ONE SAMPLE T TEST

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. Lze testovat shodu průměru a rozptylu.

Pro testování hypotéz o průměru je použito převedení na Studentovo (t) rozložení, test se nazývá **one sample t-test.**

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

$\bar{x}$ průměr testovaného souboru, s směrodatná odchylka testovaného souboru, n počet měření v testovaném souboru, μ průměr cílové populace, jednotlivé nulové hypotézy, jejich alternativy a interval spolehlivosti jsou zobrazeny v následující tabulce.

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	$\lvert t \rvert > t_{1-\alpha/2}^{(n-1)}$

Pro testování hypotéz o rozptylu je využito převedení na χ^2 (Fischer-Snedecor) rozložení, lze hovořit o one sample F testu.

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

s^2 rozptyl testovaného souboru , n počet měření v testovaném souboru, σ^2 rozptyl cílové populace, jednotlivé nulové hypotézy, jejich alternativy a interval spolehlivosti jsou zobrazeny v následující tabulce.

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	$\chi^2 > \chi^{2(n-1)}_{1-\alpha/2}$ nebo $\chi^2 < \chi^{2(n-1)}_{\alpha/2}$

PŘÍKLADY

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?

Postup řešení

H_0: $\bar{x} = \mu$ tedy two tailed test

$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{3,5 - 2,5}{1}\sqrt{25} = 5$ ➔ $t^{24}_{0,975} = 2,064$ ➔ $t > t^{24}_{1-\alpha/2}$ ➔ H_0 zamítnuta při $\alpha \leq 0,05$

2. otázka – jakou minimální odchylku $\bar{x}$ od jiné hodnoty bychom zachytili při daných hodnotách?

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

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 ?

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

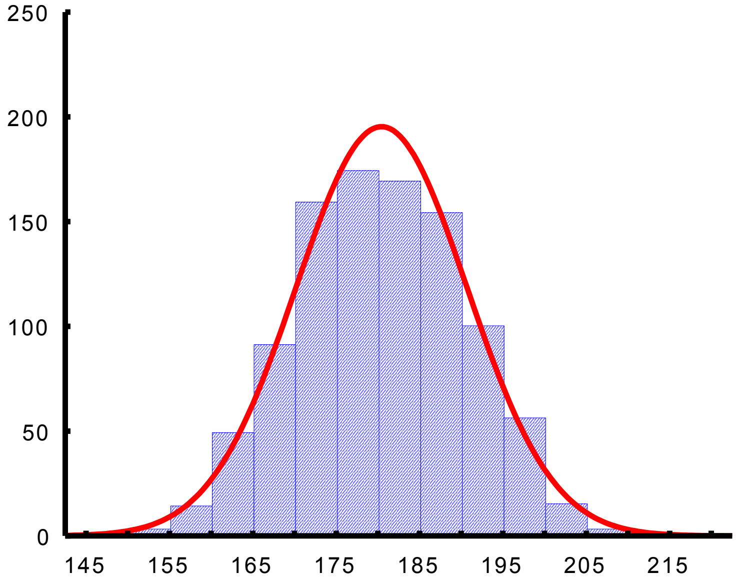
Koncentrace antibiotika v cílovém orgánu

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

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

Testy normality

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 χ^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

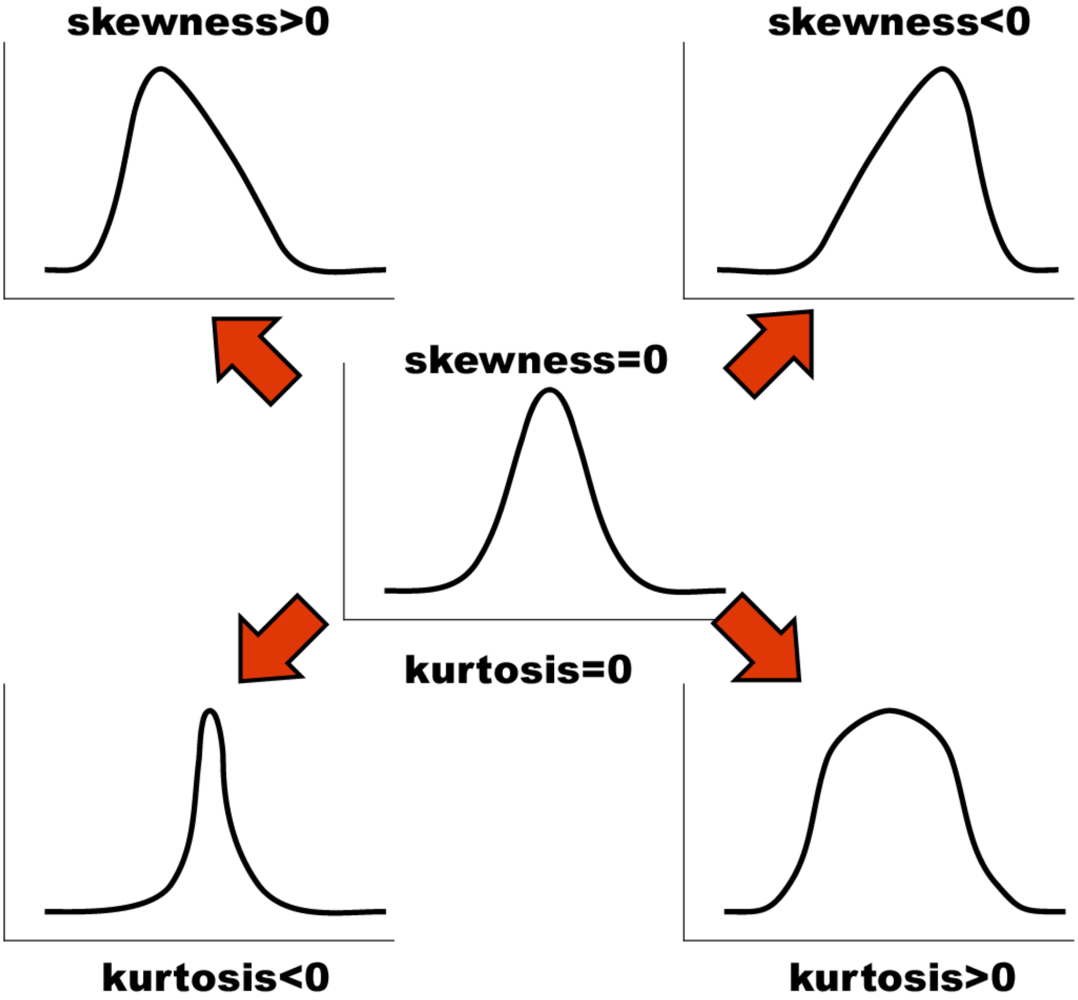
Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložení. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. 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

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



KORELACE

Pearsonův korelační koeficient – lze jej použít pouze pro normálně rozložená data bez odlehlých hodnot.

$$r = \frac{Cov(x,y)}{S_x S_y} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right]\left[\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2\right]}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Korelační koeficient je možné testovat, zda se významně liší od nuly (tedy žádná korelace). Testovací statistikou je Studentovo rozložení.

$t = \left[\frac{r}{\sqrt{1-r^2}}\right]\sqrt{n-2} = r\sqrt{\frac{n-2}{1-r^2}}$ a počet stupňů volnosti $\nu = n-2$

Pro korelační koeficient je možné spočítat i konfidenční intervaly, protože rozložení r není normální je nutné nejprve provést transformaci na normální rozložení $z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$, spočítat konfidenční interval – pro přibližně 95% jde o $z_1 = z - \frac{1,96}{\sqrt{n-3}}$ a $z_2 = z + \frac{1,96}{\sqrt{n-3}}$, který je zpětně převeden na hodnoty r pomocí exponenciální funkce $\frac{e^{2z_1}-1}{e^{2z_1}+1}$ *až* $\frac{e^{2z_2}-1}{e^{2z_2}+1}$

Další možností je porovnání dvou korelačních koeficientů – např. pro srovnání zda závislost nějakých parametrů u dvou skupin pacientů je shodná:

1. je použito konverze na normální rozložení pro oba koeficienty: $z_i = \frac{1}{2}\ln\left(\frac{1+r_i}{1-r_i}\right)$
2. je spočítána testovací charakteristika (H_0: $\rho_1 = \rho_2$; $\alpha = 0,05$): $Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$
3. výsledná hodnota je srovnána s tabulární hodnotou Z

PŘÍKLADY

1. korelační koeficient

Byl vyvinut nový způsob měření metabolismu kostí pomocí kostní fosfatázy (wBAP), otázka je, zda toto měření je korelováno s jiným, již ověřeným, ukazatelem metabolismu kostí propeptidem kolagenu typu I (PICP). Testování bylo provedeno na náhodném vzorku 46 koní.

1. prvním krokem je jako vždy specifikování nulové hypotézy – v našem případě je H_0, že mezi wBAP PICP není žádný lineární vztah a tedy ρ=0, alternativní hypotéza je ρ≠0.
2. krokem je ověření předpokladů parametrické Pearsonovy korelace – pro obě proměnné vytvoříme histogramy – ověření normality a dále vytvoříme xy graf vynášející vztah obou proměnných – ověření linearity vztahu, v případě nevhodného průběhu dat máme dvě možnosti – neparametrickou korelaci nebo transformaci proměnných.
3. Spočítáme korelační koeficient $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$, kde hodnoty x přísluší wBAP a hodnoty y PICP, výsledný koeficient je r=0,785
4. pro ověření nulové hypotézy dosadíme do vzorce $t = \left[\frac{r}{\sqrt{1-r^2}}\right]\sqrt{n-2}$ =8,41, s **46-2=44** stupni volnosti, který se používá pro testování hypotézy o významnosti korelačního koeficientu. Při srovnání výsledku s tabulární hodnotou $t_{0,975}^{44}$=2,02, kdy t> $t_{0,975}^{44}$ máme silný důkaz pro zamítnutí nulové hypotézy a můžeme říci, že obě proměnné spolu významně korelují.
5. Pro korelační koeficient zjistíme jeho 95% konfidenční intervaly. Konfidenční interval nám vlastně říká, v jakých hranicích by se tento korelační koeficient pohyboval u 95% náhodných vzorků o stejné velikosti (46 koní) odebraných z cílové populace. Při výpočtu je nejprve korelační koeficient převeden na normální rozložení $z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$(protože rozložení r není normální), jsou spočítány konfidenční intervaly $z_1 = z - \frac{1,96}{\sqrt{n-3}}$ a $z_2 = z + \frac{1,96}{\sqrt{n-3}}$ (symetrický) a převedeny zpět na rozložení r (asymetrický konfidenční interval) $\frac{e^{2z_1}-1}{e^{2z_1}+1}\ až\ \frac{e^{2z_2}-1}{e^{2z_2}+1}$, výsledkem je pak 95% konfiden4n9 interval pro r **0,64 až 0,88**.
6. závěrem lze říci, že korelační koeficient je významně odlišný od nuly a i spodní hranice jeho konfidenčního intervalu poukazuje na silnou lineární závislost, můžeme tedy prohlásit, že nové měření dobře odráží metabolismus kostní tkáně.

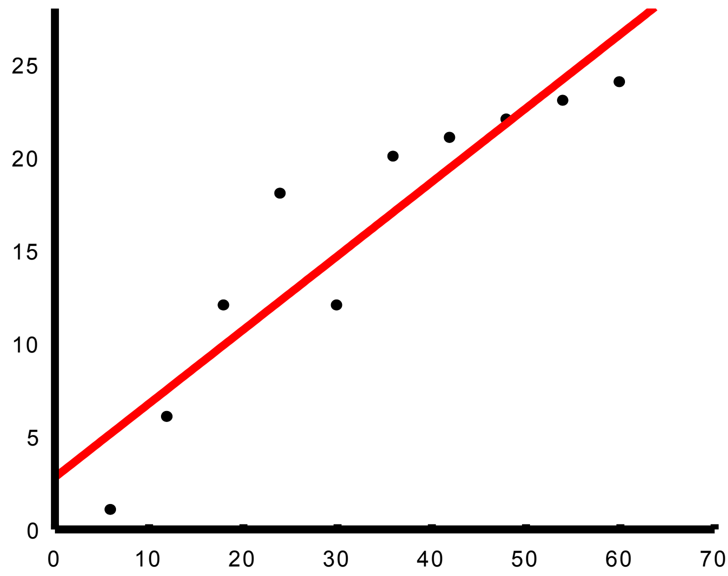
2. Srovnání korelačních koeficientů

Ve dvou skupinách pacientů byla zjišťována závislost krevního tlaku a koncentrace kyslíkových radikálů. V první skupině byla zjištěna r=0,682 při 1258 pacientech, ve skupině druhé pak r=0,402 a 462 pacientů.

1. Převedeme korelační koeficienty na normální rozložení: $z_i = \frac{1}{2}\ln\left(\frac{1+r_i}{1-r_i}\right)$, z_1=0,833, z_2=0,426
2. spočítáme testovou statistiku $Z = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = 7{,}461$
3. porovnáme tabulární hodnotou $Z_{0,975}$=1,96
4. protože Z> $Z_{0,975}$ zamítáme shodu obou korelačních koeficientů – závislost obou parametrů se ve skupinách pacientů liší

SPEARMAN RANK CORRELATION COEFFICIENT

Patří mezi neparametrické korelační koeficienty, tj. není závislý na normálním rozložení a lineární závislosti proměnných. Hodnoty proměnných jsou převedeny na jejich pořadí a výpočet je proveden na základě rozdílu pořadí spárovaných hodnot v proměnných.



	A	B	A ranks	B ranks	Rozdíl pořadí
1	6	1	1	1	0
2	12	6	2	2	0
3	18	12	3	4	-1
4	24	18	4	5	-1
5	30	11	5	3	2
6	36	20	6	6	0
7	42	21	7	7	0
8	48	22	8	8	0
9	54	23	9	9	0
10	60	24	10	10	0

$$r_S = 1 - \frac{6\sum (R_1 - R_2)^2}{n(n^2 - 1)}$$

$$r_S = 1 - \frac{6 \times 6}{10(10^2 - 1)} = 0,96$$

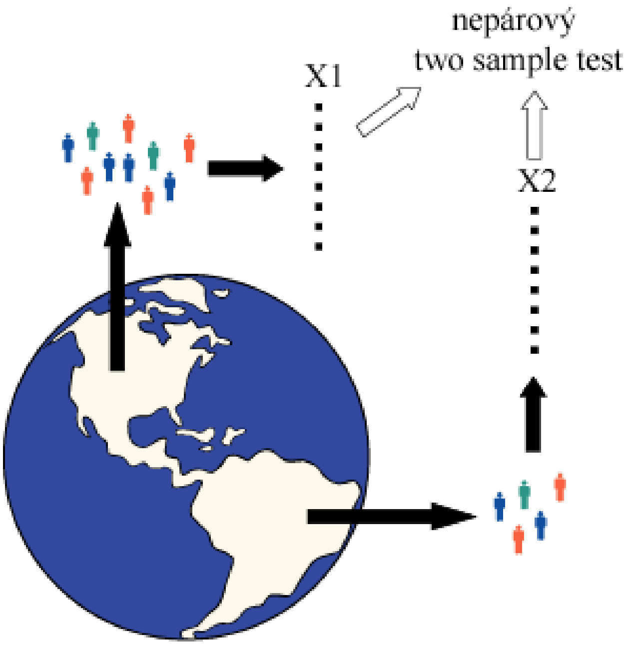
Kromě Spearmanova korelačního koeficientu existují i další neparametrické korelační koeficienty jako je Kendel τ nebo gama. Výše uvedený princip využití pořadí hodnot se používá obecně v neparametrických testech, například v obdobách two- sample t testů.

TWO SAMPLE TESTY

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

Nepárové testy

Obě skupiny hodnot jsou spojeny pouze měřeným parametrem, není vazba mezi subjekty v obou skupinách. Obě skupiny dat nemusí mít stejný počet hodnot.



Pro srovnání průměrů dvou nezávislých souborů dat se využívá two sample t-test. K jeho předpokladům patří:

- 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, budou zmíněny dále.
- Rozptyl v obou vzorcích by měl být přibližně shodný (homoscedastic). Tento předpoklad je testován několika možnými testy – Levenův test nebo níže zmíněný 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.

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

F-test pro srovnání dvou výběrových rozptylů

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

H_0	H_A	Testová statistika	Interval spolehlivosti
$\sigma_1^2 \le \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$	$F > F_{1-\alpha}^{(\upsilon_1 = n_1 - 1; \upsilon_2 = n_2 - 1)}$
$\sigma_1^2 \ge \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$	$F < F_{\alpha}^{(\upsilon_1 = n_1 - 1; \upsilon_2 = n_2 - 1)}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \ne \sigma_2^2$	$F = \frac{Max\ (s_1^2; s_2^2)}{Min\ (s_1^2; s_2^2)}$	$F > F_{1-\alpha/2}^{(\upsilon_1 = n_1 - 1; \upsilon_2 = n_2 - 1)}$

V případě ověření homogenity je testována hypotéza $\sigma_1^2 = \sigma_2^2$ (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.

3. spočítat testovou statistiku $t = \frac{Rozdíl\_průměrů}{SE(rozdílprůměrů)} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_1}\right)}}$, kde

   $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ je vážený odhad rozptylu, stupně volnosti jsou $\upsilon = n_1 + n_2 - 2$

4. výsledné t srovnáme s tabulární hodnotou t pro dané stupně volnosti a α (obvykle α=0,05)
5. spočítat interval spolehlivosti pro rozdíl průměrů (např. 95%)

   $(\overline{x}_1 - \overline{x}_2) \pm t_{0,975} SE(\overline{x}_1 - \overline{x}_2) = (\overline{x}_1 - \overline{x}_2) \pm t_{0,975} \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, kde počet stupňů volnosti a s^2 odpovídají výše uvedeným vzorcům

PŘÍKLADY

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.

1. nulová hypotéza byla stanovena, že hmotnosti ovcí v obou skupinách jsou shodné, alternativní hypotéza je, že jsou rozdílné.
2. 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 ($F = \frac{Max\ (s_1^2; s_2^2)}{Min\ (s_1^2; s_2^2)}, F > F_{1-\alpha/2}^{(\upsilon_1 = n_1 - 1; \upsilon_2 = n_2 - 1)}$) nebo Levennův test.
3. Pokud platí všechny předpoklady Two sample nepárového t-testu, můžeme spočítat testovou charakteristiku $t = \frac{Rozdíl\_průměrů}{SE(rozdíl\_průměrů)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_1}\right)}}$, kde
   $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ a $\upsilon = n_1 + n_2 - 2$, 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 zamítnut, skutečná pravděpodobnost α je pak 0,018. Rozdíl mezi skupinami je 1,59 kg ve prospěch skupiny s lepší výživou.
4. Pro rozdíl mezi oběma soubory jsou spočítány 95% konfidenční intervaly $(\bar{x}_1 - \bar{x}_2) \pm t_{0,975} SE(\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{0,975} \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ 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).

Neparametrické alternativy nepárového t-testu

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 parametrickou obdobu nepárového t-testu a z těchto neparametrických testů má nejvyšší sílu testu (95% párového t-testu).

Tied-ranks (provázaná pořadí), v případě stejné hodnoty dvou čísel je jako pořadí použito průměru pořadí, které by tato čísla dostala v případě odlišné hodnoty (např. dvě stejná čísla, která by v případě nestejnosti byla na pozici 7 a 8 by měla pořadí 7,5).

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. Filozofie testu spočívá v tom, že pokud budou původní soubory podobné, potom budou jejich čísla na střídačku větší a menší v obou souborech a pořadí hodnot budou také „cik-cak“ a tedy součet pořadí bude podobný pro oba soubory.

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.

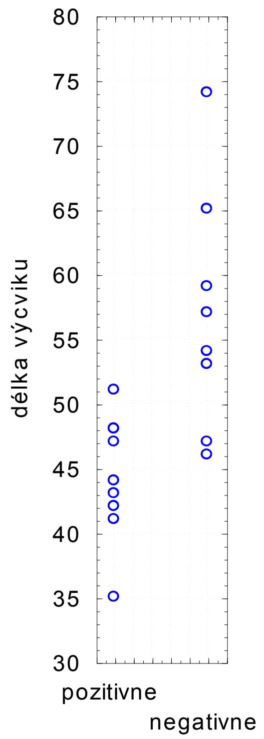
Podobným způsobem je počítán i **Wilcoxon rank sum test** (pozor, existuje ještě Wilcoxnův párový test!!!)

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		

Příklad

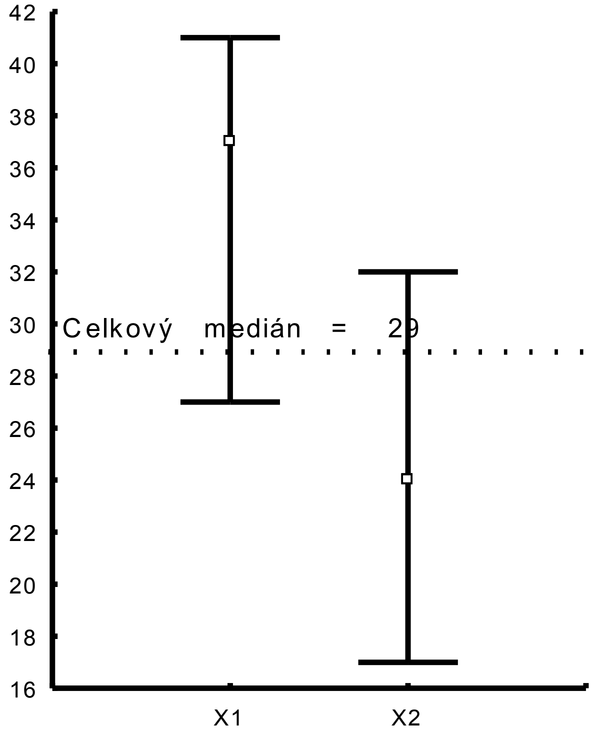
17 štěňat bylo trénováno v chození na záchod metodou pozitivního posilování (pochvala, když jde na záchod venku) nebo negativního (trest, když jde na záchod doma). Jako parametr bylo měřeno, za kolik dní je štěně vycvičeno.

1. nulová hypotéza je, že není rozdíl v metodách tréninku, tedy, že oběma metodami je štěně vycvičeno za stejnou dobu.
2. po srovnání rozložení + malý počet hodnot je vhodné použít neparametrický test
3. je vytvořeno pořadí sloučených hodnot
4. pořadí hodnot v jednotlivých skupinách dat je sečteno a menší ze součtů je použit pro srovnání s kritickou hodnotou testu
5. výsledkem testu je $p<\alpha$, nulovou hypotézu tedy zamítáme a výsledkem testu je, že pozitivní působení při výcviku štěňat dává lepší výsledky



Mediánový test

V případě mediánového testu je pro oba soubory zjištěn počet hodnot nad a pod mediánem sloučených hodnot obou souborů. Pro analýzu výsledků je použita kontingenční tabulka, kdy přepokládáme, že pokud je 50% hodnot v obou souborech nad a pod celkovým mediánem, potom není rozdíl mezi oběma soubory.

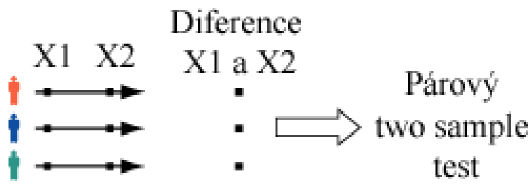


Kontingenční tabulka	X1	X2
+ medán	50	8
- medián	5	32

Tabulka je vyhodnocena a je zjištěn výsledek.

Párové testy

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.



Two sample párový 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: $t = \frac{\overline{D}}{s}\sqrt{n}$, $\upsilon = n - 1$ a provedeme srovnání s tabelární hodnotou t $?\left|t\right| > t_{1-\alpha/2}^{(n-1)}?$

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, což $\cong S_{x1-x2}$.

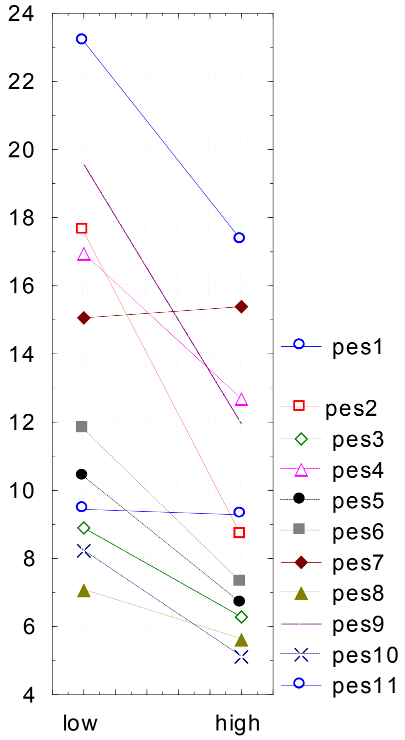
Existuje několik možných designů experimentu, stručně lze sumarizovat:

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

PŘÍKLADY

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).

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

$t = \frac{\bar{x}-\mu}{SE} = \frac{\bar{x}(diference)}{s}\sqrt{n} = 4.37$ s 10 stupni volnosti, skutečná hodnota p=0,0014 a tedy na hladině α=0,05 můžeme nulovou hypotézu zamítnout.

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.
5. Doplnění představuje výpočet 95% konfidenčního intervalu pro diference 3,8082±2,228*0,872, což znamená rozsah 1,86 až 5,75. Protože 0 neleží v tomto rozsahu, můžeme na hladině významnosti 5% zamítnout nulovou hypotézu alternativním způsobem.

Neparametrické alternativy párového t-testu

Znaménkový test

Má smysl pro data s n>10, z párových testů má nejmenší sílu testu. Je zjišťován počet kladných a záporných změn v datech. Menší z těchto čísel je srovnáno s tabulkovou kritickou hodnotou znaménkového testu a pokud je menší než tato kritická hodnota, zamítáme shodu obou souborů dat. K výpočtu je využito binomiálního rozložení, které je aproximováno na normální rozložení. Znaménkový test je možné použít i jako one tailed test, kdy zjišťujeme počty hodnot nad a pod nějakou jinou hodnotou (obdoba kladných a záporných diferencí při párovém uspořádání).

Před zásahem	Po zásahu	Změna
1	2	+
2	3	+
6	5	-
8	9	+
1	2	+
3	4	+
2	1	-
1	2	+
2	4	+
1	3	+

Příklad použití one tailed sign test

Na konferenci veterinářů bylo předneseno,že průměrný čas konzultace je 12 minut. Následovala debat, zda je lepší použít medián nebo průměr. Jede 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.

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

Příklad

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