Přednáška 7 – Testy rozdělení, testy hypotéz pro dva výběry

Testy rozdělení

V předchozí přednášce jsme mluvili o tom, že pro volbu testu hypotéz potřebujeme vědět, zda data splňují nebo nesplňují předpoklad normality - podle toho volíme mezi parametrickými testy s předpokladem normality dat a neparametrickými testy, které jsou určeny pro data bez normality. Abychom určili, zda data mají normální rozdělení nebo ne a rozhodli se mezi těmito druhy testů, musíme použít jeden z testů rozdělení.

Některé testy rozdělení jsou určeny pouze pro testování normality, jiné mohou být použity i pro testování jiných rozdělení, viz tabulka Jak zvolit test hypotéz na webu.

Pro všechny tyto testy jsou nulová a alternativní hypotézy stejné:

$$H_0: \quad \text{data pochází z testovaného rozdělení,}$$
$$H_A: \quad \text{data nepochází z testovaného rozdělení.}$$

To znamená, že pokud používáme test normality nebo jiný test na ověření normality, nulová a alternativní hypotézy zní:

$$H_0: \quad \text{data mají normální rozdělení,}$$
$$H_A: \quad \text{data nemají normální rozdělení.}$$

Rozebereme některé z testů rozdělení podrobněji.

w/s test normality

Triviální w/s test normality používá statistiku

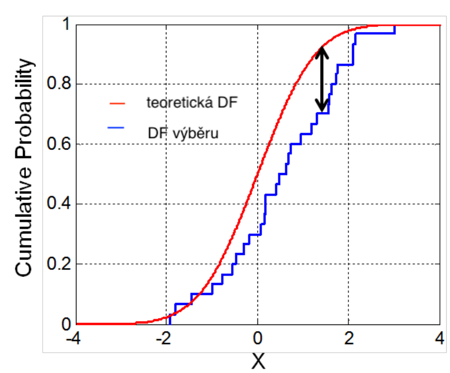
$$\frac{w}{s} = \frac{\max(X) - \min(X)}{s},$$

která je jednoduchým poměrem rozpětí výběru a výběrové směrodatné odchylky $s = \sqrt{s^2}$ a X je výběr.

Kolmogorov-Smirnovův test

Kolmogorov-Smirnovův test se používá pro ověření, zda hodnoty výběru pochází z určitého teoretického rozdělení (nemusí být pouze normální). Test porovnává vzdálenosti mezi distribuční funkcí na hodnotách výběru (z četností) a distribuční funkcí teoretického rozdělení. Statistika je maximální vzdálenost:

$$T = \max(F_{\text{výběr}}(X_i) - F_{\text{teor.}}(X_i)).$$



Zdroj: Wikipedia

Test se hodí i pro ověření, zda data ze dvou výběrů pochází ze stejného rozdělení. Kolmogorov-Smirnovův test se často používá pro testování normality, i když podle některých studií je méně spolehlivý, než Shapiro-Wilkův test nebo Anderson-Darlingův test. Tento test je pouze pravostranný.

Shapiro-Wilkův test normality

Shapiro-Wilkův test se používá pro ověření předpokladu normality výběru. Používá statistiku

$$W = \frac{b^2}{(n-1)s^2}, \quad \text{kde} \quad s^2 \quad \text{je výběrový rozptyl,}$$

$$b = a_1(r_n - r_1) + a_2(r_{n-1} - r_2) + ...,$$

kde se hodnoty a_i berou ze speciální tabulky a r_i jsou pořadí hodnot.

Shapiro-Wilkův test je jeden z nejčastěji používaných testů normality dat v různých softwarech.

Poznámka: Je ale nutné si poznamenat, že je těžké vybrat jeden univerzálně spolehlivý test normality. V praxi nabízí statistický software obvykle několik druhů takových testů, kde uživatel by měl určit, pro který test jsou data nejvíce vhodná. Například, některé testy normality fungují dobře pro malé počty dat, jiné pro velké hodnoty výběru, další pro hodnoty s menším či větším rozptylem atd. Je známo, že Shapiro-Wilkův test je méně spolehlivý při testování normality výběru s velkým množstvím identických hodnot.

Pro naše účely výuky zůstaneme u použití Shapiro-Wilkova testu ve Scilabu a Anderson-Darlingova testu v Matlabu.

Anderson-Darlingův test

Anderson-Darlingův test se používá pro ověření, zda hodnoty výběru pochází z určitého teoretického rozdělení (nemusí být pouze normální). Při testování normality spolu s testem Shapiro-Wilka je známý jako jeden z nejsilnějších statistických testů rozdělení. Statistika testu se spočte následovně:

$$A^2 = -n - S, \qquad \text{kde} \qquad S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[\ln(F(X_i)) + \ln\left(1 - F(X_{n+1-i})\right)\right],$$

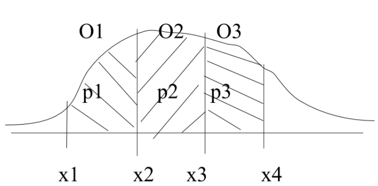
kde F je distribuční funkce pro normální rozdělení.

χ^2-test dobré shody

χ^2-test dobré shody (čte se chi-kvadrát) můžeme použít pro testování jakéhokoliv rozdělení. Statistika testu je

$$T = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

kde O_i – pozorované četnosti (angl. observed), tj., četnosti skutečně naměřených hodnot a E_i – očekávané četnosti (anlg. expected), tj., jak by vypadaly v případě teoretického rozdělení, na shodu s kterým data testujeme. Pro ověření normality χ^2-test dobré shody funguje následovně:



Máme hodnoty výběru X_i a pozorované četnosti na intervalech mezi nimi O_i. Abychom dostali očekávané četnosti E_i, spočítáme pravděpodobnosti p_i pomocí distribučních funkcí pro příslušné hodnoty:

$$p_1 = F(X_2) - F(X_1),$$

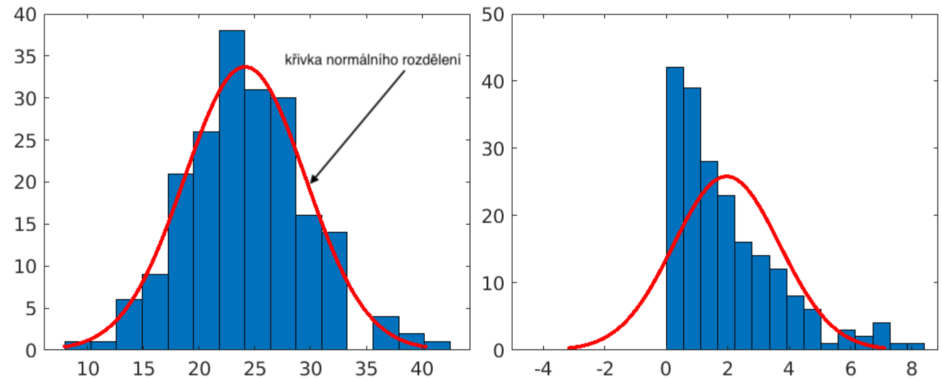
$$p_2 = F(X_3) - F(X_2),$$

$$p_3 = F(X_4) - F(X_3).$$

Dále vektor četností je $E = n * \underbrace{p}_{\text{vektor}} = [\underbrace{np_1}_{E_1}\ \underbrace{np_2}_{E_2}\ \underbrace{np_3}_{E_3}]$.

Grafické metody ověření normality

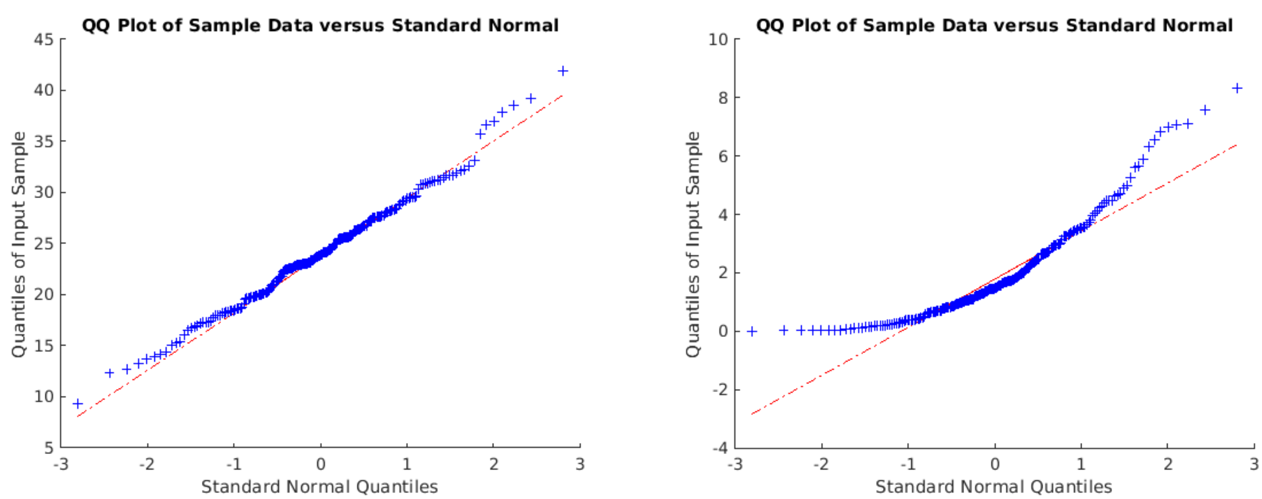
Ověření předpokladu normality je možné také pomocí grafických metod. Pokud máte dostatek dat, nejjednodušší způsob je vykreslit histogram dat z výběru, kde se můžete podívat, zda tvar histogramu odpovídá normálnímu rozdělení:



Vlevo na obrázku je vidět, že tvar histogramu odpovídá normálnímu rozdělení. Vpravo jsme vykreslili exponenciální rozdělení, které má zcela jiný histogram a křivka normálního rozdělení mu neodpovídá. Tento způsob se hodí spíš pro předběžnou analýzu, není vždy jednoznačný a není vhodný pro školní příklady s malým počtem dat.

Q-Q graf

Q-Q graf je kvantil-kvantil graf (angl. Q-Q plot), patří k grafickým metodám pro porovnání dvou rozdělení (v našem případě je jedno z nich normální) pomocí vykreslení jejich kvantilů proti sobě. Pokud data pochází z normálního rozdělení, budou hodnoty z výběru ležet na Q-Q grafu přibližně na stejné přímce.



Červená přímka na obrázcích znázorňuje vykreslené kvantily normálního rozdělení. Modrou barvou jsou označeny kvantily výběru. Na obrázku vlevo je vidět, že hodnoty z výběru leží lineárně kolem červené přímky, tedy můžeme předpokládat normalitu výběru. Na obrázku vpravo má modrá křivka nelineární povahu a neodpovídá červené přímce. V tomto případě data nepochází z normálního rozdělení.

Funkce pro Q-Q graf ve Scilabu a v Matlabu je **qqplot**.

Testy rozdělení jsou aktuální pro jakýkoliv počet výběrů – testujeme pak každý výběr zvlášť.

Testy hypotéz pro dva výběry

Teď se budeme věnovat testům hypotéz pro dva výběry, což z praktického hlediska potkáváme mnohem častěji než testy pro jeden výběr. Například, potřebujeme porovnat, zda je v Ostravě a Brně stejný počet nakažených. Nemůžeme otestovat úplně všechny, je to drahé. Vezmeme výběr v Ostravě a výběr v Brně a porovnáme dva výběry pomocí vhodného testu hypotéz. Opět bude tady velice důležité správně zvolit test. Uvidíme, že ve většině případů jsou testy pro dva výběry stejné jako pro jeden výběr, jenom je musíme umět použít pro dva výběry.

Podobně jako v případě jednovýběrových testů, volíme především mezi parametrickými testy s předpokladem normality obou výběrů a neparametrickymi testy pro výběry bez normality. To znamená, že každý výběr musíme otestovat zvlášť, a pokud jeden z nich nesplňuje tento předpoklad, musíme použít neparametrický test.

Dále ale máme pro dvouvýběrové testy další předpoklad: některé testy jsou určeny pro nepárové výběry, jiné pro párové.

Párové výběry dvou veličin X a Y znamená, že z jejich hodnot můžeme vytvořit dvojice, např., $\{X_1, Y_1\}$, $\{X_2, Y_2\}$ atd. Párové výběry mají vždy stejný počet dat. Například, denní přírůstek počtu nakažených v Praze v březnu a v květnu, měření teploty u pacienta ráno a večer, známky skupiny studentů z matematiky a fyziky, atd.

V případě nepárových výběrů se hodnoty obou výběrů měří zcela nezávisle na sobě a výběry tak mohou mít různý rozsah, např., měření rychlosti vozidel na dálnici směrem z Prahy a do Prahy, denní přírůstek počtu nakažených podle jednotlivých krajů v ČR a spolkových zemí v Německu, známky dvou skupin studentů z matematiky atd.

V následující tabulce jsou testy pro dva výběry, které budeme používat (je také dostupná na webu pod odkazem Jak zvolit test hypotéz). Dále je probereme podrobně.

Testy parametrů s předpokladem normality	Neparametrické testy bez předpokladu normality
Test o shodě středních hodnot při známém rozptylu **z_test_2** H_0: $\mu_1 = \mu_2 (>, <)$ známe σ_1^2, σ_2^2	↙↘ Pro dva nepárové výběry: Mann-Whitneyův test **mannwhit_test** Pro dva párové výběry: ↙↘ Test mediánu; Wilcoxonův test **wilcoxon_test**
Test o shodě středních hodnot při neznámém rozptylu H_0: $\mu_1 = \mu_2 (>, <)$ neznáme σ_1^2, σ_2^2 – použijeme s_1^2, s_2^2 ↙↘ Pro dva nepárové výběry: **t_test_2n** Pro dva párové výběry: **t_test_2p**	všechny levo-, pravo-, oboustranné H_0: $\tilde{X}_{0,5(1)} = \tilde{X}_{0,5(2)} (>, <)$
Test o shodě dvou rozptylů – **var_test_2** H_0: $\sigma_1^2 = \sigma_2^2 (>, <)$	
všechny levo-, pravo-, oboustranné	

Parametrické testy pro dva výběry s předpokladem normality

Abychom mohli použít tyto testy, pochopitelně předpokládáme, že oba výběry pochází z normálního rozdělení. Dále je postup práce s těmito testy stejný jako v případě jednoho výběru. Rozdíl je v tom, že tam jsme testovali, zda se parametr rovná určité předpokládané hodnotě, ale tady budeme porovnávat, zda jsou parametry obou souborů, ze kterých máme výběry, stejné na zvolené hladině významnosti.

Test o shodě středních hodnot při známých rozptylech

Příklad: Zajímá nás vliv nového léku na krevní tlak. Skupině 978 pacientů jsme podali nový lék, skupina 936 pacientů dostala placebo. Měřením tlaku dostaneme dva výběry:

$$X_{\text{lék}} = \underbrace{[120/60 \;\; 143/92 \;\; 117/92 \;\; \dots \;]}_{978 \text{ hodnot}}, \quad X_{\text{placebo}} = \underbrace{[137/86 \;\; 112/82 \;\; 111/86 \;\; \dots \;]}_{936 \text{ hodnot}}.$$

Testujeme tvrzení, že je krevní tlak po použití léku v průměru stejný jako při použití placebo.

Řešení: Podobně jako v případě jednovýběrových testů, zvolíme test podle informací v zadání:

- máme dva výběry – jsme u správné tabulky,
- předpokládáme, že podle testů rozdělení mají oba výběry normální rozdělení – jsme v parametrických testech (levý sloupec tabulky),
- potřebujeme zjistit, zda je krevní tlak v průměru stejný – jedná se o test o shodě dvou středních hodnot.
- Dále tady máme dva velké výběry 978 a 936 hodnot. U takto velkých výběrů můžeme jejich spočítané výběrové rozptyly podle limitní věty považovat za známé rozptyly souborů*.

To znamená, že můžeme použít test o shodě dvou středních hodnot při známých rozptylech, a to je **z_test_2**. Řekneme nulovou hypotézu (máte ji v tabulce pro každý test):

$$H_0 : \mu_{\text{lék}} = \mu_{\text{placebo}} \quad - \quad \text{podle tvrzení, tj., lék nemá vliv na krevní tlak.}$$

Alternativní hypotéza je opačné tvrzení:

$$H_A : \mu_{\text{lék}} \neq \mu_{\text{placebo}} \quad - \quad \text{tj., lék má vliv na krevní tlak.}$$

Všimněme si, že v tomto případě je test oboustranný. Statistika tohoto testu je

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \quad - \quad \text{pro jeden výběr je stejná, jen používá jenom } \bar{X} \text{ a } \sigma^2.$$

Dále dostaneme pomocí software p-hodnotu. Pokud bude p-hodnota menší než hladina významnosti $\alpha = 0.05$, nulovou hypotézu zamítáme. Závěr: lék má vliv. Všimněme si, že bychom mohli testovat, zda je krevní tlak po použití léku vyšší nebo nižší, než v případě placebo. V tomto případě by se jednalo o levostranný nebo pravostranný test.

Test o shodě středních hodnot při neznámých rozptylech pro dva nepárové výběry

Tento test (**t_test_2n**) použijeme v tom případě, když za splnění všech předpokladů (dva výběry, normalita, střední hodnota a neznámý rozptyl) máme dva nepárové výběry. Dost často tomuto testu říkají nepárový t-test (angl. unpaired t test). Ukážeme si, jak může skoro stejný příklad vyžadovat rozlišné testy.

Příklad: V jazykové škole mají dvouletý kurz angličtiny. Škola tvrdí, že 1.ročník má lepší výsledky.

Řešení: Tady máme dva nepárové výběry, které byly naměřeny nezávisle na sobě, tj., známky (nebo bodové hodnocení) 1.ročníku a 2.ročníku.

$$X_1 = \underbrace{[\dots\dots\dots\dots\dots]}_{\text{bodové hodnocení 1.ročníku rozsahu } n_1}, \qquad X_2 = \underbrace{[\dots\dots\dots\dots\dots]}_{\text{bodové hodnocení 2.ročníku rozsahu } n_2}.$$

*Podobná informace může také být poskytnutá výrobcem, např., povolené odchylky.

Počet dat nemusí být stejný, tj., $n_1 \neq n_2$. Nulová hypotéza zní:

$H_0 : \mu_1 = \mu_2$ – formálně a $\mu_1 > \mu_2$ – podle tvrzení, tj., bodové hodnocení studentů 1.ročníku je vyšší.

Alternativní hypotéza je opačné tvrzení:

$H_A : \mu_1 < \mu_2$ – bodové hodnocení studentů 1.ročníku je nižší, což určuje, že test je levostranný.

Dále postup je stejný. Všimněme si, že statistika tohoto testu

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \text{Studentovo rozdělení (pro jeden výběr je stejná, jen používá jenom } \bar{X} \text{ a } s^2\text{)}$$

je stejná jako pro **z_test_2** – liší se pouze použitím výběrových rozptylů místo neznámých rozptylů souborů.

Test o shodě středních hodnot při neznámých rozptylech pro dva párové výběry

Tento test (**t_test_2p**) zase použijeme za splnění všech předpokladů, ale budeme navíc mít párové výběry. Tomuto testu se také říká párový t-test (angl. paired t test). Nepatrně změníme náš příklad:

Příklad: V jazykové škole mají dvouletý kurz angličtiny. Škola tvrdí, že se studenti ve 2.ročníku zhorší.

Řešení: Zásadní rozdíl je v tom, že teď máme výběr studentů 1.ročníku a pak stejný výběr za rok, tedy výběry jsou párové.

$$X_1 = \underbrace{[\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots],}_{\text{bodové hodnocení 1.ročníku rozsahu } n} \qquad X_2 = \underbrace{[\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots].}_{\text{bodové hodnocení stejného ročníku za rok rozsahu } n}$$

Poznáme to především podle toho, že párové výběry mají vždy stejný počet dat. Nulová hypotéza zůstane stejná jako v předchozím příkladě:

$H_0 : \mu_1 = \mu_2$ – formálně a $\mu_1 > \mu_2$ – podle tvrzení, tj., bodové hodnocení studentů v 1.ročníku bylo vyšší.

Alternativní hypotéza je opačné tvrzení:

$H_A : \mu_1 < \mu_2$ – bodové hodnocení studentů 1.ročníku bylo nižší, což určuje, že test je levostranný.

Statistika tohoto testu je ale zcela jiná:

$$T = \frac{\bar{X}_d}{s_d}\sqrt{n} \sim \text{Studentovo rozdělení, kde} \quad \bar{X}_d = \frac{1}{n}\sum_{i=1}^{n}(X_{i(1)} - X_{i(2)}) \ \text{– střední hodnota rozdílů hodnot z výběrů,}$$

$$\text{a} \quad s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_{i(d)} - \bar{X}_d)^2} \ \text{– směrodatná odchylka rozdílů hodnot.}$$

Všimněme si, že na rozdíl od **z_test_2** a nepárového t-testu, používá tento párový t-test statistiku založenou na výpočtu odchylek dvojic z výběrů. Proto by ho nešlo použít pro nestejný počet dat.

Test o shodě dvou rozptylů

Tento test (**var_test_2**) je jediný pro testování shody rozptylů dvou výběrů. V textu příkladu nás na něj upozorní slova typu: výkonnost, spolehlivost, stabilita, variabilita, proměnlivost – cokoliv, co souvisí s odchylkami nebo rozmezím naměřených hodnot. Zase trochu pozměníme náš příklad:

Příklad: V jazykové škole mají dvouletý kurz angličtiny. Řeší se, jaký ročník poslat na jazykovou olympiádu. Škola tvrdí, že stabilita výsledků 1.ročníku je vyšší.

Řešení: Tady máme zase dva výběry

$$X_1 = \underbrace{[\ldots\ldots\ldots\ldots\ldots],}_{\text{bodové hodnocení 1.ročníku rozsahu } n_1} \qquad X_2 = \underbrace{[\ldots\ldots\ldots\ldots\ldots].}_{\text{bodové hodnocení 2.ročníku rozsahu } n_2}$$

Ale teď chceme porovnat variabilitu výsledků obou ročníků. Podle školy podává 1.ročník lepší výsledky stabilněji, což znamená, že odchylky mezi jejich výsledky jsou menší, tj., rozptyl rozdělení, ze kterého tento výběr pochází, je menší. Nulová hypotéza tady bude:

$H_0 : \sigma_1^2 = \sigma_2^2$ – formálně a $\sigma_1^2 < \sigma_1^2$ – podle tvrzení, tj., stabilita výsledků 1.ročníku je vyšší (rozptyl je menší).

Alternativní hypotéza je opačné tvrzení:

$H_A : \sigma_1^2 > \sigma_1^2$ – tj., stabilita výsledků 1.ročníku je nižší (rozptyl je větší), což určuje, že test je pravostranný.

Statistika testu je

$$T = \frac{s_1^2}{s_2^2} \sim \text{Fisherovo rozdělení.}$$

Neparametrické testy pro dva výběry bez předpokladu normality

Teď probereme neparametrické testy hypotéz, kdy jeden nebo oba výběry nepochází z normálního rozdělení. Podobně jako v předchozím případě rozlišujeme testy pro nepárové a párové výběry.

Mann-Whitneyův test

Mann-Whitneyův test (**mannwhit_test**) použijeme pro dva nepárové výběry bez předpokladu normality.

Příklad: Skupina pacientů byla léčena na koronavirus odlišnými metodami: buď nadějným lékem A z Ameriky nebo nadějným lékem B z Japonska. Počty dní, za kolik se pacienti uzdravili jsou zaznamenány v tabulce†. Je doba léčení stejná?

Doba léčení lékem A	14	17	23	7	11	13	...	...		
Doba léčení lékem B	28	10	16	4	19	9	14	11	6	...

Řešení: Řešení takové praktické úlohy začneme ověřením předpokladu normality. Máme dva výběry: X_A – doba léčení lékem A a X_B – lékem B. Pokud data dvou nepárových výběrů různého rozsahu nepochází z normálního rozdělení, je pro nás volba testu jednoduchá – je to Mann-Whitneyův test (podíváme se do tabulky). Řekneme si nulovou hypotézu:

$H_0 : \tilde{X}_{0,5(A)} = \tilde{X}_{0,5(B)}$ – mediány výběrů jsou stejné, tj., pacienti se vyléčili za stejnou dobu.

Alternativní hypotéza je opačné tvrzení:

$H_A : \tilde{X}_{0,5(A)} \neq \tilde{X}_{0,5(B)}$ – pacienti se nevyléčili za stejnou dobu, což určuje, že test je oboustranný.

Statistika je založena na sečtení pořadí hodnot, dále se bere minimum:

$$U_A = n_A n_B + \frac{n_A(n_A+1)}{2} - \underbrace{T_A,}_{\text{součet pořadí}} \qquad U_B = n_A n_B + \frac{n_B(n_B+1)}{2} - \underbrace{T_B,}_{\text{součet pořadí}} \qquad T = \min\{T_A, T_B\}.$$

---

†Data jsou fiktivní.

Test mediánu (znaménkový test)

V případě dvou párových výběrů bez normality můžeme použít test mediánu (znaménkový test) nebo Wilcoxonův test, se kterými jsme se už seznámili pro případ jednoho výběru. Ukážeme si na stejném příkladě použití obou testů.

Příklad: Testujeme, zda má skupina 11 studentů stejný počet vyřešených příkladů ze statistiky na začátku a na konci semestru. Nepředpokládáme normalitu dat. Data máme v tabulce:

počet příkladů na začátku semestru	1	3	4	0	1	2	3	5	4	2	3
počet příkladů na konci semestru	5	14	1	15	0	12	4	9	4	3	25

Řešení: Máme dva párové výběry X_1 – počet příkladů na začátku semestru, X_2 - na konci. Předpokládáme, že podle testů rozdělení nemají data normální rozdělení. Řekneme si nulovou hypotézu:

$$H_0 : \tilde{X}_{0,5(1)} = \tilde{X}_{0,5(2)} \quad - \text{ mediány výběrů jsou stejné, tj., počty příkladů jsou stejné.}$$

Alternativní hypotéza je opačné tvrzení:

$$H_A : \tilde{X}_{0,5(1)} \neq \tilde{X}_{0,5(2)} \quad - \text{ počty příkladů nejsou stejné, což určuje, že test je oboustranný.}$$

Statistika znaménkového testu se počítá podle rozdílu hodnot z výběrů, co můžeme znázornit v tabulce:

$X_{i(1)}$	$X_{i(2)}$	$X_{i(1)} - X_{i(2)}$	znaménka
1	5	1-5 = **- 4**	**-**
3	14	3-14 = **- 11**	**-**
4	1	4-1 = **3**	+
0	15	0-15 = **- 15**	**-**
...	...	...	...

V případě jednoho výběru se odečítá hodnota od mediánu. Dále se počet kladných a záporných znamének bere jako nový počet dat. Nulové rozdíly se vyhodí. Menší počet znamének (buď + nebo −) se označí písmenem b. Dále statistika je

$$T = \frac{2b - n}{\sqrt{n}} \sim N(0, 1).$$

Pro náš příklad p-hodnota = 0.1094, což znamená, že nezamítáme nulovou hypotézu, že studenti vyřeší stejný počet příkladů na začátku a na konci semestru. Nicméně v praxi se tento test používá spíš pro předběžnou analýzu. Je slabší, což můžeme ukázat jeho porovnáním s Wilcoxonovým testem.

Wilcoxonův test

Tento test (**wilcoxon_test**) jsme už používali na cvičení pro jeden výběr, teď se podíváme na jeho použití pro dva párové výběry bez předpokladu normality.

Použijeme stejný příklad. Nulová a alternativní hypotézy se nemění. Statistika tohoto testu se počítá podobně jako pro znaménkový test, ale s použitím pořadí hodnot:

$X_{i(1)}$	$X_{i(2)}$	$X_{i(1)} - X_{i(2)}$	$\lvert X_{i(1)} - X_{i(2)}\rvert$	znaménka	pořadí absolutních hodnot
1	5	1-5 = **- 4**	**4**	**-** (0)	2
3	14	3-14 = **- 11**	**11**	**-**(0)	3
4	1	4-1 = **3**	**3**	+(1)	1
0	15	0-15 = **- 15**	**15**	**-** (0)	4
...	...	...	...	...	...

Znaménka označíme písmenem s. Za každý $+$ bereme 1, za každý $-$ bereme 0. Pořadí absolutních hodnot označíme r. Statistika je

$$W = \sum_{i}^{n} s_i r_i,$$

což znamená, že každé pořadí vynásobíme buď 0 nebo 1. Pro tento příklad $W = 1 * 1 + 2 * 0 + 3 * 0 + 4 * 0 = 1$.

Dále se statistika porovnává s kritickou hodnotou ze speciální tabulky (nám to dělá software) a výsledná p-hodnota pro náš příklad je 0.0273. To znamená, že na hladině významnosti $\alpha = 0.05$ zamítáme nulovou hypotézu, že studenti vyřeší stejný počet příkladů na začátku a na konci semestru.

Všimneme si rozdílu oproti předchozímu příkladu, kde jsme na stejných datech nezamítli nulovou hypotézu a připomeneme si, že pouze zamítnutí hypotézy má statistický význam.