UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Neparametrické testy

Vedoucí bakalářské práce:
RNDr. et PhDr. Ivo Müller, Ph.D.
Rok odevzdání: 2014

Vypracovala:
Adéla Vrtková
AST, III. ročník

Prohlášení

Prohlašuji, že jsem vytvořila tuto bakalářskou práci samostatně za vedení RNDr. et PhDr. Iva Müllera, Ph.D. a že jsem v seznamu použité literatury uvedla všechny zdroje použité při zpracování práce.

V Olomouci dne 28. dubna 2014

Poděkování

Ráda bych poděkovala vedoucímu bakalářské práce RNDr. et PhDr. Ivu Müllerovi, Ph.D. za spolupráci i za čas a cenné rady, které mi věnoval při konzultacích.

Obsah

Úvod

Tématem práce jsou neparametrické testy. Na následujících stranách se pokusím zmíněné metody vysvětlit, ukázat jejich použití na konkrétních datech a chtěla bych se zaměřit na jejich praktickou stránku. Tedy jak ověřit předpoklady, provést výpočet a jak interpretovat výsledek testu, vzhledem k zadaným datům a zkoumané hypotéze.

Neparametrické metody jsou součástí velkého odvětví testování hypotéz. Vzhledem k rozsáhlé teorii testování, kterou nelze do rozsahu této práce zahrnout, bude práce určena již pro čtenáře, kteří mají základní poznatky z matematické statistiky. Pojmy jako náhodný výběr, distribuční funkce, nulová a alternativní hypotéza, kritický obor apod. budu zde považovat za obecnou znalost. Pokud by však bylo třeba si základy testování hypotéz osvětlit, doporučuji např. [1, kap. 7.1].

Použití neparametrických testů je v praxi široké, protože na data nekladou takové přísné nároky jako klasické parametrické testy. Jak název napovídá, parametrické testy testují některý z parametrů většinou normálního rozdělení (střední hodnotu nebo rozptyl) a k tomu je zapotřebí odhadů těchto parametrů, které vypočítáme pouze pomocí původních dat. Neparametrické testy testují pouze rozdělení, zda je symetrické, zda dva výběry pochází se stejného rozdělení apod., tedy není zde testován žádný parametr. Proto je možné se u neparametrických testů omezit pouze na pořadí veličin a původní data v podstatě nepotřebujeme. Stačí vědět, které pozorování je větší a menší. Díky této vlastnosti je okruh použití neparametrických testů mnohem větší než testů klasických. Používáme je na situace, kdy o rozdělení dat toho moc nevíme, také na ordinální data nebo na výběry s malým rozsahem. U mnoha neparametrických testů nám postačí, když náhodný výběr bude pocházet z blíže nespecifikovaného spojitého rozdělení.

Práce je rozdělena na čtyři kapitoly, kdy jejich členění je určeno povahou nulové a alternativní hypotézy. V každé kapitole bude popsána základní situace a uvedeny nejpoužívanější a nejznámější testy týkající se dané problematiky, rovněž bude na příkladu ukázán praktický postup testování. Budou dále srovnány různé možnosti testování, tedy ruční počítání, počítání pomocí aproximací a použitím statistického softwaru R. Na závěr každé kapitoly budou srovnány testy mezi sebou v závislosti na jejich použití na praktickém příkladu.

Pro účely této práce jsem pomocí dotazníků a sociálních sítí sesbírala několik sad dat, které jsem následně analyzovala. Na tomto místě bych chtěla říct, že data slouží jako konkrétní příklady obecně popsaných situací pro použití vybraných neparametrických testů a cílem práce není jakýkoliv sociologický průzkum, protože pro absolutní regulérnost by bylo zapotřebí většího výběru a hlavně větší rozmanitosti, než je k dispozici skrze sociální sítě. Data tedy slouží pro předvedení postupu testování a hlavně pro ukázku interpretace výsledků. Neznamená to však, že by data porušovala základní předpoklady nutné pro použití neparametrických testů. Pouze závěry, které z testů vyplynou, je třeba brát opatrně a negeneralizovat je na celou populaci. Přesto se ale domnívám, že přínosem práce by mohl být právě její praktický náhled na problematiku neparametrických testů.

1 Hypotéza shody dvou populací proti alternativě posunutí

Populace, neboli základní soubor, reprezentuje realitu, kterou zamýšlíme zkoumat. Populaci tvoří statistické jednotky, na kterých chceme pozorovat určitý znak, vlastnost. Vzhledem k finanční a časové náročnosti pozorování celé populace se utvoří tzv. výběr. Statistické jednotky zahrnuté do výběru nazýváme subjekty, na kterých přímo provádíme požadované pozorování nebo měření. Data, získaná pozorováním, jsou pak realizacemi náhodných veličin, protože pro dva různé výběry z téže populace můžeme stejným pozorováním získat dvě různé sady dat. Každý výzkum se pak provádí za účelem zodpovězení otázek, které již máme určené dopředu nebo mohou vyplynout během samotného sbírání dat. Tyto otázky nazýváme hypotézy.

Vysvětlíme-li na příkladu, populaci tvoří školáci základních škol celé republiky. Statistickou jednotkou je každý jednotlivý školák a chceme zjistit, neboli budeme testovat hypotézu, že průměrná hodnota IQ u školáků je 100. Provedeme tzv. dvoustupňový výběr. Nejprve ze všech základních škol republiky náhodně vybereme např. 10 škol. Na každé vybrané škole vybereme náhodně dvě třídy, které otestujeme inteligenčním testem. Z provedených pozorování získáme sadu dat, na kterou aplikujeme statistické testy. Tento příklad odpovídá situaci, kdy máme jeden náhodný výběr z jedné populace. Kdybychom chtěli vytvořit situaci dvou populací, pak bychom kromě žáků základních škol náhodně vybrali studenty středních škol a testovali bychom, že průměrné IQ u žáků základních škol je stejné jako u studentů středních škol. Dvěma náhodnými výběry se budeme dále zabývat, přičemž základní model popíšeme matematicky, čerpáno z [6].

Nechť $X_1, X_2, ..., X_m$ a $Y_1, Y_2, ..., Y_n$ jsou dva nezávislé náhodné výběry ze dvou populací, jež po řadě pocházejí z rozdělení F a G, m a n jsou pevné rozsahy výběrů. Dále budeme používat tzv. sdružený výběr $X_1, ..., X_m, Y_1, ..., Y_n$, o rozsahu $m + n$. Matematicky má nulová a alternativní hypotéza tvar:

$H_0 : F = G,$

$H_1 : G(x) = F(x - \Delta), \Delta \neq 0,$

neboli máme nulovou hypotézu o shodnosti rozdělení výběrů proti alternativě, že se jejich distribuční funkce liší v posunutí o neznámý parametr Δ. Vzhledem k faktu, že rozdělení náhodné veličiny jednoznačně určuje její distribuční funkci a naopak, nebudeme označení rozdělení a k nim příslušných distribučních funkcí rozlišovat, resp. při formulaci hypotéz je tvrzení o shodnosti rozdělení ekvivalentní se shodností odpovídajících distribučních funkcí.

V případě $\Delta \neq 0$ máme oboustrannou alternativu, která nám říká, že se distribuční funkce liší v posunutí, ale neříká jakým směrem. Lze tedy podle volby Δ formulovat i jednostrannou alternativu. Pro $\Delta > 0$ dostáváme pravostrannou alternativu a očekáváme, že hodnoty výběru $Y_1, ..., Y_n$ budou větší než hodnoty $X_1, ..., X_m$. Naopak při levostranné alternativě, kdy $\Delta < 0$, očekáváme, že hodnoty výběru $Y_1, ..., Y_n$ budou menší než hodnoty $X_1, ..., X_m$. Volba alternativy samozřejmě ovlivní tvar kritického oboru a volbu kritických hodnot. Tato problematika bude rozvedena později pro každý test zvlášť.

Dále označíme $R_1, ..., R_m$ pořadí veličin $X_1, ..., X_m$ ve sdruženém výběru seřazeném podle velikosti od nejmenší hodnoty po největší a $S_1, ..., S_n$ pořadí veličin $Y_1, ..., Y_n$. Zpravidla se označení výběrů volí tak, aby platilo $m \geq n$.

Nyní přistoupíme k praktické analýze. Popsaný model lze použít na tři základní situace:

- situace (1.a) - dva nezávislé výběry ze dvou populací, testujeme shodnost;
- situace (1.b) - dva nezávislé výběry z jedné populace - kontrola, ošetření, testujeme vliv ošetření;
- situace (1.c) - dva nezávislé výběry z jedné populace, každý s jiným ošetřením, testujeme shodnost v efektu ošetření.

Typickým příkladem situace (1.a) je výběr z populace žen a výběr z populace mužů, kdy bychom mohli zkoumat, jak velké spropitné dávají v zábavních podnicích. Sledovali bychom platby jednotlivých hostů a zapisovali si hodnotu spropitného. Nezávislost by byla v rámci výběru dosáhnuta tím, že by hosté chodili platit jednotlivě a nevěděli by, jaké spropitné nechávají ostatní. Nezávislost mezi muži a ženami by navíc byla dána tím, že mezi nimi není jakýkoli příbuzenský nebo partnerský vztah. Dále bychom testovali nulovou hypotézu, že muži a ženy nechávají stejné spropitné. Oboustranná alternativa by říkala, že výše spropitného se liší. Jednostranné alternativy by testovaly, že buď ženy dávají menší spropitné než muži, nebo naopak větší. Dalším příkladem by mohlo být sledování řidičských schopností u mužů a u žen, zkoumání životní spokojenosti u lidí žijících ve městě a lidí z venkova nebo výzkum postoje k dětem u ženatých mužů proti svobodným.

Situace (1.b) je nejlépe aplikovatelná na biomedicínské výzkumy. Řekněme, že máme skupinu pacientů s vysokým krevním tlakem a chceme testovat vliv nového léku, který by měl tlak snížit. Pacienty náhodně rozdělíme do skupiny kontrolní a do skupiny s ošetřením, kdy ošetřením rozumíme podání léku. Nulová hypotéza testuje, že ošetření nemá vliv na krevní tlak, a v tomto případě bereme jednostrannou alternativu, která říká, že ošetření krevní tlak snižuje. Stejný princip bychom použili u většiny lékařských výzkumů, kde se sleduje vliv nového léku nebo nové metody operace. Co se týče splnění předpokladů, pak nezávislosti mezi skupinami docílíme náhodným přiřazením kontroly a ošetření. Náhodným přiřazením do skupin dále lépe promícháme nezohledněné faktory, snížíme možnost ovlivňování se mezi subjekty a v některých případech budeme moct i určit kauzalitu. V rámci skupiny je nezávislost dána tím, že pacienti netuší, ve které skupině jsou, tedy se nemůžou mezi sebou ovlivňovat, a ani nemůže hrát roli placebo efekt. Rozdělení do skupin můžeme provést tak, že pomocí generátoru čísel přiřadíme každému pacientovi 0 nebo 1. Pacienti s 0 pak budou ve skupině kontrolní a bude jim podáno placebo a na pacienty s 1 se aplikuje ošetření. Obdobně bychom si mohli u každého pacienta hodit mincí a podle toho, jak padne rub a líc, je rozdělit. Zamyslíme-li se nad předpokladem o pevných rozsazích výběru, nabízí se otázka, zda tímto postupem pevný rozsah dodržíme. Hod mincí je totiž náhodný pokus a v případě, že budeme mít např. 100 pacientů, nemusíme dosáhnout požadovaných rozsahů (např. 50 na 50). Jednou možností je házet tak dlouho, dokud jedna skupina nebude naplněna a potom zbytek automaticky zařadit do druhé. Na druhou stranu, ať už přiřadíme ošetření a kontrolu jakkoli, v okamžiku, kdy začneme s daty pracovat a rozsahy neměníme, pak je už můžeme považovat za pevné.

Tento postup by šel aplikovat i pro zkoumání efektu nové učební metody u studentů nebo nového hnojiva u rostlin, kdy na část půdy by bylo použito nové hnojivo, druhá část by se nechala bez hnojení a sledoval by se vliv hnojení na růst rostlin. Dalším příkladem, který by se mohl zdát jako typický pro tuto situaci, je zkoumání kuřáků a nekuřáků, resp. vliv kouření na zdraví jedince. Ideální by bylo náhodně vybraným nekuřákům přiřadit, zda zůstanou nekuřáci, nebo jestli začnou kouřit. Věděli bychom, jak často a jak dlouho kouří a podle toho pak snadno vyvodit vliv kouření na zdraví a případně prokázat i kauzalitu. Mohli bychom zjistit, že jestliže člověk kouří, pak bude mít zdravotní problémy. Avšak z etického hlediska nelze náhodně nekuřákům přikázat, aby začli kouřit, a tak si ověřit vliv kouření, a proto je třeba postupovat jinak. Provedli bychom dva náhodné výběry, kdy jeden by byl z populace nekuřáků a druhý z populace kuřáků. Problém kuřáků a nekuřáků je proto vhodnější chápat jako situaci (1.a).

Velmi přesvědčivý příklad pro situaci (1.c) jsou náhodně vybrané dvě skupiny studentů ze stejné populace, kdy na každou skupinu aplikujeme ošetření, kterým rozumíme různé vyučovací metody. Po skončení výuky se pak každé skupině dá vědomostní test a výsledky se porovnají. Chceme tedy testovat nulovou hypotézu, že metody výuky jsou stejně účinné, oproti oboustranné alternativě, že stejně účinné nejsou. Nemáme-li předem daný odhad o tom, která metoda je lepší, oboustranná alternativa je nejvhodnější volbou, protože chceme zjistit, jestli vůbec se jejich účinnost liší. Pro testování stejné hypotézy bychom mohli provést trochu jiný způsob výběru, kdy vybereme jednu skupinu studentů, kterým pak náhodně různá ošetření přiřadíme (např. hodem mincí). Obdobně by situace (1.c) odpovídala zkoumání vlivu dvou různých hnojiv na růst rostlin. Na jednu část půdy by se aplikovalo jedno hnojivo, na druhou část půdy jiné a nulová hypotéza by testovala shodnost v efektu hnojiv oproti oboustranné alternativě, že hnojiva jsou různě účinná. Analogicky by se dala testovat účinnost dvou různých léků, metod výcviku zvířat, způsobu tréninku sportovců apod.

Předpoklady matematického modelu shrneme do následujících bodů:

- (A1.1): $X_1, ..., X_m$ a $Y_1, ..., Y_n$ jsou náhodné výběry ze spojitých rozdělení;
- (A1.2): výběry jsou nezávislé;
- (A1.3): rozsahy výběrů m a n jsou pevné.

Předpoklady platí pro všechny testy této kapitoly, nebude-li řečeno jinak. Předpoklad spojitosti je důležitý kvůli tzv. shodám, což jsou pozorování, která mají stejnou hodnotu a není možné jim jednoznačně přiřadit pořadí. Předpoklad spojitosti má zaručit, že pravděpodobnost shod je nulová. Výskyt shod totiž narušuje přesnost tabelovaných kritických hodnot a limitních aproximací u jednotlivých testů. V praxi tedy můžeme bez omezení pracovat s diskrétními daty, pokud se v nich nevyskytují shody. Někdy se však shodám nevyhneme, a potom je potřeba upravených testových statistik. Pokud máme velký výběr a shod je málo, můžeme je vypustit a bez problémů použít tabelovaných kritických hodnot a aproximace. Jinou možností je přiřazení průměrných pořadí, kdy pak můžeme pracovat s upravenou testovou statistikou a normální aproximací, je-li k dispozici, nebo použít tabelovaných kritických hodnot a aproximace pro data beze shod, ale mít na vědomí, že už je výsledek nepřesný. Formulace testů této kapitoly pochází hlavně z [2, 3].

1.1 Dvouvýběrový Wilcoxonův test

Jeden z nejstarších neparametrických testů byl představen Frankem Wilcoxonem v roce 1945 a byl formulován pro stejné rozsahy výběrů ($m = n$). O dva roky později jej pak rozšířili Henry Mann a Donald R. Whitney pro různé rozsahy ($m \neq n$) a stanovili tabulky kritických hodnot pro malé m a n, čerpáno z [3].

Označme součty pořadí náhodných výběrů

$$T_1 = \sum_{i=1}^{m} R_i, \quad T_2 = \sum_{i=1}^{n} S_i. \tag{1.1}$$

Potom zaveďme

$$U_1 = mn + \frac{1}{2}m(m+1) - T_1, \quad U_2 = mn + \frac{1}{2}n(n+1) - T_2, \tag{1.2}$$

kdy test založen na U_1 a U_2 se nazývá Mannův-Whitneyho test a test založen na T_1 je dvouvýběrový Wilcoxonův test. Dále označme

$$MW = min(U_1, U_2). \tag{1.3}$$

Hodnotu MW porovnáme s odpovídající tabelovanou kritickou hodnotou pro oboustranný test $W_c(\alpha)$, dohledatelné v [8]. Platí-li $MW \leq W_c(\alpha)$, pak nulovou hypotézu na dané hladině α zamítáme oproti oboustranné alternativě. Je možné si všimnout, že pro oboustrannou alternativu je test formulován s jednostranným kritickým oborem.

Kritické hodnoty $W_c(\alpha)$ v tabulkách jsou uváděny pro oboustranný test tak, aby dodržel hladinu α. Tabulky pro jednostranné testy speciálně uváděny nejsou. Jestliže budeme testovat proti jednostranné alternativě na hladině α, kritickou hodnotu musíme vzít z tabulek pro oboustranný test, který má hladinu 2α, tedy použijeme kritickou hodnotu $W_c(2\alpha)$. Neboli pro jednostranný test na hladině 0,05 použijeme tabulky určené oboustrannému testu s hladinou 0,1. Je třeba si dávat pozor, které tabelované kritické hodnoty používáme, protože existují i tabulky pro statistiku T_1. V tomto textu budeme však T_1 a T_2 používat jen pro výpočet U_1, U_2.

Testujeme-li pravostrannou alternativu, pak pro ni budou svědčit malé hodnoty U_2, a bude-li $U_2 \leq W_c(2\alpha)$, zamítneme nulovou hypotézu na hladině α. V případě levostranné alternativy, pokud $U_1 \leq W_c(2\alpha)$, pak zamítáme nulovou hypotézu na hladině α.

Pro větší rozsahy výběrů (obvykle $m > 10, n > 10$) je možné využít asymptotického normálního rozdělení veličiny U_1 a použít statistiku:

$$U_{MW} = \frac{U_1 - \mathrm{E}U_1}{\sqrt{\mathrm{var}U_1}}, \quad \text{kde } \mathrm{E}U_1 = \frac{1}{2}mn, \quad \mathrm{var}U_1 = \frac{1}{12}mn(m+n+1). \tag{1.4}$$

Jestliže $|U_{MW}| \geq u(1 - \frac{\alpha}{2})$, pak zamítáme nulovou hypotézu proti oboustranné alternativě na hladině testu α, kde $u(1 - \frac{\alpha}{2})$ je $(1 - \frac{\alpha}{2})$-kvantil rozdělení $N(0, 1)$. V případě pravostranné alternativy, $\Delta > 0$, budou v její prospěch svědčit větší hodnoty statistiky U_{MW} a H_0 budeme zamítat, jestliže $U_{MW} \geq u(1 - \alpha)$. Analogicky při levostranné alternativě, $\Delta < 0$, budeme zamítat H_0, když $U_{MW} \leq u(\alpha)$.

Vyskytují-li se ve výběru shody, potom jim přiřadíme průměrná pořadí a pro přesný výsledek je třeba místo U_{MW} použít upravenou statistiku

$$U_{MW}^{*} = \frac{U_1 - \frac{mn}{2}}{\sqrt{\frac{mn}{N(N-1)}\left(\frac{N^3-N}{12} - \sum_{i=1}^{r}\frac{t_i^3-t_i}{12}\right)}}, \tag{1.5}$$

kde $N = m + n$, r je počet shod a t_i je násobnost i-té shody z [1, str. 103].

Použijeme-li software R, je možné si nastavit typ alternativy a použít příslušný test podle toho, zda se ve výběru vyskytují shody. Pro výběry beze shod použijeme funkci `wilcox.test` a pro výběr se shodami `wilcox.exact`. Podle zadaných parametrů proběhne výpočet a výstupem pak bývá p-hodnota, na základě které pak rozhodujeme o zamítnutí nebo nezamítnutí nulové hypotézy. Připomeneme, že H_0 zamítáme, jestliže p-hodnota je menší než hladina testu α.

Zobecněním dvouvýběrového Wilcoxonova testu je tzv. Kruskalův-Wallisův test pro shodu více populací. Testuje hypotézu, že všech k výběrů pochází ze stejného rozdělení, oproti alternativě, že alespoň jeden pochází z jiného. Při zamítnutí nulové hypotézy pak existují metody, pomocí kterých je možno odhalit, kterých dvojic se rozdílnost týká. V případě, že bychom Kruskalův-Wallisův test použili na dva nezávislé náhodné výběry, pak test splývá s oboustranným dvouvýběrovým Wilcoxonovým testem. Veškerou teorii a odvození je možné najít např. v [3, str. 291].

Dvouvýběrový Wilcoxonův test lze sice použít i pro testy s obecnou alternativou $H_1 : F \neq G$, avšak je mnohem citlivější právě na alternativu posunutí, resp. při alternativě posunutí má mnohem větší sílu.

1.2 Van der Waerdenův test

Pro formulaci testu označme Φ distribuční funkci rozdělení $N(0,1)$ a Φ^{-1} její odpovídající kvantilovou funkci. Test je vhodný pro rozdělení velmi blízká normálnímu a testová statistika má tvar

$$VW = \sum_{j=1}^{n} \Phi^{-1}\left(\frac{S_j}{m+n+1}\right). \tag{1.6}$$

Kritické hodnoty $VW_c(\alpha)$ jsou pro $m + n \leq 50$, $|m - n| \leq 5$ tabelovány v [9]. Je-li $|VW| \geq VW_c(\alpha)$, pak zamítáme H_0 proti oboustranné alternativě. V případě jednostranného testu opět používáme kritické hodnoty $VW_c(2\alpha)$ obdobně jako u dvouvýběrového Wilcoxonova testu. Při pravostranné alternativě (resp. levostranné) zamítáme nulovou hypotézu, když $VW \geq VW_c(2\alpha)$ (resp. $VW \leq -VW_c(2\alpha)$). Dále za platnosti H_0 je rozdělení statistiky VW symetrické kolem její střední hodnoty $\mathrm{E}VW = 0$ a rozptyl je dán následovně:

$$\mathrm{var}VW = \frac{mn}{(m+n)(m+n+1)} \sum_{i=1}^{m+n} \left[\Phi^{-1}\left(\frac{i}{m+n+1}\right)\right]^2. \tag{1.7}$$

Asymptoticky, pro $m, n \to \infty$, pak platí

$$U_{VW} = \frac{VW}{\sqrt{\mathrm{var}VW}} \sim N(0,1) \tag{1.8}$$

a v případě oboustranného testu pro $|U_{VW}| \geq u(1 - \frac{\alpha}{2})$ H_0 zamítáme. Pro pravostrannou alternativu, $\Delta > 0$, svědčí větší hodnoty statistiky U_{VW} a budeme zamítat, když $U_{VW} \geq u(1 - \alpha)$. Naopak pro levostrannou alternativu, $\Delta < 0$, svědčí hodnoty menší a zamítáme pro $U_{VW} \leq u(\alpha)$. Při použití statistického softwaru zvolíme funkci `qn.test`, nastavíme si, že chceme použít Van der Waerdenův test, dále testovanou alternativu a získáme p-hodnotu, podle které rozhodneme o zamítnutí nebo nezamítnutí H_0.

V případě malého počtu shod se postup testování nemění a shodám jsou přiřazena průměrná pořadí, se kterými se dále pracuje. V případě velkého počtu shod již tabelované kritické hodnoty a normální aproximaci není možné použít. Avšak odvodit přesné rozdělení testové statistiky je náročné, a proto zde budeme řešit jen příklady, kdy je možné tabelovaných hodnot a normální aproximace využít, popř. budeme počítat s jistou nepřesností.

1.3 Mediánový test

Testová statistika má tvar

$$M = \sum_{j=1}^{n} \frac{1}{2} \left\{ sign \left[S_j - \frac{1}{2}(m+n+1) \right] + 1 \right\} \tag{1.9}$$

a za H_0 má symetrické rozdělení kolem své střední hodnoty $\mathrm{E}M = \frac{n}{2}$. Dále lze určit její rozptyl:

$$\mathrm{var}M = \begin{cases} \dfrac{mn}{4(m+n+1)}, & \text{pro } m+n \text{ sudé}, \\[2ex] \dfrac{mn}{4(m+n)}, & \text{pro } m+n \text{ liché}. \end{cases}$$

Asymptoticky, pro $m, n \to \infty$, za platnosti H_0 platí:

$$U_M = \frac{M - \mathrm{E}M}{\sqrt{\mathrm{var}M}} \sim N(0,1), \tag{1.10}$$

kdy nulovou hypotézu budeme zamítat ve prospěch oboustranné alternativy, jestliže $|U_M| \geq u(1-\frac{\alpha}{2})$. Máme-li pravostrannou alternativu, $\Delta > 0$, pak H_0 zamítáme, když $U_M \geq u(1-\alpha)$. V případě levostranné alternativy, $\Delta < 0$, zamítneme nulovou hypotézu, jestliže $U_M \leq u(\alpha)$. Lze dokázat, že statistika M má hypergeometrické rozdělení, avšak v tomto textu ji budeme používat pouze pro výpočet statistiky U_M. Co se týče softwaru R, neexistuje pro mediánový test definovaná funkce jako pro předešlé dva testy. U příkladu se tedy omezíme pouze na normální aproximaci.

Mediánový test je vhodné použít pro tzv. cenzorované výběry, což jsou výběry s extrémními hodnotami na obou koncích, o kterých víme pouze to, že přesahují určitou mez. Obecně má test vůči předešlým dvěma testům menší sílu, tedy pravděpodobnost, že zamítneme nulovou hypotézu, když neplatí, je menší, ale jak uvádí [3], v případě testování dat, které pochází z dvojitě exponenciálního rozdělení, může být použití mediánového testu vhodnější.

Souhrnem, v případě výběrů z normálního rozdělení je samozřejmostí volit klasický dvouvýběrový t-test, který je nejsilnější za předpokladu normality. V případě, kdy je větší počet odlehlých hodnot nebo je porušena normalita, je vhodnější volbou dvouvýběrový Wilcoxonův test. Van der Waerdenův test má poměrně velkou sílu pro data z rozdělení velmi blízké normálnímu, avšak pro složitost výpočtu může být Wilcoxonův test vyhovující. Mediánový test je pak mnohem silnější pro data z rozdělení s těžkými chvosty, jako je např. dvojitě exponenciální rozdělení. V případě shod je třeba použít upravených statistik a kritických hodnot. V rámci pozorování se shodám snažíme vyhnout, ale mnohdy nemáme na výběr a je třeba data se shodami analyzovat. Usnadněním pak může být použití softwaru za předpokladu správně použitých funkcí.

1.4 Aplikace testů

V této podkapitole si pro různá data předvedeme postup testování, jak pro stejné rozsahy výběrů, tak pro různé. Ukážeme si postup pro případy se shodami i beze shod a budeme testovat proti jednostranným i oboustranným alternativám. Testy budeme provádět na hladině $\alpha = 0,05$, nebude-li řečeno jinak.

Příklad 1.1. U žen a mužů bylo zkoumáno, kolik vlastní párů bot. Můžeme říci, že ženy vlastní více párů bot než muži? Data byla sesbírána vlastním dotazníkem pomocí sociálních sítí.

ženy:	20,	24,	11,	25,	18,	32,	14,	37,	9,	28
muži:	5,	7,	2,	3,	10,	4,	8,	13		

Řešení. Příklad koresponduje se situací (1.a), která byla popsána na začátku kapitoly. Potenciální porušení předpokladu (A1.1) o náhodném výběru by mohlo být způsobeno menší rozmanitostí výběru. Sociální sítě totiž nepoužívají všichni a pokud by dotazník vyplnilo pět kamarádek, které jsou zapáleny do módy a nakupují spolu, můžeme očekávat, že počty párů bot budou na sobě závislé. Dále součástí (A1.1) je předpoklad spojitosti, který je zde sice porušen, ale protože se zde nevyskytují žádné shody, porušení spojitosti nevadí. Předpoklad (A1.2) o nezávislosti mezi muži a ženami by mohl být porušen tehdy, pokud by dotazník vyplnili např. partneři nebo manželé, kteří mají společný domácí rozpočet, a nakupování je ovlivněno výší společného výdělku. Protože byl dotazník anonymní a jediné, co respondenti uváděli, bylo pohlaví, nemůžeme vědět, zda u některých dvojic hrozí závislost pozorování a budeme tedy považovat nezávislost za splněnou. Co se týče předpokladu (A1.3) o pevném rozsahu výběru, dotazníkem bylo sesbíráno více jak

50 hodnot, přičemž ženy odpovídaly častěji. Pro účely této práce bylo náhodně vybráno 10 pozorování žen a 8 pozorování mužů, aby bylo snadnější ukázat ruční počítání testových statistik. Protože si ale takto určíme rozsah výběru a jakkoli s ním nehýbeme, můžeme předpoklad (A1.3) považovat za splněný.

Označme $X_1, ..., X_{10}$ počty párů bot žen, $Y_1, ..., Y_8$ počty párů bot mužů a je zřejmé, že $m = 10, n = 8$. Slovní formulace hypotéz je následující:

$$H_0 : \text{počet párů bot u žen a mužů se neliší,}$$
$$H_1 : \text{ženy vlastní více párů bot než muži.}$$

Matematicky se jedná o levostrannou alternativu $H_1 : G(x) = F(x - \Delta), \Delta < 0$, protože očekáváme, že hodnoty $Y_1, ..., Y_8$ budou nižší.

Utvoříme sdružený výběr, který seřadíme podle velikosti a pozorováním přiřadíme pořadí. Hodnoty pozorování $X_1, ..., X_{10}$ a jim příslušná pořadí $R_1, ..., R_{10}$ jsou vyznačeny tučně.

pozorování	2	3	4	5	7	8	**9**	10	**11**
pořadí	1	2	3	4	5	6	**7**	8	**9**
pozorování	13	**14**	**18**	**20**	**24**	**25**	**28**	**32**	**37**
pořadí	10	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**

Nyní je vše nachystáno pro výpočet testových statistik a začneme dvouvýběrovým Wilcoxonovým testem. Protože máme levostranný test, nebylo by třeba počítat statistiky T_2 a U_2, ale pro ukázku jejich vypočtení uvedeme.

Pomocí (1.1) vypočteme T_1 a T_2:

$$T_1 = 7 + 9 + 11 + 12 + 13 + 14 + 15 + 16 + 17 + 18 = 132,$$
$$T_2 = 1 + 2 + 3 + 4 + 5 + 6 + 8 + 10 = 39.$$

Správný výpočet můžeme ověřit pomocí vztahu $T_1 + T_2 = \frac{1}{2}(m + n)(m + n + 1)$, který vyplývá z formulace statistik T_1, T_2. Zde $\frac{1}{2}(10 + 8)(10 + 8 + 1) = 171$, což se rovná součtu statistik T_1 a T_2. Dále užitím (1.2) vypočteme statistiky U_1 a U_2:

$$U_1 = 10 \cdot 8 + \tfrac{1}{2} \cdot 10(10 + 1) - 132 = 3,$$
$$U_2 = 10 \cdot 8 + \tfrac{1}{2} \cdot 8(8 + 1) - 39 = 77.$$

Tabelovaná kritická hodnota je $W_c(0,1) = 20$, se kterou porovnáme U_1. Vzhledem k tomu, že $U_1 < W_c(0,1)$, zamítáme nulovou hypotézu na hladině $\alpha = 0,05$ proti jednostranné alternativě. V případě použití normální aproximace využijeme (1.4) a dostaneme:

$$\mathrm{E}U_1 = 40, \quad \mathrm{var}U_1 = 126,6667, \quad U_{MW} = -3,2875.$$

Hodnotu U_{MW} porovnáme s $u(0,05) = -1,6449$, tedy $U_{MW} < u(0,05)$ a nulovou hypotézu proti jednostranné alternativě zamítáme i tímto způsobem. Konečně použijeme-li software R, resp. jeho funkci `wilcox.test`, dostaneme p-hodnotu 0,00016, která také velmi přesvědčivě zamítá H_0 proti jednostranné alternativě na hladině testu 0,05. Pokud se nevyskytují ve výběrech shody, funkce `wilcox.test` používá pro stanovení kritického oboru přesné rozdělení veličiny U_1. Nepoužívá asymptotické rozdělení, a proto je výsledek přesnější než při použití normální aproximace. Svědčí o tom i p-hodnota pro normální aproximaci, která je rovna 0.00051 a je větší oproti p-hodnotě vypočtené softwarem. To znamená, že test pomocí normální aproximace méně přesvědčivě zamítá H_0. Ručně vypočítaná p-hodnota pro statistiku U_1 vychází $P(T_1 \geq 132) = \frac{7}{\binom{18}{10}} = 0,00016$, kde 7 je počet takových součtů deseti libovolných pořadí, které jsou větší nebo rovny 132. Ručně vypočítáná p-hodnota vychází stejně jako p-hodnota vypočtená softwarem.

Pro naši situaci výsledky dvouvýběrového Wilcoxonova testu znamenají, že zamítáme shodnost v počtu párů bot u mužů a u žen a můžeme se domnívat, že ženy vlastní více párů bot.

Přistoupíme k použití Van der Waerdenova testu. Použitím vzorce (1.6) dostáváme

$$VW = \Phi^{-1}\left(\frac{1}{19}\right) + \Phi^{-1}\left(\frac{2}{19}\right) + ... + \Phi^{-1}\left(\frac{8}{19}\right) + \Phi^{-1}\left(\frac{10}{19}\right) = -5,9261.$$

Tabelované kritické hodnoty jsou k dispozici pouze pro oboustranný test s hladinou $\alpha = 0,05$, tedy jednostranný test provedeme na hladině $\alpha = 0,025$. Z tabulek vyčteme, že $VW_c(0,05) = 3,6$, a protože platí $VW < -VW_c(0,05)$, nulovou hypotézu na hladině významnosti 0,025 zamítáme. Zamítnutí na přísnější hladině testu nám říká, že bychom H_0 zamítli i pro $\alpha = 0,05$, a proto pro normální aproximaci Van der Waerdenova testu a výpočet v softwaru zůstaneme u $\alpha = 0,05$. Dále užitím (1.7) spočteme

$$\mathrm{var}VW = \frac{10 \cdot 8}{(10+8)(10+8+1)} \sum_{i=1}^{18} \left[\Phi^{-1}\left(\frac{i}{10+8+1}\right)\right]^2 = 3,1035$$

a dosazením do vzorce (1.8) dostaneme

$$U_{VW} = \frac{-5,9261}{\sqrt{3,1035}} = -3,3639.$$

Hodnotu U_{VW} porovnáme s $u(0,05) = -1,6449$, tedy $U_{VW} < u(0,05)$ a na hladině testu 0,05 zamítáme H_0 i pro normální aproximaci. Do třetice použijeme funkci `qn.test` ve statistickém softwaru. Výstupem je p-hodnota 0,0015, pro kterou zamítáme nulovou hypotézu na hladině 0,05. P-hodnota pro test s normální aproximací vyšla 0,0004 a je menší než p-hodnota daná funkcí `qn.test`. Normální aproximace zamítá nulovou hypotézu o něco přesvědčivěji.

Van der Waerdenův test zamítl nulovou hypotézu proti jednostranné alternativě a opět zamítáme shodu v počtu párů bot u mužů a u žen a přikláníme se k alternativě, že ženy vlastní více párů.

Zbývá ještě použít na data test mediánový. Užitím (1.9) dostáváme

$$M = \frac{1}{2}\left\{\left[sign(1 - \tfrac{19}{2}) + 1\right] + ... + \left[sign(10 - \tfrac{19}{2}) + 1\right]\right\} = 1.$$

Víme, že platí $\mathrm{E}M = \frac{n}{2}$ a pro $m+n$ sudé $\mathrm{var}M = \frac{mn}{4(m+n+1)}$. Dosazením dostaneme $\mathrm{E}M = 4$ a $\mathrm{var}M = 1,0526$. Podle vztahu (1.10) určíme statistiku U_M:

$$U_M = \frac{1-4}{\sqrt{1,0526}} = -2,9240.$$

Hodnotu U_M porovnáme s $u(0,05) = -1,6449$, a protože $U_M < u(0,05)$, zamítáme H_0 na hladině testu 0,05 proti jednostranné alternativě. P-hodnota vypočítaná pomocí statistiky U_M vyšla 0,0017, tedy mediánový test zamítá nulovou hypotézu méně přesvědčivěji než předešlé dva testy.

Pro tento příklad jsme nulovou hypotézu zamítli všemi třemi testy. Není proto žádný důvod se domnívat, že by ženy vlastnily stejný počet párů bot jako muži. Naopak se potvrdila domněnka, že ženy mají více párů bot.

Příklad 1.2. Máme k dispozici data s údaji o bolestivosti ramene po operaci, převzato z [7]. Jedné skupině bylo provedeno ošetření navíc, druhá skupina je kontrolní bez ošetření. Pacienti měli na stupnici od 1=malá bolestivost až po 5=velká bolestivost ohodnotit prožitek bolesti. Lze se domnívat, že ošetření snižuje bolestivost?

ošetření:	1,	2,	1,	2,	1,	2,	2,	4
kontrola:	3,	5,	1,	2,	4,	4,	3,	5

Řešení. Příklad je v souladu se situací (1.b). Předpoklad (A1.1) o náhodných výběrech ze spojitých rozdělení je splněn pouze částečně. Určitě totiž hodnoty nepochází ze spojitého rozdělení, a navíc se zde vyskytují shody. Budeme tedy přiřazovat průměrná pořadí a opatrněji zacházet s testovými statistikami. Nezávislost jak v rámci skupiny, tak mezi skupinami by mohla být potenciálně porušena tím, že by pacienti věděli, do které skupiny patří a mohli se snadněji ovlivňovat. Naopak nezávislost podporuje to, že každý jedinec má individuální prožívání bolesti. Co se týče rozsahu výběru, opět byla data převzata z většího souboru dat, ale stejně jako v minulém příkladu, můžeme rozsah považovat za pevný.

Označme $X_1, ..., X_8$ hodnoty výběru s ošetřením a $Y_1, ..., Y_8$ hodnoty kontrolního výběru. Dále je jasné, že $m = n = 8$. Slovně formulujeme nulovou a alternativní hypotézu:

$$H_0 : \text{ošetření nemá vliv na bolestivost,}$$
$$H_1 : \text{ošetření bolestivost ramene snižuje.}$$

Máme tedy pravostranný test s $H_1 : G(x) = F(x - \Delta), \Delta > 0$, protože očekáváme, že hodnoty kontrolní skupiny budou větší. Opět utvoříme sdružený výběr, seřadíme jej podle velikosti a přiřadíme průměrná pořadí. Hodnoty pozorování $X_1, ..., X_8$ a jim příslušných pořadí $R_1, ..., R_8$ jsou vyznačeny tučně.

pozorování	**1**	**1**	**1**	1	**2**	**2**	**2**	**2**
pořadí	**2,5**	**2,5**	**2,5**	2,5	**7**	**7**	**7**	**7**
pozorování	2	3	3	**4**	4	4	5	5
pořadí	7	10,5	10,5	**13**	13	13	15,5	15,5

Pořadí 2,5 pro pozorování o hodnotě 1 jsme dosáhli tak, že klasická pořadí pro první čtyři hodnoty pozorování jsou 1, 2, 3, 4 a metoda průměrných pořadí tyto hodnoty zprůměruje, a proto 2,5=(1+2+3+4)/4. Obdobně pak pro ostatní hodnoty.

Začněme opět dvouvýběrovým Wilcoxonovým testem. Spočítáme statistiky T_1, T_2, U_1, U_2 podle (1.1) a (1.2):

$$T_1 = 48,5, \quad T_2 = 87,5, \quad U_1 = 51,5, \quad U_2 = 12,5.$$

Pokud by se použil postup pro případ výběrů beze shod, pak kritická hodnota z tabulek $W_c(0,1) = 15$ se porovná s U_2. Protože $U_2 < W_c$, H_0 bychom na hladině 0,05 zamítli proti jednostranné alternativě. Při použití normální aproximace dosadíme do vzorců (1.4) a dostaneme hodnotu testové statistiky

$$U_{MW} = 2,0479,$$

kterou porovnáme s $u(0,95) = 1,6449$. Jelikož $U_{MW} > u(0,95)$, nulovou hypotézu proti jednostranné alternativě pomocí normální aproximace také zamítáme. Použijeme-li upravenou statistiku (1.5), dostaneme $U^*_{MW} = 2,1044$ a H_0 opět zamítáme. Užitím funkce `wilcox.exact`, která je určená právě pro případy se shodami, dostaneme p-hodnotu 0,0194 a na hladině testu 0,05 můžeme nulovou hypotézu zamítnout.

P-hodnota pro normální aproximaci, tedy statistiku U_{MW}, vyjde 0,0203 a je větší než p-hodnota daná softwarem. To nám naznačuje, že použitím funkce `wilcox.exact`, která je přímo definovaná pro shody ve výběru, dostáváme přesvědčivější výsledek. Konečně p-hodnota testu pomocí upravené statistiky U^*_{MW} vyjde 0,0177 a je velmi blízká p-hodnotě vypočítané softwarem a zároveň je i menší, tedy nulovou hypotézu zamítá přesvědčivěji. Ve funkci `wilcox.exact`, můžeme pak nastavit, aby použila normální aproximaci upravenou pro shody, místo přesného rozdělení, a potom vyjde p-hodnota 0,0177, která vychází stejně jako p-hodnota pro test založený na U^*_{MW}. Dvouvýběrovým Wilcoxonovým testem pokaždé zamítáme nulovou hypotézu ve prospěch jednostranné alternativy, a tedy se můžeme domnívat, že ošetření snižuje bolestivost ramene.

Nyní použijeme Van der Waerdenův test. Dosazením do vzorců (1.6), (1.7) a (1.8) dostaneme

$$VW = 3,4729, \quad U_{VW} = 2,1147,$$

kdy pro statistiku VW opět uvažujeme hladinu testu 0,025. Pro pravostranný test je kritická hodnota $VW_c(0,05) = 3,39$, a na této přísnější hladině nulovou hypotézu zamítáme. Hodnotu testové statistiky, určenou pomocí normální aproximace, porovnáme s $u(0,95) = 1,6449$. Vidíme, že $U_{VW} > u(0,95)$ a na hladině 0,05 zamítáme nulovou hypotézu proti jednostranné alternativě. Použitím statistického softwaru dostáváme p-hodnotu 0,0389 a na hladině 0,05 H_0 také zamítáme. P-hodnota pro test pomocí normální aproximace vychází 0,0172, tedy testem s normální aproximací zamítáme H_0 přesvědčivěji. Užitím Van der Waerdenova testu docházíme ke stejnému závěru jako u dvouvýběrového Wilcoxonova testu - ošetření snižuje bolestivost ramene po operaci.

Nakonec použijeme mediánový test. Pro výpočet statistiky U_M použijeme vztahy (1.9), (1.10) a dostaneme $U_M = 2,0616$. Testovou statistiku srovnáme s $u(0,95) = 1,6449$ a nulovou hypotézu na hladině 0,05 zamítáme i tímto testem. P-hodnota pro test s normální aproximací je 0,0196 a ve srovnání s předešlými testy zamítá nulovou hypotézu v podstatě stejně přesvědčivě.

Pro všechny tři testy byla nulová hypotéza zamítnuta ve prospěch pravostranné alternativy. Můžeme se tedy domnívat, že ošetření skutečně snižuje bolestivost ramene. Přičemž u testů, kde jsme nepoužili upravených statistik pro shody, je třeba počítat s větší nepřesností.

Příklad 1.3. Těsto na pečení je třeba zpracovat do určité konzistence. Pět dávek těsta bylo zpracováno mixérem A a pět jiných dávek mixérem B. Sledoval se čas, za jak dlouho mixér těsto do požadované podoby zpracuje, časy jsou uvedeny v minutách, převzato z [3, str. 282]. Je rozdíl v délce zpracování těsta?

mixér A:	7,3,	6,9,	7,2,	7,8,	7,2
mixér B:	7,4,	6,8,	6,9,	6,7,	7,1

Řešení. Tomuto příkladu lze porozumět jako situaci (1.c), kdy srovnáváme efekt dvou různých ošetření. Zde ošetřením rozumíme použití mixéru A nebo mixéru B. Tentokrát bychom čas mohli chápat jako spojitou veličinu, ale vzhledem k zaokrouhlení se objevují i shody. Takže bude třeba přiřadit průměrná pořadí a opatrně zacházet s testovými statistikami. Nezávislost mezi pozorováními v rámci jednoho mixéru by mohla být porušena nestejnými podmínkami - např. jiný poměr přísad v těstu nebo opotřebení některých součástek. Mezi mixéry by pak mohlo být porušení nezávislosti způsobeno lidskou chybou při odečítání času, např. pozorovatel mixéru A je důkladnější než pozorovatel mixéru B. Rozsahy jsou zde pevné, předpoklad (A1.3) tedy není porušen.

Označme $X_1, ..., X_5$ časy při použití mixéru A a $Y_1, ..., Y_5$ časy mixéru B. Platí, že $m = n = 5$ a slovně naformulujeme nulovou a alternativní hypotézu:

$$H_0 : \text{mixéry zpracují těsto stejně rychle,}$$
$$H_1 : \text{mixéry stejně rychle nepracují.}$$

Vzhledem k tomu, že nevíme, který mixér by mohl pracovat rychleji, budeme nulovou hypotézu testovat proti oboustranné alternativě $H_1 : G(x) = F(x - \Delta)$, $\Delta \neq 0$. Opět utvoříme sdružený výběr, seřadíme podle velikosti a přiřadíme průměrná pořadí. Hodnoty pozorování a pořadí příslušné výběru $X_1, ..., X_5$ jsou opět vyznačeny tučně.

pozorování	6,7	6,8	**6,9**	6,9	7,1	**7,2**	**7,2**	**7,3**	7,4	**7,8**
pořadí	1	2	**3,5**	3,5	5	**6,5**	**6,5**	**8**	9	**10**

Přistoupíme rovnou k dvouvýběrovému Wilcoxonovu testu. Užitím (1.1), (1.2) a (1.3) dostaneme

$$T_1 = 34,5, \quad T_2 = 20,5, \quad U_1 = 5,5, \quad U_2 = 19,5, \quad MW = 5,5$$

a platí, že $MW > W_c(0,05) = 2$, tedy nulovou hypotézu nelze na hladině 0,05 zamítnout. Pro získání testové statistiky založené na normální aproximaci použijeme (1.4) a dostáváme $U_{MW} = -1,4623$. Vidíme, že $|U_{MW}| < u(0,975) = 1,96$ a ani pro test vycházející z normální aproximace H_0 nelze zamítnout. Užitím upravené statistiky (1.5) dostáváme $U^*_{MW} = -1,4712$ a nulovou hypotézu opět nelze zamítnout. Funkce `wilcox.exact` použitím normální aproximace vypočetla p-hodnotu 0,1412 a nezamítáme ani tímto způsobem. P-hodnota pro test s upravenou statistikou U^*_{MW} je 0,1412 a opět je stejná jako p-hodnota vypočtená softwarem.

Dvouvýběrový Wilcoxonův test nezamítá na hladině 0,05 nulovou hypotézu proti oboustranné alternativě, a můžeme tvrdit, že mixér A i B pracují stejně rychle.

Nyní použijeme Van der Waerdenův test, kde dosazením do vzorců (1.6), (1.7), (1.8) dostaneme:

$$VW = -1,9222, \quad U_{VW} = -1,6171,$$

kdy pro statistiku VW je kritická hodnota oboustranného testu pro hladinu 0,05 $VW_c(0,05) = 2,6$, a protože $|VW| < VW_c(0,05)$, nulovou hypotézu nelze zamítnout. Statistiku U_{VW}, resp. její absolutní hodnotu, porovnáváme s $u(0,975) = 1,96$. Jelikož $|U_{VW}| < u(0,975)$, pak nulovou hypotézu opět na hladině 0,05 nelze zamítnout. Výstupem příkazu `qn.test` v softwaru je p-hodnota 0,1408 a ani tímto způsobem testování nezamítáme nulovou hypotézu na hladině 0,05. Van der Waerdenův test rovněž H_0 nezamítá a potvrzuje závěr, že mixéry pracují stejnou rychlostí, avšak vzhledem k shodám jsou zde nepřesnosti.

Testovou statistiku pro mediánový test určíme pomocí vzorců (1.9), (1.10) a vyjde $U_M = -1,99$, což znamená, že porovnáme-li s $u(0,975)$, pak na hladině 0,05 nulovou hypotézu velmi těsně zamítáme. Výsledek je poněkud překvapivý, protože pro většinu případů má mediánový test menší sílu. Vliv zde mohly mít shody nebo skutečné rozdělení dat. Mediánovým testem tedy nemůžeme potvrdit, že mixéry pracují stejnou rychlostí.

Dvouvýběrovým Wilcoxonovým a Van der Waerdenovým testem nulovou hypotézu nelze zamítnout a platí pro ně závěr, že mixér A i mixér B pracují stejně rychle. Mediánový test překvapivě H_0 zamítá a bylo jím stanoveno, že rychlost mixérů se liší. Protože mediánový test nejméně zohledňuje shody, na tomto místě by bylo lepší věřit výsledku dvouvýběrového Wilcoxonova nebo Van der Waerdenova testu.

Porovnáme-li použití testů pro příklady této kapitoly, pravděpodobně nejlepší volbou je dvouvýběrový Wilcoxonův test, za předpokladu, že nemáme k dispozici testy normality, ani jiný způsob jak zjistit původní rozdělení dat. Pokud bychom test normality provedli a u některé sady bychom velmi těsně zamítli, popř. nezamítli, byl by vhodnější Van der Waerdenův test, který je silnější pro rozdělení blízká normálnímu. Protože data z příkladů 1.1 a 1.2 jsou diskrétní povahy, určitě mají méně blíže normálnímu rozdělení než data z příkladu 1.3, kde bychom mohli mít podezření na normalitu. Obecně však za předpokladu neznámého spojitého rozdělení, by měl být první volbou dvouvýběrový Wilcoxonův test.

2 Hypotéza shody dvou populací proti ostatním alternativám

Stejně jako v 1. kapitole, nechť $X_1, X_2, ..., X_m$ a $Y_1, Y_2, ..., Y_n$ jsou na sobě nezávislé náhodné výběry, které po řadě pochází z rozdělení F a G a mají pevné rozsahy m a n. Budeme se i nadále držet značení z první kapitoly, takže $R_1, ..., R_m$ je pořadí veličin $X_1, ..., X_m$ v seřazeném sdruženém výběru $X_1, ..., X_m, Y_1, ..., Y_n$ a $S_1, ..., S_n$ je pořadí veličin $Y_1, ..., Y_n$. Tvar nulové hypotézy zůstává stejný a budeme ji testovat proti dvěma různým alternativám, zavedeno podle [6].

$H_0 : F = G$,

$H_1^o : F \neq G$ nebo

$H_1^m : G(x - \mu) = F(\frac{x-\mu}{\sigma}), \sigma \neq 1$.

První alternativou je obecná alternativa H_1^o a z jejího tvaru vyplývá, že budeme uvažovat pouze oboustranný test. Obecnou alternativu zvolíme, pokud nevíme, jak se můžou výběry lišit, tedy testuje libovolnou odlišnost. Druhá alternativa H_1^m se týká změny měřítka, kde μ je parametr polohy, kterým se nezabýváme a považujeme jej za stálý, a σ je parametr rozptylu, který nás zajímá. Testujeme-li proti oboustranné alternativě, pak chceme zjistit, zda se rozptyly výběrů liší. V případě jednostranné alternativy, pokud je $\sigma > 1$ (resp. $\sigma < 1$), pak očekáváme, že výběr $Y_1, ..., Y_n$ bude mít větší (resp. menší) rozptyl než $X_1, ..., X_m$.

Vzhledem k nulové hypotéze lze popsaný matematický model aplikovat na situace (1.a), (1.b), (1.c) z předešlé kapitoly a změní se jen volba alternativy, která vyplývá z našeho očekávání. Pro většinu případů se sice více hodí alternativa posunutí, ale někdy nemusíme vědět, jakým způsobem se výběry liší, a potom je vhodnější volbou obecná alternativa. Alternativu změny měřítka pak použijeme, když máme podezření na rozdílnou variabilitu výběrů.

V praxi může být obecná alternativa použita např. při zkoumání nové učební metody, u které nevíme, jak bude na studenty působit. Měli bychom tedy jednu kontrolní skupinu, která by se vyučovala klasicky, a druhou skupinu, která by se učila podle nové metody. Metoda by mohla zlepšit výsledky studentů ve smyslu, že mezi nadprůměrným a průměrným studentem bude pořád stejný rozdíl (alternativa posunutí), nebo by mohla sjednotit všechny studenty na průměrnou úroveň (alternativa změny měřítka). Např. by se nová učební metoda věnovala mnohem více podprůměrným žákům a naopak skoro vůbec těm nadprůměrným. Znalostní testy by se tedy zlepšily u podprůměrných žáků a mohly by se zhoršit u těch původně nadprůměrných. Pokud tedy nemáme odhad, jak by mohla nová metoda ovlivnit výsledky, vhodnější volbou je obecná alternativa. Analogicky bychom mohli takto zkoumat vliv nového léku, u kterého bychom si nebyli předem jisti, jakým způsobem bude působit. Testy pro obecnou alternativu by se také daly použít pro předběžné testování dat, než si stanovíme alternativu posunutí. Zejména dvouvýběrový Wilcoxonův test má menší sílu pro obecnou alternativu, a proto by mohl nezamítnout nulovou hypotézu v případě, že by se rozdělení lišila jinak než v posunutí. Samozřejmě test pro obecnou alternativu nemá takovou sílu proti alternativě změny měřítka jako test, který je přímo pro ni konstruován, to stejné platí i pro testy formulované pro alternativu posunutí.

Jako příklad pro alternativu změny variability se hodí náhodné chyby měření. Řekněme, že máme dva různé přístroje, kterými měříme určitou veličinu a po provedení měření určíme chyby. Chyby přístrojů budou tvořit dva nezávislé náhodné výběry. Uvažujeme zde náhodné chyby, kdy předpokládáme, že parametr μ je nulový. Bylo by možné uvažovat i systematické chyby za předpokladu, že bychom opět znali hodnotu parametru μ a pro oba přístroje by byla hodnota stejná, resp. přesnou hodnotu nepotřebujeme, ale musíme vědět, zda je stejná. Chceme dále testovat, zda přístroje měří stejně přesně, tedy jestli rozpětí náhodných chyb u obou přístrojů je stejné. Máme tudíž nulovou hypotézu, že výběry pocházejí ze stejného rozdělení, proti oboustranné alternativě, že se liší v měřítku. Dalším příkladem by mohlo být sledování vyhraněnosti názorů mezi muži a ženami na určitou věc. Mohl by se sestavit dotazník, který by byl vyhodnocen počtem bodů od 0 do 50. Na datech by se provedl test s alternativou změny měřítka a mohli bychom zjistit, zda jsou muži více vyhraněnější v názorech než ženy nebo naopak. Neboli pokud by muži měli hodnoty od 25 do 35 a ženy od 5 do 45, mohli bychom testovat, že ženy jsou méně vyhraněnější, tedy že hodnoty u žen jsou více rozptýlené.

Potřebné předpoklady shrnují následující body:

- (A2.1): $X_1, ..., X_m$ a $Y_1, ..., Y_n$ jsou náhodné výběry ze spojitých rozdělení;
- (A2.2): výběry jsou nezávislé;
- (A2.3): rozsahy výběrů m a n jsou pevné;
- (A2.4): při testování H_1^m je parametr μ pro rozdělení F a G stejný.

Teoretické zázemí testů této kapitoly pochází zejména z [1, 5].

2.1 Dvouvýběrový Kolmogorovův-Smirnovův test

Nechť F_m a G_n jsou empirické distribuční funkce výběrů $X_1, ..., X_m$ a $Y_1, ..., Y_n$. Platí

$$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} I[X_i \leq x], \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[Y_i \leq x], \tag{2.1}$$

kde I je indikátorová funkce a $x \in \mathbb{R}$. S rostoucím m a n se $F_m(x)$ a $G_n(x)$ blíží skutečným distribučním funkcím. Testová statistika pro dvouvýběrový Kolmogorovův-Smirnovův test má tvar

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)| \tag{2.2}$$

a používá se pro testování obecné alternativy $H_1^o : F \neq G$. Při malých m a n (doporučeno, je-li $m + n < 35$) se hodnota $D_{m,n}$ porovná s tabelovanými hodnotami $D_{m,n}(\alpha)$, které pochází z [1]. Platí-li $D_{m,n} \geq D_{m,n}(\alpha)$, potom zamítáme nulovou hypotézu na hladině α. Pro větší rozsahy se pak používá aproximace, kdy aproximativní kritickou hodnotu vypočítáme jako:

$$D_{m,n}^c(\alpha) = \sqrt{\frac{m+n}{2mn}\ln\frac{2}{\alpha}} \tag{2.3}$$

a nulovou hypotézu zamítáme, jestliže $D_{m,n} \geq D_{m,n}^c(\alpha)$. Její podrobné odvození je možno dohledat v [1, str. 105]. V statistickém softwaru je k dispozici funkce `ks.test`, která vypočte hodnotu statistiky $D_{m,n}$ a p-hodnotu.

Z formulace statistiky $D_{m,n}$ vyplývá, že vždy budou pro alternativu svědčit jen velké hodnoty statistiky. Test obecné alternativy pomocí Kolmogorovova-Smirnovova testu je tedy formulován s pravostranným kritickým oborem. V případě shod použití tabelovaných kritických hodnot nebo aproximativních vede k nepřesnému výsledku a menší síle testu.

Při použití dvouvýběrového Kolmogorovova-Smirnovova testu je dobré si vykreslit empirické distribuční funkce obou výběrů. Zamítneme-li nulovou hypotézu, může nám vzájemná poloha funkcí naznačit, zda se výběry liší spíše v poloze nebo v měřítku. Pokud by například byla funkce F_m nad funkcí G_n, můžeme mít podezření na rozdílnost v poloze. Pokud by obdobně byla funkce F_m strmější než G_n, pak bychom mohli uvažovat o rozdílnosti v měřítku. Existuje i jednovýběrová varianta Kolmogorovova-Smirnovova testu, která se pak používá k testování normality.

2.2 Siegelův-Tukeyho test

Používá se pro alternativu změny měřítka a je založen na podobném principu jako dvouvýběrový Wilcoxonův test, dokonce je i možné pro vyhodnocení použít tabulky Wilcoxonova testu pro statistiku T_1, které najdeme v [4]. Pro Siegelův-Tukeyho test je však potřeba jiného přiřazování pořadí. Opět uvažujeme sdružený výběr seřazený podle velikosti. Nyní přiřaďme pořadí 1 nejmenšímu prvku, pořadí 2 největšímu, dále pořadí 3 druhému největšímu, pořadí 4 druhému nejmenšímu atd. (bude ukázáno na příkladu). Označíme součty pořadí příslušné výběrům $X_1, ..., X_m$ a $Y_1, ..., Y_n$:

$$ST_1 = \sum_{i=1}^{m} R_i, \quad ST_2 = \sum_{i=1}^{n} S_i. \tag{2.4}$$

ST_1 je testová statistika a její rozdělení je shodné s rozdělením statistiky T_1 dvouvýběrového Wilcoxonova testu, a proto použijeme kritické hodnoty pro statistiku T_1. V tabulkách jsou uvedeny pro dané m, n a α dvě kritické hodnoty $W_{c_1}(\alpha)$ a $W_{c_2}(\alpha)$, kde $W_{c_1}(\alpha) < W_{c_2}(\alpha)$. Statistiky ST_1 a T_1 totiž nejsou symetrické kolem nuly, a proto je potřeba dvou kritických hodnot, které hlídají příliš velké i příliš malé hodnoty. Dvojice kritických hodnot $W_{c_1}(\alpha), W_{c_2}(\alpha)$ pro oboustranný test dodržují hladinu testu α. Pro jednostranné testy budeme brát pouze jednu kritickou hodnotu a budeme ji odečítat z tabulek oboustranného testu, který bude mít hladinu 2α.

V případě oboustranného testu, jestliže $ST_1 \leq W_{c_1}(\alpha)$ nebo $ST_1 \geq W_{c_2}(\alpha)$, pak zamítáme nulovou hypotézu na příslušné hladině α. V případě pravostranné alternativy ($\sigma > 1$) očekáváme, že hodnoty testové statistiky ST_1 budou větší, a tudíž zamítneme H_0 na hladině α, když $ST_1 \geq W_{c_2}(2\alpha)$. Pro levostrannou alternativu ($\sigma < 1$) budou svědčit menší hodnoty ST_1, tedy zamítneme nulovou hypotézu pro $ST_1 \leq W_{c_1}(2\alpha)$.

Výskyt shod vyžaduje úpravu testové statistiky ST_1, která je shodná s úpravou dvouvýběrového Wilcoxonova testu a je možno ji dohledat v [5]. Software R nemá definovanou funkci pro tento test, avšak je možné ručně přiřadit pozorováním pořadí, na která pak aplikovat funkci `wilcox.test` popř. `wilcox.exact` v případě shod.

2.3 Aplikace testů

Obecná alternativa a alternativa změny měřítka se v praxi vyskytují méně často než alternativa posunutí, resp. pro většinu situací se více hodí alternativa posunutí a je i snadněji interpretovatelná. Postup testování si opět ukážeme na příkladech, přičemž testy budou prováděny na hladině $\alpha = 0,05$, nebude-li řečeno jinak.

Příklad 2.1. Pro ilustraci Kolmogorovova-Smirnovova testu použijeme data z příkladu 1.1 o počtu párů bot.

ženy:	20,	24,	11,	25,	18,	32,	14,	37,	9,	28
muži:	5,	7,	2,	3,	10,	4,	8,	13		

Řešení. Budeme se držet značení z příkladu 1.1, a tedy označíme $X_1, ..., X_{10}$ pozorování výběru žen a $Y_1, ..., Y_8$ hodnoty výběru mužů, dále $m = 10, n = 8$. Formulace hypotéz je následující:

$$H_0 : \text{muži a ženy se neliší v počtu vlastněných bot},$$
$$H_1^o : \text{počet párů bot se liší libovolně}.$$

Hodnoty empirických distribučních funkcí, vypočtené podle (2.1), pro všech 18 hodnot jsou zapsány níže, tučně jsou zvýrazněny hodnoty x odpovídající výběru $X_1, ..., X_{10}$.

x:	2	3	4	5	7	8	**9**	10	**11**
$F_{10}(x)$:	0	0	0	0	0	0	1/10	1/10	2/10
$G_8(x)$:	1/8	2/8	3/8	4/8	5/8	6/8	6/8	7/8	7/8
x:	13	**14**	**18**	**20**	**24**	**25**	**28**	**32**	**37**
$F_{10}(x)$:	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	1
$G_8(x)$:	1	1	1	1	1	1	1	1	1

Např. výpočet hodnoty $F_{10}(14)$ vypadá následovně:

$$F_{10}(14) = \frac{1}{10}\sum_{i=1}^{10} I[X_i \leq 14] = \frac{1}{10}(0+0+1+0+0+0+1+0+1+0) = \frac{3}{10},$$

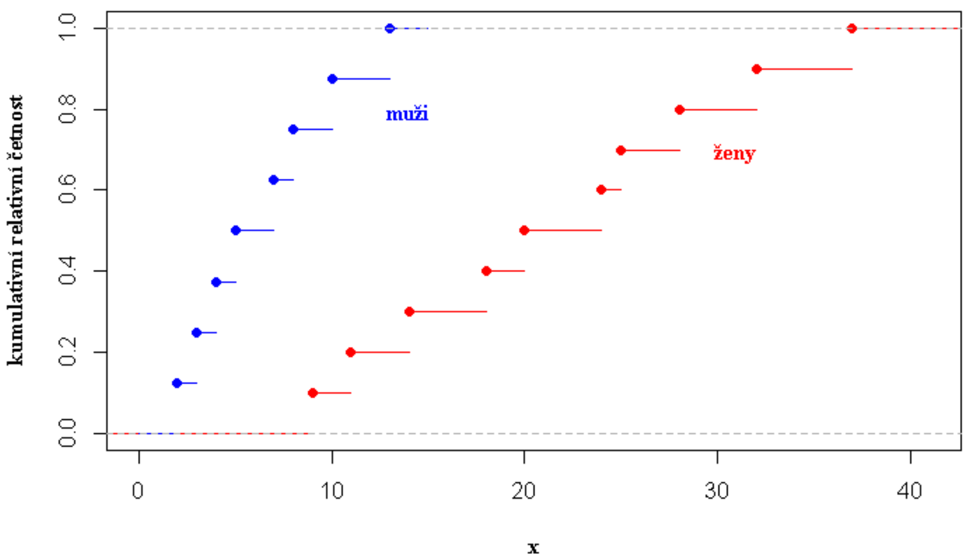
kdy identifikátorová funkce I má v tomto případě hodnotu 1 pro $X_i = 11, 14, 9$. Obdobně provedeme výpočet hodnot $G_n(x)$, např.

$$G_8(7) = \frac{1}{8}\sum_{i=1}^{8} I[Y_i \leq 7] = \frac{1}{10}(1+1+1+1+0+1+0+0) = \frac{5}{8},$$

identifikátorová funkce má hodnotu 1 pro $Y_i = 5, 7, 2, 3, 4$.

Z hodnot $F_{10}(x)$, $G_8(x)$ určíme jejich největší absolutní rozdíl, neboli určíme hodnotu testové statistiky $D_{10,8}$ podle vztahu (2.2). Největší absolutní rozdíl je pro $x = 13$, a tedy $D_{10,8} = 8/10$. $D_{10,8}$ porovnáme s tabelovanou kritickou hodnotou $D_{10,8}(0,05) = 6/10$, a protože $D_{10,8} \geq D_{10,8}(0,05)$, zamítáme nulovou hypotézu na hladině 0,05. Aproximativní kritická hodnota, kterou jsme vypočetli podle (2.3), je $D^c_{10,8}(0,05) = 0,6442$. Platí $D_{10,8} \geq D^c_{10,8}(0,05)$, takže nulovou hypotézu zamítáme i tímto způsobem. Užitím funkce `ks.test` v softwaru dostaneme p-hodnotu 0,0024 a zamítnutí nulové hypotézy souhlasí. Dvouvýběrový Wilcoxonův test zamítá v předchozí kapitole nulovou hypotézu přesvědčivěji. Mohlo by se tedy stát, že pro jistá data bychom dvouvýběrovým Wilcoxonovým testem zamítli H_0, zatímco Kolmogorovovým-Smirnovovým testem nikoli.

Podívejme se na nyní na grafy empirických distribučních funkcí. Pro vykreslení pomocí softwaru R použijeme funkci `ecdf`.



Graf 2.1: Empirické distribuční funkce

V grafu 2.1 můžeme vidět rozdíl mezi funkcemi a podezření na rozdílnost v poloze i v měřítku. V příkladu 1.1 jsme se přiklonili k alternativě změny polohy a vyzkoušíme i otestovat změnu měřítka. Formulace hypotéz je následující:

$$H_0 : \text{rozptýlenost v počtu párů bot se neliší,}$$
$$H_1^m : \text{u žen je větší rozptýlenost v počtu párů bot.}$$

Testujeme tedy alternativu $H_1 : G(x-\mu) = F(\frac{x-\mu}{\sigma}), \sigma < 1$. Protože je porušen předpoklad, že μ je pro oba výběry stejné, hodnoty pozorování vycentrujeme. Vypočteme průměr prvního výběru a ten odečteme od každého pozorování X_i, obdobně pro druhý výběr. Hodnoty a pořadí příslušné výběru $X_1, ..., X_{10}$ jsou vyznačeny tučně.

pozorování	**-12,8**	**-10,8**	**-7,8**	-4,5	**-3,8**	-3,5	-2,5	**-1,8**	-1,5
pořadí	**1**	**4**	**5**	8	**9**	12	13	**16**	17
pozorování	0,5	1,5	**2,2**	**3,2**	3,5	**6,2**	6,5	**10,2**	**15,2**
pořadí	18	15	**14**	**11**	10	**7**	6	**3**	**2**

Podle vztahu (2.4) určíme $ST_1 = 72$. Kritická hodnota $W_{c_1}(0,1) = 56$, a protože platí $ST_1 > W_{c_1}(0,1)$, nulovou hypotézu nelze na hladině 0,05 zamítnout. Přestože se podle grafu 2.1 zdálo, že by zde mohla být odlišnost i v měřítku, Siegelův-Tukeyho test ji nepotvrzuje.

Příklad 2.2. Na stabilizovaném zdroji elektrického napětí jsme nastavili pevnou hodnotu napětí, kterou jsme pak dvěma různými měřícími přístroji měřili. Na každém ze dvou přístrojů jsme provedli 5 měření a u každé změřené hodnoty jsme určili chybu. Měří přístroje stejně přesně?

přístroj A:	0,011,	0,013,	-0,007,	0,018,	-0,01
přístroj B:	0,024,	-0,015,	0,023,	0,014,	0,022

Řešení. Předpoklad (A2.1) je zde splněn, jelikož chyby měření jsou spojitou veličinou. Předpoklad náhodného výběru by mohl být teoreticky porušen, pokud by se na přístroji během měření opotřebovávala některá součástka a přístroj by vlivem poruchy mohl být např. méně citlivý. Hodnoty naměřené po poruše by pak mohly mít jiné rozdělení než ty, které jsme naměřili před poruchou. Nezávislost (A2.2) mezi přístroji je pak dána stejnými podmínkami pro měření a mohla by být porušena např. lidskou chybou nebo jiným zaokrouhlováním při výpočtu chyb, popř. výraznou změnou okolní teploty, která by mohla způsobit vychýlení měřícího přístroje. Pevné rozsahy (A2.3) jsou zde zřejmé, protože jsme si stanovili, že budeme měřit pětkrát, a tolikrát jsme měření učinili. Předpoklad (A2.4) by mohl být porušen tehdy, pokud by se vyskytovala neznámá systematická chyba na jednom z přístrojů. Předpokládáme však, že náhodné chyby mají parametr $\mu = 0$ u přístroje A i B.

Označme $X_1, ..., X_5$ hodnoty přístroje A a $Y_1, ..., Y_5$ hodnoty přístroje B a víme, že $m = n = 5$. Formulujeme nulovou a alternativní hypotézu:

$$H_0 : \text{přístroje měří stejně přesně (ve smyslu rozptýlenosti chyb)},$$
$$H_1^m : \text{přesnost měření je různá (jeden přístroj má více rozptýlené chyby)}.$$

Máme tedy oboustrannou alternativu $H_1^m : G(x-\mu) = F(\frac{x-\mu}{\sigma}), \sigma \neq 1$, protože nemáme podezření, který z přístrojů by mohl měřit přesněji. Sestavíme sdružený výběr, seřadíme podle velikosti a přiřadíme pořadí podle pravidla Siegelova-Tukeyho testu.

pozorování	-0,015	**-0,01**	**-0,007**	**0,011**	**0,013**
pořadí	1	**4**	**5**	**8**	**9**
pozorování	0,014	**0,018**	0,022	0,023	0,024
pořadí	10	**7**	6	3	2

Podle vztahu (2.4) spočteme testovou statistiku

$$ST_1 = 4 + 5 + 7 + 8 + 9 = 33, \quad ST_2 = 1 + 2 + 3 + 6 + 10 = 22.$$

Kritické hodnoty jsou $W_{c_1}(0,05) = 17$ a $W_{c_2}(0,05) = 38$, a protože $W_{c_1}(0,05) < ST_1 < W_{c_2}(0,05)$, nelze nulovou hypotézu na hladině 0,05 zamítnout. Při použití funkce `wilcox.test` přímo na přiřazená pořadí dostaneme p-hodnotu 0,3095 a nezamítnutí nulové hypotézy souhlasí. Na hladině testu 0,05 se můžeme domnívat, že přístroje měří stejně přesně.

3 Hypotéza symetrie

V první kapitole jsme měli jeden matematický model, který byl aplikovatelný na tři různé situace v praxi. V této kapitole budeme mít naopak dva možné matematické modely, které vyústí do stejného postupu, podle [5]. Model (3.a) tvoří náhodný výběr $Z_1, ..., Z_N$, který pochází z rozdělení F a má pevný rozsah N. Hypotézy pro tento model mají tvar

$H_0 : F$ je symetrické kolem 0,

$H_1 : F$ není symetrické kolem 0.

Pravostranná (resp. levostranná) alternativa pak testuje, zdali je rozdělení posunuto spíše ke kladným (resp. záporným) hodnotám. Situaci, kdy náhodný výběr $Z'_1, ..., Z'_N$ má nenulový známý medián $\widetilde{x}$ a chceme testovat hypotézu symetrie, lze vhodnou transformací převést na základní nulovou hypotézu, že rozdělení je symetrické kolem 0. Transformace spočívá v utvoření rozdílů $Z'_1 - \widetilde{x}$,...,$Z'_N - \widetilde{x}$, které dále odpovídají výběru $Z_1, ..., Z_N$ z modelu (3.a). Rozdíly tedy mají nulový medián a je možno testovat nulovou hypotézu o symetrii kolem 0. Je-li rozdělení symetrické, pak je středem symetrie medián a nulová hypotéza symetrie testuje symetrii kolem mediánu, který je roven nule.

Parametrickou obdobou hypotézy symetrie je hypotéza o mediánu (ozn. $\widetilde{x}$). Ta naopak předpokládá symetrii rozdělení a testuje hodnotu mediánu ($\widetilde{H}_0 : \widetilde{x} = \widetilde{x}_0$). Na obě možnosti se však používají stejné testy.

Matematický model (3.b) popisuje náhodný výběr $(X_1, Y_1), ...,$ (X_N, Y_N) z dvourozměrného rozdělení F_2 o pevném rozsahu N, kdy v rámci dvojice jsou pozorování obecně závislá. Nulová a alternativní hypotéza mají tvar

$H_0 : F_2(x, y) = F_2(y, x)$,

$H_1 : F_2$ není symetrické.

Levostranná alternativa testuje, že hodnoty X_i nabývají větších hodnot než příslušné Y_i, a pravostranná alternativa testuje, že větších hodnot nabývají Y_i.

Vhodně zvolenou transformací lze model (3.b) převést na model (3.a). Utvoří se rozdíly $Y_1 - X_1, ..., Y_N - X_N$ a dostáváme náhodný výběr, jehož rozdělení je za platnosti nulové hypotézy symetrické kolem 0, neboli požadovaný model (3.a).

Protože bude vždy jasné, o jaký model se jedná, budeme rozdíly značit $Z_1, ..., Z_N$ a jejich rozdělení označíme F.

V praxi lze velmi přesvědčivě model (3.a) aplikovat na chyby měření. Chtěli bychom testovat, zda přístroj měří přesně, ve smyslu, že náhodné chyby měření jsou symetrické kolem 0. Jinými slovy, symetrie pro tento příklad znamená, že pravděpodobnost, že nastane chyba $-\varepsilon$, je stejná jako pravděpodobnost, že nastane chyba ε. V případě, že by měl přístroj určitou systematickou chybu, jejíž hodnotu bychom znali, pak by se jednalo o výběr $Z'_1, ..., Z'_N$ a odečtením této hodnoty by se převedl na výběr $Z_1, ..., Z_N$. Chyby by bylo nejvhodnější otestovat proti oboustranné alternativě, protože většinou nemáme podezření, kterým směrem chyby více zanáší. Pokud bychom se domnívali, že chyby měření zanáší do kladných hodnot, pak bychom volili pravostrannou alternativu. Obdobně můžeme testovat, zda je možná symetrie v případě předčasně narozených dětí a dětí narozených po termínu. Hodnoty by se odvíjely od stanoveného termínu porodu a zaznamenali bychom např. -2, pokud by se dítě narodilo o dva dny předčasně, nebo +4, pokud by se narodilo čtyři dny po termínu. Nulová hypotéza symetrie kolem 0 by pak znamenala, že je stejně tak dětí, které se narodí předčasně, jako těch, které se narodí po termínu. Jednostrannými alternativami bychom mohli testovat, jestli je více dětí předčasně narozených (levostranná) nebo narozených po termínu (pravostranná).

Matematický model (3.b) je poněkud obsáhlejší, protože mu v praxi odpovídají tři situace, kdy pokaždé sledujeme jednu charakteristiku a dílčí situace se liší ve způsobu získávání datových dvojic:

- situace (3.b$_1$) - pozorování na jednom subjektu, ve dvou okamžicích;
- situace (3.b$_2$) - přirozené spárování - dva závislé subjekty;
- situace (3.b$_3$) - experimentální spárování - dva homogenní subjekty.

Situaci (3.b$_1$) je možné interpretovat jako sledování vlivu ošetření proti kontrole, kdy subjekt je sám sobě kontrolou. Provedeme totiž pozorování na subjektu před ošetřením a po ošetření a sledujeme změnu. Řekněme, že by se sledoval vliv nového léku na snížení hladiny cukru v krvi. Zaznamenali bychom na subjektech hodnotu před, hodnotu po a utvořili rozdíly a nulová hypotéza symetrie kolem 0 by říkala, že lék nemá vliv. V tomto případě by byla vhodnou volbou levostranná alternativa, pro kterou by svědčilo více záporných rozdílů, neboli situací, kdy před podáním byla hladina vyšší a po podání nižší.

Rozdíl oproti situaci (1.b) z 1. kapitoly je ten, že zde se zohledňují další faktory, které by mohly působit na změnu stavu, resp. v rámci transformace na model (3.a) se eliminují individuální a časově konstantní vlivy. Např. subjekt A by měl vlivem nemoci trvale zvýšený cukr a vlivem ošetření by se mu snížil na normální hladinu. Jinému subjektu B by se po podání léku hodnota z normální také snížila. Pokud bychom měli k dispozici pouze hodnoty po podání léku, pak by se mohlo zdát, že normální hodnota cukru v krvi značí, že ošetření nemělo vliv. Vzhledem k tomu, že zohledníme, že subjekt A měl předtím vysokou hladinu, potom zaznamenáme efekt ošetření. V situaci (1.b) se těmito faktory nezabýváme a v případě, že bychom brali subjekt A jako kontrolu a subjekt B jako ošetření, pak se efekt ošetření bude zdát mnohem větší, než je ve skutečnosti. Naopak pokud by subjekt B byl kontrolní a subjekt A ošetření, pak by se efekt ošetření zdál zanedbatelný, přičemž faktor nemoci by vůbec nebyl zahrnut. Další možné zkreslení by mohlo nastat tehdy, pokud by subjekt B onemocněl během výzkumu. Před podáním léku by měl normální hladinu cukru a zároveň s ošetřením by se u něj projevila choroba, která by hladinu cukru zvyšovala. Následně by se zdálo, že ošetření nemá vliv, protože by hladina byla stále stejná. Samozřejmě takovouto nemocí člověk neonemocní najednou, tedy lepším příkladem faktoru, který by takto během výzkumu mohl ovlivnit výsledky, je stres.

Situaci (3.b_2) dobře popisuje sledování síly dominantní a nedominantní ruky subjektu. Subjekt by dostal za úkol každou rukou zvlášť mačkat míček, dokud může. Počty stlačení by se zaznamenaly do dvojic (nedominantní, dominantní). Následně by se utvořily rozdíly (dominantní−nedominantní) a nulová hypotéza symetrie kolem 0 by znamenala, že mezi rukama není rozdíl. Nejvhodnější by bylo použít pravostrannou alternativu, pro kterou by svědčilo více kladných rozdílů, a znamenala by, že dominantní ruka je silnější. Přirozeně spárovanými subjekty také rozumíme dvojčata, oči, končetiny a v jistých situacích i partnery nebo sourozence. Výhodou je, že se při odečítání v rámci transformace na model (3.a), odečtou rušivé vlivy, které jsou dvojici společné.

Experimentální spárování, které odpovídá situaci (3.b$_3$) spočívá v utvoření tzv. homogenních dvojic. Řekněme, že chceme opět sledovat vliv ošetření a máme podezření na to, že by efekt mohly mít další faktory. Dvojice párujeme podle toho, které faktory chceme ohlídat. Domníváme-li se, že vliv by mohlo mít pohlaví, věk a fyzická kondice, pak podle těchto měřítek budeme párovat subjekty - např. dvojici budou tvořit muži ve věku 25-30 let, fyzicky zdatní. Jinou dvojici můžou tvořit ženy ve věku 50-55, které se sportu nevěnují. Odečtením v rámci transformace na model (3.a) pak docílíme toho, že tyto faktory se vyruší, a pokud bude prokázán vliv ošetření, pak určitě nebude skutečný efekt tkvít v těchto kontrolovaných faktorech. Tento postup představuje jakýsi ideál, jak zjišťovat vliv ošetření, avšak z časového hlediska je velmi náročný i pracný. Oproti situaci (3.b$_1$) zde nehrají roli časově proměnlivé faktory a oproti situaci (3.b$_2$) se buď subjekty neznají, a nemohou se ovlivňovat, nebo nehraje roli motivace subjektu (jako v případě stlačování míčku) a je tak dodržena větší nezávislost. Motivací rozumíme např. ztrátu zájmu vlivem únavy, popř. nepozornost, nebo naopak veliké snažení za účelem podání dobrého výkonu.

Nyní opět shrneme předpoklady nutné pro formulaci testů, jejichž teoretické zázemí bude čerpáno z [1, 3, 5].

- Model (3.a):
  - (A3.1a): $Z_1, ..., Z_N$ je náhodný výběr o pevném rozsahu N ze spojitého rozdělení;
- Model (3.b):
  - (A3.1b): $(X_1, Y_1), ..., (X_N, Y_N)$ je náhodný výběr o pevném rozsahu N ze spojitého dvourozměrného rozdělení, (X_i, Y_i jsou obecně závislé).

3.1 Znaménkový test

Tomuto nejstaršímu neparametrickému testu se datuje vznik k počátkům 18. století. V situacích, kdy je vhodné znaménkový test použít, bývají často k dispozici silnější testy, avšak hlavní výhodou znaménkového testu je jednoduchost. V principu jde o počítání kladných a záporných znamének a někdy není zapotřebí ani tabelovaných kritických hodnot.

Máme-li model (3.a), pak pracujeme s $Z_1, ..., Z_N$, popř. výběr $Z'_1, ..., Z'_N$ transformujeme. V případě modelu (3.b), pracujeme pouze s utvořenými rozdíly $Z_1, ..., Z_N$.

Testová statistika ZT je definována jako počet hodnot s kladným znaménkem ve výběru $Z_1, ..., Z_N$. Za platnosti nulové hypotézy platí, že $ZT \sim Bi(N, \frac{1}{2})$. Test je tedy současně obdobou testu o parametru binomického rozdělení, že $p = \frac{1}{2}$. Kritické hodnoty pro menší rozsah výběru jsou tabelovány např. v [8] a mají následující vlastnost

$$P(ZT \leq k_1(\alpha)) \leq \frac{\alpha}{2}, \quad P(ZT \geq k_2(\alpha)) \leq \frac{\alpha}{2}, \tag{3.1}$$

kde $k_1(\alpha)$ je největší a $k_2(\alpha)$ nejmenší celé číslo s takovou vlastností. H_0 zamítáme proti oboustranné alternativě, platí-li $ZT \leq k_1(\alpha)$ nebo $ZT \geq k_2(\alpha)$. Dvojice $k_1(\alpha)$ a $k_2(\alpha)$ jsou určeny tak, aby oboustranný test dodržel hladinu α. Kritické hodnoty pro jednostranný test na hladině α tedy bereme z tabulek pro oboustranný test s hladinou 2α.

V případě pravostranného testu svědčí pro alternativu větší počet kladných znamének, a tudíž budeme zamítat nulovou hypotézu, jestliže $ZT \geq k_2(2\alpha)$. Pro levostrannou alternativu pak svědčí více záporných znamének a nulovou hypotézu zamítáme, když $ZT \leq k_1(2\alpha)$.

Dále lze odvodit aproximativní statistiku, která má asymptoticky normální rozdělení. Odvození je založeno na vzorcích pro střední hodnotu a rozptyl veličiny, která má binomické rozdělení. Pokud $ZT \sim Bi(N, \frac{1}{2})$, pak víme, že $\mathrm{E}ZT = \frac{N}{2}$ a var$ZT = \frac{N}{4}$. Tedy pro větší rozsahy platí

$$U_{ZT} = \frac{ZT - \frac{N}{2}}{\sqrt{\frac{N}{4}}} = \frac{\frac{1}{2}(2ZT - N)}{\frac{1}{2}\sqrt{N}} = \frac{2ZT - N}{\sqrt{N}} \sim N(0,1). \tag{3.2}$$

V případě oboustranného testu hodnotu $|U_{ZT}|$ porovnáme s $u(1 - \frac{\alpha}{2})$ a bude-li platit $|U_{ZT}| \geq u(1 - \frac{\alpha}{2})$, pak nulovou hypotézu na hladině α zamítáme. Pro levostrannou alternativu budou svědčit menší hodnoty statistiky U_{ZT}, a proto nulovou hypotézu zamítáme pro $U_{ZT} \leq u(\alpha)$. V případě pravostranné alternativy očekáváme větší hodnoty statistiky U_{ZT}, a tedy zamítáme pro $U_{ZT} \geq u(1 - \alpha)$. V softwaru R je znaménkový test chápán jako test binomického rozdělení, a proto pro něj použijeme funkci `binom.test`, kdy výstupem bude opět p-hodnota.

Vyskytnou-li se v $Z_1, ..., Z_N$ nuly, zpravidla se vynechají a o jejich počet se rozsah výběru sníží, shody nám v tomto případě nevadí. Znaménkový test je vhodný zejména pro rozdělení s těžkými chvosty, např. dvojitě exponenciální rozdělení. Rovněž je jeho výhodou použitelnost pro tři druhy nulové hypotézy. Lze jím zjistit symetrii rozdělení, otestovat medián nebo zjistit, zda je parametr binomického rozdělení roven $\frac{1}{2}$. Vzhledem k menší síle testu je však pro spolehlivější výsledky potřeba většího rozsahu výběru, avšak jsou situace, kdy na data lze použít znaménkový test, ale nelze použít jednovýběrový Wilcoxonův test. Např. pokud bychom při testování symetrie u dětí narozených předčasně nebo po termínu měli k dispozici pouze jejich počty. Tedy věděli bychom, kolik dětí se narodilo předčasně (těm bychom přiřadili znaménko -) a kolik se narodilo po termínu (znaménko +).

3.2 Jednovýběrový Wilcoxonův test

Test použijeme na $Z_1, ..., Z_N$. Nejprve výběr seřadíme podle velikosti absolutních hodnot jednotlivých prvků a přiřadíme jim pořadí, které budeme značit $R_1^+, ..., R_N^+$, tedy R_i^+ bude pořadí Z_i ve výběru seřazeném podle absolutních hodnot Z_i od nejmenšího po největší. Dále zavedeme

$$V^+ = \sum_{Z_i > 0} R_i^+, \quad V^- = \sum_{Z_i < 0} R_i^+, \tag{3.3}$$

neboli veličina V^+ nám dává součet pořadí původně kladných prvků a veličina V^- součet pořadí původně záporných prvků. Hodnotu $min(V^+, V^-)$ porovnáme s tabelovanou kritickou hodnotou $V_c(\alpha)$ a platí-li $min(V^+, V^-) \leq V_c(\alpha)$, pak nulovou hypotézu na hladině α zamítáme proti oboustranné alternativě.

Pro levostrannou alternativu budou svědčit malé hodnoty statistiky V^+, a proto budeme zamítat nulovou hypotézu na hladině α pro $V^+ \leq V_c(2\alpha)$. Naopak pro pravostrannou alternativu svědčí menší hodnoty V^- a zamítáme H_0 pro $V^- \leq V_c(2\alpha)$. Pro jednostranné testy na hladině α opět použijeme tabulky pro oboustranný test s hladinou 2α, dohledatelné v [8].

Pro větší výběry se využívá asymptotického rozdělení veličiny V^+ a platí

$$\mathrm{E}V^+ = \frac{1}{4}N(N+1), \quad \mathrm{var}V^+ = \frac{1}{24}N(N+1)(2N+1). \tag{3.4}$$

Testová statistika

$$U_V = \frac{V^+ - \mathrm{E}V^+}{\sqrt{\mathrm{var}V^+}} \tag{3.5}$$

má pak asymptoticky rozdělení $N(0,1)$. Nulovou hypotézu zamítáme proti oboustranné alternativě, je-li $|U_V| \geq u(1-\frac{\alpha}{2})$. Pro levostrannou (resp. pravostrannou) alternativu svědčí menší (resp. větší) hodnoty statistiky U_V, a proto zamítáme nulovou hypotézu na hladině α, jestliže $U_V \leq u(\alpha)$ (resp. $U_V \geq u(1-\alpha)$). V softwaru použijeme už známou funkci `wilcox.test` nebo `wilcox.exact`, jsou-li přítomné shody. Výstupem je opět p-hodnota. Při ručním počítání se, stejně jako u znaménkového testu, nuly zpravidla vynechávají a o jejich počet se sníží rozsah výběru. Případným shodám se přiřadí průměrná pořadí, avšak potom jsou již tabelované hodnoty i normální aproximace nepřesné.

Funkce `wilcox.test` a `wilcox.exact` rozlišují jednovýběrový a dvouvýběrový test podle počtu zadaných vektorů. Jestliže chceme testovat dvouvýběrovým testem, zadáme dva číselné vektory, z nichž každý obsahuje data pro každý výběr zvlášť. Chceme-li testovat jednovýběrovým, pak je třeba zadat pouze jeden vektor (odpovídající výběru $Z_1, ..., Z_N$), nebo v případě modelu (3.b) můžeme zadat dva vektory, ale nastavit argument `PAIRED=TRUE`, kterým řekneme, že se jedná o spárované subjekty. Pokud máme jeden výběr s nenulovým známým mediánem $\widetilde{x}$, nastavíme ještě parametr `mu=`$\widetilde{x}$.

Z formulace testů vyplývá, že pokud výběry pochází z normálního rozdělení, je nejlepší volbou jednovýběrový t-test, popř. párový t-test pro závislé výběry. Pokud ale normalita dat není splněna, je nejvhodnější použít jednovýběrový Wilcoxonův test, protože má poměrně velkou sílu. Znaménkový test je výborný pro svoji jednoduchost, ale ve většině případů má menší sílu. Pokud by ale data pocházela z rozdělení, které má těžké chvosty, nebo by byla k dispozici pouze znaménka, je volba znaménkového testu optimální a pro menší výběry je možné vypočítat kritické hodnoty snadno i ručně.

3.3 Aplikace testů

Testy budou opět prováděny na hladině $\alpha = 0,05$, nebude-li řečeno jinak.

Příklad 3.1. Na stabilizovaném zdroji elektrického napětí byla nastavena stálá hodnota napětí, kterou jsme měřícím přístrojem desetkrát změřili. Z hodnot byly určeny chyby měření. Má přístroj systematickou odchylku?

chyby měření:	-0,02149,	-0,01189,	0,01021,	0,01571,	0,01483
	0,01491,	-0,02079,	0,01751,	-0,02459,	0,01641

Řešení. Data nemusíme jakkoli upravovat, odpovídají modelu (3.a). Chyby měření můžeme považovat za spojitou veličinu. Nezávislost mezi měřeními by potenciálně mohla být porušena, kdyby v měřícím přístroji nebo zdroji napětí se vyskytla chyba, např. špatnou součástkou, která by se během měření opotřebovávala a hodnoty by pak byly na sobě závislé. Jinou situací by mohla být porucha, která způsobí úplnou změnu naměřených hodnot, např. že přístroj bude méně citlivý, čímž se změní rozdělení, ze kterého pochází chyby.

Chyby měření označíme $Z_1, ..., Z_{10}$, tedy $N = 10$ a určíme tvar hypotéz:

$$H_0 : \text{přístroj nemá systematickou odchylku,}$$
$$H_1 : \text{přístroj má systematickou odchylku.}$$

Máme zde oboustrannou alternativu, že rozdělení, ze kterého pochází chyby, není symetrické kolem nuly a je vychýlené buď k záporným, nebo ke kladným hodnotám. Uvažujeme vychýlení ve smyslu, že je symetrické kolem jisté systematické chyby nebo je rozdělení úplně nesymetrické. Použijeme nejprve znaménkový test. Z dat vidíme, že počet kladných hodnot je $ZT = 6$. Kritické hodnoty pro rozsah 10 a hladinu testu 0,05 jsou $k_1 = 1$, $k_2 = 9$, a protože $k_1 < ZT < k_2$, nulovou hypotézu nelze zamítnout. Podle vztahu (3.2) dále spočítáme hodnotu testové statistiky, která je založena na normální aproximaci. Vychází

$$U_{ZT} = \frac{2 \cdot 6 - 10}{\sqrt{10}} = 0,6325,$$

a protože $|U_{ZT}| < u(0,975)$, nulovou hypotézu nelze zamítnout ani tímto způsobem. Použitím softwaru a funkce `binom.test` dostáváme p-hodnotu 0,7539, která velmi přesvědčivě nezamítá nulovou hypotézu. Znaménkovým testem jsme došli k závěru, že přístroj měří přesně. Jinými slovy, pravděpodobnost záporné chyby je stejná, jako pravděpodobnost kladné chyby.

Pro použití jednovýběrového Wilcoxonova testu si seřadíme hodnoty podle jejich absolutní hodnoty a přiřadíme jim pořadí. Tučně jsou vyznačeny původně záporné hodnoty a jim příslušná pořadí.

$\|Z_i\|$	0,01021	**0,01189**	0,01483	0,01491	0,01571
pořadí	1	**2**	3	4	5
$\|Z_i\|$	0,01641	0,01751	**0,02079**	**0,02149**	**0,02459**
pořadí	6	7	**8**	**9**	**10**

Podle vztahu (3.3) určíme statistiky V^+,V^-:

$$V^+ = 1 + 3 + 4 + 5 + 6 + 7 = 26, \quad V^- = 2 + 8 + 9 + 10 = 29.$$

Vidíme, že $min(V^+, V^-) = 26$. Hodnota tabelované kritické hodnoty $V_c(0,05) = 8$, a protože $min(V^+, V^-) > V_c$, nulovou hypotézu na hladině 0,05 proti oboustranné alternativě nelze zamítnout. Použitím vztahů (3.4) a (3.5) dostaneme hodnotu statistiky, která je založena na normální aproximaci. Vyjde

$$\mathrm{E}V^+ = 27,5, \quad \mathrm{var}V^+ = 96,25, \quad U_V = \frac{26 - 27,5}{\sqrt{96,25}} = -0,1529.$$

Hodnotu $|U_V|$ porovnáme s $u(0,975) = 1,96$, a protože $|U_V| < u(0,975)$, nulovou hypotézu nelze zamítnout ani tímto způsobem. Použitím softwaru dostaneme p-hodnotu 0,9219, která také velmi přesvědčivě nezamítá nulovou hypotézu.

Oběma testy se nulová hypotéza nedá zamítnout, a tedy na hladině 0,05 se můžeme domnívat, že přístroj měří přesně a nemá systematickou odchylku.

Příklad 3.2. Šest studentů přešlo na dietu, aby ztratili na váze. Byli zváženi před zahájením a po ukončení diety. Byla dieta efektivní? Převzato z [3, str. 164].

váha před (kg):	78,9,	86,6,	85,3,	82,6,	91,2,	85,3
váha po (kg):	74,8,	84,4,	83,0,	80,7,	92,1,	82,1

Řešení. Jedná se o příklad situace (3.b$_2$). Nezávislost mezi studenty by mohla být ošetřena tak, že by o sobě studenti nevěděli, a tedy by je nemohla ovlivnit motivace, kdyby se navzájem viděli, zda hubnou nebo ne. Naopak by mohla být porušena tehdy, pokud by někteří studenti byli ovlivnění reklamou nebo seminářem o zdravé výživě a jiní ne. Pokud by si studenti byli velmi podobní ve smyslu fyzické kondice nebo přístupu ke stravě, jinými slovy pokud by byli homogenní, pak musíme velmi opatrně přistupovat ke generalizaci na celou populaci, protože případné prokázání vlivu u této skupiny nemusí znamenat, že by byl vliv prokázán u jiné. Větší rozmanitost výběru tedy podporuje spolehlivější generalizaci na populaci. Váhu dále můžeme považovat za spojitou veličinu.

Označme váhu před X_i a váhu po Y_i a utvořme rozdíly $Z_i = Y_i - X_i$:

$$Z_i : \quad \text{-4,1}, \quad \text{-2,2}, \quad \text{-2,3}, \quad \text{-1,9}, \quad 0{,}9, \quad \text{-3,2}.$$

Naformulujeme nulovou a alternativní hypotézu:

$$H_0 : \text{dieta není účinná,}$$
$$H_1 : \text{dieta snižuje váhu.}$$

V tomto případě máme levostranný test, protože pro alternativu bude svědčit větší počet záporných hodnot. Testová statistika znaménkového testu je $ZT = 1$ a kritickou hodnotu bereme $k_1 = 0$, a protože $ZT > k_1$, nulovou hypotézu nelze zamítnout. Při použití aproximace dostaneme dosazením do (3.2) hodnotu statistiky $U_{ZT} = -1,633$. Protože $U_{ZT} > u(0,05) = -1,6449$, nulovou hypotézu nelze zamítnout ani pomocí normální aproximace. Software pak vypočítal p-hodnotu, která je 0,1094, a tedy znaménkovým testem nelze nulovou hypotézu zamítnout. Od pohledu bychom však očekávali jiný závěr, ale jak bylo řečeno, znaménkový test nemá takovou sílu a nezamítnutí může být způsobeno menším rozsahem výběru.

Pro použití jednovýběrového Wilcoxonova testu si hodnoty opět uspořádáme podle absolutní hodnoty a přiřadíme pořadí. Tučně vyznačíme původně záporné hodnoty a příslušná pořadí.

$\lvert Z_i \rvert$:	0,9	**1,9**	**2,2**	**2,3**	**3,2**	**4,1**
pořadí:	1	**2**	**3**	**4**	**5**	**6**

Ze vztahu (3.3) určíme $V^+ = 1$, $V^- = 20$. Pro levostranný test svědčí malé hodnoty V^+, a protože $V^+ < V_c(0,1) = 2$, nulovou hypotézu jednostranného testu na hladině 0,05 zamítáme. Užitím vztahů (3.4), (3.5) pro normální aproximaci dostaneme hodnotu $U_V = -1,9917$, která je menší než $u(0,05)$, a proto zamítáme nulovou hypotézu i tímto způsobem. Nakonec použitím statistického softwaru dostaneme p-hodnotu 0,0313 a jednovýběrovým Wilcoxonovým testem zamítáme nulovou hypotézu. Důvod, proč dal tento test jiný výsledek než znaménkový, je ten, že má větší sílu. Na hladině 0,05 se tedy můžeme domnívat, že dieta je efektivní ve smyslu, že pomáhá snížit váhu.

Příklad 3.3. Skrze vlastní dotazník a pomocí sociálních sítí byly sledovány názory rodičů na řízení auta svého potomka. Dotazník obsahoval dvě otázky, kdy první se ptala na názor otce a druhá na názor matky, zda souhlasí s tím, aby jejich potomek řídil auto. Možné odpovědi byly čtyři - od 1=souhlasím po 4=nesouhlasím. Zkoumáme, zda se názor rodičů liší, či nikoliv.

Řešení. Data byla sesbírána od 55 dvojic, kdy 40 z nich mělo stejný názor a 15 rozdílný. Protože situace odpovídá modelu (3.b), budeme zde pracovat s rozdíly

$Z_i = Y_i - X_i$, kde (X_i, Y_i) reprezentují po řadě názor otce a matky. Dostáváme tak 40 nul, které představují shodu v názorech a dalších 15 rozdílů

1, 1, -3, -1, 2, 3, 1, 1, 1, 1, 1, 1, -1, 1, 1.

Abychom mohli použít znaménkový test, musíme zde vynechat nuly. Výběr se omezí pouze na 15 nenulových rozdílů, čímž se podstatně zhorší přesnost výsledku. Použití jednovýběrového Wilcoxonova testu je zde zcela nevhodné kvůli velkému počtu nul a navíc bychom při přiřazovaní pořadí měli problém se shodami. Příklad budeme řešit pouze pomocí znaménkového testu. Formulace hypotéz je následující

$$H_0 : \text{názory rodičů jsou vyrovnané,}$$
$$H_1 : \text{názor matek nebo otců je více vyhraněný.}$$

Upřesníme nulovou hypotézu, vyrovnaností rozumíme to, že ani matky ani otcové nejsou jednostranně vyhranění ve smyslu, že by matky měly tendenci spíše nesouhlasit nebo otcové spíše souhlasit. Alternativa pak znamená, že buď matky, nebo otcové jsou více přiklonění k souhlasu nebo k nesouhlasu.

Budeme-li pracovat pouze s nenulovými rozdíly, máme rozsah výběru $N = 15$ a statistika $ZT = 12$. Protože máme oboustranný test, porovnáme ZT s $k_1(0,05) = 3$, $k_2(0,05) = 12$. Protože $ZT \geq k_2(0,05)$, zamítneme nulovou hypotézu. Zamítnutí na hladině 0,05 potvrzuje i použití aproximace, kdy $U_{ZT} = 2,3238 > u(0,975) = 1,96$ a p-hodnota ze softwaru 0,0352.

Tady můžeme vidět, jaký problém může způsobit vynechání nul. Vzhledem k výsledkům testu bychom totiž nepředpokládali zamítnutí nulové hypotézy. Interpretace oboustranné alternativy je taková, že rodiče se v názorech rozcházeli, avšak nevíme, který z nich spíše souhlasil a který naopak nesouhlasil. Pokud bychom ale testovali proti pravostranné alternativě, při které bychom rovněž zamítli H_0, mohli bychom výsledek interpretovat tak, že pokud se rodiče neshodují, pak blíže k nesouhlasu má matka.

Porovnáme-li použití testů pro konkrétní data, pak v případě, že bychom chtěli postupovat co nejjednodušeji, zvolíme znaménkový test. Jak je ale vidět u příkladu 3.2, jednovýběrový Wilcoxonův test zamítá nulovou hypotézu tam, kde znaménkový ne, což konkrétně mohlo být způsobeno menším rozsahem výběru. Obecně má totiž znaménkový test menší sílu a je silnější až pro velké rozsahy výběru. Proto pro data z příkladu 3.2 je vhodnější volbou jednovýběrový Wilcoxonův test. Dále je otázkou, zda chyby měření z příkladu 3.1 jsou normálně rozděleny a je-li pak vhodnější použít t-test. Samozřejmě i na normálně rozdělená data můžeme použít neparametrické testy, avšak mohlo by se stát, že nezamítnou nulovou hypotézu tam, kde by ji t-test zamítnul. Pro data z příkladu 3.3 je použití jednovýběrového Wilcoxonova testu značně nevhodné vzhledem k množství shod.

4 Hypotéza nezávislosti

Máme náhodný výběr $(X_1, Y_1), ...(X_N, Y_N)$ z dvourozměrného rozdělení F_2 o pevném rozsahu N. Dále přiřaďme pořadí $R_1, ..., R_N$ výběru $X_1, ..., X_N$ a pořadí $S_1, ..., S_N$ výběru $Y_1, ..., Y_N$. Testujeme hypotézu nezávislosti proti alternativní hypotéze, že veličiny nezávislé nejsou. Matematicky lze pomocí distribuční funkce zapsat

$H_0 : F_2(x, y) = F_X(x)F_Y(y),$

$H_1 : F_2(x, y) \neq F_X(x)F_Y(y),$

kde H_1 je tvar oboustranné alternativy, $F_X(x)$ je distribuční funkce výběru $X_1, ..., X_N$ a $F_Y(y)$ je distribuční funkce pro $Y_1, ..., Y_N$. Levostranná alternativa by testovala, zda se jedná o závislost negativní, tzn. čím je větší X_i, tím je menší Y_i nebo čím je menší X_i, tím je větší Y_i. Pravostranná alternativa by testovala závislost pozitivní, neboli očekáváme se zvětšujícími (resp. zmenšujícími) X_i větší (resp. menší) Y_i, podle [3].

Jak bylo popsáno v [6], tento dvourozměrný náhodný výběr si lze představit jako měření dvou charakteristik na témže subjektu. Nejedná se zde o přiřazování ošetření, nýbrž o zkoumání charakteristik, které nelze nijak přiřadit a jsou se subjektem neoddělitelně spjaté.

Příkladem by mohlo být zkoumání výšky a váhy, kdy bychom subjekty změřili a zvážili a testovali bychom nulovou hypotézu, že výška a váha spolu nesouvisí, proti pravostranné alternativě, že čím je větší výška, tím je větší váha. Obecně by se dal test použít na jakékoliv lékařské charakteristiky, u kterých chceme zkoumat, zda spolu souvisí. Např. závislost hladiny iontů v krvi s jinými krevními elementy apod. Příkladem pro testování negativní závislosti by mohlo být sledování názorů rozvedených párů, kdy bychom měli hypotézu, že jejich názory jsou nezávislé, proti alternativě, že se názorově velmi rozcházejí. Subjektem bychom zde rozuměli jeden pár. Výzkum by byl uskutečnitelný skrze dotazník, který by sledoval názor na současné dění, vztahové záležitosti apod. Vyhodnocením by byly body od 0 do 100. Očekávali bychom, že pokud má jeden z bývalých partnerů bodů hodně, druhý bude mít naopak velmi málo. Pokud by se nulová hypotéza o nezávislosti zamítla, mohli bychom tvrdit, že v názorech rozvedených párů existuje negativní závislost.

Formulace testu této kapitoly pochází z [2, 5] a potřebné předpoklady jsou shrnuty v následujících bodech:

- (A4.1): $(X_1, Y_1), ...(X_N, Y_N)$ je náhodný výběr ze spojitého dvourozměrného rozdělení;
- (A4.2): N je pevný rozsahu výběru.

4.1 Spearmanův korelační koeficient

Intuitivně bychom pro testování nezávislosti použili výběrový korelační koeficient (Pearsonův), který vyjadřuje míru lineární závislosti. Pro data z normálního rozdělení platí, že nekorelovanost je ekvivalentní s nezávislostí. Spearmanův korelační koeficient má výhodu v tom, že pracuje pouze s pořadími a nevyžaduje původní hodnoty pozorování, které někdy nemusí být známy. Má vzorec:

$$r_S = \frac{\sum R_i S_i - N\left(\frac{N+1}{2}\right)^2}{\sqrt{(\sum R_i^2 - N\left(\frac{N+1}{2}\right)^2)(\sum S_i^2 - N\left(\frac{N+1}{2}\right)^2)}} \tag{4.1}$$

a je definován jako výběrový korelační koeficient dvojic $(R_1, S_1), ..., (R_N, S_N)$. Pro praktické počítání lze odvodit nejčastěji používaný vzorec

$$r_S = 1 - \frac{6}{N(N^2-1)} \sum_{i=1}^{N} (R_i - S_i)^2. \tag{4.2}$$

V případě oboustranného testu se hodnota $|r_S|$ porovná s tabelovanými kritickými hodnotami $r_S(\alpha)$, dohledatelné v [8]. Jestliže je $|r_S| \geq r_S(\alpha)$, potom nulovou hypotézu na hladině α zamítáme. V případě jednostranných testů na hladině α musíme opět použít tabulky pro oboustranné testy na hladině 2α a zamítáme proti pravostranné (resp. levostranné) alternativě, jestliže $r_S \geq r_S(2\alpha)$ (resp. $r_S \leq -r_S(2\alpha)$).

Pro $N > 30$ je možné použít normální aproximaci a platí, že statistika

$$U_{RS} = r_S\sqrt{N-1} \tag{4.3}$$

má asymptoticky rozdělení $N(0,1)$. Nulovou hypotézu proti oboustranné alternativě budeme zamítat, jestliže $|U_{RS}| \geq u(1-\frac{\alpha}{2})$. Pokud testujeme levostrannou alternativu, pak zamítáme pro $U_{RS} \leq u(\alpha)$, a v případě pravostranné alternativy zamítáme pro $U_{RS} \geq u(1-\alpha)$. V softwaru použijeme funkci `cor.test`, kde si nastavíme Spearmanův korelační koeficient a požadovanou alternativu.

Tentokrát nám budou shody vadit pouze v rámci jednotlivých výběrů, nikoli shody mezi výběry. Pokud se shody vyskytují, přiřadí se průměrná pořadí a doporučuje se použít korigovanou statistiku, kterou je možné dohledat v [2, str. 257]. Mohli bychom samozřejmě použít i tabelovaných kritických hodnot i aproximace pro výběry beze shod, ale potom počítáme s větší nepřesností. V softwaru je třeba jako jeden z parametrů funkce `cor.test` nastavit `exact=FALSE` a software pak použije pro výpočet aproximaci na t-rozdělení.

4.2 Aplikace testu

Příklad 4.1. Pět studentů doktorského studia dělalo test o současném dění. Znamená vyšší věk studenta lepší výsledek testu? Příklad převzat z [3, str. 330].

věk:	24,	31,	38,	45,	45
skóre:	68,	85,	84,	92,	90

Řešení. Předpoklad (A4.1) o spojitosti rozdělení je zde porušen, což by nevadilo, pokud by se ve výběru nevyskytovala shoda. Musíme tedy počítat s větší nepřesností výsledků. Charakteristika výšky je neoddělitelně spjata se subjektem a skóre znalostního testu koresponduje s jakýmsi obecným přehledem jedince, což rovněž nelze považovat za přiřazenou charakteristiku.

Označme si $X_1, ..., X_5$ věk a $Y_1, ..., Y_5$ dosažené skóre, dále rozsah výběru $N = 5$ a testujeme na hladině $\alpha = 0,05$. Očekáváme, že s rostoucím věkem se bude dosažené skóre zvětšovat. Naformulujeme nulovou a alternativní hypotézu

$$H_0: \text{věk s dosaženým skórem nesouvisí},$$
$$H_1: \text{hodnoty spolu souvisí ve smyslu pozitivní závislosti.}$$

Nyní přiřadíme pozorováním pořadí.

X_i	24	31	38	45	45
R_i	1	2	3	4,5	4,5
Y_i	68	85	84	92	90
S_i	1	3	2	5	4

Upozorňujeme, že zde se přiřazuje pořadí zvlášť pozorováním X_i a zvlášť Y_i, tedy není zde použit sdružený výběr. I když se vyskytuje shoda, dovolíme si i přes nepřesnost použít vzorec (4.2):

$$r_S = 1 - \frac{6}{5(5^2-1)} \sum_{i=1}^{5} (R_i - S_i)^2 = 0,825.$$

Kritická hodnota pro je $r_S(0,1) = 0,9$, a protože platí $r_S < r_S(0,1)$, nezamítáme nulovou hypotézu na hladině 0,05. Nutno připomenout, že hodnota $r_S(0,1)$ je kritická hodnota pro oboustranný test taková, aby byla dodržena hladina 0,1. Pro jednostranný test tato kritická hodnota dodržuje hladinu 0,05.

Přestože máme malý výběr, zkusíme spočítat hodnotu aproximativní testové statistiky podle vztahu (4.3):

$$U_{RS} = 0,825 \cdot \sqrt{5-1} = 1,65.$$

Hodnotu porovnáme s $u(0,95) = 1,6449$ a vidíme, že velmi těsně $U_{RS} > u(0,95)$, a tímto způsobem zamítáme nulovou hypotézu. Nakonec použijeme statistický software a jeho funkci `cor.test`, která nám vypočte p-hodnotu 0,0269. P-hodnota je menší než hladina testu, takže zamítáme i použitím softwaru. Nezamítnutí H_0 přesnou statistikou může být způsobeno shodou ve výběru. Pomocí Spearmanova korelačního koeficientu jsme dvěma způsoby zamítli nulovou hypotézu o nezávislosti. Můžeme se tedy domnívat, že s vyšším věkem studenta se výsledek testu o současném dění zlepšuje.

Příklad 4.2. Z databáze Českého statistického úřadu [10, 11] byla převzata data o průměrné míře nezaměstnanosti (v %) a počtu sebevražd v letech 2010 až 2012 podle krajů. Souvisí míra nezaměstnanosti (MN) s počtem sebevražd (S)?

	2010		**2011**		**2012**	
	MN	**S**	**MN**	**S**	**MN**	**S**
Praha	3,50	169	3,63	174	3,85	196
Středočeský kraj	5,63	185	5,62	200	5,69	208
Jihočeský kraj	5,62	82	5,60	98	5,70	108
Plzeňský kraj	6,28	88	5,75	61	5,44	80
Karlovarský kraj	8,48	50	8,22	44	8,00	53
Ústecký kraj	9,83	125	9,66	132	10,00	146
Liberecký kraj	8,02	64	7,47	61	7,35	56
Královéhradecký kraj	5,74	75	5,48	103	5,70	105
Pardubický kraj	6,90	68	6,33	81	6,22	77
Kraj Vysočina	7,22	56	6,90	81	6,72	56
Jihomoravský kraj	7,83	144	7,47	157	7,48	168
Olomoucký kraj	8,40	117	8,04	104	8,08	107
Zlínský kraj	7,65	100	6,97	91	6,95	96
Moravskoslezský kraj	8,73	179	8,33	201	8,47	191

Řešení. Míru nezaměstnanosti můžeme považovat za spojitou veličinu, avšak v důsledku zaokrouhlování se vyskytují shody. Shody se vyskytují i mezi počty sebevražd, které za spojitou veličinu nepokládáme. Je třeba tedy počítat s větší nepřesností.

Nezávislost mezi jednotlivými kraji by byla porušena tehdy, pokud by vybrané kraje byly např. v některém projektu na snížení nezaměstnanosti nebo byly samy schopny realizovat projekty inspirované jinými kraji s menší nezaměstnaností. Nezávislost napříč jednotlivými roky by byla porušena, pokud by vedení kraje sledovalo vývoj v předešlých letech a reagovalo na něj patřičnými opatřeními, např. při poklesu nezaměstnanosti by se zaměřila na jiné problémy apod. Tuto závislost však z dat za tři roky nelze potvrdit.

Označme $X_1, ..., X_{42}$ míru nezaměstnanosti (MN) a $Y_1, ..., Y_{42}$ počty sebevražd (S), víme tedy $N = 42$ a testujeme na hladině $\alpha = 0,05$. Data budou uspořádána chronologicky, tedy pozorování $X_1, ..., X_{14}$ a $Y_1, ..., Y_{14}$ odpovídají roku 2010 apod. Rádi bychom zjistili, zda se zvyšuje počet sebevražd se zvyšující mírou nezaměstnanosti.

Formulujeme nulovou a alternativní hypotézu

H_0 : míra nezaměstnanosti s počtem sebevražd nesouvisí,

H_1 : hodnoty spolu souvisí ve smyslu pozitivní závislosti.

Seřadíme míry nezaměstnanosti a počty sebevražd zvlášť podle velikosti, přiřadíme jim pořadí R_i a S_i a provedeme výpočet statistiky r_S podle (4.2), kvůli obtížnosti provedeme pomocí softwaru,

$$r_S = 1 - \frac{6}{42(42^2-1)} \sum_{i=1}^{42} (R_i - S_i)^2 = -0,0638.$$

Kritické hodnoty pro $N = 42$ už nejsou k dispozici a navíc se v datech vyskytují shody, proto použijeme normální aproximaci (4.3)

$$U_{RS} = -0,0638 \cdot \sqrt{42-1} = -0,4083.$$

Hodnotu U_{RS} porovnáme s $u(0,95) = 1,6449$, a protože $U_{RS} < u(0,95)$, pak docela přesvědčivě nelze zamítnout nulovou hypotézu o nezávislosti. P-hodnota při použití softwaru vyšla 0,6863, tedy nezamítnutí nulové hypotézy souhlasí. Na hladině testu $\alpha = 0,05$ se můžeme domnívat, že počet sebevražd nesouvisí s mírou nezaměstnanosti v jednotlivých krajích.

Závěr

V práci byla popsána teorie vybraných neparametrických testů. Testování i interpretace byly ukázány na praktických příkladech. Kapitoly byly rozčleněny podle testovaných nulových a alternativních hypotéz. Každé hypotéze odpovídají různé situace v praxi, které byly nejdříve formulovány matematicky a pak vysvětleny na konkrétním příkladu. V rámci testování dat, které byly buď převzaté nebo vlastně sesbírané, byly použity tři postupy - ruční počítání přesné testové statistiky a aproximativní statistiky a použití funkce v softwaru R.

Přesná testová statistika vyžaduje tabelovaných kritických hodnot, které nemusíme mít vždy k dispozici, ale pro přesnost výsledku je nejlepší, za předpokladu výběru beze shod, popř. bez nul. Aproximativní statistika je na použití jednodušší, protože pro vyhodnocení se používají kvantily normovaného normálního rozdělení, které jsou už jednoduseji dohledatelné. Přesto ale je potřeba zjistit hodnotu přesné testové statistiky, která pro větší rozsahy může být obtížnější na výpočet. Software R nabízí spoustu funkcí, které automaticky provedou test a výstupem je p-hodnota, ze které velmi snadno poznáme, zda se nulová hypotéza může zamítnout. Další výhodou softwaru je, že po nastavení příslušné alternativy (oboustranné, jednostranné), si nemusíme dělat starost s kritickým oborem, protože software vše vypočítá sám podle zadaných argumentů. Na druhou stranu je předpokladem použití softwaru znalost funkcí a jejich správné použití, aby nedošlo ke špatné interpretaci výsledků. Navíc každá funkce používá jiný algoritmus výpočtu, někde je proveden výpočet pomocí aproximace, někde pomocí přesné testové statistiky.

Na závěr se domnívám, že přínosem práce byl její praktický náhled na problematiku neparametrických testů. Podrobná analýza praktických situací tak může pomoct čtenáři v lepší představě o problematice, než by bylo poskytnuto skrze matematické modely.

Literatura

[1] ANDĚL, J. *Statistické metody.* 4. vyd. Praha: MATFYZPRESS, 2007. ISBN 80-7378-003-8.

[2] ANDĚL, J. *Základy matematické statistiky.* 3. vyd. Praha: MATFYZPRESS, 2011. ISBN 978-80-7378-162-0.

[3] CONOVER, W.J. *Practical nonparametric statistics.* 3. vyd. New York: John Wiley and Sons, 1999. ISBN 978-0-471-16068-7.

[4] GLOVER, T., MITCHELL, K., *An Introduction to Biostatistics* [online]. USA: Waveland Press, 2008, s. 458-459 [cit. 2014-04-25]. ISBN 978-1-57766-580-9. Dostupné z: http://www.uwlax.edu/faculty/baggett/Math_145/HANDOUTS/wilcoxon_rank_sum_table.pdf

[5] LEHMANN, E.L. *Nonparametrics: Statistical Methods Based on Ranks.* 1. vyd. New York: Springer, 2006. ISBN 0-387-35212-0.

[6] MÜLLER, I. Neparametrické metody. (přednáška) Olomouc: Univerzita Palackého v Olomouci, únor-květen 2013.

[7] NOGUCHI, K., GEL, Y., BRUNNER, E., KONIETSCHKE, F. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. In: *Journal of Statistical Software* [online]. 2012, [cit. 2014-04-25]. Dostupné z: http://cran.r-project.org/web/packages/nparLD/nparLD.pdf.

[8] SANI, F., TODMAN, J. *Experimental Design and Statistics for Psychology: A First Course* [online]. UK: Blackwell Publishing, 2006, s. 183-196 [cit. 2014-04-25]. ISBN 978-1-4051-0023-6. Dostupné z: http://onlinelibrary.wiley.com/doi/10.1002/9780470776124.app1/pdf

[9] VAN DER WAERDEN, B.L., NIEVERGELT, R. *Tafeln zum Vergleich zweier Stichproben mittels X-test und Zeichentest*. Berlin: Springer, 1956. ISBN 978-3-540-02102-5.

[10] ČESKÝ STATISTICKÝ ÚŘAD. Tab. Průměrný podíl nezaměstnaných osob v ČR a krajích. *Český statistický úřad*. [online]. [cit. 2014-04-25]. Dostupné z: http://vdb.czso.cz/vdbvo/tabparam.jsp?voa=tabulka&cislotab=PRA6010PC_KR&vo=null

[11] ČESKÝ STATISTICKÝ ÚŘAD. Tab. Zemřelí podle příčin smrti v krajích. *Český statistický úřad*. [online]. [cit. 2014-04-25]. Dostupné z: http://vdb.czso.cz/vdbvo/tabparam.jsp?voa=tabulka&cislotab=DEM0120PU_KR&vo=null