

Plánování Průmyslových Experimentů

Část 2. Metody vyhodnocování experimentů

- Vyhodnocení experimentu: stanovení charakteristik procesu a jeho optimalizace, vyhodnocení vlastností materiálů zpravidla metodami matematické statistiky

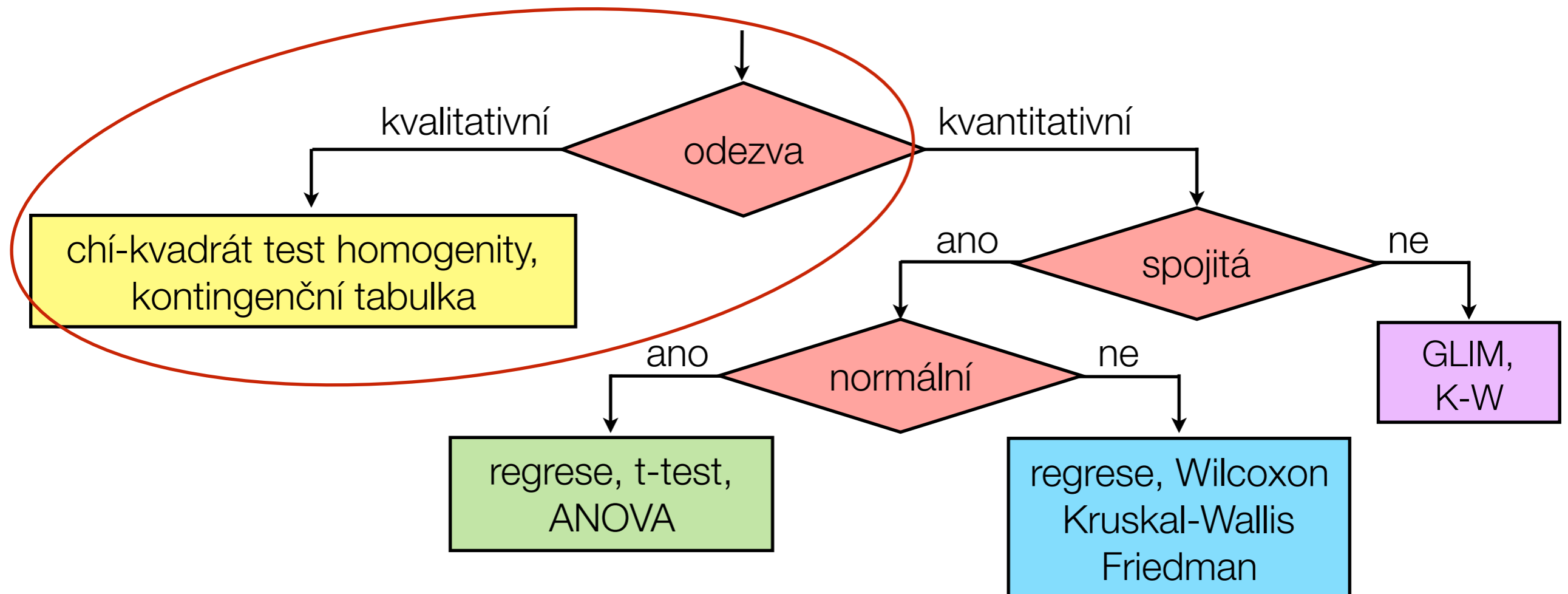
Klíčové pojmy: Testování hypotéz, odhady parametrů, regrese, korelace



2. Vyhodnocení experimentu

- Vyhodnocení experimentu: stanovení charakteristik procesu a jeho optimalizace, vyhodnocení vlastností materiálů zpravidla metodami matematické statistiky

Klíčové pojmy: Testování hypotéz, odhady parametrů, regrese, korelace



2.2. Vsuvka 1: Testování hypotéz

Máme naměřeno n replikací nějaké náhodné veličiny: X_1, X_2, \dots, X_n

$$H_0 : \mu = m_0$$

- Nulová hypotéza

$$H_A : \mu \neq m_0$$

- Alternativní hypotéza

$$T = F(X_1, X_2, \dots, X_n)$$

- Testová statistika

- má zpravidla nějaké známé rozdělení

$$\alpha \Rightarrow t_\alpha(n-1)$$

- hladina významnosti, kritická hodnota

$$|T| \geq t_\alpha(n-1)$$

- rozhodovací pravidlo

- Testy o hodnotě nějaké charakteristiky náhodné veličiny
- Testy o shodě charakteristik dvou či více souborů
- Testy o rozdělení náhodné veličiny (testy normality)

Nulová hypotéza se nepřijímá, pouze zamítá či nezamítá!

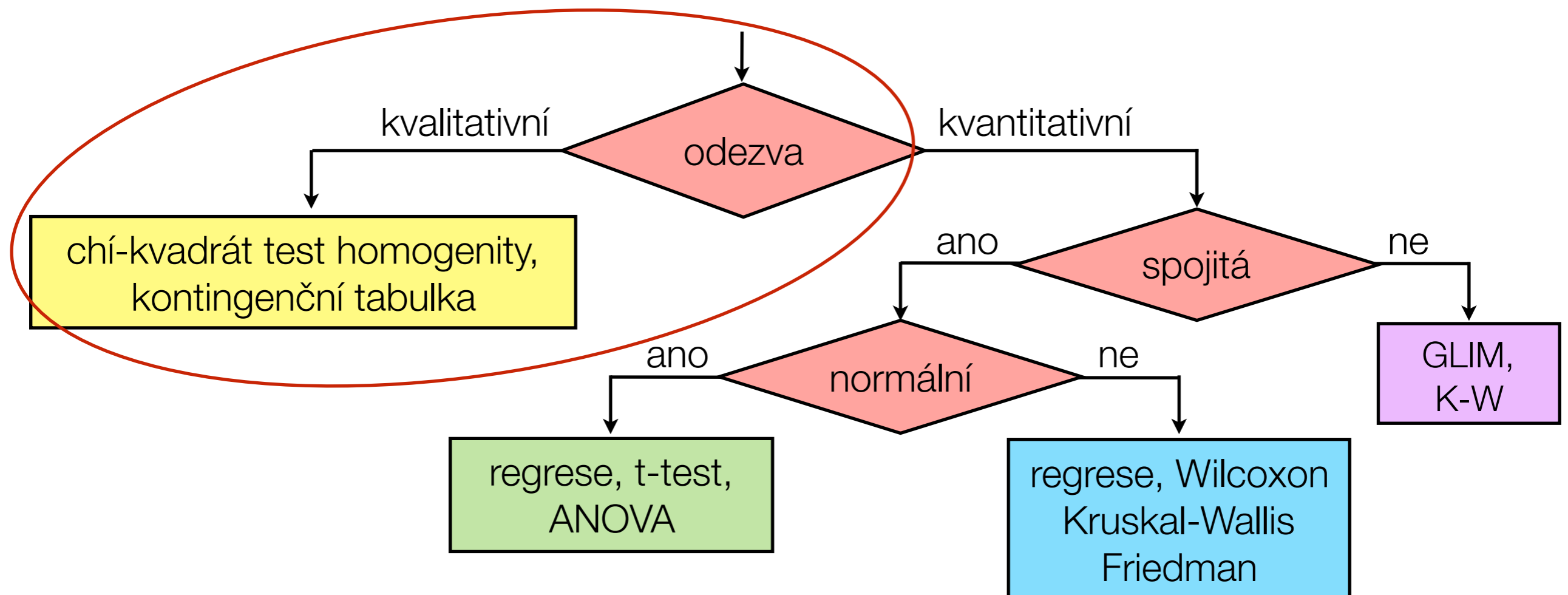
Statisticky nelze nic dokázat, pouze zamítnout!



2. Vyhodnocení experimentu



- Vyhodnocení experimentu: stanovení charakteristik procesu a jeho optimalizace, vyhodnocení vlastností materiálů zpravidla metodami matematické statistiky

Klíčové pojmy: Testování hypotéz, odhady parametrů, regrese, korelace



2.1. Kvalitativní odezva

Příklad 1: Ovlivňuje barva očí Rh faktor?

- **Problém:** Při výzkumu vlivu Rh faktoru chceme zjistit, zda barva očí ovlivňuje Rh faktor.
- **Odezva:** Rh faktor může nabývat dvou kvalitativních hodnot: +, -
- **Faktor:** Barva očí - budeme uvažovat dvě kvalitativní úrovně: modrá, hnědá
- Budeme testovat hypotézu, že barva očí neovlivňuje Rh faktor: $H_0: P(\text{Rh}|B)=P(\text{Rh})$
- Provedeme celkem 400 pozorování, 100 u modrookých, 300 u hnědookých pacientů.
- Výsledky uspořádáme do kontingenční tabulky: 
- **K vyhodnocení použijeme test nezávislosti.**
- Za předpokladu nezávislosti by (podle marginálních součtů) mělo platit: 

barva očí	Rh ⁺	Rh ⁻	součet
modrá	35	65	100
hnědá	94	206	300
součet	129	271	400

zjištěné četnosti kombinací

barva očí	Rh ⁺	Rh ⁻	součet
modrá	32,25	67,75	100
hnědá	96,75	203,25	300
součet	129	271	400

odhadnuté četnosti kombinací

Můžeme říci, že jsou tyto tabulky stejné?

Můžeme alespoň říci, že se tyto tabulky statisticky významně neliší?

Pokud ano, potom můžeme prohlásit, že na nějaké (dané) hladině významnosti α lze považovat Rh faktor nezávislý na barvě očí.



2.2.1. Vyhodnocení kvalitativní odezvy

Kontingenční tabulka

odezva má kvalitativní charakter: může nabývat r hodnot

jeden kvalitativní faktor: může nabývat s hodnot, u nichž provádíme $N > r \cdot s$ pozorování
výsledky zapisujeme do tabulky

Omezení:
očekávané četnosti
musejí být větší než
5!

Testování v kontingenční tabulce:

- test hypotézy o nezávislosti znaků (test homogeneity)
- test symetrie

Test nezávislosti: testová statistika = $\frac{\sum(\text{pozorované} - \text{očekávané})^2}{\text{očekávané}}$

označme: n_{ij} absolutní četnost v řádku i a sloupci j napozorovaná v experimentu

m_{ij} očekávaná četnost v řádku i a sloupci j za platnosti hypotézy

$$m_{ij} = \frac{R_i S_j}{N}, \text{ kde } \begin{array}{l} R_i = \text{součet četností v řádku } i \\ S_j = \text{součet četností ve sloupci } j \end{array}$$



$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

testová statistika má chí-kvadrát rozdělení o $(r-1) \times (s-1)$ stupňů volnosti



2.1. Kvalitativní odezva

Příklad 1: Ovlivňuje barva očí Rh faktor?

- Problém: Při výzkumu vlivu Rh faktoru chceme zjistit, zda barva očí ovlivňuje Rh faktor.
- Odezva: Rh faktor může nabývat dvou kvalitativních hodnot: +, -
- Faktor: Barva očí - budeme uvažovat dvě kvalitativní úrovně: modrá, hnědá
- Budeme testovat hypotézu, že barva očí neovlivňuje Rh faktor: $H_0: P(\text{Rh}|B)=P(\text{Rh})$
- Provedeme celkem 400 pozorování, 100 u modrookých, 300 u hnědookých pacientů.
- Výsledky uspořádáme do kontingenční tabulky: 
- K vyhodnocení použijeme test nezávislosti.
- Za předpokladu nezávislosti by (podle marginálních součtů) mělo platit: 

barva očí	Rh ⁺	Rh ⁻	součet
modrá	35	65	100
hnědá	94	206	300
součet	129	271	400

zjištěné četnosti kombinací

barva očí	Rh ⁺	Rh ⁻	součet
modrá	32,25	67,75	100
hnědá	96,75	203,25	300
součet	129	271	400

odhadnuté četnosti kombinací

Sestavíme testovou statistiku $\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$

$$\chi^2 = \frac{(35 - 32.25)^2}{32.25} + \frac{(65 - 67.75)^2}{67.75} + \frac{(94 - 96.75)^2}{96.75} + \frac{(206 - 203.25)^2}{203.25} \doteq 0.2289$$

$\chi_1^2(0.05) = 3.84 \Rightarrow$ Na hladině významnosti 5% nebyla prokázána závislost mezi barvou očí a Rh faktorem.



2.1. Kvalitativní odezva

Příklad 2: Ovlivňuje složení krmiva schopnost otelení krav?

- **Problém:** Chceme zjistit, zda ovlivňuje množství proteinů a energie v krmivu schopnost otelení krav
- **Odezva:** O = dojde k otelení (ano) nebo nedojde (ne) - kvalitativní: $O \in \{+, -\}$
- **Faktory:** Množství proteinů v krmivu R - kvantitativní faktor, volíme dvě úrovně, Množství energie v krmivu E - kvantitativní faktor, volíme dvě úrovně;
- Budeme testovat hypotézu, že schopnost otelení nezávisí na složení krmiva:
 $H_0: P(O|R,E)=P(O)$
- Dostáváme celkem $2 \times 2 = 4$ kombinace nastavení faktorů, pro každou z nich provedeme pozorování celkem 100 zvířat
- Výsledky uspořádáme do kontingenční tabulky:
- **K vyhodnocení použijeme test nezávislosti.**

Pro celou tabulku: testová statistika má $df = 3 - 1 = 3$ (stupně volnosti)

$$\chi^2 = 58,549, \quad \chi_{0,01}^2(3) = 11,3$$

kombinace	ano	ne
vysoká energie a vysoký protein	81	19
vysoká energie a nízký protein	88	12
nízká energie a vysoký protein	75	25
nízká energie a nízký protein	43	57



2.1. Kvalitativní odezva

Příklad 2: Ovlivňuje složení krmiva schopnost otelení krav?

Pro celou tabulku: testová statistika má $df = 3 \cdot 1 = 3$ (stupně volnosti)

$$\chi^2 = 58,549, \quad \chi_{0,01}^2(3) = 11,3$$

kombinace	ano	ne
vysoká energie a vysoký protein	81	19
vysoká energie a nízký protein	88	12
nízká energie a vysoký protein	75	25
nízká energie a nízký protein	43	57

Sloučíme-li řádky 1 + 2 (vysoká energie) a 3 + 4 (nízká energie), dostaneme tabulku 2x2 s $df = 1$ a testovou statistiku (efekt energie) $\chi_{en}^2 = 32,080, \quad \chi_{0,01}^2(1) = 6,63$

Sloučíme-li řádky 1 + 3 (vysoký protein) a 2 + 4 (nízký protein), dostaneme tabulku 2x2 s $df = 1$ a testovou statistiku (efekt proteinu) $\chi_{prot}^2 = 7,709, \quad \chi_{0,01}^2(1) = 6,63$

Odečteme-li hodnoty chí-kvadrát energie a proteinu od celkového chí-kvadrátu, dostaneme efekt interakce $\chi_{en.prot}^2 = 18,760.$



2.1. Kvalitativní odezva

Kontingenční tabulka

Test symetrie:

hypotéza: $\frac{n_{ij}}{N} = \frac{n_{ji}}{N}$

testová statistika
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^i \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}$$

má chí-kvadrát rozdělení o $r(r-1)$ stupňů volnosti

Příklad: Dědí syn barvu očí otce?

Bylo provedeno 1000 pozorování, jejichž výsledky jsou v tabulce napravo. Barva očí je zakódována: 1=sv.modrá, 2=modrozelená, 3=tm.šedá nebo sv.hnědá, 4=tm.hnědá

		barva očí syna				Σ
		1	2	3	4	
barva očí otce	1	194	70	41	30	335
	2	83	124	41	36	284
	3	25	34	55	23	137
	4	56	36	43	109	224
	Σ	358	264	180	198	1000

Dosazením do testové statistiky dostaneme hodnotu $\chi^2 = 19,56$

Kritická hodnota pro 6 stupňů volnosti a pro $\alpha=5\%$ je $\chi_6^2(0,05) = 12,59$

=> Na hladině významnosti 5% nebyla prokázána shoda barvy očí otce a syna.