Hreint meðaltal og normaldreifing

Eitt kast tíu teninga gaf eftirfarandi niðurstöðu.

Mynd 1 Útkoma úr kasti 10 teninga.

Með því að leggja saman fjölda augna á sem upp komu á teningunum og deila með 10 fæst venjulega meðaltalið

$$\frac{1+1+3+6+1+5+5+5+1+6}{10} = \frac{17}{5} = 3.4$$

Maður getur kastað þessum 10 teningum aftur og aftur og fengið fjöldann allan af meðaltölum sem öll munu tilheyra talnamenginu

$$\{1, 1.1, 1.2, 1.3, ..., 5.8,5.9,6\}$$

Með hjálp Excel getum við kannað mynstur í dreifingum slíkra meðaltalna.

Líkan fyrir 100 tíu teninga köstum með Excel

Mynd 2. sýnir meðaltöl 100 „kasta” af 10 teningum.

	A	B	C	D	E	F	G	H	I	J
1	Hreint meðaltal 100 kasta með 10 teningum									
2	1	5	4	5	3	1	6	5	3	3
3	5	1	3	5	1	6	2	2	3	4
4	3	5	3	1	5	5	4	6	4	5
5	5	5	4	3	1	6	5	5	3	1
6	5	6	1	4	5	1	6	6	4	4
7	1	6	3	1	3	5	6	1	6	3
8	1	6	3	4	3	6	3	1	6	3
9	1	6	4	6	1	2	4	3	2	5
10	3	3	6	6	6	2	2	4	4	3
11	3	5	4	4	4	3	3	3	5	3

Mynd 2 Skjámynd af Excel

Formúlan sem notuð er í hverjum reiti til að líkja eftir teningakastinu er sýnd hér að neðan. RANDBETWEEN(n;m) er fall í Excel sem kallar á handahófskennda heila tölu milli talnanna n og m. Taktu eftir að hvert kall á fallið er óháð hinum. Því getur þú haft formúlu fyrir 10 óháða teninga og alla reikninga í einum reit.

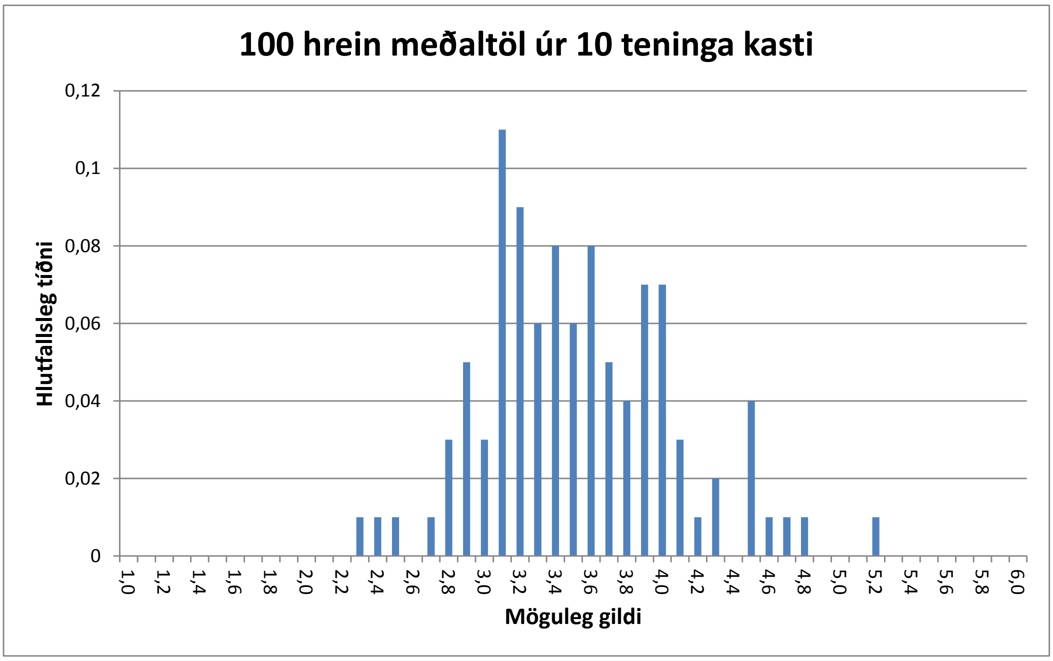
=(RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)
+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+
RANDBETWEEN(1;6)+RANDBETWEEN(1;6))/10

Formúlurnar til að vinna úr gögnunum eru sýndar á mynd 3.

K	L	M	N
Fjöldi kasta	Möguleg gildi	Tíðni	Hlutfallsleg tíðni
100	1	0	0
	1,1	0	0
	1,2	0	0
	1,3	0	0
	1,4	0	0
	1,5	0	0
	1,6	0	0
	1,7	0	0
	1,8	0	0
	1,9	0	0
	2	0	0
	2,1	0	0
	2,2	3	0,03
	2,3	0	0
	2,4	1	0,01
	2,5	1	0,01
	2,6	0	0
	2,7	2	0,02
	2,8	3	0,03
	2,9	5	0,05
	3	7	0,07
	3,1	3	0,03

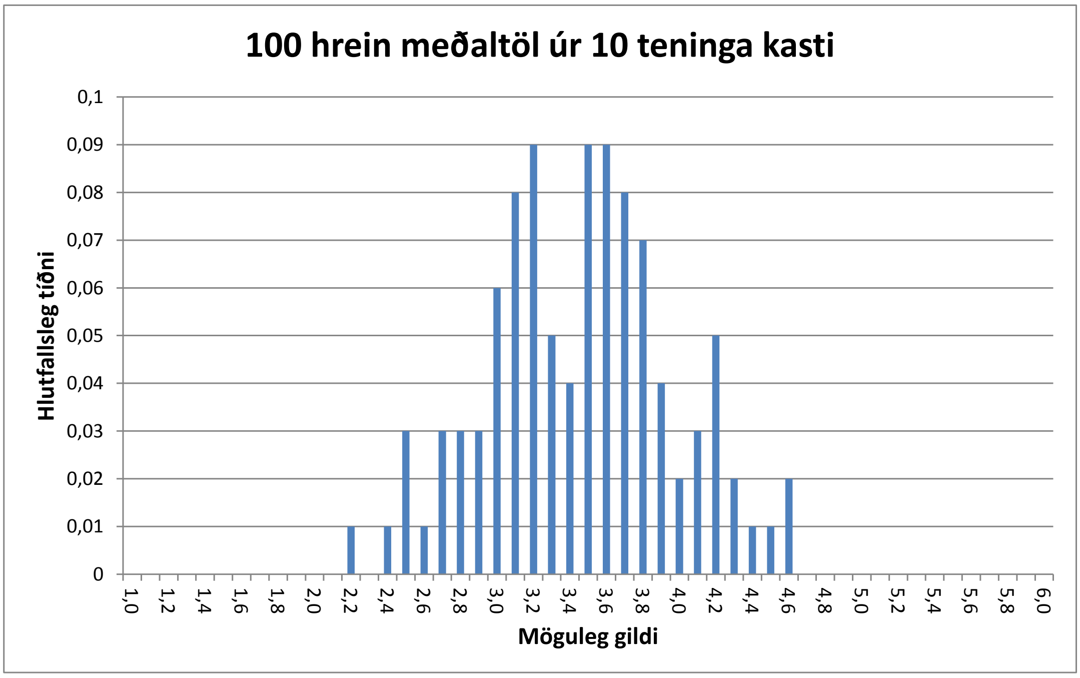
K	L	M	N
Fjöldi kasta	Möguleg gildi	Tíðni	Hlutfallsleg tíðni
= COUNT(A2:J11)	1	=COUNTIF($A$3:$J$12;L2)	=M2/$K$2
	=L2+0,1	=COUNTIF($A$3:$J$12;L3)	=M3/$K$2
	=L3+0,1	=COUNTIF($A$3:$J$12;L4)	=M4/$K$2
	=L4+0,1	=COUNTIF($A$3:$J$12;L5)	=M5/$K$2
	=L5+0,1	=COUNTIF($A$3:$J$12;L6)	=M6/$K$2
	=L6+0,1	=COUNTIF($A$3:$J$12;L7)	=M7/$K$2
	=L7+0,1	=COUNTIF($A$3:$J$12;L8)	=M8/$K$2
	=L8+0,1	=COUNTIF($A$3:$J$12;L9)	=M9/$K$2
	=L9+0,1	=COUNTIF($A$3:$J$12;L10)	=M10/$K$2
	=L10+0,1	=COUNTIF($A$3:$J$12;L11)	=M11/$K$2
	=L11+0,1	=COUNTIF($A$3:$J$12;L12)	=M12/$K$2
	=L12+0,1	=COUNTIF($A$3:$J$12;L13)	=M13/$K$2
	=L13+0,1	=COUNTIF($A$3:$J$12;L14)	=M14/$K$2

Mynd 3 Skjámyndir sem sýna formúlurnar fyrir að telja og reikna tölfræðilega stika.



Mynd 4. Súluritið sýnir tölfræðina. Súlurit er notað þar sem við vitum mögulegu (strjálu) gildin fyrir meðaltal 10 teninga.

Með töflureikninn svo uppsettan nægir að ýta á F9-takkan til að líkja eftir öðrum 100 köstum. Þú getur endurtekið oft og skoðað slembieinkenni fyrirbærisins. Það sést hvernig meðaltölin dreifast handahófskennt um möguleg gildi en einnig virðist vera mynstur í því hvernig gildin þyrpast saman í miðjunni.

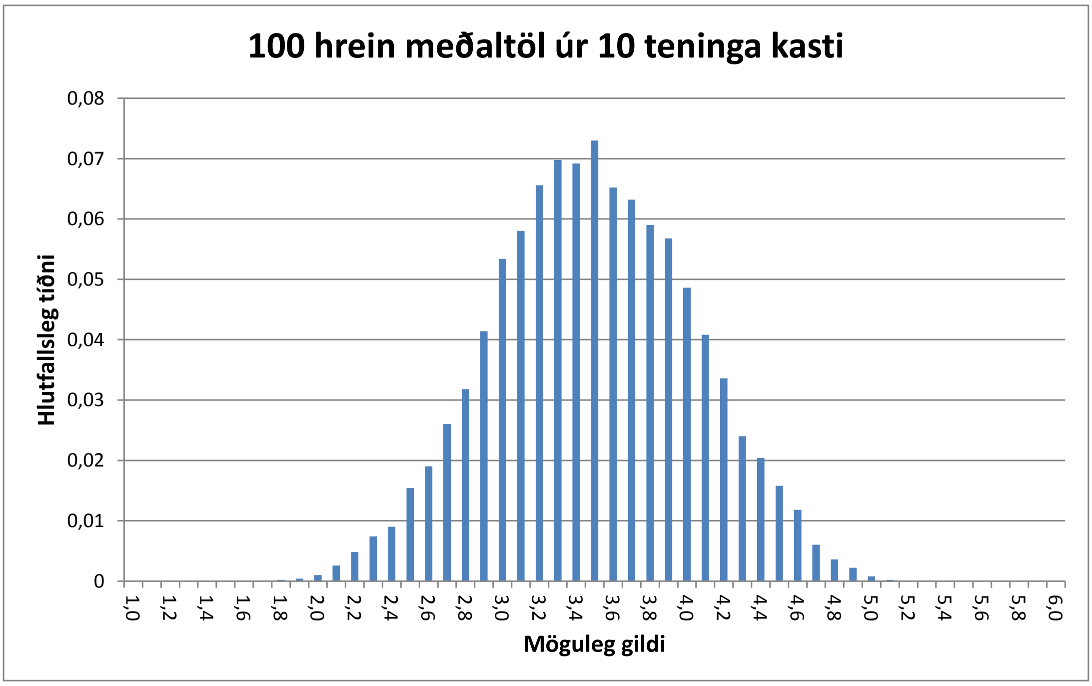


Mynd 5 Annað kast 10 teninga.

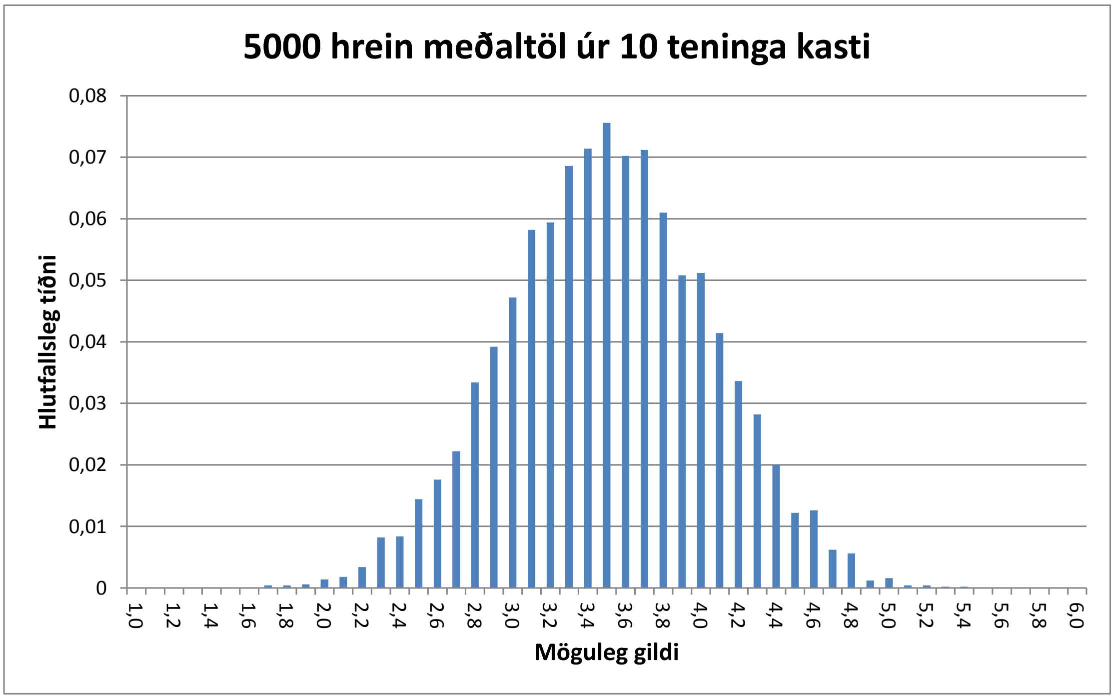
Tölvur og töflureiknar nú til dags geta höndlað gífurlegt magn gagna svo við getum auðveldlega athugað hvað gerist ef við köstum teningunum mun oftar.

Þar sem uppsetning og formúlur töflureiknisins eru nú þegar slegnar inn þá er það einungis spurning um að afrita reitina á hentug svæði..

Næstu tvær myndirnar sýna dreifingu meðaltalna fyrir 5000 köst.



Mynd 6 Súlurit sem sýnir niðurstöðu 5000 kasta.



Mynd 7 Með því að ýta á F9 kastar þú teningunum í 5000 skipti á ný.

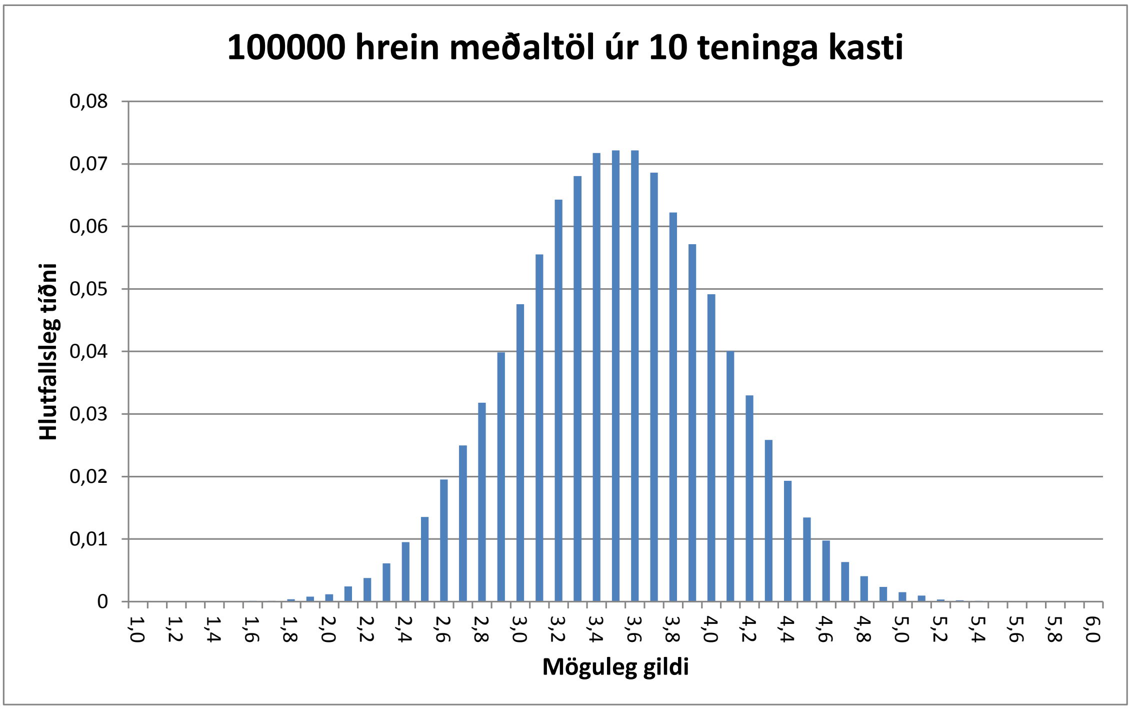
Við getum nú séð að mynstrið lítur mun reglulegra út og maður fær hugmyndir um það við hverju megi búast. Að veðja stórfé á að meðaltalið verði nálægt 1 eða 6 virðist ekki gáfulegt.

Ef til vill kann það nú vera spennandi að athuga hvað gerist ef við köstum enn oftar. Hví ekki 100.000 skipti? Það er nóg pláss til. Fjöldi raða í minni útgáfu af Excel 2007 er

$$2^{20} = 1048576$$

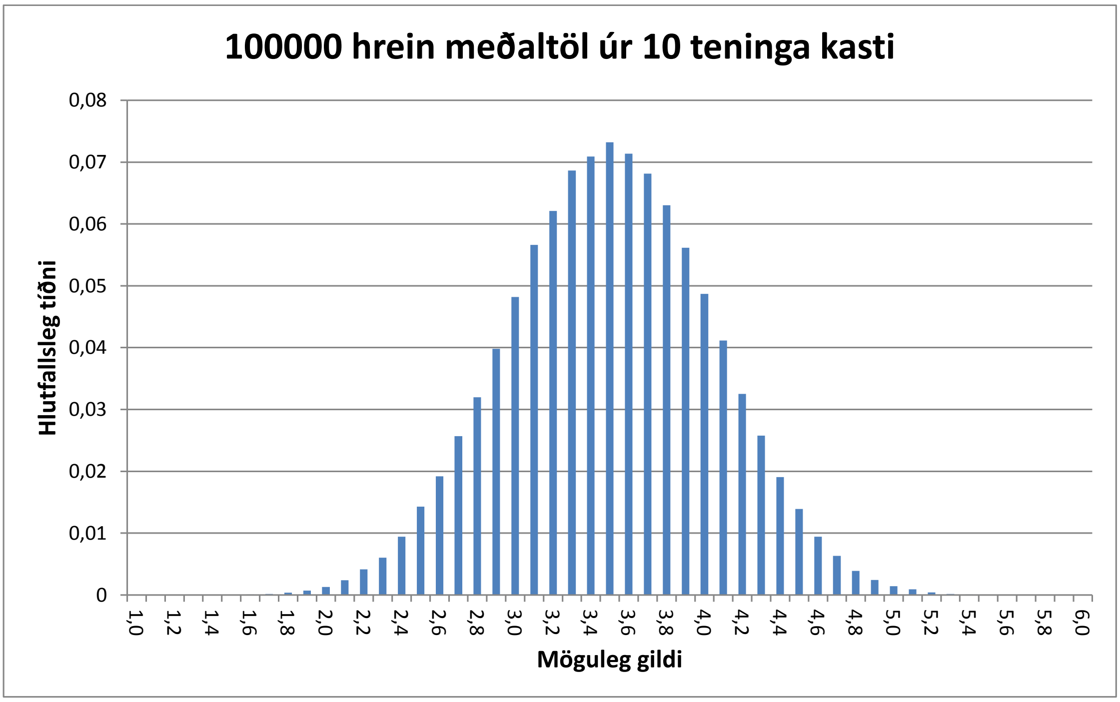
og fjödli dálka er

$$2^{14} = 16384$$



Mynd 8 Fallegt bjöllulaga mynstur af súluriti fyrir 100000 köst.

Með því að ýta á F9 fæst ný mynd (Mynd 9) sem er nauðalík mynd 8. Endurtekning kasta kallar fram álikar myndir.



Mynd 9 Önnur 100000 köst 10 teninga gefa nánast alveg eins meðaltalsdreifingu og mynd 8.

Eini munurinn í formúlunum sem notaðar voru í mynd 3. er svæðið sem talið er með COUNTIF.

K	L	M	N
Fjöldi kasta	Möguleg gildi	Tíðni	Hlutfallsleg tíðni
= COUNT(A2:J10001)	1	=COUNTIF($A$2:$J$10001;L2)	=M2/$K$2
	=L2+0,1	=COUNTIF($A$2:$J$10001;L3)	=M3/$K$2
	=L3+0,1	=COUNTIF($A$2:$J$10001;L4)	=M4/$K$2
	=L4+0,1	=COUNTIF($A$2:$J$10001;L5)	=M5/$K$2
	=L5+0,1	=COUNTIF($A$2:$J$10001;L6)	=M6/$K$2

Mynd 10 Hluti af formúlunum í því að telja 100000 köst af 10 teningum.

Hið reglulega form súluritsins í mynd 9 gefur til kynna að mögulega sé til einhver formúla sem lýsi því. Sem er rétt. Ef til vill stæði ég ekki í því að skrifa þessa grein ef ég vissi ekki fyrirfram að hún væri þegar fundin fyrir mörgum árum. Er hún vel þekkt og kallast Gauss- eða normalfallið og formið sem hún lýsir Gaussdreifing eða normaldreifing.

Innblásturinn fyrir þessa grein stafaði af vilja til þess að gefa haldbært dæmi um niðurstöðuna úr hinni svokölluðu höfuðsetningu tölfræðinnar.

Ég ætla ekki að fara nánar út formúlurnar og sannanirnar varðandi þetta viðfangsefni heldur halda mig við haldbær dæmi, svo lengi sem kalla má tölvulíkan haldbært. Ef þú vilt kynna þér hinn fræðilega bakgrunn, sjá t.d. [1] og [2]

Mörk dreifingarinnar af 10 teninga kasti

Hverju er verið að leita að? Höfuðsetning tölfræðinnar segir að meðaltalið verði nánast normaldreift með væntigildi m og dreifni $s^2/10$, þar sem m og s^2, eru væntigildi og dreifni við kast eins tenings.

Í stað þess að halda okkur við formúlur þá reikna ég þ essar tölur með Excel. Það mun gera skilgreiningarnar á þeim skýrari. Skoðaðu myndir 11 og 12.

	A	B	C	D	E
1	Fjöldi augna	Líkindi	Reiknun væntigildis	Reiknun Dreifni	
2	k	P(D=k)	k·P(D=k)	(k-m)^2·P(D=k)	
3	1	1/6	0,1667	1,0417	
4	2	1/6	0,3333	0,3750	
5	3	1/6	0,5000	0,0417	
6	4	1/6	0,6667	0,0417	
7	5	1/6	0,8333	0,3750	
8	6	1/6	1,0000	1,0417	
9			Væntigildi	Dreifni	Staðalfrávik
10			∑k·P(D=k)	∑(k-m)^2·P(D=k)	
11			m	s^2	s
12			3,5000	2,9167	1,7078

Mynd 11. Aðferð til að reikna væntigildi, dreifni og staðalfrávik eins tenings.

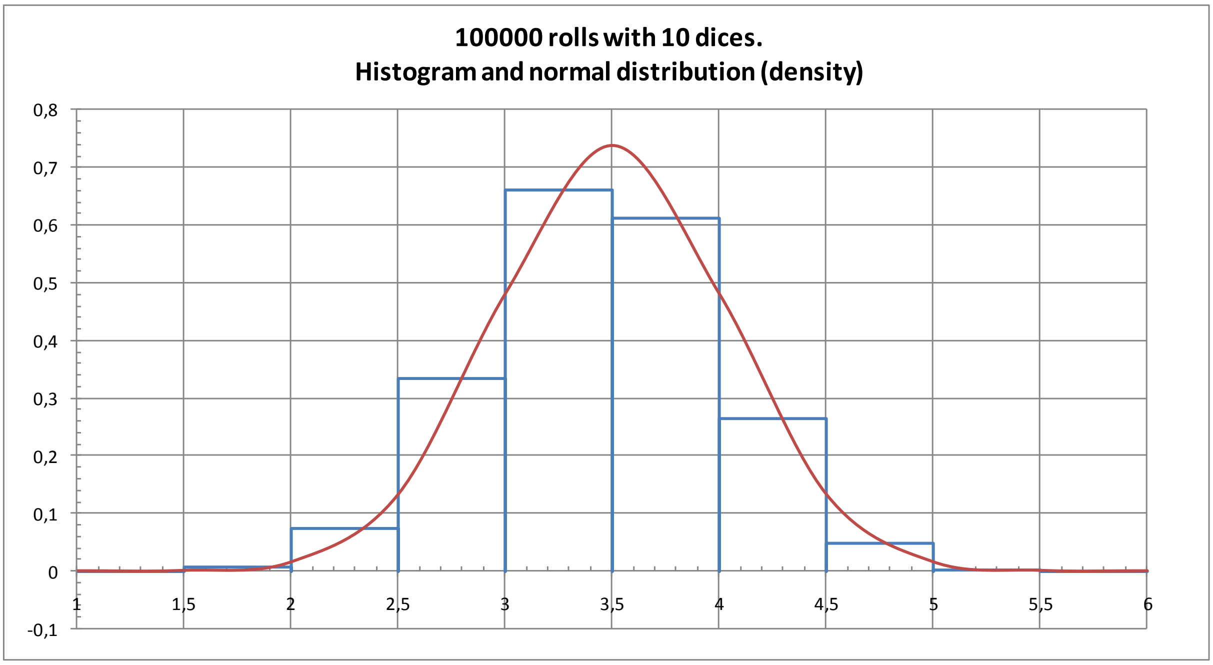
	A	B	C	D	E
1	Fjöldi augna	Líkindi	Reiknun væntigildis	Reiknun Dreifni	
2	k	P(D=k)	k·P(D=k)	(k-m)^2·P(D=k)	
3	1	1/6	=A3*B3	=(A3-$C$12)^2*B3	
4	2	1/6	=A4*B4	=(A4-$C$12)^2*B4	
5	3	1/6	=A5*B5	=(A5-$C$12)^2*B5	
6	4	1/6	=A6*B6	=(A6-$C$12)^2*B6	
7	5	1/6	=A7*B7	=(A7-$C$12)^2*B7	
8	6	1/6	=A8*B8	=(A8-$C$12)^2*B8	
9			Væntigildi	Dreifni	Staðalfrávik
10			∑k·P(D=k)	∑(k-m)^2·P(D=k)	
11			m	s^2	s
12			=SUM(C3:C8)	=SUM(D3:D8)	=SQRT(D12)

Mynd 12. Formúlur fyrir reikningana í mynd 11.

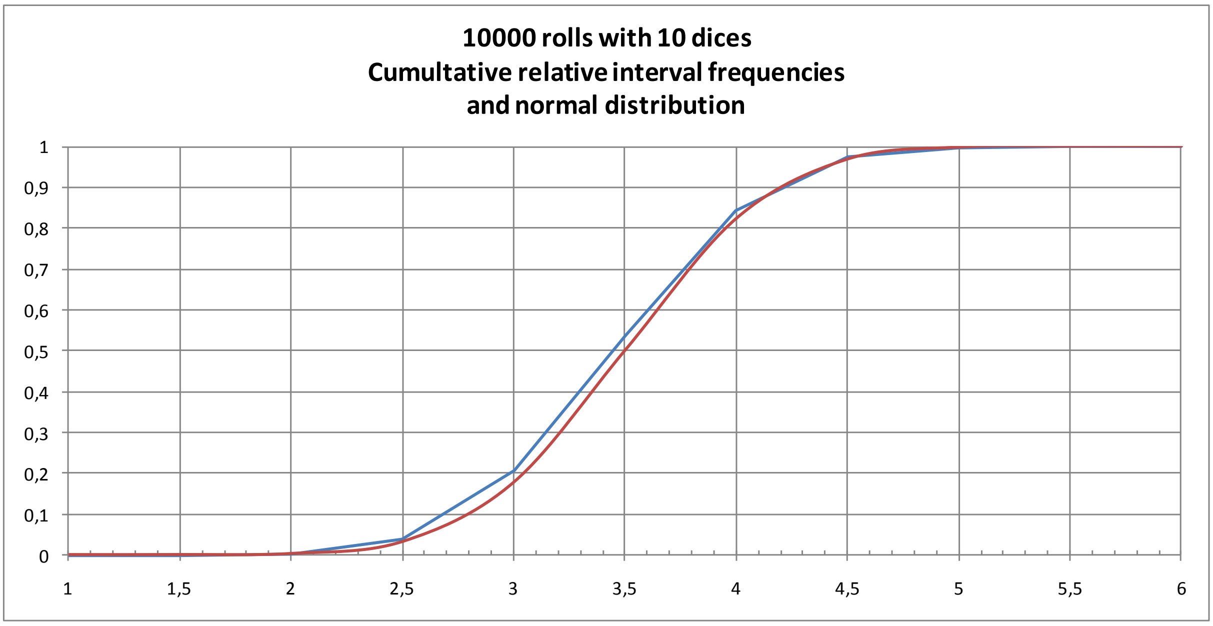
Þar sem normaldreifingin er samfelld dreifing er ekki hægt að bera súluritið úr mynd 9 við líkindaþéttleika normaldreifingar.

Ef við tökum 100000 meðaltöl sem tilvik af samfelldri breytu munum við flokka tölurnar í bil. Með því að telja fjölda meðaltalna á hverju hlutbili fæst biltíðnin. Með því að deila í þær tölur með 100000 fæst hlutfallsleg biltíðni sem er táknuð með ferhyrningum yfir hvert bil með flatarmáli jöfnu hlutfallslegu biltíðninni.

Með því að láta hvert hlutbil vera af lengd 0,5 fáum við myndirnar á myndum 13 og 14.

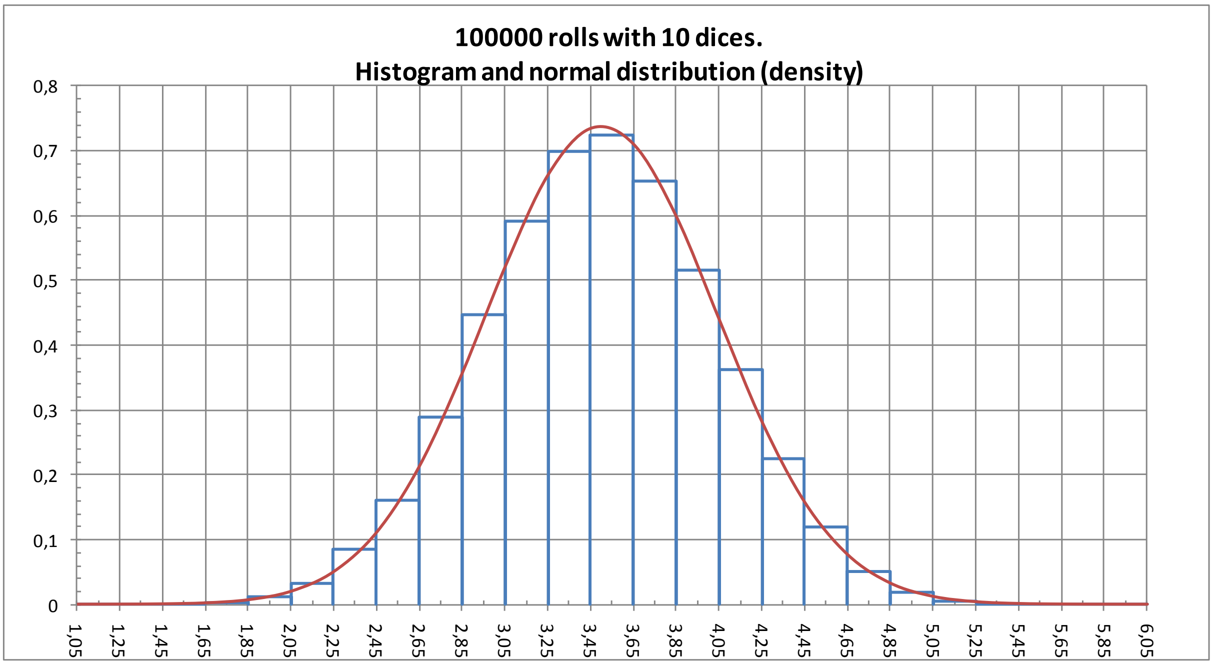


Mynd 13. Súlurit borið saman við normaldreifingu með væntigildi 3,5 og staðalfrávik 0,54.

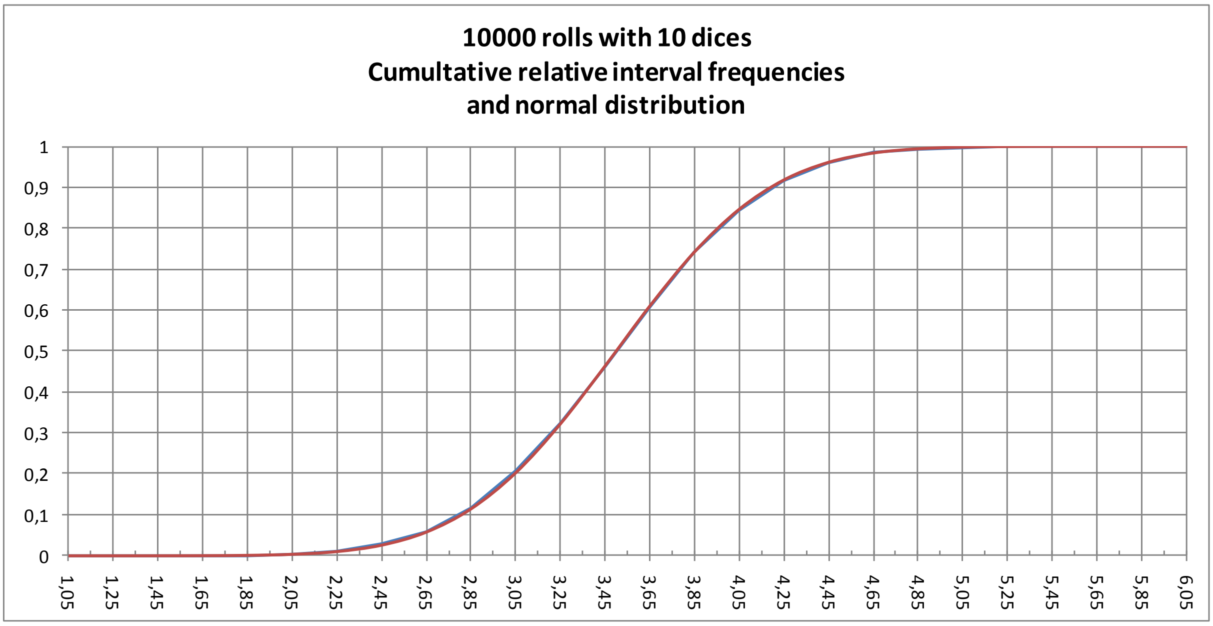


Mynd 14. Uppsöfnuð normaldreifing með sömu stika og á mynd 13 borið saman við uppsafnaðar hlutfallslegar biltíðnir.

Þar sem það eru 100000 athuganir í gagnamenginu okkar eru þær nógu margar til þess að við getum fengið nákvæmari mynd með því að velja minni hlutbil. Á myndum 14 og 15 eru bilin valin sem 1.05<1.25<1.45<...<5.85<6.05 þ.a. við höfum allar athuganir innan bilanna.



Mynd 15 Nú með minni hlutbil eins og skýrt var frá í texta.



Mynd 16 Nú með minni hlutbil eins og skýrt var frá í texta.

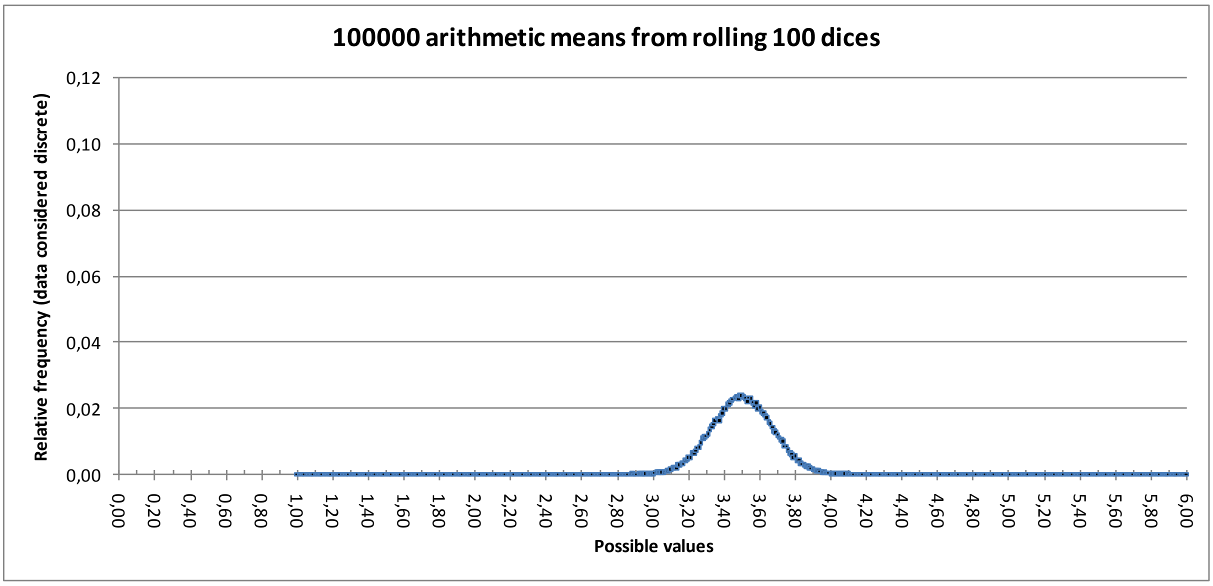
Næst reynum við að kasta 100 teningum og reikna meðaltalið. Ég nota nánast sömu uppsetningu töflureiknis og sýnd er á mynd 3.

Galdurinn við að reikna meðaltal 100 teninga er rita jöfnuna sem sést á mynd 17 og afrita hana úr WORD skjali yfir í fyrsta reit töflureiknis og síðan afrita í töflureikninum eins og venjulega.

```
=(RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETW
EEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RAN
DBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;
6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBET
WEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RA
NDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(
1;6)+RANDBETWEEN(1;6)+RANDBETWEEN(1;6))/100
```

Mynd 17 Formúla til þess að reikna meðaltal 100 teninga í einum reit. En svo stóra formúlu má samt meðhöndla sem miklu minni formúlur.

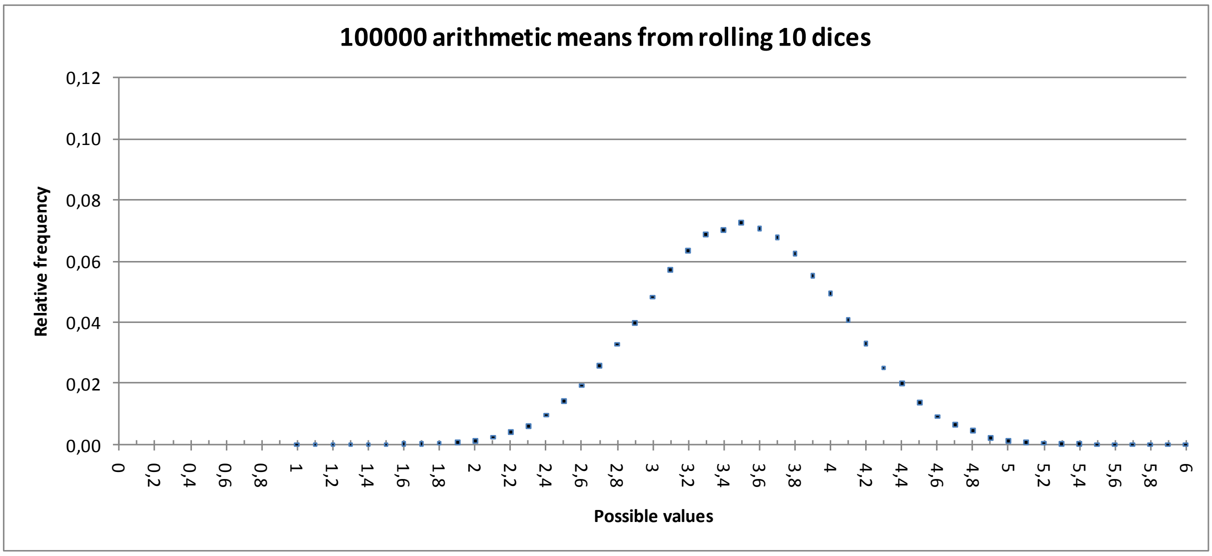
Í þetta skiptið eru 501 möguleg gildi sem meðaltalið getur tekið (1.00<1.01<...<5.99<6.00). Ef maður ætlaði nú að búa til súlurit eins og gert var t.d. á mynd 4 þá værum við með of margar súlur til þess að myndin yrði nothæf. Í stað súluritsins er notað xy-graf. Aðeins punktar eru merktir inn á og þú mátt líta á þá sem efra mark súlnanna.



Mynd 18. xy-graf notað til að lýsa súluriti.

Hver punktur á grafinu táknar súlu fyrir ofan 2-aukastafa tölu á x-ásnum.

Þú getur borið saman mynd 19 sem er xy-graf af súluritum eins og voru á myndum 8 og 9.

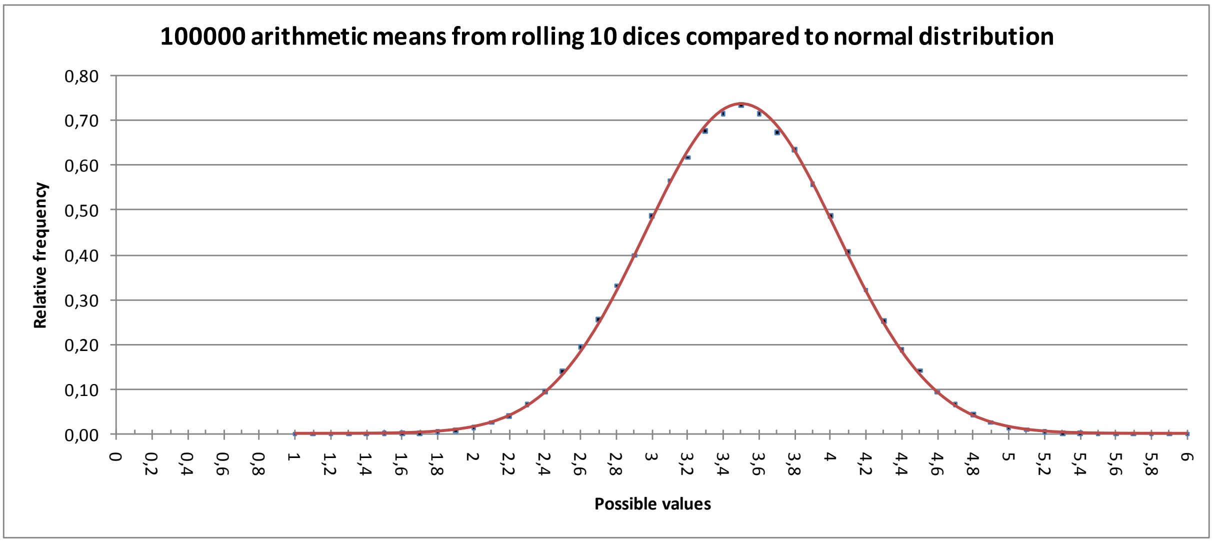


Mynd 19 xy-graf í stað súlurits eins og í mynd 8.

Hér er hver súla táknuð með efra marki sínu.

Til að bera saman við normaldreifingu þá þarft þú að hafa í huga að hlutfallslegu tíðnirnar verða að tengjast flatarmáli rétthyrninganna. Hér er bil af lengd 0.1 notað þannig að við þurfum að deila í hlutfallslegu biltíðnina með 0.1 (eða margfalda með 10).

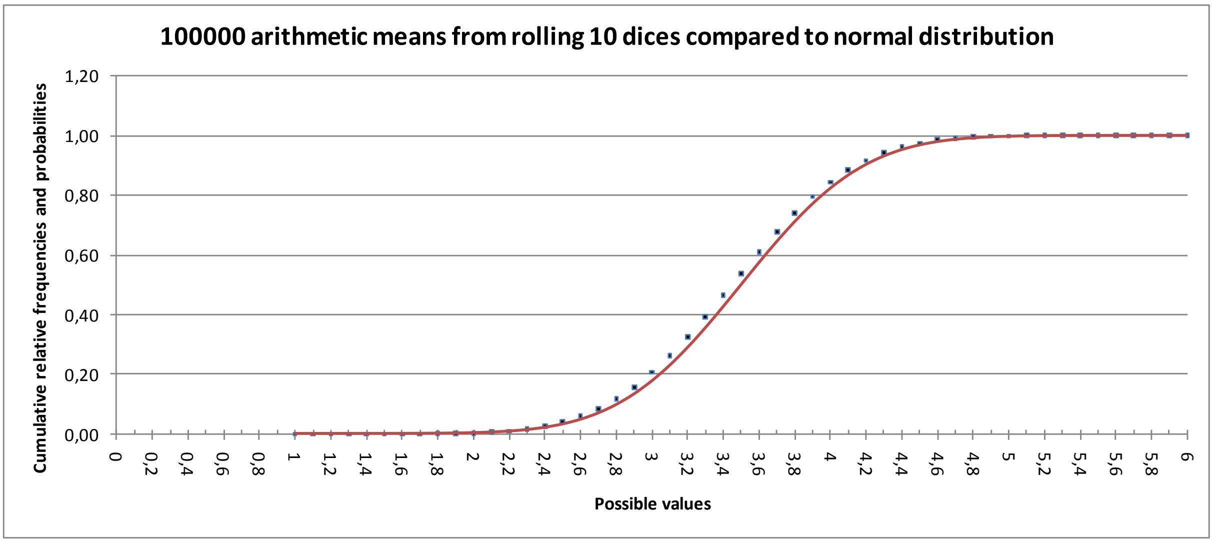
Síðan getur þú borið það saman við normaldreifingu með væntigildi 3.5 og staðalfrávik $1.7078/\sqrt{10}$ eins og mynd 20 sýnir.



Mynd 20 „Punkta súlurit“ borið saman við líkindaþéttleikafall fyrir normaldreifingu.

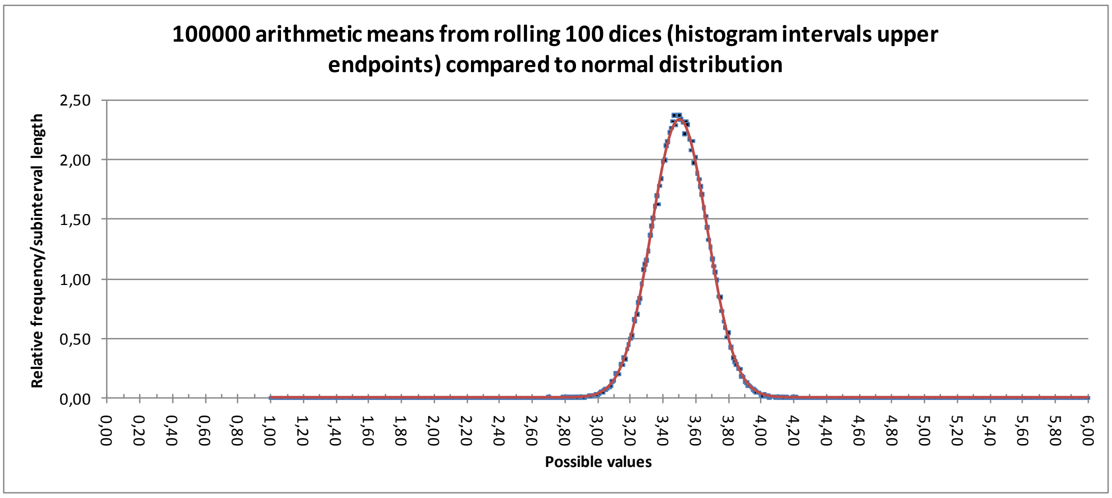
Mynd 21 sýnir svo uppsöfnun líkinda og hlutfallslegar biltíðnir í samræmi við mynd 20.

Með því að skoða myndapörin myndir 20 & 22. og myndir 21 & 23. sést hvernig frávikið fyrir meðaltal 100 teninga er mun minna en fyrir meðaltal 10 teninga. Fyrra frávikið er $\sqrt{10} \approx 3.16$ sinnum síðara frávikið.

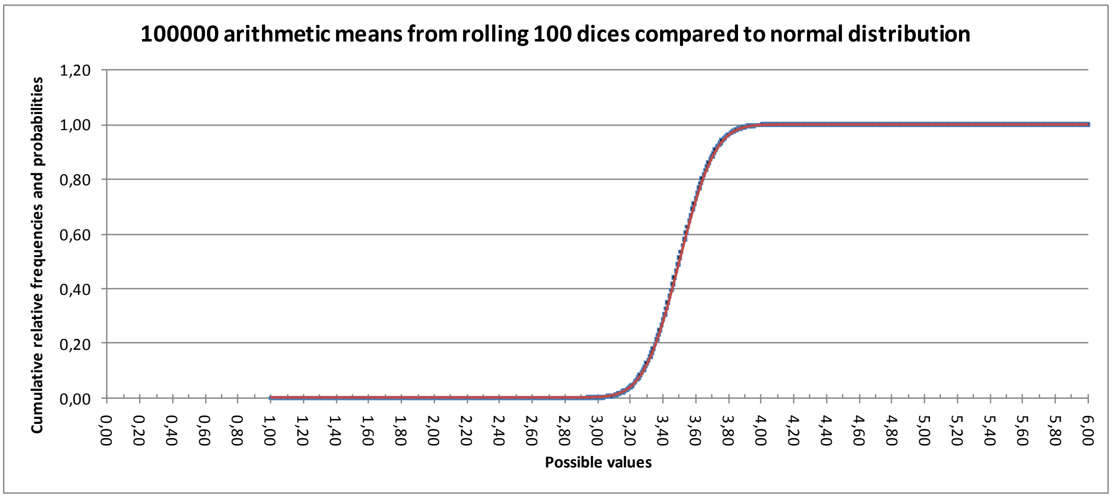


Mynd 21 Uppsöfnuð normal líkindi og uppsöfnuð hlutfallsleg biltíðni.

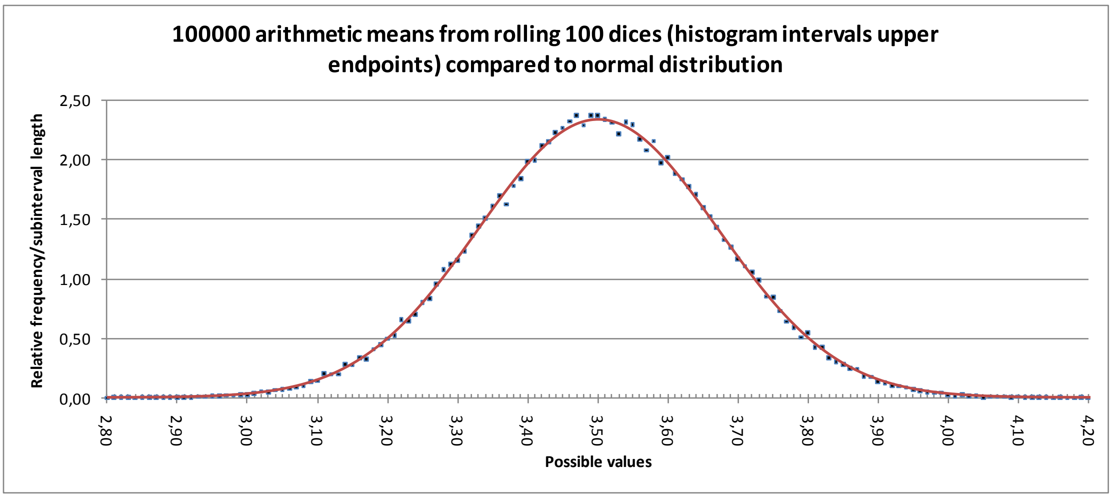
Nú koma myndirnar fyrir köstin með 100 teningum.



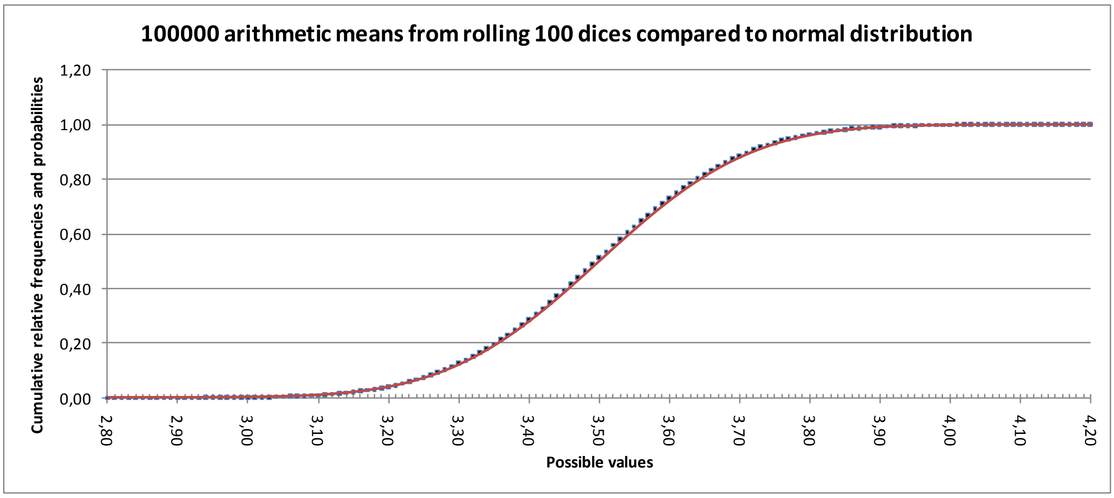
Mynd 22 „Punkta súlurit" sem samsvarar hlutbilum 1,00<1.01<1.02<....5.98<5.99<6.00 miðað við normaldreifingarþéttleika með væntigildi 3,5 og staðalfrávik $1.7078/\sqrt{100} = 0.17078$



Mynd 23 Uppsöfnuð dreifing mótsvarandi mynd 22.



Mynd 24 Mynd 22 teygð til þess að skoða smáatriði.



Mynd 25 Mynd 23 teygð til þess að skoða smáatriði.

Loka athugasemdir

Hér á undan eru sýndar mismunandi kennistærðir sem má breyta þegar verið er að gera tilraunir sem þessar.

- Breyta fjölda teninga sem kastað er.
- Breyta fjölda kasta af föstum teningafjölda.
- Breyta lengd hlutbilanna í súluritinu.
- Færa hlutbilin til til þess að skoða áhrif gildanna sem falla saman við endapunkta hlutbilanna.

Það er nokkuð flókið að skrifa nákvæma lýsingu á uppsetningu töflureiknis sem notaður var. Sum atriði eru sýnd hér að ofan, öðrum er sleppt.

.

References

[1] http://en.wikipedia.org/wiki/Central_limit_theorem (August 31, 2011)

[2] http://mathworld.wolfram.com/CentralLimitTheorem.html (August 31, 2011)