

# Simulazione della $\chi^2$ - distribuzione

John Andersen

## Introduzione

Un giorno ho tirato i 60 dadi (vedi [1])



Fig. 1 Photo of the 60 dice at my desktop

Riorganizzando un po' l'ordine per le frequenze poi ha dato questo



Fig. 2 Un modo per rappresentare le frequenze di 60 dadi

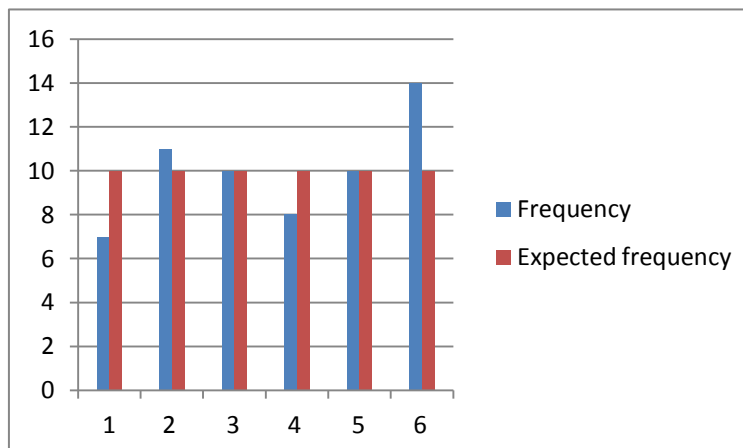


Fig. 3 Bar diagram sui dadi della Fig. 2

Con l'aiuto di un foglio elettronico che simulava il lancio di dadi 6000 ho ottenuto questo quadro dei risultati

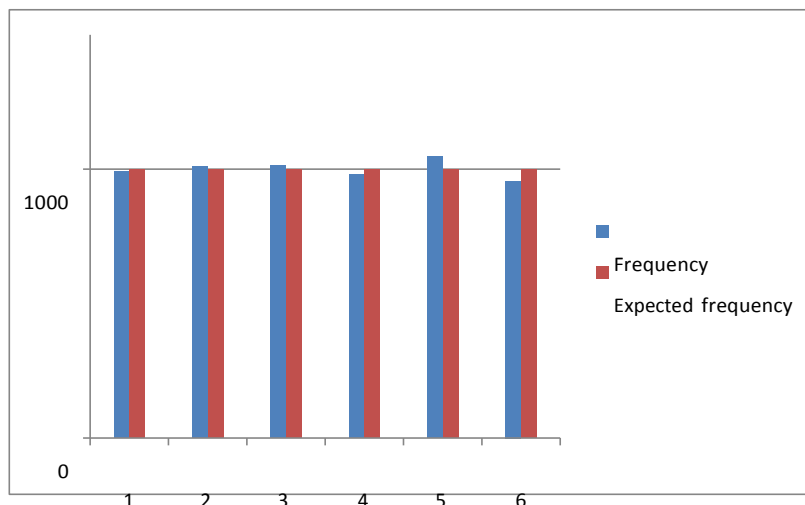


Fig. 4 Bar diagram representing a computer simulated roll of 60 dice

Come potete vedere le frequenze del secondo esperimento è molto più in conformità con le frequenze attese quando il numero di dadi è grande. Ciò è in accordo con la legge dei grandi numeri come descritto nella [2].

Ma cosa succede se si deve decidere dalla situazione in fig. 1 se la differenza tra risultato effettivo e risultato atteso è troppo grande? Dal momento che l'immagine grafica può variare molto è difficile avere un quadro semplice delle differenze. Le persone hanno sviluppato modi per rappresentare le variazioni da numeri.

## A "home made" metodo di controllo

In questo luogo si può immaginare che qualcuno ad esempio ispirato dal metodo dei minimi quadrati (si veda [3] per una spiegazione non troppo complicata) guarda somme come

$$(1) \quad \sum(\text{actual outcome} - \text{expected outcome})^2$$

Da un lato se questo è pari a zero tutti i 6 frequenze sono pari a 10, che può essere troppo bello per essere vero (forse la manipolazione con i dati?) e, dall'altro se il numero è troppo grande la differenza tra le frequenze effettive e attese sono troppo grandi quindi o si ha un risultato raro o c'è qualcosa di sbagliato con l' atteso frequenze. Alcuni carico dei dadi potrebbe essere stato eseguito. Non è affatto facile da vedere che cosa ci si può aspettare del numero. Per ottenere un grip sul quadro generale possiamo usare Excel per simulare un sacco di lanci di 60 dadi e quindi apportare alcune statistiche sui numeri calcolati da (1).

Prima di tutto: Come simulare questa situazione. Come faccio a gestire a mettere tali dati con Excel? Il foglio di diffusione è mostrato nella fig. 5 (i numeri risultanti) e Figg. 6-8 (formule).

- La colonna A indica il numero di ogni lancio di 60 dadi.
- Colonne B: BI (molti sono nascosti) dà i risultati dei dadi individuali. Riga 2 indica il numero dei dadi.
- Colonne BJ: BO dà le frequenze del numero 6 possibile di pips. (Possibilità di cui alla riga 2).
- Colonne BP: BU calcola i quadrati delle deviazioni.
- Colonna BV fornisce le somme delle deviazioni al quadrato.

	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		Pips of the 60 dices											Frequencies of pips						Squares of deviations : (frequency - 10) <sup>2</sup>						
2	Dice nr. → Toss nr. ↓	1	2	3	4	5	6	7	57	58	59	60	1	2	3	4	5	6	1	2	3	4	5	6	Sums of squares of deviations
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

Fig. 5 Spread sheet for simulating a lot of rolls with 60 dice

Tieni presente che una volta che hai impostato formule per una riga si tratta di copiare righe verso il basso per aumentare il numero di rotoli simulati. Nel caso concreto ho copiato verso il basso finché non ho avuto 20.000 rotoli.

Ci sono 220 righe da fare nella mia versione di Excel. La memoria di lavoro del mio computer si esaurisce molto prima di aver compilato in tutte le righe possibili. Quindi, se volete un numero veramente grande di rulli si dovrà passare a strumenti di programmazione più avanzati. Ma uno dei miei punti a questo proposito è che voglio esplorare quanto si può ottenere utilizzando piuttosto elementari competenze di Excel.

.Nelle successive figure 3 sono mostrate le formule dietro la Fig. 5.

Non voglio entrare nei dettagli in quanto un attento studio delle formule e delle relazioni tra loro e i riferimenti di cella può essere il modo migliore per costruire una comprensione. Si prega di zoom in se i caratteri appaiono troppo piccoli sullo schermo.

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

Fig. 6 Formulas for Fig 5 - part 1.

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

Fig. 7 Formulas for Fig 5 - part 2.

	BO	BP	BQ	BR	BS	BT	BU	BV
1								
2	6	1	2	3	4	5	6	Sums of squares of deviations
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)

Fig. 8 Formulas for Fig 5 - part 3.

Le 20000 somme di quadrati delle deviazioni dalla colonna BV (Se Fig 5), sono ora i dati che voglio avere una panoramica. Pertanto dai tradizionali metodi statistici (e Excel) traggio istogrammi e curve oltre accumulati frequenze relative intervallo. Diagrammi risultanti sono mostrati nelle Figg. 9-10.

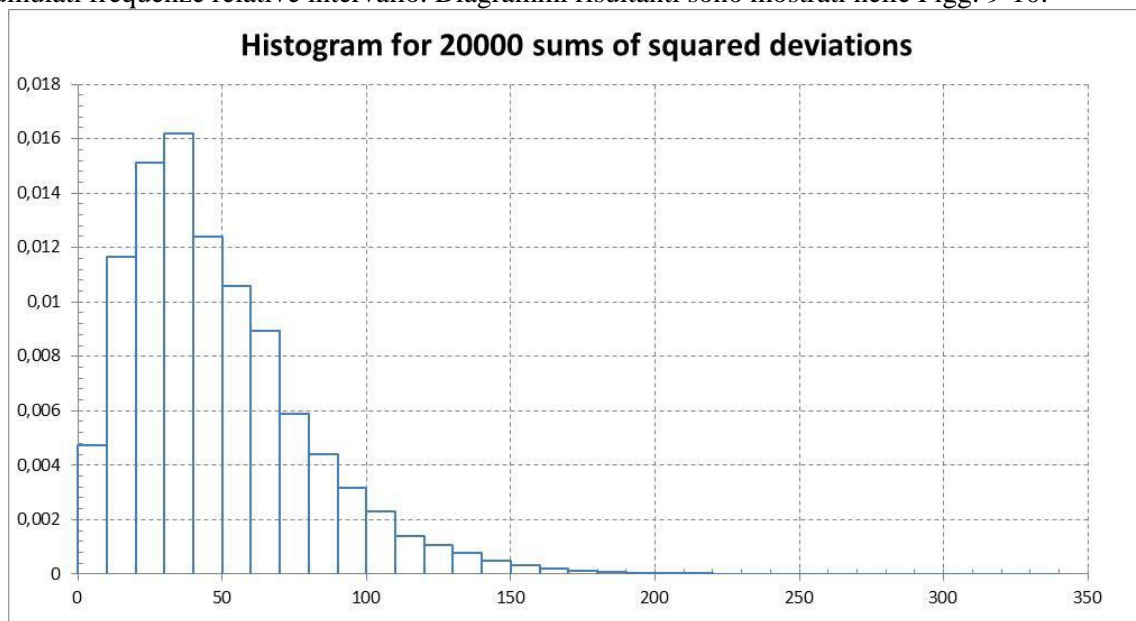


Fig. 9

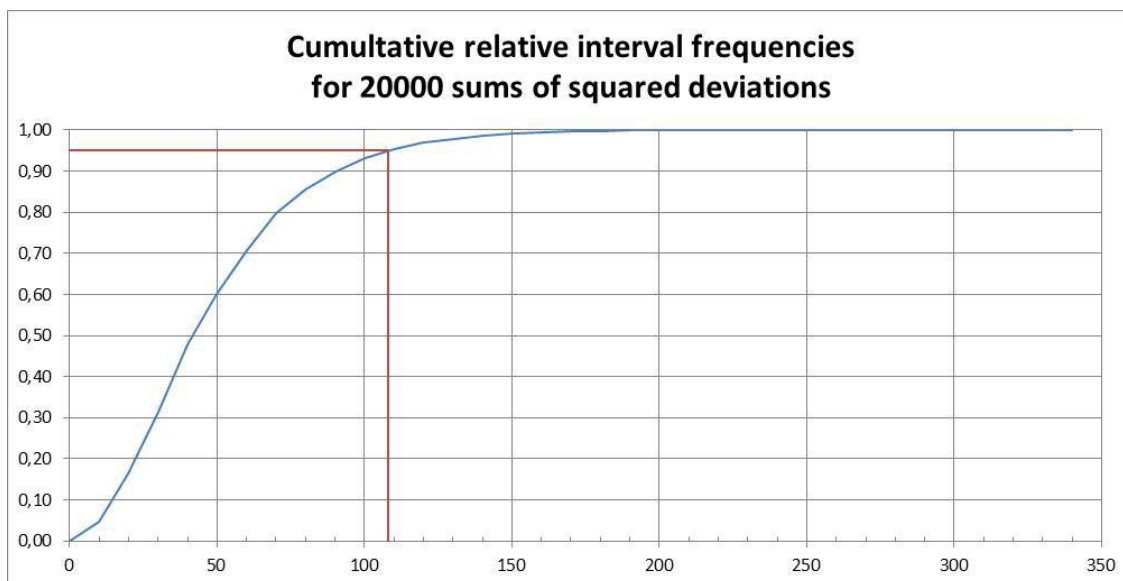


Fig. 10

Si vede che circa il 95% delle somme sono meno di 110 (108 effettivamente visto per prova ed errore nel foglio di calcolo). Quindi, solo nel 5% dei casi uno si aspetta che le somme superiori a 108. Che dire dei dadi in fig. 1? Calcolo della somma dei quadrati delle deviazioni dà il numero 30 che sembra essere un valore tipico come si vede dalla fig. 9. Quindi non c'è nulla di allarmante che rotolano. Si noti che in tutto il precedente l'unica teoria statistica utilizzata è raggruppare le osservazioni e rendendo istogramma e cumulativo relativo diagramma frequenze combinato con "forza bruta", ad esempio gestione di 20000 simulazioni di tirare 60 dadi. Hai bisogno di guardare i tuoi passi a farlo, ma è più un questione di resistenza che di sofisticate teorie matematiche.

## A standard $\chi^2$ - test

Questo tipo di problemi sono stati studiati molto prima della nascita di computer e teorie sono state sviluppate. Solitamente sono gestite mediante cosiddetto  $\chi^2$  - test. Vedi [4] e [5]. Si calcola la somma (divisione per 10 è dovuta a una sorta di normalizzazione)

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

che può essere visualizzata per essere approssimativamente distribuito come  $\chi^2$  - distribuzione con 5 gradi di libertà.

Alzando gli occhi in una tabella per questa distribuzione

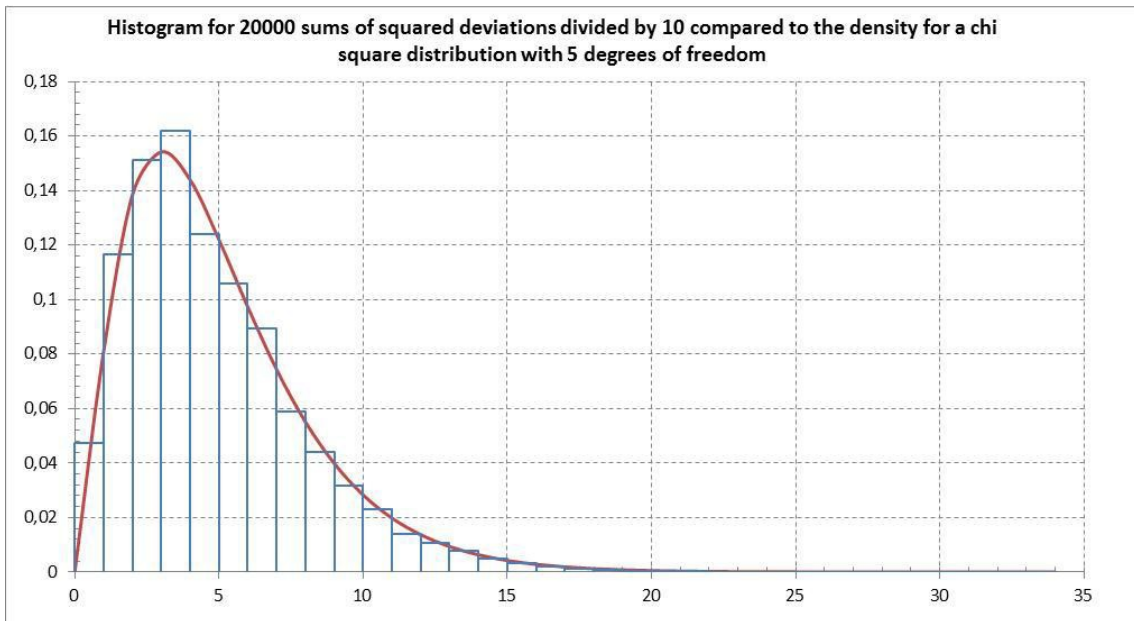
Cumulative probability	Chi square
0,99	15,09
0,98	13,39
0,97	12,37
0,96	11,64
0,95	11,07
0,94	10,60
0,93	10,19
0,92	9,84

Table 1

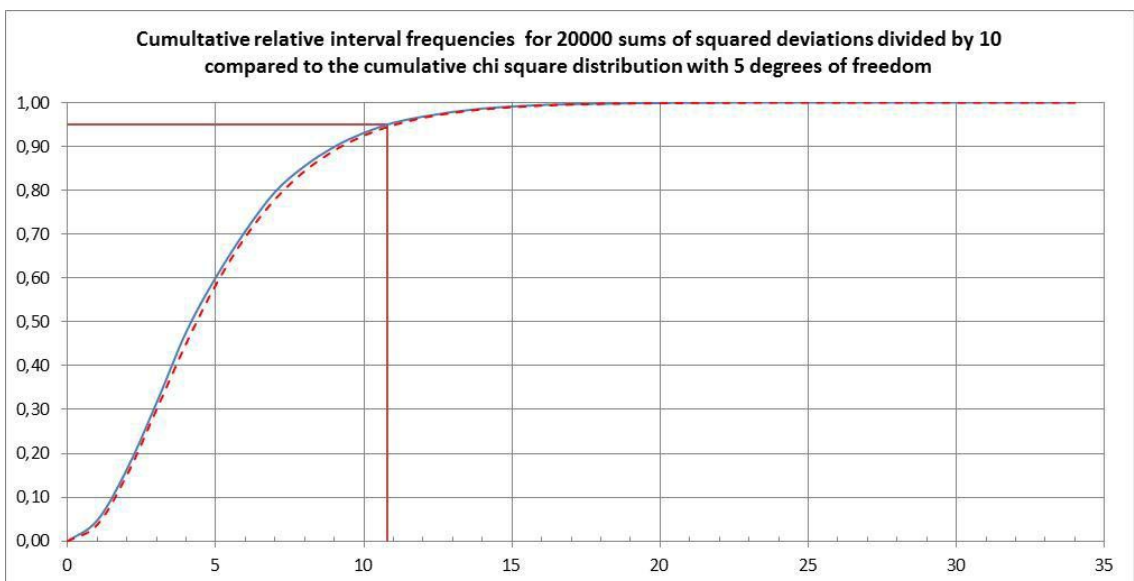
si vede che il 95% di tale distribuzione non dovrebbe superare 11,07.

Così il risultato 108 ottenuto dalla simulazione è OK (ricordare che una divisione per 10 deve essere effettuata in modo è  $10,8 \approx 11$ ) che deve essere confrontato con 11,07.

Nella successiva Figg. 11-12 I confrontare i dati simulati con  $\chi^2$  - distribuzione. Il foglio di calcolo per simulare i dati per questo è ottenuto da quello in Fig. 5 dividendo le formule nella colonna BV con 10 quindi non mi ripeto schermate del foglio elettronico.



**Fig. 11**



**Fig. 12**

## Verifica l'indipendenza delle due categorie

Esaminiamo una situazione più complicata con gli stessi mezzi.

Supponiamo di avere un paese con quattro partiti politici P, Q, R e S. Da ultime elezioni si sa che hanno 40, 30, 20 e 10% dei voti. Si vuole verificare se i redditi delle persone influenza la parte che danno il loro voto o meno. Così si qualifica reddito popoli in tre categorie IC1, IC2 e IC3. I test standard sono stati sviluppati per questa situazione - si veda [4]. Qui il problema è avvicinato simulando un sacco di serie di interviste per avere un'idea di come il  $\chi^2$  - statistica del test è distribuito. Potete trovare le immagini finali in Figg. 23-24.

Se avete intervistato 600 elettori è possibile ordinare il risultato in una tabella di seguito riportata

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

Table 1

Questa situazione è simile la situazione in Fig. 3 dove si deve decidere da un insieme di dati qualcosa è il caso. Nella situazione con i 60 dadi era come se una distribuzione prevista è accettabile come modello, nella situazione con gli elettori se i due criteri può essere considerato indipendente o meno.

L'idea è ora simile alla simulazione dei rotoli 20.000 con 60 dadi per avere un'idea di come tabelle può uscire se i criteri sono davvero indipendenti. Il compito è quello di creare un sacco di simulazioni di tabelle come quello sopra in un modo che garantisce l'indipendenza dei due criteri. Questo possiamo assicurare mediante formule.

Il lavoro è un po 'più complicata rispetto alla situazione dei dadi da gestire, ma comunque può essere fatto, come è mostrato nelle seguenti schermate da Excel. Dimostrazioni video dei passaggi effettivi saranno caricati homepage del progetto.

Il primo passo è quello di creare un foglio elettronico che riprende dove mettere ogni una nuova osservazione alla volta.

	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

Fig. 13 Placing one voter in the scheme with independence between criterias.

	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

Fig. 14 Formulas for Fig. 13

Ora dobbiamo permettere di fare un sacco di "interviste" e riassumere i risultati. Qui la stessa tecnica utilizzata in Math2Earth (vedi [6]) viene utilizzato in modo fare riferimento a quello prima di studiare le schermate successive.

	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							

Fig. 15 Accumulating "voters" simulated in Fig 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)				

Fig. 16 Formulas for Fig. 15.

Ma io voglio simulare un sacco di serie di interviste e da ogni tabella qui sopra usa più righe non è una semplice questione di copia verso il basso. Per fare questo ho riorganizzare le tabelle dalle Figg. 13 e 15 in una riga che può essere fatto da taglio (non copia!) E incollando - Excel e si regola di conseguenza i riferimenti. Forse il modo migliore per mostrare il foglio risultante spread sarebbe di avere una versione caricata .... o forse una dimostrazione video che apparirà sulla homepage del progetto. Ma ogni lettore che è sopravvissuto fino ad oggi è - quindi credo che - in grado di seguire la logica delle formule riportate di seguito in modo che siano qui, senza ulteriori commenti. Esse dovrebbero essere letto quasi come una sorta di poesia con interconnessioni logica.

	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		COLA & C		ROW 1 & 3		ROW 4		ROW 5		ROW 6												
12	Trial nr.	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	0	1	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	1	0	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

Fig. 17 Simulation a lot of independent data tables each organized in it's own row - part 1.



	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
7																			
8																			
9																			
10	$O_{i,j}$																		
11	ROW 9				ROW 10				ROW 11				ROW SUMS			COLUMN SUMS			
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

Fig. 18 Simulation a lot of independent data tables each organized in it's own row - part 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	$E_{i,j}$												$(O_{i,j} - E_{i,j})^2 / E_{i,j}$										$\chi^2$		
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	0,1	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

Fig. 19 Simulation a lot of independent data tables each organized in it's own row - part 3.

	AI	AJ	AK	AL	AM	AN	AO
10							
11							
12	ROW SUMS			COLUMN SUMS			
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

Fig. 20 Formulas for Figs. 17-19. Part 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	=\$AI13*AL13/\$A\$9	=\$AI13*AM13/\$A\$9	=\$AI13*AN13/\$A\$9	=\$AI13*AO13/\$A\$9	=\$AJ13*AL13/\$A\$9	=\$AJ13*AM13/\$A\$9	=\$AJ13*AN13/\$A\$9	=\$AJ13*AO13/\$A\$9	=\$AK13*AL13/\$A\$9	=\$AK13*AM13/\$A\$9	=\$AK13*AN13/\$A\$9
14	=\$AI14*AL14/\$A\$9	=\$AI14*AM14/\$A\$9	=\$AI14*AN14/\$A\$9	=\$AI14*AO14/\$A\$9	=\$AJ14*AL14/\$A\$9	=\$AJ14*AM14/\$A\$9	=\$AJ14*AN14/\$A\$9	=\$AJ14*AO14/\$A\$9	=\$AK14*AL14/\$A\$9	=\$AK14*AM14/\$A\$9	=\$AK14*AN14/\$A\$9
15	=\$AI15*AL15/\$A\$9	=\$AI15*AM15/\$A\$9	=\$AI15*AN15/\$A\$9	=\$AI15*AO15/\$A\$9	=\$AJ15*AL15/\$A\$9	=\$AJ15*AM15/\$A\$9	=\$AJ15*AN15/\$A\$9	=\$AJ15*AO15/\$A\$9	=\$AK15*AL15/\$A\$9	=\$AK15*AM15/\$A\$9	=\$AK15*AN15/\$A\$9
16	=\$AI16*AL16/\$A\$9	=\$AI16*AM16/\$A\$9	=\$AI16*AN16/\$A\$9	=\$AI16*AO16/\$A\$9	=\$AJ16*AL16/\$A\$9	=\$AJ16*AM16/\$A\$9	=\$AJ16*AN16/\$A\$9	=\$AJ16*AO16/\$A\$9	=\$AK16*AL16/\$A\$9	=\$AK16*AM16/\$A\$9	=\$AK16*AN16/\$A\$9
17	=\$AI17*AL17/\$A\$9	=\$AI17*AM17/\$A\$9	=\$AI17*AN17/\$A\$9	=\$AI17*AO17/\$A\$9	=\$AJ17*AL17/\$A\$9	=\$AJ17*AM17/\$A\$9	=\$AJ17*AN17/\$A\$9	=\$AJ17*AO17/\$A\$9	=\$AK17*AL17/\$A\$9	=\$AK17*AM17/\$A\$9	=\$AK17*AN17/\$A\$9
18	=\$AI18*AL18/\$A\$9	=\$AI18*AM18/\$A\$9	=\$AI18*AN18/\$A\$9	=\$AI18*AO18/\$A\$9	=\$AJ18*AL18/\$A\$9	=\$AJ18*AM18/\$A\$9	=\$AJ18*AN18/\$A\$9	=\$AJ18*AO18/\$A\$9	=\$AK18*AL18/\$A\$9	=\$AK18*AM18/\$A\$9	=\$AK18*AN18/\$A\$9
19	=\$AI19*AL19/\$A\$9	=\$AI19*AM19/\$A\$9	=\$AI19*AN19/\$A\$9	=\$AI19*AO19/\$A\$9	=\$AJ19*AL19/\$A\$9	=\$AJ19*AM19/\$A\$9	=\$AJ19*AN19/\$A\$9	=\$AJ19*AO19/\$A\$9	=\$AK19*AL19/\$A\$9	=\$AK19*AM19/\$A\$9	=\$AK19*AN19/\$A\$9
20	=\$AI20*AL20/\$A\$9	=\$AI20*AM20/\$A\$9	=\$AI20*AN20/\$A\$9	=\$AI20*AO20/\$A\$9	=\$AJ20*AL20/\$A\$9	=\$AJ20*AM20/\$A\$9	=\$AJ20*AN20/\$A\$9	=\$AJ20*AO20/\$A\$9	=\$AK20*AL20/\$A\$9	=\$AK20*AM20/\$A\$9	=\$AK20*AN20/\$A\$9
21	=\$AI21*AL21/\$A\$9	=\$AI21*AM21/\$A\$9	=\$AI21*AN21/\$A\$9	=\$AI21*AO21/\$A\$9	=\$AJ21*AL21/\$A\$9	=\$AJ21*AM21/\$A\$9	=\$AJ21*AN21/\$A\$9	=\$AJ21*AO21/\$A\$9	=\$AK21*AL21/\$A\$9	=\$AK21*AM21/\$A\$9	=\$AK21*AN21/\$A\$9

Fig. 21 Formulas for Figs. 17-19. Part 2.

	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
10													
11						$(O_{ij} - E_{ij})^2/E_{ij}$							$\chi^2$
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-AU)=(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-AU)=(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-AU)=(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-AU)=(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-AU)=(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-AU)=(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-AU)=(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-AU)=(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-AU)=(AC21-AV)=(AD21-AW)=(AE21-AX)=(AF21-AY)=(AG21-AZ)=(AH21-BA)=SUM(BB21:BM21)										

Fig. 22 Formulas for Figs. 17-19. Part 3.

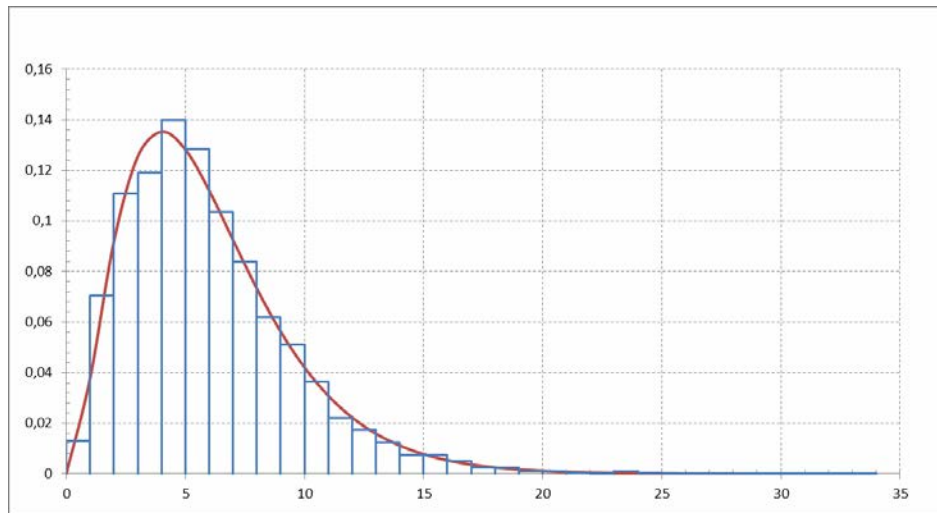


Fig. 23 Histogram for 2000 simulated interviews of 600 people compared to density for  $\chi^2$  - distribution with  $(3-1)(4-1) = 6$  degrees of freedom. The data are produced in column BN in Fig. 16.

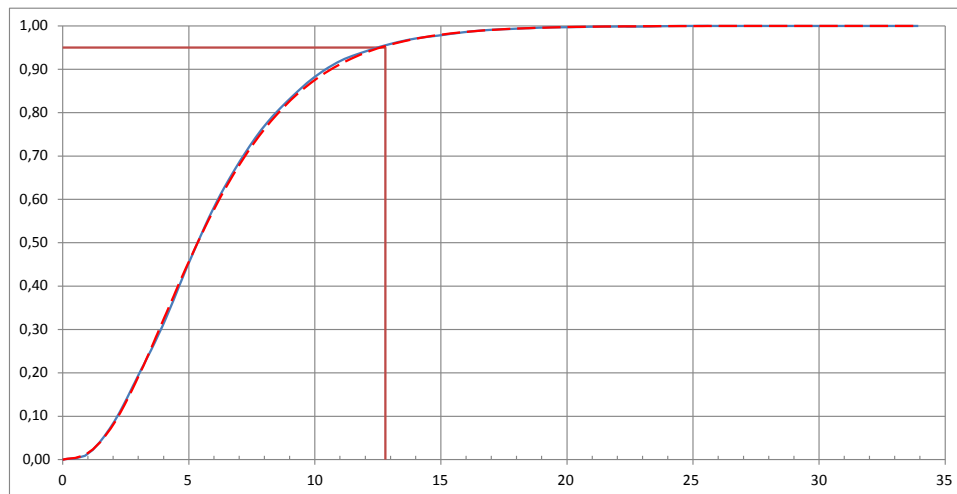


Fig. 24 Cumulated relative frequencies from Fig. 23 compared to  $\chi^2$  - distribution with 6 degrees of freedom.

## References

- [1] <http://en.wikipedia.org/wiki/Dice> (February 2012)
- [2] John Andersen: Dynamical simulation of stochastic phenomena using Excel (with link in final version)
- [3] [http://www.efunda.com/math/least\\_squares/least\\_squares.cfm](http://www.efunda.com/math/least_squares/least_squares.cfm) (December 2011)
- [4] [http://en.wikipedia.org/wiki/Pearson%27s\\_chi-squared\\_test](http://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test) (December 2011)
- [5] <http://www.merriam-webster.com/dictionary/chi> (February 2012)
- [6] John Andersen: Animation – bringing motion into mathematics <http://www.math2earth.org/> (December 2011)