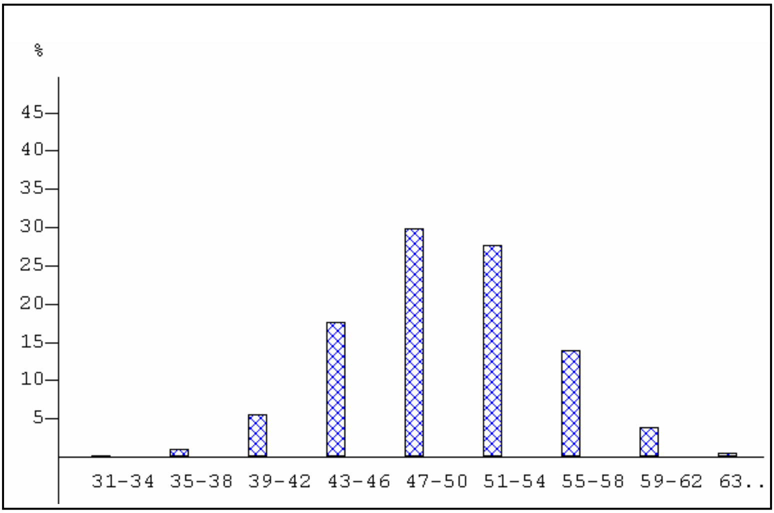
Allan C. Malmberg

MODEL og DATA

Induktiv statistik i skolen



INFA 2007

Af samme forfatter foreligger:

CHANCE – Et it-læremiljø, INFA 2002
En fagdidaktisk behandling af emner fra sandsynlighed som de kan fremtræde i en indledende undervisning. Bogen henvender sig til grundskolens matematiklærere.

Lær om chancer. Sanne og Malene går på opdagelse med computeren, INFA 2005
Bogen viser hvordan de to piger arbejder med emner fra sandsynlighedsregningen. De faglige udfordringer tages op med computeren som et vigtigt værktøj. Bogen henvender sig til elever med særlige interesser og evner for matematik.

Matematik i glimt, INFA 2005
Eksempler på matematiske ræsonnementer og bevisførelser fra forskellige områder af matematikken. Bogen henvender sig til elever med særlige interesser og evner for matematik.

Chance og Risiko, Kan det virkelig passe? INFA 2006
Henvender sig til matematiklærere i grundskolen som i deres uddannelse har stiftet bekendtskab med fagområdet sandsynlighedsregning og statistik. - Bogens emner behandles i form af udfordrende opgaver som fremlægges med løsninger.

CVU København og Nordsjælland
Titangade 11, København N

ISBN 87-91786-19-3

Forord

Denne bog behandler grundlæggende emner fra den induktive statistik: Tilrettelæggelse af *hypotesetest* og beregning af *konfidensintervaller*.

Bogen henvender sig til matematiklærere i grundskolen som i deres uddannelse har stiftet bekendtskab med emner fra sandsynlighedsregning og statistik. Fagligt set vil den også kunne give en indføring i området som kan benyttes af elever i grundskolen med særlig evne og interesse for matematik. Det gælder i særlig grad for elever som har erfaringer fra arbejdet med INFA-programmerne LOD og Kugle1.

De faglige emner i bogen er bygget op om stikprøveudtagelse med tilbagelægning og den tilhørende sandsynlighedsmodel: Binomialfordelingen.

Fremstillingen følger den pædagogiske ide der er gældende for alle INFAs materialer: *Indsigt gennem anvendelse*. Gennem brugen af faglige værktøjer som er samlet i computer-programmet INSTAT, gives en indførelse i det faglige område. Der lægges vægt på at læseren bliver fortrolig med at anvende værktøjerne og med at fortolke resultaterne fra en forelagt hypotesetest. En teoretisk behandling af de faglige emner står derfor ikke forrest, men kan eventuelt komme på tale i en senere fase.

En lang række af opgaver er knyttet til teksten, og de giver læseren gode muligheder for at afprøve den faglige indsigt.

Allan C. Malmberg

Indhold

Indholdsbeskrivelse

Afsnit 1- 4: BIN-Modeller

I bogens indledende fire afsnit gives en indføring i den matematiske model, binomialfordelingsmodellen, der kan benyttes ved beskrivelse af stikprøveudtagelse med tilbagelægning. Stikprøveudtagelsen foretages fra en æske der indeholder kugler af to slags: røde kugler og kugler der ikke er røde. Bogens første 13 afsnit bygger på denne model som benævnes en BIN-Model.

I tilknytning til indføringen introduceres INFA-programmet INSTAT. Dette program indeholder alle de matematiske værktøjer der gøres brug af i bogens behandling af de faglige emner.

Gennem en række eksempler vises hvorledes der gennem anvendelsen af en passende BIN-Model kan hentes oplysninger om chancevurderinger i tilknytning til stikprøveudtagelser. En særlig rolle spiller BIN-Modellernes midterområder, fx det midterområde der indeholder 90% af modellens sandsynligheder.

Når der udføres en stikprøveudtagelse, siger vi at der foretages en observation i tilknytning til den anvendte BIN-Model. En sådan observation kan give et antal af røde kugler der er i god overensstemmelse med den benyttede model, eller den kan give et antal der afviger fra modellen. Vi har da opnået en såkaldt signifikant observation.

I fremstillingen behandles først situationer hvor stikprøvestørrelsen er kendt, dvs. at den benyttede BIN-Model er nøje fastlagt. Her kan INSTAT beregne midterområder med givne sandsynligheder. Derefter ser vi på situationer med ukendt stikprøvestørrelse. Her vil krav til udstrækningen af testens midterområde fastlægge den nødvendige stikprøvestørrelse.

De fire afsnit afsluttes med en omfattende samling af opgaver.

Opgaverne løses ved brug af BIN-Modeller, og de vil vise hvor forskelligartede situationer der lader sig beskrive ved modeller af denne type.

Afsnit 5-6: Konfidens

I de foregående afsnit har vi haft kendskab til indholdet af den æske hvorfra stikprøveudtagelsen er foregået: Vi har vidst hvor stor en andel af kuglerne der var røde. Det svarer til at vi har kunnet opstille en BIN-Model med den rette binomialsandsynlighed.

Vi ser nu på den situation at vi udtager kugler fra en æske hvor andelen af røde kugler er ukendt. Vi skal da på grundlag af resultatet af en stikprøveudtagelse give et bud på hvad den ukendte andel af røde kugler i æsken kan være.

Vore overvejelser vil resultere i et såkaldt konfidensinterval for den ukendte andel af røde kugler i æsken. Til eksempel kan vi nå frem til at et 90%-konfidensinterval er givet ved intervallet

0.21..0.28.

I fremstillingen går vi nærmere ind på hvordan et sådant konfidensinterval kan fortolkes.

Afsnit 7-12: Testgraf

Disse afsnit udgør den centrale del af bogen. Her behandles hypotesetestningens grundbegreber, såsom godtagelse og forkastelse af den opstillede hypotese, fejl af første og anden art og enkeltsidede og dobbeltsidede test.

Behandlingen bygger på et fagligt hjælpemiddel, en testgraf. Til programmet INSTAT indtastes binomialsandsynligheden og stikprøvestørrelsen samt den accepterede risiko for fejl af første art. Programmet beregner da til de givne oplysninger en hypotesetest med angivelse af godtagelsesområde og forkastelsesområde. I tilknytning hertil udskives en graf, en testgraf, som i en grafisk fremstilling viser chancen for at hypotesen godtages ved de mulige stikprøveresultater.

I afsnit 9 undersøges sammenhængen mellem begreberne hypotesetest og konfidens. Hypotesetestning og konfidens-beregning indgår i det faglige område som kaldes Induktiv statistik.

Afsnit 13-14: BIN-Test

Der opstilles krav til en hypotesetest: Testens signifikansniveau angives, og tillige angives den maksimalt tilladte fejl af anden art for én eller to hypotesesandsynligheder.

INSTAT vil da beregne en test som opfylder de stillede krav: For testen beregnes stikprøvestørrelsen, godtagelsesområdet samt testens signifikansniveau og dens risiko for fejl af anden art.

Afsnit 15-16: To eksempler på signifikanstest

Vi afslutter arbejdet med test af hypoteser med to eksempler på test. Vi ser her på en Modeltest og en Forskelstest**.** De to test vil give et indblik i hvordan der kan foretage statistiske beslutninger ud fra forelagte sæt af data.

Ved de to test betragter vi alene fejl af 1.art. Vi opstiller altså kun krav til testens signifikansniveau og omtaler derfor de to test som signifikanstest. Det betyder at vi har styr på risikoen for at testen fejlagtigt forkaster en sand hypotese. Derimod har vi ikke nogen talværdi for sandsynligheden for at en fejlagtig hypotese godtages af testen. - Til begge test er knyttet en samling af opgaver.

BIN - Modeller

1. Oplysninger gennem stikprøver

I en æske findes der kugler. Nogle af kuglerne er røde. Fra æsken udtager du ved tilfældig udtrækning kugler én for én. Hver gang en kugle er udtaget fra æsken, noterer du om det var en rød kugle, derefter lægger du kuglen tilbage i æsken før næste udtagelse. Der er tale om "udtagelse med tilbagelægning".

Kugleudtagelsen fra æsken kan være et billede, *en model*, af mange forskellige chancesituationer. Hvis du har arbejdet med EMMA-temaet KUGLE123, har du haft lejlighed til at se mange kuglemodeller til løsning af et bredt udvalg af opgaver som vedrører chancer og risikoer.

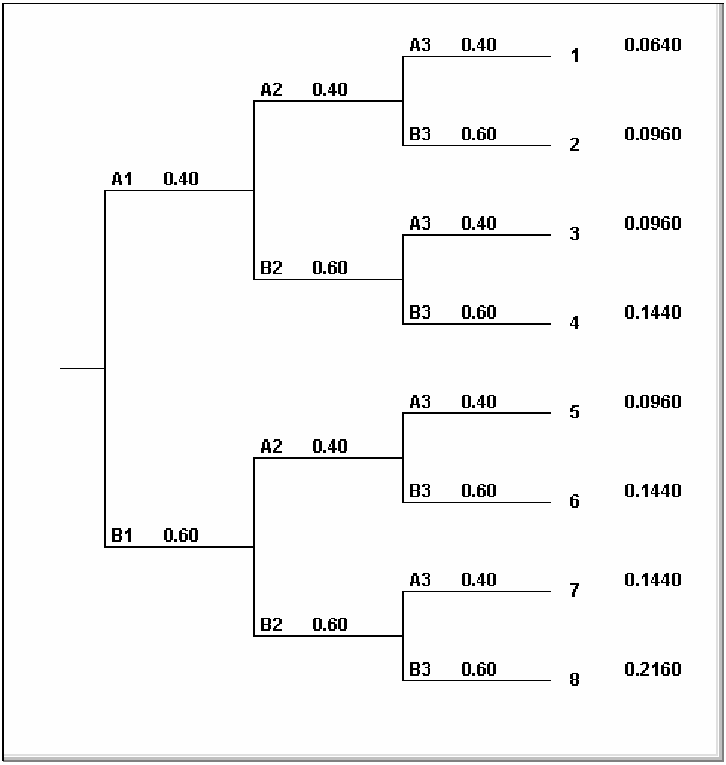
Vi skal her bruge kuglemodeller til at belyse en række spørgsmål som har at gøre med udtagelse af stikprøver. Som du skal se er der mange chancesituationer der kan opfattes som stikprøveudtagelser.

2. Fra model til data

Vi tænker os en æske hvori andelen af røde kugler er 0,4, altså 40%. Fra æsken udtager vi en lille stikprøve på kun tre kugler. Udtagelsen foregår på sædvanlig måde: kuglerne udtages én for én med tilbagelægning efter hver udtagelse.

Hvad er chancen for at vi får en stikprøve som indeholder én rød kugle? Og hvad er chancen for to røde kugler?

Vi kan besvare disse spørgsmål ved hjælp af et chancetræ. Her er billedet af et chancetræ fra INFA-programmet TRÆ:



Der er tale om et 2*2*2-træ: Ved hver af de tre kugleudtagelser er der to muligheder: Der udtages en rød kugle, eller der udtages en kugle som ikke er rød. At der udtages en rød kugle svarer til de grenstykker i træet der er afmærket med A. På disse grene finder du tallet 0.40, det er sandsynligheden for at der udtages en rød kugle.

Alt i alt indeholder træet 8 grene. Den øverste gren i træet svarer til at der udtages tre røde kugler, grenen indeholder jo tre A'er. Og der er ikke andre grene i træet som indeholder tre A'er. Vi ser at sandsynligheden for denne hændelse er 0.0640.

Denne sandsynlighed kommer vi frem til ved at multiplicere de tre sandsynligheder på grenen:

$$0,40 \cdot 0,40 \cdot 0,40 = 0,0640$$

Vi kan nu finde de grene i træet som svarer til at der udtages netop én rød kugle. Det er de grene som indeholder ét A og to B'er: gren nr. 4, nr. 6 og nr. 7. Hver af disse grene har en sandsynlighed på 0.1440. Den samlede sandsynlighed på de tre grene er da 3x0.1440 = 0.4320. Der er derfor ca. 43% chance for at stikprøven vil indeholde én rød kugle.

Vi ser herefter på de grene der svarer til to røde kugler: gren nr. 2, nr. 3 og nr. 5. Hver af disse tre grene har en sandsynlighed på 0.0960, i alt 0.2880. Der er altså ca. 29% chance for at stikprøven vil indeholde to røde kugler.

En lille oversigt

Vi kan ud fra træet opstille en tabel over de muligheder der foreligger:

Chancen for 3 røde kugler (gren 1):	0.0640
Chancen for 2 røde kugler (gren 2, 3 og 5):	0.2880
Chancen for 1 rød kugle (gren 4, 6 og 7):	0.4320
Chancen for 0 røde kugler (gren 8):	0.2160

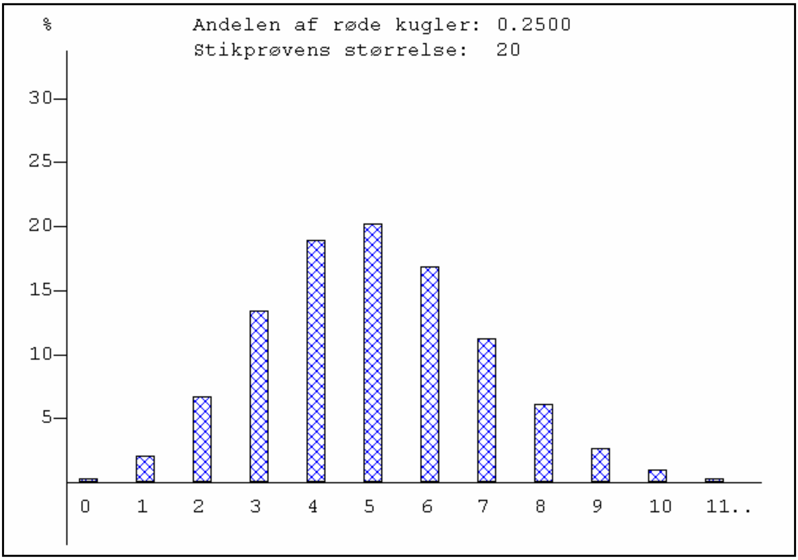
På denne måde kan de teoretiske chancer udregnes i situationer med stikprøveudtagelser. Hvis der udtages mange kugler, vil chancetræerne dog blive ganske uoverskuelige, så da må man nøjes med at tænke sig til dem.

Det er sådanne chancetræer der ligger bag programmet INSTAT som kan udregne chancer vedrørende antallet af røde kugler ved stikprøveudtagelser. Gå ind i INSTAT og vælg BIN-Model. Vælg derefter "Stikprøvestørrelse kendt"

Først skal du nu fortælle programmet hvad den kendte andel af røde kugler i æsken er. Vi kan fx vælge at se på en BIN-Model hvor andelen af røde kugler i æsken er 25%. Du indtaster derfor: 0.25.

Derefter spørger programmet om stikprøvens størrelse, dvs. hvor mange kugler der skal udtages. Du indtaster her: 20. Vi ser altså på en stikprøve der indeholder 20 kugler.

På skærmen kommer nu et grafisk billede af den valgte BIN-Model. Pindediagrammet giver en oversigt over hvad chancerne er for de forskellige antal af røde kugler i stikprøven på 20 kugler. Vi kan af diagrammet de at den største chance foreligger ved 5 røde kugler.



På skærmen kommer også følgende beregninger:

Midterområde	**Sandsynlighed**
4 .. 6	50%
3 .. 7	75%
2 .. 8	90%
2 .. 9	95%
1 .. 10	99%

Denne tabel giver os nogle nyttige oplysninger om den forelagte kugleudtagelse. Den fortæller at der er 50% chance for at antallet af røde kugler i stikprøven bliver fra 4 til 6, altså at antallet bliver 4, 5 eller 6 røde kugler.

Tabellen fortæller også at der er 90% chance for at antallet af røde kugler i stikprøven bliver fra 2 til 8. Der er altså 90% chance for at få fra 2 til 8 gevinster i de 20 spil.

Disse udregninger af gevinstområder for sandsynlighederne 50%, 75%, 90%, 95% og 99% foretages helt automatisk af programmet. Ofte er det nemlig den slags oplysninger der er mest behov for i forbindelse med en stikprøveudtagelse, så derfor har vi ladet programmet udføre disse beregninger hver gang.

Du kan imidlertid selv bestemme hvad programmet yderligere skal beregne. Du kan indtaste et område og få beregnet sandsynligheden for dette område. Benyt feltet "Indtast område".

Når du fx indtaster området **3..8** vil programmet beregne de tilhørende sandsynligheder for området fra 3 til 8, hvor både 3 og 8 regnes med til området.

Du kan fx også være interesseret i at kende sandsynligheden for at der bliver udtaget mere end 5 røde kugler. Det betyder at du vil lade programmet beregne sandsynligheden for at antallet af røde kugler bliver fra 6 og opefter (dvs. op til 20).

Du indtaster nu 6 efterfulgt af to prikker (punktummer):

6..

og programmet udskriver at sandsynligheden er 38.28%. Der er altså ca. 38% chance for at få mere end 5 røde kugler ved stikprøveudtagelsen .

Vi kan forestille os at stikprøveudtagelsen er et billede af et spil hvor der er 25% chance for at vinde. Du spiller 20 gange i spillet, det svarer til at du 20 gange udtager en kugle fra æsken. Når du udtager en rød kugle, svarer det til at du vinder.

Vi har lige set at der er ca. 38% chance for at du vil vinde mere end 5 gange i løbet af 20 spil.

Du kan nu også lade programmet beregne sandsynligheden for at du højst får 2 gevinster. Du indtaster to prikker efterfulgt af 2:

..2

og programmet svarer: 9.13%. Risikoen for at få op til 2 gevinster, dvs. 0 gevinster, 1 gevinst eller 2 gevinster; er altså ikke så stor, den er under 10%.

Hvis du vil se hvad chancen for at få lige præcis 5 gevinster er, indtaster du blot: **5**

Programmet svarer her: 20.23%. Der er altså en pæn chance for at de 20 spil lige præcis giver dig 5 gevinster.

Betegnelsen **BIN** står for **binomialfordeling.** Det er det matematiske navn for den sandsynlighedsfordeling der ligger bag **stikprøveudtagelse med tilbagelægning.**

Prøv selv
Beregn sandsynligheden for at de 20 spil giver

(1) 4 gevinster (2) 6 gevinster

(3) mere end 6 gevinster (4) højst 6 gevinster

Beregning af Midterområder
Ved anvendelse af programmet kan vi få en forklaring på hvordan programmet beregner de midterområder der udregnes automatisk.

Vi ser til eksempel på 90%-midterområdet. Det var beregnet til at være område 2..8. Det betyder følgende: I dette gevinstområde ligger **mindst** 90% af chancen for gevinst. Uden for det er der altså **højst** 10% chance for gevinst.

Men de 10% chance for gevinst uden for det angivne område skal så vidt det er muligt være "ligeligt fordelt": Der må højst være 5% chance for gevinst til venstre for området, og der må højst være 5% chance for gevinst til højre for området. Vi har altså følgende chancefordeling:

0_________2___________8____.....___________20

Højst 5% Mindst 90% Højst 5%

Når disse to betingelser opfyldes, vil der i midterområdet 2..8 være mindst 90% chance for gevinst.

Ved brug af "Indtast område" får vi:

Område 0..1 Sandsynlighed: 2.43%

Område 9..20 Sandsynlighed: 4.09%

Altså: Ikke over 5% i noget af de to områder. Området 2..8 indeholder dermed mindst 90% sandsynlighed.

Vi kunne ikke have udvidet det venstre område til 0..2. Her får vi nemlig:

Område 0..2 Sandsynlighed: 9.13%,

altså mere end 5%.

Heller ikke det højre område kunne gøres større. For området 8..20 får vi:

Område 8..20 Sandsynlighed: 10.18%,

altså igen mere end de tilladte 5%.

Når programmet derfor angiver 90% midterområdet til 2..8, så betyder det altså at der er **højst** 10% chance for at antallet af gevinster falder uden for dette område, nemlig højst 5% chance for at der bliver færre end 2 gevinster og højst 5% chance for at der bliver mere end 8 gevinster.

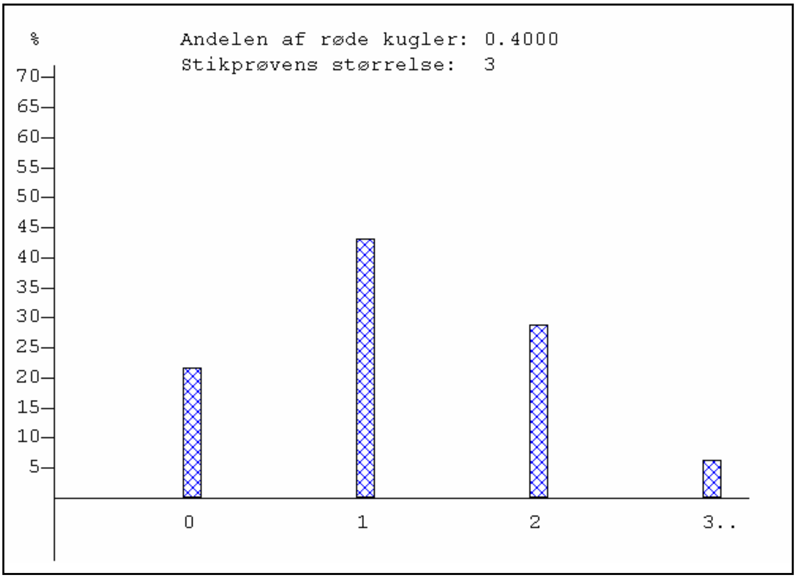
Hvis vi lader programmet udregne den nøjagtige værdi for chancen for at få et resultat i området 2..8, bliver svaret: **93.48%**. Kontroller at det er rigtigt.

Vi er nu også i stand til at fortolke de andre gevinstområder. At 95% midterområdet er 2..9, betyder at der er højst 2.5% chance for at få færre end 2 gevinster, og højst 2.5% chance for at få mere end 9 gevinster. Der er altså mindst 95% chance for at få fra 2 til 9 gevinster.

Prøv selv

1. Kontroller ved hjælp af programmet at det er korrekt at angive 50%-midterområdet til at være området 4..6. Kontroller også at 99%-midterområdet er korrekt angivet til at være området 1..10.

2. Undersøg ved hjælp af programmet de sandsynligheder vi beregnede i chancetræet i dette afsnits indledning. Du skal altså se på en stikprøveudtagelse hvor andelen af røde kugler er 0,4, og hvor stikprøvestørrelsen er 3. - Lad programmet beregne sandsynlighederne for 3, 2, 1 og 0 røde kugler.



De forventelige resultater

Stikprøvernes midterområder er ofte af særlig interesse. I mange tilfælde anvender man 90%-midterområdet som det der giver et godt svar på hvilke resultater der kan forventes i det foreliggende eksperiment.

I vort spil-eksempel hvor andelen af røde kugler er 0.25, og hvor vi udtager 20 kugler, består 90%-midterområdet af resultaterne 2..8. Det vil sige at i 9 ud af 10 tilfælde kan vi – i det lange løb - forvente at antallet af gevinster i 20 spil vil være fra 2 til 8. Dette område vil vi kalde **90%-dataområdet**. Når vi udfører eksperimentet, får vi nogle resultater, **data**, og chancen for data i området 2..8 er 90%.Vi vil derfor ikke blive overrasket hvis vi får et af disse gevinstantal når vi udfører 20 spil. Antal fra 2 til 8 hører til "det forventelige".

Får vi derimod et antal gevinster uden for dette område, vil der være indtruffet en hændelse som har en sandsynlighed på højst 10%. Vi vil da være på vagt over for spillet, og vi vil måske tage anledning til at undersøge om det er udført helt efter bogen.

Ofte vil man bruge 90%-dataområdet således: Et resultat i dette område hører til de forventelige resultater, dem der almindeligvis vil forekomme. - Et resultat som ikke ligger i 90%-dataområdet, hører ikke til de forventelige, og det kan derfor give anledning til en nærmere undersøgelse af chanceeksperimentet.

Prøv selv

Find ved hjælp af BIN-Model 90%-dataområdet ved stikprøveudtagelser fra en æske med en andel af røde kugler på 0,4. Prøv med stikprøvestørrelser på: 250, 1000 og 4000.

Observationer

50%-midterområdet vil vi kalde BIN-Modellens **kerneområde**.

Når vi udfører den stikprøveudtagelse der beskrives ved BIN-Modellen, foretager vi **en observation**.

Når vi foretager en observation, er der ca. 50% chance for at den falder i kerneområdet og ca. 50% chance for at den falder uden for området.

Vi vender tilbage til spillemaskinen hvor vi udfører 20 spil og hvor gevinstchancen i hvert spil er 0.25. Lad os se på en observation der falder uden for kerneområdet. Det kan fx forekomme at stikprøven indeholder 9 røde kugler. Vor observation er altså tallet 9.

Af oversigten over midterområder kan vi nu se hvor den forelagte observation er placeret:

Midterområde	**Sandsynlighed**
4 .. 6	50 %
3 .. 7	75 %
2 .. 8	90 %
2 .. 9	95 %
1 .. 10	99 %

Vi ser at observationen 9 er med i 95%-midterområdet, og dermed jo også i 99%-området. Men 9 er ikke med i 90%-området.

Det betyder at vi med observationen 9 har fået et stikprøveresultat som ligger uden for 90%-midterområdet. Der er dermed højst 10% chance for at vi ved stikprøveudtagelsen vil få en observation der er så "skæv" som observationen 9.

Vi kan lade computeren beregne chancen lidt nærmere:

Hvis vi indtaster 9 i feltet **"Indtast obs."**, så følger udskriften:

Observation: 9

Største midterområde uden 9:

91.8%   2..8

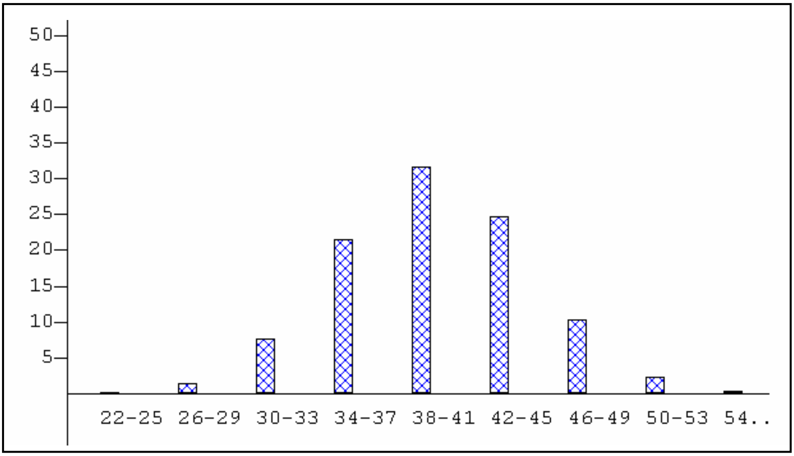
Her har vi det største midterområde som ikke indeholder observationen 9. Området har en chanceprocent på 91.8%. Det betyder at vi kan sige at der i den forelagte BIN-Model er højst 8.2% chance, nemlig 100% – 91.8%, for at få en observation så skæv som 9.

Kontroller ved indtastning af 91.9 i feltet **"Angiv procent for midterområde"** at det giver et midterområde som indeholder 9. Det største midterområde uden 9 er således det angivne område 2..8 med procenten 91.8.

Vi kan omregne observationen 9 til at svare til at andelen af røde kugler i stikprøven er 9/20 = 0.45. Men i æsken var der kun en andel af røde kugler på 0.25. Vi har altså fået en observation som giver en andel af røde kugler der afviger væsentlig fra BIN-Modellens tal.

Når vi får en observation som er meget skæv i forhold til den benyttede BIN-Model, kan vi reagere på to måder: Vi kan trøste os med at der er indtruffet en hændelse som har en lille sandsynlighed, og noget sådant kan jo ske. Eller vi kan tvivle på at vi har benyttet den rigtige model, dvs. vi sætter spørgsmålstegn ved den benyttede andel af røde kugler i BIN-Modellen.

Lad os se på en BIN-Model hvor vi udtager 100 kugler fra en æske hvor (vi tror) andelen af røde kugler er 0.40. Her er et billede af den anvendte BIN-Model:



Vi indtaster nu observationen 29. Programmet fortæller at det største midterområde som ikke indeholder 29 er:

97.0% 30 .. 51

Chancen for at få en observation der er så skæv som 29 (eller endnu skævere), er altså højst 3%.

Som sagt kan sjældne hændelser indtræffe, også dem der kun har en chance på 3%. Vi kan undersøge sagen nærmere ved nye observationer. Hvis de udviser samme skævhed, vil vi nok tage vores BIN-Model op til revision.

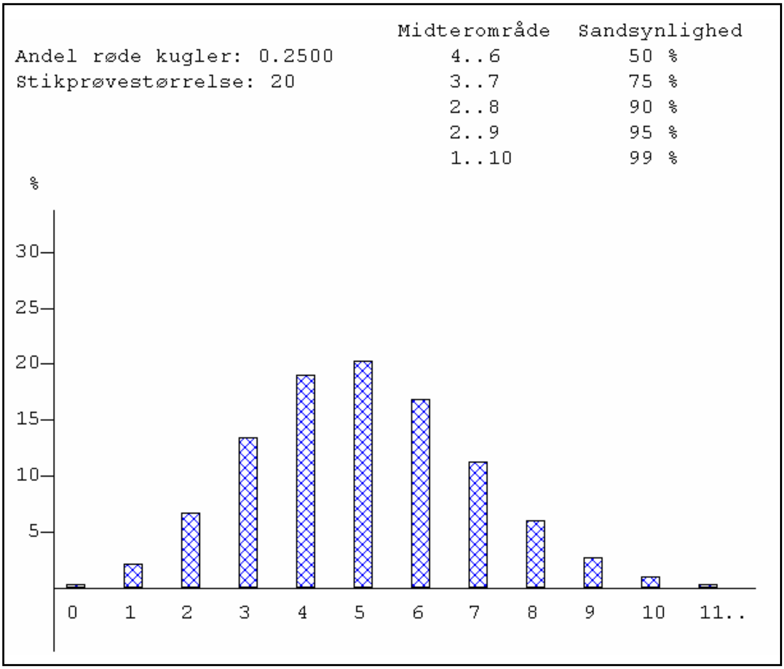
I den valgte BIN-Model vil vi betegne observationen 29 som **statistisk signifikant på 3%-nivauet.** Ordet "signifikant" kan oversættes til "af betydning" eller "påfaldende". At 29 er statistisk signifikant på 3%-niveauet betyder at der højst er en chance på 3% for at få en observation så skæv som 29 ved den valgte BIN-Model. Vi vil også kort sige at observationen 29 har et **skævniveau på 3%.**

*

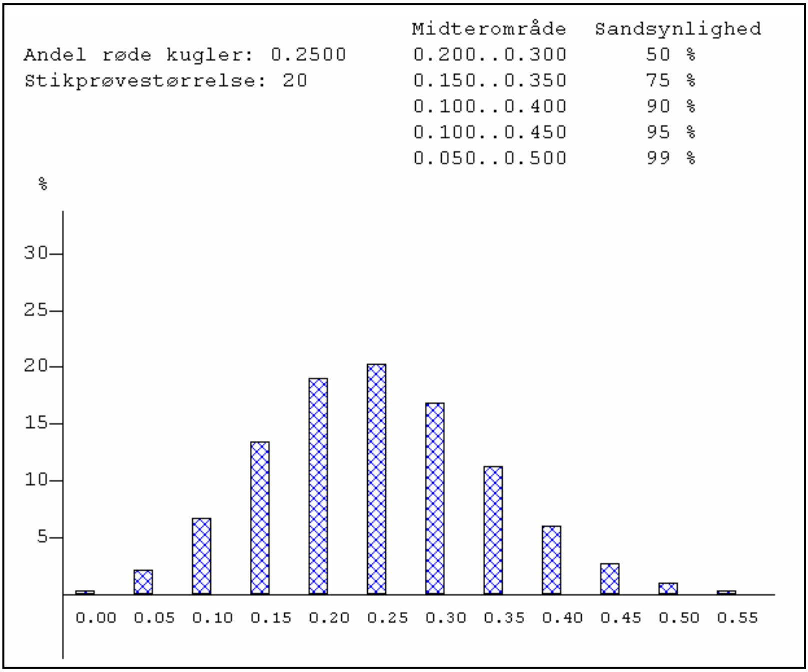
I følgende afsnit ser vi nærmere på metoder til at vurdere om en anvendt BIN-Model kan accepteres eller om den må udskiftes med en bedre model.

Antal og Andel

På skærmen vil du finde en knap hvor du kan skifte mellem antal og andel. Som standard viser programmet midterområderne angivet ved hjælp af antal røde kugler i stikprøven som til eksempel ved spillemaskinen:



Hvis du klikker på Andele-knappen får du følgende angivelse af midterområderne:



Også på den grafiske figur skiftes fra a**ntal** af røde kugler i stikprøven til **andel** af røde kugler i stikprøven.

Kontroller at midterområderne er korrekt omregnet fra antal af røde kugler til andel af røde kugler.

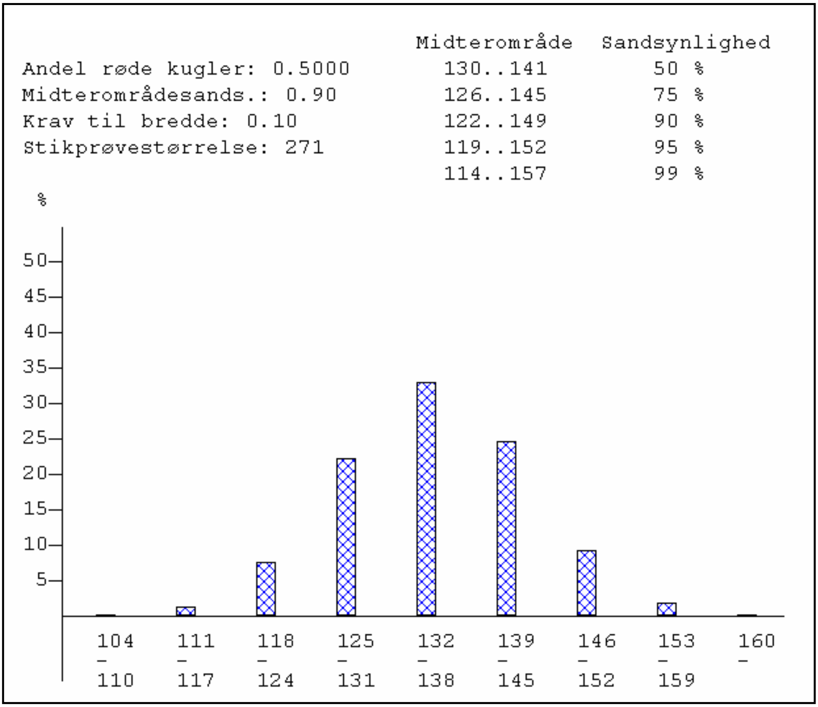
3. Stikprøvestørrelse ukendt

Vi ser nu på den anden mulighed i BIN-Model: Stikprøvestørrelsen er ikke kendt.

Vi ser på et eksempel hvor andelen af røde kugler i æsken er 0.5. Vi vil nu lade programmet beregne hvor stor en stikprøve vi skal tage for at opnå et 90%-midterområde som har en bredde på højst 0.10. Vi indtaster de tre oplysninger:

Andel røde kugler i æsken: 0.5
Midterområdesandsynlighed: 0.90
Bredde af midterområde: 0.10

På skærmen får vi nu følgende:



Programmet har beregnet den nødvendige stikprøvestørrelse til **271**, og det tilhørende 90%-midterområde er:

122 ..149.

Omregner vi området til andele af røde kugler får vi:

0.450 .. 0.550.

Altså et midterområde med en bredde på 0.100.

Vi kan oversætte eksemplet til møntkast. Hvis vi påtænker at udføre 271 møntkast, har vi en sandsynlighed på 90% for at antallet af Kronekast kommer til at ligge i området 122..149, eller med andre ord: at andelen af Kronekast vil ligge i intervallet 0.45 .. 0.55.

4. Opgaver til BIN- Model

I opgaverne i dette afsnit skal du benytte BIN-Model til at beregne teoretiske chancer. Nogle af opgaverne vil være formuleret som stikprøveudtagelser, men de fleste vil være beskrevet som chancesituationer der tilsyneladende ikke vedrører stikprøver.

Opgaverne skal give dig eksempler på de mange anvendelser der kan gøres af den model vi arbejder med, altså en model der er forelagt som udtagelse af kugler fra en æske med en kendt andel af røde kugler. Ud fra modellen kan vi beregne chancer, dvs. vi kan foretage forudsigelser om de resultater, eller de data, som kan forekomme ved stikprøveudtagelserne.

Vi går altså her ud fra den kendte model, dvs. den kendte æske med en **kendt andel** af røde kugler, og foretager beregninger om de **ukendte data**, nemlig stikprøveudtagelserne.

Inddata til programmet

Programmet BIN-Model kan arbejde med andele som kan angives som decimaltal med fire decimaler, dvs. fra 0.0001 op til 0.9999. Men du kan også indtaste brøker. Hvis der fx er tale om kast med en terning hvor chancen for sekser er 1/6, så kan du direkte indtaste denne brøk. Programmet regner den om til et decimaltal til brug for de fortsatte beregninger. Du kan dog ikke indtaste brøker der er mindre end 0.0001 eller større end 0.9999.

Programmet kan arbejde med stikprøvestørrelser fra 2 op til 50 000.

Et vigtigt spørgsmål: Er modellen brugbar?

Hvis kuglemodellen skal kunne anvendes i en opgavesituation, kræves det at der er tale om en række chanceeksperimenter hvor der ved hvert eksperiment er samme chance for gevinst, dvs. for at udtage en rød kugle. Der må ikke være nogen påvirkning fra den ene udtagelse til den næste, og der må ikke være tale om "afsmitning" fra en udtagelse til de andre. "Kugleudtagelserne skal være uafhængige af hinanden".

I et spil må man altså ikke blive dygtigere, og dermed få en bedre vinderchance efterhånden som spillet skrider frem. Der skal til stadighed være den samme chance for gevinst. Ej heller må man blive træt eller sløset efterhånden som spillene afvikles, således at ens vinderchance efterhånden bliver ringere end den var i starten.

Prøv at se kritisk på opgaverne, og overvej hver gang du løser en opgave om der foreligger en opgavesituation hvor der kan være tvivl om brugbarheden af kuglemodellen. Diskuter sådanne eksempler med dine kammerater eller med din lærer.

1. Ti elever spiller

Ti elever fra Sannes klasse udfører hver et spil på en spilleautomat hvor chancen for gevinst i hvert spil er 40%.
Undersøg hvad chancen er for at mindst fem af de ti elever får gevinst.

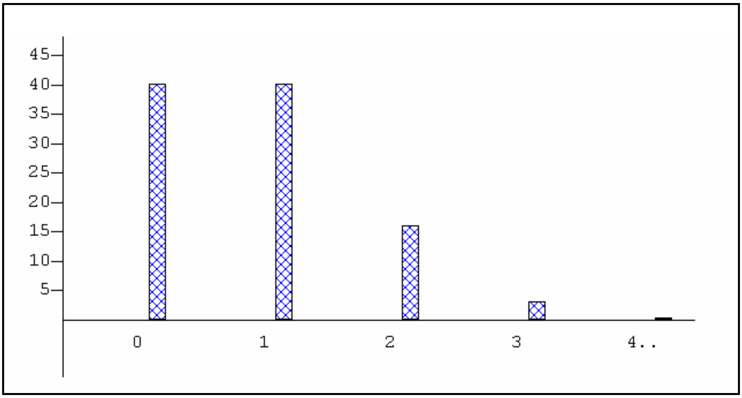
Og undersøg hvad risikoen er for at højst to af de ti elever får gevinst. – Hvad er chancen for at 4 elever får gevinst?

2. Chancen for seksere

I et spil kastes 5 terninger. Beregn følgende chancer (indtast 1/6 som brøk):

1. Chancen for at få mindst én sekser

2. Chancen for at få mindst to seksere.



3. Kronekast

I et spil kastes 10 mønter. Hvad er chancen for at få:

1. Lige præcis 5 kronekast
2. Mere end 5 kronekast.

4. Hvad er bedst: To store eller fem små chancer?

Undersøg hvor der er størst chance for at få mindst én gevinst:

(a) Ved to spil på en spillemaskine hvor chancen for en gevinst i hvert spil er 25%

(b) Ved fem spil på en spillemaskine hvor chancen for gevinst i hvert spil er 10%.

5. Hvor er chancen størst?

Beregn følgende chancer:

1. Chancen for at få mindst 3 seksere i 10 terningkast
2. Chancen for at få mindst 6 seksere i 20 terningkast
3. Chancen for at få mindst 18 seksere i 60 terningkast.

6. En prøve

Du deltager i en prøve hvor du ved hver enkelt opgave har 40% chance for at løse opgaven.

Beregn chancen for at du løser mindst halvdelen af opgaverne:

1. I en prøve med 10 opgaver
2. I en prøve med 20 opgaver.

7. Kugleudtagelser

I en pose er 40% af kuglerne røde. Der udtages 10 gange en kugle med tilbagelægning efter hver udtagelse. Beregn ved hjælp af BIN-Model:

1. Chancen for at der udtages præcis 4 røde kugler.
2. Chancen for at der udtages højst 5 røde kugler.
3. Chancen for at der udtages mindst 6 røde kugler.
4. Chancen for at der udtages højst 1 rød kugle.

8. Mesterskyttens chancer

En mesterskytte påstår at have en træfsikkerhed på 80%.

Undersøg ved hjælp af BIN-Model hvad sandsynligheden er for at han får mindst 8 træffere i 10 skud og hvad sandsynligheden er for at han får mindst 16 træffere i 20 skud?

Og hvad er sandsynligheden for at han får mindst 80 træffere i 100 skud?

9. Et spil med en mønt

Sanne og Malene udfører følgende spil: En mønt kastes. Ved Krone vinder Sanne, ved Plat vinder Malene. Den der taber, betaler 1 kr. til vinderen. Der spilles i alt 20 spil. Find ved hjælp af BIN-Model følgende chancer:

1. Chancen for at Sanne og Malene hver vinder 10 spil

2. Chancen for at Sanne får et overskud på mindst 4 kroner i de 20 spil

3. Chancen for at en af de to piger får et overskud på mindst 4 kroner.

10. En skæv mønt

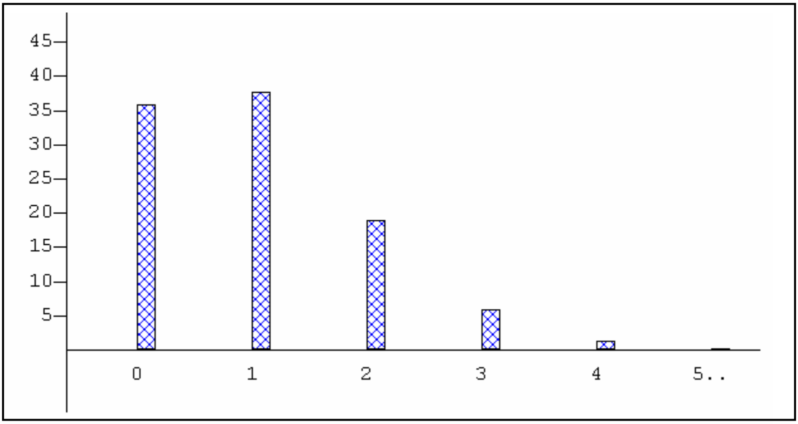
Beregn chancerne i opgave 9 når spillet udføres med en skæv mønt hvor chancen for Krone kun er 40%.

11. Farveblindhed

Man regner med at 5% af drengene i Danmark er farveblinde. Find ved hjælp af BIN-Model sandsynligheden for at der blandt 10 drenge i en klasse er mindst én der er farveblind?

Og hvad er sandsynligheden for at der blandt 20 drenge i en klasse ikke findes nogen der er farveblind?

Hvad er sandsynligheden for at der blandt 200 drenge på en skole kun findes 5 der er farveblinde?



12. En stikprøvekontrol

Ved produktionen af en vare har det vist sig at 20% af de producerede enheder er defekte. Find ved hjælp af BIN-Model sandsynligheden for følgende:

1. En stikprøve på 10 enheder indeholder højst 1 defekt.

2. En stikprøve på 20 enheder indeholder højst 2 defekte.

13. Fravær i skolen

En skolestatistik viser at i gennemsnit er ca. 10% af eleverne fraværende fra skolen. Find sandsynligheden for at der på en tilfældig dag er mindst 4 fraværende i en klasse på 20.

14. Malene som spåkone

Malene påstår at hun kan forudsige resultatet af møntkast og terningkast. Du prøver hende med 20 møntkast. Find chancen for at hun ved ren gætning får mindst 12 rigtige forudsigelser?

Og hvad er sandsynligheden for at hun vil få mindst 5 rigtige forudsigelser i 20 terningkast?

15. Stikprøver med slagside

I en pose er 40% af kuglerne røde og de øvrige kugler er hvide. Fra posen udtages 5 kugler ved stikprøveudtagelse med tilbagelægning. Beregn sandsynligheden for at stikprøven indeholder flest kugler "af den forkerte farve", altså flere røde end hvide kugler.

Foretag også undersøgelsen for en stikprøve på 15 kugler.

16. Kvalitetskontrol

Et vareparti kontrolleres ved stikprøveudtagelse af 20 enheder. Hvis stikprøven indeholder højst 3 defekte enheder godkendes varepartiet.

Hvilken chance for at blive godkendt har et vareparti hvori 10% af enhederne er defekte?

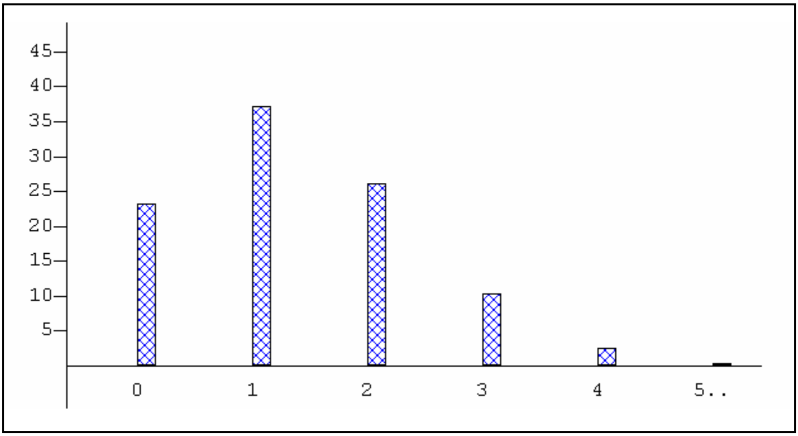
Hvilken chance for at blive godkendt har et vareparti hvori 20% af enhederne er defekte?

17. En opgave fra 1711

I en afhandling fra 1711 finder man denne opgave:

> Hvad er chancen for at få mindst to seksere i 8 kast med en terning?

Løs den ved hjælp af BIN-Model.



18. Blomsterløg

Et supermarked som sælger blomsterløg i pakninger med 10 og 20 styk, anfører at løgenes spire-evne er 90%, dvs. der er 90% chance for at et tilfældigt udvalgt løg vil spire.

Find sandsynligheden for at alle løg i en 10 styks pakning vil spire, og find sandsynligheden for at alle løg i en 20 styks pakning vil spire.

19. Hjemmecomputer

En statistik viser at 80% af hjemmene i en kommune har hjemmecomputer. Hvad er sandsynligheden for at alle elever i en tilfældig valgt klasse på 20 har hjemmecomputer?

20. Multiple-choice prøver

I en flervalgsprøve er der ved hver opgave tre svarmuligheder hvoraf én er den rigtige. I prøven indgår 20 opgaver.

Hvilken chance har du for at bestå prøven (dvs. du besvarer mindst halvdelen af opgaverne rigtigt) når du bestemmer dine svar ved hjælp af lodtrækning med lige chance for hver af de givne svarmuligheder?

Hvad er chancen for at du består når der kun er to svarmuligheder ved hver opgave?

21. Ingen dvd-afspiller ledig?

En undersøgelse på en skole viser at der i gennemsnit er ønske om at anvende dvd-afspiller i 25% af timerne. Skolen har 20 klasser.

Hvis skolen kun har 5 dvd-afspillere, hvad er da risikoen for at et ønske om dvd-afspiller ikke kan opfyldes i næste time?

Hvor mange dvd-afspillere bør skolen have når der kun må være en risiko på ca. 10% for at et ønske om dvd-afspiller ikke kan opfyldes i en tilfældig time?

22. Er terningen falsk?

Du laver din egen "kvalitetskontrol" af en terning på følgende måde: Du udfører en kasteserie. Hvis terningen ikke i 10 kast har givet mindst én sekser, kasserer du den.

Hvad er risikoen for at du ved denne test kommer til at kassere en terning som i virkeligheden er god nok?

23. Hvilket spil giver størst chance?

Sanne får tilbudt at deltage i et spil hvor der kastes med mønter. Man får gevinst hvis mindst 60% af mønterne viser Krone.

Deltagerne kan selv vælge om de vil kaste 5 mønter, 10 mønter, eller 25 mønter. I kast med 5 mønter er der altså gevinst ved mindst 3 kronekast, i kast med 10 mønter er der gevinst ved mindst 6 kronekast, og i kast med 25 mønter er der gevinst ved mindst 15 kronekast.

Hvilken af de tre muligheder bør Sanne vælge?

24. En klassisk opgave

En af historiens første lærebøger i chancelære blev udsendt i 1657 med hollænderen Christiaan Huygens som forfatter.
I denne bog finder vi følgende opgave:

> Hvor mange kast skal der udføres med en terning for at der er 50% chance for at der forekommer to seksere i kasteserien?

Løs opgaven ved hjælp af BIN-Model. Prøv dig frem indtil du finder det rigtige kasteantal. (50% skal forstås som "mindst 50%", og 2 seksere skal forstås som "mindst 2 seksere".)

25. Chancen for sekser

Hvor mange kast skal du udføre med en terning for at der er 95% chance for at du får en sekser mindst én gang?

Prøv dig frem ved hjælp af BIN-Model.

26. En lille chance

Sanne deltager i et spil hvor der kun er 1% chance for gevinst. Hvor mange spil skal hun udføre for at have 50% chance for at få gevinst mindst én gang?

Find den rigtige stikprøvestørrelse. Prøv indtil du finder en som giver en chance på 50% for mindst én gevinst.

I jagten på stikprøvestørrelsen kan du nøjes med at se på antal der ender på 0: 10, 20, 30, osv.

27. Sanne vil have flere gevinster

Sanne ønsker at vide hvor mange spil hun skal udføre med spillet i opgave 26 for at have 50% chance for at få mindst 2 gevinster.

Undersøg spørgsmålet ved hjælp af BIN-Model. Gå frem på samme måde som i opgave 26.

28. Den store gevinst

I et spil er chancen for at vinde den store gevinst kun 1 ud af 1000. - Hvor mange gange skal du mon deltage i spillet for at have en chance på 50% for mindst én gang at vinde den store gevinst?

29. Hvor mange forbindelser?

I et firma taler medarbejderne i gennemsnit i telefon i 20% af arbejdstiden. Hvis der er 10 ansatte, hvad er da risikoen for at der på et tilfældigt valgt tidspunkt er behov for flere forbindelser end de fire der er til rådighed?

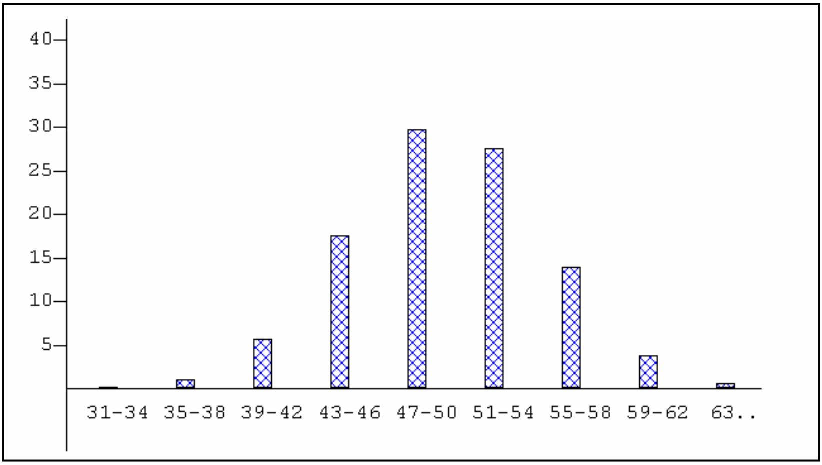
Firmaet udvides til 20 mand (som stadig taler i telefon i 20% af arbejdstiden). Hvor mange forbindelser bør man nu have til rådighed for at forholdene ikke skal være nævneværdigt dårligere end før?

30. Hvor mange møntkast skal udføres?

En lærer ønsker at lade en skoleklasse udføre så mange møntkast at man opnår et 90%-midterområde med en bredde på højst 0.06. Hvor mange kast bør han sætte klassen til at udføre?

Prøv også med et 95%-midterområde med en bredde på 0.06.

Og prøv med et 99%-midterområde med en bredde på 0.06.



31. Det bedste antal spil

Sanne udfordrer Malene: Her er en mønt hvor chancen for at få Krone kun er 45%. Du skal kaste mønten et lige antal gange, og du vinder hvis du får flere Krone end Plat. Du kan selv vælge hvor mange kast du ønsker at udføre, men du skal beslutte dig for et bestemt antal kast før du begynder at kaste.

Find ved hjælp af BIN-Model (prøv dig frem) hvor mange kast Malene skal vælge at udføre for at hun kan få så stor en vinderchance i spillet som muligt.

Hvor mange kast skal Malene vælge at udføre hvis der skal kastes et ulige antal gange?

32. Hvor mange skal udvælges?

Farveblindhed forekommer hos drenge med en sandsynlighed på 5%. Hvor mange drenge skal du udvælge tilfældigt for at du kan være 95% sikker på at finde mindst én som er farveblind?

Og hvor mange drenge skal du udvælge hvis du vil 95% sikker på at finde mindst to der er farveblinde?

33. Defekte skruer

Du køber skruer i en billig kvalitet hvor almindeligvis 20% må kasseres. Hvor mange skruer skal du købe for at være 95% sikker på at have mindst 25 brugbare skruer?

34. Hvor mange pladser tør man sælge?

Et flyselskab sælger billetter til et fly som kan tage 200 passagerer. Statistikken viser at der kun er 75% chance for at en solgt billet bliver benyttet.

Hvor mange pladser tør selskabet sælge når der højst må være 10% risiko for at man ikke kan tage alle de fremmødte passagerer med?

Hvor mange pladser tør man sælge når risikoen ikke må overstige 5%? Og hvis den ikke må overstige 1%?

35. To kontroleftersyn

I en produktion bliver hver produceret enhed udsat for to uafhængige kontroleftersyn. Ved hvert eftersyn er der 10% risiko for at en defekt enhed ikke vil blive afsløret.

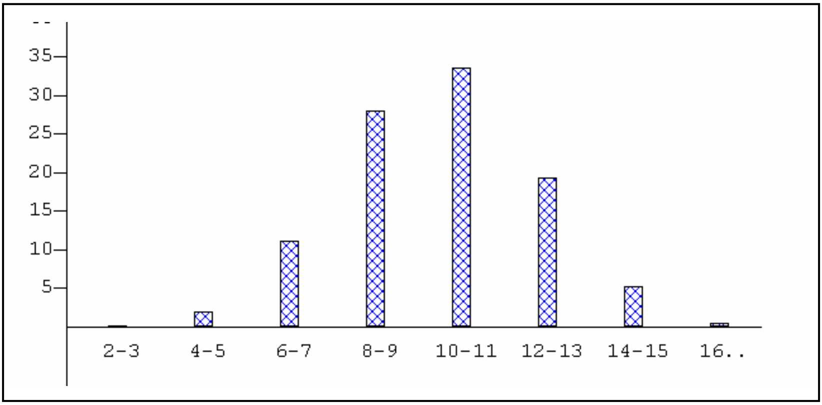
Hvad er chancen for at 10 defekte enheder alle bliver fundet ved kontroleftersynene?

Hvad er chancen for at 20 defekte enheder alle bliver fundet?

Hvad er chancen for at 50 defekte enheder alle bliver fundet?

36. Er mønten god nok?

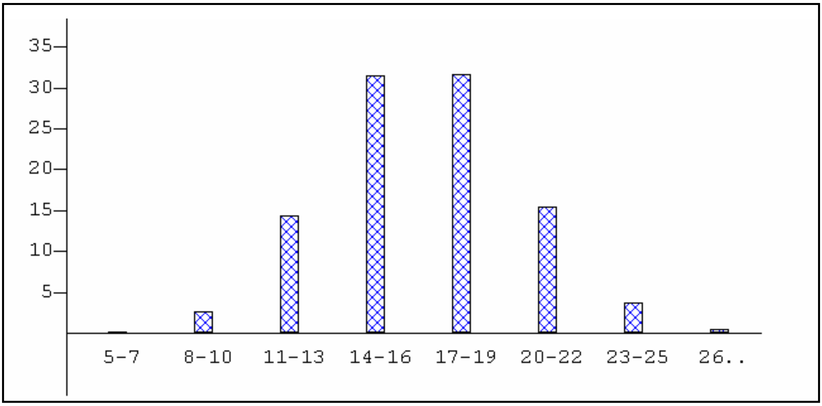
Sanne har mistillid til en mønt. Hun tror ikke at den har en sandsynlighed på 0,5 for at vise Krone. Hun tester den ved en kasteserie på 20 kast. Hvis antallet af Krone kommer til at ligge i 90%-midterområdet vil hun godkende mønten, hvis antallet af Krone ligger uden for 90%-området vil hun ikke godkende den.
Hun kaster mønten 20 gange og får kun 6 gange Krone.
Hvad bliver hendes beslutning?



37. Er terningen falsk?

Malene har mistanke om at der er manipuleret med en af hendes terninger, således at der ikke er en sandsynlighed på 1/3 for at den viser 1 eller 6. Hun vil afgøre sagen på grundlag af en kasteserie på 50 kast. Hvis det samlede antal af 1'ere og 6'ere i de 50 kast falder inden for 90%-området vil hun godtage terningen, ellers vil hun kassere den.

Hun kaster terningen 50 gange og opnår 22 gange at den viser enten 1 eller 6. Hvad bliver hendes beslutning?



38. Er der pillet ved maskinen?

En spillemaskine påstås at have en gevinstsandsynlighed på 30%. I 50 spil kommer der kun 9 gevinster. Afgør ved hjælp af et passende midterområde om dette giver grundlag for mistillid til maskinen.

39. Mystik på hospitalet?

På et dansk hospital fødes der i løbet af en måned 58 drenge, men kun 37 piger. Undersøg ved hjælp af BIN-Model om dette kan være i overensstemmelse med den danske statistik over fordelingen af drenge og piger som fortæller at 51,4% af de børn der fødes i Danmark er drenge.

40. Terningkast i klassen

I Malenes klasse udfører hver elev en serie på 20 terningkast og tæller op hvor mange kast der blev seksere.

(1) Hvad er chancen for at Malene får mindst 20% seksere i sine kast.

(2) Malene er med i en gruppe på fem elever. De samler alle deres resultater. Hvad er chancen for at de har fået mindst 20% seksere i deres 100 kast?

(3) Alle 25 elever i Malenes klasse samler deres resultater. Hvad er chancen for at klassen har fået mindst 20% seksere i deres 500 kast?

41. Lotto-trækninger

I Lotto udtrækkes hver uge 7 tal blandt tallene 1..36. Malenes lykketal er 17. Hun beslutter at holde øje med hvor tit det bliver udtrukket, så hun laver en statistik for 10 uger.

(1) Hvad er chancen for at 17 bliver udtrukket mere end 4 gange i løbet af de ti uger?

(2) Hvad er chancen for at 17 slet ikke bliver udtrukket i løbet af de ti uger?

(3) Sannes lykketal er ikke blevet udtrukket i 15 uger. Hvor stor er chancen for en sådan hændelse?

42. En terningtest

En spiller har fået konstrueret en ny terning til sine spilleturneringer. Han tester den nye terning ved at udføre 1200 kast. Hvis antallet af seksere ligger inden for området 190..210 godtager han terningen, ellers kasserer han den.

Udregn ved hjælp af BIN-Model risikoen for at han kan komme til at kassere en terning der i virkeligheden er god nok.

43. Straffespark

Straffespark-skytten på fodboldholdet påstår han har en målscorings-chance på 90%. I løbet af en turnering scorer han kun 15 mål i 23 straffespark. Undersøg ved hjælp af BIN-Model om dette kan være i overensstemmelse med hans påstand om de 90%.

44. Tre ens i Pokerterning

Chancen for at få "Tre ens" ved et kast med fem terninger er 15,4%. Beregn ved hjælp af BIN-Model:

(1) Chancen for mindst én gang "Tre ens" i en serie på 8 kast med fem terninger.

(2) Hvor mange kast skal der udføres for at der er en chance på 90% for at få "Tre ens" mindst én gang.

45. Fuldt hus i Pokerterning

I et kast med fem terninger er der 3,9% chance for at opnå "Fuldt hus", dvs. at tre terninger viser ét øjental, og de andre to terninger viser et andet øjental (eks: 66644).

Beregn ved hjælp af BIN-Model:

(1) Chancen for at få "Fuldt hus" i løbet af de første 10 kast med fem terninger.

(1) Chancen for at få "Fuldt hus" i de første 20 kast.

(2) Risikoen for ikke at få "Fuldt hus" i løbet af 50 kast.

46. Fire ens i Pokerterning

Chancen for at få "Fire ens" ved et kast med fem terninger er 1,93%. Beregn ved hjælp af BIN-Model:

(1) Chancen for mindst én gang at få "Fire ens" i en serie på 25 kast med 5 terninger.

(1) Hvor mange kast skal der udføres for at der er en chance på 90% for at få "Fire ens" mindst én gang.

47. Et klassisk problem

De Méré, fransk adelsmand og spiller, beskæftigede sig i 1650-erne med problemer vedrørende terningkast.

Hvilken af de to hændelser er mest sandsynlig:

(a) at få mindst én sekser i 4 kast med én terning

(b) at få mindst én dobbelt-sekser i 24 kast med to terninger.

48. Et atomkraftværk

På de amerikanske atomkraftværker blev en ulykke af en bestemt type vurderet til at ville indtræffe i gennemsnit én gang pr. 2000 drifts-år. Nu indtraf den første ulykke af denne type allerede efter forløbet af 400 drifts-år.

Undersøg ved hjælp af BIN-Model hvad chancen er for at der allerede forekommer en hændelse af denne art inden for de første 400 drifts-år.

49. Retsvæsenet i Aleatorien

I det lille land Aleatorien hvis vigtigste indtægtskilde er spil, har retsvæsenet indført brugen af kugleudtagelser i forbindelse med afgørelsen om en tiltalt skal kendes skyldig eller ikke-skyldig.

Når dommeren har fået forelagt alle sagens fakta, fastsætter han hvor overbevist han er om den tiltaltes skyld. Er hans overbevisning fx 60%, bliver der benyttet en æske hvori andelen af røde kugler er 60%. Fra æsken udtager de 12 nævninge hver én kugle ved udtagelse med tilbagelægning. Hvis der er flertal af røde kugler blandt de udtagne kugler bliver den tiltalte kendt skyldig, ellers bliver han frikendt.

(1) Undersøg ved hjælp af BIN-Model hvad chancen for frifindelse er i en situation hvor dommeren er 60% overbevist om tiltaltes skyld.

(2) Hvad er chancen for frifindelse når dommerens overbevisningsgrad er 50%?

(3) Hvad er chancen for frifindelse når dommerens overbevisningsgrad er 75%?

(4) Og hvad er chancen for frifindelse når dommerens overbevisningsgrad kun er 30%?

50. Loven i Aleatorien ændres

Aleatorien skærper kravet om flertal af røde kugler ved kendelsen "skyldig": Den tiltalte vil kun blive kendt skyldig hvis der blandt de 12 udtagne kugler er mindst 10 røde.

Undersøg ved hjælp af BIN-Model de fire situationer der er beskrevet i opgave 49. Hvad er nu chancerne for at den tiltalte frifindes?

51. Der skal spares på nævninge

Aleatorien skal spares på udgifterne til nævninge og sætter antallet ned fra 12 til 6. Den tiltalte vil nu blive kendt skyldig hvis der blandt udtagne 6 kugler er et flertal af røde kugler, altså mindst 4.

Undersøg hvilken betydning det har for de fire situationer der er beskrevet i opgave 49.

52. Et lykkehjul

Malene vil spille på et lykkehjul i Tivoli. Lykkehjulet har 24 numre, og Malene spiller i hvert spil på 5 af numrene.

Undersøg ved hjælp af BIN-Model hvad chancen er for at Malene vinder i løbet af de første 3 spil.

Og hvad er chancen for at hun vinder i løbet af de første 5 spil?

Hvad er risikoen for at hun ikke vinder i løbet af de første 10 spil?

53. Find den rigtige nøgle

Du har en nøglering med 7 nøgler. Kun en af nøglerne passer til låsen. Du udvælger tilfældigt en nøgle ad gangen og prøver indtil du får gevinst.

Hvad er chancen for at du kan klare dig med 3 forsøg hvis afprøvningen foregår "med tilbagelægning"?

54. Udvælg et primtal

I talområdet 1..100 er der 24 primtal. Du udvælger tal fra talområdet, ét ad gangen, indtil du får et primtal. Udvælgelsen foregår ved lodtrækning med tilbagelægning.

Hvad er chancen for at du kan nøjes med at tage 3 tal?
Hvad er risikoen for at du skal udtage mindst 7 tal for at få et primtal?

55. Roulette

Du spiller på en roulette hvor der er 37 mulige resultater: 0..36.
I hvert spil sætter du din indsats på dit lykketal nr. 13.

Find ved hjælp af BIN-Model:

1. Chancen for at dit lykketal kommer ud i de første 10 spil.

2. Chancen for at dit lykketal kommer ud i de første 20 spil.

3. Risikoen for at dit lykketal ikke kommer ud i løbet af de første 50 spil.

56. Toldeftersyn

I tolden bliver 5% af de rejsende udvalgt til et nærmere toldeftersyn.

I løbet af en sæson passerer du tolden 20 gange. Hvad er chancen for at du ikke nogen af de 20 gange bliver udvalgt til toldeftersyn?

57. Hvornår kommer sedlen ud?

I et lotteri er der 20% chance for at din lotteriseddel bliver udtrukket.

Beregn ved hjælp af en BIN-Model:

1. Chancen for at din seddel kommer ud i løbet af 3 lotterier.

2. Chancen for at din seddel kommer ud i løbet af 6 lotterier.

3. Risikoen for at din seddel ikke kommer ud i løbet af 10 lotterier.

58. En fifty-fifty chance

I et spil er der 2% chance for gevinst. Hvor mange spil skal du deltage i for at have 50% chance for at få en gevinst?

59. Bliver cyklen stjålet?

I et boligkvarter bliver 28% af cyklerne stjålet inden for en periode på et år.

Hvad er chancen for at en tilfældig cykel kan overleve 5 år uden at blive stjålet?

60. Kontrol i S-toget

I S-togene foretages der stikprøvekontrol af billetterne. Til kontrollen udtages 5% af de rejsende.

Hvad er sandsynligheden for at du ikke kommer ud for en kontrol i løbet af de næste 10 ture du tager med S-tog?

Og hvad er sandsynligheden for at du ikke kommer ud for en kontrol i løbet af 50 ture?

61. Skolebørn kommer til skade

Af Danmarks skolebørn kommer hvert skoleår 10% til skade i klasseværelset, i skolegården eller på vej til og fra skole.

Beregn ved hjælp af BIN-Model:

1. Chancen for at ingen af eleverne i en klasse på 15 elever kommer til skade i løbet af næste skoleår.

2. Chancen for at ingen af eleverne i en klasse på 20 elever kommer til skade i løbet af næste skoleår.

3. Risikoen for at mere end fem elever i en klasse på 20 elever kommer til skade i løbet af næste skoleår.

62. Hastighedskontrol

Chancen for at en bilist der kører ad motorvejen fra Lyngby til Helsingør kommer ud for en af politiets hastighedskontroller, er 1%.

Hvor mange ture skal bilisten foretage mellem de to byer for at der er 50% chance for at han kommer ud for en hastigheds-kontrol mindst én gang?

63. En sjælden hændelse

En sjælden hændelse vurderes til at indtræffe i gennemsnit én gang pr. 100 år.

Beregn ved hjælp af BIN-Model:

1. Chancen for at hændelsen indtræffer (dvs. forekommer mindst én gang) inden for de næste 50 år.

2. Chancen for at hændelsen indtræffer inden for de næste 100 år.

3. Chancen for at hændelsen ikke indtræffer inden for de næste 400 år.

64. En meget sjælden hændelse

En hændelse vurderes til at indtræffe i gennemsnit én gang pr. 1000 år.

Beregn ved hjælp af BIN-Model:

1. Chancen for at hændelsen indtræffer inden for 100 år.

2. Chancen for at hændelsen indtræffer inden for 500 år.

3. Chancen for at hændelsen ikke indtræffer inden for 2000 år.

65. Tre seksere

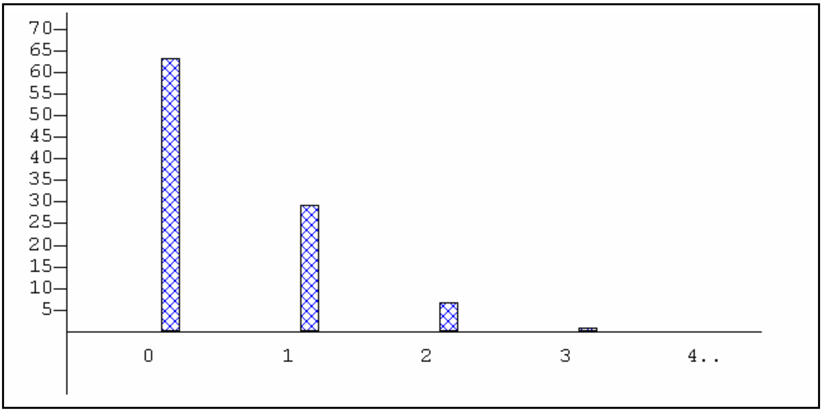
I et kast med tre terninger er chancen for "3 seksere" 1 ud af 216.

Beregn ved hjælp af BIN-Model:

1. Chancen for at du kan slå "3 seksere" i løbet af 25 kast.

2. Risikoen for at du må kaste mere end 100 kast før du opnår "3 seksere".

3. Hvor mange kast du skal udføre for at have 50% chance for mindst én gang at få "3 seksere".



66. Enarmet tyveknægt

På en spilleautomat har du en chance på 3 ud af 8000 for at få den store gevinst, "jackpot".

Undersøg ved hjælp af BIN-Model:

1. Chancen for jackpot i løbet af 500 spil.
2. Chancen for jackpot i løbet af 1000 spil.
3. Risikoen for ikke at få jackpot i løbet af 4000 spil.

67. Yatzy

Chancen for "Fem ens" i et kast med fem terninger er 1/1296.

(1) Hvad er chancen for at få "Fem ens" i løbet af 100 kast?

(2) Hvad er chancen for at få "Fem ens" i løbet af 500 kast?

(3) Hvor mange kast skal du udføre for at have 50% chance for at få "Fem ens".

68. Er 13 et uheldigt tal?

I et lottospil udtrækkes 6 tal blandt tallene 1 - 49. En statistik over 1212 spil viser at tallet 13 kun er udtrukket 121 gange.

Benyt en BIN-Model med 1212 udtagelser og en andel af røde kugler på 6/49.

Undersøg hvad chancen er for en observation så skæv som 121.

Konfidens

5. Fra data til model
Vi har i det foregående set hvordan vi kommer fra en model til data. Med udgangspunkt i en BIN-Model har vi beregnet sandsynligheder for de data der kan forekomme ved eksperimentet.

Vi har altså taget udgangspunkt i at vi kender indholdet af den æske hvorfra kuglerne udtages, vi ved hvad andelen af røde kugler er. Herudfra kan vi ved hjælp af BIN-Model fx beregne 90%-midterområdet for stikprøveudtagelsen.

En anden problemstilling
Vi ser nu på en anden problemstilling. Fra en æske udtages 100 kugler ved udtagelse med tilbagelægning, og vi noterer os at 20 af kuglerne er røde. Spørgsmålet er: **Hvad er andelen af røde kugler i æsken?**

Strengt taget kan vi ikke udelukke at æsken fx kun indeholder 5% røde kugler. Selv i en sådan situation kan det jo forekomme at vi får 20 røde kugler eller endnu flere i 100 udtagelser. Men det vil være en hændelse med en lav sandsynlighed.

Ligeledes kan det forekomme at vi får 20 røde kugler eller endnu færre når vi udtager100 kugler fra en æske som indeholder 50% røde kugler. Men heller ikke det vil have nogen stor sandsynlighed.

Lad os forestille os at vi har 99 æsker med kugler. I den første æske er andelen af røde kugler 0.01, i den næste æske er andelen af røde kugler 0.02, i den næste æske 0.03, osv.
Vi ser fx på den æske hvor andelen af røde kugler er 0.30. Fra denne æske udtager vi 100 kugler ved udtagelse med tilbagelægning.

Vi lader nu BIN-Model beregne midterområder:

Andel af røde kugler i æsken: 0.30

Midterområde	Sandsynlighed
27 .. 33	50 %
25 .. 35	75 %
23 .. 38	**90 %**
21 .. 39	95 %
19 .. 42	99 %

Vi ser at 90%-dataområdet er 23..38. Det betyder at der er 90% chance for at antallet af røde kugler i stikprøven vil være fra 23 til 38.

Dette vil vi fortolke sådan at kun 20 røde kugler i stikprøven på de 100 kugler ikke er særlig sandsynlig. Vi vil derfor ikke have stor tiltro til at vor ukendte æske skulle indeholde 30% røde kugler.

Vi ser nu på en æske hvor andelen af røde kugler er 0.25. For 90%-midterområdet får vi her:

Andel af røde kugler i æsken: 0.25

Midterområde	Sandsynlighed
22 .. 28	50 %
20 .. 30	75 %
18 .. 32	**90 %**
17 .. 34	95 %
14 .. 37	99 %

Denne gang indeholder 90%-midterområdet fra 18 til 32 røde kugler. Da vor stikprøve indeholdt 20 røde kugler, vil vi ikke afvise at den kan stamme fra denne æske hvor andelen af røde kugler er 0.25.

Her er nøglen til vor løsning af problemet:
Hvad er andelen af røde kugler i den æske vi udtog de 100 kugler fra og hvor vi fik 20 røde kugler?

Svaret er:
Vi vil finde de æsker hvor et antal på 20 røde kugler i stikprøven hører med til 90%-midterområdet.

Vi kan prøve os frem med æsker hvor andelen af røde kugler er 0.01, 0.02, 0.03 osv. indtil vi opnår at 20 er med i 90%-midterområdet.

Det viser sig at vi får gevinst første gang ved en andel på 0.14. Her er 90%-midterområdet: 9..20. Vi får fra BIN-Model:

Andel af røde kugler i æsken: 0.14

Midterområde	Sandsynlighed
12 .. 16	50 %
10 .. 18	75 %
9 .. 20	**90 %**
8 .. 21	95 %
6 .. 24	99 %

Hvis vi fortsætter med højere andele, viser det sig at 20 er med i 90%-midterområdet indtil andelen er 0.27. Her bliver 90%-midterområdet: 20..34. For højere andele er 20 ikke mere med i 90%-midterområdet:

Andel af røde kugler i æsken: 0.27

Midterområde	Sandsynlighed
24 .. 30	50 %
22 .. 32	75 %
20 .. 34	**90 %**
19 .. 36	95 %
16 .. 39	99 %

Alt i alt har vi altså at 20 er med i 90%-midterområdet for alle æsker med en andel fra 0.14 til 0.27.

Dette interval fra 0.14 til 0.27 vil vi kalde et **90%-konfidensinterval** for den ukendte andel af røde kugler. Vi kender jo ikke den rigtige andel af røde kugler i den æske vi udtog de 100 kugler fra , men vi har nu opstillet et skøn over hvad andelen kan tænkes at være. Og her vil vi opfatte intervallet fra 0.14 til 0.27 som et fornuftigt skøn.

Du kan nu prøve programmet **Konfidens**. Her skal du blot indtaste antallet af kugler i stikprøven og antallet eller andelen af røde kugler. Du indtaster altså 100 og 20. Programmet melder nu:

Konfidens
Stikprøvestørrelse: 100
Antal røde kugler : 20

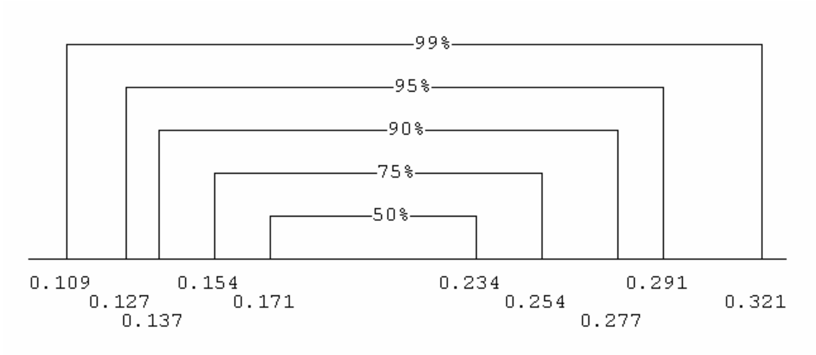
Konfidensniveau	Konfidensinterval
50%	0.171 .. 0.234
75%	0.154 .. 0.254
90%	**0.137 .. 0.277**
95%	0.127 .. 0.291
99%	0.109 .. 0.321

Du kan se at 90%-konfidensintervallet her angives til:

0.137 ... 0.277

Programmet angiver intervallerne med tre decimaler hvor vi før kun brugte to. Du kan let ved hjælp af BIN-Model kontrollere at det angivne interval er korrekt.

INSTAT kan også give et grafisk billede af **konfidensområderne**:



Ordet konfidens

Ordet **konfidens** kommer fra latin og betyder **tillid** eller **tiltro**. Med det angivne 90%-konfidensinterval har vi givet et skøn over den ukendte andel af røde kugler: "Vi har 90% tillid til at den ukendte andel findes i det givne interval fra 0.137 til 0.277".

De 90% henviser til at vi ved konfidensberegningen har benyttet **en metode** som i 90% af tilfældene vil føre til et interval der indeholder den ukendte andel af røde kugler. De 90% beskriver sikkerheden i vor metode**: I et stort antal anvendelser kan vi gå ud fra at metoden i 90% af tilfældene fører til et resultat hvor det beregnede interval indeholder den ukendte andel**.

Advarsel: Selv om metoden i 90% af tilfældene fører til et korrekt resultat, så kan vi ikke vide om et forelagt konkret resultat er korrekt eller ej. Vi bruger en god metode, men vi kan ikke vide hvornår den fører til et korrekt resultat.

Prøv selv

1. Syv gevinster. Du deltager 20 gange i et spil og får gevinst 7 gange. Angiv et 90%-konfidensinterval for gevinstchancen i spillet.

2. Hvad er andelen af røde kugler? Du udtager 200 kugler fra en æske og finder at 147 af dem er røde. Angiv et 90%-konfidensinterval for andelen af røde kugler i æsken.

3. Lotto. Du spiller på 200 lottorækker og får gevinst på 4 af dem. Angiv et 90%-konfidensinterval for gevinstchancen i lotto.

4. En politisk meningsmåling. Blandt 1100 adspurgte vælgere fortæller 285 at de vil stemme på Venstre. Angiv et 90%-konfidensinterval for partiets tilslutning blandt vælgerne.

5. Møntkast. I 100 kast med mønt opnår du 63 kronekast. Angiv et 90%-konfidensinterval for kronechancen med den pågældende mønt.

6. Terningkast. I 100 kast med en terning opnår du 12 seksere. Hvad fortæller det om sekserchancen med den pågældende terning?

7. Ingen gevinst. Du spiller 20 gange i et spil og får ingen gevinster. Hvad fortæller det om gevinstchancen i spillet?

8. Gevinst hver gang. Du spiller 20 gange i et spil og får gevinst hver gang. Hvad fortæller det om gevinstchancen i spillet?

En afprøvning af konfidensområderne

Vi afprøver beregningen af konfidensområder ved at lade computeren udtage kugler fra en æske hvor andelen af røde kugler er 0.20. Computeren optæller antallet af røde kugler der udtages og beregner de tilhørende konfidensområder.

Derefter undersøger den om de beregnede konfidensområder indeholder værdien 0.20.

Her er en oversigt over 100 serier af udtagelser med 100 udtagelser pr. serie.

```
Antal serier: 100
Antal udtagelser i hver serie: 100
Andel af røde kugler i den æske
der udtages fra: 0.20
```

Konfidensområde:	Antal serier hvor 0.20 ligger i dette område:
K1(50%)	55
K2(75%)	86
K3(90%)	93
K4(95%)	97
K5(99%)	98

Af listen ser vi at 55 af de 100 serier gav et 50%-konfidensområde som indeholdt tallet 0.20. Og at 93 af serierne gav et 90 % -konfidensområde som indeholdt 0.20. I 93 tilfælde ud af 100 indeholder 90%-konfidensintervallet altså tallet 0.20. Men vi må også bemærke at i 7 tilfælde ud af de 100 fik vi et 90%-konfidensområde som ikke indeholdt tallet 0.20.
Tabellen bekræfter dog at vi med god ret kan benytte konfidensområderne når vi ønsker at opstille et skøn over den ukendte værdi af andelen af røde kugler i den æske hvorfra vi udtager kugler.

Konfidensintervallernes bredde

Vi ser på en stikprøve på 100 kugler fra en æske. Lad os antage at stikprøven indeholder 50 røde kugler. Vi kan nu beregne et 90%-konfidensinterval for den ukendte andel af røde kugler i æsken:

90% 0.414 .. 0.586

Vi ser at bredden af intervallet er 0.586 – 0.414 = 0.172.

Vi prøver nu med andre stikprøveresultater, nemlig med stikprøver som indeholder 30 røde kugler, 20 røde kugler, 10 røde kugler og 5 røde kugler. Vi samler vore resultater i en oversigt:

Andel røde kugler i stikprøven	**90%-konfidens-interval**	**Bredde af interval**
0.50	0.414...0.586	0.172
0.30	0.225...0.384	0.159
0.20	0.137...0.277	0.140
0.10	0.056...0.163	0.107
0.05	0.020...0.102	0.082

Vi ser at det bredeste interval får vi ved en andel på 0.50 røde kugler i stikprøven.

Dette gælder for alle stikprøver med en bestemt størrelse, fx for stikprøver på 100 kugler: Det bredeste konfidensinterval forekommer når andelen af røde kugler i stikprøven er 0.50. Jo mere andelen afviger fra 0.50, jo kortere bliver konfidensintervallet.

Prøv selv
Brug en stikprøvestørelse på 250 og beregn 90%-konfidens-intervaller for følgende andele af røde kugler i stikprøven:

0.50, 0.20, 0.10, 0.80 og 0.90

Prøv også med en stikprøvestørrelse på 1000 og de samme andele af røde kugler i stikprøven.

Et konfidensområde af en given bredde

Der er en sammenhæng mellem konfidensintervallernes bredde og bredden af stikprøvernes midterområder. Vi ser på et eksempel: Vi ønsker et 90%-midterområde som har en bredde på 0.10. Som du har set, kan vi beregne det ved hjælp af BIN-Model: Gå ind i BIN-Model Stikprøvestørrelse ukendt. Indtast:

Andel af røde kugler i æsken:	0.50
Midterområdets sandsynlighed:	0.90
Bredde af midterområde:	0.10

Programmet beregner nu en stikprøvestørrelse på 271, og det tilhørende 90%-midterområde beregnes til 122..149 røde kugler, svarende til 0.450..0.550 i andele af røde kugler i stikprøven.
Vi har altså her et midterområde med en bredde på 0.10.

Prøver vi nu at beregne et 90%-konfidensområde med en stikprøvestørrelse på 271 og en andel af røde kugler i stikprøven på 0.50, får vi:

90%   0.451 .. 0.553

altså et 90%-konfidensinterval med bredden 0.102, dvs. kun en smule bredere end de 0.10.

Ønsker vi et 90%-konfidensinterval som ikke er bredere end de 0.10, kan vi uanset andelen af røde kugler i stikprøven benytte en stikprøvestørrelse på 271. Ved en andel af røde kugler på 0.50 bliver bredden af konfidensintervallet 0.102, og ved enhver anden andel af røde kugler i stikprøven vil konfidensintervallet blive snævrere, som vi så før.

Hvis vi ønsker et konfidensområde som har bredden 0.100, kan vi prøve med lidt større stikprøvestørrelser end 271. Det viser sig at 278 er nok til at give en bredde på 0.100 ved en andel af røde kugler i stikprøven på 50%. Kontroller det ved brug af Konfidens.

6. Opgaver til Konfidens

1. Hvad er andelen af røde kugler i posen?

Fra en pose udtages med tilbagelægning 50 kugler. Blandt dem er der 15 røde kugler. Opstil 90%-, 95%- og 99%-konfidensintervaller for den ukendte procentdel af røde kugler i posen.

2. En mønt kastes

Af 300 møntkast resulterer de 128 i Krone. Angiv et 90%-konfidensinterval for sandsynligheden for Krone i kast med den pågældende mønt.

3. En skæv mønt

I 100 kast med en skæv mønt opnås 67 kronekast. Angiv et 90%- konfidensinterval for sandsynligheden for Krone.

4. Hvor mange rygere?

En undersøgelse viser at der er 267 rygere blandt 1000 tilfældigt udvalgte danskere i aldersgruppen fra 15 år.

Giv et 90%-konfidensinterval for procenten af rygere i Danmark.

5. En politisk meningsmåling

En politisk meningsmåling der bygger på svar fra 1200 personer giver følgende tilslutning til tre af folketingets partier:

Socialdemokratiet	:	346
Konservative folkeparti	:	118
Radikale venstre	:	71

Beregn 90%-konfidensintervaller for de tre partiers tilslutning.

6. En forbrugerundersøgelse

I en undersøgelse omfattende 100 tilfældigt udvalgte danske husstande indsamles følgende data:

92	husstande har køleskab
63	husstande har dvd-afspiller
38	husstande har mikrobølgeovn
8	husstande har telefax

Opstil 90%-konfidensintervaller for udbredelsen af de anførte tekniske hjælpemidler i danske husstande.

7. Folkets røst

En formiddagsavis spørger fem tilfældigt udvalgte om de er for eller imod Danmarks overgang til Euroen. Tre af dem er imod. Hvad kan vi slutte heraf om danskernes holdning til Euroen?

8. Møntkast

Udfør 50 kast med en mønt og optæl antallet af kronekast. Lad Konfidens beregne et 90%-konfidensinterval for møntens kronesandsynlighed.

9. Terningkast

Udfør 50 kast med en terning og optæl antallet af seksere. Lad Konfidens beregne et 90%-konfidensinterval for terningens seksersandsynlighed.

10. Chancen for en dreng

Et år fødes der i København 4910 børn, af dem var 2530 drenge. Opstil et 90%-konfidensinterval for "drengesandsynligheden" i København. Er der grund til at tro at drengesandsynligheden er større end 50%?

11. En klar ændring?

Skaf dig tallene fra en af de nyeste politiske meningsmålinger og sammenlign partiernes tilslutninger med resultaterne fra det seneste folketingsvalg. Læg konfidensintervaller omkring tallene fra meningsmålingen og undersøg ved hjælp af dem om nogle af partierne har haft en klar fremgang eller tilbagegang siden valget.

12. Forcensur

Når folkeskolens afgangsprøver i regning og matematik har været afholdt, skaffer undervisningsministeriet sig et hurtigt overblik over den forventede karakterfordeling ved at udtage en stikprøve af elever og lade deres besvarelser rette før det øvrige rettearbejde går i gang. I stikprøven indgår elever med et bestemt nr. på klasselisten. (Hvorfor kan man ikke blot lade censor selv udvælge en elev fra klassen?)

Et år viser stikprøven på 1834 elever at 276 vil få en karakter på 10 eller derover, og at 115 vil få en karakter under 6. Angiv et 90%-konfidensinterval for hvor stor en procentdel af årets eksaminander der vil få en karakter på 10 eller derover, og hvor stor en procentdel der vil få en karakter under 6.

13. En stavetest pr. stikprøve

Prøv en af dine kammeraters stavefærdighed på følgende måde: Du udvælger tilfældige ord fra retskrivningsordbogen, læser dem op for din kammerat og lader ham fortælle hvordan ordene staves.

Prøv din kammerats stavefærdighed ved en stikprøve på mindst 25 ord.

Angiv et konfidensinterval for andelen af ord i retskrivningsord-bogen som du mener din kammerat kan stave rigtigt.

Byt derefter roller, og lad din kammerat teste dig.

14. Dit kendskab til engelske gloser

Tilrettelæg på samme måde som i opgave 13 en test til under-søgelse af din kammerats kundskaber i engelsk. Udvælg ord fra en engelsk-dansk ordbog, og undersøg hvor mange han kan oversætte korrekt.

15. En tombola

Sanne trækker 20 sedler i en tombola og får gevinst på de 6. Hvad kan der siges om tombolaens gevinstsandsynlighed? Er der tilstrækkelig grund til at tvivle på at der er gevinst på mindst 50% af sedlerne?

16. Er terningen falsk?

I 15 kast med en terning opnår du overhovedet ingen sekser. Er der grundlag for at påstå at terningen er falsk?

17. En meningsmåling på din skole

Lav en stikprøveundersøgelse på din skole. Vælg et spørgs-mål som kan besvares med ja eller nej, og spørg 20-30 tilfæl-digt valgte elever. Overvej først hvordan du skal foretage ud-vælgelsen af de elever der skal spørges.

Opstil på grundlag af svarene et 90%-konfidensinterval for ja-svar på din skole.

18. Hvor mange rygere?

Lav en undersøgelse i din skoles 7-9 klasse af hvor mange procent af eleverne der er rygere. Vælg en passende stikprøve og beregn 90%-konfidensintervaller på grundlag af dine resul-tater fra stikprøven.

Undersøg også om der ser ud til at pigernes rygerprocent er forskellig fra drengenes.

19. Venstrehåndet

Lav som i opgave 18 en undersøgelse af hvad procenten af venstrehåndede kan være blandt eleverne på din skole. Lad undersøgelsen omfatte elever fra alle klassetrin.

20. Hvordan ændres bredden af konfidensintervallerne?

Find ud af hvilken sammenhæng der er mellem stikprøvens omfang og konfidensintervallets bredde. Vælg fx en andel af røde kugler i stikprøven på 0.4 og find 90%-konfidensintervallernes længde for følgende stikprøvestørrelser:

n=25, n=100, n=400, n=1600.

Hvilken sammenhæng finder du?

21. Andre andele af røde kugler

Gentag undersøgelsen fra opgave 20 med nogle andre valg for andelen af røde kugler.

22. Hvor mange skal spørges?

I en politisk meningsmåling om tilslutningen til de politiske partier skal inddrages så mange personer at de 90%-konfidensintervaller der kan opstilles for partiernes vælgertilslutning ikke har en længde på mere end 0.05, uanset hvad partiets tilslutningsprocent er.

Find ved hjælp af Konfidens hvor mange personer der skal indgå i meningsmålingen for at det stillede krav kan opfyldes.

23. En mønt testes

Malene kaster en mønt 50 gange for at undersøge dens sandsynlighed for at vise Krone. I de 50 kast får hun Krone 19 gange. Vurder dette resultat ved hjælp af Konfidens.

24. En terning testes

Sanne kaster en terning 100 gange for at undersøge dens sandsynlighed for at give sekser. I de 100 kast får 10 seksere. Vurder dette resultat ved hjælp af Konfidens. Tror du der er tale om en ægte terning med en seksersandsynlighed på 1/6?

25. Ligevægtede udfald?

Sanne kaster en tegnestift 50 gange. I 35 tilfælde falder den med spidsen opad. Angiv ved hjælp af Konfidens et skøn over den ukendte sandsynlighed for "spidsen opad".

26. Hvor mange træffere?

En mesterskytte påstår at hans træfsikkerhed er 80%. I 25 skud opnår han 15 træffere. Vurder hans påstand ved hjælp af Konfidens.

27. For små karakterer?

Ved en afgangsprøve hvor man bruger 13-skalaen er der normalt 20% der får en karakter på 6 eller derunder. I en klasse med 25 elever viser det sig at der er 8 elever som får 6 eller derunder. Giver dette grund til at tro at prøven har været for svær? Besvar spørgsmålet ved hjælp af Konfidens.

28. Er kvaliteten i orden?

I en produktion er normalt 60% af de producerede enheder af 1. kvalitet. I en stikprøve på 50 enheder findes kun 24 enheder af 1. kvalitet. Giver dette grund til bekymring?

29. En kvalitetskontrol

Ved en kvalitetskontrol undersøges om produktionen højst indeholder 10% defekte enheder. I en stikprøve på 50 enheder er der 8 defekte enheder. Kan produktionen godtages? I en stikprøve på 100 enheder er der 16 defekte enheder. Kan produktionen godtages?

30. Et kedeligt spil
Malene spiller 100 gange i et spil. Hun får kun gevinst i ét af de 100 spil. Hvad fortæller det om gevinstchancen i spillet?

31. Hvor mange billedmærker?
Sanne får tilbudt en sæk med brugte frimærker. Det påstås at 40% af frimærkerne i sækken er billedmærker. Sanne udtager en håndfuld på 150 frimærker hvoriblandt der viser sig at være 50 billedmærker. Kan hun tro på sælgerens påstand om 40% billedmærker i sækken?

32. En terning testes
Ved 500 kast med en terning testes at seksersandsynligheden er 1/6. Der opnås 63 seksere. Hvad er din konklusion?

33. Er kvaliteten faldet?
En produktion garanteres at indeholde mindst 60% enheder af 1. kvalitet. En stikprøve på 100 enheder indeholder 52 enheder af 1. kvalitet. Hvor godt stemmer det med garantien?

En stikprøve på 250 enheder indeholder 135 enheder af 1. kvalitet. Hvad er din konklusion?

Testgraf

7. En testsituation

Lad os se på et eksempel. En spillehal ønsker at anskaffe en ny spillemaskine. Spillemaskinefirmaet påstår at maskinen giver spillerne en gevinstchance på 30%. Spillehallen ønsker at undersøge denne påstand ved et eksperiment.

Man opstiller den hypotese at firmaet har ret: Maskinens gevinstchance er 30%. Et testeksperiment skal så afgøre om hypotesen skal forkastes eller om den kan godtages. Eksperimentet kan fx bestå i at der udføres 50 spil. I et sådant antal spil skal der i gennemsnit forekomme 30% gevinstspil, dvs. 15 gevinstspil. Spillehallen er tilfreds hvis antallet af gevinster bliver fra 10 til 20, men under 10 og over 20 gevinster vil man ikke acceptere.

Den opstillede hypotese belyses ved en test med **godtagelsesområdet 10 .. 20.**

Det betyder følgende:

> Når eksperimentet giver et resultat som falder i godtagelsesområdet, accepteres hypotesen.
>
> Når eksperimentet giver et resultat som falder uden for godtagelsesområdet, forkastes hypotesen.

Testen fører altså til en af de to afgørelser:

(1) Hypotesen godtages

(2) Hypotesen forkastes

Nu er det desværre ikke sikkert at testen resulterer i den rigtige afgørelse. Det kan jo ske at testen fører til at hypotesen forkastes selv om hypotesen i virkeligheden er sand. Og det kan også ske at testen fører til at hypotesen godtages selv om den er falsk. - Der er risiko for to typer af fejl:

Fejl af 1. art: Hypotesen forkastes, men er sand.

Fejl af 2. art: Hypotesen godtages, men er falsk.

En fejl af 1. art vil her bestå i at hypotesen om gevinstchancen bliver forkastet selv om gevinstchancen i virkeligheden er 30% som påstået. En fejl af 2. art vil bestå i at påstanden om en gevinstchance på 30% godtages selv om gevinstchancen i virkeligheden ikke har denne værdi.

En fejl af 1. art vil altså være til skade for spillemaskinefirmaet. Det får jo underkendt spillemaskinen til trods for at den er god nok. En fejl af 2. art er til skade for køberen af spillemaskinen. Han godtager en maskine som ikke er i overensstemmelse med de givne specifikationer.

I denne situation kan vi sige at fejl af 1. art udgør ***sælgers risiko*****,** medens fejl af 2. art er ***købers risiko***.

I andre forbindelser beskrives fejl af 1. art som en **negativ-falsk** beslutning: Hypotesen forkastes, men denne beslutning er falsk. Vi siger nej til hypotesen, men det er en falsk beslutning.

Tilsvarende beskrives fejl af 2. art som en **positiv-falsk** beslutning: Hypotesen godtages, men denne beslutning er falsk. Vi siger ja til hypotesen, men det er en falsk beslutning.

I medicinsk sammenhæng kan en negativ-falsk beslutning betyde at en test viser at patienten ikke er syg, men dette er ikke rigtigt. En positiv-falsk beslutning betyder at testen viser at patienten har sygdommen, men dette er ikke rigtigt.
Den der gør brug af en test, er selvfølgelig interesseret i at vide hvor stor en risiko der er for at testen fører til forkerte

afgørelser: Hvad er risikoen for en fejl af 1. art? Og hvad er risikoen for en fejl af 2. art?

Fejl af 1. art. Testens signifikansniveau
Det første af de to spørgsmål kan let besvares. Når der er opstillet en test med et godtagelsesområde, kan vi beregne den sandsynlighed der er knyttet til godtagelsesområdet når den opstillede hypotese er sand. Har vi fx et godtagelsesområde som har en sandsynlighed på 90%, vil risikoen for fejl af 1. art højst være 10%. At godtagelsesområdet har en sandsynlighed på 90% betyder jo:

> **Når hypotesen er sand, er der mindst 90% chance for at testeksperimentet giver et resultat som falder i godtagelsesområdet.**

Når hypotesen er sand, er der derfor mindst 90% chance for at testen fører til at hypotesen godtages og dermed højst 10% risiko for at hypotesen forkastes. Benytter vi en test med et 90%-godtagelsesområde er risikoen for fejl af 1. art altså 10%.

En test som har en risiko på højst 10% for fejl af 1. art, kaldes i statistikken for en test med et ***signifikansniveau*** på 10%. Når man benytter en test med et signifikansniveau på 10%, er der altså højst en risiko på 10% for at testen fører til fejlagtig forkastelse af den opstillede hypotese, dvs. forkastelse af en hypotese som er sand.

Hvis testeksperimentet med den forelagte test giver et resultat som fører til forkastelse af hypotesen - altså et resultat uden for godtagelsesområdet - siger man at dette **resultat er statistisk signifikant på 10%-niveauet**. Eller at resultatet har **et skævniveau på 10%**.

Fejl af 2. art.
Hvad nu med fejl af 2. art, altså fejl som består i at testen fører til godtagelse af en falsk hypotese? Her er situationen ikke så

enkel. Sagen er jo den at en hypotese kan være falsk på mange måder.

I eksemplet med spillemaskinen kan den opstillede antagelse om en gevinstchance på 30% være falsk fordi den rigtige chance er 40%. Men det kan også være at den rigtige chance er 25%, eller 20%, eller 50%. Der er mange muligheder, og til hver af dem svarer en risiko for fejl af 2. art.

Vi kan derfor ikke give én bestemt risiko for fejl af 2. art, vi bliver nødt til at tage hensyn til hvilken af de mulige "falskheder af hypotesen" der tænkes på.

Er modellen korrekt?
I øvrigt kunne det også tænkes at det er falsk at spillene på spillemaskinen kan beskrives ved den model vi har benyttet, altså modellen for kugleudtagelse med tilbagelægning. Denne model ville jo ikke kunne benyttes hvis der fx var en påvirkning fra det ene spil til det andet, så spillene ikke var uafhængige af hinanden.

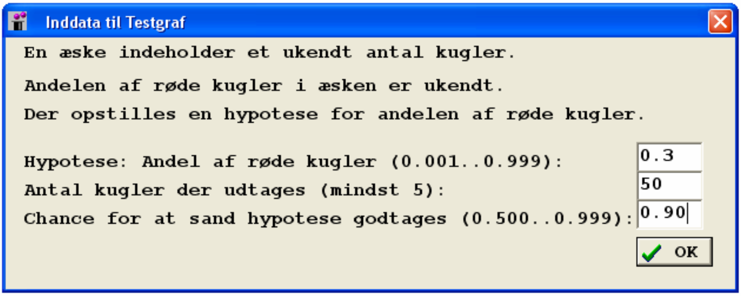
Vi vil i det følgende gå ud fra at kugleudtagelses-modellen kan benyttes som model for spillene på spillemaskinen. En sådan model kaldes også en ***binomialfordelings-model***, og den testtype vi skal se på, kaldes en ***binomialtest***.

I arbejdet med binomialtest benytter vi programmet **INSTAT**. Her findes en række underprogrammer. Vi ser nu nærmere på underprogrammet **Testgraf**.

8. Testgraf: Dobbeltsidede test

Til behandling af spørgsmål vedrørende opstilling af binomialtest vil vi hente hjælp i underprogrammet **Testgraf**, og dette program kan klare alle de udregninger vi har brug for ved overvejelser over fejl af 2. art i binomialtest. Det kan derfor også besvare de spørgsmål vi opstillede i forbindelse med den fore-slåede test for spillemaskinen.

Lad os hvad Testgraf kan. Når du åbner dette program får du følgende tekst på skærmen:



Vi har indtastet tallene fra testen af spillemaskinen: Her er hy-potesen at andelen af røde kugler (= gevinstsandsynligheden) er 0.30. Vi udfører 50 spil, og vi ønsker en test hvor der er mindst 90% chance for at hypotesen godtages når den er sand.

Programmet beregner nu tre typer af godtagelsesområder:

1. område (midterområde): 10 .. 20

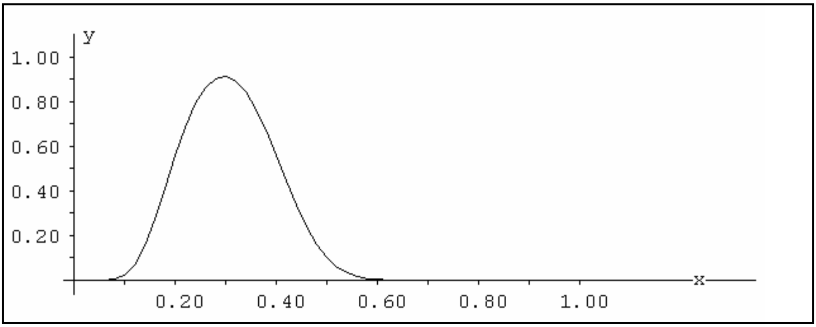
2. område (venstreområde): 0 .. 19

3. område (højreområde): 11 .. 50

Vi kan vælge et af disse tre testområder, og programmet vil da tegne et grafisk billede, den såkaldte **testgraf**.

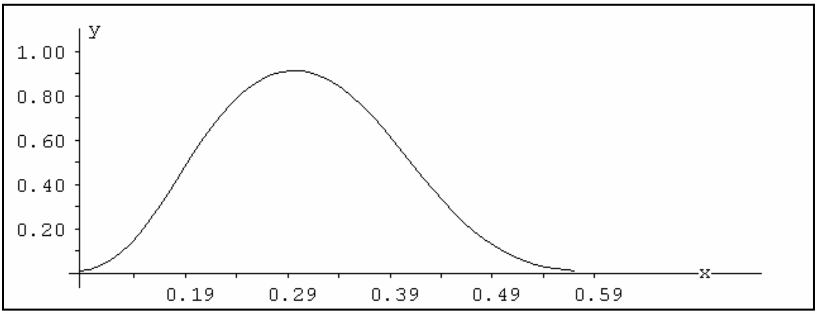
Vi er interesseret i en test med et midterområde som godtagel-sesområde, og vi vælger derfor vælger vi at få tegnet testgra-fen for område 1. Vi klikker derfor på "midter". Vi godtager altså hypotesen såfremt vi får at antallet af gevinster ligger i det beregnede midterområde, dvs. i området 10..20

Vi får nu følgende figur på skærmen:

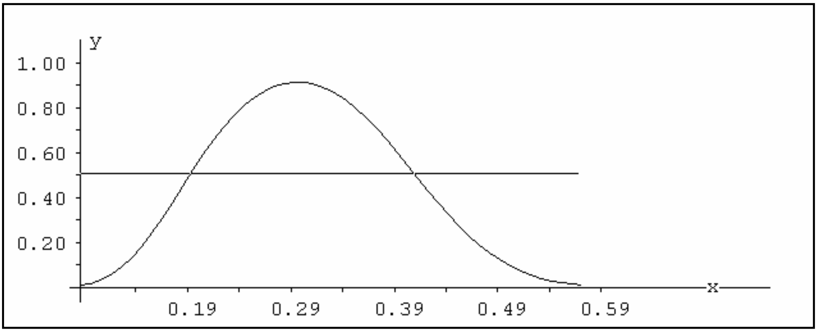


Tegningen viser testens *acceptkurve* eller dens *testgraf*. Den fortæller i et grafisk billede hvad sandsynligheden er for at testen fører til godtagelse af den opstillede hypotese.

Figuren viser hele området fra 0 til 1 på x-aksen. Hvis vi vil gå nærmere på tegningen, kan vi klikke på knappen **Detaljer**. Vi får da et nærbillede af figuren:



Hvis du vil foretage aflæsninger, så venstreklik et sted på figuren. Nu viser der sig en lineal. Den kan være til hjælp når du fx vil se hvor kurven har en højde på 0.5. Linealen kan fjernes igen ved et højreklik.



Da det kan være svært at aflæse nøjagtigt på figurerne, giver programmet mulighed for at vi kan indtaste x-værdier og få udskrevet de tilhørende y-værdier for kurvepunkterne. Indtast fx værdien x = 0.3 og tryk Enter. På skærmen udskrives nu: y = 0.912. Det betyder at når gevinstchancen har en værdi på 0.3, så er der en sandsynlighed på 0.912 for at testen fører til godtagelse af den opstillede hypotese. Testen opfylder altså det stillede krav: **Når hypotesen er sand, skal der være mindst 90% chance for at testen fører til godtagelse af hypotesen.**

Lad os prøve nogle flere udregninger. Du kan finde dem til højre på skærmen når du indtaster de tre x-værdier:

```
x= 0.200   y=  0.556
x= 0.400   y=  0.560
x= 0.500   y=  0.101
```

Når maskinens gevinstchance er 0.2 i stedet for 0.3, er der en chance på 0.556 for at testen fører til godtagelse af hypotesen om at gevinstchancen er 0.3. *Der er altså en risiko på over 55% for at begå en fejl af 2. art når maskinens virkelige gevinstchance er nede på 20%*.

Dette viser os netop svagheden ved en signifikanstest: Testen opfylder ganske vist vort krav om at risikoen for fejl af 1. art er begrænset til 10%, men risikoen for fejl af 2. art kan langt overstige dette tal. Hvis vi anvender den foreliggende test, er der over 55% risiko for at vi godtager den opstillede hypotese om en gevinstsandsynlighed på 30% selv om den i virkeligheden er nede på 20%.

Også ved en gevinstchance på 40% er der en høj risiko for fejl af 2. art, nemlig 0.560. Først ved en gevinstchance på 50% er risikoen for fejl af 2. art nede på et tåleligt niveau, nemlig kun ca. 10%.

Du kan på figuren se at testgrafen har sin top for x-værdier i nærheden af 0.30. Sådan bør det også være: *Når x er nær ved 0.30, skal der være stor sandsynlighed for at testen fører til godtagelse af hypotesen x=0.30*.

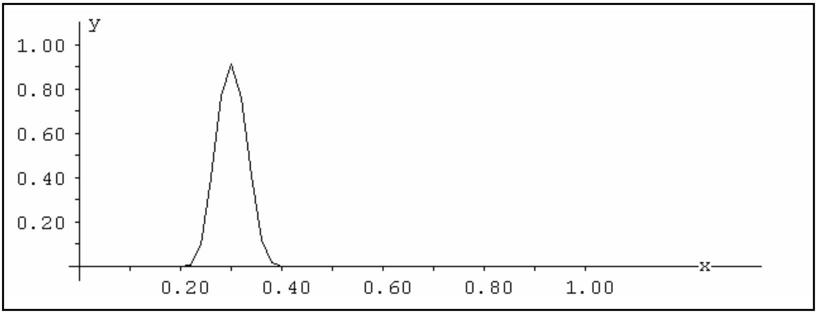
For x-værdier langt fra 0.30 er der kun lille sandsynlighed for at testen fører til godtagelse af hypotesen x=0.30. Også dette er helt som det bør være: *jo mere den virkelige x-værdi afviger fra 0.30, jo mindre sandsynlighed må der være for at testen godtager hypotesen x=0.30*.

Hvad testgrafen fortæller

Den ideelle testgraf ville være en graf hvor kurven har en højde på 1 ved hypotesesandsynligheden (i vort eksempel ved x=0.3) og en højde på 0 ved alle andre x-værdier. Så ville risikoen for fejl være helt fjernet. Jo mere testgrafen ligner denne ideelle figur, jo mere følsom en test er der tale om, dvs. jo bedre er testen til at afsløre små afvigelser fra hypotesesandsynligheden.

En testgraf kan give os et hurtigt overblik over hvor god den opstillede test er. Jo stejlere og smallere kurven er, jo mere følsom er testen. Desværre kan vi jo ikke frit fastlægge testgrafens udseende. Så snart en bestemt test er valgt, dvs. så snart et testeksperiment og en godtagelsessandsynlighed er valgt, så er testgrafen i princippet fastlagt. Den vi her har set på, svarede til et testeksperiment på 50 spil og som godtagelsesområde et midterområde med en sandsynlighed på 90%. Hvis vi med denne godtagelsessandsynlighed vil have en mere fintmærkende test, må vi udvide testeksperimentet så det kommer til at omfatte et større antal udtagelser.

Her har vi udtaget 500 kugler i stedet for 50:

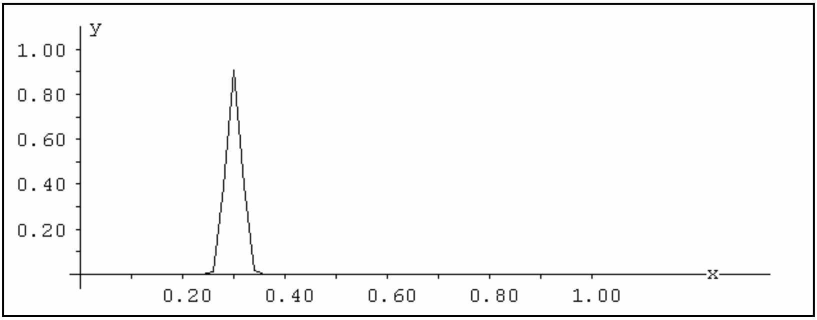


Vi ser at testgrafen nu er smallere og ligger mere koncentreret i området omkring x=0.30.

Kommentar: Du vil se at den nye kurve næsten har samme top som den kurve vi tegnede før, men den nye kurve er smallere. For en given x-værdi som ikke ligger lige omkring 0.30, er risikoen for fejl af 2. art derfor mindre ved den nye test med 100 spil end ved den gamle test med 50 spil.

Testen der omfatter 500 spil er ***stærkere***, dvs. mere følsom eller fintmærkende end testen der kun omfatter 50 spil. De to test har begge et 90%-godtagelsesområde, eller hvad der er det samme: de har begge et signifikansniveau på 10%. Men testen med de 500 spil har større chance for at afsløre en falsk hypotese. Testene har altså samme risiko for fejl af 1. art, nemlig ca. 10%, men ved testen der omfatter 500 spil er der mindre risiko for fejl af 2. art end ved testen der omfatter 50 spil.

Havde vi opstillet en test der fx omfattede 2000 spil, ville vi have fået en endnu stærkere test. Det gælder alment at ***jo større testomfang, jo stærkere test***, dvs. jo bedre er den til at afsløre falske hypoteser. Her er testgrafen for en test med 2000 spil:



Denne test har et godtagelsesområde fra 566 til 634. Det svarer til at andelen af røde kugler der udtages, er fra 28,3% til 31,7%. Det betyder at hvis testen resulterer i 28% røde kugler, så forkastes spillemaskinen. Men måske er afvigelsen mellem de påståede 30% og de opnåede 28% uden praktisk betydning for spillemaskine-køberen, så han alligevel godtager maskinen:

> **En afvigelse kan godt være statistisk signifikant og dog være uden praktisk betydning.**

Ved vurderingen af en test der fører til forkastelse af den opstillede hypotese bør man derfor først undersøge om den observerede afvigelse er af praktisk betydning. Hvis den ikke er det, er der ingen grund til at forkaste hypotesen selv om testresultatet er statistisk signifikant.

Prøv selv
Vi ser nærmere på spillemaskinens testgraf (50 udtagelser)

1. Hvad er risikoen for fejl af 2. art når den rigtige gevinst-chance er 45%?

2. Find for hvilke x-værdier risikoen for fejl af 2. art er mindre end 50%. (Prøv dig frem ved indtastning af x-værdier).

3. Hvordan kan du ved hjælp af testgrafen finde testens risiko for fejl af 1. art, dvs. testens signifikansniveau?

4. Lad Testgraf opstille en test for spillemaskinen der omfatter 200 spil. Benyt igen en test med midterområde som godtagelsesområde og en chance på 90% for at en sand hypotese godtages. - Undersøg hvad risikoen for fejl af 2. art er ved denne test når den rigtige gevinst-sandsynlighed for spillemaskinen er 20%, og når den er 40%.

5. Opstil en test som gør brug af 2000 spil og beregn risiko-en for fejl af 2. art når gevinstchancen er 20%, og når den er 40%.

*

De beregninger vi har foretaget over risikoen for fejl af 2. art, viser hvor farligt det kan være ukritisk at gøre brug af signifi-kanstest hvor man alene interesserer sig for fejl af 1. art, altså for testens signifikansniveau.

Især må man være forsigtig med sine konklusioner når en sig-nifikanstest *fører til godtagelse* af den opstillede hypotese. Når man ikke har styr på risikoen for fejl af 2. art, kan man i realiteten intet sige om hypotesens trolighed i denne situation. Det gælder især hvis testens omfang er beskedent.

Hypotesen kan altid forkastes?

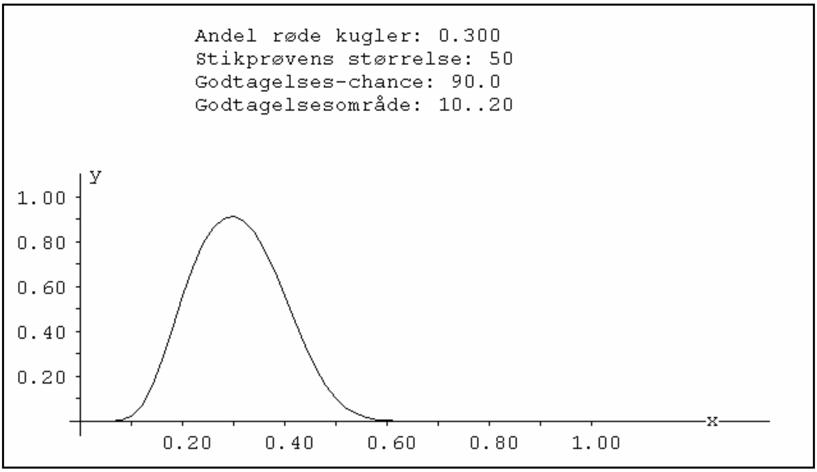
Undertiden siger man at en hypotese der udtaler sig om at hypotesesandsynligheden har en bestemt værdi, fx 0.30, altid vil kunne forkastes. Blot man opstiller en test med et stort omfang.

I virkelighedens verden er der jo ingen sandsynlighed der vil være lige netop 0.30, der vil altid være små praktiske afvigelser. Og sådanne afvigelser vil blive afsløret af en test hvor vi gør brug af et stort antal udtagelser.

9. Konfidens og hypotesetest

Vi har nu to metoder til at vurdere om en opstillet BIN-Model ligger bag de observationer vi foretager, altså bag de data vi indhenter ved stikprøveudtagelserne. Til eksempel ser vi på den hypotese at spillemaskinens vinderchance er 0.3, og vi udfører et eksperiment der består af 50 spil.

Vi kan da angive et 90%-godtagelsesområde. Det beregnes af Testgraf til at være området 10..20:



Hvis vi opnår en observation i området fra 10 til 20, vil vi godtage den opstillede hypotese. Fx vil vi godtage hypotesen hvis vi opnår 12 gevinster i de 50 spil.

Vi beregner nu et 90%-Konfidensområde for observationen 12:

90%   0.145 .. 0.359

Vi ser at 0.3 ligger i konfidensområdet. Vi vil derfor godtage at den ukendte vinderchance på spillemaskinen kan være 0.3.

Tilsvarende vil observationen 20 føre til godtagelse af den opstillede hypotese. Og for denne observation bliver 90%-Konfidensområdet:

90%   0.284 .. 0.526

Altså igen et konfidensområde der indeholder værdien 0.3.

Prøver vi med værdien 21 der ligger uden for godtagelsesområdet, får vi et 90%-Konfidensinterval der ikke indeholder 0.3:

90%   0.302 .. 0.546

Alt i alt vil der gælde at hvis hypotesetesten fører til godkendelse af hypotesessandsynligheden, så vil konfidensberegningen give et interval der indeholder denne sandsynlighed.

Og hvis hypotesetesten fører til forkastelse af hypotesesandsynligheden, så vil konfidensberegningen give et interval som ikke indeholder denne sandsynlighed.

De to metoder vil derfor føre til samme resultat.

Der gælder endvidere at hvis vi får en observation som ligger yderligt i godtagelsesområdet, så vil vi også få et konfidensområde hvor hypotesesandsynligheden ligger yderligt.

Hvis vi gør brug af **testgrafen** ved hypotesetesten, så får vi nogle oplysninger om niveauet for fejl af 2. art, altså risikoen for at hypotesen godtages selv om den er falsk. Sådanne oplysninger får vi ikke fra konfidensberegningerne.

De statistiske værktøjer **konfidensberegninger** og **hypotesetest** hører hjemme i et område af statistik som kaldes **induktiv statistik.** Her har man som udgangspunkt en opstillet statistisk model. Derefter indsamles observationer, og ved hjælp af de statistiske værktøjer afgøres det om de forelagte observationer peger på at den opstillede model kan godtages eller om den må forkastes.

10. Enkeltsidede test

Hidtil har vi kun set eksempler på test hvor godtagelsesområdet har været et midterområde. Ved sådanne test er området uden for godtagesesområdet, det såkaldte *skævområde*, opdelt i to områder: ét til venstre for godtagelsesområdet og ét til højre for godtagelsesområdet. Sådan en test med et midterområde som godtagelsesområde kaldes en ***dobbeltsidet test***. Den vil føre til forkastelse af den opstillede hypotese både når der observeres en for lille værdi (en værdi i skævområdet til venstre), og når der observeres en for stor værdi (en værdi i skævområdet til højre).

Det var rimeligt at bruge en sådan dobbeltsidet test da vi afprøvede spillemaskinen. Vi var jo her på vagt over for både små gevinstsandsynligheder og store gevinstsandsynligheder.

Ofte er man i testsituationer kun interesseret i at sikre sig mod afvigelser fra den opstillede hypotese som ligger i én bestemt retning. Man kan da opstille en ***enkeltsidet test***. Vi ser på et par eksempler.

Test med venstreområde som godtagelsesområde

I en produktion påstås højst 20% af de producerede enheder at være defekte. Opstil en test til belysning af spørgsmålet. Lad os tænke os at vi udtager en stikprøve på 50 vareenheder, og at vi fastlægger hvor mange af disse 50 enheder der er defekte. Ved denne test vil vi ikke være bekymrede hvis der i stikprøven er meget få defekte enheder (eller måske slet ingen). Vi er alene på vagt over for at der skulle være for mange defekte enheder i stikprøven.

Vi opstiller derfor her en test med et godtagelsesområde som strækker sig fra 0 og op til en sådan værdi at der er 90% chance for at testen giver et resultat i godtagelsesområdet når den givne hypotese er sand. Altså når der kun er 20% defekte enheder i produktionen.

Nu kan vi lade Testgraf overtage beregningerne. Vi indtaster:

Hypotese: Andel af røde kugler:	0.20
Antal kugler der udtages:	50
Chance for at sand hypotese godtages:	0.90

Programmet beregner nu følgende tre områder:

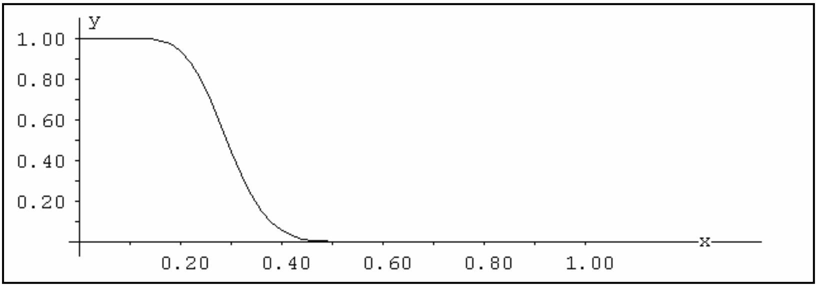
1. 1. område (midter) : 6..15

2. område (venstre) : 0..14

3. område (højre) : 6..50

Denne gang er vi interesseret i godtagelsesområde 2, venstreområdet. Det omfatter værdierne fra 0 til 14. Det betyder at vi vil godtage hypotesen (om en produktion med højst 20% defekte enheder) når antallet af defekte enheder i stikprøven er højst 14. Får vi mere end 14 defekte enheder, forkaster vi den opstillede hypotese. Vi har her en test hvor godtagelsesom-rådet er et venstreområde.

Vi lader nu programmet tegne testgrafen:



Vi ser at ved små x-værdier er der stor chance for at testen vil føre til godtagelse af den opstillede hypotese:

x= 0.10	y= 0.9999
x= 0.20	y= 0.9393

Men for større x-værdier falder chancen for godtagelse. For eksempel kan vi lade programmet beregne:

x=0.30	y=0.447
x=0.40	y=0.054
x=0.50	y=0.001

Når den rigtige fejlprocent i produktionen er 30%, er der altså en risiko på ca. 45% for fejl af 2. art, altså en risiko på 45% for at testen anerkender produktionen som værende god nok.

Når fejlprocenten i produktionen stiger til 40%, falder risikoen for fejl af 2. art til ca. 5%. Og når fejlprocenten er helt oppe på 50%, er risikoen for fejl af 2. art kun 0,1%.

Test med højreområde som godtagelsesområde

I et spil garanteres en gevinstsandsynlighed på mindst 40%. En deltager der har mistillid til denne angivelse, ønsker at undersøge spillet ved hjælp af en test.

I denne situation er spilleren på vagt over for et lille antal af gevinster. Derimod giver det ikke anledning til bekymring hvis antallet af gevinster skulle blive stort. Spilleren ønsker derfor en test med et godtagelsesområde som er et højreområde. Lad os opstille en test hvor der udføres 50 spil.

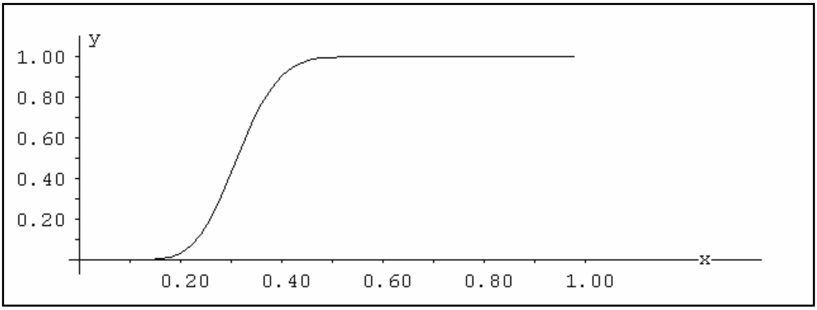
I Testgraf indtaster vi:

Hypotese: Andel af røde kugler:	0.40
Antal kugler der udtages:	50
Chance for at sand hypotese godtages:	0.90

Programmet beregner herefter følgende tre områder:

1. område (midter): 14..26
2. område (venstre): 0..24
3. område (højre): 16..50

Vi vælger her godtagelsesområde 3, højreområdet. Det omfatter værdierne fra 16 til 50. Det betyder at vi godtager gevinstgarantien på mindst 40% hvis antallet af gevinster i de 50 spil bliver 16 eller derover.



Vi lader Testgraf tegne acceptkurven for område 3: Vi ser at for små værdier af x er der kun lille chance for at testen fører til godtagelse af den opstillede hypotese. Vi beregner fx:

x=0.20 y=0.031

x=0.25 y=0.163

x=0.30 y=0.431

x=0.40 y=0.904

Når den rigtige gevinstsandsynlighed er 30%, er der en risiko på ca. 43% for at testen godkender spillet. Men når gevinstsandsynligheden er på 25%, er risikoen for fejl af 2. art kun ca. 16%. - Af tabellen ser vi at når den opstillede hypotese er sand, dvs. når gevinstsandsynligheden er 40%, så er der over 90% chance for at testen fører til godtagelse af hypotesen.

Vi har her set to eksempler på enkeltsidede test: En test med et venstreområde som godtagelsesområde og en test med et højreområde som godtagelsesområde.

Valget mellem en enkeltsidet test og en dobbelt-sidet test er bestemt af den foreliggende opgave.

Såfremt vi ved testen alene er på vagt over for stikprøveresultater som "ligger for langt ude i den ene retning", skal der vælges en enkeltsidet test. Vil man derimod sikre sig mod afvigelser der kan gå i begge retninger, må en dobbeltsidet test anvendes, altså en test med et midterområde som godtagelsesområde.

11. Binomialtest som værktøj

Ved fastlæggelse af en binomialtest kan du gå frem på følgende måde:

1. En hypotese opstilles.

2. Stikprøvens størrelse vælges.

3. Testens signifikansniveau fastsættes.

4. Godtagelsesområdet beregnes.

De tre første punkter er overladt til din afgørelse. Punkt 4 kan Testgraf klare for dig, men også her må du tage en beslutning om hvilken type af godtagelsesområde du ønsker:

midterområde , venstreområde, højreområde.

Når testgrafen foreligger, kan du derefter undersøge om den opstillede test opfylder dine ønsker med hensyn til risikoer for fejl af 2. art. Hvis testen ikke er god nok, kan du gå tilbage til punkt 2 og vælge en stærkere test, dvs. en test med et mere omfattende testeksperiment. Det er helt overladt til dig at afgøre om en forelagt test er for grov, eller om den har en tilstrækkelig følsomhed.

Brug værktøjet med omtanke

Som du vil forstå er en binomialtest ikke et sikkert værktøj der fejlfrit fører frem til den rigtige afgørelse om den opstillede hypotese. Det kan ske at testen fører til forkastelse af en sand hypotese, og det kan ske at testen fører til godtagelse af en falsk hypotese. Der er risiko for fejl af 1. art, og der er risiko for fejl af 2. art.

Medens risikoen for fejl af 1. art kan begrænses ved valg af et passende godtagelsesområde, kan risikoen for fejl af 2. art ikke kontrolleres helt så let. Den kan dog - som du har set - nedsættes hvis testens omfang forøges. Men uanset hvilken binomialtest der benyttes, vil der altid være en risiko for forkerte afgørelser.

Vi må derfor være forsigtige med de konklusioner vi drager af en udført test. Lettest er det at fortolke en situation hvor testen har ført til forkastelse af den opstillede hypotese. Hvis testen har et signifikansniveau på 5%, kan vi sige at der foreligger et testresultat som kun har en sandsynlighed på 5% hvis den opstillede hypotese er sand. Dette vil vi i fagsprog udtrykke ved at sige at vi har **forkastet hypotesen på 5%-niveauet.**.

Nogle vil i denne situation sige at "*de er 95% sikre på at hypotesen er falsk*". Denne angivelse af en overbevisning på 95% har intet med statistisk sandsynlighed at gøre, det er alene et subjektivt udtryk for graden af tiltro til den foretagne beslutning.

Du må hele tiden huske på at en hypotesetest ikke giver dig svar på om en opstillet hypotese er sand eller om den er falsk. Ej heller kan den svare på hvad chancen er for at hypotesen er sand. Det eneste spørgsmål du kan få svar på er:

Hvis den opstillede hypotese er sand, hvad er da risikoen for at testen vil føre til forkastelse af hypotesen?

Hvis testen ikke fører til forkastelse af den opstillede hypotese, må vi nøjes med at sige at testeksperimentet ikke har givet så klare signaler at vi er i stand til at forkaste hypotesen. *Hypote-*

sen er derfor ***godkendt indtil videre***, måske indtil ny information i form af nye observationer foreligger.

Godtagelse af den opstillede hypotese kan derfor siges at svare til afgørelsen "ikke skyldig" i retssalen. Her er princippet jo at den anklagede anses for at være ”ikke skyldig” indtil det modsatte er sandsynliggjort. I denne sammenhæng betyder en fejl af 1. art en fejlagtig domfældelse, medens en fejl af 2. art betyder en fejlagtig frifindelse.

Når testeksperimentet giver et resultat i godtagelsesområdet, vil vi altså **ikke** konkludere at hypotesen er sand:

En godtaget hypotese er ikke nødvendigvis en sand hypotese, men måske blot en hypotese vi endnu ikke har haft held til at afsløre som falsk.

Ligeledes er en forkastet hypotese ikke nødvendigvis en falsk hypotese. Der er altid en risiko for at testen har ført til et "statistisk justitsmord".

Når vi benytter en test med et signifikansniveau på 5%, ved vi at der højst er 5% risiko for at testen fører til forkastelse af en hypotese der er sand. De 5% er en risiko som er knyttet til hypotesetestnings-metoden: I det lange løb - altså ved mange anvendelser af testen - kan vi regne med at vi i højst 5% af de tilfælde *hvor hypotesen er sand* vil komme til at forkaste den. Når testen anvendes på sande hypoteser, vil den altså i 95% af tilfældene føre til godtagelse af hypotesen.

Der er intet i vejen for at vi bruger disse 95% som et udtryk for den tillid vi har til en afgørelse der er foretaget med testen. Vi skal blot gøre os klart at de 95% knytter sig til *den metode* vi gør brug af, ikke til den forelagte specielle anvendelsessituation. – Det var den samme fortolkning vi benyttede i forbindelse med beregning af konfidensområder.

Når der foreligger en observation der fører til **forkastelse** af den opstillede hypotese, vil man ofte kunne høre: "Jeg er 95% sikker på at testen har ført til den rigtige afgørelse", eller "Der er højst 5% chance for at hypotesen er sand ". Men det korrekte vil være at sige: "Der er højst 5% chance for at få et så skævt resultat når hypotesen er sand". Om hypotesen er sand eller falsk, kan vi ikke vide. Og vi kan ikke sætte chancer på.

Når testen fører til **godtagelse** af hypotesen, kan tilsvarende udsagn ikke fremsættes. I dette tilfælde vil du ikke med nogen ret kunne sige at du er 95% sikker på at hypotesen er sand. Selv når hypotesen er falsk, kan der jo være stor risiko for at testen fører til godtagelse af hypotesen. Men hvis du kender testens testgraf kan du dog give nogle udtalelser om sikkerhed i de situationer hvor testen fører til godtagelse af hypotesen:

Af testgrafen kan du måske aflæse at for x = 0.50 er der ca. 25% risiko for at den anvendte test fører til godtagelse af den opstillede hypotese om at den rette sandsynlighed er 0.30. Hvis du anvender testen, og den fører til godtagelse af hypotesen, kan du derfor sige at du er 75% sikker på at hypotesen ikke er "så falsk" at den rigtige x-værdi er helt oppe på 0.50.

Man kan undertiden se følgende tommelfingeregler fremsat om troværdigheden af de afgørelser der træffes på grundlag af test af hypoteser:

> Når en hypotese **godtages**, er det bedst at det er sket ved en **følsom** test, dvs. ved en test med et stort omfang.
>
> Når en hypotese **forkastes**, er det bedst at det er sket ved en **grov** test, dvs. ved en test med et lille omfang.

Begrundelsen for disse regler er følgende: En test med et stort omfang er følsom over for små afvigelser mellem model og observation. Når en sådan test godtager hypotesen, og altså ikke finder nogen signifikant afvigelse, så er der god grund til at tro at godtagelsen er den rigtige beslutning.

En test med et lille omfang er ikke så følsom over for afvigelser mellem model og observation. Når en sådan test forkaster hypotesen, og altså finder en signifikant afvigelse, så er der god grund til at tro at forkastelse er den rigtige beslutning.

12. Opgaver til Testgraf

1. En mønt testes

Hypotesen at sandsynligheden for Krone er 0.5 skal testes ved en test med et 90%-midterområde som godtagelsesområde. Testen skal have et omfang på 50.

Fastlæg testen ved hjælp af Testgraf, og beregn risikoen for fejl af 2. art når den rigtige kronesandsynlighed er:

(1) 0.55 (2) 0.60 (3) 0.65

2. En terning testes

Hypotesen at sandsynligheden for sekser er 1/6 skal testes ved en test med et 90%-midterområde som godtagelsesområde. Testen skal have et omfang på 100.

Fastlæg testen ved hjælp af Testgraf, og beregn risikoen for fejl af 2. art når den rigtige seksersandsynlighed er

(1) 0.20 (2) 0.25 (3) 0.10.

3. Ligevægtede udfald?

Opstil en test til undersøgelse af om de to mulige udfald ved kast med en tegnestift kan betragtes som ligevægtede. Testen skal have et 90%-midterområde som godtagelsesområde og et omfang på 25. - Udfør derefter test-eksperimentet ved kast med en tegnestift.

4. Test af LOD

Opstil en test til undersøgelse af om INFA-programmet LOD kan antages at frembringe cifferet 7 med en sandsynlighed på 0.1. Lad testen have et omfang på mindst 100 og benyt et 90%-midterområde som godtagelsesområde.

Udfør dernæst test-eksperimentet: Lad LOD udskrive det nødvendige antal tal fra talområdet 0..9.

5. En svag test

Opstil en test for hypotesen p=0.40 med et 90%-midterområde som godtagelsesområde og med et omfang på 10.

Undersøg hvor meget testen forbedres når testomfanget sættes op til 20. Sammenlign de to testgrafer.

6. En ynkelig test

Malene ønsker at kontrollere en mønts kronesandsynlighed ved at udføre 5 kast med mønten.

Opstil en dobbeltsidet test med signifikansniveau 5%, og undersøg hvad testens godtagelsesområde er.

Hvad er risikoen for fejl af 1. art ved den opstillede test, og hvad er risikoen for fejl af 2. art når kronesandsynligheden kun er 0.25?

7. En mesterskytte

Opstil en test til undersøgelse af mesterskyttens påstand om at hans træfsikkerhed er mindst 80%.
Benyt en test med signifikansniveau 10% og med et omfang på 25. Angiv risikoen for at mesterskyttens påstand godtages selv om hans træfsikkerhed kun er 60%.

Opstil derefter en test med et omfang på 50, og beregn også her risikoen for at hans påstand godtages selv om træfsikkerheden kun er 60%.

8. For små karakterer?

Ved en afgangsprøve er der normalt 20% der får en karakter på 6 eller derunder. I en klasse med 25 elever viser det sig at der er 8 elever som får 6 eller derunder. Giver dette grund til at tro at prøven har været for svær? - Besvar spørgsmålet ved hjælp af en test med et venstreområde som godtagelses-område.

9. Er kvaliteten i orden?

I en produktion er normalt 60% af de producerede enheder af 1. kvalitet. I en stikprøve på 50 enheder findes kun 25 enheder af 1. kvalitet. Giver dette grund til bekymring over produktionens kvalitet? Benyt en test med et højreområde som godtagelsesområde.

10. En kvalitetskontrol

Ved en kvalitetskontrol undersøges om produktionen højst indeholder 10% defekte enheder.

Hvor mange defekte enheder må der være i en stikprøve på 50 enheder for at produktionen kan godtages?

Hvor mange defekte enheder må der være i en stikprøve på 100 enheder?

11. Er terningen god nok?

Du kaster en terning 3000 gange og opnår 460 seksere. Giver dette dig anledning til at tvivle på terningens ægthed?

12. Hvilket signifikansniveau?

Du kaster en mønt 2000 gange og opnår 949 kronekast. Dette får dig til at forkaste den hypotese at møntens sandsynlighed for krone er 0.5.

Hvilket signifikansniveau skal du have testet på for at kunne træffe den beslutning når der er tale om en dobbeltsidet test?

13. Hvor mange billedmærker?

Sanne får tilbudt en sæk med brugte frimærker. Det påstås at mindst 40% af frimærkerne i sækken er billedmærker. Sanne udtager en håndfuld på 150 frimærker, hvoriblandt der viser sig at være 50 billedmærker. Kan hun tro på sælgerens påstand om 40% billedmærker i sækken?

14. Skal dvd'en produceres?

En dvd-klub påtænker at udsende en ny dvd, men ønsker først ved en stikprøveundersøgelse at sikre sig at der er tilstrækkelig interesse for pladen blandt klubbens medlemmer.

Opstil en test til undersøgelse af om der er mindst 70% af medlemmerne der vil købe dvd'en. Testen skal have et signifikansniveau på 5% og et omfang på 50.

Hvilken risiko er der for at testen godtager hypotesen når kun 50% af klubbens medlemmer vil købe dvd'en?

Opstil dernæst en test med et omfang på 100. Hvilken risiko er der nu for fejl af 2. art når kun 50% af klubbens medlemmer vil købe dvd'en?

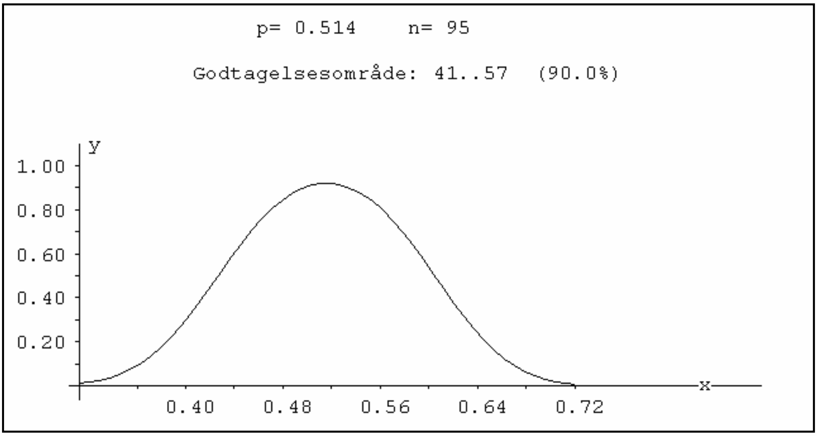
15. Sælgers og købers risiko

Producenten af en vare påstår at højst 25% af varerne er af 2. sortering. Opstil en test til undersøgelse af hans påstand. Testen skal have et omfang på 50, og testens signifikansniveau må ikke overstige 10%.

Hvad er risikoen for at testen godtager et vareparti som indeholder 40% enheder af 2. sortering?

16. Drenge og piger

På et hospital fødes en måned 58 drenge og kun 37 piger. Den danske fødselsstatistik siger at her i landet udgør drengene 51.4% af alle nyfødte. Giv en statistisk kommentar til hospitalets tal.



17. En elev med særlige evner?

En lærer udfører følgende eksperiment med en klasse. Der udføres 50 møntkast. Forud for hvert kast nedskriver hver elev sit gæt på om kastet giver Krone eller Plat. Blandt klassens 20 elever er der én der har opnået 32 rigtige gæt. Af Testgraf fremgår at dette resultat fører til forkastelse af hypotesen om en gættesandsynlighed på 50%, når der benyttes en dobbelt-sidet test med et signifikansniveau på 10%.

Giver eksperimentet grundlag for at tillægge den pågældende elev særlige gætteevner?

18. En terning testes

Ved 500 kast med en terning testes at seksersandsynligheden er 1/6. Opstil dobbeltsidede test med følgende signifikansniveauer:

(1) 10% (2) 5% (3) 1%

Beregn for hver af de tre test risikoen for fejl af 2. art når seksersandsynligheden kun er 0.12.

19. Er kvaliteten faldet?

En produktion indeholder normalt mindst 60% enheder af 1. kvalitet. Opstil en test til undersøgelse af om produktion er blevet forringet. Testen skal have et signifikansniveau på 10%. Opstil enkeltsidede test med omfanget

(1) 50 (2) 100 (3) 250

og beregn for hver af de tre test risikoen for fejl af 2. art når procentdelen af enheder af 1. kvalitet er gået ned på 50%.

20. Endnu en kvalitetskontrol

Ved en kvalitetskontrol udtages stikprøver på 80 enheder til kontrol af at andelen af defekte enheder ikke overstiger 10%. Som godtagelsesområde for testen anvendes venstreområdet 0..10.

Beregn testens signifikansniveau og dens risiko for fejl af 2. art når andelen af defekte enheder er 20%.

21. En annoncekampagne

Et marketingfirma får til opgave at gøre en bestemt målgruppe i befolkningen opmærksom på et nyt produkt. Hvis det lykkes firmaet at gøre mere end 40% af målgruppen bekendt med produktet udbetales en præmie fra producenten.

Producenten kontrollerer firmaets indsats gennem en stikprøveundersøgelse der omfatter 200 tilfældigt udvalgte personer fra målgruppen.

Opstil en test som kan fortælle producenten om marketingfirmaet har fortjent præmien. Testen skal have et sådant godtagelsesområde at der kun må være 1% risiko for at præmien udbetales uberettiget.

Hvor stor er risikoen ved den opstillede test for at præmien ikke kommer til udbetaling selv om marketingfirmaet har gjort det så godt at 50% af målgruppen er bekendt med det nye produkt?

Undersøg også hvad denne risiko ville være hvis testen havde et omfang på 500 i stedet for 200.

BIN-Test

13. Den optimale test

Du har set hvordan Testgraf kan fastlægge en test efter opstillede krav. Du oplyser hvilken hypotese der skal testes og hvilket omfang testen skal have. Endvidere angiver du den sandsynlighed der skal være for at testen fører til godtagelse af hypotesen når den er sand.

Derefter giver Testgraf dig valget mellem tre godtagelsesområder og dermed tre test: En midterområde-test, en test med venstreområde og en test med højreområde. Hvilken du skal vælge afhænger af den problemstilling der foreligger. Men uanset dit valg, så vil testen have det forlangte signifikansniveau.

Ved hjælp af Testgraf-programmet kan du undersøge om den fundne test er stærk nok, dvs. om den er tilstrækkelig følsom til at afsløre om den opstillede hypotese er falsk. Hvis testen ikke er stærk nok, kan du forøge testens omfang indtil du finder en tilfredsstillende test. En test med et større omfang vil være stærkere og altså bedre til at skelne mellem sande og falske hypoteser.

Det betyder dog ikke at man altid opstiller test med store omfang. Det kan jo ske at testeksperimentet derved bliver så krævende med hensyn til tid og penge at det vil være praktisk umuligt at gennemføre det. I almindelighed er der derfor tale om en løsning hvor der vælges en test af et moderat omfang. Derved bliver udførelsen af testeksperimentet overkommelig, og testens risiko for fejl af 1. og 2. art bliver af tålelig størrelse.

Vi skal nu se hvordan computeren kan hjælpe os til at opstille en test som opfylder de krav vi stiller. I programmet INSTAT finder du et underprogram **BIN-Test**. Dette program kan fastlægge en test ud fra de krav vi stiller til testens risiko for fejl af 1. og 2. art.

Vi ser på et eksempel hvor vi igen benytter spillemaskinen fra de tidligere afsnit.

Vi ønsker at teste at maskinens gevinstsandsynlighed er 0.3.

Til testen stiller vi følgende krav:

(1) Når spillemaskinens gevinstsandsynlighed er 0.3, skal der være mindst 95% chance for at testen fører til godtagelse af spillemaskinen.

(2) Når spillemaskinens gevinstsandsynlighed kun er 0.2, må der højst være 10% chance for at testen fører til godtagelse af spillemaskinen.

(3) Når spillemaskinens gevinstsandsynlighed er oppe på 0.4, må der også højst være 10% chance for at testen fører til godtagelse af spillemaskinen.

De tre krav kan vi også formulere sådan:

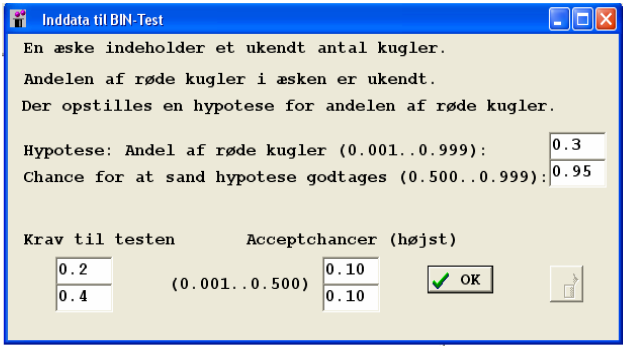
(1) Testens signifikansniveau skal være 5%, dvs. risikoen for fejl af 1. art må højst være 5%.

(1) Risikoen for fejl af 2. art må højst være 10% når den rigtige gevinstsandsynlighed er 0.2.

(2) Risikoen for fejl af 2. art må højst være 10% når den rigtige gevinstsandsynlighed er 0.4.

Som du ser stilles der ingen krav til testens omfang. Dette omfang vil blive beregnet af BIN-Test. Programmet finder også selv ud af hvilken type af godtagelsesområde der er tale om.

I BIN-Test indtaster vi nu følgende inddata:



Vi har i disse inddata oplyst at hypotesen siger at andelen af røde kugler er 0.3, dvs. at gevinstsandsynligheden er 0.3. Endvidere har vi fortalt at der skal være 95% chance for at testen godkender maskinen når den opstillede hypotese er sand. - Endelig fortæller vi at der er opstillet yderligere to krav til testen.

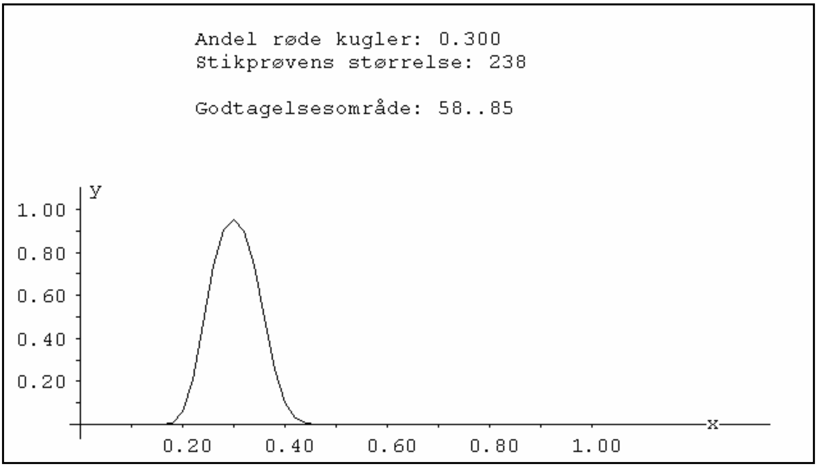
Vi indtaster her:

0.2 0.10

0.4 0.10

Disse tal indeholder jo de to krav vi har opstillet til testens risiko for fejl af 2. art.

BIN-Test leverer nu følgende skærmbillede:



Vi skal altså udføre 238 spil på maskinen, og hvis vi blandt disse spil får fra 58 til 85 gevinster, skal vi godtage maskinen. Får vi færre end 58 eller flere end 85 gevinster, skal maskinen ikke godtages.

Til kontrol har vi indtastet x-værdierne 0.3, 0.2 og 0.4. Vi får her følgende udskrifter:

x = 0.3   y = 0.953

x = 0.2   y = 0.057

x = 0.4   y = 0.099

Af disse tre aflæsninger ser vi at de stillede krav er opfyldt:

Testens signifikansniveau er 1 – 0.953 = 0.047, altså under 5%, og risikoen for fejl af 2. art ligger både for x = 0.2 og for x = 0.4 under 10%.

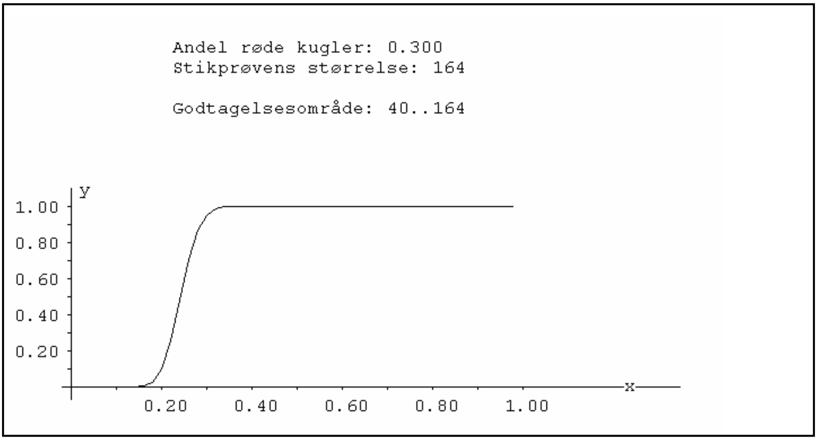
BIN-Test søger at fastlægge "den billigste test" der opfylder de stillede krav, dvs. den test som med det krævede signifikansniveau har mindst omfang og som opfylder de krav der er stillet til risikoen for fejl af 2. art.

Prøv selv

1. Undersøg hvad det betyder hvis du ændrer det andet krav til: Risikoen for fejl af 2. art må højst være 5% når den rigtige gevinstsandsynlighed er 0.4. Kontroller ved aflæsning på test-grafen at de stillede krav er opfyldt.

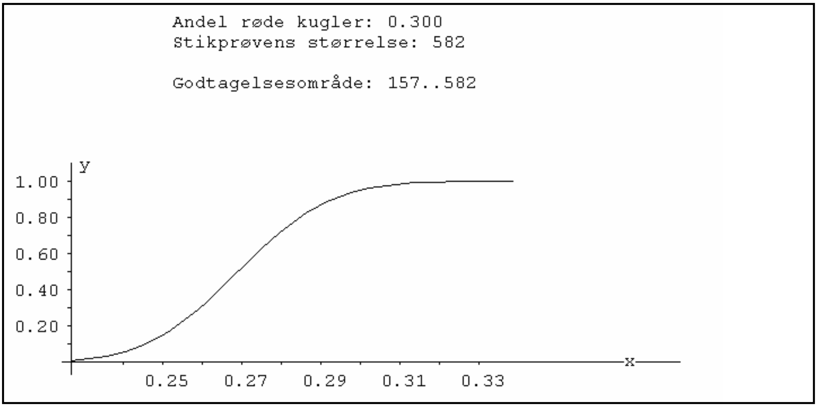
2. Du tester at en ternings seksersandsynlighed er 1/6. Du ønsker en test med et signifikansniveau på 5% og med en risiko for fejl af 2. art på højst 5% når terningens seksersandsynlighed er 0.15 og når den er 0.18. Lad BIN -Test fastlægge testen.

Hvis vi kun opstiller én betingelse vedrørende fejl af 2. art, vil programmet fastlægge en enkeltsidet test. Vi anvender eksemplet fra før, men stryger betingelsen der vedrører en gevinstsandsynlighed på 0.4. Figuren nedenfor viser den beregnede test.



Programmet melder at der skal udtages 164 kugler og godtagelsesområdet er beregnet til 40..164. Der er altså tale om en test med et højreområde som godtagelsesområde.

Hvis vi opstiller to betingelser som "ligger på samme side" af hypotesesandsynligheden, vil det også resultere i en enkeltsidet test. Vi kan fx kræve at sandsynligheden for fejl af 2.art er 0.10 ved en x-værdi på 0.20 og 0.15 ved en x-værdi på 0.25.



Begrænsninger i BIN-Test

BIN-Test fastlægger ikke test som har en stikprøvestørrelse under 10. Så selv om de stillede krav kunne opfyldes af en test med et mindre omfang end 10, vil BIN-Test angive en test med n=10. - Opadtil er der ikke nogen begrænsning i BIN-Test. Hvis du formulerer meget strenge krav til den ønskede test, vil du få beregnet et testomfang som vil være ganske urealistisk i praktisk sammenhæng.

Det vil blive belyst gennem opgaverne i næste afsnit.

14. Opgaver til BIN-Test

1. En mønt testes
Opstil ved hjælp af BIN-Test en test af kronesandsynligheden 0.5 for en mønt. Testens signifikansniveau skal være 5% og dens risiko for fejl af 2. art når møntens kronesandsynlighed er 0.4 eller 0.6 skal begge være 10%.

Lad BIN-Test tegne testgrafen og kontroller ved aflæsning på grafen at de stillede krav er opfyldt.

2. En terning testes
En spiller har mistillid til en terning. Han mener ikke at dens seksersandsynlighed er stor nok.

Opstil en test som med et signifikansniveau på 5% undersøger om terningens seksersandsynlighed er 1/6. Der skal være 90% chance for at testen afslører om terningens virkelige seksersandsynlighed er nede på 0.125.

3. Er der 25% eller 50% røde kugler?
I en krukke findes der røde og hvide kugler. Det oplyses, at krukken enten indeholder 25% røde kugler eller 50% røde kugler.

Opstil en test til afklaring af sagen. Testens risiko for fejl af 1. art og 2. art skal begge være 5%.

4. En medicinsk behandling
Ved en medicinsk behandling har der vist sig komplikationer i 40% af tilfældene. Med en ny teknik påstås det at der højst vil være komplikationer i 25% af tilfældene.

Opstil en test for hypotesen p=0.4. Testen skal have et signifikansniveau på 5% og dens risiko for fejl af 2. art skal være 10% når p har værdien 0.2.

5. Gætning

A påstår at han kan forudsige resultatet af et møntkast med 70% sikkerhed. B påstår at A kun kan gætte tilfældigt, dvs. forudsige resultatet med 50% sikkerhed.

Fastlæg en test til prøvning af A's påstand. Opstil hypotese og stil krav til risiko for fejl af 1. og 2. art.

6. Skal festen afholdes?

En forening ønsker at arrangere en fest for medlemmerne. Det udregnes at 60% af medlemmerne må deltage for at festen økonomisk kan svare sig. Foreningen er villig til at bære et underskud, dog må mindst 30% af medlemmerne deltage.

Ved forespørgsel hos medlemmer, der udvælges tilfældigt fra medlemslisten, danner man sig et skøn over den formodede mødeprocent.

Opstil en test som kan hjælpe foreningen i afgørelsen. Testen skal have et signifikansniveau på 10%, og dens risiko for fejl af 2. art må ikke overstige 2% når kun 30% af medlemmerne ønsker at deltage.

7. Kvalitetskontrol

En vare kvalitetskontrolleres ved stikprøveudtagelse. Producenten påstår at højst 10% af varerne er defekte, aftageren forlanger at indholdet af defekte enheder ikke må overstige 20%.

Opstil en test af at defektindholdet er 0.1. Testen skal have en risiko for fejl af 1. art (sælgers risiko) på højst 5%, og en risiko for fejl af 2. art (købers risiko) på højst 10% når defektindholdet er 0.2.

8. Skal dvd'en produceres?

En dvd-klub overvejer at udsende en ny dvd, men ønsker ved forudgående forespørgsler til tilfældigt udvalgte medlemmer at sikre sig, at der er tilstrækkelig interesse for serien, dvs. at mindst 40% af medlemmerne vil bestille dvd'en.

Opstil en test med et signifikansniveau på 5%, og med en risiko på højst 5% for at pladen produceres når kun 25% af medlemmerne ønsker at købe.

9. Et kongresarrangement

Et rejsebureau arrangerer udflugter i forbindelse med en kongres, der afholdes med deltagelse af 8000 udenlandske gæster. De foregående års erfaringer har vist at ca. 75% af deltagerne ønsker at være med i disse arrangementer.

Rejsebureauet ønsker nu ved forudgående forespørgsler at afgøre hvor stort behovet er, idet man på forhånd vil sikre transport og andre arrangementer til de lavest mulige priser.

Som hypotese antages at tilslutningsprocenten er mindst 70. Rejsebureauet kan tillade at tilslutningen går ned til 60%, medens yderligere nedgang ikke kan bæres økonomisk.

Opstil en test med signifikansniveau 5% og med en risiko på højst 2% for at hypotesen godtages, når den virkelige tilslutningsprocent kun er 60.

10. Hvor mange kast?

En mand med god tid til rådighed ønsker ved et eksperiment at kontrollere kronesandsynligheden for en mønt. Han ønsker en test med et signifikansniveau på 1% og med en risiko på højst 1% for at mønten godtages når kronesandsynligheden er 0.49 eller 0.51. - Lad BIN-Test opstille en test for ham.

11. Over eller under spærregrænsen?

Et politisk parti ønsker gennem en meningsmåling at undersøge om det ligger over spærregrænsen, dvs. om dets tilslutning blandt vælgerne er over 2%.

Opstil en test for den hypotese at tilslutningen til partiet er mindst 2%. Testen skal have følgende kvaliteter:

1. Såfrem tilslutningen til partiet er mindst 2% skal der være 90% sikkerhed for at testen godtager den opstillede hypotese.

2. Såfremt vælgertilslutningen til partiet er nede på 1.5% skal testen afsløre dette med 99% sikkerhed.

Undersøg hvordan testomfanget ændres hvis kravet under punkt 2 skærpes fra 1.5% til 1.75%.

Prøv endvidere at ændre procenten under punkt 2 fra 1.5% til 1.9%.

To eksempler på signifikanstest

Vi afslutter arbejdet med test af hypoteser med to eksempler på test. Vi ser her på en **Modeltest** og en **Forskelstest.** De to test vil give dig et indblik i hvordan du kan foretage statistiske beslutninger ud fra forelagte sæt af data.

Ved de to test betragter vi alene fejl af 1.art. Vi opstiller altså kun krav til testens signifikansniveau og omtaler derfor de to test som **signifikanstest**. Det betyder at vi har styr på risikoen for at testen fejlagtigt forkaster en sand hypotese. Derimod har vi ikke nogen talværdi for sandsynligheden for at en fejlagtig hypotese godtages af testen.

15. Modeltest: Stemmer data med modellen?

1. Er terningen god nok?

Sanne udfører et eksperiment der består i 60 kast med en terning. Hun fører regnskab med hvilke øjental der forekommer i de 60 kast. Her er hendes tabel over kasteresultaterne:

Øjental	1	2	3	4	5	6
Hyppighed	6	12	9	13	13	7

Hun er lidt forbavset over disse tal. Hun ved jo at der ved et kast med en terning er samme chance for hvert af de seks øjental, det undrer hende derfor at der er så stor forskel på hvor tit de enkelte øjental forekommer: Øjentallet 1 er kun forekommet 6 gange, men øjentallene 4 og 5 er begge forekommet 13 gange i de 60 kast.

Sanne spekulerer på om der kan være noget galt med terningen. Hvis den er skæv på en eller anden måde, så kan det jo være forklaringen på at tallene afviger så meget fra det forventede.

Hvis de seks øjental har samme chance for at forekomme, ville Sanne vente at alle de observerede hyppigheder vil ligge tæt på 10. Der er jo 60 kast, og hvis de skulle fordele sig ligeligt på de seks muligheder, så måtte der være 10 af hver slags.

Sanne arbejder her med en tænkt model, et billede af hvordan kastene ville fordele sig i en ideel virkelighed. Men hun ved godt fra tidligere eksperimenter med terningkast at virkeligheden kan afvige en del fra modellen. Kaster hun seks kast, er det jo ikke særligt ofte at det vil forekomme at hvert af de seks øjental forekommer lige præcis én gang (chancen herfor er kun 1.5%). Og i en serie på 60 kast er det selvfølgelig heller ikke at forvente at hvert øjental forekommer lige præcis 10 gange.

Men hvor meget må virkelighedens data nu afvige fra modellen før Sanne bør få mistanke om at der er noget galt?

Til at besvare dette spørgsmål gør vi brug af en statistisk test, *en modeltest*. Ved hjælp af den kan vi få oplysning om hvor godt tallene fra eksperimentet stemmer med den opstillede model.

Vi gentager her Sannes tabel over de opnåede resultater, men vi tilføjer de tal som vi ville forvente efter den opstillede model:

Øjental	1	2	3	4	5	6
Hyppighed	6	12	9	13	13	7
Model	10	10	10	10	10	10

Der er her tale om en model som giver samme chance for alle de foreliggende resultater. En sådan model kalder vi en ***jævn fordeling.***

Når vi skal vurdere hvor meget Sannes tal afviger fra den opstillede model, kigger vi på forskellen mellem de observerede antal og dem der forudsiges af modellen. Ved øjentallet 1 tager vi altså forskellen mellem 6 og 10, ved øjentallet 2 tager vi forskellen mellem 12 og 10, osv.

Disse forskelle kvadreres og adderes:

$$(6-10)^2 + (12-10)^2 + ... + (7-10)^2 = 48$$

Vi kan se at hvis Sanne havde fået tal der præcist passede med modellen, så ville alle tal i parenteserne være 0, og så ville den samlede sum også være 0.

Vi bruger denne sum af de kvadrerede tal som et udtryk for hvor meget de opnåede resultater afviger fra modellens tal. Jo større summen er, desto større er afvigelsen mellem de observerede data og modellens tal.

Vi anvender nu modeltesten på Sannes tal. Den foreligger som et edb-program, ***Modeltest***, så vi er fri for at foretage beregninger ved håndkraft.

Sanne fortæller først programmet at der er tale om seks forskellige resultattyper. Der åbnes nu seks felter på skærmen, og Sanne indtaster her de kasteantal hun opnåede ved eksperimentet: 6, 12, 9, 13, 13, 7.

Derefter fortæller hun at den model der skal testes er en ***jævn fordeling***. I modellinjen indsætter programmet nu tallet 10 i hver rubrik. Hun klikker herefter på "Udfør test".

Programmet giver herefter oplysninger om hvor godt de observerede data stemmer med en opstillede model:

***Testresultat*:**

Ingen signifikant afvigelse mellem model og data.

Når programmet melder "Ingen signifikant afvigelse" så betyder det at chancen for et resultat der er så skævt som det foreliggende, er over 10%. - Programmet melder således her at der intet påfaldende er ved de foreliggende data, og der er derfor ikke grund til at tvivle på den opstillede model.

Vi kan fortælle dig at der ved en kasteserie af den foreliggende art er ca. 45% chance for at få resultater som afviger lige så meget fra den jævne fordeling - eller endnu mere - end dem Sanne fik. Med andre ord: I næsten hver anden kasteserie kan der forventes en så stor afvigelse som den der var i Sannes kasteserie.

Modeltesten viser således at Sanne ikke har nogen grund til at mistænke sin terning for at være skæv. Afvigelser som den hun fik, er ganske almindelige.

Modellen forkastes

Sanne har mistanke til en terning som ser ud til at være lidt ekstra tung i den ene ende. Hun tester den ved en kasteserie på 60 kast. Her er hendes resultater:

Øjental	1	2	3	4	5	6
Hyppighed	16	5	9	7	6	17

Som model anvender Sanne en jævn fordeling.

En modeltest giver følgende resultat:

```
Modeltest: Antal forskellige talværdier: 6

Data:
    16       5       9       7       6      17

Model:
    10.0    10.0    10.0    10.0    10.0    10.0
```

Testresultat: (Modeltest)

Signifikant afvigelse mellem model og data: p= 1.84%

Ud fra dette konkluderer Sanne at der nok er tale om en skæv terning. Hvis terningen var ægte, ville der jo kun være 1.84% chance for at få så stor en afvigelse som den der findes i de givne data. Den opstillede model, en jævn fordeling, forkastes. Men nye data kan selvfølgelig ændre beslutningen.

Modeltestens beregninger

Vi kan forestille os at modeltesten beregner testchancerne på følgende måde: Programmet udfører 10000 eksperimenter af den foreliggende art (her 60 kast med en terning), og det opstiller en statistik over de afvigelser fra den givne model (her en jævn fordeling) der forekommer i de 10000 eksperimenter.

Hvis det fx viser sig at der er 500 eksperimenter af de 10000 der giver en afvigelse som er af mindst samme størrelse som den der er i de givne data, så fortæller programmet at der er en chance på 500/10000, dvs. 5% for at få en så stor afvigelse som den der foreligger.

Hvis der blev udført en ny serie på 10000 eksperimenter, kunne vi få lidt andre tal, men i praksis ville der ikke være store forskelle. Så vi opfatter modeltestens chancer som nogle der er fastlagt på grundlag af et stort antal af eksperimenter, der er altså tale om statistisk sandsynlighed.

Hvad kan vi opnå med en modeltest?

Når vi anvender en modeltest kan vi bruge den på følgende måde:

1. Hvis testen viser at afvigelsen mellem model og data er ***signifikant stor***, vil vi være på vagt over for den opstillede model: Vi er i tvivl om den opstillede model kan accepteres.

   Jo lavere testresultatets *signifikans-niveau* er, jo mere overbevist er vi om at den opstillede model ikke er den korrekte.

2. Hvis testen viser at der ***ingen signifikant afvigelse*** er mellem model og data, så har vi ingen grund til at tvivle på den opstillede model.

I tilfælde 1 kan vi måske ønske os nye data indsamlet før vi tager endelig stilling.

Prøv selv

Udfør en modeltest på følgende resultater fra 60 terningkast:

Øjental	1	2	3	4	5	6
Hyppighed	9	11	9	11	11	9

Hvad er din konklusion?

Andre modeller

Ved modeltesten kan vi også teste modeller der ikke bygger på en jævn fordeling. Lad os tænke os at vi ved terningkast slår øjentallene 1 og 6 sammen i én gruppe, og lader de andre øjental udgøre hver sin gruppe.

Med tallene fra en af Sannes kasteserier har vi da:

Øjental	1 og 6	2	3	4	5
Hyppighed	13	12	9	13	13
Model	20	10	10	10	10

Ved anvendelse af modeltesten skal vi først meddele at der er tale om 5 grupper af data. Dernæst indtaster vi de opnåede hyppigheder og dernæst de tal der er anført under Model. Modellen er jo denne gang *ikke* en jævn fordeling, der forventes ikke samme antal kast i de fem grupper: Første gruppe forventes at indeholde 20 kast, de øvrige fire hver 10 kast.

Modeltesten giver følgende resultat:

```
Modeltest: Antal forskellige talværdier: 5
```

Data:

```
     13      12       9      13      13
```

Model:

```
     20      10      10      10      10
```

Testresultat: (Modeltest)

Ingen signifikant afvigelse mellem model og data: p>10%

Modeltesten viser altså at der ikke er nogen statistisk grund til at forkaste den opstillede model.

Modellens tal må ikke være under 5

Ved opstilling af en model skal du indtaste modellens tal i rubrikkerne i nederste linje. Disse tal må gerne indtastes som decimaltal, de angiver jo de forventede gennemsnitstal i et stort antal forsøg og behøver derfor ikke at være hele tal.

Af hensyn til modeltestens beregninger bør tallene i **model-rubrikkerne** ikke have værdier under 5. Hvis du opstiller en model med en eller flere rubrikker som indeholder en talværdi under 5, vil programmet opfordre dig til at slå nogle grupper sammen, således at du i alle rubrikker kommer op på værdier der ligger på 5 og derover.

Opgaver til Modeltest

I arbejdet med opgaverne bør du indlede med at tage et overblik over de givne data og give dit gæt på om der foreligger en signifikant afvigelse mellem model og data. Først derefter anvender du modeltesten.

Når du bliver mere øvet, kan du prøve at gætte på hvor stærk en signifikans der vil være tale om ved det foreliggende datasæt.

1. Dit eget eksperiment

Udfør 60 kast med en terning og lav en tabel over forekomsten af øjental.

Undersøg dine data ved hjælp af en modeltest.

2. Gentag eksperimentet

Gentag eksperimentet fra opgave 1 og udfør også en modeltest på det nye datasæt.

Slå derefter de to datasæt fra opgave 1 og 2 sammen til ét sæt og udfør en modeltest på dette sæt.

3. Et opdigtet datasæt

Få en kammerat til at digte et datasæt som viser hvordan 60 terningkast fordeler sig på de seks mulige øjental.

Undersøg datasættet ved en modeltest.

4. En falsk terning

En terning er forfalsket således at der er en chance på 50% for at få en sekser ved et terningkast, medens der er en chance på 10% for hvert af de øvrige øjental.

Terningen undersøges i en kasteserie på 100 kast. Der opnås følgende øjental:

Øjental	1	2	3	4	5	6
Hyppighed	6	9	6	10	8	61

Undersøg om disse data er overensstemmelse med den givne model.

5. Data fra LOD

Lad programmet LOD udskrive 100 tal fra området 1..10.

Undersøg ved en modeltest om de givne resultater er i overensstemmelse med en jævn fordeling over de ti mulige talværdier.

6. Er fagene lige populære?

I en undersøgelse deltager 300 skoleelever. De bliver spurgt om hvilket skolefag de synes bedst om. Svarene fordeler sig på fem fag:

Dansk	65
Matematik	88
Idræt	41
Engelsk	54
Fysik/kemi	52

Undersøg ved en modeltest om disse data stemmer med at der ikke er nogen væsentlig forskel på fagenes popularitet.

7. At slå To ens øjental

Tag dine resultater fra et eksperiment hvor du har undersøgt hvor svært det er at slå *To ens øjental* med to terninger.

Undersøg ved en modeltest om dine resultater passer med følgende model:

Antal kast	1-2	3-5	6-10	over 10
Chance	30%	30%	24%	16%

8. En anden model

Undersøg ligesom i foregående opgave om dit datasæt fra undersøgelsen af hvor svært det er at slå *To ens øjental* passer med følgende model:

Antal kast	1-3	4-6	7-10	over 10
Chance	44%	26%	18%	12%

9. Kast med en tegnestift

Ved kast med en tegnestift kan der opnås to resultater:

Tegnestiften ender "på ryg" eller den ender "på spids".

Udfør mindst 50 kast med en tegnestift og undersøg ved en modeltest om det kan afvises at de to udfald har samme chance for at forekomme.

10. Er modellen god nok?

Når man kaster med to mønter, kan der forekomme følgende tre resultater:

Begge mønter viser krone
Begge mønter viser plat
En mønt viser krone, en viser plat

Udfør 60 kast med to mønter (brug små mønter i et raflebæger) og noter dine resultater i kasteserien.

Undersøg ved en modeltest om resultaterne stemmer med en jævn fordeling over de tre mulige resultater.

Udfør dernæst yderligere 60 kast og anvend en modeltest på det samlede datasæt fra de to kasteserier.

11. Tal fra virkeligheden

En optælling af cifferfordelingen i befolkningstallene for Danmarks amter giver følgende fordeling:

1: 6	2: 16	3: 10	4: 14	5: 8
6: 16	7: 10	8: 7	9: 4	

Cifferet 1 forekommer altså 6 gange, cifferet 2 forekommer 16 gange, osv. - Cifferet 0 er ikke taget med, da det jo ikke kan stå forrest i et tal. Det kan de øvrige cifre.

Undersøg ved en modeltest om der kan være tale om en jævn cifferfordeling for cifrene 1..9.

12. Fysikkens tal

En optælling af cifferfordelingen i atomvægtene for de første 12 grundstoffer i det periodiske system giver følgende fordeling:

1: 13	2: 8	3: 2	4: 5	5: 3
6: 3	7: 4	8: 5	9: 12	

Cifferet 0 er ikke taget med, da det jo ikke kan være første ciffer i et tal.
Undersøg ved en modeltest om der kan være tale om en jævn cifferfordeling.

13. En afvigelse på 10%

Undersøg ved en modeltest om der kan være tale om en jævn fordeling af krone og plat når der observeres en fordeling af krone- og platkast som ligger 10% fra det forventede, dvs. når der forekommer 40% krone og 60% plat:

Antal kast: 50	Krone: 20	Plat: 30
Antal kast: 100	Krone: 40	Plat: 60
Antal kast: 150	Krone: 60	Plat: 90
Antal kast: 200	Krone: 80	Plat: 120

Hvilken konklusion kan du give vedrørende testens følsomhed over for en afvigelse på en fast procentdel?

14. En afvigelse på 10

Undersøg ved en modeltest om der kan være tale om en jævn fordeling af krone og plat når der observeres en fordeling af krone- og platkast som ligger 10 fra det forventede, dvs. når antallet af kronekast er 10 mindre end halvdelen af kasteantallet.

Antal kast: 50	Krone: 15	Plat: 35
Antal kast: 100	Krone: 40	Plat: 60
Antal kast: 150	Krone: 65	Plat: 85
Antal kast: 200	Krone: 90	Plat: 110

Hvilken konklusion kan du give vedrørende testens følsomhed over for en afvigelse på et fast antal?

15. Hvis skævheden forstørres

I 60 kast med en terning opnås følgende resultater:

Øjental	1	2	3	4	5	6
Hyppighed	7	13	9	13	12	6

Undersøg ved en modeltest om disse resultater er i overensstemmelse med en jævn fordeling.

Vi tænker os nu at der udføres 60 ekstra kast med præcis den samme fordeling: "Vi ganger de første resultater med 2". Vi har herefter 120 kast der fordeler sig således på de enkelte øjental:

Øjental	1	2	3	4	5	6
Hyppighed	14	26	18	26	24	12

Undersøg fordelingen ved en modeltest.

Prøv derefter at gange de første resultater med 3 og udfør en modeltest på denne fordeling.

16. Hjælper det med træning?

24 skoleelever deltager i en konditest. Derefter går de ind i en periode med træning for at forsøge at forbedre deres kondital, og efter træningsperioden bliver deres kondital målt igen.

Det viser sig 16 af de 24 elever har forbedret deres kondital, medens 8 elever får et ringere resultat end ved første prøve.

Anvend en modeltest på disse tal: Er der nogen signifikant afvigelse fra en jævn fordeling mellem bedre og ringere kondital?

16. Forskelstest:

Er der en væsentlig forskel mellem de to datasæt?

1. Et eksempel

Vi skal nu se på en test som kan hjælpe os med at vurdere om der er en væsentlig talmæssig forskel på to datasæt.

Vi ser på et eksempel:

Tre piger og fem drenge deltager i en matematikprøve. Deres point i prøven er følgende:

Pigerne: 15, 16, 30

Drengene: 18, 23, 27, 31, 33

Vi vil nu anvende en forskelstest på dette materiale. Den kræver ikke beregninger af gennemsnit eller andre statistiske mærketal. Den bygger alene på *en* ***rangorden*** af de opnåede pointtal.

Her er alle pointtal for de 8 elever opstillet efter størrelse:

15	16	18	23	27	30	31	33
P	P	D	D	D	P	D	D

I tilknytning til tallene har vi anført om pointtallet er opnået af en dreng eller en pige.

Et overblik over fordelingen af point på drenge og piger giver det indtryk at pigernes og drengenes placering i rækkefølgen ikke ser ud til at være ganske tilfældig. Det ser ud som om der er en skævhed med størst koncentration af pigerne på de lavere pointtal: "Pigerne klarer sig dårligere ved prøven end drengene."

Ved forskelstesten vil vi undersøge om der er statistisk grundlag for at fremsætte denne påstand.

Hvert af pointtallene giver vi nu det nummer, "den rang", det har i den opstillede rækkefølge:

15	16	18	23	27	30	31	33	Pointtal
P	P	D	D	D	P	D	D	
1	2	3	4	5	6	7	8	Rang

Herefter ser vi ikke mere på de foreliggende pointtal, men udelukkende på rangtallene, tallene i den nederste linje.

Vi ser at pigernes pointtal har rangene 1, 2 og 6. Disse tal har summen 9. Det vil sige at pigernes ***rangsum*** i den forelagte rækkefølge er 9.

Det er dette tal vi vil se nærmere på i forskelstesten. Vi vil undersøge om denne rangsum på 9 kan siges at være udtryk for en skævhed. En skævhed som er så påfaldende, så *signifikant*, at den får os til at forkaste den hypotese at pigernes og drengenes placering i rangordenen er tilfældig.

Hvis det er rigtigt at pigerne klarer sig dårligere ved prøven end drengene, vil vi forvente at pigernes pointtal vil findes i den nedre ende af opstillingen, vi vil altså vente at finde at pigernes rangsum er "for lille".

Oversigt over rangsummerne

Testen udfører vi ved at sammenligne den observerede rangsum på 9 med alle de rangsummer der kan forekomme for de tre pigers pointtal.

Den laveste rangsum der kunne forekomme for de tre piger, er 6. Den forekommer jo hvis de tre piger er placeret på de tre første pladser:

Placering: 1, 2, 3   Rangsum: 1+2+3 = 6.

Den højeste rangsum der kan forekomme er 21. Den forekommer hvis de tre piger er placeret på de tre øverste pladser:

Placering: 6, 7, 8   Rangsum: 6+7+8 = 21

Rangsummerne kan altså i det forelagte tilfælde variere fra 6 op til 21:

6____21

Den rangsum pigerne opnåede ved prøven var 9. Spørgsmålet er nu hvor skævt en rangsum på 9 er placeret i forhold til alle mulige rangsummer.

Vi laver en lille oversigt over rangsummerne fra 6 op til 9:

Rangsum 6: 1+2+3

Rangsum 7: 1+2+4

Rangsum 8: 1+2+5, 1+3+4

Rangsum 9: 1+2+6, 1+3+5, 2+3+4

Ved optælling kan vi se at der er 7 tilfælde hvor rangsummen er på 6, 7, 8 eller 9.

Vi har altså at der ligger 7 rangsummer i området fra 6 op til 9:

6_______9________________________________21

Spørgsmålet er nu: Hvor stor en del udgør de 7 af alle de rangsummer der findes i det samlede felt fra 6 op til 21.

Dette antal kan bestemmes ved en tællemetode: På hvor mange måder kan der udvælges 3 tal blandt 8. Hver gang vi opstiller en rangsum tager vi jo tre af tallene fra talmængden 1,2,3, 4,5,6,7,8.

Det søgte antal er givet ved udtrykket:

$$K(8,3) = \frac{8*7*6}{1*2*3} = 56.$$

Der er altså i alt 56 mulige måder at beregne rangsummen på i det forelagte eksempel.

Hver af disse 56 kombinationer af pigernes range tillægger vi nu samme sandsynlighed. Hvis vi antager at piger og drenge er lige dygtige til prøven, så vil ethvert sæt af range for de tre piger have samme chance for at forekomme, hvad enten det er rangene (1,2,3), rangene (2,4,6) eller fx rangene (6,7,8). Så hvis den opstillede hypotese er sand, vil enhver af de 56 mulige rangsæt have sandsynligheden 1/56.

De 7 muligheder i området fra rangsum 6 til rangsum 9 skal nu sammenlignes med det samlede antal på 56.

Vi kan derfor sige: Chancen for at få en rangsum i området fra 6 til 9 er 7/56 = 1/8 = 12.5%. Testen har dermed beregnet en testchance på 12.5%, dvs. der er 12.5% chance for at få en rangsum som er så lille som den foreliggende eller endnu mindre.

En testchance på 12.5% er almindeligvis ikke noget der giver grund til forkastelse af en opstillet hypotese. Vi vil derfor konkludere at der ikke er opnået så skævt et resultat at der er grund til at forkaste den opstillede hypotese:

```
Forskelstest: (Enkeltdata)

A :      1      2      6

B :      3      4      5      7      8

Testresultat:
```

Ingen signifikant afvigelse mellem de to datasæt A og B: p>10%

Tallene giver os altså ikke grundlag for at hævde at prøven viser at drengene er bedre end pigerne.

Hvis pigerne i stedet havde været placeret som nr. 1, 2 og 4, så havde deres rangsum været 7. Af oversigten ser vi at der kun er to måder hvorpå vi kan få en rangsum på 7 eller der under. Chancen for en rangsum i området fra 6 til 7 er derfor: 2/56 = 3.57%.

Denne chance på 3.57% ville måske få os til at sige at den opstillede hypotese skulle forkastes. Vi kan derfor i denne situation konkludere: Prøveresultatet tyder på at drenge og piger ikke klarer sig lige godt ved prøven. Da pigernes resultat ligger i det venstre område, "deres rangsum er for lille", kan vi mere tydeligt sige: *Det ser ud til at pigerne klarer sig dårligere end drengene*.

Testen ville i dette tilfælde føre til resultatet:

```
Forskelstest: (Enkeltdata)


A :      1      2      4

B :      3      6      5      7      8

Testresultat:
```

B har signifikant højere range end A: p= 3.57%

Det betyder at der er mindre end 4% chance for at et så godt resultat for drengene kan forekomme når der foretages en tilfældig fordeling af range til de to grupper.

Vi anvender kun rangene

Den opstillede forskelstest kaldes også en ***rangsumtest***. Den bygger jo sine beregninger på de rangsummer der kan forekomme.

Som du har set, benytter en rangsumtest kun rangene for de forelagte data, derimod ikke talværdierne. Det gør rangsumtesten til et lettilgængeligt værktøj. Men dette er selvfølgelig ikke uden konsekvenser. Ved kun at benytte rangene for de forelagte data, går en række informationer tabt, idet der jo ofte ligger væsentlige oplysninger om de givne data i deres talværdier.

Der findes da også forskelstest som gør brug af de forelagte talværdier og ikke kun af deres range. Nogle af disse test har imidlertid en række forudsætninger som ikke altid kan opfyldes, og det kan i øvrigt være svært at afgøre om de pågældende forudsætninger er opfyldt eller ej.

Man vil derfor ofte stille sig tilfreds med at benytte en rangsumtest som alene forudsætter at der foreligger data der meningsfyldt kan opstilles i en rangorden.

Programmet Forskelstest

Vi skal nu se hvordan programmet *Forskelstest* kan benyttes ved undersøgelse af om der foreligger signifikante forskelle mellem to datasæt.

I programmet åbner der sig en menu med følgende muligheder:

Enkeltdata
Gruppedata

Vi vælger punktet *Enkeltdata*. Programmet beder nu om oplysninger om hvor mange data der er i de to sæt. I vort eksempel består det ene sæt af data for pigerne, det andet sæt af data for drengene.

Vi svarer derfor at det ene sæt består af 3 data og det andet af 5 data. Herefter spørger programmet hvilke data der foreligger for de to sæt. Vi svarer med først at indtaste tallene for pigerne: 15, 16 og 30. Derefter indtaster vi tallene for drengene: 18, 23, 27, 31 og 33.

Vi går nu til *Udfør test*. Her får vi følgende oversigt på skærmen:

```
Forskelstest: (Enkeltdata)

A :      15      16      30

B :      18      23      27      31      33

Testresultat:
```

Ingen signifikant afvigelse mellem de to datasæt A og B: p>10%

Vi anvender herefter programmet på den situation at pigernes pointtal ligger som nr. 1, 2 og 4 i rangfølgen. Vi kan fx benytte pointtallene 15, 16 og 20 for pigerne, medens vi for drengene bruger de samme pointtal som før.

Vi får da følgende udskrift fra programmet:

```
Forskelstest: (Enkeltdata)

A :      15      16      20

B :      18      23      27      31      33

Testresultat:
```

B har signifikant højere range end A: p= 3.57%

Vi kan så overveje om vi ud fra dette tør konkludere at pigerne klarer sig dårligere ved prøven end drengene. Men nu har vi i hvert fald et statistisk grundlag for at fremsætte en sådan påstand.

Programmet benytter betegnelserne A og B for de to datasæt. På skærmen kan du se hvilke datasæt A og B står for.

En forskelstest på de to datasæt A og B kan føre til tre mulige resultater:

Ingen signifikant forskel mellem A og B

A har signifikant højere range end B

B har signifikant højere range end A

Prøv selv

1. Lad programmet foretage en beregning for den situation hvor de tre piger havde opnået pointtallene 19, 20 og 25, medens drengene har de samme pointtal som før.

2. Lad programmet foretage en beregning for den situation hvor de tre piger havde opnået rangene 3, 7 og 8. Digt selv nogle pointtal der passer til disse range. Hvilken konklusion ville du give i dette tilfælde om drengenes og pigernes dygtighed?

Sammenfald af talværdier

Lad os antage at pointtallene for de tre piger er 15, 17 og 30, og drengenes pointtal er 17, 23, 27, 31 og 33. I dette tilfælde er der pointsammenfald, idet pointtallet 17 er opnået af to elever.

I en sådan situation fordeler man rangene således at de point der har pointsammenfald deler de pågældende range:

15	17	17	23	27	30	31	33
P	P	D	D	D	P	D	D
1	2.5	2.5	4	5	6	7	8

De to elever der begge har opnået 17 point har delt rangene 2 og 3 og derfor begge fået 2.5 som rang.

Rangsummen for pigerne er i denne situation: 1 + 2.5 + 6 = 9.5.

Prøv selv
Afprøv programmet: Vælg *Enkeltdata* og indtast de 8 pointtal der er nævnt ovenfor.

*

Programmet kan klare alle de pointsammenfald der måtte forekomme. Hver gang fordeles de pågældende range mellem de elever der har pointsammenfald, og beregningerne foregår derefter på samme måde som i eksemplet vi lige har set på.

Prøv selv
Afprøv en situation med nogle flere pointsammenfald, og lad programmet udføre forskelstesten.

Gruppedata
Under dette menupunkt gemmer sig en mulighed for en let indtastning af data som foreligger med en stor mængde af pointsammenfald.

Lad os se på et eksempel hvor 40 elever deltager i et forsøg med en ny undervisningsmetode i matematik. 40 andre elever udgør kontrolgruppen. I en afsluttende prøve inddeles besvarelserne i 4 kategorier:

Under middel
Middel
God
Særdeles god

Vi kan fx tildele talværdierne 1, 2, 3 og 4 til disse fire kategorier. Den afsluttende prøve giver følgende resultater:

	Kat. 1	Kat. 2	Kat. 3	Kat. 4
Forsøgsgruppen	5	7	16	12
Kontrolgruppen	7	9	15	9

Det kunne se ud som om forsøgsgruppen har klaret sig bedre end kontrolgruppen. Vi kan nu anvende en rangsumtest på disse data. Talværdien 1 er tildelt 5+7 = 12 personer. Det betyder at rangene 1-12 må deles mellem disse 12 personer. De får derfor hver rangen 6.5, nemlig (1+12)/2 = 6.5.

På tilsvarende måde tildeles range til de øvrige personer i skemaet.

Denne tildeling af range foretager programmet for os når vi går ind under menupunktet *Gruppedata*.

Her spørges først om hvor mange grupper der er tale om. I vort eksempel er der 4 grupper svarende til de fire kategorier ved prøven. Vi svarer derfor: 4.

Dernæst beder programmet om de antal der er placeret i hvert af skemaets felter. Når de er indtastet, ser skærmen således ud:

Antal i datasæt A	5	7	16	12
Antal i datasæt B	7	9	15	9

Derefter går vi til *Udfør test*. Her får vi nu følgende udskrift fra programmet:

```
Forskelstest: (Gruppedata, stigende
rangorden) Antal forskellige talværdier: 4

A :      5      7     16     12

B :      7      9     15      9
```

Testresultat:

Ingen signifikant forskel mellem A og B: p>10%

Vi kan nævne at testchancen kan beregnes til knap 17%. Det betyder at hvis det er rigtigt at de to grupper er lige dygtige, så er der ca. 17% chance for at opnå data med en skævhed som den foreliggende.

Nu vil en testchance på 17% ikke føre til forkastelse af den opstillede hypotese. Vi må derfor konstatere at der ikke i de opnåede resultater er tilstrækkeligt statistisk grundlag for at tro at forsøgsgruppen klarer sig bedre til prøven end kontrolgruppen. Med andre ord: Den nye undervisningsmetode giver ikke signifikant bedre resultater end den der er anvendt i kontrolgruppen.

I et skema som det der foreligger i denne undersøgelse, er det ligegyldigt om vi tildeler andre talværdier til de fire kategorier. I stedet for 1, 2, 3 og 4 kunne vi fx have benyttet 10, 20, 50 og 100. Vi kunne også have taget karakterer fra den nye karakterskala: 4, 7, 10 og 12. Det eneste afgørende er at til bedre præstationer svarer større talværdier. Talværdierne skal altså afspejle en rangorden. (Vi kan også vende sagen om og lade bedre præstationer svare til mindre talværdier. Der skal blot være sådan at bedre præstationer svarer til at talværdierne ændres i én og samme retning hver gang). I programmet antages det at data indtastes i stigende rangorden, altså således at det første felt svarer til de laveste range. – Hvis du indtaster i modsat rangorden, kan du blot fjerne tegnet i feltet som fastlægger den stigende rangorden.

Hvis vi havde lavet en undersøgelse i en klasse der viste hvor mange drenge og piger der havde matematik, engelsk og dansk som deres bedste fag, kunne vi fx få et skema som ser sådan ud:

	Matematik	Engelsk	Dansk
Drenge	10	5	7
Piger	12	7	3

Her er imidlertid ikke nogen naturlig rangorden mellem de tre grupper af data. Vi kunne således lige så godt have placeret Dansk eller Matematik i midten af skemaet. Det har derfor ingen mening at anvende en rangsumtest på de foreliggende data.

Ved et skema med kun to grupper er der ingen forvirrende ombytningsmuligheder. Enten står den ene gruppe forrest eller også den anden, og i begge tilfælde vil en rangsumtest føre til samme resultat. I et skema af den art kan derfor godt udføre en rangsumtest, selv om der ikke umiddelbart foreligger nogen naturlig rangfølge af de to grupper af data.

Prøv selv

I en klasse er der 10 drenge og 12 piger der har matematik som deres bedste fag, og der er 8 drenge og 3 piger der har dansk som deres bedste fag. Undersøg disse data ved en forskelstest.

Opgaver til Forskelstest

I opgaverne forudsættes det at der opstilles den hypotese at de to sæt af data stammer fra samme materiale, og at den fordeling af rangene der foreligger i to datasæt kan tænkes at være foretaget ved lodtrækning. Forskelstesten vil give en statistisk vurdering af denne antagelse.

Når testen fører til et signifikant resultat, kan du konkludere hvilket datasæt "der er bedst". Men du kan ikke sige noget om "hvor meget bedre" det ene sæt er i forhold til det andet. Bemærk at det ikke altid er datasættet med de højeste range der er bedst. Hvis undersøgelsen fx drejer sig om bilisters reaktionstid, så er det jo bedst at have korte reaktionstider, dvs. lave range i en rangordnet opstilling.

Når testens resultat er *Ingen signifikant forskel*, vil du ikke have nogen statistisk grund til at forkaste den opstillede hypotese om at de to datasæt kunne være fremkommet ved tilfældig udtagelse fra samme talmateriale. Eller med andre ord: *Der er ikke påvist nogen væsentlig forskel mellem de to grupper*.

Opstil i hver af opgaverne den hypotese du vil teste, og giv din fortolkning af testens resultat.

Et forslag: Inden du benytter testen, så tag et overblik over de givne data og giv dit gæt på hvad testen vil føre frem til.

1. Konditest

Ved en konditest af to grupper af idrætsmænd opnås følgende kondital for gruppens medlemmer:

Gruppe 1: 63, 71, 74, 69, 69, 76, 71, 72, 70.

Gruppe 2: 58, 72, 61, 67, 59, 69, 71, 61.

Undersøg de forelagte data ved hjælp af en forskelstest.

2. Mål i fodboldkampe

I en spillerunde i en fodboldturnering scores der følgende antal mål i kampene i to divisioner:

1. division: 3, 2, 1, 0, 2, 3, 2, 0, 1, 4.

2. division: 2, 5, 2, 4, 3, 1, 4, 2, 0, 3.

Undersøg de forelagte data ved hjælp af en forskelstest.

3. Driftssikkerhed

Maskiner af to fabrikater undersøges for deres driftssikkerhed. Her er antallet af dage med fejlfri kørsel:

Type 1:

195, 178, 250, 238, 252, 236, 234, 200, 212, 237, 214.

Type 2:

262, 242, 253, 212, 276, 266, 158, 301, 140.

Undersøg de forelagte data ved hjælp af en forskelstest.

4. Fejltastninger

To tasteoperatørers fejlprocenter undersøges. Der foreligger følgende data i 12 prøver:

Operatør 1:

2.05, 2.18, 3.04, 2.35, 2.37, 2.23, 2.23, 2.56, 1.02, 1.57, 1.45, 1.88.

Operatør 2:

1.31, 2.39, 0.24, 1.93, 2.38, 1.22, 0.35, 1.14, 1.43, 1.93, 2.48, 1.59.

Undersøg de forelagte data ved hjælp af en forskelstest.

5. C-vitamin

I en klasse på 20 skoleelever udvælges ved lodtrækning 10 som hver får et stort tilskud af C-vitamin. I løbet af skoleåret noteres følgende antal sygedage i de to grupper:

C-vitamin: 0, 0, 2, 3, 5, 10 13, 18, 20, 22.

De øvrige: 0, 2, 3, 6, 9, 15, 17, 18, 21, 24.

Undersøg de forelagte data ved hjælp af en forskelstest.

6. En færdighedsprøve

I en færdighedsprøve opnår to elevgrupper følgende point ud af 100:

Hold 1: 81, 71, 81, 69, 78, 53, 63, 45, 83, 81.

Hold 2: 85, 84, 86, 98, 94, 68, 93, 78, 74, 73.

Undersøg de forelagte data ved hjælp af en forskelstest.

7. Reaktionstider

I et psykologisk eksperiment foreligger følgende reaktionstider for forsøgsgruppen og for en kontrolgruppe:

Forsøgsgruppen:
12, 15, 14, 15, 17, 18, 8, 16, 20, 14, 19, 18, 7, 12.

Kontrolgruppen:
15, 12, 21, 16, 11, 20, 16, 14, 13, 18, 27, 23, 12, 22.

Undersøg de forelagte data ved hjælp af en forskelstest.

8. Et eksperiment

Lad to elever udføre følgende eksperiment: Der afmærkes et punkt på gulvet. Fra 3 meters afstand kastes en mønt som skal komme så tæt på det afmærkede punkt som muligt. Lad hver elev få 5-10 kast (lige mange til hver), og mål op hvor langt mønten lander fra målet.

Undersøg ved en forskelstest om der er noget der tyder på at de to elever ikke er lige dygtige.

9. Højre eller venstre hånd

Udfør eksperimentet fra opgave 8 således at du lader samme person foretage kast med skiftevis venstre og højre hånd.

Undersøg de forelagte data med en forskelstest: Er der noget der tyder på at personen rammer bedre med den ene hånd end med den anden?

10. Kopimaskiners sammenbrud

Her er antallet af timer mellem tilkaldelse af service til to fabrikater af kopimaskiner:

Maskine 1: 408, 127, 215, 555, 510.

Maskine 2: 88, 93, 214, 102, 365, 515.

Undersøg de forelagte data ved en forskelstest.

11. Hvem reagerer hurtigst?

I en måling af reaktionstider deltager to grupper af idrætsudøvere: en gruppe af sprintere og en gruppe af bordtennisspillere.

Sprintere:
0.215, 0.234, 0.239, 0.243, 0.248, 0,248, 0.249, 0.252, 0.259, 0.268, 0.273, 0.296.

Bordtennisspillere:
0.202, 0.211, 0.216, 0.229, 0.230, 0.241, 0.242, 0.246, 0.249, 0.256, 0.276, 0.285.

Undersøg de forelagte data ved hjælp af en forskelstest.

12. Antal kilometer pr. liter

For biler af to fabrikater måles benzinforbruget ved kørsel på en motorbane.

Fabrikat P:
16.10, 16.12, 16.20, 16.22, 16.30, 16.32, 16.50, 16.53, 16.67, 16.70, 16.75, 17.10, 17.15, 17.32, 17.40, 17.90.

Fabrikat R:
15.99, 16.02, 16.03, 16.10, 16.12, 16.14, 16.23, 16.25, 16.51, 16.67, 16.78, 16.92, 17.14, 17.16, 17.25, 17.70.

Undersøg de forelagte data ved en forskelstest.

13. Reaktionstid

I INFA-programmet BILER er indlagt et program til måling af reaktionstider. Lad to personer afprøve programmet og afgør ved hjælp af en test om der er en signifikant forskel på de to personers reaktionstider. Benyt 10 reaktionstider for hver person.

14. En prøve
Ved en prøve foreligger følgende resultater:

	Ikke bestået	Bestået
Gruppe 1	5	12
Gruppe 2	9	4

Undersøg de forelagte data ved en forskelstest.

15. En klar forskel?
I en klasse udtrækkes ved lodtrækning 3 piger og 3 drenge.

De deltager i samme opgaveløsning. Her er resultaterne:

	Løser ikke opgaven	Løser opgaven
Drenge	3	0
Piger	0	3

Undersøg disse data ved en forskelstest.

16. En klarere forskel?
I en klasse udtrækkes ved lodtrækning 4 piger og 4 drenge.
De deltager i samme opgaveløsning: Her er resultaterne:

	Løser ikke opgaven	Løser opgaven
Drenge	4	0
Piger	0	4

Undersøg disse data ved en forskelstest.

17. Tre grupper af data

I en prøve deltager hold 1 og 2. Deres præstationer bliver opdelt i tre kategorier: Lav, Middel, Høj:

	Lav	Middel	Høj
Hold 1:	10	12	4
Hold 2:	6	10	12

Undersøg de forelagte data ved hjælp af en forskelstest.