Párový neparametrický test

- Wilcoxonův test: obdobně seřadíme diference a testujeme proti kritické hodnotě daného testu

- Výpočet pomocí STATISTICY

Dieta laboratorních krys

- Máme dva typy diety. Zkontrolujte předpoklad párovosti a vhodnosti použití parametrického testu (r=0,98)

- můžeme použít oba testy – nicméně, zde použijeme nepárovou variantu testu.

[korelace ✓, Normalita diferencí ✓
(nicméně 0.094 je vcelku málo, šly by použít oba testy),
parametrický: p=0.102 ⇨ nezamítáme H0
Neparametricky: Wilcoxon : p=0.056 ⇨ nezamítáme H0
Znaménkový test: p=0.080 ⇨ nezamítáme H0]

Obdoba – jednovýběrový neparametrický test

- Wilcoxonův znaménkový test
- ve STATISTICE není naimplementovaný
- nicméně porovnáváme hodnotu mediánu jednoho výběru proti nějaké hodnotě
- můžeme využít Wilcoxonův test pro párové uspořádání testu tak, že druhý výběr bude sestávat pouze z hodnoty, s kterou chceme porovnávat náš původní výběr

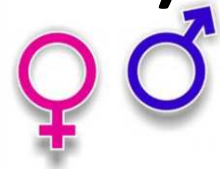
Vrtačka

- *Vrtacka.sta*
- Testujte na hladině významnosti 0,05 , že výdrž jednoho vrtáku ve vrtačce je 500 otáček

Kontingenční tabulky

Typy dat - opakování

- **Kvalitativní (kategoriální) data:**
  - Binární data



  - Nominální data

  - Ordinální data



- **Kvantitativní data:**
  - Intervalová data

  - Poměrová data

Kontingenční tabulka

- Frekvenční sumarizace dvou binárních, nominálních nebo ordinálních proměnných.
- Obecně: **R x C kontingenční tabulka** (R – počet kategorií jedné proměnné, C – počet kategorií druhé proměnné).
- Speciální případ: 2 × 2 tabulka = čtyřpolní tabulka.

gen \ †	Ano	Ne	Σ
Ano	20	82	102
Ne	10	54	64
Σ	30	136	166

Kontingenční tabulky – hypotézy

- **Nezávislost** (Pearsonův chí-kvadrát test)
  - Jeden výběr, dvě charakteristiky – obdoba nepárového uspořádání
  - Př.: pacienti s AD – pohlaví × vzdělání (VŠ, SŠ, ZŠ)
- **Shoda struktury** (Pearsonův chí-kvadrát test)
  - Více výběrů, jedna charakteristika – obdoba nepárového uspořádání
  - Př.: pacienti s AD v několika nemocnicích × věková struktura
- **Symetrie** (McNemarův test)
  - Jeden výběr, opakovaně jedna charakteristika – obdoba párového uspořádání
  - Př.: Známky z testu A a z testu B (Jsou testy stejně obtížné?)

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

POZOROVANÉ ČETNOSTI

X \ Y	1	2	Σ
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
Σ	$n_{.1}$	$n_{.2}$	n

OČEKÁVANÉ ČETNOSTI

X \ Y	1	2	Σ
1	e_{11}	e_{12}	$e_{1.}$
2	e_{21}	e_{22}	$e_{2.}$
Σ	$e_{.1}$	$e_{.2}$	e

Pearsonův chí-kvadrát test

- Výpočet testové statistiky:

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

- Nulovou hypotézu o nezávislosti dvou kategoriálních proměnných zamítáme na hladině významnosti α, když

$$\mathrm{X}^2 \geq \chi^2_{(1-\alpha)}(r-1)\left(c-1\right)$$

- *r – počet řádků*
  c – počet sloupců

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!

Příklad I. nezávislost

- Máme 74 pacientů s krevní skupinou A0. Naším cílem je zjistit, zda u těchto pacientů je věk nezávislý na přítomnosti sledovaného onemocnění.

Pozorované četnosti

A0	<25	25-40	>40	Celkem
nemoc	9	16	6	31
zdravý	13	25	5	43
Celkem	22	41	11	74

Očekávané četnosti

A0	<25	25-40	>40	Celkem
nemoc	9,22	17,18	4,61	-
zdravý	12,78	23,82	6,39	-
Celkem	-	-	-	-

$$X^2 = \frac{(9-9.22)^2}{9.22} + \frac{(16-17.18)^2}{17.18} + \cdots + \frac{(5-6.39)^2}{6.39} \cong 0.87$$

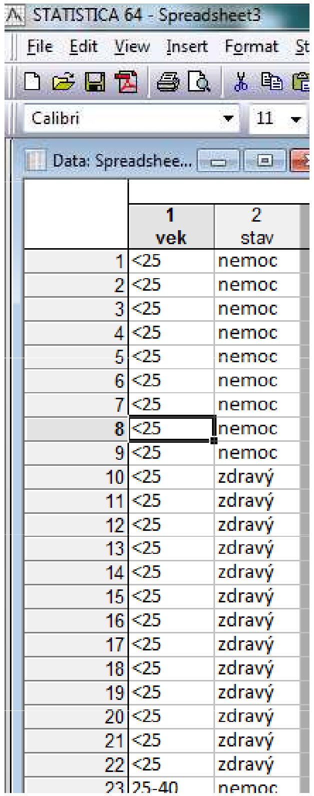
$$\chi^2_{(1-0,05)}(2) \cong 5,99$$

Nezamítáme nulovou hypotézu.
Tedy u těchto pacientů je věk nezávislý na daném onemocnění.

Řešení STATISTICA

- 2 možné vstupy

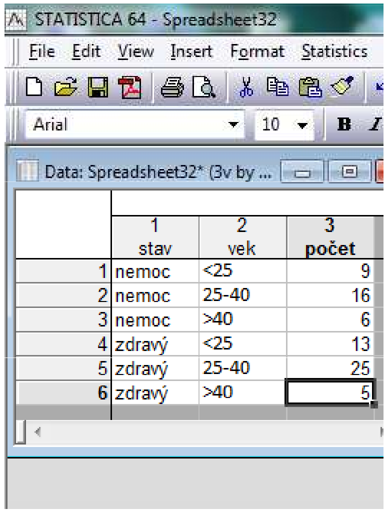
I



	1 vek	2 stav
1	<25	nemoc
2	<25	nemoc
3	<25	nemoc
4	<25	nemoc
5	<25	nemoc
6	<25	nemoc
7	<25	nemoc
8	<25	nemoc
9	<25	nemoc
10	<25	zdravý
11	<25	zdravý
12	<25	zdravý
13	<25	zdravý
14	<25	zdravý
15	<25	zdravý
16	<25	zdravý
17	<25	zdravý
18	<25	zdravý
19	<25	zdravý
20	<25	zdravý
21	<25	zdravý
22	<25	zdravý
23	25-40	nemoc

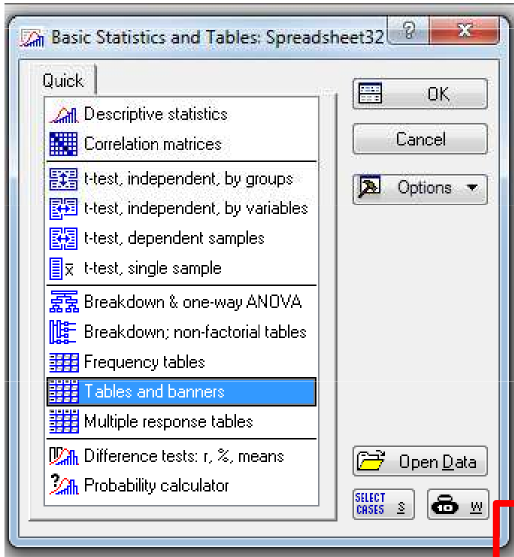
každý případ z tabulky odpovídá jednomu řádku

II

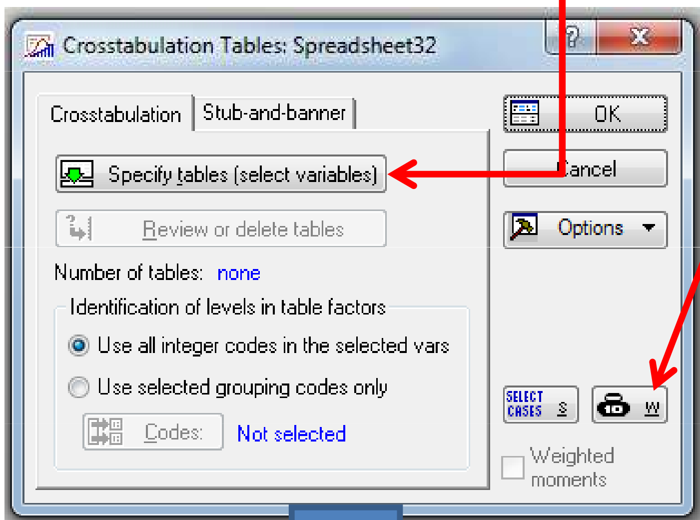


	1 stav	2 vek	3 počet
1	nemoc	<25	9
2	nemoc	25-40	16
3	nemoc	>40	6
4	zdravý	<25	13
5	zdravý	25-40	25
6	zdravý	>40	5

každá kombinace má přiřazený počet výskytů

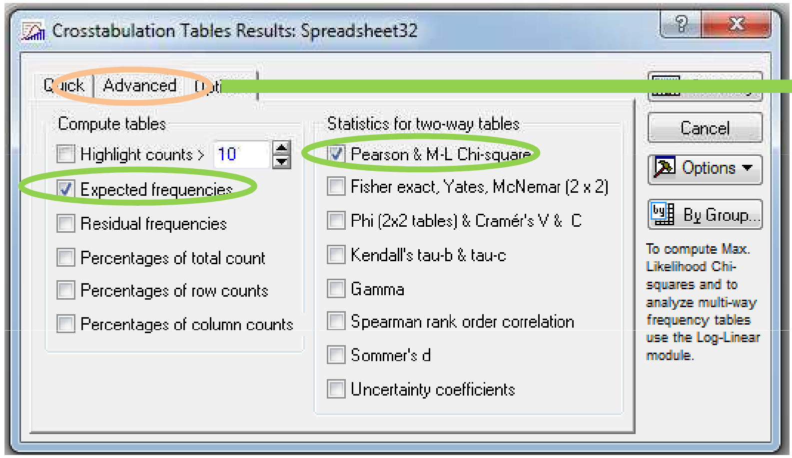


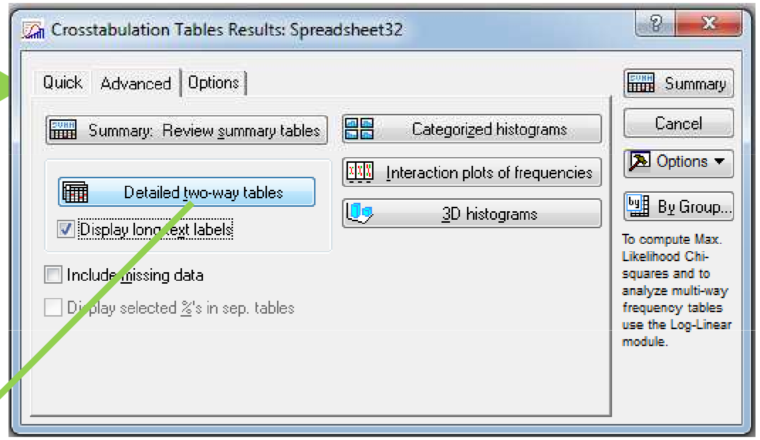
zvolit proměnné (vek a stav)



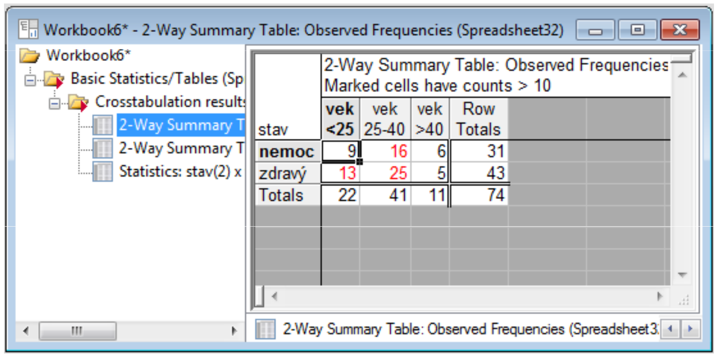
V případě, že je vstupem tabulka II,pak je třeba jako váhu nastavit počet výskytů dané kombinace kategorií

Řešení STATISTICA





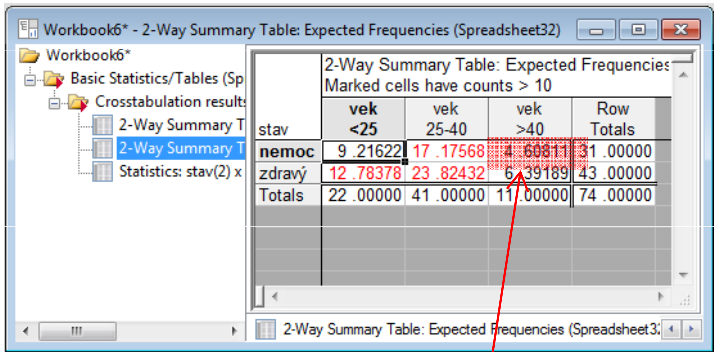
1.tabulka – pozorované četnosti



2-Way Summary Table: Observed Frequencies
Marked cells have counts > 10

stav	vek <25	vek 25-40	vek >40	Row Totals
nemoc	9	16	6	31
zdravý	13	25	5	43
Totals	22	41	11	74

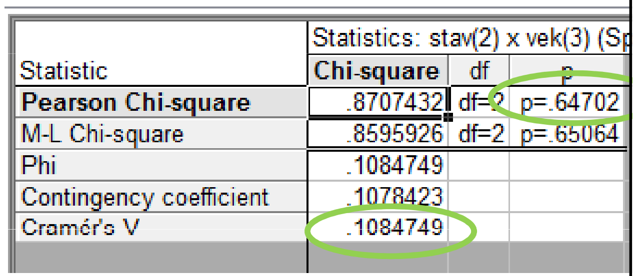
2.tabulka – očekávané četnosti



2-Way Summary Table: Expected Frequencies
Marked cells have counts > 10

stav	vek <25	vek 25-40	vek >40	Row Totals
nemoc	9.21622	17.17568	4.60811	31.00000
zdravý	12.78378	23.82432	6.39189	43.00000
Totals	22.00000	41.00000	11.00000	74.00000

3.tabulka – statistiky



Statistics: stav(2) x vek(3) (Sp

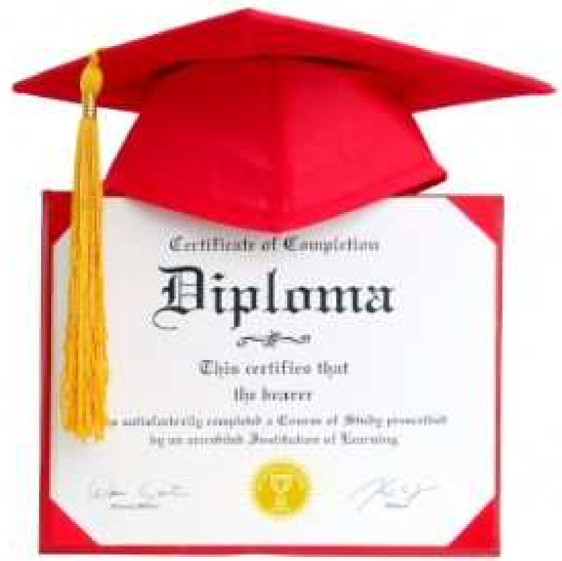
Statistic	Chi-square	df	p
Pearson Chi-square	.8707432	df=2	p=.64702
M-L Chi-square	.8595926	df=2	p=.65064
Phi	.1084749		
Contingency coefficient	.1078423		
Cramér's V	.1084749		

ZKONTROLOVAT PŘEDPOKLADY !!!

Příklad II. nezávislost

- Na hladině významnosti 0,05 testujte hypotézu o nezávislosti pedagogické hodnosti a pohlaví

pohlaví	pedagogická hodnost		
	odb. asistent	docent	profesor
muž	32	15	8
žena	34	8	3



[X^2 =3,5 ; p=0,1739]

Příklad III.

- Máme data z dotazníku od 100 mužů, kteří odpovídali na to, jaké sporty dělají. Zjistěte, zda se preference pro americký fotbal dají srovnávat s preferencemi baseballu.

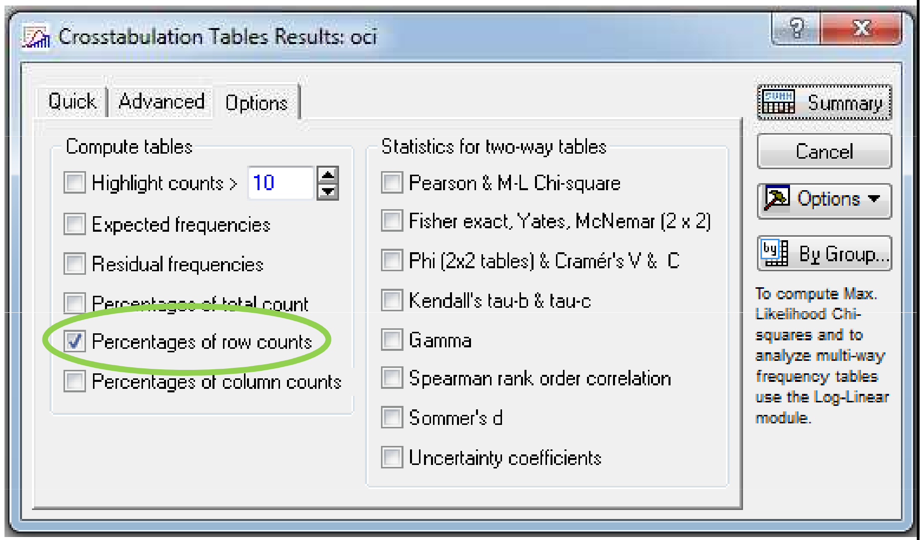
[nejsou splněny podmínky dobré aproximace – posloučování (Always + Usually), (Sometimes+Never) -> X^2 =33,15 ; p<0.001]

Příklad I. shoda struktury

- Máme údaje o barvě očí u třech ročníku. Rozhodněte, zda se na hladině významnosti 0,05 liší procentuální zastoupení jednotlivých barev očí v těchto ročnících.

	modre	hnede	zelene
1 rocnik	15	25	15
2 rocnik	10	18	15
3 rocnik	10	26	16

vizualizace procentualni zastoupeni:



[X^2 =1,54 ; p=0,819]

Příklad II. shoda struktury

- Zjistěte, zda se liší intenzita kouření s rozdílným postavením ve firmě…

	Nekuřák	Lehký	Střední	Těžký
Sr. Manager	8	4	6	4
Jr. Manager	8	6	14	8
Sr. Empl	25	10	12	4
Jr. Empl	18	24	33	13
Secretar	10	6	7	2

[X^2 =18,88 ; p=0,092]

McNemarův test – test symetrie

- obdoba párového testu – pouze čtyřpolní tabulky
- Testová statistika pro čtyřpolní tabulku:

$$\mathrm{X}^2 = \frac{(b-c)^2}{b+c}$$

Veličina *X*	Veličina *Y*		
	Y = 1	*Y* = 2	Celkem
X = 1	*a*	*b*	*a* + *b*
X = 2	*c*	*d*	*c* + *d*
Celkem	*a* + *c*	*b* + *d*	*n*

- Zaměřuje se pouze na pozorování, u kterých jsme při opakovaném měření zaznamenali rozdílné výsledky – za platnosti H_0 by jejich četnosti (označeny b a c) měly být stejné.

- Testová statistika pro obecnou čtvercovou kontingenční tabulku:

$$\mathrm{X}^2 = \sum_{i<j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

Příklad I.

- Liší se tlak před podáním a po podání léku?

		po	
		v normě	zvýšený
před	v normě	15	8
	zvýšený	16	23

$$X^2 = \frac{(8-16)^2}{8+16} \cong 2{,}67$$

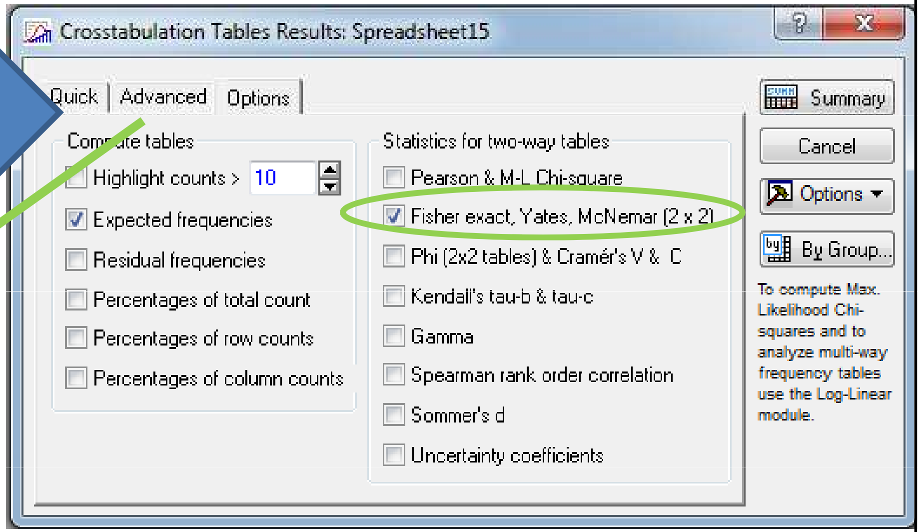
$$\chi^2_{(1-0{,}05)}(1) \cong 3{,}84$$

Výpočet ve STATISTICE

Data: Spreadsheet15* (3v by 1...

	1 Var1	2 Var2	3 Var3
1	norma	norma	15
2	norma	zvyseny	8
3	zvyseny	norma	16
4	zvyseny	zvyseny	23
5			
6			
7			
8			
9			
10			

stejně jako obyčejný chí test,
pouze v možnostech
zaškrtneme McNemara



Statistic	Statistics: Var1(2) x Var2(2) (Chi-square	df	p
Pearson Chi-square	3.386845	df=1	p=.06572
M-L Chi-square	3.427852	df=1	p=.06411
Yates Chi-square	2.488294	df=1	p=.11470
Fisher exact, one-tailed			p=.05691
two-tailed			p=.11382
McNemar Chi-square (A/D)	1.289474	df=1	p=.25614
(B/C)	2.041667	df=1	p=.15304

STATISTICA používá korekci pro nespojitost,
proto trochu jiný výsledek …

Příklad II.

- Máme zjistit, zda požití alkoholu ovlivňuje schopnost řidičů projet nějakou trasu.

Před požitím alkoholu	Po požití alkoholu		$n_{j.}$
	bezchybně	s chybami	
bezchybně	45	35	80
s chybami	15	5	20
$n_{.k}$	60	40	100

[$X^2 = 8$; $p \approx 0{,}005$]

Příklad III.

- Zjišťování přítomnosti onemocnění před a po provedení léčby

Disease	After: present	After: absent	Row total
Before: present	101	121	222
Before: absent	59	33	92
Column total	160	154	314

[Chi-square = 20.67, p<0.001]

Malý počet

- Pokud čtyřpolní tabulka – Fisherův exaktní test

- Jinak pokusit se kategorie smysluplně poslučovat

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
NÚ I	ne	6	4

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

0	5
8	2

1	4
7	3

2	3
6	4

3	2
5	5

4	1
4	6

5	0
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**

Příklad I. – Fisherův exaktní test

- V náhodném výběru 50 obézních dětí byla zjišťována obezita rodičů. X – obezita matky, Y-obezita otce. Na hladině významnosti 0,05 ověřte, zda lze zamítnout hypotézu o nezávislosti veličin X a Y.

X	Y		$n_{j.}$
	ano	ne	
ano	15	9	24
ne	7	19	26
$n_{.k}$	22	28	50

Výpočet pomocí systému STATISTICA:

Vytvoříme datový soubor o třech proměnných X, Y (varianty 0 – neobézní, 1 – obézní) a četnost a čtyřech případech:

	1 X	2 Y	3 četnost
1	obézní	obézní	15
2	obézní	neobézní	9
3	neobézní	obézní	7
4	neobézní	neobézní	19

Statistiky – Základní statistiky/tabulky – OK – Specif. Tabulky – List 1 X, List 2 Y – OK, zapneme proměnnou vah četnost – OK, Výpočet – na záložce Možnosti zaškrtneme Fisher exakt., Yates, McNemar (2x2). Dostaneme výstupní tabulku:

	Statist. : X(2) x Y(2) (obezita rodicu)		
Statist.	Chí-kvadr.	sv	p
Pearsonův chí-kv.	6,410777	df=1	p=,01134
M-V chí-kvadr.	6,548348	df=1	p=,01050
Yatesův chí-kv.	5,048207	df=1	p=,02465
Fisherův přesný, 1-str.			p=,01188
2-stranný			p=,02163
McNemarův chí-kv. (A/D)	,2647059	df=1	p=,60691
(B/C)	,0625000	df=1	p=,80259

Vidíme, že p-hodnota pro Fisherův exaktní oboustranný test je 0,02163, tedy na hladině významnosti 0,05 zamítáme hypotézu, že obezita matky a otce spolu nesouvisí.