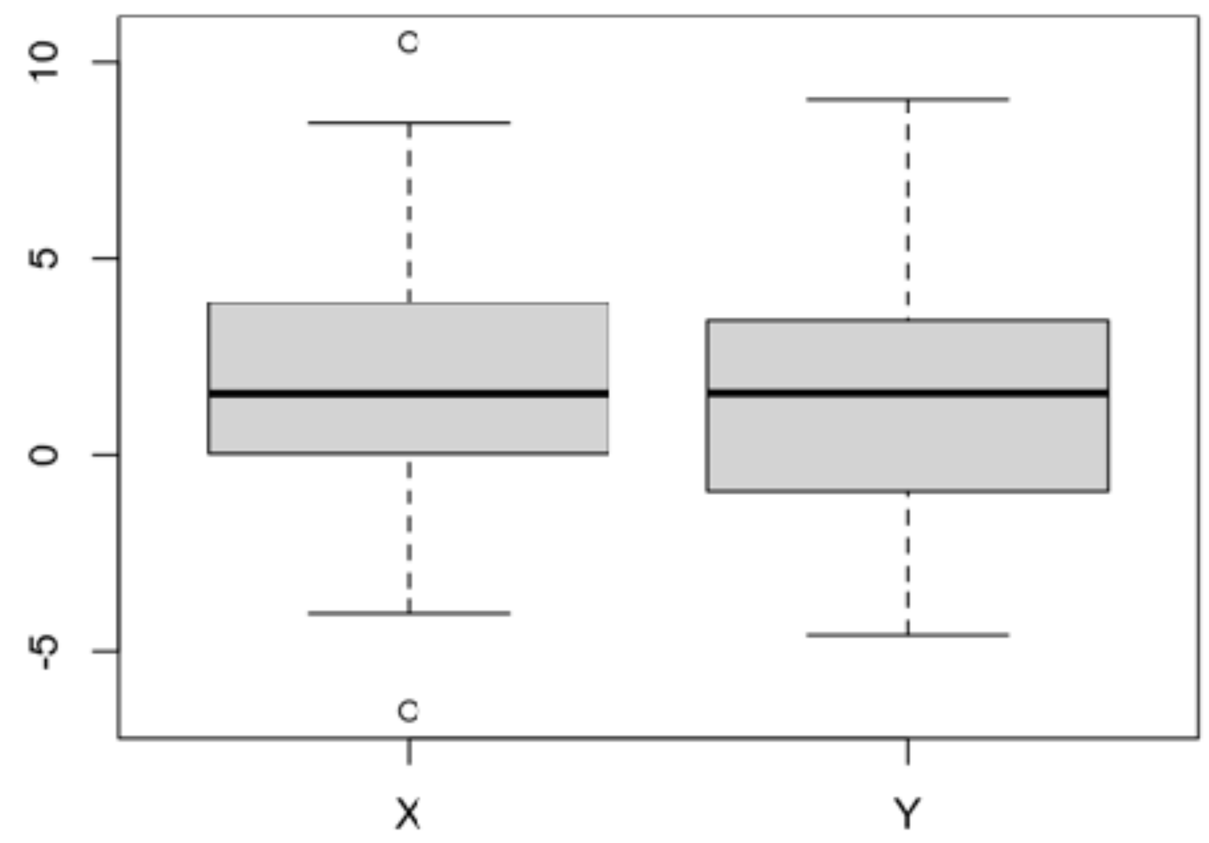
Srovnání dvou náhodných veličin

Příklad: 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%?

1) Vizualizace dat: Box&Whiskers diagram

```
> boxplot(c(x),c(y), names=c("X","Y"))
```



2) Test normality:

```
> 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
```

Srovnání dvou náhodných veličin

Příklad: 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%?

3) Srovnání rozptylů: F-test

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

    F test to compare two variances

data:  x and y
F = 0.8712, 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

Srovnání dvou náhodných veličin

Příklad: 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%?

4) Srovnání středních hodnot: dvouvýběrový t-test se shodnými rozptyly

```
> 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

Srovnání dvou náhodných veličin

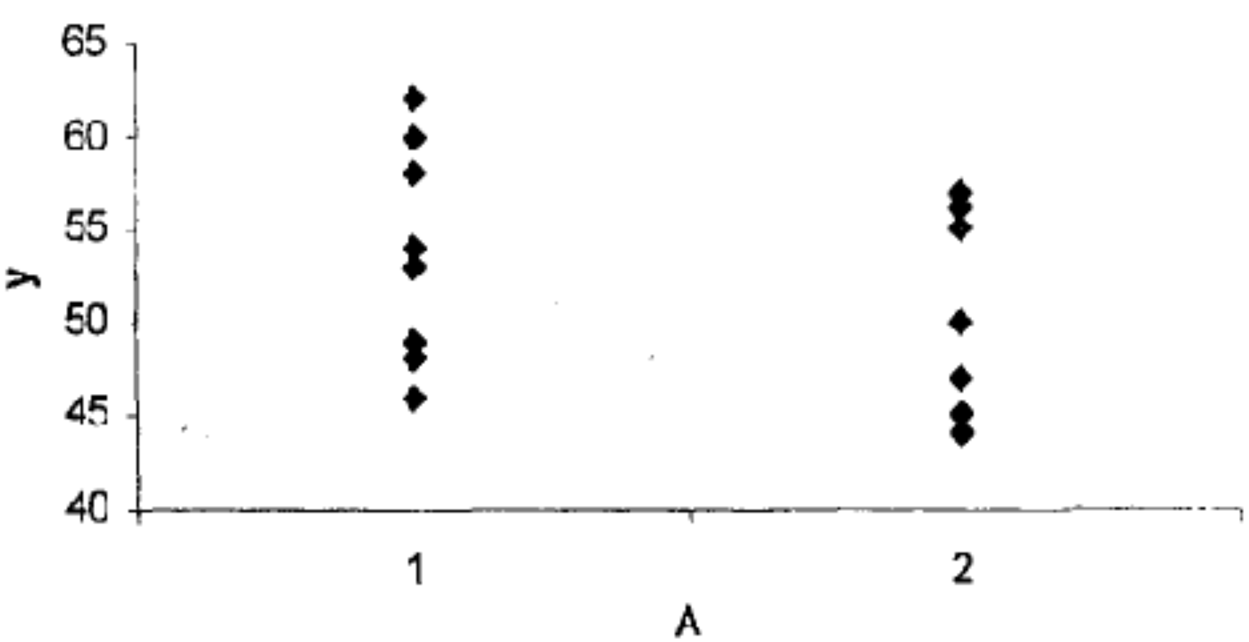
Příklad: Bylo zjišťováno, zda množství vyrobených kusů za hodinu (výkonnost) závisí na typu šicího stroje.

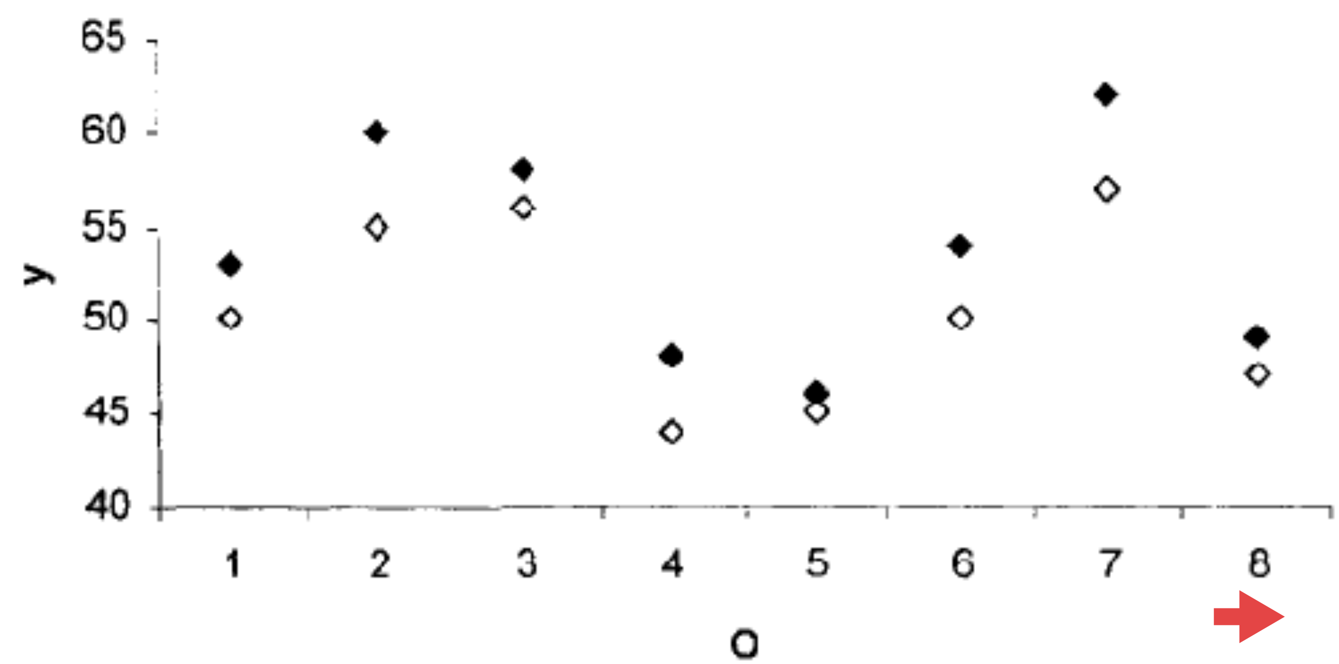
Navrhneme jednofaktorový experiment, 2 úrovně, 8 replikací, 8 bloků => 16 měření

1. 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

2. Grafické znázornění dat:





Srovnání dvou náhodných veličin

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

3. Párový t-test:

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

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

kde

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

t-Test: Párový test pro střední hodnoty

	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

Srovnání dvou náhodných veličin

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

1) Data:

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

```
> pred_cvicenim
  [1] 12.666378  7.322789 15.021706 13.616913 10.970712  5.464451
  [7]  9.999636 15.693764 13.771444 17.065310  6.940708 15.860749
 [13] 18.019348  6.326531 20.647763 23.005369 14.619170 20.787108
 [19] 14.238225  9.674337 14.763170  9.613791  9.727326  9.146292
 [25] 21.246960 16.200128 15.466065 13.691879  9.032113 10.558392
 [31] 18.258896 14.992416 14.722569 10.579842 10.758363  8.894299
 [37] 13.502299 12.994734 14.775563  9.818535 18.208089  8.438143
 [43]  8.282819 11.090392 15.174881  7.704479  8.917742 10.275903
 [49] 11.488700 16.572150 18.892428 13.544225  9.309845 13.713258
 [55] 12.904993  8.951567  9.041688 10.222305 14.136072  9.222289
 [61] 15.208694 14.627659 15.287092 11.389052  7.716052 14.307632
 [67] 14.647653 18.705963 13.665201  8.025347 13.157791 14.336731
 [73]  9.548584 12.522605 11.876452 12.241549 12.944160 17.637175
 [79]  9.854223 17.877400 15.892081  9.893356  7.791175 11.901961
 [85] 15.605362 13.464186 12.451922 16.090626  8.907932 16.333859
 [91] 13.554146 19.586575 11.765020  9.981692  5.325750 20.168371
 [97] 12.485393 14.349888 14.198229  7.315012 16.787920 10.998550
[103] 10.377856 13.531181 12.258939 11.346062 12.998020  8.498104
[109] 14.195263 15.372914 11.698431 12.929311 11.232474 21.551867
[115] 10.436798 14.430260 18.836296 14.838428 14.450987 10.879682
```

Srovnání dvou náhodných veličin

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

1) Data:

> y <- data.matrix(read.table("po_cviceni.txt"))

```
> 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
```

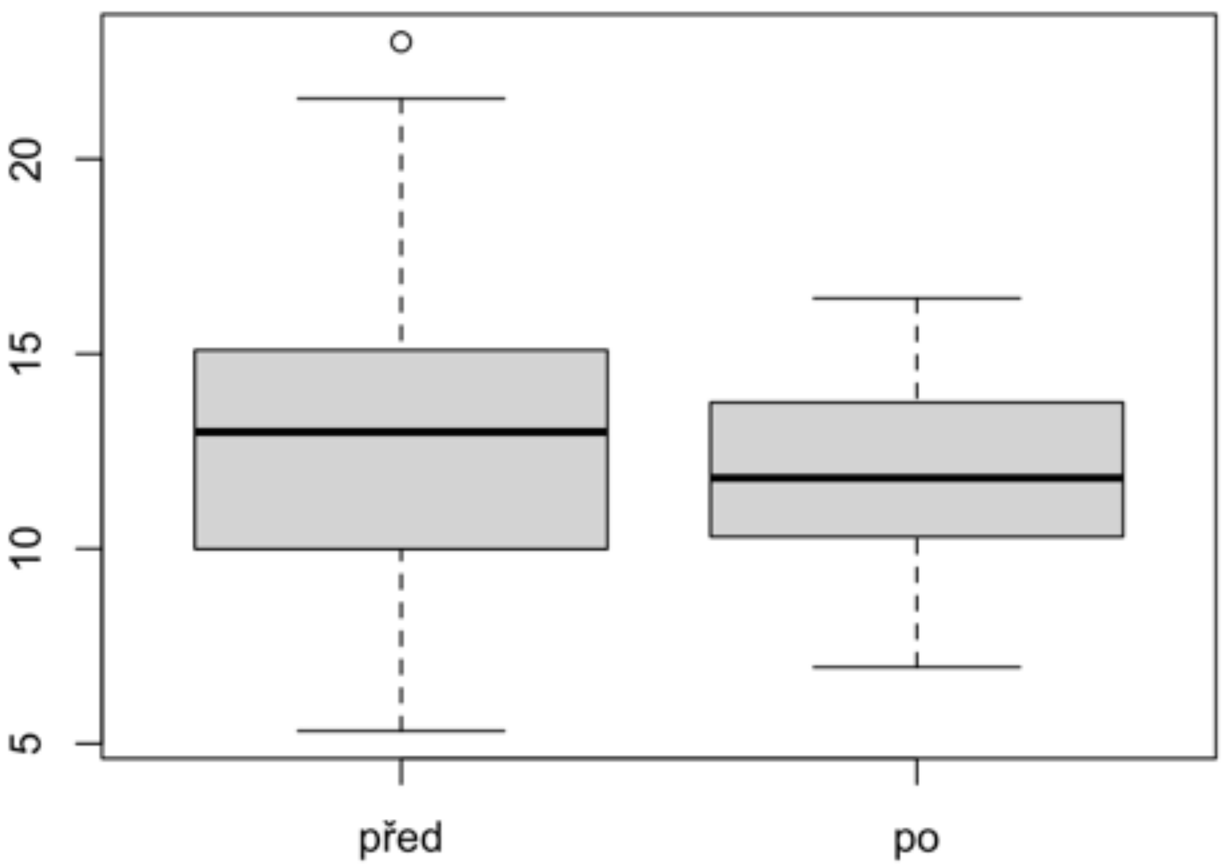
Srovnání dvou náhodných veličin

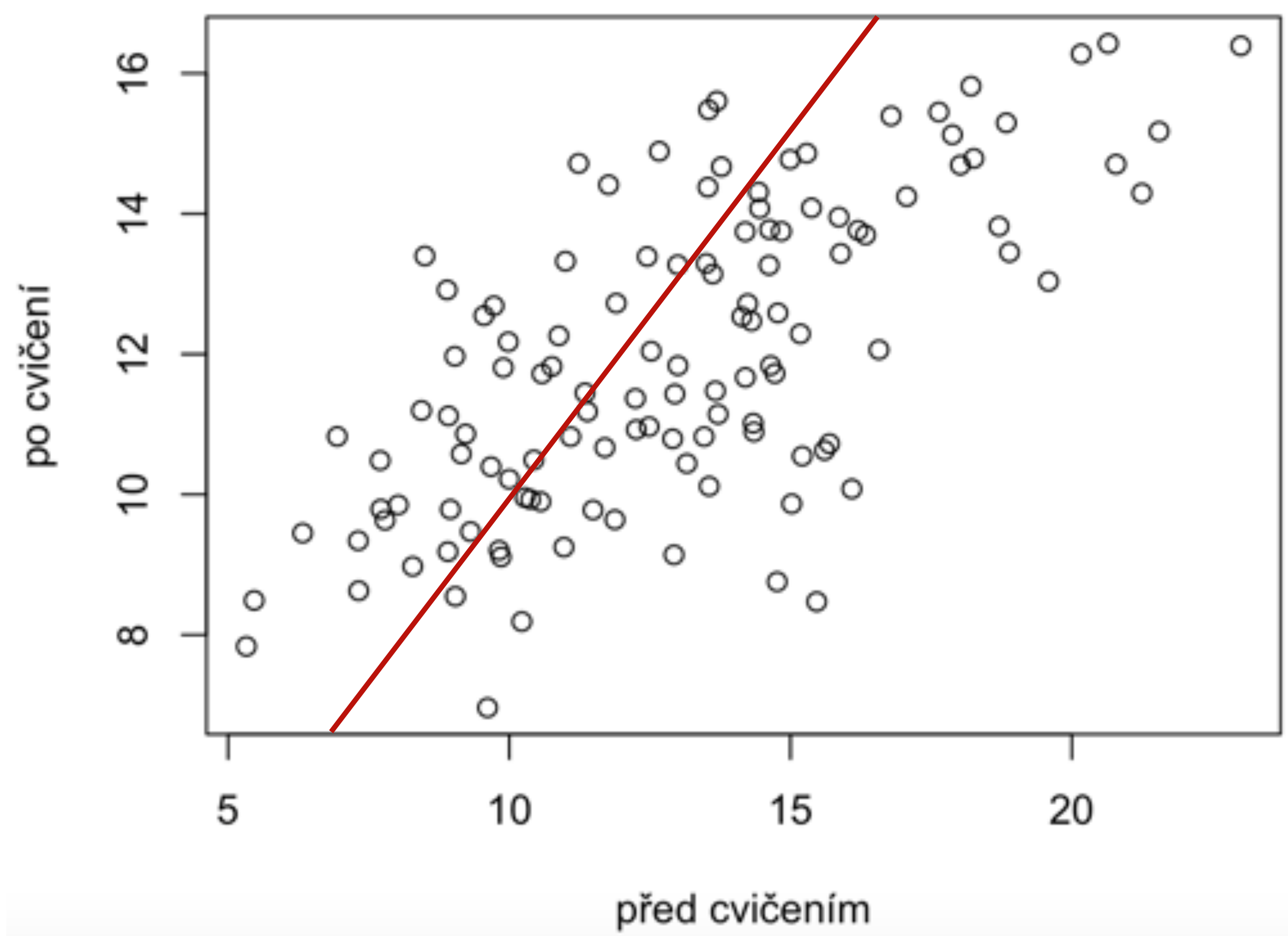
Příklad: Byla měřena rychlost reakce operátorů před a po speciálním cvičení v sekundách. Mělo cvičení statisticky významný vliv na rychlost?

```
> cviceni<- data.frame(x, y, x-y)
```

2) Grafické zobrazení

```
> plot(cviceni[,1], cviceni[,2], xlab="před cvičením", ylab="po cvičení")
> lines(c(6:18), c(6:18), col="red")
> boxplot(cviceni[,1],cviceni[,2])
```



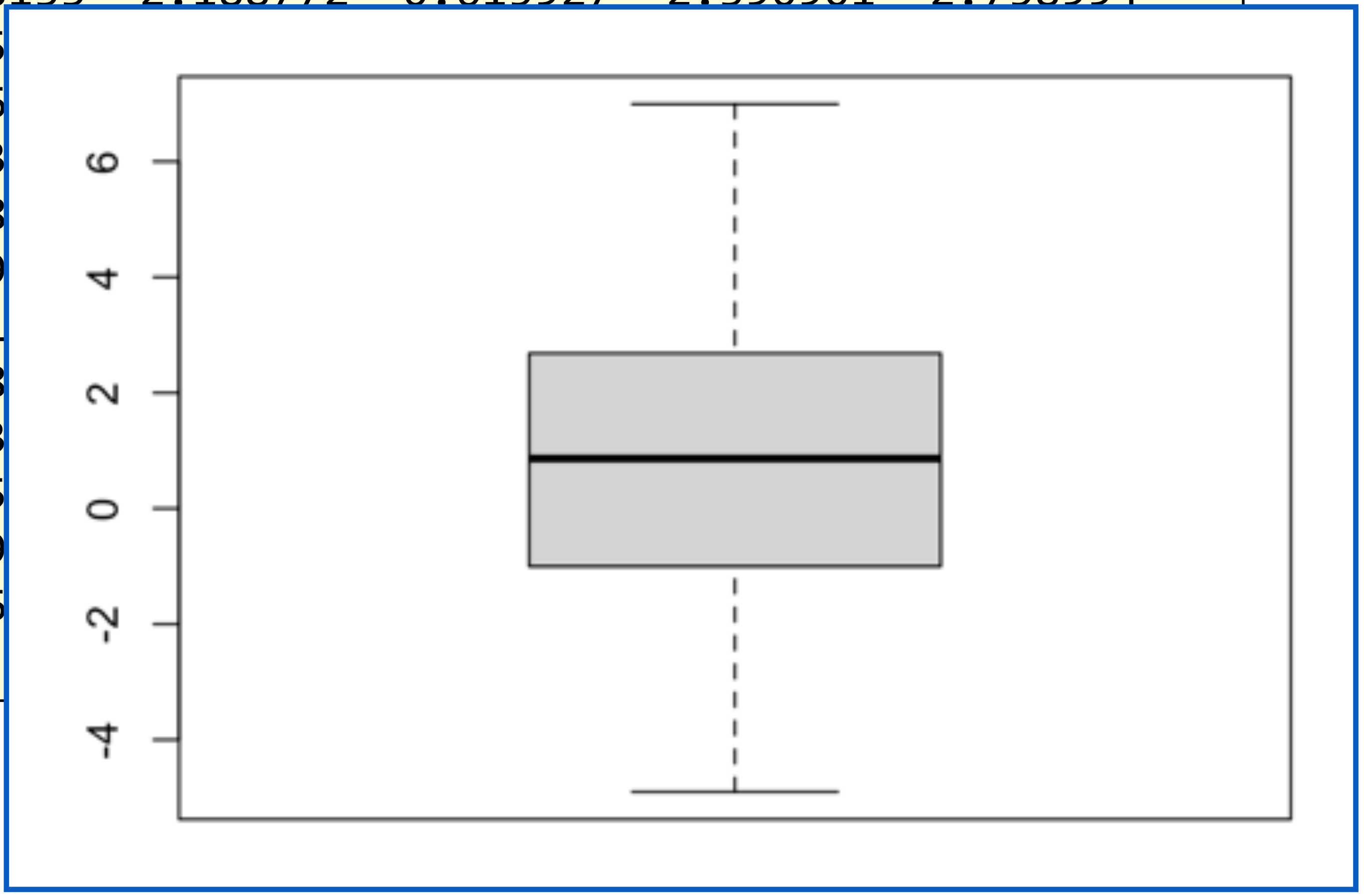


Srovnání dvou náhodných veličin

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

3) Rozdíl:

```
> cviceni[,3]
  [1] -2.223001 -1.304823  5.154251  0.475745  1.721590 -3.026323 -0.217654
  [8]  4.969361 -0.898006  2.821366 -3.886197  1.909228  3.325947 -3.123031
 [15]  4.221875  6.612680  1.353696  6.082114  1.520118 -0.721048  6.006894
 [22]  2.652270 -2.961171 -1.432050  6.952896  2.437096  6.993741 -1.913374
 [29] -2.936823  0.661108  3.470691  0.219038  2.999233 -1.139622 -1.066044
 [36] -4.020186  0.210494 -0.278133  2.188772  0.615927  2.390901 -2.758994
 [43] -0.691591  0.266450  2.885
 [50]  4.510066  5.442456 -1.936
 [57]  0.494108  2.033358  1.603
 [64]  0.208384 -2.074414  1.838
 [71]  2.716901  3.321174 -2.999
 [78]  2.188111  0.744171  2.751
 [85]  4.976839  2.639712 -0.938
 [92]  6.550508 -2.647074 -2.193
 [99]  2.529940 -2.025255  1.395
[106] -0.102258  1.161859 -4.899
[113] -3.484147  6.378183 -0.056
[120] -1.381456
```




```
> boxplot(cviceni[,3])
```

Srovnání dvou náhodných veličin

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

4) Párový t-test:

```
> t.test(cviceni[,3], mu=0)

	One Sample t-test

data:  cviceni[, 3]
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 zamítáme, cvičení mělo vliv a rychlost reakce se statisticky významně zvýšila

Srovnání dvou náhodných veličin

a) dvě nezávislá měření
b) párová měření

Dvě nezávislá měření

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

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

Neparametrické testy pro srovnání dvou souborů dat

- jsou založeny na **pořadových statistikách:**

1)Dvouvýběrový Wilcoxonův test

2)Párový (jednovýběrový) Wilcoxonův test

3)Znaménkový test

Srovnání dvou náhodných veličin

Dvouvýběrový Wilcoxonův test $H_0 : \mu_X = \mu_Y$ $H_A : \mu_X \neq \mu_Y$

- 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_l^Y = pořadí l-tého pozorování Y ve spojeném souboru
  - R^X = součet všech R_k^X pro k=1, …, n a R^Y= součet všech R_l^Y pro l=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 (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 $E(R^X) = \frac{n(m+n+1)}{2}, \quad var(R^X) = \frac{mn(m+n+1)}{12}$

Tato statistika má přibližně N(0,1) rozdělení.

Srovnání dvou náhodných veličin

Dvouvýběrový Wilcoxonův test $H_0 : \mu_X = \mu_Y$ $H_A : \mu_X \neq \mu_Y$

Příklad: 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%?

```
> x <- data.matrix(read.table("dodavatelX.txt"))
> y <- data.matrix(read.table("dodavatelY.txt"))
> wilcox.test(x, y, alternative = „two.sided")
```

```
	Wilcoxon rank sum test with continuity correction

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

Srovnání dvou náhodných veličin

Dvouvýběrový Wilcoxonův test $H_0 : \mu_X = \mu_Y$ $H_A : \mu_X \neq \mu_Y$

Příklad: 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%?

```
> x <- data.matrix(read.table("dodavatelX.txt"))
> y <- data.matrix(read.table("dodavatelY.txt"))
> wilcox.test(x, y, alternative = „two.sided"), mu = 0, paired = FALSE,
    exact = NULL, correct = TRUE, conf.int = FALSE, conf.level = 0.95)

    Wilcoxon rank sum test with continuity correction

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

Srovnání dvou náhodných veličin

Párový Wilcoxonův test $H_0 : \mu_X = \mu_Y$ $H_A : \mu_X \neq \mu_Y$

- 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.

Srovnání dvou náhodných veličin

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

```
> x <- data.matrix(read.table("pred_cvicenim.txt"))
> y <- data.matrix(read.table("po_cviceni.txt"))
> wilcox.test(x, y, alternative = "two.sided", mu = 0, paired = TRUE,
       correct = TRUE, conf.level = 0.95)

    Wilcoxon signed rank test with continuity correction

data:  x and y
V = 4973, p-value = 0.0004384
alternative hypothesis: true location shift is not equal to 0
```