

Eksperimentel statistik i matematikundervisningen

STX A og STX B

Find al materialet her:
www.kortlink.dk/2kban

17.3.2023
 Matematikdidaktikkens dag

Janus Lylloff, Mulernes Legatskole

1

Dagens program (på 20 min)

Binomialfordelingen

- Terningekast
- Teoridannelse
- Undersøg teori

Binomialtest

Normalfordelings approx.

Find al materialet her:
www.kortlink.dk/2kban

C. Simulering af fødselsparadokset

E. Simulering af V2 bomber over London

D. Minecraft og Binomialfordelingen

F. Undersøgelse af normalfordelingen

Janus Lylloff, Mulernes Legatskole

2

Op Dette er en PDF-version af en Nspire aktivitet.

Aktiviteten er tænkt som en Nspire fil, med en del interaktive elementer, disse er ikke med her, men også Layoutet kan drille i PDF-versionen, da det ikke er tænkt som en sådan fra starten.

Mvh Janus

Binomialfordelingen

Et eksempel på eksperimentel tilgang til statistik opdelt i 3 trin

- 1) Eksperiment (simulering af terninger).
- 2) Udvikling af en generel teori på baggrund af eksperimentet.
- 3) Undersøg teorien ved forskellige eksperimenter.

Arket er en samling af 3 forskellige aktiviteter, der bygger oven på hinanden, men man kan sagtens tage andre elementer ind imellem hver af de 3 elementer. Opsamlinger mellem hver del er helt sikkert en god ide. **BEMÆRK: Ikke klar til elever – der er svar på mange af opgaverne undervejs.**

Janus Lylloff, Marts 2023

Eksperiment kast med 6 terninger

Centrale kommandoer ved simulering:

For at forstå simuleringen på næste side (og alle andre simuleringer) skal man kende til disse ting – se evt. på simuleringen inden du læser dette:

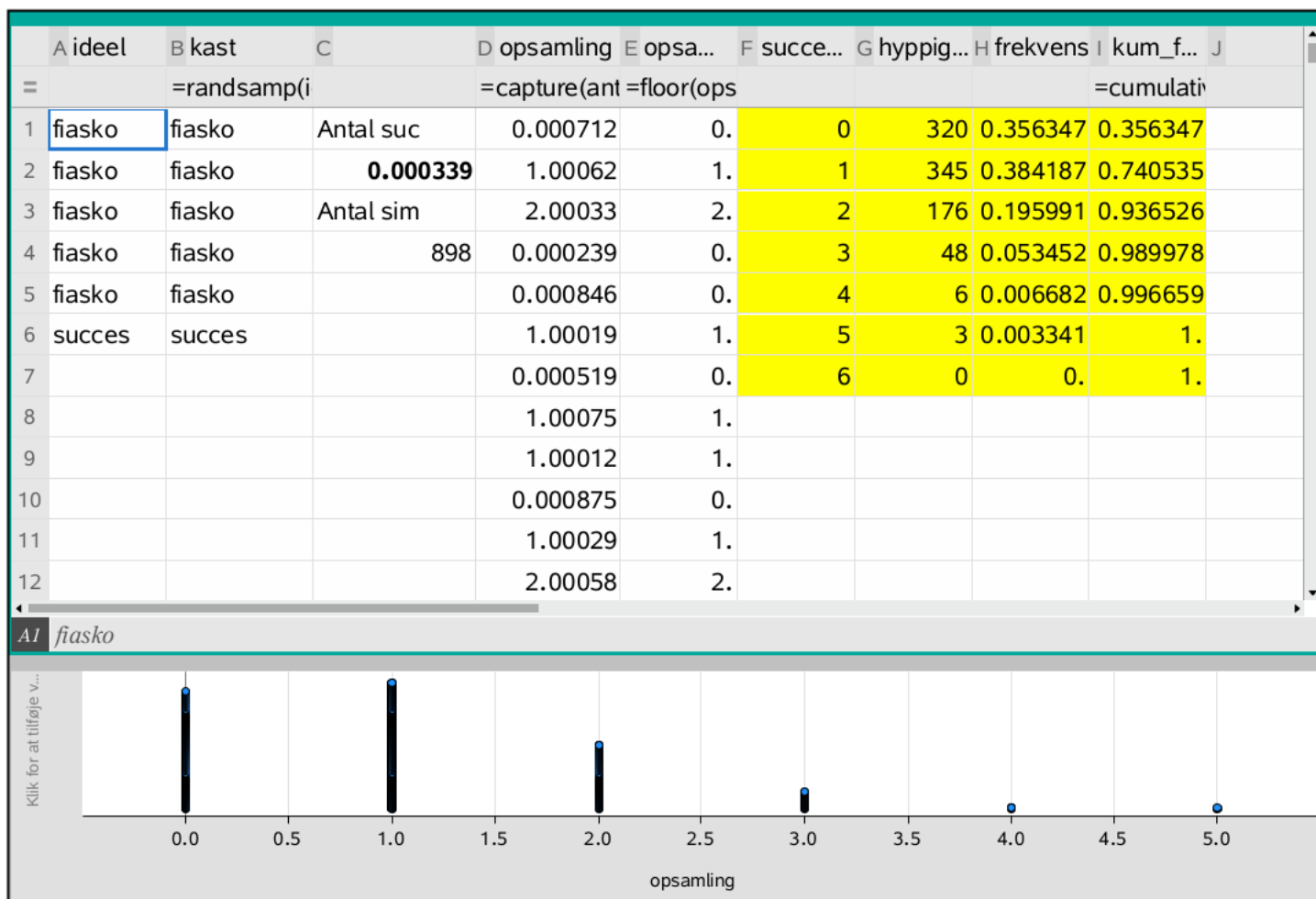
=randsamp(liste,n), udtager n tilfældige elementer ud af listen med tilbagelægning. Bemærk at alle computere laver den samme tilfældighed første gang (det er svært at få en computer til at lave tilfældige ting). **CTRL+R** giver en ny tilfældig prøve.

=capture(var,{1,0}) kommandoen finder man ved at højre klikke i formelfeltet i regnearket og vælge datafangst. Kommandoen gemmer variabelen **var 1: automatisk** (dvs. variabelen gemmes kun når den ændrer værdi) **0: manuelt** variabelen gemmes kun når man trykker **CTRL+"PUNKTUM"**.

Vær opmærksom på at man i simuleringer, hvor den variable ofte antager samme værdi (som her) begår man en **stor metodisk fejl, hvis man bruger automatisk dataopsamling**. Så enten skal man bruge manuel dataopsamling (trykke skiftevis CTRL+R og CTRL+"PUNKTUM") eller indlægge støj på den variabel man opsamler. Det kan f.eks. gøres ved at lægge

rand()
1000 ▶ 0.000944 til variabelen. man kan let fjerne støjen med floor kommandoen – som det hele på næste side

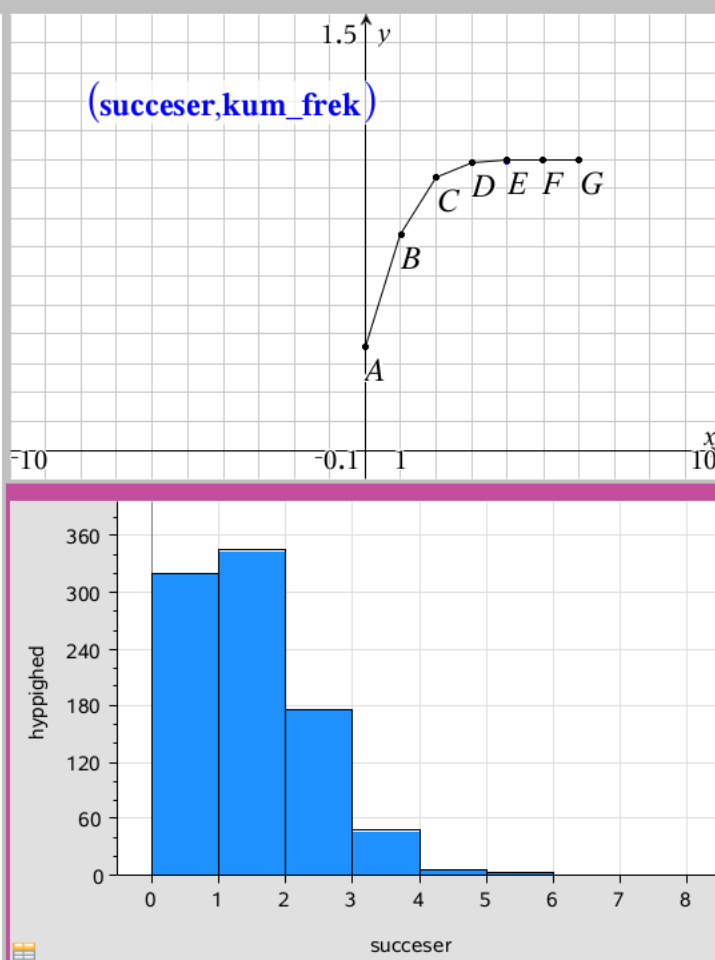
Start simuleringen forfra ved at højreklikke på søjlen med datafangsten og vælg "ryd data"



HVIS MAN VIL kan man opsamle data som i den gule tabel på forrige side, det kan åbne op for at man laver deskriptiv statistik og de velkendte plots med de simulerede data.

Det er mest oplagt med større n -værdi i forbindelse med tilnærmelse med normalfordelingen (se Nspire filen simulering af 24 terningekast)

[BEMÆRK: Den gule tabel skal være på samme side som simuleringen, hvis den skal blive ved med at være dynamisk]



START HER!

Hvis du vil taste med så starter vi i denne opgave (der ender med at blive lige som den forrige)

Hvis man ikke gider taste – så se resultatet ovenfor

	A ideel	B kast	C	D opsamling	E opasmling2	F	G	H	I	J
=										
1			Antal suc							
2										
3			Antal simu...							
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										

Teori

b) Hvilken sammenhæng er der mellem kombinationerne med 2 seksere og kombinationerne med 4 seksere? **Det er det samme blot "spejlet"**

OPGAVE: Sandsynlighederne

a) Hvor mange forskellige udfald er der ialt? (tæl eller rgn) **SVAR: $2^6 \rightarrow 64$**

b) Hvor mange forskellige udfald er dere med de forskellige antal seksere? Tæl

c) Kan vi regne det ud? [Hint: permutationer eller kombinationer]

SVAR:

0 1 $nCr(6,0) \rightarrow 1$

1 6 $nCr(6,1) \rightarrow 6$

2 15 $nCr(6,2) \rightarrow 15$

3 20 $nCr(6,3) \rightarrow 20$

4 15 $nCr(6,4) \rightarrow 15$

5 6 $nCr(6,5) \rightarrow 6$

6 1 $nCr(6,6) \rightarrow 1$

$nPr(6,0) \rightarrow 1$

$nPr(6,1) \rightarrow 6$

$nPr(6,2) \rightarrow 30$

	A	B	C	D	E	F	G	H
=								
1	kast	1	2	3	4	5	6	
2	0 seksere..	0	0	0	0	0	0	
3	1 seksere..	1	0	0	0	0	0	
4			1					
5				1				
6					1			
7						1		
8							1	
9	2 seksere..	1	1					
10		1		1				
11		1			1			
12		1				1		
13		1					1	
14			1	1				
15			1		1			
16			1			1		
17			1				1	
18				1	1			

Linje 1 (1 1 0 0 0 0) kan oversættes til:

Terning 1 er en 6'er OG

Terning 2 er en 6'er OG

Terning 3 er ikke en 6'er OG

Terning 4 er ikke en 6'er OG

Terning 5 er ikke en 6'er OG

Terning 6 er ikke en 6'er

OPGAVE: Forklar at sandsynligheden for dette udfald

$$\text{er } \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6}$$

Linje 2 (1 0 1 0 0 0) kan oversættes til:

Terning 1 er en 6'er OG....

[Udfyld selv]

Sandsynligheden for *netop* dette udfald er:

[Udfyld selv]

OPGAVE: Forklar at sandsynligheden for at få netop 2 seksere kan udregnes med formlen

$$15 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6}$$

[Hint, hvis man får 2 seksere er det enten som i linje 1 ELLER linje 2 ELLER linje 3...]

	A	B	C	D	E	F	G	H
=								
1		1	2	3	4	5	6	
2	2 seksere..	1	1	0	0	0	0	
3		1	0	1	0	0	0	
4		1	0	0	1	0	0	
5		1				1		
6		1					1	
7			1	1				
8			1					
9			1			1		
10			1				1	
11				1	1			
12				1		1		
13				1			1	
14					1	1		
15					1		1	
16						1	1	
17								
18								

Sandsynligheden for 1 sekser er:

Sandsynligheden for 2 seksere er:

Sandsynligheden for 3 seksere er:

Sandsynligheden for 4 seksere er:

Sandsynligheden for 5 seksere er:

Sandsynligheden for 6 seksere er:

OPGAVE: Tjek at dette stemmer med at den samlede sandsynlighed er 1.

Sandsynligheden for at få præcis 0 seksere skrives $P(X=0)$, $P(X=1)$ angiver sandsynligheden for at få præcis 1 sekser osv.

SÆTNING: For en binomialfordelt stokastisk variabel, hvor p er sandsynligheden for succes og dermed er $(1-p)$ sandsynligheden for fiasko, er sandsynligheden for at få r succeser ud af n forsøg givet ved:

$$P(X = r) = K(n, r) \cdot p^r \cdot (1 - p)^{n-r}$$

Det ser lidt indviklet ud, men Nspire har en kommando for det. Hvis vi skal regne sandsynligheden for 3 seksere ud af 6 kast ud gør man således: $\text{binomPdf}\left(6, \frac{1}{6}, 3\right) \rightarrow 0.053584$ svarende til 5,36%.

[TIP find syntaksen til kommandoen binompdf i kataloget til højre]

OPGAVE: Brug kommandoen til at udregne sandsynlighederne for de øvrige udfald som du regnede ovenfor.

Undersøg n, p og grafen

Sandsynligheden for succes



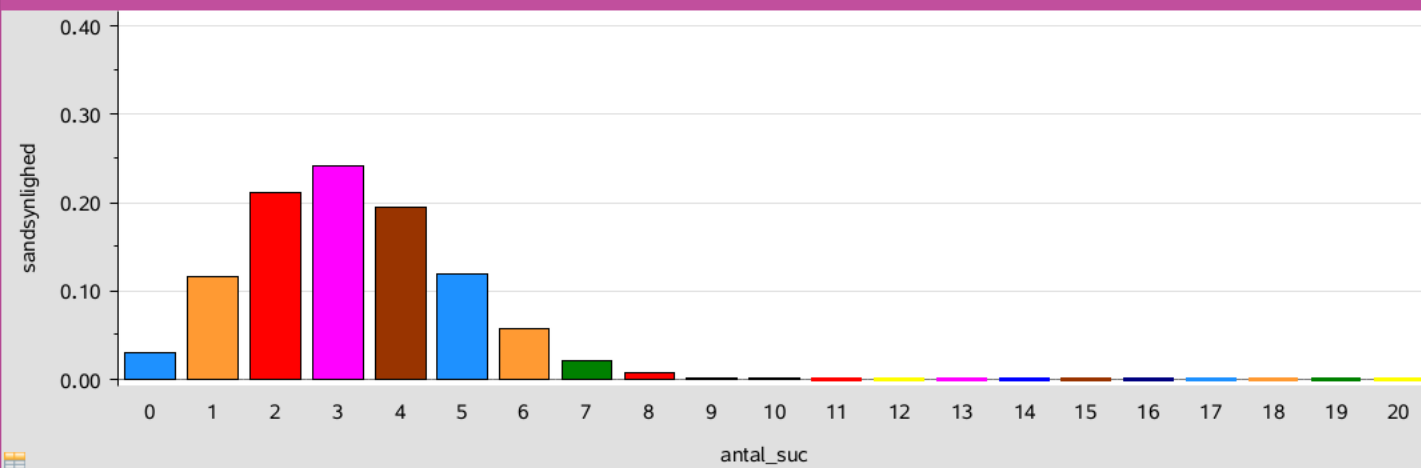
OPGAVE: Varier n og p mens i betragter grafen nedenfor, og svar på disse spørgsmål. Gode svar indeholder beskrivelser af antallet af søjler, deres højde, placering af den højeste søjle etc...

a) Hvordan ændres grafen når p bliver større eller mindre.

b) Hvordan ændres grafen når n bliver større eller mindre?

[Teknisk kommentar: 1) Diagrammet nedenfor er lavet med en kategorisk x-akse, for at få zoom til at følge med værdien af n det er derfor at søjlerne har forskellige farver 2) Regnearket til højre er maskinrummet, du må gerne gå på opdagelse derinde, men pas på ikke at ødelægge noget]

	A antal_...	B sands...	C
=	=seq(i,i,0,'=binompdf		
1	0	0.03059	
2	1	0.116535	
3	2	0.210873	
4	3	0.240998	
5	4	0.195093	
6	5	0.118914	
7	6	0.056626	



Undersøg middelværdi og spredning

Spredningen er $\sigma = \sqrt{n \cdot p \cdot (1-p)} \rightarrow 5.04975$

De lodrette streger på grafen angiver $my \rightarrow 30$, $my - \sigma \rightarrow 24.9502$ og $my + \sigma \rightarrow 35.0498$

OPGAVE:

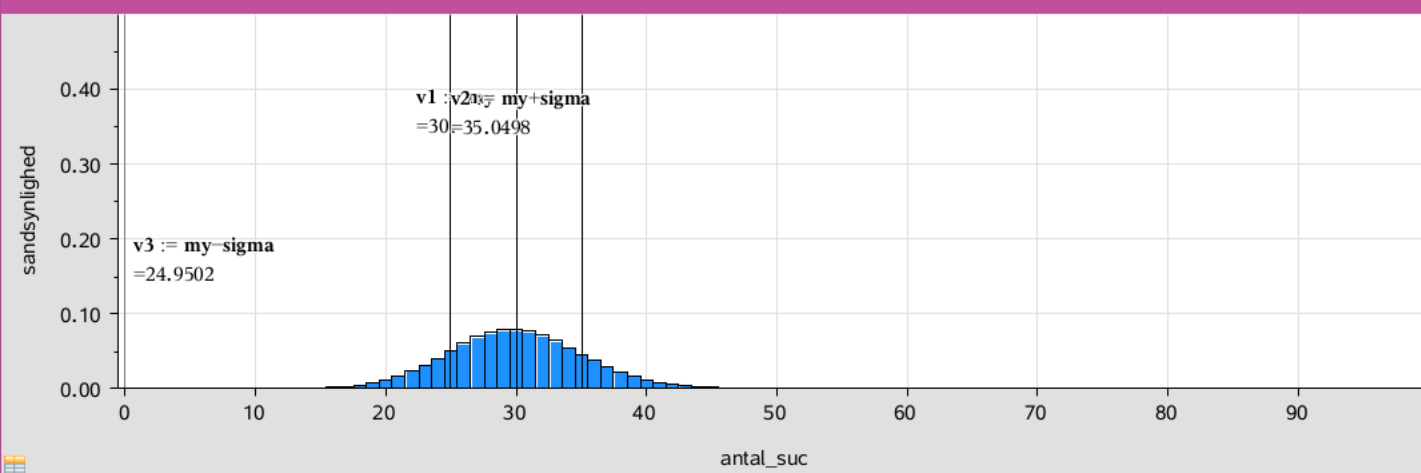
a) Sæt n=100 og eksperimenter med p. Hvornår er spredningen størst og mindst? Hvordan ser man på grafen om spredningen er stor eller lille?

b) Hvad er n's indvirkning på spredningen?

c) Hvad kan man sammenhængende sige man skal gøre for at få en så lille spredning som muligt?

[Teknisk kommentar: Diagrammet nedenfor er lavet med en almindelig numerisk x-akse (som man plejer), zoomet følger ikke med, men vi kan plote værdierne middelværdi og spredning]

	A antal_...	B sands...	C
=	=seq(i,i,0,'=binompdf		
1	0	7.65218...	
2	1	2.70077...	
3	2	4.74223...	
4	3	5.52331...	
5	4	4.8004E...	
6	5	3.32075...	
7	6	1.90455...	



Binomialtest+ Normalfordelingsapproximation

Eksperimentel tilgang til binomialtest ender med normalfordelingsapproximationen (kan undlades).

- 1) Forforståelsen afdækkes: Hvad ved vi og hvad forventer vi.
- 2) Opstil nulhypotese H_0 .
- 3) Simulering af kast med 24 terninger med antagelse af H_0 .
- 4) Afkræft eller underbyg H_0 .
- 5) Tilnærmelse med normalfordelingen, tæthedsfunktionen og fordelingsfunktionen dukker nærmest op...

Jeg vil bruge denne på to forskellige måder, enten med hele klassen som "tast med" i begyndelsen af et forløb (mat A), eller som udleveret dokument hvor dataopsamlingerne er slettet og de bare simulerer (mat B).

Bemærk: Den systematiske fejl man begår med automatisk opsamling er her forsvindende lille, fordi det vi undersøger har en lille sandsynlighed (9 seksere på 24 slag). Derfor er der ikke tilføjet støj – vi lever med fejlen (for ikke at overkomplicere det). Janus Lylloff, Marts 2023

24 kast med en terning – kan det give 9 seksere?

Du møder en skummel mand på et skummelt værtshus, der vil spille terninger om penge.

Du er som altid venlig men mistroisk og vil undersøge terningen – er det en snyde terning eller...?!

Du kaster med terningen 24 gange og får 9 seksere. **Hvad skal du tro?**



Opgave 1:

Hvad er mest sandsynlige udfald

Du kaster med terningen 24 gange og får 9 seksere. **Hvad skal du tro?** Du ved at det var mest sandsynligt at få $\frac{1}{6} \cdot 24 = 4$. seksere, men på den anden side er det jo et stokastisk (tilfældigt) fænomen og det kan jo godt forekomme at man slår 9 seksere ud af 24 slag.

Spørgsmålet er:

Hvor sandsynligt er det at slå 9 seksere ud af 24 slag?

Vi vil prøve at kaste lys over dette spørgsmål ved at lave en simulering af terningekast. Metoden er at vi starter med at antage følgende hypotese – kaldet H_0 ("h nul") hypotesen.

H_0 : Terningen er en ideel terning – dvs. der er $\frac{1}{6}$ sandsynlighed for at slå en 6'er.

Basiseksperimentet: Kast med en terning.

Altså er basissandsynligheden $\frac{1}{6}$ og $n=24$

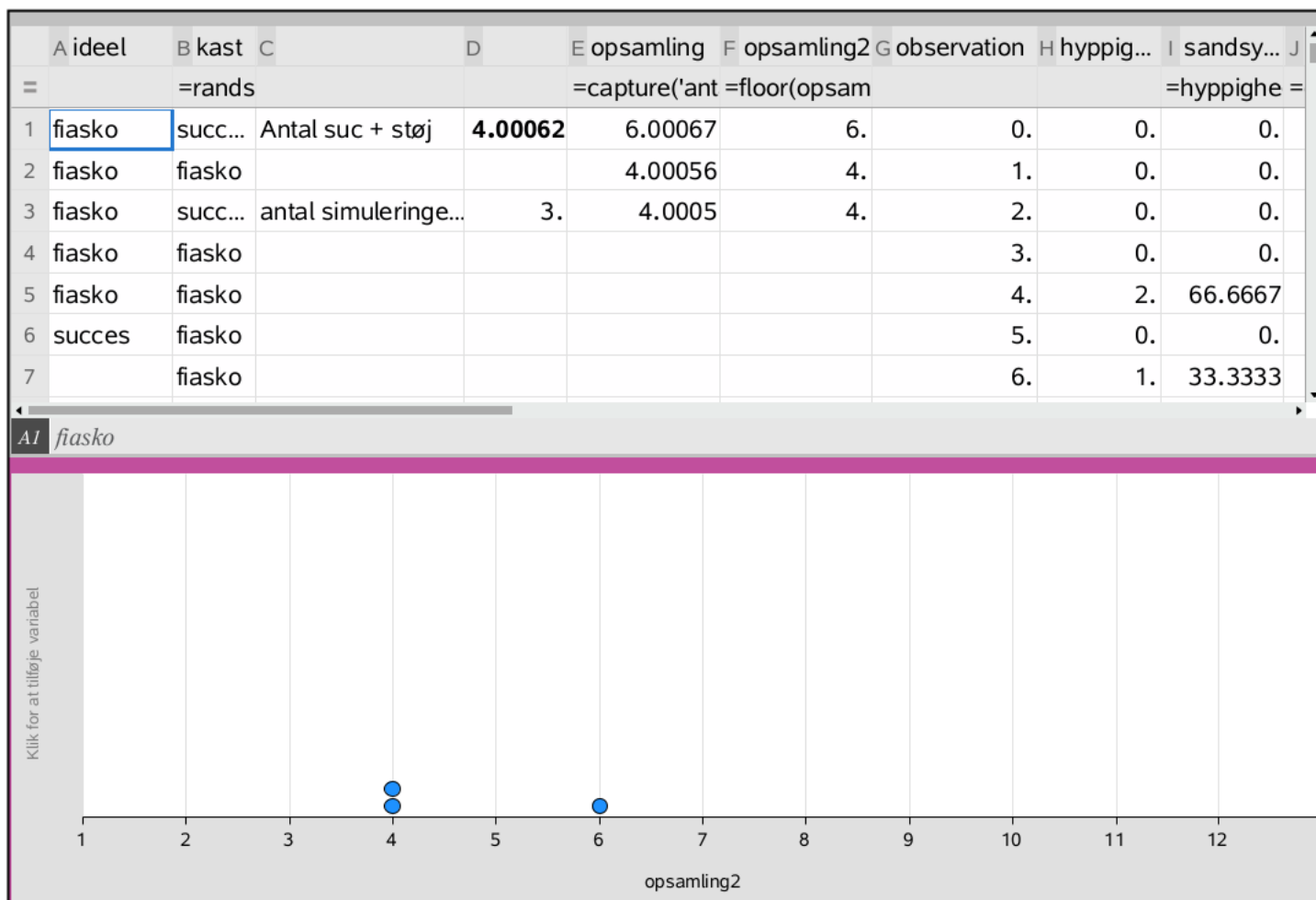
Simulering af terninge kast

Søjlen ideel repræsenterer en ideel terning, dvs. en terning hvor sandsynlighedene er $1/6$ for at kaste en 6'er. Da vi er interesseret i antallet af 6'ere kalder vi udfaldet 6 for en succes og alle andre udfald kaldes fiasko. Således har den ideelle søjle 5 fiaskoer og 1 succes.

I søjlen kast får vi Nspire til at udtage 24 værdier fra søjlen ideel på en helt tilfældig måde svarende til at N spire kaster med en ideel terning 24 gange. Dette gøres vha. kommandoen `=randsamp` ("tilfældig prøve"). Herved produceres 24 output der kan være fiasko eller succes, og sandsynligheden for succes er $1/6$.

Vi får Nspire til at tælle succeserne vha `=countif` ("tæl hvis"). Hvis man trykker på CTRL og R **omrører** Nspire dvs. den slår igen 24 gange med terningen. Vha datafangst og `=capture` får vi Nspire til at gemme antallet af 6'ere for hvert slag i søjlen opsamling. **Vælg hurtiggraf** og se hvordan slagene udvikler sig når du får Nspire til at simulere terningens slag.

Bemærk, hvis du vil køre simuleringen igen (og vise den til din mor) skal du blot højreklikke formelfeltet i opsamling og vælge "slet data"



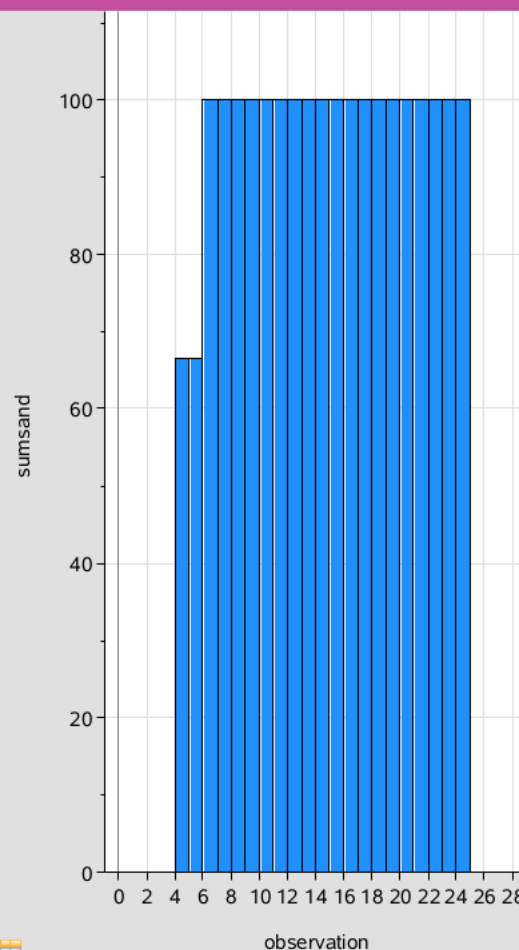
efterfølgende) sandsynlighed for at få 9 seksere ud af 24 slag. Men vi vil se på alle udfald der er mindst lige så skæve som udfaldet 9 seksere (dvs. 10, 11, 12... seksere også tæller med), og det ses at det alt i alt er $100\% - (98.294)\% = 0.01706 = 1,7\%$

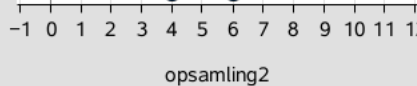
En anden måde at indse dette på er at lave nedenstående søjlediagram. Der minder om en sumkuve, men er det ikke da det jo ikke kan få observationer mellem de hele tal som strengen ved sumkurven antyder (vi taler kun om sumkurver for grupperede data).

Aflæsning er dog som på en sumkurve og vi kan se at man i 98,2% af kastene får 8 eller færre 6'ere, dvs. at vi i 2% af tilfældene får et "skævt" udfald på 9 seksere eller flere (i tabellen kan vi se mere præcist at det er ca. 1.97%)

Normalt beslutter man inden forsøget, ved hvilken sandsynlighed man vil forkaste H_0 hypotesen – den værdi kaldes **signifikansniveauet** og sættes ofte til 5% (men også til tider til 10%, 0,1% eller 1 %)

Konklusion: Vores hypotese kan forkastes med et





samlede antal prikker angiver 100% og hvis der er 19 prikker i en søjle er sandsynligheden for denne observation altså

$$\frac{19}{\text{"samlede antal"}} \cdot 100\%$$

Hvis vi først ændre prik diagrammet til et **søjlediagram** (højreklik) får vi nu antallet udtrykt ved et areal – (19 "kvadrat enheder"). Hvis vi dernæst højreklikker og vælger **skala og vælger densitet** omregner nspire arealet til en sandsynlighed, således at det samlede areal giver 1.

Det ligner næsten en normalfordelingskurve! Vi kan få Nspire til at tegne den bedst normalfordelings kurve til disse data, ved at vælge **analyse** og derefter **vis normalpdf**.

Nspires indbyggede normalfordelingskurve hedder normalpdf(x,μ,σ), og vi kan se at middelværdien for vores data er μ=3.97 og spredningen er σ=1.947.

Bemærk hvorvidt det er realistisk at antage at data bør følge en normalfordeling bør man diskutere.

Bemærk at man kan beregne sandsynligheden som normalfordelingen giver under analyse og så skraver areal den kan endda skraver fra -uendelig til +uendelig!

der skal indtræffe som succes. X betegner antal gange vi får basishændelsen.

Sandsynligheden for H er basissandsynligheden p . Sandsynligheden for X udfald af hændelsen H er da givet ved:

$$P(X = r) = \frac{n!}{r! \cdot (n-r)!} \cdot p^r \cdot (1-p)^{n-r}$$

Bemærk: $(1-p)$ angiver sandsynligheden for fiasko

Altså er den teoretiske sandsynlighed for 9 seksere ud af 24 kast med en ideel terning

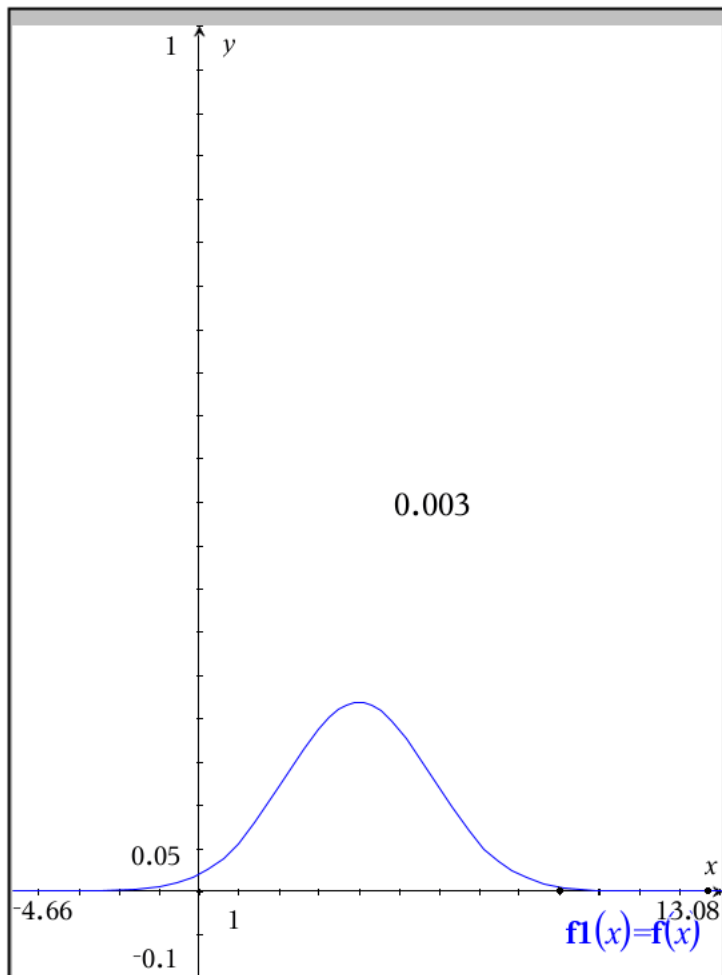
$$P(X = 9) = \frac{24!}{9! \cdot (24-9)!} \cdot \left(\frac{1}{6}\right)^9 \cdot \left(1 - \frac{1}{6}\right)^{24-9} = \frac{24!}{9! \cdot (24-9)!} \cdot \left(\frac{1}{6}\right)^9 \cdot \left(1 - \frac{1}{6}\right)^{24-9} \rightarrow 0.008421$$

Eller kort $\text{binomPdf}\left(24, \frac{1}{6}, 9\right) \rightarrow 0.008421$

Vi ser at dette skulle svare til den eksperimenteller værdi på 1.28% som vi fandt i regnearket. Ønsker man at bestemme sandsynligheden for "udfald mindst lige så skæve som 9 seksere" skal man bestemme $P(X \geq 9)$ der gøres således

$$\text{binomCdf}\left(24, \frac{1}{6}, 9, 24\right) \rightarrow 0.01176 \text{ svarende til værdien i eksperimentet der gav 1.97\%}$$

Opgave 2



$n:=24 \rightarrow 24.$

$p:=\frac{1}{6} \rightarrow 0.166667$

$my:=p \cdot n \rightarrow 4.$

$sigma:=\sqrt{n \cdot p \cdot (1-p)} \rightarrow 1.82574$

$f(x):=\frac{1}{\sqrt{2 \cdot \pi \cdot sigma}} \cdot e^{\frac{-1}{2} \cdot \left(\frac{x-4}{sigma}\right)^2} \rightarrow \text{Udført}$

$\int_9^{\infty} \left(\frac{1}{\sqrt{2 \cdot \pi \cdot sigma}} \cdot e^{\frac{-1}{2} \cdot \left(\frac{x-4}{sigma}\right)^2} \right) dx$
 $\rightarrow 0.003085 \text{!}$

Altså 0,3 % sandsynlighed for at slå 9
 seksere på 24 slag.

Fødselsdagsproblemet

"Hvad er sandsynligheden for at ingen herinde har fødselsdag samme dag?"

Her er to eksempler på forskellige tilgange til fødselsdagsproblemet. Det første er en simulering og den anden er en mere teoretisk beregning. Den teoretiske tilgang bruger jeg tit som en (avanceret) icebreaker første matematik time i 1.g: Jeg starter timen med at lave vædemålet: Hvis ingen herinde har fødselsdag samme dag giver jeg kage, hvis nogen 2 eller flere har fødselsdag samme dag giver de kage. Vi repeterer først kort lidt sandsynlighedsregning (og/eller samt komplementære hændelser), derefter taster alle med. De kommer godt rundt i deres nye CAS værktøj, og det lykkedes faktisk hver gang for langt de fleste...

Janus Lylloff, Marts 2023

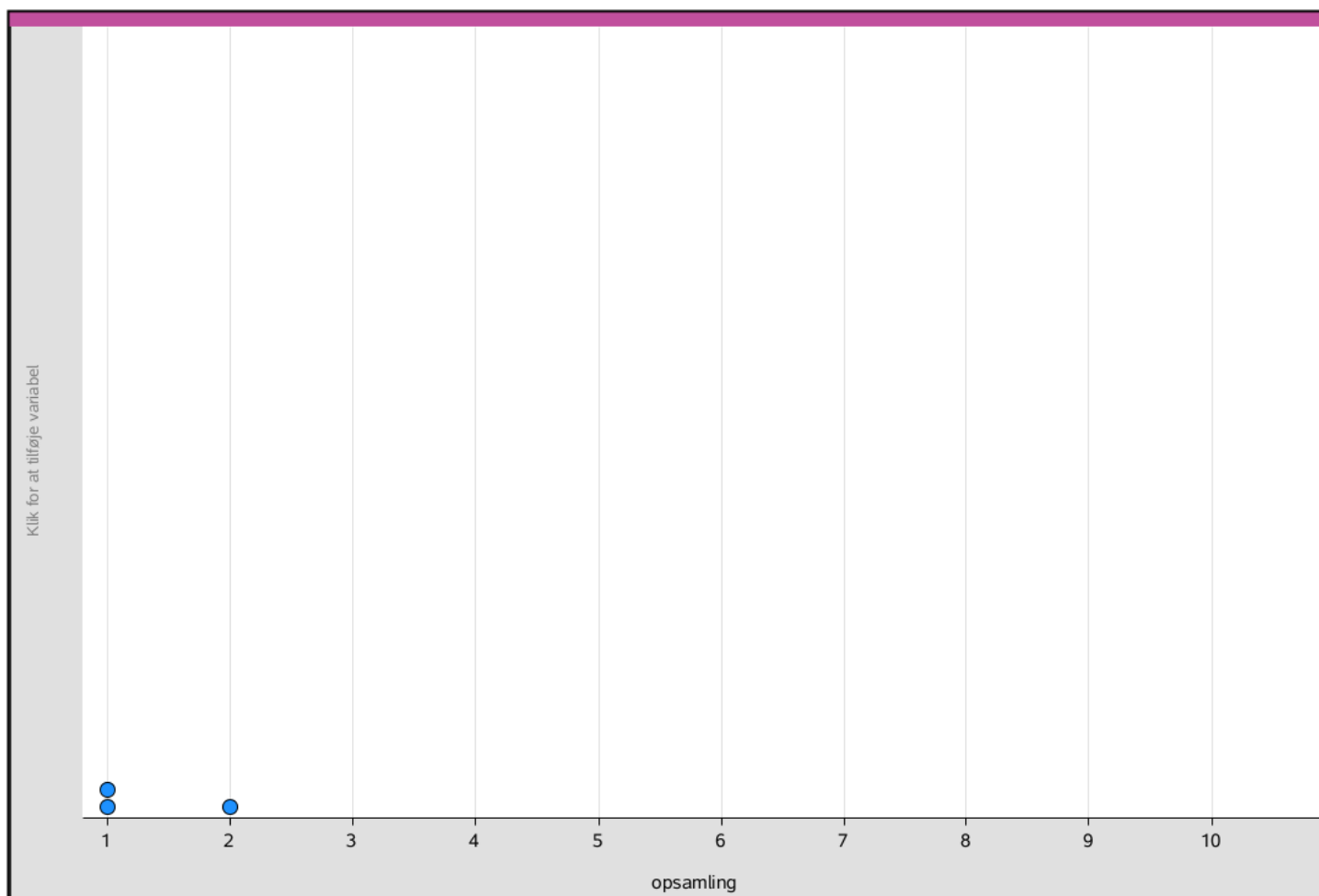
Om simuleringen

Umiddelbart er den ikke så svær at forstå, men teknikken kan drille en smule. I simuleringen skal vi opsamle den største hyppighed for fødselsdag og den er oftest 1 eller 2. Bruger man automatisk opsamling, vil den ignorere resultatet hver gang det samme tal kommer to gange (eller flere) efter hinanden. Det kan man komme omkring ved at tilføje lidt støj til den variabel man samler op. Se næste side hvordan det gøres.

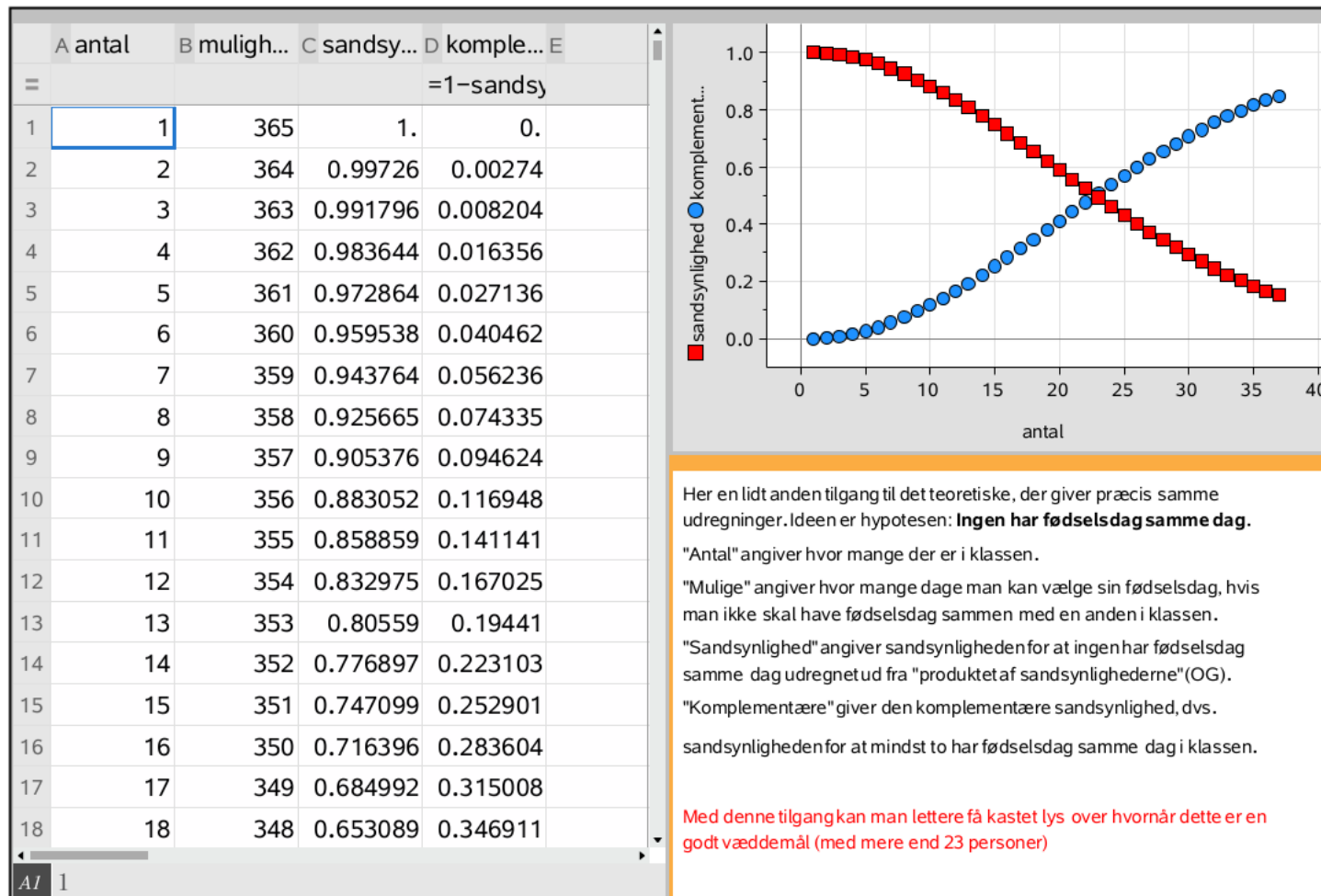
	A dato	B klasse	C hyppighed	D	E opsamling	F	G
=	=seq(i,i,1,365)	=randsamp(datc	=frequency(klas		=capture(maks,1)		
1	1	136	0	Max hyppighed + støj	2.0004		
2	2	323	0	1.00012	1.00059		
3	3	256	0	Antal simuleringer	1.00035		
4	4	255	0	3	1.00099		
5	5	5	0	Antal gentagelser			
6	6	7	0	1			
7	7	29	0				
8	8	222	0	%-del med dubletter			
9	9	219	0	33.3333			
10	10	137	0				
11	11	312	1				

A1 =1

- 1) I søjlen "dato" er alle årets dage nummereret fra 1 til 365.
- 2) I søjlen "klasse" genereres en tilfældig klasse med 30 elever.
- 3) I søjlen "hyppighed" er hyppigheden for den enkelte dag.
- 4) "Maks hyppighed + støj" tæller op hvad den højeste hyppighed er. **Støj:** Vi er her nød til at tilføje støj, hvis vi vil bruge automatisk dataopsamling. Automtisk dataopasmling (angives af 1 i capture kommandoen) samler kun data en når tallet ændres. Da vi i de fleste simuleringer får



Teoretisk beregning



Minecraft og binomialfordelingen

Kan man bevise snyd med matematik?

Matt Parker er en matematiker der laver "Stand-up Maths" om mange forskellige emner. I februar 2021 lavede han en film hvor han ser på Minecraft med matematiske briller. Se filmen nedenfor, og svar på spørgsmålene undervejs.

1. Produkt

Fremlæg for klassen hvad dette handler om. I kan gennemgå opgaverne med klassen, som minimum skal I vise jeres udregninger. I må gerne fylde mere på end der er i spørgsmålene nedenfor.

2. Filmen og spørgsmålene

Til filmen er lavet en række spørgsmål, der passer til filmens forskellige dele. I overskriften står tidskoden for hvor i filmen spørgsmålene behandles. *Læs spørgsmålene inden i ser de pågældende dele af filmen.*



<https://www.youtube.com/watch?v=8Ko3TdPy0TU>

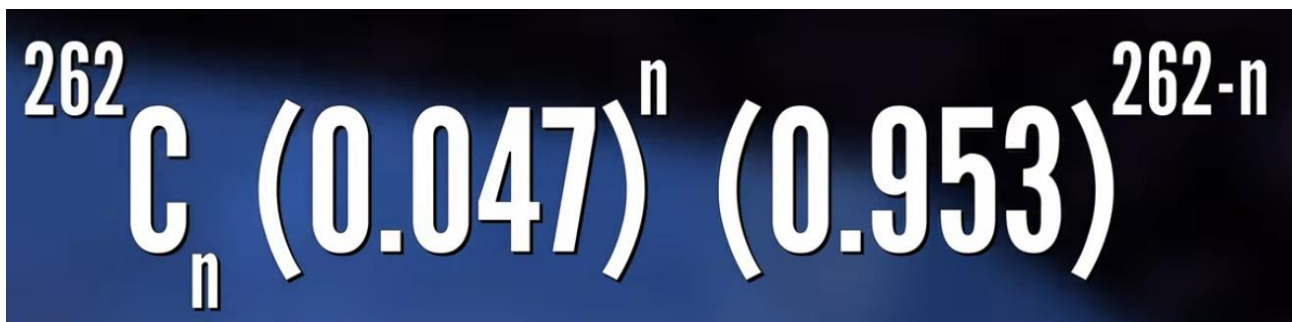
3. Minecraft (forforståelse) [0:00-4:46]

Mange kender nok Minecraft som et spil hvor man bygger ting op i samarbejde med andre, men det kan spilles på mange forskellige måder. Det er faktisk også muligt at gennemføre Minecraft...

- a) Hvordan gennemfører man Minecraft? Fokuser på "Ender Pearl" og "Blaze rods".
- b) Hvad er speedruns?
- c) Hvem er "Dream" og hvad går kontroversen i "Minecraft verdenen" ud på?

4. Minecraft og matematik [4:46-7:35]

- a) Hvorfor kan man betragte Dreams speedrun som en et Binomialfordelt stokastisk eksperiment?
- b) I filmen dukker nedenstående formel op og faktoren ${}^{262}C_n$ kaldes for "choose function" – hvad hedder den på dansk og hvordan skriver vi den? (Lyt godt til hans forklaring af det, eller kig grundigt på resten af formlen...)
- c) Regn ud hvad "choose funktionen" er i tilfældet $n=42$ (brug Nspires indbyggede kommando).
- d) Forklar hvordan man kommer frem til formlen på baggrund af Dreams speedrun.


$${}^{262}C_n (0.047)^n (0.953)^{262-n}$$

5. Dreams speedrun (beregninger der ikke er i filmen)

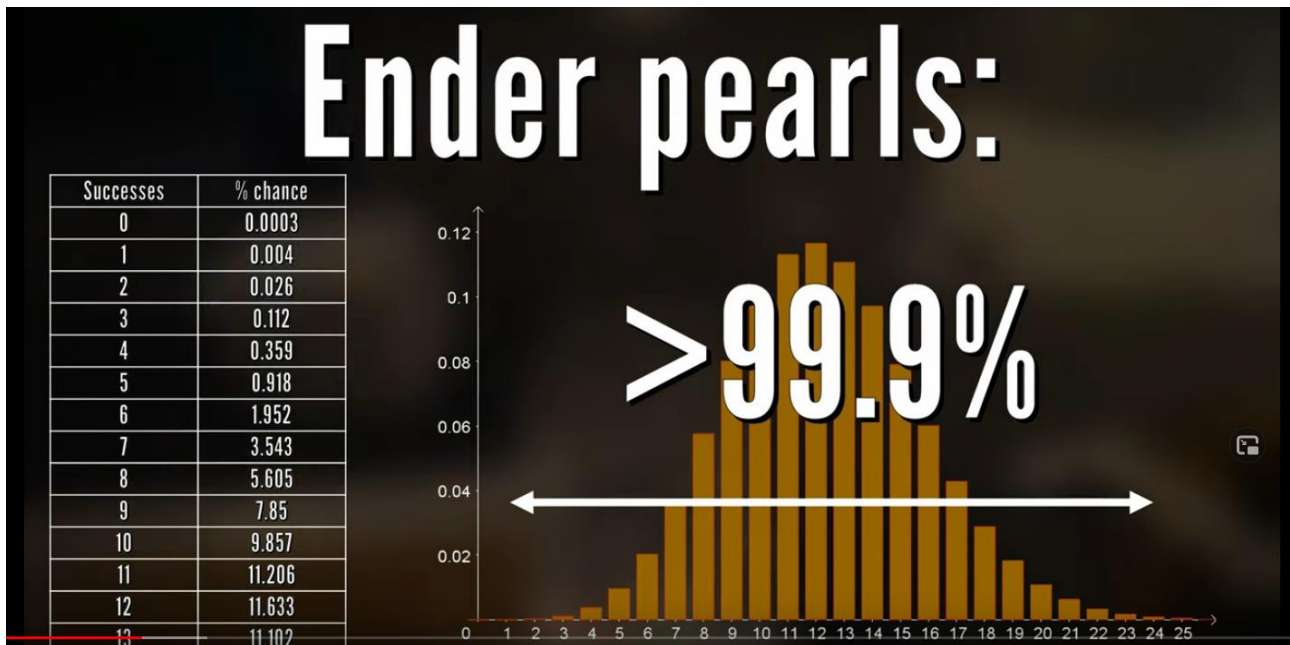
Vi kan altså betragte Dreams indsamling af Ender Pearls, som et binomialfordelt stokastisk eksperiment med 42 succeser.

- a) Hvad er n og p ?
- b) Hvad er middelværdien og spredningen for sådan et binomialfordelt eksperiment?
- c) Hvad er de kritiske værdier for et sådant eksperiment?

Bemærk: 42 Ender Pearls ud af 262 er en kritisk værdi, og hvis vi lavede en binomialtest, ville vi naturligvis få en sandsynlighed på langt under 5% for mindst 42 succeser. Det må man også forvente når man har at gøre med en verdensrekord, da verdensrekorder forekommer meget sjældent. I det følgende vil vi derfor undersøge nærmere hvor sandsynligt dette er.

6. Videre undersøgelse [7:35-9.20]

a) Forklar denne tabel og graf.



- b) Matt Parker beregner sandsynligheden for netop af få 42 Ender Pearls til 0,0000000004% - tjek beregningen i Nspire.
- c) Hvorfor er det mere korrekt at finde sandsynligheden for 42 eller mere Ender Pearls? Prøv at lave denne beregning i Nspire (Nspire giver ikke helt det samme tal som Matt får – vi bruger Matts tal i det efterfølgende).
- d) Hvordan kommer man fra sandsynligheden på $5.6 \cdot 10^{-12}$ til odds 1 i $1.8 \cdot 10^{11}$?
- e) Hvorfor er det lettere at forstå en sandsynlighed når den omregnes til ods?
- f) Hvad er sandsynligheden for at få mindst 211 *Blaze rods* ud af 305 forsøg? Tjek Matts udregning.

7. Den samlede sandsynlighed [9:20-9:40]

- a) Hvad er den samlede sandsynlig for Dreams rekordløb?
- b) Hvorfor ganges de to sandsynligheder samme? (Hvilken regel er det der bruges).

I afsnit 8 skal vi prøve at begribe tallet 1 ud af $2 \cdot 10^{22}$, dette er et meget vigtigt tal...

... Utrolig mange detaljer [9:40-18:22]

Se dette og se hvor meget I forstår (**må også springes over**)

Konklusionen er at det ikke er nødvendigt med alle de detaljer...

... Dreams svar [18:23-23:10]

Dream får lavet et matematisk modbevis, der er meget usammenhængende og uordentligt, det er sjovt at se hvordan noget sådan pilles fra hinanden (**men må gerne springes over**).

8. Sammenligning med andre spil [23:10-28:54]

Her viser Matt Parker en meget elegant måde at illustrere tallet

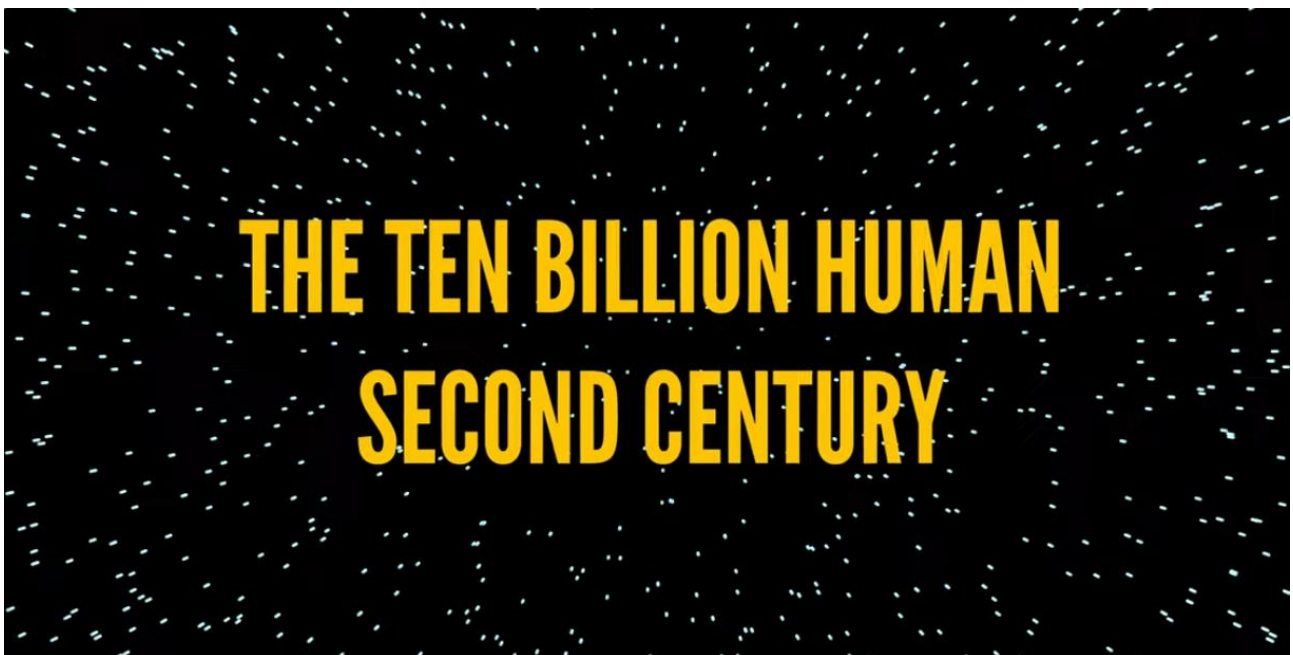
1 ud af $2 \cdot 10^{22}$.

- a) På Læn jer tilbage og nyd det – kan I genfortælle det i overskrifter for klassekammeraterne?

9. Sammenligning med jordens befolkning [28:54-35:10]

Matt undersøger hvilke ting der KAN ske for mennesker, ved at introducerer enheden 10 mia. mennesker sekund århundrede (bemærk billion på engelsk svarer til en dansk milliard).

- a) Skriv en sammenfatning af hvad den enhed kan bruges til.
b) Hvordan kan dette bruges i forbindelse med Minecraft?



10. Konklusion [35:10-]

- a) Hvad er konklusionen på Matts undersøgelse? Vær meget præcis her...
b) Hvordan kan man affeje argumentet om at Dream har trænet meget og derfor er så hurtig?

Bomber over London

Dette er lavet med elever i grupper, og er blot inspiration. Det stammer fra "Hvad er matematik? 1" Side 370–

Historien ultrakort:

Under anden verdens krig blev London ramt af 537 V2 bomber, og hvis leder efter et mønster i bombernes nedslag, skal man bruge noget statistik...

London deles ind i 576 lige store felter, og man tæller hvor mange bomber hvert felt er blevet ramt af.

Bemærk: Der er lige lidt under en bombe pr. felt.

Det viser sig at $\frac{229}{576} \cdot 100 \rightarrow 39.7569\%$ af felterne ikke rammes af bomber og $\frac{43}{576} \cdot 100 \rightarrow 7.46528\%$

% af felterne rammes af mindst 3 bomber.

Det store spørgsmål, kunne de (onde) tyskere styre deres bomber og dermed gå efter bestemte mål, eller var der tale om tilfældige (stokastiske) hændelser der hvor bomberne slog ned. Vi undersøger sagen og indfører en H_0 hypotese: Bomberne rammer felterne tilfældigt

Denne hypotese omsættes til at alle felter har en sandsynlighed for at blive ramt af en bombe på

$\frac{1}{576} \cdot 100 \rightarrow 0.173611\%$

Vi simulerer det...

Øvelse 9.2

- Opret en liste med numrene 1 til 576 svarende til felterne, som vi har inddelt det sydlige London i.
- Opret en ny liste på 537 celler, hvor hver celle indeholder et tilfældigt valgt tal fra den første liste. Dette tal angiver således, hvor pågældende bombe landede.
- Opret en tabel og et diagram, der viser antallet af felter, der slap helt fri, antallet der blev ramt netop 1 gang, antallet der blev ramt netop 2 gange osv.
- Sammenhold den simulerede fordeling med den observerede fordeling.
Konklusion: Ligner den observerede fordeling en typisk tilfældig fordeling af bomberne?
- Da du har oprettet listen med de 537 ved hjælp af en random-funktion, kan du let gentage forsøget.

	A felter	B bomber	C felt_hy...	D træffere	E antal_...	F	G	H	I	J	K
=	=randsam										
1	1	133	2	0	216						
2	2	45	0	1	235						
3	3	240	1	2	84						
4	4	90	0	3	30						
5	5	434	1	4	11						
6	6	410	1	5	0						
7	7	25	0	6	0						

Vi regner lidt på det...

Sandsynligheden for at et felt blive ramt hvis der kastes en bombe (der er 576 felter)?

$$p(\text{bombe}) = \frac{1}{576} \cdot 100 \rightarrow 0.173611 \%$$

Sandsynligheden for at et felt ikke bliver ramt.

$$p(\text{ikke bombe}) = \frac{575}{576} \cdot 100 \rightarrow 99.8264 \%$$

Bemærk: $p(\text{ikke bombe}) = 1 - p(\text{bombe})$

$$\left(1 - \frac{1}{576}\right) \cdot 100 \rightarrow 99.8264 \%$$

Ikke bombe er **den komplementære hændelse** til bombe.

Hvad er sandsynligheden for at man IKKE bliver ramt af 2 ud af to 2 bomber?

Da der er tale om to afhængige hændelser **ganges sandsynlighederne sammen** (jf. sætning 2 side 230), man rammes ikke af 1. bombe **OG** ikke af 2. bombe. $P(\text{ikke bombe}) \cdot p(\text{ikke bombe})$

$$\left(1 - \frac{1}{576}\right)^2 \cdot 100 \rightarrow 99.6531 \% \text{ eller } \left(\frac{575}{576}\right)^2 \cdot 100 \rightarrow 99.6531 \%$$

Hvad er sandsynligheden for at man ikke bliver ramt af 537 bomber?

$$\left(\frac{575}{576}\right)^{537} \cdot 100 \rightarrow 39.3332 \%$$

Overraskende... Når der nu er ca. 1 bombe pr. felt...

Sandsynligheden for netop at blive ramt en gang?

$$p(\text{rammes første gang}) = \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{536} \cdot 100 \%$$

$$p(\text{rammes anden gang}) = \frac{575}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \%$$

Altså er sandsynligheden for at blive ramt af netop en bombe den samme lige meget hvilken bombe det er. Man kan altså blive ramt af 1 bombe på 537 forskellige måder, da der er tale om en **ELLER** situation skal **sandsynlighederne lægges sammen** (jf. sætning 2 side 230).

$$p(\text{rammes netop en gang}) = 537 \cdot p(\text{rammes første gang})$$

$$p(\text{rammes netop en gang}) = 537 \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{536} \cdot 100 \rightarrow 36.7338 \%$$

Sandsynligheden for netop at blive ramt to gange?

For at kunne holde styr på denne opgave, er det en god ide at tænke på det som om at bomberne er nummereret fra 1 til 537 og de så bliver kastet i en tilfældig rækkefølge (altså at nummeret ikke er kaste rækkefølgen)

Vi ser på mulige kombinationer ved at være lidt systematiske

1) På hvor mange måder kan man blive ramt af to bomber, **hvis man bliver ramt af den første?:**

$$p(1. \text{ og } 2. \text{ bombe}) = \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 0.000119 \%$$

$p(1. \text{ og } 3. \text{ bombe}), p(1. \text{ og } 4. \text{ bombe}) \dots p(1. \text{ og } 537. \text{ bombe})$ [Disse giver alle præcis det samme som ovenfor]
der er altså 536 af disse kombinationer, så

$$p(\text{rammes af } 1. \text{ bombe og en anden}) = 536 \cdot \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 0.063766 \% (*)$$

2) På hvor mange måder kan man blive ramt af to bomber, **hvis man bliver ramt af den anden bombe?:**

$$p(2. \text{ og } 1. \text{ bombe}) = \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 0.000119 \%$$

Bemærk: Den har vi også med i punkt 1. men for at holde systemet tager vi lige denne med her også.

$p(2. \text{ og } 3. \text{ bombe}), p(2. \text{ og } 4. \text{ bombe}) \dots p(2. \text{ og } 537. \text{ bombe})$

der er altså 536 af disse kombinationer så $p(\text{rammes af } 2. \text{ bombe og en anden})$ er den samme som (*).

3) Dette skal man blive ved med for hver af de 537 bomber, hvor man i hver af dem kommer til at gentage netop en af de tidligere kombinationer.

Nu har vi set hvor mange forskellige måder et felt kan blive ramt af netop 2 bomber. Den samlede sandsynlighed for at blive ramt af netop 2 bomber kan så bestemmes vha. sætning 2 (side 390 eller næste side her), da feltet enten bliver ramt på den ene eller anden eller... der er altså 537 af de samme sandsynligheder, og derfor bliver det 537 gange sandsynligheden (*),

$$537 \cdot 536 \cdot \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 34.2423 \%$$

dog skal vi huske at vi i hvert af de 537 skridt er kommet til at gentage en kombinationsmulighed, og skal derfor dele med 2, så sandsynligheden for at blive ramt af netop 2 bomber er

$$p(\text{rammes netop to gange}) = \frac{537 \cdot 536}{2} \cdot \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 17.1211 \%$$

Kort fortalt: Første bombe kan vælges mellem 537 bomber og anden bombe kan vælges på 536 måder. Om man først bliver ramt af bombe 402 og derefter af bombe 305 eller omvendt er ligemeget derfor deles med 2.

Bemærk: Tallet $\frac{537 \cdot 536}{2} \rightarrow 143916$ angiver på hvor mange måder man kan udtage 2 bomber ud af 537 bomber.

Det er et vigtigt tal lige om lidt...

Sandsynligheden for at blive ramt af mindst 3 bomber:

Sandsynligheden for at blive ramt af 0, 1 eller 2 bomber er den komplementære hændelse

Sandsynligheden herfor er

$$p(\text{ingen bomber}) + p(\text{netop en bombe}) + p(\text{netop 2 bomber}) = 39.3 + 36.7 + 17.1 \rightarrow 93.1\%$$

$$p(\text{mindst 3 bomber}) = (1 - 0.931) \cdot 100 \rightarrow 6.9\%$$

Dette er over et 5% signifikansniveau og altså får vi underbygget H_0 hypotesen.

Sætning 2: Sandsynligheder ved kombination af uafhængige hændelser

1) Hvis to hændelser M og N ikke har noget fælles udfald, så gælder der om sandsynlighederne:

$$P(\text{enten } M \text{ eller } N) = P(M) + P(N)$$

2) Hvis to hændelser A og B er uafhængige, så gælder der om sandsynlighederne:

$$P(\text{både } A \text{ og } B) = P(A) \cdot P(B)$$

Binomialfordelingen

Kan man ikke gøre alt dette meget lettere??

Jo man kan!

Alt det tælleri kan gøres vha.

binomialkoefficienten, der er defineret til højre.

Antallet af måder hvorpå man kan udtage (ramme) to bomber ud af 536 bomber beregnes således

$$\frac{537!}{2! \cdot (537-2)!} \rightarrow \frac{537!}{2 \cdot 535!}$$

Nspire kan ikke regne det om til decimaltal, men det kan vi heldigvis, fordi vi ved hvad fakultet betyder.

$$\frac{537 \cdot 536 \cdot 535 \cdot 534 \cdot \dots \cdot 1}{2 \cdot 535 \cdot 534 \cdot \dots \cdot 1} = \frac{537 \cdot 536}{2}$$

Hov det er jo den samme som vi så før...!

Definition: Binomialkoefficienter $K(n, r)$

Antallet af måder, hvorpå vi kan vælge en stikprøve på r individer ud af en population på n , betegnes $K(n, r)$. Tallene kaldes binomialkoefficienter. De indgår i Pascals trekant. Af og til anvendes symbolet $\binom{n}{r}$ for $K(n, r)$.

Sætning 3: Formlen for binomialkoefficienten

$$K(n, r) = \binom{n}{r} = \frac{n!}{r! \cdot (n-r)!}$$

Med denne smarte omskrivning kan vi regne tallet ud som Nspire ikke vil.

$$\frac{537 \cdot 536}{2} \rightarrow 143916.$$

Nspire kan dog godt regne det ud med denne kommando.

$$nCr(537, 2) \rightarrow 143916$$



Så giver er det ret let at begribe hvorfor at binomialfordelingens sandsynlighedsfordeling ser således ud. Vi bruger den på næste side.

Sætning 4: Binomialfordelingens sandsynlighedsfordeling

Hvis den stokastiske variabel X er binomialfordelt $X \sim b(n, p)$ så udregnes sandsynligheden for, at der optræder held k gange ud af de i alt n mulige således:

$$P(X=k) = K(n, r) \cdot p^k \cdot (1-p)^{n-k} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Antallet af måder
hvor på udfaldet
kan kombineres.

Sandsynlighed
for succes

Sandsynlighed
for fiasko
den komplementære
hændelse

Sandsynligheden for at blive ramt af netop 2 bomber er bemærk notationen.

$$P(X=2) = \frac{537 \cdot 536}{2} \cdot \frac{1}{576} \cdot \frac{1}{576} \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 17.1211 \%$$

Dette kan også let bestemmes således

$$P(X=2) = nCr(537, 2) \cdot \left(\frac{1}{576}\right)^2 \cdot \left(\frac{575}{576}\right)^{535} \cdot 100 \rightarrow 17.1211 \%$$

Eller endnu smartere

$$\text{binomPdf}\left(537, \frac{1}{576}, 2\right) \cdot 100 \rightarrow 17.1211 \%$$

Den kummlerede binomialfordeling

Som vi så er man ikke altid interesseret i kun et enkelt udfald, men måske er spørgsmål som: Hvad er sandsynligheden for at blive ramt af højst 3 bomber. Dette kan man beregne som

$$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

Der på Nspire'sk ser således ud

$$\text{binomPdf}\left(537, \frac{1}{576}, 0\right) + \text{binomPdf}\left(537, \frac{1}{576}, 1\right) + \text{binomPdf}\left(537, \frac{1}{576}, 2\right) + \text{binomPdf}\left(537, \frac{1}{576}, 3\right) \rightarrow 0.984981$$

Det bliver hurtigt besværligt, og derfor kan man med fordel benytte den kummlerede binomialkoefficient

$$\text{binomCdf}\left(537, \frac{1}{576}, 0, 3\right) \rightarrow 0.984981 \text{ hvor } 0 \text{ er angiver den nedre grænse og } 3 \text{ den øvre grænse.}$$

På lignende vis kan man bestemme sandsynligheden for at blive ramt af mellem 2 og 7 bomber

$$\text{binomCdf}\left(537, \frac{1}{576}, 2, 7\right) \rightarrow 0.239325$$

Middelværdi og spredning

Sætning 5: Middelværdi og spredning for binomialfordelinger

Hvis den stokastiske variabel X er binomialfordelt $X \sim b(n, p)$, så kan middelværdi og spredning beregnes ud fra formlerne:

$$\mu = E(X) = n \cdot p$$

$$\sigma(X) = \sqrt{n \cdot p \cdot (1-p)}$$

Middelværdien for eksemplet med bomber over London er $537 \cdot \frac{1}{576} \rightarrow 0.932292$ altså vil hvert felt

i gennemsnit blive ramt af 0.93 bomber.

Spredningen der er mere interessant er

$$\sqrt{537 \cdot \frac{1}{576} \cdot \left(1 - \frac{1}{576}\right)} \rightarrow 0.964714$$

OPGAVE: Hvad vil et normalt og hvad vil et exceptionelt udfald være for dette "forsøg" (gense definitionen side 386). Skriv en forklarende tekst til.

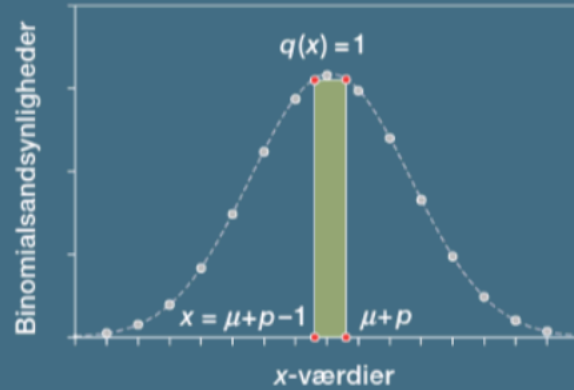
Opgave

Hvad er det mest sandsynlige antal bomber, som et felt bliver ramt af? Som vi så før kan man ikke (helt) bruge middelværdien, da den jo gave et decimaltal (under 1). Nogen har tænkt sig om og er kommet frem til denne sætning:

Sætning 6: Det mest sandsynlige udfald for en binomialfordeling

Hvis $\mu + p$ ikke er et helt tal, er det mest sandsynlige udfald det største hele tal under $\mu + p$.

Hvis $\mu + p$ er et helt tal, er de to mest sandsynlige udfald tallene $\mu + p$ og $\mu + p - 1$.



Brug den og svar på spørgsmålet, om hvad den mest sandsynlige hændelse er?

STANDARD NORMALFORDELINGEN

$$f1(x) := \text{normPdf}(x)$$

a) Træk i punktet P og forklar hvad der sker.

b) Opstil et integral der beregner arealet

a ▶ 0.130618 (du får brug for punktets P's x-koordinat, der er gemt som xx ▶ -1.12348

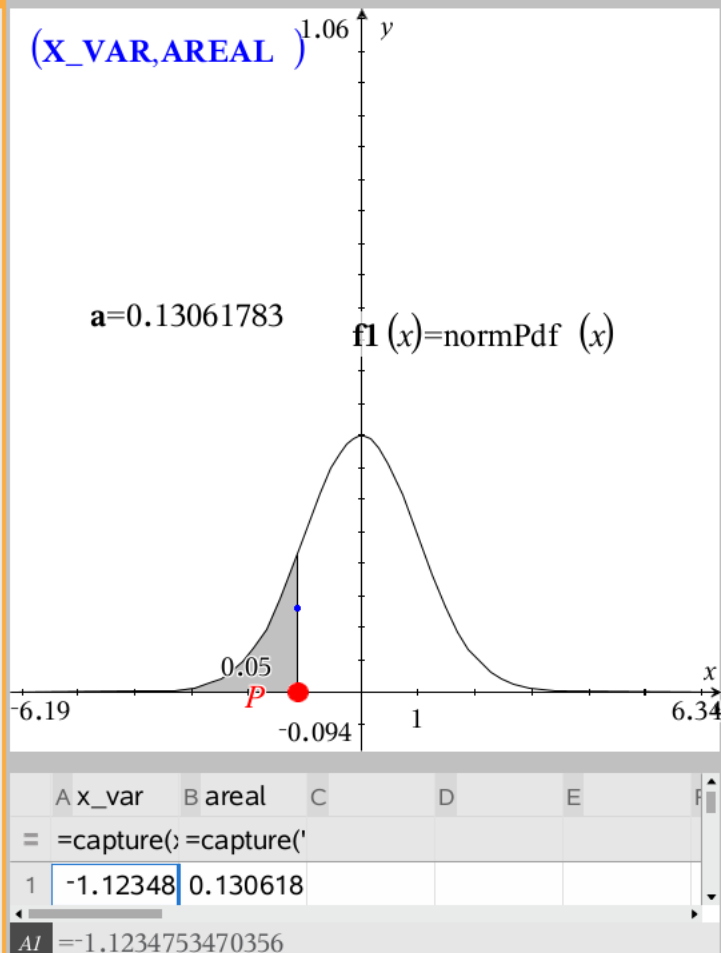
Tillykke! Dette bestemte integral en den funktion, der kaldes **fordelingsfunktionen**.

c) Grafen for fordelingsfunktionen, kan laved med lidt datafangst. Sæt n=1, og træk i punktet P

< > n = 1.

d) Scroll ned og se de tomme linjer, og se hvorledes man let kan regne det i Nspire

$$\text{normCdf}(-\infty, xx)$$



På vej mod normalfordelingen

STANDARD NORMALFORDELINGEN

Tæthedsfunktionen er

$$f(x) := \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{1}{2} \cdot x^2} \quad \text{▶ Udført}$$

Vi finder vendepunkterne

HUSK: Et vendepunkt er hvor tangentens hældning går fra at vokse til at aftage eller omvendt.

Dvs. Det er der hvor $f'(x) = 0$

$$ddf(x) := \frac{d^2}{dx^2}(f(x)) \quad \text{▶ Udført}$$

$$\text{solve}(ddf(x)=0, x) \quad \text{▶ } x=-1 \text{ or } x=1$$

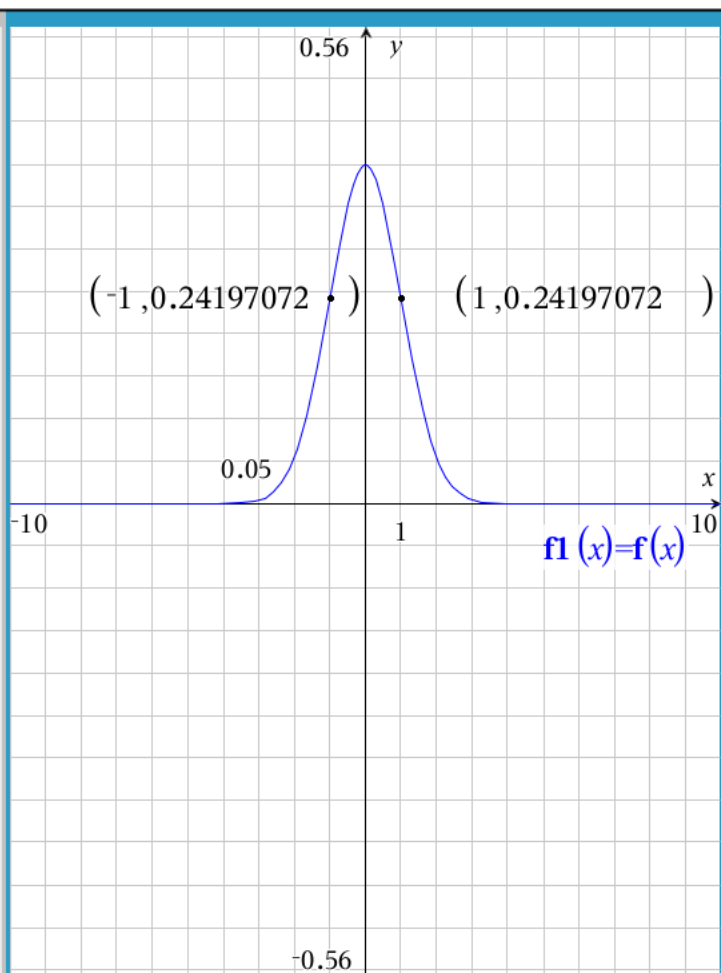
Bestem arealet mellem de to vendepunkter:

$$\int_{-1}^1 f(x) dx \quad \text{▶ } 0.682689 \text{ Svarende til } 68,2\%$$

Hey halloj! 1 er altså standardafvigelsen....

Det samlede areal under kurven er

$$\int_{-\infty}^{\infty} f(x) dx \quad \text{▶ } 1.$$



Betydning af μ (øvelse 8.24)

Betrakt $f_2(x) := f(x - \mu)$ ▶ Udført

b) My forskyder hele grafen vandret.

My angiver middelværdien, dvs. udfaldet med størst sandsynlighed. **My angiver altså en vandret forskydning...**

Vi finder vendepunkterne

HUSK: Et vendepunkt er hvor tangentens hældning går fra at vokse til at aftage eller omvendt.

Dvs. Det er der hvor $f'(x) = 0$

$$ddf_2(x) := \frac{d^2}{dx^2}(f_2(x)) \quad \text{▶ Udført}$$

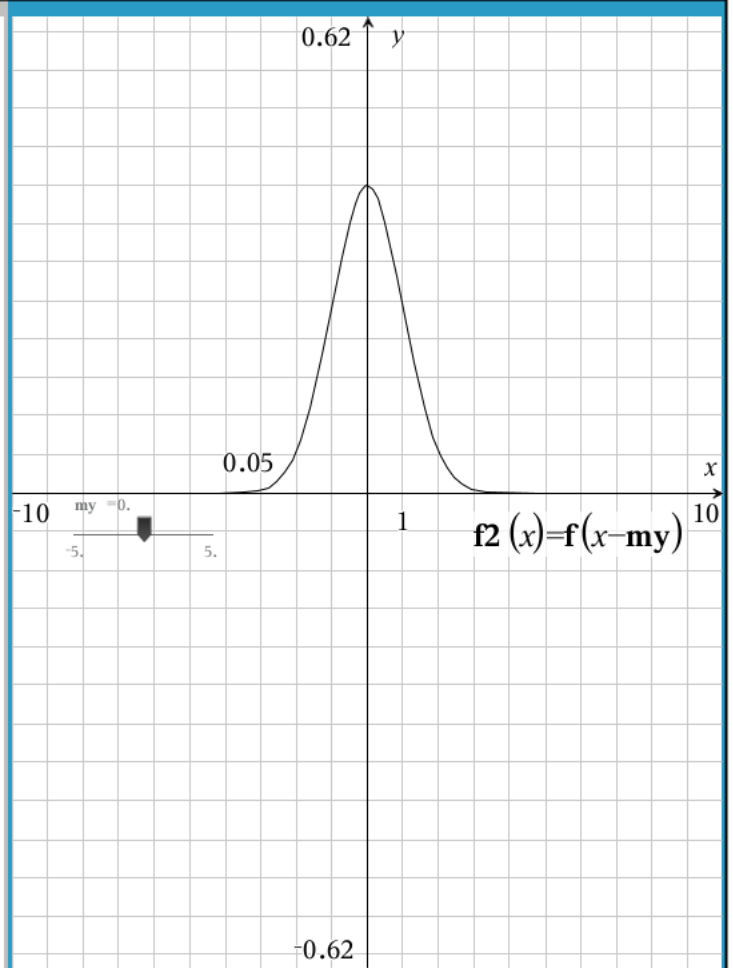
$$\text{solve}(ddf_2(x) = 0, x) \quad \text{▶ } x = -1, \text{ or } x = 1.$$

Vendepunkter forskudt ± 1 ifht. middelværdien μ .

Vendepunkter forskudt ± 1 ifht. middelværdien.

Det samlede areal under kurven er

$$\int_{-\infty}^{\infty} f_2(x) dx \quad \text{▶ } 1.$$



Betydning af μ og σ (øvelse 8.25)

$$f_3(x) := f\left(\frac{x - \mu}{\sigma}\right) \quad \text{▶ Udført}$$

At vi dividerer med σ , svarer til at man skallerer x -værdien

a) Vi finder vendepunkterne

HUSK: Et vendepunkt er hvor tangentens hældning går fra at vokse til at aftage eller omvendt.

Dvs. Det er der hvor $f'(x) = 0$

$$ddf_3(x) := \frac{d^2}{dx^2}(f_3(x)) \quad \text{▶ Udført}$$

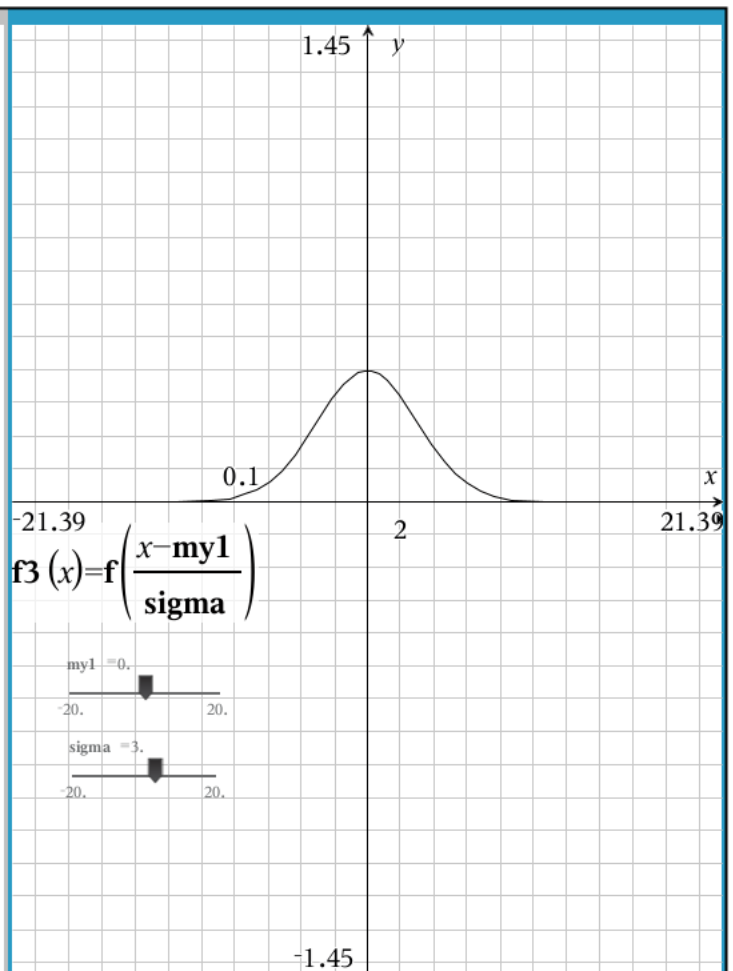
$$\text{solve}(ddf_3(x) = 0, x) \quad \text{▶ } x = -3, \text{ or } x = 3.$$

Vendepunkterne (dvs. spredningen) bliver længere væk fra middelværdien (symmetriaksen).

Det samlede areal under kurven er

$$\int_{-\infty}^{\infty} f_3(x) dx \quad \text{▶ } 3.$$

Arealet bliver større....



Betydning af μ og σ (øvelse 8.25 fortsat...)

$$f_4(x) := \frac{1}{\sigma} \cdot f\left(\frac{x - \mu_1}{\sigma}\right) \rightarrow \text{Udført}$$

b) Vi finder vendepunkterne

HUSK: Et vendepunkt er hvor tangentens hældning går fra at vokse til at aftage eller omvendt.

Dvs. Det er der hvor $f''(x) = 0$

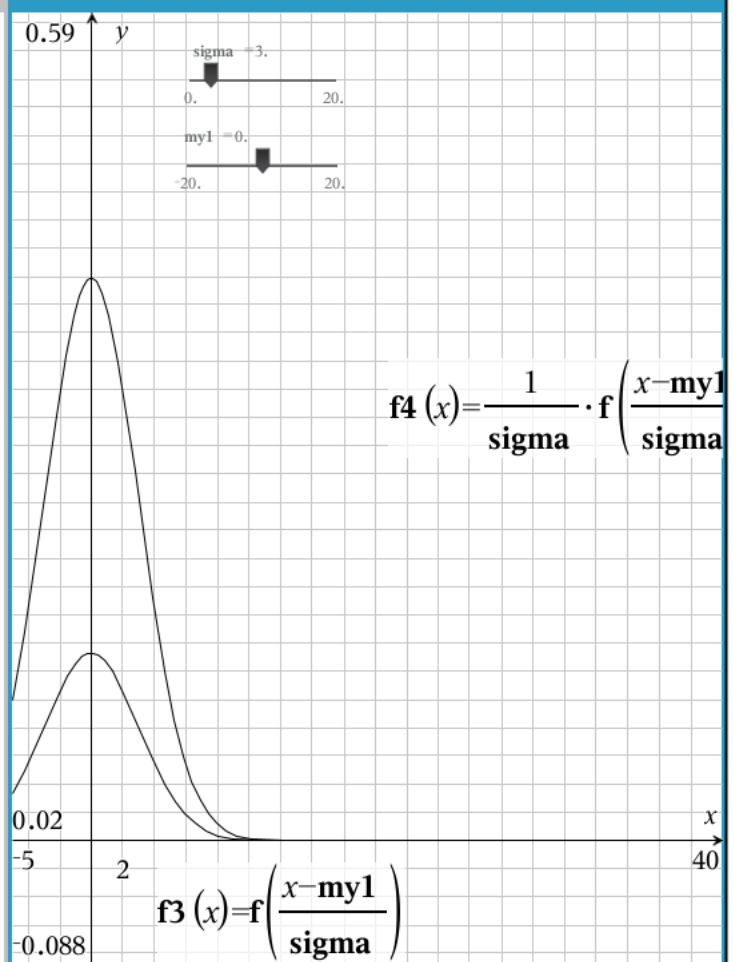
$$ddf_4(x) := \frac{d^2}{dx^2}(f_4(x)) \rightarrow \text{Udført}$$

$$\text{solve}(ddf_4(x) = 0, x) \rightarrow x = -3. \text{ or } x = 3.$$

Spredningen (vendepunkter) ændres ikke.

$$\int_{-\infty}^{\infty} f_4(x) dx \rightarrow 1.$$

Arealet under denne kurve er 1!!!! HURRA, det er en tæthedsfunktion



Opgave 3

7.D2.1 En normalfordelt stokastisk variabel X er givet ved $X \sim N(10, 4)$ og fordelingsfunktion F .

- Tegn grafen for tæthedsfunktionen f for X .
- Bestem $F(7)$, og forklar betydningen af dette tal.

a) Grafen for tæthedsfunktionen tegnes

$$f(x) := \frac{1}{\sigma} \cdot f\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x - 10}{4}\right)^2} \rightarrow \text{Udført}$$

DEN SMARTE MÅDE:

$$f1(x) := \text{normPdf}(x, 10, 4) \rightarrow \text{Udført}$$

b) Bestem $F(7)$. Bemærk: "STORE F" – angiver fordelingsfunktionen.

$F(7)$ er arealet til højre dvs. sandsynligheden for at få højst 7 er 22,66%

$$\text{ved beregning: } \int_{-\infty}^7 f1(x) dx \rightarrow 0.226627$$

