

# 23. Statistika I.

Vladimír Znojil

## 1. Úvod

Vzhledem k velkým rozdílům mezi lidmi nelze z jediného případu usoudit na výsledky léčby, k posouzení její účinnosti, případně obecněji k formulaci vztahů mezi skupinami pokusných subjektů, proto musíme použít statistických metod. Škála statistických metod užívaných v medicíně je velice široká, jedny metody používáme například k porovnání dvou poměrně malých skupin pacientů, jiné a složitější při zpracování vysoce komplexních dat (například faktorovou nebo shlukovou analýzu) a další při statistických šetřeních na velkých populacích (do této skupiny můžeme zařadit i tabulky četností). Mnoho (hlavně moderních metod) lze jen těžko správně použít bez pomoci odborníka.

Základním pojmem při statistickém hodnocení výsledků je **nulová hypotéza**. Nulovou hypotézou může být v nejjednodušších případech předpoklad, že léčbou se stav pacienta nezlepšil, nebo že mezi studovanými skupinami není zjistitelný rozdíl, nebo že dvojice získaných veličin nejsou ve vzájemném vztahu cílem statistického testování je nulovou hypotézu vyvrátit. Přesněji řečeno prokázat, že pravděpodobně neplatí.

Při použití všech statistických metod je nezbytné vědět, že žádný ze získaných výsledků není dogmatem (pravdivý sám o sobě), každý výsledek získaný vědeckými metodami včetně statistických je správný jen do určité (a v případě statistických výsledků navíc kvantifikované) míry. Samozřejmě, pravděpodobnost správnosti většiny základních poznatků je velmi blízká 1, hodnota 1 však zůstává nedosažitelným limitem. V přírodních vědách a medicíně se dost vžilo používání dvou hraničních hodnot pravděpodobnosti omylu: pod 0.05 (v tomto případě je obvykle požívaným termínem "významný" efekt), nebo pod 0.01 ("velmi významný" efekt).

V této části se budeme zabývat jen nejjednoduššími metodami, kterými můžeme řešit mnoho běžně se vyskytujících problémů:

- zjištění, zda došlo k nějaké změně v hodnotě měřeného parametru (například krevního tlaku po zátěžovém testu)
- zda se hodnoty sledovaného parametru u dvou skupin mezi sebou liší (například výsledky určitých laboratorních vyšetření u zdravých a nemocných)
- jestli existuje souvislost mezi dvěma studovanými veličinami.

Tyto (a řadu dalších) problémy lze řešit více cestami. V některých případech můžeme použít již známých informací (například o typu rozdělení dané veličiny) k utvoření vstupních předpokladů, za nichž lze příslušnou úlohu řešit jednodušší nebo citlivější metodou. Takovým metodám zahrnujícím v sobě předpoklady o vlastnostech měřených veličin říkáme **parametrické**, těm metodám, při nichž téměř žádné vstupní předpoklady neuvádíme, pak **neparametrické**. Čím se

tyto dvě skupiny od sebe liší? Parametrické metody mohou mít (při dobrém splnění předpokladů, za nichž jsou odvozeny) poněkud větší senzitivitu a dovolují popsat zjištěné poznatky kvantitativně (například velikostí změny nebo rozdílu a chybou této veličiny, případně rovnicí uvádějící dvě studované veličiny do matematického vztahu), neparametrické metody jsou celkově robustnější. Téměř všechny běžně používané parametrické metody vycházejí z předpokladu **normálního rozdělení** veličin. Ověření jejich předpokladů není příliš jednoduché, a proto jsou v současné době používány velice často metody neparametrické.

Neparametrické metody nepočítají s naměřenými hodnotami, ale pouze s výsledky jednoduchých porovnání, případně s pořadím těchto hodnot. Nejjednodušším z testů je takzvaný **znaménkový test**, při kterém například určujeme úspěšnost léčby znakem "+", neúspěšnost znakem "-". Za prokázanou úspěšnost považujeme stav, kdy je počet případů s "-" natolik malý, že je jeho pravděpodobnost menší než 5%.

Jako příklad vezměme pokus s 12 subjekty, který skončil s výsledkem 10-krát "+" a 2-krát "-". Pokud chceme spočítat onu pravděpodobnost výskytu "-" (neúspěchu léčby), musíme nejprve určit počet všech takových situací, při kterých se vyskytne maximálně 2-krát "-" (tzn., kdy je všech 12 "+" nebo je 11-krát "+" a 1-krát "-" nebo 10-krát "+" a 2-krát "-").

- situace, že bychom měli samá "+" může vzniknout jen jediným způsobem
- situace s jedním "-" 12 způsoby (celkem je 12 subjektů, kterýkoliv z nich může mít "-").
- situace se dvěma "-" vznikne 12\*11 způsoby (poprvé se vybírá ze 12 subjektů, podruhé se vybírá z 11 zbývajících), protože se však jedná o kombinace a nezáleží na pořadí (který byl vybrán první a který druhý), je skutečný počet možností poloviční, tedy 66.

K získanému (nebo příznivějšímu) výsledku se tedy dostaneme  $1+12+66 = 79$  způsoby. Formálně tento postup zapíšeme jako součet kombinačních čísel

$$N = \sum_{k=0}^m \binom{n}{k} = \sum_{k=0}^m \frac{n!}{k!(n-k)!}$$

kde **n** je celkový počet pokusných objektů (zde 12), **m** je počet vybraných (zde 2). Jaký však je celkový počet všech možných situací (bez ohledu na počty "+" a "-")? Tento počet získáme z uvedené rovnice, pokud bychom za **m** dosadili **n**, tzn. vypočetli bychom součet všech kombinačních čísel od 0 do **n**. Lze snadno spočítat, je to  $2^n$ , tedy v našem případě 4096. Hledaná pravděpodobnost je pak prostým podílem  $79/4096 = 0.019$ , tedy v tomto příkladu méně než "požadovaných" 5%.

Podobným způsobem pracuje i mediánová ANOVA. Výhodou těchto metod je, že umožňují srovnávat i neúplně kvantifikovatelná data, mají však dosti malou citlivost. Nejčastěji používanými metodami jsou proto citlivější metody pracující s pořadími získaných hodnot, pokud mají některé údaje stejná pořadí, použijeme jejich průměrné hodnoty (například z pořadí 12, 13 hodnotu 12.5, podobně z pořadí 10, 11 a 12 hodnotu 11).

Úloha uvedená v minulém odstavci vede na Wilcoxonův test. Jeho technické provedení je jednoduché: spočteme rozdíly změřených hodnot před léčbou ( $x_i$ ) a po léčbě ( $y_i$ ) jako

$$d_i = x_i - y_i$$

a vytvoříme pořadí hodnot  $d_i$  bez ohledu na znaménko. Spočteme součet pořadí všech kladných hodnot  $d_i$ , a tím získáme hodnotu testovacího kritéria  $T$ . Meze testovacího kritéria  $T$  jsou většinou udány pro hodnoty pravděpodobnosti  $\alpha < 0.5$ , pro hodnoty  $\alpha > 0.5$  lze spočítat meze testovacího kritéria

$$T_{1-\alpha} = \frac{n \cdot (n + 1)}{2} - T_{\alpha}$$

rozdělení je tedy symetrické (je to zcela logické, setřídění od největšího čísla a setřídění od nejmenšího musí dát - při patřičné úpravě nulové hypotézy - stejnou odpověď).

Takto provedený test je jednostranný, odpovídá tedy na nulovou hypotézu typu: skupina A dává nižší nebo stejné hodnoty jako skupina B. Pokud chceme testovat hypotézu, že skupiny se od sebe neliší, musíme testovat nejen situaci již uvedenou, ale i situaci, že A dává vyšší, nebo stejné hodnoty jako B. Na každý z těchto testů musíme použít polovinu přípustné hodnoty pravděpodobnosti omylu, testujeme tedy tak, že místo  $\alpha$  použijeme pravděpodobnost  $\alpha/2$  a test provedeme také pro hodnotu pravděpodobnosti  $1-\alpha/2$ . Například pro 12 pokusů najdeme pro pravděpodobnost 0.01 (1%) hodnoty  $T_{0.005} = 7$  a  $T_{0.995} = 78 - 7 = 71$ . Pokud nám tedy vyjde součet pořadí mezi 7 a 71 nepodařilo se nám rozdíly na úrovni pravděpodobnosti 0.01 prokázat.

Častěji musíme porovnat dvě skupiny získané na  $m$  a na  $n$  nezávislých objektech (porovnáváme například dva způsoby léčby).

V tomto případě sestrojíme společné pořadí hodnot  $r_i$  a  $s_j$ , z těchto pořadí spočteme pro skupiny o velikostech  $m$  a  $n$  součty

$$W_1 = \sum_{i=1}^m r_i$$

$$W_2 = \sum_{j=1}^n s_j$$

tedy součet pořadí z jedné a/nebo z druhé skupiny. Častěji se proto používají hodnoty

$$U_1 = \frac{mn - m(m+1)}{2} - W_1, \quad U_2 = \frac{mn - n(n+1)}{2} - W_2$$

Při použití takto zavedených kritérií mluvíme o **Mann-Whitneyově testu** (je původně definovanému postupu - **dvojvýběrovému Wilcoxonovu testu** - zcela ekvivalentní).

Pro jednostranný test používáme jen jednu z hodnot  $U_1, U_2$ . Tabulky kritických hodnot  $U_{\alpha}$  jsou obvykle tabelovány pro  $m \leq n$ , za prvou skupinu proto obvykle bereme tu menší. Obdobně jako při Wilcoxonově testu platí

$$U_{\alpha} = mn - U_{1-\alpha}$$

Protože na pořadí skupin nezáleží, můžeme (pochopitelně s uvažováním správného směru porovnání) použít vztahu  $U = \min(U_1, U_2)$ . Pro oboustrannou hypotézu musíme opět použít hraničních hodnot  $\alpha/2$  místo  $\alpha$  (stejně jako u Wilcoxonova testu).

Třetím velmi používaným testem je test vzájemné souvislosti mezi dvěma kvantitativními veličinami. Lze provést více způsoby, nejběžněji se z nich v současné době používá výpočet Spearmanova koeficientu korelace. Při jeho výpočtu postupujeme tak, že si utvoříme dvě pořadí: jedno pro veličiny  $x_i$ , druhé pro  $y_i$ . V ideálním případě budou obojí pořadí v příslušných dvojicích měřených hodnot stejná (pokud hodnoty  $y$  s rostoucím  $x$  vesměs rostou), nebo opačná (pokud klesají): Postupujeme tak, že pro každou dvojici hodnot  $x_i, y_i$  spočteme rozdíl jejich pořadí  $d_i$ . Spearmanův koeficient korelace  $r_s$  potom je

$$U_{\alpha} = mn - U_{1-\alpha}$$

pro test významnosti koeficientu  $r_s$  lze pak použít vztahu ( $n$  je vesměs počet dvojic):

$$t = r_s * \sqrt{\frac{n-2}{1-r_s^2}}$$

hodnoty mezních testovacích kritérií pravděpodobnost  $t$  najdeme v tabulkách Studentova rozdělení s  $n-2$  stupni volnosti.

**Důležitá poznámka:** Takto provedený test není zcela korektní, pro malé počty  $n$  je lépe použít tabulek mezních hodnot Spearmanova koeficientu z větších statistických tabulek.

Dalším problémem je výskyt stejných pořadí při třídění souborů (viz poznámka výše). Tyto situace by se rozhodně neměly ve studovaných datech vyskytovat příliš často, výsledky získané z dat tohoto typu jsou nespolehlivé, spočtené hodnoty se od skutečných pravděpodobností mohou (i značně) lišit. Na "nedodržení" možnosti vytvořit jednoznačná pořadí (bez opakování) není příliš citlivý ani Wilcoxonův test (často se též u nás užívá ve spojení jednovýběrový Wilcoxonův test),

ani Mann-Whitneyův test, Spearmanův test je však na výskyt stejných pořadí v údajích  $x_i$  i  $y_i$  velice citlivý. Pokud nejsou tyto případy velmi ojedinělé, musíme na jejich výskyt výpočet hodnoty  $r_s$  korigovat, pro hodnotu  $r_s$  pak platí

$$r_s = \frac{A + B - D^2}{2\sqrt{AB}}$$

v tomto vztahu jsou hodnoty, kde

$$A = \frac{n^3 - n}{12} - T_x, B = \frac{n^3 - n}{12} - T_y, D = \sum_{i=0}^n d_i^2$$

korekční členy na stejná pořadí jsou v hodnotách  $x$   $T_x$  a v hodnotách  $y$   $T_y$ , obojí hodnoty spočteme jako

$$T = \sum \frac{t^3 - t}{12}$$

kde jednotlivá  $t$  jsou počty stejných hodnot příslušné veličiny (pro dvojici tedy 7/12, pro trojici 26/12 a podobně).

I tyto vztahy jsou však jen přibližné, pokud nabývají měřené veličiny jen několika málo hodnot (třeba 3-4) jsou uvedené postupy velmi nespolehlivé. Posoudit věrohodnost výsledků jde pochopitelně i v těchto případech (celou řadou metod, vesměs dost komplikovaných), nejjednodušší z nich jsou metody četnostních (kontingenčních) tabulek. Kromě nejjednodušších z nich (2x2 případy) jsou obvykle pravděpodobnosti výsledků počítány aproximativními metodami nebo statistickým modelováním (obvykle metodou Monte-Carlo).

## 2. Mnohonásobné srovnání

Při použití statistických metod na složitější úlohy je třeba brát v úvahu korekci na mnohonásobné srovnání. Mnohonásobné srovnání vzniká například tehdy, když určujeme ve které ze tří skupin má sledovaná veličina nejvyšší hodnotu. Čím více porovnání děláme, tím je pravděpodobnější, že (často zcela náhodně) najdeme nějaký rozdíl. Pro tyto situace je obvykle používán Bonferoniho postup: limitní zvolenou hranici významnosti  $P$  dělíme počtem použitých srovnání. Jinou, modernější, možnost nabízí Holmův postup. Při tomto postupu seřadíme významnosti od nejvyšší (s nejmenší hodnotou  $P$ ) po nejméně významnou (tedy vzestupně dle  $P$ ). Při  $N$  srovnáních vynásobíme prvou hodnotu číslem  $N$ , druhou  $N-1$ , třetí  $N-2$  a tak dále, až předposlední 2 a poslední již nenásobíme. Poté postupujeme od počátku tabulky a porovnáme znásobené hodnoty se zadanou mezí (např. 0.05). Pokud jsou porovnávané hodnoty menší než mez, jsou jednotlivá jim příslušná tvrzení významná, pokud je vynásobená hodnota větší již nelze příslušné tvrzení a všechna tvrzení následující považovat za významná.

Problémem téměř všech metod korekce je obvykle nesplněný předpoklad, že jednotlivé porovnávané hodnoty jsou na sobě nezávislé, často existují vztahy mezi skupinami, korelace nebo asociace jednotlivých srovnávaných hodnot. Zanedbání těchto vztahů vede ve většině případů (ne však nutně!) k tomu, že korigované pravděpodobnosti jsou "překorigovány", a že jsou tedy větší, než by být měly (experimentátor je tím tedy "poškozen"). Námitky vůči korektnosti takto korigovaných srovnání jsou proto vzácné.

### 3. Design experimentu

Jednou z nejobtížnějších oblastí experimentální práce je návrh studie. Nesčetné námitky často provázejí (i dosti vtípná) zpracování více-méně náhodně sebraných medicínských údajů. Pokud má mít studie skutečnou hodnotu je nutné získávat údaje přesně definovanými postupy na dobře definovaných (a standardizovaných) souborech pacientů. Právě proti této zásadě se mnoho studií prohřešuje: je dosti běžné, že při srovnávání nové a starší léčebné metody není výběr pacientů náhodný a tím mohou být výsledky značně zkreslené (nejlepšímu chirurgovi umře nejvíce pacientů - operuje většinu těžkých případů). Mimořádnou pozornost je věnovat pokusům, při nichž je například posuzována kvalita života - za této situace se může velmi silně uplatnit "placebo efekt", zkreslení výsledků jím vyvolané může vést až k jejich bezcennosti (některé výzkumy, pokud mají být vůbec provedeny, se často blíží k hranicím etiky - pacient se nemá dozvědět, jak je léčen). Při hodnocení stavu pacientů může být nejen pacient, ale i lékař ovlivněn svými názory na lék, nebo použitý postup). Pokusům, při nichž pacient ani lékař nevědí do ukončení experimentu, do které skupiny léčených pacient náleží, říkáme "dvojitě slepé studie". Pro řadu problémů medicíny ale asi představují jediný dostatečně spolehlivý přístup k získání adekvátních výsledků.

I v nejjednodušších studiích se rozhodně vyplatí mít srovnávané soubory věkově, pohlavím a případně i dalšími charakteristikami dobře vyvážené. Design studií začátečníků by měl být také dost jednoduchý, aby odpovídal na několik málo přesně definovaných otázek.

## 24. Statistika II.

Vladimír Znojil

V kapitole “Statistika I.” bylo uvedeno několik nejjednodušších statistických metod (vesměs neparametrických) a některé poznámky k jejich výběru a konstrukci pokusů. Nemohlo pochopitelně jít o vyčerpávající přehled, ale spíše o ilustraci k jejich pochopení. Tato část se zabývá parametrickými metodami, opět jen v užším výběru skupinou metod, která je odvozena pro normální rozdělení sledovaných hodnot.

Důvodem proč používáme parametrické metody dnes již nebývá snadnější provádění statistických testů (řada statistických programů pracuje spolehlivě s neparametrickými testy), ani jejich vyšší citlivost (někdy dost sporná, například při Laplaceovsky rozdělených datech je neparametrický Mann-Whitnevův test asymptoticky optimální) jako spíše možnost získat příslušné vztahy “v číselném vyjádření”. Právě možnost kvantifikace spojená s touto skupinou testů je jejich hlavní výhodou. Popisy odvozené z normálního rozdělení proto často používáme i tehdy, když vlastní testy provedeme neparametricky.

Hustota normálního rozložení je symetrickou křivkou, popsanou vztahem

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

parametr  $\mu$  v tomto vztahu udává střední hodnotu (průměr), parametr  $\sigma^2$  rozptyl (druhou mocninu střední vzdálenosti od průměru).

Je prokázáno, že pokud působí aditivně na měřenou hodnotu mnoho různých malých vlivů a rušivých faktorů, bude mít tato veličina rozdělení blízké se normálnímu (Ljapunovova věta). V medicíně (a v biologii) se však bohužel příliš často setkáváme s případy, kdy předpoklad jednoduché aditivity působících faktorů není splněn a proto má většina veličin rozdělení blízké se spíše logaritmicko-normálnímu (normální rozdělení nemá samotná veličina ale její logaritmus), někdy se normálnímu rozdělení blíží spíše druhé odmocniny z měřených hodnot (například u systolického i diastolického krevního tlaku). Jak je ale možné posoudit tvar rozdělení a jeho vlastnosti?

Na úvod této části nejprve několik pojmů: *průměr* souboru je součet všech hodnot v souboru dělený jejich počtem. *Medián* je hodnota určená tak, že polovina hodnot v souboru je menších a polovina větších (při lichém počtu prvků je rovna prvku s pořadím  $(n + 1) / 2$ , při sudém počtu průměru mezi prvky s pořadím  $n / 2$  a následujícím). *Modus* je nejčastěji se vyskytující hodnota. Mimo tyto hodnoty se používá pojmu “*geometrický průměr*” pro n-tou odmocninu ze součinu n



hodnot a podobně.

Je zřejmé, že se těmito způsoby určené hodnoty sobě vzájemně nerovnaj (jak by tomu při normálním rozdělení mělo být. K optickému posouzení charakteru distribuce pozorovaných hodnot můžeme použít metody *histogramu*: celé rozmezí měřených hodnot si rozdělíme na několik stejných úseků a spočteme počty případů v jednotlivých úsecích (počet úseků by měl při menším počtu měření být velmi zhruba asi odmocninou z počtu hodnot, (histogram z malého počtu údajů, tedy asi 10-20 má ale malou výpovědní hodnotu). Histogram by měl být pěknou symetrickou zvonovitou funkcí, podobnou hustotě pravděpodobnosti normálního rozdělení. Pro posouzení “normality” získaných dat se obvykle používá buď Kolmogorov-Smirnovův test, nebo  $\chi^2$ -test; jejich společnou nevýhodou je malá citlivost při malých počtech pokusů. Nás ale ve skutečnosti nezajímá, za se pozorované rozdělení *významně liší* od normálního, ale zda můžeme pozorované rozdělení *považovat* za normální aniž by byl náš další výpočet zatížen *podstatnější metodickou chybou*. K tomu účelu je možné k posouzení normality použít posouzení momentů. Z hodnot  $x_i$  si spočteme součty

$$\mu_m = \sum_{i=1}^n \frac{(x_i - x_0)^m}{n}$$

kde  $x_0$  je průměr hodnot  $x_i$ . Hodnotu  $\mu_2$  nazýváme *rozptyl*,  $\sigma = \sqrt{\mu_2}$  nazýváme *směrodatnou odchylkou*, nebo také *standardní deviací*.

**Důležité upozornění:** Takto spočtený rozptyl je *vnitřním rozptylem hodnot v rámci měřené řady*. Jako *rozptyl* bývá obvykle uváděn “*vnější*” rozptyl, který je opraven o vliv chyby průměru, spočteme jej stejným způsobem s tím že v uvedeném vzorečku nahradíme počet  $n$  upraveným počtem  $n-1$  (totéž platí pro *standardní deviaci*).

Pro následující výpočet vyšších momentů použijeme hodnoty  $n$  a spočteme normované momenty  $\eta_m$  ze vztahu

$$\eta_m = \frac{\mu^m}{\sigma^m}$$

pro hodnoty  $m = 3, 4$

Hodnotu  $\eta_3$  nazýváme *asymetrií*,  $\eta_4$  *špičatostí*. Pro normální rozložení je asymetrie rovna 0, špičatost 3. Pro parametrická zpracování by proto měla být asymetrie a špičatost co nejbližší těmto hodnotám - absolutní hodnota asymetrie by měla být menší než 2, pro některé “citlivější” parametrické testy spíše pod 1; špičatost by měla být menší než 4 (hodnoty kolem 1 - 2 obvykle nevadí). Je ovšem jasné, že pro malý počet měřených hodnot (zvlášť pokud nevíme, zda bývají normálně rozděleny) je lépe použít metod neparametrických, nebo méně citlivých metod odhadu požadovaných veličin.



Parametrickou analogií Wilcoxonova testu je *Studentův párový t-test*. Při tomto testu si spočteme rozdíly mezi sledovanými hodnotami před a po pokusu u všech pokusných osob:

$$z_i = x_i - y_i$$

pro jednotlivé pokusné osoby. Spočteme průměr hodnot  $z_i$  a označíme jej  $z$ . Očekávaná chyba průměru je

$$SE = \sqrt{\frac{\mu_2}{n-1}}$$

(při  $\mu_2$  počítaném výše uvedeným vzorcem).

Hodnota testovacího kritéria je

$$T = \frac{z}{SE}$$

tedy poměr mezi průměrným rozdílem hodnot a jeho chybou. Výsledek testujeme pomocí t-testu s počtem *stupňů volnosti*  $v = n - 1$ ; tedy počtu měření zmenšeného o 1. Pokud se nad smyslem počtu stupňů volnosti trochu zamyslíte je jasné, že toto číslo udává počet hodnot, z nichž může být patřičná veličina (v tomto případě SE) určena. Jediné měření nám může poskytnout jen velmi hrubý odhad průměru veličiny, o její chybě nevíme nic. K velmi přibližnému určení SE potřebujeme nejméně 2 měření, čím máme měření víc, tím je odhad SE přesnější. Z této úvahy také vyplývá tvar hustoty pravděpodobnosti t-rozdělení: při malých počtech měření je dost vysoká pravděpodobnost chybného odhadu SE, významné je hlavně riziko jeho *podcenění*, proto má toto rozdělení při malém počtu stupňů volnosti výrazná vysoká “křídla” k větším kladným i záporných hodnotám (a poněkud užší maximum než normální rozdělení).

Parametrickou obdobou Mann-Whitneyova testu je *nepárový Studentův t-test*. Při tomto testu se pracuje s průměry hodnot měřené veličiny  $x$  a  $y$  v jednotlivých skupinách samostatně, spočteme tedy  $SE_x$  a  $SE_y$  dle počtů pokusů  $n_x$ ,  $n_y$ . K hlavním předpokladům provedení tohoto testu patří nejen normalita distribucí v obou skupinách, musejí mít také přibližně stejné rozptyly a počty měření. Počet stupňů volnosti je v tomto případě dán vztahem

$$v = n_x + n_y - 2$$

a celková hodnota SE je dána vztahem

$$SE = \sqrt{\frac{[(n_x - 1) * n_x * SE_x^2 + (n_y - 1) * n_y * SE_y^2] * (n_x + n_y)}{n_x * n_y * (n_x + n_y - 2)}}$$

pokud lze rozptyly považovat za zcela stejné, případně pokud mají obě skupiny stejný počet měření, lze uvedený vzorec zjednodušit (upravíte dosazením  $SE_x = SE_y$  a  $n_x = n_y$ ). Pokud se počet stupňů volnosti (pokusů) mezi skupinami podstatně mezi sebou liší, je nutno použít korekce počtu stupňů volnosti

$$v = \frac{(SE_x^2 + SE_y^2)^2}{\frac{SE_x^4}{n_x - 1} + \frac{SE_y^4}{n_y - 1}}$$

Ze vzorců a z korekce je patrné, že může být výhodné pořídít více měření ve skupině, v níž jsou (z jakýchkoliv důvodů) měření méně přesná. Pro výpočet SE se dost často používá přibližný alternativní vzorec

$$SE = \sqrt{SE_x^2 + SE_y^2}$$

Pokud rozdělení měřených hodnot se od normálního rozdělení velmi liší, není nepárový t-test použitelný.

Protože rozdíly hodnot (i ve skupinách) bývají blíže normálnímu rozdělení než veličiny samotné (důsledek Ljapunovovy věty, viz výše) bývá výhodné sdružit pokusné objekty s kontrolními do dvojic (při složitějším uspořádání pokusu i trojic atd.) tak aby jejich vlastnosti u nichž soudíme, že budou vzhledem ke sledované veličině relevantní (věk, pohlaví, BMI a pod.) byly si co nejpodobnější, a při vyhodnocování pokusu zacházíme s celou dvojicí jako s jediným individuem, můžeme pak použít jednovýběrových postupů. Velmi často tím dosáhneme významnějších výsledků, než při jednodušším provedení experimentu.

Dosti často používanou metodou studia vzájemné závislosti dvou (i více) studovaných veličin je metoda nejmenších čtverců. Tato metoda je z hlediska svých možností mnohem obecnější, než *metody neparametrické korelace*, které jsou nezávislé na matematickém popisu vzájemného vztahu, odpovídající parametrické metody již s předpoklady o kvantitativních vztazích mezi nimi pracují. Nejjednodušším předpokladem vzájemného vztahu měřených veličin je vztah *lineární*, který jde vyjádřit rovnicí

$$y_i = y + a * (x_i - x)$$

v níž jsou  $x_i$  a  $y_i$  měřené hodnoty a  $a$  a  $b$  určované koeficienty. Pro určení významnosti vzájemného vztahu a hodnot koeficientů spočteme průměry měřených hodnot  $x$  a  $y$  a hodnoty jejich momentů  $\mu_2$  (pro obě veličiny, dále je proto označíme  $\mu_x$  a  $\mu_y$ ). Spočteme *smíšený moment* měřených dat ze vztahu

$$\mu_{xz} = \sum_{i=1}^n \frac{(x_i - x) * (y_i - y)}{n}$$

Korelační koeficient  $r_{xy}$  spočteme vztahem

$$r_{xy} = \frac{\mu_{xy}}{\sqrt{\mu_x * \mu_y}}$$

z výše uvedených vzorců lze odvodit, že korelační koeficient může být v uzavřeném intervalu  $\langle -1, 1 \rangle$ . Pokud má hodnotu nula, vzájemná souvislost mezi veličinami chybí, při kladných hodnotách obě veličiny spolu rostou (nebo klesají), při záporných hodnotách při vzrůstu jedné veličiny druhá klesá; pokud je rovna -1, 1 je vazba mezi hodnotami přesně lineární. Korelační koeficient můžeme testovat proti nule, pokud se od nuly významně liší můžeme vzájemnou souvislost mezi měřenými hodnotami považovat za prokázanou (pro testování použijeme tabulky kritických hodnot korelačního koeficientu).

Na rozdíl od neparametrických korelací je tento parametrický test zaměřen na lineární vztah mezi měřenými veličinami, pokud tento vztah není lineární, může být parametrický korelační koeficient v absolutní hodnotě menší, než neparametrický. Pro hodnotu  $a$  v uvedeném vztahu platí

$$a = \mu_{xy} \sqrt{\frac{\mu_y}{\mu_x}}$$

Pro strmost sklonu pro vztah  $x$  v závislosti na  $y$  pak platí též vzorec se výměnou hodnot  $\mu_y$  a  $\mu_x$ . Je zřejmé, že strmosti vztahu veličiny  $y$  vůči  $x$  a  $x$  vůči  $y$  nejsou vzájemně převrácenými veličinami, jak jsme z popisu lineárních závislostí v matematice zvyklí. Lze pochopitelně také odhadnout chyby těchto hodnot (jsou zhruba úměrné  $\sqrt{(1 - r_{xy})^2}$ ).

Hodnota korelačního koeficientu není příliš citlivá na normalitu výchozích dat, zásadní problémy nastávají při distribucích s “*odlehými hodnotami*”, tedy při vysoké špičatosti rozdělení alespoň jedné z korelovaných veličin (na tuto situaci nejsou neparametrické metody příliš citlivé);

rovnorné rozdělení hodnot jedné z veličin výsledky téměř neovlivní (lze tedy použít klasického postupu, kdy jednu z veličin nastavujeme (raději na více různých hodnot) a druhou zaznamenáváme (například souvislost mezi dávkou léku a léčebným efektem).

**Závěrečná upozornění:** Na tomto místě (i když se tato poznámka statistických testů netýká) považujeme za vhodné upozornit na některé další vlastnosti normálně rozdělených veličin.

O tom, že součet i rozdíl dvou normálně rozdělených veličin mají normální rozdění byla zmínka v úvodu. Tuto vlastnost však normálně rozdělené veličiny při násobení ani dělení nemají (logaritmus výsledku má však normální rozdění tehdy, když logaritmy násobených/dělených veličin jsou normálně rozdělené). Pokud jsou chyby (rozptyly) měřených veličin malé vůči měřeným hodnotám (asi do 10% až 30%, v souvislosti s požadovanou spolehlivostí výpočtu) není rozdíl mezi normálním a logaritmickeo-normálním rozdělením příliš velký a vznikající zkreslení můžeme zanedbat, v opačném případě musíme výpočty provádět s transformovanými daty, v těchto případech je vhodné výsledky takových měření konzultovat. Z tohoto hlediska je velmi problematická především práce s *indexy* (tedy s vyjádřením výsledku poměrem).