MUNI
MED

MIAM021p(s) **Analýza a management dat pro zdravotnické obory – přednáška a cvičení** (jaro 2023)

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)

MUNI MED

Organizace výuky

- 21. 2. – Excel: opakování, příprava dat, základní vzorce
- 28. 2. – Základy popisné statistiky
- 7. 3. – Základní rozdělení pravděpodobnosti, testování hypotéz
- 14. 3. – Parametrické testy
- 21. 3. – Neparametrické testy
- 28. 3.
- 4. 4. – Analýza kontingenčních tabulek, testy dobré shody
- 11. 4. – Základy korelační analýzy + opakování vybraných témat
- 18. 4. – Ukončení předmětu, test

MUNI MED

MUNI
MED

Základy korelační analýzy

Korelace
Pearsonův korelační koeficient
Spearmanův korelační koeficient

Proč hodnotit vztah dvou spojitých proměnných?

Vztah mezi dvěma spojitými veličinami zjišťujeme, když:

- chceme zjistit, jestli mezi nimi **existuje vztah** – např. jestli vyšší hodnoty jedné veličiny znamenají nižší hodnoty jiné veličiny;
- chceme **predikovat hodnoty** jedné veličiny na základě znalosti hodnot jiné veličiny;
- chceme **kvantifikovat vztah** mezi dvěma spojitými veličinami; např. pro použití jedné veličiny na místo druhé veličiny.

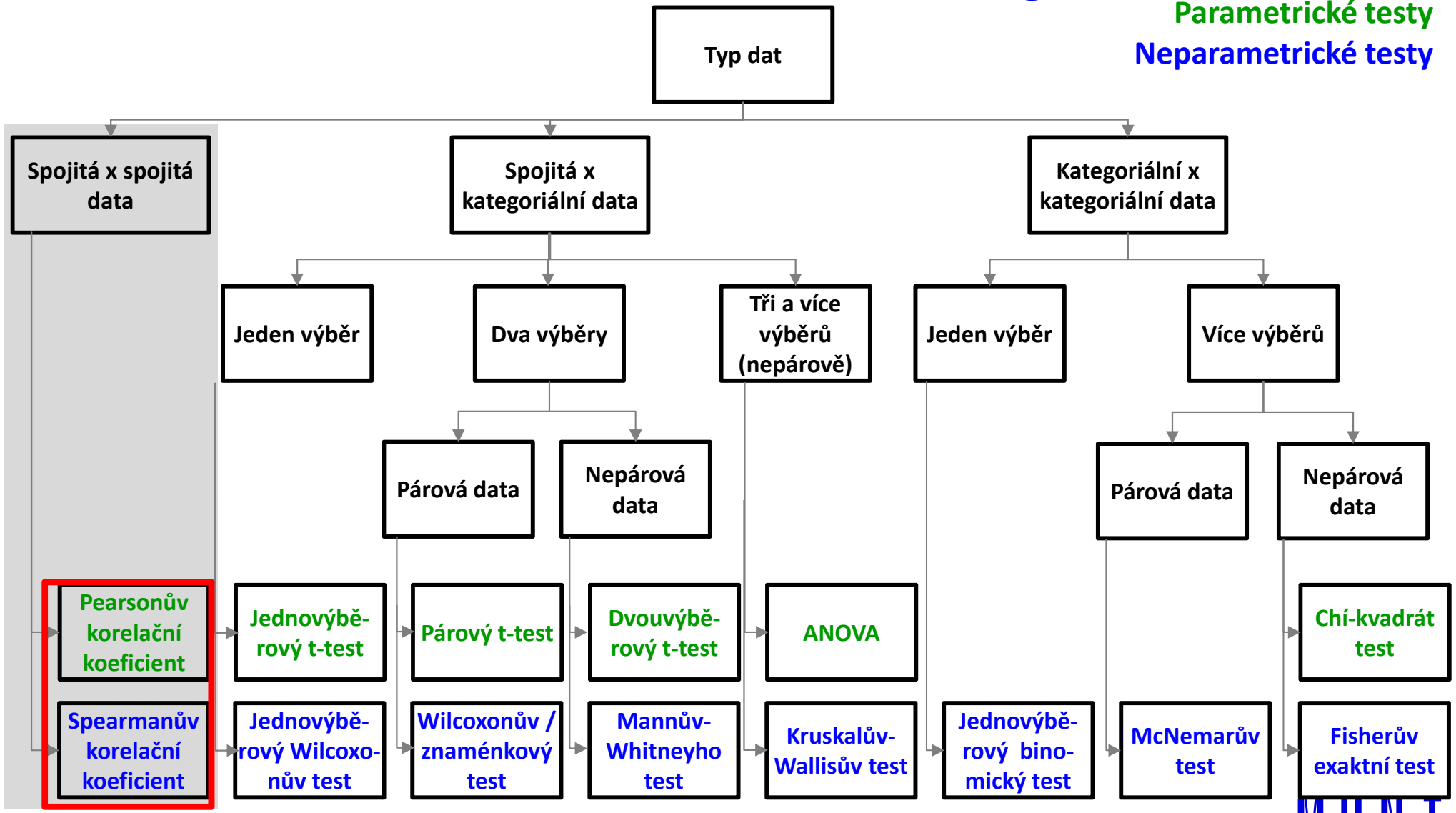
Korelační a regresní analýza

- **Korelační analýza** je využívána pro vyhodnocení míry **vztahu** dvou spojitých proměnných. Obdobně jako jiné statistické metody, i korelace mohou být **parametrické** nebo **neparametrické**.

- **Regresní analýza** vytváří **model vztahu** dvou nebo více proměnných, tedy jakým způsobem jedna proměnná (vysvětlovaná) závisí na jiných proměnných (prediktorech). Regresní analýza je obdobně jako ANOVA nástrojem pro vysvětlení variability hodnocené proměnné.

Základní statistické testy

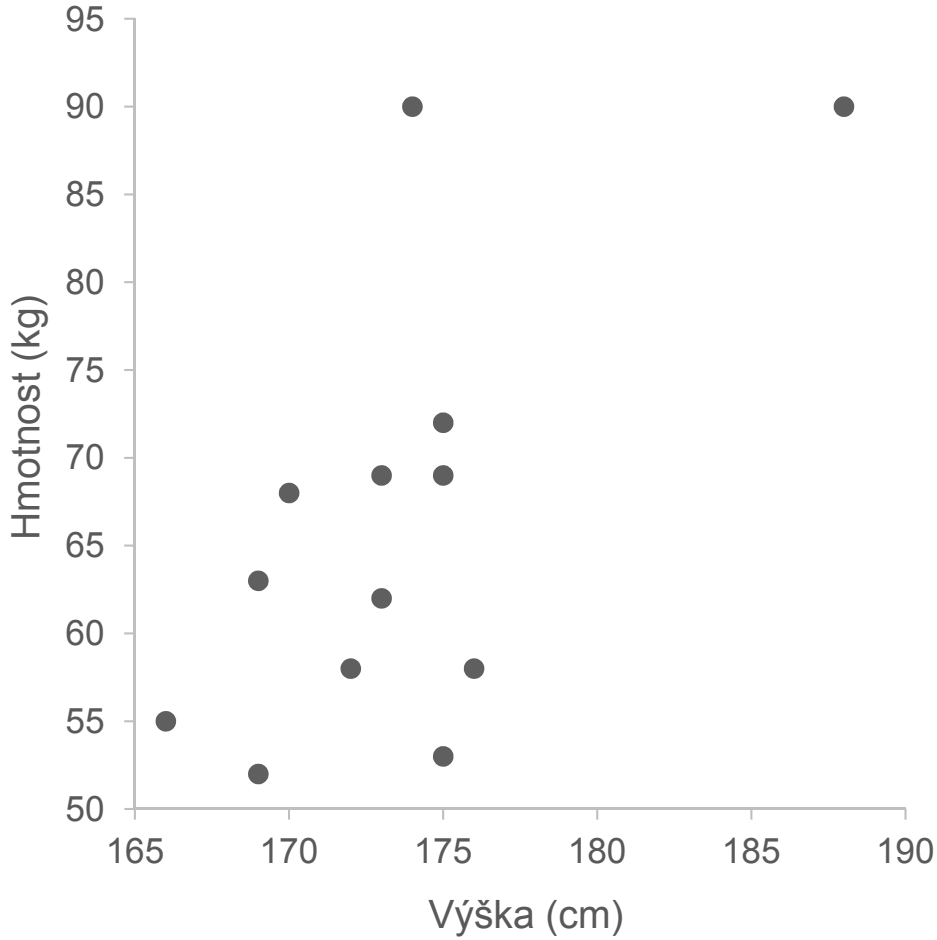
Parametrické testy
Neparametrické testy



MUNI MED

Bodový graf – vizualizace vztahu dvou spojitých proměnných

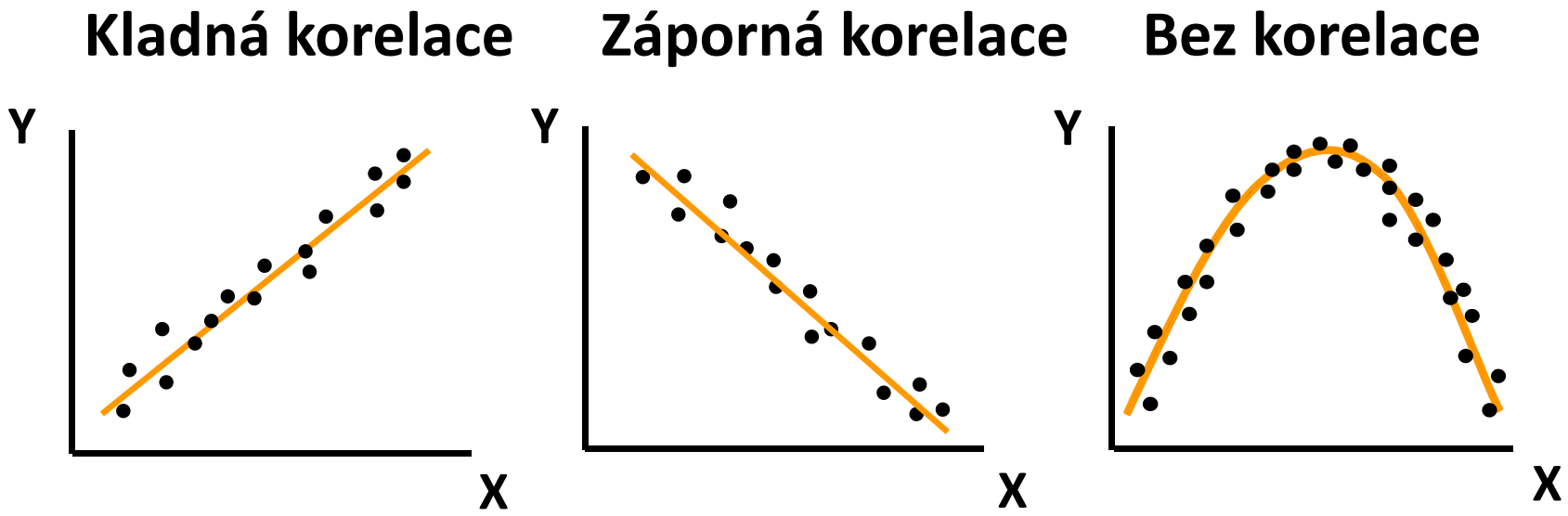
- Nejjednodušší formou je **bodový graf** (XY graf), tzv. *scatterplot*.
- Vztah výšky a hmotnosti studentů Biostatistiky (jaro 2010).



MUNI MED

Korelace

Korelace = vztah (závislost) dvou znaků (parametrů)



MUNI MED

Korelační koeficienty

- **Korelační koeficient** (*r*) – kvantifikuje míru vztahu mezi dvěma spojitými veličinami *X* a *Y*.
- **Pearsonův korelační koeficient** je parametrický; hodnotí míru lineární závislosti mezi dvěma spojitými proměnnými.
  Předpoklad: proměnné pocházejí z tzv. dvourozměrného **normálního rozdělení** (pro každou hodnotu X má proměnná Y normální rozdělení a pro každou hodnotu Y má proměnná X normální rozdělení)
- **Spearmanův korelační koeficient** je neparametrický; hodnotí míru závislosti pořadí hodnot dvou spojitých proměnných.
- Hodnota ***r*** je **kladná**, když vyšší hodnoty *X* souvisí s vyššími hodnotami *Y*. Naopak hodnota ***r*** je **záporná**, když nižší hodnoty *X* souvisí s vyššími hodnotami *Y*.

MUNI MED

Statistická významnost korelačního koeficientu

- Korelační koeficient nabývá hodnot od -1 do 1

  r = 0 → nekorelované veličiny

  r > 0 → kladně korelované veličiny

  r < 0 → záporně korelované veličiny

- Testujeme hypotézu o nezávislosti spojitých proměnných:

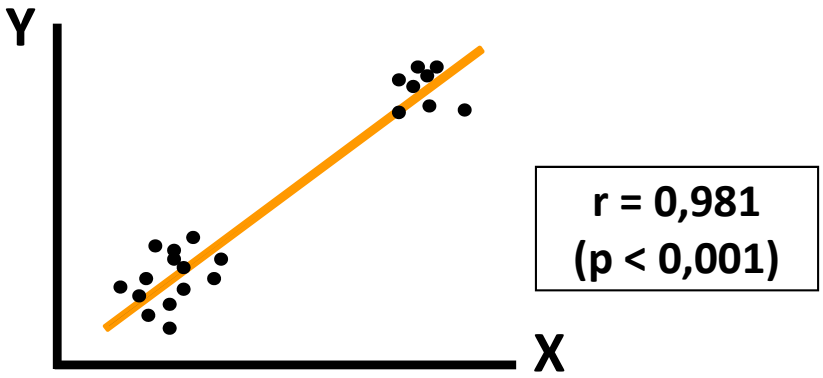
  H_0: proměnné X a Y jsou nezávislé náhodné veličiny; $r = 0$

  H_A: proměnné X a Y nejsou nezávislé náhodné veličiny; $r \neq 0$

- Testování pomocí intervalu spolehlivosti nebo výpočet testové statistiky a srovnání s kritickou hodnotou nebo výpočet p-hodnoty.

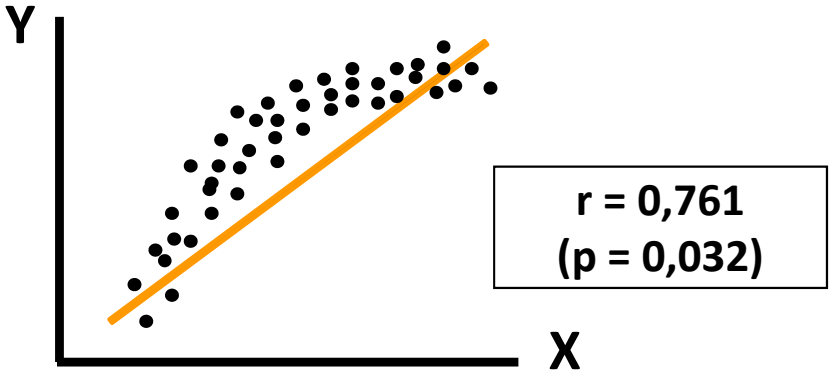
Možné problémy s výpočtem *r*

Problém více skupin



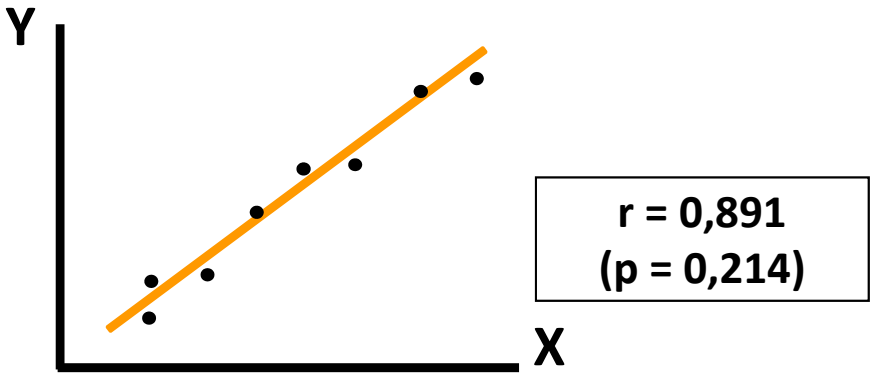
r = 0,981
(p < 0,001)

Nelineární vztah

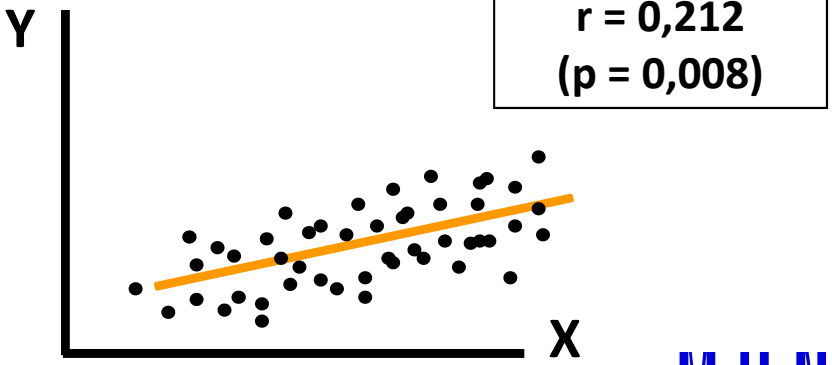


r = 0,761
(p = 0,032)

Problém velikosti výběru



r = 0,891
(p = 0,214)



r = 0,212
(p = 0,008)

MUNI MED

MUNI
MED

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 ID	2 Pohlavi	3 Vek	4 Etiologie	5 Lokalizace	6 Terapie	7 Komorbid	8 Barthel_inc	9 Kategorie_zavislosti_p	10 Ukoncen
1	1	muž	82	okluze neb	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 neb	mozkové tepny	jiná farmakolog	0	35	vysoce závislý	propuště
4	4	žena	46	embolie	mozkové tepny	intravenózní tro	0	20	vysoce závislý	propuště
5	5	muž	76	okluze neb	mozkové tepny	jiná farmakolog	0	45	částečně soběstačný	propuště
6	6	muž	72	okluze neb	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 neb	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 neb	mozkové tepny	jiná farmakolog	0	40	vysoce závislý	propuště
12	12	žena	92	okluze neb	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

MUNI MED

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**.

MUNI
MED

Pearsonův korelační koeficient

Úkol č. 1 – Pearsonův korelační koeficient

Zadání: „ U pacientů hospitalizovaných s mozkovým infarktem bylo při propuštění vyhodnoceno zlepšení míry soběstačnosti vyjádřené diferencí hodnot indexu Barthelové. Zjistěte, zda má věk vliv na úspěšnost terapeutické a rehabilitační péče. Jinými slovy, určete, zda věk koreluje s diferencí indexu Barthelové.“

Postup:

1. Ověříme předpoklady použití Pearsonova korelačního koeficientu (normalita rozložení věku a diferencí BI).

MUNI MED

Úkol č. 1 – Pearsonův korelační koef.

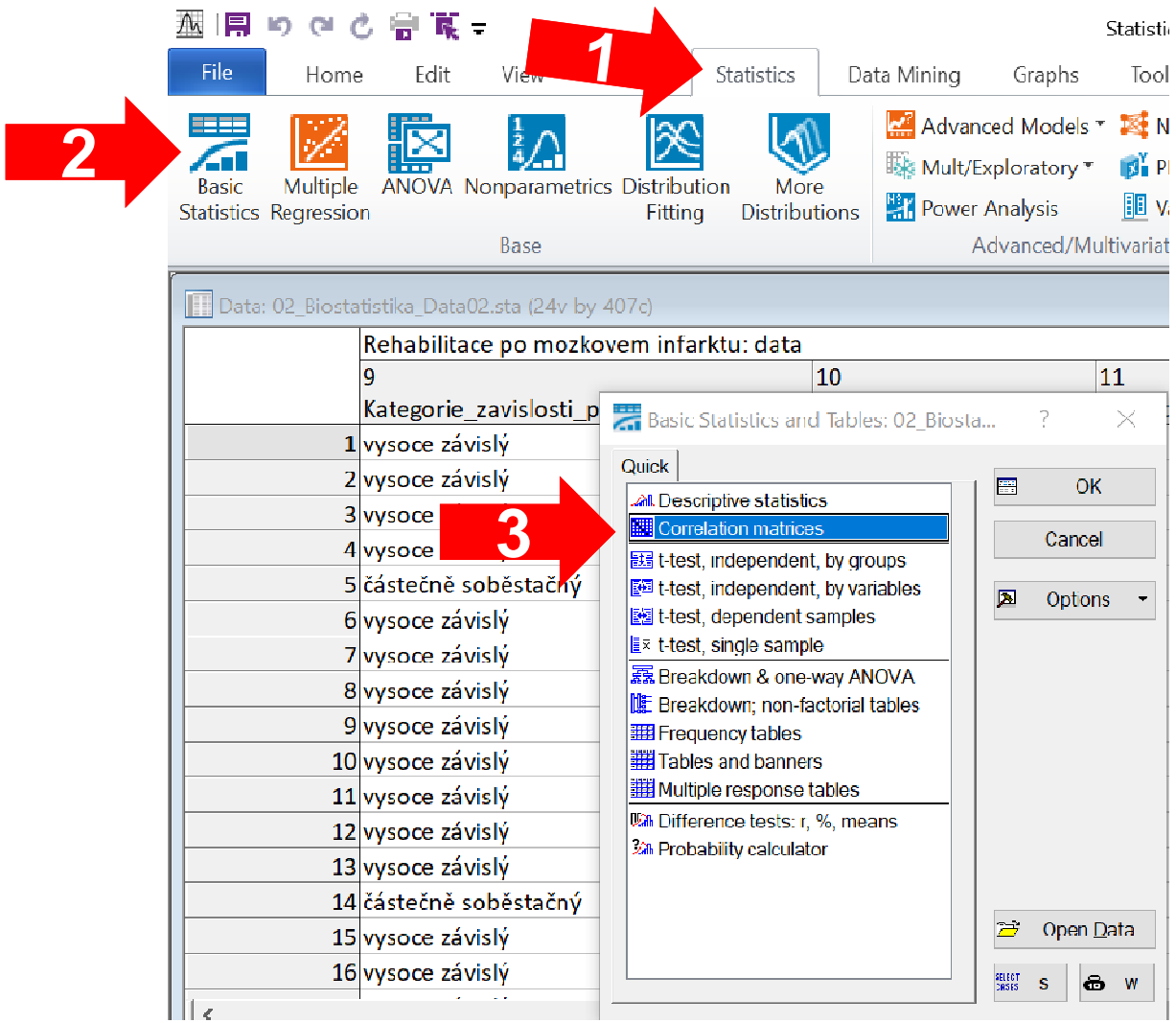
Postup (po ověření předpokladů):

1. Na hladině významnosti α = 0,05 testujeme hypotézu **H_0**: $r = 0$ proti **H_A**: $r \neq 0$
2. Graficky znázorníme závislost obou proměnných pomocí bodového XY grafu.
3. Vypočítáme hodnotu **korelačního koeficientu r** a odpovídající **p-hodnotu**: $r = 0{,}099 \Rightarrow p = 0{,}046$
4. Porovnáme p-hodnotu s hladinou významnosti α = 0,05.
5. Je-li **p-hodnota ≤ α** ➡ **zamítáme H_0**. **Věk pacienta má vliv na zlepšení míry soběstačnosti po léčbě mozkového infarktu. Pozitivní korelace značí, že u starších pacientů je zlepšení menší** (diference jsou vypočítány tak, že nižší hodnoty odpovídají většímu zlepšení).

MUNI MED

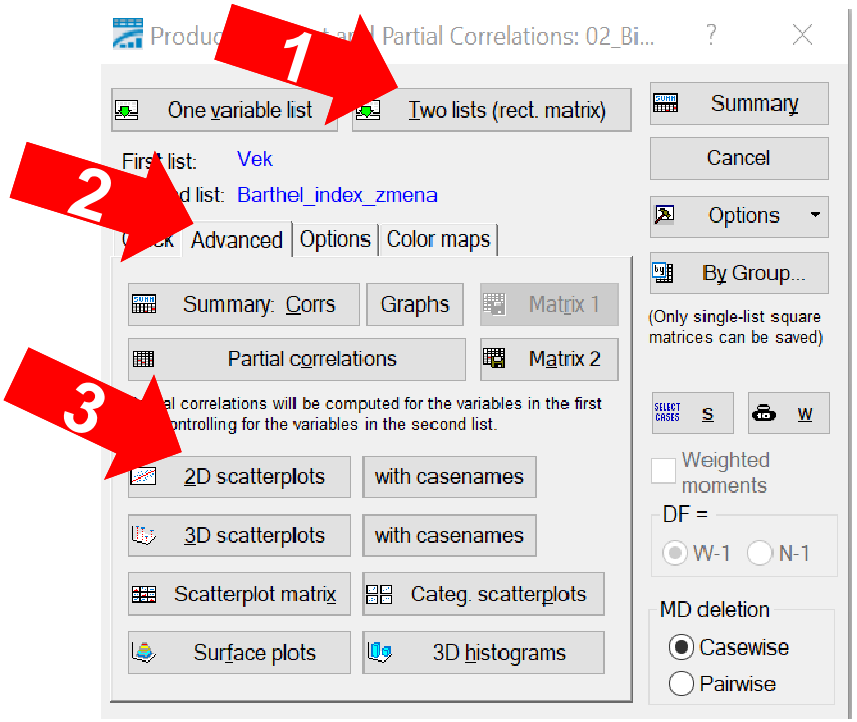
Úkol č. 1 – Řešení v programu Statistica

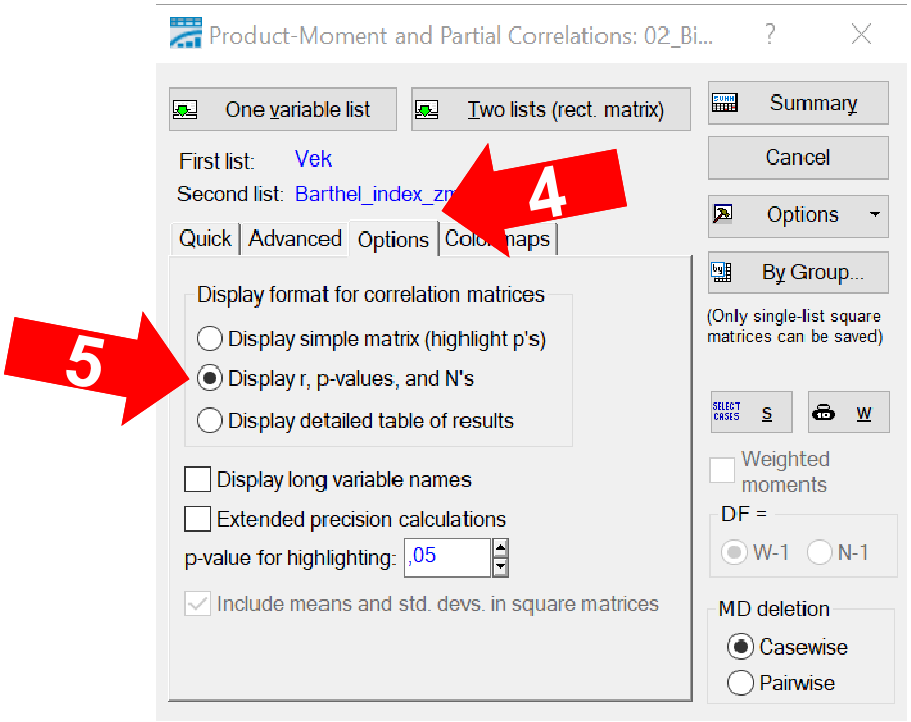
- V menu ***Statistics*** zvolíme ***Basic statistics***, vybereme ***Correlation matrices***.



MUNI MED

Úkol č. 1 – Řešení v programu Statistica





- Vybereme obě proměnné, které chceme testovat (***Two lists***)
- V záložce ***Advanced*** kliknutím na ***2D scatterplots*** získáme grafické znázornění závislosti vybraných proměnných.

- Poté v záložce ***Options*** zvolíme možnost ***Display r, p-values, and N´s*** a přes ***Summary*** zobrazíme výsledky.

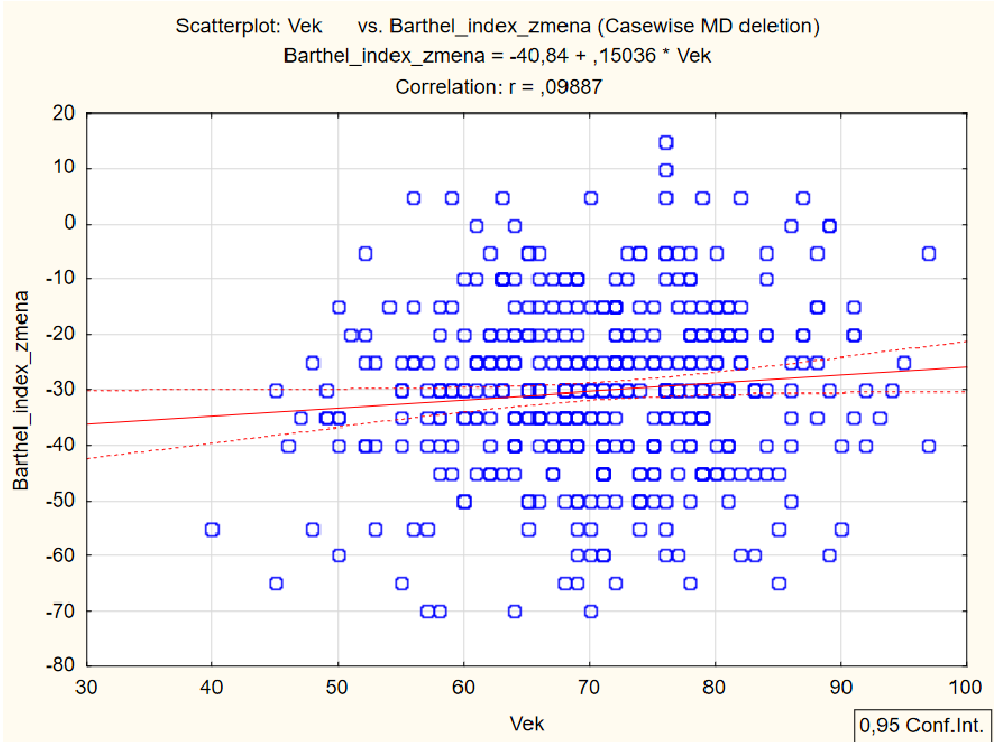
MUNI MED

Úkol č. 1 – Výsledky v Statistica

① Z grafu sice není nikterak výrazná závislost přímo patrná, nicméně je možné, že je přítomen mírně pozitivní trend.

	Correlations (02_Biostatist Marked correlations are si N=407 (Casewise deletion
Variable	**Barthel_index_zmena**
Vek	,0989 p=,046

Korelační koeficient a p-hodnota



P-hodnota statistické významnosti korelace je p = 0,046, což na hladině významnosti 0,05 značí **významný výsledek** a ze získaných dat jsme tedy **prokázali, že věk pacienta má vliv na zlepšení míry soběstačnosti po léčbě mozkového infarktu.** Přesto je potřeba výsledek interpretovat s opatrností, neboť samotná korelace je velmi slabá (0,099).

MUNI MED

MUNI
MED

Spearmanův korelační koeficient

Úkol č. 2 – Spearmanův korelační koeficient

Zadání: „ U pacientů hospitalizovaných s mozkovým infarktem bylo při propuštění vyhodnoceno zlepšení míry soběstačnosti vyjádřené diferencí hodnot indexu Barthelové. Zjistěte, zda má věk vliv na úspěšnost terapeutické a rehabilitační péče. Jinými slovy, určete, zda věk koreluje s diferencí indexu Barthelové.“

MUNI MED

Úkol č. 2 – Spearmanův korelační koef.

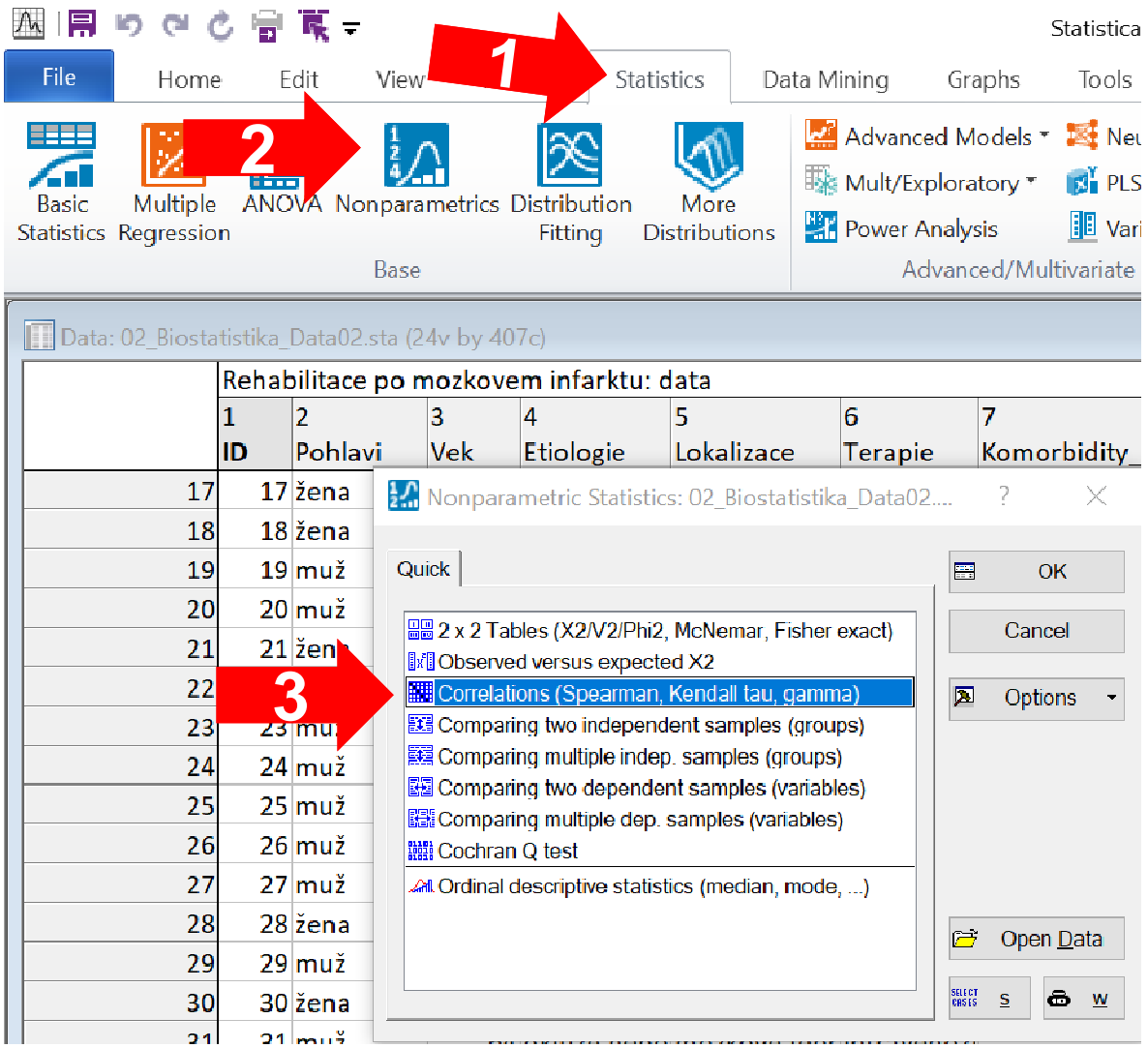
Postup (po nemožnosti použít Pearsonův korelační koeficient):

1. Na hladině významnosti α = 0,05 testujeme hypotézu **H_0**: $r = 0$ proti **H_A**: $r \neq 0$
2. Graficky znázorníme závislost obou proměnných pomocí bodového XY grafu.
3. Vypočítáme hodnotu **korelačního koeficientu r_s** a odpovídající **p-hodnotu**: $r_s = 0{,}074 \Rightarrow p = 0{,}136$
4. Porovnáme p-hodnotu s hladinou významnosti α = 0,05.
5. Je-li **p-hodnota > α** ➡ **nezamítáme H_0**. Neprokázali jsme, že by **věk pacienta měl vliv na zlepšení míry soběstačnosti po léčbě mozkového infarktu.**

MUNI MED

Úkol č. 2 – Řešení v programu Statistica

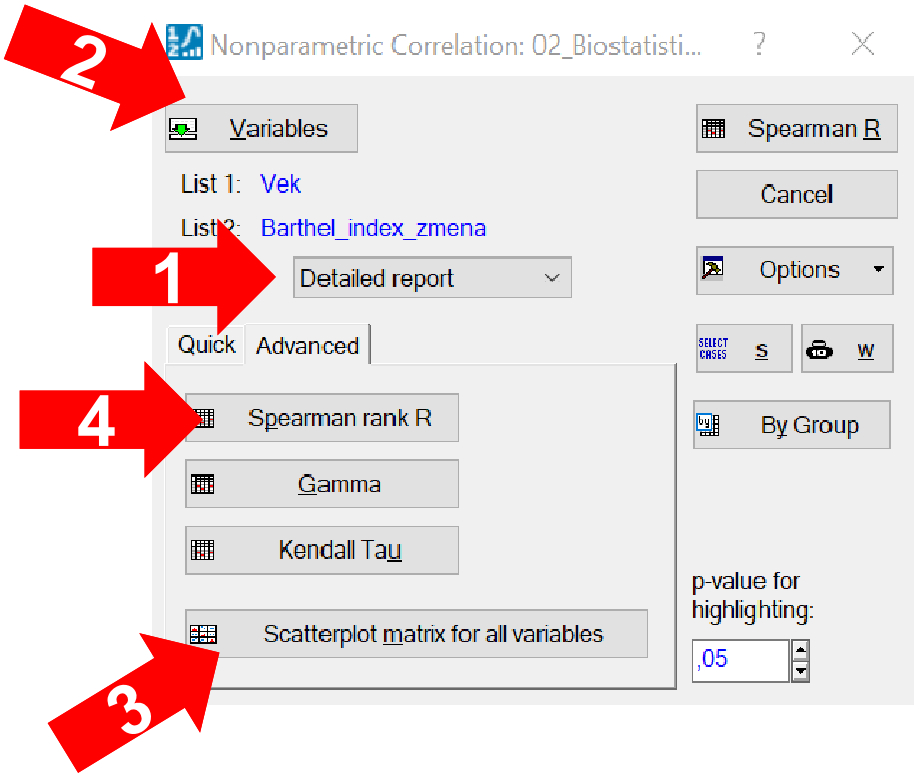
- V menu ***Statistics*** zvolíme ***Nonparametrics***, vybereme ***Correlation (Spearman, …)***.



MUNI MED

Úkol č. 2 – Řešení v programu Statistica

- V možnostech ***Compute:*** vybereme ***Detailed report***.
- Vybereme jednotlivé proměnné, které chceme testovat (***Variables***).
- V záložce ***Advanced*** kliknutím na ***Scatterplot matrix*** získáme grafické znázornění závislosti vybraných proměnných.
- Poté přes ***Spearman rank R*** zobrazíme výsledky.



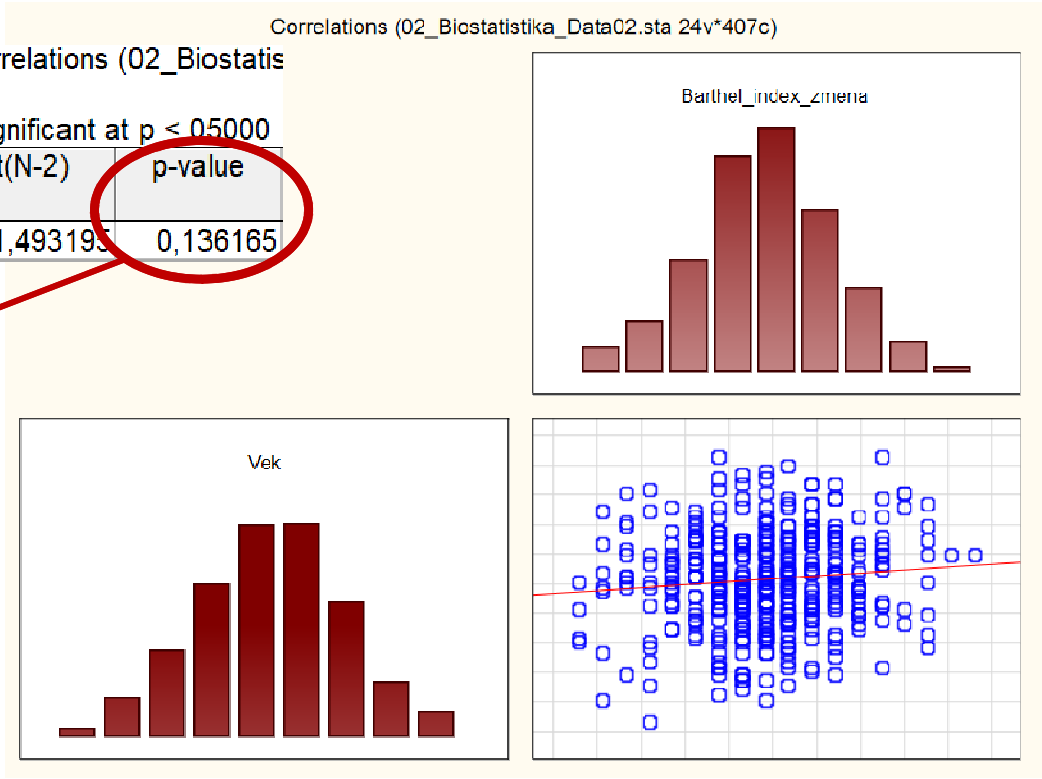
MUNI MED

Úkol č. 2 – Výsledky v Statistica

Spearman Rank Order Correlations (02_Biostatis
MD pairwise deleted
Marked correlations are significant at p < 05000

Pair of Variables	Valid N	Spearman R	t(N-2)	p-value
Vek & Barthel_index_zmena	407	0,073994	1,493193	0,136165

Korelační koeficient a p-hodnota



① Z grafu není nikterak výrazná závislost patrná, nicméně je možné, že je přítomen mírně pozitivní trend.

② P-hodnota statistické významnosti korelace je p = 0,136, což na hladině významnosti 0,05 značí **nevýznamný výsledek** a ze získaných dat jsme tedy **neprokázali, že by věk pacienta měl vliv na zlepšení míry soběstačnosti po léčbě mozkového infarktu.**

MUNI MED