

20. Kontingenční tabulky

20.1 Úvodní informace

V praxi je velmi častá situace, kdy vyšetřujeme najednou dva statistické znaky, které jsou svou povahou diskrétní kvantitativní (mají přesně stanovený konečný počet všech možností); spojité kvantitativní, ale s hodnotami sloučenými do skupin nebo kvalitativní. Pro účely této kapitoly není podstatné, zda jde o znaky typu kvantitativního nebo kvalitativního. Takováto situace vzniká například tehdy, když na jednom objektu sledujeme najednou dva znaky. Jestliže je například základním souborem množina studentů, potom jeden znak může být pohlaví (2 možnosti), druhý znak znalosti z matematiky (tyto znalosti popisujeme slovně – nicméně tento popis je typu ordinárního).

Pro další potřebu popíšeme naši situaci následovně: Máme k dispozici celkem výběr o n prvcích. Předpokládejme, že zkoumáme tento výběr pomocí statistického znaku X , který může nabývat hodnot $1, 2, \dots, c$ a statistického znaku Y , který může nabývat hodnot $1, 2, \dots, r$. Označíme

$$P_{ij} = P(X=i; Y=j) \quad (20.1)$$

Dále označíme

$$p_i = P(X=i) = \sum_{j=1}^r p_{ij}, \quad p_j = \sum_{i=1}^c p_{ij} \quad (20.2)$$

Výše uvedený výběr lze popsat multinomickým rozdělením (viz kapitola o chi – kvadrát metodě) o $r \times c$ třídách tvořených dvojicemi (i, j) . Označíme – li n_{ij} jako počet těch případů, kdy v našem výběru je $X = i$ a zároveň $Y = j$, můžeme výsledky výběru shrnout do níže uvedené tabulky:

Y	X				celkem
	1	2	...	c	
1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
...	...	...	...	...	...
r	n_{r1}	n_{r2}	...	n_{rc}	$n_{r.}$
celkem	$n_{.1}$	$n_{.2}$	...	$n_{.c}$	n

Tabulka 1 - Rozvržení statistických znaků X, Y

V poslední řádce a posledním sloupci jsou uvedeny následující hodnoty

$$n_{i.} = \sum_{j=1}^r n_{ij}, \quad n_{.j} = \sum_{i=1}^c n_{ij}, \quad n = \sum_{i=1}^c n_{i.} = \sum_{j=1}^r n_{.j} = \sum_{i=1}^c \sum_{j=1}^r n_{ij} \quad (20.3)$$

výše uvedené hodnoty $n_{i.}$ a $n_{.j}$ se nazývají **marginální četnosti**. Pro další práci není podstatné, že máme jednotlivé třídy označeny čísly. Tato čísla mají jen pomocný charakter. V obecném případě mohou mít jednotlivé třídy běžné názvy, značení čísly podržíme z důvodů větší přehlednosti.

Při práci s kontingenčními tabulkami se velmi často setkáváme s problémy, kdy je třeba rozhodnout o nezávislosti obou statistických znaků, o případné shodnosti struktury

obou znaků (homogenitě) nebo o symetrii struktury u čtvercových kontingenčních tabulek. Právě tyto základní problémy budeme řešit v následujících jednotlivých částech.

20.2 Test o nezávislosti jednotlivých statistických znaků

Jednou z nejčastějších úloh je provedení testu hypotézy , že náhodné veličiny X a Y jsou na sobě nezávislé. Základním tvrzením , které budeme dále využívat je následující.

Věta 20.1

Veličiny X a Y jsou nezávislé právě , když platí $p_{ij} = p_{i.} \cdot p_{.j}$, pro všechny přípustné hodnoty i, j .

Důkaz :

Ponechávám čtenáři.

Znamená to , že hypotéza o nezávislosti je tedy ekvivalentní s hypotézou:

$$H_0 : p_{ij} = p_{i.} \cdot p_{.j} , i = 1, \dots, c ; j = 1, \dots, r \quad (20.4)$$

Celkový počet neznámých parametrů není tedy $r \times c$, ale je menší než toto číslo , je roven $r + c$. Ovšem jednotlivé pravděpodobnosti $p_{i.}$; $p_{.j}$ nejsou nezávislé viz (20.2) .Je tedy celkový počet neznámých parametrů roven $r + c - 2$, protože jeden z příslušných neznámých parametrů lze ze vztahu (20.2) spočítat! Přitom se samozřejmě omezujeme na případy , že hodnota marginálních pravděpodobností je kladná , kdyby tomu tak nebylo , mohli bychom takové řádky nebo sloupce vynechat.

Za předpokladu platnosti hypotézy H_0 je možno odvodit , že mezi jednotlivými marginálními četnostmi a marginálními pravděpodobnostmi platí následující vztahy

$$\begin{aligned} \frac{n_{i.}}{p_{i.}} - \frac{n_{c.}}{p_{c.}} &= 0, \quad i = 1, \dots, c \\ \frac{n_{.j}}{p_{.j}} - \frac{n_{.r}}{p_{.r}} &= 0, \quad j = 1, \dots, r \end{aligned} \quad (20.5)$$

Z těchto vztahů je možno odvodit , že

$$n = \frac{n_{c.}}{p_{c.}}, \quad n = \frac{n_{.r}}{p_{.r}} \quad (20.6)$$

Odtud získáme odhad pro $p_{c.}$ a $p_{.r}$

$$\hat{p}_{c.} = \frac{n_{c.}}{n} \text{ a } \hat{p}_{.r} = \frac{n_{.r}}{n} \quad (20.7)$$

Z odhadů (20.7) získáme dosazením do (20.5) odhady pro jednotlivé pravděpodobnosti

$$\begin{aligned} \hat{p}_{i.} &= \frac{n_{i.}}{n}, \quad i = 1, \dots, c \\ \hat{p}_{.j} &= \frac{n_{.j}}{n}, \quad j = 1, \dots, r \end{aligned} \quad (20.8)$$

Podle kapitoly o testech typu chí – kvadrát , je následující náhodná veličina

$$\chi^2 = \sum \frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}} = \sum_{i=1}^c \sum_{j=1}^r \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n} \right)^2}{\frac{n_{i.} \cdot n_{.j}}{n}} \quad (20.9)$$

asymptoticky rozdělení χ^2 , jehož počet stupňů volnosti je roven $r \times c - (r+c-2) - 1 = r \times c - r - c - 1 = (r-1) \times (c-1)$.

Tedy závěrem : Při zjišťování nezávislosti dvou statistických znaků (náhodných veličin) používáme χ^2 - test , jestliže je hodnota testové veličiny $\chi^2 \geq \chi^2_{(r-1),(c-1)}(\alpha)$, zamítáme hypotézu H_0 o nezávislosti náhodných veličin X a Y . Protože tento proces funguje asymptoticky , požadujeme podobně jako v klasickém χ^2 - testu , aby všechny očekávané četnosti (teoretické četnosti) $\frac{n_{i.} \cdot n_{.j}}{n}$ byly větší než **5**. Není – li tato podmínka splněna , spojujeme řádky a sloupce tak , abychom tuto podmínku splnili.

Příklad 20.2

U 6800 mužů byla zjišťována barva očí a barva vlasů (viz Yule a Kendall 1950) . Výsledky jsou uvedeny v následující tabulce:

Barva očí	Barva vlasů				celkem
	Světlá	Kaštanová	Černá	Zrzavá	
Světle modrá	1768	807	189	47	2811
Šedá nebo zelená	946	1387	746	53	3132
Tmavohnědá	115	438	288	16	857
Celkem	2829	2632	1223	116	6800

Tabulka 2 - Barva očí a vlasů

Zjistěte , zda barva očí a vlasů u mužů jsou nezávislé statistické znaky!

Řešení:

K tabulce 2 vytvoříme tabulku teoretických (očekávaných četností):

Barva očí	Barva vlasů				celkem
	Světlá	Kaštanová	Černá	Zrzavá	
Světle modrá	1169,459	1088,0224	505,5666	47,95235	2811
Šedá nebo zelená	1303,004	1212,2682	563,2994	53,42824	3132
Tmavohnědá	356,5372	331,70941	154,134	14,61941	857
Celkem	2829	2632	1223	116	6800

Tabulka 3 - Očekávané četnosti

Protože žádná z očekávaných četností není menší než 5 , můžeme přistoupit k výpočtu jednotlivých členů v součtu (20.9) . Pro jednoduchost je opět uvedeme v tabulce:

Barva očí	Barva vlasů			
	Svělá	Kaštanová	Černá	Zrzavá
Světle modrá	306,33978	72,584504	198,222	0,018914
Šedá nebo zelená	97,81392	25,185177	59,25713	0,003432
Tmavohnědá	163,63011	34,058995	116,2632	0,130376

$$\chi^2 = 1073,5076$$

Tabulka 4 - Vypočtené hodnoty testové statistiky

Hodnota testové statistiky $\chi^2=1073,507564$. Hodnota kritická je rovna 12,591577. Protože hodnota testové statistiky leží v kritickém oboru zamítáme hypotézu H_0 o nezávislosti obou statistických znaků.

Příklad 20.3

V náhodném výběru 50 studentů byl zjišťován vztah mezi známkou ze statistiky a znalostí práce na počítači. Výsledky jsou opět pro přehlednost uvedeny v následující tabulce:

		Známka ze statistiky				celkem
		1	2	3	4	
Ovládání výpočetní techniky	vynikající	7	3	0	1	11
	průměrné	2	5	6	2	15
	podprůměrné	1	3	7	0	11
	žádné	1	0	4	8	13
	Celkem	11	11	17	11	50

Tabulka 5 - Data pro příklad 20.3

Řešení:

Opět podobně jako v předchozím případě vytvoříme postupně dvě tabulky s pomocnými výpočty:

		Známka ze statistiky				celkem
		1	2	3	4	
Ovládání výpočetní techniky	vynikající	2,42	2,42	3,74	2,42	11
	průměrné	3,3	3,3	5,1	3,3	15
	podprůměrné	2,42	2,42	3,74	2,42	11
	žádné	2,86	2,86	4,42	2,86	13
	Celkem	11	11	17	11	50

Tabulka 6 - Očekávané četnosti v tabulce 5

Poslední tabulkou, kterou vytvoříme pro tento příklad je tabulka hodnot, které se objevují v součtu (20.9). Tentokrát ale je zřejmé, že hodnoty očekávané jsou menší než 5, proto budeme některé sloupce a řádky sjednocovat.

Nejdříve jsme sloučili jen hodnoty sloupců 1,2 a dále hodnoty sloupců 3,4. Tím jsme získali níže uvedenou tabulku .

		Známka ze statistiky		celkem
		1 - 2	3 - 4	
Ovládání výpočetní techniky	vynikající	4,84	6,16	11
	průměrné	6,6	8,4	15
	podprůměrné	4,84	6,16	11
	žádné	5,72	7,28	13
Celkem		22	28	50

Tabulka 7 - Sloučení dvou sloupců

Protože však hodnoty očekávaných četností v této tabulce 7 i nadále v některých buňkách nepřesahují 5 , musíme ještě provést sloučení řádek – konkrétně řádky vynikající a průměrné , a dále řádek podprůměrné a žádné. Získáme následující tabulku:

		Známka ze statistiky		celkem
		1 - 2	3 - 4	
Ovládání výpočetní techniky	vyn - prům	11,44	14,56	26
	podpr - žádné	10,56	13,44	24
Celkem		22	28	50

Tabulka 8 - Maximální sloučení tříd

Nyní již můžeme spočítat hodnotu testové funkce , ještě uvedeme tabulku skutečných četností po sloučení:

		Známka ze statistiky		celkem
		1 - 2	3 - 4	
Ovládání výpočetní techniky	vyn - prům	17	9	26
	podpr - žádné	5	19	24
Celkem		22	28	50

Tabulka 9 - Skutečné četnosti po sloučení

Nyní již můžeme snadno zjistit hodnotu testové statistiky $\chi^2 = 10,05297$, protože hodnota 5% kvantilu rozdělení chí kvadrát o jednom stupni volnosti je roven 3,8415.

20.3 Testy o shodnosti struktury jednotlivých statistických znaků

V mnoha případech se zajímáme o možnost vyšetřovat shodnost struktury jednoho statistického znaku (například X) za různých podmínek , které jsou vyjádřeny třídami (kategoriemi) druhého statistického znaku Y. Například vyšetřujeme soubor mužů a žen v specifických podmínkách a studujeme možnost stejné reakce u jednoho statistického znaku. I v tomto případě lze dokázat , že výraz (20.9) je za předpokladu hypotézy o shodnosti struktury daného statistického znaku asymptoticky roven rozdělení χ^2 o $(r - 1) \cdot (c - 1)$ stupňů volnosti. Tedy jestliže je hodnota testové veličiny $\chi^2 \geq \chi^2_{(r-1),(c-1)}(\alpha)$, zamítáme hypotézu H_0 o shodnosti struktury náhodné veličiny X.

Příklad 20.4

V průběhu sociologického výzkumu byla dotázána skupina 420 lidí na problematiku životního prostředí. Statistický znak X – věk, byl rozdělen na třídy nedospělí (10 – 18 let), mladí lidé (19 – 34 let), produktivní věk (35 – 60 let), post produktivní věk (> 60 let). Statistický znak význam životního prostředí byl rozčleněn na prioritní, vysoký, průměrný, malý, žádný (tento postoj byl hodnocen dotazníkem a pak přeformulován do této odpovědi). Zjistěte, zda se liší postoje jednotlivých věkových skupin lidí k životnímu prostředí!

		věk				celkem
		nedospělí	mladí lidé	produktivní věk	postproduktivní věk	
význam	prioritní	12	28	26	6	72
	vysoký	18	23	35	26	102
	průměrný	18	15	12	6	51
	malý	2	3	6	9	20
	žádný	0	7	12	21	40
celkem		50	76	91	68	285

Tabulka 10 - Výsledky šetření

Řešení:

Nejdříve jako u příkladů z předchozí části zjistíme očekávané četnosti jednotlivých postojů věkových skupin. Postupovat budeme stejně jako v předchozí části, výsledkem je následující tabulka očekávaných četností:

		věk				celkem
		nedospělí	mladí lidé	produktivní věk	postproduktivní věk	
význam	prioritní	12,6315789	19,2	22,9894737	17,17894737	72
	vysoký	17,8947368	27,2	32,5684211	24,33684211	102
	průměrný	8,94736842	13,6	16,2842105	12,16842105	51
	malý	3,50877193	5,33333333	6,38596491	4,771929825	20
	žádný	7,01754386	10,6666667	12,7719298	9,543859649	40
celkem		50	76	91	68	285

Tabulka 11 - Teoretické - očekávané četnosti

Protože na řádce malý jsou četnosti menší než 5, provedeme sloučení tříd malý a žádný. Výsledná tabulka sloučených četností je uvedena dále:

		věk				celkem
		nedospělí	mladí lidé	produktivní věk	postproduktivní věk	
význam	prioritní	12,6315789	19,2	22,9894737	17,17894737	72
	vysoký	17,8947368	27,2	32,5684211	24,33684211	102
	průměrný	8,94736842	13,6	16,2842105	12,16842105	51
	malý-žádný	10,5263158	16	19,1578947	14,31578947	60
celkem		50	76	91	68	285

Tabulka 12 - Sloučené teoretické četnosti

Z těchto hodnot budeme nyní počítat hodnotu testové statistiky chí – kvadrát, pro naše potřeby uvedeme tabulku jak s hodnotami jednotlivých sčítanců (20.9), tak i s hodnotou testové statistiky a kritickou hodnotou rozdělení chí – kvadrát s 9 stupni volnosti.

		věk			
		nedospělí	mladí lidé	produktivní věk	postproduktivní věk
význam	prioritní	0,03157895	4,03333333	0,39423559	7,274535604
	vysoký	0,0006192	0,64852941	0,18154322	0,113658714
	průměrný	9,15913313	0,14411765	1,12713231	3,126898561
	malý-žádný	6,90631579	25	33,0370156	48,374613

testová statistika= 139,55326

hodnota chi - kvadrát : 16,9189602

Tabulka 13 - Vypočítané hodnoty testové statistiky

Ve speciálním případě čtyřpolní tabulky (tabulka , kde $r = c = 2$) , uvedenou na obrázku níže :

n_{11}	n_{12}	$n_{1.}$
n_{21}	n_{22}	$n_{2.}$
$n_{.1}$	$n_{.2}$	n

Tabulka 14 - Čtyřpolní kontingenční tabulka

Ve čtyřpolní tabulce platí:

$$\chi^2 = n \cdot \frac{(n_{11} \cdot n_{22} - n_{12} \cdot n_{21})^2}{n_{1.} \cdot n_{2.} \cdot n_{.1} \cdot n_{.2}} \quad (20.10)$$

Důkaz si proveděte dosazením hodnot z tabulky 14 do vztahu (20.9). Při případných testech v čtyřpolní tabulce budeme tedy hodnotu testové statistiky počítat podle vztahu (20.10) . Tuto hodnotu budeme porovnávat hodnotou kvantitu α chí – kvadrát rozdělení s 1 stupněm volnosti.

Výraz (20.10) lze ještě dále upravit na následující tvar

$$\chi^2 = \frac{n_{1.} \cdot n_{2.} \cdot n}{n_{.1} \cdot n_{.2}} \cdot \left(\frac{n_{11}}{n_{1.}} - \frac{n_{21}}{n_{.1}} \right)^2 \quad (20.11)$$

Z tohoto vzorce přímo vyplývá vztah χ^2 testu vzhledem k zhodnocení homogenity dvou binomických rozdělení. Platí tedy následující tvrzení

Věta 20.5

Test χ^2 v čtyřpolní tabulce je ekvivalentní s oboustranným testem homogenity dvou binomických rozdělení , který je založen na statistice znormalizování.

Důkaz :

Je zřejmý.

Na datech uspořádaných do čtyřpolní tabulky můžeme proto ověřovat zároveň hypotézu o nezávislosti , tak i hypotézu o homogenitě struktury. Problémy s užitím χ^2 testu nastávají podobně jako v obecném použití tohoto testu , jestliže jsou některé z očekávaných

četností v čtyřpolní tabulce menší než 5 . Navíc se násobí tím , že v případě takovéto tabulky již nemůže provádět slučování jednotlivých tříd. Budeme proto požadovat v případě čtyřpolní tabulky , aby žádná očekávaná četnost nebyla menší než 5 a rozsah výběru byl aspoň 40.

V případě , že nejsou tyto předpoklady splněny , provádíme formálně dva způsoby řešení . První způsob používá korekci na spojitost rozdělení chí – kvadrát , jde o tzv. **Yatesovu korekci**

$$\chi^2 = \frac{n}{n_{1.} \cdot n_{2.} \cdot n_{.1} \cdot n_{.2}} \cdot \left(|n_{11} \cdot n_{22} - n_{21} \cdot n_{12}| - \frac{n}{2} \right)^2 \quad (20.12)$$

Výsledkem této korekce je zmenšení hodnoty testové statistiky. Tím se současně zmenší riziko chyby prvního druhu , ale zároveň se zvětšuje chyba druhého druhu.

Druhou možností je provedení tzv. **Fischerova exaktního testu (faktoriálový test)** , který se používá v případě , že v čtyřpolní tabulce jsou velmi malé četnosti. Hypotézu H_0 stanovujeme stejně jako v (20.4) tedy $p_{ij} = p_{i.} \cdot p_{.j}$. Alternativní hypotéza H_1 je stanovena buď jednostranná nebo oboustranná pomocí tzv. logaritmické interakce. Ta je stanovena jako

logaritmický poměr šancí $\delta = \ln \left(\frac{p_{11} \cdot p_{22}}{p_{12} \cdot p_{21}} \right)$, kde $\ln$ je přirozený logaritmus . Odhad této

hodnoty se stanovuje jako $d = \ln \left(\frac{n_{11} \cdot n_{22}}{n_{12} \cdot n_{21}} \right)$. Tedy hypotéza H_1 je stanovena buď jako $\delta > 0$

(resp. $\delta < 0$) – pak je jednostranná , nebo jako $H_1 : \delta \neq 0$, pak je oboustranná. Vlastní výpočet spočívá v tom , že zjistíme všechny tabulky , které mají stejné marginální četnosti jako tabulka 20. Pro tyto tabulky vyšetříme hodnotu logaritmické interakce a pravděpodobnosti , za jakých je jich nabýváno . Tato pravděpodobnost je pro tabulku typu tabulka 14 rovna

$$P = \frac{n_{1.}! \cdot n_{2.}! \cdot n_{.1}! \cdot n_{.2}!}{n! \cdot n_{11}! \cdot n_{12}! \cdot n_{21}! \cdot n_{22}!} \quad (20.13)$$

Podle typu hypotézy H_1 sčítáme pravděpodobnosti těch tabulek , které této hypotéze vyhovují. Tedy pro případ $\delta > 0$ sčítáme pravděpodobnosti těch tabulek , jejichž logaritmické interakce jsou větší nebo rovny hodnotě d . V případě $\delta < 0$ sčítáme hodnoty pravděpodobností těch tabulek , které mají logaritmické interakce menší nebo rovny d . V případě oboustranné alternativní hypotézy H_1 sčítáme pravděpodobnosti těch tabulek , které mají logaritmické interakce v absolutní hodnotě větší nebo rovny číslu $|d|$.

Příklad 20.6

Výběr celkem 25 studentů (20 děvčat a 5 chlapců) zaujalo postoj k způsobu řešení určitého problému.

		chlapci	děvčata	celkem
Odpověď	ano	1	15	16
	ne	4	5	9
	celkem	5	20	25

Tabulka 15 data pro příklad 20.6

Ověřte , zda pohlaví a způsob řešení problému v tomto případě na sobě nezáleží.

Řešení:

V tomto případě existuje celkem šest různých tabulek se stejnými marginálními četnostmi. V následujícím textu jsou tyto tabulky uvedeny spolu s jejich logaritmickou interakcí a pravděpodobností spočtenou podle vztahu (20.13).

1	15	16	d=	-1,07918
4	5	9	P=	0,037945
5	20	25		

2	14	16	d=	-0,54407
3	6	9	P=	0,189723
5	20	25		

0	16	16	d=	-nek.
5	4	9	P=	0,002372
5	20	25		

3	13	16	d=	-0,09275
2	7	9	P=	0,379447
5	20	25		

4	12	16	d=	0,425969
1	8	9	P=	0,3083
5	20	25		

součet= 0,084585

5	11	16	d=	+nek.
0	9	9	P=	0,082213
5	20	25		

tabulka 16 Tabulky se stejnými marginálními četnostmi

Protože alternativní hypotéza je stanovena jako oboustranná, musíme zjistit součet pravděpodobností u těch tabulek, které mají absolutní hodnotu logaritmické interakce větší než tabulka 15. Jde tedy o druhou a čtvrtou tabulku, součet pravděpodobností je v tomto případě roven 0,084585. Na hladině významnosti 0,05 bychom tedy zamítli nezávislost způsobu řešení na pohlaví.

20.4 Testy o symetrii kontingenční tabulky

V případě, že je daná kontingenční tabulka čtvercová, zajímáme se mnohdy o symetrii pravděpodobností. Stanovujeme tedy hypotézu $H_0 : p_{ij} = p_{ji}$, pro všechny hodnoty $i, j = 1, \dots, r$. Za předpokladu platnosti hypotézy H_0 je počet neznámých parametrů v matici (p_{ij}) roven počtu členů na diagonále + prvky, které leží nad diagonálou, s výjimkou například prvku p_{rr} , který lze dopočítat ze symetrie a znalosti ostatních prvků. Celkově je proto počet neznámých parametrů roven $N = 2 + 3 + \dots + r = \frac{r \cdot (r-1)}{2}$. Jako testovou statistiku zvolíme přirozeně

$$\chi^2 = \sum_{i < j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \quad (20.14)$$

, tato statistika je asymptoticky rozdělení chí – kvadrát s $\frac{r \cdot (r-1)}{2}$ stupni volnosti.

Hypotézu symetrie kontingenční tabulky na hladině významnosti α zamítáme, jestliže $\chi^2 \geq \chi^2_{\frac{r \cdot (r-1)}{2}}(\alpha)$.

V případě čtyřpolní tabulky se test symetrie nazývá **Mc Nemarův**. V součtu (20.14) se vyskytuje jen jeden, který je roven

$$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \quad (20.15)$$

Tato statistika je asymptoticky rozdělení chí – kvadrát s 1 stupněm volnosti.

Příklad 20.7

Při výuce anglického jazyka byla použita speciální metoda pro zrychlení reakce v konverzaci. Tato metoda byla aplikována na vzorek 100 studentů a posléze vyhodnocena do tabulky:

		Po speciální metodě		Celkem
		Ihned	průměrně	
Před speciální metodou	ihned	24	28	52
	průměrně	12	36	48
Celkem		36	64	100

Tabulka 17 - Data příkladu 20.7

Řešení:

Jde o čtyřpolní tabulku, použijeme proto MC Nemarův test, hodnotu testové statistiky spočteme podle (20.15), ta je rovna **6,4**. Kritická hodnota chí – kvadrát o 1 stupni volnosti na hladině významnosti $\alpha = 0,05$ je rovna **3,841455**. Zamítáme tedy na hladině významnosti 0,05, že speciální metoda nemá vliv na zrychlení reakce při konverzaci.

Příklad 20.8

Anglický přírodovědec F.Galton sledoval v celkem 1000 případech barvu očí otce a syna. Výsledky šetření jsou uvedeny v následující tabulce:

		Barva očí otce				Celkem
		světle modrá	modrozelená nebo šedá	tmavě šedá nebo světle hnědá	tmavě znědá	
Barva očí syna	světle modrá	194	70	41	30	335
	modrozelená nebo šedá	83	124	41	36	284
	tmavě šedá nebo světle hnědá	25	34	55	23	137
	tmavě znědá	56	36	43	109	244
Celkem		358	264	180	198	1000

tabulka 18 - Data příkladu 20.8

Zjistěte , zda barva očí otce ovlivňuje barvu očí syna.

Řešení:

Uvedené hodnoty dosadíme do vztahu (20.14) . Tím získáme hodnotu testové statistiky , v tomto případě je rovna číslu 19,55777. Protože je kritická hodnota χ^2 – kvadrát s 6 stupni volnosti rovna 12,59 zamítáme na hladině významnosti 0,05 hypotézu o symetrii v tomto případě.