

MUNI
MED

MNBS081 Biostatistika **(jaro 2022)**

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: 14:00–15:40, počítačová učebna F01B1/709
- Materiály v IS
- Software: Microsoft Office - Excel, Statistica
- Pro získání zápočtu/kolokvia je třeba:

1) Účast – povoleny jsou 2 absence

- Při větší absenci – splnění písemky na konci semestru (teoretická část + řešení příkladů na počítači)

2) Domácí úkoly – povoleno max 1 neodevzdání

- za účelem procvičení, dostanete zpětnou vazbu, na dalším cvičení se vrátíme, kdyby byl problém

3) Závěrečný úkol – datový soubor – praktické úkoly

Organizace výuky

- 15. 2. – Excel: opakování, příprava dat, základní vzorce
- 1. 3. – Základy popisné statistiky (**pozor! 22. 2. setkání nebude**)
- 15. 3. – Základní rozdělení pravděpodobnosti, testování hypotéz (**pozor! 8. 3. setkání nebude**)
- 22. 3. – Parametrické testy
- 29. 3. – Neparametrické testy
- 5. 4. – Analýza kontingenčních tabulek, testy dobré shody
- 12. 4. – Korelační analýza + základy lineární regrese
- 19. 4. – Volitelné sezení (návrat k některým tématům)
- 3. 5. – Ukončení předmětu, test
- 10. 5. – Vyhodnocení testu

Analýza kontingenčních tabulek

Kontingenční tabulky

Pearsonův chí-kvadrát test (test dobré shody)

Fisherův exaktní test

McNemarův test

Kontingenční tabulka

- Sumarizuje **vztah** dvou **kategoriálních proměnných**.
- Řádky (r) jsou tvořeny hodnotami (kategoriemi) prvního znaku, sloupce (c) hodnotami druhého znaku.
- V příslušné buňce tabulky je uveden počet případů s hodnotou prvního znaku odpovídající příslušnému řádku a druhého znaku s hodnotou odpovídající příslušnému sloupci.

	y_1	...	y_c	
x_1	n_{11}	...	n_{1c}	$n_{1.}$ → Marginální četnost
...
x_r	n_{r1}	...	n_{rc} → Absolutní četnost	$n_{r.}$
	$n_{.1}$ → Marginální četnost	...	$n_{.c}$	N → Celkový počet

Ukázka kontingenční tabulky

Vztah pohlaví a výskytu onemocnění

	Nemocný	Zdravý	Celkem
Muž	45	11	56
Žena	25	6	31
Celkem	70	17	87



**Jsou více nemocní
muži nebo ženy?**

	Nemocný	Zdravý	Celkem
Muž	a	b	a + b
Žena	c	d	c + d
Celkem	a + c	b + d	a + b + c + d

Absolutní četnost

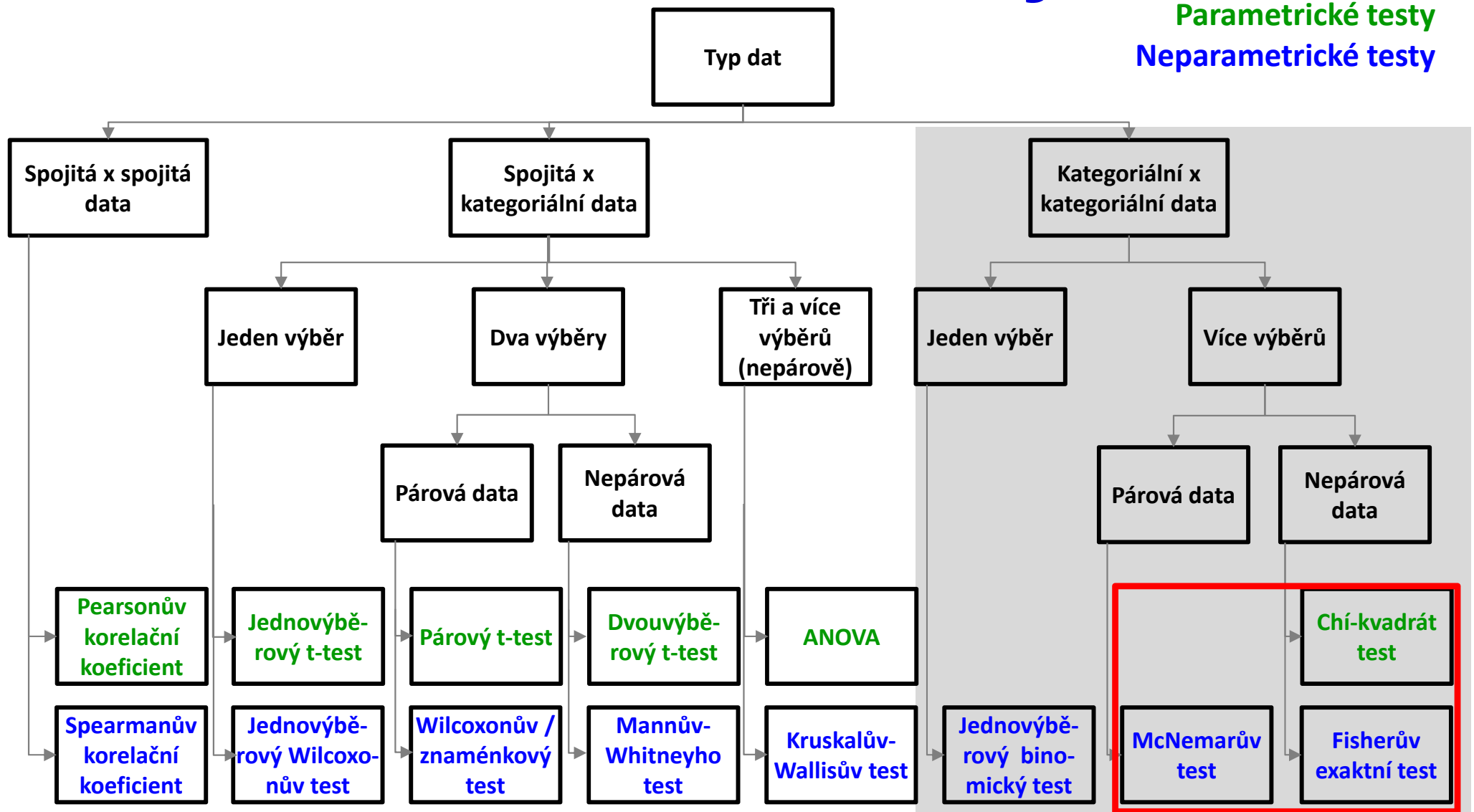
Marginální četnost

Celkový počet

Analýza kontingenčních tabulek

- Analýza kontingenčních tabulek umožňuje analyzovat **vazbu mezi dvěma kategoriálními proměnnými**. Základním způsobem testování je tzv. **chí-kvadrát test**, který **srovnává pozorované četnosti kombinací kategorií oproti očekávaným četnostem**, které vychází z teoretické situace, kdy je vztah mezi proměnnými náhodný.
- Test dobré shody je využíván také pro **srovnání pozorovaných četností proti očekávaným četnostem daným určitým pravidlem**
- Specifickým typem výstupů odvozených z kontingenčních tabulek jsou tzv. **poměry šancí a relativní rizika**, využívaná často v medicíně pro identifikaci rizikových skupin pacientů.

Základní statistické testy



Test dobré shody – princip

- Srovnání pozorovaných četností oproti očekávaným četnostem, které vychází z teoretické situace, kdy je vztah mezi proměnnými náhodný.

- Testová statistika $\chi^2 = \sum \frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}}$

$$\chi^2 = \underbrace{\frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}}}_{\text{1. jev}} + \underbrace{\frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}}}_{\text{2. jev}} + \dots$$

Test dobré shody – příklad


- **Příklad:** 10 000 lidí hází mincí. V 4 000 případech padne rub a v 6 000 případech padne líc. Lze výsledek považovat za statisticky významně odlišný od očekávaného poměru 1 : 1?
- H_0 : Výskyt jevů rub a líc nastává v poměru 1 : 1.
 H_A : Výskyt jevů rub a líc nenastává v poměru 1 : 1.

$$\chi^2 = \sum \frac{\left(\begin{array}{c} \text{pozorovaná} \\ \text{četnost} \end{array} - \begin{array}{c} \text{očekávaná} \\ \text{četnost} \end{array} \right)^2}{\text{očekávaná četnost}}$$

$$\chi^2 = \frac{(4000 - 5000)^2}{5000} + \frac{(6000 - 5000)^2}{5000} = 400$$

Tabulková hodnota:

$$\chi_{(0,95)}^2(1) = 3,84$$

- Vypočítaná hodnota $\chi^2 \geq \chi_{(0,95)}^2(1)$  zamítáme H_0 .

Analýza kontingenčních tabulek

1. Hypotéza o nezávislosti

test: Pearsonův chí-kvadrát test, Fisherův exaktní test

- Jeden výběr, 2 charakteristiky – obdoba nepárového uspořádání
- *Příklad: existence vztahu mezi krevní skupinou a výskytem nemoci*

2. Hypotéza o shodě struktury (tzv. test homogeneity)

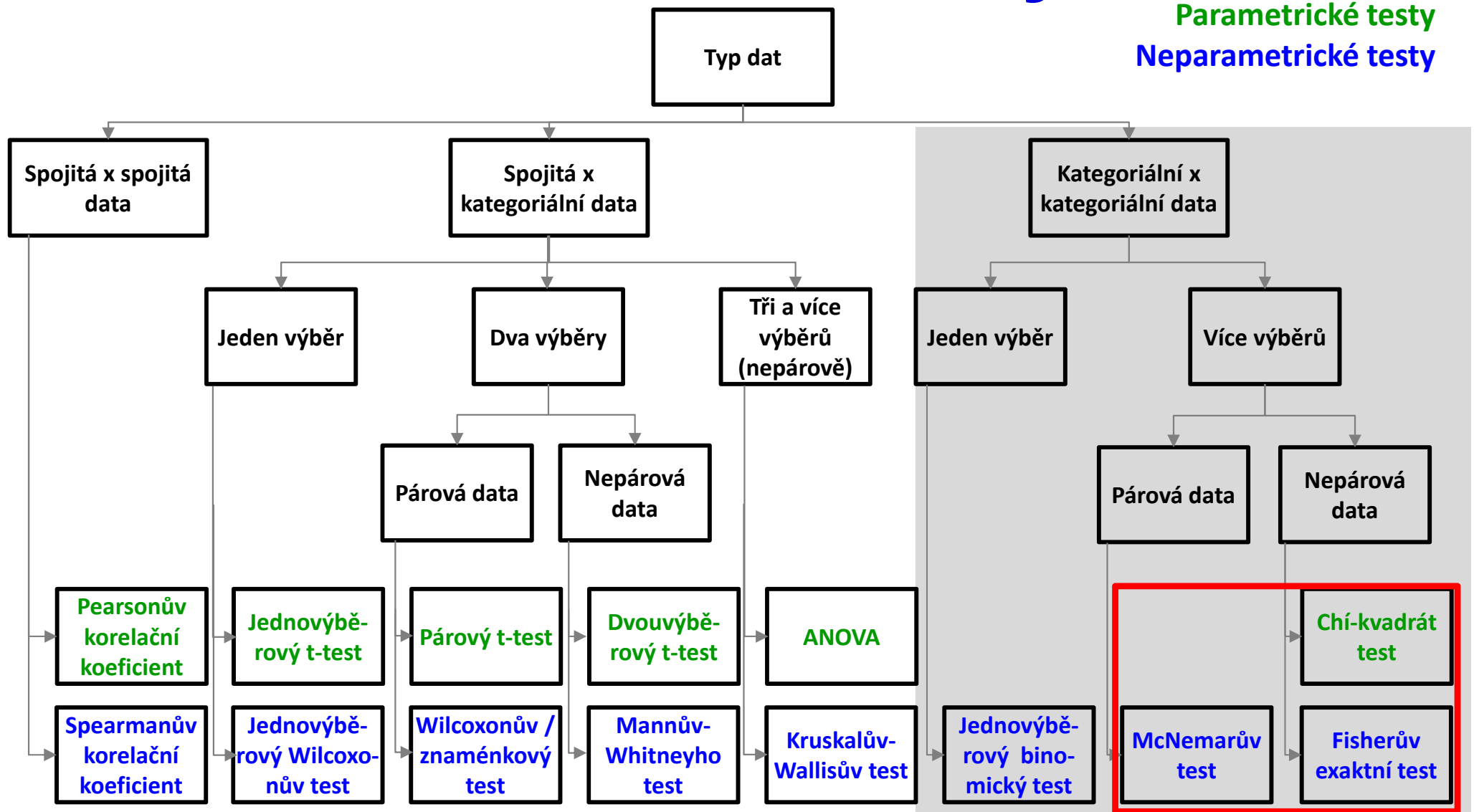
test: Pearsonův chí-kvadrát test, Fisherův exaktní test

- Více výběrů, jedna charakteristika – obdoba nepárového uspořádání
- *Příklad: věková struktura pacientů s diabetem v K nemocnicích*

3. Hypotéza o symetrii – McNemarův test

- Jeden výběr, opakovaně měřena jedna charakteristika – obdoba párového uspořádání
- *Příklad: posouzení výskytu bolesti před a po léčbě*

Základní statistické testy



Testování nezávislosti – Pearsonův chí-kvadrát test

- **Hypotéza o nezávislosti:** Souvisí spolu výskyt dvou nominálních znaků měřených na jediném výběru?
Příklad: Barva očí (modrá, zelená, hnědá) a barva vlasů (hnědá, černá, blond) u vybraných 95 studentů jsou nezávislé.
- H_0 : Znaky X a Y jsou nezávislé náhodné veličiny.
- H_A : Znaky X a Y jsou závislé náhodné veličiny.

– Test: **Pearsonův chí-kvadrát**

$$K = \sum_{j=1}^r \sum_{k=1}^c \frac{(n_{jk} - e_{jk})^2}{e_{jk}} \approx \chi^2((r-1)(c-1))$$

Očekávané teoretické četnosti: $e_{jk} = \frac{n_{j.} \cdot n_{.k}}{n}$

- H_0 zamítáme na hladině významnosti α , pokud

$$K \geq \chi_{1-\alpha}^2((r-1)(c-1))$$

Testování nezávislosti – Pearsonův chí-kvadrát test

Předpoklady Pearsonova chí-kvadrát testu:

1. Jednotlivá pozorování jsou nezávislá (tj. každý prvek patří jen do jedné buňky kontingenční tabulky)

2. Podmínka dobré aproximace

Očekávané (teoretické) četnosti jsou aspoň v 80 % případů větší nebo rovné 5 a ve 100 % případů nesmí být pod 2 (pokud není tento předpoklad splněn, je vhodné sloučit kategorie s nízkými četnostmi).

Měření síly závislosti: Cramérův koeficient

Význam hodnot: 0 – zanedbatelná závislost 1 – silná závislost

Testování nezávislosti – příklad

- **Příklad:** Souvisí pohlaví s výskytem nemoci?
- H_0 : Pohlaví a výskyt nemoci jsou nezávislé veličiny.
 H_A : Pohlaví a výskyt nemoci nejsou nezávislé veličiny.

Pozorované četnosti

	Nemocný	Zdravý	
Muž	45	11	56
Žena	25	6	31
	70	17	87

Očekávané četnosti

	Nemocný	Zdravý	
Muž	45,1 _{70*56/87}	10,9 _{17*56/87}	56
Žena	24,9 _{70*31/87}	6,1 _{17*31/87}	31
	70	17	87

$$\chi^2 = 0,001 \quad df = 1 \quad p = 0,974 \quad \rightarrow \quad \text{nezamítáme } H_0$$

Testování shody struktury – Pearsonův chí-kvadrát test

- **Hypotéza o shodě struktury:** Zajímá nás výskyt nominálního znaku u r nezávislých výběrů.
Příklad: Je zájem o sport stejný u děvčat jako u chlapců?
- H_0 : Pravděpodobnostní rozdělení kategoriální proměnné je stejné v různých populacích.
- Test: **Pearsonův chí-kvadrát test.**

	Zájem o sport ANO	Zájem o sport NE	Celkem
Dívky	a	b	a + b
Chlapci	c	d	c + d
Celkem	a + c	b + d	N

Některé marginální četnosti (buď sloupcové nebo řádkové) jsou předem pevně stanoveny

Fisherův exaktní test

- Využití ve čtyřpolní tabulce s nízkými četnostmi, které znemožňují použití Pearsonova chí-kvadrát testu.
- Patří mezi **neparametrické testy** pracující s daty na nominální škále, v nejjednodušší podobě ve dvou třídách: pozitivní/negativní, úspěch/neúspěch apod.
- Nulová hypotéza H_0 předpokládá rovnoměrné zastoupení sledovaného znaku u dvou nezávislých souborů.
- Slovo exaktní (přímý) znamená, že se přímo vypočítává pravděpodobnost odmítnutí, resp. platnosti nulové hypotézy.

Fisherův exaktní test

- Výpočet přesné p-hodnoty jako pravděpodobnosti, s jakou dostaneme za předpokladu platnosti nulové hypotézy tabulku stejně nebo více odlišnou od nulové hypotézy.

Sledovaný jev	Kontrolní skupina	Experimentální skupina	Celkem
Ano	a	b	a + b
Ne	c	d	c + d
Celkem	a + c	b + d	N

1. Spočítá se parciální pravděpodobnost čtyřpolní tabulky p_1
2. Spočítá se p_o všech možných tabulek při zachování marginálních četností (řádkové a sloupcové součty). Výsledná p-hodnota je součtem p_o menších nebo stejných jako p_1 , která přísluší pozorované tabulce.

Testování symetrie – McNemarův test

- **Hypotéza o symetrii:** Opakovaně sledujeme binární proměnnou a zajímá nás, zda došlo ke změně jejího rozdělení.
Příklad: Výskyt bolesti před a po užití léku.
- $H_0: n_{ij} = n_{ji}$ (pokus nemá vliv na výskyt daného znaku)

Četnost	Po: ANO	Po: NE	
Před: ANO	a	b	a + b
Před: NE	c	d	c + d
	a + c	b + d	N

Teoretická pravděpodobnost	Po: ANO	Po: NE	
Před: ANO	n_{11}	n_{12}	$n_{1.}$
Před: NE	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	

- Testová statistika: $\chi^2 = \frac{(|b - c| - 1)^2}{b + c}$ Pokud je větší než kritická hodnota χ^2 rozdělení o jednom stupni volnosti (vhodné pro počty údajů $b + c > 8$), pak nulovou hypotézu zamítáme.

**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 mozkovém 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

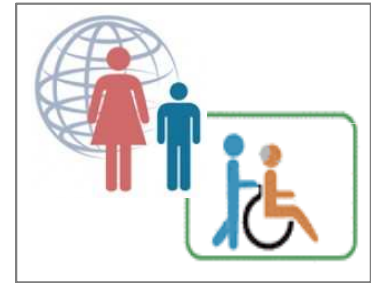
Sbírané informace:

- 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. Pearsonův chí-kvadrát test

Úkol č. 1 – Pearsonův chí-kvadrát test

Zadání: „Stupeň soběstačnosti pacientů po mozgovém infarktu lze pomocí indexu Barthelové vyjádřit také kategoriálně. Např. pro definici vysoce závislých pacientů bylo stanoveno rozmezí 0 až 40 bodů. Zjistěte, zda je u žen a mužů stejné procento alespoň částečně soběstačných pacientů (45 až 100 bodů) a zda je tento rozdíl statisticky významný.“




Úkol č. 1 – Pearsonův chí-kvadrát test

Postup:

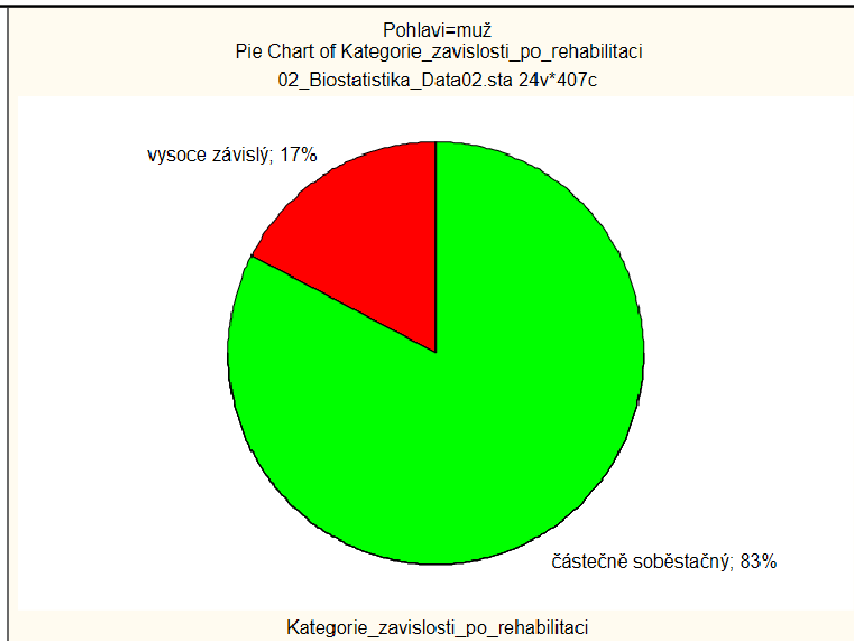
1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu
 H_0 : „Stupeň soběstačnosti nezávisí na pohlaví“ proti
 H_A : „Stupeň soběstačnosti a pohlaví jsou závislé veličiny.“
2. Vypočítáme očekávané a pozorované četnosti v kategoriích.
3. Vypočítáme **testovou statistiku K** a odpovídající **p-hodnotu**:

$$K = \sum_{j=1}^r \sum_{k=1}^s \frac{(n_{jk} - e_{jk})^2}{e_{jk}} = \frac{(205 - 200)^2}{200} + \frac{(123 - 128)^2}{128} + \frac{(43 - 48)^2}{48} + \frac{(36 - 31)^2}{31} = 1,74 \Rightarrow p = 0,187$$

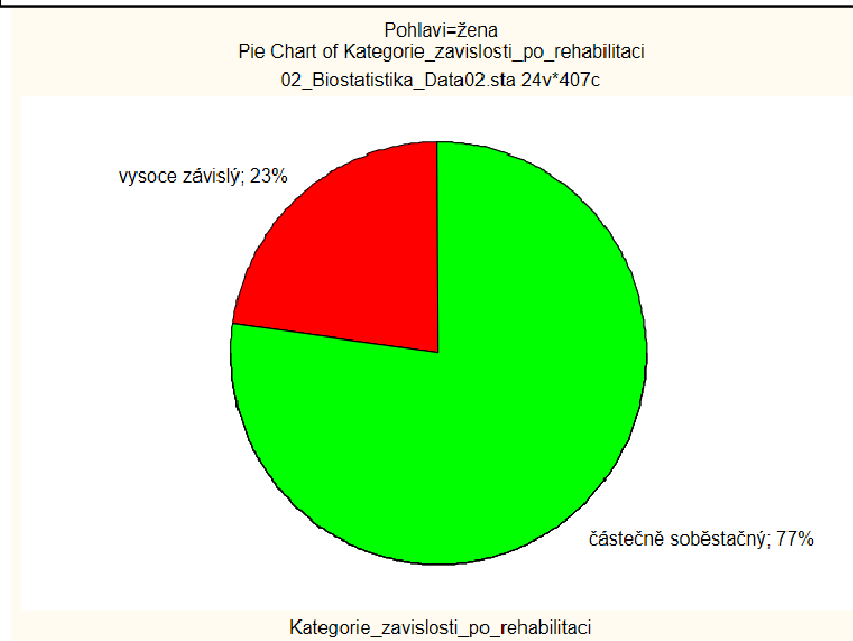
4. Testovou statistiku porovnáme s kritickou hodnotou nebo porovnáme p-hodnotu s hladinou významnosti $\alpha = 0,05$.
5. Je-li **p-hodnota $> \alpha$**  **nezamítáme H_0 . Stupeň soběstačnosti nezávisí na pohlaví (tj. výsledná míra soběstačnosti se u žen a u mužů neliší).**

Úkol č. 1 – Popis dat

Zastoupení částečně soběstačných a vysoce závislých mužů



Zastoupení částečně soběstačných a vysoce závislých žen



① Ze základního popisu je patrný mírný rozdíl v procentu částečně soběstačných pacientů na konci hospitalizace. U žen je podíl těchto pacientů 77 % oproti 83 % u mužů.

Úkol č. 1 – Řešení v programu Statistica

- V menu **Statistics** zvolíme **Basic statistics**, vybereme **Tables and banners** (v češtině **Kontingenční tabulky**).

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Basic Statistics and Tables' dialog box is displayed. A red arrow labeled '1' points to the 'Statistics' menu. A red arrow labeled '2' points to the 'Basic Statistics' icon in the ribbon. A red arrow labeled '3' points to the 'Tables and banners' option in the 'Quick' list of the dialog box. The background shows a data table with columns 'ID', 'Pohlavi', and 'Vek'.

	1	2	3
	ID	Pohlavi	Vek
	1	1 muž	
	2	2 žena	
	3	3 muž	
	4	4 žena	
	5	5 muž	
	6	6 muž	
	7	7 muž	
	8	8 muž	
	9	9 žena	
	10	10 muž	
	11	11 muž	
	12	12 žena	
	13	13 žena	
	14	14 muž	

Úkol č. 1 – Řešení v programu Statistica

- Na záložce **Stub-and-banner** vybereme **proměnné**, které chceme testovat, a potvrdíme **OK**.

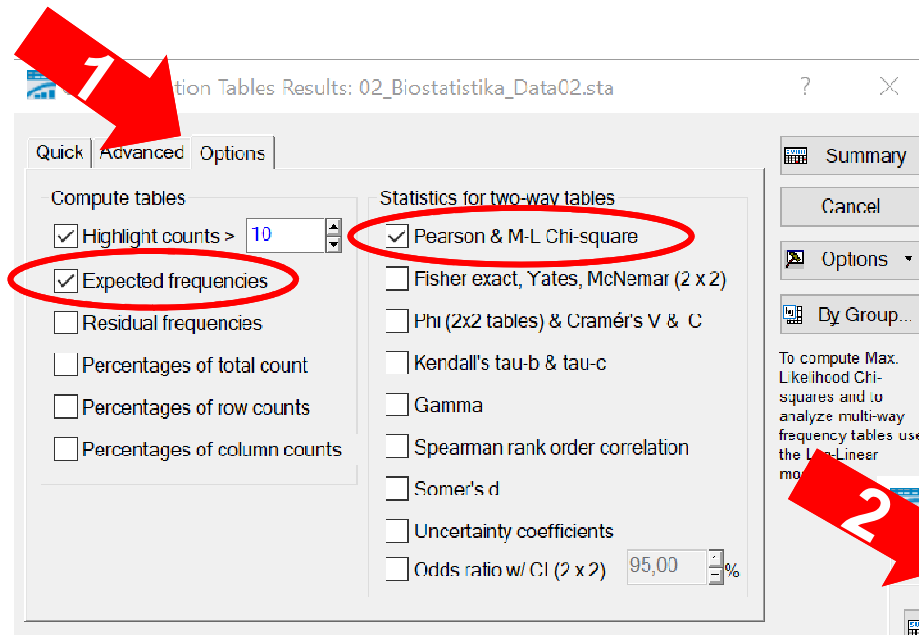
The screenshot shows the Statistica software interface. The 'Statistics' ribbon is active, and the 'Stub-and-banner' dialog box is open. The dialog box has two tabs: 'Crosstabulation' and 'Stub-and-banner'. The 'Stub-and-banner' tab is selected. The dialog box contains the following options:

- Specify tables (select variables):** A button to select variables.
- No. of 2-way tables:** none
- Identification of levels in table factors:**
 - Use all integer codes in the selected vars
 - Use selected grouping codes only
- Codes:** Not selected
- Weighted moments:**

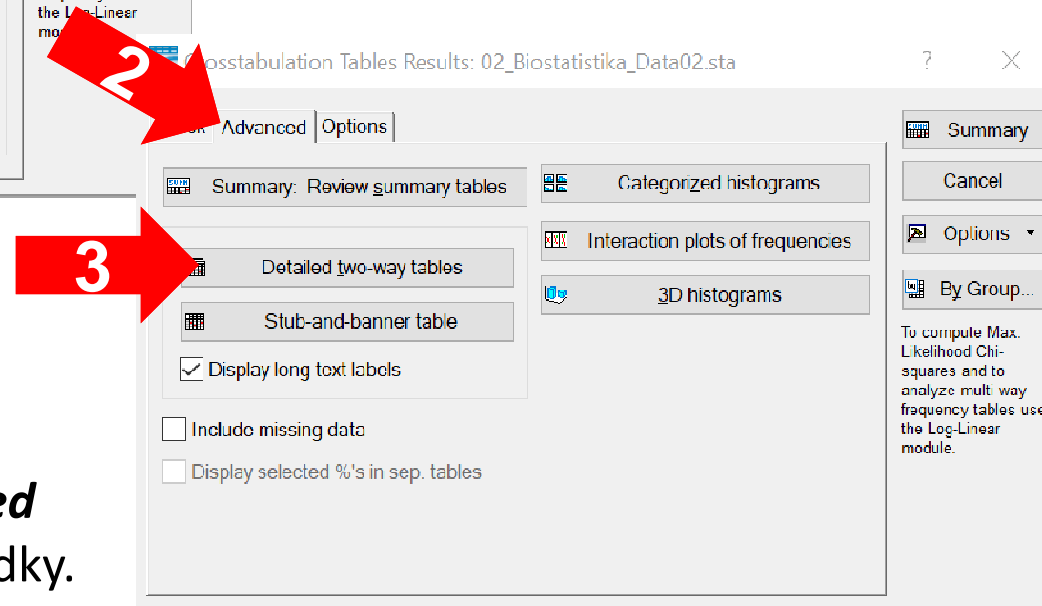
Red arrows point to the 'Stub-and-banner' tab (1), the 'Specify tables' button (2), and the 'OK' button (3).

Rehabilitace po mozkovem infarktu: data							
	1	2	3	4	5	6	7
	ID	Pohlavi	Vek	etiologie	Lokalizace	T	Komorbidity
1	1	muž					
2	2	žena					
3	3	muž					
4	4	žena					
5							
6	6	muž					
7	7	muž					
8	8	muž					
9	9	žena					
10	10	muž					
11	11	muž					
12	12	žena					
13	13	žena					
14	14	muž					

Úkol č. 1 – Řešení v programu Statistica



- Na záložce **Options** zaškrtneme **Expected frequencies (Očekávané četnosti)** potřebné k ověření podmínek dobré aproximace) a **Pearsonův chí-kvadrát**.



- Poté se vrátíme na záložku **Advanced** a přes volbu **Detailed two-way tables** získáme výsledky.

Úkol č. 1 – Výsledky v Statistica

Pozorované četnosti

2-Way Summary Table: Observed Frequencies (02_Biostat)
Marked cells have counts > 10

Pohlavi	Kategorie_zavislosti_ po_rehabilitaci částečně soběstačný	Kategorie_zavislosti_ po_rehabilitaci vysoce závislý	Row Totals
	muž	205	
žena	123	36	159
Totals	328	79	407

Očekávané četnosti

2-Way Summary Table: Expected Frequencies (02_Biostat)
Marked cells have counts > 10

Pohlavi	Kategorie_zavislosti_ po_rehabilitaci částečně soběstačný	Kategorie_zavislosti_ po_rehabilitaci vysoce závislý	Row Totals
	muž	199,8624	
žena	128,1376	30,86241	159,0000
Totals	328,0000	79,00000	407,0000

① Z předchozího popisu je patrný mírný rozdíl mezi muži a ženami (u žen je podíl částečně soběstačných pacientů 77 % oproti 83 % u mužů).

Očekávané četnosti jsou 200, 48, 128 a 31, což jsou dostatečně vysoké počty a podmínka dobré aproximace pro použití chí-kvadrát testu je tedy splněna.

Statistics: Pohlavi(2) x Kategorie_za

Statistic	Chi-square	df	p
Pearson Chi-square	1,741617	df=1	p=,18693
M-L Chi-square	1,720067	df=1	p=,18693

p-hodnota
Pearsonova
chí-kvadrát testu

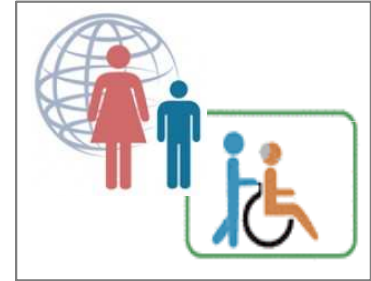


P-hodnota statistické významnosti pozorované závislosti je $p = 0,187$, což na hladině významnosti 0,05 značí **nevýznamný výsledek** a ze získaných dat tedy **nelze říct, že by míra soběstačnosti souvisela s pohlavím.**

Úkol 2. Fisherův exaktní test

Úkol č. 2 – Fisherův exaktní test

Zadání: „Stupeň soběstačnosti pacientů po mozgovém infarktu lze pomocí indexu Barthelové vyjádřit také kategoriálně. Např. pro definici vysoce závislých pacientů bylo stanoveno rozmezí 0 až 40 bodů. Zjistěte, zda je u žen a mužů léčených mechanickou trombektomií stejné procento alespoň částečně soběstačných pacientů (45 až 100 bodů) a zda je tento rozdíl statisticky významný.“




Úkol č. 2 – Fisherův exaktní test

Postup (po nemožnosti použít Pearsonův chí-kvadrát test):

1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu
 H_0 : „Stupeň soběstačnosti nezávisí na pohlaví“ proti
 H_A : „Stupeň soběstačnosti a pohlaví jsou závislé veličiny.“
2. Spočítá se parciální pravděpodobnost (p_o) všech možných tabulek při zachování marginálních četností. Výsledná p-hodnota je součtem p_o menších nebo stejných jako pravděpodobnost, která přísluší námi pozorované tabulce.

$$\Rightarrow p = 0,700$$

3. Vypočítané p porovnáme s hladinou významnosti $\alpha = 0,05$.
4. Je-li p-hodnota $> \alpha$  nezamítáme H_0 . **Stupeň soběstačnosti nezávisí na pohlaví (tj. výsledná míra soběstačnosti se u žen a u mužů podstupujících mechanickou trombektomii neliší).**

Úkol č. 2 – Popis dat

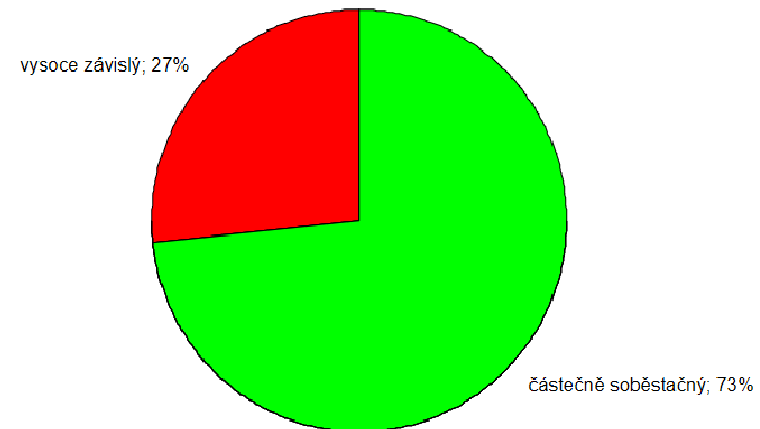
Zastoupení částečně soběstačných a vysoce závislých mužů

Pohlaví=muž, Terapie=mechanická trombektomie
Pie Chart of Kategorie_zavislosti_po_rehabilitaci
02_Biostatistika_Data02.sta 24v*407c



Zastoupení částečně soběstačných a vysoce závislých žen

Pohlaví=žena, Terapie=mechanická trombektomie
Pie Chart of Kategorie_zavislosti_po_rehabilitaci
02_Biostatistika_Data02.sta 24v*407c



① Ze základního popisu je patrný mírný rozdíl v procentu částečně soběstačných pacientů na konci hospitalizace. U žen je podíl těchto pacientů 73 % oproti 64 % u mužů.

Úkol č. 2 – Řešení v programu Statistica

- V menu **Statistics** zvolíme **Basic statistics**, vybereme **Tables and banners** (v češtině **Kontingenční tabulky**).

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Basic Statistics and Tables' dialog box is displayed. A red arrow labeled '1' points to the 'Statistics' menu, a red arrow labeled '2' points to the 'Basic Statistics' icon, and a red arrow labeled '3' points to the 'Tables and banners' option in the dialog box. The background shows a data table with columns 'ID', 'Pohlavi', and 'Vek'.

	1	2	3
	ID	Pohlavi	Vek
	1	1 muž	
	2	2 žena	
	3	3 muž	
	4	4 žena	
	5	5 muž	
	6	6 muž	
	7	7 muž	
	8	8 muž	
	9	9 žena	
	10	10 muž	
	11	11 muž	
	12	12 žena	
	13	13 žena	
	14	14 muž	

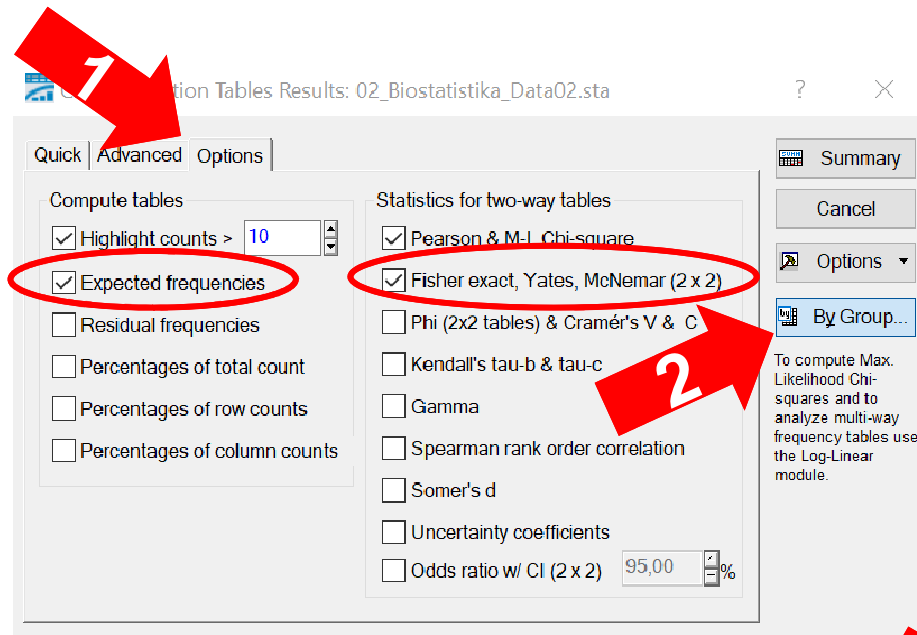
Úkol č. 2 – Řešení v programu Statistica

- Na záložce **Stub-and-banner** vybereme **proměnné**, které chceme testovat, a potvrdíme **OK**.

The screenshot shows the Statistica software interface. The main window displays a data table titled "Rehabilitace po mozkovem infarktu: data" with 14 rows and 7 columns. The columns are labeled: 1 ID, 2 Pohlavi, 3 Vek, 4 etiologie, 5 Lokalizace, 6 T, and 7 Komorbidit. The data rows contain numerical IDs and gender labels (muž, žena). A dialog box titled "Crosstabulation" is open, showing the "Stub-and-banner" tab. The dialog has a "Specify tables (select variables)" section, a "No. of 2-way tables" set to "none", and an "Identification of levels in table factors" section with two radio buttons: "Use all integer codes in the selected vars" (selected) and "Use selected grouping codes only". There is also a "Codes" field set to "Not selected" and a "Weighted moments" checkbox. Red arrows point to the "Stub-and-banner" tab (labeled 1), the "Specify tables" section (labeled 2), and the "OK" button (labeled 3).

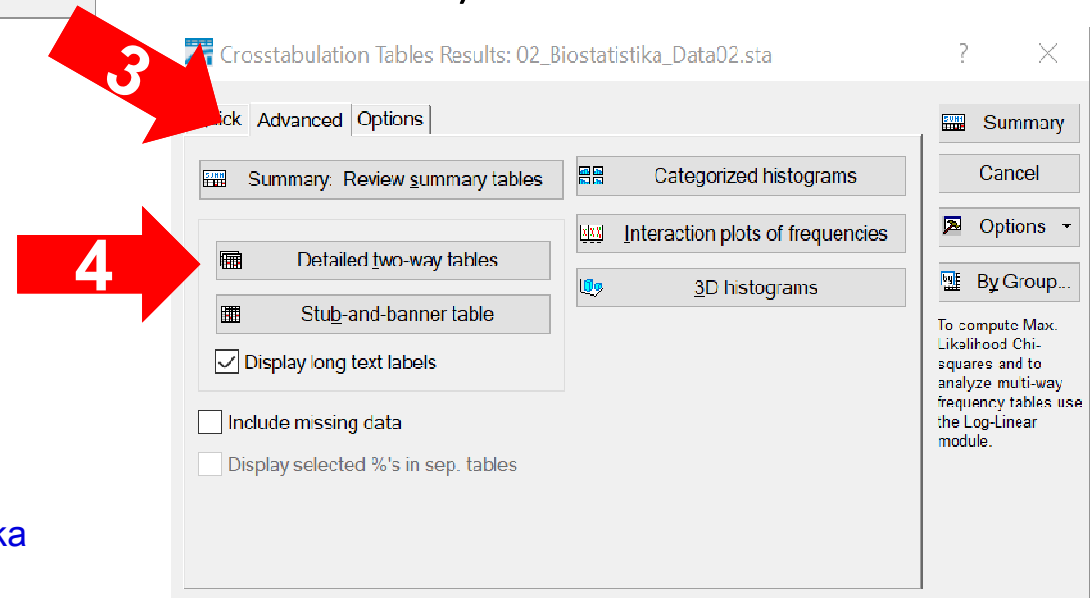
	1	2	3	4	5	6	7
	ID	Pohlavi	Vek	etiologie	Lokalizace	T	Komorbidit
1	1	muž					
2	2	žena					
3	3	muž					
4	4	žena					
5							
6	6	muž					
7	7	muž					
8	8	muž					
9	9	žena					
10	10	muž					
11	11	muž					
12	12	žena					
13	13	žena					
14	14	muž					

Úkol č. 2 – Řešení v programu Statistica



- Na záložce **Options** zaškrtneme **Expected frequencies (Očekávané četnosti)** potřebné k ověření podmínek dobré aproximace) a **Fisher exact**.
- V nastavení **By Group** vybereme jako třídící proměnnou terapii (analýza se tak provede pro všechny druhy terapie samostatně).

- Poté se vrátíme na záložku **Advanced** a přes volbu **Detailed two-way tables** získáme výsledky.



Úkol č. 2 – Výsledky v Statistica

Pozorované četnosti

Terapie=mechanická trombektomie
2-Way Summary Table: Observed Frequencies (02_Biost)
Marked cells have counts > 10

Pohlavi	Kategorie_zavislosti_ po_rehabilitaci vysoce závislý	Kategorie_zavislosti_ po_rehabilitaci částečně soběstačný	Row Totals
muž	5	9	14
žena	4	11	15
Totals	9	20	29

Očekávané četnosti

Terapie=mechanická trombektomie
2-Way Summary Table: Expected Frequencies (02_Biostat)
Marked cells have counts > 10

Pohlavi	Kategorie_zavislosti_ po_rehabilitaci vysoce závislý	Kategorie_zavislosti_ po_rehabilitaci částečně soběstačný	Row Totals
muž	4,344828	9,65517	14,00000
žena	4,655172	10,34483	15,00000
Totals	9,000000	20,00000	29,00000

① Z předchozího popisu je patrný mírný rozdíl mezi muži a ženami (u žen je podíl částečně soběstačných pacientů 73 % oproti 64 % u mužů).

Očekávané četnosti jsou 4, 10, 5 a 10, což nejsou dostatečně vysoké počty a místo chí-kvadrát testu je tedy vhodné použít Fisherův exaktní test.

Terapie=mechanická trombektomie
Statistics: Pohlavi(2) x Kategorie_zavislosti

Statistic	Chi-square	df	p
Pearson Chi-square	,2769577	df=1	p=,59870
M-L Chi-square	,2771859	df=1	p=,59855
Yates Chi-square	,0155357	df=1	p=,90081
Fisher exact, one-tailed			p=,44998
two-tailed			p=,69985
	1,562500	df=1	p=,11130
	1,230769	df=1	p=,26726

**p-hodnota
Fisherova exaktního testu**



P-hodnota statistické významnosti pozorované závislosti je $p = 0,700$, což na hladině významnosti 0,05 značí **nevýznamný výsledek** a ze získaných dat tedy **nelze říct, že by míra soběstačnosti souvisela s pohlavím.**

Úkol 3. McNemarův test

Úkol č. 3 – McNemarův 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 pomocí indexu Barthelové (BI) jako *vysoce závislý* (0 až 40 bodů) nebo *částečně soběstačný* (45 až 100 bodů) a byli přeloženi na rehabilitační oddělení. Po dvou týdnech byl stejně vyhodnocen stupeň soběstačnosti dle BI. Zjistěte, zda poskytnutá rehabilitační péče vedla ke zvýšení podílu alespoň částečně soběstačných pacientů.“



Úkol č. 3 – McNemarův test

Postup:

1. Na hladině významnosti $\alpha = 0,05$ testujeme hypotézu H_0 : „Počet zhoršených případů je stejný jako počet zlepšení“ proti H_A : „Počet zhoršených případů není stejný jako počet zlepšení.“

2. Vypočítáme pozorované četnosti měnících se stavů.

3. Vypočítáme **testovou statistiku K** a odpovídající **p-hodnotu**:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} = \frac{(|280 - 0| - 1)^2}{280 + 0} = 278 \quad \Rightarrow \quad p < 0,001$$

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 . Během rehabilitace se podařilo změnit míru soběstačnosti pacientů.**

Úkol č. 3 – Řešení v programu Statistica

- V menu **Statistics** zvolíme **Basic statistics**, vybereme **Tables and banners** (v češtině **Kontingenční tabulky**).

The screenshot shows the Statistica software interface. The 'Statistics' menu is open, and the 'Basic Statistics and Tables' dialog box is displayed. A red arrow labeled '1' points to the 'Statistics' menu. A red arrow labeled '2' points to the 'Basic Statistics' icon in the ribbon. A red arrow labeled '3' points to the 'Tables and banners' option in the 'Quick' list of the dialog box. The background shows a data table with columns 'ID', 'Pohlavi', and 'Vek'.

	1	2	3
	ID	Pohlavi	Vek
	1	1 muž	
	2	2 žena	
	3	3 muž	
	4	4 žena	
	5	5 muž	
	6	6 muž	
	7	7 muž	
	8	8 muž	
	9	9 žena	
	10	10 muž	
	11	11 muž	
	12	12 žena	
	13	13 žena	
	14	14 muž	

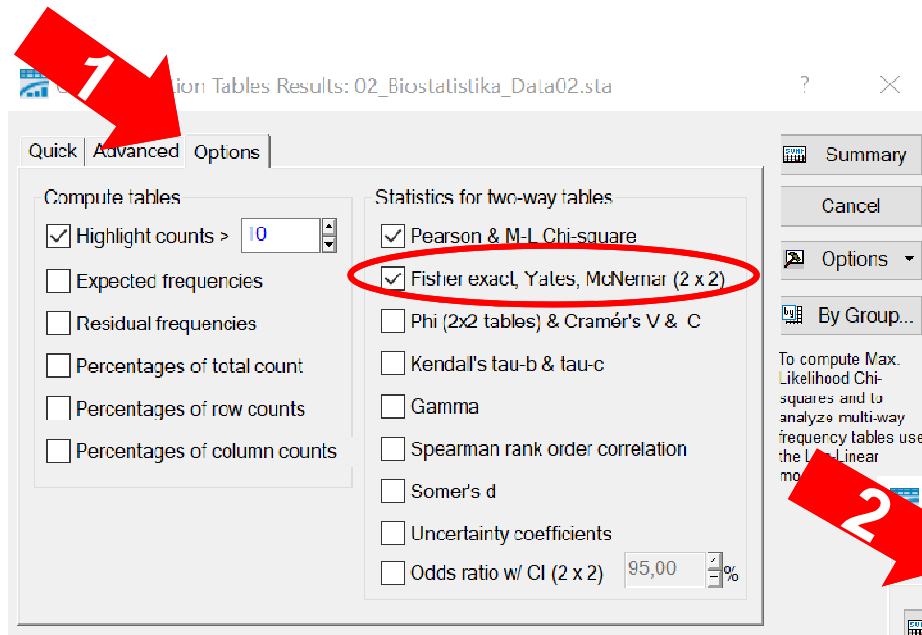
Úkol č. 3 – Řešení v programu Statistica

- Na záložce **Stub-and-banner** vybereme **proměnné**, které chceme testovat, a potvrdíme **OK**.

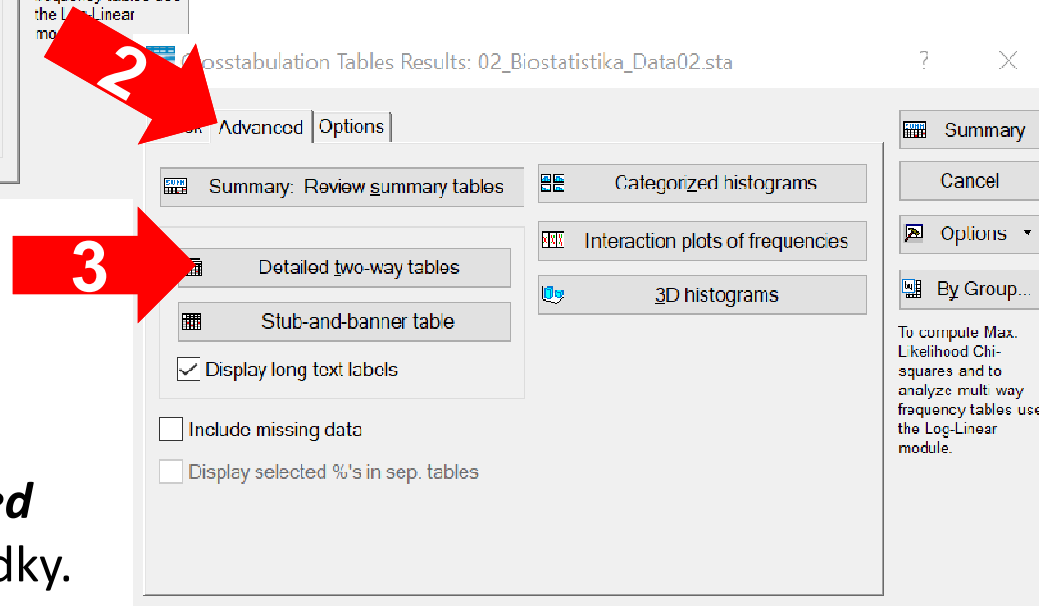
The screenshot shows the Statistica software interface. The main window displays a data table titled "Rehabilitace po mozkovem infarktu: data" with 14 rows and 7 columns. The columns are labeled: 1 ID, 2 Pohlavi, 3 Vek, 4 Etiologie, 5 Lokalizace, 6 Tření, and 7 Komorbidita. The data rows contain numerical IDs and gender labels (muž, žena). A dialog box titled "Crosstabulation" is open, showing the "Stub-and-banner" tab. The dialog has a "Specify tables (select variables)" section, a "No. of 2-way tables" set to "none", and an "Identification of levels in table factors" section with two radio buttons: "Use all integer codes in the selected vars" (selected) and "Use selected grouping codes only". There is also a "Codes" field set to "Not selected" and a "Weighted moments" checkbox. Three red arrows point to the "Specify tables" section (labeled 2), the "Stub-and-banner" tab (labeled 1), and the "OK" button (labeled 3).

	1	2	3	4	5	6	7
	ID	Pohlavi	Vek	Etiologie	Lokalizace	Tření	Komorbidita
1	1	muž					
2	2	žena					
3	3	muž					
4	4	žena					
5	5						
6	6	muž					
7	7	muž					
8	8	muž					
9	9	žena					
10	10	muž					
11	11	muž					
12	12	žena					
13	13	žena					
14	14	muž					

Úkol č. 3 – Řešení v programu Statistica



- Na záložce *Options* zaškrtneme *McNemar (2x2)*.



- Poté se vrátíme na záložku *Advanced* a přes volbu *Detailed two-way tables* získáme výsledky.

Úkol č. 3 – Výsledky v Statistica

Pozorované četnosti

2-Way Summary Table: Observed Frequencies (02_Biostatistics)
Marked cells have counts > 10

Kategorie_zavislosti_pred_rehabilitaci	Kategorie_zavislosti_po_rehabilitaci částečně soběstačný	Kategorie_zavislosti_po_rehabilitaci vysoce závislý	Row Totals
vysoce závislý	A 280	B 79	359
částečně soběstačný	C 48	D 0	48
Totals	328	79	407

Statistics: Kategorie_zavislosti_pred

Statistic	Chi-square	df	p
Pearson Chi-square	13,10673	df=1	p=,00029
M-L Chi-square	22,21371	df=1	p=,00000
Yates Chi-square	11,73772	df=1	p=,00061
Fisher exact, one-tailed			p=,00002
two-tailed			p=,00002
McNemar Chi-square (A/D)	278,0036	df=1	p=0,0000
(B/C)	7,086614	df=1	p=,00777



**p-hodnota
McNemarova testu**

Dvě hodnoty testových statistik a p-hodnoty podle toho, kde jsou ve výstupní kontingenční tabulce uloženy četnosti, u kterých jsme při opakovaném měření zaznamenali rozdílné výsledky (A/D nebo B/C).

① Počet pacientů, u kterých došlo ke změně z vysoce závislého stavu do částečně soběstačného je 280. Naopak ke zhoršení nedošlo u žádného pacienta. Počty změn jsou v kontingenční tabulce na pozicích A a D.

② P-hodnota statistické významnosti pozorované změny je $p < 0,001$, což na hladině významnosti 0,05 značí **významný výsledek** a ze získaných dat jsme **prokázali, že během rehabilitace se podařilo změnit míru soběstačnosti pacientů v denních aktivitách.**