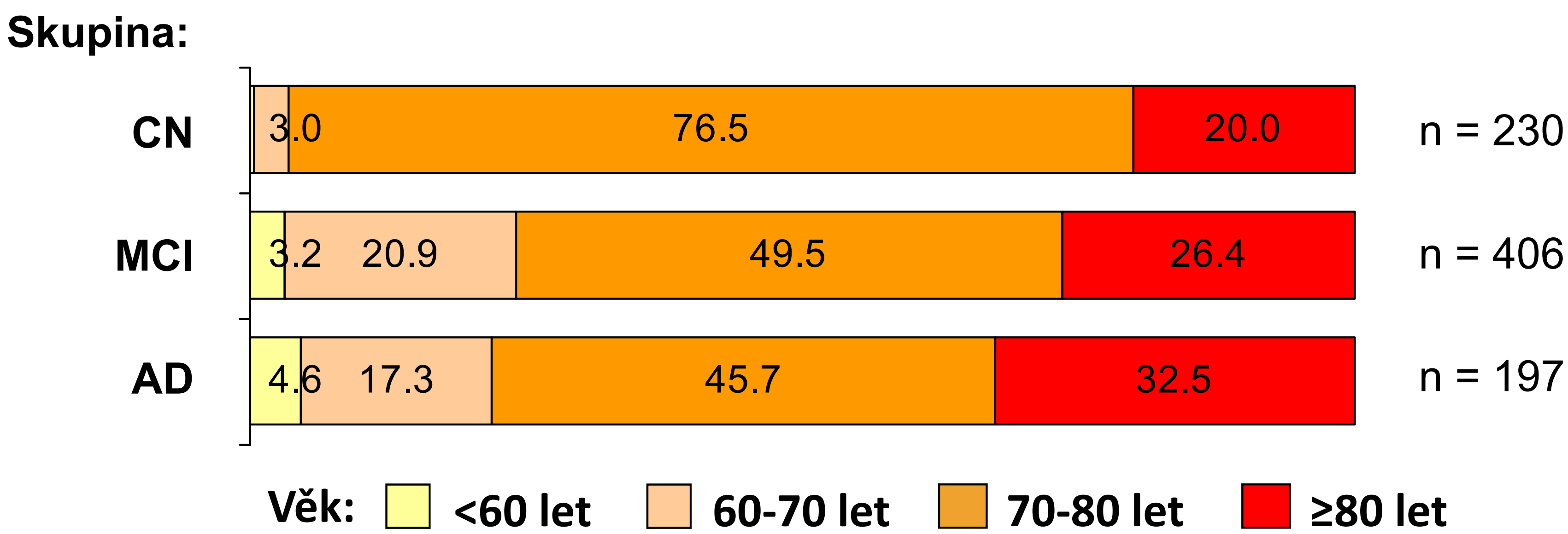
Kontingenční tabulky – ukázka finálního popisu a vizualizace

Skupina	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN	1 (0,4%)	7 (3,0%)	176 (76,5%)	46 (20,0%)	230 (100,0%)
MCI	13 (3,2%)	85 (20,9%)	201 (49,5%)	107 (26,4%)	406 (100,0%)
AD	9 (4,6%)	34 (17,3%)	90 (45,7%)	64 (32,5%)	197 (100,0%)
Celkem	23 (2,8%)	126 (15,1%)	467 (56,1%)	217 (26,1%)	833 (100,0%)



Kontingenční tabulky – hypotézy

- Kontingenční tabulky umožňují testování různých hypotéz:
- **Nezávislost a shoda struktury** (Pearsonův chí-kvadrát test, Fisherův exaktní test)
  - Jeden výběr, dvě charakteristiky nebo více výběrů, jedna charakteristika – obdoba nepárového uspořádání
  - Př.: pacienti s AD – pohlaví × vzdělání (VŠ, SŠ, ZŠ); pacienti s AD v několika nemocnicích × věková struktura
- **Symetrie** (McNemarův test)
  - Jeden výběr, opakovaně jedna charakteristika – obdoba párového uspořádání
  - Př.: MMSE v normě a pod normou na začátku studie a dva roky po zahájení studie

Pearsonův chí-kvadrát test

- Založen na myšlence srovnání pozorovaných a očekávaných četností kategorií dvou proměnných.
- Pozorované četnosti jednotlivých kategorií první proměnné a druhé proměnné nám vyjadřují n_{ij}.
- Očekávané četnosti jednotlivých kategorií lze vypočítat pomocí:

$$e_{ij} = \frac{n_{i.} n_{.j}}{n}$$

($n_{i.}$ je součet hodnot v řádku, $n_{.j}$ je součet hodnot ve sloupci)

Typ onemocnění	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1.}$
MCI	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2.}$
AD	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3.}$
Celkem	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	n

- Výpočet testové statistiky:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

- Nulovou hypotézu o nezávislosti dvou kategoriálních proměnných zamítáme na hladině významnosti α, když $X^2 \geq \chi^2_{(1-\alpha)}(r-1)(c-1)$

Pearsonův chí-kvadrát test: Chceme zjistit, jestli existuje vztah mezi typem onemocnění a věkovými kategoriemi v našem souboru.

Tabulka pozorovaných četností:

Typ onemocnění	Věk <60 let	Věk 60-70 let	Věk 70-80 let	Věk ≥80 let	Celkem
CN	1	7	176	46	230
MCI	13	85	201	107	406
AD	9	34	90	64	197
Celkem	23	126	467	217	833

Tabulka očekávaných četností:

Typ onemocnění	Věk <60 let	Věk 60-70 let	Věk 70-80 let	Věk ≥80 let	Celkem
CN					230
MCI					406
AD					197
Celkem	23	126	467	217	833

Testová statistika:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} =$$

Pearsonův chí-kvadrát test

Příklad: Chceme zjistit, jestli existuje vztah mezi typem onemocnění a věkovými kategoriemi v našem souboru.

Postup:

Tabulka pozorovaných četností:

Typ onemocnění	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN	1	7	176	46	230
MCI	13	85	201	107	406
AD	9	34	90	64	197
Celkem	23	126	467	217	833

Tabulka očekávaných četností:

Typ onemocnění	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN	6,4	34,8	128,9	59,9	230
MCI	11,2	61,4	227,6	105,8	406
AD	5,4	29,8	110,4	51,3	197
Celkem	23	126	467	217	833

$$e_{11} = \frac{23 \cdot 230}{833} = 6{,}4$$

$$e_{21} = \frac{23 \cdot 406}{833} = 11{,}2$$

$$e_{12} = \frac{126 \cdot 230}{833} = 34{,}8 \quad \ldots$$

Testová statistika: $$\mathrm{X}^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(1-6{,}4)^2}{6{,}4} + \frac{(7-34{,}8)^2}{34{,}8} + \ldots = 69{,}4$$

$\mathrm{X}^2 = 69{,}4 \geq \chi^2_{(0,95)}(3-1)(4-1) = \chi^2_{(0,95)}(6) = 12{,}6$ → **zamítáme H_0** o nezávislosti → Vztah mezi typem onemocnění a věkovými kategoriemi je statisticky významný.

Předpoklady Pearsonova chí-kvadrát testu

- Nezávislost jednotlivých pozorování
- Alespoň 80 % buněk musí mít očekávanou četnost (e_{ij}) větší než 5
- 100 % buněk musí mít očekávanou četnost (e_{ij}) větší než 2

- Může nám pomoci slučování kategorií, ale můžeme slučovat jen slučitelné kategorie!

Fisherův exaktní test

- Určen pro čtyřpolní tabulky, **je vhodný i pro tabulky s malými četnostmi** – pro ty, které nesplňují předpoklad Pearsonova chí-kvadrát testu.
- Založen na výpočtu „přesné" p-hodnoty (pravděpodobnosti, s jakou bychom dostali stejný nebo ještě extrémnější výsledek při zachování součtu řádků i sloupců v tabulce).
- **Příklad**: Chceme ověřit vztah dvou typů nežádoucích účinků, které jsou sumarizovány následující tabulkou:

		NÚ II ano	NÚ II ne
NÚ I	ano	2	3
	ne	6	4

- **Postup:** Všechny varianty tabulky při zachování součtu řádků a sloupců:

0	5	1	4	2	3	3	2	4	1	5	0
8	2	7	3	6	4	5	5	4	6	3	7

Pravděpodobnosti výskytu jednotlivých tabulek:

0,007 0,093 0,326 0,392 0,163 0,019

Oboustranná p-hodnota (sečtení pravděpodobností stejných nebo menších než je pravděpodobnost pozorované varianty):

p = 0,326 + 0,093 + 0,007 + 0,163 + 0,019 = **0,608**

Fisherův x Pearsonův test

- Pearsonův chí-kvadrát test lze použít na jakoukoliv kontingenční tabulku, ALE je nutné hlídat předpoklady: 100% očekávaných četností větších než 2 a 80 % očekávaných četností větších než 5 – u čtyřpolní tabulky to znamená, že všechny očekávané četnosti musí být větší než 5.

- **Nedodržení předpokladů pro Pearsonův chí-kvadrát test může stejně jako u t-testu a analýzy rozptylu vést k nesmyslným závěrům!**

- **Pro hodnocení čtyřpolních tabulek je Fisherův exaktní test standardem v klinických analýzách.**