

Analýza dat v neurologii

XXIV. Vybrané příklady hodnocení kategoriálních dat

V tomto díle seriálu pokračujeme ve výkladu hodnocení kategoriálních dat a v sérii příkladů navážeme na předchozí části. Příklady nejsou jen opakováním již vysvětlených témat, ale ukazují i nové přístupy k analýze kontingenčních tabulek (díl XXI). Tyto podněty dodali sami čtenáři našeho seriálu, a tudíž jednotlivá témata probereme systémem otázek a odpovědí. Příklady přibližují analýzy, které většinou nejsou běžně dostupné v nabídce statistických software anebo jejich aplikace vyžaduje určitou znalost a invenci na straně uživatele. Proto jsme výpočty v příkladech doplnili i o matematické vztahy, jež umožní vyzkoušet si výpočet v nějakém dostupném nástroji. Na své si ale, doufejme, přijdou i čtenáři, kteří nemají ambici dané problémy sami hodnotit. Každý příklad je doplněn zadáním problému, zdůvodněním výpočtu a interpretací výsledku.

Otázka 1. Běžné tabulky četností v podstatě dávají do kontrastu výskyt dvou znaků a sledují, zda je mezi kategoriemi určitý vztah, nebo ne. Můžeme ale nějak kvantifikovat odhad četností kategorií a rozdílů mezi nimi?

Odpověď je jednoznačně ano. Nezapomeňme, že tabulku četností (např. o velikosti 2×2) získáváme výběrem N jedinců, u kterých sledujeme výskyt dvou znaků. Jde tedy o „klasický“ náhodný výběr z cílové populace, a cokoli z tabulky spočítáme,

je statistický odhad skutečnosti v oné cílové populaci. I zde platí, že čím máme větší velikost vzorku (N), tím je náš odhad přesnější, a tedy s větší pravděpodobností vystihuje realitu. V tab. 1 to dokumentujeme na příkladu sledování výskytu dýchacích problémů u kuřáků a nekuřáků. Při samotném výběru N jedinců ovšem nesmíme předjímat žádný vztah obou proměnných a výběr musí být zcela náhodný; obě proměnné jsou si rovnocenné.

Kromě obecného hodnocení tabulky pomocí χ^2 testu má samozřejmě význam i parciální analýza výskytu dýchacích problémů mezi kuřáky a nekuřáky. Anebo naopak, srovnání relativního zastoupení kuřáků mezi osobami zdravými a osobami s problémy. Tabulka poskytuje číselné odhady těchto relativních četností. Odborně hovoříme o bodových odhadech. Pro jejich rozdíl je pak možné získat interval spolehlivosti, který má vysokou interpretační hodnotu. Šířka intervalu indikuje spolehlivost provedených odhadů a umožňuje i srozumitelnou prezentaci zjištěných rozdílů. Výpočet takových intervalů přibližuje příklad 1a, b.

Otázka 2. V klinické praxi jsou často mezi sebou srovnávány nežádoucí účinky léčebných metod. Ty mohou být sice hodnoceny jako binární proměnné (ano/ne, presence/absence), ale častější je jejich hodnocení na určité škále (grade).

L. Dušek, T. Pavlík,
J. Jarkovský, 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

Existují pro tento problém statistické testy?

Předpokládejme, že se jedná o srovnání v tzv. párovém uspořádání experimentu. Pacientovi je podáván lék 1 a jsou sledovány nežádoucí účinky, poté je lék vysazen a stejné sledování následuje u léku 2. Jde o běžný model uplatňovaný v klinických studiích, např. fáze I nebo II. Příkladem může být sledování ospalosti nebo únavy u léků na sezónní alergie. Pokud u nežádoucích účinků sledujeme prostý výskyt (ano/ne), pak lze pro srovnání léků doporučit tzv. Q-test podle Cochran, jehož výhodou je, že výpočet přijme i více než dva srovnávané léky. Cochranův test jsme již představili v předchozím díle seriálu jako rozšíření McNemarova testu. Zde uvádíme další, více klinickou aplikaci výpočtu (příklad 2).

Avšak v případech, že jsou nežádoucí účinky zaznamenávány jako stupnice více hodnot odrážející obtížnost problémů (grade), již nevystačíme s McNemarovým

Tab. 1. Příklad tabulky četností dvou binárních proměnných nabývajících hodnot + nebo –.

		Proměnná II dýchací problémy: ano/ne		Řádkové součty	Kromě vlastního hodnocení vztahu proměnných lze z tabulky počítat sloupcové a řádkové relativní četnosti, např.: $p_{a a} = a/(a + c)$ nebo $p_{a a} = a/(a + b)$. Tyto odhady relativních četností lze vzájemně srovnávat a pro jejich rozdíly hodnotit intervaly spolehlivosti.
		+	–		
Proměnná I kuřák: ano/ne	+	a	b	a + b	
	–	c	d	c + d	
Sloupcové součty		a + c	b + d	N = a + b + c + d	

ani s Cochranovým testem. Oba testy vyžadují na vstupu binární neboli dichotomickou proměnnou. Naproti tomu škála vyjadřující stupňovitě míru potíží vytváří ordinální proměnnou. Pro párové srovnání založené na ordinální proměnné byl navržen tzv. Bowkerův test. Jeho postup zpracovává $K \times K$ tabulku četností, kde jsou obě proměnné vzájemně závislé a jsou zaznamenávány více než dvěma kategoriemi. Označením $K \times K$ zdůrazňujeme, že musí jít o tabulku „čtvercovou“, tedy se stejným počtem řádků a sloupců. Při $K = 2$ je Bowkerův test redukován na již vysvětlený McNemarův test. Výpočet Bowkerova testu přibližujeme v příkladu 3.

Všechny výše zmíněné testy jsou pro hodnocení klinických dat velmi užitečné. Léčebné postupy jsou často sledovány vzájemně závisle na stejných kohortách pacientů. Hodnocení jejich účinnosti a bezpečnosti je běžně prováděno na nominální nebo ordinální škále, nejruznější skóre „severity“ existují ve většině klinických disciplin. Přesto tyto testy nejsou běžně dostupné v nabídce řady počítačových programů. Jak ale ukazují příklady 2 a 3, výpočet není složitý a lze jej zvládnout i bez specializovaných softwarových nástrojů. Dobře využitelné kalkulátory obou testů se také dají nalézt na internetu, kde v anglickém jazyce hledáme postupy vyhovující tomuto popisu: „Hypothesis tests evaluating changes in matched paired data with binomial or multinomial responses“. V našem jazyce ještě doplníme, že Cochranův i Bowkerův test jsou tzv. testy symetrie tabulky četností.

O prof. W. G. Cochranovi (1909–1980) jsme se již zmínili v předchozím díle seriálu. Rovněž jsme uvedli, že McNemarův test byl publikován v roce 1947. Bowkerův test, nazvaný po svém autorovi, významném americkém statistikovi A. H. Bowkerovi, následoval bezprostředně poté v roce 1948. I přes své stáří nejsou ale tyto testy uzavřenou kapitolou. V recentní literatuře stále probíhá diskuze nad novými aplikacemi těchto postupů, řeší se jejich uplatnění v klinických studiích s různými vzorkovacími plány a jsou vyvíjeny počítačové simulace umožňující podchytit i vzdálenost mezi body ordinální stupnice v Bowkerově testu. Některé zajímavé od-

kazy uvádíme níže v přehledu literatury. Mimochodem sám autor testu k tomuto výzkumu dlouhá léta přispíval. Test totiž publikoval jako relativně mladý vědec ve věku 29 let. Prof. Bowker zemřel nedávno, v roce 2008, ve věku 88 let.

Otázka 3. Můžeme z tabulky četností hodnotící výskyt dvou znaků usuzovat na jejich korelaci?

Tomuto problému jsme se krátce věnovali již v díle XXI, nicméně pouze výčtem využitelných metod. Korelace je pojem široký, jehož platnost nelze omezit jen na kategoriální data, týká se zejména dat spojitých. Obecně pokud zkoumáme vztah dvou nominálních znaků (ano/ne), hovoříme spíše o významné závislosti nebo asociaci v jejich výskytu. Pojem korelace značí již kvantitativní vyjádření síly takového vztahu a používá se spíše při hodnocení vztahu dvou ordinálních znaků. K měření síly vztahu dvou ordinálních znaků slouží tzv. pořadové koeficienty korelace, kterým se budeme podrobněji věnovat v připravovaných dílech seriálu. Zde jmenujme pouze Kendallův koeficient tau neboli Kendallův korelační koeficient jako míru často využívanou při analýze kontingenčních tabulek. Ukázkou jeho výpočtu přináší příklad 4.

Otázka 4. Existuje vzájemná zastupitelnost testů používaných pro hodnocení kontingenčních tabulek?

Jistá zastupitelnost existuje, vždy ale pouze při řešení stejné „třídy“ problémů. Nelze zaměňovat testy určené pro párové (závislé) sledované proměnné s testy pro zcela nezávislé vzorkování výskytu dvou proměnných. Tedy například Fisherův exaktní test (nezávislé uspořádání) nelze v žádném případě zaměnit za McNemarův test (párové uspořádání). A to i přesto, že jsou často ve statistických počítačových programech nabízeny spolu a lze je spočítat současně na stejných datech. Rovněž nelze zaměňovat testy určené pro binární (binomická) data a pro ordinální stupnice. Pokud tato základní pravidla respektujeme, lze někdy pro ověření stanovené hypotézy zvolit i více testů, které jsou alternativně použitelné na stejných datech. Vždy je však nutné sledovat, jakou hypotézu testy hodnotí, a jaká

je tedy jejich faktická interpretace; málokdy jde o skutečnou alternativu testování se stejným významem.

Využijeme tohoto problému k menšímu připomenutí tzv. neparametrických testů pro hodnocení dvou pokusných zásahů (díle XVII seriálu). Tyto metody totiž také umožňují pracovat s tabulkami četností vzešlymi z nezávislého (Mannův-Whitneyho U test) nebo z párového (Wilcoxonův test) uspořádání experimentu. Problém ale redukuje na srovnání výskytu jedné proměnné o dvou kategoriích, jejichž vliv hodnotíme jako „pokusný zásah“. Příklady 5 a 6 připomínají způsob výpočtu těchto testů.

Otázka 5. Je McNemarův test využitelný pro vzájemné srovnání účinnosti laboratorních metod?

Jednoznačně ano, pokud tuto účinnost testujeme striktně párově (tedy obě metody jsou paralelně testovány na stejné sadě vzorků) a sledovanou proměnnou hodnotíme jako binární kód (pozitivita/negativita výsledku; záchyt 1/0; překročena nějaká patologicky kritická hranice ano/ne; apod.). McNemarův test je především test symetrie tabulky četností a sleduje tedy, zda se metody vzájemně liší v četnosti výsledků, kde se neshodují. V příkladu 7 uvádíme ještě alternativní výpočet pomocí znaménkového testu a modelu binomického rozdělení, které doplňují pravděpodobnost, s níž bychom daný výsledek pozorovali zcela náhodně. Můžeme tak přímo posoudit, zda je neshoda obou metod pravděpodobnostně výjimečná, nebo ne.

Literatura

Bowker AH. A Test for Symmetry in Contingency Tables. *J Am Stat Assoc* 1948; 43(244): 572–574.
 McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; 12(2): 153–157.
 Krampe A, Kunt S. Bowker's test for symmetry and modifications within the algebraic framework. *Comput Stat Data Anal* 2007; 51(9): 4124–4142.
 May WL, Johnson WD. Symmetry in square contingency tables: tests of hypotheses and confidence interval construction. *J Biopharm Stat* 2001; 11(1–2): 23–33.
 Von Eye A, Spiel C. Standard and Nonstandard Log-Linear Symmetry Models for Measuring Change in Categorical Variables. *Am Stat* 1996; 50(4): 300–305.

XXIV. VYBRANÉ PŘÍKLADY HODNOCENÍ KATEGORIÁLNÍCH DAT

Tabulka četností 2 × 2

Prostředí \ Myši	Problémy ano	Problémy ne	Celkem
toxické látky +	15 (62,5 %)	9 (37,5 %)	24 (100,0 %)
toxické látky –	10 (40,0 %)	15 (60,0 %)	25 (100,0 %)
Celkem	25 (51,0 %)	24 (49,0 %)	49 (100,0 %)

Pozorování v tabulce jsou zcela nezávislá, hovoříme tedy o tzv. nezávislém uspořádání experimentu. Oba znaky byly při náhodném výběru $N = 49$ myší uvažovány jako rovnocenné. Jejich možný vztah je předmětem výzkumu.

- U dvou skupin myší je testován vztah mezi životem v toxickém prostředí a výskytem určitých zdravotních problémů. Data mohou být příkladem řízeného experimentu i sledování zvířat odchycených v přírodě na různých lokalitách.
- Pomocí binomického testu pro dva výběry testujeme hypotézu shody podílu jedinců s problémy mezi oběma pokusnými skupinami. Nulová hypotéza: obě skupiny myší se neliší. Rozdíl v zastoupení jedinců s problémy je popsán a doplněn o interval spolehlivosti. Jde o interval spolehlivosti odhadu rozdílu relativních četností.
- Binomický test pro dva výběry je dán vztahem:

$$Z = \frac{|p_1 - p_2|}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}}$$

$$\bar{p} = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}$$

kde p_1 a p_2 jsou relativní četnosti jedinců s problémy v experimentálních skupinách (0,625; 0,400), n_1 a n_2 počet jedinců v experimentálních skupinách (24; 25).

- Po dosazení do vztahu v bodě 3 vypočítáme hodnotu testové statistiky $Z = 1,575$.
- Testová statistika má standardizované normální rozdělení, vypočtená hodnota testové statistiky je porovnána s kritickou hodnotou testu pro $\alpha = 0,05$: $Z = 1,96$.

- Výsledná p-hodnota statistické významnosti je $p = 0,115$ – nezamítáme tedy H_0 , experimentální skupiny se statisticky významně neliší v relativní četnosti jedinců se zdravotními problémy.

- Rozdíl v srovnávaných relativních četnostech je 0,225; interval spolehlivosti (IS) pro tento odhad je dán vztahem:

$$(p_1 - p_2) \pm Z_{(1-\alpha/2)} \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}$$

- Po dosazení získáme rozdíl a jeho 95% IS: 0,225 (–0,051; 0,458).
- Závěr:** 95% interval spolehlivosti pro odhad rozdílu relativních četností zahrnuje hodnotu 0, toto zjištění je ekvivalentní s nezamítnutím nulové hypotézy na hladině významnosti $\alpha = 0,05$.

Příklad 1a. Odhad rozdílu relativních četností a jeho interval spolehlivosti.

Prostředí \ Myši	Problémy ano	Problémy ne	Celkem
toxické látky +	30 (76,9 %)	9 (23,1 %)	39 (100,0 %)
toxické látky –	20 (40,0 %)	30 (60,0 %)	50 (100,0 %)
Celkem	50 (56,2 %)	39 (43,8 %)	89 (100,0 %)

- Zadání i postup výpočtu jsou zde stejné jako v příkladu 1a, pouze hodnoty jsou navoleny tak, aby mezi oběma srovnávanými skupinami byl ve výsledku větší rozdíl. Výsledný interval spolehlivosti pro odhad rozdílu relativních četností tak neobsahuje nulu, což indikuje statistickou významnost rozdílu experimentálních skupin.

- Po dosazení do vztahu pro binomický test pro dva výběry (příklad 1a/bod3) získáme hodnotu testové statistiky $Z = 3,481$. Tato hodnota vysoce převyšuje kritickou hodnotu testu pro $\alpha = 0,05$ ($Z = 1,96$), a tudíž dovoluje zamítnout nulovou hypotézu. Experimentální skupina myší žijících v toxickém prostředí má statisticky významně vyšší podíl jedinců s problémy.

- Závěr v bodě 3 potvrzuje i výpočet intervalu spolehlivosti. Bodový rozdíl v relativních četnostech obou skupin je 0,369; jeho 95% interval spolehlivosti je 0,163–0,531 (viz výpočet v příkladu 1a). Interval spolehlivosti pro rozdíl relativních četností nezahrnuje hodnotu 0, toto zjištění je ekvivalentní k zamítnutí nulové hypotézy na hladině významnosti $\alpha = 0,05$.

Příklad 1b. Odhad rozdílu relativních četností a jeho interval spolehlivosti.

Popis problému: Srovnáváme výskyt nežádoucích účinků (NÚ) u dvou léků u $N = 8$ pacientů. Testované léky jsou dle protokolu studie aplikovány u stejných pacientů, hodnocení tedy není nezávislé. Testujeme zde shodu rozdělení u dvou proměnných, které jsou vzájemně provázány osobou pacienta. Hodnocená proměnná je binární (dichotomická: 1/0).

	Lék 1	Lék 2	Součet řádků
pacient 1	0	1	1
pacient 2	0	1	1
pacient 3	1	0	1
pacient 4	0	1	1
pacient 5	1	1	2
pacient 6	0	0	0
pacient 7	1	0	1
pacient 8	1	1	2
Součet sloupců	4	5	9

1. Srovnáváme výskyt nežádoucích účinků (NÚ) u dvou léků. Sledovanou proměnnou je binární kód (1 = nežádoucí účinek se vyskytl; 0 = bez nežádoucího účinku). H_0 : léky se neliší ve výskytu nežádoucích účinků.

2. Pomocí Cochranova testu srovnáme rozdělení sledované proměnné mezi pokusnými zásahy (léky). Cochranův test je dán vztahem:

$$T = \frac{k(k-1) \sum_{j=1}^k \left(X_{.j} - \frac{b}{k} \right)^2}{\sum_{i=1}^N X_{i.} (k - X_{i.})}$$

kde k je počet sledovaných léků, N je počet pacientů v experimentu, $X_{.j}$ je součet hodnot „1“ u j -té léčby, $X_{i.}$ je četnost hodnot „1“ u i -tého pacienta, b je celkový součet četností hodnot „1“.

3. Po dosažení do vztahu v bodě 2 vypočítáme hodnotu testové statistiky:

$$T = \frac{2(2-1) \left[\left(4 - \frac{9}{2}\right)^2 + \left(5 - \frac{9}{2}\right)^2 \right]}{1(2-1) + 1(2-1) + 1(2-1) + 1(2-1) + 2(2-2) + 0(2-0) + 1(2-1) + 2(2-2)} = 0,2$$

4. Vypočtená hodnota testové statistiky je porovnána s kritickou hodnotou testu pro hladinu významnosti $\alpha = 0,05$:

$$\chi^2_{(0,95)}^{(v=1)} = 3,84.$$

5. Výsledná hodnota statistické významnosti je: $T < \chi^2_{(0,95)}^{(v=1)}$, $p > 0,05$; nezamítáme tedy H_0 , výskyt NÚ u srovnávaných léků se statisticky významně neliší.

Příklad 2. Hodnocení výskytu nežádoucích účinků u pacientů při postupném nasazení dvou léků (párové uspořádání, Cochranův test).

Popis problému: Srovnáváme tíži nežádoucích účinků (NÚ) na ordinální škále 1–4 u dvou léků, celkem u $N = 71$ pacientů. Testované léky jsou dle protokolu studie aplikovány u stejných pacientů s určitým odstupem, hodnocení tedy není nezávislé.

Lék 1 – tíže NÚ (i)	Lék 2 – tíže NÚ (j)			
	1	1	3	4
1	5	5	8	11
2	1	5	6	10
3	2	5	4	2
4	3	2	1	1

Kontingenční tabulka vzniká přepsáním primárních dat

Lék 1 – tíže NÚ	Lék 2 – tíže NÚ
1	1
1	1
1	2
1	3
1	3
1	4
2	1
2	1
2	1
2	1
2	1
2	2
...	...

1. Srovnáváme tíži nežádoucích účinků (NÚ) u dvou léků. Sledovanou proměnnou je ordinální stupnice tíže (grade) nežádoucích účinků (1–4). H_0 : Léky se ve výskytu různě závažných nežádoucích účinků neliší.

2. Pomocí Bowkerova testu (znám také pod názvem McNemar-Bowker test) srovnáme rozdělení sledované proměnné (tíže NÚ) mezi pokusnými zásahy (typy léčby).

3. Bowkerův test testuje symetrii čtvercové tabulky kolem diagonály a je dán vztahem

$$Q_B = \sum_{i < j} \sum \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

kde i a j jsou indexy reprezentující řádky a sloupce kontingenční tabulky, n_{ij} je počet pacientů v dané buňce kontingenční tabulky

4. Po dosažení do vztahu v bodě 3 vypočítáme hodnotu testové statistiky:

$$Q_B = 16,596$$

5. Testová statistika má přibližně χ^2 rozložení s $r(r-1)/2$ stupni volnosti, kde r je počet ordinálních kategorií u sledované veličiny.

6. Vypočtená hodnota testové statistiky je porovnána s kritickou hodnotou testu pro $\alpha = 0,05$: $\chi^2_{(0,95)}^{(v=6)} = 12,591$.

7. Výsledná hodnota statistické významnosti je: $Q_B > \chi^2_{(0,95)}^{(v=6)}$, $p < 0,05$ a tedy zamítáme H_0 , testované léky se liší v tíži nežádoucích účinků

Poznámka:

Bowkerův test je rozšířením McNemaraova testu pro $K \times K$ tabulku četností. Test tedy vyžaduje čtvercovou tabulku četností – ordinální škála pro hodnocení NÚ u obou léků je stejná. Při $K = 2$ je výpočet Bowkerova testu stejný jako u testu dle McNemara.

Příklad 3. Hodnocení tíže nežádoucích účinků u pacientů při postupném nasazení dvou léků (párové uspořádání, Bowkerův test).

XXIV. VYBRANÉ PŘÍKLADY HODNOCENÍ KATEGORIÁLNÍCH DAT

Tabulka četností	Anémie	Tíže průběhu neurologického onemocnění		
		I	II	III
	bez anémie	16	6	8
	lehká anémie	3	24	27
	těžká anémie	3	17	40

Tabulka četností vzniká přepsáním z primárních dat

Ukázka primárních dat	Tíže onemocnění	Anémie
	I	bez anémie
	I	bez anémie
	I	lehká anémie
	I	těžká anémie
	II	bez anémie
	II	bez anémie
	II	lehká anémie
	II	těžká anémie
	II	těžká anémie
	III	bez anémie
	III	lehká anémie
	III	těžká anémie
	III	těžká anémie

1. Hodnotíme vztah dvou ordinálních znaků: tíže anémie ve třech stupních vs obtížnost průběhu neurologického onemocnění měřená na ordinální škále I–III.
2. Nulová hypotéza: není vztah mezi tíží anémie a tíží neurologického onemocnění.
3. Pro výpočet využijeme Kendallův korelační koeficient:

$$\tau_A = \frac{n_s - n_d}{\frac{1}{2}n(n-1)}$$
 kde n_s je počet souhlasných dvojic subjektů (se stejným pořadím hodnot v obou znacích), n_d je počet nesouhlasných dvojic subjektů (s rozdílným pořadím hodnot v obou znacích) a n je celkový počet dvojic subjektů. Více viz díl XXI seriálu.
4. Pro testování H_0 je pohodlnější použít testovou statistiku z_A , jejíž rozdělení lze aproximovat standardizovaným normálním rozdělením:

$$z_A = \frac{3(n_s - n_d)}{\sqrt{\frac{n(n-1)(2n+5)}{2}}}$$
 kde n_s je počet souhlasných dvojic subjektů (se stejným pořadím hodnot v obou znacích), n_d je počet nesouhlasných dvojic subjektů (s rozdílným pořadím hodnot v obou znacích) a n je celkový počet dvojic subjektů.
5. Korelační koeficient nabývá hodnoty $\tau_A = 0,348$ (odpovídající statistika $z_A = 3,3$), což odpovídá hodnotě $p < 0,001$.
6. Na hladině významnosti $\alpha = 0,05$ zamítáme hypotézu o nezávislosti tíže anémie a tíže neurologického onemocnění.
7. **Závěr:** Existuje statisticky významný vztah mezi tíží anémie a tíží neurologického onemocnění.

Příklad 4. Test vztahu dvou ordinálních znaků pomocí Kendallova korelačního koeficientu.

Skóre kvality života	Před léčbou	Po léčbě
1	10	1
2	13	2
3	9	3
4	8	5
5	9	5
6	4	9
7	5	8
8	5	8
9	4	12
10	1	15

Tabulka četností vzniká přepisem primárních dat

Skóre před léčbou	Skóre po léčbě
1	1
1	2
1	2
1	3
1	3
1	3
1	3
1	4
1	4
1	4
1	4
2	4
2	5
2	5
2	5
2	5
...	...

- Hodnotíme skóre kvality života na ordinální škále 1–10 dotazováním pacientů před a po léčbě; $n = 68$. Nulová hypotéza (oboustranná): po léčbě nedošlo ke změně skóre kvality života oproti stavu před léčbou.
- Vzhledem k ordinální škále skóre kvality života a párovému uspořádání sledování je vhodným testem Wilcoxonův párový test.
- Je spočítán rozdíl mezi hodnotami skóre kvality života před a po léčbě (nulové rozdíly jsou z výpočtu vyřazeny).
- Absolutní hodnotě rozdílů je přiřazeno pořadí; vyskytnou-li se stejné hodnoty, je jim přiřazen průměr pořadí, která na ně připadají. Následně sečteme pořadí připadající na kladné a záporné rozdíly:

$$T_+ = 2211$$

$$T_- = 0$$
- Pro porovnání s kritickou hodnotou testu je vybrána menší z hodnot T_+ a T_- , v tomto případě $T_- = 0$. Hodnota T je aproximována na standardní normální rozdělení Z , zde $Z = 7,062$.
- Kritická hodnota testové statistiky Z pro $\alpha = 0,05$ je $Z_{0,975} = 1,96$.
- Hodnota $Z = 7,062$ je větší než nalezená kritická hodnota $Z_{0,975} = 1,96$, což vede k zamítnutí nulové hypotézy.
- Závěr:** Na hladině významnosti $\alpha = 0,05$ byla prokázána statisticky významná změna ve skóre kvality života po realizaci hodnocené léčby.

Poznámka. Výpočet testu je detailněji popsán v díle XVII tohoto seriálu. Test je běžnou součástí nabídky statistických softwarových nástrojů.

Příklad 6. Využití Wilcoxonova párového testu pro srovnání dvou ordinálních proměnných v párové tabulce četností.

Potíže při chůzi	Operováni	Léčení neoperačně
I	8	19
II	7	16
III	16	12
IV	21	4

↑
Tabulka četností vzniká přepisem primárních dat

Skupina dle léčby	Potíže při chůzi
operovaný	I
operovaný	II
operovaný	II
operovaný	III
operovaný	IV
léčený neoperačně	I
léčený neoperačně	II
léčený neoperačně	III
léčený neoperačně	III
léčený neoperačně	IV
...	...

1. U pacientů léčených dvěma způsoby (operačně a neoperačně) byly na čtyřbodové ordinální stupnici hodnoceny potíže při chůzi ($n_1 = 52$ u operačně léčených a $n_2 = 51$ u léčených neoperačně). Uspořádání sledování je nezávislé, obě skupiny jsou tvořeny různými pacienty.

2. Nulová hypotéza: výskyt potíží při chůzi se mezi oběma skupinami pacientů neliší.

3. Výpočet testu přiřazuje k stupni potíží při chůzi (sledovaná proměnná) celkové pořadí bez ohledu na typ léčby, ze kterého daná hodnota pochází. Vyskytnou-li se stejné hodnoty, je jim přiřazen průměr pořadí, která na ně připadají. Pro pacienty s různou léčbou jsou potom pořadí sečtena odděleně: R_1 (operováni) = 3 328; R_2 (léčení neoperačně) = 2 028.

4. Počty měření a sumy pořadí v jednotlivých experimentech jsou dosazeny do vztahu:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

5. Výsledná hodnota testové statistiky je $U = 702$.

6. Hodnota U je aproximována na standardizované normální rozdělení pomocí statistiky Z , přičemž $Z = 4,11$. Hodnota $Z = 4,11$ je větší než kritická hodnota $Z_{0,975} = 1,96$ což vede k zamítnutí nulové hypotézy.

7. **Závěr:** Na hladině významnosti $\alpha = 0,05$ byl prokázán statisticky významný rozdíl ve výskytu různých potíží při chůzi mezi různými léčenými skupinami pacientů

Poznámka: Výpočet Mannova-Whitneyho U testu je detailněji popsán v díle XVII tohoto seriálu. Výpočet lze pro tento typ problému doporučit zejména pokud sledovaná ordinální škála dosahuje velkého množství hodnot (stupňů). Test je běžnou součástí nabídky statistických softwarových nástrojů.

Příklad 5. Využití Mannova-Whitneyho U testu pro srovnání dvou ordinálních proměnných v tabulce četností.

DETEKCE ANTIGENU	Metoda II		
	Metoda I	Před léčbou	Po léčbě
ne		60	9
ano		1	42

■ Nesouhlas metod v detekci antigenu

1. Pro detekci antigenu (v experimentu je ve vzorku vždy přítomen) byly srovnávány dvě metody.

2. Pomocí McNemarova testu a znaménkového testu testujeme hypotézu symetrie rozdílů mezi metodami detekce antigenu. H_0 : Obě metody se neliší v relativním výskytu nesouhlasných výsledků detekce antigenu.

3. Testová statistika McNemarova testu je dána vztahem:

$$\chi^2_{(c)} = \frac{(|M_1 - M_2| - 1)^2}{M_1 + M_2}$$

kde M_1 je počet případů, kdy je antigen detekován pouze metodou I, M_2 je počet případů, kdy je antigen detekován pouze metodou II.

4. Po dosažení do vztahu v bodě 3 vypočítáme hodnotu testové statistiky $\chi^2 = 4,90$. Testová statistika má χ^2 rozložení, vypočtená hodnota je porovnána s kritickou hodnotou testu $\chi^2_{(0,95)}^{v=1} = 3,84$. Výsledná hodnota statistické významnosti je $p = 0,027$, a tedy zamítáme H_0 . Obě metody se statisticky významně liší v záchytu antigenu.

5. Alternativou k McNemarovu testu je testování symetrie nesouhlasných výsledků testů pomocí znaménkového testu. Vstupem do výpočtu je zde počet detekcí antigenu pouze metodou I (1 případ) a počet detekcí antigenu pouze metodou II (9 případů).

6. V případě, že není žádný rozdíl mezi metodami detekce, je podíl hodnot detekce pouze jednou z metod shodně 0,5 (50 %); pro další testování je využit binomický test pro binomické rozdělení s hodnotou $\pi = 0,5$ a $n = 10$.

7. Hledáme pravděpodobnost náhodného výskytu binomického rozložení s $n = 10$ a $\pi = 0,5$, pokud je počet výskytu jevu ≤ 1 (detekce pouze metodou I) nebo ≥ 9 (detekce pouze metodou II):

$$p(X \leq 1 \text{ nebo } X \geq 9) = 0,00098 + 0,00977 + 0,00977 + 0,00098 = 0,02$$

8. Hodnota $p = 0,02$ představuje pravděpodobnost, že zcela náhodně vznikne stejné nebo extrémnější rozmístění nesouhlasů metod v detekci antigenu než v našem příkladu. **Hodnota $p = 0,02$ je menší než $\alpha = 0,05$; což vede k zamítnutí nulové hypotézy H_0 . Ve shodě se závěrem McNemarova testu i zde uzavíráme, že se obě metody statisticky významně liší v detekci antigenu.**

Poznámka. Znaménkový test zde vystupuje v roli plnohodnotné alternativy McNemarova testu. Jako standardní postup pro daný problém lze zejména při malé velikosti vzorku doporučit McNemarův test. Aplikace binomického rozdělení přináší jistou přidanou hodnotu pro prezentaci výsledku tím, že kalkuluje, s jakou pravděpodobností bychom pozorovaný výsledek zachytili zcela náhodně.

Příklad 7. Znaménkový test jako alternativa McNemarova testu při hodnocení rozdílů relativních četností u párových sledování.