

Simulation af χ^2 - fordeling

John Andersen

Introduktion

En dag kastede jeg 60 terninger



Fig. 1 Billede af de 60 terninger på mit skrivebord

For at danne mig et billede af hyppighederne flyttede jeg rundt på terningerne som vist på figur 2:



Fig. 2 En måde at repræsentere resultatet af kast med 60 terninger

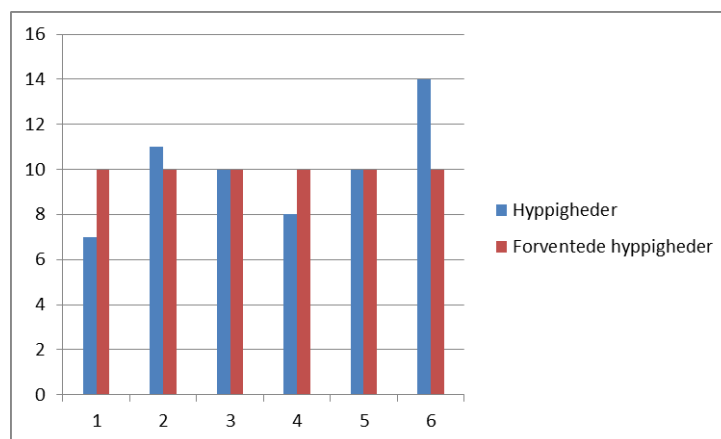


Fig. 3 Søjlediagram over terningerne fra Fig. 2

Ved hjælp af et Excel-regneark simulerede jeg derpå et kast med 6000 terninger og fik søjlediagrammet på figur 3:

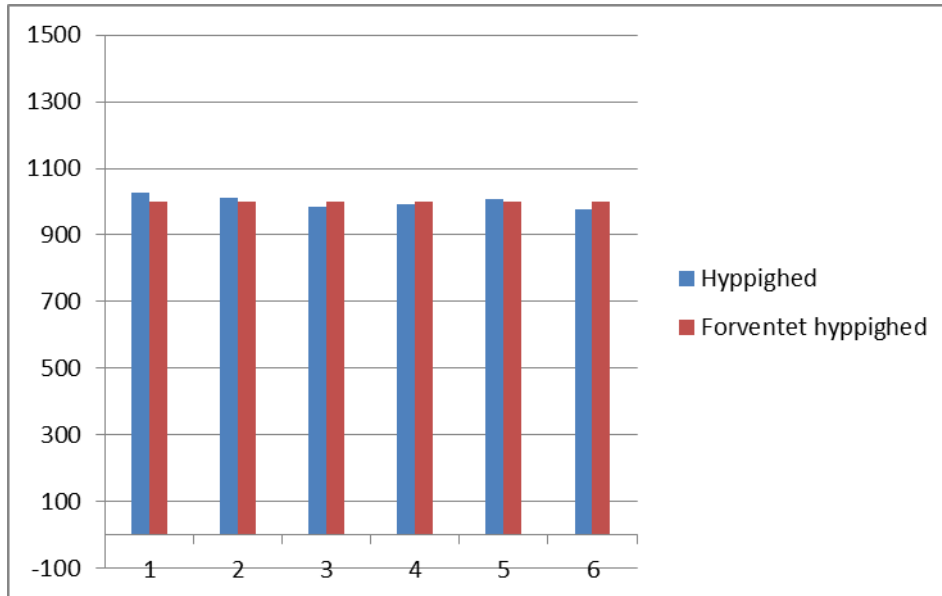


Fig. 4 Søjlediagram der repræsenterer resultatet af en computersimulering af kast med 6000 terninger

Som man ser stemmer hyppighederne fra det kastet med 6000 terninger godt overens med en forventning om at der er ca. en sjettedel af terningerne der viser "1", cirka en sjettedel der viser "2" osv. Jo flere terninger desto bedre overensstemmelse med forventningerne hvilket er helt i overensstemmelse med de store tals lov som beskrevet i [2].

Men hvad nu, hvis man ud fra situationen med kun 60 terninger skulle afgøre om afvigelsen fra det forventede var passende eller for stor? Her kan en brug af grafikken komme til kort, for hvordan ser man om variationerne er for store eller inden for normalområdet? Vi skal i det følgende se på metoder til at behandle den slags vurderinger.

En „hjemmelavet“ metode

Hvis man kender "mindste kvadraters metode" fra lineær regression eller lignende kan man måske føle sig inspireret til at se på en sum som denne

$$\sum (faktisk\ udfald - forventet\ udfald)^2$$

eller for at være mere præcis

$$(1) \quad \sum_{k=1}^6 (hyppighed\ af\ k - 10)^2$$

Hvis denne sum er nul svarer det til at alle seks hyppigheder er 10 hvilket nok er for godt til at være sandt (der sikkert blevet manipuleret med data). Hvis summen (1) derimod er alt for stor kan det tyde på at hyppighederne afviger for meget fra det forventede til at vi tør tro på at terningen er ægte.

For at få en idé om, hvad man kan forvente af summer som (1) kan man simulere en masse kast med 60 ægte terninger med tilhørende beregninger af summen (1). Så kan man prøve at få et overblik over fordelingen af disse summer ved hjælp af metoder fra den deskriptive statistik.

Nedenfor på figur 5 ses et udsnit af et regneark hvor en sådan simulering er foretaget. Figurene 6-8 viser formlerne. Nedenstående punkter forklarer princippet bag opbygningen af regnearket:

- Kolonne A viser numrene på kastene med de 60 terninger
- Kolonnerne B:BI (af pladshensyn er en del kolonner skjult i visningen) viser resultaterne for de enkelte terninger. Række 2 viser nummeret på hver af de 60 terninger.
- Kolonnerne BJ:BO viser hyppighederne for de 6 mulige øjental. Mulighederne ses i række 2.
- Kolonnerne BP:BU beregner de kvadrerede afvigelser mellem hyppighederne og de forventede 10.
- Kolonne BV viser summerne af de kvadrerede afvigelser – altså summerne svarende til (1)

	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		Øjne af 60 terninger											Hyppigheder af øjne						Kvadrater på afvigelser : (hyppighed - 10) ²						
2	Terning nr. → Kast nr. ↓	1	2	3	4	5	6	7	57	58	59	60	1	2	3	4	5	6	1	2	3	4	5	6	Summer af kvadrater på afvigelser
3	1	3	4	6	5	5	3	3	2	1	5	1	12	8	11	9	11	9	4	4	1	1	1	1	12
4	2	2	6	6	1	2	6	6	6	3	2	5	8	13	8	4	9	18	4	9	4	36	1	64	118
5	3	1	5	3	6	4	2	6	4	1	3	6	8	8	13	7	6	18	4	4	9	9	16	64	106
6	4	5	2	2	2	6	5	3	2	6	5	3	10	13	11	9	10	7	0	9	1	1	0	9	20
7	5	1	5	4	2	1	4	4	4	3	4	6	9	8	10	14	12	7	1	4	0	16	4	9	34
8	6	1	1	6	2	5	1	1	4	3	1	5	19	5	10	9	7	10	81	25	0	1	9	0	116
9	7	5	6	1	6	4	1	4	3	3	2	6	9	13	4	13	10	11	1	9	36	9	0	1	56
10	8	2	3	4	6	2	6	3	3	1	2	1	8	14	13	11	4	10	4	16	9	1	36	0	66
11	9	6	1	1	1	2	2	3	2	1	4	1	11	9	9	12	7	12	1	1	1	4	9	4	20
12	10	2	5	3	5	3	3	5	1	2	1	1	7	14	8	7	14	10	9	16	4	9	16	0	54
13	11	6	5	6	5	2	5	2	6	3	4	1	2	17	10	9	6	16	64	49	0	1	16	36	166

Fig. 5 Regneark med simulering af en masse kast med 60 terninger.

Læg mærke til at så snart man har indtastet formlerne i række 3 så kan resten af regnearket opbygges ved at kopiere række 3 nedad. I det konkrete tilfælde har jeg kopieret nedad så jeg får simuleret 20000 kast med 60 terninger hvilket må formodes at give et rimeligt indtryk af hvorledes summerne fordeler sig.

Der er 2²⁰ rækker at gøre godt med i min version af Excel og min computers arbejdshukommelse slipper op længe før jeg har udnyttet alle disse rækker. Så hvis man vil simulere millioner af kast må man ty til andre programmer end Excel, men lad os nu se hvor langt vi kan komme med Excel. Mon ikke det er tilstrækkeligt til at vi kan fatte pointen?

På de næste tre figurer ses formlerne bag figur 5.

Jeg gå ikke yderligere i detaljer med formlerne. Nærlæsning af formlerne skulle gerne være tilstrækkeligt til at den interesserede læser opdager tankegangen bag opbygningen. Forstør billedet hvis du har svært ved at se hvad der står.

	A	B	C	D	E	F
1					Øjne af 60 terninger	
	Terning nr. → Kast nr. ↓					
2		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 Formler til Fig. 5 – første del.

	BF	BG	BH	BI	BJ	BK	BL	BM
1					Hyppigheder af øjne			
2	=BE2+1	=BF2+1	=BG2+1	=BH2+1	1	2	3	4
3	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B3:\$B3;BJS\$2)	=TÆL.HVVIS(\$B3:\$B3;BKS\$2)	=TÆL.HVVIS(\$B3:\$B3;BLS\$2)	=TÆL.HVVIS(\$B3:\$B3;BMS\$2)			
4	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B4:\$B4;BJS\$2)	=TÆL.HVVIS(\$B4:\$B4;BKS\$2)	=TÆL.HVVIS(\$B4:\$B4;BLS\$2)	=TÆL.HVVIS(\$B4:\$B4;BMS\$2)			
5	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B5:\$B5;BJS\$2)	=TÆL.HVVIS(\$B5:\$B5;BKS\$2)	=TÆL.HVVIS(\$B5:\$B5;BLS\$2)	=TÆL.HVVIS(\$B5:\$B5;BMS\$2)			
6	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B6:\$B6;BJS\$2)	=TÆL.HVVIS(\$B6:\$B6;BKS\$2)	=TÆL.HVVIS(\$B6:\$B6;BLS\$2)	=TÆL.HVVIS(\$B6:\$B6;BMS\$2)			
7	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B7:\$B7;BJS\$2)	=TÆL.HVVIS(\$B7:\$B7;BKS\$2)	=TÆL.HVVIS(\$B7:\$B7;BLS\$2)	=TÆL.HVVIS(\$B7:\$B7;BMS\$2)			
8	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B8:\$B8;BJS\$2)	=TÆL.HVVIS(\$B8:\$B8;BKS\$2)	=TÆL.HVVIS(\$B8:\$B8;BLS\$2)	=TÆL.HVVIS(\$B8:\$B8;BMS\$2)			
9	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B9:\$B9;BJS\$2)	=TÆL.HVVIS(\$B9:\$B9;BKS\$2)	=TÆL.HVVIS(\$B9:\$B9;BLS\$2)	=TÆL.HVVIS(\$B9:\$B9;BMS\$2)			
10	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B10:\$B10;BJS\$2)	=TÆL.HVVIS(\$B10:\$B10;BKS\$2)	=TÆL.HVVIS(\$B10:\$B10;BLS\$2)	=TÆL.HVVIS(\$B10:\$B10;BMS\$2)			
11	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B11:\$B11;BJS\$2)	=TÆL.HVVIS(\$B11:\$B11;BKS\$2)	=TÆL.HVVIS(\$B11:\$B11;BLS\$2)	=TÆL.HVVIS(\$B11:\$B11;BMS\$2)			
12	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B12:\$B12;BJS\$2)	=TÆL.HVVIS(\$B12:\$B12;BKS\$2)	=TÆL.HVVIS(\$B12:\$B12;BLS\$2)	=TÆL.HVVIS(\$B12:\$B12;BMS\$2)			
13	=SLUMP=SLUMP=SLUMP=SLUMPMELLEM(1;6)	=TÆL.HVVIS(\$B13:\$B13;BJS\$2)	=TÆL.HVVIS(\$B13:\$B13;BKS\$2)	=TÆL.HVVIS(\$B13:\$B13;BLS\$2)	=TÆL.HVVIS(\$B13:\$B13;BMS\$2)			

Fig. 7 Formler til Fig. 5 – anden del. (Forstør for at se detaljer)

	BO	BP	BQ	BR	BS	BT	BU	BV
1		Kvadrater på afvigelses : (hyppighed - 10) ²						Summer af kvadrater på afvigelser
2	6	1	2	3	4	5	6	
3	=TÆL.HVVIS(\$B3:\$B3;BO\$2)	=(B3-10)^2	=(BK3-10)^2	=(BL3-10)^2	=(BM3-10)^2	=(BN3-10)^2	=(BO3-10)^2	=SUM(BP3:BU3)
4	=TÆL.HVVIS(\$B4:\$B4;BO\$2)	=(B4-10)^2	=(BK4-10)^2	=(BL4-10)^2	=(BM4-10)^2	=(BN4-10)^2	=(BO4-10)^2	=SUM(BP4:BU4)
5	=TÆL.HVVIS(\$B5:\$B5;BO\$2)	=(B5-10)^2	=(BK5-10)^2	=(BL5-10)^2	=(BM5-10)^2	=(BN5-10)^2	=(BO5-10)^2	=SUM(BP5:BU5)
6	=TÆL.HVVIS(\$B6:\$B6;BO\$2)	=(B6-10)^2	=(BK6-10)^2	=(BL6-10)^2	=(BM6-10)^2	=(BN6-10)^2	=(BO6-10)^2	=SUM(BP6:BU6)
7	=TÆL.HVVIS(\$B7:\$B7;BO\$2)	=(B7-10)^2	=(BK7-10)^2	=(BL7-10)^2	=(BM7-10)^2	=(BN7-10)^2	=(BO7-10)^2	=SUM(BP7:BU7)
8	=TÆL.HVVIS(\$B8:\$B8;BO\$2)	=(B8-10)^2	=(BK8-10)^2	=(BL8-10)^2	=(BM8-10)^2	=(BN8-10)^2	=(BO8-10)^2	=SUM(BP8:BU8)
9	=TÆL.HVVIS(\$B9:\$B9;BO\$2)	=(B9-10)^2	=(BK9-10)^2	=(BL9-10)^2	=(BM9-10)^2	=(BN9-10)^2	=(BO9-10)^2	=SUM(BP9:BU9)
10	=TÆL.HVVIS(\$B10:\$B10;BO\$2)	=(B10-10)^2	=(BK10-10)^2	=(BL10-10)^2	=(BM10-10)^2	=(BN10-10)^2	=(BO10-10)^2	=SUM(BP10:BU10)
11	=TÆL.HVVIS(\$B11:\$B11;BO\$2)	=(B11-10)^2	=(BK11-10)^2	=(BL11-10)^2	=(BM11-10)^2	=(BN11-10)^2	=(BO11-10)^2	=SUM(BP11:BU11)
12	=TÆL.HVVIS(\$B12:\$B12;BO\$2)	=(B12-10)^2	=(BK12-10)^2	=(BL12-10)^2	=(BM12-10)^2	=(BN12-10)^2	=(BO12-10)^2	=SUM(BP12:BU12)
13	=TÆL.HVVIS(\$B13:\$B13;BO\$2)	=(B13-10)^2	=(BK13-10)^2	=(BL13-10)^2	=(BM13-10)^2	=(BN13-10)^2	=(BO13-10)^2	=SUM(BP13:BU13)

Fig. 8 Formler til Fig. 5 – tredje del. (Forstør for at se detaljer)

De 20000 summer af kvadrerede afvigelses der står i kolonne BV (se figur 5) udgør nu det datamateriale jeg vil have et overblik over. Derfor tegner jeg ved hjælp af Excel histogram og sumkurve over dette observationsæt. Se figurerne 9 og 10.

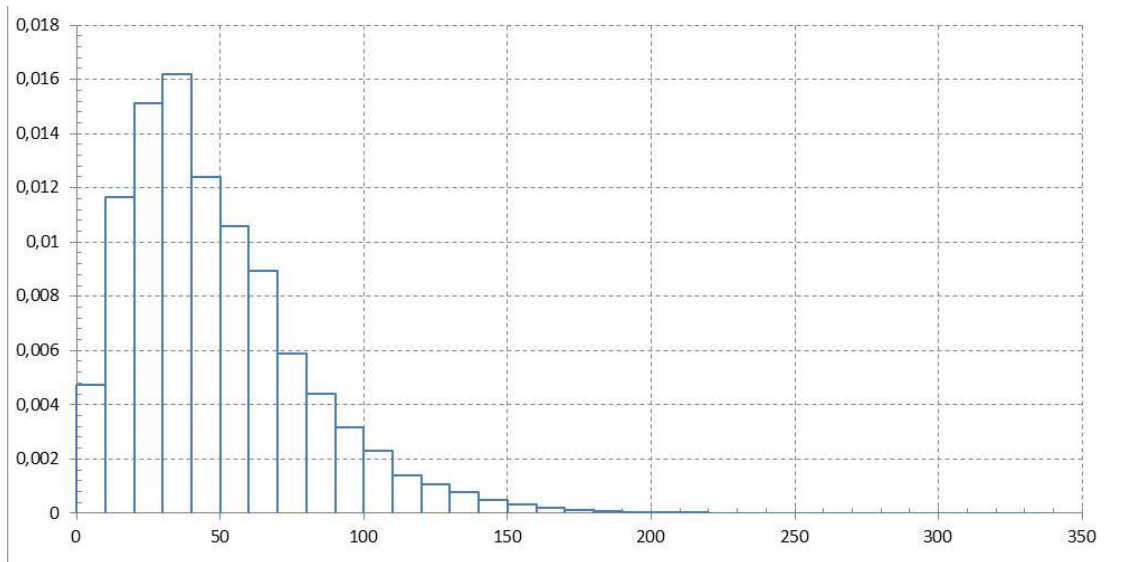


Fig. 9 Histogram over 20000 summer af kvadratiske afvigelser.

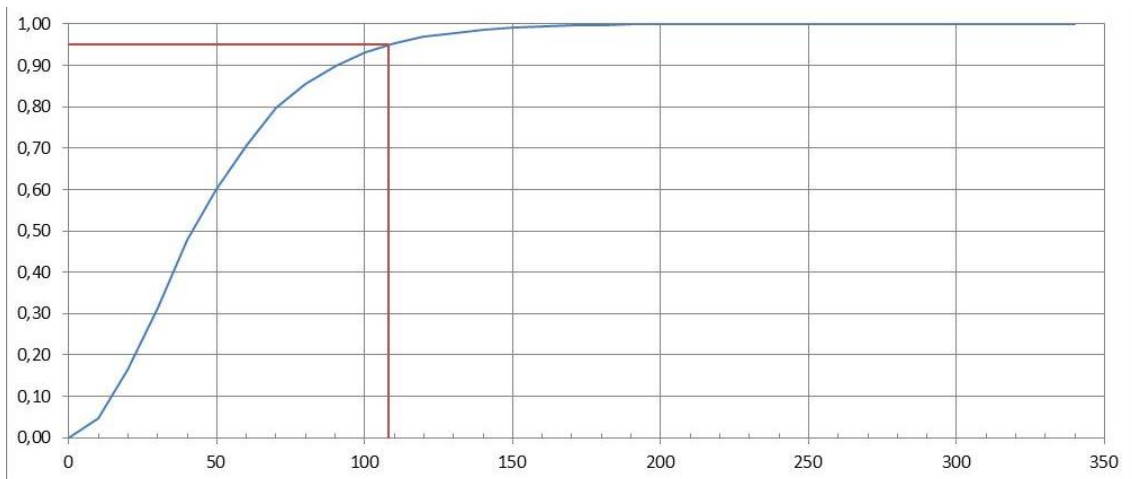


Fig. 10 Sumkurve over 20000 summer af kvadratiske afvigelser.

Man kan se at cirka 95% af summerne er mindre end 110 (108 helt præcis hvilket ses ved at prøve sig frem i regnearket). Så kun i 5% af alle tilfælde vil vi forvente summer over 108.

Hvad med terningerne i Fig. 1? For det kast kan man beregne summe (1) til af være 30. Det er øjensynlig en meget typisk værdi for den slags summer så der er vist ikke noget alarmerende ved det kast. Så de terninger er formodentlig i orden. Med mindre man har mistanke om at der alligevel er noget lusk ved fordelingen mellem 1-ere og 6-ere. F.eks. kunne man vælge at gå nærmere ind på at se på 6-erne separat, men det vil jeg ikke komme ind på i nærværende fremstilling.

Bemærk at i det foregående er stort set kun brugt metoder fra den deskriptive statistik – nærmere betegnet histogram og sumkurve – kombineret med ”rå regnekraft”. Man skal selvfølgelig være omhyggelig når regnearket bygges op, men der er mere et spørgsmål om udholdenhed end om brug af sofistikerede matematiske metoder. Det er netop en af pointerne ved denne fremstilling.

Standardiseret χ^2 - test

Problemstillinger af ovenstående slags har været undersøgt længe for computere var opfundet. Der er udviklet metoder som ikke kræver så meget regnekraft. Til gengæld er matematikken bagved noget mere krævende. Det drejer sig bl.a. om de såkaldte χ^2 - test. Se [4] og [5]. Her vil jeg justere ovenstående så det kommer i overensstemmelse med standardmetoderne.

Det har vist sig at det letter den bagved liggende matematik at man ser på følgende sum

$$\sum \frac{(\text{faktisk udfald} - \text{forventet udfald})^2}{\text{forventet udfald}}$$

så man i stedet for summen (1) ovenfor skal beregne summen

$$(1.2) \quad \sum_{k=1}^6 \frac{(\text{hyppighed af } k - 10)^2}{10}$$

Man kan vise at den tilnærmelsesvis er χ^2 - fordelt med 5 frihedsgrader. Tabelopslag giver følgende:

Brøkdel	Invers χ^2 - fordeling
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

Tabel 1

Man ser at 95 % af sådanne summer skal ligge under 11,07.

Det stemmer fint med de 108 vi fik tidligere ved simulering, for divideres 108 med 10 får vi netop $10,8 \approx 11 \approx 11,07$.

På figurerne 11 og 12 sammenlignes simulerede data med χ^2 - fordelingen. Regnearket der er brugt ligner det der tidligere er brugt bortset fra at der i kolonne BV er divideret med 10, så der henvises til figurerne 5-8 men henblik på detaljerne.

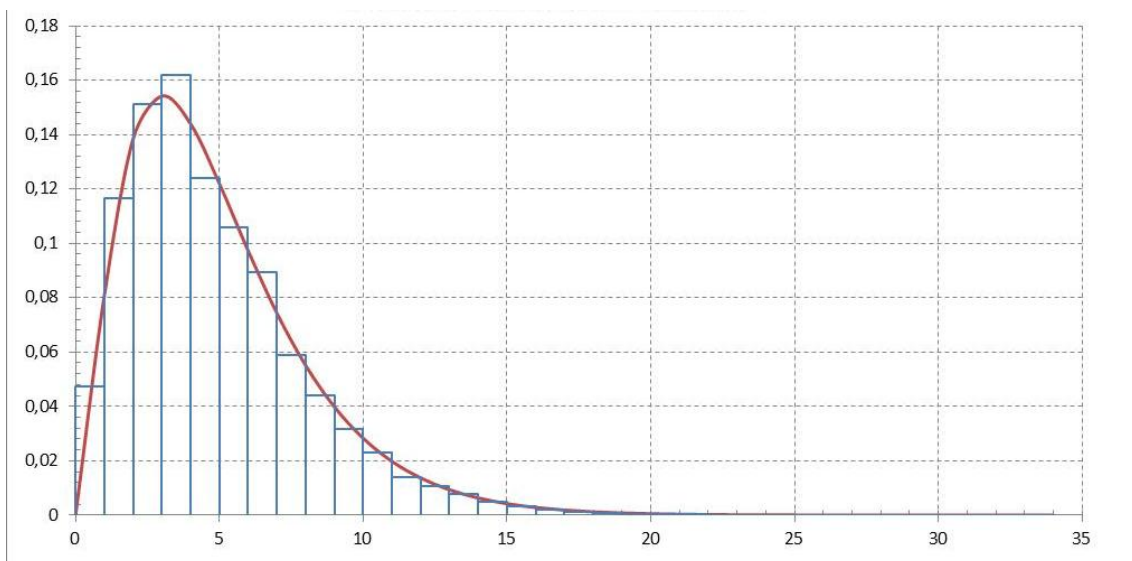


Fig. 11 Histogram over 20000 kvadratiske afvigelser divideret med 10 sammenlignet med sandsynlighedstætheden for χ^2 - fordelingen med 5 frihedsgrader.

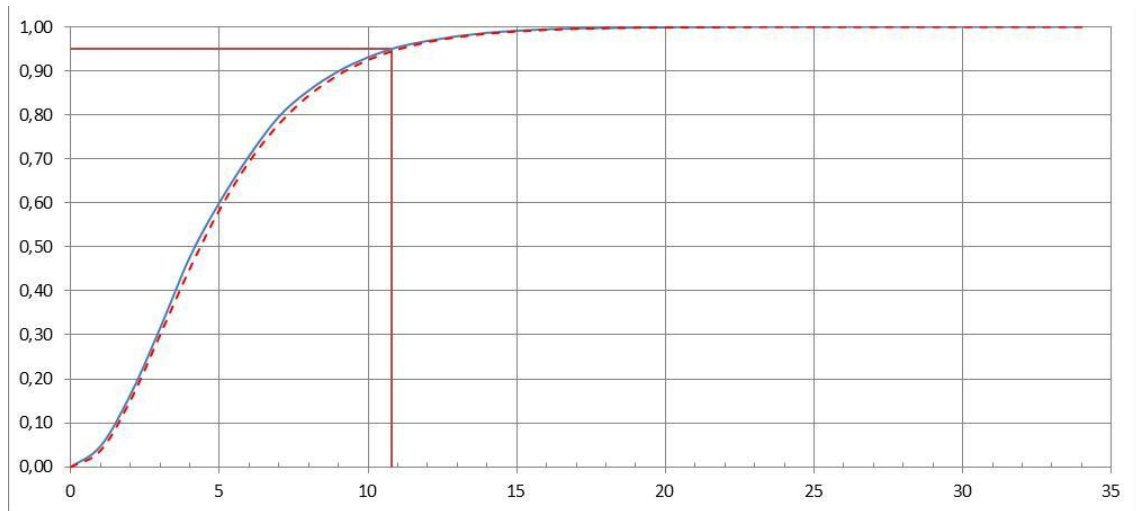


Fig. 12 Sumkurve for 20000 kvadratiske afvigelser divideret med 10 sammenlignet med fordelingsfunktionen for χ^2 – fordelingen med 5 frihedsgrader.

Test af uafhængighed af to kategorier

Vi skal nu se på et lidt mere kompliceret eksempel.

Forestil dig et land med fire politiske partier P, Q, R og S. Ved sidste valg fik de 40, 30, 20 og 10 % af stemmerne. Vi vil undersøge om folks indkomst har indflydelse på hvilket parti de stemmer på. Vi deler befolkningen op i tre indkomstkategorier I1, I2 og I3. Der er udviklet standardtest til at tjekke en sådan situation – se f.eks. [4]. Her behandles problemstillingen ved simulering i regneark.

Et analyseinstitut sender konsulenter ud for at interviewe 600 tilfældigt udvalgte stemmeberettigede borgere og resultatet samles i Tabel 1:

	P	Q	R	S
I1	110	96	53	23
I2	70	60	41	22
I3	42	41	18	24

Tabel 1

Hvordan kan vi nu ud fra disse data afgøre om indkomsterne har indflydelse på stemmeprocenterne?

Vi griber det an ved at skaffe os en ide om, hvordan data typisk vil fordele sig, hvis indkomsterne ingen indflydelse har på procenterne.

Det gør vi ved at simulere 2000 interviewundersøgelser af 600 personer hvor vi har sikret os at det drejer sig om uafhængige kategorier.

Først laver jeg et regneark der simulerer indplaceringen af en tilfældigt udvalgt vælger.

I eksemplet antager jeg at fordelingen mellem kategorierne P, Q, R og S er 40%, 30%, 20% og 10% medens fordelingen mellem kategorierne IC1, IC2 og IC3 er 50%, 30% og 20%.

Jeg bruger to af hinanden uafhængige SLUMPMELLE(1;100) i cellerne C2 og B3. Det tilfældige tal i C2 bruges til at placere ”vælgeren” i en tilfældig af kategorierne P, Q, R og S i overensstemmelse med procenterne for disse kategorier. Der skrives ”1” i den relevante af cellerne D3:G3 og ”0” i de tre andre. Uafhængigt af dette placeres ”vælgeren” i en af kategorierne IC1, IC2 og IC3 ud fra det tilfældige tal i B3 og resultatet markeres med ”1” i den relevante celle blandt C4:C6.

Se detaljerne i Figur 13 og 14.

	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 Placering af en vælger i skemaet ud fra uafhængighed mellem kategorierne.

	A	B	C	D	E	F	G
1				0,4	0,3	0,2	0,1
2			=SLUMPMELLE(1;100)	P	Q	R	S
3		=SLUMPMELLE(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 Formler til Fig. 13

Næste skridt er at justere regnearket så man kan opsummerer resultaterne af de enkelte interviews i tabellen. Her bruges samme teknik som i afsnittet "Kast med 5 terninger" i [2] så læs lige dette afsnit igen.

	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 Opsamling af de enkelte "interviews". Her er vi nået til person nr. 336 (celle A11)

	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 Formler til den del af Fig. 15 der ikke er med i Fig. 13 og 14.

Hvert interview simuleres med tryk på F9. Der skal således 600 tryk for at fremstille en tabel svarende til Tabel 1. Hvis jeg vil have en ide om hvordan tabeller som denne typisk vil se ud må jeg så gennemføre 600 "interview" mange gange – flere tusinde hvis jeg vil have glæde af de store tals love så jeg kan få solide statistikker. Det holder jo ikke en meter, så jeg må justere mit regneark for at kunne simulere et stort antal interviewserier á 600 interviews i ét hug.

For at kunne lave det vælger jeg at bryde beregningerne i regnearket i Fig. 15 op og placere formlerne til hver tabel på en lang række, så jeg kan kopiere nedad og få simuleret en masse tabeller i en fart.

Det sker ved at jeg klipper (ikke kopierer) formlerne til Fig. 15 og indsætter dem på række. Når jeg vælger "klip" og "indsæt" justerer Excel selv cellehenvisningerne. Det er meget svært at forklare detaljerne i dette med ord. Derfor henviser jeg den interesserede – og vedholdende – læser det færdige regneark [3].

	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	0	1	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

Fig. 17 Simulering af en masse uafhængige datatabeller, hver især organiseret på en række – del 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 Simulering af en masse uafhængige datatabeller, hver især organiseret på en række – del 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}												$(O_{i,j} - E_{i,j})^2 / E_{i,j}$										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

Fig. 19 Simulering af en masse uafhængige datatabeller, hver især organiseret på en række – del 3.

	AI	AJ	AK	AL	AM	AN	AO
10							
11	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 Formler til Figureerne. 17-19 - del 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/\$AS9	=\$AI13*AM13/\$AS9	=\$AI13*AN13/\$AS9	=\$AI13*AO13/\$AS9	=\$AJ13*AL13/\$AS9	=\$AJ13*AM13/\$AS9	=\$AJ13*AN13/\$AS9	=\$AJ13*AO13/\$AS9	=\$AK13*AL13/\$AS9	=\$AK13*AM13/\$AS9	=\$AK13*AN13/\$AS9
14	=\$AI14*AL14/\$AS9	=\$AI14*AM14/\$AS9	=\$AI14*AN14/\$AS9	=\$AI14*AO14/\$AS9	=\$AJ14*AL14/\$AS9	=\$AJ14*AM14/\$AS9	=\$AJ14*AN14/\$AS9	=\$AJ14*AO14/\$AS9	=\$AK14*AL14/\$AS9	=\$AK14*AM14/\$AS9	=\$AK14*AN14/\$AS9
15	=\$AI15*AL15/\$AS9	=\$AI15*AM15/\$AS9	=\$AI15*AN15/\$AS9	=\$AI15*AO15/\$AS9	=\$AJ15*AL15/\$AS9	=\$AJ15*AM15/\$AS9	=\$AJ15*AN15/\$AS9	=\$AJ15*AO15/\$AS9	=\$AK15*AL15/\$AS9	=\$AK15*AM15/\$AS9	=\$AK15*AN15/\$AS9
16	=\$AI16*AL16/\$AS9	=\$AI16*AM16/\$AS9	=\$AI16*AN16/\$AS9	=\$AI16*AO16/\$AS9	=\$AJ16*AL16/\$AS9	=\$AJ16*AM16/\$AS9	=\$AJ16*AN16/\$AS9	=\$AJ16*AO16/\$AS9	=\$AK16*AL16/\$AS9	=\$AK16*AM16/\$AS9	=\$AK16*AN16/\$AS9
17	=\$AI17*AL17/\$AS9	=\$AI17*AM17/\$AS9	=\$AI17*AN17/\$AS9	=\$AI17*AO17/\$AS9	=\$AJ17*AL17/\$AS9	=\$AJ17*AM17/\$AS9	=\$AJ17*AN17/\$AS9	=\$AJ17*AO17/\$AS9	=\$AK17*AL17/\$AS9	=\$AK17*AM17/\$AS9	=\$AK17*AN17/\$AS9
18	=\$AI18*AL18/\$AS9	=\$AI18*AM18/\$AS9	=\$AI18*AN18/\$AS9	=\$AI18*AO18/\$AS9	=\$AJ18*AL18/\$AS9	=\$AJ18*AM18/\$AS9	=\$AJ18*AN18/\$AS9	=\$AJ18*AO18/\$AS9	=\$AK18*AL18/\$AS9	=\$AK18*AM18/\$AS9	=\$AK18*AN18/\$AS9
19	=\$AI19*AL19/\$AS9	=\$AI19*AM19/\$AS9	=\$AI19*AN19/\$AS9	=\$AI19*AO19/\$AS9	=\$AJ19*AL19/\$AS9	=\$AJ19*AM19/\$AS9	=\$AJ19*AN19/\$AS9	=\$AJ19*AO19/\$AS9	=\$AK19*AL19/\$AS9	=\$AK19*AM19/\$AS9	=\$AK19*AN19/\$AS9
20	=\$AI20*AL20/\$AS9	=\$AI20*AM20/\$AS9	=\$AI20*AN20/\$AS9	=\$AI20*AO20/\$AS9	=\$AJ20*AL20/\$AS9	=\$AJ20*AM20/\$AS9	=\$AJ20*AN20/\$AS9	=\$AJ20*AO20/\$AS9	=\$AK20*AL20/\$AS9	=\$AK20*AM20/\$AS9	=\$AK20*AN20/\$AS9
21	=\$AI21*AL21/\$AS9	=\$AI21*AM21/\$AS9	=\$AI21*AN21/\$AS9	=\$AI21*AO21/\$AS9	=\$AJ21*AL21/\$AS9	=\$AJ21*AM21/\$AS9	=\$AJ21*AN21/\$AS9	=\$AJ21*AO21/\$AS9	=\$AK21*AL21/\$AS9	=\$AK21*AM21/\$AS9	=\$AK21*AN21/\$AS9

Fig. 21 Formler til Figurene 17-19 - del 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-AV)=(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-AV)=(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-AV)=(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-AV)=(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-AV)=(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-AV)=(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-AV)=(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-AV)=(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-AV)=(AE21-AX)=(AF21-AY)=(AG21-AZ)=(AH21-BA)=SUM(BB21:BM21)										

Fig. 22 Formler til Figurene 17-19 - del 3.

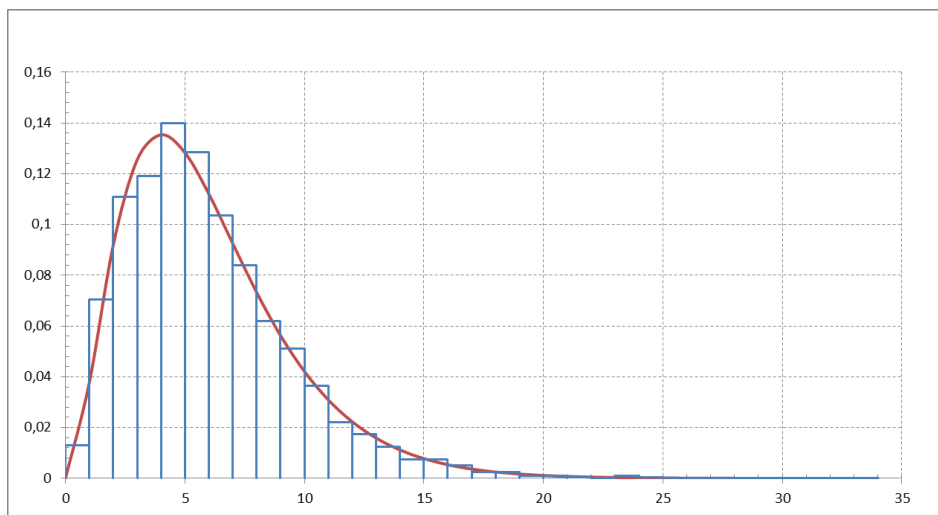


Fig. 23 Histogram over 2000 simuleringer af interviews med 600 personer sammenlignet med tæthedsfunktionen for χ^2 – fordelingen med $(3-1)(4-1) = 6$ frihedsgrader. Data til histogrammet fremkommer i kolonne BN på Fig. 16.

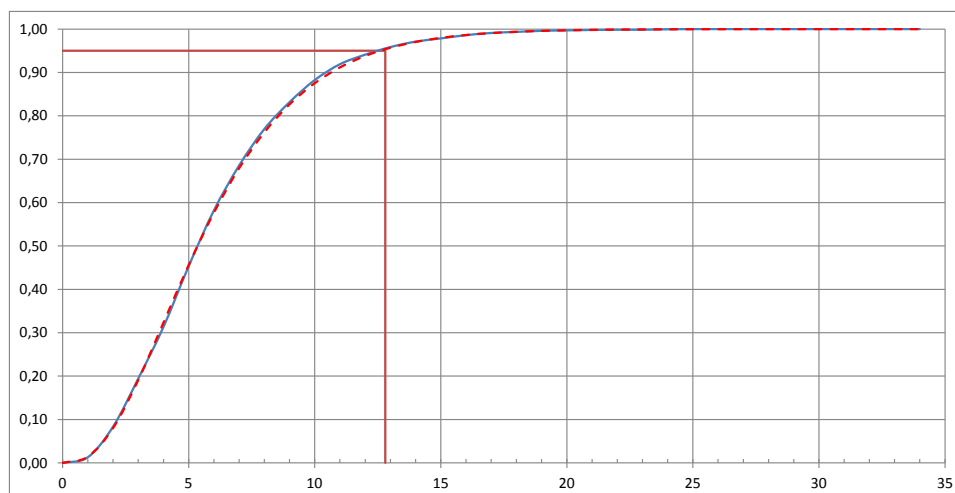


Fig. 24 Summerede frekvenser fra Fig. 23 sammenlignet med χ^2 – fordelingen med 6 frihedsgrader.

References

- [1] <http://en.wikipedia.org/wiki/Dice> (February 2012)
- [2] John Andersen: Simulering af stokastiske fænomener med Excel
- [3] <http://www.efunda.com/math/leastquares/leastquares.cfm> (December 2011)
- [4] 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)