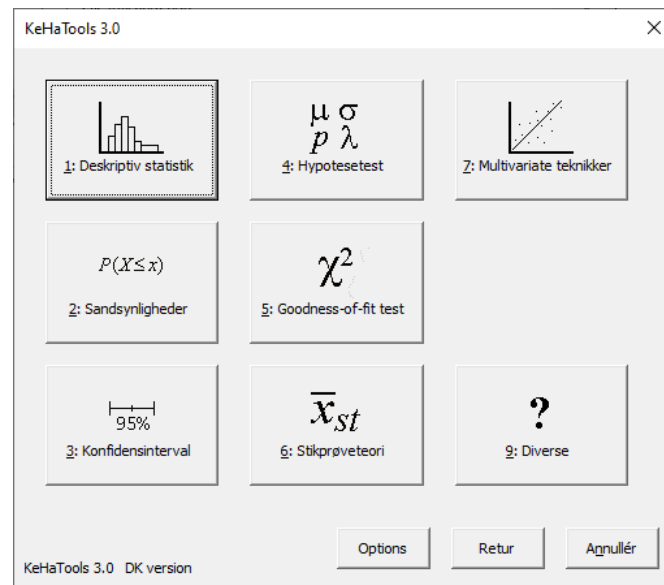


KeHaTools 3.0.2

En vejledning



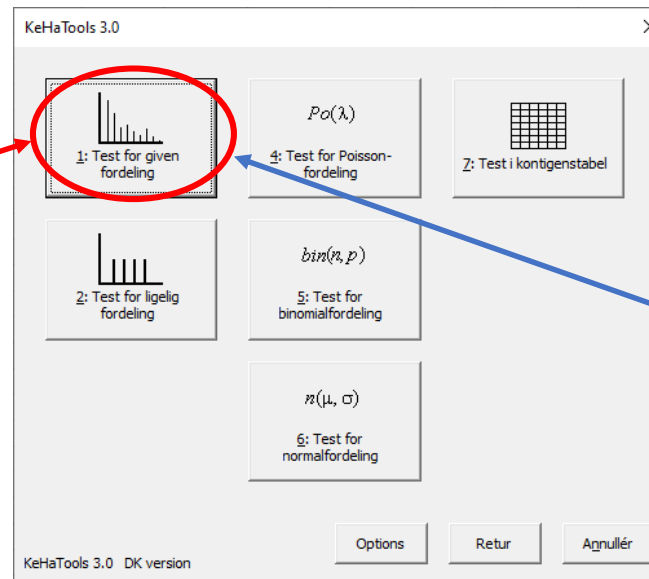
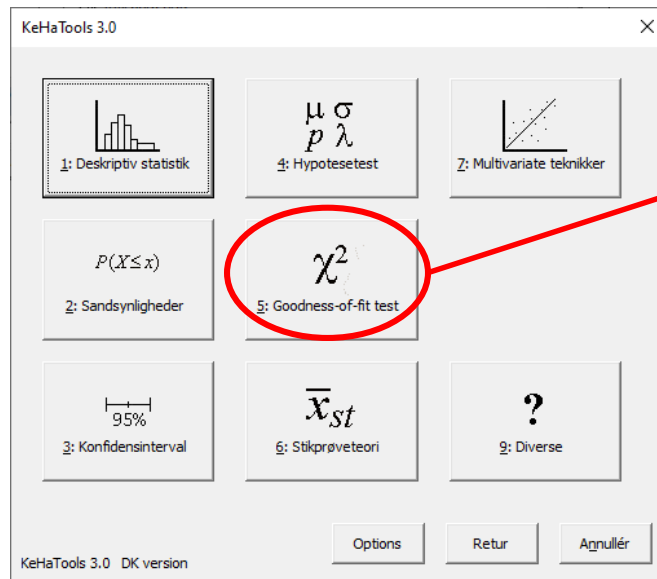
Version af 3/11 2020

Versionshistorik

- KeHaTools 3.0
 - frigivet sommeren 2020
- KeHaTools 3.0.1
 - der er rettet en del fejl i forhold til 3.0
 - du bør derfor afinstallere 3.0 og geninstallere 3.0.1
- keHaTools 3.0.2
 - der er rettet en enkelt, ubetydelig fejl i forhold til 3.0.1
 - (ved beregning af stikprøvestørrelse for en andel kommer der en dialogboks, som man bare kan lukke)
 - det er ikke nødvendigt at opdatere fra 3.0.1 til 3.0.2

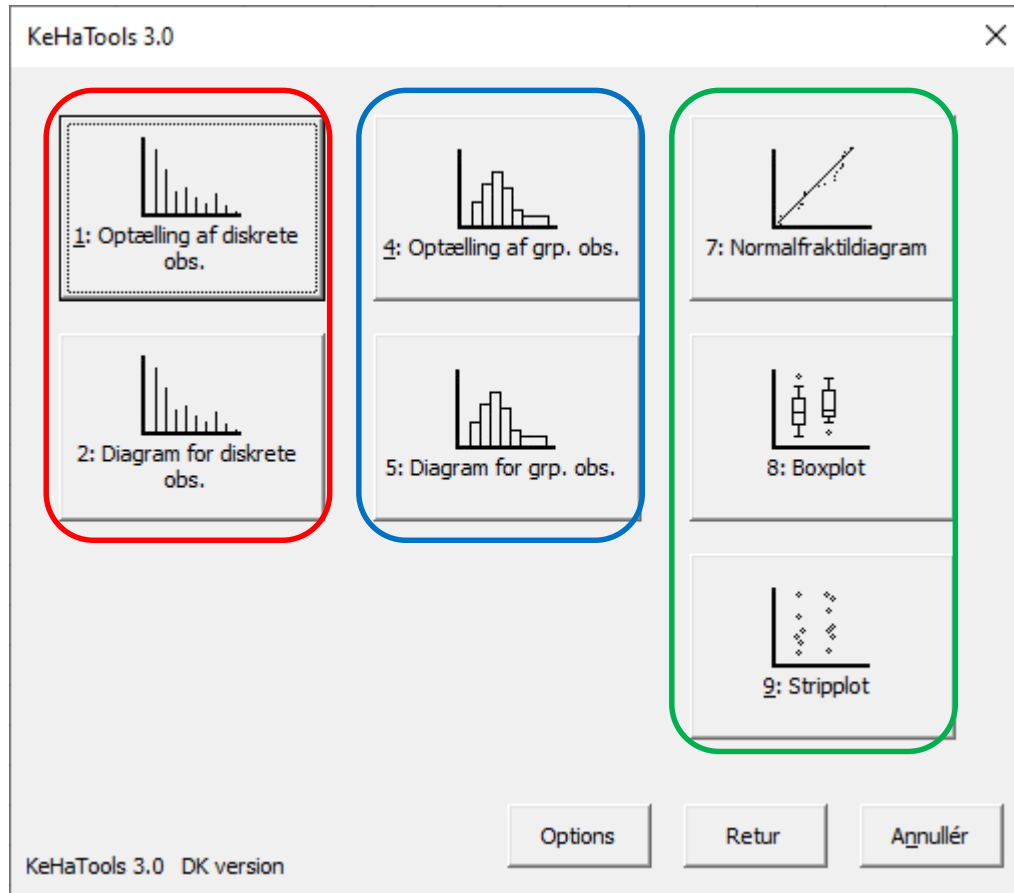
Opbygning af dette dokument

- For hver enkelt beregningstype angives der et eksempel på grafisk vis.
- Til venstre nogle typiske tal i Excel, i midten inputtet til KeHaTools, og til højre outputtet fra KeHaTools
- Alle beregninger i KeHaTools har en talkode, efter knapperens numre:



Knap 5.1

1. Deskriptiv statistik



Diskrete observationer
optælling, pindediagram

Grupperede (kontinuerte) observationer
optælling og gruppering, histogram

Specialiserede diagrammer
normalfraktildiagram (er data normalfordelte?)
boxplot
stripplot

1.1 Optælling af diskrete observationer

2	3	5
3	4	3
6	5	2
4	3	11
0	6	-2
0	2	
3	8	

Optælling af diskrete (heltallige) observationer

Observationer: 'Ark7!\$I\$3:\$K\$9 -

Minimumsværdi: 0

Maksimumsværdi: 20

OK

Annullér

Disse værdier styrer størrelsen af tabellen. Observationer, som falder udenfor dette interval, vil blive ignoreret.

Værdi	Hyppighed
0	2
1	0
2	3
3	5
4	2
5	2
6	2
7	0
8	1
9	0
10	0
11	1
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0

1.2 Diagram for diskrete observationer

Værdi	Hypighed
0	2
1	0
2	3
3	5
4	2
5	2
6	2
7	0
8	1
9	0
10	0

Deskriptiv statistik for diskrete observationer

Observerede værdier:

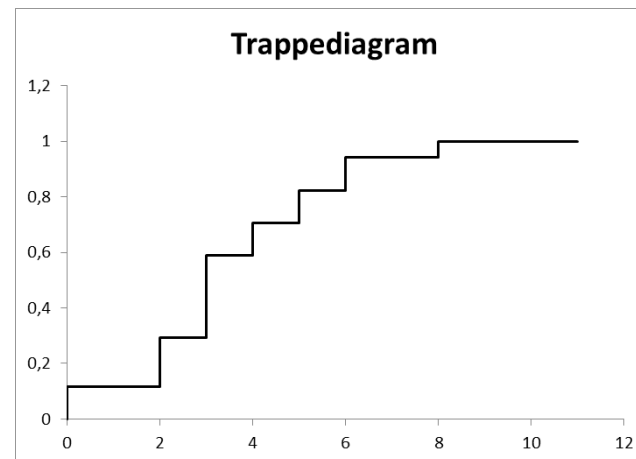
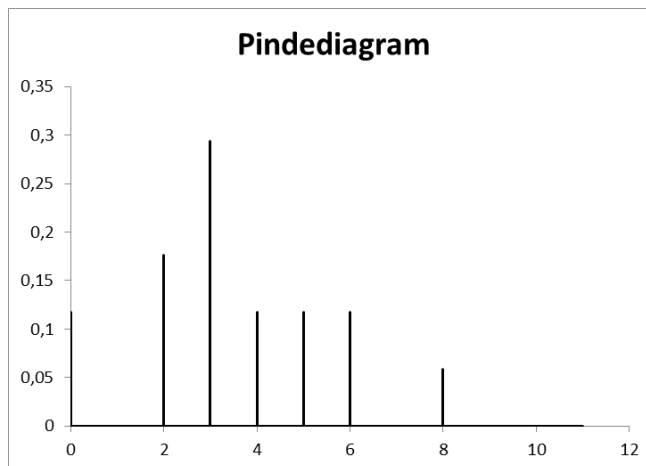
Hypigheder:

OK

Annullér

Middelvæ	3,470588
Varians	4,389706
Standarda	2,095162

Observati	Hypighed	Kum. hypi	Frekvens	Kum. frek
0	2	2	11,8%	11,8%
1	0	2	0,0%	11,8%
2	3	5	17,6%	29,4%
3	5	10	29,4%	58,8%
4	2	12	11,8%	70,6%
5	2	14	11,8%	82,4%
6	2	16	11,8%	94,1%
7	0	16	0,0%	94,1%
8	1	17	5,9%	100,0%
9	0	17	0,0%	100,0%
10	0	17	0,0%	100,0%



1.4 Optælling af grupperede observationer

5,23	40,62	21,24	51,96	0
65,80	22,39	31,31	49,77	10
37,99	73,05	56,32	69,67	25
27,24	66,11	37,76	68,69	30
70,00	19,66	59,46	56,22	40
51,22	27,05	62,40	94,13	50
69,26	45,74	30,19	60,25	60
44,87	49,26	51,19	77,16	70
57,80	46,25	72,95	16,23	90
7,16	87,23	22,36	11,32	100
75,51	4,89	14,39	23,67	
40,24	39,64	32,97	63,72	
31,96	50,35	34,30	71,22	
44,37	47,78	89,62	36,36	
60,44	29,15	79,73	22,62	

Optælling af grupperede observationer

Observationer → Ark1!\$A\$1:\$D\$15

Inddeling → Ark1!\$F\$1:\$F\$11

OK

Annullér

Nedre	Øvre	Hypighe
0	10	3
10	25	9
25	30	3
30	40	9
40	50	9
50	60	8
60	70	10
70	90	8
90	100	1
100		

Dette giver en inddeling på:

0-10 10-25 25-30 30-40 40-50
50-60 60-70 70-90 90-100

Intervallerne behøver *ikke* at være
lige brede

Dette betyder, at der er
9 værdier mellem 40 og 50

Denne værdi er med af hensyn
til histogram (se 1.5)

1.5 Diagram for grupperede observationer

Optælling af grupperede observationer		
Nedre	Øvre	Hypigheder
0	10	3
10	25	17
25	30	28
30	40	43
40	50	22
50	60	19
60	70	23
70	90	8
90	100	2
100		

Deskriptiv statistik for grupperede observationer

Intervalgrænser: [rk7!\$A\$4:\$A\$13] -

Hypigheder: [rk7!\$C\$4:\$C\$12] -

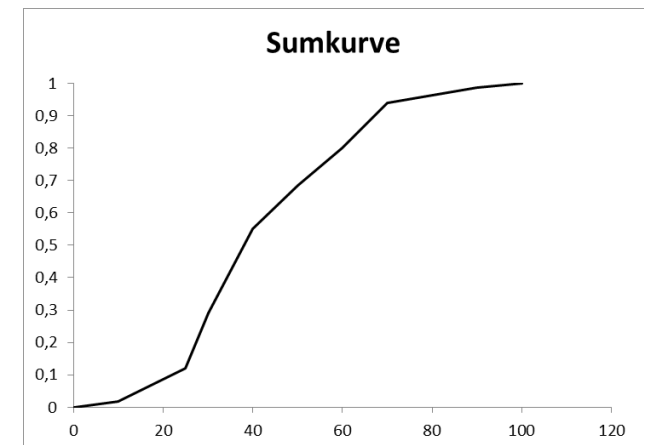
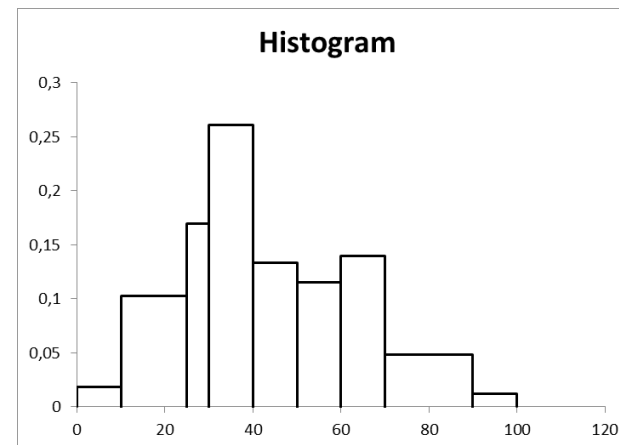
OK

Annullér

Husk øvre grænse!

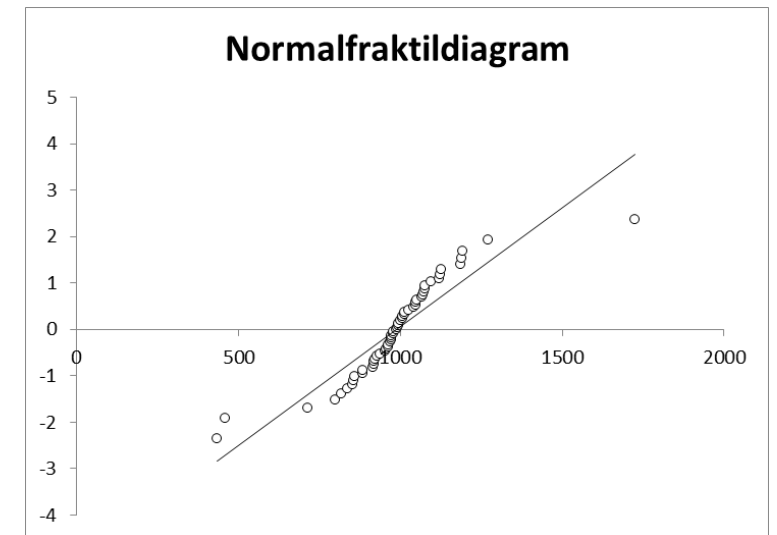
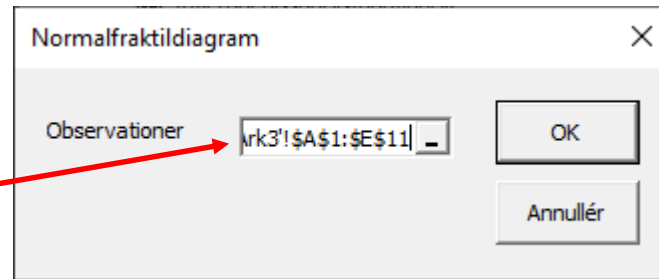
Inputtet kan med fordel være et output fra optællingen i 1.4

Deskriptiv statistik for grupperede data					
Middelvæ	42,10606				
Varians	335,667				
Standardda	18,32122				
Observationssættet					
Nedre	Øvre	Hypigheder	Kum. hyp	Frekvens	Kum. frekvens
0	10	3	3	1,8%	1,8%
10	25	17	20	10,3%	12,1%
25	30	28	48	17,0%	29,1%
30	40	43	91	26,1%	55,2%
40	50	22	113	13,3%	68,5%
50	60	19	132	11,5%	80,0%
60	70	23	155	13,9%	93,9%
70	90	8	163	4,8%	98,8%
90	100	2	165	1,2%	100,0%



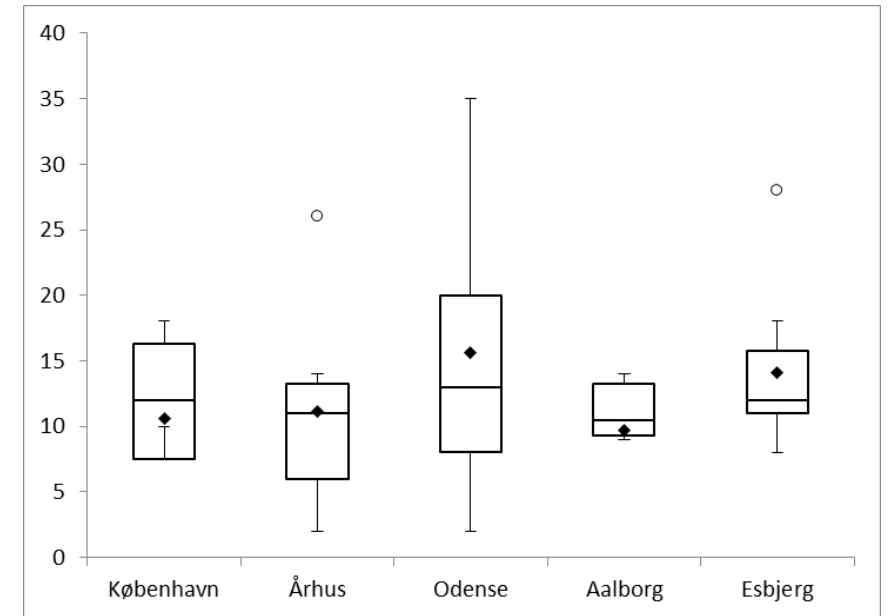
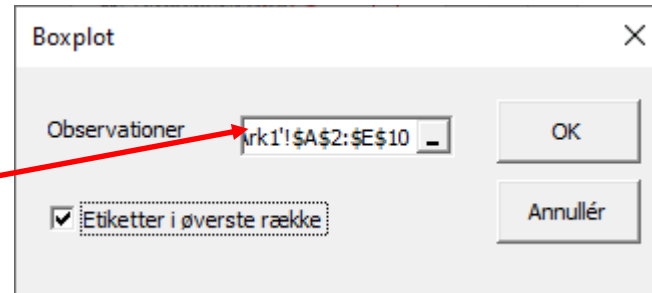
1.7 Normalfraktildiagram

1007,3	458,0	973,0	956,3	1186,1
919,0	970,3	1024,5	852,4	1000,6
964,4	914,6	1049,6	979,8	836,5
970,9	1093,9	963,9	1127,3	1189,4
1119,6	936,8	432,7	1272,6	978,5
1013,7	1076,1	1072,5	883,7	920,4
953,0	857,2	1075,2	1040,3	712,8
992,3	1047,5	1122,0	994,4	883,2
987,4	928,9	858,6	798,5	1068,0
971,9	995,5	919,8	1013,2	1047,0
1191,5	1069,3	1005,8	818,0	1724,1



1.8 Boxplot

Årligt forbrug af ost (kg pr. år)				
København	Århus	Odense	Aalborg	Esbjerg
0	6	20	10	28
17	26	8	14	18
10	13	35	0	8
13	10	13	14	12
11	14	2	9	12
16	6		11	15
18	2			12
0	12			8



1.9 Stripplot (uden og med jitter)

Årligt forbrug af ost (kg pr. år)				
København	Århus	Odense	Aalborg	Esbjerg
0	6	20	10	28
17	26	8	14	18
10	13	35	0	8
13	10	13	14	12
11	14	2	9	12
16	6		11	15
18	2			12
0	12			8

Stripplot

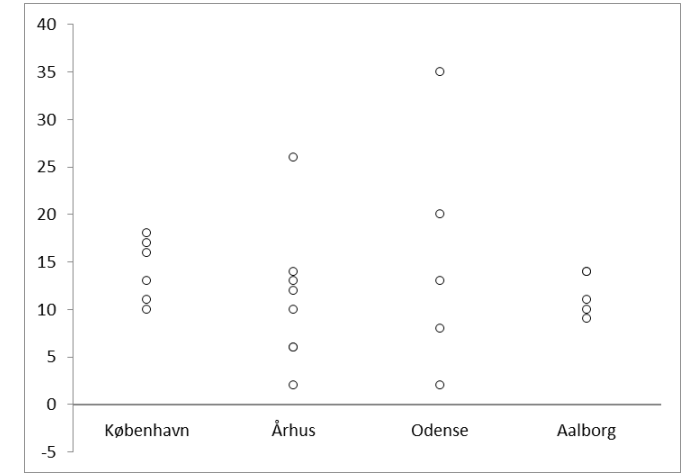
Observationer: ark1!\$A\$2:\$E\$10

Etiketter i øverste række

Medtag jitter

Parameter:

OK Annullér



Stripplot

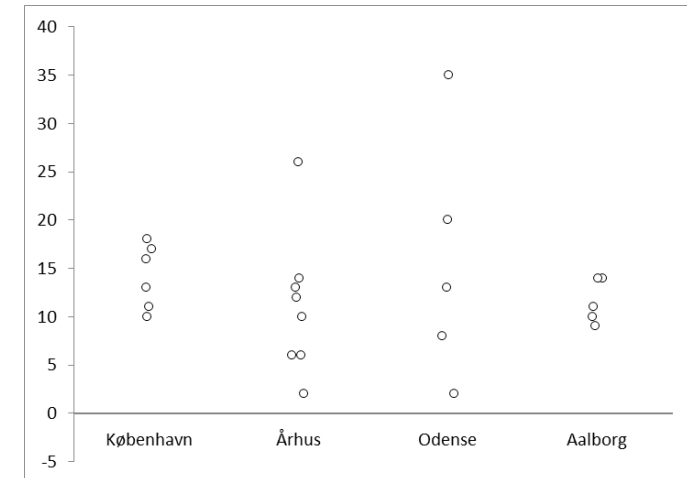
Observationer: ark1!\$A\$2:\$E\$10

Etiketter i øverste række

Medtag jitter

Parameter:

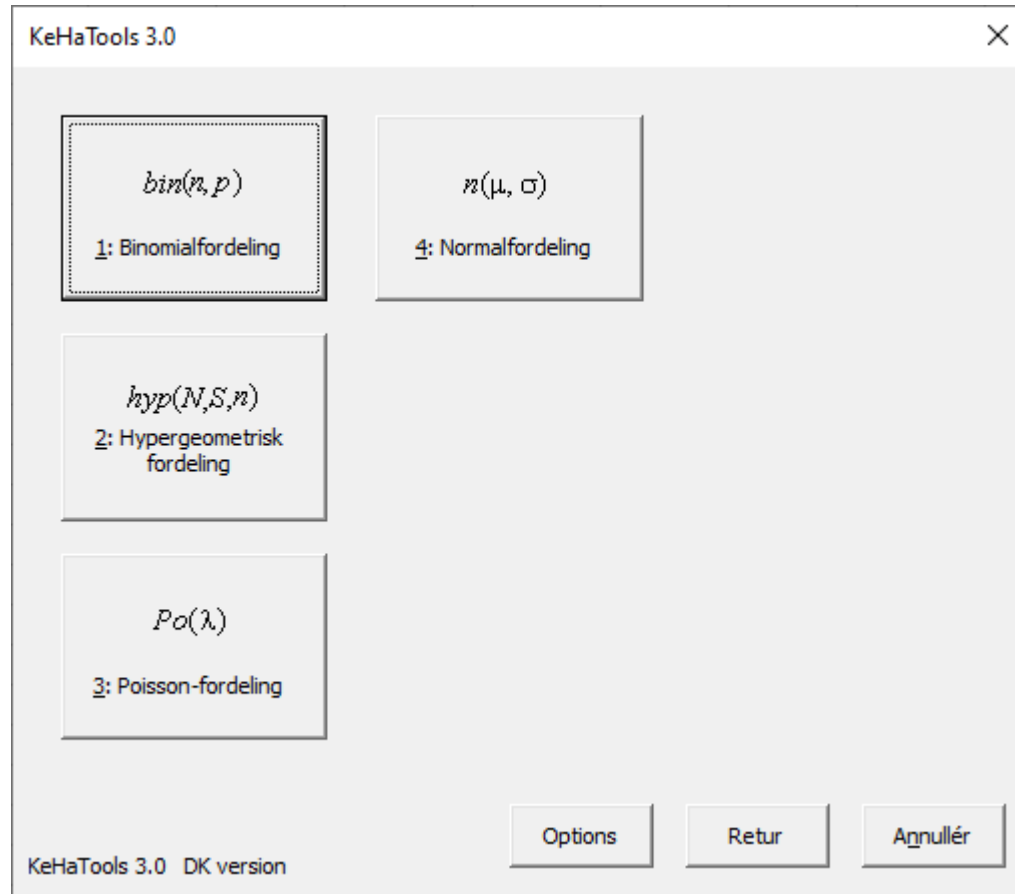
OK Annullér



Denne parameter kontrollerer spredningen af punkterne
Jo større, desto mere spredt

Jitter gør det nemmere at se ved store observationssæt, hvor mange observationer der er, idet datapunkterne ikke ligger oven i hinanden

2. Sandsynligheder



Beregner forskellige størrelser for de givne fordelinger
Genererer sandsynlighedstabeller

Bemærk, at alle beregninger er via Excel-formler ud fra de givne parametre. Man kan derfor anvende Målsøgning eller Solver (Problemløser) til at lave 'omvendt sandsynlighedsregning'.

2.1 (beregninger i) Binomialfordeling

Tabulering af bin(10; 0,2)-fordelingen

Beregninger i binomialfordelingen

Antal gentagelser

Basis-sandsynlighed

OK

Annullér

Beregninger i Binomialfordelingen			
	Antal gentagelser	10	
	Basis-sandsynlighed	0,2	
	Middelværdi	2,0000	
	Varians	1,6000	
	Standardafvigelse	1,2649	
Tabel over sandsynligheder			
k	P(X<=k)	P(X=k)	P(X>=k)
0	0,1074	0,1074	1,0000
1	0,3758	0,2684	0,8926
2	0,6778	0,3020	0,6242
3	0,8791	0,2013	0,3222
4	0,9672	0,0881	0,1209
5	0,9936	0,0264	0,0328
6	0,9991	0,0055	0,0064
7	0,9999	0,0008	0,0009
8	1,0000	0,0001	0,0001
9	1,0000	0,0000	0,0000
10	1,0000	0,0000	0,0000

2.2 (beregninger i) Hypergeometrisk fordeling

Tabulering af $\text{hyp}(100; 7; 5)$ -fordelingen

Beregninger i den hypergeometriske fordeling

Populationsstørrelse

Antal defekte

Stikprøvestørrelse

OK

Annullér

Beregninger i den hypergeometriske fordeling			
	Populatio	100	
	Antal defe	7	
	Stikprøve	5	
	Middelvæ	0,3500	
	Varians	0,3123	
	Standarda	0,5589	
Tabel over sandsynligheder			
k	P(X≤k)	P(X=k)	P(X≥k)
0	0,6903	0,6903	1,0000
1	0,9618	0,2715	0,3097
2	0,9980	0,0362	0,0382
3	1,0000	0,0020	0,0020
4	1,0000	0,0000	0,0000
5	1,0000	0,0000	0,0000

2.3 (beregninger i) Poisson-fordelingen

Tabulering af Po(4)-fordelingen

Beregninger i Poisson-fordelingen

Intensitet OK

Startværdi Annullér

Slutværdi

Da Poisson-fordelingen kan antage alle mulige heltallige, positive værdier, skal man begrænse tabellen. Standardværdierne på 0 og 20 plejer at være tilstrækkelige.

Beregninger i Poisson-fordelingen			
	Intensitet	4	
	Middelvæ	4,0000	
	Varians	4,0000	
	Standarda	2,0000	
Tabel over sandsynligheder			
k	P(X<=k)	P(X=k)	P(X>=k)
0	0,0183	0,0183	1,0000
1	0,0916	0,0733	0,9817
2	0,2381	0,1465	0,9084
3	0,4335	0,1954	0,7619
4	0,6288	0,1954	0,5665
5	0,7851	0,1563	0,3712
6	0,8893	0,1042	0,2149
7	0,9489	0,0595	0,1107
8	0,9786	0,0298	0,0511
9	0,9919	0,0132	0,0214
10	0,9972	0,0053	0,0081
11	0,9991	0,0019	0,0028
12	0,9997	0,0006	0,0009

2.4 (beregninger i) Normalfordelingen

Beregninger i normalfordelingen

Middelværdi

Standardafvigelse

Beregn gennemsnit af flere uafhængige normalfordelte variable

Antal variable

Nedre grænse

Øvre grænse

OK

Annullér

$X \sim n(20;10)$

$P(X \leq 18)$

$P(X \geq 18)$

$P(18 \leq X \leq 20)$

Beregninger i normalfordelingen	
Middelværdi	20
Standardafvigelse	10
Beregninger af sandsynligheder	
a =	18
P(X <= a)	0,42074
P(X >= a)	0,57926
a =	18
b =	25
P(a <= X <= b)	0,270722

Anvendes hvis man har et antal uafhængige, ens fordelte, normalfordelte stokastiske variable.

2.4 Omvendt sandsynlighedsregning

X er normalfordelt med $\sigma=20$.
Bestem μ , således at $P(X \geq 45) < 5\%$

Bare skriv noget, målsøgning vil ændre tallet alligevel

Skriv det samme som a ovenfor

Beregninger i normalfordelingen

Middelværdi: 50

Standardafvigelse: 20

Beregn gennemsnit af flere uafhængige normalfordelte variable

Antal variable: []

Nedre grænse: 45

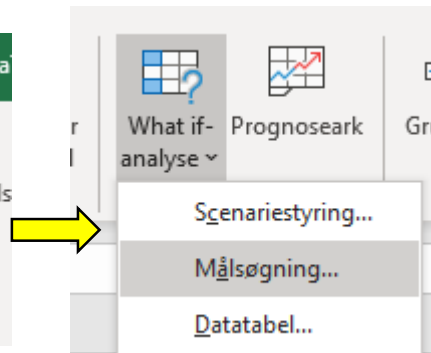
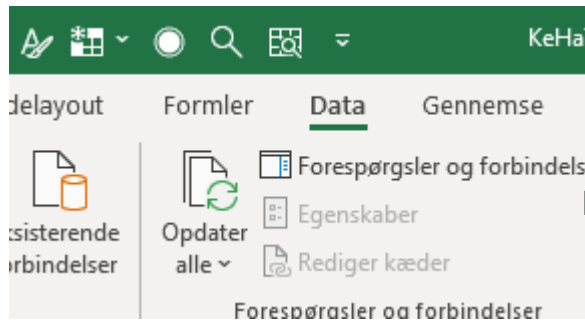
Øvre grænse: 45

OK

Annullér

Beregninger i normalfordelingen	
Middelvæ	50
Standarda	20
Beregninger af sandsynlighede	
a =	45
P(X <= a)	0,401294
P(X >= a)	0,598706
a =	45
b =	45
P(a <= X <	0

Beregninger i normalfordelingen	
Middelvæ	12,28884
Standarda	20
Beregninger af sandsynlighede	
a =	45
P(X <= a)	0,949034
P(X >= a)	0,050966
a =	45
b =	45
P(a <= X <	0



Målsøgning

Angiv celle: \$B\$12

Til værdi: 5%

Ved ændring af celle: \$B\$4

OK

Annuller

Målsøgningstatus

Målsøgning med celle B12 fandt en løsning.

Målværdi: 0,05

Aktuel værdi: 0,050966071

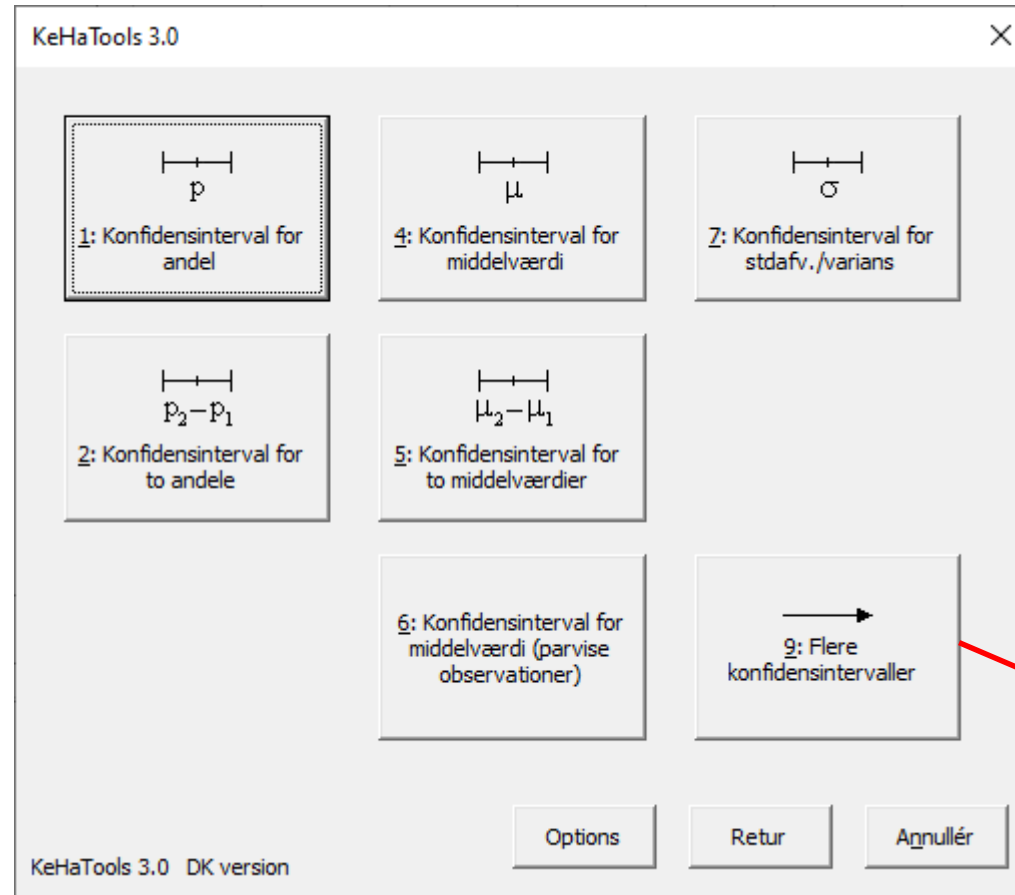
Trin

Afbryd midlertidigt

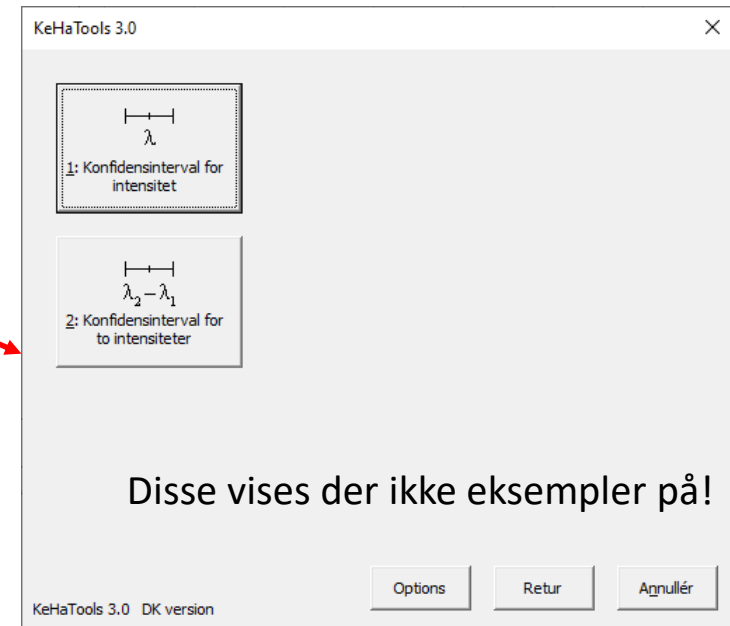
OK

Annuller

3. Konfidensintervaller



I alle beregninger er konfidensniveauet som standard sat til 95%. Man kan manuelt ændre denne værdi ved at overskrive de 95%.



3.1 Konfidensinterval for andel

Konfidensinterval for andel

Stikprøvestørrelse: 100

Antal gunstige: 25

Populationsstørrelsen kendes

Populationsstørrelse: []

Konfidensniveau: 95%

OK, Annullér

Approximativt konfidensinterval for andel

	x
Størrelse	100
Antal gunstige	25
Estimeret andel	25,00%
Population	N/A
Konfidensniveau	95,00%
Forudsætning	18,8
Nedre grænse	16,51%
Øvre grænse	33,49%

Ukendt populationsstørrelse

Approximativt konfidensinterval for andel

	x
Størrelse	100
Antal gunstige	25
Estimeret andel	25,00%
Population	N/A
Konfidensniveau	95,00%
Forudsætning	18,8
Nedre grænse	16,51%
Øvre grænse	33,49%

Populationsandelen estimeres til 25,00%, Med 95% sikkerhed ligger populationsandelen mellem 16,51% og 33,49%.

Forudsætning: $n \cdot p \cdot (1-p)$ skal være mindst 9

Konfidensinterval for andel

Stikprøvestørrelse: 100

Antal gunstige: 25

Populationsstørrelsen kendes

Populationsstørrelse: 100000

Konfidensniveau: 95%

OK, Annullér

Approximativt konfidensinterval for andel

	x
Størrelse	100
Antal gunstige	25
Estimeret andel	25,00%
Population	100000
Konfidensniveau	95,00%
Forudsætning	18,8
Nedre grænse	16,52%
Øvre grænse	33,48%

Kendt populationsstørrelse

$n \cdot p \cdot (1-p)$

3.2 Konfidensinterval for to andele

Konfidensinterval for to andele

Stikprøvestørrelse	Stikprøve 1 100	Stikprøve 2 250	OK
Antal gunstige	25	140	Annullér
Konfidensniveau	95%		

Konfidensinterval for forskellen på to andele

	x	y
Størrelse	100	250
Antal gunstige	25	140
Estimeret andel	25,00%	56,00%
Forudsætning	18,75	61,6
Konfidensniveau	95%	
Forskel på andele	31,00%	
Nedre grænse	20,52%	
Øvre grænse	41,48%	

Punktestimater for de to andele

Estimat for $andel_2 - andel_1$

Forudsætning:
 $n \cdot p \cdot (1-p)$ skal være mindst 9
for begge stikprøver

Forskellen mellem andel 2 og andel 1
estimeres til 31,00%.
Med 95% sikkerhed ligger denne forskel
mellem 20,52% og 41,48%.

Husk:
stikprøve 2 - stikprøve 1,
ikke omvendt

3.3 Konfidensinterval for middelværdi

9,28
9,73
10,69
8,05
10,42
11,63
9,18
10,62
10,75

Konfidensinterval for middelværdi

Observationer Resumé-data

Observationer:

Stikprøvestørrelse:

Gennemsnit:

Standardafvigelse:

Populationsstørrelsen kendes

Populationsstørrelse:

Konfidensniveau:

z-KI t-KI

OK Annullér

	x
Størrelse	9
Gennemsnit	10,04
Standardafvigelse	1,08
Konfidensniveau	95%
Population	0
fg	8
Nedre grænse	9,33
Øvre grænse	10,74

Middelværdien estimeres til **10,04**.

Med 95% sikkerhed ligger middelværdien mellem **16,51%** og **33,49%**.

Forudsætninger:

z-KI: Stikprøven er stor (mindst 30)

t-KI: Stikprøven er normalfordelt

3.5 Konfidensinterval for to middelværdier

Stikprøve 1	Stikprøve 2
9,28	12,53
9,73	11,01
10,69	11,17
8,05	14,33
10,42	10,91
11,63	12,47
9,18	12,81
10,62	
10,75	

Konfidensinterval for to middelværdier

Observationer Resumé-data

Stikprøve 1: Stikprøve 2:

Størrelse:

Gennemsnit:

Standardafvigelse:

Konfidensniveau:

z-KI t-KI t-KI (Welch)

OK Annullér

Konfidensinterval for to middelværdier (z-fordeling)		
	x	y
Størrelse	9	7
Gennemsnit	10,04	12,18
Standardafvigelse	1,08	1,24
Konfidensniveau	95%	
fg	0	
Forskel på middelværdier	2,14	
Nedre grænse	0,98	
Øvre grænse	3,29	

Husk:
stikprøve 2 - stikprøve 1,
ikke omvendt

Forudsætninger:

- z-KI: Stikprøverne er store (begge mindst 30)
- t-KI: Stikprøverne er normalfordelte og har *samme* standardafvigelse (se test 4.8)
- Welch: Stikprøverne er normalfordelte og har *forskellige* standardafvigelse (se test 4.8)

Middelværdi 2 minus middelværdi 1 estimeres til 2,14.
Med 95% sikkerhed ligger forskellen på middelværdierne mellem 0,98 og 3,29.

3.6 Konfidensinterval for middelværdi, parvise obs.

Stikprøve 1	Stikprøve 2
8,8	12,6
10,2	15,1
11,4	13,1
10,3	13,8
10,6	13,5
9,3	13,5
10,9	11,3
11,6	13,2
9,2	14,1
11,2	11,8
12,5	12,7

Konfidensinterval for middelværdi - parvise observationer

Observationer Resumé-data

Stikprøve 1: Stikprøve 2:

Størrelse:

Stikprøvegennemsnit for forskellene:

Standardafvigelse for forskellene:

Konfidensniveau:

z-KI t-KI

OK Annullér

	y-x
Størrelse	11
Gennemsnit af forskellene	2,62
Standardafvigelse for forskellene	1,77
Konfidensniveau	95%
fg	10
Nedre grænse	1,43
Øvre grænse	3,80

Husk:
stikprøve 2 - stikprøve 1,
ikke omvendt

Forudsætninger:
z-KI: Stikprøven er stor (mindst 30)
t-KI: Differencerne er normalfordelte

Ændringen i middelværdi estimeres til 2,62.
Med 95% sikkerhed ligger ændringen i middelværdierne mellem 1,43 og 3,80.

3.7 Konfidensinterval for standardafvigelse

9,28
9,73
10,69
8,05
10,42
11,63
9,18
10,62
10,75

Konfidensinterval for standardafvigelse / varians

Observationer Resumé-data

Observationer: k12!\$H\$2:\$H\$10

Stikprøvestørrelse: []

Standardafvigelse: []

Konfidensniveau: 95%

OK

Annullér

Konfidensinterval for standardafvigelse	
	x
Størrelse	9
Standardafvigelse	1,08
Varians	1,16
Konfidensinterval	95%
fg	8
Standardafvigelse	
Nedre grænse	0,73
Øvre grænse	2,07
Varians	
Nedre grænse	0,53
Øvre grænse	4,27

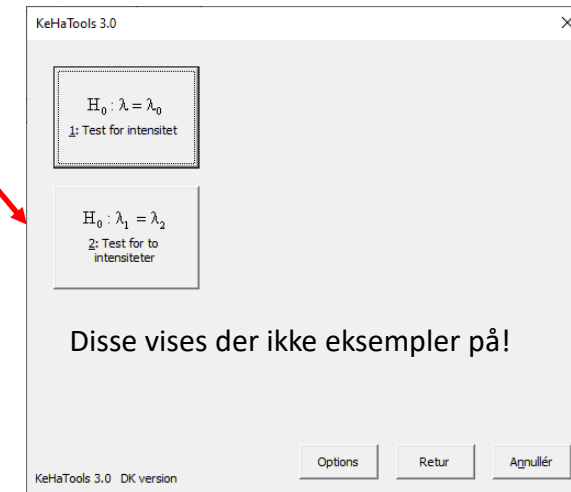
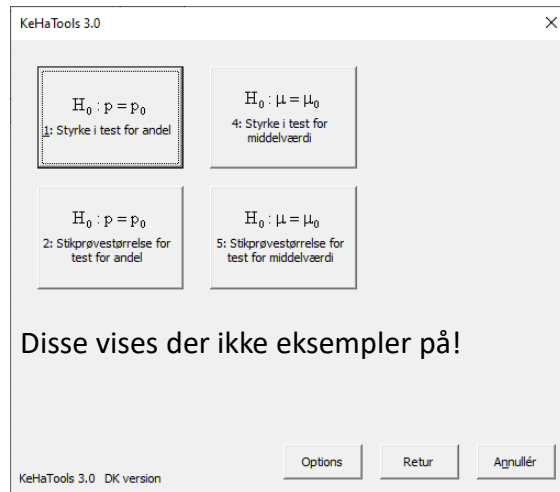
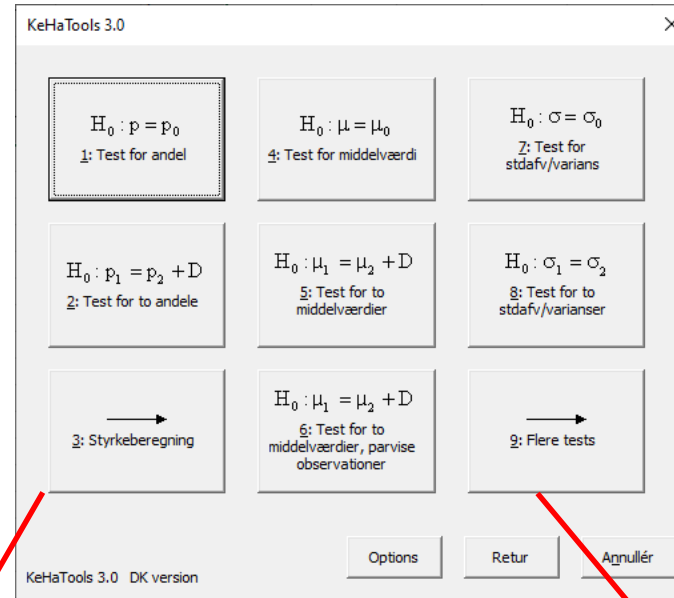
Forudsætning:
Stikprøven skal være normalfordelt

Standardafvigelsen estimeres til 1,08. Med 95% sikkerhed er standardafvigelsen mellem 0,73 og 2,07.

Variansen estimeres til 1,16. Med 95% sikkerhed er variansen mellem 0,53 og 4,27.

4. Hypotesetest

Alle p -værdier er vist som procenttal med 2 decimaler. Hvis der står 0,00% så betyder det ikke, at p -værdien er 0, men blot at den er meget lille. Du kan ændre formateringen af cellen hvis du vil kende den præcise værdi.



4.1 Test for andel

Test for andel

Stikprøvestørrelse

Antal gunstige

Nul-værdi

Tilnærmet test (z-test) Eksakt test

Venstresidet test Tosidet test Højresidet test

OK

Annullér

$$H_0 : p = 0,20$$
$$H_A : p \neq 0,20$$

Approx. z-test for andel - tosidet	
	x
Størrelse	100
Antal gunstige	25
Estimeret andel	25,00%
Nul-værdi	20,00%
Forudsætning	16,00
Teststørrelse	1,25
p-værdi	21,13%

Forudsætning:
 $n \cdot p \cdot (1-p)$ skal være mindst 9

Høj p -værdi, så nulhypotesen kan ikke forkastes.
Der er altså ingen beviser imod, at andelen er på 0,20.

4.2 Test for to andele (D=0)

Forudsætning:
 $n \cdot p \cdot (1-p)$ skal være mindst 9
for begge stikprøver

Test for to andele

Stikprøvestørrelse	Stikprøve 1	Stikprøve 2	
	100	250	OK
Antal gunstige	25	140	Annullér
Difference (D)	0		

Venstresidet test Tosidet test Højresidet test

$$H_0 : p_1 \geq p_2$$
$$H_A : p_1 < p_2$$

Approximativ test for to andele - venstresidet

	x	y
Størrelse	100	250
Antal gunstige	25	140
Estimeret andel	25,0%	56,0%
Difference (D)	0,0%	
Forudsætninger	18,8	61,6
Teststørrelse	-0,06	
p-værdi	47,54%	

Høj p -værdi, så nulhypotesen kan ikke forkastes.
Det kan altså ikke konkluderes, at andelen fra population 1 er mindre end andelen fra population 2.

4.2 Test for to andele ($D \neq 0$)

Forudsætning:
 $n \cdot p \cdot (1-p)$ skal være mindst 9
for begge stikprøver

Test for to andele

	Stikprøve 1	Stikprøve 2
Stikprøvestørrelse	100	250
Antal gunstige	25	140
Difference (D)	-30%	

Venstresidet test Tosidet test Højresidet test

$$H_0 : p_1 \geq p_2 - 30\%$$
$$H_A : p_1 < p_2 - 30\%$$

Det undersøges, om andelen i population 2 er mindst 30% større end i population 1

Approximativ test for to andele - venstresidet

	x	y
Størrelse	100	250
Antal gunstige	25	140
Estimeret andel	25,0%	56,0%
Difference (D)	-30,0%	
Forudsætninger	18,8	61,6
Teststørrelse	-0,19	
p-værdi	42,58%	

Høj p -værdi, så nulhypotesen kan ikke forkastes.
Det kan altså ikke konkluderes, at andelen fra population 2 er mindst 30% større end andelen fra population 1.

4.4 Test for middelværdi

9,28
9,73
10,69
8,05
10,42
11,63
9,18
10,62
10,75

Test for middelværdi

Observationer Resumé-data

Observationer: k12!\$H\$2:\$H\$10

Stikprøvestørrelse: []

Gennemsnit: []

Standardafvigelse: []

Nul-værdi: 12

z-test t-test

Venstresidet test Tosidet test Højresidet test

OK

Annulér

Stikprøven er lille, men et normalfraktildiagram afslører, at der er tale om en normalfordelt stikprøve

Forudsætninger:
z-test: Stikprøven er stor (mindst 30)
t-test: Stikprøven er normalfordelt

$$H_0 : \mu = 12$$
$$H_A : \mu \neq 12$$

	x	
Størrelse	9	
Gennemsnit	10,04	
Standardafvigelse	1,08	
Nul-værdi	12	
fg	8	
teststørrelse	-5,45	
p-værdi	0,06%	

Meget lav p -værdi, så nulhypotesen kan forkastes. Det kan altså konkluderes, at middelværdien ikke er 12.

4.5 Test for to middelværdier (D=0)

Stikprøve 1	Stikprøve 2
9,28	12,53
9,73	11,01
10,69	11,17
8,05	14,33
10,42	10,91
11,63	12,47
9,18	12,81
10,62	
10,75	

Test for to middelværdier

Observationer Resumé-data

Stikprøve 1: Stikprøve 2:

Observationer:

Stikprøvestørrelse:

Gennemsnit:

Standardafvigelse:

Difference (D):

z-test t-test t-test (Welch's test)

Venstresidet test Tosidet test Højresidet test

OK Annullér

t-test: To stikprøver med ens standardafvigelse - venstresidet		
	x	y
Størrelse	9	7
Gennemsnit	10,04	12,18
Standardafvigelse	1,08	1,24
Difference (D)	0	
fg	14	
Fælles standardafvigelse	1,15	
teststørrelse	-3,68	
p-værdi	0,12%	

Forudsætninger:
Normalfraktildiagrammer og beregningen til test 4.8

Forudsætninger:
z-test: Stikprøverne er store (begge mindst 30)
t-test: Stikprøverne er normalfordelte og har *samme* standardafvigelse (se test 4.8)
Welch: Stikprøverne er normalfordelte og har *forskellige* standardafvigelse (se test 4.8)

$$H_0 : \mu_1 \geq \mu_2$$

$$H_A : \mu_1 < \mu_2$$

Meget lav p-værdi, så nulhypotesen kan forkastes. Det kan altså konkluderes, at middelværdi 2 er større end middelværdi 1.

4.5 Test for to middelværdier ($D \neq 0$)

Stikprøve 1	Stikprøve 2
9,28	12,53
9,73	11,01
10,69	11,17
8,05	14,33
10,42	10,91
11,63	12,47
9,18	12,81
10,62	
10,75	

Test for to middelværdier

Observationer Resumé-data

Stikprøve 1: Stikprøve 2:

Observationer:

Stikprøvestørrelse:

Gennemsnit:

Standardafvigelse:

Difference (D):

z-test t-test t-test (Welch's test)

Venstresidet test Tosidet test Højresidet test

OK Annullér

	x	y
Størrelse	9	7
Gennemsnit	10,04	12,18
Standardafvigelse	1,08	1,24
Difference (D)	-3	
fg	14	
Fælles standardafvigelse	1,15	
teststørrelse	1,49	
p-værdi	7,94%	

Samme setup som på foregående slide, men hypoteserne er lidt anderledes:
 "Er middelværdi 2 højst 3 mindre end middelværdi 1.

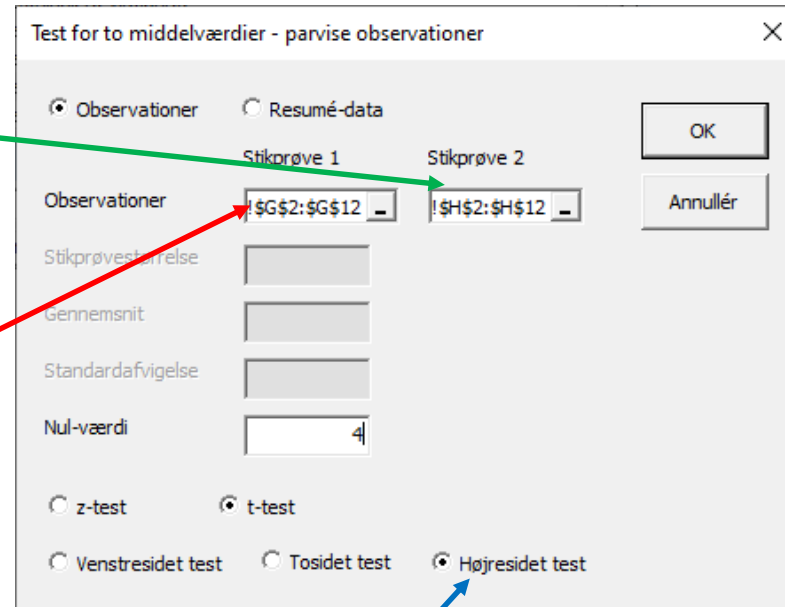
$$H_0 : \mu_1 \leq \mu_2 - 3$$

$$H_A : \mu_1 > \mu_2 - 3$$

Relativt p -værdi, så nulhypotesen kan ikke forkastes.
 Det kan altså ikke konkluderes, at middelværdi 2 er højst 3 mindre end middelværdi 1.
 Da p -værdien er tæt på 5%, så er denne konklusion ikke specielt velbegrunderet

4.6 Test for middelværdi, parvise obs.

Stikprøve 1	Stikprøve 2
8,8	12,6
10,2	15,1
11,4	13,1
10,3	13,8
10,6	13,5
9,3	13,5
10,9	11,3
11,6	13,2
9,2	14,1
11,2	11,8
12,5	12,7



p-værdien er moderat lille, mellem 1% og 5%.
 Nulhypotesen kan forkastes, og der er derfor belæg for påstanden om, at ændringen i middelværdi er mindst 4.

Er ændringen i middelværdi fra population 1 til population 2 mindst 4?

$$H_0 : \mu_d \leq 4$$

$$H_A : \mu_d > 4$$

	y - x
Størrelse	11
Gennemsnit for forskel	2,62
Standardafvigelse for forskel	1,77
Nul-værdi	4
fg	10
Teststørrelse	-2,60
<i>p</i> -værdi	1,33%

4.7 Test for standardafvigelse

9,28
9,73
10,69
8,05
10,42
11,63
9,18
10,62
10,75

Test for standardafvigelse / varians

Observationer Resumé-data

Observationer: k12!\$H\$2:\$H\$10

Stikprøvestørrelse: []

Standardafvigelse: []

Nul-værdi: 1

Venstresidet test Tosidet test Højresidet test

OK

Annullér

$$H_0 : \sigma = 1$$
$$H_A : \sigma \neq 1$$

Forudsætning:
Stikprøven skal være normalfordelt

ChiKvadrat-test for standardafvigelse - tosid

	x	
Størrelse	9	
Standardafvigelse	1,08	
Nul-værdi	1,00	
fg	8	
Teststørrelse	9,32	
p-værdi	63,25%	

p-værdien er høj, så nulhypotesen kan ikke forkastes. Det kan altså ikke konkluderes, at standardafvigelsen ikke er 1.

4.8 Test for to standardafvigelser

Stikprøve 1	Stikprøve 2
9,28	12,53
9,73	11,01
10,69	11,17
8,05	14,33
10,42	10,91
11,63	12,47
9,18	12,81
10,62	
10,75	

Test for to standardafvigelser / varianser

Observationer Resumé-data

Stikprøve 1: Stikprøve 2:

Observationer:

Stikprøvestørrelse:

Standardafvigelse:

Venstresidet test Tosidet test Højresidet test

OK Annullér

Forudsætning:
Begge stikprøver skal være normalfordelt

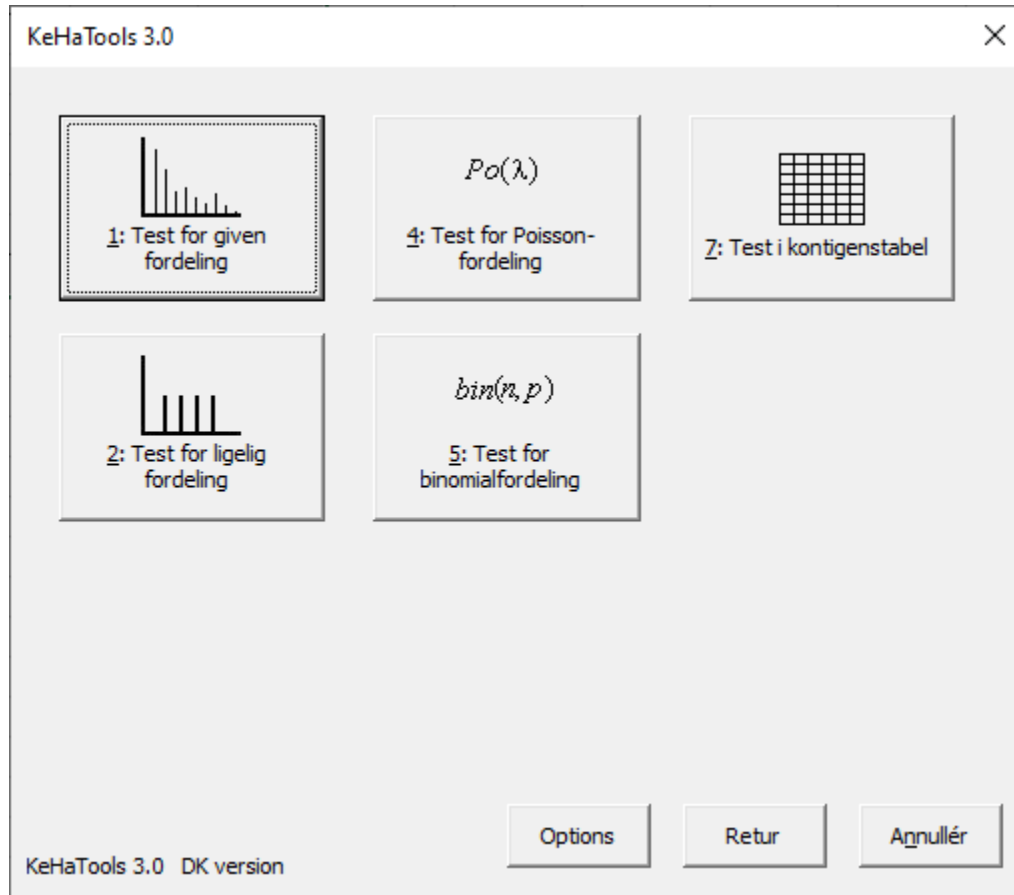
$$H_0 : \sigma_1 = \sigma_2$$

$$H_A : \sigma_1 \neq \sigma_2$$

	x	y
Størrelse	9	7
Standardafvigelse	1,08	1,24
fg	8	6
teststørrelse	0,76	
p-værdi	69,64%	

p-værdien er høj, så nulhypotesen kan ikke forkastes.
Der er altså intet, der taler imod, at de to standardafvigelser kan være ens.
(Varianshomogenitet - en forudsætning til test 4.5)

5. Goodness-of-fit tests



En generel forudsætning for disse tests er, at det forventede antal i hver celle skal være over 5.

Hvis dette ikke er tilfældet, så må man enten anskaffe sig flere observationer eller begynde at slå celler sammen.

Forudsætning:

Alle forventede værdier ≥ 5 .

5. Goodness-of-fit tests: Heatmaps

Alle disse beregninger tillader brugen af heatmaps

Goodness-of-fit test for ligelig fordeling				
Kategori	pi	Ei	Oi	Std. residu
mandag	14,29%	248,86	280	1,97
tirsdag	14,29%	248,86	230	-1,20
onsdag	14,29%	248,86	256	0,45
torsdag	14,29%	248,86	308	3,75
fredag	14,29%	248,86	332	5,27
lørdag	14,29%	248,86	209	-2,53
søndag	14,29%	248,86	127	-7,72
Sum	1	1742	1742	113,42
Frihedsgr	6			
Teststørre	113,42			
p-værdi	0,00%			

De standardiserede residualer, som måler forskellen mellem den observerede og den forventede værdi, farves efter værdien:

blå: signifikant for få observationer i forhold til det forventede

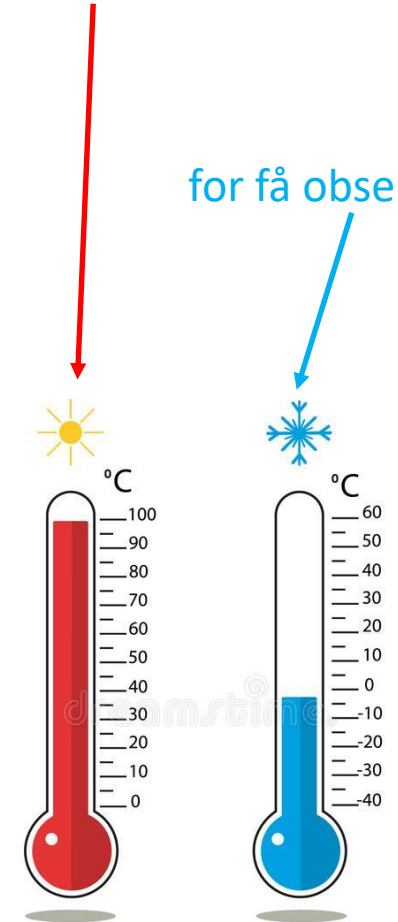
hvid: ingen signifikant forskel mellem det observerede og det forventede antal

rød: signifikant for mange observationer i forhold til det forventede

(rød: over 1,96; blå: under -1,96)

for mange observerede

for få observerede



5.1 Test for en given fordeling

Antal trafikulykker fordelt på ugens syv dage

Ugedag	Antal ulyk
mandag	280
tirsdag	230
onsdag	256
torsdag	308
fredag	332
lørdag	209
søndag	127

Goodness-of-fit test for ligelig fordeling

Etiketter for kategorier

Etiketter: 10. 1! \$A \$2: \$A \$8

Observerede antal: 10. 1! \$B \$2: \$B \$8

Heatmap for residualer

OK

Annullér

Goodness-of-fit test for ligelig fordeling				
Kategori	pi	Ei	Oi	Std. residu
mandag	14,29%	248,86	280	1,97
tirsdag	14,29%	248,86	230	-1,20
onsdag	14,29%	248,86	256	0,45
torsdag	14,29%	248,86	308	3,75
fredag	14,29%	248,86	332	5,27
lørdag	14,29%	248,86	209	-2,53
søndag	14,29%	248,86	127	-7,72
Sum	1	1742	1742	113,42
Frihedsgr	6			
Teststørre	113,42			
p-værdi	0,00%			

p -værdien er meget lav, så nulhypotesen om en ligelig fordeling forkastes. Det ses, at der er for mange ulykker på mandage, torsdage og fredage end forventet, og at der er for få ulykker på lørdage og især søndage end forventet.

Forudsætning:
Alle forventede værdier ≥ 5 .

5.2 Test for given fordeling

Ugedag	Antal ulykker	Forventet andel
mandag	280	11%
tirsdag	230	15%
onsdag	256	20%
torsdag	308	15%
fredag	332	20%
lørdag	209	12%
søndag	127	7%
	1742	100%

Goodness-of-fit test for given fordeling				
Kategori	pi	Ei	Oi	Std. residual
mandag	11%	191,62	280	6,38
tirsdag	15%	261,30	230	-1,94
onsdag	20%	348,40	256	-4,95
torsdag	15%	261,30	308	2,89
fredag	20%	348,40	332	-0,88
lørdag	12%	209,04	209	0,00
søndag	7%	121,94	127	0,46
Sum	1	1742	1742	78,35
Frihedsgrader	6			
Teststørrelse	78,35			
p-værdi	0,00%			

Goodness-of-fit test for given fordeling

Etiketter for kategorier

Etiketter: 10.1!'\$A\$2:\$A\$8

Forventede andele: 10.1!'\$C\$2:\$C\$8

Observerede antal: 10.1!'\$B\$2:\$B\$8

Heatmap for residualer

OK

Annullér

Forudsætning:
Alle forventede værdier ≥ 5 .

p -værdien er meget lav, så nulhypotesen om, at antal ulykker følger den givne fordeling, kan forkastes. Det ses, at der er for mange ulykker på mandage og torsdage, men at der observeres for få ulykker på onsdage end forventet. (og egl. også tirsdage, men det fanges ikke af det anvendte heatmap.)

5.4 Test for Poisson-fordeling

Antal solgte biler pr. døgn	Antal døgn
0	12
1	22
2	29
3	16
4	9
5	0
6	3
I alt	91

Goodness-of-fit test for Poisson-fordeling

Intensiteten estimeres ud fra stikprøve

Intensitet:

Værdier:

Observerede antal:

Heatmap for residualer

OK

Annullér

Forudsætningsbrud:
Slå cellerne 4, 5, 6 antal
sammen og prøv igen!

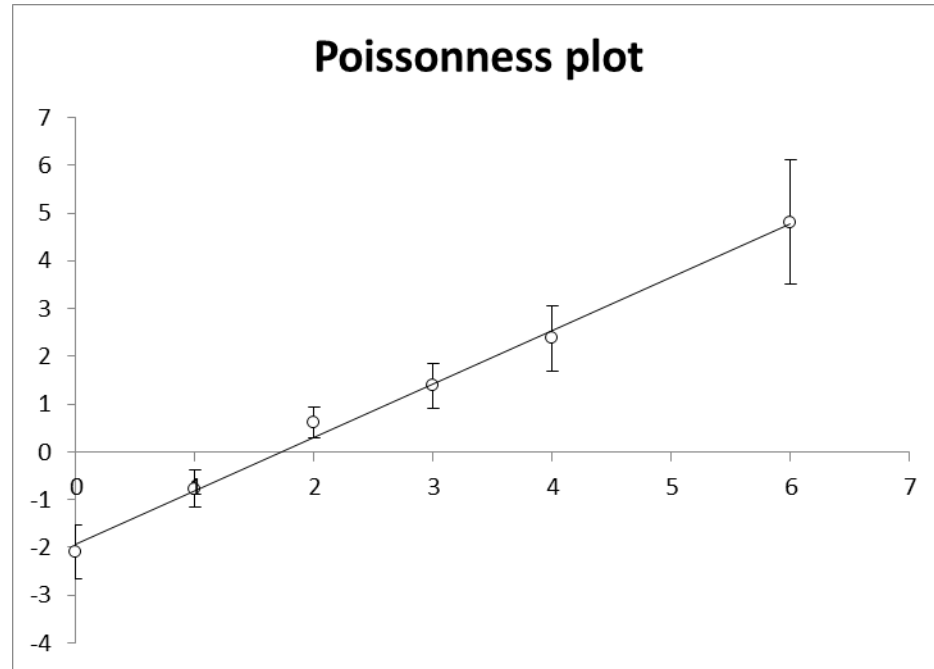
i	pi	Ei	Oi	Std. residu
0	13,53%	12,32	12	-0,09
1	27,07%	24,63	22	-0,53
2	27,07%	24,63	29	0,88
3	18,04%	16,42	16	-0,10
4	9,02%	8,21	9	0,28
5	3,61%	3,28	0	-1,81
6	1,66%	1,51	3	1,22
Sum	1	91	91	5,91
Intensitet	2,00			
Frihedsgr	6			
Teststørre	5,91			
p-værdi	43,30%			

Følger antallet af solgte biler pr. døgn en Poisson-fordeling med $\lambda = 2$?

p -værdien er høj, så vi kan ikke forkaste nulhypotesen, og der er altså ikke noget som taler imod, at antallet af solgte biler pr. døgn er Poisson-fordelt med.
Forudsætningen for testen er dog ikke opfyldt!

Forudsætning:
Alle forventede værdier ≥ 5 .

5.4 Test for Poisson-fordeling



I outputtet er der også et 'Poissonness-plot'. Dette er en specialiseret graftype, som kan bruges til at vise, om data følger en Poisson-fordeling. Hvis datapunkterne ligger tæt på den rette linje (dvs. mere eller mindre indenfor de lodrette fejlinjer), så er der tale om en Poisson-fordeling.

Fordelen ved dette plot er, at forudsætningen om forventede værdier mindst 5 ikke er aktuel!

Forudsætning:
Alle forventede værdier ≥ 5 .

5.5 Test for binomialfordeling

Følger antallet af knuste æg i en 6-pakkers æggebakke en binomialfordeling?

Antal knuste æg	Antal bakker
0	606
1	318
2	67
3	8
4	0
5	1
6	0

Goodness-of-fit test for binomialfordeling

Basissandsynligheden estimeres ud fra stikprøve

Basissandsynlighed:

Antalsparameter:

Værdier:

Observerede antal:

Heatmap for residualer

OK Annullér

Forudsætningsbrud:
Man skal slå kategorierne 3, 4, 5 og 6 sammen og prøve igen!

Goodness-of-fit test for binomialfordeling				
i	pi	Ei	Oi	Std. residu
0	60,57%	605,70	606	0,01
1	31,67%	316,73	318	0,07
2	6,90%	69,01	67	-0,24
3	0,80%	8,02	8	-0,01
4	0,05%	0,52	0	-0,72
5	0,00%	0,02	1	7,26
6	0,00%	0,00	0	-0,02
Sum	1	1000	1000	53,33

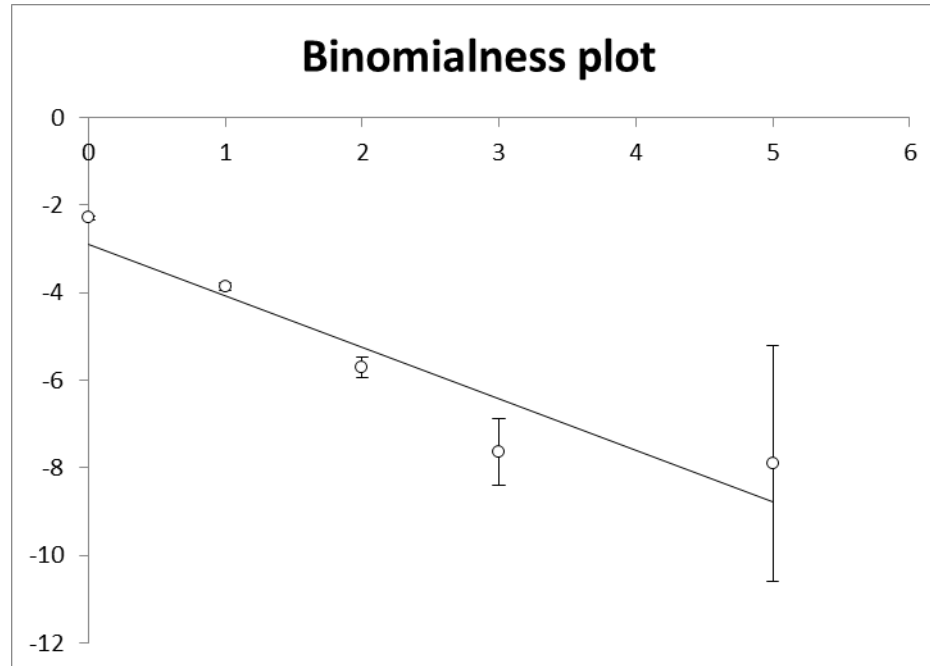
Antal	6
Basissand	0,08
Frihedsgr	5
Teststørre	53,33
p-værdi	0,00%

Forudsætning:
Alle forventede værdier ≥ 5 .

p -værdien er meget lav, så vi kan forkaste nulhypotesen.

Men forudsætningen er ægge opfyldt, så konklusionen skal tages med et gran salt.

5.5 Test for binomialfordeling



I outputtet er der også et 'binomialness-plot'. Dette er en specialiseret graftype, som kan bruges til at vise, om data følger en binomialfordeling. Hvis datapunkterne ligger tæt på den rette linje (dvs. mere eller mindre indenfor de lodrette fejlinjer), så er der tale om en binomialfordeling.

Fordelen ved dette plot er, at forudsætningen om forventede værdier mindst 5 ikke er aktuel!

Forudsætning:
Alle forventede værdier ≥ 5 .

5.7 Test i kontingenstabel

Er der uafhængighed mellem flyrute og kundetilfredshed?

	Flyrute			I alt
	Kbh-Tir	Kbl-Aal	Kbl-Kar	
Utilfreds	88	64	29	181
Neutral	15	35	10	60
Tilfreds	37	35	67	139
I alt	140	134	106	380

Goodness-of-fit test i kontingenstabel

Observerede antal

Første række og søjle indeholder etiketter

Heatmap for residualer

OK Annullér

p -værdien er meget lav, så nulhypotesen forkastes, og det konkluderes, at der er sammenhæng mellem flyrute og kundetilfredshed. Mønsteret er lidt kompliceret, men kunderne er mest tilfredse på Kbh-Kar og mest utilfredse på Kbh-Tir.

Godness-of-fit-test i kontingenstabel				
Observerede	Kbh-Tir	Kbl-Aal	Kbl-Kar	Sum
Utilfreds	88	64	29	181
Neutral	15	35	10	60
Tilfreds	37	35	67	139
Sum	140	134	106	380
Forventede	Kbh-Tir	Kbl-Aal	Kbl-Kar	Sum
Utilfreds	66,68	63,83	50,49	181
Neutral	22,11	21,16	16,74	60
Tilfreds	51,21	49,02	38,77	139
Sum	140	134	106	380
Residualer	Kbh-Tir	Kbl-Aal	Kbl-Kar	Sum
Utilfreds	2,61	0,02	-3,02	
Neutral	-1,51	3,01	-1,65	
Tilfreds	-1,99	-2,00	4,53	
Sum				58,51
Antal frihedsgrader	4			
Teststørrelse	58,51			
p -værdi	0,00%			

Forudsætning:
Alle forventede værdier ≥ 5 .

6. Stikprøvet teori

Stikprøvestørrelser for
simpelt stokastisk
udvalgte stikprøver

The screenshot shows the KeHaTools 3.0 interface with three main columns of options. The first column, highlighted with a red box, contains two options: '1: Stikprøvestørrelse for KI for andel' (with a p symbol) and '2: Stikprøvestørrelse for KI for middelværdi' (with a μ symbol). The second column, highlighted with a green box, contains three options: '4: Andel: Konfidensinterval' (with \hat{p}_{st}), '4: Andel: Allokering' (with \hat{p}_{st}), and '6: Andel: Stikprøvestørrelse' (with \hat{p}_{st}). The third column, highlighted with a blue box, contains three options: '7: Middelværdi: Konfidensinterval' (with \bar{x}_{st}), '8: Middelværdi: Allokering' (with \bar{x}_{st}), and '9: Middelværdi: Stikprøvestørrelse' (with \bar{x}_{st}). At the bottom, there are buttons for 'Options', 'Retur', and 'Annullér'. The text 'KeHaTools 3.0 DK version' is visible in the bottom left corner.

Beregninger for stratificerede
stikprøver for en andel

Beregninger for stratificerede
stikprøver for en middelværdi

6.1 Stikprøvestørrelse for KI for andel

Har man ingen idé om andelens størrelse, så behold 50%. Ellers skriv et estimat på andelen.

Stikprøvestørrelse for konfidensinterval for andel

Estimeret andel

Ønsket fejlmargen

Populationsstørrelsen kendes

Populationsstørrelse

Konfidensniveau

OK

Annullér

Stikprøvestørrelse for konfidensinterval for andel			
	x		
Estimeret andel	50,0%		
Ønsket fejlmargen	3,0%		
Konfidensniveau	95%		
Population	100000		
Stikprøvestørrelse	1056		

Der skal anvendes en stikprøvestørrelse på 1056 enheder

6.2 Stikprøvestørrelse for KI for middelværdi

Angiv et
estimat for
standard-
afvigelsen

Stikprøvestørrelse for konfidensinterval for middelværdi

Standardafvigelse

Ønsket fejlmargen

Populationsstørrelsen kendes

Populationsstørrelse

Konfidensniveau

OK

Annullér

Stikprøvestørrelse for konfidensinterval for middelvæ			
	x		
Standardafvigelse	40,00		
Ønsket fejlmargen	5,00		
Konfidensniveau	95%		
Population	0		
Stikprøvestørrelse	246		

Der skal anvendes en stikprøve-
størrelse på 246 enheder

Generelt for stratificerede stikprøver

Oftest anvender der en tabel, hvor hvert enkelt stratum svarer til en række, og hver søjle svarer til en bestemt variabel.

Den øverste række og subtotaler som 'I alt' bør ikke medtages.

Stratum	Population	Stikprøve	Gns	Standardafvigelse
A	200	20	820	564
B	600	60	304	209
C	1200	120	26	17
I alt	2000	200	–	–

Stratum B

Populations-
størrelserne
for de enkelte
strata.

Stikprøvernes standardafvigelser,
stratum for stratum

6.4 Stratificeret andel: Konfidensinterval

	Populatio	Stikprøve	Antal fejl
over 40 år	4000	80	5
25 - 40 år	5000	100	2
under 25 år	8000	160	8

Konfidensinterval for andel i stratificeret stikprøve

Etiketter for strata

Etiketter: 11.5!\$A\$2:\$A\$4

Populationsstørrelser, fordelt på strata: 11.5!\$B\$2:\$B\$4

Stikprøvestørrelser, fordelt på strata: 11.5!\$C\$2:\$C\$4

Antal gunstige, fordelt på strata: 11.5!\$D\$2:\$D\$4

Konfidensniveau: 95%

OK

Annullér

Kategori	Populatio	Stikprøve	Antal gunstige	Andel
over 40 år	4000	80	5	6,3%
25 - 40 år	5000	100	2	2,0%
under 25 år	8000	160	8	5,0%
Konfidensniveau	95%			
Strat. andel	4,41%			
Nedre grænse	2,26%			
Øvre grænse	6,57%			

Andelen estimeres til **4,41%**, og ligger med 95% sikkerhed mellem **2,26%** og **6,57%**

6.5 Stratificeret andel: Allokering

	Populatio	Andel
over 40 år	4000	6,25%
25 - 40 år	5000	2%
under 25 år	8000	5%

Allokering i stratificeret stikprøve for andel

Etiketter for strata

Etiketter: 11.5!\$A\$2:\$A\$4

Populationsstørrelser, fordelt på strata: '11.5!\$B\$2:\$B\$4

Andele, fordelt på strata: 11.5!\$C\$2:\$C\$4

Stikprøvestørrelse: 100

OK

Annullér

Allokering i stratificeret stikprøve for andel

Stratum	Populatio	Andel	Proportio	Optimal allokering
over 40 år	4000	0,0625	24	28
25 - 40 år	5000	0,02	29	21
under 25 år	8000	0,05	47	51
Total			100	100

Proportional allokering

Optimal allokering

6.6 Stratificeret andel: Stikprøvestørrelse

	Populatio	Andel
over 40 år	4000	6,25%
25 - 40 år	5000	2%
under 25 år	8000	5%

Stikprøvestørrelse for andel i stratificeret stikprøve

Etiketter for strata

Etiketter: 11.5' \$A\$2:\$A\$4

Populationsstørrelser, fordelt på strata: 11.5' \$B\$2:\$B\$4

Andele, fordelt på strata: 11.5' \$C\$2:\$C\$4

Konfidensniveau: 95%

Ønsket fejlmargen: 1%

OK

Annullér

Kategori	Populatio	Andel
over 40 år	4000	6%
25 - 40 år	5000	2%
under 25 år	8000	5%
Konfidensniveau	95%	
Fejlmargen	1%	
Proportional allokering	1471	
Optimal allokering	1414	

Proportional allokering:
1471 elementer er krævet

Optimal allokering:
1414 elementer er krævet

6.7 Stratificeret middelværdi: Konfidensinterval

Stratum	Population	Stikprøve	Gns	Standardafvigelse
A	200	20	820	564
B	600	60	304	209
C	1200	120	26	17
I alt	2000	200	-	-

Konfidensinterval for middelværdi i stratificeret stikprøve

Etiketter for strata

Etiketter: Ark17!\$A\$2:\$A\$4

Populationsstørrelser, fordelt på strata: Ark17!\$B\$2:\$E\$4

Stikprøvestørrelser, fordelt på strata: Ark17!\$C\$2:\$C\$4

Stikprøvegennemsnit, fordelt på strata: Ark17!\$D\$2:\$D\$4

Standardafvigelser, fordelt på strata: Ark17!\$E\$2:\$E\$4

Konfidensniveau: 95%

OK

Annullér

Kategori	Populatio	Stikprøve:	Gennemsi	Standardafvigelse
A	200	20	820	564
B	600	60	304	209
C	1200	120	26	17
Konfidens	95%			
Strat. genl	188,80			
Nedre græn	160,88			
Øvre græn	216,72			

Middelværdien estimeres til **188,80**.
Med 95% sikkerhed ligger denne middelværdi mellem **160,88** og **216,72**

6.8 Stratificeret middelværdi: Allokering

Kategori	Populatio	Standarda
A	200	564
B	600	209
C	1200	17

Allokering i stratificeret stikprøve for middelværdi

Etiketter for strata

Etiketter: 18!\$D\$17:\$D\$19

Populationsstørrelser, fordelt på strata: 18!\$E\$17:\$E\$19

Standardafvigelse, fordelt på strata: 18!\$F\$17:\$F\$19

Stikprøvestørrelse: 100

OK Annullér

Allokering i stratificeret stikprøve for middelværdi				
Stratum	Populatio	Standarda	Proportio	Optimal allokering
A	200	564	10	44
B	600	209	30	48
C	1200	17	60	8
Total			100	100

Proportional allokering

Optimal allokering

6.9 Stratificeret middelværdi: Stikprøvestørrelse

Kategori	Populatio	Standarda
A	200	564
B	600	209
C	1200	17

Stikprøvestørrelse for middelværdi i stratificeret stikprøve

Etiketter for strata

Etiketter: 18!\$D\$17:\$D\$19

Populationsstørrelser, fordelt på strata: 18!\$E\$17:\$E\$19

Standardafvigelser, fordelt på strata: 18!\$F\$17:\$F\$19

Konfidensniveau: 95%

Ønsket fejlmargin: 10

OK Annullér

Kategori	Populatio	Standardafvigelse
A	200	564
B	600	209
C	1200	17
Konfidens	95%	
Fejlmargi	10	
Proportio	929	
Optimal a	345	

Proportional allokering:
929 elementer er krævet

Optimal allokering:
345 elementer er krævet

7. Multivariate teknikker

The screenshot shows the KeHaTools 3.0 software interface. The window title is "KeHaTools 3.0". The interface contains several buttons for statistical analysis, each with an icon and a label:

- 1: Lineær regression**: Represented by a scatter plot with a regression line.
- 2: Forudsigelsesinterval**: Represented by a horizontal line with vertical end caps and the letter 'y' below it.
- 3: Test for hældning og skæring**: Represented by the equation $H_0 : b = b_0$.
- 4: Korrelationsanalyse**: Represented by a grid of squares.
- 5: Logistisk regression**: Represented by a sigmoid curve.
- 7: Variansanalyse (en-vejs)**: Represented by the text "ANOVA".
- 8: Variansanaluse (to-vejs)**: Represented by the text "ANOVA".

Annotations with arrows point from external text to these buttons:

- A blue arrow points from "Lineær regression" to button 1.
- A green arrow points from "Korrelationsanalyse" to button 4.
- A red arrow points from "Variansanalyse" to button 7.
- A purple arrow points from "Logistisk regression" to button 5.

At the bottom of the window, there are three buttons: "Options", "Retur", and "Annullér". The text "KeHaTools 3.0 DK version" is visible in the bottom left corner.

7.1 Lineær regression

x1	x2	x3	y
1	2	-1	-3
3	5	5	16
5	3	3	12
3	6	4	12
5	2	6	22
6	6	2	10

Lineær regression

x-værdier:

y-værdier:

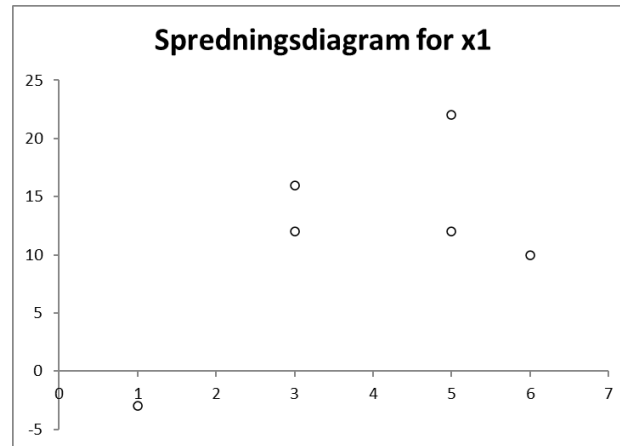
Etiketter i første række Forudsætninger

RESUMEOUTPUT						
Regressionsstatistik						
Multipel F	1,00					
R-kvadrer	99,78%					
Justeret R	99,44%					
Standardf	0,619					
Observati	6					
ANOVA						
	fg	SK	MK	F	Signifikans F	
Regressio	3	342,734	114,245	298,227	0,33%	
Residual	2	0,766	0,383			
I alt	5	343,500				
	Koefficient	Standardf	t-stat	p-værdi	Nedre 95%	Øvre 95%
Skæring	-0,566	0,79	-0,72	54,59%	-3,948	2,815
x1	1,058	0,17	6,18	2,52%	0,321	1,795
x2	-0,336	0,15	-2,24	15,49%	-0,983	0,311
x3	2,954	0,12	23,65	0,18%	2,417	3,492

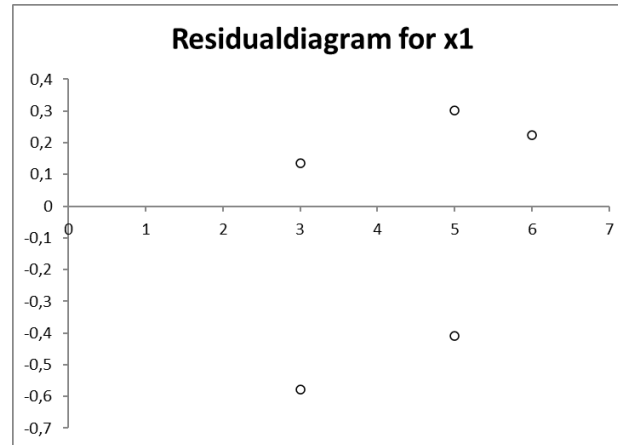
Det anbefales at medtage etiketterne, idet outputtet bliver meget mere læseligt!

7.1 Lineær regression

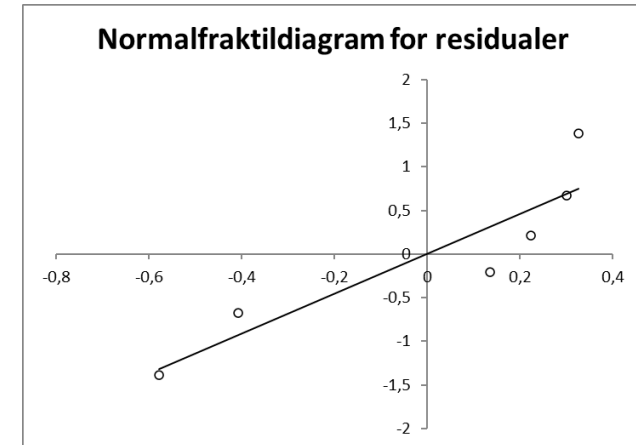
Vælges Forudsætninger i foregående beregning, så produceres der en lang række grafer.



Spredningsdiagram
ét for hver variabel



Residualdiagram
ét for hver variabel



Normalfraktildiagram for residualer
ét enkelt

7.2 Forudsigelsesinterval (simpel lineær regression)

RESUMEOUTPUT						
Regressionsstatistik						
Multipel F	0,56					
R-kvadrer	31,68%					
Justeret R	14,60%					
Standardf	3,340					
Observati	6					
ANOVA						
	fg	SK	MK	F	Signifikans F	
Regressio	1	20,700	20,700	1,855	24,49%	
Residual	4	44,634	11,158			
I alt	5	65,333				
	Koefficient	Standardf	t-stat	p-værdi	Nedre 95%	Øvre 95%
Skæring	2,416	3,41	0,71	51,73%	-7,041	11,872
x1	1,109	0,81	1,36	24,49%	-1,152	3,369

Lineær regression: Forudsigelsesinterval

Model:

Ny x-værdi:

Konfidensniveau:

Forudsigelse af enkelt y-værdi
 Forudsigelse af gennemsnitlig y-værdi

Denne model kan med fordel laves vha. beregningen i 7.1.

Forudsigelsesinterval i lineær regression		
	x	
Hældning	1,108911	
Skæring	2,42	
Konfidens	95%	
Forudsagt	7,96	
Nedre grænse	-2,40	
Øvre grænse	18,32	

7.3 Test for hældning og skæring

RESUMEOUTPUT					
Regressionsstatistik					
Multipel F	0,56				
R-kvadrer	31,68%				
Justeret R	14,60%				
Standardf	3,340				
Observati	6				
ANAVA					
	fg	SK	MK	F	Signifikans F
Regressio	1	20,700	20,700	1,855	24,49%
Residual	4	44,634	11,158		
I alt	5	65,333			
	Koefficient	Standardf	t-stat	p-værdi	Nedre 95% Øvre 95%
Skæring	2,416	3,41	0,71	51,73%	-7,041 11,872
x1	1,109	0,81	1,36	24,49%	-1,152 3,369

Lineær regression: Test for skæring eller hældning

Test for hældning Test for skæring

Model: Ark9!\$T\$1:\$Z\$18

Nul-værdi: 2

Venstresidet test Tosidet test Højresidet test

OK Annullér

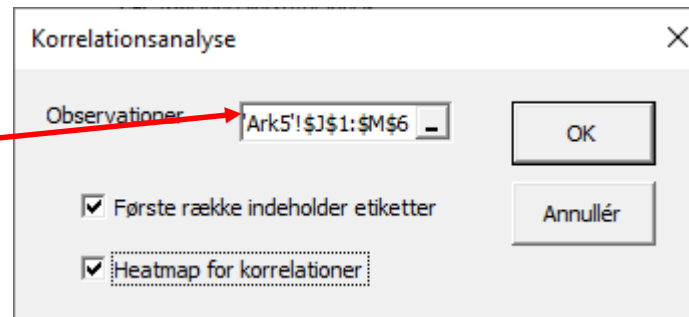
Test for hældning - tosidet test	
Stikprøvestørrelse	6
Skæring	1,11
Standardfejl	0,81
Nul-værdi	2
Teststørrelse	-1,09
p-værdi	33,52%

Denne model kan med fordel laves vha. beregningen i 7.1.

p-værdien er stor, så nulhypotesen kan ikke forkastes. Det kan altså godt antages, at den sande hældning er lig med 2.

7.4 Korrelationsanalyse

x1	x2	x3	x4	x5
1	2	1	-2	-1
2	5	4	-5	-4
3	3	6	-4	-6
4	1	5	-1	-5
5	6	3	-6	-3
6	7	4	-7	-4



Korrelationsanalyse					
	x1	x2	x3	x4	x5
x1	1				
x2	0,587	1			
x3	0,341	-0,049	1		
x4	-0,577	-0,985	-0,058	1	
x5	-0,341	0,049	-1,000	0,058	1

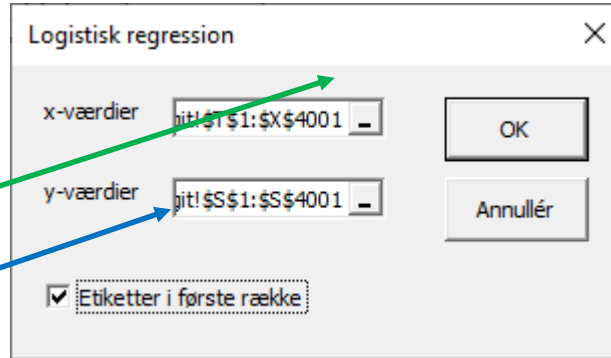
Det anbefales at medtage etiketterne, idet outputtet bliver meget mere læseligt!

Vælges muligheden *Heatmap*, så farves de enkelte celler i matricen efter størrelse:

over 0,5	blå
mellem 0,3 og 0,5	lyseblå
mellem -0,3 og 0,3	hvid
mellem -0,5 og -0,3	lyserød
under 0,5	rød

7.5 Logistisk regression

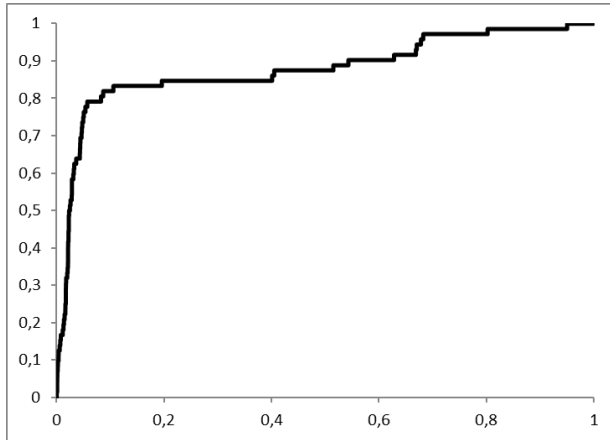
Default	WC/TA	RE/TA	EBIT/TA	ME/TL	S/TA
0	0,500799	0,306846	0,043373	0,956271	0,334774
0	0,54778	0,322214	0,051843	1,064545	0,334591
0	0,451001	0,22515	0,026813	0,804096	0,245585
0	0,306887	0,191936	0,030058	0,387011	0,253438
0	0,447246	0,217368	0,032458	0,791639	0,275531
0	0,457619	0,224064	0,031838	1,288137	0,318193
0	0,006034	-0,0284	0,006764	0,11027	0,254988
0	-0,11473	-0,11866	0,030892	0,149645	0,315722
0	0,061305	-0,1095	0,039947	0,41015	0,286025
0	0,050641	-0,0967	0,045355	0,253354	0,342056
0	0,121053	-0,11404	0,037298	0,457177	0,312122
...



RESUMEOUTPUT				
Regressionsstatistik (logistisk regression)				
Pseudo-R ²	22,2%			
InL	-280,526			
InL0	-360,6			
LR-test	160,148			
p-værdi	0,0%			
AUC (ROC)	87,9%			
Observati	4000			
Antal itera	12			
	Koefficient	Standardf	t-stat	p-værdi
Konstant	-2,543	0,266	-9,56	0,00%
WC/TA	0,414	0,572	0,72	46,92%
RE/TA	-1,454	0,229	-6,34	0,00%
EBIT/TA	-7,999	2,702	-2,96	0,31%
ME/TL	-1,594	0,323	-4,93	0,00%
S/TA	0,620	0,349	1,77	7,59%

Det anbefales at medtage etiketterne, idet outputtet bliver meget mere læseligt!

7.5 Logistisk regression - Diagnostik



Diagnostik									
Cutoff	TP	FN	FP	TN	TPR	FPR	PPV	NPV	ACC
0,00	72	0	3928	0	1,0000	1,0000	0,0180		0,0180
0,05	56	16	215	3713	0,7778	0,0547	0,2066	0,5000	0,9423
0,10	17	55	64	3864	0,2361	0,0163	0,2099	0,5000	0,9703
0,15	12	60	35	3893	0,1667	0,0089	0,2553	0,5000	0,9763
0,20	9	63	22	3906	0,1250	0,0056	0,2903	0,5000	0,9788
0,25	8	64	14	3914	0,1111	0,0036	0,3636	0,5000	0,9805
0,30	7	65	10	3918	0,0972	0,0025	0,4118	0,5000	0,9813
0,35	6	66	9	3919	0,0833	0,0023	0,4000	0,5000	0,9813
0,40	6	66	7	3921	0,0833	0,0018	0,4615	0,5000	0,9818
0,45	5	67	5	3923	0,0694	0,0013	0,5000	0,5000	0,9820

ROC-kurve. Bruges til at beregne AUC

For hvert muligt cutoff beregnes:

Confusion matrix		Faktisk værdi	
		Sand	Falsk
Forudsagt værdi	Sand	TP	FP
	Falsk	FN	TN

$$TPR = TP / (TP + FN)$$

$$FPR = FP / (FP + TN)$$

$$PPV = TP / (TP + FP)$$

$$NPV = TN / (FN + TN)$$

$$ACC = (TP + TN) / (TP + FN + FP + TN)$$

7.7 ANOVA (en-vejs)

DynamoU	MicroOma	MegaClea	MultiVask
259	277	264	283
275	267	258	270
276	271	257	281
278	267	264	261
283	279	269	268
262	272	257	269
272	287	265	270
275	269	264	279
269	262	257	277
274	261	268	278

Variansanalyse (ANOVA) - envejs

Observationer: 'Ark21'!\$S\$27

Konfidensniveau: 95%

Første række indeholder etiketter

OK

Annullér

Det anbefales at medtage etiketterne, idet outputtet bliver meget mere læseligt!

p -værdien er lav, så vi forkaster nulhypotesen om, at alle middelværdier er ens. Se diffogrammet næste slide for yderligere fortolkning.

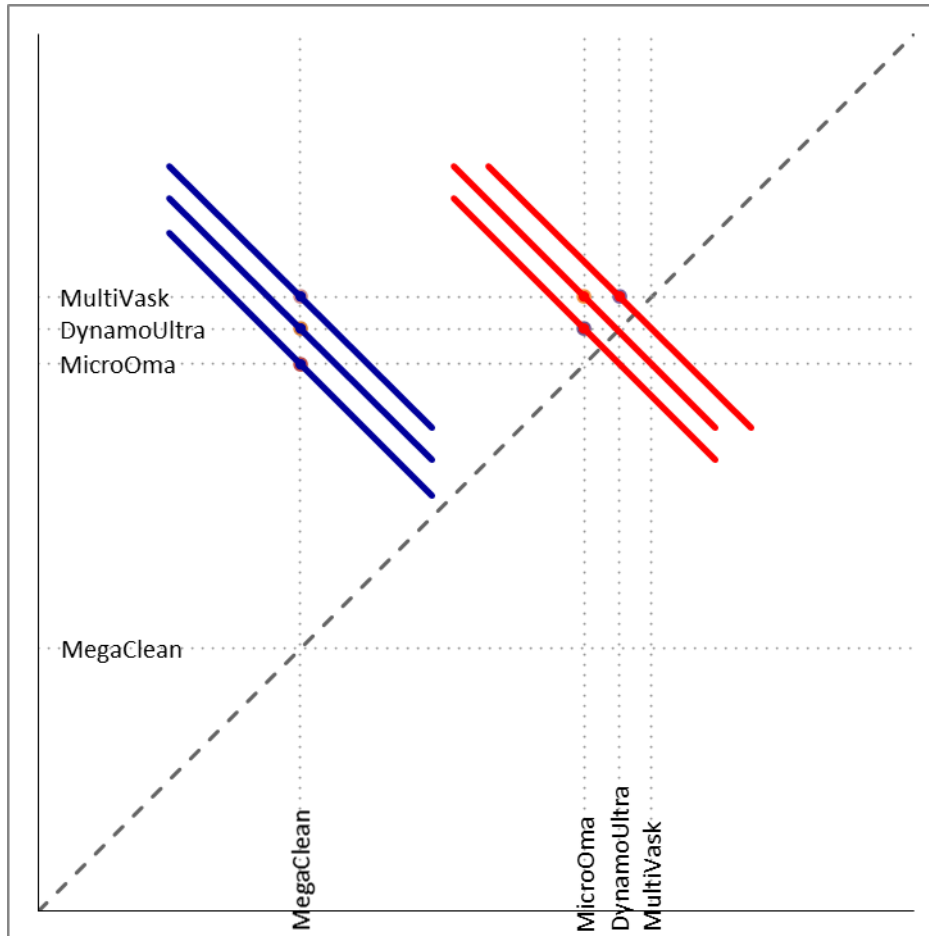
Grupper	Antal	Sum	Gennemsnit	Varians
DynamoU	10	2723,00	272,30	52,46
MicroOma	10	2712,00	271,20	63,73
MegaClea	10	2623,00	262,30	21,79
MultiVask	10	2733,00	273,30	46,90

VariationsSK	fg	MK	F	p -værdi
Mellem gr	767,08	3	255,69	5,53
Indenfor gr	1663,900	36	46,219	
I alt	2430,975	39		

Test for varianshomogenitet	
	Teststørrelse p -værdi
Bartlett	2,49 47,70%

p -værdien for Bartlett's test er høj, så vi kan ikke forkaste nulhypotesen om varianshomogenitet.

7.7 ANOVA (en-vejs) - diffogrammet



Diffogrammet viser, om der er signifikante forskelle mellem de enkelte middelværdier.

Hvis de skrå linjer krydser den stiplede linje (og er røde), så er der ikke nogen signifikant forskel.

Hvis den skrå linje ikke krydser den stiplede linje (og er blå), så er der en signifikant forskel.

Der er ingen signifikante forskelle mellem middelværdierne for MicroOma, DynamoUltra og MultiVask, hvorimod middelværdien for MegaClean er signifikant forskellig fra de andre middelværdier.

7.8 ANOVA (to-vejs)

	XXX	YYY	ZZZ
AAA	7,92	14,53	18,15
	3,35	14,85	20,21
	5,95	16,72	15,81
BBB	18,69	34,82	38,42
	25,22	30,96	36,89
	25,56	26,20	39,11
CCC	10,51	26,76	38,83
	17,13	25,72	34,77
	13,53	28,37	37,05
DDD	20,18	30,14	44,16
	24,97	29,66	41,44
	22,86	29,52	41,64

Variansanalyse (ANOVA) - tovejs

Observationer:

Antal rækker pr. stikprøve:

OK

Annullér

Anava: To-faktor med gentagelse				
RESUME	XXX	YYY	ZZZ	I alt
AAA				
Antal	3	3	3	9
Sum	17,22122	46,10771	54,1645	117,4934
Gennemsnit	5,740407	15,36924	18,05483	13,05482
Varians	5,254655	1,399938	4,854267	34,32369
BBB				
Antal	3	3	3	9
Sum	69,46671	91,97873	114,4155	275,861
Gennemsnit	23,15557	30,65958	38,13852	30,65122
Varians	14,96823	18,6482	1,286079	50,81728
CCC				
Antal	3	3	3	9
Sum	41,17226	80,84553	110,6503	232,6681
Gennemsnit	13,72409	26,94851	36,88343	25,85201
Varians	10,99424	1,780688	4,127419	105,4685
DDD				
Antal	3	3	3	9
Sum	68,01021	89,32182	127,2354	284,5674
Gennemsnit	22,67007	29,77394	42,4118	31,6186
Varians	5,777029	0,105926	2,299007	77,03504
I alt				
Antal	12	12	12	36
Sum	195,8704	308,2538	406,4657	910,5899
Gennemsnit	16,32253	25,68782	33,87215	25,29416
Varians	62,83029	44,75467	97,84384	117,4411

ANAVA						
VariationsSK	fg	MK	F	p-værdi	F krit	
Stikprøve	1969,283	3	656,4277	110,1763	0,00%	3,008787
Kolonner	1850,723	2	925,3613	155,3148	0,00%	3,402826
Interaktio	147,4424	6	24,57373	4,124513	0,55%	2,508189
Indenfor	142,9914	24	5,957973			
I alt	4110,439	35				

7.8 ANOVA (to-vejs) - interaktionsgrafer

	XXX	YYY	ZZZ
AAA	7,92	14,53	18,15
	3,35	14,85	20,21
	5,95	16,72	15,81
BBB	18,69	34,82	38,42
	25,22	30,96	36,89
	25,56	26,20	39,11
CCC	10,51	26,76	38,83
	17,13	25,72	34,77
	13,53	28,37	37,05
DDD	20,18	30,14	44,16
	24,97	29,66	41,44
	22,86	29,52	41,64

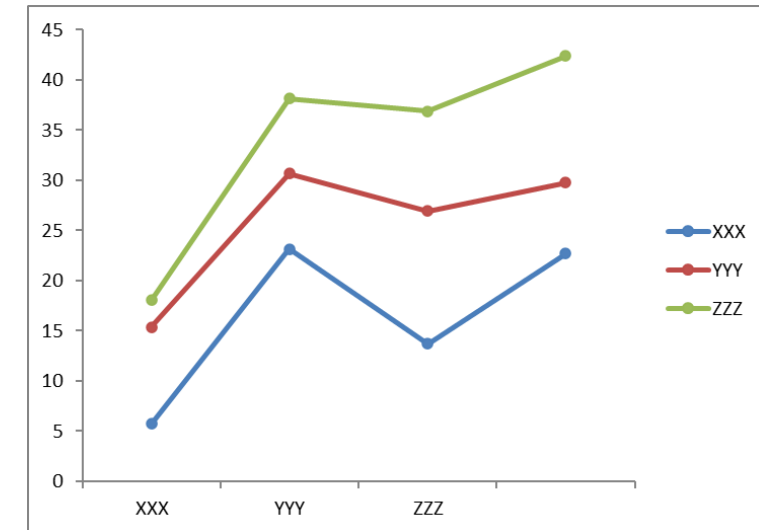
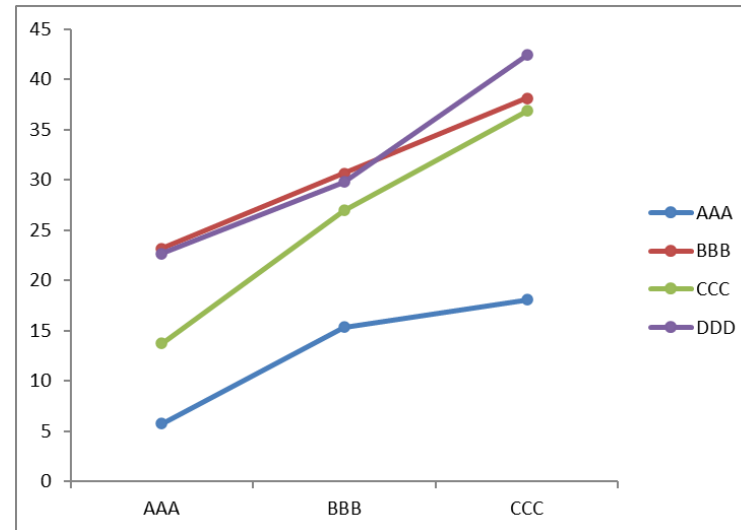
Variansanalyse (ANOVA) - tovejs

Observationer: rk2!\$N\$1:\$Q\$13

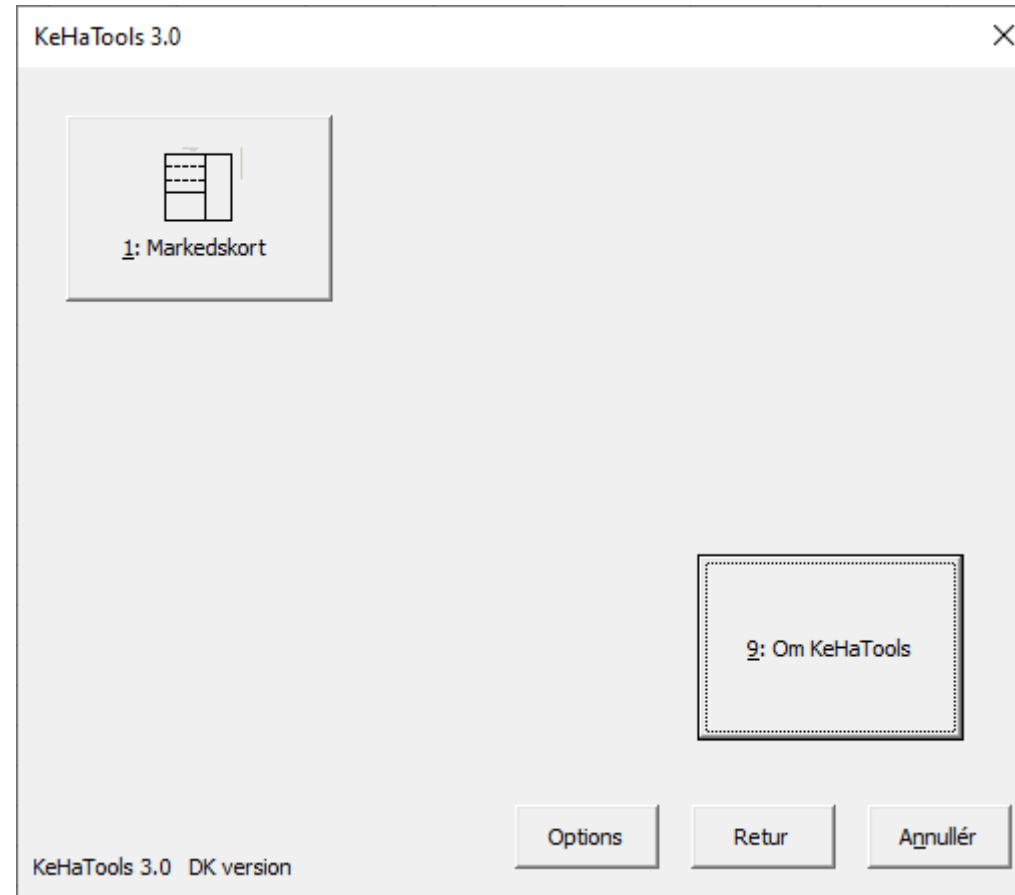
Antal rækker pr. stikprøve: 3

OK

Annullér



9. Diverse



9.1 Markedskort

Anvendes indenfor markedsanalyse

Markedskort

Indtast antal personer, eller andel, som:

Ikke kender

Kender, men ikke har prøvet

Kender, men bruger ikke

Kender, og bruger indimellem

Kender, og fortrinsvist

OK

Annullér

	Antal	Andel
Kender ikke	40	40,00%
Har ikke prøvet	20	20,00%
Forkaster	20	20,00%
Indifferent	15	15,00%
Foretrækker	5	5,00%
Kendskabsgrad		0,6
Prøveandel		0,666667
Forkastningsgrad		0,5
Indifferensgrad		0,375
Præferencegrad		0,125
Absolut forventning		0,05

Foretrækker: 5%	Kender ikke: 40%
Indifferent: 15%	
Forkaster: 20%	Har ikke prøvet: 20%

9.9 Om KeHaTools

