proLékaře.cz | 28.3.2024

Analýza dat v neurologii

XVI. Zlatý standard statistického testování: *t*-test

L. Dušek, T. Pavlík, J. Koptíková
Institut biostatistiky a analýz
Masarykova univerzita, Brno

✉
doc. RNDr. Ladislav Dušek, Dr.
Institut biostatistiky a analýz
Masarykova univerzita, Brno
e-mail: dusek@cba.muni.cz

Asi bychom mezi čtenáři těžko hledali někoho, kdo by neznal *t*-test jako standard pro srovnávání dvou průměrů. *T*-test využívající testovou statistiku, která má Studentovo *t* rozdělení, určitě patří mezi nejpoužívanější testy v biostatistice. V tomto díle seriálu tedy nemáme ambici přinést pro čtenáře něco zcela nového, spíše se pokusíme výklad o *t*-testu pragmaticky a přehledně uspořádat. I tento jednoduchý test má své předpoklady a pravidla, jejichž nerespektování může vést k vážným zkreslením.

Nejprve si ale musíme uvědomit, že pod souhrnným názvem „*t*-test" se v podstatě skrývá několik variant testu s rozdílným významem a interpretací. Zdůrazňování tohoto faktu je zvlášť důležité ve věku osobních počítačů, kdy uživatel statistického programu spouští proceduru jediným stiskem klávesy „enter". Může tak lehce opominout důležité předpoklady nebo přímo vybrat nesprávný test, který s danými daty buď nepůjde vůbec realizovat, anebo povede ke zcela nesmyslnému závěru. Nadto různé *t*-testy bývají často na různých místech nabídky software, a uživatel tak musí vědět, který chce zvolit, a aktivně si jej vyhledat.

Popis variant *t*-testu můžeme využít ke krátkému seznámení s jednovýběrovým a dvouvýběrovým experimentálním uspořádáním (obr. 1). Jednovýběrové testování pracuje s jedním náhodným výběrem o *n* opakováních, který vede k výběrovému rozdělení s určitou střední hodnotou. Jednovýběrový *t*-test ověřuje, zda odhad této střední hodnoty nabývá určité konkrétní hodnoty (dáno jako konstanta). Předpokladem je normální rozdělení sledované veličiny. Dvouvýběrový *t*-test pracuje se dvěma normálními rozděleními a ověřuje, zda mají stejné střední hodnoty. Toto uspořádání předpokládá provedení dvou náhodných výběrů, které jsou vzájemně nezávislé a mohou mít i různou velikost. Zvláštním typem uspořádání je srovnávání dvojic hodnot, které k sobě mají nějaký vztah (např. hodnota parametru X měřená u téhož pacienta před operací a po ní). U dvojic takto provázaných hodnot logicky sledujeme jejich rozdíly, resp. posuzujeme, zda střední hodnota jejich rozdílů je rovna určitému číslu. Je-li rovna nule, původní dvojice se v průměru neliší. U takového výpočtu hovoříme o párovém *t*-testu.

Jak vidno, existují tři základní varianty *t*-testu. Každý hodnotí zcela jinou experimentální situaci a jinou vstupní nulovou hypotézu (obr. 1, tab. 1). Párový *t*-test vede k výpočtu, který pracuje s rozdíly původně naměřených dvojic hodnot, a konečný výpočet je tak shodný s jednovýběrovým testem. Všechny varianty *t*-testu spojuje řada charakteristik:

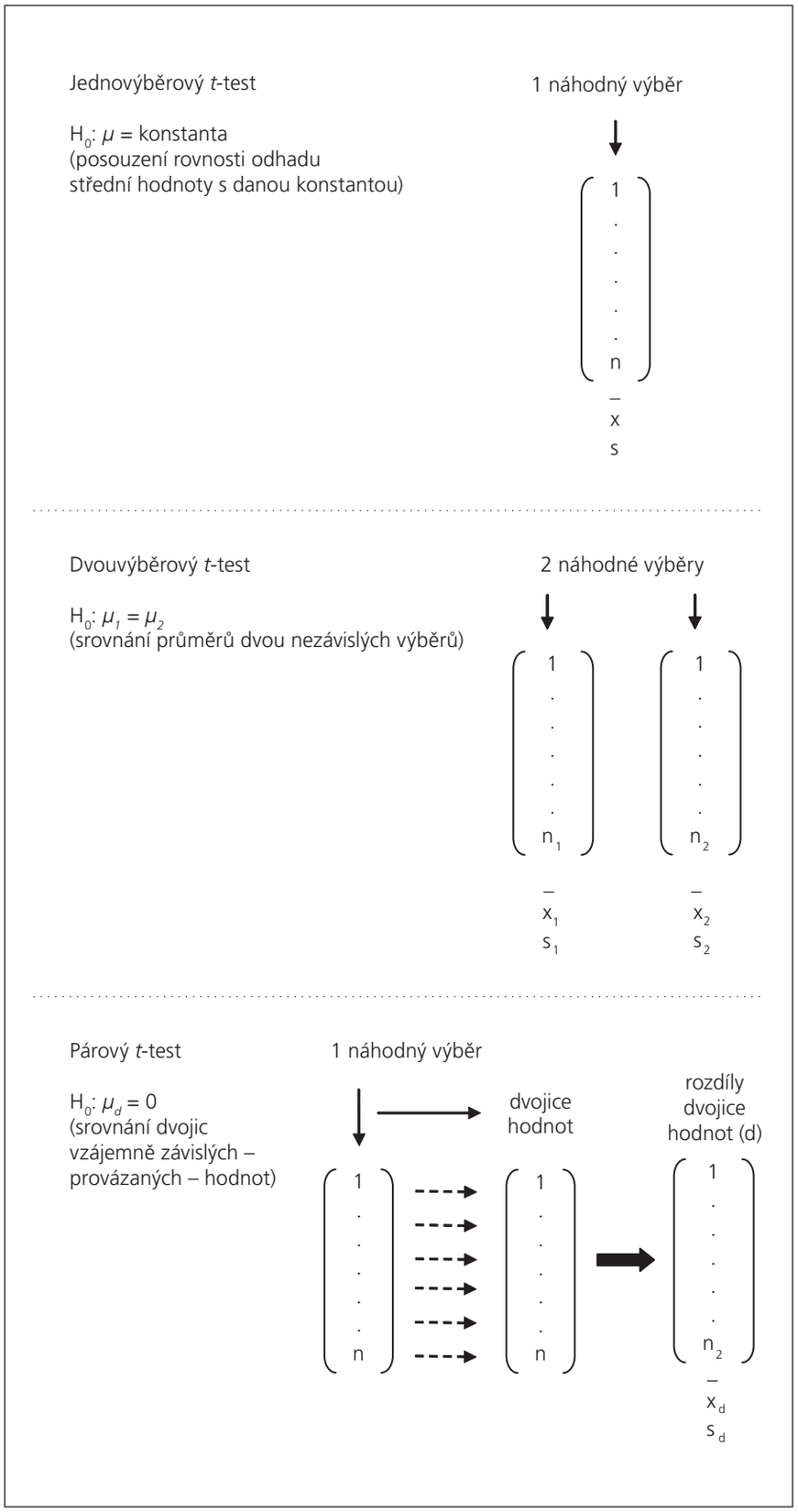
- Univerzálním předpokladem je provedení náhodného výběru z normálního rozdělení. Tento předpoklad je nutné ověřovat především u malých vzorků, kde hrozí i zvýšený vliv odlehlých hodnot.
- Vzhledem k předpokladu normálního rozdělení jsou všechny varianty *t*-testu parametrické. Pracujeme s aritmetickým průměrem jako odhadem střední hodnoty normálního rozdělení. Aritmetický průměr má sám o sobě normální výběrové rozdělení. V předchozích dílech jsme dokládali, jaký dopad má na výpočet aritmetického průměru asymetrie výběrového rozdělení nebo přítomnost odlehlých extrémních hodnot. V takových případech nemá už sám aritmetický průměr interpretační význam, a nemá tudíž smysl ani jakýkoli test o něm.
- *T*-testy mají svůj název odvozen od konečné testové statistiky, která má Studentovo *t* rozdělení. Výpočet testovacího kritéria *t* vychází z odhadů parametrů μ a σ u výběrových souborů: $\bar{x}$ a *s*. Vypočtené testovací kritérium porovnáme s tabulkovou kritickou hodnotou (1 – α/2 kvantil Studentova rozdělení pro dané ν a zvolené α).

Tab 1. Přehled variant t-testu.

t-test	Nulová hypotéza (H_0)	Stupně volnosti statistiky *t**	Komentář
Jednovýběrový	μ = konstanta	$\nu = n - 1$	Různé varianty *t*-testu vycházejí z různých experimentálních situací a testují rozdílné hypotézy (obr. 1)
Dvouvýběrový	$\mu_1 = \mu_2$	$\mu = n_1 + n_2 - 2$	
Párový	μ_d = konstanta	$\mu = n - 1$	

* Stupně volnosti jsou jedním ze základních statistických konceptů. Jednoduše je počítáme jako velikost vzorku (*n*) minus počet odhadovaných parametrů. Máme-li si vymyslet libovolných 10 čísel, máme 10 stupňů volnosti. Máme-li vygenerovat 10 čísel s nějakým daným průměrem, máme 10 – 1, tedy 9 stupňů volnosti. Stupně volnosti jsou jediným parametrem Studentova *t* rozdělení, podle kterého také hledáme kvantily tohoto rozdělení.

proLékaře.cz | 28.3.2024



Obr. 1. Schéma experimentálních uspořádání, které vedou k různým variantám *t*-testu.

Hodnota ν označuje stupně volnosti daného rozdělení *t* (tab. 1) a je jediným parametrem Studentova rozdělení (jednoduše řečeno podle ν hledáme v tabulkách). Pokud pracujeme s velkými výběry (*n* v řádu stovek a více), můžeme namísto kritických hodnot rozdělení *t* použít přímo kritické hodnoty normálního rozdělení *z*.

Jsme přesvědčeni, že i ve světě osobních počítačů a samostatně pracujících statistických programů má smysl znát postup výpočtu nejčastěji používaných testů. Proto na příkladech 1 až 3 demonstrujeme výpočet jednotlivých variant *t*-testu. V dalším textu shrnujeme některé podstatné metodické poznámky.

Jednovýběrový *t*-test:

- V tomto případě stojíme před úkolem srovnat průměrnou hodnotu *n* měření (jeden náhodný vzorek) s danou hodnotou, která představuje nějakou normativně danou hranici nebo předpoklad. V podstatě tak ověřujeme, zda provedený výběr patří svou střední hodnotou do základního souboru se středem, který určuje daná konstanta.
- V praxi jde o typický příklad, kdy ověřujeme, zda se opakované měření nějakého znaku v průměru neodchyluje od stanovené hodnoty (konstanty). Jiným příkladem může být test, zda se průměr *n* měření nějaké látky neliší od normou dané hodnoty.

Dvouvýběrový *t*-test (dva nezávislé výběry):

- Úkolem je srovnat průměry dvou náhodných výběrů, které byly získány zcela nezávisle jeden na druhém. Tedy výběr jedinců (subjektů) v jednom výběru nijak nesouvisí s výběrem v druhém výběru, a to ani pořadím, jak byli vybíráni. Oba výběry mohou mít rozdílný počet měření ($n_1 \neq n_2$); příliš velké rozdíly ve velikosti výběru se ale nedoporučují, neboť takové odhady průměru jsou obtížně srovnatelné z hlediska variability.
- Předpokladem je, že náhodné výběry mají normální rozdělení, a tento předpoklad je nutno ověřit u každého z obou výběrů samostatně, neboť do výpočtu vstupují jako nezávislé vzorky s odhadem aritmetického průměru a směrodatné odchylky (příklad 2).
- Příklad 2 popisuje situaci, kdy se rozptyly obou výběrů neliší. Pokud by variabilita

proLékaře.cz | 28.3.2024

v srovnávaných výběrech byla významně rozdílná, je nutné použít jiný postup, který podrobně popíšeme v dalším díle našeho seriálu.

Párový *t*-test:

- Test srovnávající posun hodnot ve dvojici vzájemně nějak souvisejících (závislých) výběrů. V podstatě jde o stejné srovnání jako u dvouvýběrového testu, i zde máme na vstupu dva výběry. Zásadní rozdíl je ale v tom, že u párového uspořádání nebyl druhý vzorek vybírán

Příklad 1. Jednovýběrový *t*-test pro srovnání průměrné hmotnosti souboru pacientů s danou populační hodnotou.

Nulová hypotéza H_0: Průměrná hmotnost sledovaných pacientů se neliší od předpokládané hodnoty 90 kg ($\mu = 90$ kg).
Alternativní hypotéza H_1: Průměrná hmotnost sledovaných pacientů se liší od předpokládané hodnoty 90 kg ($\mu \neq 90$ kg).
Zvolíme hladinu významnosti testu, např. $\alpha = 0{,}05$.
Hodnotou μ zde označujeme danou konstantu, která je průměrem základní populace (referenční hodnota pro srovnání)

Naměřená data (kg):

102,3	97,5	101,7	94,9	88,1	94,4	92,6	88,7	89,9	100,3

Pro výpočet testové statistiky *t* potřebujeme popisné charakteristiky souboru:
$N = 10$ pacientů Výběrový průměr $\bar{x} = 95{,}0$ kg Výběrová směrodatná odchylka $s = 5{,}3$ kg

Výpočet testové statistiky: $T = \dfrac{\bar{x} - \mu}{s/\sqrt{N}} \rightarrow T = \dfrac{95{,}0 - 90{,}0}{5{,}3/\sqrt{10}} = 2{,}98$

Pro vyhodnocení testu je nutné absolutní hodnotu testové statistiky srovnat s kvantilem Studentova *t* rozdělení pravděpodobnosti, který přísluší oboustrannému testu na zvolené hladině významnosti.

Při počtu stupňů volnosti $\nu = N - 1 = 9$
a hladině významnosti $\alpha = 0{,}05$ je kvantil $t_{1-\alpha/2}(\nu) = t_{0{,}975}(9) = 2{,}26$

Vzhledem k tomu, že absolutní hodnota testové statistiky je větší než příslušný kvantil a jedná se tedy o výsledek více vzdálený od nulové hypotézy než výsledek příslušný zvolené hladině významnosti testu, **zamítáme na hladině významnosti $\alpha = 0{,}05$ nulovou hypotézu o rovnosti průměrné hmotnosti souboru pacientů hodnotě 90 kg**.

Příklad 2. Dvouvýběrový *t*-test pro srovnání hmotnosti dvou souborů pacientů.

Nulová hypotéza H_0: Průměrná hmotnost pacientů v prvním a druhém souboru se neliší ($\mu_1 = \mu_2$).
Alternativní hypotéza H_1: Průměrná hmotnost pacientů v prvním a druhém souboru se liší ($\mu_1 \neq \mu_2$).
Zvolíme hladinu významnosti testu, např. $\alpha = 0{,}05$.

Naměřená data (kg):

Soubor 1	126,3	116,7	102,6	90,1	87,2	113,3	106,9	103,8	102,6	111,0
Soubor 2	106,1	74,4	85,5	94,6	91,6	83,7	87,4	95,8	104,1	114,6

Pro výpočet testové statistiky potřebujeme popisné charakteristiky obou souborů:
$N_1 = 10$ pacientů Výběrový průměr $\bar{x}_1 = 106{,}1$ kg Výběrová směrodatná odchylka $s_1 = 11{,}7$ kg
$N_2 = 10$ pacientů Výběrový průměr $\bar{x}_1 = 93{,}8$ kg Výběrová směrodatná odchylka $s_2 = 12{,}0$ kg

Výpočet testové statistiky:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad \text{kde} \quad s_{x_1x_2} = \sqrt{\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}} = 11{,}9. \quad \text{Tedy} \quad T = \frac{106{,}1 - 93{,}8}{11{,}9\sqrt{\frac{1}{10} + \frac{1}{10}}} = 2{,}31$$

Pro vyhodnocení testu je opět nutné absolutní hodnotu testové statistiky srovnat s kvantilem Studentova *t* rozdělení pravděpodobnosti, který přísluší oboustrannému testu na zvolené hladině významnosti.

Počtu stupňů volnosti $\nu = N_1 + N_2 - 2 = 18$ a hladině významnosti $\alpha = 0{,}05$ přísluší kvantil $t_{1-\alpha/2}(\nu) = t_{0{,}975}(18) = 2{,}10$

Vzhledem k tomu, že absolutní hodnota testové statistiky je větší než příslušný kvantil a jedná se tedy o výsledek více vzdálený od nulové hypotézy než výsledek příslušný zvolené hladině významnosti testu, **zamítáme na hladině významnosti $\alpha = 0{,}05$ nulovou hypotézu o rovnosti průměrné hmotnosti srovnávaných souborů pacientů**.

proLékaře.cz | 28.3.2024

Příklad 3. Párový *t*-test pro srovnání hmotnosti souboru pacientů před a po redukční dietě.

Nulová hypotéza H_0: Průměrná hmotnost pacientů před dietou se neliší od hmotnosti pacientů po dietě ($\mu_1 = \mu_2$).
Alternativní hypotéza H_1: Průměrná hmotnost pacientů před dietou se liší od hmotnosti pacientů po dietě ($\mu_1 \neq \mu_2$).
Zvolíme hladinu významnosti testu, např. $\alpha = 0{,}05$.
Srovnání párových dat (dvojic naměřených hodnot před a po dietě) přechází na srovnání střední hodnoty jejich rozdílů s nulou, H_0 lze tedy přepsat jako $\mu_1 - \mu_2 = 0$ a H_1 jako $\mu_1 - \mu_2 \neq 0$.
Konečný výpočet testu je stejný jako u jednovýběrového t-testu (obr. 1).

Naměřená data (kg):

Číslo pacienta	1	2	3	4	5	6	7	8	9	10
Váha před dietou	113,7	100,0	112,6	112,8	104,1	91,5	122,4	110,7	129,8	109,7
Váha po dietě	101,0	89,5	105,5	95,1	106,4	85,3	101,5	111,1	115,8	107,1
Úbytek váhy	12,7	10,5	7,1	17,7	–2,3	6,2	20,9	–0,4	14,0	2,6

Pro výpočet testové statistiky budeme potřebovat popisné charakteristiky váhových úbytků pacientů před a po dietě (označujeme *d* jako diference):
$N = 10$ pacientů Výběrový průměr $\bar{x}_d = 8{,}9$ kg Výběrová směrodatná odchylka $s_d = 7{,}6$ kg

Výpočetní vzorec testové statistiky: $T = \frac{\bar{x}_d - 0}{s_d / \sqrt{N}} \rightarrow T = \frac{8{,}9 - 0}{7{,}6 / \sqrt{10}} = 3{,}70$

Pro vyhodnocení testu je nutné absolutní hodnotu testové statistiky srovnat s kvantilem Studentova *t* rozdělení pravděpodobnosti, který přísluší oboustrannému testu na zvolené hladině významnosti.

Při počtu stupňů volnosti $\nu = N - 1 = 9$
a hladině významnosti $\alpha = 0{,}05$ přísluší kvantil $t_{1-\alpha/2}(\nu) = t_{0{,}975}(9) = 2{,}26$

Vzhledem k tomu, že absolutní hodnota testové statistiky je větší než příslušný kvantil a jedná se tedy o výsledek více vzdálený od nulové hypotézy než výsledek příslušný zvolené hladině významnosti testu, **zamítáme na hladině významnosti $\alpha = 0{,}05$ nulovou hypotézu o rovnosti průměrné hmotnosti pacientů před a po redukční dietě**.

náhodně, ale v souvislosti s prvním měřením. Základní otázka zde je, „o kolik" se liší tyto dvojice čísel a výpočet směřuje k jejich odečtení a hodnocení jejich rozdílů (obr. 1, příklad 3).

- Předpoklad normálního rozdělení zde tedy platí až pro rozdíly původních hodnot, protože i samotný výpočet pracuje pouze s těmito rozdíly.
- Typickým příkladem párového uspořádání je měření parametru na témže jedinci před nějakou změnou a po ní (experimentu), před léčebným výkonem a po něm apod. Dalším příkladem mohou být plánovitá měření na párových orgánech, které srovnáváme vždy na témže jedinci.

Je patrné, že o tom, jaký bude použit *t*-test, rozhoduje již sám experimentátor tím, jaký zvolí postup. Není to primárně odpovědnost statistika, naopak. A také bohužel platí, že mají-li dva výběry stejný počet opakování ($n_1 = n_2$), můžeme na nich současně provést výpočet dvouvýběrového i párového *t*-testu. Velmi pravděpodobně ovšem s různým výsledkem, protože každý testuje něco jiného:

- Dvouvýběrový test srovnává aritmetické průměry dvou zcela nezávislých populací. Pracujeme zde přímo s primárními hodnotami a ve výpočtu srovnáváme dva nezávislé odhady průměrů. Do výpočtu přímo vstupují výběrové odhady směrodatných odchylek s_1 a s_2. Konečná statistika *t* má počet stupňů volnosti $\nu = n_1 + n_2 - 2$.
- Párový test sleduje, o kolik se změnila průměrná hodnota v téže populaci po nějaké změně. Hodnoty naměřené před touto událostí a po ní jsou odečteny a je hodnocen průměr takto získaných rozdílů. Variabilita původních měření do výpočtu nevstupuje, pracujeme se směrodatnou odchylkou rozdílů (s_d), která bude v převážné většině případů menší než směrodatná odchylka původních měření. Počet stupňů volnosti pro konečnou statistiku *t* počítáme stejně jako u jednovýběrového *t*-testu, $\nu = n - 1$.

Příklad 4 znázorňuje situaci, kdy na data z jednoho experimentu můžeme současně použít dvouvýběrový i párový *t*-test. Zatímco dvouvýběrový test neprokazuje statisticky významný rozdíl mezi dvěma výběrovými průměry, párový *t*-test ukazuje, že průměrný rozdíl obou výběrů je významně rozdílný od nuly. Jelikož experimentální uspořádání je v příkladu 4 jednoznačně párové, správné je hodnocení párovým testem. Dvouvýběrový test nelze použít již vzhledem k tomu, že druhé měření není nezávislé na prvním; je provedeno na stejných jedincích po realizovaném experimentu. Bohužel uživatel statistického programu, který tyto skutečnosti nezná, může lehce zvolit chybný test. Snad i tento náš seriál přispěje k tomu, že se to čtenářům ČSNN nestane ☺.

Závěrem nelze odolat pokušení alespoň se zmínit o tom, komu vlastně vděčíme za *t*-test, který je také někdy nazýván „Studentův *t*-test" (jelikož využívá Studentovo *t* rozdělení). Jeho autor se totiž nejmenoval Student a nejde ani o název vě-

proLékaře.cz | 28.3.2024

Příklad 4. Modelové srovnání dvouvýběrového *t*-testu a párového *t*-testu na stejných datech.

Vstupní data: Koncentrace látky X měřená v krvi před a po určité léčbě. Sledování celkem u $n = 8$ pacientů.

Naměřená data:

Číslo pacienta	1	2	3	4	5	6	7	8
Měření před léčbou	38	10	84	36	50	35	73	48
Měření po léčbě	32	16	57	28	55	12	61	29
Diference (d)	–6	6	–27	–8	5	–23	–12	–19

n = 8 dvojic měření
Výběrový průměr $\bar{x}_1 = 46{,}8$; $\bar{x}_2 = 36{,}2$
Výběrová směrodatná odchylka $s_1 = 23$; $s_2 = 19$
Výběrový průměr diferencí $\bar{x}_d = -10{,}5$
Výběrová směrodatná odchylka diferencí $s_d = 12$

Vzhledem k tomu, že $n_1 = n_2 = 8$, můžeme současně vypočítat dvouvýběrový i párový *t*-test. Dvouvýběrový test je zde ovšem použit nesprávně, neboť jde o typické párové uspořádání experimentu a druhá z dvojice hodnot není naměřena nezávisle na první.

Výsledky výpočtu dvou variant *t*-testu:

Varianta *t*-testu	Statistika *t*	Stupně volnosti (ν)	Hodnota p
Dvouvýběrový test (použit nesprávně)	0,989	14	0,339
Párový test (použit správně)	–2,434	7	0,045

Je patrné, že dvouvýběrový test vedl k opačnému výsledku než test párový. Správný závěr nabízí test párový a můžeme tedy konstatovat, že koncentrace dané látky se po sledované léčbě statisticky významně snižuje. Dvouvýběrový test nedokázal rozdíl mezi dvojicemi hodnot určit jako významný mimo jiné i proto, že výpočet zahrnuje variabilitu původních hodnot a ta je zde viditelně vysoká. Párový test počítá pouze s rozdíly původně naměřených hodnot.

novaný altruisticky studentům statistiky. Autorem je chemik a statistik W. S. Gosset (1876–1937), který pracoval v pivovaru Guinness. Zde měl za úkol srovnávat kvalitu jednotlivých várek piva a vyvíjet procedury, které by zaručovaly standardní testování kvality. *T*-test vyvinul jako postup pro porovnávání obsahu kvasinek v jednotlivých várkách s nastaveným standardem. A jelikož si vedení pivovaru nepřálo veřejné publikace výzkumných výsledků svých zaměstnanců, pilný pan Gosset sdělil své objevy lidstvu pod pseudonymem „Student". Jak je zřejmé, vědecké poznání si vždy cestu najde. Čtenářům tedy na konci této kapitoly můžeme doporučit skvělý vychlazený Guinness jako odměnu za jejich trpělivost při studiu ☺.

www.geriatrickarevue.cz