

MUNI
MED

MIAM021p(s) Analýza a management dat pro zdravotnické obory – přednáška a cvičení (jaro 2024)

MICHAL SVOBODA

Institut biostatistiky a analýz LF MU
svoboda@iba.muni.cz

Osnova

- Excel: opakování, příprava dat, základní vzorce
- Základy popisné statistiky
- Základní rozdělení pravděpodobnosti, testování hypotéz
- Parametrické testy
- Neparametrické testy
- Analýza kontingenčních tabulek
- Základy korelační analýzy a lineární regrese

Důležité informace

- Výuka: 11:00–13:30, D29/347-RCX2
- Materiály v IS
- Software: Microsoft Office - Excel, Statistica
- Pro získání zápočtu/kolokvia je třeba:
 - 1. Účast – povoleny jsou 2 absence**
 - 2. Domácí úkoly – povoleno 1 neodevzdání**
 - za účelem procvičení, dostanete zpětnou vazbu, na dalším cvičení se vrátíme, kdyby byl problém
 - 3. Závěrečný úkol** – praktické úkoly (povoleny materiály)

Organizace výuky

- 20. 2. – Excel: opakování, příprava dat, základní vzorce
- 27. 2. – Základy popisné statistiky
- 19. 3. – Základní rozdělení pravděpodobnosti, testování hypotéz
- 26. 3. – Parametrické testy
- 2. 4. – Neparametrické testy
- 9. 4. – Analýza kontingenčních tabulek, korelační analýza
- 16. 4. – Ukončení předmětu, test

Parametrické testy

- Mají **předpoklady** o rozložení vstupních dat (normální rozložení).
- Při stejném počtu pozorování (N) a dodržení předpokladů mají vyšší sílu testu než testy neparametrické.
- **Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být chybný.**



Proč nemusí parametrický a neparametrický test vyjít stejně?

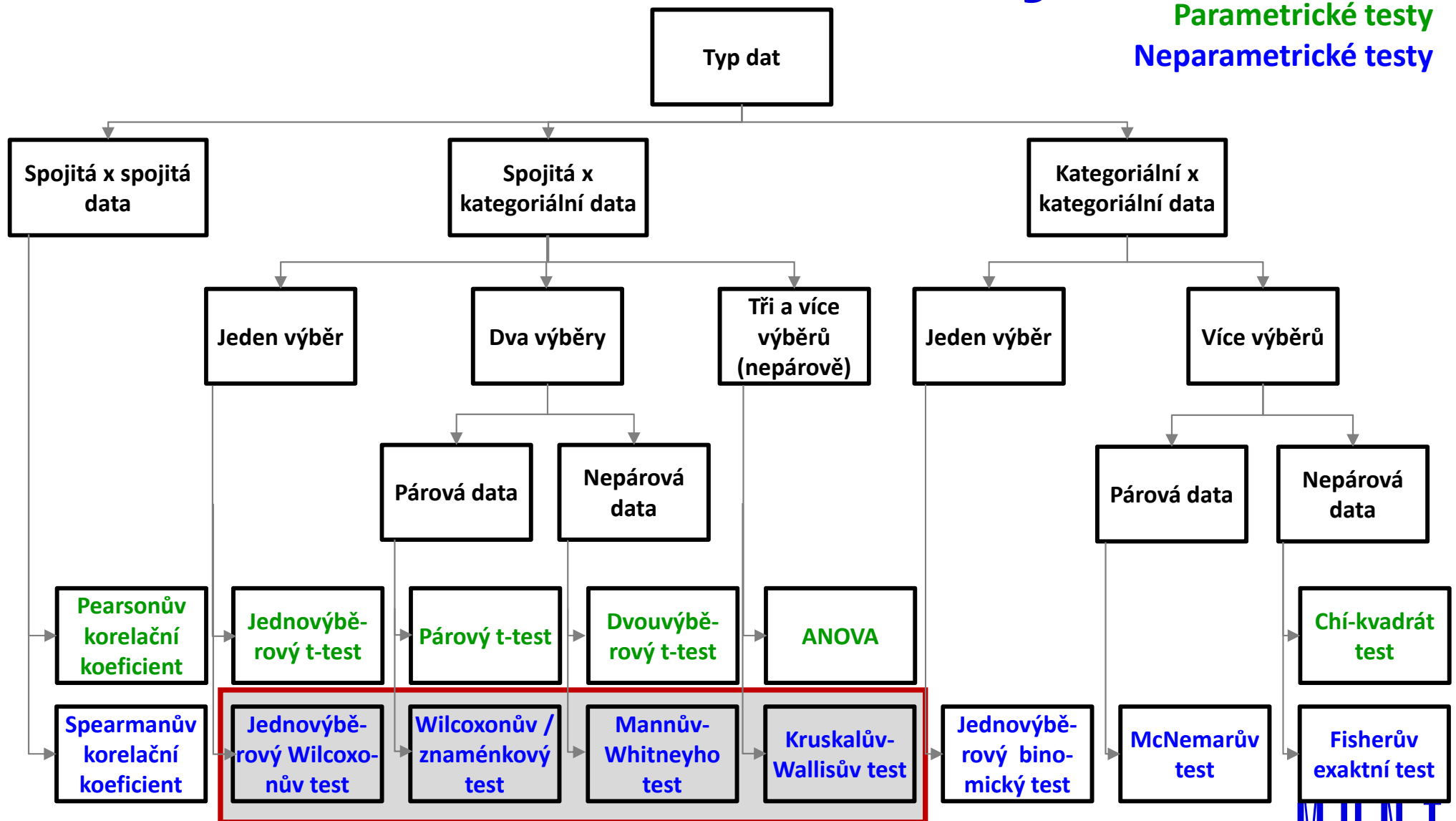
Neparametrické testy

- Vyžadují splnění **méně předpokladů** o rozložení vstupních dat, lze je tedy použít i při asymetrickém rozložení, přítomnosti odlehlých hodnot, či nedetekovatelném rozložení.
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí.
- Používají se také při hodnocení souborů s nízkým počtem pozorování (N ; malé soubory), kdy nejsme schopni normalitu dat spolehlivě ověřit.



Neparametrické testy

Základní statistické testy



Statistické testy o parametrech jednoho výběru

Jednovýběrový Wilcoxonův test
Jednovýběrový znaménkový test

Jednovýběrový test

Jednovýběrový Wilcoxonův test

- Předpokladem je **symetrické rozdělení** dat kolem mediánu.

Jednovýběrový znaménkový test

- Lze použít v situaci, kdy není splněn předpoklad symetrie rozdělení kolem mediánu.
- Oba testy testují, zde je **medián** jednoho výběru roven hodnotě c (v případě párového designu je $x_{0,5}$ reprezentováno mediánem rozdílu hodnot).

$$H_0: x_{0,5} = c \quad \text{proti } H_A: x_{0,5} \neq c$$

Jednovýběrový Wilcoxonův test

Postup:

1. Určíme rozdíly hodnot výběru s testovanou hodnotou mediánu.
2. Absolutní hodnoty rozdílů uspořádáme vzestupně a přiřadíme jim pořadí.
3. Spočítáme statistiky S_w^+ a S_w^- , které odpovídají **součtu pořadí kladných (S_w^+) a záporných rozdílů (S_w^-)**. Jako finální hodnotu testové statistiky bereme minimum z S_w^+ a S_w^- . Nulovou hypotézu zamítáme, pokud je hodnota testové statistiky menší nebo rovna tabelované kritické hodnotě (při dané hladině významnosti a počtu nenulových rozdílů), nebo když příslušná p-hodnota \leq zvolená hladina významnosti.

Nebo: Pro $N > 30$ lze využít asymptotické normality statistiky S_w^+ .

Jednovýběrový Wilcoxonův test

Ukázka výpočtu: U 15 pacientů byla vyhodnocena doba, kterou museli strávit v čekárně u lékaře. Zjistěte, zda medián čekací doby je roven půl hodině.

| ID | Doba čekání | Medián | Rozdíl | Rozdíl | Pořadí |
|----|-------------|--------|--------|--------|--------|
| 1 | 1 | 30 | -29 | 29 | 15 |
| 2 | 45 | 30 | 15 | 15 | 10 |
| 3 | 25 | 30 | -5 | 5 | 3,5 |
| 4 | 15 | 30 | -15 | 15 | 10 |
| 5 | 34 | 30 | 4 | 4 | 2 |
| 6 | 19 | 30 | -11 | 11 | 8 |
| 7 | 31 | 30 | 1 | 1 | 1 |
| 8 | 25 | 30 | -5 | 5 | 3,5 |
| 9 | 8 | 30 | -22 | 22 | 14 |
| 10 | 12 | 30 | -18 | 18 | 12 |
| 11 | 20 | 30 | -10 | 10 | 6 |
| 12 | 15 | 30 | -15 | 15 | 10 |
| 13 | 40 | 30 | 10 | 10 | 6 |
| 14 | 20 | 30 | -10 | 10 | 6 |
| 15 | 10 | 30 | -20 | 20 | 13 |



➔ $S_w^+ = 19$ $S_w^- = 101$

$\min(S_w^+, S_w^-) = 19$

Kritická hodnota $w_{15(0,05)} = 25$

Hodnota testové statistiky je menší než kritická hodnota

➔ **zamítáme H_0 .**

Jednovýběrový znaménkový test

Postup:

1. Spočítáme rozdíly hodnot výběru s testovanou hodnotou mediánu.
2. Spočítáme statistiku S_z^+ , která odpovídá počtu kladných rozdílů → **test nevyužívá hodnot pořadí původních dat, ale pouze informaci, zda se hodnota realizuje nad nebo pod mediánem** → dochází ke snížení síly testu.
3. Nulovou hypotézu zamítáme, pokud statistika S_z^+ realizuje v kritickém oboru hodnot $W = (0, k_1) \cup (k_2, n)$, kde n odpovídá počtu nenulových rozdílů a hodnoty k_1 a k_2 lze dohledat v matematických tabulkách; nebo když příslušná p-hodnota \leq zvolená hladina významnosti.
Nebo: Pro $N > 20$ lze využít asymptotické normality statistiky S_z^+ .

Jednovýběrový znaménkový test

Ukázka výpočtu: U 15 pacientů byla vyhodnocena doba, kterou museli strávit v čekárně u lékaře. Zjistěte, zda medián čekací doby je roven půl hodině.

| ID | Doba čekání | Medián | Rozdíl | Větší než medián? |
|----|-------------|--------|--------|-------------------|
| 1 | 1 | 30 | -29 | Ne |
| 2 | 45 | 30 | 15 | Ano |
| 3 | 25 | 30 | -5 | Ne |
| 4 | 15 | 30 | -15 | Ne |
| 5 | 34 | 30 | 4 | Ano |
| 6 | 19 | 30 | -11 | Ne |
| 7 | 31 | 30 | 1 | Ano |
| 8 | 25 | 30 | -5 | Ne |
| 9 | 8 | 30 | -22 | Ne |
| 10 | 12 | 30 | -18 | Ne |
| 11 | 20 | 30 | -10 | Ne |
| 12 | 15 | 30 | -15 | Ne |
| 13 | 40 | 30 | 10 | Ano |
| 14 | 20 | 30 | -10 | Ne |
| 15 | 10 | 30 | -20 | Ne |



➔ $S_z^+ = 4$

Kritický obor:

$W = (0,3) \cup (12,15)$

Hodnota statistiky se realizuje mimo kritický obor hodnot

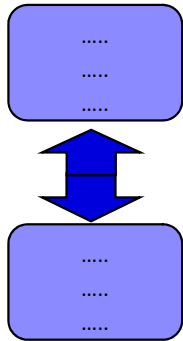
➔ **nezamítáme H_0 .**

Statistické testy o parametrech dvou výběrů

Nepárový Mannův-Whitneyův test
Párový Wilcoxonův a znaménkový test

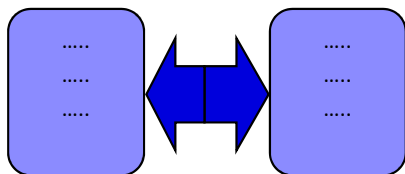
Párový a nepárový test

- Srovnání dvou **nezávislých** výběrů:



Nepárový Mannův-Whitneyův U test

- Srovnání dvou **závislých** výběrů:



Párový Wilcoxonův test
Párový znaménkový test

Mannův-Whitneyův U test

- Neparametrická alternativa dvouvýběrového t-testu
- **Počítá s pořadím hodnot** namísto s původními daty.
- Testuje **nulovou hypotézu o shodě rozdělení**, ze kterého pocházejí porovnávané výběry.
- Když chceme interpretovat výsledek testu jako test o poloze (střední hodnoty jsou stejné), musíme **předpokládat**, že tvar rozdělení je v obou skupinách stejný.
- **Poznámka:** test lze použít i pro ordinální data (např. hodnocení zdravotního stavu na stupnici 1-5 apod.).

Mannův-Whitneyův U test

Postup:

1. Stanovíme nulovou a alternativní hypotézu ($F(x)$ – distribuční funkce): $H_0: F(x_1) = F(x_2)$ $H_A: F(x_1) \neq F(x_2)$.
2. Hodnoty obou výběrů (skupin) jsou sloučena a je určeno jejich pořadí v tomto sloučeném souboru.
3. Pro oba výběry zvlášť je spočítán součet pořadí (T_1 a T_2).
4. Ze součtů pořadí ve skupinách je určena finální hodnota testové statistiky U .

Mannův-Whitneyův U test

5. Hodnotu testové statistiky U porovnáme s kritickou hodnotou testu, pokud je tato hodnota menší než kritická hodnota testu, zamítáme nulovou hypotézu shody distribučních funkcí obou skupin.
6. Pro **velká n_1 a $n_2 (> 30)$** lze využít asymptotické normality statistiky U .

Mannův-Whitneyův U test

| ID | Délka výcviku | Skupina | Pořadí |
|----|---------------|-----------|--------|
| 1 | 35 | pozitivne | 1 |
| 2 | 41 | pozitivne | 2 |
| 3 | 43 | pozitivne | 4 |
| 4 | 44 | pozitivne | 5 |
| 5 | 47 | pozitivne | 7,5 |
| 6 | 48 | pozitivne | 9,5 |
| 7 | 48 | pozitivne | 9,5 |
| 8 | 51 | pozitivne | 11 |
| 9 | 42 | negativne | 3 |
| 10 | 46 | negativne | 6 |
| 11 | 47 | negativne | 7,5 |
| 12 | 53 | negativne | 12 |
| 13 | 54 | negativne | 13 |
| 14 | 57 | negativne | 14 |
| 15 | 59 | negativne | 15 |
| 16 | 65 | negativne | 16 |
| 17 | 74 | negativne | 17 |

Ukázka výpočtu: 17 štěňat bylo trénováno k hygienickým návykům pomocí pozitivní (8 štěňat) nebo negativní motivace (9 štěňat). Zjistěte, zda se tyto dva přístupy liší.



→ $T_1 = 49,5 \quad T_2 = 103,5$

$U_1 = 58,5 \quad U_2 = 13,5$

$\min(U_1, U_2) = 13,5$

Kritická hodnota $U_{(8,9;0,05)} = 15$

Hodnota testové statistiky je menší než kritická hodnota

→ **zamítáme H_0 .**

Párový Wilcoxonův test a znaménkový test

- Vycházíme z rozdílů párových hodnot a přecházíme na design jednovýběrových testů.
- Testuje, zda je **medián diferencí (D)** párových hodnot roven hodnotě 0.

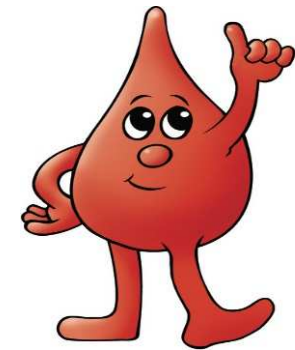
$$H_0: D_{0,5} = 0 \quad H_A: D_{0,5} \neq 0$$

- Dále postupujeme stejně jako u jednovýběrových testů výpočtem testové statistiky S_w^+ a S_w^- (u Wilcoxonova testu), resp. S_z^+ (u znaménkového testu) a jejich porovnáním s kritickou hodnotou, resp. s kritickým intervalem (nebo pro větší vzorky použijeme aproximaci normálním rozdělením).

Párový Wilcoxonův test

Ukázka výpočtu: U 10 pacientů byla zjištěna hodnota krevního parametru před a po podání léku. Zjistěte, zda se hodnoty před a po podání léku liší.

| ID | Před | Po | Rozdíl | Rozdíl | Pořadí |
|----|------|-----|--------|--------|--------|
| 1 | 142 | 138 | 4 | 4 | 4,5 |
| 2 | 140 | 136 | 4 | 4 | 4,5 |
| 3 | 144 | 147 | -3 | 3 | 3 |
| 4 | 144 | 139 | 5 | 5 | 7 |
| 5 | 142 | 143 | -1 | 1 | 1 |
| 6 | 146 | 141 | 5 | 5 | 7 |
| 7 | 149 | 143 | 6 | 6 | 9,5 |
| 8 | 150 | 145 | 5 | 5 | 7 |
| 9 | 142 | 136 | 6 | 6 | 9,5 |
| 10 | 148 | 146 | 2 | 2 | 2 |



➡ $S_w^+ = 51$ $S_w^- = 4$

$\min(S_w^+, S_w^-) = 4$

Kritická hodnota $w_{10(0,05)} = 8$

Hodnota testové statistiky je menší než kritická hodnota

➡ **zamítáme H_0 .**

Statistické testy o parametrech tří a více výběrů

Kruskalův-Wallisův test

Kruskalův-Wallisův test

- Neparametrická alternativa analýzy rozptylu (ANOVA)
- Zobecnění Mannova-Whitneyova U testu pro více než dvě srovnávané skupiny.
- **Počítá s pořadím hodnot** v souborech namísto s původními daty.
- Nulová hypotéza předpokládá stejné rozdělení pravděpodobnosti veličiny ve všech skupinách.
- **Předpoklad:** tvar rozdělení je ve všech skupinách stejný.
- **Poznámka:** test lze použít i pro ordinální data (např. hodnocení zdravotního stavu na stupnici 1-5 apod.).

Kruskalův-Wallisův test

Postup:

1. Stanovíme nulovou a alternativní hypotézu pro k skupin ($F(x)$ – distribuční funkce):
$$H_0: F(x_1) = F(x_2) = \dots = F(x_k)$$
$$H_A: \text{alespoň jedna } F(x_i) \text{ se liší od ostatních}$$
2. Hodnoty všech výběrů (skupin) jsou sloučena a je určeno jejich pořadí v tomto sloučeném souboru.
3. Pro všechny výběry zvlášť je spočítán součet pořadí (T_1, \dots, T_k).
4. Ze součtu pořadí ve skupinách je určena finální hodnota testové statistiky Q :

Kruskalův-Wallisův test

5. Pokud je $Q \geq \chi^2 (k-1)$, nebo když příslušná p-hodnota \leq zvolená hladina významnosti, zamítáme nulovou hypotézu. Pro malé velikosti výběrů určujeme kritický obor z tabulek pro Kruskalův-Wallisův test.
6. V případě zamítnutí nulové hypotézy pokračujeme dále hledáním lišících se dvojic pomocí **metod mnohonásobného porovnávání**.

**M U N I
M E D**

Praktické cvičení v programu Statistica



Datový soubor

Rehabilitace po mozkovém infarktu

Data: 02_Biostatistika_Data02.sta* (24v by 407c)

| Rehabilitace po mozkovem infarktu: data | | | | | | | | | | |
|---|----|---------|-----|------------|----------------|------------------|----------|-------------|------------------------|----------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | ID | Pohlavi | Vek | Etiologie | Lokalizace | Terapie | Komorbid | Barthel_inc | Kategorie_zavislosti_p | Ukoncen |
| 1 | 1 | muž | 82 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 25 | vysoce závislý | propuště |
| 2 | 2 | žena | 81 | embolie | mozkové tepny | jiná farmakolog | 2 | 20 | vysoce závislý | přeložen |
| 3 | 3 | muž | 55 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 35 | vysoce závislý | propuště |
| 4 | 4 | žena | 46 | embolie | mozkové tepny | intravenózní trc | 0 | 20 | vysoce závislý | propuště |
| 5 | 5 | muž | 76 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 45 | částečně soběstačný | propuště |
| 6 | 6 | muž | 72 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 25 | vysoce závislý | přeložen |
| 7 | 7 | muž | 62 | trombóza | mozkové tepny | jiná farmakolog | 0 | 40 | vysoce závislý | propuště |
| 8 | 8 | muž | 64 | trombóza | přívodní tepny | jiná farmakolog | 0 | 15 | vysoce závislý | propuště |
| 9 | 9 | žena | 82 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 10 | vysoce závislý | přeložen |
| 10 | 10 | muž | 58 | trombóza | mozkové tepny | jiná farmakolog | 0 | 25 | vysoce závislý | propuště |
| 11 | 11 | muž | 84 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 40 | vysoce závislý | propuště |
| 12 | 12 | žena | 92 | okluze nek | mozkové tepny | jiná farmakolog | 0 | 30 | vysoce závislý | propuště |
| 13 | 13 | žena | 79 | embolie | mozkové tepny | jiná farmakolog | 1 | 40 | vysoce závislý | propuště |
| 14 | 14 | muž | 69 | trombóza | mozkové tepny | jiná farmakolog | 3 | 45 | částečně soběstačný | propuště |

Rehabilitace po mozkovém infarktu

- Cvičný datový soubor obsahuje záznamy o **celkem 407 pacientech hospitalizovaných pro mozkový infarkt** na neurologickém oddělení akutní péče, kde jim byla poskytnuta terapie pro obnovu krevního oběhu v postižené části mozku.
- Po zvládnutí akutní fáze byl u pacientů vyhodnocen stupeň soběstačnosti v základních denních aktivitách (ADL) pomocí tzv. **indexu Barthelové (BI)** a byli přeloženi na **rehabilitační oddělení**.
- Po dvou týdnech byl opět dle BI vyhodnocen stupeň soběstačnosti a pacienti byli buď propuštěni do ambulantní péče, nebo přeloženi na oddělení následné péče.

Rehabilitace po mozkovém infarktu

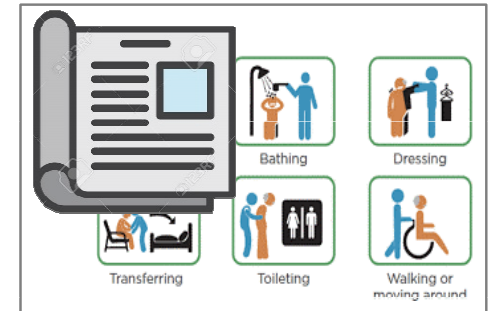
Sbírané informace:

- základní demografické údaje (**pohlaví a věk**),
- informace o samotné diagnóze mozkové příhody (**etiologie a lokalizace uzávěru cévy**),
- informace o léčbě (typ indikované **terapie a výskyt komplikací**)
- informace o **způsobu ukončení rehabilitace**.
- Stupeň soběstačnosti před rehabilitací byl dodatečně zjištěn z neurologie a na konci rehabilitace byl vyplněn nový dotazník pro určení výsledného **indexu Barthelové**.

Úkol 1. Jednovýběrový Wilcoxonův test

Úkol č. 1 – Jednovýběrový Wilcoxonův test

Zadání: „V podobné zahraniční studii byla publikovaná střední hodnota indexu Barthelové na konci akutní rehabilitace po mozkovém infarktu ve výši 64,4. Zjistěte, zda výsledné dosažení stupně soběstačnosti dle BI ve vašich datech je stejné nebo jiné než v této studii.“



Postup:


1. Ověříme předpoklady testu: Normalita rozložení hodnot indexu Barthelové na konci rehabilitace (ověříme vizuálně i statistickým testem – Shapiro-Wilkův test).

Úkol č. 1 – Jednovýběrový Wilcoxonův test

Postup (po nemožnosti použít jednovýběrový t-test):

1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu
 $H_0: Me = 64,4$ proti $H_A: Me \neq 64,4$
2. Původní hodnoty Barthelové indexu převedeme na pořadí (určené podle absolutní hodnoty rozdílu oproti referenci).
3. Vypočítáme **testovou statistiku S_w nebo Z** a odpovídající **p-hodnotu**.

$$S_w = 41\,099 \quad Z = 0,17 \Rightarrow p = 0,861$$

4. Vypočítané statistiky porovnáme s kritickou hodnotou, nebo porovnáme p-hodnotu s hladinou významnosti $\alpha = 0,05$.
5. Je-li **p-hodnota $> \alpha$**  **nezamítáme H_0 . Výsledná soběstačnost pacientů v našem souboru se neliší od výsledků publikovaných v porovnávané studii.**

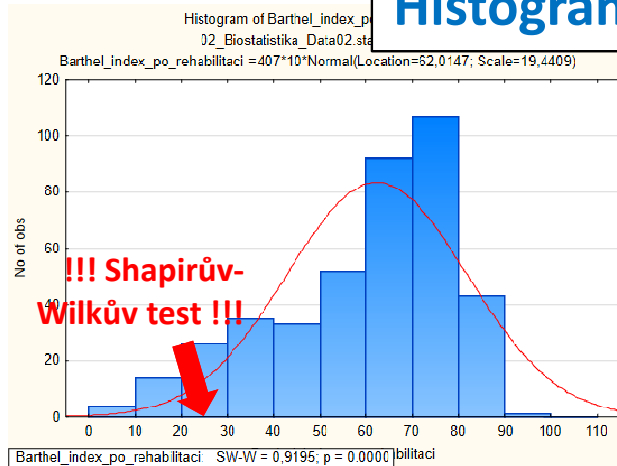
Úkol č. 1 – Ověření normality a popis dat

① Průměr a medián se výrazně liší (průměr 62 bodů, medián 70 bodů. Data jsou nejspíše asymetrická.

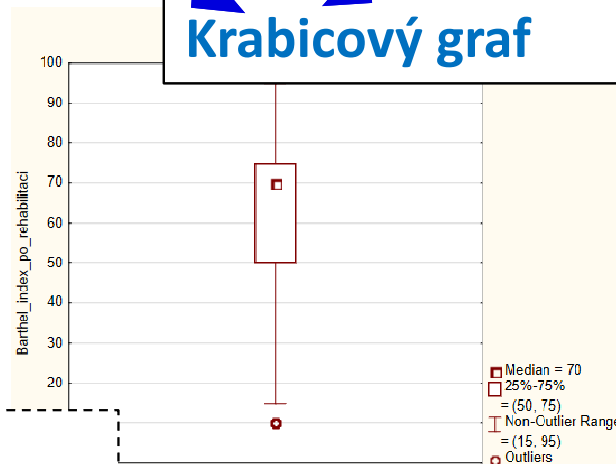
Srovnání průměru a mediánu

| Variable | Descriptive Statistics (02_Biostatistik) | | |
|-------------------------------|--|----------|----------|
| | Valid N | Mean | Median |
| Barthel_index_po_rehabilitaci | 407 | 62,01474 | 70,00000 |

Histogram



Krabicový graf



Diagnostický N-P graf



② Asymetrie je patrná i z krabicového grafu a histogramu. Z histogramu je navíc zřetelně vidět odlišnost od normálního rozdělení. Odchyłky od normality jsou patrné i z N-P grafu.

③ Na základě p-hodnoty $< 0,001$ zamítáme nulovou hypotézu o normalitě (tj. zamítáme, že není rozdíl mezi pozorovanými daty a teoretickým normálním rozdělením, ... tj. data nejsou normálně rozdělená).

Úkol č. 1 – Řešení v programu Statistica

- Do datové tabulky je potřeba přidat sloupec obsahující konstantní hodnotu reference, se kterou porovnáváme naše výsledky.
- V menu **Statistics** zvolíme **Nonparametrics**, vybereme **Comparing two dependent samples (groups)**.

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Nonparametrics' option is selected. The 'Comparing two dependent samples (variables)' option is highlighted. A red arrow labeled '1' points to the 'Statistics' menu, a red arrow labeled '2' points to the 'Nonparametrics' option, and a red arrow labeled '3' points to the 'Comparing two dependent samples (variables)' option.

| Rehabilitace po mozkovém infarktu: data | | | | |
|---|------------------------|------------------------|---------|--------------------|
| | 21 | 22 | 23 | 24 |
| | Barthel_index_po_rehal | Kategorie_zavislosti_p | Barthel | Barthel_index_refe |
| 1 | | 85 částečně soběstačný | -60 | 64,4 |
| 2 | | 40 zcela závislý | 20 | 64,4 |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |

Úkol č. 1 – Řešení v programu Statistica

- Vybereme proměnné (*Variables*), které chceme testovat (testovaný parametr a reference).
- Kliknutím na *Wilcoxon matched pair test* získáme výsledky.

The screenshot shows the Statistica software interface. The main window displays a data table titled "Rehabilitace po mozkovém infarktu: data". The table has columns for patient ID (21-24), Barthel index after rehabilitation, category of dependence, Barthel index reference, and Barthel index reference. A dialog box titled "Comparing two variables: 02_Biostatistik..." is open over the table. The dialog box has a "Variables" section with "List 1: Barthel_index_po_rehabilitaci" and "List 2: Barthel_index_reference". The "Quick" section has three options: "Sign test", "Wilcoxon matched pairs test", and "Box & whisker plots for all variables". The "Sign test" option is selected. A red arrow labeled "1" points to the "Variables" section, and another red arrow labeled "2" points to the "Wilcoxon matched pairs test" option. The "Options" section has a "By Group" button and a "p-value for highlighting:" field set to ".05".

| | 21 | 22 | 23 | 24 |
|----|------------------------|------------------------|----------|---------------------|
| | Barthel_index_po_rehal | Kategorie_zavislosti_p | Barthel_ | Barthel_index_refer |
| 1 | | 85 částečně soběstačný | -60 | 64,4 |
| 2 | | | | 64,4 |
| 3 | | | | 64,4 |
| 5 | | | | 64,4 |
| 6 | | | | 64,4 |
| 7 | | | | 64,4 |
| 8 | | | | 64,4 |
| | | | | 64,4 |
| | | | | 64,4 |
| 11 | | | | 64,4 |
| 12 | | | | 64,4 |
| 13 | | | | 64,4 |
| 14 | | | | 64,4 |
| 15 | | 35 vysoce závislý | -10 | 64,4 |
| 16 | | 70 částečně soběstačný | 20 | 64,4 |

Úkol č. 1 – Výsledky v Statistica

Rozsah výběru

Hodnota testové statistiky S_w a Z

| Pair of Variables | Wilcoxon Matched Pairs Test (02_Biostatistika_) | | | |
|---|---|----------|----------|----------|
| | Valid N | T | Z | p-value |
| Barthel_index_po_rehabilitaci & Barthel_index_reference | 407 | 41099,00 | 0,174762 | 0,861267 |

p-hodnota
Wilcoxonova testu



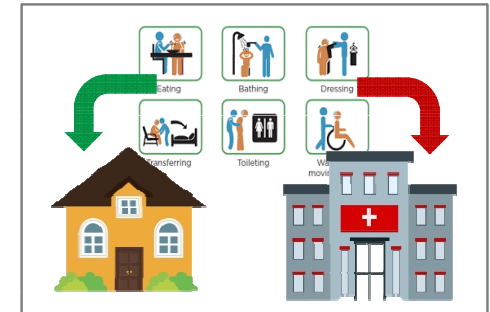
⚠️ Pozorovaný výsledný medián Barthelové indexu je 70 bodů, což je oproti výsledku 64,4 bodů v porovnávané studii **lepší výsledný stav o 5,6 bodů**.

② P-hodnota statistické významnosti tohoto pozorovaného rozdílu je ale $p = 0,861$, což na hladině významnosti 0,05 značí **nevýznamný rozdíl**, a z dostupných dat tedy nelze prokázat, že by výsledná soběstačnost pacientů léčených s mozkovým infarktem v našem souboru byla **odlišná od výsledků publikovaných v porovnávané studii**.

Úkol 2. Dvouvýběrový Mannův-Whitneyův test

Úkol č. 2 – Dvouvýběrový Mannův-Whitneyův test

Zadání: „U pacientů hospitalizovaných pro mozkový infarkt by po úspěšné terapii a absolvování akutní rehabilitace měl následovat přesun do ambulantní péče nebo na následné lůžko k pokračování v další rehabilitaci. Při správném managementu péče by do následné lůžkové péče měli pokračovat pouze pacienti, u kterých dosud nebylo dosaženo dostatečné rekonvalescence. Zkontrolujte, zda pacienti překládaní na následné lůžko mají skutečně horší míru soběstačnosti v základních denních aktivitách (ADL) vyjádřenou indexem Barthelové určenou v době propouštění.“



Úkol č. 2 – Dvouvýběrový Mannův-Whitneyův test

Postup (po nemožnosti použít dvouvýběrový t-test):

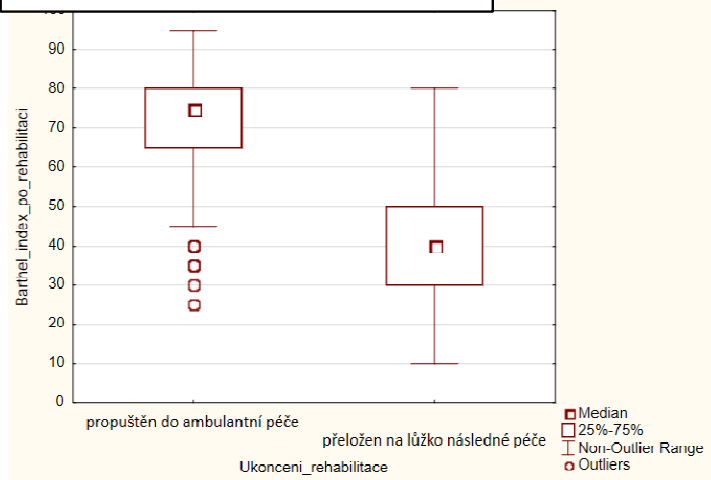
1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu
 $H_0: F(x_1) = F(x_2)$ proti $H_A: F(x_1) \neq F(x_2)$
2. Původní hodnoty Barthelové indexu převedeme na pořadí v celém souboru.
3. Vypočítáme **testovou statistiku U nebo Z** a odpovídající **p-hodnotu**.
$$U = 1\,998 \quad Z = 14 \quad \Rightarrow \quad p < 0,001$$
4. Vypočítané statistiky porovnáme s kritickou hodnotou, nebo porovnáme p-hodnotu s hladinou významnosti $\alpha = 0,05$.
5. Je-li p-hodnota $\leq \alpha$ **➡ zamítáme H_0 . Aktuální soběstačnost pacientů je určující pro jejich další pokračování v systému zdravotní péče.**

Úkol č. 2 – Ověření normality a popis dat

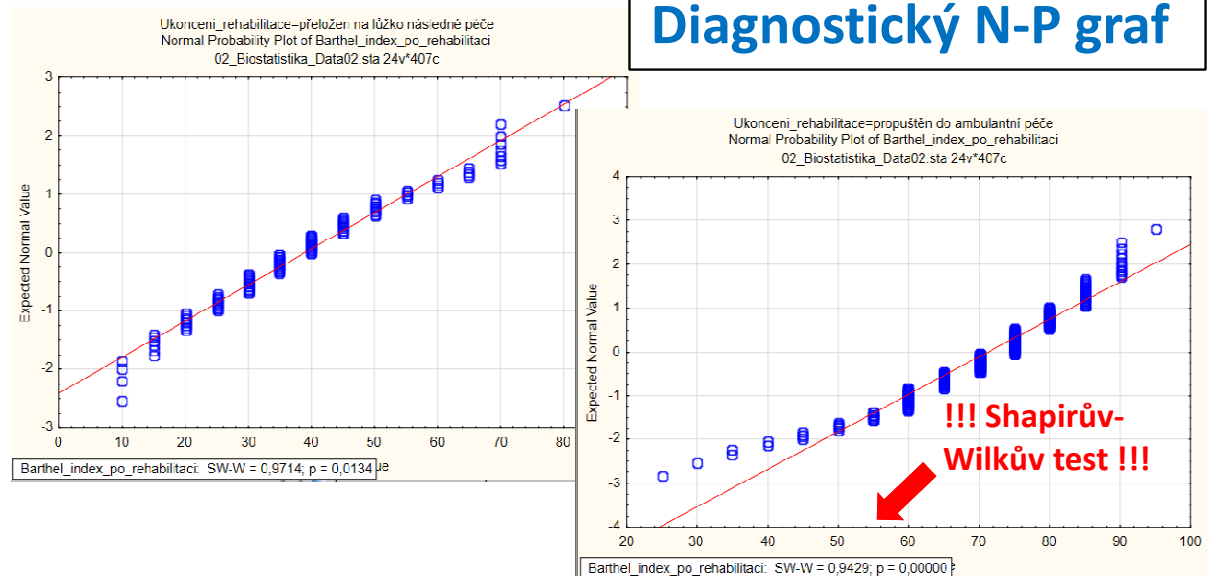
Popis dat

| Aggregate Results | | | | |
|--|---------------------------------|---------|----------|----------|
| Descriptive Statistics (02_Biostatistika_Data02.sta) | | | | |
| Variable | Ukocneni_rehabilitace | Valid N | Mean | Median |
| Barthel_index_po_rehabilitaci | propuštěn do ambulantní péče | 290 | 71,34483 | 75,00000 |
| Barthel_index_po_rehabilitaci | přeložen na lůžko následné péče | 117 | 38,88889 | 40,00000 |

Krabicový graf



Diagnostický N-P graf



① Základní popis i grafické srovnání ukazuje výrazný rozdíl mezi skupinami (soběstačnost při propuštění do ambulantní péče je v mediánu 75 bodů, ale pacienti pokračující do následné péče mají medián pouze 40 bodů).

② Normalitu dat zamítáme u obou skupin ($p = 0,013$ a $p < 0,001$) a přinejmenším u pacientů propuštěných domů je výrazné porušení normality patrné graficky i z N-P grafu.

Úkol č. 2 – Řešení v programu Statistica

- V menu **Statistics** zvolíme **Nonparametrics**, vybereme **Comparing two independent samples (groups)**.

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Nonparametrics' option is selected. A red arrow labeled '1' points to the 'Statistics' menu, and another red arrow labeled '2' points to the 'Nonparametrics' option. The 'Nonparametric Statistics' dialog box is open, and the 'Comparing two independent samples (groups)' option is selected. A red arrow labeled '3' points to this option. The background shows a data table with columns 'Ukonceni_rehabilitace', 'BI_dotaznik_po_rehab_1', and 'BI_dotaznik_po'.

| | Rehabilitace po mozkovem infarktu: data | | |
|----|---|------------------------|----------------|
| | 10 | 11 | 12 |
| | Ukonceni_rehabilitace | BI_dotaznik_po_rehab_1 | BI_dotaznik_po |
| 1 | propu | | |
| 2 | přelož | | |
| 3 | propu | | |
| 4 | propu | | |
| 5 | propu | | |
| 6 | přelož | | |
| 7 | propu | | |
| 8 | přelož | | |
| 9 | přelož | | |
| 10 | propu | | |
| 11 | propu | | |
| 12 | propu | | |
| 13 | propu | | |
| 14 | propu | | |
| 15 | přelož | | |

Úkol č. 2 – Řešení v programu Statistica

- Vybereme proměnnou, kterou chceme testovat (*dependent*) a proměnnou obsahující skupiny, které srovnáváme (*grouping*).
- Kliknutím na **Mann-Whitney U test**, nebo na M-W U test získáme výstupy.

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Comparing Two Groups' dialog box is displayed. The dialog box is titled 'Comparing Two Groups: 02_Biostatistika_...' and has a 'Variables' section where 'Barthel_index_po_rehabilitaci' is selected as the dependent variable and 'Ukonceni_rehabilitace' is selected as the grouping variable. The 'Codes for:' section shows 'Group 1: "propuštěn do..."' and 'Group 2: "přeložen na lůž..."'. The 'Quick' section has several test options, with 'Mann-Whitney U test' selected. The 'Apply continuity correction' checkbox is checked. The 'p-value for highlighting:' is set to .05. A red arrow labeled '1' points to the 'Variables' section, and a red arrow labeled '2' points to the 'Mann-Whitney U test' option.

Úkol č. 2 – Výsledky v Statistica

Hodnota testové statistiky *U* a *Z*

Rozsahy výběru obou skupin

| Mann-Whitney U Test (w/ continuity correction) (02_Biostatistika_Data02.sta) | | | | | | | | | |
|--|---------------------------------------|--|----------|----------|---------|------------|---------|--------------------------------------|---|
| By variable Ukonceni_rehabilitace | | | | | | | | | |
| Marked tests are significant at p <,05000 | | | | | | | | | |
| variable | Rank Sum propuštěn do ambulantní péče | Rank Sum přeložen na lůžko následné péče | U | Z | p-value | Z adjusted | p-value | Valid N propuštěn do ambulantní péče | Valid N přeložen na lůžko následné péče |
| Barthel_index_po_rehabilitaci | 74127,00 | 8901,000 | 1998,000 | 13,93442 | 0,0000 | 14,00397 | 0,0000 | 290 | 117 |

Z předchozího popisu je patrný výrazný rozdíl mezi skupinami (soběstačnost při propuštění do ambulantní péče je v mediánu 75 bodů, ale pacienti pokračující do následné péče mají medián pouze 40 bodů).

p-hodnota
Mannova-Whitneyova testu



P-hodnota statistické významnosti tohoto pozorovaného rozdílu je $p < 0,001$, což na hladině významnosti 0,05 značí **významný rozdíl**, a ze získaných dat tedy lze říct, že **aktuální soběstačnost pacientů souvisí s jejich dalším pokračováním v systému zdravotní péče**.

Úkol 3. Párový Wilcoxonův test


Úkol č. 3 – Párový Wilcoxonův test

Zadání: „Pacientům hospitalizovaným s mozkovým infarktem byla na lůžku akutní péče poskytnuta terapie pro obnovu krevního oběhu v postižené části mozku. Po zvládnutí akutní fáze byl u pacientů vyhodnocen stupeň soběstačnosti v základních denních aktivitách (ADL) pomocí indexu Barthelové (BI) a byli přeloženi na rehabilitační oddělení. Po dvou týdnech byl opět vyhodnocen stupeň soběstačnosti dle BI. Zjistěte, zda poskytnutá rehabilitační péče vedla ke zlepšení soběstačnosti ADL.“



Úkol č. 3 – Párový Wilcoxonův test

Postup (po nemožnosti použít párový t-test):

1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu o diferencích párových hodnot. $H_0: Me = 0$, $H_A: Me \neq 0$
2. Původní hodnoty vypočítaných diferencí obou měření převedeme na pořadí (určené podle jejich absolutní hodnoty).
3. Vypočítáme **testovou statistiku S_w nebo Z** a odpovídající **p-hodnotu**. $S_w = 198,5$ $Z = 17,29 \Rightarrow p < 0,001$
4. Vypočítané statistiky porovnáme s kritickou hodnotou, nebo porovnáme p-hodnotu s hladinou významnosti $\alpha = 0,05$.
5. Je-li p-hodnota $\leq \alpha$  zamítáme H_0 . **Během rehabilitace se podařilo změnit soběstačnost pacientů v denních aktivitách.** Ke stejnému závěru jsme došli při použití parametrického t-testu.

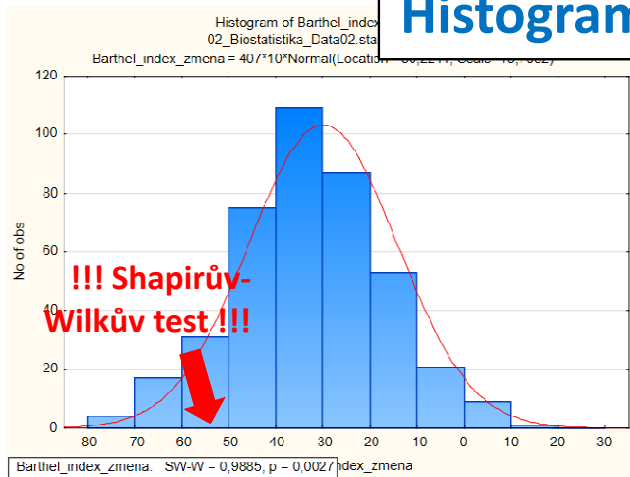
Úkol č. 3 – Ověření normality diferencí

① Průměr a medián jsou v podstatě shodné (cca -30) a data jsou tedy nejspíš alespoň symetrická.

Srovnání průměru a mediánu

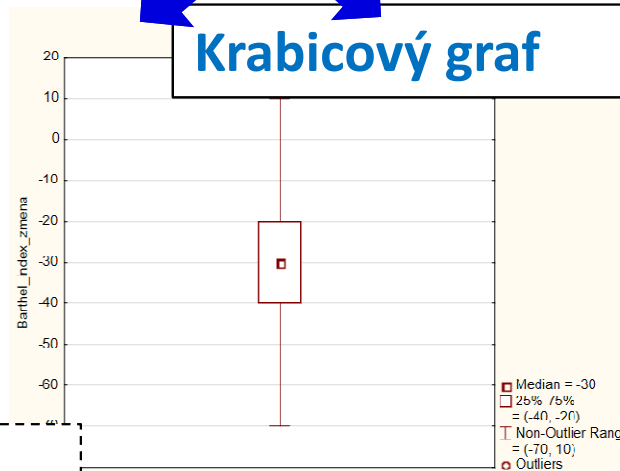
| Variable | Descriptive Statistics (02_Biostatistika) | | |
|---------------------|---|----------|----------|
| | Valid N | Mean | Median |
| Barthel_index_zmena | 407 | -30,2211 | -30,0000 |

Histogram



Změna BI

Krabicový graf



Diagnostický N-P graf



② Symetrie je patrná i z krabicového grafu. Navíc histogram je svým průběhem velmi podobný normálnímu rozdělení. Z N-P grafu také nejsou patrné odchylky od normality.

③ Na základě p-hodnoty 0,003 zamítáme nulovou hypotézu o normalitě (tj. zamítáme, že není rozdíl mezi pozorovanými daty a teoretickým normálním rozdělením, ... tj. data formálně dle testu nejsou normálně rozdělená).

Úkol č. 3 – Řešení v programu Statistica

- V menu **Statistics** zvolíme **Nonparametrics**, vybereme **Comparing two dependent samples (groups)**.

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Nonparametrics' option is selected. A red arrow labeled '1' points to the 'Statistics' menu, and another red arrow labeled '2' points to the 'Nonparametrics' option. A third red arrow labeled '3' points to the 'Comparing two dependent samples (variables)' option in the 'Nonparametric Statistics' dialog box. The dialog box is titled 'Nonparametric Statistics: 02_Biostatistika_Data02...' and has a 'Quick' tab selected. The 'Quick' tab contains a list of options, with 'Comparing two dependent samples (variables)' highlighted. The background shows a data table with columns for 'Rehabilitace po mozkovem infarktu: data' and rows for 'Barthel_index_po_rehal', 'Kategorie_zavislosti_p', 'Barthel', and 'Barthel_index_refe'.

| | 21 | 22 | 23 | 24 |
|------------------------|----|------------------------|-----|------|
| Barthel_index_po_rehal | | | | |
| Kategorie_zavislosti_p | | | | |
| Barthel | | | | |
| Barthel_index_refe | | | | |
| 1 | | 85 částečně soběstačný | -60 | 64,4 |
| 2 | | 40 zcela závislý | 20 | 64,4 |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |

Úkol č. 3 – Řešení v programu Statistica

- Vybereme proměnné (*Variables*), které chceme testovat.
- Kliknutím na *Wilcoxon matched pair test* získáme výsledky.

The screenshot shows the Statistica software interface. The main window displays a data table titled "Rehabilitace po mozkovem infarktu: data". The table has three columns: "Barthel_index_pred_rehabilitaci", "Kategorie_zavislosti_po_rehabilitaci", and "Barthel". The data rows are numbered 1 to 16. A dialog box titled "Comparing two variables: 02_Biostatistik..." is open over the table. The dialog box has a "Variables" section with "List 1: Barthel_index_pred_rehabilitaci" and "List 2: Barthel_index_po_rehabilitaci". The "Quick" section has three options: "Sign test", "Wilcoxon matched pairs test", and "Box & whisker plots for all variables". The "Options" section has a "By Group" button and a "p-value for highlighting:" field set to ".05". Two red arrows point to the "Variables" section (labeled "1") and the "Wilcoxon matched pairs test" option (labeled "2").

| | 21 | 22 | 23 |
|----|-------------------------------|--------------------------------------|---------|
| | Barthel_index_po_rehabilitaci | Kategorie_zavislosti_po_rehabilitaci | Barthel |
| 1 | | 85 částečně soběstačný | |
| 2 | | 40 vvsoco závislý | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |
| 16 | | 70 částečně soběstačný | |

Úkol č. 3 – Výsledky v Statistica

Rozsah
výběru

Hodnota testové
statistiky S_w a Z

Wilcoxon Matched Pairs Test (02_Biostatistika)
Marked tests are significant at $p < ,05000$

| Pair of Variables | Valid N | T | Z | p-value |
|---|------------|----------|----------|---------|
| Barthel_index_pred_rehabilitaci & Barthel_index_po_rehabilitaci | 402 | 198,5000 | 17,28938 | 0,0000 |



p-hodnota
Wilcoxonova testu

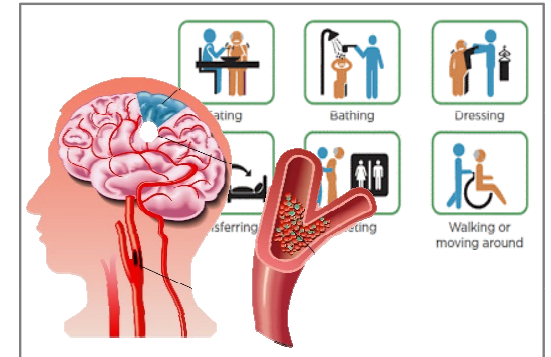
① Pozorovaný medián
zlepšení Barthelové
indexu na začátku a po
rehabilitaci je 30 bodů.

② P-hodnota statistické významnosti této pozorované změny je $p < 0,001$, což na hladině významnosti 0,05 značí **významný rozdíl**, a lze tedy prohlásit, že **stupeň soběstačnosti v základních denních aktivitách se viditelně během péče zlepšil**.

Úkol 4. Kruskalův-Wallisův test

Úkol č. 4 – Kruskalův-Wallisův test

Zadání: „Zjistěte, zda etiologie vzniku mozko-
vého infarktu (deficit způsobený embolií,
trombózou nebo neurčenou okluzí/stenózou)
je potenciálním prediktivním faktorem
výsledného stupně soběstačnosti v základních
denních aktivitách (ADL) vyjádřeného indexem Barthelové. Tj.,
liší se pacienti s různým typem vzniku mozkového infarktu ve
výsledné soběstačnosti?“



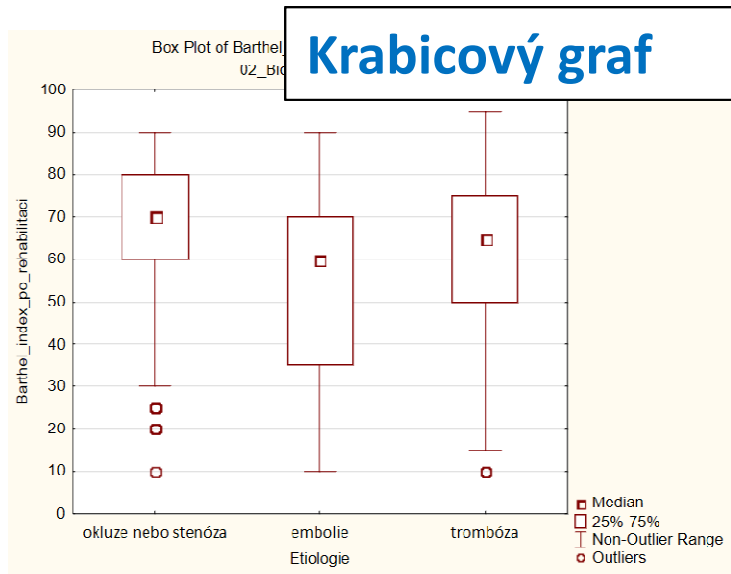
Úkol č. 4 – Kruskalův-Wallisův test

Postup (po nemožnosti použít ANOVA test):

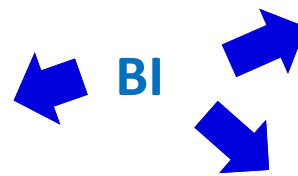
1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu $H_0: F(x_1) = F(x_2) = F(x_3)$ proti H_A : **alespoň jedna dvojice $F(x_k)$ se liší.**
2. Původní hodnoty Barthelové indexu převedeme na pořadí v celém souboru.
3. Vypočítáme **testovou statistiku Q** a odpovídající **p-hodnotu.**
$$Q = 23,63 \Rightarrow p < 0,001$$
4. Testovou statistiku porovnáme s kritickou hodnotou nebo porovnáme p-hodnotu s hladinou významnosti $\alpha = 0,05$.
5. Je-li **p-hodnota $\leq \alpha$** **➡ zamítáme H_0 .** **Existuje alespoň jedna dvojice způsobu vzniku mozkového infarktu, která se liší v následné soběstačnosti pacientů.**

Úkol č. 4 – Ověření normality a popis dat

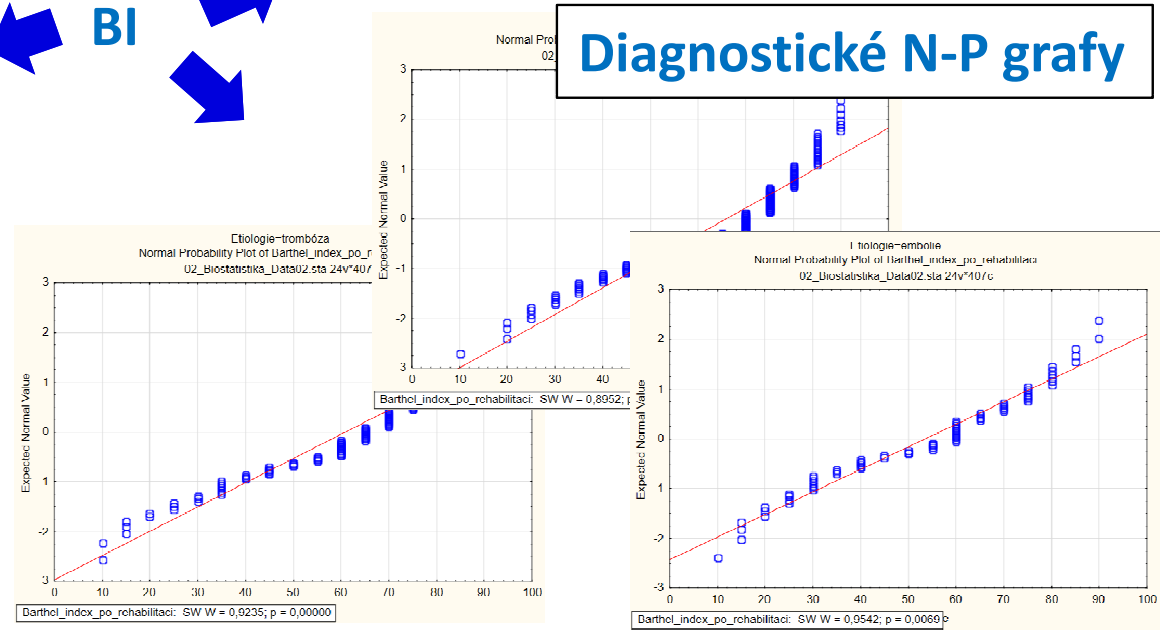
Srovnání průměru a mediánu



| Variable | Descriptive Statistics (02_Biostatistika_Data02.sta) | | | |
|-------------------------------|--|---------|----------|----------|
| | Etiologie | Valid N | Mean | Median |
| Barthel_index_po_rehabilitaci | trombóza | 128 | 60,89844 | 65,00000 |
| Barthel_index_po_rehabilitaci | okluze nebo stenóza | 201 | 65,99502 | 70,00000 |
| Barthel_index_po_rehabilitaci | embolie | 78 | 53,58974 | 60,00000 |



Diagnostické N-P grafy



① Základní popis i grafické srovnání ukazuje možný rozdíl mezi skupinami (soběstačnost po embolii je v mediánu 60 bodů, po trombóze 65 bodů a po neurčené okluzi nebo stenóze 70 bodů).

② Normalitu dat zamítáme u všech tří skupin ($p < 0,001$, $p < 0,001$ a $p = 0,007$) s tím, že u všech je porušení normality patrné graficky i z N-P grafu.

Úkol č. 4 – Řešení v programu Statistica

- V menu **Statistics** zvolíme
zvolíme
Nonparametrics,
vybereme **Comparing
multiple indep. samples
(groups)**.

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Nonparametrics' option is selected. The 'Comparing multiple indep. samples (groups)' option is highlighted in the list. Red arrows labeled 1, 2, and 3 indicate the steps: 1 points to the 'Statistics' menu, 2 points to the 'Nonparametrics' option, and 3 points to the 'Comparing multiple indep. samples (groups)' option. The background shows a data table with columns 'Barthel_index_po_rehabilitaci' and 'Kategorie_zavislosti_po_rehabilitaci'.

| | Rehabilitace po mozkovem infarktu: data | | |
|----|---|--------------------------------------|-----|
| | 21 | 22 | 23 |
| | Barthel_index_po_rehabilitaci | Kategorie_zavislosti_po_rehabilitaci | Bar |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |

Úkol č. 4 – Řešení v programu Statistica

- Vybereme proměnnou, kterou chceme testovat (*dependent*) a proměnnou obsahující skupiny, které srovnáváme (*grouping*).
- Kliknutím na **Multiple comparisons of mean ranks for all groups** získáme výstupy (celkové srovnání ale také mnohonásobné porovnání mezi všemi skupinami).

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Kruskal-Wallis ANOVA and Median Test' dialog box is displayed. The dialog box has the following settings:

- Dependent variables: Barthel_index_po_rehabilitaci
- Grouping variable: Etiologie
- Codes: none
- Summary: Kruskal-Wallis ANOVA & Median test
- Multiple comparisons of mean ranks for all groups (selected)
- Box & whisker
- Categorized histogram
- p-value for highlighting: ,05

Red arrows labeled '1' and '2' indicate the steps described in the text: clicking on 'Variables' and selecting 'Multiple comparisons of mean ranks for all groups'.

Úkol č. 4 – Výsledky v Statistica

| Depend.: | okluze nebo stenóza | embolie | trombóza |
|-------------------------------|---------------------|----------|----------|
| Barthel_index_po_rehabilitaci | R:228,81 | R:154,65 | R:195,11 |
| okluze nebo stenóza | | 0,000007 | 0,033834 |
| embolie | 0,000007 | | 0,050002 |
| trombóza | 0,033834 | 0,050002 | |

Multiple Comparisons p values (2-tailed); Barthel_index_po_rehabilitaci
Independent (grouping) variable: Etiologie
Kruskal-Wallis test: H (2, N= 407) =23,63471 p =,0000

Souhrnná p-hodnota
Kruskalova-Wallisova
testu



p-hodnoty mnohonásobného
porovnání všech skupin

① Z předchozího popisu je patrný možný rozdíl mezi skupinami (soběstačnost po embolii je v mediánu 60 bodů, po trombóze 65 bodů a po neurčené okluzi nebo stenóze 70 bodů).

② Souhrnná p-hodnota statistické významnosti tohoto pozorovaného rozdílu je $p < 0,001$, což na hladině významnosti 0,05 značí **významný rozdíl** a ze získaných dat tedy lze říct, že **existuje alespoň jedna dvojice způsobu vzniku mozkového infarktu, která se liší v následné soběstačnosti pacientů (tj. etiologie souvisí s další soběstačností).**

③ Mnohonásobným porovnáním jsme navíc prokázali významný rozdíl mezi embolií a okluzí/stenózou a mezi trombózou a okluzí/stenózou (rozdíl mezi embolií a trombózou významný není). Jinými slovy, **výsledný stupeň soběstačnosti je významně lepší u pacientů s okluzí/stenózou oproti embolii i trombóze.**