VII. DVOJ VÝBEROVÉ TESTY

Anotace

Jedným z najčastejších úloh štatistickej analýzy dát je porovnanie spojitých dát vo dvoch skupinách pacientov. Na výber je celá škála testov, výber konkrétneho testu sa potom odvíja od toho, či ide o porovnanie párové alebo nepárové a či je vhodné použiť test parametrický (má predpoklady o rozložení dát) alebo neparametrický (nemá predpoklady o rozložení dát, ale má nižšiu vypovedaciu silu).

Najznámejšími testami z tejto skupiny sú tzv. t-testy používané na porovnanie priemerov dvoch skupín hodnôt

Nepárový vs. párový design

NEPÁROVÝ DESIGN

- Skupiny porovnávaných dát sú na sebe úplne nezávislé (tiež nezávislý, independent design), napr. Ľudia z rôznych zemí, nezávislé skupiny pacientov s odlišnou liečbou atď.
- Pri výpočte je nevyhnutné brať v úvahu charakteristiky oboch skupín dát

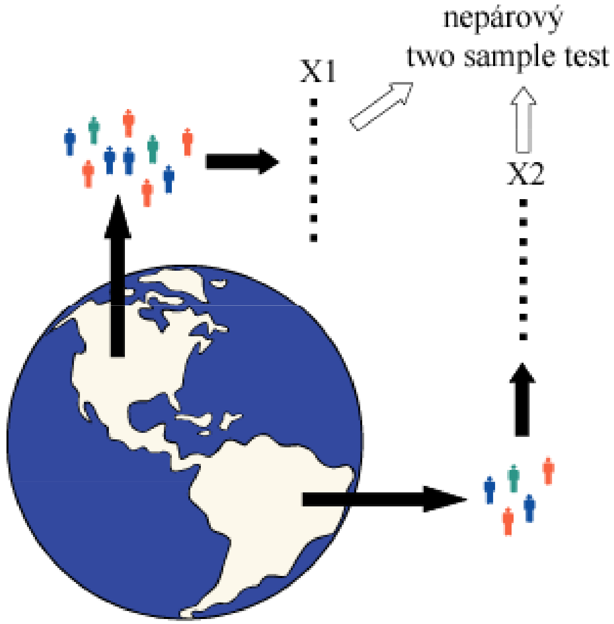
PÁROVÝ DESIGN

- Medzi objektami v porovnávaných skupinách existuje väzba, daná napr. človekom pred a po operácii, reakcie rovnakého kmeňa krýs atď.
- Väzba môže byť buď priamo daná alebo len predpokladaná (v tom prípade je nutné ju overiť)
- Test je v podstate uskutočňovaný na diferenciach skupín, nie na ich pôvodných dátach

Typ srovnání	Parametrický test	Neparametrický test
2 skupiny dat nepárově:	Nepárový t-test	Mann Whitney test
2 skupiny dat párově:	Párový t-test	Wilcoxon test, znaménkový test
Více skupin nepárově:	ANOVA	Kruskal- Wallis test
Korelace:	Pearsonův koeficient	Spearmanův koeficient

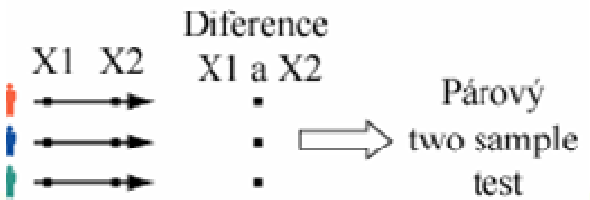
Dvojvýberové testy: párové a nepárové I

Pri použití two sample testov porovnávame spolu dve rozloženia. Ich základným delením je podľa designu experimentu na testy párové a nepárové.



Základným testom na porovnávanie dvoch nezávislých rozložení spojitých čísel je nepárový two-sample t-test

Základným testom na porovnanie dvoch závislých rozložení spojitých čísel je párový two-sample t-test



Dvojvýberové testy: párové a nepárové II

Dáta

X_1 X_2

Nezávislé usporiadanie

X_1 X_2

$H_0 : \mu_1 = \mu_2$

n_1 n_2

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

s_1^2 s_2^2

Párové usporiadanie

$X_1 - X_2 = D$

n

$\bar{D}$

s_D^2

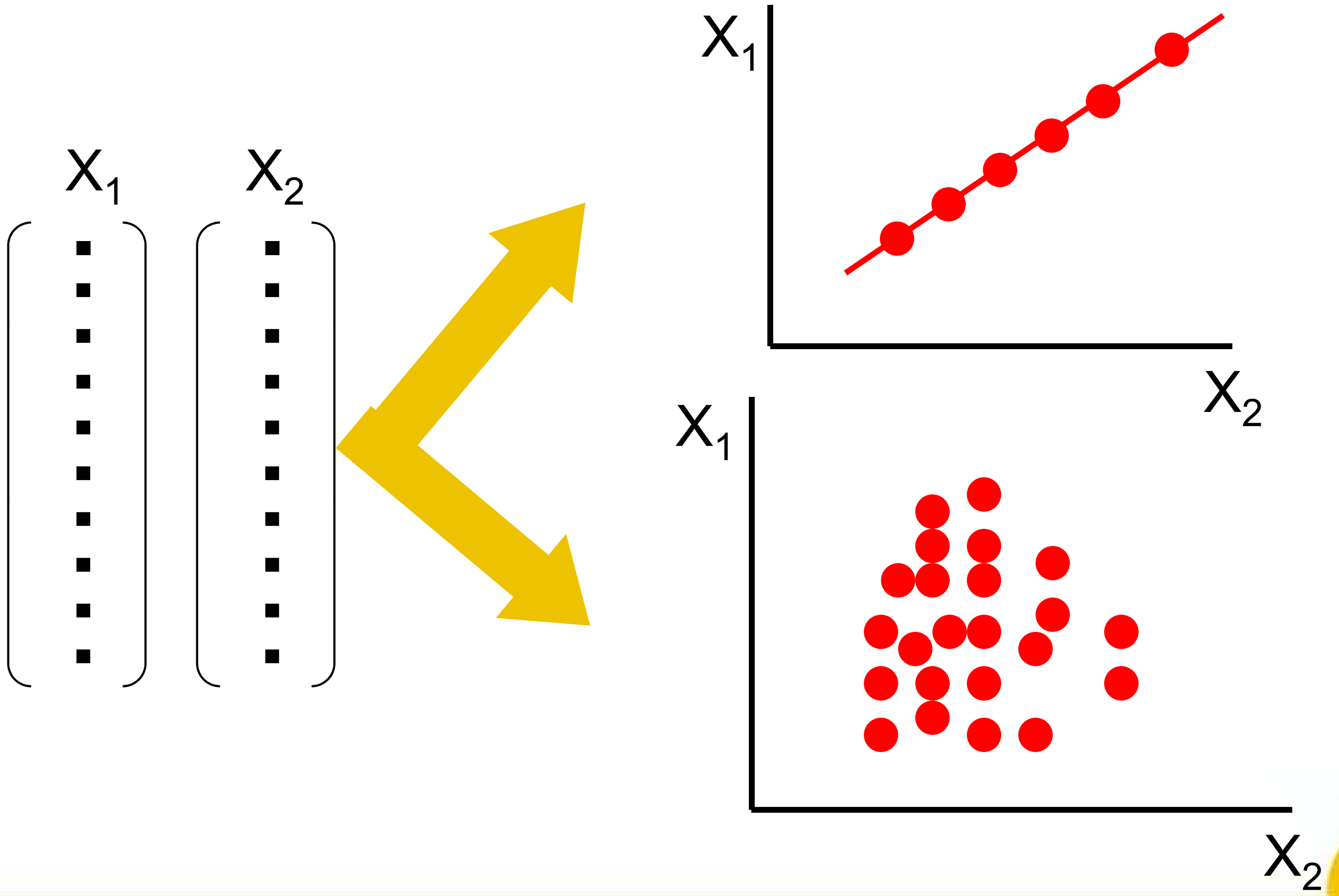
$H_0 : \bar{D} = 0$

$(n = n_2 = n_1)$

Design usporiadania
zásadne ovplyvňuje interpretáciu
parametrov

Dvojvýberové testy: párové a nepárové III

Identifikácia párovitosti (Korelácia, Kovariancia)



r = 0,954
(p < 0,001)

r = 0,218
(p < 0,812)

Predpoklady nepárového dvojvýberového t-testu

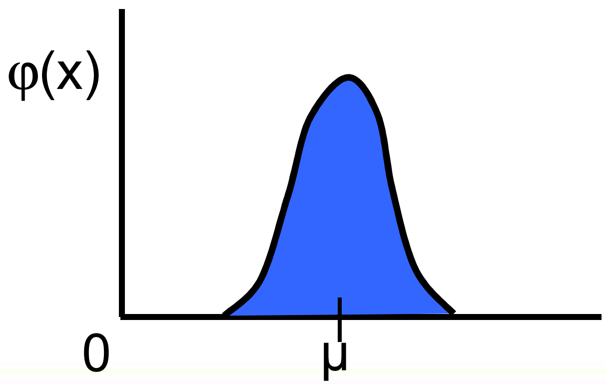
Náhodný výber subjektov jednotlivých skupín z ich cieľových populácií

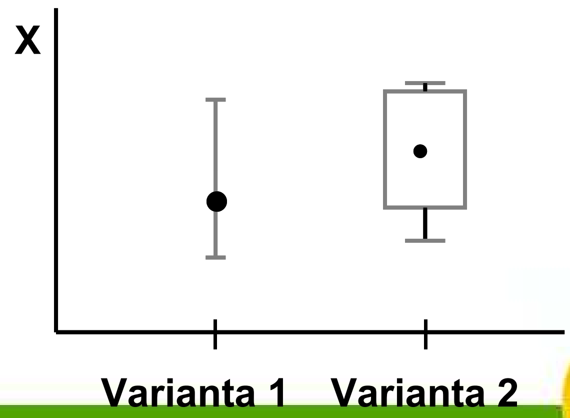
Nezávislosť oboch porovnávaných vzoriek

Približne **normálne rozloženie** premennej vo vzorkách, drobné odchýlky od normality však nie sú kritické, test je robustný proti drobným odchýlkam od tohto predpokladu, normalita môže byť testovaná testami normality

Rozptyl v oboch vzorkách by mal byť približne **zhodný** (homoscedastic). Tento predpoklad je testovaný niekoľkými možnými testami – Levenov test alebo F-test.

Vždy je vhodné prezrieť si histogramy premennej v jednotlivých vzorkách pre okometrické porovnanie a overenie predpokladov normality a homogenity rozptylu – nenahradí štatistické testy, ale poskytne prvotnú predstavu.





Nepárový dvojvýberový t-test – výpočet I

1. nulová hypotéza: priemery oboch skupín sú zhodné, alternatívna hypotéza je, že nie sú zhodné, two tailed test
2. prezrieť priebeh dát, priemer, medián apod. pre zistenie odchýlok od normality a nehomogenitu rozptylu, uskutočniť F –test

H_0	H_A	Testová štatistika
$\sigma_1^2 \le \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \ge \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$	$F = \frac{\max(s_1^2; s_2^2)}{\min(s_1^2; s_2^2)}$

F-test na porovnanie dvoch výberových rozptylov

Používa sa na porovnanie rozptylu dvoch skupín hodnôt, často za účelom overenia homogenity rozptylu týchto skupín dát.

V prípade overenia homogenity je testovaná hypotéza zhody rozptylov (two tailed); v prípade zhodných rozptylov je všetko v poriadku a je možné pokračovať vo výpočte t-testu, v opačnom prípade nie je vhodné test počítať.

Nepárový dvojvýberový t-test – výpočet II

3. Výpočet testovej štatistiky (stupne voľnosti sú $\upsilon = n_1 + n_2 - 2$):

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

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

4. výsledné t porovnávame s tabuľkovou hodnotou t pre dané stupne voľnosti a α (obvykle α=0,05)
5. Je možné spočítať interval spoľahlivosti pre rozdiel priemerov (napr. 95%), počet stupňov voľnosti a s^2 zodpovedajú predchádzajúcim vzorcom

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

Dvojvýberový t-test - príklad

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 alternatívy nepárového t-testu

Mann Whitney U-test

Rovnako ako rada iných neparametrických testov počíta i tento test s poradím dát v súboroch namiesto s originálnymi dátami. Ide o neparametrickú obdobu nepárového t-testu a z týchto neparametrických testov má najvyššiu silu testu (95% párového 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		

V prípade Mann-Whitney testu sú najskôr čísla oboch súborov zlúčené a je vytvorené ich poradie v tomto zlúčenom súbore, potom sú hodnoty vrátené do pôvodných súborov a naďalej sa pracuje už len s ich poradím.

Pre obidva súbory je teda vytvorený súčet poradí a menší z oboch súčtov je porovnaný s kritickou hodnotou testu, pokiaľ je táto hodnota menšia než kritická hodnota testu, zamietame nulovú hypotézu zhody distribučných funkcií oboch skupín.

Podobným spôsobom je počítaný i

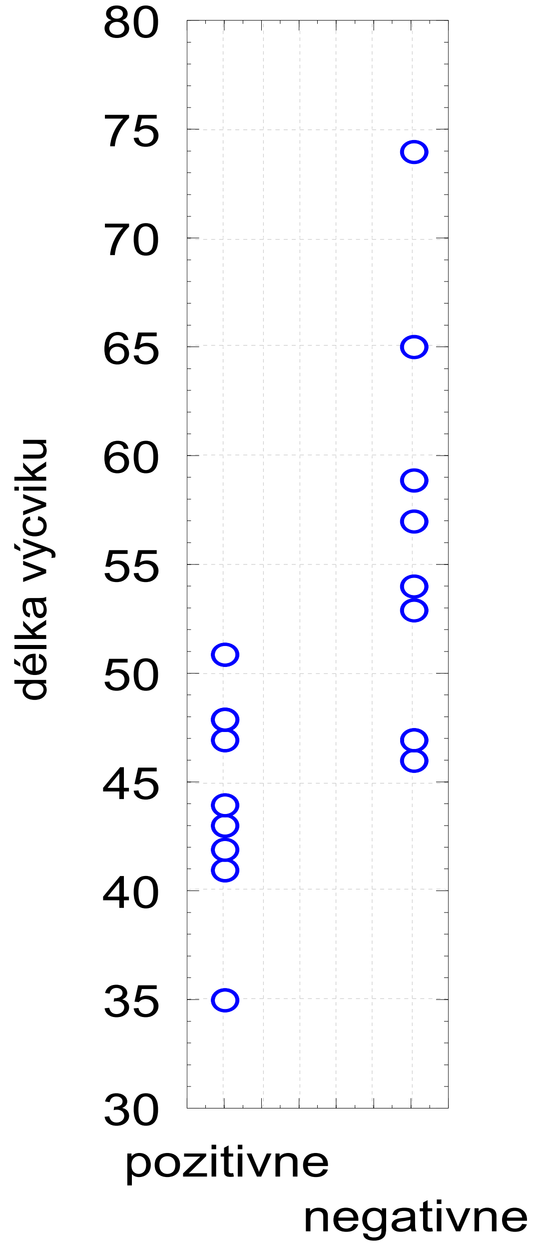
Wilcoxon rank sum test

(pozor, existuje ešte

Wilcoxnov párový test!!!)

Mann – Whitney U test - príklad

- 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 dvojvýberové testy – predpoklady

- Skupiny dát sú spojené cez objekt merania, príkladom môže byť meranie parametrov pacienta pred liečbou a po liečbe (nemusí ísť priamo o rovnaký objekt, ďalším príkladom môžu byť napr. krysy z rovnakej línie).
- Obidva súbory musia mať rovnaký počet hodnôt, pretože všetky merania v jednom súbore musia byť spárované s meraním v druhom súbore. Pri vlastnom výpočte sa potom počíta so zmenou hodnôt (diferencií) subjektov v obidvoch súboroch.
- Pred párovým testom je vhodné overiť si či existuje väzba medzi obomi skupinami – vynesenie do grafu, korelácia.

Existuje niekoľko možných designov experimentu, stručne je možné sumarizovať:

1. pokus je párový a ako párový sa prejaví
2. párové prevednie pokusu – párovo sa neprejaví
    - možná párovosť nie je
    - nesprávne uskutočnený pokus – malé n, veľká variabilita, nesprávny výber jedincov
3. čakali sme nezávislé a sú
4. čakali sem nezávislé a nie sú
    - väzba
    - náhoda

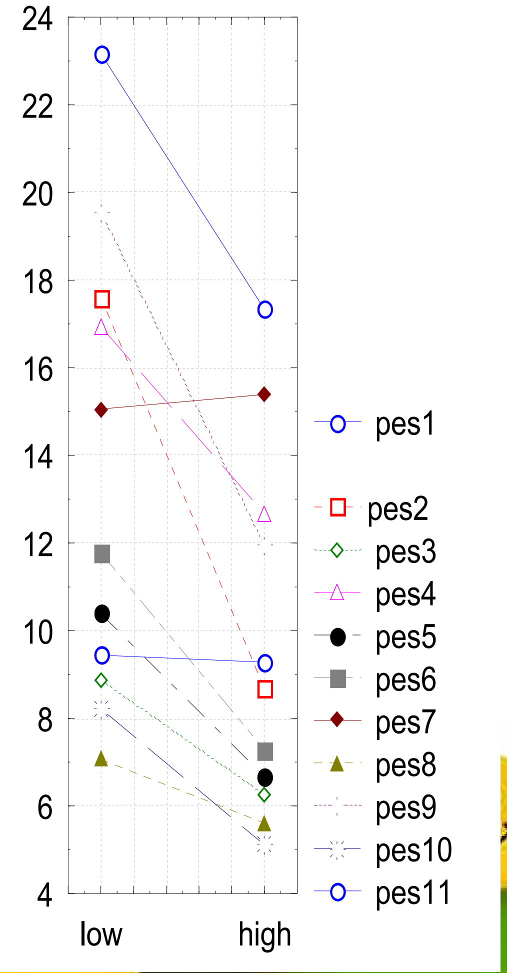
Párový dvojvýberový t-test

- Tento test nemá žiadne predpoklady o rozložení vstupných dát, pretože je počítaný až na základe ich diferencií.
- Tieto diferencie by mali byť normálne rozložené a otázkou v párovom t-teste je, či sa priemerná hodnota diferencií rovná nejakému číslu, typicky ide o porovnanie s nulou ako dôkaz neexistencie zmeny medzi obidvomi spárovanými skupinami.
- V podstate ide o one sample t-test, kde namiesto rozdielu priemeru vzorku a cieľovej populácie je uvedený priemer diferencií a porovnávané číslo (0 v prípade otázky, či nie je rozdiel medzi vzorkami).
- Na porovnanie s 0 (testovou štatistikou je t rozloženie): $t = \frac{\bar{D}}{s}\sqrt{n}$ $\upsilon = n - 1$
- Niekedy je obtiažne rozhodnúť, či ide alebo nejde o párové usporiadanie, párový test by mal byť použitý len v prípade, že môžeme potvrdiť väzbu (korelácie, vynesenie do grafu), jedným z dôvodov prečo toto overovať je fakt, že v prípade párového t-testu nie je nutné brať ohľad na variabilitu pôvodných dvoch súborov, tento predpoklad však platí len v prípade väzby medzi premennými. Výpočet obidvoch typov testov sa vlastne líši v použitej s, raz ide o s diferencií, v druhom prípade o zložený odhad rozptylov obidvoch súborov.
- Či je párové usporiadanie efektívnejšie je možné určiť na základe:
  - Sily väzby
  - Ak je s_D výrazne menšie než s_{x1-x2}
- Závislosť je možné rozpísať pomocou vzorca: $s_D^2 \cong \sigma_{x_1}^2 + \sigma_{x_2}^2 - 2Cov(x_1; x_2)$

v prípade Cov=0, teda v prípade neexistencie väzby potom s_D^2 odpovedá súčtu pôvodných rozptylov, teda približne S_{x1-x2}.

Párový dvojvýberový t-test – príklad

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.

Neparametrická obdoba párového t-testu

Wilcoxon test

- Sú vytvorené diferencie medzi súbormi, je vytvorené ich poradie bez ohľadu na znamienko a potom je sčítané poradie kladných a poradie záporných rozdielov. Menší z týchto dvoch hodnôt je porovnaná s kritickou hodnotou testu a pokiaľ je menšia než kritická hodnota testu, potom zamietame hypotézu zhody oboch súborov hodnôt. Pre test existuje aproximácia na normálne rozloženie, ale len pre veľké n>25.

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

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

Wilcoxonov 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…….parameter krvi pred podaním lieku

B…….parameter krvi po podaní lieku

W_+ …… © poradie kladných rozdielov = 51

W_- …… = 4

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

počet párů = n = 10

Pokiaľ je W menšie než kritická hodnota testu,
potom zamietame hypotézu zhody distribučných funkcií obidvoch skupín.

Wilcoxonov test – príklad 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

Znamienkový test – príklad I

Párovo usporiadaný experiment pre nominálne dáta

I. Dva preparáty, každý na ½ listu

- sledovaná veličina: počet škvŕn (hodnotené len ako rozdiel)

	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 – väčší; M – menší

n = 10 listov s rozdielnymi výsledkami

jav → A je väčší: + $n_+ = 7$

jav → B je menší: - $n_- = 3$

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

II. dve protilátky z rôznych zdrojov (A;B)

– aplikované na vzorku s antigénom

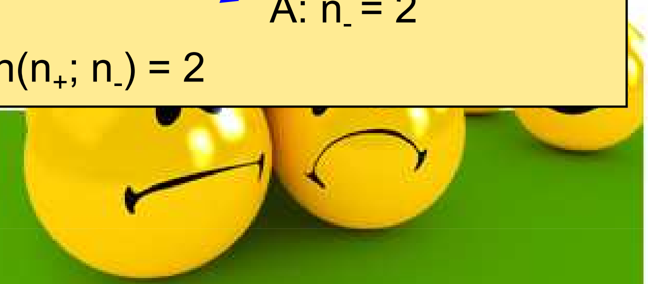
n = 10

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

n – nenulových rozdielov: 6 → A: $n_+ = 4$

n – nenulových rozdielov: 6 → A: $n_- = 2$

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



Znamienkový test – príklady II

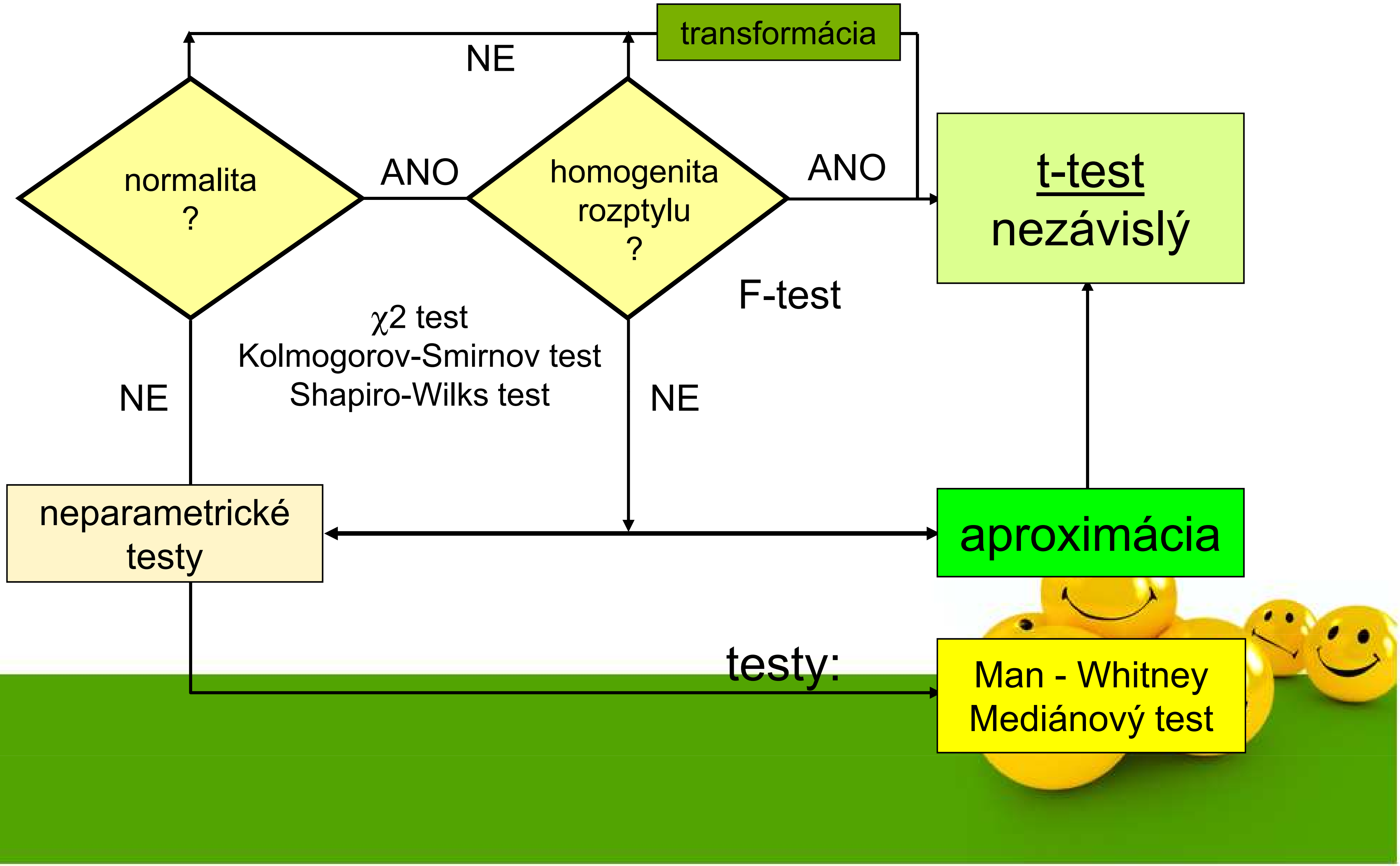
- Na konferenci veterinářů bylo předneseno,že průměrný čas konzultace je 12 minut. Následovala debata, zda je lepší použít medián nebo průměr. Jeden z nich se rozhodl ověřit teorii, že průměrná konzultace trvá 12 minut na vlastní praxi a zaznamenal si trvání svých 43 konzultací. K otestování hypotézy,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

Dvojvýberové testy: schéma analýzy

Nezávislé usporiadanie



Dvojvýberové testy: schéma analýzy

Párové usporiadanie