

6 Doplnky k testování statistických hypotéz

6.1 Párové testy

V mnoha situacích jsou k dispozici data v párové struktuře. Znamená to, že pro každou statistickou jednotku je zaznamenána dvojice hodnot (odtud „párový test“). Číselná párová data získáváme tedy v situacích, kdy je pro každou statistickou jednotku měření provedeno dvakrát, a to buď

- za týchž podmínek v různém čase (nejčastěji před a po nějaké intervenci);
- nebo za různých podmínek (např. během dvou odlišných experimentů);
- nebo současně na dvou různých částech (např. porovnání hodnoty z pravé a levé paže).

V případě, kdy lze předpokládat, že se sledované veličiny řídí nějakým typem rozdělení (např. normálním), můžeme testovat tvrzení o parametru tohoto rozdělení (např. o střední hodnotě) a pak se jedná o **parametrický** přístup. Pokud podobný předpoklad nelze učinit, volíme **neparametrický** přístup.

6.1.1 Párový t-test

Specifickou podmínkou aplikace tohoto typu párového testu je, aby se jednalo o hodnoty veličiny spojitého typu s předpokladem normálního rozdělení. Jde o parametrický přístup.

Příklad 1: Pro každého pacienta-hypertonika je zaznamenán krevní tlak po podání placebo a po podání testované látky Hydrochlorothiazid (HCT) na snížení tlaku. Aplikace placebo a HCT musí proběhnout u každého pacienta s dostatečným odstupem, aby vždy odezněl efekt látky aplikované jako první, v náhodném pořadí a tzv. zaslepeně (aby pacient nevěděl, kdy dostává kterou z obou látek). Příklad a naměřené hodnoty uvedené v tabulce jsou převzaty z učebnice Zvárová J., *Základy statistiky pro biomedicínské obory* (s. 136).

Placebo	211	210	210	203	196	190	191	177	173	170	163
HCT	181	172	196	191	167	161	178	160	149	119	156

Příklad 2: U každé z 20 pacientek po ablaci prsu byl zaznamenán parametr rozsahu pohybu paže v zevní rotaci (ve stupních) ve dvou po sobě jdoucích etapách terapie (hodnoty značené jako ZR1 v první etapě, ZR2 ve druhé, viz tabulka, data převzata z výzkumu K. Vaníkové, FZS UJEP, Katedra fyzioterapie a ergoterapie).

Pac. č.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
ZR1	70	68	60	62	70	76	84	62	70	64
ZR2	70	70	60	68	70	76	84	62	74	64

Pac. č.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
ZR1	68	56	82	50	64	66	70	82	84	70
ZR2	68	60	82	56	64	68	72	82	86	70

V obou příkladech je cílem potvrdit efekt nějaké intervence (efekt léčby pomocí HCT, resp. efekt terapie po ablaci) ověřením, že rozdíl mezi oběma zaznamenávanými veličinami je statisticky významný. Jde tedy na první pohled o podobný problém jako u dvouvýběrového

testu, ale tam nebyly zaznamenávané hodnoty spárované, jednalo se vždy o dva vzájemně nezávislé výběry.

Z obou ukázek snad již jasně vyplývá, že při zpracování je stěžejní rozdílová veličina. To, co bychom mohli testovat, je tvrzení, že mezi oběma hodnotami je, resp. není významný rozdíl (diference).

$H_0: \mu_d = 0$ (aneb střední rozdíl není významný)

$H_1: \mu_d \neq 0$ (aneb střední rozdíl je významný)

kde μ_d značí střední hodnotu rozdílové veličiny („střední rozdíl“) ve dvojicích měření. Jde o test s tzv. **oboustrannou alternativou**. „Významný“ zde znamená „nenulový“, čili v případě, kdy zamítneme H_0 , by se podařilo prokázat významnou rozdílnost mezi účinkem placebo a HCT, resp. významnou rozdílnost rozsahu pohybu v obou etapách, ale bez specifikace, která z obou srovnávaných středních hodnot je vlastně větší.

V obou uvedených ilustračních situacích bude zřejmě logičtější použít test s tzv. **jednostrannou alternativou**:

$H_0: \mu_d = 0$ (aneb střední rozdíl není významný)

$H_1: \mu_d > 0$ (aneb střední rozdíl je významný)

Zde by se v případě, kdy zamítneme H_0 , podařilo prokázat významně vyšší tlak po podání placebo, aneb prokázal by se významný efekt HCT jako účinné látky na snížení tlaku. Možnost opaku, tedy zvýšení tlaku po podání HCT, se nepřipouští. Ve druhém ilustračním případě znamená „významný“ významně vyšší rozsah pohybu ve druhé etapě terapie. To, že by mohlo dojít naopak ke snížení rozsahu pohybu, je a priori vyloučeno.

Poznámka 1: V případě jednostranné alternativy samozřejmě nezáleží na tom, jestli odečítáme podle vzorce A–B nebo naopak B–A (A značí první hodnotu, B druhou hodnotu v páru), musíme ale odečítat u všech statistických jednotek (pacientů) dle téhož vzorce. Zvolený vzorec ovlivní to, že by v alternativě byla případně obrácená nerovnost (<).

Poznámka 2: Testovaný střední rozdíl nemusí být nulový. Mohli bychom např. testovat tvrzení, že účinná látka snižuje krevní tlak o 20 jednotek, dvojice hypotéz by pak vypadala takto: $H_0: \mu_d = 20$, $H_1: \mu_d \neq 20$. V Excelu pak zvolenou testovanou hodnotu rozdílu zadáme jako „Hypotetický rozdíl středních hodnot“, pokud tuto volbu vynecháme, testujeme automaticky jeho nulovost.

Provedení pomocí nástroje v Excelu ilustruje obrázek. Hodnoty tlaku po podání placebo mějme v buňkách A3 až A13, po podání HCT v buňkách B3 až B13. Z nabídky Data / Analýza dat vybereme nástroj pojmenovaný v češtině „Dvouvýběrový párový t-test na střední hodnotu“. Pokud jako součást zadání „Vstup“ zadáme i buňky s názvy obou veličin (první soubor tedy včetně buňky A2, druhý včetně buňky B2), zaškrtneme v menu položku „Popisky“. Jako hodnotu alfa ponecháváme standardní volbu 0,05. Výstup ponecháme na „Nový list“.

Výstup se zobrazí na novém listu. Pro obě veličiny zvlášť (placebo a HCT) zobrazuje vypočtené průměry (v překladu „Stř.hodnota“), výběrové rozptyly („Rozptyl“) a počty hodnot („Pozorování“). Míru shody dvojice veličin popisuje korelační koeficient („Pears.korelace“). Zadaný testovaný rozdíl najdeme na řádce „Hyp. rozdíl stř. hodnot“ (standardně je zde nula, viz Poznámka 2 výše). Hodnota pojmenovaná „Rozdíl“ vyjadřuje ve skutečnosti tzv. počet stupňů volnosti příslušného t-testu, který určujeme jako počet párů mínus jedna. Běžné značení je DF (degrees of freedom). Při provedení t-testu počítáme z naměřených dat hodnotu tzv. testové statistiky („t Stat“, 6,08), kterou porovnáváme s příslušnou kritickou hodnotou. Kritická hodnota je jiná při oboustranné alternativě (na výstupu „t krit (2)“, 2,23) a při jednostranné (na výstupu „t krit (1)“, 1,81). Pokud testová statistika překročí kritickou hodnotu, zamítáme H_0 ve prospěch příslušné alternativy. V ilustračním příkladu bychom tedy H_0 při jednostranném (ale i oboustranném) přístupu zamítli a léčebný efekt HCT se tak podařilo prokázat.

	A	B	C	D	E	F	G	H
1	zdroj: Zvárová J., Základy statistiky pro biomedicínské obory (ISBN: 80-7184-786-0, s.136)							
2	Placebo	Hydrochlorothiazid						
3	211	181						
4	210	172						
5	210	196						
6	203	191						
7	196	167						
8	190	161						
9	191	178						
10	177	160						
11	173	149						
12	170	119						
13	163	156						
14								

Dvouvýběrový párový t-test na střední hodnotu

Vstup

1. soubor:

2. soubor:

Hypotetický rozdíl středních hodnot:

Popisky

Alfa:

Možnosti výstupu

Výstupní oblast:

Nový list:

Nový sešit

	Placebo	Hydrochlorothiazid
Stř. hodnota	190,3636	166,36
Rozptyl	303,2545	458,85
Pozorování	11	11
Pears. korelace	0,791776	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	10	
t Stat	6,07998	
P(T<=t) (1)	5,94E-05	
t krit (1)	1,812461	
P(T<=t) (2)	0,000119	
t krit (2)	2,228139	

V posledních letech se využívá toho, že SW dokáže rychle určit tzv. p-hodnotu, což je vlastně pravděpodobnost odpovídající (ne)překročení příslušné kritické hodnoty zjištěnou

testovou statistikou. Pro oboustranný test je tato p-hodnota na výstupu značena jako „ $P(T \leq t)(2)$ “ (0,000119), pro jednostranný jako „ $P(T \leq t)(1)$ “ ($5,94E-05 = 5,94 \cdot 10^{-5} = 0,0000594$). Podle principu rozhodování o testované H_0 pomocí p-hodnoty zamítneme v našem příkladu H_0 ve prospěch jednostranné (ale i oboustranné) alternativy, neboť výsledná p-hodnota je menší než hladina 0,05.

Pokud nepotřebujeme všechny výše uvedené výpočty, ale stačí nám pouze rozhodnout o H_0 pomocí p-hodnoty, můžeme místo výše předvedeného nástroje použít příkaz:

=TTEST(matice1;matice2;chvosty;typ)

kde zadáváme

matice1...hodnoty první veličiny (zde odkaz na oblast buněk A3:A13);

matice2... hodnoty druhé veličiny (zde odkaz na oblast buněk B3:B13);

chvosty...zadáme hodnotu 1 pro jednostrannou, resp. 2 pro oboustrannou alternativu;

typ...zadáme hodnotu 1 pro volbu „párový test“.

Výsledkem provedení příkazu je p-hodnota (samozřejmě musí vyjít stejně jako odpovídající p-hodnota určená pomocí nástroje).

Spuštění stejné úlohy pomocí SW STATISTICA ilustruje obrázek. Při výběru „Proměnných“ zadáme např. placebo jako "1. seznam proměnných" a HCT jako "2. seznam proměnných ". Výsledky pak nalezneme pod nabídkou "Výpočet". Oproti výsledkům z nástroje v Excelu zde chybí hodnota korelačního koeficientu a test je uveden pouze v oboustranné variantě. Kromě p-hodnoty ale obsahují výsledky také interval spolehlivosti pro střední rozdíl. V menu navíc můžeme získat grafické znázornění obou proměnných volbou "Křabicový graf".

	1 placebo	2 HCT
1	211	181
2	210	172
3	210	196
4	203	191
5	196	167
6	190	161
7	191	178
8	177	160
9	173	149
10	170	119
11	163	156

Pro procvičení si čtenář může samostatně analyzovat data z příkladu 2.

6.1.2 Párový Wilcoxonův test

Tento typ testu používáme v případě, kdy nejsou splněny předpoklady pro aplikaci párového t-testu: data nevykazují normalitu, nebo je dvojic hodnot příliš málo (obvykle se jako dostatečný počet uvádí 30 nebo více). Jako příklad poslouží opět data z učebnice Zvárová, J.,

Základy statistiky pro biomedicínské obory (s. 147). K dispozici máme počty buněk T4 a T8 v mm³ ve vzorcích krve u 10 pacientů s Hodgkinsovou nemocí.

T4	396	568	1212	171	554	1104	257	435	295	397
T8	236	786	311	449	811	686	412	286	336	936

V tomto případě testujeme tvrzení $H_0: M = 0$ proti $H_1: M \neq 0$, kde M značí medián rozdílů (situace je tedy obdobná jako u párového t-testu, zde je μ_d nahrazeno M). Jelikož M nepředstavuje parametr rozdělení, jedná se o neparametrický typ testu. V Excelu není k dispozici předpřipravený nástroj nebo příkaz, bylo by nutno postupovat podle vzorce pro výpočet testové statistiky a tu porovnat s kritickou hodnotou ze statistické tabulky pro tento typ testu. Při výpočtu testové statistiky bychom museli určit rozdíly v jednotlivých dvojicích hodnot, rozdíly v absolutní hodnotě seřadit podle velikosti a následně k pořadím vrátit původní znaménka rozdílů. Test se proto nazývá také **signed-rank**. Zpracování je snadnější pomocí SW STATISTICA (modul Neparametrické statistiky, volba „Porovnání dvou závislých vzorků“), postup je zřejmý z obrázku.

Wilcoxonův párový test (Tema6-Wilcox)				
Označené testy jsou významné na hladině $p < ,05000$				
Dvojice proměnných	Počet platných	T	Z	p-hodn.
T4 & T8	10	24,00000	0,356753	0,721277

Ve spodní části obrázku vidíme výstup pro „Wilcoxonův párový test“. Analyzovány byly proměnné T4 a T8, k dispozici bylo 10 párů hodnot, testová statistika T má hodnotu 24. Kritickou hodnotu výstup bohužel neuvádí, museli bychom ji vyhledat ve statistických tabulkách (pro 10 dvojic má hodnotu 8; v případě Wilcoxonova testu rozhodujeme tak, že pokud testová statistika T překročí kritickou hodnotu, nelze H_0 zamítnout). Výstup místo toho transformuje T do hodnoty tzv. normální Z-aproximace (0,3568). Jde o přibližnou testovou statistiku, která se asymptoticky (zjednodušeně řečeno s vyšším počtem pozorování stále lépe a lépe) řídí normálním rozdělením. Výhodou této Z-aproximace je, že k rozhodnutí nepotřebujeme speciální tabulku, neboť kritickou hodnotou je běžně známý kvantil normovaného normálního rozdělení (při 5% hladině významnosti hodnota 1,96). V případě

testové Z statistiky rozhodujeme tak, že pokud nepřekročí kritickou hodnotu (1,96), nelze H_0 zamítnout. Nejjednodušší rozhodnutí je ovšem s využitím uvedené p-hodnoty (0,721). Jelikož p-hodnota překračuje hodnotu 0,05, nelze zamítnout hypotézu H_0 . Data neprokázala statisticky významný rozdíl v počtech buněk T4 a T8 u daných pacientů.

Poznámka 1: Na internetu jsou k dispozici různé kalkulatory, doporučit lze např. tento:

<http://www.socscistatistics.com/tests/signedranks/Default2.aspx>

Použití je zcela intuitivní, do sloupců „Treatment 1“ a „Treatment 2“ jednoduše vkopírujeme popořadě data (jako dva sloupce), hladinu významnosti („Significance level“) ponecháme standardní 0,05 a v případě, kdy požadujeme test s jednostrannou alternativou, zvolíme místo přednastavené volby „Two-tailed“ variantu „One-tailed“. Po pokynu „Calculate“ obdržíme výstup, v němž je patrný postup výpočtu testové statistiky (určení pořadí včetně jejich znamének), která je zde namísto T značena W (podle názvu testu). V části výstup „Result 1“ vidíme přímo Z-aproximaci (bez ohledu na znaménko 0,3568) a jí odpovídající p-hodnotu (0,71884). Až v části „Result 2“ najdeme testovou statistiku stejnou (až na označení) jako podle SW STATISTICA ($W=24$), přičemž v tomto případě je rozhodnutí pro změnu činěno právě pomocí tabulkové kritické hodnoty (8), taktéž zde uvedené. Tak či tak je výsledkem rozhodnutí samozřejmě stejné jako podle SW STATISTICA: nelze zamítnout hypotézu H_0 , data neprokázala statisticky významný rozdíl v počtech buněk T4 a T8 u daných pacientů.

Poznámka 2: Pro případ, kdy nelze měřit některou (příp. ani jednu) z párových hodnot, ale lze určit, zda došlo ke zlepšení či zhoršení, lze použít **Wilcoxonův znaménkový (sign) test**. Pro zpracování pomocí SW STATISTICA stačí zaznamenat situaci např. dvojicí hodnot 1-2 vždy, když změnu hodnotíme jako nárůst a 1-0 naopak vždy, když ji hodnotíme jako pokles). Odpovídající webová aplikace je např. zde:

<http://www.socscistatistics.com/tests/signtest/Default.aspx>

6.2 ANOVA – speciální případy

V kapitole 5 jsme se seznámili s testem ANOVA, a to v situaci tzv. jednoduchého třídění, kdy jsme uvažovali jediný faktor. V rámci této kapitoly se seznámíme s modely pro **dvojně třídění** (budeme analyzovat vliv dvou faktorů) a ANOVA pro **opakovaná měření** (jakési zobecnění párového testu, ale pro libovolný počet opakování, nejen pro dvojice hodnot).

6.2.1 ANOVA – dvojně třídění bez opakování

Předpokládejme, že sledujeme závislou veličinu Y spojitého typu, normálně rozdělenou. Chování veličiny může být různé v závislosti na dvou kategoriálních faktorech. Příkladem by mohla být závislost platu (Y) na faktorech vzdělání (rozlišujeme kategorie ZŠ, SŠ, VŠ) a pohlaví (muž, žena). V této situaci obvykle chceme testovat hned několik tvrzení najednou: vliv samotného faktoru č.1 (závisí plat na vzdělání?) a vliv samotného faktoru č.2 (závisí plat na pohlaví?). Označme obecně počet kategorií prvního faktoru jako r, počet kategorií druhého faktoru jako s. Pokud pro každou kombinaci faktorů zaznamenejme jedinou hodnotu Y, jde o

data **bez opakování** (angl. „without replications“), která můžeme zaznamenat do tabulky, v níž r značí počet řádků, s počet sloupců. Testované hypotézy mají potom tvar:

pro faktor 1

$H_0^{(1)}$: $\mu_{1\cdot} = \dots = \mu_{r\cdot}$ (střední hodnota veličiny Y nezávisí na faktoru 1)

$H_1^{(1)}$: non $H_0^{(1)}$ (střední hodnota veličiny Y závisí na faktoru 1)

a pro faktor 2

$H_0^{(2)}$: $\mu_{\cdot 1} = \dots = \mu_{\cdot s}$ (střední hodnota veličiny Y nezávisí na faktoru 2)

$H_1^{(2)}$: non $H_0^{(2)}$ (střední hodnota veličiny Y závisí na faktoru 2)

„Tečkovaná“ symbolika se běžně používá k rozlišení, zda hovoříme o údajích pro řádky či sloupce (např. $\mu_{1\cdot}$ značí hodnotu odpovídající prvnímu řádku, zatímco $\mu_{\cdot 1}$ značí hodnotu odpovídající prvnímu sloupci).

Data (<http://archive.bio.ed.ac.uk/jdeacon/statistics/tress8.html>) zaznamenávají 12 vzájemně nezávislých kultivací při čtyřech různých teplotách (faktor 1) a 3 různých pH (faktor 2). Sledovanou veličinou Y byla „optical density“. Pro každou kombinaci teploty a pH máme jedinou zaznamenanou hodnotu, viz tabulka.

Temp oC	pH 5.5	pH 6.5	pH 7.5
25	10	19	40
30	15	25	45
35	20	30	55
40	15	22	40

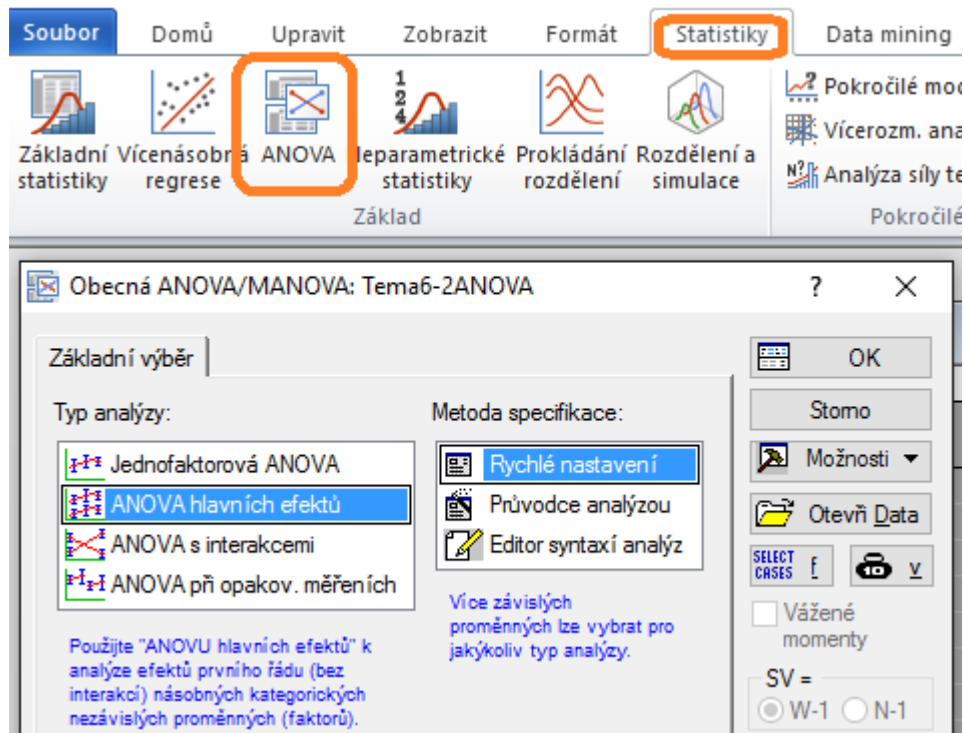
V Excelu pro takto zapsaná data spustíme z nabídky Data / Analýza dat nástroj „Anova: dva faktory bez opakování“. Pokud při zadávání do „Vstupní oblasti“ zahrneme i nadpisy řádků a sloupců (první bílý sloupec, resp. řádek z tabulky), zatrhneme možnost „Popisky“. Jako „Výstupní oblast“ je nejjednodušší zvolit „Nový list“. Výsledkem je přehled deskriptivních charakteristik pro oba faktory (pro každou kategorii zvlášť počet hodnot, jejich součet, průměr a rozptyl) a dále tzv. tabulka ANOVY.

Položky „Řádky“ se týkají faktoru 1, který byl v datech zapsán do řádků (tedy pro analyzovaná data teplota), „Sloupce“ se týkají sloupcového faktoru 2 (tedy zde pH). Pro oba faktory je v tabulce ANOVA popořadě spočten tzv. součet čtverců (SS sum of squares), počet stupňů volnosti (DF aneb počet kategorií faktoru mínus 1, v české verzi nevýstižně označeno jako „Rozdíl“), průměrný součet čtverců (MS mean sum = SS/DF). Následuje testová statistika F , což je pro každý faktor hodnota jeho MS vydělená hodnotou MS pro položku „Chyba“ (pro „řádkový“ faktor určíme F výpočtem $79,556/4,556=17,463$, pro „sloupcový“ faktor určíme F výpočtem $948/4,556=208,098$). Testovou statistiku F porovnáme s tabulkovou kritickou hodnotou „ F krit“ (poslední sloupec), nebo použijeme k rozhodnutí o příslušném faktoru odpovídající p -hodnotu (předposlední sloupec). Oběma způsoby a pro oba faktory zamítáme příslušnou H_0 . Jak teplotu, tak pH lze vyhodnotit jako statisticky významné faktory, ovlivňující sledovanou veličinu Y („optical density“).

Anova: dva faktory bez opakování					
Faktor	Počet	Součet	Průměr	Rozptyl	
25	3	69	23	237	
30	3	85	28,33333	233,33333	
35	3	105	35	325	
40	3	77	25,66667	166,33333	
pH 5.5	4	60	15	16,66667	
pH 6.5	4	96	24	22	
pH 7.5	4	180	45	50	

ANOVA						
Zdroj variace	SS	Rozdíl (SIC!)	MS	F	Hodnota P	F krit
Řádky	238,66667	3	79,55556	17,463415	0,00228	4,757063
Sloupce	1896	2	948	208,09756	2,870E-006	5,143253
Chyba	27,333333	6	4,555556			
Celkem	2162	11				

V SW STATISTICA v modulu ANOVA volíme metodu „ANOVA hlavních efektů“, což je označení pro naši situaci, v níž sledujeme efekt každého faktoru zvlášť.



Pracujeme s daty v základním tvaru, tj. že závislá veličina Y (zadávaná jako „seznam závislých proměnných“) i oba faktory (zadávané jako „kategor. nezáv. prom. (faktory)“) tvoří tři datové sloupce, každé statistické jednotce odpovídá samostatný řádek. Na dalším obrázku vidíme první tři řádky z databáze, vedle je už předvedeno zadání proměnných pro ANOVU.

	1 temp	2 OptDens	3 pH
1	25	10	pH 5.5
2	30	15	pH 5.5
3	35	20	pH 5.5

Vyberte závislé proměnné a kategor. nezávislé proměnné (faktory):

1 - temp 2 - OptDens 3 - pH	1 - temp 2 - OptDens 3 - pH
-----------------------------------	-----------------------------------

Výsledná tabulka ANOVY (získáme ji volbou „Všechny efekty“) je prakticky identická s výstupem z Excelu. Chybí zde pouze poslední sloupec s kritickou hodnotou (p-hodnota je sama o sobě pro rozhodování o hypotézách postačující). Navíc je zde řádek „Abs.člen“, pomocí nějž testujeme nulovost celkové střední hodnoty Y (bez ohledu na jakékoli kategorie).

Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	9408,000	1	9408,000	2065,171	0,000000
temp	238,667	3	79,556	17,463	0,002280
pH	1896,000	2	948,000	208,098	0,000003
Chyba	27,333	6	4,556		

Specifickou situací, v níž využíváme tento model, je ANOVA v případě jednoho faktoru, ale tzv. s „opakovanými měřeními“ (angl. „repeated measures“, nezaměňovat s pojmem „replications“). Následující příklad je převzatý z knihy Neter - Wasserman - Kutner, Applied Linear Statistical Models (s. 1041). Každý ze šesti ochutnávačů ochutnal čtyři různé druhy vína. Sledovanou veličinou Y je chuťové skóre (možné maximum činí 40).

		j=1	j=2	j=3	j=4
ochutnávač 1	i=1	20	24	28	28
ochutnávač 2	i=2	15	18	23	24
ochutnávač 3	i=3	18	19	24	23
ochutnávač 4	i=4	26	26	30	30
ochutnávač 5	i=5	22	24	28	26
ochutnávač 6	i=6	19	21	27	25

Degustace probíhala ze strany ochutnávačů zaslepeně, v náhodném pořadí, takže až dodatečně bylo příslušné skóre přiřazeno vždy k příslušnému druhu vína ($j=1, \dots, 4$). Cílem je posoudit, zda jsou mezi jednotlivými druhy vína významné rozdíly v chuti. Chybně by bylo aplikovat na uvedená data test ANOVA s jediným faktorem (s rozlišením sloupců). Problém je v tom, že jsme ve 4 porovnávaných kategoriích neměli celkem 24 vzájemně nezávislých hodnot Y. Hodnoty na každém řádku ($i=1, \dots, 6$) patří ke stejné statistické jednotce (k danému ochutnávači) a nemůžou být tedy považovány za vzájemně nezávislé. Správný přístup je považovat data za výsledek situace se dvěma faktory: faktor 1 (řádkový) je osoba ochutnávače, faktor 2 (sloupcový) je druh vína.

ANOVA						
Zdroj variabilit	SS	Rozdíl	MS	F	Hodnota P	F krit
Řádky	173,333333	5	34,6666667	32,5	1,5494E-07	2,90129454
Sloupce	184	3	61,3333333	57,5	1,85377E-08	3,2873821
Chyba	16	15	1,06666667			
Celkem	373,333333	23				

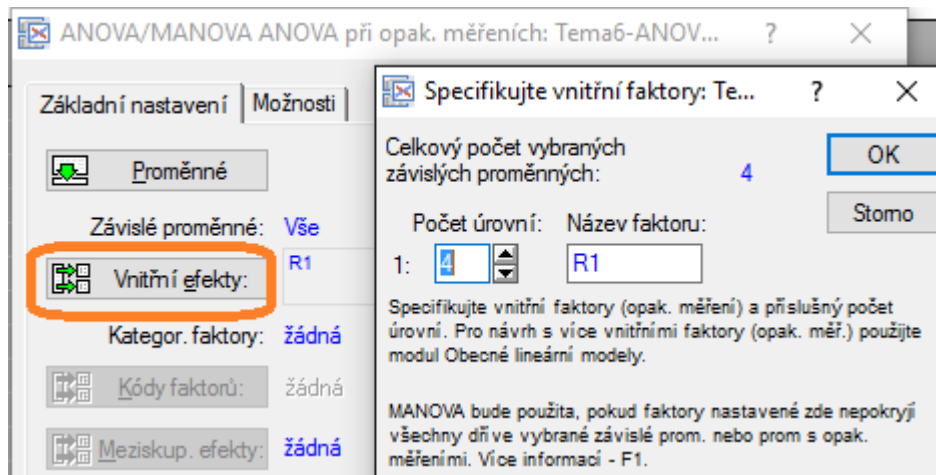
Spuštěním nástroje v Excelu „Anova: dva faktory bez opakování“ získáme tabulku, podle níž (viz údaje ze sloupce HodnotaP) je jak řádkový, tak sloupcový efekt významný. (Pozn: České označení „bez opakování“ může působit v dané situaci pro „repeated measures“ jako matoucí, je tím ale míněno ono anglické „without replications“.) Díky tomu lze konstatovat, že jak mezi jednotlivými ochutnávači, tak mezi jednotlivými druhy vína byly shledány významné rozdíly v hodnocení chuti.

The screenshot shows the 'Obecná ANOVA/MANOVA: Tema6' dialog box. On the left, a data table is visible with columns for wine types (1-4) and rows for judges (judge1-6). The 'ANOVA při opakov. měřeních' option is selected under 'Typ analýzy:'.

K provedení pomocí SW STATISTICA lze využít specifickou volbu „ANOVA při opakovaných měřeních“. Při zadávání proměnných představují sloupce wine1 až wine4 závislé proměnné. (Nepovinný kategoriální faktor bychom zde zadali pouze v případě, že bychom u šesti degustátorů např. rozlišovali jejich pohlaví.)

The screenshot shows the variable selection dialog. The 'Seznam závislých proměnných:' list contains '1-4'. The 'Kategor. nezáv. prom. (faktory):' list is empty.

Stejně jako v Excelu, i zde musíme označit, kolik sloupců představuje "vnitřní efekty", tedy kolik bylo těch "repeated measures" (zde 4); bez tohoto vyznačení by SW situaci analyzoval jako nezávislé soubory. „Název faktoru“ se automaticky vytváří jako R1 (mohl by zde být ale přejmenován např. na „druh vína“).



Výsledná tabulka ANOVY (po volbě „Všechny efekty“) obsahuje v řádce R1 stejné výsledky jako v Excelu (pro srovnání zvýrazněna hodnota součtu čtverců SČ nebo testové statistiky F). Nízké p-hodnoty STATISTICA neuvádí ve formátu jako Excel (např. E-08), výraz 0,000000 zde označuje fakt, že p-hodnota vyšla menší než 10^{-6} .

Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	13442,67	1	13442,67	387,7692	0,000006
Chyba	173,33	5	34,67		
R1	184,00	3	61,33	57,5000	0,000000
Chyba	16,00	15	1,07		

Poznámka k zamyšlení: Pokud bychom v předešlé úloze uvažovali sloupcový faktor pouze se dvěma kategoriemi (tedy pokud bychom v předešlé úloze porovnávali jen dva druhy vína), o jakou situaci by se vlastně jednalo, jakou jinou metodou bychom pak mohli postupovat? Nápověda: Pročtěte si krátký úvodní odstavec k celé kapitole 6.2.

6.2.3 ANOVA – dvojné třídění s opakováním

Situace je obdobná jako v případě dvojného třídění bez opakování, ale zde pro každou kombinaci faktorů zaznamenáváme více hodnot Y. Pro jednoduchost uvažujeme pouze tzv. vyváženou situaci, tedy případ, kdy je počet hodnot v každé kombinaci faktorů shodný.

Jako ilustrační data použijeme příklad z internetového odkazu J. H. McDonald, Handbook of Biological Statistics (<http://www.biostathandbook.com/twowayanova.html>). Celkem 24 různých jedinců téhož živočišného řádu (Amphipoda - různonožci) bylo rozlišováno podle tří různých genotypů (faktor 1) a pohlaví (faktor 2). V každé kombinaci obou faktorů jsme měli shodně po čtyřech živočiších.

Na obrázku v jeho levé části je standardní zápis původních dat. Každému živočichu odpovídá samostatný řádek. Sledovaná závislá veličina Y (hodnoty ve sloupci se zkratkou MPI) je následována sloupcem s rozlišením faktoru 1 - genotyp (kategorie FF, FS, SS), v posledním sloupci je zaznamenán faktor 2 - pohlaví (kategorie f, m).

animal n.	enzym.aktivita mannose-6-phosphate isomerase (MPI)	MPI genotyp	male / female	RE-ARRANG.:	female	male	
1	2,838	FF	f		2,838	1,884	
2	4,216	FF	f		4,216	2,283	
3	2,889	FF	f		2,889	4,939	
4	4,198	FF	f		4,198	3,486	
5	1,884	FF	m		3,550	2,396	
6	2,283	FF	m		4,556	2,956	
7	4,939	FF	m		3,087	3,105	
8	3,486	FF	m		1,943	2,649	
9	3,550	FS	f		3,620	2,801	
10	4,556	FS	f		3,079	3,421	
11	3,087	FS	f		3,586	4,275	
12	1,943	FS	f		2,669	3,110	
13	2,396	FS	m				
14	2,956	FS	m				
15	3,105	FS	m				
16	2,649	FS	m	Anova: dva faktory s opakováním			
17	3,620	SS	f				
18	3,079	SS	f	Faktor	female	male	Celkem
19	3,586	SS	f	FF			
20	2,669	SS	f	Počet	4	4	8
21	2,801	SS	m	Součet	14,141	12,592	26,733
22	3,421	SS	m	Průměr	3,53525	3,148	3,341625
23	4,275	SS	m	Rozptyl	0,602151583	1,889282	1,110603696
24	3,110	SS	m				

Pro využití připraveného nástroje je v Excelu nejprve nutné data upravit do podoby znázorněné na tomtéž obrázku v pravé horní části. V řádcích pod sebou uvedeme hodnoty Y ze stejné kategorie faktoru 1. Třídění podle faktoru 2 tvoří samostatné sloupce. Pro takto připravená data můžeme spustit nástroj „Anova: dva faktory s opakováním“ (v anglické verzi „with replications“). Jako „vstupní oblast“ zadáme data v upravené podobě, včetně nadpisů. Jako hodnotu „řádků na výběr“ zadáme zde číslo 4 (obecně počet měření pro každou kombinaci faktorů). Výsledkem jsou podrobné deskriptivní charakteristiky pro jednotlivé kategorie (část je viditelná na stejném obrázku vpravo). Stěžejní pro analytický účel je ale až poslední část výstupu – tabulka ANOVA. Struktura tabulky ANOVA je obdobná jako v případě dvojnásobného třídění bez opakování. Místo „řádků“ se zde pro faktor 1 objevuje pojem „výběr“, pro faktor 2 zůstává označení „sloupce“. Díky opakováním v rámci každé kombinace faktorů však nyní kromě faktorů samotných lze navíc posoudit také vliv tzv. **interakce**, kdy Y může záviset na jednotlivých kombinacích obou faktorů (v ilustračním příkladu zmíněném v úvodu kapitoly 6.2.1 bychom tak u platu mohli zjišťovat nejen efekt vzdělání a efekt pohlaví, ale i to, zda jsou rozdíly mezi vzdělanostními kategoriemi jiné u mužů než u žen). Pro analyzovaná data vyšly všechny p-hodnoty větší než 0,05. Znamená to, že se nepodařilo prokázat vliv ani faktoru 1, ani faktoru 2, ani jejich vzájemné interakce na sledovanou veličinu Y. U různonožců tedy nebyl zjištěn statisticky významný vliv ani genotypu, ani pohlaví, ani jejich vzájemné interakce na hodnoty MPI.

ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Výběr	0,483851083	2	0,241925542	0,331750147	0,721962541	3,554557146
Sloupce	0,356728167	1	0,356728167	0,489177872	0,49323432	4,413873419
Interakce	0,511610583	2	0,255805292	0,350783313	0,708843965	3,554557146
Dohromady	13,1263235	18	0,729240194			
Celkem	14,47851333	23				

Při aplikaci SW STATISTICA pracujeme s daty v původním tvaru (tedy na 24 řádcích). V menu volíme metodu „ANOVA s interakcemi“. Dle výstupu můžeme zhodnotit efekt jak každého faktoru zvlášť (v tabulce řádky „genotyp“, resp. „mf“), tak i efekt jejich vzájemné interakce (řádek „genotyp*mf“).

Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	250,4930	1	250,4930	343,4986	0,000000
genotyp	0,4839	2	0,2419	0,3318	0,721963
mf	0,3567	1	0,3567	0,4892	0,493234
genotyp*mf	0,5116	2	0,2558	0,3508	0,708844
Chyba	13,1263	18	0,7292		

Pro samostatné vyzkoušení necht' čtenář analyzuje data z výzkumu Z. Čerovského (FZS UJEP, Laboratoř pro studium pohybu). Bylo provedeno celkem 54 vzájemně nezávislých měření přesností kalibrace. Sledovanou mírou přesnosti Y byla hodnota tzv. reziduálu. Předpokládané faktory, které mohou přesnost kalibrace ovlivnit, jsou metoda kalibrace (V nebo R) a dále počet použitých kamer (5, 8 nebo 11). Má některý z nich, nebo jejich interakce, významný vliv na přesnost?

	reziduál	5 kamer	8 kamer	11 kamer
kalibrace V	0,727514	0,580964	0,582265	
	0,657894	0,570824	0,604557	
	0,775170	0,564699	0,616352	
	0,686344	0,525669	0,515885	
	0,622298	0,584919	0,599785	
	0,700272	0,522976	0,544850	
	0,683006	0,572774	0,482972	
	0,695102	0,593798	0,499552	
	0,721288	0,595784	0,496679	
kalibrace R	0,884018	0,670863	0,718277	
	0,896706	0,655316	0,738539	
	0,850614	0,704349	0,729805	
	0,829632	0,658905	0,706587	
	0,823400	0,652056	0,694139	
	0,883152	0,658378	0,669165	
	0,798450	0,714216	0,632251	
	0,794676	0,652140	0,702103	
	0,828720	0,640940	0,677002	

6.3 Velikost efektu a síla testu

Vraťme se k úloze s ochutnávkou vína. Metodou ANOVA pro opakovaná měření jsme zjistili významný efekt druhu vína, jinak řečeno, mezi jednotlivými druhy jsou významné rozdíly v chuti. Výstup ze SW STATISTICA umožňuje změřit i **velikost efektu** (angl. „effect size“). Mírou je zde tzv. Eta-kvadrát, které se interpretuje v procentech celkové variability dat. Cohenovo pravidlo (též „rule of thumb“) říká, že hodnota vyšší než

- 0,02 odpovídá nízkému efektu,
- 0,13 odpovídá středně velkému efektu,
- 0,26 odpovídá velkému efektu.

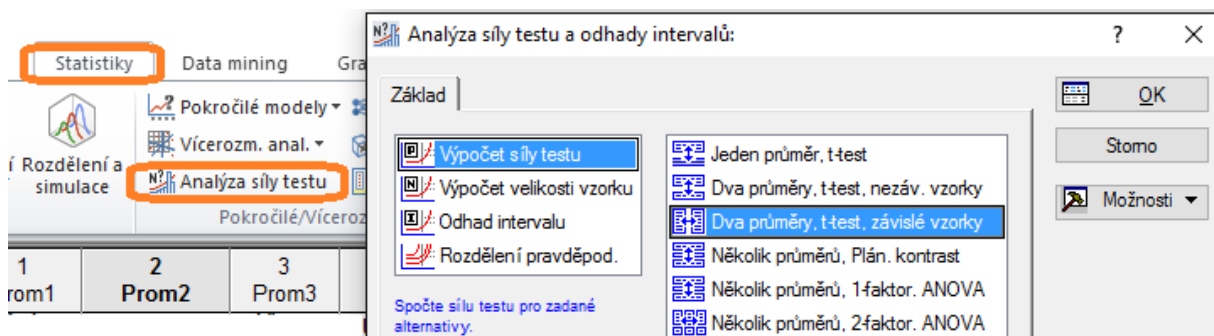
(Přehled pravidel pro různé míry lze najít např. zde: <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>.) V úloze s ochutnáváním vína vysvětlují odlišnosti mezi jednotlivými čtyřmi druhy vína celkovou variabilitu dat z 92 %. Jde tedy skutečně o podstatný faktor.

	p	Parciál. éta-kvadr.	Výstřednost	Pozor. síla (alfa=0,05)
	0,000006	0,987270	387,7692	1,000000
	0,000000	0,920000	172,5000	1,000000

Další charakteristikou, kterou SW STATISTICA poskytuje, je **síla testu**. Je definovaná jako pravděpodobnost správného zamítnutí H_0 :

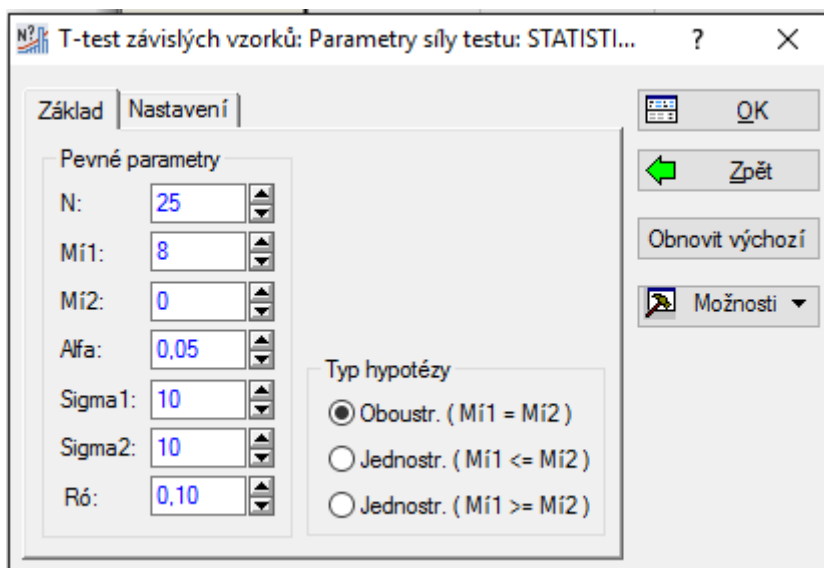
$$1-\beta = 1-P(\text{chyby 2. druhu}) = P(\text{zamítnutí } H_0 \mid \text{non}H_0)$$

Při pevně zvolené hladině významnosti by správně zvolený test měl být ten nejsilnější možný, takže výsledná 100% síla zde není nic mimořádného.

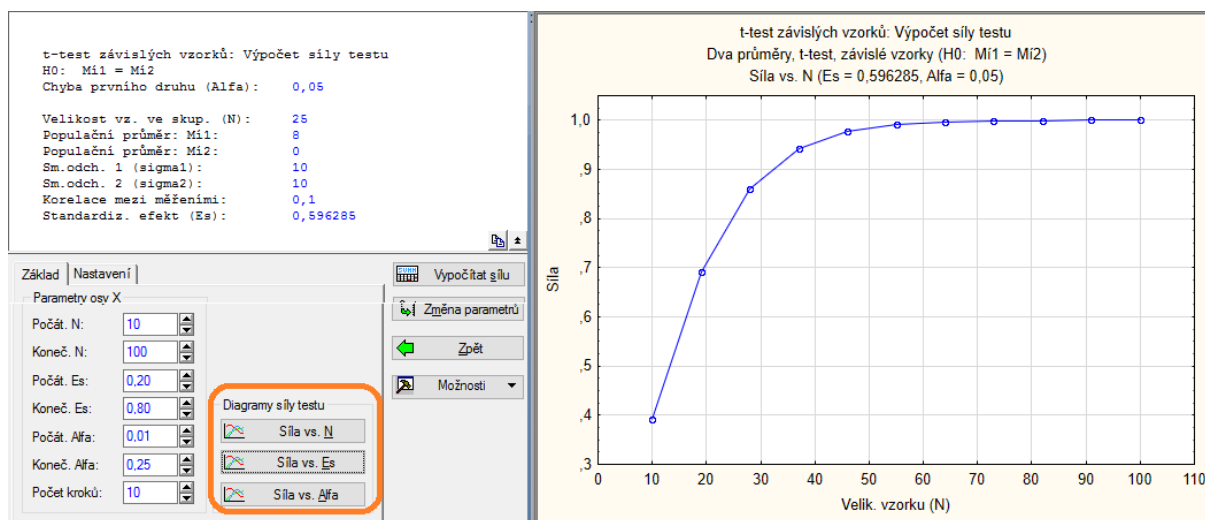


STATISTICA umožňuje v modulu „Analýza síly testu“ spočítat sílu testu pro mnoho různých statistických metod. Na obrázku je naznačen způsob spuštění v situaci, kdy bychom chtěli určit sílu v případě párového t-testu. Po otevření menu jsou přednastaveny parametry, které můžeme dle potřeby měnit (viz další obrázek): počet statistických jednotek (N), předpokládaná střední hodnota pro první veličinu (M_1) a pro druhou veličinu (M_2), hladina významnosti (Alfa), směrodatná odchylka pro první veličinu (Σ_1) a pro druhou veličinu (Σ_2), vzájemná korelace mezi oběma veličinami (Ró) a formou výběru také požadovaný typ hypotézy (oboustranná, resp. obě možné jednostranné).

Poznámka: Ve značení podle kapitoly 6.1.1 je střední rozdíl μ_d nyní nahrazen dvojicí M_1 a M_2 , přičemž (až na znaménko) $\mu_d = M_1 - M_2$.



Po zadání parametrů spustíme (OK) výpočet. Výsledky nabídnou kromě jiného grafické znázornění závislosti síly na počtu statistických jednotek (N), na velikosti efektu (Es) nebo na hladině významnosti (alfa). V praxi bývá jako minimální doporučená síla testu hodnota 0,7. Z výsledného obrázku vidíme, že za daných předpokladů (tedy při hodnotách, které jsme zadali jako parametry) bude párový t-test dostatečně silný již při zhruba N=20. Podobné úvahy provádíme ještě před samotnými experimenty ve fázi jejich přípravy, abychom naplánovali dostatečné rozsahy pro statistické šetření (nebo naopak abychom nenaplánovali zbytečně rozsáhlá šetření). Předpokládá to ovšem mít alespoň hrubou představu o chování analyzovaných veličin, abychom vůbec mohli zadat potřebné parametry.



V praxi často využíváme tzv. **Cohenovo delta** (d). Tato charakteristika slouží ke změření velikosti rozdílů v číselné veličině mezi dvěma porovnávanými skupinami. Obvykle rozhodujeme podle pravidla, že výsledná hodnota nad

- 0,20 odpovídá nízkému efektu,
- 0,50 odpovídá středně velkému efektu,
- 0,80 odpovídá velkému efektu.

K výpočtu lze doporučit např. tuto webovou aplikaci:

<http://www.socscistatistics.com/effectsize/Default3.aspx>

(kde jsou uvedeny i další alternativní míry pro různé situace).

Velikost efektu chceme změřit i v jiných případech, než je porovnání střední hodnoty číselné veličiny ve dvou či více skupinách. U kategoriálních veličin, kde pracujeme s četnostmi výskytů sledovaných jevů, lze využít **poměr šancí** („odds ratio“, OR). Např. při házení klasickou hrací kostkou označme jako jev A hození šestky s pravděpodobností $P(A)=1/6$ a jako jev B hození sudého čísla s pravděpodobností $P(B)=3/6=1/2$. **Šance** („odds“) výskytu jevu A je obecně definována jako

$$O(A) = P(A) : (1-P(A)),$$

takže konkrétně pro hození šestky bude po dosazení a zkrácení

$$O(A) = (1/6) : (5/6) = 1 : 5,$$

čteme „šance jedna ku pěti“. V podmíněné verzi sledujeme šanci daného jevu za podmínky, že nějaký jiný jev nastal:

$$O(A|B) = P(A|B) : (1-P(A|B)).$$

Jelikož konkrétně pro již definovaná A a B dostáváme $P(A|B) = 1/3$ (pokud víme, že na kostce bylo hozeno sudé číslo, pak pravděpodobnost, že se jednalo právě o šestku, je 1/3), takže po dosazení a zkrácení bude:

$$O(A|B) = (1/3) : (1-1/3) = (1/3) : (2/3) = 1 : 2.$$

Pokud tedy na kostce hodíme sudé číslo (podmínka B), pak šance na to, že oním hozeným číslem byla právě šestka (jev A) je jedna ku dvěma. Máme-li zaveden pojem podmíněná šance, poměr šancí (OR) můžeme definovat jako

$$OR(A) = O(A|B) / O(A|nonB).$$

Příklad (Zvárová, J., *Základy statistiky pro biomedicínské obory*): Díky epidemiologickým studiím bylo zjištěno, že šance výskytu rakoviny plic u kuřáků je 5 : 4, zatímco šance výskytu rakoviny plic u nekuřáků je 1 : 8. Jak velký efekt má kouření na vznik rakoviny plic?

Řešení: Sledovaným jevem A je zde výskyt rakoviny plic, podmínkou (jev B) je kuřáctví (nekuřáctví je tudíž jeho doplněk nonB). Zadané údaje odpovídají podmíněným šancím:

$$O(A|B) = 5 : 4, \quad O(A|nonB) = 1 : 8$$

takže poměr šancí bude po dosazení a úpravě složeného zlomku:

$$OR(A) = (5 : 4) / (1 : 8) = 10.$$

U kuřáků (za podmínky B) je tedy šance na vznik rakoviny plic 10-krát vyšší než u nekuřáků (za absence podmínky B).

Poznámka: Obdobnou charakteristikou je tzv. **věrohodnostní poměr** („likelihood ratio“), definovaný nikoli jako poměr podmíněných šancí, ale jako poměr podmíněných pravděpodobností.

Máme-li dvojici kategoriálních veličin zpracovanou formou kontingenční tabulky, můžeme vzájemnou nezávislost obou veličin otestovat metodou chí-kvadrát (příslušnou testovou charakteristiku označme jako χ^2). Jejich vzájemnou závislost pak změříme např. pomocí **Pearsonova koeficientu (C)**:

$$C = \sqrt{(\chi^2 / (n + \chi^2))}$$

Připomeňme, že vzájemnou závislost spojitých veličin lze měřit pomocí **korelačních koeficientů** (klasický korelační koeficient se také nazývá Pearsonův). Obecně pro tyto míry platí, že hodnoty poblíž 0 odpovídají vzájemné nezávislosti a naopak hodnoty blíží se k maximu, tedy k jedné (bez případného ohledu na znaménko), odpovídají závislosti analyzovaných proměnných.

6.4 Test rovnosti parametrů u dvou alternativních rozdělení

Alternativním rozdělením se řídí veličina (označovaná též jako dichotomická), u které rozlišujeme pouze dvě kategorie – sledovaný jev buď nastal, nebo nenastal. Příkladem mohou být veličiny kuřáctví (kategorie kuřák versus nekuřák, sledovaným jevem je výskyt kuřáka), nebo pohlaví (kategorie muži versus ženy, za sledovaný jev lze považovat např. výskyt muže). Někdy uvedené situace zkombinujeme a sledujeme výskyt nějakého jevu ve dvou podpopulacích, např. výskyt kuřáků u mužů a u žen. Označme jako P_1 pravděpodobnost sledovaného jevu (kuřáctví) v první podpopulaci (mezi muži) a jako P_2 pravděpodobnost téhož sledovaného jevu (kuřáctví) ve druhé podpopulaci (mezi ženami). Testovanou nulovou hypotézou je pak test rovnosti pravděpodobnostního parametru

$$H_0: P_1 = P_2$$

oproti oboustranné (případně lze i jednostranné) alternativě $H_1: P_1 \neq P_2$. Jinak řečeno, ověřujeme, zda je či není výskyt sledovaného jevu (kuřáctví) stejný v obou porovnávaných podpopulacích (u mužů a u žen).

Příklad (zdroj J. Zvárová, *Základy statistiky pro biomedicínské obory*, s.133): Na základě výsledků populační studie (data viz tabulka) ověřte na 5% hladině významnosti, zda je procentuální výskyt osob s antigenem HLA-DR4 stejný u nemocných progresivní polyartritidou a u zdravých kontrolních jedinců.

antigen HLA-DR4:	pozitivní	negativní	CELKEM
progres. polyartritida	46	28	74
zdraví (kontroly)	50	184	234
CELKEM	96	212	308

Označme jako p relativní četnost sledovaného jevu (v příkladu tedy výskyt antigenu) v celé populaci, p_1 v první podpopulaci (u nemocných polyartritidou) a p_2 ve druhé podpopulaci (u kontrolní skupiny):

$$p = 96/308 = 0,31 \quad p_1 = 46/74 = 0,62 \quad p_2 = 50/234 = 0,21$$

Dále označme jako n_1 velikost první a n_2 druhé podpopulace ($n_1 = 74$, $n_2 = 234$). Test výše uvedené nulové hypotézy lze provést prostřednictvím Z -statistiky (tzv. **Z-test**):

$$Z = (p_1 - p_2) \cdot \sqrt{[n_1 n_2 / (p(1-p)(n_1 + n_2))]}.$$

Kritickou hodnotou je kvantil normovaného normálního rozdělení, obvykle hodnota 1,96.

Příklad (pokračování): Dosazením dostaneme

$$Z = (0,62 - 0,21) \cdot \sqrt{[74 \cdot 234 / (0,31(1 - 0,31)(74 + 234))]} = 6,60.$$

Jelikož výsledná Z-statistika 6,60 překročila kritický kvantil 1,96, zamítáme na 5% hladině hypotézu H_0 o rovnosti parametru v obou podpopulacích. Pravděpodobnost výskytu sledovaného antigenu se v obou skupinách významně liší.

Poznámka 1: Předpokladem funkčnosti Z-testu je, aby jak hodnota $n_1p_1(1-p_1)$, tak $n_2p_2(1-p_2)$ byla větší než 9.

Poznámka 2: K výpočtu lze doporučit opět webovou aplikaci <http://www.socscistatistics.com/tests/ztest/Default.aspx>

Zpracovávaná data byla zaznamenána ve tvaru tzv. kontingenční tabulky. Ta se obvykle vyskytuje v souvislosti s **testem chí-kvadrát nezávislosti**. To není náhoda. Hypotézu, zda se výskyt antigenu shoduje či naopak liší u nemocných polyartritidou a u zdravých kontrol, můžeme přece formulovat také tak, že nás zajímá, zda výskyt antigenu nezávisí, resp. závisí na výskytu sledovaného onemocnění.

Poznámka 3: Aplikace testu chí-kvadrát pro tatáž data musí skončit zcela stejným rozhodnutím jako aplikace Z-testu. Pokud k rozhodnutí nepoužijeme přímo p-hodnotu, ale získáme testovou statistiku, bude mít tato přibližně hodnotu $T=43,61$. Kdybychom výpočty nezaokrouhlovali, přesvědčili bychom se, že mezi statistikou Z-testu (Z) a statistikou testu chí-kvadrát (T) platí přesně vztah $T=Z^2$. Analogický vztah platí i pro odpovídající kritické hodnoty: v situaci se čtyřpolní kontingenční tabulkou je touto hodnotou kvantil rozdělení chí-kvadrát při 1 stupni volnosti, tedy hodnota 3,84, přičemž platí: $3,84=1,96^2$.

Při porovnávání parametrů dvou alternativních rozdělení můžeme změřit velikost efektu vypočtením **intervalu spolehlivosti pro rozdíl populačních proporcí**. Zaveďme

$$s = \sqrt{[p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2]}$$

Hledaným oboustranným intervalem spolehlivosti je potom

$$(p_1-p_2) \pm 1,96 \cdot s$$

Příklad (dokončení): Dosazením dostaneme

$$s = \sqrt{[0,62(1-0,62)/74 + 0,21(1-0,21)/234]} = 0,06$$

$$(0,62-0,21) \pm 1,96 \cdot 0,06 = (0,29 ; 0,53)$$

Pravděpodobnost výskytu antigenu HLA-DR4 je s 95% spolehlivostí o 29-53 % (procentních bodů) vyšší u nemocných progresivní polyartritidou než u zdravých kontrolních jedinců.

Pro samostatné procvičení může čtenář zpracovat data, s úpravou převzatá opět z knihy J. Zvárová, *Základy statistiky pro biomedicínské obory* (s.135). Dobrovolníci byli před chřipkovou epidemií očkovaní (zaslepeně) buď testovanou vakcínou, nebo kontrolně jen placebem. Následně bylo zjištěno, u koho chřipka vypukla a u koho nikoli, viz tabulka. Zjistěte, zda má vakcína protektivní efekt a pokud ano, jak velký.

chřipka:	ano	ne	
vakcína	20	220	240
placebo	80	140	220
	100	360	460