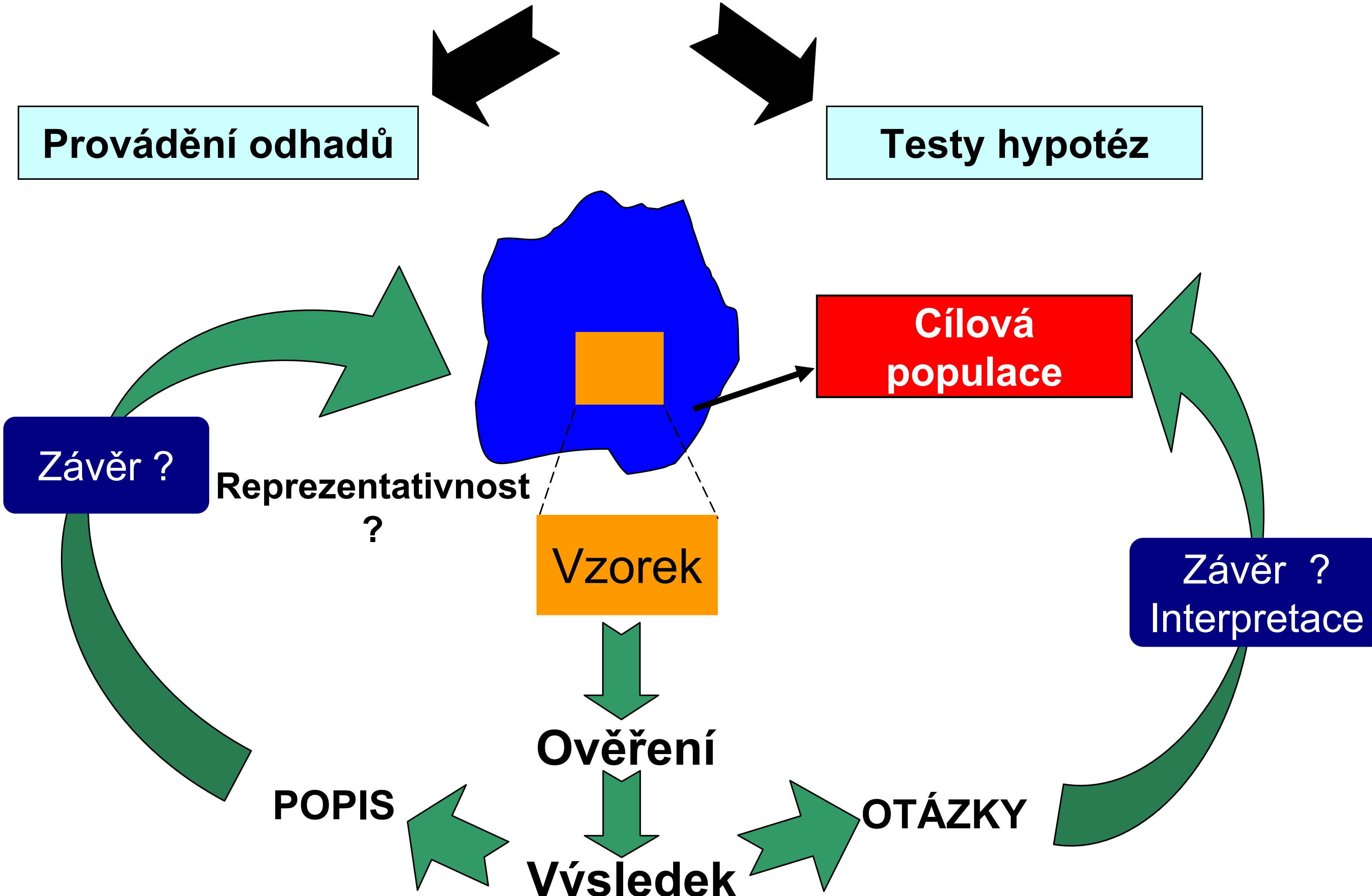
Testy hypotéz - úvod

Statistika v průzkumném studiu



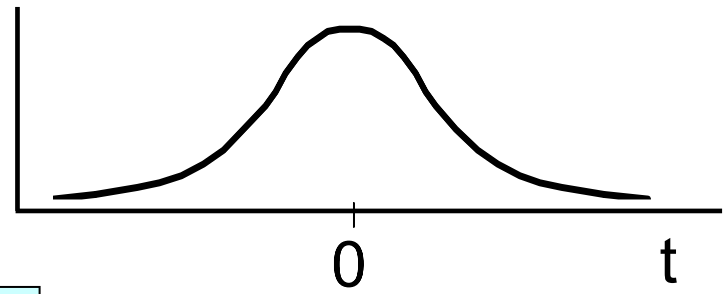
Elementární prvky statistických testů

Nulová hypotéza H_O

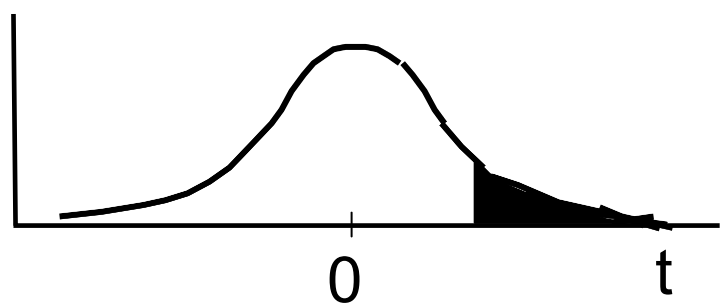
Alternativní hypotéza H_A

Testová statistika

$$t = \frac{\overline{x} - \mu}{s}\sqrt{n}$$



Kritický obor testované statistiky



Experimentální design

Optimální velikost vzorku nebo počet opakování

Efektivní uspořádání experimentů

Účelná minimalizace chyb

		Závěr testu	
		Platí	**Neplatí**
Skutečnost	**Platí**	**1-**α	α
	Neplatí	β	**1-**β

Experimentální design

Pravděpodobnost chyby 1. druhu

α → **P nesprávného zamítnutí nulové hypotézy**

Pravděpodobnost chyby 2. druhu

β → **P nerozpoznání platné nulové hypotézy**

Síla testu

$1-\beta$ → **Pravděpodobnostně vyjádřená schopnost rozpoznat neplatnost hypotézy**

Předpoklady a pojmy statistických testů

Parametrické vs. neparametrické testy

Parametrické testy

- Mají předpoklady o rozložení vstupujících dat (např. normální rozložení)
- Při stejném N a dodržení předpokladů mají vyšší sílu testu než testy neparametrické
- Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být zcela chybný a nesmyslný

Neparametrické testy

- Nemají předpoklady o rozložení vstupujících dat, lze je tedy použít i při asymetrickém rozložení, odlehlých hodnotách, či nedetekovatelném rozložení
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí

One-sample vs. two sample testy

One – sample testy

- Srovnávají jeden vzorek (one sample, jednovýběrové testy) s referenční hodnotou (popřípadě se statistickým parametrem cílové populace)
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace)
- Otázka položená v testu může být vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek

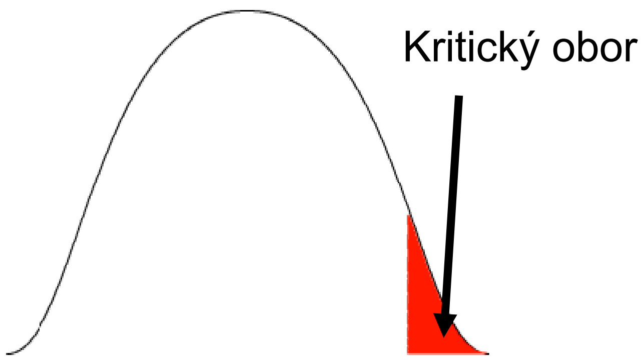
Two – sample testy

- Srovnávají navzájem dva vzorky (two sample, dvouvýběrové vzorky)
- V testu jsou srovnávány dvě rozložení hodnot
- Otázka položená v testu může být opět vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek
- Kromě testů pro dvě skupiny hodnot existují samozřejmě i testy pro více skupin dat

One-tailed vs. Two-tailed testy

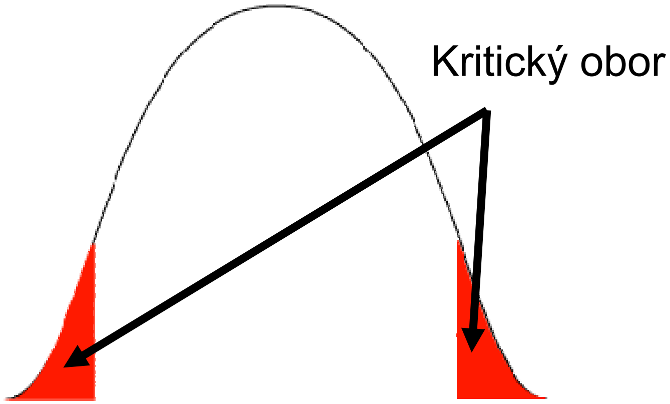
One – tailed testy

- Hypotéza testu je postavena asymetricky, tedy ptáme se na **větší než/ menší než**
- Test může mít pouze dvojí výstup – jedna z hodnot je větší (menší) než druhá a všechny ostatní případy



Two – tailed testy

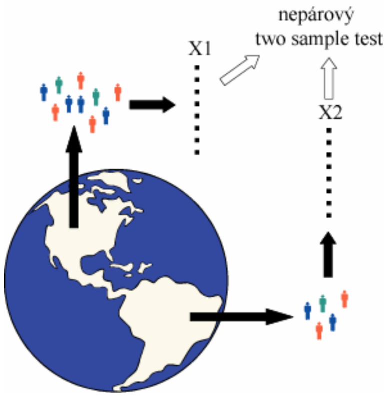
- Hypotéza testu se ptá na otázku **rovná se/nerovná se**
- Test může mít trojí výstup – **menší - rovná se – větší než**
- Situace **nerovná se** je tedy souhrnem dvou možných výstupů testu (**menší+větší**)



Nepárový vs. Párový design

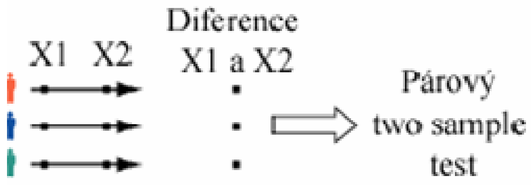
Nepárový design

- Skupiny srovnávaných dat jsou na sobě zcela nezávislé (též nezávislý, independent design), např. lidé z různých zemí, nezávislé skupiny pacientů s odlišnou léčbou atd.
- Při výpočtu je nezbytné brát v úvahu charakteristiky obou skupin dat

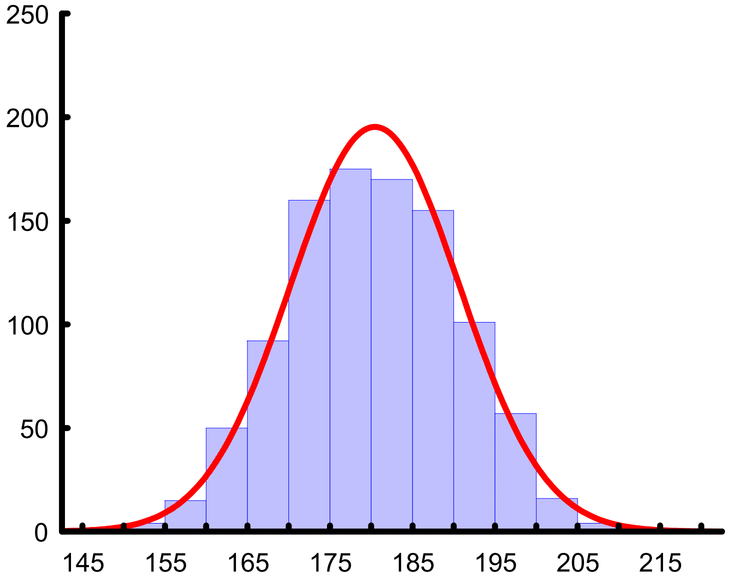


Párový design

- Mezi objekty v srovnávaných skupinách existuje vazba, daná např. člověkem před a po operaci, reakce stejného kmene krys atd.
- Vazba může být buď přímo dána nebo pouze předpokládána (v tom případě je nutné ji ověřit)
- Test je v podstatě prováděn na diferencích skupin, nikoliv na jejich původních datech



- Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.



•Test dobré shody

V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí $\chi 2$ testu dobré shody. Test dává dobré výsledky, ale je náročný na n, tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

•Kolgomorov Smirnov test

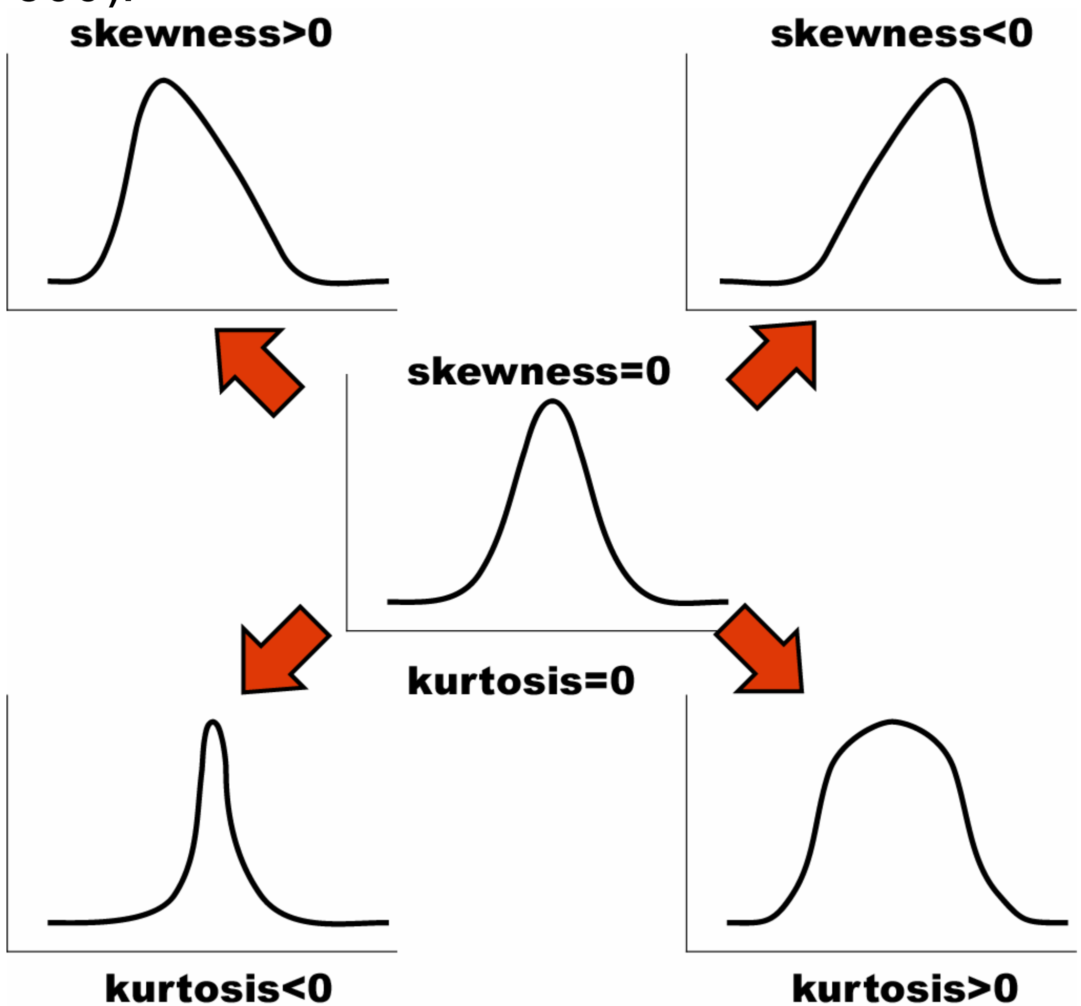
Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložení. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Měl by být počítán pouze v případě, že známe průměr a směrodatnou odchylku hypotetického rozložení, pokud tyto hodnoty neznáme, měla by být použita jeho modifikace – Lilieforsův test.

•Shapiro-Wilk`s test

Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou sílou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.

Šikmost a špičatost jako testy normality

- Parametry normálního rozložení, skewness a kurtosis mohou být využity pro testování normality, ale pouze pro velké vzorky (šikmost – 100, špičatost – 500).



Statistické testy o parametrech jednoho výběrů

"One sample" testy

V případě one sample testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení.

Průměr – cílová vs. výběrová populace

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

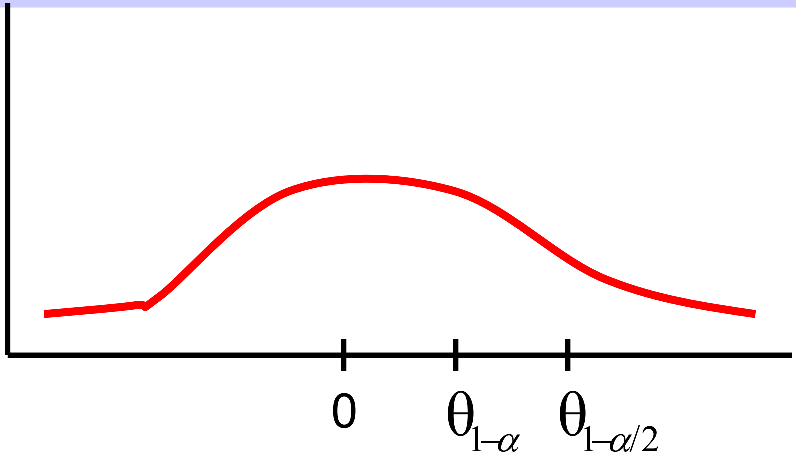
H_0	H_A	Testová statistika	Interval spolehlivosti
$\bar{x} \leq \mu$	$\bar{x} > \mu$	t	$t > t_{1-\alpha}^{(n-1)}$
$\bar{x} \geq \mu$	$\bar{x} < \mu$	t	$t < t_{\alpha}^{(n-1)}$
$\bar{x} = \mu$	$\bar{x} \neq \mu$	t	$\lvert t \rvert > t_{1-\alpha/2}^{(n-1)}$

Rozptyl – cílová vs. výběrová populace

$$\chi^2 = \frac{(n-1).s^2}{\sigma^2}$$

H_0	H_A	Testová statistika	Interval spolehlivosti
$s^2 \leq \sigma^2$	$s^2 > \sigma^2$	χ^2	$\chi^2 > \chi^{2\ (n-1)}_{1-\alpha}$
$s^2 \geq \sigma^2$	$s^2 < \sigma^2$	χ^2	$\chi^2 < \chi^{2\ (n-1)}_{\alpha}$
$s^2 = \sigma^2$	$s^2 \neq \sigma^2$	χ^2	$\chi^2 > \chi^2_{1-\alpha/2}$ **nebo** $\chi^2 < \chi^2_{\alpha/2}$

“One sample“ testing vs one/two tailed



α = 0,05

Pokud “two – tailed” test vyjde významný tak, že P $\leq$ 0,05, pak dobře zvolený “one – tailed” test je významný při P $\leq$ 0,025.

. . . tzn.že testová charakteristika > $\theta_{1-\alpha/2}$ a “one – tailed” testy na hladině jsou v podstatě zbytečné.

Pokud je pro “two – tailed” test P = 0,1, pak lze na hladině α = 0,05 prokázat nerovnost srovnávaných parametrů vhodně voleným “one – tailed” testem.

. . . tzn. že testová charakteristika $\in \left\langle \theta_{1-\alpha} ; \theta_{1-\alpha/2} \right\rangle$

Koncentrace antibiotika v cílovém orgánu

Při 1000 měřeních antibiotika byla zjištěna v cílovém orgánu průměrná koncentrace 202,5 jednotek a směrodatná odchylka 44 jednotek.

Požadovaná koncentrace antibiotika je 200 jednotek.

1) Je daný rozdíl 2,5 významný vzhledem k variabilitě znaku na hladině významnosti 5%?

2) Jaká je skutečná hladina významnosti?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{2,5}{44}\sqrt{1000} = 1,797$$

Srovnání odhadu průměru s předpokládanou hodnotou II

Aktivita enzymu v buňkách

Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

1. otázka zní, zda se naměřené hodnoty našeho vzorku liší od výsledků dřívější rozsáhlé studie zaměřené na celou cílovou populaci, kde byla zjištěna průměrná aktivita 2,5 jednotky?

H0: x=μ tedy two tailed test

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{3,5 - 2,5}{1}\sqrt{25} = 5$$

$$t_{0,975}^{24} = 2,064 \Rightarrow t > t_{1-\alpha/2}^{24} \Rightarrow$$ **H0 zamítnuta při $\alpha \leq 0,05$**

od jiné hodnoty bychom zachytili při daných hodnotách?

2. otázka – jakou minimální odchylku X od jiné hodnoty bychom zachytili při daných hodnotách?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \rightarrow d = \frac{t_{1-\alpha/2}^{\nu}}{\sqrt{n}}s \rightarrow d = \frac{2,064}{5}1$$

3. za předpokladu, že z praktického hlediska je významná odchylka již 0,2 jednotky, jaký minimální počet měření musíme provést, abychom ji byli schopni prokázat ?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \rightarrow n = \left(\frac{t_{1-\alpha/2}^{\nu}}{d}s\right)^2$$

Srovnání odhadu průměru s předpokládanou hodnotou III

x: Aktivita enzymu v buňkách
n = 25; $\bar{x}$ = 3,5; s = 1
μ: Hodnota zjištěná při předcházejícím, dlouhodobém průzkumu

H_0 : $\bar{x} = \mu$

$$t = \frac{3{,}5 - 2{,}5}{1}\sqrt{25} = 5$$

Kvantil $t_{0,975}^{(24)} = 2{,}064$

$$t > t_{1-\alpha/2}^{(24)}$$

H_0 zamítnuta při $\alpha \leq 0{,}05$

Situace: Odhad průměrné hodnoty znaku X

? Jakou minimální odchylku X od nějaké jiné hodnoty zachytíme jako významnou při daném n, a, b ?

$$d = \sqrt{\frac{s^2}{n}}\left(t^{\nu}_{1-\alpha/2} + t^{\nu}_{1-\beta}\right)$$

Nechť $\alpha = 0{,}05$; $\beta = 0{,}10$; n = 25 ; $s^2 = 1{,}5682$

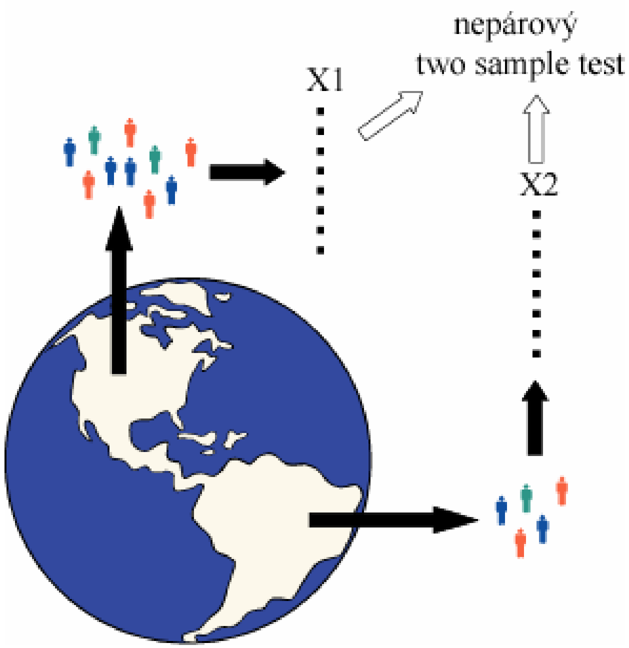
$t_{1-\alpha/2}^{(24)} = 2{,}064$ $t_{1-\beta}^{(24)} = 1{,}318$

$$d = \sqrt{\frac{1{,}5682}{25}}(2{,}064 + 1{,}318) = \underline{\underline{0{,}85}}$$

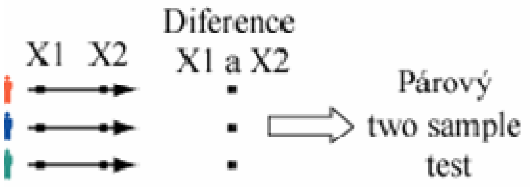
Statistické testy o parametrech dvou výběrů

Two sample testy

- Při použití two sample testů srovnáváme spolu dvě rozložení. Jejich základním dělením je podle designu experimentu na testy párové a nepárové.



- Základním testem pro srovnání dvou nezávislých rozložení spojitých čísel je **nepárový two-sample t-test**



- Základním testem pro srovnání dvou závislých rozložení spojitých čísel je **párový two-sample t-test**

Srovnání dvou pokusných variant – obecné schéma zapojených testů I.

Data

X_1 X_2

Nezávislé uspořádání

X_1 X_2

n_1 n_2

$\bar{x}_2$ $\bar{x}_2$

s_1^2 s_2^2

$$H_0: \mu_1 = \mu_2$$

Párové uspořádání

$X_1 - X_2 = D$

n

$\bar{D}$

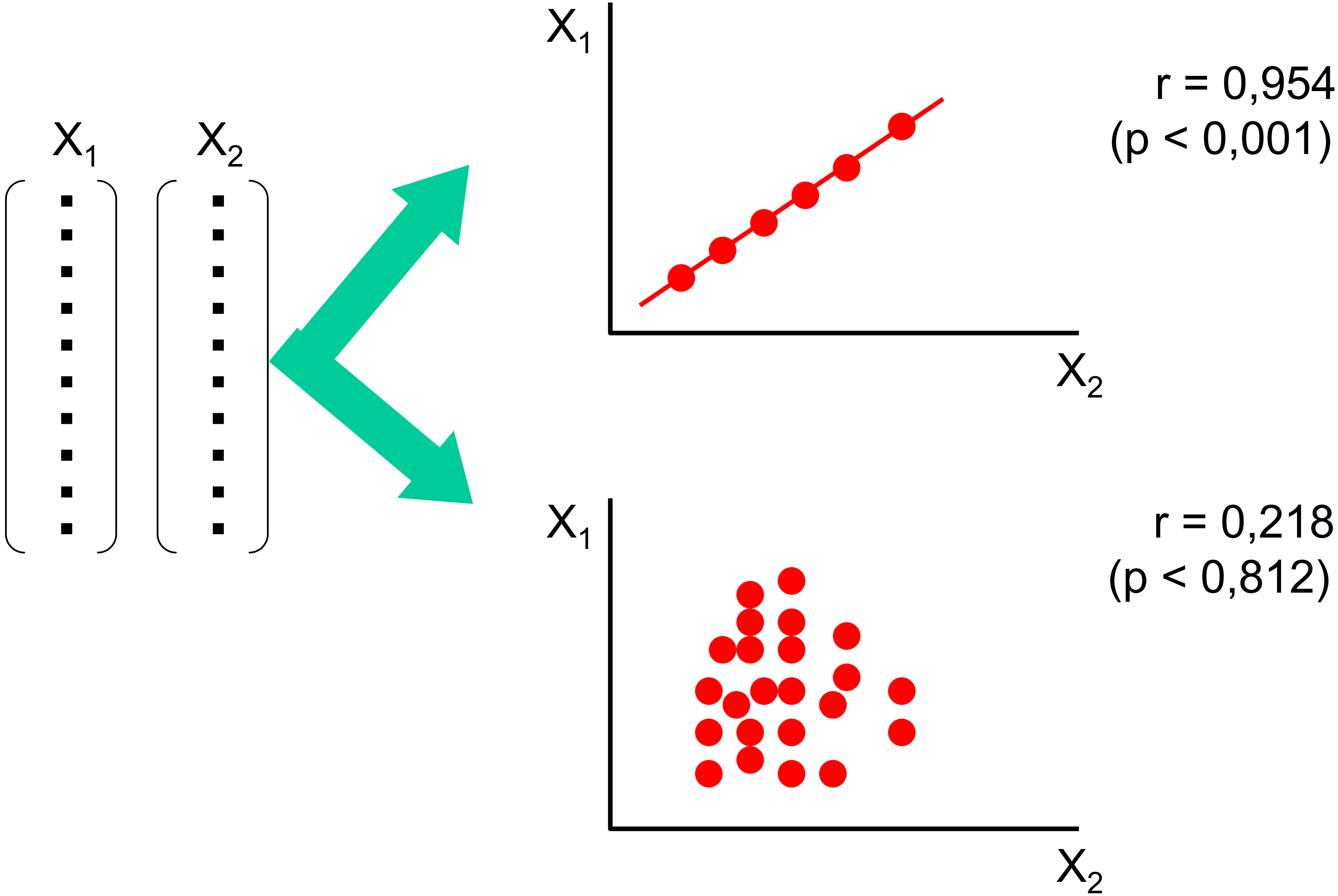
s_D^2

$$H_0: \bar{D} = 0$$

($n = n_2 = n_1$)

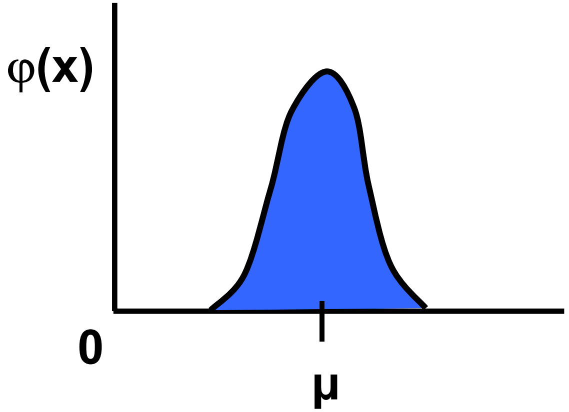
Design uspořádání
zásadně ovlivňuje interpretaci parametrů

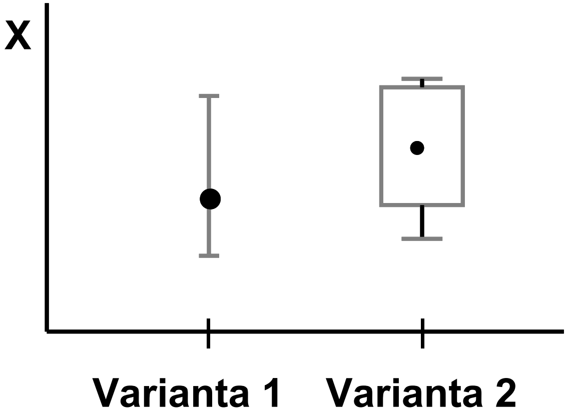
Identifikace párovitosti (Korelace, Kovariance)



Předpoklady nepárového two sample t-testu

- Náhodný výběr subjektů jednotlivých skupin z jejich cílových populací
- Nezávislost obou srovnávaných vzorků
- Přibližně normální rozložení proměnné ve vzorcích, drobné odchylky od normality ovšem nejsou kritické, test je robustní proti drobným odchylkám od tohoto předpokladu, normalita může být testována testy normality
- Rozptyl v obou vzorcích by měl být přibližně shodný (homoscedastic). Tento předpoklad je testován několika možnými testy – Levenův test nebo F-test.
- Vždy je vhodné prohlédnout histogramy proměnné v jednotlivých vzorcích pro okometrické srovnání a ověření předpokladů normality a homogenity rozptylu – nenahradí statistické testy, ale poskytne prvotní představu.





Nepárový two sample t-test – výpočet 1

1. nulová hypotéza: průměry obou skupin jsou shodné, alternativní hypotéza je, že nejsou shodné, two tailed test
2. prohlédnout průběh dat, průměr, medián apod. pro zjištění odchylek od normality a nehomogenita rozptylu, provést F –test

H_0	H_A	Testová statistika
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\max\left(s_1^2; s_2^2\right)}{\min\left(s_1^2; s_2^2\right)}$

F-test pro srovnání dvou výběrových rozptylů

- Používá se pro srovnání rozptylu dvou skupin hodnot, často za účelem ověření homogenity rozptylu těchto skupin dat.

- V případě ověření homogenity je testována hypotéza shody rozptylů (two tailed); v případě shodných rozptylů je vše v pořádku a je možné pokračovat ve výpočtu t-testu, v opačném případě není vhodné test počítat.

Nepárový two sample t-test – výpočet 2

3. Výpočet testové statistiky (stupně volnosti jsou $\upsilon = n_1 + n_2 - 2$):

$$t = \frac{Rozdíl\_průrůmě}{SE(rozdílprůo\ ěrů)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ vážený odhad rozptylu

4. výsledné t srovnáme s tabulární hodnotou t pro dané stupně volnosti a α (obvykle α=0,05)
5. Lze spočítat interval spolehlivosti pro rozdíl průměrů (např. 95%), počet stupňů volnosti a s^2 odpovídají předchozím vzorcům

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,975} SE(\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{0,975} \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Test homogenity rozptylů: Two sample F test

H_0	H_A	F
$\sigma_1^2 \le \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F_1 = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \ge \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F_2 = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$	$F_3 = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\alpha/2}^{(\nu_1;\nu_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2}^{(\nu_2;\nu_1)}$$

Two sample testing – nezávislý t-test

$$H_0 = \overline{X}_1 ? \overline{X}_2$$

1) $\sigma_1^2 = \sigma_2^2 \Rightarrow s_p^2 = \dfrac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2} = \dfrac{ss_1 + ss_2}{\nu_1 + \nu_2}$

2) $s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}} = s_p \cdot \sqrt{\dfrac{n_2 + n_1}{n_1 \cdot n_2}}$

3) $t = \dfrac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$

$$\mu_1 - \mu_2 : \overline{X}_1 - \overline{X}_2 \pm t_{1-\alpha/2}^{(n_1+n_2-2)} \cdot s_{\overline{X}_1 - \overline{X}_2}$$

....

Pokud $\sigma_1^2 \neq \sigma_2^2$: nelze vyjádřit s_p^2

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Two sample t-test - příklad

Průměrná hmotnost ovcí v čase páření byla srovnávána pro kontrolní skupinu a skupinu krmenou zvýšenou dávkou potravy. Kontrolní skupina obsahuje 30 ovcí, skupina se zvýšeným příjmem potravy pak 24 ovcí.

- Vlastní experiment byl prováděn tak, že na začátku máme 54 ovcí (ideálně stejného plemene, stejně staré atd.), které náhodně rozdělíme do dvou skupin (náhodné rozdělování objektů do pokusných skupin je objektem celého specializovaného odvětví statistiky nazývaného randomizace). Poté co experiment proběhne, musíme nejprve ověřit teoretický předpoklad pro využití nepárového t-testu. Pro obě proměnné jsou vykresleny grafy (můžeme též spočítat základní popisnou statistiku), na kterých můžeme posoudit normalitu a homogenitu rozptylu, kromě okometrického pohledu můžeme pro ověření normality použít testy normality, pro ověření homogenity rozptylu pak F-test
- Pokud platí všechny předpoklady Two sample nepárového t-testu, můžeme spočítat testovou charakteristiku, výsledné t je 2,43 s 52 stupni volnosti, podle tabulek je a $t_{0,975\ (52)}$= 2,01, tedy t> $t_{0,975\ (52)}$= a nulovou hypotézu můžeme zamítnout, skutečná pravděpodobnost je pak 0,018. Rozdíl mezi skupinami je 1,59 kg ve prospěch skupiny s lepší výživou.

$$t=\frac{Rozdíl\_průrůmě}{SE(rozdílprůo\ ěrů)}=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \qquad s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \qquad \upsilon=n_1+n_2-2$$

- Pro rozdíl mezi oběma soubory jsou spočítány 95% konfidenční intervaly jako 1,59±2.01*(0,655) kg, což odpovídá rozsahu 0,28 až 2,91 kg. To, že konfidenční interval nezahrnuje 0 je dalším potvrzením, že mezi skupinami je významný rozdíl – jde o další způsob testování významnosti rozdílů mezi skupinami dat – nulovou hypotézu o tom, že rozdíl průměrů dvou skupin dat je roven nějaké hodnotě zamítáme v případě, kdy 95% konfidenční interval rozdílu nezahrnuje tuto hodnotu (v tomto případě 0).

$$(\bar{x}_1-\bar{x}_2)\pm t_{0,975}SE(\bar{x}_1-\bar{x}_2)=(\bar{x}_1-\bar{x}_2)\pm t_{0,975}\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Neparametrické alternativy nepárového t-testu

X1	X2	ALL	Rank ALL	X1 rank	X2 rank
27	25	25	5	6	5
35	29	29	7,5	11	7,5
38	31	31	9	13	9
37	23	23	4	12	4
39	18	18	2	14	2
29	17	17	1	7,5	1
41	32	32	10	15	10
	19	19	3		3
		27	6		
		35	11		
		38	13		
		37	12		
		39	14		
		29	7,5		
		41	15		

Mann Whitney U-test

- Stejně jako řada jiných neparametrických testů počítá i tento test s pořadím dat v souborech namísto s originálními daty. Jde o neparametrickou obdobu nepárového t-testu a z těchto neparametrických testů má nejvyšší sílu testu (95% párového t-testu).
- V případě Mann-Whitney testu jsou nejprve čísla obou souborů sloučena a je vytvořeno jejich pořadí v tomto sloučeném souboru, pak jsou hodnoty vráceny do původních souborů a nadále se pracuje již jen s jejich pořadím.
- Pro oba soubory je tedy vytvořen součet pořadí a menší z obou součtů je porovnán s kritickou hodnotou testu, pokud je tato hodnota menší než kritická hodnota testu, zamítáme nulovou hypotézu shody distribučních funkcí obou skupin.
- Podobným způsobem je počítán i **Wilcoxon rank sum test** (pozor, existuje ještě Wilcoxnův párový test!!!)

Man – Whitney test

***Změna počtu buněk po aplikaci preparátu: A**

$27^{6}, 35^{11}, 38^{13}, 37^{12}, 39^{14}, 29^{7,5}, 41^{15}$

***Kontrolní skupina: B**

$25^{5}, 29^{7,5}, 31^{9}, 23^{4}, 18^{2}, 17^{1}, 32^{10}, 19^{3}$

R_A ……. součet pořadí pro skupinu **A = 78,5**

R_B ……. **= 41,5**

$$U_A = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_A = 7 \cdot 8 + \frac{7 \cdot 8}{2} - R_A = 5{,}5$$

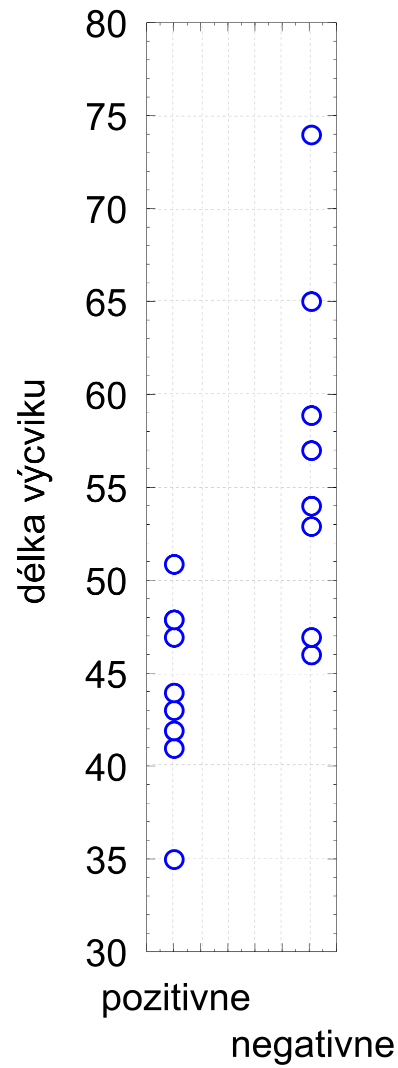
$$U_A + U_B = n_1 \cdot n_2 \Rightarrow U_B = 50{,}5$$

$$\min(U_A; U_B) = 5{,}5 \qquad [n_1 = 7; n_2 = 8]$$

Pokud je **min(U_A; U_B)** menší než kritická hodnota testu, pak zamítáme hypotézu shody distribučních funkcí obou skupin

Mann – Whitney test - příklad

- 17 štěňat bylo trénováno v chození na záchod metodou pozitivního posilování (pochvala, když jde na záchod venku) nebo negativního (trest, když jde na záchod doma). Jako parametr bylo měřeno, za kolik dní je štěně vycvičeno.
- nulová hypotéza je, že není rozdíl v metodách tréninku, tedy, že oběma metodami je štěně vycvičeno za stejnou dobu.
- po srovnání rozložení + malý počet hodnot je vhodné použít neparametrický test
- je vytvořeno pořadí sloučených hodnot
- pořadí hodnot v jednotlivých skupinách dat je sečteno a menší ze součtů je použit pro srovnání s kritickou hodnotou testu
- výsledkem testu je p<α, nulovou hypotézu tedy zamítáme a výsledkem testu je, že pozitivní působení při výcviku štěňat dává lepší výsledky



Párové two sample testy – předpoklady

- Skupiny dat jsou spojeny přes objekt měření, příkladem může být měření parametrů pacienta před léčbou a po léčbě (nemusí jít přímo o stejný objekt, dalším příkladem mohou být např. krysy ze stejné linie).
- Oba soubory musí mít shodný počet hodnot, protože všechna měření v jednom souboru musí být spárována s měřením v druhém souboru. Při vlastním výpočtu se potom počítá se změnou hodnot (diferencí) subjektů v obou souborech.
- Před párovým testem je vhodné ověřit si zda existuje vazba mezi oběma skupinami – vynesení do grafu, korelace.

Existuje několik možných designů experimentu, stručně lze sumarizovat:

1. pokus je párový a jako párový se projeví
2. párové provedení pokusu – párově se neprojeví
   - možná párovost není
   - špatně provedený pokus – malé n, velká variabilita, špatný výběr jedinců
3. čekali jsme nezávislé a jsou
4. čekali jsem nezávislé a nejsou
   - vazba
   - náhoda

Párový two sample t-test

- Tento test nemá žádné předpoklady o rozložení vstupních dat, protože je počítán až na základě jejich diferencí.
- Tyto diference by měly být normálně rozloženy a otázkou v párovém t-testu je, zda se průměrná hodnota diferencí rovná nějakému číslu, typicky jde o srovnání s nulou jako důkaz neexistence změny mezi oběma spárovanými skupinami.
- V podstatě jde o one sample t-test, kde místo rozdílu průměru vzorku a cílové populace je uveden průměr diferencí a srovnávané číslo (0 v případě otázky, zda není rozdíl mezi vzorky).
- Pro srovnání s 0 (testovou statistikou je t rozložení): $t = \frac{\bar{D}}{s}\sqrt{n} \qquad \upsilon = n - 1$
- Někdy je obtížné rozhodnout, zda jde nebo nejde o párové uspořádání, párový test by měl být použit pouze v případě, že můžeme potvrdit vazbu (korelace, vynesení do grafu), jedním z důvodů proč toto ověřovat je fakt, že v případě párového t-testu není nutné brát ohled na variabilitu původních dvou souborů, tento předpoklad však platí pouze v případě vazby mezi proměnnými. Výpočet obou typů testů se vlastně liší v použité s, jednou jde o s diferencí, v druhém případě o složený odhad rozptylu obou souborů.
- Zda je párové uspořádání efektivnější lze určit na základě:
  - Síly vazby
  - Je-li s_D výrazně menší než s_{x1-x2}
- Závislost je možné rozepsat pomocí vzorce: $s_D^2 \cong \sigma_{x_1}^2 + \sigma_{x_2}^2 - 2Cov(x_1; x_2)$
- v případě Cov=0, tedy v případě neexistence vazby pak s_D^2 odpovídá součtu původních rozptylů, tedy přibližně S_{x1-x2}.

Two sample testing: paired design

$$H_0 : \mu_d = 0 \qquad t = \frac{\overline{D}}{s_{\overline{D}}} \cong \frac{\overline{D} - \mu_d}{s_{\overline{D}}}$$

$$\mu_d : \overline{D} \pm t_{1-\alpha/2}^{(n-1)} s_{\overline{D}}$$

Paired **Independent**

$$\underbrace{\sigma_D^2 \;?\approx\; \sigma_{\overline{X}_1 - \overline{X}_2}^2}$$

$$\sigma_{\overline{X}_1 - \overline{X}_2}^2 = \sigma_{\overline{X}_1}^2 + \sigma_{\overline{X}_2}^2 - 2Cov(\overline{X}_1; \overline{X}_2)$$

Mathematically:

$$\sigma_D \sim 2\sigma$$

$$\sigma_D^2 \sim 2\sigma^2$$

1. Evaluate experiment as paired and as independent

$$\sigma_D^2 \sim 2\sigma^2 \qquad s_p^2$$

2. $$2\hat{\sigma}^2 = 2s_p^2 - \frac{2s_p^2 - s_D^2}{(2n-1)}$$

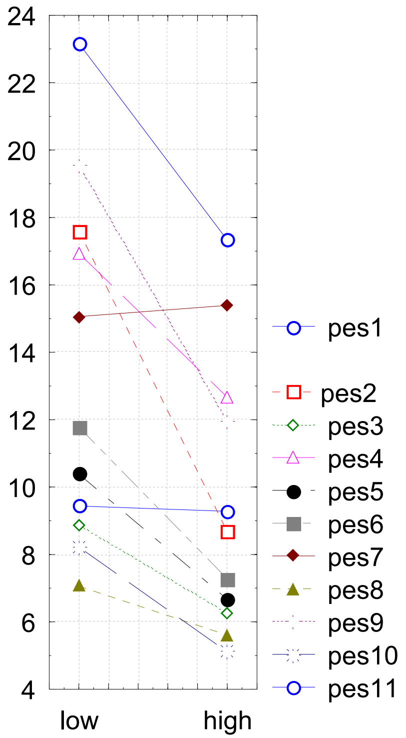
Párový two sample t-test – příklad

Byl prováděn pokus s dietou 11 diabetických psů, každý pes byl vystaven dvěma dietám s odlišným typem sacharidů (snadno vstřebatelné X pozvolna se rozkládající na glukózu), hodnoty krevní glukózy v průběhu jednotlivých diet mají být srovnány pro zjištění vlivu diety na hladinu krevní glukózy. Protože každý pes absolvoval obě diety, jde o párové uspořádání, kdy výsledky hodnoty v obou pokusech jsou spojeny přes pokusné zvíře.

1. Nulová hypotéza zní, že skutečný průměrný rozdíl mezi oběma dietami je 0, alternativní hypotéza zní, že to není 0.
2. Pro každého psa je spočítán rozdíl mezi jeho hladinou glukózy při obou dietách a měly by být ověřeny předpoklady pro one sample t-test – tedy alespoň přibližně normální rozložení.
3. Je spočítána testová charakteristika, výpočet vlastně probíhá jako one-sample t-test, kde je zjišťována významnost průměru diferencí obou souborů jako rozdíl mezi touto hodnotou a nulou (nula je hodnota, kterou by průměrná diference měla nabývat, pokud platí nulová hypotéza). T=4.37 s 10 stupni volnosti, skutečná hodnota p=0,0014 a tedy na hladině p=0,05 můžeme nulovou hypotézu zamítnou

$$t = \frac{rozdíl\_průměru\_vzorku\_a\_populace}{SE(průměru)} = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{\overline{x}-\mu}{s}\sqrt{n}$$

4. Závěrem můžeme říci, že nulová hypotéza neexistence rozdílu mezi oběma dietami byla zamítnuta, což znamená, že high-fibre dieta má významný vliv na snížení hladiny krevní glukózy.



Paired?

A one-tailed t test for the hypotheses H_0: $\mu \geq 0$ and $H_A < 0$

Máme hodnoty hmotnostních změn u lidí, seřazené po užívání drog, které mají za následek ztrátu hmotnosti. Každá změna hmotnosti (v kg) je hmotnost po mínus hmotnost před užitím drogy.

0,2
-0,5
-1,3
-1,6
-0,7
0,4
-0,1
0,0
-0,6
-1,1
-1,2
-0,8

$$n = 12$$

$$\overline{X} = -0,61\ kg$$

$$s^2 = 0,4008\ kg^2$$

$$s_{\bar{x}} = \sqrt{\frac{0,4008\ kg^2}{12}} = 0,18\ kg$$

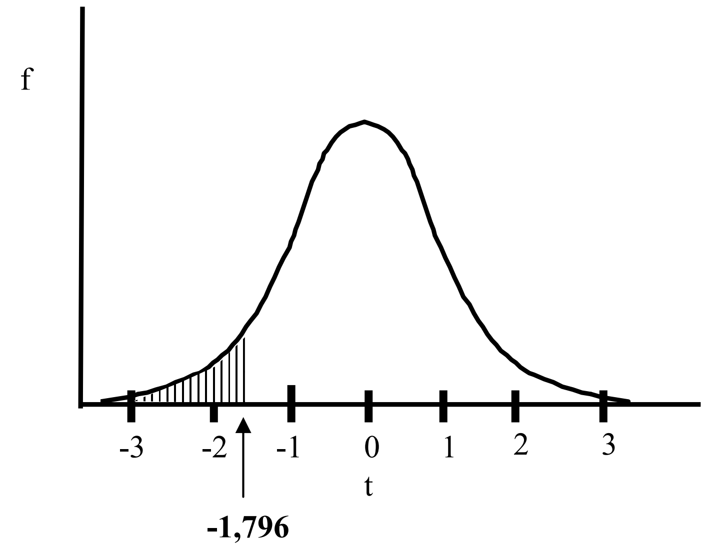
$$t = \frac{\overline{X} - \mu}{s_{\bar{x}}} = \frac{-0,61kg}{0,18kg} = -3,389$$

$$v = n - 1 = 11$$

$$t_{0,05(1),11} = 1,796$$

Když $t \leq -1,796$ **, zamítáme H_0.**

$$0,0025 < P(t \leq -3,389) < 0,005$$



The distribution of t for v=11, showing the critical region (shaded area) for a one-tailed test using α=0,05. (The critical value of t is –1,796.)

Wilcoxon test

- Jsou vytvořeny diference mezi soubory, je vytvořeno jejich pořadí bez ohledu na znaménko a poté je sečteno pořadí kladných a pořadí záporných rozdílů. Menší z těchto dvou hodnot je srovnána s kritickou hodnotou testu a pokud je menší než kritická hodnota testu, pak zamítáme hypotézu shody obou souborů hodnot. Pro test existuje aproximace na normální rozložení, ale pouze pro velká n>25.

$$t = \frac{Menší\_suma\_diferencí - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Před zásahem	Po zásahu	Změna	Absolutní pořadí
6	2	4	10
2,5	3	-0,5	1,5
6,3	5	1,3	6
8,1	9	-0,9	5
1,5	2	-0,5	1,5
3,4	4	-0,6	3
2,5	1	1,5	8
1,11	2	-0,89	4
2,6	4	-1,4	7
1	3	-2	9

Wilcoxonův test – příklad I

člověk	A	B	diference	pořadí
1	142	138	4	4,5
2	140	136	4	4,5
3	144	147	-3	3
4	144	139	5	7
5	142	143	-1	1
6	146	141	5	7
7	149	143	6	9,5
8	150	145	5	7
9	142	136	6	9,5
10	148	146	2	2

A…….parametr krve před podáním léku

B…….parametr krve po podání léku

$\mathbf{W_+}$ …… Σ pořadí kladných rozdílů = 51

$\mathbf{W_-}$ …… = 4

W = min(W_+;W_-) = 4

počet párů = n = 10

Pokud je **W** menší než kritická hodnota testu, pak zamítáme hypotézu shody distribučních funkcí obou skupin.

Wilcoxonův test – příklad II

Byla testována nová dieta pro laboratorní krysy, při pokusu byl zjišťován její vliv na různých liniích krys, bylo proto zvoleno párové uspořádání kdy krysy v obou dietách jsou spojeny přes svoji linii, tj. na začátku byly dvojice krys stejné linie, jedna z nich byla náhodně přiřazena k dietě, druhá z dvojice pak do druhé diety.

1. nulová hypotéza je, že váha krys není ovlivněna použitou dietou, alternativní, že ovlivnění dietou existuje
2. spočítáme diference – tyto diference jsou nenormální a proto je vhodné využít neparametrický test
3. Spočítáme sumu pořadí kladných a záporných diferencí, zde je menší suma záporných diferencí – 31
4. výsledkem výpočtu je p>0,05 a tedy nemáme dostatečné důkazy pro zamítnutí nulové hypotézy, nelze říci, že by nová dieta byla efektivnější než stará
5. pro doplnění výsledků je vhodné zjistit také skutečnou velikost rozdílu hmotností ve skupinách, např. ve formě mediánu

Znaménkový test – příklady I

Párově uspořádaný experiment pro nominální data

I. Dva preparáty, každý na ½ listu

- sledovaná veličina: počet skvrn (hodnoceno pouze jako rozdíl)

	Počet skvrn									
A	V	V	M	V	V	M	M	V	V	V
B	M	M	V	M	M	V	V	M	M	M

V – větší; M – menší

n = 10 listů s rozdílnými výsledky

jev

A je větší: + $n_+ = 7$

B je menší: - $n_- = 3$

$\min(n_+; n_-) = 3$

II. dvě protilátky z různých zdrojů (A;B)

– aplikované na vzorek s antigenem

n = 10

A	+	+	-	+	-	+	-	+	+	-
B	-	-	+	-	+	+	-	-	+	-

n – nenulových rozdílů: 6

A: $n_+ = 4$

A: $n_- = 2$

$\min(n_+; n_-) = 2$

Znaménkový test – příklady II

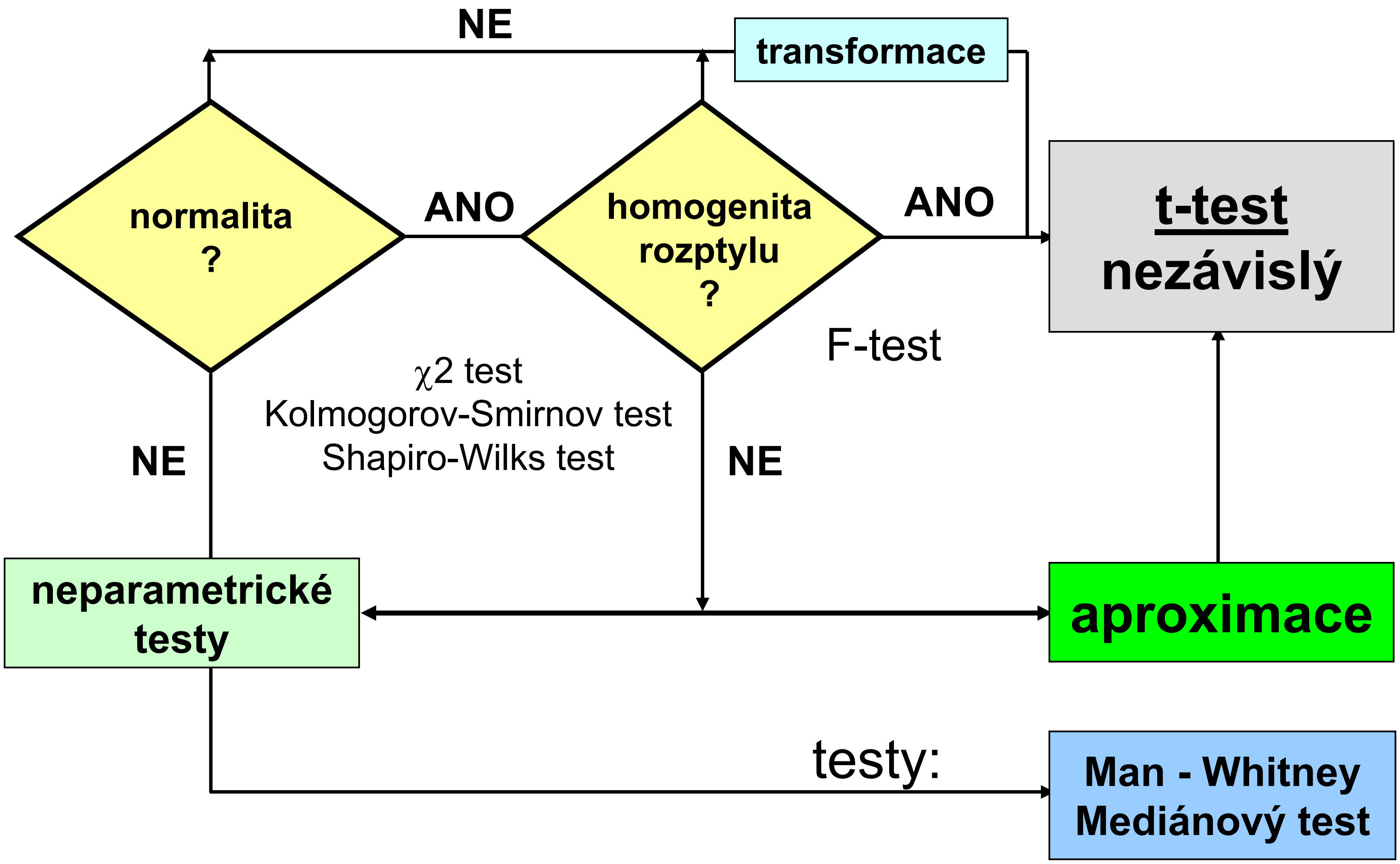
- Na konferenci veterinářů bylo předneseno,že průměrný čas konzultace je 12 minut. Následovala debata, zda je lepší použít medián nebo průměr. Jeden z nich se rozhodl ověřit teorii, že průměrná konzultace trvá 12 minut na vlastní praxi a zaznamenal si trvání svých 43 konzultací. K otestování hypotézy, že podíl konzultací kratších a delších než 12 minut použil znaménkový test.

Délka konzultace	Počet
<12	22
12	6
>12	15
Celkem	43

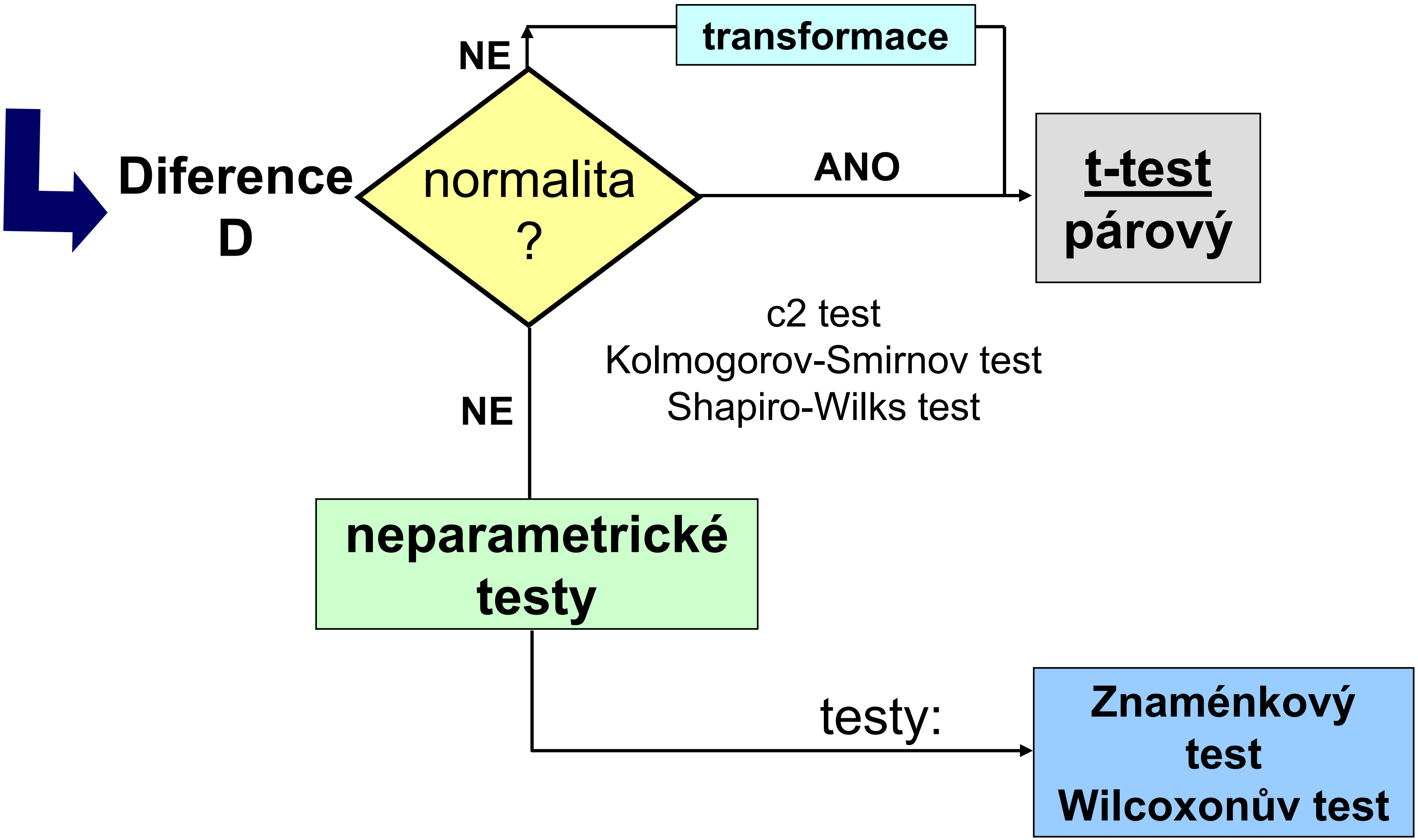
Další výpočet probíhá obdobně jako v případě klasického znaménkového testu na diferencích dvou skupin dat.

Srovnání dvou pokusných zásahů - obecné schéma zapojených testů III

Nezávislé uspořádání



Párové uspořádání



Principy statistického testování lze využít pro různé typy dat

Testování – typ dat

- Spojitá čísla
  - T test, Mann-Whitney test, Wilcoxon test, Znaménkový test atd.

- Binární data?
- Kategoriální data?
  - Výše zmíněné testy nelze použít
  - Základní přístupy testování lze ovšem použít i na tato data
    - Nulová a alternativní hypotéza
    - One sample a two sample testy

- Analýzy na binomickém rozložení
- Analýzy na Poissonově rozložení
- Analýza kontingenčních tabulek