C2

Uforudsigelig variation kan forudsiges af gennemsnitstal

C1	Fra bunke til bundt - mangfoldighed, bundtning & stakning
C2	Uforudsigelig variation kan forudsiges af gennemsnitstal
A1	Sammenstakning af konkrete og abstrakte stakke
A2	Sammenlægning af per-tal
T1	Opstakning og afstakning, fremadregning og tilbageregning
T2	Stakke i tid, konstant og forudsigelig variation
S1	Stakke i rum, geometri
S2	Stakke i gitre, koordinatgeometri
PoMo	*Mængde-matematik eller mangfoldigheds-matematik*
KL	*Kvantitativ litteratur, Algebra: Opsamle & opdele*
GE	*Geometri: Jordmåling*

MATH**e**CADEMY: Matematik *nedefra*

2.1 Statistik ”bagud-siger” det uforudsigelige

En mangfoldighed, som varierer uforudsigeligt, siges at udføre stokastisk variation, og at være en stokastisk variabel, i modsætning til en deterministisk variabel, som kan forudsiges, fordi dens variation er determineret af en vækstligning (f.eks. $\Delta K = a$: $K = Ko+a*n$).

Typisk vil en stokastisk variabel optræde i forbindelse med spørgeskemaer, hvor vi spørger mennesker om f.eks. alder, afstand til skole, mening, mm., eller hvor vi spørger naturen om f.eks. kyllingers vægt to uger efter fødslen. Altså situationer, hvor vi ikke kan forudsige det næste svar.

Kan svaret ikke forudsiges, kan det til gengæld oftest ”bagud-siges” ved at lave statistik på de tidligere svar.

Eksempel 1: Ikke-grupperede observationer. I en sportsklub har vi spurgt medlemmerne: ”Hvor gammel er du?” Vi kunne ikke forudsige det næste svar, men da alle 50 svar var indløbet kunne vi opstille vore observationer i en tabel.

Observation Alder/år	Hyppighed	Frekvens	Summeret frekvens	Summeret alder	Afvigelse fra gennemsnit	Summeret afvigelse
8	4	4/50 = 0,08 = 8%	8%	8*4 = 32	10.1 - 8 = 2,1	2,1*4 = 8.4
9	12	12/50 = 0,24 = 24%	32%	9*12 = 108	10.1 – 9 = 1.1	1,1*12 = 13.2
10	16	16/50 = 0,32 = 32%	64%	10*16 = 160	10.1 – 10 = 0.1	0,1*16 = 1.6
11	10	10/50 = 0,20 = 20%	84%	11*10 = 110	11 - 10.1 = 0.9	0.9*10 = 9
12	8	8/50 = 0,16 = 16%	100%	12*8 = 96	12 - 10.1 = 1.9	1.9*8 = 15.2
Total	50	1,00 = 100%	-	506	-	47.4
Gennemsnit, middeltal:				506/50 = 10.1		47.4/50 = 0.9

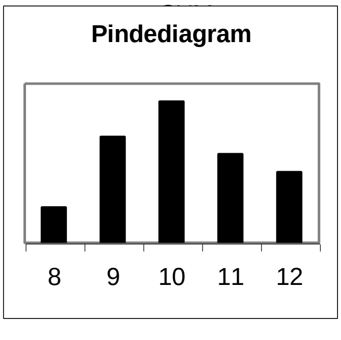
De indløbne svar kaldes observationer. Hyppigheden fortæller hvor ofte de enkelte svar forekommer. Frekvensen angiver de enkelte svars procentandel af det samlede antal svar.

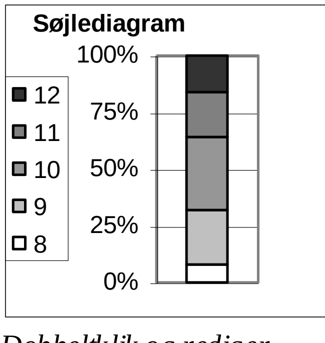
For at finde den gennemsnitlige alder, opsummeres alle aldre og fordeles ligeligt på de 50 svar. At gennemsnitsalderen er 10.1 er således et fiktivt tal, som blot siger ”hvis alle observationer var ens, ville de være 10.1”. I virkeligheden er aldrene ikke ens. Nogle afviger meget, andre lidt fra gennemsnittet. Vi kan derfor også udregne den gennemsnitlige afvigelse til at være 0.9. Igen et fiktivt tal, som blot siger: ”Hvis alle observationer afveg lige meget fra gennemsnittet, ville de afvige med 0.9. I virkeligheden er afvigelserne ikke ens. Nogle afvigelser afviger meget, andre lidt fra gennemsnittet.

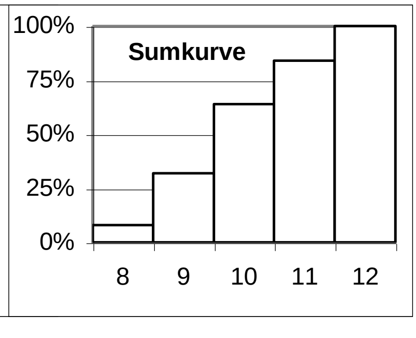
Den gennemsnitlige afvigelse minder meget om et mere avanceret statistisk tal, som kaldes spredningen. I praksis vil det være sådan, at ca. 95% af alle svar vil ligge i konfidens-intervallet: gennemsnit ± dobbelt gennemsnitlig afvigelse. Selvom vi står over for en uforudsigelig variation, kan vil alligevel lave en forudsigelse, ikke en tal-forudsigelse, men en interval-forudsigelse: Der er 95% sandsynlighed for at næste svar ligger i konfidens-intervallet 10.1 ± 2*0.9 år = 10.1 ± 1.8 år.

Hyppigheder kan indtegnes i et pindediagram. De summerede hyppigheder kan indtegnes i et søjlediagram eller en sumkurve, som fremkommer ved at forskubbe søjlediagrammets dele. Bliver sumkurven en ret linie på normalfordelingspapir, er der tale om en normalfordeling: En symmetrisk pukkel, hvor 68% ligger inden for spredningen, og 95% inden for den dobbelte spredning. Af sumkurven kan kvartilsættet aflæses: 1. kvartil = 9 år, da de laveste 25% af svarene går op til og med 9. Og 2. kvartil (medianen) = 10, da de laveste 50% af svarene går op til og med 10. Og 3. kvartil = 11, da de laveste 75% af svarene går op til og med 11. Typetallet er den observationen, som har størst hyppighed, dvs. 10.

OBS	**HYP**
8	**4**
9	**12**
10	**16**
11	**10**
12	**8**







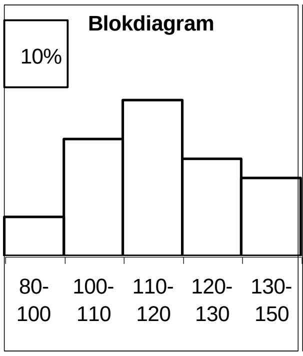
Dobbeltklik og rediger

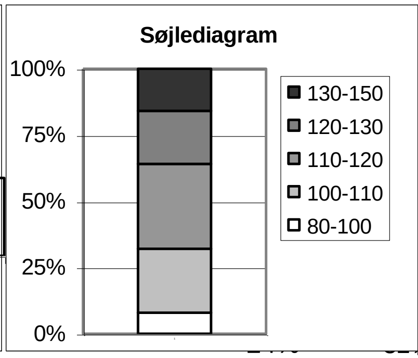
Eksempel 2: Grupperede observationer. Hvis der er mange forskellige svar, samles disse i grupper. Vi spurgte klassen: Hvor mange minutter læser du i gennemsnit lektier om ugen?

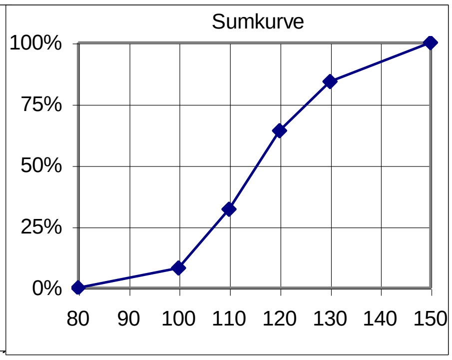
Observation minutter	Hyppighed	Frekvens	Summeret frekvens	Summeret varighed	Afvigelse fra gennemsnit	Summeret afvigelse
80-100	4	4/50 = 0,08 = 8%	8%	90*4 = 360	117 - 90 = 27	27*4 = 108
100-110	12	12/50 = 0,24 = 24%	32%	105*12 = 1260	117 – 105 = 12	12*12 = 144
110-120	16	16/50 = 0,32 = 32%	64%	115*16 = 1840	117 – 115 = 2	2*16 = 32
120-130	10	10/50 = 0,20 = 20%	84%	125*10 = 1250	125 – 117 = 8	8*10 = 80
130-150	8	8/50 = 0,16 = 16%	100%	140*8 = 1120	140 - 117 = 23	23*8 = 184
I alt	50	1,00 = 100%	-	5830	-	548
Gennemsnit, middeltal:				5830/50 = 117		548/50 = 11

Af sumkurven ses, at 1. kvartil = 108, 2. kvartil = median = 116, 3. kvartil = 125

OBS	HYP
80-100	4
100-110	12
110-120	16
120-130	10
130-150	8







Dobbeltklik og rediger

2.2 Sandsynlighedsregning forudsiger det uforudsigelige

Statistik ”bagud-siger” det uforudsigelige, sandsynlighedsregning forud-siger det uforudsigelige.

2.2.1 Stikprøveudtagning

Vi udtager på tilfældig måde 1 kort af en kortbunke (en stikprøve på 1). Vi beskriver udtagningen ved en udfaldstabel som afhænger af hvad vi interessere os for. Er vi interesseret i udfaldet sort/rødt kort har vi to udfald som er lige sandsynlige, fordi de to udfald er symmetriske og må derfor have ens sandsynligheder (symmetri argument), eller fordi De to udfald statistisk udgør ½ hver, og derfor vil forekomme ½ af gangene hver i det lange løb (statistisk argument)

Udfald	Rød	Sort
Sandsynlighed	½	½

Spiller vi på rød vil vi vinde hver anden gang. Odds (gevinst/indsats) skal derfor være 2/1 hvis spillet skal være retfærdigt. *Odds er således den reciprokke sandsynlighed.*

Interesserer vi os for kortets kulør, bliver udfaldstabellen:

Udfald	Hjerter	Ruder	Klør	Spar
Sandsynlighed	¼	¼	¼	¼

Interesserer vi os for om kortet er ruder eller ikke, bliver udfaldstabellen:

Udfald	Ruder (Gevinst G)	Ikke-Ruder (Ikke-Gevinst G)
Sandsynlighed	¼	¾

Interesserer vi os for om kortet er et es eller ikke, bliver udfaldstabellen:

Udfald	G = es	G = Ikke-es
Sandsynlighed	4/52	48/52

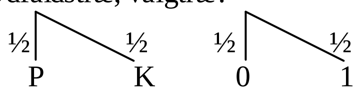
Kaster vi en tegnestift er der to mulige udfald. Sandsynligheden p findes ved at lave statistik på mange kast, dvs. ved bagudsigelse.

Udfald	Spidsen op	Spidsen ned
Sandsynlighed	p	1 - p

2.2.2 Gentaget kast med en symmetrisk mønt

1 kast med en symmetrisk mønt giver 2 lige sandsynlige udfald, plat(P) og krone(K), hver med sandsynligheden ½, og som i det lange løb hver vil forekomme ½ af gangene.

Udfaldstræ, valgtræ:



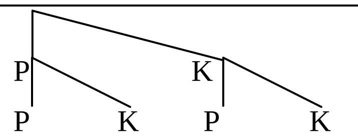
Vi opstiller en udfaldstabel

Udfald	P	K
Sandsynlighed	p(P) = ½	p(K) = ½

Hvis Gevinst = Krone (G = K) kan vi opstille en Gevinst-tabel:

Antal Gevinst-gange	0	1
Sandsynlighed	p(0) = ½	p(1) = ½

2 kast med en symmetrisk mønt giver 4 lige sandsynlige udfald, PP, PK, KP og KK, hver med sandsynligheden ¼ , og som i det lange løb hver vil forekomme ¼ af gangene.



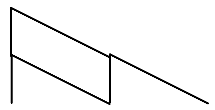
Udfald:	PP	PK	KP	KK
Antal gevinster:	0	1	1	2

I dette tilfælde vil udfaldstabellen få udseendet:

Kast nr.				
1	P		K	
2	P	K	P	K
Udfald	PP	PK	KP	KK
Sandsynlighed	p(PP) = ¼	p(PK) = ¼	p(KP) = ¼	p(KK) = ¼

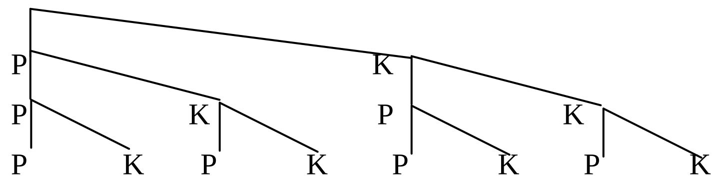
I dette tilfælde vil gevinsttabellen få udseendet:

Antal G	0	1	2
sandsynlighed	p(0) = ¼	p(1) = 2* ¼ = ½	p(2) = ¼



Antal Gevinster	0	1	2
Gennemløb	*1*	*2*	*1*

3 kast med en symmetrisk mønt giver 2*2*2=8 lige sandsynlige udfald, hver med sandsynligheden 1/8, og som i det lange løb hver vil forekomme 1/8 af gangene.



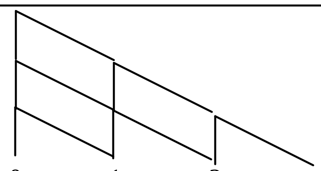
Udfald:	PPP	PPK	PKP	PKK	KPP	KPK	KKP	KKK
Antal gevinster:	0	1	1	2	1	2	2	3

I dette tilfælde vil udfaldstabellen få udseendet:

Kast nr.								
1	P				K			
2	PP		PK		KP		KK	
3	PPP	PPK	PKP	PKK	KPP	KPK	KKP	KKK
Antal G	0	1	1	2	1	2	2	3

I dette tilfælde vil gevinsttabellen få udseendet:

Antal G	0	1	2	3
sandsynlighed	p(0) = 1/8	p(1) = 3/8	p(2) = 3/8	p(3) = 1/8



Antal Gevinster	0	1	2	3
Gennemløb	*1*	*3*	*3*	*1*

4 kast med en symmetrisk mønt giver 2^4=16 lige sandsynlige udfald, hver med sandsynligheden 1/16.
Gennemløbstabellen for antal gevinster bliver som Pascals trekant (binomialkoefficienter):

Kast nr.					
1					
2	1	1			
3	1	2	1		
4	1	3	3	1	
	1	4	6	4	1
Antal G	0	1	2	3	4

I dette tilfælde vil Gevinst-tabellen få udseendet:

Antal G	0	1	2	3	4
sandsynlighed	p(0) = 1/16	p(1) = 4/16	p(2) = 6/16	p(3) = 4/16	p(4) = 1/16

5 kast med en symmetrisk mønt giver 2^5=32 lige sandsynlige udfald, hver med sandsynligheden 1/32 . I dette tilfælde vil gevinsttabellen få udseendet:

Antal G	0	1	2	3	4	5
sandsynlighed	p(0) = 1/32	p(1) = 5/32	p(2) = 10/32	p(3) = 10/32	p(4) = 5/32	p(5) = 1/32

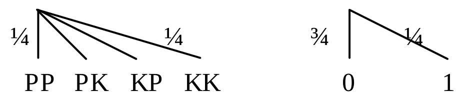
K(n,r) eller $\binom{n}{r}$ står for binomial-tallet i Pascals trekant i række n og søjle r kan vi skrive:

Med n = 5 er p(3) = K(5,3)*½^5.
Med n = n er p(3) = K(n,3)*½^n.
Eller generelt: p(r) = K(n,r)*½^n.

2.2.3 Gentaget kast med en usymmetrisk mønt

1 kast med en dobbeltmønt mønt giver 4 lige sandsynlige udfald, PP,PK,KP og KK, hver med sandsynligheden ¼, og som i det lange løb hver vil forekomme ¼ af gangene.

Udfaldstræ:



I dette tilfælde vil udfaldstabellen få udseendet:

Udfald	PP	PK	KP	KK
Sandsynlighed	p(PP) = ¼	p(PK) = ¼	p(KP) = ¼	p(KK) = ¼

Hvis gevinsten er KK (G = KK) vil 1 kast med 2 symmetriske mønter svare til 1 kast med 1 usymmetrisk mønt, hvis udfaldstabel har udseendet:

Udfald	0	1
Sandsynlighed	p(0) = ¾	p(1) = ¼

2 kast med en dobbeltmønt giver 4*4=16 lige sandsynlige udfald hver med sandsynligheden 1/16

Kast#																
1	PP				PK				KP				KK			
2	PP	PK	KP	KK	PP	PK	KP	KK	PP	PK	KP	KK	PP	PK	KP	KK
Udfald	PPPP	PPPK	PPKP	PPKK	PKPP	PKPK	PKKP	PKKK	KPPP	KPPK	KPKP	KPKK	KKPP	KKPK	KKKP	KKKK
Antal G	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	2
Sands.	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16	1/16

Vi kan også beskrive eksperimentet uf fra om der er gevinst 1 (G = KK) eller ikke gevinst 0.

Kast nr.				
1	0 (p=¾)		1 (p= ¼)	
2	0 (p=¾)	1 (p= ¼)	0 (p= ¾)	1 (p= ¼)
Udfald	00	01	10	11
Antal G	0	1	1	2
sandsynlighed	p(00) = 9/16 = ¾* ¾	p(01)=3/16= ¾* ¼	p(10) = 3/16=¼*¾	p(11) = 1/16=¼*¼

Vi ser at p(00) = 9/16 = ¾* ¾ = p(0) * p(0). Ved et gentaget eksperiment er den samlede sandsynlighed produktet af enkeltsandsynlighederne, svarende til at ¾ af ¾ af mange gentagelser giver 00.

I dette tilfælde vil Gevinst-tabellen få udseendet:

Antal G	0	1	2
sandsynlighed	p(0)=9/16	p(1)=2* 3/16=6/16	p(2)=1/16

¾ ¼
¾ ¼

Antal Gevinster	0	1	2
Gennemløb	*1*	*2*	*1*

3 kast med en dobbeltmønt giver 4*4*4 = 64 lige sandsynlige udfald hver med sandsynligheden 1/64. Igen kan vi nøjes med at interessere os for antal gevinster G = KK

Kast nr.

1	0 (p=3/4)				1 (p=1/4)			
2	0 (p=3/4)		1 (p=1/4)		0 (p=3/4)		1 (p=1/4)	
3	0 (p=3/4)	1 (p=1/4)	0 (p=3/4)	1 (p=1/4)	0 (p=3/4)	1 (p=1/4)	0 (p=3/4)	1 (p=1/4)
Udfald	000	001	010	011	100	101	110	111
AntalG	0	1	1	2	1	2	2	3
Sands.	¾* ¾*¾	¾*¾* ¼	¾*¾* ¼	¾*¼*¼	¾*¾* ¼	¾*¼*¼	¾*¼*¼	¼*¼*¼

Eller, hvis vi samler samme antal G:

Kast nr.

1	0	1						
2	0	1	0	1				
3	0	1	0	1	0	1		

Antal G	0	1	2	3
Antal gennemløb	1	3	3	1
Sands. pr. gennemløb	¾* ¾*¾	¾*¾* ¼	¾*¼*¼	¼*¼*¼
Samlet sandsynlighed	1 * ¾* ¾*¾	3 * ¾*¾* ¼	3 * ¾*¼*¼	1 * ¼*¼*¼

I dette tilfælde vil gevinsttabellen få udseendet:

Antal G	0	1	2	3
Sandsynlighed	p(0) = 27/64 = 1*¾* ¾*¾	p(1) = 27/64 = 3 *¾*¾* ¼	p(2) = 9/64 = 3 *¾*¼*¼	p(3) = 1/64 = 1* ¼*¼*¼

Igen ser vi fremkomsten af Pascals trekant. Hvis vi benytter skrivemåden K(n,r) for tallet i Pascals trekant i række n og søjle r kan vi skrive:

Hvis n = 3 og p(G) = ¼ er p(1) = K(3,1) * ¼ ^ 1 * ¾ ^2
Hvis n = n og p(G) = ¼ er p(1) = K(n,1) * ¼ ^ 1 * ¾ ^ (n-1)
Hvis n = n og p(G) = ¼ er p(r) = K(n,r) * ¼ ^ r * ¾^ (n-r) , eller generelt:
Hvis n = n og p(G) = p er p(r) = K(n,r) * p ^ r * (1-p) ^ (n-r)

En sådan fordeling kaldes en binomialfordeling. Den er tabellagt i binomialfordelingstabeller op til n = 50. Ellers kan tallene findes tilnærmelsesvis i en normalfordelingstabel.

Da vi aldrig kan forudsige det næste kast, er møntkast et eksempel på en stokastisk variabel. Vi kan vise, at for en binomialfordeling er middeltallet m = n*p, og spredningen s = √(n*p*(1-p)).

Eksempel. To mønter kastes samtidigt med gevinst ved udfaldet KK. Ved 12 kast kan vi vinde fra 0-12 gange. Gennemsnit = n*p = 12*0.25 = 3. Spredning = √(n*p*(1-p)) = √(12*0.25*0.75) = 1.5. Der er derfor 95% sandsynlighed for, at antal gevinstgange vil ligge i konfidens-intervallet 3±2*1.5. Dvs. der er under 5% sandsynlighed for at vinde mere end 6 gange.

Sandsynligheden for at vinde højst 4 gange = p(x<=4) = 0.8424
Sandsynligheden for at vinde netop 4 gange = p(x=4) = p(x<=4) - p(x<=3) = 0.8424 – 0.6488 = 0.1936
Sandsynligheden for at vinde mindst 4 gange = p(x>=4) = 1 - p(x<=3) = 1– 0.6488 = 0.3512

Tal i en binomialfordeling kan også findes med tilnærmelse ved en normalfordelings-approksimationen p(x<= r) = Φ((r + ½ -m)/s) enten i tabel eller på normalfordelingspapir:

Sandsynligheden for at vinde højst 4 gange = p(x<=4) = Φ((4 + ½ -3)/1.5) = Φ(1) = 0.84134

2.3 Kombinatorik

2.3.1Rækkefølgeordning

5 personer A,B,C,D,E kan rækkefølgeordnes (permuteres) på 5*4*3*2*1 = 5! måder: Thi der er 5 kandidater til plads 1. Og for hver af disse 5 muligheder er der 4 kandidater til plads 2. Og for hver af disse 5*4 muligheder er der 3 kandidater til plads 3. Og for hver af disse 5*4*3 muligheder er der 2 kandidater til plads 4. Og for hver af disse 5*4*3*2 muligheder er der 1 kandidat til plads 5. Dvs. der er i alt 5*4*3*2*1 = 5! muligheder. ! kaldes fakultet eller udråbstegn. Tallet 5! Kan findes i en tabel eller på en lommeregner.

2.3.2 Pårække-udtagning

Blandt 7 personer kan udtages en rækkefølge på 5 personer på R(7,5) måder. Der er 7 kandidater til plads 1. For hver af disse 7 muligheder er der 6 kandidater til plads 2. For hver af disse 7*6 muligheder er der 5 kandidater til plads 3. For hver af disse 7*6*5 muligheder er der 4 kandidater til plads 4. For hver af disse 7*6*5*4 muligheder er der 3 kandidat til plads 5. Dvs. der er i alt 7*6*5*4*3 muligheder. Vi omskriver: 7*6*5*4*3 = 7*6*5*4*3*(2*1)/(2*1) = 7!/2!. Tilsvarende: Blandt 7 personer kan der udtages en rækkefølge på r personer på 7!/(7-r)! måder. Og: Blandt n personer kan der udtages en rækkefølge på r personer på R(n,r) = n!/(n-r)! måder.

2.3.3 Klump-udtagning

Blandt 7 personer kan udtages en klump (kombination) på 5 personer på K(7,5) måder. Hver klump på 5 personer kan rækkefølgeordnes på 5! måder. Resultatet af <u>først</u> at vælge en klump og <u>dernæst</u> vælge en rækkefølge bliver en pårækkeudtagning af 5 personer blandt 7, hvilket kan gøres på R(7,5) måder. Dvs. K(7,5)*5! = R(7,5). Heraf ses, at

K(7,5) = R(7,5)/5! = 7!/(5!*2!)

K(n,5) = R(n,5)/5! = n!/(5!*(n-5)!)

K(n,r) = R(n,r)/r! = n!/(r!*(n-r)!)

K-tallene kan findes i en tabel eller på en lommeregner.

2.4 Gentaget stikprøveudtagning med og uden tilbagelægning

I en kasse befinder der sig 12 kugler, hvoraf 4 er Gunstige (f.eks. G = hvid), og 8 ugunstige (~~G~~ = ikke-hvid). Der udtages på tilfældig måde en stikprøve på 3 udtag.

2.4.1 Med tilbagelægning

Hvis kuglen lægges tilbage efter hvert udtag, er en stikprøveudtagning en gentagelse af samme eksperiment, og kan derfor behandles som om det var gentaget kast med en usymmetrisk mønt. Stikprøveudtagningen med tilbagelægning kan da beskrives som en binomialfordeling.

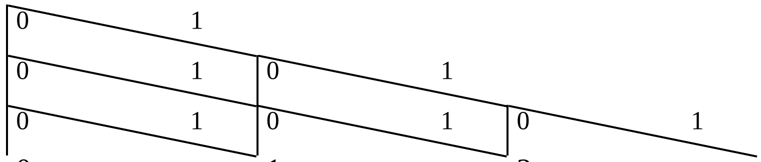
2.4.2 Uden tilbagelægning

Hvis kuglen ikke lægges tilbage efter hvert udtag, er en stikprøveudtagning ikke en gentagelse af samme eksperiment, og kan derfor ikke behandles som om det var gentaget kast med en usymmetrisk mønt. Stikprøveudtagningen uden tilbagelægning kan da ikke beskrives som en binomialfordeling, men som en såkaldt "hypergeometrisk fordeling".

Hver gang der udtages en kugle får kassen en ny sammensætning og dermed nye sandsynligheder:

Udtag nr.									
1	0 (p=8/12)				1 (p=4/12)				
2	0 (p=7/11)		1 (p=4/11)		0 (p=8/11)		1 (p=3/11)		
3	0 (p=6/10)	1 (p=4/10)	0 (p=7/10)	1 (p=3/10)	0 (p=7/10)	1 (p=3/10)	0 (p=8/10)	1 (p=2/10)	
Udfald		000	001	010	011	100	101	110	111
Antal G		0	1	1	2	1	2	2	3
Sand-synlighed		$\frac{8*7*6}{12*11*10}$	$\frac{8*7*4}{12*11*10}$	$\frac{8*4*7}{12*11*10}$	$\frac{8*4*3}{12*11*10}$	$\frac{4*8*7}{12*11*10}$	$\frac{4*8*3}{12*11*10}$	$\frac{4*3*8}{12*11*10}$	$\frac{4*3*2}{12*11*10}$

Eller, hvis vi samler samme antal G:
Kast nr.
1
2
3



Antal G	0	1	2	3
Antal gennemløb	1	3	3	1
Sands. pr. gennemløb	(8*7*6)/(12*11*10)	(8*7*4)/(12*11*10)	(8*4*3)/(12*11*10)	(4*3*2)/(12*11*10)
Samlet sandsynlighed	1 *(8*7*6)/(12*11*10)	3 *(8*7*4)/(12*11*10)	3 *(8*4*3)/(12*11*10)	1*(4*3*2)/(12*11*10)

I dette tilfælde vil gevinsttabellen få udseendet:

Antal G	0	1	2	3
Sandsynlighed	p(0)=(8*7*6)/(12*11*10)	p(1)=3*(8*7*4)/(12*11*10)	p(2)=3*(8*4*3)/(12*11*10)	p(3)=(4*3*2)/(12*11*10)

Igen ser vi fremkomsten af Pascals trekant. Hvis vi benytter skrivemåden K(n,p) for tallet i Pascals trekant i række n og søjle p kan vi skrive:

Hvis n = 3 og n(G, ~~G~~) = (4,8) er p(1) = K(3,1) * R(8,2) * R(4,1) / R(12,3)
p(1) = K(3,1) * R(8,(3-1)) * R(4,1) / R((8+4),3)
Hvis n = n og n(G, ~~G~~) = (4,8) er p(1) = K(n,1) * R(8,(n-1)) * R(4,1) / R(12,n)
Hvis n = n og n(G, ~~G~~) = (s,t) er p(1) = K(n,1) * R(t,(n-1)) * R(s,1) / R(s+t,n)
Hvis n = n og n(G, ~~G~~) = (s,t) er p(r) = K(n,r) * R(t,(n-r)) * R(s,r) / R(s+t,n)

Da R(n,r) = K(n,r) * r! kan dette omformes til

R(t,(n-r)) * R(s,r) / R(s+t,n) = K(t,(n-r)) * K(s,r) / K(s+t,n) * (n-r)! * r!/n!
= K(t,(n-r)) * K(s,r) / K(s+t,n) / K(n,r) dvs.

p(r) = K(n,r) * R(t,(n-r)) * R(s,r) / R(s+t,n) = K(n,r) * K(t,(n-r)) * K(s,r) / K(s+t,n) / K(n,r)
p(r) = K(t,(n-r)) * K(s,r) / K(s+t,n)

Denne formel kunne vi også have tænkt sig til direkte. En stikprøve på n blandt s+t er en klump, der kan udtages på K(n,s+t) måder. De r gunstige kan udtages blandt s mulige på K(s,r) måder. De n-r ugunstige kan udtages blandt t mulige på K(t,n-r) måder. En stikprøven med r gunstige og n-r ugunstige kan da udtages på K(s,r)*K(t,n-r) måder blandt de i alt K(n,s+t) måder, dvs. med sandsynligheden p(r) = K(t,(n-r))*K(s,r) / K(s+t,n).

Hypergeometriske sandsynligheder kan findes i en tabel eller på en lommeregner. Hvis kugletallet er stort og udtaget lille, er der ikke den store forskel på udtagning med og uden tilbagelægning. Så binomialfordelingen bruges ofte som approksimation til den hypergeometriske fordeling.

2.5 Dobbeltfordeling, krydstabeller, betinget sandsynlighed

Indenfor statistik/sandsynlighedsregning kaldes en mangfoldighed for en population. En population kan ofte optælles på to forskellige måder (som ved dobbeltbundtning) og beskrives i en krydstabel.

Eksempel 1. Et sædvanligt kortspil

	Ruder	Ikke-ruder	Total
Billedkort	3	9	12
Talkort	10	30	40
Total	13	39	52

<u>Gennemsnitlig procent/sandsynlighed:</u>

Procenten-af/sandsynligheden-for billedkort blandt alle = p(B I alle) = p(B) = 12/52 = 23.1%

<u>Betinget procent/sandsynlighed:</u>

Procenten-af/sandsynligheden-for billedkort blandt ruderne = p(B I Ruder) = 3/13 = 23.1%

Da p(B I Ruder) = p(B) siger vi: ”Ruderne har en procentandel af billedkort svarende til gennemsnittet”, eller ”Blandt ruderne er billedkortene normalt repræsenteret”.

De betingede procenter/sandsynligheder kan vises i en fordelingstabel, evt. oprettet i Excel som krydstabel (pivot-tabel):

I efter kulør	Ruder	Ikke-ruder	Total
Billedkort	23.1%	23.1%	23.1%
Talkort	76.9%	76.9%	76.9%
Total	100%	100%	100%

II efter talværdi	Ruder	Ikke-ruder	Total
Billedkort	25%	75%	100%
Talkort	25%	75%	100%
Total	25%	75%	100%

Eksempel 2. Et kortspil uden sorte konger

	Ruder	Ikke-ruder	Total
Billedkort	3	7	10
Talkort	10	30	40
Total	13	37	50

Vi opstiller de to krydstabeller:

I efter kulør	Ruder	Ikke-ruder	Total
Billedkort	23.1%	18.9%	20%
Talkort	76.9%	81.1%	80%
Total	100%	100%	100%

II efter talværdi	Ruder	Ikke-ruder	Total
Billedkort	30%	70%	100%
Talkort	25%	75%	100%
Total	26%	74%	100%

På baggrund af krydstabel I kan vi sige:

”Ruderne har en procentandel af billedkort over gennemsnittet”, eller ”Blandt ruderne er billedkortene over-repræsenteret”. Modsat for talkortene.

”Ikke-ruderne har en procentandel af billedkort under gennemsnittet”, eller ”Blandt ikke-ruderne er billedkortene under-repræsenteret”. Modsat for talkortene.

På baggrund af krydstabel II kan vi sige:

” Billedkortene har en procentandel af ruder over gennemsnittet”, eller ”Blandt billedkortene er ruderne over-repræsenteret”. Modsat for ikke-ruderne.

” Talkortene har en procentandel af ruder under gennemsnittet”, eller ”Blandt talkortene er ruderne under-repræsenteret”. Modsat for ikke-ruderne.

Eksempel 3. To klasser fordelt efter køn og ryger-status

	Ryger	Ikke-ryger	Total
Piger	3	7	10
Drenge	10	30	40
Total	13	37	50

Vi opstiller de to krydstabeller:

I efter rygerst.	Ryger	Ikke-ryger	Total
Piger	23.1%	18.9%	20%
Drenge	76.9%	81.1%	80%
Total	100%	100%	100%

II efter køn	Ryger	Ikke-ryger	Total
Piger	30%	70%	100%
Drenge	25%	75%	100%
Total	26%	74%	100%

På baggrund af krydstabel I kan vi sige:

”Rygerne har en procentandel af piger over gennemsnittet”, eller ”Blandt rygerne er pigerne over-repræsenteret”. Modsat for drengene.

”Ikke-rygerne har en procentandel af piger under gennemsnittet”, eller ”Blandt ikke-rygerne er pigerne under-repræsenteret”. Modsat for drengene.

På baggrund af krydstabel II kan vi sige:

”Pigerne har en procentandel af rygere over gennemsnittet”, eller ”Blandt pigerne er rygerne over-repræsenteret”. Modsat for ikke-rygerne.

”Drengene har en procentandel af rygere under gennemsnittet”, eller ”Blandt drengene er rygerne under-repræsenteret”. Modsat for ikke-rygerne.

Eksempel 4. En stor population fordelt efter køn og ryger-status.

Hvis vi ikke kan undersøge hele populationen udtager vi en stikprøve, hvis procenter så er behæftet med en usikkerhed u = 2*√(p*(1-p)/n).

Eksempel: p(Piger | Ryger) = 27.3%, n = 550, u = 2*√(27.3%*72.7%/550) = 3.8%

Vi opstiller antalstabellen med usikkerhederne samt de to krydstabeller:

	Ryger	Ikke-ryger	Total	*u*
Piger	150	270	420	*4,7%*
Drenge	400	500	900	*3,3%*
Total	550	770	1320	*2,7%*
u	*3,8%*	*3,4%*	*2,6%*	

I efter køn	Ryger	Ikke-ryger	Total
Piger	27,3%	35,1%	31,8%
Drenge	72,7%	64,9%	68,2%
Total	100,0%	100,0%	100,0%

II efter rygerst.	Ryger	Ikke-ryger	Total
Piger	35,7%	64,3%	100,0%
Drenge	44,4%	55,6%	100,0%
Total	41,7%	58,3%	100,0%

Dobbeltklik og rediger i antalstabellen

En afvigelse kaldes signifikant hvis den er større end usikkerheden.

På baggrund af krydstabel I kan vi sige:

"Rygerne har en procentandel af piger som er signifikant under gennemsnittet (31.8%-27.3% = 4.5% > 3.8%)", eller "Blandt rygerne er pigerne signifikant under-repræsenteret".

"Ikke-rygerne har en procentandel af piger som ikke er signifikant over gennemsnittet (35.1%-31.8% = 3.3% < 3.4%)", eller "Blandt ikke-rygerne er pigerne ikke signifikant over-repræsenteret".

"Populationen har en procentandel af piger som er signifikant under halvdelen (50%-31.8% = 18.2% > 2.6%)", eller "I populationen er pigerne signifikant under-repræsenteret".

Hvad kan siges om drengene?

På baggrund af krydstabel II kan vi sige:

"Pigerne har en procentandel af rygere som er signifikant under gennemsnittet (41.7%-35.7% = 6% > 4.7%)", eller "Blandt pigerne er rygerne signifikant under-repræsenteret".

"Drengene har en procentandel af rygere som ikke er signifikant over gennemsnittet (44.4%-41.7% = 2.7% < 3.3%)", eller "Blandt drengene er rygerne ikke signifikant over-repræsenteret".

"Populationen har en procentandel af rygere som er signifikant under halvdelen (50%-41.7% = 8.3% > 2.7%)", eller "I populationen er rygerne signifikant under-repræsenteret".

Hvad kan siges om ikke-rygerne?

Eksempel 5. Hypotesetest

En hypotese som f.eks. "Halvdelen af populationen er piger" (p = 50%) kan testes ved at beregne procentandelen af piger i en stikprøve. Hvis afvigelsen fra hypotesen er signifikant, forkastes hypotesen. Hvis afvigelsen fra hypotesen ikke er signifikant, accepteres hypotesen. Da usikkerhedsformlen bygger på, at der er 5% chance for at afvige mere end den dobbelte spredning, siges hypotesetesten at være gennemført på signifikansniveauet 5%. Hypotesetest og signifikans er således blot andre ord for bestemmelse af en procentandels konfidens-interval.

C2 OPGAVER
Spørgsmål: Hvordan kan vi regne os frem til sluttallet, hvis tilvæksten er uforudsigelig? Svaret er indførelse af statistik og sandsynlighedsregning: Hvor statistikken kan bagud-sige, at gennemsnitstallet er 8.2 og spredningen er 2.3, så kan sandsynligheden forudsige, at der er 95% sandsynlighed for at det næste tal vil ligge i intervallet: gennemsnit ± 2*spredning, dvs. 8.2 ± 4.6.

C21.1 Faglige opgaver
1. Færdiggør Excel-arket udfaldstræ svarende til en stikprøveudtagning på 8 udtag. Byg evt. et sømbrædt med 8 rækker så glaskugler bliver stoppet 8 gange under deres gennemløb.

2. Hent en tabel i statistisk tiårsoversigt, og vis hvordan fordelingssøjlen ændrer sig henover et tiår. Brug Excel.

3. Løs opgave 3 (skrabspil) i eksamenssættet fra maj 2003 (husk at læse 5. ”Hvad er chancen” i forberedelsesmaterialet, side 10-21).

4. Bevis formlen for middelværdi og spredning for en binomialfordeling i tilfældet n = 2 og evt. n = 3.

C21.2 Rutineopgaver
1. Opstil og løs en række opgaver inden for ikke-grupperede observationer. Brug Excel-arket som facitliste.

2. Opstil og løs en række opgaver inden for grupperede observationer. Brug Excel-arket som facitliste.

3. Opstil og løs en række opgaver inden for binomialfordeling. Brug Excel-arket som facitliste. Regn også opgaverne ved normalfordelingsapproksimation, og ved normalfordelingspapir.

4. Regn nedenstående træningsopgaver i statistik og binomialfordeling.

5. Find og løs nogle eksamensopgaver fra HF matematik fællesfag.

C21.3 Didaktisk opgave
1. Spørgeskema undersøgelse ved brug af kort

Tag et almindeligt spil kort. Fjern billedkortene, de røde 4 og 6ere, de sorte 5 og 7ere. Udtag på tilfældig måde et kort og spørg det om følgende:

1) Antal fætre og kusiner: <talværdi>. 2) Boligform: Landet <klør>, parcelhus <spar>, rækkehus <ruder>, lejlighed <hjerter>. 3) Holdning til rygeforbud på skolen: Ja <flertal af retvendte figurer>, Nej <mindretal af retvendte figurer>, Ved ikke <samme antal>. 4) Afstand til skole i km. <maksimal ”kort frigang” i cm>, 5) daglig indkomst (lommepenge og fritidsarbejde) <maksimal ”lang frigang” i mm>. 6) Køn: <pige-primtal, dreng-ellers> 7) Klasse: A <kort 1-25>, B <kort 26-45>, C <kort 46-60>

Eksempel: Spar 6 giver følgende svar: 1) 6 fætre og kusiner, 2) parcelhus, 3) ja (nej hvis kortet var lagt modsat på bordet) 4) 1,1 km (afstand mellem to vandrette spartegn) 5) 16 kr. (afstand mellem to lodrette spartegn), 6) dreng, 7) B (kort nr. 32)

Læg kortet tilbage bland kortene og udtag det næste. Gentag dette 60 gange (3 klasser).

Opstil svarene på en nummereret liste, og beskriv undersøgelsens resultat ved hjælp af forskellige statistiske metoder.

Afprøv opgaven på dig selv og på en anden. Rapporter dine observationer af hvad forsøgspersonen gør og tænker/siger (handling og refleksion). Vær især opmærksom på eksempler på genkendelse og ny erkendelse (assimilering og akkommodering).

2. Undervisningsforløb i udtag af 2 og 3 kugler med og uden tilbagelægning

Du skal udarbejde et forslag til et 4*3-lektioners undervisningsforløb i stikprøveudtagning i 9. klasse.

Forslaget skal bl.a. indeholde 1) planer for hver uge (3 timer), herunder forslag til hjemmeopgaver,
2) referencer til relevant didaktisk og pædagogisk teori, 3) mulighed for tværfaglig undervisning med et eller flere andre fag.

Træningsopgaver:

						MID		SPR	KVT
	1	x	h	p	? p	x·p	|x-M|	(x-M)^2·p	
10	30	10-30	3						
30	40	30-40	5						
40	50	40-50	9						
50	60	50-60	4						
60	70	60-70	2						

svar:		MID		SPR	KVT
p	? p	x·p	|x-M|	(x-M)^2·p	
0,130	0,130	2,6	23,0	69,3	
0,217	0,348	7,6	8,0	14,1	35,5
0,391	0,739	17,6	2,0	1,5	43,9
0,174	0,913	9,6	12,0	24,9	50,6
0,087	1,000	5,7	22,0	41,9	
1,000		43,0		151,6	
MID±2·SPR:		*18,4*	*67,7*	12,3	

						MID		SPR	KVT
	2	x	h	p	? p	x·p	|x-M|	(x-M)^2·p	
0	10	0-10	3						
10	20	10-20	9						
20	30	20-30	12						
30	40	30-40	11						
40	60	40-60	4						

		MID		SPR	KVT
p	? p	x·p	|x-M|	(x-M)^2·p	
0,077	0,077	0,4	21,5	35,7	
0,231	0,308	3,5	11,5	30,7	17,5
0,308	0,615	7,7	1,5	0,7	26,3
0,282	0,897	9,9	8,5	20,2	34,8
0,103	1,000	5,1	23,5	56,5	
1,000		26,5		143,8	
MID±2·SPR:		*2,6*	*50,5*	12,0	

Dobbeltklik og rediger

Binomialfordeling			*Netop 3*	*Højst 2*	*Mindst 2*	*Mi1&Hø3*
3	**x**	**∑p**	**x 3**	**x 2**	**x 2**	**1 x 3**
	0					
n	1					
4	2					
p	3					
0,3	4					

Svar:	*Netop 3*	*Højst 2*	*Mindst 2*	*Mi1&Hø3*
∑p	**x = 3**	**x ≤ 2**	**x ≥ 2**	**1≤x≤3**
0,2401				-0,2401
0,6517			-0,6517	
0,9163	-0,9163	0,9163		
0,9919	0,9919			0,9919
1,0000			1,0000	
	0,0756	**0,9163**	**0,3483**	**0,7518**

			Netop 2	*Højst 3*	*Mindst 4*	*Mi2&Hø4*
4	**x**	**∑p**				
n						
5						
p						
0,6						

	Netop 2	*Højst 3*	*Mindst 4*	*Mi2&Hø4*
∑p	**x = 2**	**x ≤ 3**	**x ≥ 4**	**2≤x≤4**
0,0102				
0,0870	-0,0870			-0,0870
0,3174	0,3174			
0,6630		0,6630	-0,6630	
0,9222				0,9222
1,0000			1,0000	
	0,2304	**0,6630**	**0,3370**	**0,8352**