

Simulácia Chí-kvadrát rozdelenia

John Andersen

Preložili: Ján Beňačka, Soňa Čeretková, Andrej Svorad, Daniel Hamran

Úvod

Jedného dňa sme hodili šesťdesiatimi hracími kockami.

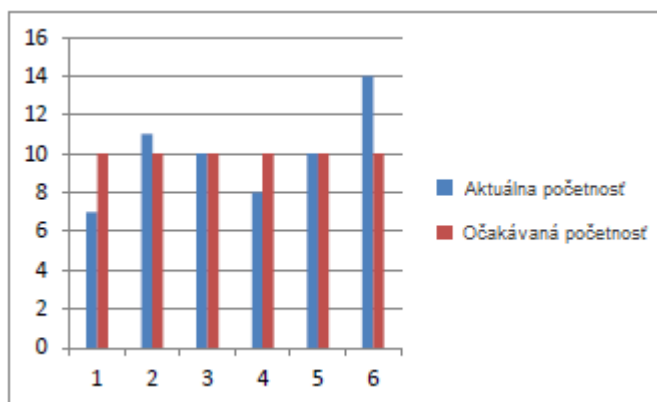


Obr. 1 Fotografia 60 hracích kociek

Po usporiadaní kociek tak, aby sa dala ľahko určiť početnosť možných výsledkov sme získali toto

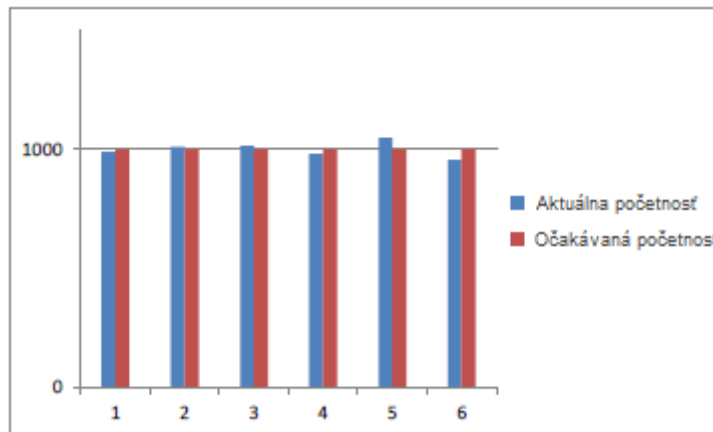


Obr. 2 Jeden spôsob reprezentovania početností výsledkov u 60 kociek



Obr. 3 Stĺpcový graf výsledkov z obr. 2

Pomocou tabuľkového kalkulátora sme simulovali hod so 6000 kockami a získali sme tento výsledok



Obr. 4 Stĺpcový graf výsledkov počítačovej simulácie hodu so 60 kockami

Ako vidieť, početnosti z druhého experimentu sú v omnoho väčšej zhode s očakávanými početnosťami, pretože počet kociek je omnoho väčší. To je v súlade so zákonom veľkých čísel, ako je to opísané v [1].

Čo však, ak sa musíme na základe obr. 1 rozhodnúť, či rozdiel medzi aktuálnym a očakávaným výsledkom nie je príliš veľký? Nakoľko situácia na obrázkoch sa môže značne rôzniť, je ťažké zobrazit' tieto rozdiely pomocou jednoduchého obrázku. Preto ľudia vyvinuli spôsob ako reprezentovať tieto rozdiely pomocou čísel.

„Domáca“ kontrolná metóda

Človeka by najskôr asi napadlo, že niekto, inšpirovaný napr. metódou najmenších štvorcov (nie príliš zložitú vysvetlenie pozri v [2]) použije nasledovnú sumáciu

$$\sum (\text{aktuálny výsledok} - \text{očakávaný výsledok})^2,$$

alebo konkrétnejšie pre 60 kociek,

$$(1) \quad \sum_{k=1}^6 (\text{početnosť } k - 10)^2.$$

Na jednej strane, ak je výsledok nula, potom všetkých 6 početností je rovných 10, čo je príliš dobrý výsledok na to, aby to bola pravda (snáď manipulácia s údajmi?), na druhej strane, ak je výsledok príliš veľké číslo, potom rozdiely medzi aktuálnou a očakávanou početnosťou sú príliš veľké; takže buď máme veľmi ojedinelý výsledok alebo niečo s očakávanými početnosťami nie je v poriadku. Je možné, že kocky boli upravené tak, aby niektoré číslo padalo častejšie.

Odhadnúť výsledok nie je jednoduché. Pre lepšie pochopenie môžeme použiť Excel na simulovanie veľkého počtu hodov 60 kockami, urobiť štatistické vyhodnotenie výsledkov rovnice (1) a znázorniť ich na obrázku.

Predovšetkým: Ako nasimulovať túto situáciu?. Ako dokážeme vytvoriť tieto dáta v Exceli? Tabuľka je na obr. 5 (výsledné hodnoty) a na obr. 6-8 (vzorce).

- Stĺpec A obsahuje číslo hodu so 60 kockami.
- Stĺpce B:BI (väčšina je skrytá) obsahuje výsledok hodu každej jednej kocky. V riadku 2 sú uvedené čísla jednotlivých kociek.
- Stĺpce BJ:BO obsahujú početnosti jednotlivých možných čísel (možnosti sú v riadku 2).
- V stĺpcoch BP:BU sú počítané kvadratické odchýlky.
- V stĺpci BV sú sumy kvadratických odchýlok.

	A	B	C	D	E	F	G	H	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV
1		Počet hodov so 60 kockami										Početnosť hodov						Kvadratické odchýlky: (početnosť-10) ²							
2	č. kocky → č. hodů ↓	1	2	3	4	5	6	7	57	58	59	60	1	2	3	4	5	6	1	2	3	4	5	6	Súčty štvorcov odchýlok
3	1	5	6	2	6	2	1	1	1	5	4	2	11	12	9	10	10	8	1	4	1	0	0	4	10
4	2	2	3	5	6	4	6	2	5	6	1	5	11	10	7	6	13	13	1	0	9	16	9	9	44
5	3	2	5	5	4	2	3	4	4	4	2	6	5	16	11	14	5	9	25	36	1	16	25	1	104
6	4	2	3	6	3	4	1	2	3	6	4	3	11	10	10	17	7	5	1	0	0	49	9	25	84
7	5	1	6	3	4	3	1	4	2	6	5	5	12	6	9	10	12	11	4	16	1	0	4	1	26
8	6	2	2	6	6	6	3	1	6	4	1	2	12	11	8	6	13	10	4	1	4	16	9	0	34
9	7	5	6	5	3	6	1	1	6	3	4	5	7	13	10	8	10	12	9	9	0	4	0	4	26
10	8	5	3	6	4	5	5	2	1	2	1	5	5	10	11	12	11	11	25	0	1	4	1	1	32
11	9	2	5	5	1	1	5	3	3	3	4	1	12	5	14	8	13	8	4	25	16	4	9	4	62
12	10	1	5	2	6	2	6	2	2	6	5	3	8	12	9	7	10	14	4	4	1	9	0	16	34
13	11	2	6	1	3	2	4	6	2	6	5	2	9	13	9	6	7	16	1	9	1	16	9	36	72

Obr. 5 Tabuľka pre simuláciu veľkého počtu hodov so 60 kockami

Uvedomme si, že ako náhle máme v jednom riadku zadané potrebné vzorce, potom zväčšiť počet simulovaných hodov je len otázkou kopírovania tohto riadku nadol. V danom prípade sme vzorce kopírovali nadol, až kým sme nedostali 20000 hodov.

Keby sme chceli vzorce skopírovať do všetkých riadkov, veľmi skoro narazíme na problém nedostatku pamäti. Preto, ak niekto chce získať skutočne veľký počet hodov, musí použiť pokročilejšie programovacie techniky. Jedným z cieľov tejto aplikácie je aj to, že chceme zistiť, ako ďaleko sa môžeme dostať s použitím len elementárnych zručností z Excelu.

Na nasledujúcich troch obrázkoch sú ukázané vzorce použité na obr. 5.

Nebudeme zachádzať do detailov, pretože dôkladné štúdium vzorcov a vzťahov medzi nimi a odkazmi v bunkách je asi najlepší spôsob ako danú problematiku pochopiť. V prípade, že znaky vo vzorcoch sú príliš malé, treba obrázok zväčšiť.

	A	B	C	D	E	F
1						
2	č. kocky → č. hodů ↓	1	=B2+1	=C2+1	=D2+1	=E2+1
3	1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
4	=A3+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
5	=A4+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
6	=A5+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
7	=A6+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
8	=A7+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
9	=A8+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
10	=A9+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
11	=A10+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
12	=A11+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)
13	=A12+1	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)	=SLUMPMELEM(1;6)

Obr. 6 Vzorce pre obr. 5 – časť 1

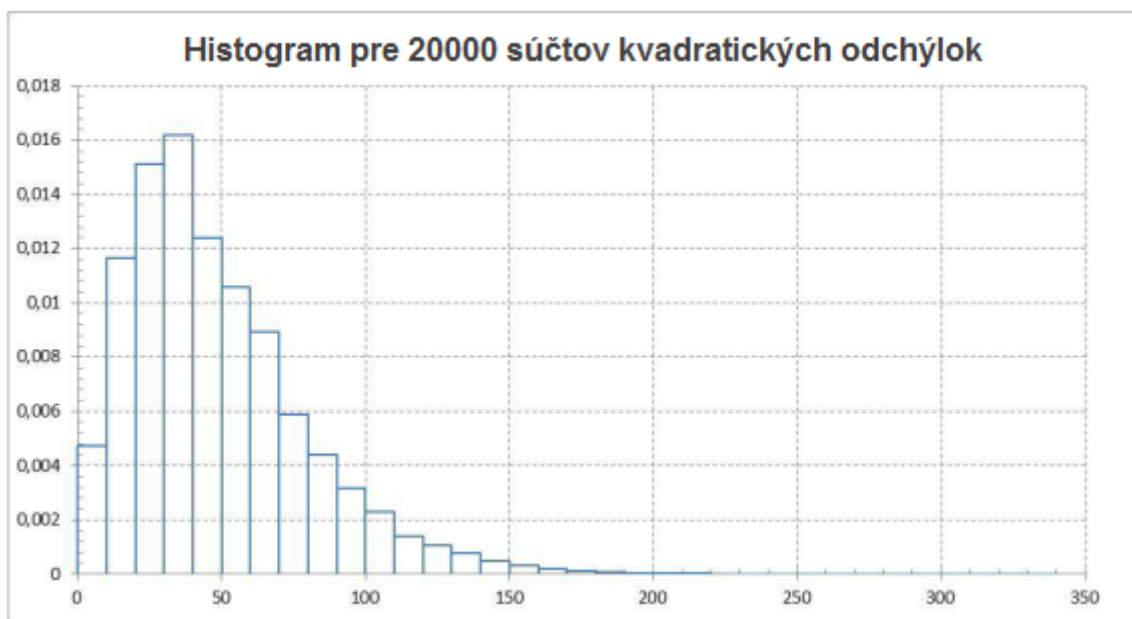
	BI	BJ	BK	BL	BM
1			Početnosť hodov		
2	=BH2+1	1	2	3	4
3	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B3:\$BI3;BJ\$2)	=TÆL.HVIS(\$B3:\$BI3;BK\$2)	=TÆL.HVIS(\$B3:\$BI3;BL\$2)	=TÆL.HVIS(\$B3:\$BI3;BM\$2)
4	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B4:\$BI4;BJ\$2)	=TÆL.HVIS(\$B4:\$BI4;BK\$2)	=TÆL.HVIS(\$B4:\$BI4;BL\$2)	=TÆL.HVIS(\$B4:\$BI4;BM\$2)
5	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B5:\$BI5;BJ\$2)	=TÆL.HVIS(\$B5:\$BI5;BK\$2)	=TÆL.HVIS(\$B5:\$BI5;BL\$2)	=TÆL.HVIS(\$B5:\$BI5;BM\$2)
6	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B6:\$BI6;BJ\$2)	=TÆL.HVIS(\$B6:\$BI6;BK\$2)	=TÆL.HVIS(\$B6:\$BI6;BL\$2)	=TÆL.HVIS(\$B6:\$BI6;BM\$2)
7	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B7:\$BI7;BJ\$2)	=TÆL.HVIS(\$B7:\$BI7;BK\$2)	=TÆL.HVIS(\$B7:\$BI7;BL\$2)	=TÆL.HVIS(\$B7:\$BI7;BM\$2)
8	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B8:\$BI8;BJ\$2)	=TÆL.HVIS(\$B8:\$BI8;BK\$2)	=TÆL.HVIS(\$B8:\$BI8;BL\$2)	=TÆL.HVIS(\$B8:\$BI8;BM\$2)
9	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B9:\$BI9;BJ\$2)	=TÆL.HVIS(\$B9:\$BI9;BK\$2)	=TÆL.HVIS(\$B9:\$BI9;BL\$2)	=TÆL.HVIS(\$B9:\$BI9;BM\$2)
10	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B10:\$BI10;BJ\$2)	=TÆL.HVIS(\$B10:\$BI10;BK\$2)	=TÆL.HVIS(\$B10:\$BI10;BL\$2)	=TÆL.HVIS(\$B10:\$BI10;BM\$2)
11	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B11:\$BI11;BJ\$2)	=TÆL.HVIS(\$B11:\$BI11;BK\$2)	=TÆL.HVIS(\$B11:\$BI11;BL\$2)	=TÆL.HVIS(\$B11:\$BI11;BM\$2)
12	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B12:\$BI12;BJ\$2)	=TÆL.HVIS(\$B12:\$BI12;BK\$2)	=TÆL.HVIS(\$B12:\$BI12;BL\$2)	=TÆL.HVIS(\$B12:\$BI12;BM\$2)
13	=SLUMPMELEM(1;6)	=TÆL.HVIS(\$B13:\$BI13;BJ\$2)	=TÆL.HVIS(\$B13:\$BI13;BK\$2)	=TÆL.HVIS(\$B13:\$BI13;BL\$2)	=TÆL.HVIS(\$B13:\$BI13;BM\$2)

Obr. 7 Vzorce pre obr. 5 – časť 2

	BO	BP	BQ	BR	BS	BT	BU	BV
1								
2	6	1	2	3	4	5	6	Súčty štvorcov odchýlok
3	=TÆL.HVIS(\$B3:\$BI3;BO\$2)	=(BJ3-10)^2	=(BK3-10)^2	=(BL3-10)^2	=(BM3-10)^2	=(BN3-10)^2	=(BO3-10)^2	=SUM(BP3:BU3)
4	=TÆL.HVIS(\$B4:\$BI4;BO\$2)	=(BJ4-10)^2	=(BK4-10)^2	=(BL4-10)^2	=(BM4-10)^2	=(BN4-10)^2	=(BO4-10)^2	=SUM(BP4:BU4)
5	=TÆL.HVIS(\$B5:\$BI5;BO\$2)	=(BJ5-10)^2	=(BK5-10)^2	=(BL5-10)^2	=(BM5-10)^2	=(BN5-10)^2	=(BO5-10)^2	=SUM(BP5:BU5)
6	=TÆL.HVIS(\$B6:\$BI6;BO\$2)	=(BJ6-10)^2	=(BK6-10)^2	=(BL6-10)^2	=(BM6-10)^2	=(BN6-10)^2	=(BO6-10)^2	=SUM(BP6:BU6)
7	=TÆL.HVIS(\$B7:\$BI7;BO\$2)	=(BJ7-10)^2	=(BK7-10)^2	=(BL7-10)^2	=(BM7-10)^2	=(BN7-10)^2	=(BO7-10)^2	=SUM(BP7:BU7)
8	=TÆL.HVIS(\$B8:\$BI8;BO\$2)	=(BJ8-10)^2	=(BK8-10)^2	=(BL8-10)^2	=(BM8-10)^2	=(BN8-10)^2	=(BO8-10)^2	=SUM(BP8:BU8)
9	=TÆL.HVIS(\$B9:\$BI9;BO\$2)	=(BJ9-10)^2	=(BK9-10)^2	=(BL9-10)^2	=(BM9-10)^2	=(BN9-10)^2	=(BO9-10)^2	=SUM(BP9:BU9)
10	=TÆL.HVIS(\$B10:\$BI10;BO\$2)	=(BJ10-10)^2	=(BK10-10)^2	=(BL10-10)^2	=(BM10-10)^2	=(BN10-10)^2	=(BO10-10)^2	=SUM(BP10:BU10)
11	=TÆL.HVIS(\$B11:\$BI11;BO\$2)	=(BJ11-10)^2	=(BK11-10)^2	=(BL11-10)^2	=(BM11-10)^2	=(BN11-10)^2	=(BO11-10)^2	=SUM(BP11:BU11)
12	=TÆL.HVIS(\$B12:\$BI12;BO\$2)	=(BJ12-10)^2	=(BK12-10)^2	=(BL12-10)^2	=(BM12-10)^2	=(BN12-10)^2	=(BO12-10)^2	=SUM(BP12:BU12)
13	=TÆL.HVIS(\$B13:\$BI13;BO\$2)	=(BJ13-10)^2	=(BK13-10)^2	=(BL13-10)^2	=(BM13-10)^2	=(BN13-10)^2	=(BO13-10)^2	=SUM(BP13:BU13)

Obr. 8 Vzorce pre obr. 5 – časť 3

Údaje, ktorými sa budeme zaoberať je 20000 súčtov kvadratických odchýlok v stĺpci BV. Pomocou tradičných štatistických metód (a Excelu) nakreslíme stĺpcový graf ich relatívnych početností v rámci intervalov šírky 10, a tiež čiarový graf ich kumulovaných relatívnych početností v rámci týchto intervalov. Výsledok je na obr. 9 a 10.



Obr. 9



Obr. 10

Môžeme vidieť, že asi 95 % týchto súčtov je menších ako 110 (v skutočnosti 108, ako bolo zistené metódou pokus a omyl v tabuľke). Takže len v 5 % prípadov môžeme očakávať súčet väčší než 108.

Ako je to konkrétne s kockami na obr. 1 a 2? Výpočtom sa ľahko presvedčíme, že súčet kvadratických odchýlok je 30, čo je veľmi typický výsledok, ako ukazuje obr. 9.

Všimnime si, že jediná štatistická metóda použitá v predchádzajúcej časti je triedenie do skupín a vytváranie stĺpcových grafov relatívnych početností a čiarových grafov kumulatívnych relatívnych početností v kombinácii s "hrubou silou", ako je vyprodukovanie 20000 simulovaných hodov 60 kockami a ich spravovanie. To je skôr otázkou výdrže a trpezlivosti než dômyselnej matematickej teórie.

Štandardná Chí-kvadrát štatistika

Problémy tohto typu boli študované dávno pred vynájdením počítačov, kedy boli vypracované i príslušné teórie. Zvyčajne sú riešené takzvaným Chí-kvadrát testom (pozri [3]).

Ak počítame súčet (delíme s 10 kvôli normalizácii)

$$\sum_{k=1}^6 \frac{(\text{početnosť } k - 10)^2}{10},$$

dá sa dokázať, že má Chí-kvadrát rozdelenie s 5 stupňami voľnosti.

Po vyhľadání v tabuľke tohto rozdelenia

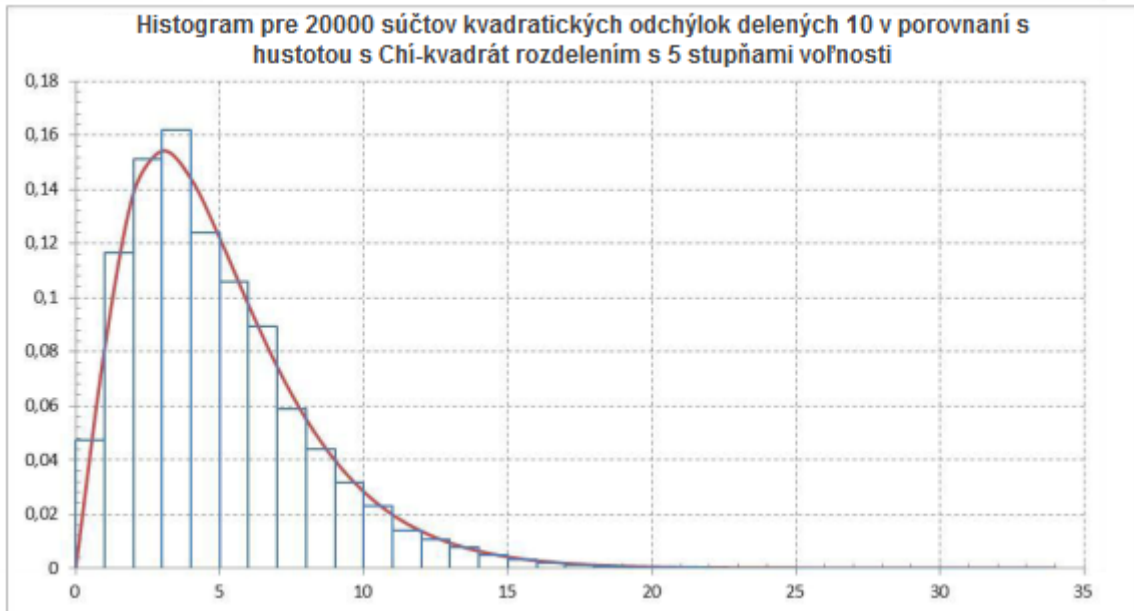
Kumulatívna pravdepodobnosť	Chí-kvadrát
0,99	15,09
0,98	13,39
0,97	12,37
0,96	11,64
0,95	11,07
0,94	10,60
0,93	10,19
0,92	9,84

Tabuľka 1

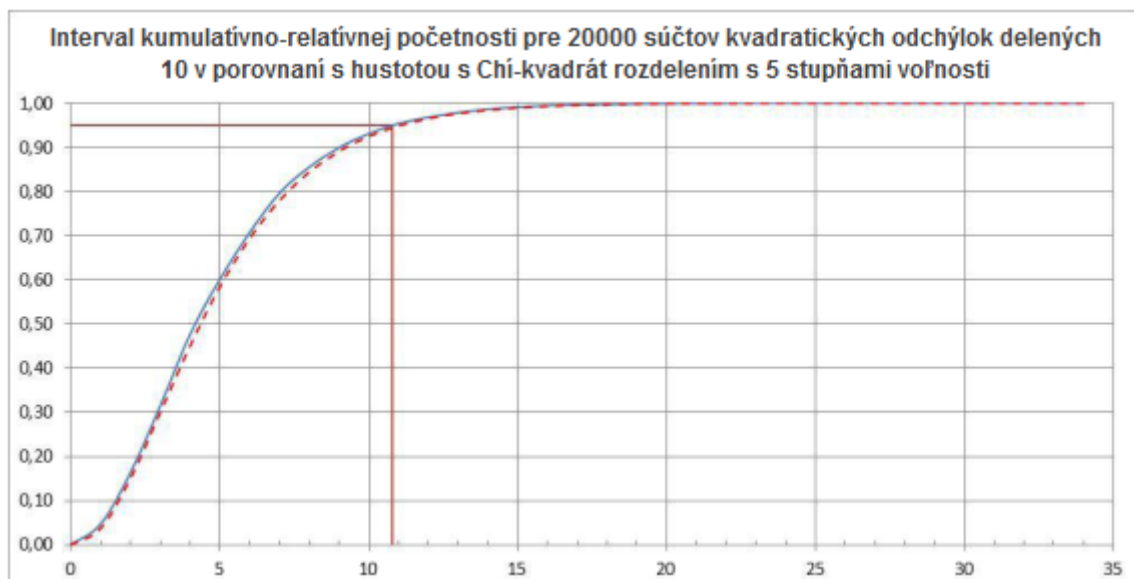
môžeme vidieť, že 95 % z tohto rozdelenia by nemalo presiahnuť 11,07.

Výsledok získaný našou simuláciou je 108. Musíme ho vydeliť s 10, čo dáva $10,8 \approx 11$. To je hodnota, ktorá je porovnateľná s 11,07. Takže náš výsledok 108 je v poriadku.

Na obr. 11 a 12 sú porovnané simulované dáta s Chi-kvadrát rozdelením. Hárok je rovnaký ako na obr. 5 s tým, že vzorce v stĺpci BV boli vydelené a 10. Obrázky preto uvádzame.



Obr. 11



Obr. 12

Kontrola nezávislosti kategórií

Preskúmame rovnakými prostriedkami omnoho zložitejšiu situáciu.

Majme krajinu so štyrmi politickými stranami P, Q, R a S. Vieme, že v posledných voľbách získali 40, 30, 20 a 10 percent hlasov. Chceme zistiť, či príjmy ľudí majú vplyv na to, ktorej strane odovzdajú svoj hlas. Za týmto účelom roztriedime príjmy ľudí do troch kategórií IC1, IC2 and IC3. Pre túto

situáciu boli vyvinuté štandardné testy (pozri [3]). Aby sme získali predstavu o tom, aké je rozdelenie Chí-kvadrát testu, riešime problém simulovaním veľkého počtu pohovorov s voličmi. Výsledok je na obr. 23 a 24.

V nižšie uvedenej tabuľke je usporiadaný výsledok odpovedí 600 voličov

	P	Q	R	S
IC1	110	96	53	23
IC2	70	60	41	22
IC3	42	41	18	24

Tabuľka 1

Táto situácia pripomína situáciu na obr. 3, kde sa na základe jednej množiny údajov musíme pre čosi rozhodnúť. V prípade so 60 kockami to bola otázka, či očakávané rozdelenie je akceptovateľné ako model, v prípade s voličmi je to otázka, či uvedené dve kritériá môžu byť považované za nezávislé alebo nie.

Nápad je podobný simulovaniu 20000 hodov so 60 kockami. Úlohou je vytvoriť veľký počet simulácií tabuliek takých ako je uvedená vyššie, a to spôsobom, ktorý zaručí nezávislosť uvedených dvoch kritérií. Toto zaistíme použitím patričných vzorcov.

Úloha je omnoho komplikovanejšia než problém s kockami. Predsa však, je riešiteľná, ako je ukázané na nasledujúcich obrázkoch. Videá demonštrujúce jednotlivé kroky riešenia budú umiestnené na domovskej stránke projektu.

Prvým krokom je vytvoriť hárok, ktorý ukazuje, kam vložiť nové informácie, jednu za druhou.

	A	B	C	D	E	F	G
1				40%	30%	20%	10%
2			68	P	Q	R	S
3		14		0	1	0	0
4	50%	IC1	1	0	1	0	0
5	30%	IC2	0	0	0	0	0
6	20%	IC3	0	0	0	0	0

Obr. 13 Umiestnenie jedného voliča v schéme s nezávislosťou kritérií

	A	B	C	D	E	F	G
1				0,4	0,3	0,2	0,1
2			=SLUMPMELEM(1;100)	P	Q	R	S
3		=SLUMPMELEM(1;100)		=HVIS(\$C\$2<=40;1;0)	=HVIS(OG(\$C\$2>40;\$C\$2<=70);1;0)	=HVIS(OG(\$C\$2>70;\$C\$2<=90);1;0)	=HVIS(\$C\$2>90;1;0)
4	0,5	IC1	=HVIS(\$B\$3<=50;1;0)	=D\$3*\$C4	=E\$3*\$C4	=F\$3*\$C4	=G\$3*\$C4
5	0,3	IC2	=HVIS(OG(50<\$B\$3;\$B\$3<=80);1;0)	=D\$3*\$C5	=E\$3*\$C5	=F\$3*\$C5	=G\$3*\$C5
6	0,2	IC3	=HVIS(\$B\$3>80;1;0)	=D\$3*\$C6	=E\$3*\$C6	=F\$3*\$C6	=G\$3*\$C6

Obr. 14 Vzorce pre obr. 13

Teraz musíme umožniť vykonať veľký počet "pohovorov" a zosumarizovať výsledky. Použijeme rovnakú techniku ako v Math2Earth (pozri [4]), preto, ešte pred tým ako začneme študovať nasledujúce obrázky, si ju treba preštudovať.

	A	B	C	D	E	F	G
1				40%	30%	20%	10%
2			53	P	Q	R	S
3		2		0	1	0	0
4	50%	IC1	1	0	1	0	0
5	30%	IC2	0	0	0	0	0
6	20%	IC3	0	0	0	0	0
7							
8	Reset (0/1)			P	Q	R	S
9	1		IC1	67	44	31	18
10	N		IC2	37	27	28	13
11	336		IC3	25	16	18	12
12							
13			N	336			
14							

Obr. 15 Akumulovaní voliči simulovaní na obr. 13

	A	B	C	D	E	F	G
7							
8	Reset (0/1)			P	Q	R	S
9	1	IC1	=HVIS(\$A\$9=1;D9+TÆL.HVIS(D4;1);0)	=HVIS(\$A\$9=1;E9+TÆL.HVIS(E4;1);0)	=HVIS(\$A\$9=1;F9+TÆL.HVIS(F4;1);0)	=HVIS(\$A\$9=1;G9+TÆL.HVIS(G4;1);0)	
10	N	IC2	=HVIS(\$A\$9=1;D10+TÆL.HVIS(D5;1);0)	=HVIS(\$A\$9=1;E10+TÆL.HVIS(E5;1);0)	=HVIS(\$A\$9=1;F10+TÆL.HVIS(F5;1);0)	=HVIS(\$A\$9=1;G10+TÆL.HVIS(G5;1);0)	
11	=HVIS(A9=1;A11+1;0)	IC3	=HVIS(\$A\$9=1;D11+TÆL.HVIS(D6;1);0)	=HVIS(\$A\$9=1;E11+TÆL.HVIS(E6;1);0)	=HVIS(\$A\$9=1;F11+TÆL.HVIS(F6;1);0)	=HVIS(\$A\$9=1;G11+TÆL.HVIS(G6;1);0)	
12							
13			N	=SUM(D9:G11)			

Obr. 16 Vzorce pre obr. 15

Nakoľko ale potrebujeme simulovať veľké množstvo pohovorov, a každá vyššie uvedená tabuľka obsahuje viacero riadkov, vytvorenie simulácií nie je len otázkou kopírovania nadol. Musíme prerobiť tabuľky na obr. 13 a 15 tak, aby obsahovali len jeden riadok. Toto môžeme vykonať presunutím buniek (nie kopírovaním!), pričom Excel automaticky upraví odkazy na bunky. Snáď najlepším spôsobom ako ukázať výsledný hárok je stiahnuť si ho zo stránky projektu, resp. stiahnuť si k nemu demonštračné video. Predpokladáme však, že čitateľ, ktorý sa dopracoval až do tohto bodu je schopný pochopiť logiku vzorcov uvedených nižšie, preto ich uvádzame bez ďalších komentárov. Môžu byť čítané a chápané ako istý druh poézie so vzájomnými logickými väzbami.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
7	1																					
8	N																					
9	993																					
10																						
11		Stĺpec A & C		Riadok 1 & 3				Riadok 4				Riadok 5				Riadok 6						
12	Pokus č.	50	30	20		40	30	20	10	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	
13	1	11	1	0	0	70	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
14	2	74	0	1	0	83	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
15	3	100	0	0	1	66	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
16	4	1	1	0	0	85	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
17	5	3	1	0	0	55	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
18	6	8	1	0	0	46	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
19	7	44	1	0	0	98	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
20	8	22	1	0	0	80	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
21	9	7	1	0	0	13	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Obr. 17 Simulácia množstva nezávislých dátových tabuliek, každá organizovaná v jednom rade – časť 1

	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
7																			
8																			
9																			
10	O_{ij}																		
11	Riadok 9				Riadok 10				Riadok 11				Súčet riadkov			Súčet stĺpcov			
12	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	RS1	RS2	RS3	CS1	CS2	CS3	CS4
13	178	160	91	43	118	86	69	41	75	69	32	31	472	314	207	371	315	192	115
14	203	143	106	56	118	81	54	36	76	55	43	22	508	289	196	397	279	203	114
15	190	142	98	50	111	104	68	31	82	58	40	19	480	314	199	383	304	206	100
16	198	149	95	48	135	80	48	29	92	62	42	15	490	292	211	425	291	185	92
17	185	141	107	55	123	92	64	38	80	61	31	16	488	317	188	388	294	202	109
18	195	172	97	46	96	86	58	24	86	70	45	18	510	264	219	377	328	200	88
19	194	138	98	38	138	88	76	37	64	64	37	21	468	339	186	396	290	211	96
20	186	154	110	50	105	89	62	35	81	60	41	20	500	291	202	372	303	213	105
21	199	142	100	44	120	92	59	34	84	54	47	18	485	305	203	403	288	206	96
22	210	140	113	50	136	79	61	26	65	55	44	14	513	302	178	411	274	218	90
23	201	145	78	55	113	79	51	34	97	75	45	20	479	277	237	411	299	174	109
24	201	138	92	43	126	98	64	34	79	65	37	16	474	322	197	406	301	193	93

Obr. 18 Simulácia množstva nezávislých dátových tabuliek, každá organizovaná v jednom rade – časť 2

	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
7																									
8																									
9																									
10																									
11	E_{ij}												$(O_{ij} - E_{ij})^2 / E_{ij}$										χ^2		
12	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	
13	176	150	91,3	54,7	117	99,6	60,7	36,4	77,3	65,7	40	24	0,02	0,7	0	2,49	0	1,86	1,13	0,59	0,07	0,17	1,61	2,06	10,70287
14	203	143	104	58,3	116	81,2	59,1	33,2	78,4	55,1	40,1	22,5	0	0	0,04	0,09	0,05	0	0,44	0,24	0,07	0	0,21	0,01	1,163858
15	185	147	99,6	48,3	121	96,1	65,1	31,6	76,8	60,9	41,3	20	0,13	0,17	0,02	0,06	0,84	0,64	0,13	0,01	0,36	0,14	0,04	0,05	2,595311
16	210	144	91,3	45,4	125	85,6	54,4	27,1	90,3	61,8	39,3	19,5	0,65	0,2	0,15	0,15	0,8	0,36	0,75	0,14	0,03	0	0,18	1,06	4,492981
17	191	144	99,3	53,6	124	93,9	64,5	34,8	73,5	55,7	38,2	20,6	0,17	0,08	0,6	0,04	0,01	0,04	0	0,29	0,58	0,51	1,37	1,04	4,742738
18	194	168	103	45,2	100	87,2	53,2	23,4	83,1	72,3	44,1	19,4	0,01	0,07	0,32	0,01	0,18	0,02	0,44	0,02	0,1	0,08	0,02	0,1	1,359655
19	187	137	99,4	45,2	135	99	72	32,8	74,2	54,3	39,5	18	0,29	0,01	0,02	1,16	0,06	1,22	0,22	0,55	1,4	1,72	0,16	0,51	7,317601
20	187	153	107	52,9	109	88,8	62,4	30,8	75,7	61,6	43,3	21,4	0,01	0,01	0,07	0,16	0,15	0	0	0,58	0,37	0,04	0,13	0,09	1,611609
21	197	141	101	46,9	124	88,5	63,3	29,5	82,4	58,9	42,1	19,6	0,02	0,01	0	0,18	0,12	0,14	0,29	0,69	0,03	0,4	0,57	0,13	2,592167
22	212	142	113	46,5	125	83,3	66,3	27,4	73,7	49,1	39,1	16,1	0,03	0,02	0	0,26	0,97	0,23	0,42	0,07	1,02	0,7	0,62	0,28	4,622283
23	198	144	83,9	52,6	115	83,4	48,5	30,4	98,1	71,4	41,5	26	0,04	0	0,42	0,11	0,02	0,23	0,12	0,42	0,01	0,19	0,29	1,39	3,257867
24	194	144	92,1	44,4	132	97,6	62,6	30,2	80,5	59,7	38,3	18,5	0,27	0,22	0	0,04	0,24	0	0,03	0,49	0,03	0,47	0,04	0,33	2,168131
25	196	155	99,7	53,9	114	90,3	58	31,4	75,4	59,6	38,3	20,7	0,17	0,95	0,87	0,16	0,05	0,5	0,63	0,08	0,15	0,49	0,28	0,08	4,401406

Obr. 19 Simulácia množstva nezávislých dátových tabuliek, každá organizovaná v jednom rade – časť 3

	AI	AJ	AK	AL	AM	AN	AO
10							
11		Súčet riadkov			Súčet stĺpcov		
12	RS1	RS2	RS3	CS1	CS2	CS3	CS4
13	=W13+X13+Y13+Z13	=AA13+AB13+AC13+AD13	=AE13+AF13+AG13+AH13	=W13+AA13+AE13	=X13+AB13+AF13	=Y13+AC13+AG13	=Z13+AD13+AH13
14	=W14+X14+Y14+Z14	=AA14+AB14+AC14+AD14	=AE14+AF14+AG14+AH14	=W14+AA14+AE14	=X14+AB14+AF14	=Y14+AC14+AG14	=Z14+AD14+AH14
15	=W15+X15+Y15+Z15	=AA15+AB15+AC15+AD15	=AE15+AF15+AG15+AH15	=W15+AA15+AE15	=X15+AB15+AF15	=Y15+AC15+AG15	=Z15+AD15+AH15
16	=W16+X16+Y16+Z16	=AA16+AB16+AC16+AD16	=AE16+AF16+AG16+AH16	=W16+AA16+AE16	=X16+AB16+AF16	=Y16+AC16+AG16	=Z16+AD16+AH16
17	=W17+X17+Y17+Z17	=AA17+AB17+AC17+AD17	=AE17+AF17+AG17+AH17	=W17+AA17+AE17	=X17+AB17+AF17	=Y17+AC17+AG17	=Z17+AD17+AH17
18	=W18+X18+Y18+Z18	=AA18+AB18+AC18+AD18	=AE18+AF18+AG18+AH18	=W18+AA18+AE18	=X18+AB18+AF18	=Y18+AC18+AG18	=Z18+AD18+AH18
19	=W19+X19+Y19+Z19	=AA19+AB19+AC19+AD19	=AE19+AF19+AG19+AH19	=W19+AA19+AE19	=X19+AB19+AF19	=Y19+AC19+AG19	=Z19+AD19+AH19
20	=W20+X20+Y20+Z20	=AA20+AB20+AC20+AD20	=AE20+AF20+AG20+AH20	=W20+AA20+AE20	=X20+AB20+AF20	=Y20+AC20+AG20	=Z20+AD20+AH20
21	=W21+X21+Y21+Z21	=AA21+AB21+AC21+AD21	=AE21+AF21+AG21+AH21	=W21+AA21+AE21	=X21+AB21+AF21	=Y21+AC21+AG21	=Z21+AD21+AH21

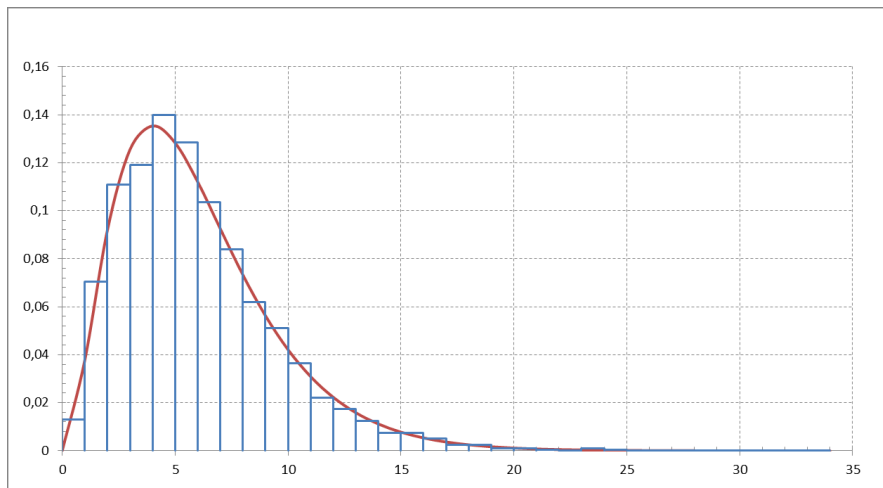
Obr. 20 Vzorce pre obr. 17 – 19, časť 1

	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	
10												
11					E _{ij}							
12	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	
13	=SAI13*AL13/SA\$9	=SAI13*AM13/SA\$9	=SAI13*AN13/SA\$9	=SAI13*AO13/SA\$9	=SAI13*AL13/SA\$9	=SAI13*AM13/SA\$9	=SAI13*AN13/SA\$9	=SAI13*AO13/SA\$9	=SAI13*AL13/SA\$9	=SAI13*AM13/SA\$9	=SAI13*AN13/SA\$9	
14	=SAI14*AL14/SA\$9	=SAI14*AM14/SA\$9	=SAI14*AN14/SA\$9	=SAI14*AO14/SA\$9	=SAI14*AL14/SA\$9	=SAI14*AM14/SA\$9	=SAI14*AN14/SA\$9	=SAI14*AO14/SA\$9	=SAI14*AL14/SA\$9	=SAI14*AM14/SA\$9	=SAI14*AN14/SA\$9	
15	=SAI15*AL15/SA\$9	=SAI15*AM15/SA\$9	=SAI15*AN15/SA\$9	=SAI15*AO15/SA\$9	=SAI15*AL15/SA\$9	=SAI15*AM15/SA\$9	=SAI15*AN15/SA\$9	=SAI15*AO15/SA\$9	=SAI15*AL15/SA\$9	=SAI15*AM15/SA\$9	=SAI15*AN15/SA\$9	
16	=SAI16*AL16/SA\$9	=SAI16*AM16/SA\$9	=SAI16*AN16/SA\$9	=SAI16*AO16/SA\$9	=SAI16*AL16/SA\$9	=SAI16*AM16/SA\$9	=SAI16*AN16/SA\$9	=SAI16*AO16/SA\$9	=SAI16*AL16/SA\$9	=SAI16*AM16/SA\$9	=SAI16*AN16/SA\$9	
17	=SAI17*AL17/SA\$9	=SAI17*AM17/SA\$9	=SAI17*AN17/SA\$9	=SAI17*AO17/SA\$9	=SAI17*AL17/SA\$9	=SAI17*AM17/SA\$9	=SAI17*AN17/SA\$9	=SAI17*AO17/SA\$9	=SAI17*AL17/SA\$9	=SAI17*AM17/SA\$9	=SAI17*AN17/SA\$9	
18	=SAI18*AL18/SA\$9	=SAI18*AM18/SA\$9	=SAI18*AN18/SA\$9	=SAI18*AO18/SA\$9	=SAI18*AL18/SA\$9	=SAI18*AM18/SA\$9	=SAI18*AN18/SA\$9	=SAI18*AO18/SA\$9	=SAI18*AL18/SA\$9	=SAI18*AM18/SA\$9	=SAI18*AN18/SA\$9	
19	=SAI19*AL19/SA\$9	=SAI19*AM19/SA\$9	=SAI19*AN19/SA\$9	=SAI19*AO19/SA\$9	=SAI19*AL19/SA\$9	=SAI19*AM19/SA\$9	=SAI19*AN19/SA\$9	=SAI19*AO19/SA\$9	=SAI19*AL19/SA\$9	=SAI19*AM19/SA\$9	=SAI19*AN19/SA\$9	
20	=SAI20*AL20/SA\$9	=SAI20*AM20/SA\$9	=SAI20*AN20/SA\$9	=SAI20*AO20/SA\$9	=SAI20*AL20/SA\$9	=SAI20*AM20/SA\$9	=SAI20*AN20/SA\$9	=SAI20*AO20/SA\$9	=SAI20*AL20/SA\$9	=SAI20*AM20/SA\$9	=SAI20*AN20/SA\$9	
21	=SAI21*AL21/SA\$9	=SAI21*AM21/SA\$9	=SAI21*AN21/SA\$9	=SAI21*AO21/SA\$9	=SAI21*AL21/SA\$9	=SAI21*AM21/SA\$9	=SAI21*AN21/SA\$9	=SAI21*AO21/SA\$9	=SAI21*AL21/SA\$9	=SAI21*AM21/SA\$9	=SAI21*AN21/SA\$9	

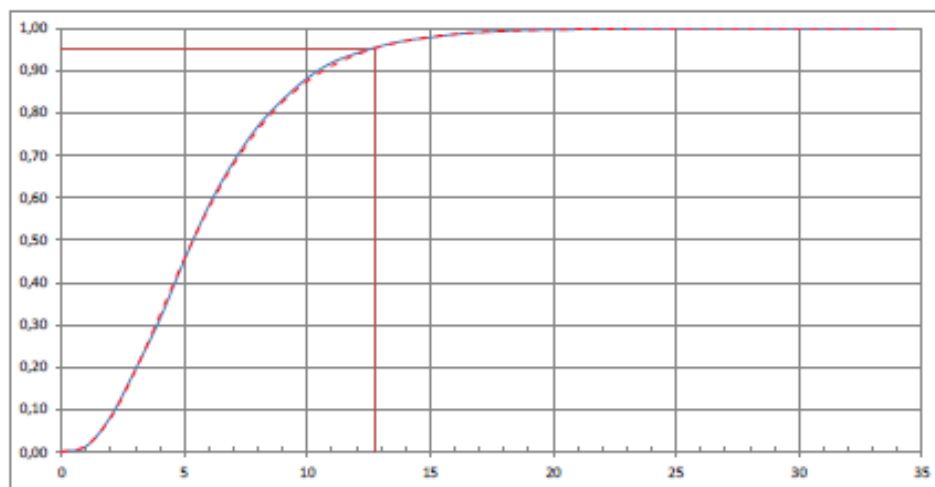
Obr. 21 Vzorce pre obr. 17 – 19, časť 2

	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
10													
11				(O _{ij} - E _{ij}) ² /E _{ij}									χ ²
12	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	(1,1)	(1,2)	(1,3)	(1,4)	
13	= (W13-AP13)^2/AP13	= (X13-AQ13)^2/AQ13	= (Y13-AR1) = (Z13-AS1) = (AA13-AT) = (AB13-AL) = (AC13-AV) = (AD13-AW) = (AE13-AX) = (AF13-AY) = (AG13-AZ) = (AH13-BA) = SUM(BB13:BM13)										
14	= (W14-AP14)^2/AP14	= (X14-AQ14)^2/AQ14	= (Y14-AR1) = (Z14-AS1) = (AA14-AT) = (AB14-AL) = (AC14-AV) = (AD14-AW) = (AE14-AX) = (AF14-AY) = (AG14-AZ) = (AH14-BA) = SUM(BB14:BM14)										
15	= (W15-AP15)^2/AP15	= (X15-AQ15)^2/AQ15	= (Y15-AR1) = (Z15-AS1) = (AA15-AT) = (AB15-AL) = (AC15-AV) = (AD15-AW) = (AE15-AX) = (AF15-AY) = (AG15-AZ) = (AH15-BA) = SUM(BB15:BM15)										
16	= (W16-AP16)^2/AP16	= (X16-AQ16)^2/AQ16	= (Y16-AR1) = (Z16-AS1) = (AA16-AT) = (AB16-AL) = (AC16-AV) = (AD16-AW) = (AE16-AX) = (AF16-AY) = (AG16-AZ) = (AH16-BA) = SUM(BB16:BM16)										
17	= (W17-AP17)^2/AP17	= (X17-AQ17)^2/AQ17	= (Y17-AR1) = (Z17-AS1) = (AA17-AT) = (AB17-AL) = (AC17-AV) = (AD17-AW) = (AE17-AX) = (AF17-AY) = (AG17-AZ) = (AH17-BA) = SUM(BB17:BM17)										
18	= (W18-AP18)^2/AP18	= (X18-AQ18)^2/AQ18	= (Y18-AR1) = (Z18-AS1) = (AA18-AT) = (AB18-AL) = (AC18-AV) = (AD18-AW) = (AE18-AX) = (AF18-AY) = (AG18-AZ) = (AH18-BA) = SUM(BB18:BM18)										
19	= (W19-AP19)^2/AP19	= (X19-AQ19)^2/AQ19	= (Y19-AR1) = (Z19-AS1) = (AA19-AT) = (AB19-AL) = (AC19-AV) = (AD19-AW) = (AE19-AX) = (AF19-AY) = (AG19-AZ) = (AH19-BA) = SUM(BB19:BM19)										
20	= (W20-AP20)^2/AP20	= (X20-AQ20)^2/AQ20	= (Y20-AR2) = (Z20-AS2) = (AA20-AT) = (AB20-AL) = (AC20-AV) = (AD20-AW) = (AE20-AX) = (AF20-AY) = (AG20-AZ) = (AH20-BA) = SUM(BB20:BM20)										
21	= (W21-AP21)^2/AP21	= (X21-AQ21)^2/AQ21	= (Y21-AR2) = (Z21-AS2) = (AA21-AT) = (AB21-AL) = (AC21-AV) = (AD21-AW) = (AE21-AX) = (AF21-AY) = (AG21-AZ) = (AH21-BA) = SUM(BB21:BM21)										

Obr. 22 Vzorce pre obr. 17 – 19, časť 3



Obr. 23 Stĺpcový graf 2000 simulovaných interview so 600 ľuďmi porovnaný s hustotou pre Chí-kvadrát rozdelenie s (3-1)(4-1) = 6 stupňami voľnosti. Údaje sú zo stĺpca BN na obr. 16



Obr. 24 Kumulované relatívne početnosti z obr. 23 porovnané s Chí-kvadrát rozdelením so 6 stupňami voľnosti

Literatúra

- [1] John Andersen Dynamical simulation of stochastic phenomena using Excel
- [2] http://www.efunda.com/math/least_squares/least_squares.cfm (December 2011)
- [3] http://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test (December 2011)
- [4] John Andersen: Animation – bringing motion into mathematics <http://www.math2earth.org/> (December 2011)

Z angličtiny preložil Ján Beňačka