

Indblik i Excel - Statistiske problemstillinger

Niels Petersen



Niels Petersen

Indblik i Excel

– Statistiske Problemstillinger

BusinessSumup

Indblik i Excel - Statistiske Problemstillinger
© 2008 Niels Petersen & BusinessSumup
ISBN 978-87-7681-314-7

Indholdsfortegnelse

	Indledning	6
1.	Excels terminologi	7
2.	Genvejstaster	8
3.	Funktioner	10
4.	Vejledninger	16
4.1	Problemløser	16
4.2	Envejstabel	18
4.3	Tovejstabel	19
5.	Statistiske problemstillinger	22
5.1	Sandsynlighedsregning	22
5.1.1	Binomialfordelingen	22
5.1.2	Poissonfordelingen	22
5.1.3	Normalfordelingen	22
5.2	Deskriptiv statistik	23
5.2.1	Histogram	23
5.2.2	Normalfraktildiagram	25
5.2.3	Boxplot - Kassediagram	27



 **MTHøjgaard**

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



5.2.4	Beskrivende statistik	29
5.3	Regressionsanalyse	32
5.3.1	Model	34
5.3.2	Forudsætninger	34
5.3.3	Hypotese vedr. alle forklarende variable	39
5.3.4	Teststørrelse vedr. alle forklarende variable	39
5.3.5	Afgørelse vedr. alle forklarende variable	39
5.3.6	Reducering af modellen - Backward	40
5.3.7	Hypotese vedr. enkelt forklarende variabel	40
5.3.8	Teststørrelse vedr. enkelt forklarende variabel	40
5.3.9	Afgørelse vedr. enkelt forklarende variabel	40
5.4	Enkeltsidet variansanalyse	41
5.4.1	Model	42
5.4.2	Forudsætninger	42
5.4.3	Konstant varians grafisk	46
5.4.4	Hypotese vedr. konstant varians	46
5.4.5	Teststørrelse vedr. konstant varians	46
5.4.6	Autofilter	46
5.4.7	Pivottabel	48
5.4.8	Afgørelse vedr. konstant varians	51
5.4.9	Hypotese vedr. ens middelværdier (alle grupper)	51
5.4.10	Teststørrelse vedr. ens middelværdier (alle grupper)	52
5.4.11	Afgørelse vedr. ens middelværdier (alle grupper)	52
5.4.12	Hypotese vedr. to ens middelværdier	52
5.4.13	Teststørrelse vedr. to ens middelværdier	52
5.4.14	Afgørelse vedr. to ens middelværdier	52



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.

banedanmark



Indledning

Bogen er skrevet som en hjælp til folk, der i vid udstrækning bruger Excel til at løse henholdsvis økonomiske og statistiske problemstillinger.

Indledningsvis indeholder bogen en række oversigter over genvejstaster, funktioner og andet, der har generel interesse. Disse er meget generelle for Excel og kan også anvendes ud over de to fagområder. Herefter følger en gennemgang af løsningsvejledninger til en række økonomiske problemstillinger, hvor fremgangsmåder er beskrevet i detaljer, og endelig tilsvarende vejledninger til en række statistiske-excel problemstillinger.

Bogen kan enten anvendes til læsning for at skabe indsigt i Excels anvendelsesmuligheder inden for fagområderne økonomi og indledende statistik eller som opslagsværk i forbindelse med løsning af opgaver inden for et af de to fagområder.

Kompendiet tager udgangspunkt i "Erhvervsøkonomiske modeller med Excel" af Jørgen Meyer, derudover er der, til kapitlet om statistiske problemstillinger, hentet inspiration i "Teoretisk statistik – En erhvervsøkonomisk tilgang" 3. Udgave af Jens Overø & Gorm Gabrielsen.

1. Excels terminologi

Omkring Excel anvendes en masse faste termer. Flere af disse er umiddelbart indlysende, men i dette afsnit findes en kort forklaring af de enkelte begreber.

Begreb	Forklaring
<i>Mappe/Arbejdsbog</i>	En Excel-fil som kan bestå af et til flere ark.
<i>Ark</i>	En 2-dimensionel matrix af felter.
<i>Celle</i>	Et af alle felterne i arket. En celle defineres af dens adresse
<i>Adresse</i>	Et entydigt navn, der knytter sig til én specifik celle i et ark. Navnet er bestemt af cellens placering i kolonne og række
<i>Reference/ henvisning</i>	Et peg på eller en henvisning til en celle ud fra dens adresse.
<i>Område</i>	En nærmere defineret række celler samlet i et rektangel. Området defineres af cellen i øverste venstre hjørne og nederste højre, og de to celleadresser adskilles af et kolon.
<i>Tal</i>	Er et tal.
<i>Tekst</i>	Er en tekststreng. Tal eller formler kan fremstå som tekststreng, hvis tegnet ' er tilføjet før tal eller formel.
<i>Formel</i>	Indledes med tegnet = og er en udregning eller en funktion, der typisk returnerer et tal. Nogle funktioner kan dog også returnere tekststreng eller andet.
<i>Funktion</i>	Excel rummer flere hundrede funktioner. Det er dog kun et fåtal, der er interessante i forhold til at lave økonomiske kalkuler. Disse er gennemgået i afsnittet funktioner. En funktion modtager et antal parametre og returnerer så typisk et tal eller tekst.
<i>Parameter</i>	Et nærmere defineret input med en række begrænsninger. Det kan f.eks. skulle være et tal.

2. Genvejstaster

I Excel anvendes tastatur og mus side om side i noget større grad end i f.eks. Word. Arbejdet i Excel kan dog lettes en del ved at bruge tastaturet mere end det umiddelbart virker logisk. Det kræver dog, at en række genvejstaster indlæres. Hvis disse først bliver en fast del af brugen af Excel, vil det resultere i en lettere og dermed langt hurtigere arbejdsgang. De mest anvendelige genvejstaster er gennemgået herunder.

Tast(er)	Funktion
Enter	Afslut redigering af celle og gem.
Esc	Afslut redigering af celle uden at gemme.
Ctrl+(pile)	Flytter til første/sidste celle i række/kolonne med indhold ellers enden af arket.
PgDn/PgUp	Flytter et skærbillede ned/op.
Alt+PgDn/PgUp	Flytter et skærbillede til højre/venstre.
Ctrl+Home	Flytter markøren til øverste venstre hjørne (Celle A1).
Ctrl+End	Flytter markøren til nederste højre hjørne af området med udfyldte celler.
Ctrl+PgDn/Up	Skifter mellem arkene i arbejdsbogen.



CISO Conference
Produced by **Inspired**

**Apollo Hotel 1, Groenlandsekade
Vinkeveen, Amsterdam, NL
Dec 5th 2019**

**Listen, learn & build relationships with our
Network of CISOs & Cyber Security Leaders**

Inspired

F2/(dobbeltklik)	Aktiver allerede udfyldt celle til redigering.
Ctrl+Z	Fortryd sidste handling (f.eks. redigering af celle).
Shift+(pile)	Marker område.
Ctrl+Shift+(pil)	Marker til enden af tekstområde.
Shift+Space	Markerer rækken hvori markøren står.
Ctrl+Space	Marker kolonne hvori markøren står.
Ctrl+C	Kopier.
Ctrl+X	Klip.
Ctrl+V	Sæt ind (Her kan også trykkes Enter – i så fald kan kun kopieres én gang).
Ctrl++	Indsætter et område svarende til det markerede (f.eks. en række).
Ctrl+-	Sletter det markerede område.
Ctrl+F	Formaterer området med fed skrift.
Ctrl+K	Formaterer området med kursiv skrift.
Ctrl+U	Formaterer området med understreget skrift.
F4	Låser cellehenviisning med \$-tegn. Der kan trykkes op til 3 gange for at låse henholdsvis række, kolonne eller begge dele. Anvendes for efterfølgende at kunne kopiere celler med formler som ønsket.
Σ-knap	Genvej til SUM (+ MIN, MAKS, TÆL og MIDDEL). Foreslår selv område.
f_x-knap	Genvej til at finde alle funktioner, og hvor der samtidig er vejledning til hvilke parametre, der skal benyttes i funktionen.

3. Funktioner

Excel rummer over 200 forskellige funktioner. Det vil føre alt for vidt at gennemgå alle disse her, og i stedet er kun de få funktioner, der har relevans i forhold til at løse økonomiske og statistiske problemstillinger gennemgået. Det er nemlig et fåtal, der er nødvendige i dette henseende. For de enkelte funktioner er forklaret navn på både dansk og engelsk, hvilke parametre, der skal anvendes, hvilken kategori funktionen er placeret i samt en beskrivelse af funktionaliteten. Funktionerne er opstillet i alfabetisk rækkefølge på baggrund af de danske titler.

Funktion	BINOMIALFORDELING(<i>Tal;Forsøg;Sandsynlighed;Kumulativ</i>)
Engelsk	BINOMDIST
Parametre	<i>Tal</i> : Antallet af "succeser" <i>Forsøg</i> : Antallet af forsøg <i>n</i> <i>Sandsynlighed</i> : SSH for "succes" (<i>p</i>) <i>Kumulativ</i> : Er en logisk værdi, vælges SAND returneres den kumulerede SSH. Vælges FALSK returneres punkt SSH.
Placering	<i>Statistisk</i>
Beskrivelse	Returnerer enten punkt SSH eller den kumulerede SSH i en given binomialfordeling

Funktion	ELLER(<i>betingelse1;betingelse2;...</i>)
Engelsk	OR
Parametre	<i>betingelse1</i> & 2: En betingelse, der typisk er et udtryk med en sammenligningsoperator (=, <, >=, ...).
Placering	<i>Logisk</i>
Beskrivelse	Returnerer sandt hvis en eller flere betingelser er sande – ellers falsk. Benyttes typisk "inde" i en HVIS-formel.

Funktion	FFORDELING(<i>X;Frihedsgrader1;Frihedsgrader2</i>)
Engelsk	FDIST
Parametre	<i>X</i> : Er den værdi vi ønsker funktionen evalueret for <i>Frihedsgrader1</i> : Er antallet af frihedsgrader til tælleren <i>Frihedsgrader2</i> : Er antallet af frihedsgrader til nævneren
Placering	<i>Statistik</i>
Beskrivelse	Returnerer sandsynligheden for at <i>F</i> er større end eller lig med den angivne værdi Dvs. $P(F \geq x)$

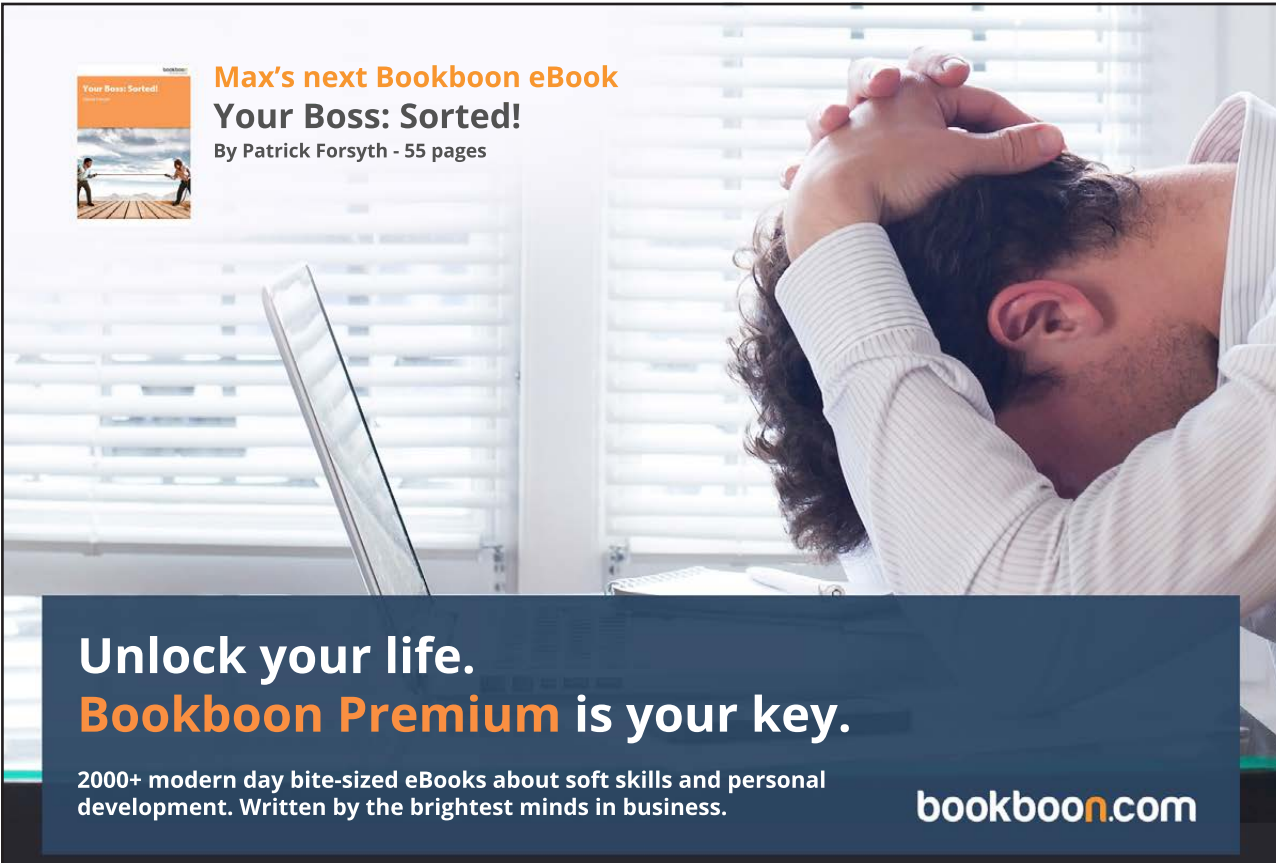
Funktion	HVIS(<i>betingelse</i> ; <i>formel_sand</i> ; <i>formel_falsk</i>)
Engelsk	IF
Parametre	<i>betingelse</i> : Er typisk et udtryk med en sammenligningsoperator (=, <, >=, ...). <i>formel_sand</i> : En vilkårlig formel der resulterer i en værdi eller tekst. Det kan også bare være et tal eller en tekst. Formlen returneres hvis betingelsen er sand. <i>formel_falsk</i> : En vilkårlig formel der resulterer i en værdi eller tekst. Det kan også bare være et tal eller en tekst. Formlen returneres hvis betingelsen er falsk.
Placering	<i>Logisk</i>
Beskrivelse	Returnerer sandt eller falsk ud fra logisk udsagn.


Funktion	LOPSLAG(<i>værdi</i> ; <i>tabel</i> ; <i>kolonneindeks</i> ; <i>sand/falsk</i>)
Engelsk	VLOOKUP
Parametre	<i>værdi</i> : Den værdi, der skal slås op i første kolonne. Kan være både tal eller tekst. <i>tabel</i> : Tabelområdet hvor opslaget foretages (Eks. B2:E10). <i>kolonneindeks</i> : Den kolonne hvorfra resultatet skal returneres. <i>sand/falsk</i> : Ved falsk returneres #I/T, hvis værdien ikke findes i tabellen. Ved sand forudsættes at tabellens første kolonne er sorteret stigende i værdi, og så returneres fra den række, hvor værdien i første kolonne er tættest på men mindre end den angivne.
Placering	<i>Opslag og reference</i>
Beskrivelse	Lodret opslag. Slår en given værdi op i første kolonne i et tabelområde, og returnerer værdien i samme række i det angivne kolonneindeks.

Funktion	MAKS(<i>tal1</i> ; [<i>tal2</i>]; ...)
Engelsk	MAX
Parametre	<i>tal1</i> & 2: Kan være både et tal, en formel, en celle eller et område.
Placering	<i>Statistisk</i>
Beskrivelse	Finder maksimumværdien af en række tal.

Funktion	MIN(<i>tal1</i> ;[<i>tal2</i>];...)
Engelsk	MIN
Parametre	<i>tal1</i> & 2: Kan være både et tal, en formel, en celle eller et område.
Placering	<i>Statistisk</i>
Beskrivelse	Finder minimumværdien af en række tal.

Funktion	NORMFORDELING(<i>X</i> ;Middelværdi;Standardafv;Kumulativ)
Engelsk	NORMDIST
Parametre	<i>X</i> : Er den værdi funktionen ønskes evalueret for <i>Middelværdi</i> : Er middelværdien (gns.) for vores normalfordeling <i>Standardafv</i> : Er standardafvigelsen for vores normalfordeling <i>Kumulativ</i> : Er en logisk værdi, vælges SAND returneres den kumulerede SSH. Vælges FALSK returneres punkt SSH.
Placering	<i>Statistisk</i>
Beskrivelse	Returnerer enten punkt SSH eller den kumulerede SSH i en given Normalfordeling





Max's next Bookboon eBook
Your Boss: Sorted!
 By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Funktion	OG(betingelse1;betingelse2;...)
Engelsk	AND
Parametre	<i>betingelse1</i> & 2: En betingelse, der typisk er et udtryk med en sammenligningsoperator (=, <, >=,...).
Placering	<i>Logisk</i>
Beskrivelse	Returnerer sandt hvis alle betingelser er sande – ellers falsk. Benyttes typisk "inde" i en HVIS-formel.

Funktion	POISSON(X;Middelværdi;Kumulativ)
Engelsk	POISSON
Parametre	<i>X</i> : Antal hændelser <i>Middelværdi</i> : Forventede antal hændelser <i>Kumulativ</i> : Er en logisk værdi, vælges SAND returneres den kumulerede SSH. Vælges FALSK returneres punkt SSH.
Placering	<i>Statistisk</i>
Beskrivelse	Returnerer enten punkt SSH eller den kumulerede SSH i en given Poissonfordeling

Funktion	RUND.NED(<i>tal</i> ;Antal cifre)
Engelsk	ROUNDDOWN
Parametre	<i>Tal</i> : Er et vilkårligt tal, som skal rundes ned <i>Antal cifre</i> : Er antallet af decimaler tallet skal rundes ned til. Vælges 0 angives et helt tal
Placering	<i>Mat og Triq</i>
Beskrivelse	Runder det angivne tal ned til et bestemt antal cifre

Funktion	STANDARDNORMFORDELING(Z)
Engelsk	NORMSDIST
Parametre	<i>Z</i> : Er den værdi vi ønsker evalueret
Placering	<i>Statistik</i>
Beskrivelse	Returnerer den kumulerede sandsynlighed for standardnormalfordelingen – altså den normalfordeling med middelværdi 0 og varians 1

Funktion	STANDARDNORMINV(<i>Sandsynlighed</i>)
Engelsk	NORMSINV
Parametre	<i>Sandsynlighed</i> : Er den SSH der knytter sig til standardnormalfordelingen
Placering	<i>Statistik</i>
Beskrivelse	Returnerer den Z-fraktil der knytter sig til sandsynligheden i standardnormalfordelingen – altså den normalfordeling med middelværdi 0 og varians 1

Funktion	SUM(<i>tal1</i> ;[<i>tal2</i>];...)
Engelsk	SUM
Parametre	<i>tal1</i> & 2: Kan være både et tal, en formel, en celle eller et område.
Placering	<i>Mat og trig</i>
Beskrivelse	Finder summen af en række tal.



 MTHøjgaard

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



Funktion	TFORDELING(<i>X;Frihedsgrader;Haler</i>)
Engelsk	TDIST
Parametre	<i>X</i> : Er den numeriske værdi vi ønsker funktionen evalueret for <i>Frihedsgrader</i> : Er antallet af frihedsgrader <i>Haler</i> : Angiver om sandsynligheden skal findes enkelt eller dobbeltsiddet
Placering	<i>Statistik</i>
Beskrivelse	Funktionen er lidt drilsk i Excel sammenlignet med andre statistikprogrammer som SPSS og SAS. Derfor opfordres til at læse Excels hjælp til denne funktion

Funktion	TÆL(<i>tal1</i> ;[<i>tal2</i>];...)
Engelsk	COUNT
Parametre	<i>tal1</i> & 2: Kan være både et tal, en formel, en celle eller et område.
Placering	<i>Statistik</i>
Beskrivelse	Tæller antallet af tal. Her defineres typisk et område.

Funktion	VOPSLAG(<i>værdi</i> ; <i>tabel</i> ;rækkeindeks; <i>sand/falsk</i>)
Engelsk	HLOOKUP
Parametre	<i>værdi</i> : Den værdi, der skal slås op i første række. Kan være både tal eller tekst. <i>tabel</i> : Tabelområdet hvor opslaget foretages (Eks. B2:E10). <i>rækkeindeks</i> : Den række hvorfra resultatet skal returneres. <i>sand/falsk</i> : Ved falsk returneres #I/T, hvis værdien ikke findes i tabellen. Ved sand forudsættes at tabellens første række er sorteret stigende i værdi, og så returneres fra den kolonne, hvor værdien i første række er tættest på men mindre end den angivne.
Placering	<i>Opslag og reference</i>
Beskrivelse	Vandret opslag. Slår en given værdi op i første række i et tabelområde, og returnerer værdien i samme kolonne i det angivne rækkeindeks.

4. Vejledninger

Dette afsnit indeholder en række vejledninger, der har almen interesse, og kan bruges i forbindelse med forskellige typer opgaver. Der vil i gennemgangen af de forskellige opgavetyper være henvisninger til netop disse vejledninger, såfremt de kan anvendes med fordel.

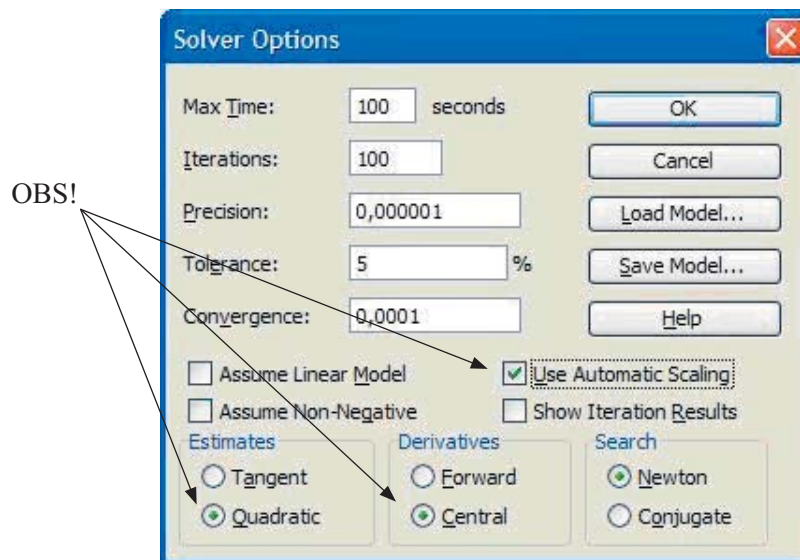
4.1 Problemløser

Problemløseren er en yderst anvendelig funktionalitet i Excel, der kan bruges til at optimere enkelte celler mod minimum, maksimum eller en bestemt værdi ved at ændre på andre angivne celler. For overhovedet at kunne anvende problemløseren er det imidlertid nødvendigt både at installere og indstille den korrekt. Hvis den er installeret ligger den som et menupunkt i menuen *Funktioner* og hedder enten *Problemløser* eller *Solver*. Hvis den ikke er installeret gøres det således.

1. Gå i menuen *Funktioner*
2. Vælg *Tilføjelsesprogrammer (Add-ins)*
3. Sæt et hak ud for *Tilføjelsesprogrammet Problemløser (Solver Add-in)*
4. Tryk *OK*

Hvis der arbejdes i en engelsk udgave af Excel, vil problemløseren og andre tilføjelsesprogrammer stadig være på engelsk, selvom Excel er indstillet til at køre på dansk. Derfor er de engelske navne på knapper og menupunkter skrevet med i vejledningen i parentes. Derudover skal problemløseren også indstilles til at optimere via den rigtige metode. Det gøres således.

1. Gå i menuen *Funktioner*
2. Vælg *Problemløser (Solver)*
3. Tryk på knappen *Indstillinger (Options)*
4. Sørg for at indstillingerne er som på skærmdumpet herunder
5. Tryk *OK*
6. Tryk *Luk (Close)*



Problemløseren er nu installeret, sat op og klar til brug en gang for alle. Den kan nu anvendes som beskrevet herunder igen og igen.

1. Gå i menuen *Funktioner*
2. Vælg *Problemløser (Solver)*
3. Indtast den celle, der skal optimeres i feltet *Angiv målcelle (Set Target Cell)*. Dette gøres lettest ved at sørge for, at markøren står i feltet og derefter trykke på den ønskede celle med musen. De følgende indtastninger af celler kan alle gøres på samme måde.
4. Vælg hvorvidt cellen skal optimeres mod maksimum, minimum eller en bestemt værdi, der så indtastes.
5. Indtast den eller de celler, der må justeres for at optimere, i feltet *Ved redigering af cellerne (By Changing Cells)*.

Nu er der i princippet klar til at optimere. Nogen gange er der dog visse begrænsninger for ændring i cellerne, der skal indlægges, og det gøres som følger. Hvis der ingen begrænsninger er, kan der springes direkte til punkt 12.

6. Tryk *Tilføj*
7. Indtast den celle, for hvilken der er begrænsninger i feltet *Celle reference (Cell reference)*
8. Vælg hvilken type begrænsning, der skal indlægges. De tre relevante, der kan vælges i mellem er \leq , $=$ og \geq .
9. Indtast begrænsning i feltet *Begrænsning (Constraint)*. Dette kan være et tal, men er typisk en henvisning til en celle indeholdende netop dette tal.
10. Punkt 6 til 9 gentages indtil alle begrænsninger er indlagt
11. Tryk *OK*



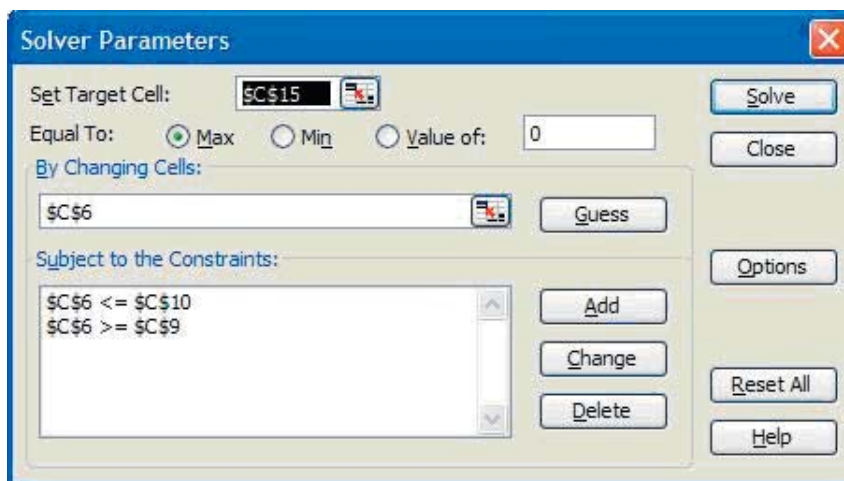
Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.

banedanmark





12. Tryk *Løs*
13. Tryk *OK*

Inden der trykkes *OK* til sidst, kan der vælges at beholde de originale værdier af cellerne i stedet for de nye fundne, såfremt det skulle ønskes.

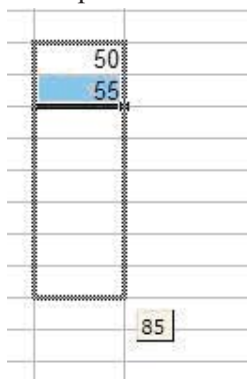
4.2 Envejstabel

Envejstabeller kan anvendes til at lave en række simuleringer. De kan anvendes til at illustrere, hvorledes ændringer på én variable påvirker en eller flere andre variable. Med andre ord er der én uafhængig variabel og en eller flere afhængige variable. De kan typisk anvendes til at illustrere, hvordan forskellige prissætninger påvirker efterspørgsel, dækningsbidrag indtjeningsbidrag eller lignende. En envejstabel kan typisk se ud, som herunder.

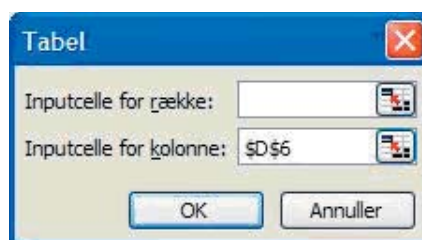
Pris	Total PB
	379,4028125
50	50,1
55	111,6625
60	166,675
65	215,01875
70	257,175
75	293,08125
80	322,7375
85	346,14375
90	363,3
95	374,3625
100	379,175
105	377,1875
110	368,375
115	353,34375
120	332,125
125	304,78125
130	271,25
135	231,46875
140	185,5
145	133,40625
150	75,125

I det illustrerede eksempel udregnes det totale produktbidrag på baggrund af en række priser. Forudsætningen for at lave en sådan tabel er, at arbejdsbogen er opbygget, så produktbidraget udregnes på baggrund af en række inddata – og heriblandt selvfølgelig prisen. Hvis disse forudsætninger er opfyldt oprettes tabellen således.

1. Skriv overskrifter øverst i kolonner
2. Efterlad et blankt felt i første kolonne under overskriften
3. Udfyld resten af første kolonne med de ønskede tal. Hvis de 2 første tal skrives med rette interval imellem og derefter markeres, kan man ved at trække i den lille firkant nede i højre hjørne af markeringen udfylde hele kolonnen med tal med samme interval imellem, som vist på skærmdumpet herunder.



4. Udfyld øverste celle i anden kolonne med en reference til det ønskede resultat fra økonomiberegningerne (I dette tilfælde kolonnen *I alt* for *Produktbidrag*)
5. Hvis flere økonomiske resultater ønskes udregnet i tabellen gøres tilsvarende for tredje, fjerde osv. kolonne.
6. Marker hele tabelområdet – dog ikke overskrifterne.
7. Gå i menuen *Data*.
8. Vælg *Tabel*
9. Indlæg reference (klik på cellen med musen) til *Inddata* for *Pris* i feltet *Inputcelle for kolonne* (her kan selvfølgelig simuleres for andre variable end pris, men det er klart den hyppigst anvendte).



10. Tryk *OK*

Så er tabellen oprettet og i vores eksempel er det totale produktbidrag udregnet for hver af de angivne prissætninger i første kolonne.

4.3 Tovejstabel

Tovejstabeller minder i princippet meget om envejstabeller, og måden, hvorpå de oprettes, er også meget ens. Selvfølgelig er der dog også forskelle. Den vigtigste forskel er, at der her simuleres med to forskellige variable, mens der kun kan ses på hvordan det påvirker én enkelt variabel. Dvs. der er to uafhængige variable og her altid kun én afhængig variabel. En typisk tovejstabel vil se ud som herunder.

MF-bidrag		Marketingomkostninger i 1000 kr.															
	2840,8	80	82	84	86	88	90	92	94	96	98	100	102	104	106	108	
Pris	210	964,8	974,5	984,1	993,7	1003	1013	1023	1032	1042	1051	1061	1071	1080	1090	1099	
	220	1067	1078	1089	1100	1111	1122	1133	1144	1155	1166	1177	1188	1199	1210	1220	
	230	1163	1175	1188	1200	1213	1225	1238	1250	1262	1275	1287	1300	1312	1325	1337	
	240	1254	1268	1282	1296	1310	1323	1337	1351	1365	1379	1393	1406	1420	1434	1447	
	250	1341	1356	1371	1386	1401	1417	1432	1447	1462	1477	1493	1508	1523	1538	1552	
	260	1421	1438	1455	1471	1488	1505	1521	1538	1554	1571	1588	1604	1621	1637	1653	
	270	1497	1515	1533	1551	1569	1587	1605	1623	1641	1659	1677	1695	1713	1731	1749	
	280	1568	1587	1607	1626	1645	1665	1684	1704	1723	1742	1762	1781	1801	1820	1840	
	290	1633	1654	1675	1695	1716	1737	1758	1779	1800	1820	1841	1862	1883	1904	1925	
	300	1693	1715	1738	1760	1782	1804	1826	1849	1871	1893	1915	1938	1960	1982	2004	
	310	1748	1772	1795	1819	1843	1866	1890	1913	1937	1961	1984	2008	2031	2055	2078	
	320	1798	1823	1848	1873	1898	1923	1948	1973	1998	2023	2048	2073	2098	2123	2148	
	330	1842	1869	1895	1922	1948	1974	2001	2027	2054	2080	2106	2133	2159	2186	2212	
	340	1882	1909	1937	1965	1993	2021	2049	2076	2104	2132	2160	2188	2215	2243	2271	
	350	1916	1945	1974	2003	2033	2062	2091	2120	2150	2179	2208	2237	2266	2296	2325	
	360	1945	1975	2006	2037	2067	2098	2128	2159	2190	2220	2251	2281	2312	2343	2373	
	370	1968	2000	2032	2064	2096	2128	2161	2193	2225	2257	2289	2321	2353	2385	2417	
	380	1987	2020	2054	2087	2121	2154	2187	2221	2254	2288	2321	2355	2388	2421	2454	
	390	2000	2035	2070	2105	2140	2174	2209	2244	2279	2314	2348	2383	2418	2453	2488	

I det illustrerede eksempel udregnes markedsføringsbidraget på baggrund af en række priser og en række marketingomkostninger. Forudsætningen for at lave en sådan tabel er ligesom for envejstabellen, at arbejdsbogen er opbygget, så markedsføringsbidraget udregnes på baggrund af en række inddata – og heriblandt selvfølgelig prisen og marketingomkostningerne. Hvis disse forudsætninger er opfyldt oprettes tabellen således.

1. Opret reference til den afhængige variabel i øverste venstre hjørne af tabellen (i eksemplet markedsføringsbidraget)
2. Udfyld første kolonne med tal for den ene afhængige variabel (i eksemplet prisen). En hurtig metode til at få skrevet tallene er gennemgået i forrige afsnit om envejstabeller.
3. Udfyld første række med tal for den anden afhængige variabel (i eksemplet marketingomkostningerne)



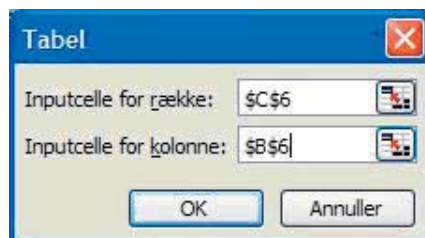
CISO Conference
Produced by **Inspired**

**Apollo Hotel 1, Groenlandsekade
Vinkeveen, Amsterdam, NL
Dec 5th 2019**

**Listen, learn & build relationships with our
Network of CISOs & Cyber Security Leaders**

Inspired

4. Marker hele tabelområdet
5. Gå i menuen *Data*
6. Vælg *Tabel*
7. Indlæg reference (klik på cellen med musen) til *Inddata* for *Marketingomkostninger* i feltet *Inputcelle for række*
8. Indlæg reference (klik på cellen med musen) til *Inddata* for *Pris* i feltet *Inputcelle for kolonne*



9. Tryk *OK*

Så er tabellen oprettet og i vores eksempel er markedsføringsbidraget udregnet for alle kombinationer af de angivne prissætninger i første kolonne og marketingomkostninger i første række.

5. Statistiske problemstillinger

5.1 Sandsynlighedsregning

Her gennemgås kort, hvordan sandsynlighedsregning udføres i tre forskellige fordelinger: Binomialfordelingen, Poissonfordelingen og Normalfordelingen.

5.1.1 Binomialfordelingen

Der udføres 35 forsøg og sandsynligheden for succes er 0,4.

1. Hvad er sandsynligheden for 14 succeser? $P(X = 4)$
2. Hvad er sandsynligheden for 10 eller færre succeser? $P(X \leq 10)$
3. Hvad er sandsynligheden for flere end 15 succeser? $P(X > 15) = P(X \geq 16) = 1 - P(X \leq 15)$

Svarene udregnes i Excel som

1. =BINOMIALFORDELING(14;35;0,4;FALSK) = 0,1366
2. =BINOMIALFORDELING(10;35;0,4;SAND) = 0,1123
3. =1-BINOMIALFORDELING(15;35;0,4;SAND) = 0,2997

5.1.2 Poissonfordelingen

Antag at middelværdien $\lambda = 6$ (Lampda = 6)

1. Hvad er sandsynligheden for 3 hændelser? $P(X = 3)$
2. Hvad er sandsynligheden for 4 eller færre hændelser? $P(X \leq 4)$
3. Hvad er sandsynligheden for flere end 5 hændelser? $P(X > 5) = P(X \geq 6) = 1 - P(X \leq 5)$

Svarene udregnes i Excel som

1. =POISSON(3;6;FALSK) = 0,0892
2. =POISSON(4;6;SAND) = 0,2851
3. =1-POISSON(5;6;SAND) = 0,5543

5.1.3 Normalfordelingen

Antag at middelværdien $\mu = 100$ (My = 100) og standardafvigelsen $\sigma = 7$ (Sigma = 7)

1. Hvad er sandsynligheden for en hændelse på 80? $P(X = 80)$
2. Hvad er sandsynligheden for en hændelse på 90 eller mindre? $P(X \leq 90)$
3. Hvad er sandsynligheden for en hændelse større end 120? $P(X > 120) = 1 - P(X \leq 120)$

Svarene udregnes i Excel som

1. $P(X = 80) = 0$ Punktsandsynligheden er 0 i en kontinuert fordeling.
2. =NORMFORDELING(90;100;7;SAND) = 0,0766
3. =1-NORMFORDELING(120;100;7;SAND) = 0,0021

5.2 Deskriptiv statistik

For at udføre beskrivende statistik, histogram og normalfraktildiagram tages udgangspunkt i nedenstående datasæt, der angiver tiden for et bilsyn målt i minutter.

	C
1	
2	Tiden for et bilsyn
3	30
4	28
5	21
6	19
7	23
8	29
9	29
10	26
11	18
12	22
13	23
14	21
15	22
16	27
17	24
18	22
19	21
20	31
21	28
22	26

5.2.1 Histogram

For at histogrammet skal blive godt kræver Excel en smule forarbejde. Vi skal nemlig selv inddele datamaterialet i intervaller af passende størrelse.

Intervallængden udregnes først og den findes som $(\text{maks}-\text{min})/\text{kvrod}(n)$. Altså som den største observation minus den mindste observation og divideret med kvadratroden af antallet af observationer. Det tal man dividerer med skal dog være et helt tal, men det findes der heldigvis midler for i Excel. Her bliver bare vist en af metoderne.

Forestil dig at datasættet med tiden for bilsyn er placeret i C2 til C22 (altså vores 20 observationer). Formlen til at bestemme intervallængden er hermed:

$\text{=(MAKS(C3:C22)-MIN(C3:C22))/RUND.NED(KVROD(TÆL(C3:C22)));0)}$.

Formlen kan ved første øjekast virke en smule uoverskuelig, specielt når vi kun har 20 observationer. Men forestil dig du sidder med 6000 observationer. Resultatet af formlen bliver 3,25, hvilket er vores intervallængde.

Selve intervallerne udføres nu først ved at bruge vores minimumsformel for at bestemme den mindste værdi i datasættet da Excel skal vide hvor det første interval skal starte. Derefter findes de resterende intervaller ved hele tiden at summere det forrige interval med intervallængden.

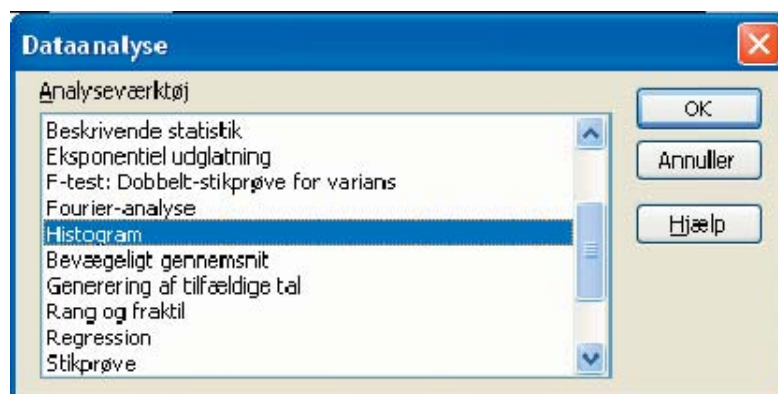
Processen fortsættes indtil vi når den højeste observation som man vil ramme præcist, hvis man har benyttet ovenstående teknik til udregning af intervallængden. Resultatet ses nedenfor:


Interval	Intervallængde
18	3,25
21,25	
24,5	
27,75	
31	

Første kolonne: 1. Række: =MIN(C3:C22)

Første kolonne 2. Række: =J4+\$K\$4 (Hvor J4 er 1.række i første kolonne og K4 er intervallængden)

Nu kan vi tegne selve histogrammet dette gøres ved at vælge Funktioner – Dataanalyse – Histogram





Max's next Bookboon eBook

Your Boss: Sorted!

By Patrick Forsyth - 55 pages

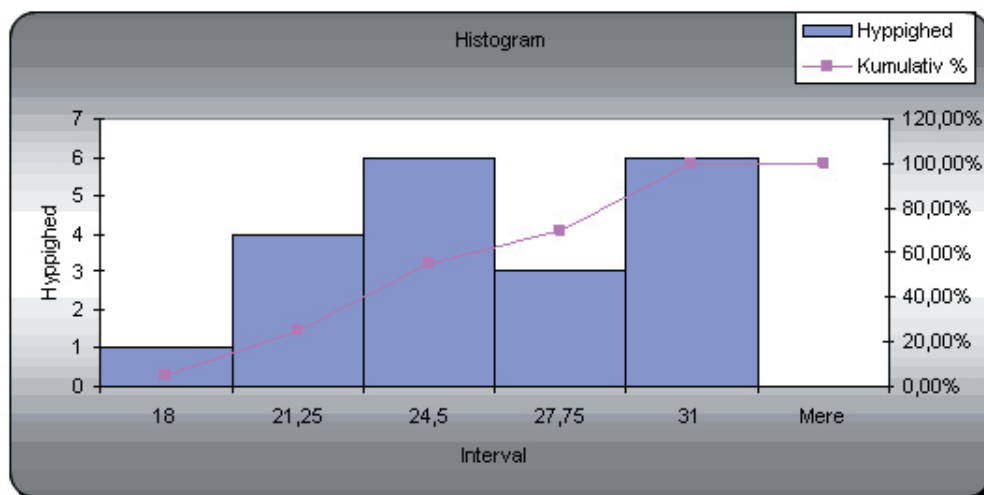
Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com

Og inputtet vælges som vist nedenfor

Og histogrammet vil så efter lidt små justeringer se således ud:



Her er det vigtigt, at kigge efter formen på histogrammet og den kumulerede %. Selve hyppigheden skal have en klokkeform, hvis data er normalfordelte. Den kumulerede % skal følge et S der ligger ned. Det kan være svært at se hvorvidt data er normalfordelte på baggrund af dette plot, men det har nu alligevel sin ret.

5.2.2 Normalfraktildiagram

Dette diagram udføres med et simpelt XY-diagram, men inden vi kan tegne er der nogle få beregninger vi skal igennem. Det færdige skema ser ud som nedenstående:

Obsnr. (i)	$(i-0,5)/n$	Tiden sorteret stigende	$\Phi^{-1}((i-0,5)/n)$
1	0,025	18	-1,959964
2	0,075	19	-1,4395315
3	0,125	21	-1,1503494
4	0,175	21	-0,9345893
5	0,225	21	-0,755415
6	0,275	22	-0,5977601
7	0,325	22	-0,4537622
8	0,375	22	-0,3186394
9	0,425	23	-0,1891184
10	0,475	23	-0,0627068
11	0,525	24	0,06270678
12	0,575	26	0,18911843
13	0,625	26	0,31863936
14	0,675	27	0,45376219
15	0,725	28	0,59776013
16	0,775	28	0,75541503
17	0,825	29	0,93458929
18	0,875	29	1,15034938
19	0,925	30	1,43953147
20	0,975	31	1,95996398

Skemaet burde mere eller mindre være selvsigende, men det kræver nok et par enkelte forklaringer.

Første kolonne er blot observationsnummeret, som her er fra 1 til 20.

Anden kolonne udregner det der står øverst nemlig $(i-0,5)/n$. Her er $n=20$ altså antallet af observationer. Og i er det respektive observationsnummer. Eksempelvis vil man i første række i anden kolonne finde udregningen $(1-0,5)/20$ som giver 0,025. Sådan fortsættes nedefter.



 MTHøjgaard

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



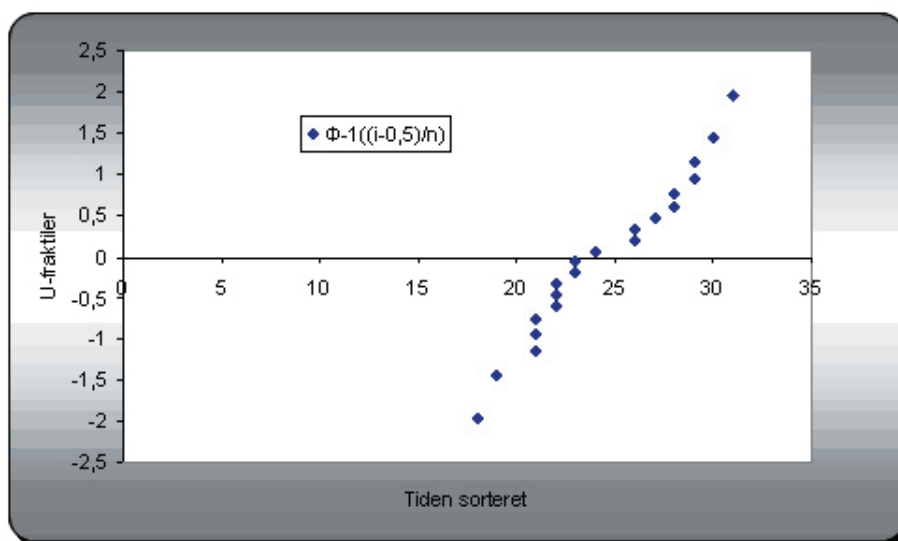
Tredje kolonne indeholder tiden sorteret stigende, dette udføres i Excel ved at markere det man ønsker



sorteret og så trykke på knappen sorter stigende som findes i værktøjslinien Standard. Kan også findes ved at vælge menuen "DATA" og "SORTER.."

Fjerde kolonne indeholder Excel funktionen =STANDARDNORMINV(F3). Her henviser F3 til anden kolonne. Dvs. der reelt i første række står standardnorminv(0,025) og resultatet bliver ca. -1,96.

Nu tegnes XY-diagrammet, på sædvanlig vis, hvor vi har tiden sorteret ud af x-aksen og u-fraktilerne (fjerde kolonne) op ad y-aksen. Det færdige diagram ser ud som nedenfor:

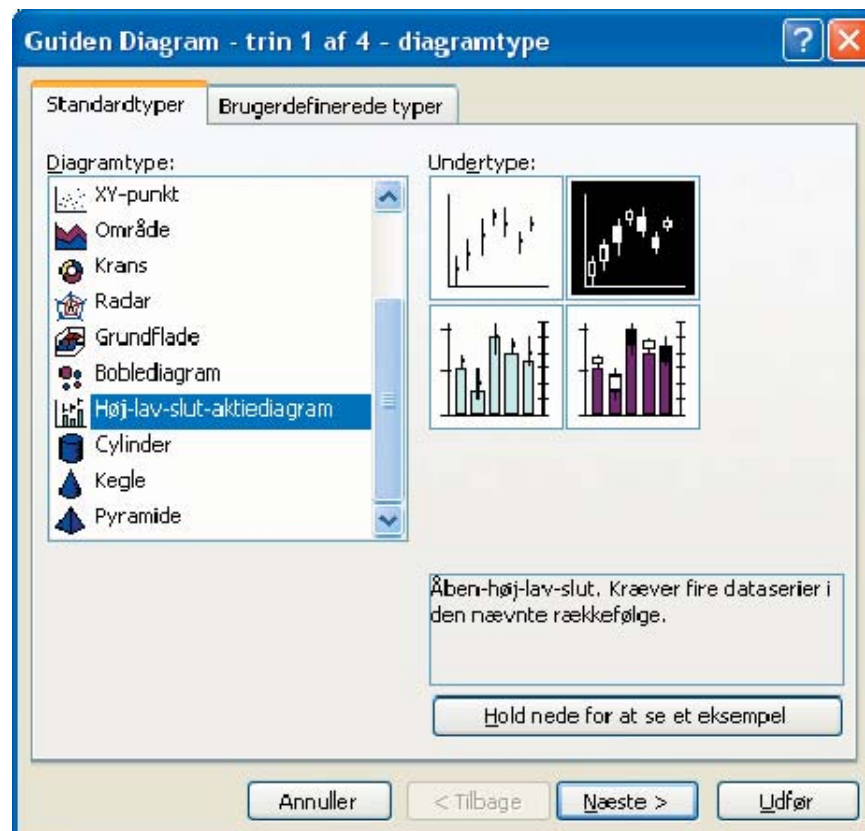


Her ses det, at punkterne snor sig pænt om en ret linie, hvorfor antagelsen om normalfordelte data er opfyldt.

5.2.3 Boxplot - Kassedigram

Excel har desværre ikke nogen direkte funktion til at tegne et boxplot (kassedigram) i den forstand vi kender det. Så vi må snyde lidt for at få det bedste ud af det.

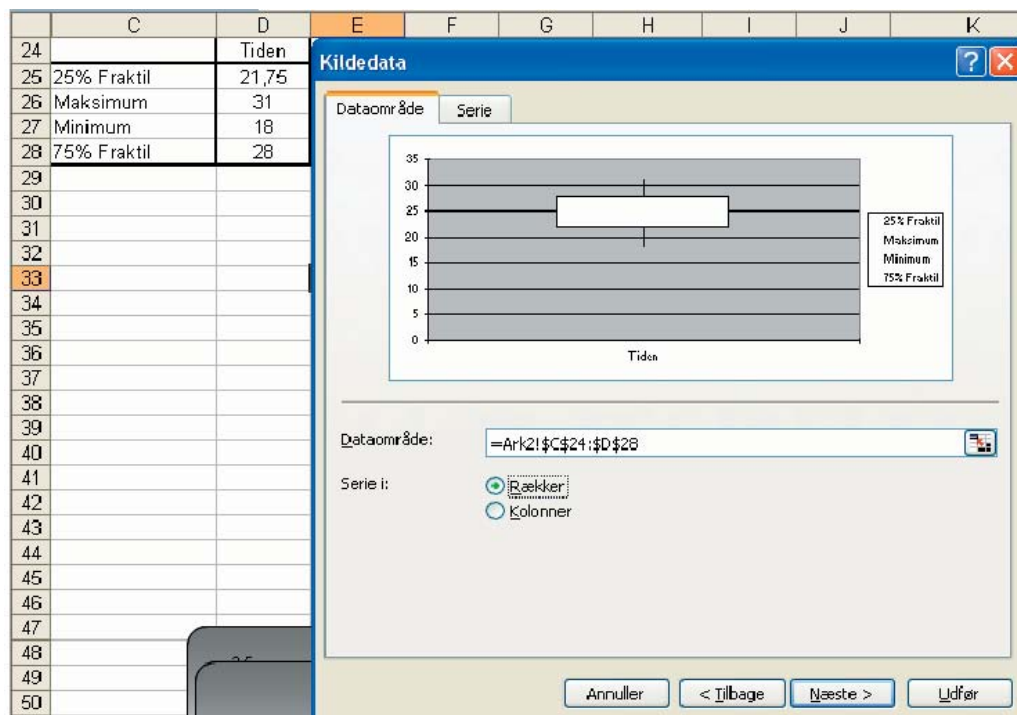
Vi benytter et "Åben-Høj-Lav-Slut" diagram



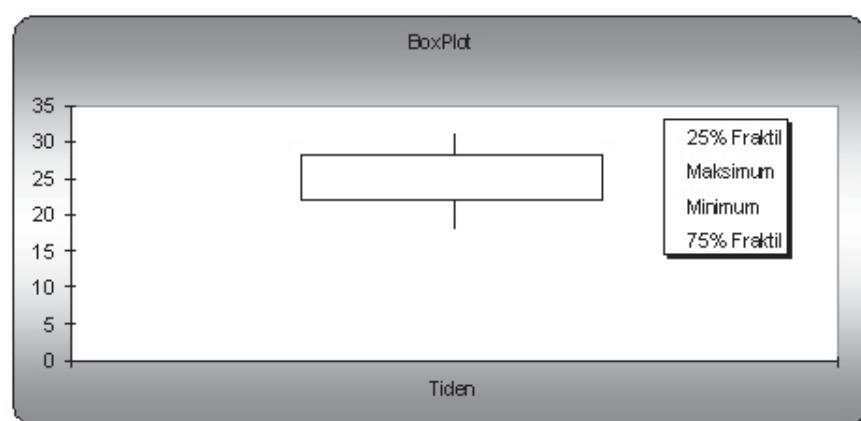
Der kræver 4 dataserier som skal angives i ovenstående rækkefølge. Her betyder det, at vi skal opstille nedenstående tabel:

	Tiden			
25% Fraktil	21,75	=FRAKTIL(C3:C22;0,25)		
Maksimum	31	=MAKS(C3:C22)		
Minimum	18	=MIN(C3:C22)		
75% Fraktil	28	=FRAKTIL(C3:C22;0,75)		

Og BoxPlottet tegnes nu som vist nedenfor



Det færdige resultat ser således ud



5.2.4 Beskrivende statistik

Resten af den beskrivende statistik udføres i Excel ved at vælge "Funktioner – Dataanalyse – Beskrivende statistik" som vist.



Og der vælges nu efter ønske som vist

Output fra analysen vil se således ud:

Tiden for et bilsyn	
Middelværdi	24,5
Standardfejl	0,859926557
Median	23,5
Tilstand	21
Standardafvigelse	3,845708476
Stikprøvevarians	14,78947368
Kurtosis	-1,178695067
Skævhed	0,111044887
Område	13
Minimum	18
Maksimum	31
Sum	490
Antal	20
Konfidensniveau(95,0%)	1,799846966

Der skal knyttes nogle enkelte kommentarer til outputtet.

Middelværdien (Forventningen)

Er 24,5, dvs. gennemsnittstiden for et bilsyn er 24,5 minutter og derfor forventer vi at den næste bil der skal synes vil tage 24,5 minutter at syne. Vi kan dog tydeligt se i materialet, at ikke alle syn tager 24,5 minut, vores middelværdi er altså behæftet med usikkerhed.

Varians og Standardafvigelse:

Varians er reelt blot en mellemregning for at komme frem til det egentlige mål for spredning, nemlig standardafvigelsen.

$$\text{Varians} = \hat{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x - \bar{x})^2$$

$$StdAfv = \hat{\sigma} = s = \sqrt{s^2}$$

Så når Excel har udregnet variansen til at være 14,79, så kan vi finde standardafvigelsen blot ved at tage kvadratroden af 14,79. Excel har dog allerede gjort det for os og resultatet er 3,85.

Standardafvigelsen er et mål for hvor langt vi forventer at ligge fra middelværdien i vores stikprøve.

Standardfejl:

Er vores estimat på den fejl vi begår, når vi estimerer middelværdien. Vi siger at vores standardfejl på gennemsnittet kan udregnes som

$$Std.Error = \frac{s}{\sqrt{n}}$$

Og denne størrelse får vi brug for når vi skal udregne et konfidensinterval for middelværdien.

Skævhed

En fordeling (datasæt) kan i grove træk siges at være symmetrisk, højreskæv eller venstreskæv. Det udtrykkes ved beregningen om skævhed, og der gælder følgende

Skævhed negativ ($SKH < 0$) så er fordelingen venstreskæv.

Skævhed positiv ($SKH > 0$) så er fordelingen højreskæv.

Skævhed er 0 ($SKH = 0$) så er fordelingen symmetrisk.

Konfidensinterval for middelværdien¹:

Findes ved hjælp af formlen

$$\hat{\mu} \pm t_{1-\alpha/2}(fg = n-1) \cdot \frac{s}{\sqrt{n}}$$



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.

banedanmark



Excel har dog allerede udregnet en del af denne formel for os i tallet udfor konfidensniveau (1,7998)

er udregningen for $t_{1-\alpha/2}(fg = n - 1) \cdot \frac{s}{\sqrt{n}}$. Vores konfidensinterval (95%) bliver altså

Nedre grænse: $24,5 - 1,7998 = 22,7002$

Øvre grænse: $24,5 + 1,7998 = 26,2998$

Ud fra stikprøven kan vi altså sige, at den sande værdi af middelværdien med 95% sandsynlighed vil ligge mellem nedre grænse og øvre grænse.

5.3 Regressionsanalyse

Her vil udelukkende blive gennemgået en multipel regression, idet processen med en simpel regression er præcis den samme (dog undersøges ikke for multicollinearitet ved simpel regressionsanalyse).

Dataene til modellen er som vist nedenfor

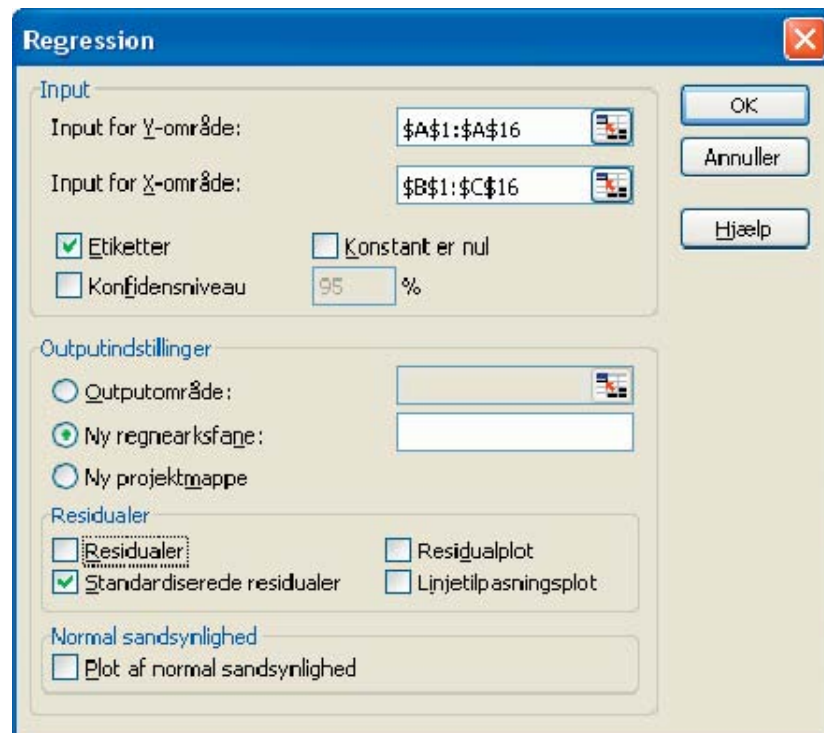
	A	B	C
1	Prisen 1000 kr	Størrelse kvm	Forbedringer i 1000 kr.
2	1192,107384	56	200
3	2539,112238	72	25
4	2100,661895	59	75
5	1954,282702	50	0
6	1780,392431	49	370
7	1878,483708	80	110
8	2411,91713	85	270
9	879,8937688	34	34
10	1543,74532	97	65
11	2004,474873	100	140
12	2171,507554	96	213
13	1623,007215	102	107
14	2562,861418	109	140
15	1512,886165	54	60
16	1387,234925	48	29

I kolonne A har vi Prisen målt i 1000 kr. for 15 lejligheder i København, dette er vores responsvariabel (afhængige variabel). i kolonne B har vi størrelsen på lejligheden målt i kvadratmeter (uafhængig variabel). Og kolonne C indeholder forbedringer der er foretaget på lejlighederne i løbet af de sidste 10 år (uafhængig variabel).

Regressionsanalysen udføres i Excel ved at vælge "Funktioner – Dataanalyse – Regression" som vist



Og der vælges nu efter ønske



The image shows the 'Regression' dialog box in Microsoft Excel. It is divided into several sections: 'Input', 'Outputindstillinger', 'Residualer', and 'Normal sandsynlighed'. In the 'Input' section, 'Input for Y-område:' is set to '\$A\$1:\$A\$16' and 'Input for X-område:' is set to '\$B\$1:\$C\$16'. The 'Etiketter' checkbox is checked, and 'Konstant er nul' is unchecked. The 'Konfidensniveau' is set to '95 %'. In the 'Outputindstillinger' section, 'Ny regnearksfane:' is selected. In the 'Residualer' section, 'Standardiserede residualer' is checked, while 'Residualer', 'Residualplot', and 'Linjetilpasningsplot' are unchecked. The 'Normal sandsynlighed' section has 'Plot af normal sandsynlighed' unchecked. On the right side, there are buttons for 'OK', 'Annuller', and 'Hjælp'.

Outputtet består nu af 2 dele: 1. Resume output og 2. Residualoutput og de ser ud som vist



CISO Conference
Produced by **Inspired**

**Apollo Hotel 1, Groenlandsekade
Vinkeveen, Amsterdam, NL
Dec 5th 2019**

**Listen, learn & build relationships with our
Network of CISOs & Cyber Security Leaders**

Inspired

RESUMEOUTPUT						
Regressionsstatistik						
Multipel R	0,562970666					
R-kvadreret	0,31693597					
Justeret R-kvadreret	0,203091965					
Standardfejl	436,6508638					
Observationer	15					
ANOVA						
	fg	SK	MK	F	Signifikans F	
Regression	2	1061597,799	530798,9	2,78395	0,101570712	
Residual	12	2287967,722	190663,98			
I alt	14	3349565,521				
	Koefficienter	Standardfejl	t-stat	P-værdi	Nedre 95%	Øvre 95%
Skæring	992,7567036	375,1143146	2,6465444	0,02131	175,4528233	1810,0606
Størrelse kvm	10,8082694	4,884627978	2,2127109	0,04705	0,165579303	21,450959
Forbedringer i 1000 kr.	0,467571415	1,151020134	0,4062235	0,69173	-2,04028602	2,9754288

Inden vi kigger på det output, som Excel frembringer skal vi først have opstillet model for materialet og kontrolleret forudsætningerne.

5.3.1 Model²

$Y_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \dots, n = 15$ Hvor
 $\mu_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2}$

Hvor

Y angiver prisen målt i 1000 kr.

X1 angiver størrelsen i kvadratmeter

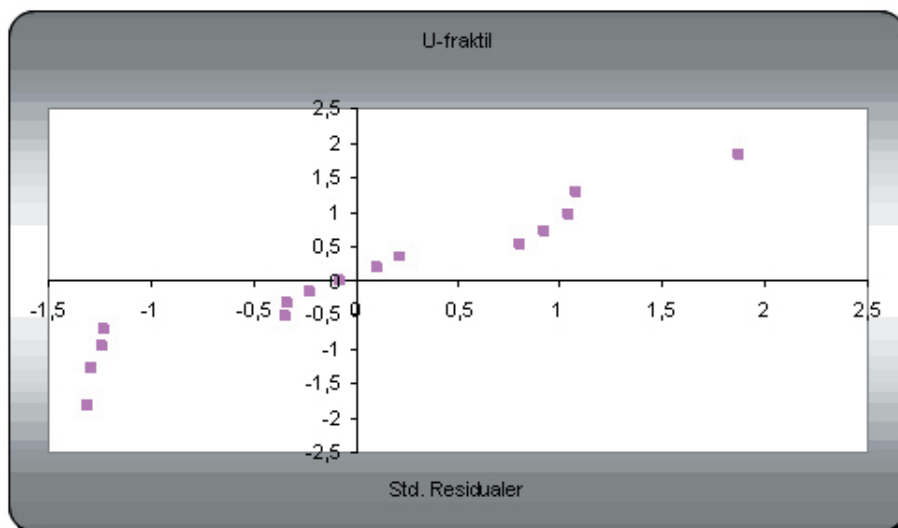
X2 angiver forbedringer målt i 1000 kr.

5.3.2 Forudsætninger³

1. Y_i er normalfordelt dvs. $e_i \sim N(0, \sigma^2)$
 2. Y_i har middelværdi $E(Y_i) = \mu_i$, som er en lineær funktion af x. Dvs. $E(e_i) = 0$
 3. Y_i har varians $Var(Y_i) = \sigma^2$ altså konstant varians. Dvs. $Var(e_i) = \sigma^2$
 4. Y 'erne er stokastisk uafhængige. Dvs. $cov(e_i, e_j) = 0$ for $i \neq j$
- Ekstra forudsætning: Multikollinearitet

Forudsætning 1

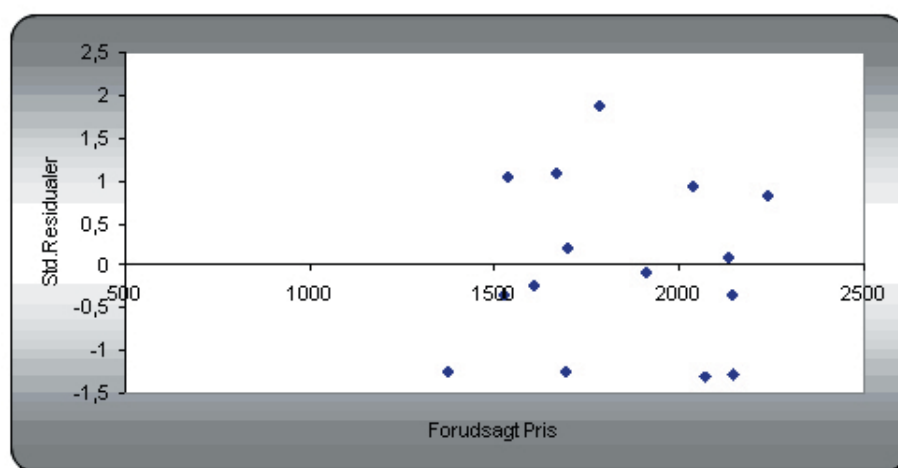
Normalitet kontrolleres ved et normalfraktildiagram, vær opmærksom på at det er et normalfraktildiagram udført på de standardiserede residualer, men processen er den samme som beskrevet under normalfraktildiagram.



Der kan være lidt problemer med denne forudsætning, men jo færre observationer vi benytter jo mere fleksibel skal man være, når man vurderer diagrammet.

Forudsætning 2:

Linearitet kontrolleres i et diagram med forudsagte værdier ud af x-aksen og standardiserede residualer op ad y-aksen.

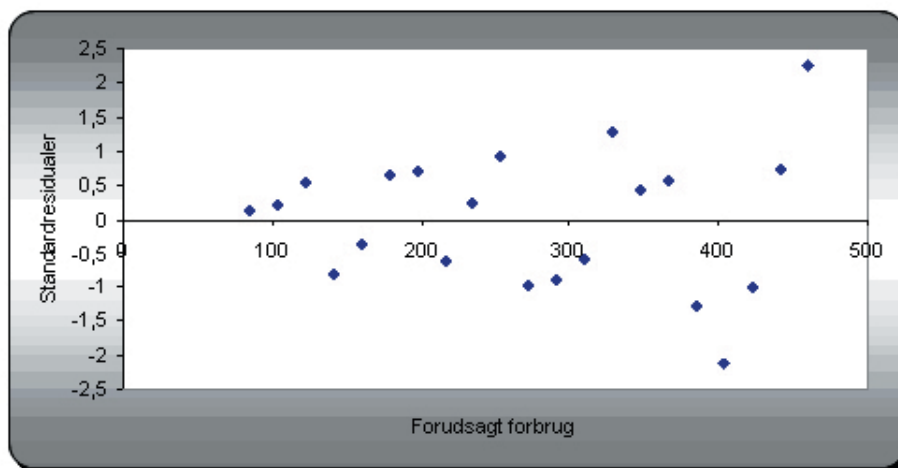


Her skal punkterne ligge usystematisk omkring 0. Der må eksempelvis ikke være tendens til kurve mv.

Her er der ingen tegn på brud på forudsætningen

Forudsætning 3:

Konstant varians kontrolleres i samme diagram som forudsætning 2. Og nu må der ikke bære "trompet" eller "Kile" i nogle af enderne. Nedenfor er vist et eksempel, hvor der er brud på denne forudsætning. Vær opmærksom på, at det ikke er fra denne opgave. Her ses tydeligt en "trompet"



I opgaven stillet her, er der dog ingen tegn på trompet og derfor er denne forudsætning opfyldt.

Forudsætning 4:

Stokastisk uafhængighed, kan kun kontrolleres såfremt data er en tidsserie, eller der er en logisk rækkefølge i observationerne. Dette er ikke tilfældet i denne opgave, hvorfor forudsætningen ikke kan kontrolleres.

Såfremt data er en tidsserie, kontrolleres forudsætningen i et diagram med observationsnummeret (eller tiden) ud af x-aksen og de standardiserede residualer op ad y-aksen. Der må ikke være systematik i punkterne (dvs. ingen systematik i fortegnsskift). Nedenfor er vist et eksempel, med brud på denne forudsætning – det kaldes autokorrelation.

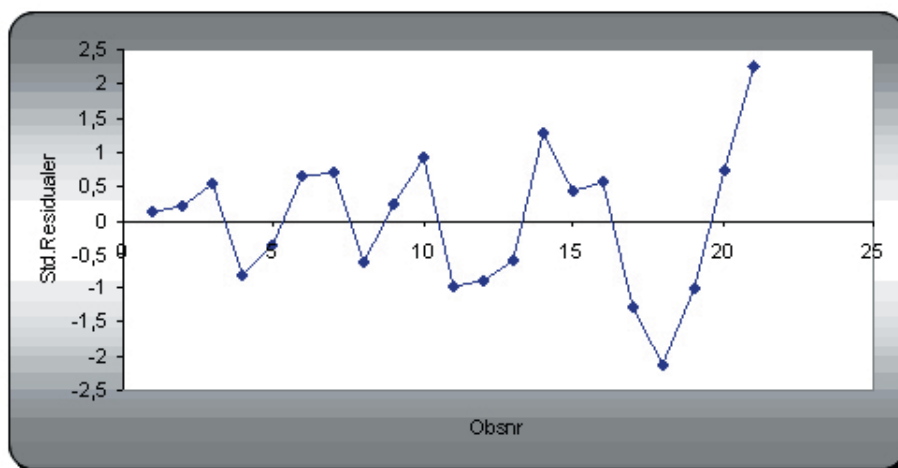
Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com





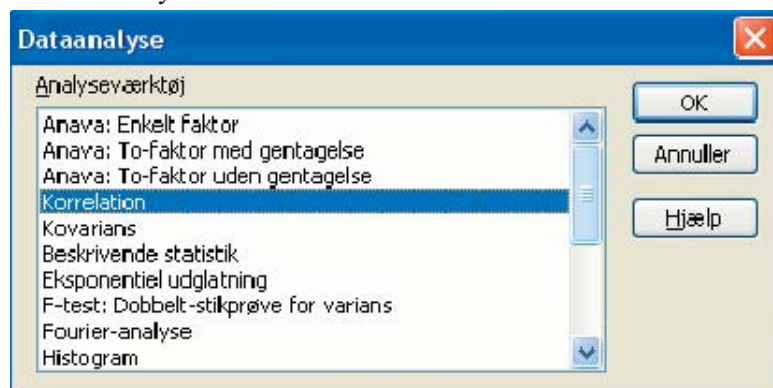
Ekstra forudsætning: Multicollinearitet:

De forklarende (uafhængige) variable skal helst være lineært uafhængige, da det ellers kan skade estimationen af β 'erne og dermed forringe modellens evne til forudsigelse.

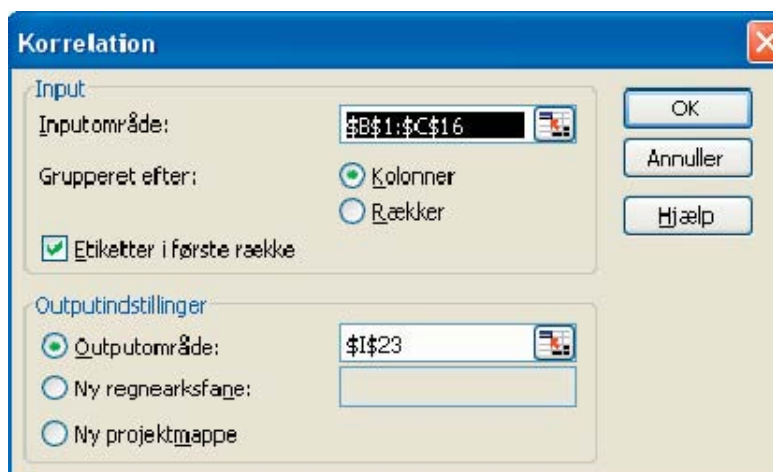
Der er forskellige metoder til at kontrollere om der er multikollinearitet, det være både grafiske og algebraiske metoder. Her vil vi dog blot benytte en enkelt metode, hvor Excel vil være til stor hjælp.

Udregn korrelationen mellem de enkelte forklarende variable.

Vælg "Funktioner – Dataanalyse – Korrelation" som vist



Og input vælges som følgende



Og output bliver

	Størrelse kvm	Forbedringer i 1000 kr.
Størrelse kvm	1	
Forbedringer i 1000 kr.	0,180736812	1

Det vil sige at korrelationen mellem størrelse og forbedringer er 0,1807. Der gælder at jo tættere denne er på 0, jo mindre multicollinearitet er der. Her er korrelationen ikke af væsentlig størrelse hvorfor der ikke er tegn på multicollinearitet.

Lad os nu vende tilbage til selve resumeoutputtet og se hvad vi bliver oplyst med her. For nemmere at kunne overskue outputtet skrives beregningerne i nedenstående output.

RESUMEOUTPUT						
Regressionsstatistik						
Multipel R	=Kvrod(R-kvadreret)					
R-kvadreret	= SKreg/SKres (Forklaringsgrad)					
Justeret R-kvadreret	$= \frac{S_T^2 - S_{res}^2}{S_T^2}$		$S_T^2 = \frac{SK}{n - 1}$		Teststørrelse for om alle forklarende variable kan udelades af modellen	
Standardfejl	=Kvrod(MKres)					
Observationer	n					
ANAVA						
	fg	SK	MK	F	Signifikans F	
Regression	k= antal forklarende variable	=Sum((Ypred-Ygns)^2)	=SK/fg=S ² _{reg}	=MK(reg)/MK(res)	0,101570712	
Residual	=n-k-1	=Sum((Yi-Ypred)^2)	=SK/fg=S ² _{res}			
I alt	=n-1	=Sum((Yi-Ygns)^2)				
Koefficienter						
		Standardfejl	t-stat	P-værdi	Nedre 95%	Øvre 95%
Skæring	Skæring med Y-aksen (Beta0)	$\hat{SD}(\hat{\beta}_0)$	2,64654444	0,021314631	175,4528233	1810,060584
Størrelse kvm	Marginale ændring for x1 (hældning) (Beta1)	$\hat{SD}(\hat{\beta}_1)$	2,212710865	0,047050765	0,165579303	21,45095949
Forbedringer i 1000 kr.	Marginale ændring for x2 (hældning) (Beta2)	$\hat{SD}(\hat{\beta}_2)$	0,406223489	0,691729094	-2,040286018	2,975428849
t-stat og tilhørende P-værdi. her testes om de enkelte Beta'er kan være 0. Dvs. om de enkelte forklarende variable kan udelades af modellen, uden at miste signifikant forklaringsgrad. t-stat er teststørrelsen og P-værdi er signifikanssandsynligheden ved dobbeltsiddet test						
Her angives nedre og øvre grænse for et 95% konfidensinterval omkring de enkelte beta'er						

Model kan nu angives med de estimerede parametre som

$$\hat{y}_i = 1992,76 + 10,81 \cdot x_{i1} + 0,468 \cdot x_{i2}$$

Og forklaringsgraden er udregnet til ca. 32% (0,3169). Den udtrykker hvor stor en del af variationen i Prisen (Afhængige variabel), der kan forklares ved Størrelsen og Forbedringer – Altså de forklarende variable.

Nu efterses om modellen kan reduceres. For at kontrollere om samtlige forklarende variable kan udelades af modellen på en gang testes som nedenfor.⁴

5.3.3 Hypotese vedr. alle forklarende variable

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : H_0^c$$

Nulhypotesen angiver at ingen af de forklarende variable bidrager signifikant til forklaringsgraden. Alternativhypotesen angiver, at mindst én bidrager signifikant til forklaringsgraden.

5.3.4 Teststørrelse vedr. alle forklarende variable⁵

$$V = F = \frac{S_{regression}^2}{S_{residual}^2} \sim F(k, n - k - 1)$$

I denne opgave vil teststørrelsen få værdien

$$V = F = \frac{530.799}{190.664} = 2,78395 \sim F(2,12)$$

5.3.5 Afgørelse vedr. alle forklarende variable

Excel har allerede udregnet signifikanssandsynligheden for os, men for den interesserede læser findes den som.

$$Sig.SSH = P(F \geq 2,78395) = 1 - P(F \leq 2,78395) = 1 - 0,8943 = 0,10157$$



 **MTHøjgaard**

BEDRE LØSNINGER

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



I Excel udregnes det som =FFORDELING(2,78395;2;12).

Vi kan altså acceptere vores nulhypotese og kan dermed ikke afvise, at ingen af de forklarende variable bidrager til forklaringen.

Vi ender reelt set allerede nu i normalfordelingen, der kan dog være problemer med udelukkende at afgøre sin hypotese ved ovenstående test.

Normalt udelades de forklarende variable successivt (dvs. én af gangen), indtil de resterende forklarende variable alle bidrager signifikant til forklaringsgraden.

5.3.6 Reducering af modellen - Backward

Nu undersøges om de enkelte forklarende variable, kan udelades af modellen, det gøres ved et T-test og den forklarende variabel med højest signifikanssandsynlighed udelades først.

5.3.7 Hypotese vedr. enkelt forklarende variabel

$H_0 : \beta_i = 0$
 $H_1 : \beta_i \neq 0$ i vores tilfælde starter vi med β_2 (Altså om forbedringer kan udelades af modellen)

5.3.8 Teststørrelse vedr. enkelt forklarende variabel⁶

$$T = \frac{\hat{\beta}_i - \beta_{0i}}{SD(\hat{\beta}_i)} \sim T(fg = n - k - 1)$$

$$t = \frac{0,46757 - 0}{1,15102} = 0,406223 \sim t(fg = 15 - 2 - 1)$$

Altså nøjagtig som fundet i Exceludskriften

5.3.9 Afgørelse vedr. enkelt forklarende variabel

Excel har allerede fundet signifikanssandsynligheden for os, men for den interesserede læser findes den som:⁷

$$Sig.SSH = 2 \cdot P(T \geq 0,406223) = 2 \cdot (1 - P(T \leq 0,406223)) = 2 \cdot (1 - 0,6541) = 0,691729$$

I Excel findes størrelsen som

$$=TFORDELING(0,406223;12;2)$$

Idet signifikanssandsynligheden er større end $\alpha (=0,05)$ accepteres nulhypotesen. Dvs. "Forbedringer" bidrager ikke signifikant til forklaringsgraden.

På denne måde afgøres for hver forklarende variabel og den med størst signifikanssandsynlighed (her er det "Forbedringer") fjernes fra modellen og en ny regressionsanalyse udføres. Prøv selv at frembringe nedenstående udskrift.

RESUMEOUTPUT						
<i>Regressionsstatistik</i>						
Multipel R	0,554565454					
R-kvadreret	0,307542842					
Justeret R-kvadreret	0,254276907					
Standardfejl	422,3952414					
Observationer	15					
ANOVA						
	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>	
Regression	1	1030134,901	1030134,9	5,773725	0,03191383	
Residual	13	2319430,62	178417,74			
I alt	14	3349565,521				
	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	1023,965641	355,175585	2,88298432	0,01282	256,655441	1791,276
Størrelse kvm	11,16689653	4,647340124	2,4028576	0,031914	1,126928611	21,20686

5.4 Enkeltsidet variansanalyse

I en undersøgelse vedrørende lønninger er nedenstående data indsamlet:

	A	B	C	D
1	Obs.nr.	Månedsløn i 1000 kr	Arbejdssted P=1, O=0	Alder
2	1	17	Privat	18
3	2	23	Offentlig	24
4	3	18,4	Offentlig	21
5	4	26	Offentlig	29
6	5	25	Privat	26
7	6	16,5	Privat	24
8	7	27	Privat	33
9	8	31	Privat	36
10	9	26	Privat	30
11	10	26	Offentlig	38
12	11	23	Offentlig	37
13	12	25	Offentlig	31
14	13	23	Offentlig	33
15	14	31	Offentlig	40
16	15	36	Offentlig	55
17	16	62	Privat	60
18	17	42	Privat	52
19	18	25	Privat	46
20	19	35	Privat	57
21	20	25	Offentlig	54
22	21	27	Offentlig	43

I den første analyse ønskes udarbejdet en enkeltsidet variansanalyse, hvor månedslønnen er den afhængige variabel og alder er den uafhængige variabel. Dog ønskes at alderen inddeles i grupper hvor:

- Gruppe 1 er alderen 18 til og med 29
- Gruppe 2 er alderen 30 til og med 39
- Gruppe 3 er alderen 40 og alt derover.

Udfør analysen.

Løsning:

Der ønskes altså her en sammenligning af 3 normalfordelinger, for at vurdere hvorvidt der er forskel i middelværdierne – altså i lønningerne imellem grupperne.

5.4.1 Model⁸

$Y_{ij} \sim N(\mu_i, \sigma_2)$ for $i = 1, 2, \dots, n_j$ s.u. Hvor
 $j = 1, 2, 3$

$$\mu_j = \beta_0 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3$$

Grunden til at leddet $\beta_1 \cdot X_1$ er udeladt er, at første gruppe vælges som referencegruppe og de resterende middelværdier beregnes ud fra referencen.

5.4.2 Forudsætninger⁹

1. Stokastisk uafhængighed
 - a. Er vanskelig at kontrollere grafisk, hvorfor den som oftest afgøres ved brug af sund fornuft, eller blot antages opfyldt.
2. Normalitet
 - a. Her udføres fraktildiagrammer og/eller histogrammer, som blev udført under multipel regressionsanalyse. Det gøres derfor ikke her, så kig evt. under regressionsanalysen for udførelse.
3. Varianshomogenitet.
 - a. Denne testes ved hjælp af Bartlett's test. Eller kontrolleres grafisk ved et XY-diagram med forudsagte værdier ud af x-aksen og standardiserede residualer op ad y-aksen.



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.

banedanmark



Excel:

Først skal alder opdeles i de ovenstående grupper, dette gøres ved hjælp af HVIS-funktion som =HVIS(D2<=29;1;HVIS(D2<=39;2;3)), derefter kopieres formelen det ønskede antal gange. Så ender man op med følgende input til modellen.

	F	G
1	Månedsløn i 1000 kr	Aldersgrupper
2	17	1
3	23	1
4	18,4	1
5	26	1
6	25	1
7	16,5	1
8	27	2
9	31	2
10	26	2
11	26	2
12	23	2
13	25	2
14	23	2
15	31	3
16	36	3
17	62	3
18	42	3
19	25	3
20	35	3
21	25	3
22	27	3

Excel kan, dog desværre ikke forstå en faktor med 3 niveauer, som aldersgrupper her er. Derfor er det faktisk ikke tilstrækkeligt, for at kunne udføre analysen.

Excel kan derimod godt forstå en dummy-variabel, som kan sammenlignes med en Tænd/Sluk knap. En dummy kan nemlig udelukkende antage værdien 0 (Sluk) eller 1 (Tænd). Derfor frembringes nu nedenstående input.

Hvor metoden er simpel – Eksempelvis gælder i søjlen Aldersgruppe1, at hvis søjlen Aldersgruppe er 1 skal den antage værdien 1 ellers 0. I søjlen aldersgruppe2 gælder tilsvarende, at hvis søjlen aldersgruppe er 2, skal den antage værdien 1 ellers 0.

Formlen er som du nok allerede har gættet af formen: =HVIS(G2=1;1;0).

Denne proces kunne naturligvis være klaret direkte ud fra søjlen alder, så dette er blot gjort for intuitionens skyld.

	F	G	I	J	K
1	Månedsløn i 1000 kr	Aldersgrupper	Aldersgruppe1	Aldersgruppe2	Aldersgruppe3
2	17	1	1	0	0
3	23	1	1	0	0
4	18,4	1	1	0	0
5	26	1	1	0	0
6	25	1	1	0	0
7	16,5	1	1	0	0
8	27	2	0	1	0
9	31	2	0	1	0
10	26	2	0	1	0
11	26	2	0	1	0
12	23	2	0	1	0
13	25	2	0	1	0
14	23	2	0	1	0
15	31	3	0	0	1
16	36	3	0	0	1
17	62	3	0	0	1
18	42	3	0	0	1
19	25	3	0	0	1
20	35	3	0	0	1
21	25	3	0	0	1
22	27	3	0	0	1

Nu kan analysen udføre som en regression, dog skal man huske, at søjlen aldersgruppe slet ikke bruges, og at søjlen aldersgruppe1 ligeledes er overflødig, idet den bliver brugt som reference i modellen.

Regression

Input

Input for Y-område:

Input for X-område:

☒ Etiketter ☐ Konstant er nul

☐ Konfidsniveau %

Outputindstillinger

☐ Outputområde:

☒ Ny regnearksfane:

☐ Ny projektmappe

Residualer

☐ Residualer ☐ Residualplot

☒ Standardiserede residualer ☐ Linjetilpasningsplot

Normal sandsynlighed

☐ Plot af normal sandsynlighed

OK Annuller Hjælp

Følgende output fremkommer nu

RESUMEOUTPUT						
<i>Regressionsstatistik</i>						
Multipel R	0,624800909					
R-kvadreret	0,390376176					
Justeret R-kvadreret	0,322640196					
Standardfejl	8,133403265					
Observationer	21					
ANOVA						
	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>	
Regression	2	762,497619	381,24881	5,7632	0,011629402	
Residual	18	1190,74048	66,152249			
I alt	20	1953,2381				
	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	20,98333333	3,32044798	6,3194284	5,9E-06	14,007331	27,95934
Aldersgruppe2	4,873809524	4,52500785	1,0770831	0,29567	-4,63287918	14,3805
Aldersgruppe3	14,39166667	4,3925398	3,2763885	0,00419	5,16328301	23,62005

Den eneste forskel fra outputtet omkring regressionsanalysen er tolkning af koefficienterne og hvad der reelt testes for i F-testet og t-testene.

Der gælder nu, at tallet 20,983 udfor skæring under koefficienter er middelværdien (gennemsnitslønnen) i aldersgruppe1 (18-29 år). Og tallet 4,8738 udfor Aldersgruppe2 (30-39 år) under koefficienter er forskellen i middelværdi set i forhold til referencen (aldersgruppe1). Det vil sige, at middelværdien for Aldersgruppe 2 er 25,857 (20,983+4,8738).

Da analysen nu er udført vendes tilbage til forudsætning 3



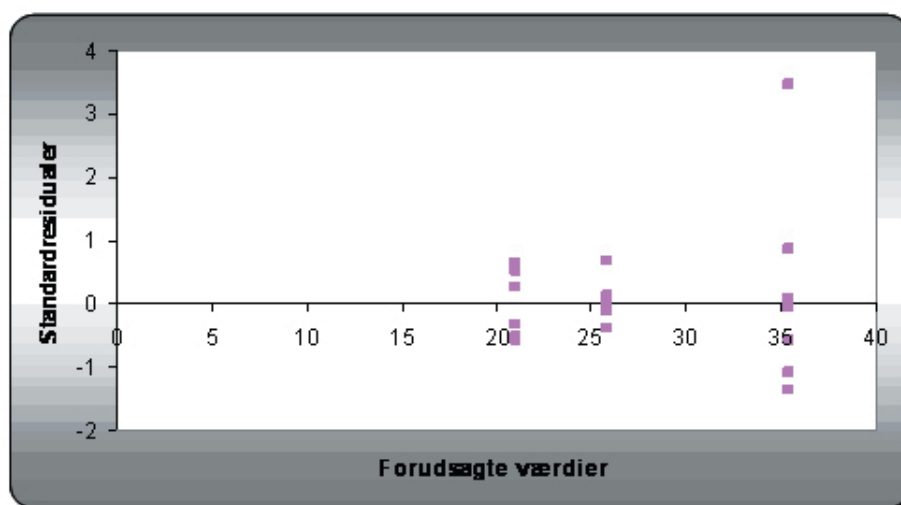
CISO Conference
Produced by **Inspired**

**Apollo Hotel 1, Groenlandsekade
Vinkeveen, Amsterdam, NL
Dec 5th 2019**

**Listen, learn & build relationships with our
Network of CISOs & Cyber Security Leaders**

Inspired

5.4.3 Konstant varians grafisk



Umiddelbart kunne det tyde på, at variansen i gruppe 3 er større end variansen i de 2 andre grupper, men det kan være vanskeligt at afgøre grafisk. Derfor testes.

5.4.4 Hypotese vedr. konstant varians

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$

$$H_1 : H_0^c$$

5.4.5 Teststørrelse vedr. konstant varians¹⁰

$$Q_B = fg \cdot \ln(S^2) - \sum_{j=1}^k fg_j \cdot \ln(S_j^2) \sim \chi^2(fg = k - 1)$$

Hvor

- k er antal af grupper (normalfordelinger - her 3)
- fg er antallet af frihedsgrader hørende til residualvariansen ($fg=n-k$ dvs. $21-3=18$)
- S_2 er residualvariansen og den er et vejet gennemsnit af gruppevarianserne, kan dog findes direkte i output til 66,1522
- fg_j er antallet af frihedsgrader hørende til gruppevariansen ($fg_j = n_j - 1$)
- S_j^2 er variansen hørende til gruppe j .

Så det står hurtigt klart, at de eneste ukendte er gruppevarianserne, så de skal først estimeres. Der er flere måder, at bestemme varianserne på, men en god og simpel metode er ved at bruge autofilter, en bedre metode er at bruge Pivottabel. Begge metoder vil blive gennemgået nedenfor.

5.4.6 Autofilter

Der findes muligvis nemmere metoder til at udregne varianserne, man kunne evt. sortere efter ønskede kriterier, så grunden til at autofilter bruges i denne sammenhæng er for at give et indtryk af hvad man kan med autofilter og hvad man ikke kan.

Start med at stille dig i rækken med overskrifter. Vælg ”Data – Filter – Autofilter” herefter fremkommer nogle drop-down pile, hvorfra brugeren kan vælge at sortere efter fastlagte kriterier.

	F	G
1	Månedsløn i 1000	Aldersgruppe
2	17	Sorter stigende
3	23	Sorter faldende
4	18,4	(Alle)
5	26	(De 10 øverste...)
6	25	(Bruger...)
7	16,5	1
8	27	2
9	31	3
10	26	2
11	26	2
12	23	2
13	25	2
14	23	2
15	31	3
16	36	3
17	62	3
18	42	3
19	25	3
20	35	3
21	25	3
22	27	3

Prøv nu at sortere efter kriteriet 1 – herefter skulle du gerne ende ud med nedenstående billede.

	F	G
1	Månedsløn i 1000	Aldersgruppe
2	17	1
3	23	1
4	18,4	1
5	26	1
6	25	1
7	16,5	1

I dette tilfælde kan vi nu direkte udregne variansen, vha. Excel funktionen ”varians” Men gør dig selv den tjeneste, at kopiere F2:G7 og indsæt disse tal, som værdier (Indsæt speciel – Værdier) et andet sted, inden du regner variansen. Manøvren skal udføres, såfremt der er spring i rækkerne, da vi ellers vil få et forkert estimat på variansen.

Her gælder, at variansen til Gruppe1 er givet ved: =VARIANS(F2:F7) og resultatet er 17,602. Prøv nu selv ved samme fremgangsmåde at frembringe varianserne til gruppe2 og gruppe3.

- Varians(Gruppe2)=7,476
- Varians(Gruppe3)=151,125

Nu kan man så udregne teststørrelsen

5.4.7 Pivottabel

Der er små forskelle i metoden afhængigt af hvilken version af Excel man har, metoden der bliver gennemgået her er fra Excel 2003.

Start med at stille dig et tilfældigt sted inde i din datatabel (ikke uden for tabellen). Vælg ”Data – Pivottabel og pivotdiagram” Herefter fremkommer en 3 trins guide.



Max's next Bookboon eBook
Your Boss: Sorted!
By Patrick Forsyth - 55 pages

Unlock your life.
Bookboon Premium is your key.

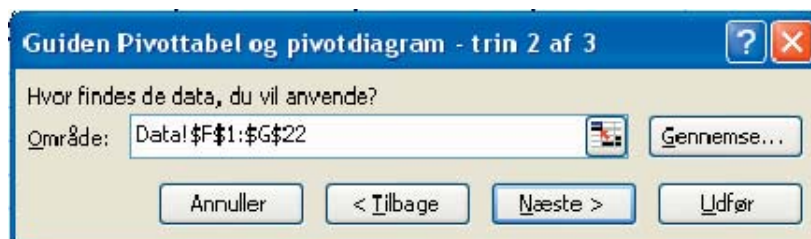
2000+ modern day bite-sized eBooks about soft skills and personal development. Written by the brightest minds in business.

bookboon.com



Kontroller om du har samme indstillinger og vælg næste. Excel skulle nu gerne selv komme med et forslag på hvad den skal medbringe af data til tabellen. Hvis Excel ikke har valgt rigtigt så vælg selv dataområde (Husk overskrifter).

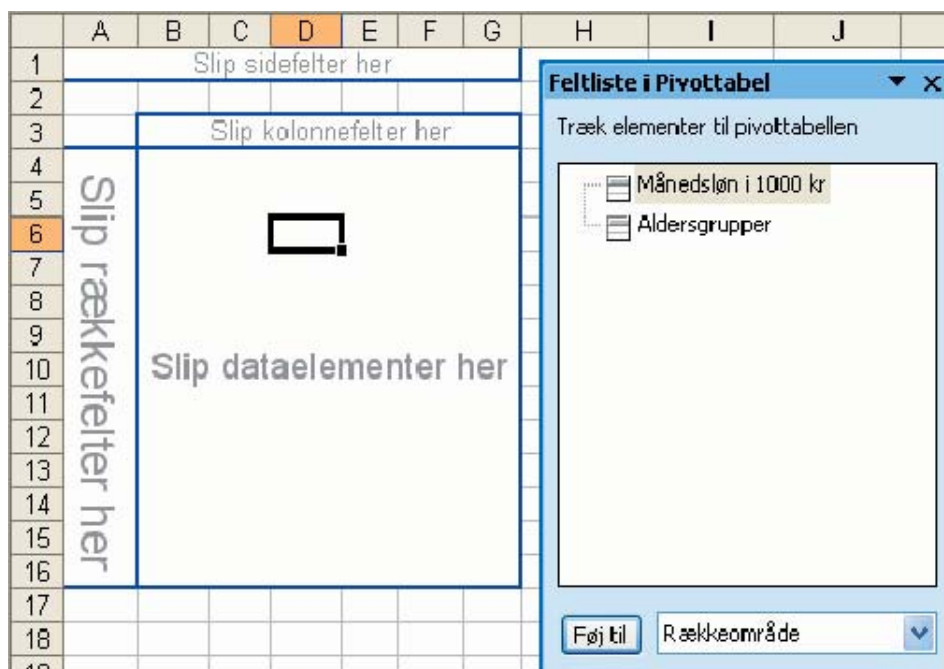
Her ønsker vi blot søjlen med Løn og søjlen med aldersgrupper.



Vælg herefter næste hvorefter sidste trin i processen fremkommer



Trin 3, kræver vist ikke nærmere forklaring, så vi afslutter med udfør. Nu fremkommer nedenstående tabel.



De manøvrer der nu udføres kunne man have udført i selve 3. trins guiden.

Først trækkes Aldersgrupper over i feltet "Slip rækkefelter her" og månedsløn trækkes over i "Slip dataelementer her"

Det vil som udgangspunkt medføre nedenstående tabel

	A	B
1	Slip sidefelter her	
2		
3	Sum af Månedsløn i 1000 kr.	
4	Aldersgrupper	Total
5		1 125,9
6		2 181
7		3 283
8	Hovedtotal	589,9

Tabellen har godt nok inddelt månedslønnen i de 3 grupper, men den udregner lige nu summen af lønningerne og vores interesse ligger i variansen. For at ændre dette dobbeltklikkes nu i cellen "Sum af Månedsløn i 1000 kr." Vælg derefter varians i den fremkomne hjælpeboks og klik OK



Derefter skulle din tabel gerne se således ud

	A	B
1		
2		
3	Varians af Månedsløn i 1000 kr.	
4	Aldersgrupper	Total
5		1 17,60166667
6		2 7,476190476
7		3 151,125
8	Hovedtotal	97,66190476

Og til vores store glæde er der nu fremkommet gruppernes varianser.

Nu kan teststørrelsen for konstant varians beregnes

Teststørrelse vedr. konstant varians fortsat:

$$Q_B = (21 - 3) \cdot \ln(66,1522) - ((6 - 1) \cdot \ln(17,6017) + (7 - 1) \cdot \ln(7,47619) + (8 - 1) \cdot \ln(151,125))$$

$$Q_B = 13,9182$$

5.4.8 Afgørelse vedr. konstant varians

$Sig.SSH = P(Q \geq 13,9182 | Q \sim \chi^2 (fg = 3 - 1))$ dvs.
 $Sig.SSH = 1 - P(Q \leq 13,9182) = 1 - 0,99905 = 0,00095$

I Excel udregnes det som: =CHIFORDELING(13,9182;2)

Da Sig.SSH er mindre end niveauet på 5% forkaster vi nulhypotesen, stikprøven tyder altså på, at der ikke er konstant varians. Det er jo et problem eftersom det var en forudsætning for modellen. Derfor ville man ofte forsøge at transformere data for at opnå konstant varians. Her vil vi blot konstatere, at der er problemer med denne forudsætning, men vi fortsætter processen.

Nu undersøges om alle middelværdierne kan være ens – dvs. om vi kan slå alle grupper sammen til én gruppe (om modellen forklarer noget af variationen i månedslønnen). Sagt på en anden måde, undersøges om lønnen er uafhængig af hvilken aldersgruppe man tilhører.

5.4.9 Hypotese vedr. ens middelværdier (alle grupper)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$$

$$H_1 : H_0^c$$

Dette svarer til at teste hypotesen

$$H_0 : \beta_2 = \beta_3 = 0$$

$$H_1 : H_0^c$$

Idet β 'erne jo nu er forskellen i middelværdien – dvs. der testes om forskellen er 0.



MTHøjgaard

**BEDRE
LØSNINGER**

I MT Højgaard insisterer vi på, at der findes en bedre løsning. Vi udvikler og anvender metoder og teknologier, der sætter nye standarder for bygge- og anlægsbranchen. Vi har fokus på hele tiden at videreudvikle vores medarbejdere, så vi gennem nye teknologier og nye samarbejdsformer kan transformere bygge- og anlægsbranchen. Vil du med på holdet?

mth.dk/vorestilgang



5.4.10 Teststørrelse vedr. ens middelværdier (alle grupper)¹¹

$$V = F = \frac{S_{\text{Regression}}^2}{S_{\text{Residual}}^2} = \frac{S_{\text{Betwen}}^2}{S_{\text{Within}}^2} \sim F(fg1 = k - 1, fg2 = n - k)$$

$$V = F = \frac{381,24881}{66,1522} = 5,7632, \text{ som kan aflæses direkte i outputtet fra Excel.}$$

5.4.11 Afgørelse vedr. ens middelværdier (alle grupper)

$\text{Sig.SSH} = P(V \geq 5,7632 | V \sim F(fg1 = 3 - 1, fg2 = 21 - 3))$ dvs.

$$\text{Sig.SSH} = 1 - P(V \leq 5,7632) = 1 - 0,98837 = 0,01163$$

I Excel findes den som: =FFORDELING(5,7632;2;18)

Da Sig.SSH er mindre end niveauet på 5% forkastes nulhypotesen, stikprøven tyder altså på, at der er mindst en middelværdi, der skiller sig ud.

Man kunne så overveje om enkelte middelværdier kunne være ens, så vi på denne måde kunne ligge enkelte grupper sammen.

Eksempelvis kan man undersøge om der er signifikant forskel i lønnen i gruppe1 og gruppe2.

5.4.12 Hypotese vedr. to ens middelværdier

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Hvilket svarer til

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

5.4.13 Teststørrelse vedr. to ens middelværdier¹²

$$T = \frac{\hat{\beta}_j}{\hat{SD}(\hat{\beta}_j)} \sim t(fg = n - k)$$

$$t = \frac{4,8738}{4,525} = 1,07708 \text{ som kan aflæses direkte fra Excel outputtet.}$$

5.4.14 Afgørelse vedr. to ens middelværdier

$\text{Sig.SSH} = 2 \cdot P(T \geq 1,07708 | T \sim t(fg = 21 - 3))$ dvs.

$$\text{Sig.SSH} = 2 \cdot (1 - P(T \leq 1,07708)) = 2 \cdot (1 - 0,85217) = 0,29566$$

Som også kan aflæses direkte i Excel-outputtet.

I Excel udregnes det som: =TFORDELING(1,0770831;18;2)

Da Sig.SSH er større end niveauet på 5%, accepteres nulhypotesen, man kan altså ikke afvise, at middelværdierne i gruppe1 og gruppe2 kan være ens.

Fodnoter

¹ Jf. Overø s. 189

² Jf. Overø s. 254

³ Jf. Overø s. 222

⁴ Denne nulhypotese og alternativhypotese kan udvides til vilkårlig mange forklarende variable

⁵ Jf. Overø s. 266

⁶ Jf. Overø s. 267

⁷ Vær opmærksom på, at udregningen udelukkende gælder såfremt større end værdien angivet i nulhypotesen – Altså her større end nul. Såfremt værdien er mindre end nul, vendes tegnet ”større end” så det bliver til ”mindre end”

⁸ Jf. Overø s. 272

⁹ Jf. Overø s. 276

¹⁰ Jf. Overø s. 277

¹¹ Jf. Overø s. 279

¹² Vær opmærksom på at denne teststørrelse kun kan bruges såfremt den ene gruppe er referencegruppen

– Jf. Overø s. 279



Ses vi til DSE-Aalborg?

Kom forbi vores stand den
9. og 10. oktober 2019.

Vi giver en is og fortæller
om jobmulighederne hos
os.

banedanmark

