

# Statistiske tests ved



*Omrøring...*



*eller Stølestropning*

## Statistiske test ved brug af DataMeter.

Metoderne i klassisk matematisk statistik er kun forståelige med et omfattende matematisk begrebsapparat som baggrund.

I de senere år er det med computerens udvikling blevet realistisk at udføre en række test "eksperimentelt" ved at udnytte computerens regnekraft. Ideen hertil er af ældre dato, faktisk lige så gammel som de klassiske metoder, men nu er den blevet praktisk gennemførlig. For en række klassiske test findes der nu parallelle "eksperimentelle" test, som ikke kræver nær samme matematiske baggrund, og som derfor kan anvendes på gymnasieniveau. Metoderne går på engelsk under betegnelsen *resampling*. Det, jeg kalder *omrøringstest*, kaldes også *permutationstest*, og på engelsk anvendes flere forskellige navne.

Vi vil med nogle eksempler belyse et par hovedtyper af test, *omrøringstest* og *test ved at trække sig selv op ved håret* (engelsk *bootstrap*, altså egentlig *at trække sig selv op ved støvlestroppen*).

### Eksempel på statistisk test ved omrøring.

Engang var et hold, 2z, til terminsprøve og siden til skriftlig eksamen i matematik. Vi vil undersøge, om terminsprøven havde en gunstig virkning (i hvert fald på karaktererne!), således at eksamen gik bedre end terminsprøven.

Karaktererne blev dengang givet efter 13-skalaen og var som følger:

#### Terminsprøven:

00	03	5	6	7	8	9	10	11	13
0	7	5	2	0	2	0	1	0	0

#### Eksamen:

00	03	5	6	7	8	9	10	11	13
0	4	5	2	0	2	3	1	0	0

Som nulhypotese har vi, at holdet klarede terminsprøve og eksamen på samme måde, således at forskelle kan forklares ved tilfældigheder.

Vi indskrives dem i datameter som ét datasæt, karakterer, og knytter kategorivariablen tidspunkt



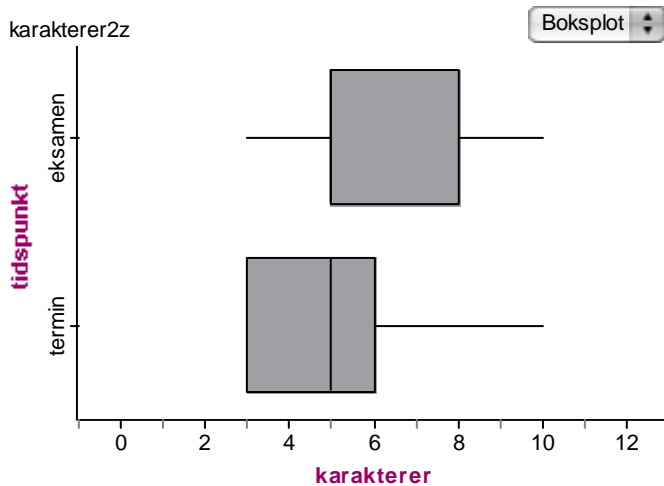
med værdierne "termin" og "eksamen" hertil:

karakterer2z

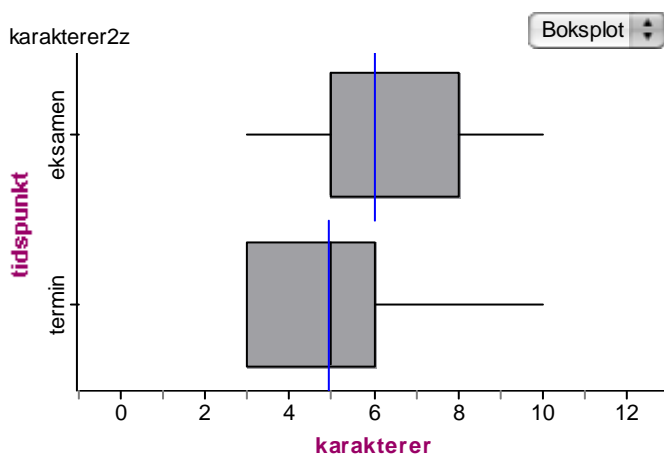
2z karakterer

	karakterer	tidspunkt	<ny>
1		3 termin	
2		3 termin	
3		3 termin	
4		3 termin	
5		3 termin	
6		3 termin	
7		3 termin	
8		5 termin	
9		5 termin	
10		5 termin	
11		5 termin	
12		5 termin	
13		6 termin	
14		6 termin	
15		8 termin	
16		8 termin	
17		10 termin	
18		3 eksamen	
19		3 eksamen	
20		3 eksamen	
21		3 eksamen	
22		5 eksamen	
23		5 eksamen	
24		5 eksamen	
25		5 eksamen	
26		5 eksamen	
27		6 eksamen	
28		6 eksamen	
29		8 eksamen	
30		8 eksamen	
31		9 eksamen	
32		9 eksamen	
33		9 eksamen	
34		10 eksamen	

I et grafvindue trækker vi de variable ind og beder om et boksplot:



Det ser lidt pudsigt ud hvilket bl.a. hænger sammen med, at medianen i begge tilfælde er 5! Det kunne dog godt se ud som om, eksamen ligger lidt bedre end terminsprøven; men vi må nok sammenligne på middelværdien i stedet for medianen. Vi indlægger middelværdi i boksplottet (højreklik; plot værdi; ”middel” findes under statistiske funktioner eller indskrives simpelthen):



| middel ( ) = 5,47059

Vi tager også en opsummering (beregningsværktøjet):

karakterer2z

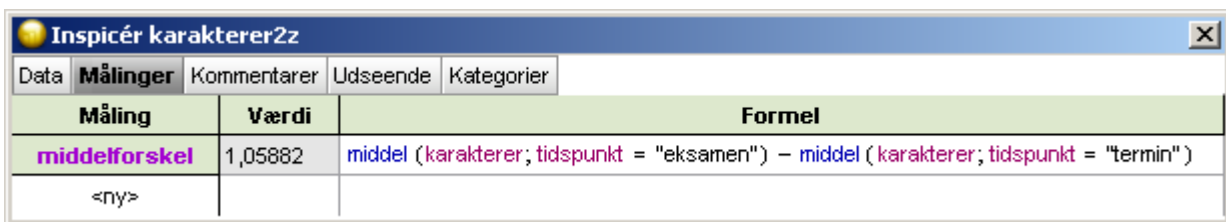
tidspunkt	eksamen	6
	termin	4,9411765
Søjle total		5,4705882

R1 = middel (karakterer)

Forskellen i middelværdi er 1,05882.

Vi foretager en måling af forskellen mellem middelværdierne i datasættets inspektør:

Under Måling i inspektøren defineres middelforskel som  $\text{middel}(\text{karakterer}; \text{tidspunkt} = \text{"eksamen"}) - \text{middel}(\text{karakterer}; \text{tidspunkt} = \text{"termin"})$ .



Måling	Værdi	Formel
middelforskel	1,05882	$\text{middel}(\text{karakterer}; \text{tidspunkt} = \text{"eksamen"}) - \text{middel}(\text{karakterer}; \text{tidspunkt} = \text{"termin"})$
<ny>		

Nu omrøres datasættet (højreklik på datasættet og vælg: "rør rundt i en variabel"):

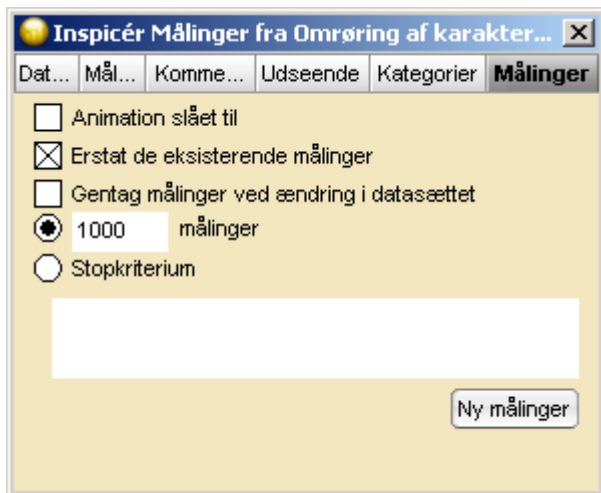


## Omrøring af karakterer2z

I inspektøren kan vi under "omrøring" vælge at omrøre kategorivariablen "tidspunkt" (det siges at være det smarteste).



Vi højreklikker først på Omrøring af karakterer 2z og bestiller Gentagne målinger. Så inspiceres Måling af Omrøring af karakterer2z, og vi bestiller gentagne omrøringer med passende valg: animation slået fra, erstat de eksisterende målinger, 1000 målinger.

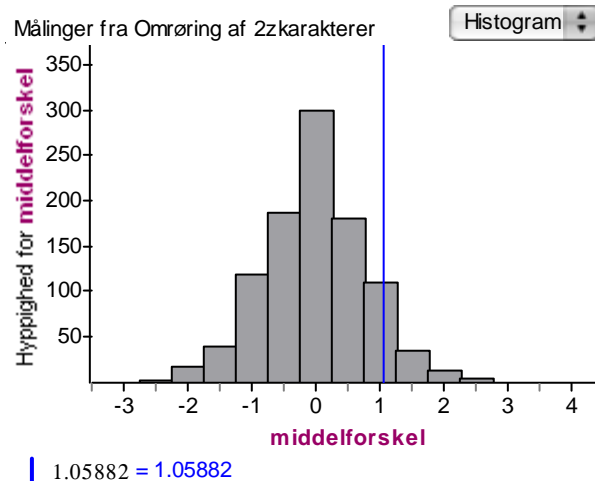


Vi kan nu efter at have trukket en tabel ned se målingerne af middelforskel i nogle af de omrørte datasæt:

Målinger fra Omrøring af 2z karakterer

	middelforskel	<ny>
1	-0.352941	
2	0.588235	
3	0.117647	
4	-0.588235	
5	1.41176	
6	0	
7	-0.352941	
8	0.705882	
9	0	
10	0.941176	

Nu trækker vi en graf ned og vi ser på hele middelforskel-fordelingen i et histogram, hvor vi indlægger (Plot værdi) vores allerførste midelforskel 1,05882:



Vi må nu indrømme, at vores middelforskel ikke er så usandsynlig, hvis alle 34 karakterer fordeles tilfældigt på terminsprøve og eksamen.

Vi kan få talt, hvor mange af de 1000 omrøringer, der giver en middelforskel større end 1,05882:

Målinger fra Omrøring af 2z karakt...

middelforskel	91
---------------	----

R1 = tæl (middelforskel > 1.05882)

altså 91 af de 1000, og det er jo 9,1%. Tallet er for stort (f.eks. klart større end 5%, som vi vælger som signifikansniveau)

til at vi har kunnet påvise, at eksamen for holdet gik bedre end terminsprøven!

### Statistisk test ved ”trækken sig selv op ved håret”, bootstrap.

Testet tager udgangspunkt i et simpelt eksempel fra Mogens Lønborg: ”Grundbog i statistik”, Forlaget Lønborg, 2.udg., 3. opl., 1993, side 151. En virksomhed i genbrugsbranchen tømmer hver 14. dag en bestemt container med glas. Indholdet vejes i kg, og over 18 perioder på hver 14 dage har man fået følgende datamateriale, som indskrives i datameter:

	Vægt	<ny>
1	1660	
2	1820	
3	1590	
4	1440	
5	1730	
6	1680	
7	1750	
8	1720	
9	1900	
10	1570	
11	1700	
12	1900	
13	1800	
14	1770	
15	2010	
16	1580	
17	1620	
18	1690	

Firmaet ønsker at vide, hvor meget glas, der gennemsnitligt indsamles, og det er kun tilfreds, hvis middelværdien er større end 1600 kg.

Ideen i testet er, at hvis middelværdien faktisk er højst 1600, må en del (f.eks. mindst 5%) af middelværdierne ved gentagne stikprøver (med tilbagelægning og alle af størrelse 18 ligesom datamaterialet) udtaget af det givne datamateriale være højst 1600.

Dobbeltklik i glasindholds inspektør og lad under målinger *middel* angive middel(Vægt).

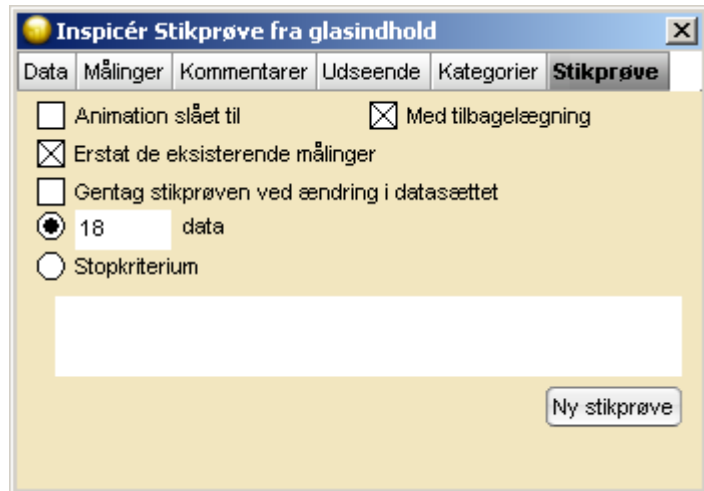
Måling	Værdi	Formel
middel	1718,33	middel (Vægt)
<ny>		

Efter et højreklik vælges ”udtag stikprøve”.



Stikprøve fra glasindhold

Vi inspicerer og vælger passende:

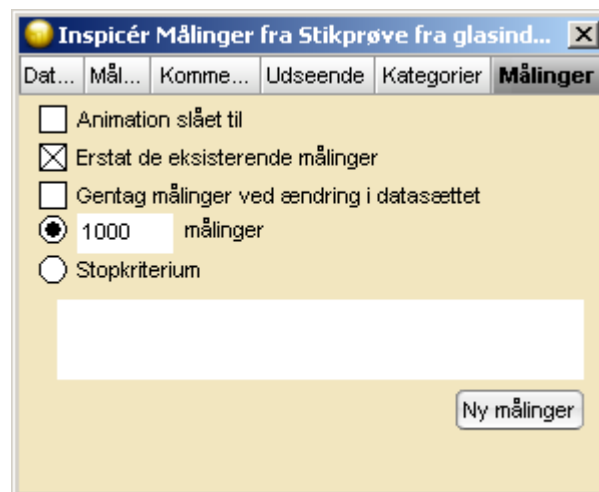


Højreklik på stikprøve fra glasindhold følges af ”udfør gentagne målinger”

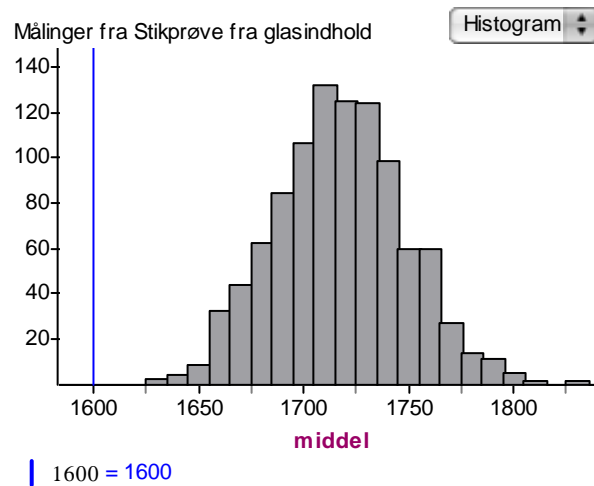


Målinger fra Stikprøve fra glasindhold

Så inspektion:



Vi ser nu på histogrammet for de 1000 målinger af middelværdi:



Der er ingen middelværdier under 1600, så vi forkaster nulhypotesen. Vi kontrollerer lige med en optælling:

Målinger fra Stikprøve fra glasindhold

	middel
	0

R1 = tæl (middel < 1600)

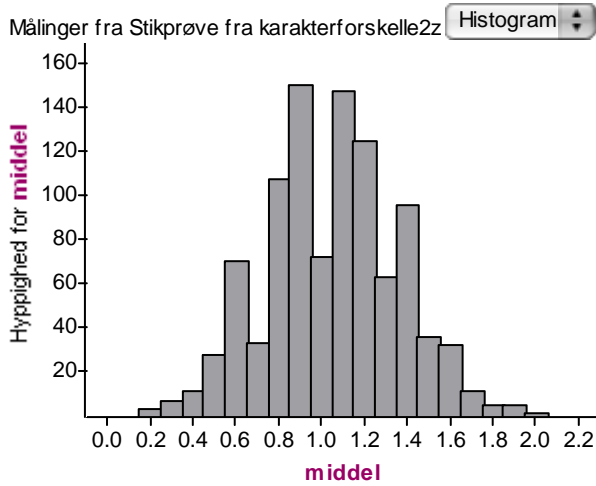
Firmaet kan igen trygt fortsætte!

Denne testtype stiller vores første test på terminsprøve- og eksamenskaraktererne i et nyt lys. I første omgang sammenlignede vi de to sæt karakterer uden at tage hensyn til, hvem der havde fået hvad. Så kunne vi ikke sige, at der var en signifikant forskel. Men naturligvis ved vi godt, hvordan det gik de enkelte elever, og det fremgår af følgende skema:

Termin	8	6	03	5	10	03	03	8	03	6	5	03	5	03	03	5	5
Eksamen	8	9	03	8	10	6	03	9	03	9	6	5	5	03	5	5	5

Her vil det være rimeligt, at danne et datamateriale bestående af de 17 differenser Eksamen-Termin: 0, 3, 0, 3, 0, 3, 0, 1, 0, 3, 1, 2, 0, 0, 2, 0, 0.

Her kan vi så ved bootstrapping teste nulhypotesen, som må være, at middelværdien er 0. 1000 gange udtager vi en stikprøve på 17 med tilbagelægning og måler hver gang middelværdien. Det giver følgende histogram og optælling:



Målinger fra Stikprøve fra kara...

	<b>middel</b>
	0
R1 = tæl (middel = 0)	

Der er ingen middelværdier på 0 (med dette talmateriale kan vi ikke få negative middelværdier, men i andre tilfælde kunne det naturligvis sagtens ske). Så når vi ser på de enkelte elevers karakterer, må vi sige, at eksamen er gået bedre end terminsprøven. Årsagen *kan* være, at terminsprøven har haft en gavnlig virkning, men det har vi ikke bevist; andre faktorer kunne være i spil. Men vi kan da glæde os sammen med de elever, det faktisk gik bedre!

# Chi i anden test ved



*Omrøring*

$\chi^2$

## Eksempler på eksperimentelle $X^2$ -test.

Eksemplerne er hentet i Mogens Lønborg: "Grundbog i statistik", Forlaget Lønborg 1993, kapitel 3, side 195-226.

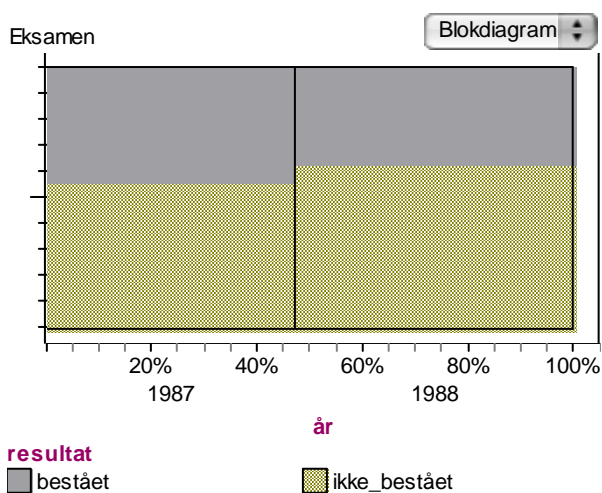
### Eksempel 1.

På en lærestanstalt blev i 1980'erne gennemført en besparelsesrunde, som ramte statistikundervisningen. Man frygtede en nedgang i det faglige niveau! Med spænding sammenlignede man eksamensresultatet i 1988, det første efter besparelsen, med resultatet året før. Ved eksamen i 1987 deltog 527, hvoraf de 238 bestod. I 1988 deltog 583, hvoraf 225 bestod. Vi vil nu teste, om eksamen gik dårligere i 1988 end i 1987.

Vi begynder med at indtaste vores data i en Datametertabel. Som det ses, kan man med fordel bruge formler for "år" og "resultat":

Eksamen			
	år	resultat	<ny>
=	<code>hvis (indeks ≤ 527) {</code> 1987 1988	<code>ombyt ("resultat")</code> { (indeks ≤ 238) : "bestået" (indeks ≤ 527) : "ikke_bestået" (indeks ≤ 752) : "bestået" ellers : "ikke_bestået"	
1	1987	bestået	
2	1987	bestået	
3	1987	bestået	
4	1987	bestået	

Vi kan lige se på et blokdiagram (træk "graf" ned, træk "år" til førsteaksen, og slip "resultat" i grafrummet; hold shift-tasten nede ved "år", så det bliver en kategorisk variabel):



Det kunne jo se ud som om, der var færre beståede i 1988, men er det statistisk signifikant? Vores nulhypotese er, at der ikke er forskel på de to år, altså at den forskel, vi kan se, skyldes tilfældigheder.

Vores teststørrelse er inspireret af  $X^2$ -testen.

Vi trækker de to variable ind i en beregningsboks, og vi anvender funktionen "Forventet" på tabellen (den ligger under specielle funktioner):

		resultat		Række total
		bestået	ikke_bestået	
år	1987	238	289	527
	219.82072	307.17928	527	
1988	225	358	583	
	243.17928	339.82072	583	
Søjle total		463	647	1110
		463	647	1110

R1 = tæl ( )

R2 = Forventet

Efter et højreklik i krydstabellen vælges "overfør celler til nyt datasæt". Det skaffer os nu en tabel:

Celler fra oversigtstabel for Eksamen

	år	resultat	R1	R2
1	1987	bestået	238	219.821
2	1987	ikke_best...	289	307.179
3	1988	bestået	225	243.179
4	1988	ikke_best...	358	339.821

og ud fra den kan vi beregne teststørrelsen:

Celler fra oversigtstabel for Eksamen

	4.910863

$$R1 = \text{sum} \left( \frac{(R1 - R2)^2}{R2} \right)$$

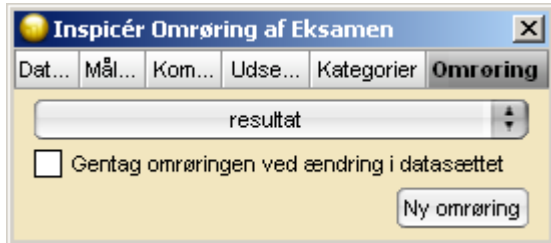
(Teknisk trak vi overskriften "Celler fra oversigtstabel for Eksamen" ind i en beregningsboks og redigerede formlen for R1.) Spørgsmålet er nu, om teststørrelsen 4.910863 er stor nok til, at vi kan forkaste nulhypotesen.



Vi højreklikker på den første skatkiste og vælger ”rør rundt i en variabel”:

Omrøring af Eksamen

Vi inspicerer:



Vi foretrækker at omrøre ”resultat”.

Vi ser på en tabel:

Omrøring af Eksamen

	år	resultat
1	1987	bestået
2	1987	bestået
3	1987	ikke_bestået
4	1987	bestået

og vi danner den tilsvarende krydstabel:

Omrøring af Eksamen

		resultat		Række total
		bestået	ikke_bestået	
år	1987	201	326	527
	219.82072	307.17928	527	
1988	262	321	583	
	243.17928	339.82072	583	
Søjle total		463	647	1110
		463	647	1110

R1 = tæl ( )

R2 = Forventet

Som før overfører vi celler til et nyt datasæt:



Celler fra oversigtstabel for Omrøring af Eksamen

som omsættes i en tabel:

Celler fra oversigtstabel for Omrøring af Eksamen

	år	resultat	R1	R2
1	1987	bestået	201	219.821
2	1987	ikke_best...	326	307.179
3	1988	bestået	262	243.179
4	1988	ikke_best...	321	339.821

og en beregning:

Celler fra oversigtstabel for Omrøring af Eksamen

	5.2635287

$$R1 = \text{sum} \left( \frac{(R1 - R2)^2}{R2} \right)$$



Nu vender vi tilbage til [Celler fra oversigtstabel for Omrøring af Eksamen](#) og i inspektøren bestiller vi målinger:

Måling	Værdi	Formel
Chi_i_anden	5.26353	$\text{sum} \left( \frac{(R1 - R2)^2}{R2} \right)$
<ny>		

Efter et højreklik på skatkisten bestilles ”udfør gentagne målinger”, som giver:



Målinger fra Celler fra oversigtstabel for Omrøring af Eksamen

og her stiller vi inspektøren til 1000 målinger:

Dat...	Mål...	Komme...	Udseende	Kategorier	Målinger
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="button" value="Ny målinger"/>					

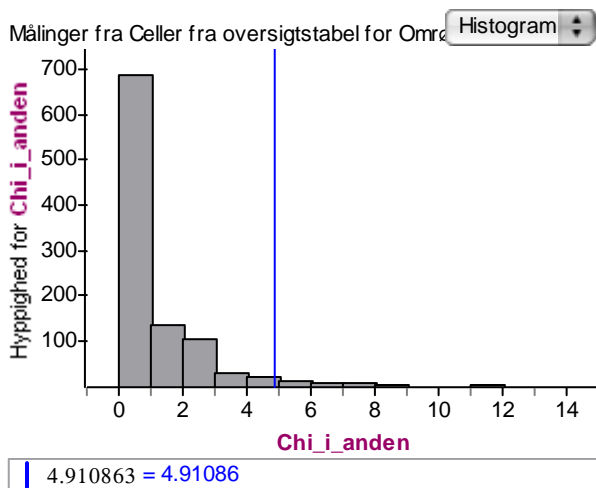
Herved foretages 1000 omrøringer med tilhørende beregning af teststørrelsen.

Vi ser på en tabel over målingerne (man kan se efter, at der faktisk er 1000!):

Målinger fra Celler fra oversigtstabel for Omrøring af Eksamen

	Chi_i_anden	<ny>
1	0.908862	
2	0.345325	
3	1.43315	
4	0.0492596	
5	0.118229	

I en graf vælger vi histogram:



Det spændende er nu, hvor meget af fordelingen, der ligger over 4.910863. Vi beregner:

Målinger fra Celler fra oversigtstabel for Omrøring af Eksamen

	0.029

$$R1 = \frac{\text{tæl}(\text{Chi}_i\text{anden} \geq 4.910863)}{\text{tæl}(\text{Chi}_i\text{anden})}$$

så svaret er 2.9 %.

Med et signifikansniveau på 5 % forkaster vi nulhypotesen.

Vi må dermed sige, at eksamen gik dårligere i 1988 end i 1987 (men vi har ikke vist, hvad årsagen var!).

Hvis man kender til en klassisk  $\chi^2$ -test, kan man bruge Datameters testværktøj:

Test fra Eksamen Test af uafhængighed

Første variabel (kategoriseret): resultat  
 Anden variabel (kategoriseret): år

		resultat		Række total
		bestået	ikke_bestået	
år	1987	238 (219.8)	289 (307.2)	527
	1988	225 (243.2)	358 (339.8)	583
Søjle total		463	647	1110

Første variabel: **resultat**  
 Antal kategorier: **2**  
 Anden variabel: **år**  
 Antal kategorier: **2**  
 Alternativ hypotese: Der er en sammenhæng mellem **resultat** og **år**

Teststørrelsen, chi-i-anden, er **4.911**. Der er **1** frihedsgrader (antallet af rækker minus én ganget med antallet af søjler minus én).

Hvis det var sandt at **resultat** var uafhængig af **år** (nulhypotesen), og stikprøven blev gentaget mange gange, så ville sandsynligheden for at få en værdi for chi-i-anden, der var mindst lige så stor være **0.027**.

Tallene i parentes i tabellen er de forventede antal.

Her bliver sandsynligheden for at få en mere ekstrem værdi af teststørrelsen 2.7 %, så resultatet bliver det samme.

### Eksempel 2.

I en markedsundersøgelse ønsker man at analysere sammenhængen mellem kundernes erhvervsmæssige stilling og deres mærkeloyalitet over for et bestemt produkt. 230 har svaret; de er fordelt på selvstændig, lønmodtager og andet. Loyaliteten måles som høj, moderat eller lav. Nulhypotesen er, at der ikke er nogen sammenhæng mellem den erhvervsmæssige stilling og mærkeloyaliteten.

Datamaterialet er følgende:

	Høj	Moderat	Lav	I alt
Selvstændig	30	42	18	90
Lønmodtager	14	20	31	65
Andet	34	25	16	75
I alt	78	87	65	230

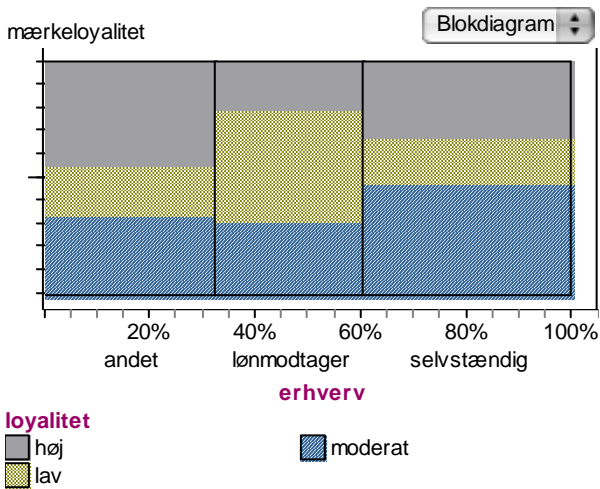
Kun hovedresultaterne vises.

Ved indtastningen bruges med fordel ombyt-funktionen (under betingede funktioner):

mærkeloyalitet

	erhverv	loyalitet
=	ombyt ("erhverv") { <ul style="list-style-type: none"> <li>(indeks ≤ 90) : "selvstændig"</li> <li>(indeks ≤ 155) : "lønmotager"</li> <li>ellers : "andet"</li> </ul>	ombyt (loyalitet) { <ul style="list-style-type: none"> <li>(indeks ≤ 30) : "høj"</li> <li>(indeks ≤ 72) : "moderat"</li> <li>(indeks ≤ 90) : "lav"</li> <li>(indeks ≤ 104) : "høj"</li> <li>(indeks ≤ 124) : "moderat"</li> <li>(indeks ≤ 155) : "lav"</li> <li>(indeks ≤ 189) : "høj"</li> <li>(indeks ≤ 214) : "moderat"</li> <li>ellers : "lav"</li> </ul>
1	selvstændig	høj
2	selvstændig	høj
3	selvstændig	høj

Et boksdiagram viser en tydelig forskel; er den stor nok til at forkaste nulhypotesen?



mærkeloyalitet

		loyalitet			Række total
		lav	høj	moderat	
erhverv	lønmodtager	31 18.369565	14 22.043478	20 24.586957	65 65
	andet	16 21.195652	34 25.434783	25 28.369565	75 75
	selvstændig	18 25.434783	30 30.521739	42 34.043478	90 90
Søjle total		65	78	87	230
		65	78	87	230

R1 = tæl ( )

R2 = Forventet

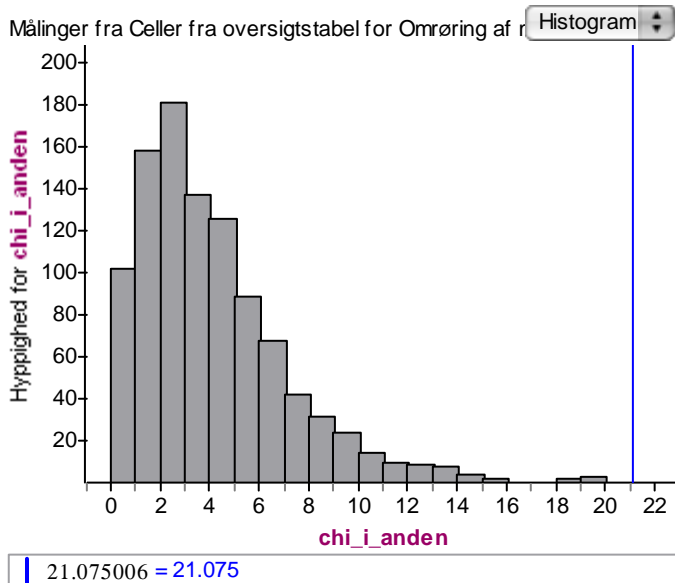
Vi ender med teststørrelsen:

Celler fra oversigtstabel for mærkeloyalitet

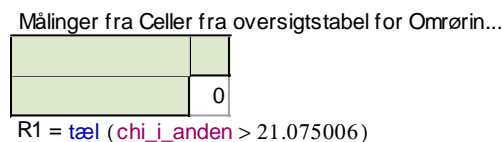
	21.075006

$$R1 = \text{sum} \left( \frac{(R1 - R2)^2}{R2} \right)$$

Derefter omrører vi 1000 gange og måler hver gang teststørrelsen. Histogrammet bliver:

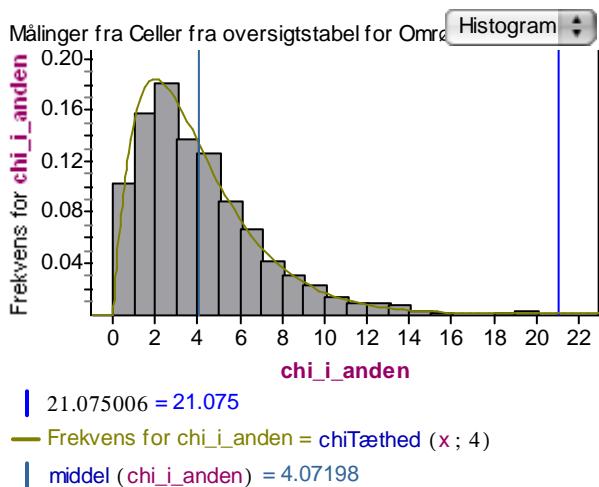


Vores teststørrelse synes at ligge helt til højre for fordelingen, og det bekræftes ved en beregning:



Vi må klart forkaste nulhypotesen. Der synes altså at være en sammenhæng mellem erhverv og mærkeloyalitet.

Vi prøver at højreklikke i histogrammet og som histogramskala vælge Frekvens. Så kan vi tegne tæthedsfunktionen for  $\chi^2$ -fordelingen med 4 frihedsgrader, og den passer jo fint. Middelværdien ligger tæt ved antallet af frihedsgrader.



Et klassisk test bekræfter vores resultat:

Test fra mærkeloyalitet Test af uafhængighed

Første variabel (kategoriseret): loyalitet  
Anden variabel (kategoriseret): erhverv

		loyalitet			Række total
		høj	lav	moderat	
erhverv	andet	34 (25.4)	16 (21.2)	25 (28.4)	75
	lønmodtager	14 (22.0)	31 (18.4)	20 (24.6)	65
	selvstændig	30 (30.5)	18 (25.4)	42 (34.0)	90
Søjle total		78	65	87	230

Første variabel: **loyalitet**  
Antal kategorier: **3**  
Anden variabel: **erhverv**  
Antal kategorier: **3**  
Alternativ hypotese: Der er en sammenhæng mellem **loyalitet** og **erhverv**

Teststørrelsen, chi-i-anden, er **21.08**. Der er **4** frihedsgrader (antallet af rækker minus én ganget med antallet af søjler minus én).

Hvis det var sandt at **loyalitet** var uafhængig af **erhverv** (nulhypotesen), og stikprøven blev gentaget mange gange, så ville sandsynligheden for at få en værdi for chi-i-anden, der var mindst lige så stor være **0.00031**.

Tallene i parentes i tabellen er de forventede antal.

### Eksempel 3.

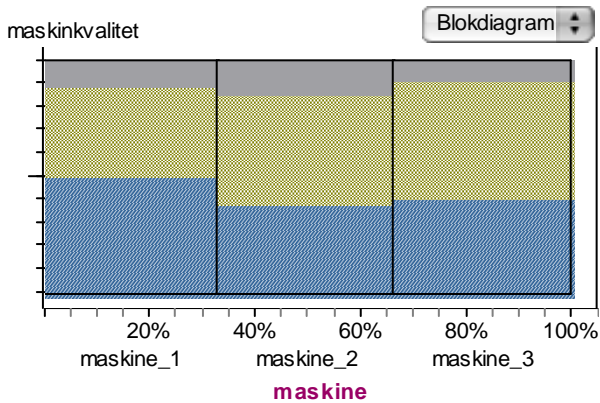
Tre maskiner kvalitetstestes; ved en stikprøve på 150 produkter fra hver maskine er kategorierne

1. og 2. sortering samt kassabel. Data er følgende:

	1. sortering	2. sortering	Kassabel	I alt
Maskine 1	58	74	18	150
Maskine 2	70	56	24	150
Maskine 3	75	60	15	150
I alt	203	190	57	450

Nulhypotesen er, at maskinerne producerer ens med hensyn til kvalitet.

maskinkvalitet		maskine	kvalitet	<ny>
=	ombyt ("maskine")	{ (indeks ≤ 150) : "maskine_1" (indeks ≤ 300) : "maskine_2" ellers : "maskine_3"	ombyt ("kvalitet")                     { (indeks ≤ 58) : "sort1" (indeks ≤ 132) : "sort2" (indeks ≤ 150) : "kassabel" (indeks ≤ 220) : "sort1" (indeks ≤ 276) : "sort2" (indeks ≤ 300) : "kassabel" (indeks ≤ 375) : "sort1" (indeks ≤ 435) : "sort2" ellers : "kassabel"	
1		maskine_1	sort1	
2		maskine_1	sort1	
3		maskine_1	sort1	



Er forskellen stor nok til at forkaste nulhypotesen?

maskinkvalitet		kvalitet			Række total
		kassabel	sort1	sort2	
maskine	maskine_1	18	58	74	150
	maskine_2	24	70	56	150
	maskine_3	15	75	60	150
Søjle total		57	203	190	450
		57	203	190	450

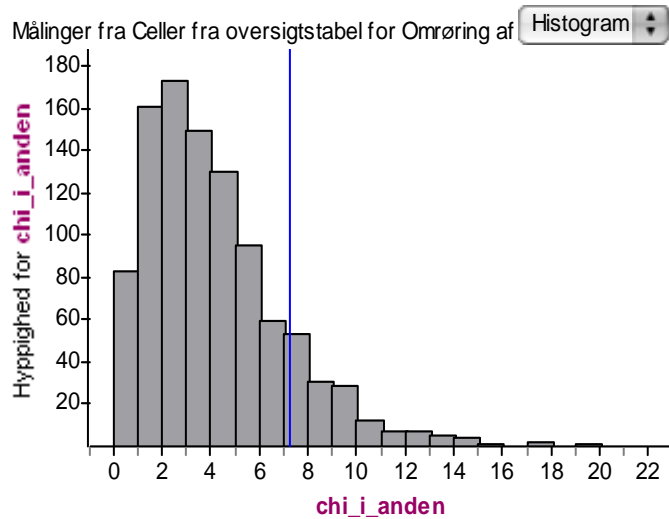
Celler fra oversigtstabel for maskinkvalitet

	7.2877366

R1 = tæl ( )  
 R2 = Forventet

$$R1 = \text{sum} \left( \frac{(R1 - R2)^2}{R2} \right)$$

Teststørrelsen er altså 7.2877366. Så omrøres og måles og resultatet bliver:



| 7.2877366 = 7.28774

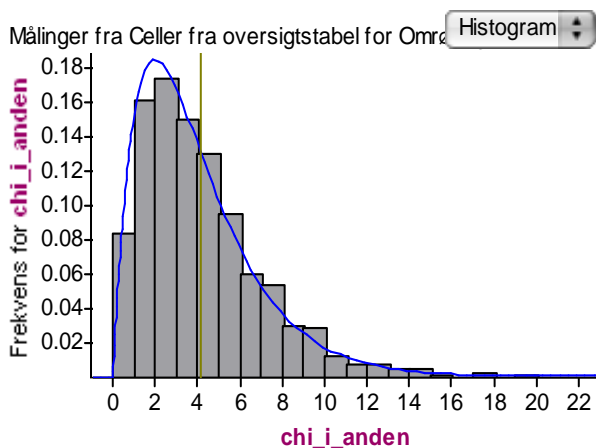
Målinger fra Celler fra oversigtstabel for Omrørin...

	0.131

$$R1 = \frac{\text{tæl}(\chi_{i\_anden} \geq 7.2877366)}{\text{tæl}(\chi_{i\_anden})}$$

13.1 % af fordelingen ligger over teststørrelsen, så nulhypotesen kan ikke forkastes. Vi må regne med, at der ikke er forskel på maskinerne.

Som i foregående eksempel tilføjer vi en tæthedsfunktion og middelværdi samt et klassisk test:



— Frekvens for  $\chi_{i\_anden} = \chi_{i\_anden}$  (x; 4)

| middel ( $\chi_{i\_anden}$ ) = 4.14329

Test fra maskinkvalitet

Test af uafhængighed

Første variabel (kategoriseret): kvalitet

Anden variabel (kategoriseret): maskine

		kvalitet			Række total
		kassabel	sort1	sort2	
maskine	maskine_1	18 (19.0)	58 (67.7)	74 (63.3)	150
	maskine_2	24 (19.0)	70 (67.7)	56 (63.3)	150
	maskine_3	15 (19.0)	75 (67.7)	60 (63.3)	150
Søjle total		57	203	190	450

Første variabel: **kvalitet**Antal kategorier: **3**Anden variabel: **maskine**Antal kategorier: **3**Alternativ hypotese: Der er en sammenhæng mellem **kvalitet** og **maskine**

Teststørrelsen, chi-i-anden, er **7.288**. Der er **4** frihedsgrader (antallet af rækker minus én ganget med antallet af søjler minus én).

Hvis det var sandt at **kvalitet** var uafhængig af **maskine** (nulhypotesen), og stikprøven blev gentaget mange gange, så ville sandsynligheden for at få en værdi for chi-i-anden, der var mindst lige så stor være **0.12**.

Tallene i parentes i tabellen er de forventede antal.

#### Eksempel 4.

En eksperimentel test kan også gennemføres ved et såkaldt univariat datamateriale.

(Lønborg side 133.)

I en produktionsproces klassificeres de færdige produkter i 4 kvalitetskategorier: 1., 2., 3. og 4.

sortering. Enheder i 4. sortering kasseres, medens enheder af 3. sortering sendes til reparation.

Øvrige enheder sælges umiddelbart. Med henblik på at undersøge produktionsprocessen udtages en

stikprøve på i alt 1301 produkter. Idet vi betegner de 4 kategorier ved  $x_1$ ,  $x_2$ ,  $x_3$  og  $x_4$ , har vi

følgende datamateriale:

$X_1$	$X_2$	$X_3$	$X_4$	I alt
773	231	238	59	1301

Hvis maskinen er korrekt indstillet, skal fordelingen på de fire kategorier være som forholdstallene

9 : 3 : 3 : 1. Nulhypotesen er, at maskinen er indstillet sådan.

Allerførst udregnes teststørrelsen inspireret af  $\chi^2$ -testen:

teststørrelsen

	antal	forv	<ny>
1	773	731.8130	
2	231	243.9380	
3	238	243.9380	
4	59	81.3125	

teststørrelsen

	9.271433

$$R1 = \text{sum} \left( \frac{(\text{antal} - \text{forv})^2}{\text{forv}} \right)$$

Testen gennemføres nu ved, at vi indlæser et datamateriale med 9 ettaller, 3 totaller, 3 tretaller og et firtal, altså en "ideel" stikprøve på 16 fra maskinen:

kvalitet

	produktkvalitet	<ny>
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	1	
8	1	
9	1	
10	2	

Med højreklik på ikonen bestilles en stikprøve (med tilbagelægning!) på 1301:

Stikprøve fra kvalitet

	produktkvalitet	<ny>
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7	4	
8	1	
9	1	
10	4	
11	4	
12	1	
13	1	
14	2	
15	2	

I inspektøren bestilles en måling af teststørrelsen:

Måling	Værdi	Formel
chi_i_anden	1.97906	$\frac{(\text{teel}(\text{produktkvalitet} = 1) - 731.8)^2}{731.8} + \frac{(\text{teel}(\text{produktkvalitet} = 2) - 243.9)^2}{243.9} + \frac{(\text{teel}(\text{produktkvalitet} = 3) - 243.9)^2}{243.9} + \frac{(\text{teel}(\text{produktkvalitet} = 4) - 81.3)^2}{81.3}$
<ny>		

og vi foretager 1000 målinger ("udfør gentagne målinger" på stikprøven)

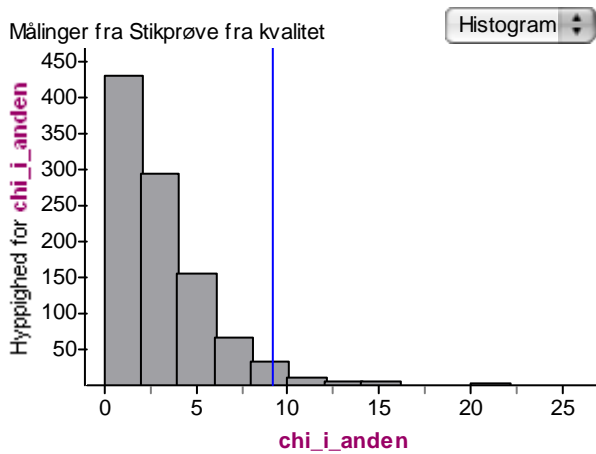
Data	Målinger	Kommentarer	Udseende	Kategorier	Målinger
<input type="checkbox"/>	Animation slået til				
<input checked="" type="checkbox"/>	Erstat de eksisterende målinger				
<input type="checkbox"/>	Gentag målinger ved ændring i datasættet				
<input checked="" type="radio"/>	1000 målinger				
<input type="radio"/>	Stopkriterium				

De skrives i en tabel:

Målinger fra Stikprøve fra kvalitet

	chi_i_anden	<ny>
1	3.82165	
2	4.06133	
3	0.531789	
4	2.61073	
5	2.60065	
6	2.37544	
7	3.36326	
8	3.45592	
9	0.795004	
10	3.80277	

som grafisk giver



| 9.2 = 9.2

Målinger fra Stikprøve fra kvalitet

<b>chi_i_anden</b>	29
--------------------	----

R1 = tæl (**chi\_i\_anden** ≥ 9.271433)

og beregning

Sandsynligheden for at få en større værdi end teststørrelsen 9.271433 er altså 2.9 %. På 5 % signifikansniveau forkaster vi nulhypotesen, og maskinen kan ikke anses for at være korrekt indstillet.

# Variansanalyse ved



*Omrøring...*

## Variansanalyse i Datameter.

I variansanalyse sammenligner man datamateriale, der er opdelt i flere grupper. Er der kun to grupper, bliver der tale om et almindeligt omrørningstest, så vi vil se på eksempler med flere grupper. I klassisk statistik gør man forskellige antagelser om materialet, bl.a. at der er tale om normalfordelt materiale. I første omgang ser vi på det, der i klassisk statistik hedder **ensidet variansanalyse**. Problemstillingen belyses bedst ved eksempler, så vi kaster os ud i det.

### Eksempel 1.

(Fra Mogens Lønborg: "Grundbog i Statistik", Forlaget Lønborg 1993, side 274).

På en gødningsfabrik har man fra hver af 4 produktionsgrene udtaget en stikprøve af færdige gødningsprodukter og målt det procentvise indhold af kali. Resultatet blev:

Produktionsgren	Kali-indhold i %					
1	11.9	9.0	10.2	8.5	13.6	11.3
2	15.9	17.2	21.0	17.1	19.0	23.2
3	13.9	14.6	16.1	14.9	13.7	12.4
4	15.6	18.3	14.6	16.9	14.7	

Spørgsmålet om variabelsammenhæng skal vurderes på grundlag af en sammenligning mellem det gennemsnitlige kaliindhold for hver af de fire produktionsgrene.

Nulhypotesen er, at der ikke er forskel, altså middelværdierne er ens.

Den klassiske teststørrelse er

$$\frac{\sum n_i * (\bar{y}_i - \bar{y})^2 / (k - 1)}{\sum (y_{ij} - \bar{y}_i)^2 / (n - k)}$$

$n$  er antal observationer i alt,  $n_i$  er antallet i produktionsgren  $i$ .  $k$  er antal

grene. Her er  $k = 4$  og  $n = 23$  og  $n_i$ 'erne er 6, 6, 6 og 5.  $k-1$  og  $n-k$  er kun divideret på af hensyn til teststørrelsens fordeling, så vi vil se bort fra dem og altså bruge

$$\frac{\sum n_i * (\bar{y}_i - \bar{y})^2}{\sum (y_{ij} - \bar{y}_i)^2}$$

En stor tæller betyder stor forskel på gruppegennemsnittene (gennemsnit for

hver produktionsgren) og totalgennemsnittet, og det trækker væk fra nulhypotesen. En lille nævner betyder lille variation inden for grupperne, altså et homogent materiale i grupperne, hvorved man måler forskellene skarpere, og det trækker også væk fra nulhypotesen.

Altså: jo større brøk, jo værre for nulhypotesen.

Man kan også opfatte det sådan, at testen måler variationen mellem grupperne i forhold til variationen inden for grupperne.

Vi udregner først teststørrelsen:

Kali

	grene	kali	gruppemiddel	kvadrater
=			ombyt ( gruppemiddel) { ( grene = "gren1" ) : middel ( kali; grene = "gren1" ) ( grene = "gren2" ) : middel ( kali; grene = "gren2" ) ( grene = "gren3" ) : middel ( kali; grene = "gren3" ) ellers : middel ( kali; grene = "gren4" )	( kali – gruppemiddel ) <sup>2</sup>
1	gren1	11.9	10.75	1.3225
2	gren1	9.0	10.75	3.0625
3	gren1	10.2	10.75	0.3025
4	gren1	8.5	10.75	5.0625
5	gren1	13.6	10.75	8.1225
6	gren1	11.3	10.75	0.3025
7	gren2	15.9	18.9	9
8	gren2	17.2	18.9	2.89
9	gren2	21.0	18.9	4.41
10	gren2	17.1	18.9	3.24
11	gren2	19.0	18.9	0.01
12	gren2	23.2	18.9	18.49
13	gren3	13.9	14.2667	0.134444
14	gren3	14.6	14.2667	0.111111
15	gren3	16.1	14.2667	3.36111
16	gren3	14.9	14.2667	0.401111
17	gren3	13.7	14.2667	0.321111
18	gren3	12.4	14.2667	3.48444
19	gren4	15.6	16.02	0.1764
20	gren4	18.3	16.02	5.1984
21	gren4	14.6	16.02	2.0164
22	gren4	16.9	16.02	0.7744
23	gren4	14.7	16.02	1.7424

Kali

	2.8129397

$$R1 = \frac{\text{sum}((\text{gruppemiddel} - \text{middel}(\text{kali}))^2)}{\text{sum}(\text{kvadrater})}$$

Så teststørrelsen er 2.8129397.

Man er nu nødt til at slette de variable gruppemiddel og kvadrater. Så omrøres kali, og i omrøringen genindføres gruppemiddel og kvadrater med samme formler (uden den manøvre følger de ikke med i omrøringsprocessen!):


  
Omrøring af Kali




Omrøring af Kali

	grene	kali	gruppemiddel	kvadrater
=			$\text{ombyt ( gruppemiddel) } \begin{cases} (\text{grene} = \text{"gren1"}): \text{middel (kali; grene} = \text{"gren1"}) \\ (\text{grene} = \text{"gren2"}): \text{middel (kali; grene} = \text{"gren2"}) \\ (\text{grene} = \text{"gren3"}): \text{middel (kali; grene} = \text{"gren3"}) \\ \text{ellers} & : \text{middel (kali; grene} = \text{"gren4"}) \end{cases}$	$(\text{kali} - \text{gruppemiddel})^2$
1	gren1	21.0	15.4167	31.1736
2	gren1	11.3	15.4167	16.9469
3	gren1	13.9	15.4167	2.30028
4	gren1	8.5	15.4167	47.8403
5	gren1	14.6	15.4167	0.666944
6	gren1	23.2	15.4167	60.5803

Vi måler ved hver omrøring teststørrelsen:

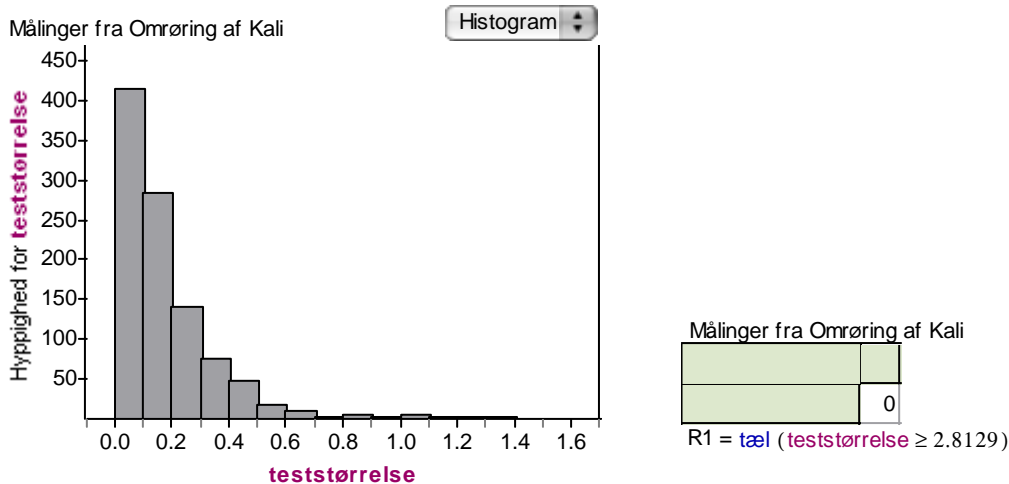


og vi måler 1000 gange:



Målinger fra Omrøring af Kali

	teststørrelse
1	0.267158
2	0.0758135
3	0.211936
4	0.0775543



Der er ingen tvivl: Nulhypotesen forkastes, så der er forskel på kaliindholdet i de forskellige produktionsgrene.

Vi prøver lige et klassisk test (her skal teststørrelsen jo være  $\frac{19}{3} * 2.8129397 = 17.81528477$ ):

Test fra Kali

Variansanalyse

Responsevariabel (numerisk): kali  
Grupperingsvariabel (kategoriseret): grene

Kilde	Friheds- grader	Kvadrat- sum	Middel- kvadrat	F- statistik	p- værdi
<b>Gruppe</b>	3	207.978	69.3261	17.815	0.0000
<b>Fejl</b>	19	73.936	3.8914		
<b>Total</b>	22	281.915			

Alternativ hypotese: Populationsmiddelværdierne for **kali** grupperet efter **grene** er ikke lige store.

Hvis det var sandt at populationsmiddelværdierne for **kali** var lige store (nulhypotesen) og stikprøverne blev gentaget mange gange, så ville sandsynligheden for at få en værdi for F der var mindst lige så stor som teststørrelsen **17.8153** være **< 0.0001**.

Resultatet bliver det samme.

## Eksempel 2.

Til højspændingsledninger skal der anvendes ensartede kabler med høj trækstyrke. Hvert kabel består af wirer, der fremstilles i én bestemt længde. I følgende tabel er anført trækstyrken i kg af wirerne i 5 kabler, der hver består af 12 wirer. Det skal undersøges, om de 5 kabler kan antages at have samme trækstyrke.

Kabel i	$X_{ij}$											
1	347	341	345	340	350	346	345	342	340	339	330	338
2	341	340	335	336	339	340	342	345	341	338	346	347
3	339	340	342	341	336	342	347	345	341	340	336	342
4	339	340	347	345	350	348	341	342	337	346	340	345
5	342	346	347	348	355	351	333	347	350	347	348	341

Nulhypotesen er, at middelværdierne i de fem grupper er ens. (Vi skønner, at grupperne har ensartede varianser; i klassisk variansanalyse er der en særlig test herfor.)

Vi indtaster i Datameter og beregner teststørrelsen:

trækstyrke		kabel	styrke	gruppemiddel	kvadrater	
=				ombyt ( gruppemiddel ) { ( kabel = "kabel1" ) : middel ( styrke; kabel = "kabel1" ) ( kabel = "kabel2" ) : middel ( styrke; kabel = "kabel2" ) ( kabel = "kabel3" ) : middel ( styrke; kabel = "kabel3" ) ( kabel = "kabel4" ) : middel ( styrke; kabel = "kabel4" ) ellers : middel ( styrke; kabel = "kabel5" )	( styrke – gruppemiddel ) <sup>2</sup>	
1	kabel1	347			341.917	25.8403
2	kabel1	341			341.917	0.840278
3	kabel1	345			341.917	9.50694
4	kabel1	340			341.917	3.67361

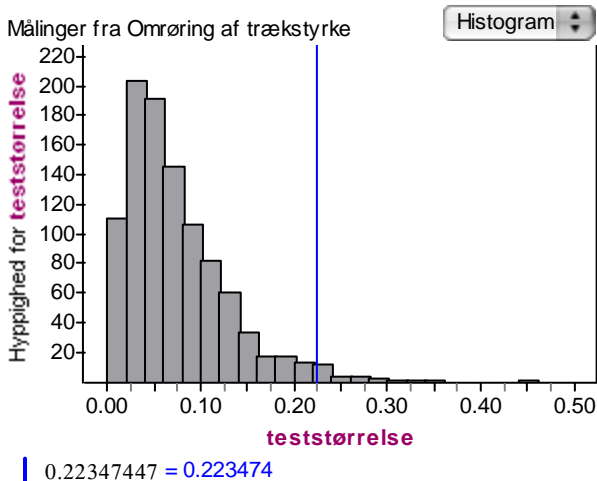
trækstyrke	
	0.22347447

$$R1 = \frac{\text{sum} ( ( \text{gruppemiddel} - \text{middel} ( \text{styrke} ) )^2 )}{\text{sum} ( \text{kvadrater} )}$$

Teststørrelsen er 0.22347447.

Vi omrører styrke 1000 gange, samtidig med at vi måler teststørrelsen hver gang (husk det tekniske med, at gruppemiddel og kvadrater skal fjernes inden omrøringen og efter første omrøring genindføres med samme formler; ellers deltager de ikke i omrøringsprocessen.)

Resultatet bliver for teststørrelsens vedkommende:



Målinger fra Omrøring af trækstyrke

	19

R1 = tæl (teststørrelse ≥ 0.22347447)

Kun 1.9 % bliver større end den oprindelige teststørrelse. Vi må derfor regne med, at kablernes trækstyrker ikke er ens.

Også her prøver vi lige et klassisk test (teststørrelsen skal blive  $0.22347447 * 55/4 = 3.073$ ):

Test fra trækstyrke

Responsevariabel (numerisk): styrke  
Grupperingsvariabel (kategoriseret): kabel

Kilde	Frihedsgrader	Kvadratsum	Middelkvadrat	F-statistik	p-værdi
Gruppe	4	243.23	60.8083	3.073	0.0235
Fejl	55	1088.42	19.7894		
Total	59	1331.65			

Alternativ hypotese: Populationsmiddelværdierne for **styrke** grupperet efter **kabel** er ikke lige store.

Hvis det var sandt at populationsmiddelværdierne for **styrke** var lige store (nulhypotesen) og stikprøverne blev gentaget mange gange, så ville sandsynligheden for at få en værdi for F der var mindst lige så stor som teststørrelsen **3.07277** være **0.023**.

Resultatet bliver det samme.

### Eksempel 3.

Det tredje eksempel er fra ornitologien. Man har en efterhånden velbegrunnet antagelse om, at gøge specialiserer sig i valget af adoptivforældre til udrugning af æg og opfostring af deres unger. Gøge af en bestemt gruppe lægger normalt æg i en bestemt fuglearts reder, og deres æg har udseende og størrelse tilpasset deres valg af adopterende fuglearter.

I tabellen nedenfor er gengivet nogle målinger af bredden af gøgeæg fundet i reder fra fire forskellige fuglearter. Det skal undersøges, om disse data støtter ornitologernes antagelse.

Nulhypotesen er, at der ikke er forskel på ægstørrelsen hos de forskellige arter.

Adopterende fugleart	Bredde i halve millimeter					
	30	31	32	33	34	35
Gærdesmutte	3	11	19	7	10	4
Tornsanger	0	5	13	11	6	1
Kærsanger	0	3	14	11	9	0
Rødstjert	2	2	17	19	22	2

I Datameter bliver det:

gøgeæg

	fugl	æg	gruppemiddel	kvadrater
=	ombyt ("fugl") { (indeks ≤ 54) : "gærdesmutte" (indeks ≤ 90) : "tornsanger" (indeks ≤ 127) : "kærsanger" ellers : "rødstjert"		ombyt (gruppemiddel) { (fugl = "gærdesmutte") : middel (æg; fugl = "gærdesmutte") (fugl = "tornsanger") : middel (æg; fugl = "tornsanger") (fugl = "kærsanger") : middel (æg; fugl = "kærsanger") ellers : middel (æg; fugl = "rødstjert")	(æg - gruppemiddel) <sup>2</sup>
1	gærdesmutte	30	32.4074	5.79561
2	gærdesmutte	30	32.4074	5.79561
3	gærdesmutte	30	32.4074	5.79561
4	gærdesmutte	31	32.4074	1.9808
5	gærdesmutte	31	32.4074	1.9808

gøgeæg

	0.043076219

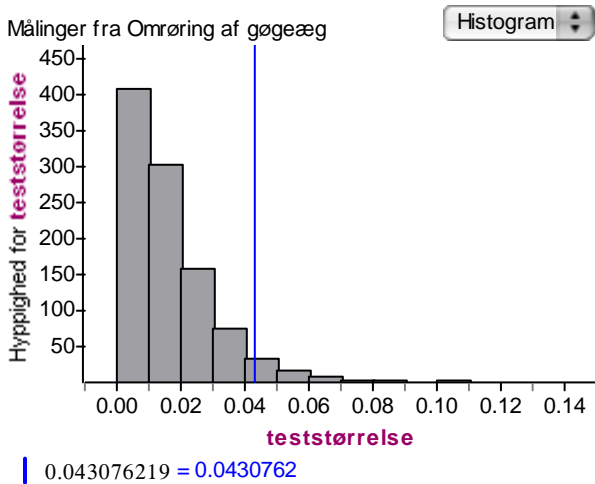
$$R1 = \frac{\text{sum}((\text{gruppemiddel} - \text{middel}(\text{æg}))^2)}{\text{sum}(\text{kvadrater})}$$

Teststørrelsen:

Æg omrøres (gruppemiddel og kvadrater væk og ind igen i omrøringen!), og teststørrelsen måles:

Inspicér Omrøring af gøgeæg		
Data		
Måling	Værdi	Formel
teststørrelse	0.005790...	$\frac{\text{sum}((\text{gruppemiddel} - \text{middel}(\text{æg}))^2)}{\text{sum}(\text{kvadrater})}$
<ny>		

Resultatet bliver:



Målinger fra Omrøring af gøgeæg

	44

R1 = tæl (teststørrelse  $\geq$  0.043076219)

Altså er 4.3 % af teststørrelserne større end den oprindelige. På 5 % signifikansniveau vil vi forkaste nulhypotesen, så vi må sige, at talmaterialet støtter antagelsen om, at gøgen specialiserer sig i valget af adoptivforældre.

Vi ser også det klassiske test (teststørrelsen skal blive  $0.043076219 \cdot 187/3 = 2.685084310$ ).

Test fra gøgeæg

Responsevariabel (numerisk): æg

Grupperingsvariabel (kategoriseret): fugl

Kilde	Frihedsgrader	Kvadratsum	Middelkvadrat	F-statistik	p-værdi
Gruppe	3	10.274	3.4246	2.685	0.0480
Fejl	187	238.501	1.2754		
Total	190	248.775			

Alternativ hypotese: Populationsmiddelværdierne for **æg** grupperet efter **fugl** er ikke lige store.

Hvis det var sandt at populationsmiddelværdierne for **æg** var lige store (nulhypotesen) og stikprøverne blev gentaget mange gange, så ville sandsynligheden for at få en værdi for F der var mindst lige så stor som teststørrelsen **2.68508** være **0.048**.

4.8 % er stadig under 5 %, så resultatet bliver det samme.

Der tilføjes et par billeder og et lille uddrag fra "Natur og Museum", marts 2004, hvor biologen Poul Hansen skriver om nyere gøgeforskning. I det pågældende nummer kan der læses meget mere om emnet.

Senere blev rekorden slået eftertrykkeligt uden menneskelig hjælp af en gøgehun i Oxfordshire i England. Hun opererede i et område med 36 rørsangerpar og lagde på et år 25 æg i reder hos 24 af parrene. Et af parrene måtte tåle hendes snylten to gange.

Selvom Chance's eksperimenter i dag ville være helt utilladelige, må vi beundre hans ihærdighed og den viden, han opnåede under sin rekordjagt. I øvrigt var Chance den første til at dokumentere, hvordan gøgehunnes æglægning foregår. Han satte fotoskjul op ved sandsynlige nye æglægningsreder og ventede i skjulet med et filmsapparat på, at en gøgehun skulle ankomme. Det lykkedes ham i flere tilfælde at forevige æglægningen.

#### Gøgeæggets udseende

Edgar Chance's studier viste, at hans to gøgehunner foretrak at lægge æg i engpiberreder, selvom der i området var masser af reder af andre arter, som kunne være brugt. Senere studier har bekræftet, at gøgehunner helst holder sig til en bestemt værtsfugleart. Man

taler ligefrem om gøgestammer: Engpibergøge, rørsangergøge, rødstjertegøge, jernspurvegøge osv. I Europa er der mindst 15 forskellige gøgestammer. Æggene hos engpibergøge, rørsangergøge og rødstjertegøge ligner oftest værtsfuglernes temmelig godt bortset fra en lille størrelsesforskel.

Når æggene hos de forskellige gøgestammer ligner værtsfuglernes så meget, må det skyldes, at værtsfuglene ser med kritiske øjne på afvigende æg. Opdager en rørsanger et æg, der afviger for meget fra dens egne, vil det blive smidt ud, eller også forlades reden, hvorefter parret begynder på et nyt kuld et andet sted. Kun gøgehunner, som lægger æg, der ikke stikker for meget i øjnene på værtsfuglene, vil få afkom. Omvendt vil værtsfugle, der er i stand til at opdage og afvise et gøgeæg, få mere afkom, end mere ukritiske værtsfugle. Hvis den kritiske sans nedarves, vil værtsfuglene med tiden blive mere kritiske, mens gøgehunnerne, hvis ægfarven nedarves, vil lægge mere og mere velligende æg (se næste side).

En undtagelse fra reglen er jernspurvegøgene. Deres æg ser helt anderledes ud end jernspurvens smukt blå æg. Alligevel accepterer jernspurven æggene. Allerede Chaucer nævner jernspurven som gøgevært i 1382. Noget tyder altså på, at det kan tage mange generationer at udvikle en kritisk sans over for afvigende gøgeæg. En vigtig detalje i den forbindelse er, i hvor høj grad jernspurven udsættes for snylten fra gøgen. Hvis kun en lille procentdel af jernspurvebestanden parasiteres, vil en mutant, som er mere kritisk over for afvigende æg, spredes meget langsomt i populationen. De afvigende æg hos jernspurvegøgene skyldes i øvrigt ikke en manglende evne hos gøgen til at producere en blå ægfarve. Rødstjertegøgene i Finland lægger smukt blå æg, der ligner rødstjertens til forveksling.

▼ Gøgeæg i reder af rørsanger, jernspurv og rødstjert. Jernspurven accepterer helt afvigende æg. Fotos: N. B. Davies, W. B. Carr og J. Rutila.



## Tosidet variansanalyse.

I tosidet variansanalyse er data inddelt efter to kriterier. I klassisk statistik gøres forskellige antagelser om normalfordelt materiale med ensartede varianser. Variansanalysen består her af flere på hinanden følgende test. Vi må som sædvanligt bruge de traditionelle teststørrelser. De har komplicerede matematiske udledninger, men man kan godt se på dem, at de måler på det ”rigtige”. Ligesom ved ensidet variansanalyse er nogle af konstanterne i teststørrelserne egentlig overflødige ved et omrøringstest, men de bevares her, så man umiddelbart kunne sammenligne med en klassisk test.

Det hele belyses bedst ved et eksempel, så vi går i gang.

### Eksempel 4.

Vi bruger et godt gammelt eksempel fra A. Hald: ”Statistiske metoder”, 1968 (men 1. oplag var fra 1948), side 365. Data består af logaritmen til permeabiliteten (vandgennemtrængningstiden) i sekunder for 81 plader udtaget gennem 9 dage fra 3 maskiner (at logaritmen tages skyldes, at man så erfaringsmæssigt får normalfordelt materiale; det ville ikke være nødvendigt ved en omrøringstest). Hvert tal er middeltal af 8 målinger, og vi ved også, at de tre maskiner tilføres råmaterialer fra et fælles lager. Vi har altså to inddelingskriterier, dage og maskiner, og der er 3 data i hver gruppe, i alt altså 81 data:

Dage	Maskiner		
	1	2	3
1	1.404	1.306	1.932
	1.346	1.628	1.674
	1.618	1.410	1.399
2	1.447	1.241	1.426
	1.569	1.185	1.768
	1.820	1.516	1.859
3	1.914	1.506	1.382
	1.477	1.575	1.690
	1.894	1.649	1.361
4	1.887	1.673	1.721
	1.485	1.372	1.528
	1.392	1.114	1.371

5	1.772	1.227	1.320
	1.728	1.397	1.489
	1.545	1.531	1.336
6	1.665	1.404	1.633
	1.539	1.452	1.612
	1.680	1.627	1.359
7	1.918	1.229	1.328
	1.931	1.508	1.802
	2.129	1.436	1.385
8	1.845	1.583	1.689
	1.790	1.627	2.248
	2.042	1.282	1.795
9	1.540	1.636	1.703
	1.428	1.067	1.370
	1.704	1.384	1.839

Først lægges materialet ind i Datameter. Ved hjælp af ”dage” og ”maskiner” kan man identificere data inden for hver gruppe:

Permeabilitet

	dage	maskiner	perm	<ny>
1	dag1	maskine1	1.404	
2	dag1	maskine1	1.346	
3	dag1	maskine1	1.618	
4	dag1	maskine2	1.306	
5	dag1	maskine2	1.628	
6	dag1	maskine2	1.410	
7	dag1	maskine3	1.932	
8	dag1	maskine3	1.674	
9	dag1	maskine3	1.399	
10	dag2	maskine1	1.447	
11	dag2	maskine1	1.569	
12	dag2	maskine1	1.820	
13	dag2	maskine2	1.241	
14	dag2	maskine2	1.185	
15	dag2	maskine2	1.516	
16	dag2	maskine3	1.426	
17	dag2	maskine3	1.768	
18	dag2	maskine3	1.859	
19	dag3	maskine1	1.914	
20	dag3	maskine1	1.477	
21	dag3	maskine1	1.894	
22	dag3	maskine2	1.506	

Det man først tester er vekselvirkning mellem de to kriterier, altså er der en vekselvirkning mellem dage og maskiner? Den måles ved størrelsen

$$\sum_{i=1}^k \sum_{j=1}^m N(\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2 .$$

Her er N=81 (antal data), k=9 (antal dage) og m=3 (antal maskiner). Overstregning angiver gennemsnit. Kvadratet kaldes i Datameter "veksel"

$\bar{X}_{ij}$  er derfor gennemsnittet inden for den enkelte gruppe, og kaldes nu gruppemiddel.

$\bar{X}_{i.}$  og  $\bar{X}_{.j}$  er henholdsvis gennemsnittet inden for den enkelte dag på tværs af maskinerne ("dagegennemsnit") og gennemsnittet for den enkelte maskine på tværs af dagene ("maskinergennemsnit").

$\bar{X}$  er det totale gennemsnit.

Denne størrelse skal måles i forhold til variationen inden for de enkelte grupper, nærmere bestemt

$$\sum_{i=1}^k \sum_{j=1}^m \sum_{v=1}^n (X_{ijv} - \bar{X}_{ij})^2 .$$

n er antal data i hver gruppe, her altså 3. Kvadratet kaldes i Datameter ”igrup”.

Det ser nu sådan ud:

Omrøring af Permeabilitet

	dage	maskiner	perm	grupmiddel	igrup	dagemiddel	maskinermiddel	veksel
1	dag1	maskine1	1.673	1.64467	0.000802...	1.56578	1.57904	0.00474738
2	dag1	maskine1	1.612	1.64467	0.001067...	1.56578	1.57904	0.00474738
3	dag1	maskine1	1.649	1.64467	1.87778e...	1.56578	1.57904	0.00474738
4	dag1	maskine2	1.583	1.47667	0.0113068	1.56578	1.56148	0.0066493
5	dag1	maskine2	1.229	1.47667	0.0613388	1.56578	1.56148	0.0066493

Nogle af formlerne er ret store, så de gengives separat:

”grupmiddel”-formel:

**Formel for grupmiddel**

grupmiddel =

```

ombyt ( grupmiddel )
(( (dage = "dag1") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag1") og (maskiner = "maskine1"))))
(( (dage = "dag1") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag1") og (maskiner = "maskine2"))))
(( (dage = "dag1") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag1") og (maskiner = "maskine3"))))
(( (dage = "dag2") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag2") og (maskiner = "maskine1"))))
(( (dage = "dag2") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag2") og (maskiner = "maskine2"))))
(( (dage = "dag2") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag2") og (maskiner = "maskine3"))))
(( (dage = "dag3") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag3") og (maskiner = "maskine1"))))
(( (dage = "dag3") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag3") og (maskiner = "maskine2"))))
(( (dage = "dag3") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag3") og (maskiner = "maskine3"))))
(( (dage = "dag4") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag4") og (maskiner = "maskine1"))))
(( (dage = "dag4") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag4") og (maskiner = "maskine2"))))
(( (dage = "dag4") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag4") og (maskiner = "maskine3"))))
(( (dage = "dag5") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag5") og (maskiner = "maskine1"))))
(( (dage = "dag5") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag5") og (maskiner = "maskine2"))))
(( (dage = "dag5") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag5") og (maskiner = "maskine3"))))
(( (dage = "dag6") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag6") og (maskiner = "maskine1"))))
(( (dage = "dag6") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag6") og (maskiner = "maskine2"))))
(( (dage = "dag6") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag6") og (maskiner = "maskine3"))))
(( (dage = "dag7") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag7") og (maskiner = "maskine1"))))
(( (dage = "dag7") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag7") og (maskiner = "maskine2"))))
(( (dage = "dag7") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag7") og (maskiner = "maskine3"))))
(( (dage = "dag8") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag8") og (maskiner = "maskine1"))))
(( (dage = "dag8") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag8") og (maskiner = "maskine2"))))
(( (dage = "dag8") og (maskiner = "maskine3")) : middel (perm; ((dage = "dag8") og (maskiner = "maskine3"))))
(( (dage = "dag9") og (maskiner = "maskine1")) : middel (perm; ((dage = "dag9") og (maskiner = "maskine1"))))
(( (dage = "dag9") og (maskiner = "maskine2")) : middel (perm; ((dage = "dag9") og (maskiner = "maskine2"))))
ellers : middel (perm; ((dage = "dag9") og (maskiner = "maskine3")))

```

Mindst mulig

ellers

7 8 9 + = Datavariable  
4 5 6 - < Funktioner  
1 2 3 x > Ikoner  
0 . ^ ÷ () Målinger  
1/x √x ↑ ikke og Parametre  
|x| ← ↓ → eller Specielle

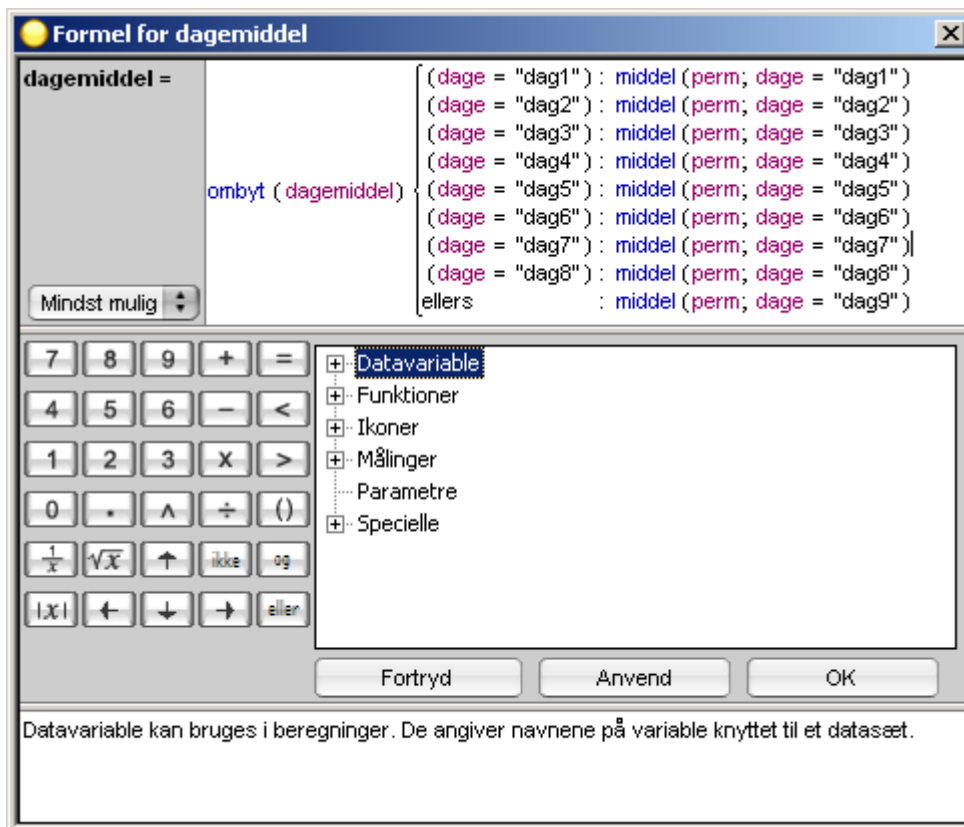
Fortryd Anvend OK

Datavariable kan bruges i beregninger. De angiver navnene på variable knyttet til et datasæt.

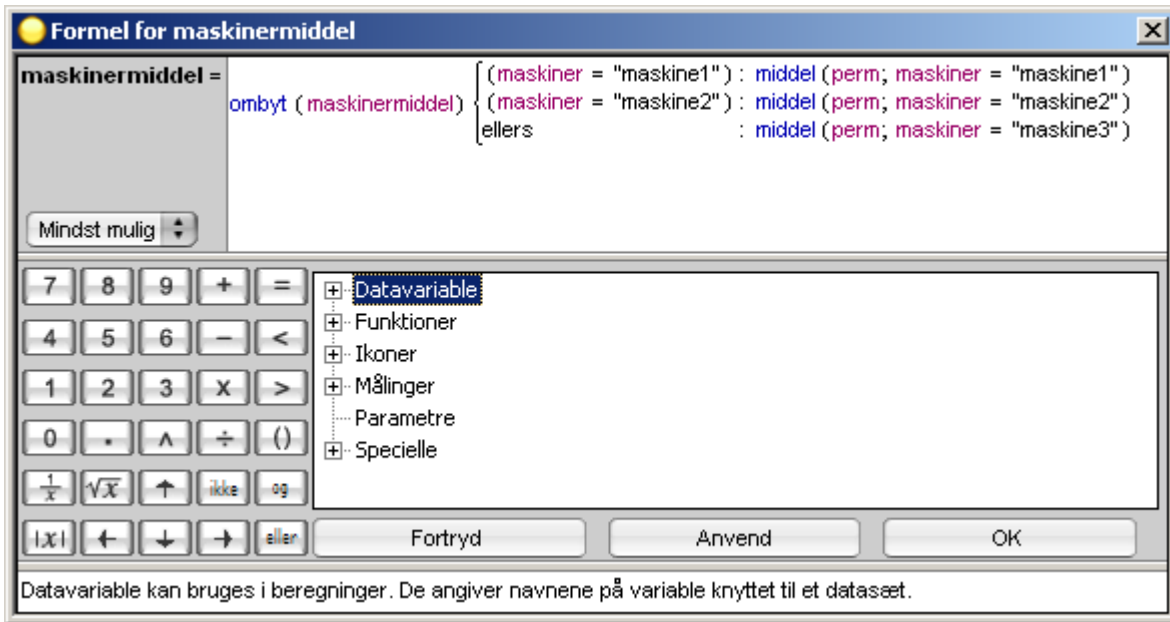
Formel for igrup:



Formel for ”dagemiddel”:



Formel for ”maskinermiddel”:



Formel for "veksel":



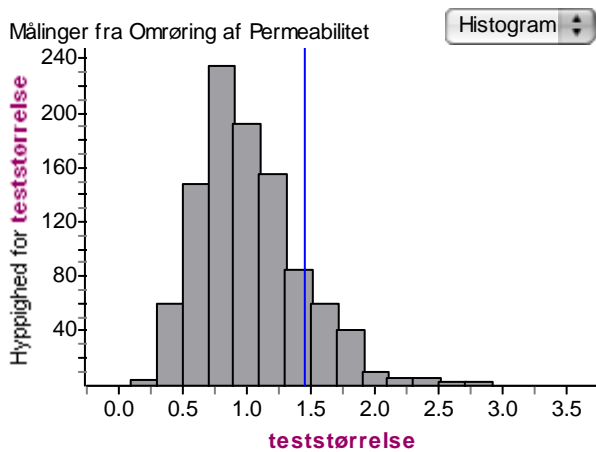
Så er vi klar til at finde teststørrelsen, hvor den samlede vekselvirkning måles i forhold til variationen inden for grupper. Konstanterne  $(k-1)(m-1) = 8*2 = 16$  og  $km(n-1) = 9*3*2 = 54$  divideres på, så det bliver den klassiske teststørrelse (der er tale om de såkaldte frihedsgrader):

Permeabilitet	
	1.4499612
$R1 = \frac{\text{sum}(\text{veksel})}{16} \div \frac{\text{sum}(\text{igrup})}{54}$	

Teststørrelsen er altså 1.4499612.

Nulhypotesen er, at det ikke er vekselvirkning mellem dage og maskiner, og ligesom ved ensidet variansanalyse er det en stor teststørrelse, der er kritisk for hypotesen.

Som ved ensidet variansanalyse omrøres nu "perm" (husk det tekniske: alle de afhængige talvariable fjernes og sættes ind igen i omrøringen, så de deltager i omrøringen). 1000 gange måles teststørrelsen, og resultatet blev her:



| 1.4499612 = 1.44996

Målinger fra Omrøring af Permeabilitet	
	144
$R1 = \text{tæl}(\text{teststørrelse} \geq 1.4499612)$	

Vi kan se, at 14.4 % af teststørrelserne ligger over den oprindelige, som altså ikke er ekstremt stor. På 5 % signifikansniveau kan vi ikke forkaste nulhypotesen, så vi vil mene, at dage og maskiner ikke vekselvirker.

Det er forudsætningen for at gennemføre de to næste test, der tester den såkaldte rækkevirkning og søjlevirkning, her altså spørgsmålet, om der er signifikant forskel, når vi sammenligner dagene, henholdsvis når vi sammenligner maskinerne.

I begge tilfælde måles i forhold til summen af kvadratsummerne fra "igrup" og "veksel" divideret med  $54+16 = 70$  (Samlet antal "frihedsgrader"). Så den finder vi lige:

Permeabilitet

	0.041153056

$$R1 = \frac{\text{sum}(\text{veksel}) + \text{sum}(\text{igrup})}{70}$$

Der skal altså måles op imod 0.041153056 .

[Note. Ved en revision af noterne bør denne størrelse ikke være konstant; dens formel bør optræde i nævneren, som ændrer sig under omrøringen.]

Tællerne i brøkerne for test af virkning mellem dage (rækkevirkning) og virkning mellem maskiner (søjlevirkning) er

$$\sum_{i=1}^k mn (\bar{X}_{i.} - \bar{X})^2 .$$

og

$$\sum_{j=1}^m kn (\bar{X}_{.j} - \bar{X})^2 .$$

Nulhypothesen er i første runde, at der ikke er ”dagevirkning” altså der er ikke forskel på dagene. Som vanligt bliver en stor brøk kritisk for nulhypotesen.

De indgående middelværdier er udregnet i datameter, så vi supplerer med ”dagevirkning” (og ”maskinervirkning” i næste runde):

Permeabilitet

	dage	maskiner	perm	dagemiddel	dagevirkning
=					$(\text{dagemiddel} - \text{middel}(\text{perm}))^2$
1	dag1	maskine1	1.404	1.52411	0.00201945
2	dag1	maskine1	1.346	1.52411	0.00201945
3	dag1	maskine1	1.618	1.52411	0.00201945
4	dag1	maskine2	1.306	1.52411	0.00201945
5	dag1	maskine2	1.628	1.52411	0.00201945

Teststørrelsen er:

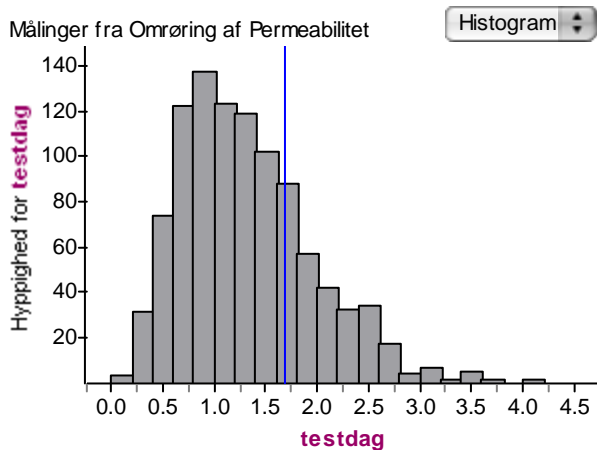
Permeabilitet

	1.6810309

$$R1 = \frac{\text{sum}(\text{dagevirkning})}{8} = \frac{0.041153056}{0.041153056}$$

Teststørrelsen: 1.6810309.

Så kører omrøringen af ”perm”, mens teststørrelsen måles 1000 gange:



1.6810309 = 1.68103

Målinger fra Omrøring af Permeabilitet

	248

R1 = tæl (testdag  $\geq$  1.6810309)

24.8 % af teststørrelserne ligger over den oprindelige, så vi opretholder hypotesen om, at der ikke er forskel på dagene.

Maskinerne da? Vi prøver:

Permeabilitet

	dage	maskiner	perm	maskinermiddel	maskinervirkning
=					$(\text{maskinermiddel} - \text{middel}(\text{perm}))^2$
1	dag1	maskine1	1.404	1.68552	0.0135651
2	dag1	maskine1	1.346	1.68552	0.0135651
3	dag1	maskine1	1.618	1.68552	0.0135651
4	dag1	maskine2	1.306	1.42833	0.019801
5	dag1	maskine2	1.628	1.42833	0.019801

Permeabilitet

	11.138388

sum (maskinervirkning)

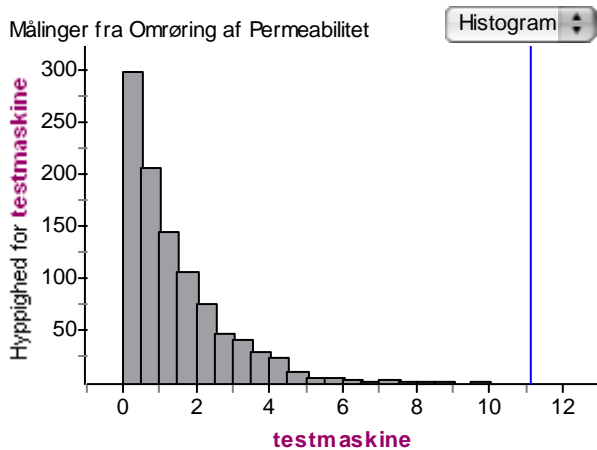
R1 =  $\frac{\text{sum}(\text{maskinervirkning})}{2}$

0.041153056

med teststørrelsen

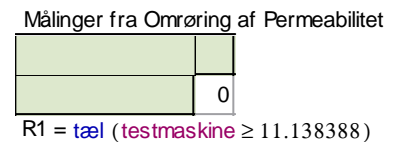
altså 11.138388.

Så følger 1000 omrøringer af ”perm” med tilhørende målinger af teststørrelsen:



11.138388 = 11.1384

og



Ingen måling kom over 11.138388, så det er en ekstremt stor teststørrelse. Denne gang forkaster vi nulhypotesen om forsvindende forskel mellem maskinerne; vi må regne med, at der er forskel på deres produktioner.

Den samlede konklusion bliver, at der ikke er vekselvirkning mellem tidspunkter og maskiner, at variationen fra dag til dag må anses for tilfældig, men at de tre maskiner producerer signifikant forskelligt. Et nærmere blik på talmaterialet tilsiger, at det er maskine 2, der producerer mindre vandtætte plader end de to andre maskiner.

### Eksempel 5.

Vi ser på et lidt simplere eksempel, hvor der kun er én observation i hver gruppe, men altså stadig to inddelingskriterier. Det er nu ikke muligt at teste hypotesen om forsvindende vekselvirkning (der er ingen variation inden for grupperne, man kan måle op imod). I sådanne tilfælde kan man alligevel ofte ud fra erfaringen eller ved at se på materialet tillade sig at gå ud fra, at der ikke er vekselvirkning. Man kan så teste for række- og søjlevirkning.

Vi prøver igen et af Halds gamle eksempler ("Statistiske metoder", side 371). Der er foretaget 16 kontrolvejninger af 50-kg sække fra 6 automatiske veje- og pakkemaskiner:

Stikprøver	Maskiner					
	1	2	3	4	5	6
1	50.6	50.5	50.7	50.2	50.5	50.4
2	50.1	50.4	50.5	50.2	50.1	50.4
3	50.1	50.6	50.8	50.4	50.7	50.8
4	50.2	50.4	51.0	50.2	50.0	50.5
5	50.1	50.5	50.7	50.1	50.3	50.3
6	50.4	50.3	50.8	50.3	49.9	50.4
7	50.2	50.4	51.0	50.4	50.2	50.0
8	50.0	50.4	50.8	50.4	50.1	50.3
9	50.0	50.5	50.4	50.3	50.3	50.6
10	50.0	50.2	50.0	49.6	49.3	50.2
11	50.1	50.1	50.3	49.9	50.1	49.8
12	49.8	50.1	50.0	49.9	49.7	50.1
13	49.8	50.1	50.2	50.0	49.7	50.1
14	49.9	50.3	49.9	50.1	49.8	50.2
15	50.1	50.3	50.3	50.3	50.0	50.1
16	49.2	49.3	49.3	49.2	49.3	49.0

De må ind i Datameter:

sækkevægt

	stikprøve	maskine	vægt	<ny>
1	stikprøve1	maskine1	50.6	
2	stikprøve1	maskine2	50.5	
3	stikprøve1	maskine3	50.7	
4	stikprøve1	maskine4	50.2	
5	stikprøve1	maskine5	50.5	
6	stikprøve1	maskine6	50.4	
7	stikprøve2	maskine1	50.1	
8	stikprøve2	maskine2	50.4	
9	stikprøve2	maskine3	50.5	
10	stikprøve2	maskine4	50.2	

og de nødvendige variable tilføjes:

sækkevægt

	stikprøve	maskine	vægt	stikprøvemiddel	maskinemiddel	veksel	stikprøvevirkning	maskinevirkning	<ny>
1	stikprøve1	maskine1	50.6	50.4833	50.0375	0.0625	0.0976562	0.0177778	
2	stikprøve1	maskine2	50.5	50.4833	50.275	0.00765625	0.0976562	0.0108507	
3	stikprøve1	maskine3	50.7	50.4833	50.4187	0.000976562	0.0976562	0.0614627	
4	stikprøve1	maskine4	50.2	50.4833	50.0938	0.0425391	0.0976562	0.00594184	

med disse formler:

**Formel for stikprøvemiddel**

stikprøvemiddel =

ombyt (stikprøvemiddel)

```
(stikprøve = "stikprøve1") : middel (vægt; stikprøve = "stikprøve1")
(stikprøve = "stikprøve2") : middel (vægt; stikprøve = "stikprøve2")
(stikprøve = "stikprøve3") : middel (vægt; stikprøve = "stikprøve3")
(stikprøve = "stikprøve4") : middel (vægt; stikprøve = "stikprøve4")
(stikprøve = "stikprøve5") : middel (vægt; stikprøve = "stikprøve5")
(stikprøve = "stikprøve6") : middel (vægt; stikprøve = "stikprøve6")
(stikprøve = "stikprøve7") : middel (vægt; stikprøve = "stikprøve7")
(stikprøve = "stikprøve8") : middel (vægt; stikprøve = "stikprøve8")
(stikprøve = "stikprøve9") : middel (vægt; stikprøve = "stikprøve9")
(stikprøve = "stikprøve10") : middel (vægt; stikprøve = "stikprøve10")
(stikprøve = "stikprøve11") : middel (vægt; stikprøve = "stikprøve11")
(stikprøve = "stikprøve12") : middel (vægt; stikprøve = "stikprøve12")
(stikprøve = "stikprøve13") : middel (vægt; stikprøve = "stikprøve13")
(stikprøve = "stikprøve14") : middel (vægt; stikprøve = "stikprøve14")
(stikprøve = "stikprøve15") : middel (vægt; stikprøve = "stikprøve15")
ellers : middel (vægt; stikprøve = "stikprøve16")
```

Mindst mulig

7 8 9 + =

4 5 6 - <

1 2 3 x >

0 . ^ ÷ ()

1/x √x ↑ ikke og

|x| ← ↓ → eller

Fortryd Anvend OK

Datavariabel kan bruges i beregninger. De angiver navnene på variable knyttet til et datasæt.

**Formel for maskinemiddel**

maskinemiddel = `ombyt (maskinemiddel)`

Mindst mulig

```
(maskine = "maskine1") : middel (vægt; maskine = "maskine1")
(maskine = "maskine2") : middel (vægt; maskine = "maskine2")
(maskine = "maskine3") : middel (vægt; maskine = "maskine3")
(maskine = "maskine4") : middel (vægt; maskine = "maskine4")
(maskine = "maskine5") : middel (vægt; maskine = "maskine5")
ellers : middel (vægt; maskine = "maskine6")
```

7 8 9 + =

4 5 6 - <

1 2 3 x >

0 . ^ ÷ ()

$\frac{1}{x}$   $\sqrt{x}$  ↑ ikke og

|x| ← ↓ → eller

- + Datavariabel
- + Funktioner
- + Ikoner
- ... Målinger
- ... Parametre
- + Specielle

Fortryd Anvend OK

Datavariabel kan bruges i beregninger. De angiver navnene på variable knyttet til et dataset.

**Formel for vækst**

Mindst mulig `(vægt - stikprøvemiddel - maskinemiddel + middel (vægt))2`

Styrer fontstørrelsen i formleredatoren

4 5 6 - <

1 2 3 x >

0 . ^ ÷ ()

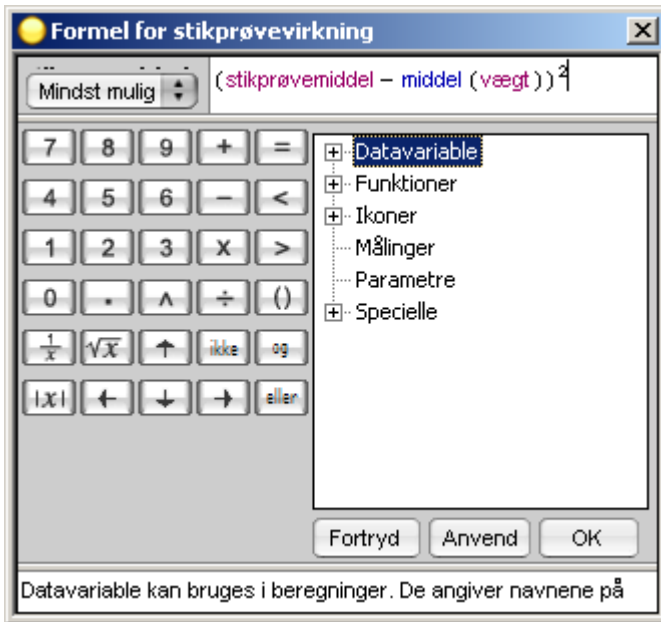
$\frac{1}{x}$   $\sqrt{x}$  ↑ ikke og

|x| ← ↓ → eller

- + Funktioner
- + Ikoner
- ... Målinger
- ... Parametre
- + Specielle

Fortryd Anvend OK

Datavariabel kan bruges i beregninger. De angiver navnene på variable knyttet til et



Nu udregnes teststørrelsen til test af rækkevirkning:

sækkevægt

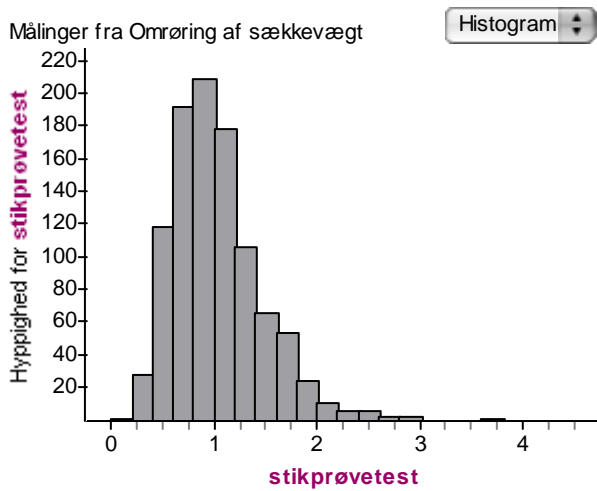
	18.115183

$\text{sum}(\text{stikprøvevirkning})$

$$R1 = \frac{15}{\text{sum}(\text{veksel})}$$

75

15 er antal stikprøver -1 og 75 er  $15 \cdot 5$ , hvor 5 er antal maskiner -1 (igen er der tale om frihedsgrader). Nu følger omrøringen 1000 gange af "vægt" med tilhørende måling af teststørrelsen. Resultat:



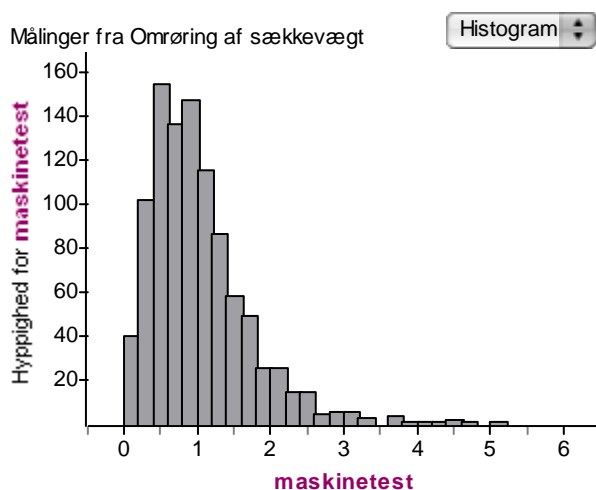
Den oprindelige teststørrelse 18.115183 falder tydeligt langt til højre for alle omrøringsteststørrelserne, så den er ekstremt stor. Vi forkaster nulhypotesen om forsvindende rækkevirkning, så det tyder stærkt på, at der er forskel på stikprøverne.

Nu kommer helt samme procedure til test af søjlevirkning:

sækkevægt	
	11.520704

$\frac{\text{sum}(\text{maskinevirkning})}{5}$

$$R1 = \frac{\text{sum}(\text{veksel})}{75}$$



og

Igen er det tydeligt, at 11.520704 ligger ekstremt langt til højre, så vi vil regne med, at der også er søjlevirkning.

Vi må således alt i alt antage, at såvel variationerne i varen som i justeringen af maskinerne er væsentlige for resultatet af vejningerne.

Professor Halds eksempler er ingeniør-orienterede. Vi ser derfor på endnu et – lidt anderledes -

### Eksempel 6.

Et nyere eksempel på tosidet variansanalyse er hentet fra Ernst Hansen: ”Introduktion til Matematisk Statistik”, 2. udgave, Afdeling for anvendt matematik og statistik, Københavns Universitet, 2005, side 489ff.

I et dyrkningsforsøg med tomater, gennemført i det vestlige Nigeria, har man undersøgt udbyttet af forskellige tomatsorter. Udover valget af tomatsort, har forsøget været rettet mod at finde ud af hvilken effekt tilplantningstætheden har på udbyttet. Data anføres i følgende tabel.

Der indgår tre tomatsorter i forsøget med produktionsnavnene Harvester, Ife No.1 og Pusa Early Dwarf, og fire dyrkningstætheder. Enheden kpla/ha betyder kilo planter per hektar, hvilket betyder, at de faktiske dyrkningstætheder i forsøget er henholdsvis 10000, 20000, 30000 og 40000 planter per hektar. ”Responserne” er udbyttet opgjort i tons per hektar.

	10 kpla/ha	20 kpla/ha	30 kpla/ha	40 kpla/ha
Harvester	10.5	12.8	12.1	10.8
	9.2	11.2	12.6	9.1
	7.9	13.3	14.0	12.5
Ife No. 1	8.1	12.7	14.4	11.3
	8.6	13.7	15.4	12.5
	10.1	11.5	13.7	14.5
Pusa Early Dwarf	16.1	16.6	20.8	18.4
	15.3	19.2	18.8	18.9
	17.5	18.5	21.0	17.2

Da testene forløber på samme måde som professor Halds permeabilitetseksempel (Eksempel 4), gennemgås det her summarisk.

Data indtastes i Datameter :

## Tomatdyrkning

	tæthed	tomatsort	udbytte	grupmiddel	igrup	tæthedmiddel	tomatsortmiddel	veksel	<ny>
1	tiklaprha	Harvester	10.5	9.2	1.69	11.4778	11.3333	0.09	
2	tiklaprha	Harvester	9.2	9.2	3.15544e...	11.4778	11.3333	0.09	
3	tiklaprha	Harvester	7.9	9.2	1.69	11.4778	11.3333	0.09	
4	tiklaprha	lfeNo1	8.1	8.93333	0.694444	11.4778	12.2083	0.708403	
5	tiklaprha	lfeNo1	8.6	8.93333	0.111111	11.4778	12.2083	0.708403	
6	tiklaprha	lfeNo1	10.1	8.93333	1.36111	11.4778	12.2083	0.708403	
7	tiklaprha	PEDw arf	16.1	16.3	0.04	11.4778	18.1917	0.293403	
8	tiklaprha	PEDw arf	15.3	16.3	1	11.4778	18.1917	0.293403	
9	tiklaprha	PEDw arf	17.5	16.3	1.44	11.4778	18.1917	0.293403	
10	tyveklaprha	Harvester	12.8	12.4333	0.134444	14.3889	11.3333	0.38716	
11	tyveklaprha	Harvester	11.2	12.4333	1.52111	14.3889	11.3333	0.38716	
12	tyveklaprha	Harvester	13.3	12.4333	0.751111	14.3889	11.3333	0.38716	
13	tyveklaprha	lfeNo1	12.7	12.6333	0.004444...	14.3889	12.2083	0.002785...	
14	tyveklaprha	lfeNo1	13.7	12.6333	1.13778	14.3889	12.2083	0.002785...	
15	tyveklaprha	lfeNo1	11.5	12.6333	1.28444	14.3889	12.2083	0.002785...	
16	tyveklaprha	PEDw arf	16.6	18.1	2.25	14.3889	18.1917	0.324267	
17	tyveklaprha	PEDw arf	19.2	18.1	1.21	14.3889	18.1917	0.324267	
18	tyveklaprha	PEDw arf	18.5	18.1	0.16	14.3889	18.1917	0.324267	
19	trediveklaprha	Harvester	12.1	12.9	0.64	15.8667	11.3333	0.151235	
20	trediveklaprha	Harvester	12.6	12.9	0.09	15.8667	11.3333	0.151235	
21	trediveklaprha	Harvester	14.0	12.9	1.21	15.8667	11.3333	0.151235	

Formlerne er:

**Formel for grupmiddel**

grupmiddel =

```

(( (tæthed = "tiklaprha") og (tomatsort = "Harvester"))) : middel (udbytte; (( (tæthed = "tiklaprha") og (tomatsort = "Harvester"))))
(( (tæthed = "tiklaprha") og (tomatsort = "lfeNo1"))) : middel (udbytte; (( (tæthed = "tiklaprha") og (tomatsort = "lfeNo1"))))
(( (tæthed = "tiklaprha") og (tomatsort = "PEDwarf"))) : middel (udbytte; (( (tæthed = "tiklaprha") og (tomatsort = "PEDwarf"))))
(( (tæthed = "tyveklaprha") og (tomatsort = "Harvester"))) : middel (udbytte; (( (tæthed = "tyveklaprha") og (tomatsort = "Harvester"))))
(( (tæthed = "tyveklaprha") og (tomatsort = "lfeNo1"))) : middel (udbytte; (( (tæthed = "tyveklaprha") og (tomatsort = "lfeNo1"))))
(( (tæthed = "tyveklaprha") og (tomatsort = "PEDwarf"))) : middel (udbytte; (( (tæthed = "tyveklaprha") og (tomatsort = "PEDwarf"))))
(( (tæthed = "trediveklaprha") og (tomatsort = "Harvester"))) : middel (udbytte; (( (tæthed = "trediveklaprha") og (tomatsort = "Harvester"))))
(( (tæthed = "trediveklaprha") og (tomatsort = "lfeNo1"))) : middel (udbytte; (( (tæthed = "trediveklaprha") og (tomatsort = "lfeNo1"))))
(( (tæthed = "trediveklaprha") og (tomatsort = "PEDwarf"))) : middel (udbytte; (( (tæthed = "trediveklaprha") og (tomatsort = "PEDwarf"))))
(( (tæthed = "fyrreklaprha") og (tomatsort = "Harvester"))) : middel (udbytte; (( (tæthed = "fyrreklaprha") og (tomatsort = "Harvester"))))
(( (tæthed = "fyrreklaprha") og (tomatsort = "lfeNo1"))) : middel (udbytte; (( (tæthed = "fyrreklaprha") og (tomatsort = "lfeNo1"))))
ellers : middel (udbytte; (( (tæthed = "fyrreklaprha") og (tomatsort = "PEDwarf"))))

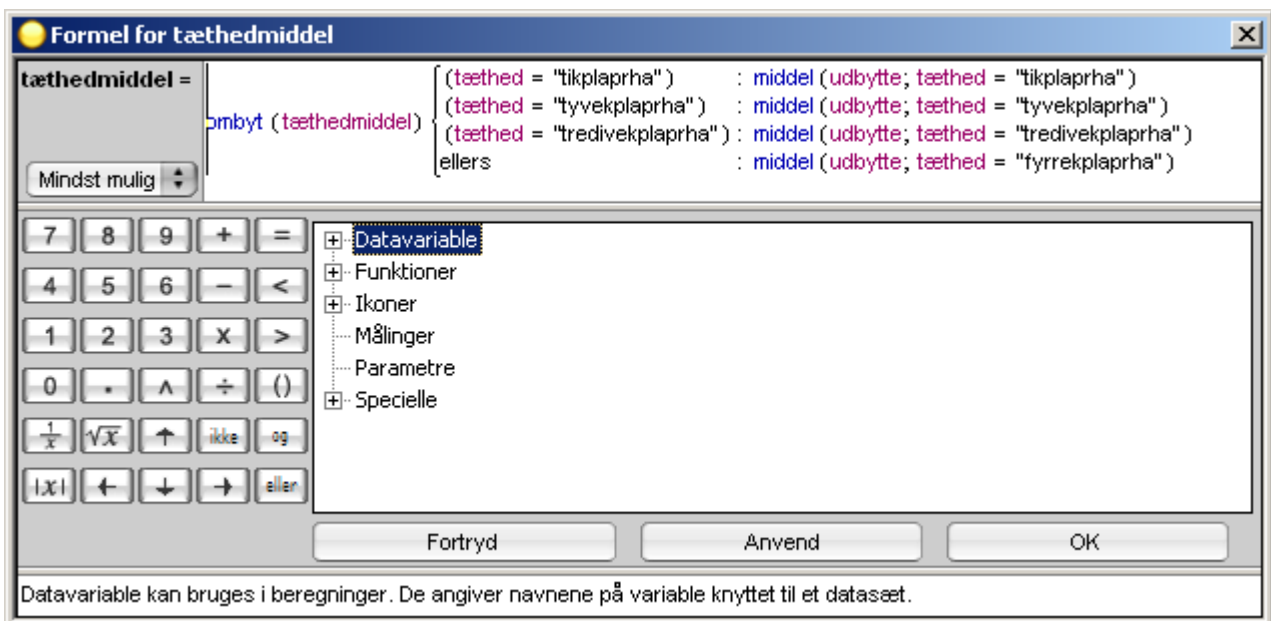
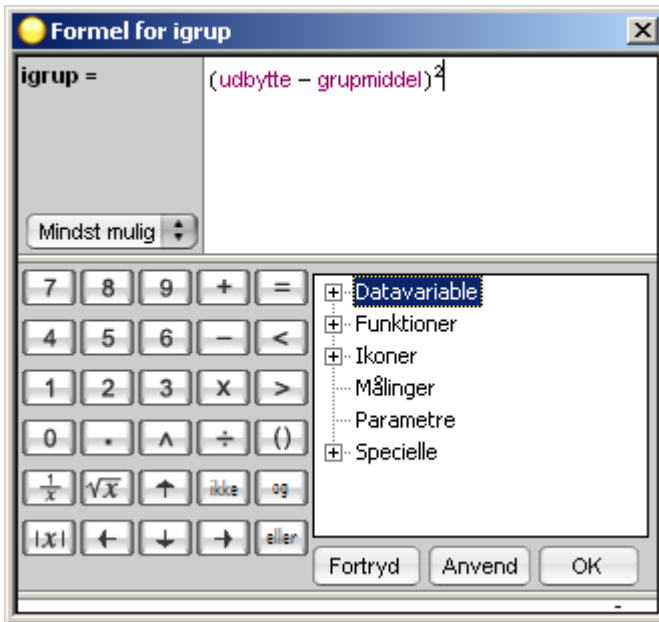
```

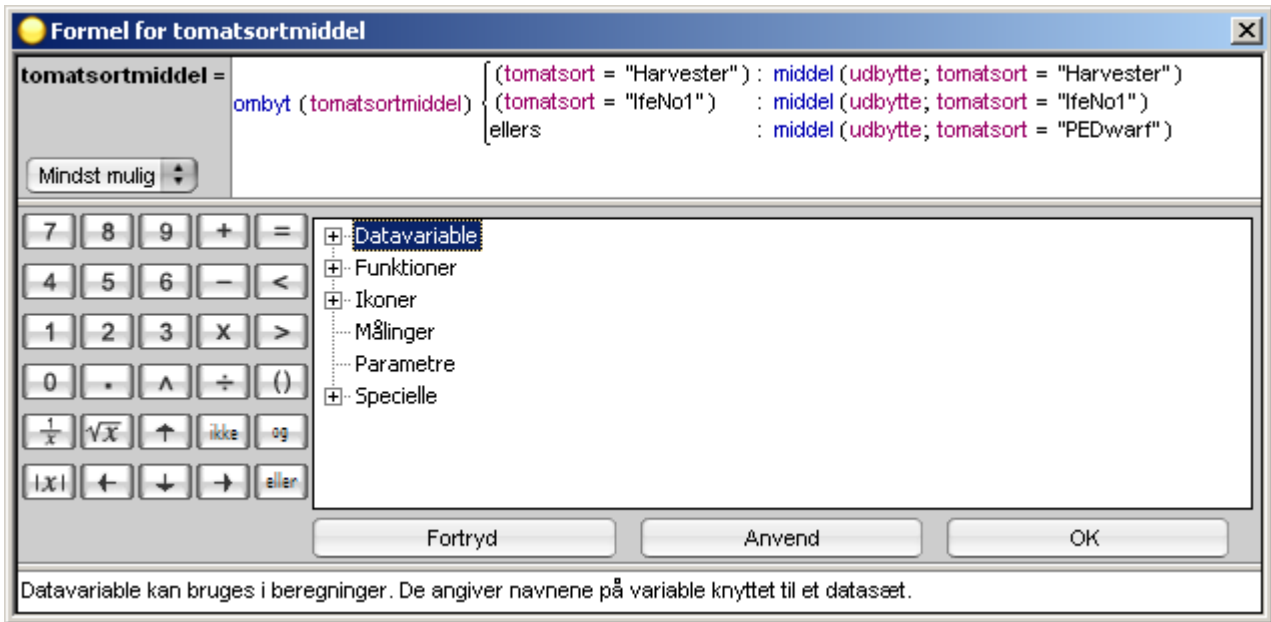
Mindst mulig +

Datavariable  
 Funktioner  
 Ikoner  
 Målinger  
 Parametre  
 Specielle

Fortryd      Anvend      OK

Datavariable kan bruges i beregninger. De angiver navnene på variable knyttet til et datasæt.





Testene kører nu som i Halds eksempel med permeabilitet.

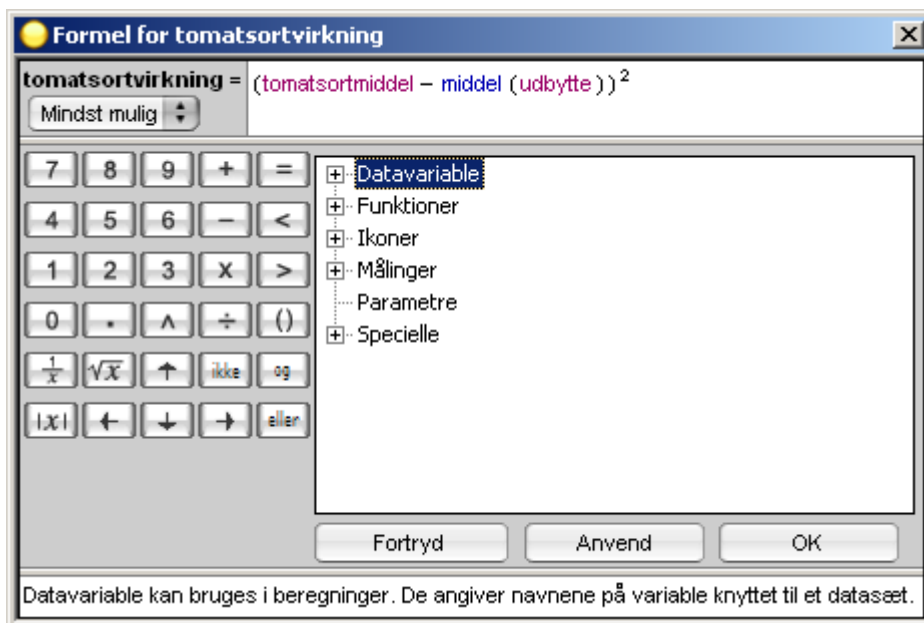
Der bliver ikke vekselvirkning mellem tæthed og tomatsort, så man kan fortsætte med test for virkning mellem tæthed og tomatsorter. Der bliver ikke signifikant forskel på tæthedsvirkningerne, så det spændende bliver spørgsmålet om sorter.

Omrøringen kører på

Omrøring af Tomatdyrkning

	tæthed	tomatsort	udbytte	tomatsortmiddel	tomatsortvirkning	<ny>
1	tikplapra	Harvester	16.1	13.275	0.404637	
2	tikplapra	Harvester	10.8	13.275	0.404637	
3	tikplapra	Harvester	9.2	13.275	0.404637	
4	tikplapra	lfeNo1	11.5	14.6583	0.558341	
5	tikplapra	lfeNo1	12.8	14.6583	0.558341	
6	tikplapra	lfeNo1	18.4	14.6583	0.558341	
7	tikplapra	PEDw arf	14.4	13.8	0.0123457	
8	tikplapra	PEDw arf	13.7	13.8	0.0123457	
9	tikplapra	PEDw arf	12.7	13.8	0.0123457	
10	tyvekpla...	Harvester	10.1	13.275	0.404637	

hvor formelen for tomatvirkning er



**Formel for tomatvirkning**

tomatsortvirkning = (tomatsortmiddel - middel (udbytte))^2

Mindst mulig

7 8 9 + =

4 5 6 - <

1 2 3 x >

0 . ^ ÷ ()

1/x √x ↑ ikke og

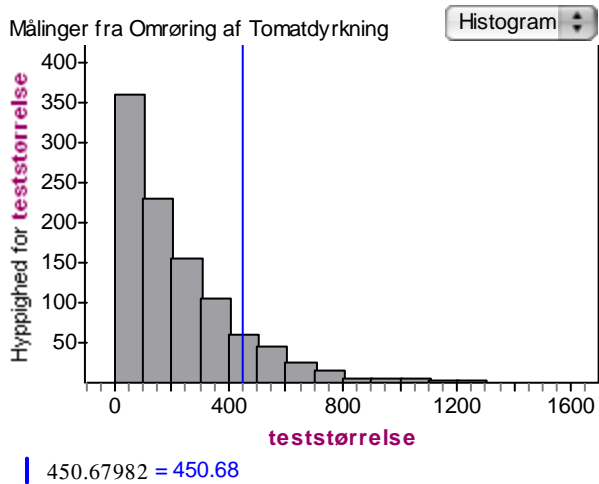
|x| ← ↓ → eller

- + Datavariabel
- + Funktioner
- + Ikoner
- + Målinger
- Parametre
- + Specielle

Fortryd Anvend OK

Datavariabel kan bruges i beregninger. De angiver navnene på variable knyttet til et datasæt.

Resultatet bliver



Målinger fra Omrøring af Tomatdyrkning

teststørrelse	116
---------------	-----

R1 = tæl (teststørrelse  $\geq$  450.67982)

altså 11.6 % sandsynlighed for en højere værdi end teststørrelsen. Det lykkes altså heller ikke at påvise signifikant forskel på tomatarterne. Tallene er dog sådan, at det nok er rimeligt at gennemføre nogle flere undersøgelser af sagen.

# Et par eksempler på Statistiske tests ved



## Omrøring...

### Tre statistikeksempler.

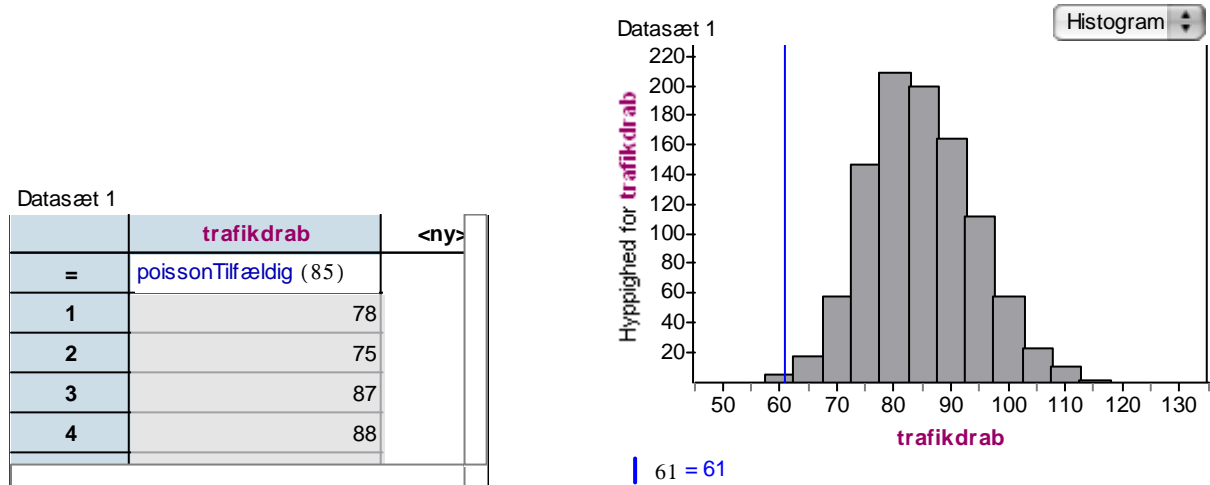
#### Eksempel 1.

(Lønborg, side 245). I et område havde man et år 85 trafikdræbte. En kampagne sættes i værk, og året efter er der 61 trafikdræbte. Er nedgangen signifikant, eller kan det være en tilfældighed?

Med så få observationer må man gøre en antagelse om fordelingen. Vi antager, at sådanne observationer i det enkelte år er Poisson-fordelte. Det første år er vores bedste bud, at fordelings

middelværdi = parameteren = 85. Vores nulhypotese er, at der ikke er forskel på de to år, så 61 tilhører den samme fordeling. Hvor sandsynligt er det?

Vi danner et datasæt med 1000 tal fra Poissonfordelingen med parameter 85 (funktionen findes under Tilfældige talgeneratorer):



61 synes at ligge yderligt. Nærmere bestemt:

Datasæt 1	
	4

R1 = tæl (trafikdrab ≤ 61)

så sandsynligheden for en observation mindre eller lig 61 er 0.4 %.

Vi forkaster nulhypotesen, så nedgangen i trafikdrab er signifikant.

### Eksempel 2.

(Lønborg, side 267). Over to år har man målt forrentningsprocenterne for 11 tilfældigt udvalgte landbrugsbedrifter på Vestfyn:

Forrentningsprocent	
1. år	2. år
7.5	11.4

10.2	13.5
2.2	4.4
10.8	8.3
6.0	16.0
3.4	4.6
5.2	4.2
4.5	5.7
6.7	7.7
6.6	11.8
4.9	7.7

Da der selvfølgelig er afhængighed, når tallene er fra samme bedrifter, danner man differenstabellen og tester for middelværdi = 0.

Forskel i forrentningsprocent mellem 2. og 1. år
3.9
3.3
2.2
-2.5
10.0
1.2
-1.0
1.2
1.0
5.2
2.8

forrentning

	merforrentning	<ny>
1	3.9	
2	3.3	
3	2.2	
4	-2.5	
5	10.0	
6	1.2	
7	-1.0	
8	1.2	
9	1.0	
10	5.2	
11	2.8	



Stikprøve fra forrentning

Vi tester for middelværdi 0 ved bootstrap.

**Inspicér Stikprøve fra forrentning**

Dat... Mål... Kom... Udse... Kategorier **Stikprøve**

Animation slået til  Med tilbagelægning

Erstat de eksisterende målinger

Gentag stikprøven ved ændring i datasættet

11 data

Stopkriterium

Ny stikprøve

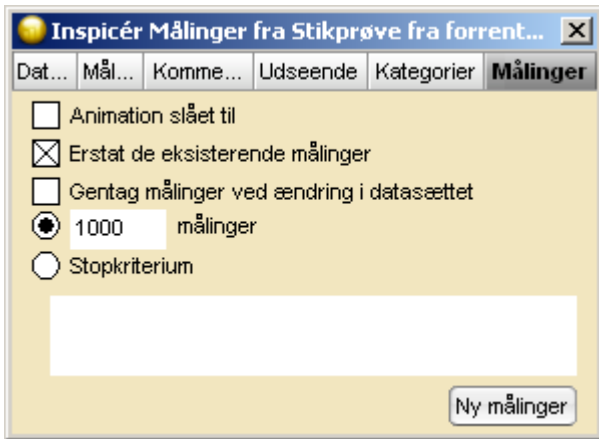
Vi måler:

**Inspicér Stikprøve fra forrentning**

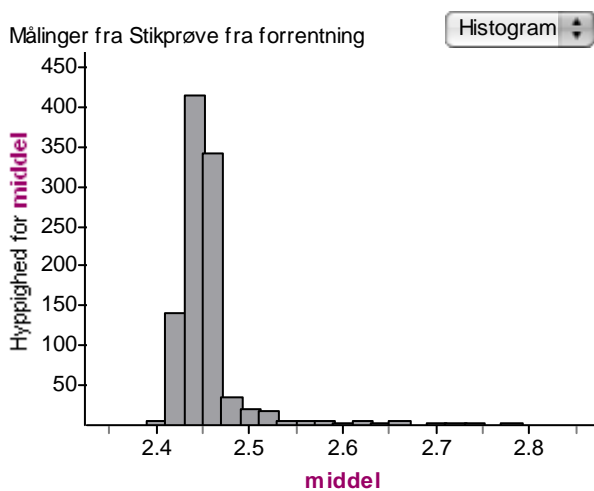
Data **Målinger** Kommentarer Udse... Kat... Sti...

Måling	Værdi	Formel
middel	2.4642	middel (merforrentning)
<ny>		

og gennemfører det 1000 gange:



Målinger fra Stikprøve fra forrentning		
	midde	<ny>
1	2.52644	
2	2.55	
3	2.47615	



Målinger fra Stikprøve fra forrentning	
	0
R1 = tæl (midde ≤ 0)	

Der er ingen middelværdier mindre end eller lig 0, så vi forkaster nulhypotesen (middelværdi 0).

Der er signifikant forskel på forrentningsprocenten 1. og 2. år. Den er højere 2. år.

### Eksempel 3.

Vi tester, om et datamateriale **følger en bestemt fordeling**, i dette tilfælde en Poissonfordeling.

I et supermarked har man undersøgt ankomsten af kunder ved kasseapparaterne. I 60 tidsintervaller af hvert et minut har man registreret antallet af kundeankomster. Resultatet fremgår af nedenstående tabel, hvor x angiver antallet af kundeankomster:

X	0	1	2	3	4	5	6	I alt
	8	17	18	13	2	1	1	60

Vi indtaster materialet, men trækker ”højre hale” sammen. Det skyldes, at testen ifølge teorien kræver mindst 5 observationer på hver værdi i *den forventede fordeling*.

Datasæt 1

	ankomster	antalank	psands	forventet
=			$\text{hvis}(\text{indeks} \leq 4) \begin{cases} \text{poissonPunkt}(\text{indeks} - 1; 1.85) \\ 0.116874 \end{cases}$	60psands
1	0	8	0.157237	9.43423
2	1	17	0.290889	17.4533
3	2	18	0.269072	16.1443
4	3	13	0.165928	9.95567
5	>=4	4	0.116874	7.01244

Observationernes middelværdi er 1.85, som er poissonfordelingens estimerede parameter. ”psands” er punktsandsynlighederne for denne poissonfordeling; det sidste interval har simpelthen fået ”resten” af sandsynligheden (1 minus summen af de foregående).

Vi beregner teststørrelsen i en Chi-i-anden test:

Datasæt 1

	2.6681306

$$R1 = \text{sum} \left( \frac{(\text{antalank} - \text{forventet})^2}{\text{forventet}} \right)$$

Vi udvælger nu en stikprøve på 60 fra den samme poissonfordeling:

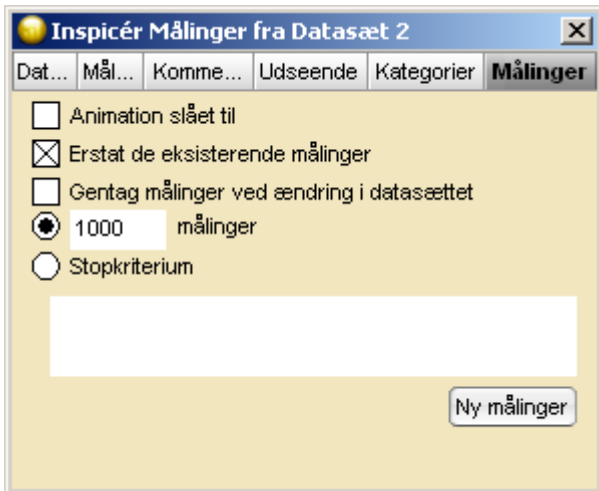
Datasæt 2

	pois
=	poissonTilfældig (1.85)
1	1
2	0
3	1
4	1
5	0
6	2

og måler Chi-i-anden teststørrelsen:

Måling	Værdi	Formel
chi_i_anden	0.614303	$\frac{(\text{tæl}(\text{pois} = 0) - 9.43423)^2}{9.43423} + \frac{(\text{tæl}(\text{pois} = 1) - 17.4533)^2}{17.4533} + \frac{(\text{tæl}(\text{pois} = 2) - 16.1443)^2}{16.1443} + \frac{(\text{tæl}(\text{pois} = 3) - 9.95567)^2}{9.95567} + \frac{(\text{tæl}(\text{pois} \geq 4) - 7.01244)^2}{7.01244}$
<ny>		

Dette gentages 1000 gange:

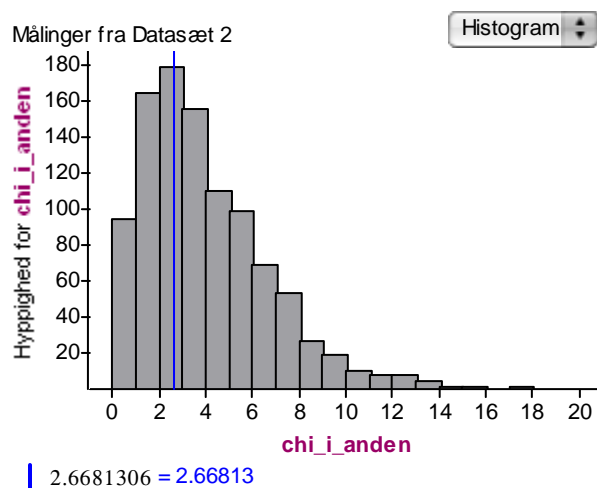


Vi omsætter målingerne i en tabel:

Målinger fra Datasæt 2

	chi_i_anden	<ny>
1	1.73587	
2	4.00583	
3	8.4716	
4	2.88055	
5	3.64143	
6	7.48755	
7	2.89394	
8	1.89724	
9	1.74319	
10	1.44041	

og et histogram:



Vores teststørrelse ligger så pænt i "midten", at vi slet ikke kan forkaste nulhypotesen om, at vores data fulgte poissonfordelingen med parameter 1.85. Det kan vi altså regne med, at de gør!