

## Dvouvýběrové testy

### Parametrický případ

#### Dvouvýběrový t-test

Máme dva nezávislé náhodné výběry, první pochází z rozložení  $N(\mu_1, \sigma_1^2)$  a má rozsah  $n_1 \geq 2$ , druhý pochází z rozložení  $N(\mu_2, \sigma_2^2)$  a má rozsah  $n_2 \geq 2$ .

Označme  $M_1, M_2$  výběrové průměry,  
 $S_1^2, S_2^2$  výběrové rozptyly,

$$S_*^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{vážený průměr výběrových rozptylů.}$$

Na hladině významnosti  $\alpha$  testujeme  $H_0: \mu_1 - \mu_2 = c$  proti  $H_1: \mu_1 - \mu_2 \neq c$  (často volíme  $c = 0$ ).

$$\text{Testová statistika: } T_0 = \frac{(M_1 - M_2) - c}{S_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \quad \text{Platí-li } H_0, T_0 \sim t(n_1 + n_2 - 2).$$

$$\text{Kritický obor: } W = \left(-\infty, -t_{1-\alpha/2}(n_1 + n_2 - 2)\right) \cup \left(t_{1-\alpha/2}(n_1 + n_2 - 2), \infty\right)$$

$T_0 \in W \Rightarrow H_0$  zamítáme na hladině významnosti  $\alpha$ .

Před provedením dvouvýběrového t-testu doporučují někteří autoři ověřit shodu rozptylů **F-testem**.

Na hladině významnosti  $\alpha$  testujeme  $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$  proti  $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ .

Testová statistika:  $T_0 = \frac{S_1^2}{S_2^2}$ . Platí-li  $H_0$ ,  $T_0 \sim F(n_1-1, n_2-1)$ .

Kritický obor:  $W = (0, F_{\alpha/2}(n_1-1, n_2-1)) \cup (F_{1-\alpha/2}(n_1-1, n_2-1), \infty)$

$T_0 \in W \Rightarrow H_0$  zamítáme na hladině významnosti  $\alpha$ .

Jiná možnost – použít dvouvýběrový t-test se samostatnými odhady rozptylů.

**Upozornění:** t-testy jsou při větších rozsazích výběrů (nad 30) robustní vůči porušení předpokladu normality. Pro výběry malých rozsahů lze použít např. Boxovu – Coxovu transformaci nebo je možné provést některý z neparametrických testů.

## Neparametrický případ

### Dvouvýběrový Wilcoxonův test

Máme dva nezávislé náhodné výběry, první pochází ze spojitého rozložení s mediánem  $x_{0,50}$  a má rozsah  $n$ , druhý pochází ze spojitého rozložení s mediánem  $y_{0,50}$  a má rozsah  $m$ . Předpokládáme, že distribuční funkce těchto dvou rozložení se mohou lišit pouze posunutím. Testujeme hypotézu, že distribuční funkce těchto rozložení jsou shodné (neboli mediány jsou shodné) proti alternativě, že jsou rozdílné.

Všech  $n + m$  hodnot uspořádáme vzestupně podle velikosti. Zjistíme součet pořadí hodnot 1. výběru a označíme ho  $T_1$ . Součet pořadí hodnot 2. výběru označíme  $T_2$ .

Vypočteme statistiky  $U_1 = mn + n(n+1)/2 - T_1$ ,  $U_2 = mn + m(m+1)/2 - T_2$ .

Přitom platí  $U_1 + U_2 = mn$ .

Pokud  $\min(U_1, U_2) \leq$  tabelovaná kritická hodnota (pro dané rozsahy výběrů  $m$ ,  $n$  a dané  $\alpha$ ), pak nulovou hypotézu o totožnosti obou distribučních funkcí zamítáme na hladině významnosti  $\alpha$ .

Pro velká  $n$ ,  $m$  (prakticky  $n, m > 30$ ) lze využít asymptotické normality statistiky  $U_1$ . V případě platnosti  $H_0$

má statistika  $U_0 = \frac{U_1 - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}}$  asymptoticky rozložení  $N(0,1)$ . Kritický obor pro oboustrannou alternativu

má tvar:  $W = (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty)$ . (Analogicky pro jednostranné alternativy.)  $H_0$  zamítáme na asymptotické hladině významnosti  $\alpha$ , když  $U_0 \in W$ .

## Dvouvýběrový Kolmogorovův - Smirnovův test

Máme dva nezávislé náhodné výběry ze dvou spojitých rozložení, jejichž distribuční funkce se mohou lišit nejenom posunutím, ale také tvarem. Testujeme hypotézu, že distribuční funkce těchto rozložení jsou shodné, tj., že všech  $n+m$  veličin pochází z téhož rozložení proti alternativě, že distribuční funkce jsou rozdílné.

Označme  $F_1(x) = \frac{1}{n} \text{card}\{i; X_i \leq x\}$  výběrovou distribuční funkci 1. výběru,

$F_2(y) = \frac{1}{m} \text{card}\{i; Y_i \leq y\}$  výběrovou distribuční funkci 2. výběru.

Testová statistika:  $D = \max_{-\infty < x < \infty} |F_1(x) - F_2(x)|$ .

$H_0$  zamítáme na hladině významnosti  $\alpha$ , když  $D \geq D_{n,m}(\alpha)$ , kde  $D_{n,m}(\alpha)$  je tabelovaná kritická hodnota.

Pro větší rozsahy  $n, m$  lze kritickou hodnotu aproximovat vzorcem  $\sqrt{\frac{n+m}{2nm} \ln \frac{2}{\alpha}}$ .

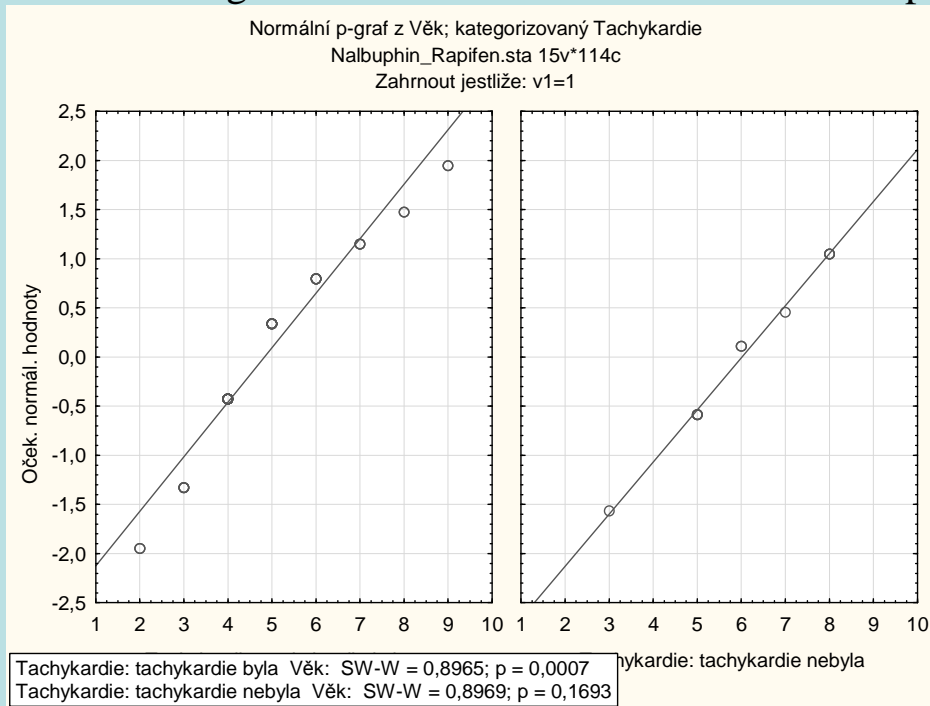
## Ukázka dvouvýběrových testů

Uvažme pacienty s Nalbuphinem.

$H_0$ : Rozložení věku je stejné ve skupinách pacientů bez tachykardie a s tachykardií.

$H_1$ : Rozložení věku je rozdílné ve skupinách pacientů bez tachykardie a s tachykardií

Pomocí N-P grafu ověříme normalitu dat v obou skupinách.



Pro skupinu pacientů, u nichž se vyskytla tachykardie, S-W test zamítá hypotézu o normalitě rozložení věku. Tečky se však od přímky odchyľují jen málo a rozsah souboru je velký (45), data budeme považovat za normálně rozložená.

## Výsledky dvouvýběrového t-testu:

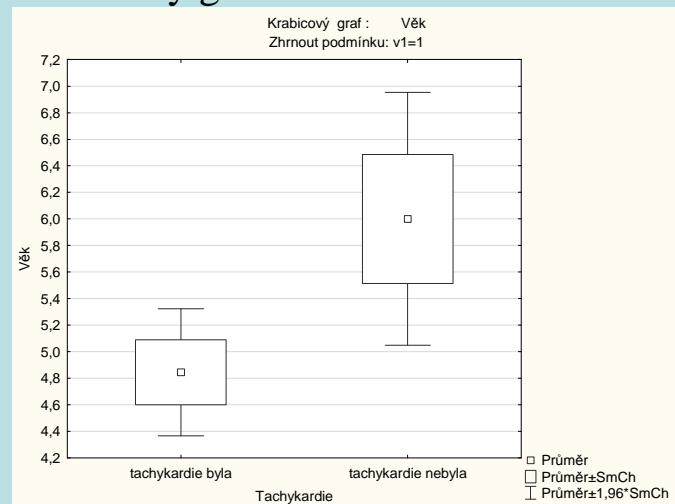
t-testy; grupováno: Tachykardie (Nalbuphin_Rapifen.sta)											
Skup. 1: tachykardie byla											
Skup. 2: tachykardie nebyla											
Zhrnout podmínku: v1=1											
Proměnná	Průměr tachykardie byla	Průměr tachykardie nebyla	t	sv	p	Poč.plat tachykardie byla	Poč.plat tachykardie nebyla	Sm.odch. tachykardie byla	Sm.odch. tachykardie nebyla	F-poměr Rozptyly	p Rozptyly
Věk	4,844444	6,000000	-2,10450	54	0,040005	45	11	1,637009	1,612452	1,030692	1,000000

Pacientů s tachykardií bylo 45, průměrný věk pacientů byl 4,84 roku a směrodatná odchylka činila 1,64 roku. Pacientů bez tachykardie bylo 11, průměrný věk pacientů byl 6 let roku a směrodatná odchylka činila 1,03 roku.

F-test nezamítá hypotézu o shodě rozptylů na hladině významnosti 0,05 (testová statistika = 1,0307, p-hodnota = 1).

Dvouvýběrový t-test zamítá hypotézu o shodě středních hodnot věku na hladině významnosti 0,05 (testová statistika = -2,1045, p-hodnota = 0,04).

### Krabicový graf

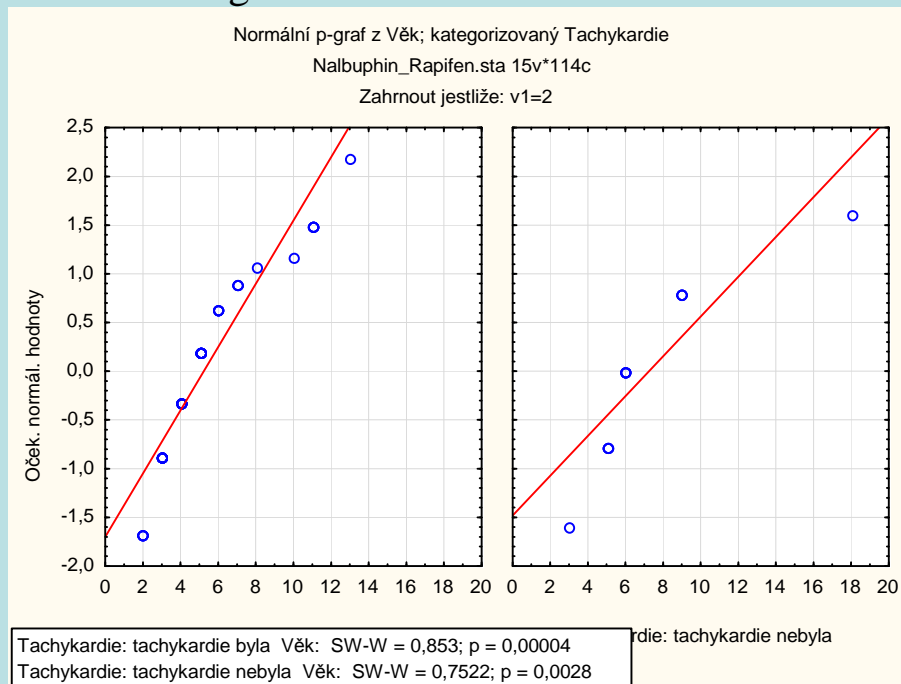


Uvažme pacienty s Rapifenem.

$H_0$ : Rozložení věku je stejné ve skupinách pacientů bez tachykardie a s tachykardií.

$H_1$ : Rozložení věku je rozdílné ve skupinách pacientů bez tachykardie a s tachykardií

Pomocí N-P grafu ověříme normalitu dat v obou skupinách.



Pro obě skupiny pacientů S-W test zamítá hypotézu o normalitě rozložení věku. Tečky se v obou případech odchyľují od přímky výrazněji. Použijeme raději neparametrické testy.

### Výsledek dvouvýběrového Wilcoxonova testu:

Mann-Whitneyův U Test (w/ oprava na spojitost) (Nalbuphin_Rapifen.sta)										
Dle proměn. Tachykardie										
Označené testy jsou významné na hladině $p < ,05000$										
Zhrnout podmínku: $v1=2$										
Proměnná	Sčt poř. tachykardie byla	Sčt poř. tachykardie nebyla	U	Z	p-hodn.	Z upravené	p-hodn.	N platn. tachykardie byla	N platn. tachykardie nebyla	2*1str. přesné p
Věk	1243,000	468,0000	162,0000	-2,17867	0,029357	-2,20654	0,027347	46	12	0,028026

U pacientů s Rapifenem se tachykardie vyskytla ve 46 případech, ve 12 případech nikoliv.

Zajímá nás p-hodnota označená jako 2\*1str. přesné p (používá se pro rozsahy výběrů pod 30). V našem případě  $p = 0,028$ , tedy na asymptotické hladině významnosti 0,05 zamítáme hypotézu, že rozložení věku je stejné ve skupinách pacientů s tachykardií a bez tachykardie.

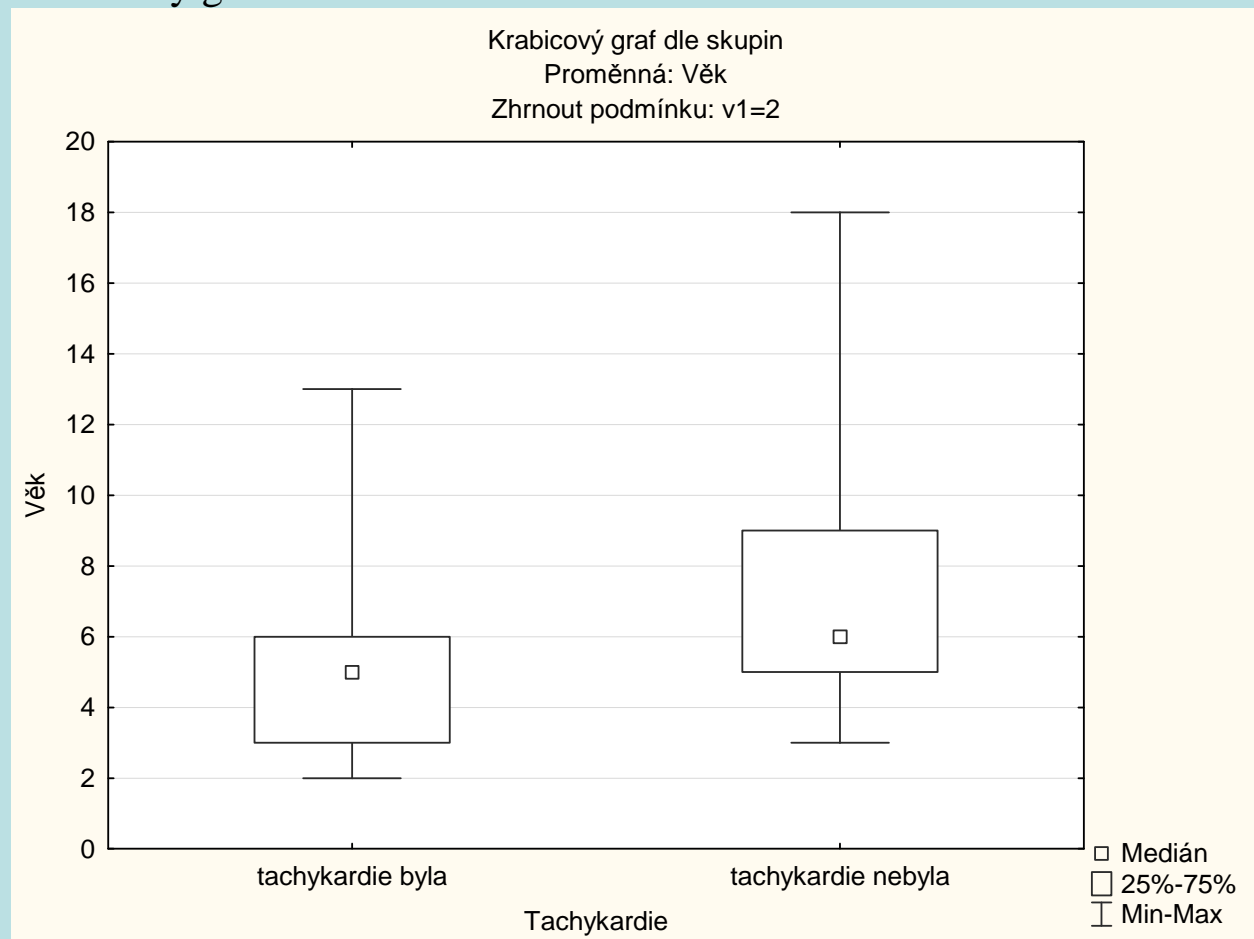
### Výsledek dvouvýběrového K- S testu:

Kolmogorov-Smirnovův test (Nalbuphin_Rapifen.sta)									
Dle proměn. Tachykardie									
Označené testy jsou významné na hladině $p < ,05000$									
Zhrnout podmínku: $v1=2$									
Proměnná	Max záp rozdíl	Max klad rozdíl	p-hodn.	Průměr tachykardie byla	Průměr tachykardie nebyla	Sm.odch. tachykardie byla	Sm.odch. tachykardie nebyla	N platn. tachykardie byla	N platn. tachykardie nebyla
Věk	-0,373188	0,047101	$p > .10$	5,260870	7,250000	2,727920	3,864171	46	12

Odpovídající p-hodnota je větší než 0,1, tedy K – S test na hladině významnosti 0,05 nezamítá hypotézu o shodném rozložení věku v obou skupinách pacientů.



## Krabicový graf



## Cohenův koeficient věcného účinku (doplnění významu dvouvýběrového t-testu)

Cohenův koeficient slouží k posouzení velikosti rozdílu průměrů, který je standardizován pomocí odmocniny z váženého průměru výběrových rozptylů. Jedná se o tzv. **věcnou významnost** neboli **velikost účinku** skupiny na variabilitu hodnot sledované náhodné veličiny.

Počítá se podle vzorce:  $d = \frac{|m_1 - m_2|}{s_*}$ , kde  $m_1, m_2$  jsou výběrové průměry a  $s_*^2$  je vážený průměr výběrových rozptylů.

Velikost účinku hodnotíme podle následující tabulky:

Hodnota d	účinek
aspoň 0,8	velký
mezi 0,5 až 0,8	střední
mezi 0,2 až 0,5	malý
pod 0,2	zanedbatelný

Vypočteme Cohenův koeficient pro věk pacientů s Nalbuphinem, u nichž se vyskytla resp. nevyskytla tachykardie.

	1	2	3	4	5	6	7
	n1	n2	m1	m2	s1	s2	d
1	45	11	4,844444	6,000000	1,637009	1,612452	0,70784888

Protože  $d = 0,71$ , můžeme vliv skupiny na variabilitu věku hodnotit jako středně silný.