Appendiks 2

# Teknikker til analyse af tal med Excel

Dette appendiks indeholder mange gentagelser fra kapitel 10, afsnit 4 "Teknikker til analyse af tal" i *Den skinbarlige virkelighed*) dog med den forskel, at appendikset indeholder vejledning i, hvordan disse teknikker kan benyttes ved hjælp af regnearket i Microsoft Excel, Microsoft Office Professional plus 2010.

## Univariat analyse

Betragter du alene én variabel og dens fordeling på undersøgelsesenhederne, er du i gang med den simpleste form for kvantitativ analyse, det der betegnes **univariat analyse**.

### Frekvensfordeling

Det kan være en **frekvensfordeling**, du kan udtrykke i absolutte tal eller procenter. Det kan fx være en opgørelse over, hvor mange i en spørgeskemaundersøgelse der har svaret i de enkelte svarkategorier på hvert spørgsmål. Denne analyseform har et rent beskrivende formål.

På næste side ser du et rådatasæt, der er organiseret i et Excel-regneark:

**Figur 1**: Udsnit af regneark med respondenters besvarelser på en spørgeskemaundersøgelse af trivsel i en større dansk virksomhed.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR | AS | AT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Svar | Køn | samlet antal år ansat i Virksomheden | alder | stilling i Virksomheden | Hvor er du ansat | Hvor tilfreds er du med at være ansat i Virksomheden | Hvor tilfreds er du med din nuværende funktion i Virksomheden | Hvor tilfreds er du med de køretøjer du bruger i dit arbejde | Hvor tilfreds er du med det øvrige materiel du bruger i dit arbejde | Hvor tilfreds er du med de tekniske/elektroniske hjælpemidler, du br | Hvor tilfreds er du med de fysiske forhold på din arbejdsplads | Hvor tilfreds er du med den uddannelse, Virksomheden har givet dig i | Hvor tilfreds er du med dine muligheder for efter- og videreuddannel | Hvor tilfreds er du med Virksomhedens nuværende områdestruktur | Hvor tilfreds er du med den måde denne strukturændring er blevet genn | Hvor tilfreds er du med Virksomhedens nuværende servicecentralstrukt | Hvor tilfreds er du med den måde denne strukturændring er blevet gen | Hvor tilfreds er du med det selvstændige ansvar, du har i det arbejd | Hvor tilfreds er du med den generelle information, du får om Virksom | Hvor tilfreds er du med den generelle information, du får gennem Vir | Hvor tilfreds er du med den lokale information du får om Virksomhede | Hvor enig er du i følgende påstand: Jeg føler, jeg har de nødvendige | Hvor enig er du i følgende påstand: de forslag og synspunkter, jeg ko | Hvor enig er du i følgende påstand: de forslag og synspunkter, jeg k | Hvor enig er du i følgende påstand: jeg ved klart, hvilke befejelser | Hvor enig er du i følgende påstand: jeg synes, at min nærmeste leder | Hvor enig er du i følgende påstand: jeg føler at min nærmeste leder | Hvor enig er du i følgende påstand: jeg får ros af min leder for vel | Hvor enig er du i følgende påstand: jeg er tilfreds med den måde min | Hvor enig er du i følgende påstand: jeg føler, jeg har tilstrækkelig | Hvor enig er du i følgende påstand: jeg føler, jeg har for meget bes | Hvor enig er du i følgende påstand: sammenholdet mellem mine nærmest | Hvor enig er du i følgende påstand: på min arbejdsplads er omgangsto | Hvor enig er du i følgende påstand: jeg synes der er en klar linie i | Hvor enig er du i følgende påstand: jeg synes mig mere tilfreds med | Hvor enig er du i følgende påstand: jeg føler at kvaliteten af den s | Hvor enig er du i følgende påstand: jeg føler at den service jeg per | Hvor enig er du i følgende påstand: jeg føler mig tryg i min ansætte | Hvor enig er du i følgende påstand: selvom jeg havde mulighed for de | Hvor godt er samarbejdet mellem dig og – servicegruppen | Hvor godt er samarbejdet mellem dig og – deltidsansatte | Hvor godt er samarbejdet mellem dig og –ansatte på servicecentralern | Hvor godt er samarbejdet mellem dig og –nærmeste ledelse og administ | Hvor godt er samarbejdet mellem dig og – regional ledelse og adminis | Hvor godt er samarbejdet mellem dig og – central ledelse og administ | A |
| 2 | 1 | 1 | 2 | 3 | 16 | 47 | 1 | 1 | 1 | 1 | 0 | 2 | 2 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 0 | 0 | 2 | 2 | 2 | 2 | 1 | 3 | 1 | 2 | 2 | 3 | 1 | 2 | 1 | 1 | 5 | 1 | 1 | 3 | 3 | |
| 3 | 2 | 2 | 4 | 4 | 15 | 26 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 4 | 4 | 4 | 4 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 4 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | | |
| 4 | 3 | 2 | 1 | 2 | 15 | 32 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 4 | 4 | 3 | 3 | 1 | 2 | 4 | 2 | 3 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 5 |
| 5 | 4 | 2 | 3 | 3 | 10 | 0 | 2 | 2 | 2 | 0 | 3 | 3 | 2 | 3 | 4 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 1 | 2 | 2 | 3 | 4 | 4 | 2 | 3 | 3 | 2 | 4 | 1 | 2 | 3 | 3 |
| 6 | 5 | 2 | 2 | 3 | 15 | 30 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 5 | 5 |
| 7 | 6 | 2 | 4 | 4 | 15 | 30 | 1 | 3 | 2 | 3 | 2 | 2 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 2 | 2 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 4 | 3 | 1 | 1 | 1 | |
| 8 | 7 | 2 | 2 | 2 | 15 | 29 | 1 | 2 | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 3 | 3 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 5 | 3 | 3 | 3 | 3 |
| 9 | 8 | 1 | 2 | 2 | 2 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | 3 | 3 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 1 | 2 | 2 | 2 | 3 | 0 | 0 | 2 | 4 | 5 | 5 | 5 | 3 | 2 | 2 |
| 10 | 9 | 2 | 4 | 4 | 15 | 44 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 2 | 2 | 4 | 4 | 4 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 4 | 4 | 4 | 1 | 3 | 3 | 0 | 0 | 0 | 0 | 5 | 5 | |
| 11 | 10 | 2 | 1 | 2 | 18 | 35 | 2 | 2 | 0 | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 5 | 1 | 5 | 5 | |
| 12 | 11 | 2 | 1 | 1 | 19 | 2 | 1 | 1 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 1 | 1 | 1 | 1 | 4 | 2 | 1 | 2 | 1 | 3 | 2 | 2 | 3 | 1 | 1 | 5 | 1 | 1 | 5 | 5 |
| 13 | 12 | 2 | 2 | 2 | 15 | 47 | 1 | 1 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 4 | 4 | 3 | 1 | 2 | 2 | 3 | 4 | 3 | 1 | 2 | 2 | 2 | 5 | 2 | 3 | 5 | 5 |
| 14 | 13 | 2 | 2 | 3 | 15 | 47 | 1 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 3 | 2 | 4 | 4 | 3 | 3 | 2 | 2 | 1 | 2 | 3 | 2 | 3 | 1 | 2 | 1 | 2 | 2 | 3 | 3 | 5 | 5 |
| 15 | 14 | 2 | 2 | 3 | 15 | 29 | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 1 | 2 | 2 | 3 | 1 | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 3 | 1 | 3 | 2 | 2 | 5 | 1 | 2 | 5 | 5 |
| 16 | 15 | 2 | 2 | 2 | 15 | 29 | 2 | 2 | 3 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 3 | 4 | 0 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 5 | 5 |
| 17 | 16 | 2 | 1 | 3 | 5 | 1 | 1 | 1 | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 3 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 3 | 1 | 1 | 1 | 3 | 2 | 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 5 |
| 18 | 17 | 2 | 4 | 0 | 13 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 3 | 2 | 2 | 3 | 1 | 1 | 2 | 3 | 3 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 5 |
| 19 | 18 | 2 | 3 | 3 | 15 | 33 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 3 | 3 | 4 | 4 | 3 | 2 | 4 | 3 | 1 | 1 | 3 | 3 | 4 | 1 | 3 | 2 | 1 | 1 | 1 | 3 | 5 | 5 |
| 20 | 19 | 2 | 3 | 2 | 15 | 31 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 4 | 1 | 2 | 2 | 3 | 4 | 3 | 2 | 4 | 4 | 3 | 3 | 2 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 3 | 2 | 3 | 5 | 5 |
| 21 | 20 | 2 | 2 | 2 | 15 | 29 | 3 | 1 | 2 | 3 | 4 | 3 | 2 | 2 | 4 | 4 | 3 | 3 | 1 | 1 | 1 | 3 | 4 | 2 | 3 | 3 | 4 | 2 | 3 | 1 | 2 | 3 | 1 | 1 | 4 | 4 | 4 | 4 | 2 | 4 | 3 | 1 | 5 | 2 | 4 | 4 | 4 |
| 22 | 21 | 2 | 3 | 3 | 15 | 32 | 2 | 1 | 4 | 3 | 2 | 4 | 3 | 3 | 2 | 2 | 3 | 4 | 2 | 1 | 1 | 4 | 2 | 3 | 3 | 1 | 3 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 2 | 4 | 2 | 4 | 4 | 2 | 1 | 3 | 2 | 2 | 5 | |

Datasættet stammer fra en trivselsundersøgelse i en større dansk virksomhed: Tallene i matrixen repræsenterer svar på et spørgeskema om trivsel. Kolonne A angiver svarpersonens nummer. Der er i alt 5418 besvarelser, men af pladshensyn er kun de 22 første besvarelser medtaget. Svarene i kolonne B står for køn. 1 betyder kvinde, og 2 betyder mand. Svarene fx i kolonne G er svarene på spørgsmålet.: "Hvor tilfreds er du med at være ansat i virksomheden?" Her står

1 for "Meget tilfreds"
2 for "Overvejende tilfreds"
3 for "Overvejende utilfreds"
4 for "Meget utilfreds"

Disse svarkategorier angiver en ordinalskala (jf. kap. 4, afsnit 5: Måleniveauer for data). En univariat analyse af disse trivselsdata kan bestå i at give en oversigt over,

hvor mange personer, der har svaret på de forskellige spørgsmål fordelt på kategorier. Det vil se således ud for spørgsmålet i kolonne G:

**Figur 2**: Frekvensfordeling af respondenters besvarelse af spørgsmålet: "Hvor tilfreds er du med at være ansat i virksomheden?"

| Tilfredshed fordelt på svarkategorier | | | | | |
|---|---|---|---|---|---|
| | Meget tilfreds | Overvejende tilfreds | Overvejende utilfreds | Meget utilfreds | Hovedtotal |
| Hvor tilfreds er du med at være ansat i Virksomheden? | 2192 | 2885 | 279 | 26 | 5382 |

Du kan se, at der i alt er 5382 af de 5418, der har svaret på dette spørgsmål. Samtidig kan du se, at de overvejende har svaret i de to første svarkategorier.

Hvis du beregner den procentvise fordeling på svarkategorierne, får du følgende tabel:

**Figur 3**: Procentvis fordeling af respondenters besvarelse af spørgsmålet: "Hvor tilfreds er du med at være ansat i virksomheden?"

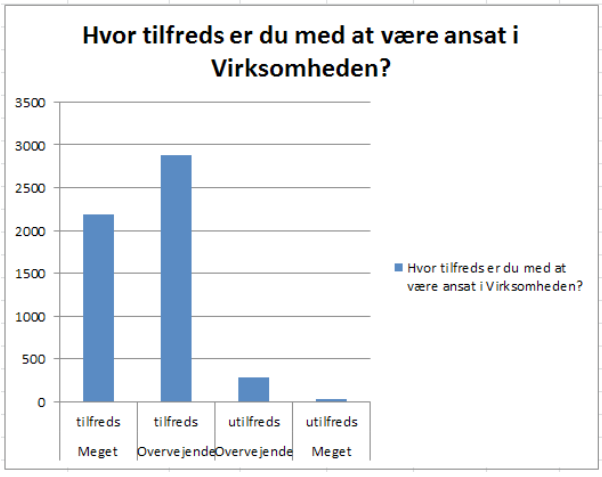
| Tilfredshed procentvis fordelt på svarkategorier | | | | | |
|---|---|---|---|---|---|
| | Meget tilfreds | Overvejende tilfreds | Overvejende utilfreds | Meget utilfreds | Hovedtotal |
| Hvor tilfreds er du med at være ansat i Virksomheden? | 40,7 | 53,6 | 5,2 | 0,5 | 100,0 |

Med andre ord svarer mere end 94 % af respondenterne positivt på dette spørgsmål.

*Søjlediagram*

Grafisk kan du udtrykke dette i et søjlediagram:

**Figur 4**: Frekvensfordeling af respondenters besvarelse: Søjlediagram



Dette kan ske i Excel ved at markere tabellens indmad inklusive kolonne og rækkenavne og derefter klikke på fanen "Indsæt", vælge funktionen "Diagrammer" og derefter klikke på "Søjle".

## Gennemsnit

Hvis du har tal, der er skalerbare på mindst intervalskalaniveau, kan resultatet af en univariat analyse være et gennemsnitstal. Se på nedenstående datasæt:

**Figur 5**: Påbegyndte boliger og salg fordelt over kvartaler

| År | Kvartal | Påbeg t-1 | Salg t |
|---|---|---|---|
| 1 | 1 | 11300 | 142 |
| | 2 | 9000 | 98 |
| | 3 | 6000 | 90 |
| | 4 | 5200 | 70 |
| 2 | 1 | 4800 | 78 |
| | 2 | 7000 | 93 |
| | 3 | 7000 | 90 |
| | 4 | 8000 | 111 |
| 3 | 1 | 11000 | 114 |
| | 2 | 14000 | 110 |
| | 3 | 7500 | 96 |
| | 4 | 8000 | 94 |
| 4 | 1 | 9000 | 99 |
| | 2 | 9000 | 90 |
| | 3 | 8500 | 100 |
| | 4 | 9000 | 99 |
| 5 | 1 | 9200 | 93 |
| | 2 | 9200 | 100 |

Kilde: Frit efter Vejrup-Hansen (2008).

Matrixen viser en oversigt over en virksomheds salg (Y) fordelt over kvartaler i 4,5 år. X-variablen viser et indekstal for antal påbegyndte boliger kvartalet før. Hvis vi i første omgang kun ser på salget, kan vi se, at det varierer over hele perioden mellem 70 som det laveste og 142 som det højeste salg. Vi kan udregne det gennemsnitlige salg over hele perioden ved at tage et aritmetisk gennemsnit af salgstallene ved at lægge alle salgstallene sammen og dividere med antallet af kvartaler. Så når vi frem til det gennemsnitlige salg pr. kvartal, som er 98,17 ved hjælp af formlen:

gennemsnit = $\Sigma x_i/n$, hvor $x_i$ er alle salgstal, som lægges sammen og divideres med antal observationer n (=18).

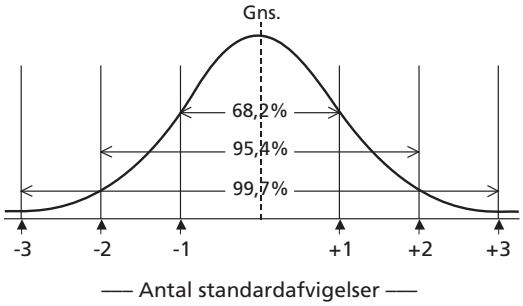
## Spredning

Hvis vi ønsker at vide, hvor meget salgstallene spreder sig om dette gennemsnit, kan vi beregne spredningen. Formlen for spredning er følgende:

$$\hat{\sigma} = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

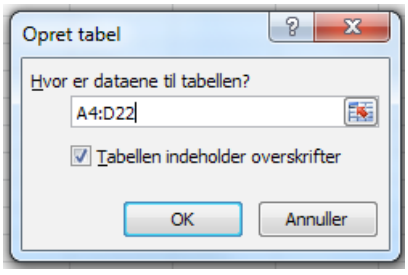
hvor "n" er antallet af observationer – i dette tilfælde 18. $x_i$ angiver de 18 forskellige observationer af "salg t", og x-streg er gennemsnittet, som vi fandt ovenfor. Formlen siger, at differencerne mellem gennemsnittet og de enkelte salgsobservationer beregnes og kvadreres, hvorefter de lægges sammen og divideres med n-1(=17), og endelig tages kvadratroden af den fremkomne størrelse. I dette tilfælde kommer vi frem til en spredning på 15,26. Tager vi spredningen og ganger med 2 – og henholdsvis trækker den fra og lægger den til gennemsnittet - får vi følgende to værdier: 67,65 og 128,69. Hvis fordelingen er tilnærmelsesvis normalfordelt, ved vi, at vi har godt 95 af den samlede fordeling mellem disse to tal – se nedenstående illustration:

**Figur** 6: Spredningen (også kaldet standardafvigelsen) i en normalfordeling




Kilde: Per Vejrup-Hansen (2012).

I Excel beregnes gennemsnit og spredning lettest ved at omdanne datasættet til en Excel-tabel. Vælg fanen "Indsæt" og klik på ikonet "Tabel" ude til venstre i fanen. Du får nu følgende dialogboks frem:



Du angiver i feltet, hvor tabellen skal hente sine data. Husk at medtage overskrifter og sæt flueben i "Tabellen indeholder overskrifter". Klik på "OK": Vi får nu en Excel-tabel, der ser sådan ud:

**Figur 7**: Påbegyndte boliger og salg fordelt over kvartaler: Tabel

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | År | Kvartal | Påbeg t- | Salg | |
| 2 | 1 | 1 | 11300 | 142 | |
| 3 | | 2 | 9000 | 98 | |
| 4 | | 3 | 6000 | 90 | |
| 5 | | 4 | 5200 | 70 | |
| 6 | 2 | 1 | 4800 | 78 | |
| 7 | | 2 | 7000 | 93 | |
| 8 | | 3 | 7000 | 90 | |
| 9 | | 4 | 8000 | 111 | |
| 10 | 3 | 1 | 11000 | 114 | |
| 11 | | 2 | 14000 | 110 | |
| 12 | | 3 | 7500 | 96 | |
| 13 | | 4 | 8000 | 94 | |
| 14 | 4 | 1 | 9000 | 99 | |
| 15 | | 2 | 9000 | 90 | |
| 16 | | 3 | 8500 | 100 | |
| 17 | | 4 | 9000 | 99 | |
| 18 | 5 | 1 | 9200 | 93 | |
| 19 | | 2 | 9200 | 100 | |
| 20 | | | | | |
| 21 | | | | | |

Det anbefales generelt, at du fra begyndelsen omdanner din matrix til en tabel. Det medfører bl.a., at afgrænsningen bliver dynamisk, så du fx kan indsætte eller slette rækker og kolonner uden at skulle revidere formler og funktioner.

Anbring nu din markør i en tilfældig tabelcelle. Der fremkommer nu en fane "Tabelværktøjer – Design". Vælg fanen og sæt flueben i "Rækken total" lige nedenfor. Nu tilføjes en række med "Total". Anbring din markør i denne række i kolonnen "Salg" sådan her:

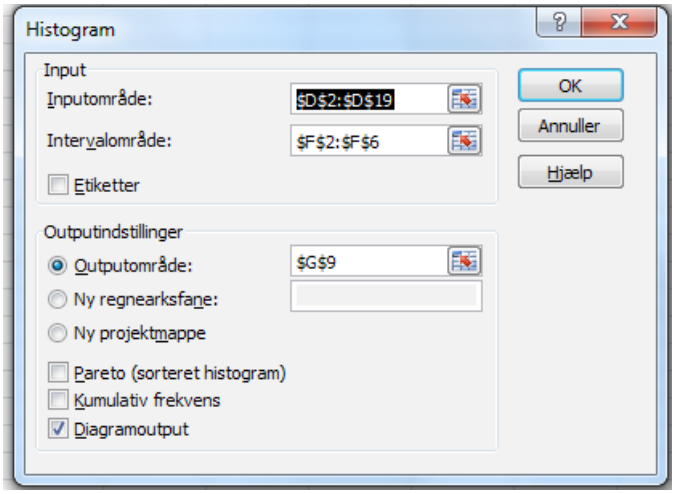
| | | | | |
|---|---|---|---|---|
| 17 | | 4 | 9000 | 99 |
| 18 | 5 | 1 | 9200 | 93 |
| 19 | | 2 | 9200 | 100 |
| 20 | Total | | | |
| 21 | | | | |

Tryk på den lille pil til højre for cellen. Du får nu en række muligheder. Tryk på hhv. "Gennemsnit" og "Stdafv". Den sidste angiver spredningen omkring gennemsnittet. Prøv evt. selv nogen af de øvrige beregningsfunktioner i menuen for at se, hvad de indeholder.

## Histogram

Ønsker du at udtrykke salget grafisk, kan du bruge funktionen "Histogram". Gå ind i fanen "Data" og vælg funktionen "Dataanalyse" helt ude til højre. Hvis denne funktion ikke kan ses i din fane, skal du aktivere den. Du finder fremgangsmåden til denne aktivering her: [http://office.microsoft.com/da-dk/excel-help/hurtig-start-aktivere-og-bruge-et-tilfojelsesprogram-HA010370161.aspx].

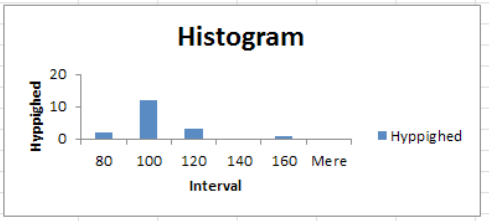
Men inden du bruger funktionen "Histogram" skal Excel have at vide, i hvilke intervaller dine data skal ordnes. Ved siden af "Salg t" skiver du "Interval", og nedenunder angiver du intervalinddelingen. Vælg fanen "Data", dernæst "Dataanalyse", marker "Histogram" og klik "OK". Følgende dialogboks kommer op:




Angiv først salgstallene i "Inputområde", angiv "Intervalområde", sæt flueben i "Diagramoutput" og klik "OK". Du får nu et histogram som her:

**Figur 8**: Påbegyndte boliger og salg: Histogram

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | År | Kvartal | Påbeg t- | Salg | | Intervaller | | | | |
| 2 | 1 | 1 | 11300 | 142 | | 80 | | Interval | Hyppighed | |
| 3 | | 2 | 9000 | 98 | | 100 | | 80 | 2 | |
| 4 | | 3 | 6000 | 90 | | 120 | | 100 | 12 | |
| 5 | | 4 | 5200 | 70 | | 140 | | 120 | 3 | |
| 6 | 2 | 1 | 4800 | 78 | | 160 | | 140 | 0 | |
| 7 | | 2 | 7000 | 93 | | | | 160 | 1 | |
| 8 | | 3 | 7000 | 90 | | | | Mere | 0 | |
| 9 | | 4 | 8000 | 111 | | | | | | |
| 10 | 3 | 1 | 11000 | 114 | | | | | | |
| 11 | | 2 | 14000 | 110 | | | | | | |
| 12 | | 3 | 7500 | 96 | | | | | | |
| 13 | | 4 | 8000 | 94 | | | | | | |
| 14 | 4 | 1 | 9000 | 99 | | | | | | |
| 15 | | 2 | 9000 | 90 | | | | | | |
| 16 | | 3 | 8500 | 100 | | | | | | |
| 17 | | 4 | 9000 | 99 | | | | | | |
| 18 | 5 | 1 | 9200 | 93 | | | | | | |
| 19 | | 2 | 9200 | 100 | | | | | | |
| 20 | Total | | | | | | | | | |
| 21 | | | | | | | | | | |




Beregning af gennemsnit, spredning samt den grafiske fremstilling af tallene giver en god oversigt over salgsfordelingen over kvartalerne.

## Bivariat analyse

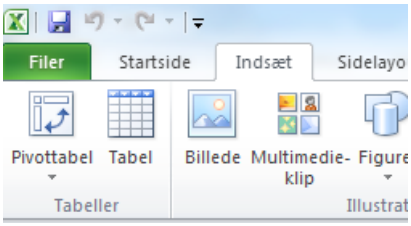
Den **bivariate analyse** beskæftiger sig med to variabler og deres indbyrdes samvariation.

### Krydstabeller

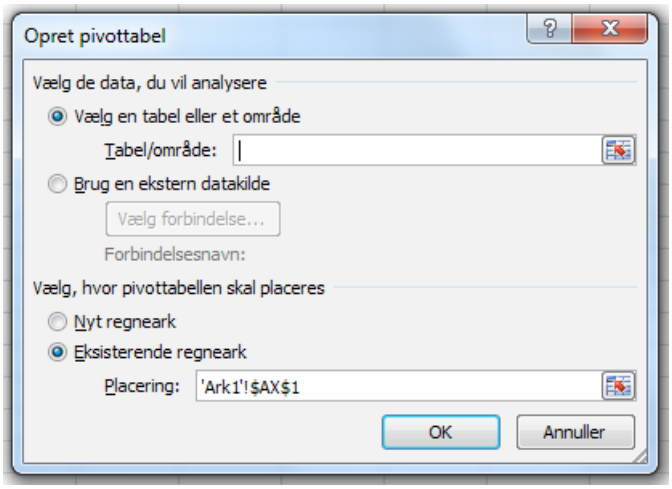
For at undersøge en sådan samvariation kan du krydse de to variabler. Det sker ved at sammenholde værdierne på de to variabler for hver enkelt undersøgelsesenhed. Resultaterne kan angives i en **krydstabel**.

Du kan tage udgangspunkt i et Excel-regneark eller et regneark, der er omdannet til Excel-tabel til at danne krydstabeller mellem de forskellige variabler to og to med den funktion, der hedder "Pivottabel". Tager du udgangspunkt i den ovenfor refererede matrix med trivselsdata, kan du se, at der er i alt 45 spørgsmål i spørgeskemaet. Det vil sige, at du kan producere 45x44/2 krydstabeller (n(n-1)/2) = 990 krydstabeller. Det er et meget stort arbejde, hvis dette skal gøres i Excel uden et særligt redskab. Men her kan du betjene dig af den funktion i Excel, der hedder "Pivottabel".

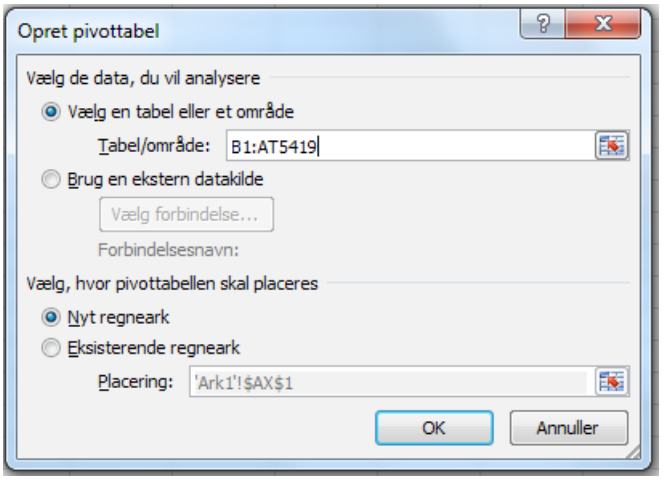
I Excel finder du pivottabel-funktionen i fanen "Indsæt" helt ude til venstre:




Klik på ikonet ”Pivottabel” og følgende dialogboks ”Opret pivottabel” kommer op:



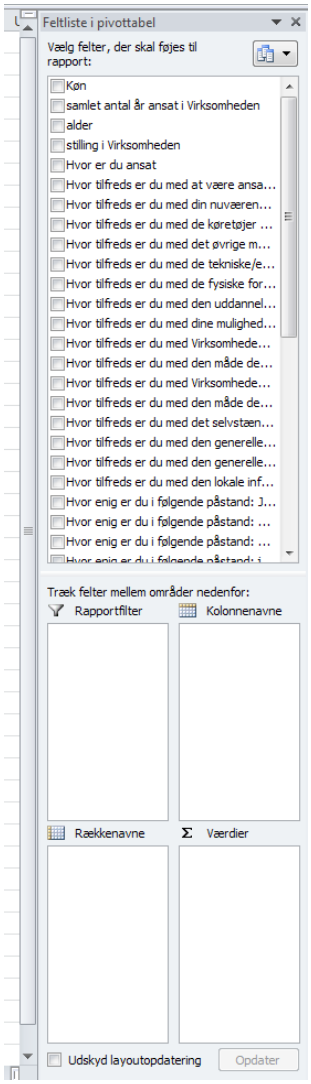

Du skal nu vælge Tabel/område, hvor Pivottabellen skal hente sine data. Du kan med fordel vælge hele datasættet fra variablen ”Køn” til den sidste variabel i kolonne ”AT”. Så du vælger området fra B2 til AT5419. Derefter skal du vælge, om pivottabellen skal dannes i det ”Eksisterende regneark” eller i et ”Nyt regneark”. Du vælger ”Nyt regneark”. Nu ser dialogboksen sådan ud:




Vælg"OK". Nu kommer følgende skærmbillede op i venstre side af skærmen:


Dette skærmbillede viser, hvor tabellen bliver dannet. Her vil du kunne følge, hvordan din tabel ser ud og ændre indhold, alt efter hvilke variabler du vælger hhv. til tabellens forspalte (rækker) og tabellens hoved (kolonner).

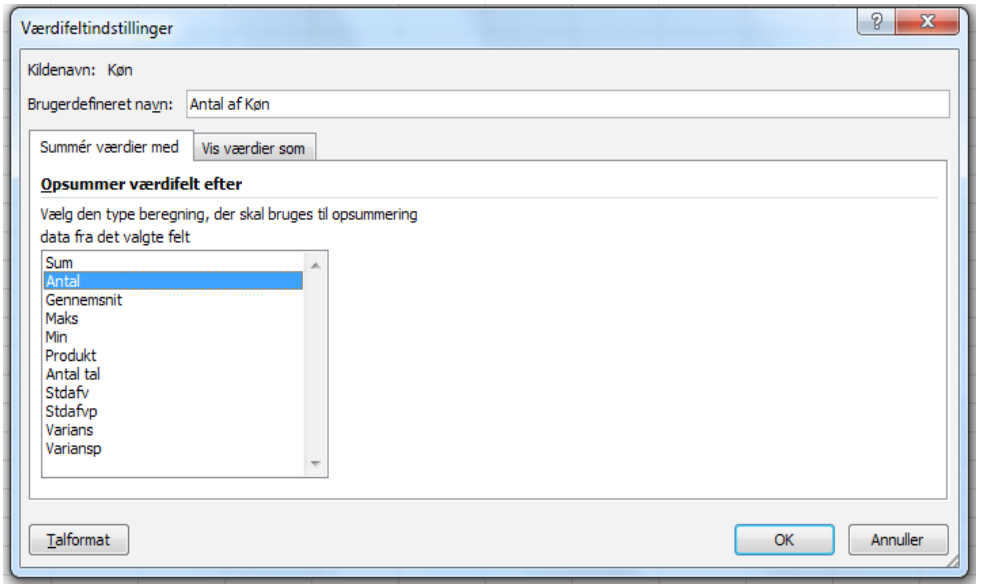
I højre side af skærmen er der en menu, der ser sådan ud:




Dette område angiver de variabler, du kan placere hhv. i forspalten (”Rækkenavne”) og i tabelhovedet (”Kolonnenavne”). Du vælger variabler ved at trække dem fra variabeloversigten ned til felterne hhv. ”Rækkenavne” og ”Kolonnenavne”. Du vælger variablen ”Køn” og trækker den ned i feltet ”Rækkenavne”, og du vælger variablen nr. seks fra oven: ”Hvor tilfreds er du med at være ansat i virksomheden”, og trækker den ned i feltet ”Kolonnenavne”. Til sidst skal du vælge en vilkårlig variabel, som du trækker ned i ”∑ Værdier”. Hvis datamatrixen indeholder tal som variabelværdier, som de gør her, skal du også vælge, om værdierne skal lægges sammen, eller om det

blot er antallet af hver værdi der skal findes. Excel tager automatisk sumværdien, men du er interesseret i antallet. Derfor må du klikke på den lille pil ved siden Sum-variablen.


Her er valgt køn som sumvariabel. Hvis du klikker på den lille pil til højre, får du en dialogboks op. Du klikker på det sidste emne i dialogboksen "Værdiindstillinger…". Herved fremkommer en ny dialogboks som ser således ud:




Du klikker på "Antal" og derefter "OK". Du får nu en pivottabel frem, der ser således ud:

**Figur 9**: Køn fordelt på trivsel (1)

| Antal af Køn | Kolonnenavne | | | | | |
|---|---|---|---|---|---|---|
| Rækkenavne | 0 | 1 | 2 | 3 | 4 | Hovedtotal |
| 0 | 5 | 51 | 62 | 6 | 1 | 125 |
| 1 | 7 | 275 | 316 | 43 | 8 | 649 |
| 2 | 24 | 1866 | 2507 | 229 | 17 | 4643 |
| 3 | | | | 1 | | 1 |
| Hovedtotal | 36 | 2192 | 2885 | 279 | 26 | 5418 |

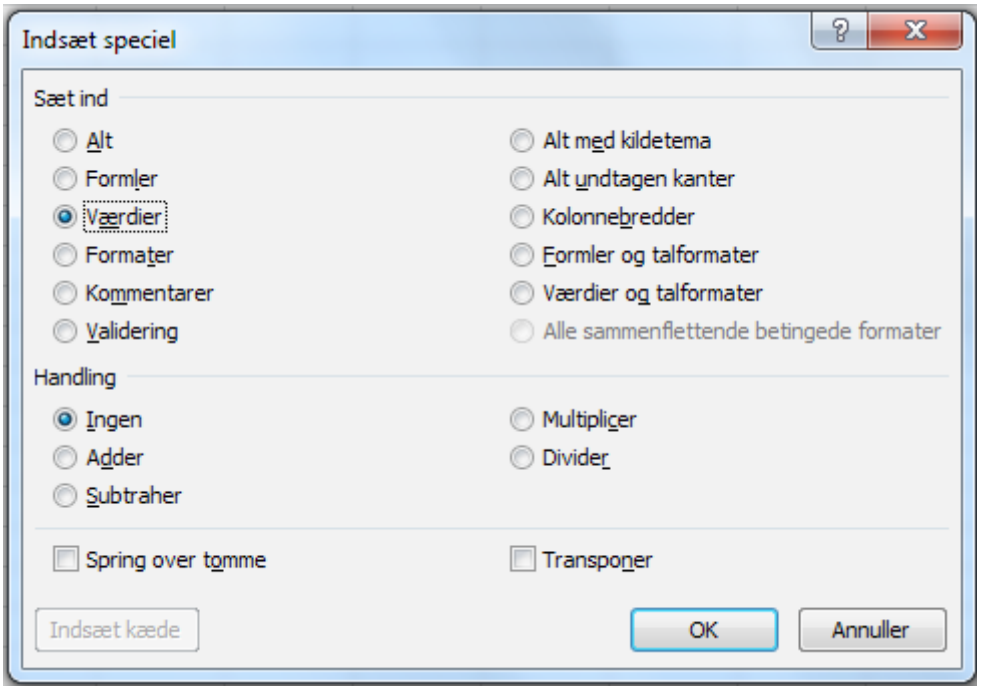
0

Her angiver 0 angiver, at spørgsmålet ikke er besvaret. Der alt 125 personer, der ikke har svaret på spørgsmålet om køn. Vi sletter derfor denne række. Det fremgår, at en person har kønnet 3, men det eksisterer ikke. Det må være en fejl ved udfyldelse eller kodning. Så den række sletter vi også. Det fremgår af tabellens kolonne ”0”, at der er 36 personer, der ikke har svaret på trivsels-spørgsmålet. Du fjerner rækker og kolonner ved at trykke på den lille pil ud for rækkenavne hhv. kolonne navne. Der fremkommer så en menu, hvor alle rækker hhv. kolonner er angivet med et flueben. Fjerner du fluebenet i dem, du ikke ønsker at have med fjerner du hhv. uønskede rækker og kolonner. Du klikker bare i fluebenet, så forsvinder det. Derefter klikker du på ”OK”. Du får nu en tabel, der ser sådan ud:

**Figur 10**: Køn fordelt på trivsel (2)

| Antal af Køn | Kolonnenavne | | | | |
|---|---|---|---|---|---|
| Rækkenavne | 1 | 2 | 3 | 4 | Hovedtotal |
| 1 | 275 | 316 | 43 | 8 | 642 |
| 2 | 1866 | 2507 | 229 | 17 | 4619 |
| Hovedtotal | 2141 | 2823 | 272 | 25 | 5261 |

Du er ikke helt tilfreds med tabellen. Hvis du vil redigere række- og kolonnenavne, skal du tage en kopi af pivottabellen og indsætte den med ”Indsæt speciel”. Den funktion finder du, hvis du højreklikker på det felt, hvor tabellen skal indsættes. Du får en dialogboks op, hvor du vælger ”Indsæt speciel”. Der kommer nu en ny dialogboks op, som ser sådan ud:




Du vælger "Værdier", som du ser ovenfor. Derefter klikker du på "K". Du har nu en tabel, som ikke har bindinger til den oprindelige pivottabel. Der for kan du redigere tabellen som du vil. Eventuelt med et sådant udseende:

**Figur 11**: Køn fordelt på trivsel (3)

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Køn fordelt på trivsel** | | | | | |
| 2 | | | | **Trivsel** | | |
| 3 | | Meget | Overvejende | Overvejende | Meget | **Hovedtotal** |
| 4 | **Køn** | tilfreds | tilfreds | utilfreds | utilfreds | |
| 5 | Kvinde | 275 | 316 | 43 | 8 | **642** |
| 6 | Mand | 1866 | 2507 | 229 | 17 | **4619** |
| 7 | **Hovedtotal** | **2141** | **2823** | **272** | **25** | **5261** |

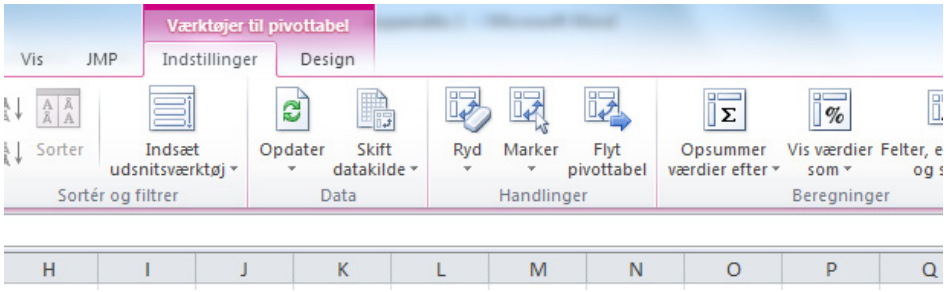
Tabellens forspalte angiver køn, og tabellens hoved angiver de fire svarmuligheder på det stillede spørgsmål om tilfredshed. Variablen "køn" udgør en nominalskala, og variablen "tilfredshed" udgør en ordinalskala (jf. kap. 4, afsnit 5). Tabellen viser, at der i alt er 5.261 personer med i tabellen (de resterende 157 er internt bortfald). Heraf er 642 kvinder og 4.619 mænd. Tabellen viser, hvordan kvinder og mænds besvarelser fordeler sig over de fire mulige svarkategorier. Hvis du ønsker at under-

søge, om der er forskel på kvinder og mænds besvarelser, kan det fx gøres ved at se på den procentvise fordeling af de to køn over de forskellige svarmuligheder. Det gøres ligeledes nemt i Excel ved hjælp af formelindsætning eller ved hjælp af "Værktøjer til pivottabel". Bruger du den første fremgangsmåde, skal du først tage en almindelig kopi af ovenstående tabel. Derefter skal du placere markøren i den kopierede tabels celle i første række og kolonne. Du kan nu bruge formlen "=B5/$F5*100 og trække igennem hele den kopierede tabel. Så får den en tabel, der ser sådan ud, hvis du har begrænset decimalerne til to og tilføjet "procentvis" i tabeloverskriften:

**Figur 12**: Køn procentvis fordelt på trivsel

| Køn procentvis fordelt på trivsel | | | | | |
|---|---|---|---|---|---|
| | Trivsel | | | | |
| | Meget | Overvejende | Overvejende | Meget | **Hovedtotal** |
| **Køn** | tilfreds | tilfreds | utilfreds | utilfreds | |
| Kvinde | 42,83 | 49,22 | 6,70 | 1,25 | 100 |
| Mand | 40,40 | 54,28 | 4,96 | 0,37 | 100 |
| **Hovedtotal** | 40,70 | 53,66 | 5,17 | 0,48 | 100 |

Bruger du den anden fremgangsmåde, skal du gå tilbage til din pivottabel. Marker tabellen og fane-menuen vil ændre sig, så der fremkommer en fane med "Værktøjer til pivotabel. Sådan her:



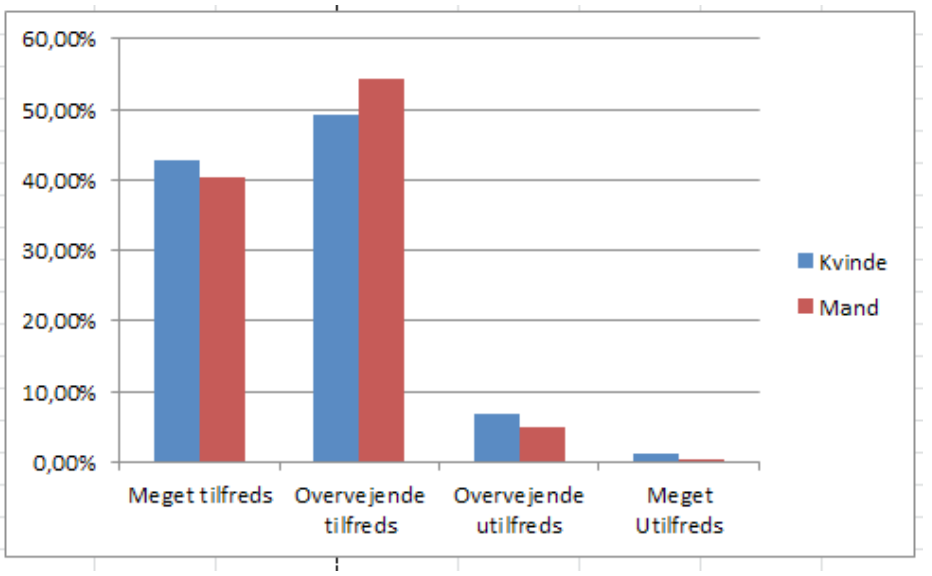

Du vælger fanen "Indstillinger" og klikker på ikonet med %-tegnet (det sidste ikon på ovenstående billede). Din pivottabel vil blive fyldt med de sammen procenter, som tabellen ovenfor. Det er meget hurtigt og nemt. Men vil du have samme design på din tabel som ovenstående, skal du igennem hele redigeringsprocessen igen. Det slipper du for ved at bruge formelmetoden.

Umiddelbart ser det ud som om, at de to køn svarer meget i overensstemmelse med hverandre.

### Søjlediagram i bivariat analyse

Dette kan tydeligt illustreres grafisk med et søjlediagram. I Excel markerer man blot tabellens indmad, køn og svarmuligheder for derefter at klikke på "Søjle" i funktionen "Diagrammer" i fanen "Indsæt". Herved fremkommer følgende søjlediagram:

**Figur 13**: Køn fordelt på trivsel: Søjlediagram




Diagrammet viser ligeledes temmelig tydeligt en meget lille variation mellem kvinder og mænds besvarelse. Samtidig viser diagrammet, at kvinders svar fordeler sig mere spredt end mændenes svar.

### Statistiske mål for samvariation mellem to variabler

Der findes en lang række statistiske mål for samvariation mellem to variabler, som du kan udtrykke med en talstørrelse eller en sandsynlighed. Der findes grundlæggende to former for disse mål: *ikke-parametriske mål* og *parametriske mål*.

**Husk, at et statistisk samvariationsmål intet siger om årsagssammenhænge!**

#### *Ikke-parametriske mål*

Disse mål betegnes også som fordelingsfrie *mål* (se kapitel 10 i *Den skinbarlige virkelighed*). Et hyppigt anvendt mål er χ2-testet. Dette test kan udføres uanset hvilken skala, vi arbejder med (nominal-, ordinal-, interval- eller forholdstalsskala). χ2-testet kan gennemføres i Excel, og den $\chi^2$-størrelse, som $\chi^2$-testet i Excel angiver, er sandsynligheden for, at der **ingen** sammenhæng er i tabellen. Med andre ord sandsynligheden for, at der er en fuldstændig ligelig fordeling mellem kvinder og mænds svar.

Excel har brug for to tabeller til at gennemføre dette test. Den ene tabel er den, der er angivet i den tidligere figur, der indeholder de absolutte tal for "Køn fordelt på trivsel". Den anden tabel er en konstrueret tabel, der angiver, hvordan fordelingen burde se ud, såfremt der var fuldstændig ligelig svarfordeling mellem kønnene (også kaldet "forventede værdier"). Denne tabel beregnes ud fra "i alt" værdierne for hhv. rækker og kolonner. For hver tabelcelle ganges deres kolonnes "Hovedtotal"-værdi med deres rækkes "Hovedtotal"-værdi – og dette tal divideres med det totale antal besvarelser (i dette tilfælde 5261 besvarelser). Dette gøres for hver celle i tabellen. I Excel kan man bruge formelfunktionen til denne beregning (jf. appendiks 1). Nedenfor vises tabellen over de forventede værdier:

**Figur 14**: Køn fordelt på trivsel: forventede værdier

| | | | | | | |
|---|---|---|---|---|---|---|
| 41 | **Køn fordelt på trivsel: forventede værdier** | | | | | |
| 42 | | | | | | |
| 43 | | **Hvor tilfreds er du med at være ansat i virksomhed X** | | | | |
| 44 | **Køn** | **Meget tilfreds** | **Overvejende tilfreds** | **Overvejende utilfreds** | **Meget Utilfreds** | **Hovedtotal** |
| 45 | Kvinde | 261,3 | 344,5 | 33,2 | 3,1 | **642** |
| 46 | Mand | 1879,7 | 2478,5 | 238,8 | 21,9 | **4619** |
| 47 | **Hovedtota** | **2141** | **2823** | **272** | **25** | **5261** |
| 48 | | | | | | |
| 49 | | | | | | |

I Excel finder vi $\chi^2$-testet på følgende måde: Gå ind i fanen "Formler" og vælg "Flere funktioner" og vælg derefter "Statistik" og endelig "chi2.test". Der fremkommer nu en dialogboks, hvor du skal angive områderne for henholdsvis "observerede værdier" og "forventede værdier", og tryk på "OK". Excel vil nu beregne sandsynligheden for, at der ingen samvariation er. Beregningen giver en værdi på 0,00116. Det vil sige, at der er godt 1 promilles sandsynlighed for, at der ikke er samvariation mellem køn og tilfredshed. Alternativt må du så slutte, at der må være en eller anden form for samvariation, men $\chi^2$-testet fortæller dig ikke, hvad det er for en eventuel samvariation, der er mellem de to variabler.

Det er jo lidt overraskende, at $\chi^2$-testet antyder, at der er en samvariation, når vi på procentfordeling og søjlediagram vurderede, at der næsten var ligelig fordeling mellem kønnenes besvarelser. Dette afslører en svaghed ved $\chi^2$-testet. Når man har så mange observationer (i dette tilfælde 5216), vil $\chi^2$-testet vise meget små sandsynligheder for, at der ikke er nogen samvariation mellem de to variabler. Havde du taget en stikprøve på 10 % af dem, der har besvaret spørgeskemaet fx hver tiende respondent (i alt 526 besvarelser), ville $\chi^2$-testet have givet en sandsynlighed på 0,6246. Altså godt 62 % sandsynlighed for, at der ingen samvariation er mellem de to variabler.

Der findes andre ikke-parametriske mål for samvariation mellem to variabler. Disse forudsætter som regel en rangordnet skala (ordinalskala). Svarkategorierne i trivselsvariabelen ovenfor udgør en ordinalskala. Disse mål beregnes ved at give data rang-værdier. Den mindste værdi får rang 1, den næstmindste rang 2 og så videre.

Der er ikke nødvendigvis lige så langt mellem 1 og 2, som der er mellem 2 og 3 osv. Dette er ikke-parametrisk statistik beregnet til at håndtere. Et meget almindeligt anvendt mål er Kendall rangordens korrelation. Som regel betegnet med det græske bogstav τ (Tau). Dette mål kan beregnes for to variabler, der er målt på ordinalskalaniveau. Målet varierer fra -1 til +1 og angiver hhv. perfekt negativ samvariation og perfekt positiv samvariation mellem to variabler. τ = 0 angiver, at der *ingen* samvariation er mellem de to variable. τ kan ikke umiddelbart beregnes ved hjælp af Excel regneark. (brug SPSS eller SAS). Hvis du ønsker at bruge dette sammenhængsmål, bør du henvende dig til en ekspert på området. Kendall τ svarer lidt til Pearson produkt-moment korrelationskoefficient "r" (se efterfølgende).

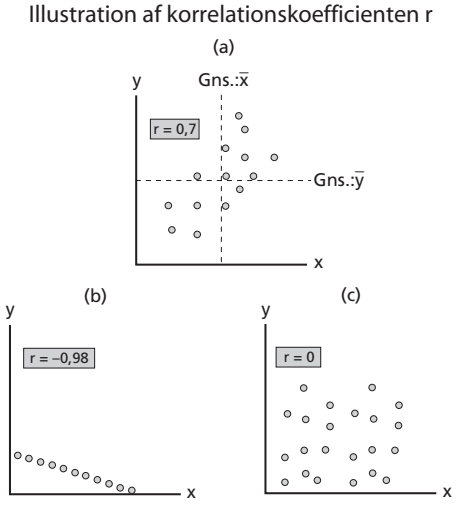
*Parametriske mål*

Der findes andre samvariationsmål, der bygger på statiske fordelinger (almindeligvis normalfordelingen).

*Korrelationsmål*

Det mest anvendte mål, hvis to stokastiske variabler x og y er normalfordelt og målt på mindst intervalniveau, er (Pearson) korrelationskoefficienten, som almindeligvis angives med bogstavet "R" eller "r". Dette mål er et mål for lineær samvariation mellem de to variabler. r varierer på samme måde som Kendall fra -1 til +1. Fortegnet angiver, om sammenhængen er positiv eller negativ. Tallet angiver, hvor godt variabelmålingerne på x og y samler sig om en ret linje. Se illustrationen nedenfor:

**Figur 15**: Grafisk illustration af forskellige korrelationskoefficienter



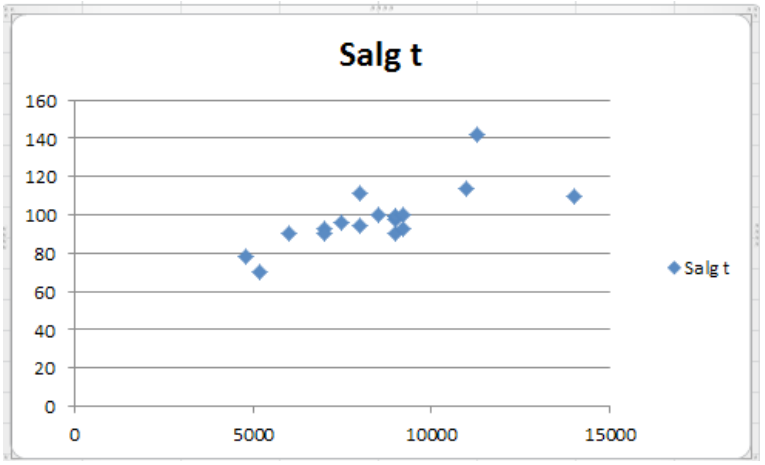
Denne koefficient er let at beregne i Excel regneark. Tag udgangspunkt i den tidligere viste tabel som her:

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | År | Kvartal | Påbeg t-1 | Salg | |
| 2 | 1 | 1 | 11300 | 142 | |
| 3 | | 2 | 9000 | 98 | |
| 4 | | 3 | 6000 | 90 | |
| 5 | | 4 | 5200 | 70 | |
| 6 | 2 | 1 | 4800 | 78 | |
| 7 | | 2 | 7000 | 93 | |
| 8 | | 3 | 7000 | 90 | |
| 9 | | 4 | 8000 | 111 | |
| 10 | 3 | 1 | 11000 | 114 | |
| 11 | | 2 | 14000 | 110 | |
| 12 | | 3 | 7500 | 96 | |
| 13 | | 4 | 8000 | 94 | |
| 14 | 4 | 1 | 9000 | 99 | |
| 15 | | 2 | 9000 | 90 | |
| 16 | | 3 | 8500 | 100 | |
| 17 | | 4 | 9000 | 99 | |
| 18 | 5 | 1 | 9200 | 93 | |
| 19 | | 2 | 9200 | 100 | |
| 20 | | | | | |
| 21 | | | | | |

Kilde: Per Vejrup-Hansen, 2012.

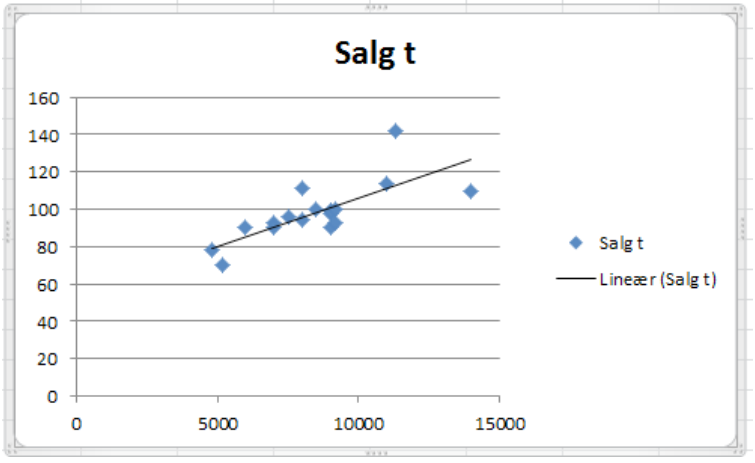
Ønsker du at se, hvordan sammenhængen er mellem salg og påbegyndt antal boliger kvartalet før, kan du grafisk gengive dette i et Excel punktdiagram. Marker de to kolonner med hhv. "Påbeg t-1" samt "Salg t". Gå ind i fanen "Indsæt", gå til "Diagrammer" og klik på "Punktdiagram". Der fremkommer nu nogle valgmuligheder for punktdiagrammer. Klik på det punktdiagram, der kun består af punkter (og ingen streger og kurver). Nu får du følgende diagram frem:

**Figur 16**: Påbegyndte boliger og salg: Punktdiagram



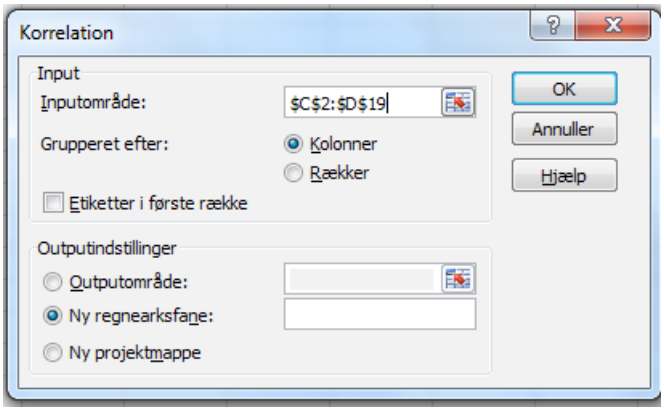

Dette diagram viser samhørende værdier for salg og antal påbegyndte boliger i kvartalet før. Der er tilsyneladende en vis positiv samvariation mellem de to variabler. Du kan lægge en tendenslinje ind i diagrammet ved at højreklikke på en af de blå punkter i diagrammet. Der fremkommer nu nogle forskellige muligheder. Du klikker på "Tilføj tendenslinje" og du får følgende diagram frem:

**Figur 17**: Påbegyndte boliger og salg: Punktdiagram med tendenslinje




Excel placerer automatisk tendenslinjen, så den bedst passer til de forskellige punkter, hvis der er en lineær samvariation.

Du kan beregne Pearsons korrelationskoefficient sådan her: Gå ind i fanen "Data", vælg "Dataanalyse", derefter "Korrelation" og klik på "OK". Derefter viser følgende dialogboks sig:




Hvor du angiver dine datas inputområde og klikker på "OK". Herefter viser Excel følgende beregning:

**Figur 18:** Korrelationsanalyse

| | A | B | C |
|---|---|---|---|
| 1 | | *Kolonne 1* | *Kolonne 2* |
| 2 | Kolonne 1 | 1 | |
| 3 | Kolonne 2 | 0,745064 | 1 |
| 4 | | | |

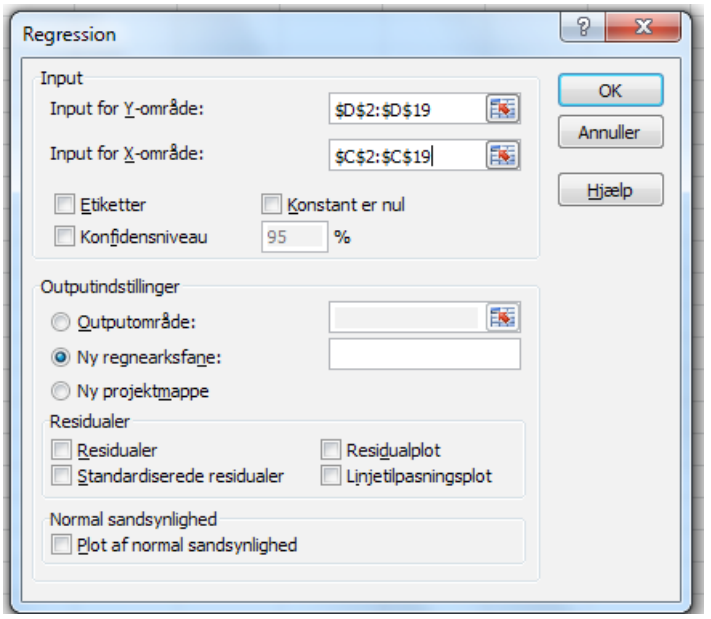
Hvor tallet ud for kolonne 1 og 2 angiver Pearsons r = 0,745064, hvilket antyder, at der er en god positiv samvariation mellem de to variable, svarende til det, vi så på punktdiagrammet.

Du kan læse mere om Pearsons og Kedalls korrelationsmål i kapitel 10 i *Den skinbarlige virkelighed*.

*Regression*

Tendenslinjens hældning kan du beregne ved simpel lineær regressionsanalyse, som også er temmelig nemt at foretage i Excel.

Gå ind i fanen "Data", vælg "Dataanalyse", derefter "Regression" og klik"OK". Derefter viser følgende dialogboks:




Hvor du indsætter områderne for Y og X, hvor X angiver den såkaldte uafhængige variabel og Y den afhængige variable, hvilket vil sige, at din antagelse er, at antal påbegyndte boliger forårsager et bestemt salg plus minus tilfældige variationer. Klikker du på "OK", får du følgende beregning:

**Figur 19:** Regressionsanalyse

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | RESUMEOUTPUT | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *Regressionsstatistik* | | | | | | | | |
| 4 | Multipel R | 0,745064 | | | | | | | |
| 5 | R-kvadreret | 0,555121 | | | | | | | |
| 6 | Justeret R-kvadreret | 0,527316 | | | | | | | |
| 7 | Standardfejl | 10,49388 | | | | | | | |
| 8 | Observationer | 18 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | ANAVA | | | | | | | | |
| 11 | | *fg* | *SK* | *MK* | *F* | *ignifikans F* | | | |
| 12 | Regression | 1 | 2198,557 | 2198,557 | 19,96484 | 0,000388 | | | |
| 13 | Residual | 16 | 1761,943 | 110,1215 | | | | | |
| 14 | I alt | 17 | 3960,5 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Koefficientetandardfe* | | *t-stat* | *P-værdi* | *Nedre 95%* | *Øvre 95%* | *Iedre 95,0%* | *Øvre 95,0%* |
| 17 | Skæring | 54,54349 | 10,07147 | 5,415643 | 5,72E-05 | 33,19292 | 75,89405 | 33,19292 | 75,89405 |
| 18 | X-variabel 1 | 0,005142 | 0,001151 | 4,468203 | 0,000388 | 0,002703 | 0,007582 | 0,002703 | 0,007582 |
| 19 | | | | | | | | | |

Vi skal kun koncentrere os om tallene ud for række 4 "Multipel R", 17 "Skæring" og "X-variabel 1". Multipel R svarer nøjagtig til den korrelationskoefficient r, som vi fandt tidligere.

Tendenslinjen er en ret linje, som matematisk kan beskrives med denne formel:

Y = a + bX eller Salg t = a + b·Påbeg t-1, hvor a er det stykke linjen afskærer af Y-aksen, og b er hældningskoefficienten. Skæringen = 54,54349 er altså der, hvor linjen skærer Y-aksen og X-variabel 1 = 0,005142 er linjen hældning. Du kan nu udtrykke salgets afhængighed af påbegyndte boliger med denne ligning:

Slag t = 54,54 + 0,00514 · Påbeg t-1. Det vil sige, at vi kan beregne en cirka-værdi for salget, fx hvis Pabeg t-1 er fx 18000 ved at indsætte i ligningen. Dette vil så give et cirka salg på ca. 147.

Hældningskoefficienten er et udtryk for styrken af sammenhængen mellem de to variabler.

Regressionsanalyse forudsætter, at de to variabler er kontinuerte og minimum er målt på intervalskalaniveau. Ligeledes forudsættes det, at den afhængige variabel skal være normalfordelt og har samme spredning på Y for alle X.

*Logistisk regression*

I den lineære regressionsanalyse ovenfor opfattes den afhængige variabel Y som et observerbart tal. Hvis den afhængige variabel kun har to udfald fx "føler stress" og "føler ikke stress", kan vi anvende logistisk regression. Der fører for vidt her at forklare i detaljer, hvordan sådan en analyse gennemføres. En logistisk regressionsanalyse kan ikke umiddelbart gennemføres i Excel. Vil du gennemføre en sådan analyse, bør du henvende dig til en ekspert på området.

## Multivariat analyse

**Multivariat analyse** anvender du, når du ønsker at undersøge samvariationen imellem flere variabler på samme tid. Udgangspunktet for en sådan analyse er ønsket om at sandsynliggøre en eller flere årsagssammenhænge for at kunne opstille en *kausalmodel* (en model, der angiver årsagssammenhængene mellem et sæt af variabler).

### Korrelationsanalyse

En af de metoder, du kan anvende i den multiple analyse er korrelationsanalyse, som tilfældet var i den bivariate analyse. Nedenstående tabel angiver som ovenfor samvariationen mellem "Påbegyndte antal boliger t-1" og "Salg t", men nu indeholder tabellen foruden disse to variabler en tredje variabel: "Antal kvadratmeter påbegyndt boligareal t-1" illustreret i tabellen ved "KM t-1".

**Figur 20**: Påbegyndte boliger, påbegyndte kvadratmeter og salg (konstrueret)

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | År | Kvartal | Påbeg t-1 | KM t-1 | Salg t |
| 2 | 1 | 1 | 11300 | 5010 | 142 |
| 3 | | 2 | 9000 | 4200 | 98 |
| 4 | | 3 | 6000 | 3050 | 90 |
| 5 | | 4 | 5200 | 2800 | 70 |
| 6 | 2 | 1 | 4800 | 2650 | 78 |
| 7 | | 2 | 7000 | 3700 | 93 |
| 8 | | 3 | 7000 | 3900 | 90 |
| 9 | | 4 | 8000 | 4600 | 111 |
| 10 | 3 | 1 | 11000 | 5000 | 114 |
| 11 | | 2 | 14000 | 7050 | 110 |
| 12 | | 3 | 7500 | 3100 | 96 |
| 13 | | 4 | 8000 | 4500 | 94 |
| 14 | 4 | 1 | 9000 | 4300 | 99 |
| 15 | | 2 | 9000 | 4100 | 90 |
| 16 | | 3 | 8500 | 4500 | 100 |
| 17 | | 4 | 9000 | 4400 | 99 |
| 18 | 5 | 1 | 9200 | 4200 | 93 |
| 19 | | 2 | 9200 | 4300 | 100 |

Du vælger nu at afprøve følgende kausalmodel:

Påbeg t-1 KM t-1 Salg t

Det vil sige, at antal påbegyndte boliger i kvartalet før forårsager et bestemt antal kvadratmeter påbegyndt boligareal i dette kvartal, som igen forårsager et bestemt salg i det efterfølgende kvartal.

Du bruger nu igen korrelationsanalysen til at undersøge denne sammenhæng nu med tre variabler. Du markerer de tre variabelkolonner i tabellen og anvender "korrelation" i "Dataanalyse", som tidligere. Du får nu en korrelationsmatrix, der ser således ud:

**Figur 21:** Korrelationsmatrix

| | A | B | C | D |
|---|---|---|---|---|
| 1 | | *Kolonne 1* | *Kolonne 2* | *Kolonne 3* |
| 2 | Kolonne 1 | 1 | | |
| 3 | Kolonne 2 | 0,936972 | 1 | |
| 4 | Kolonne 3 | 0,745064 | 0,673545 | 1 |

Af denne matrix kan du se, at der er en meget fin positiv lineær samvariation mellem "Påbeg t-1" og "KM t-1" med en korrelationskoefficient på 0,937. Samvariationen mellem "KM t-1" og "Salg t" er på 0,674. Samvariationen mellem "Påbeg t-1" og "Salg t" er, som du tidligere har fundet frem til, 0,745. Det vil sige, at hvis vi vil forudsige salget på basis af disse de to variable, der her er nævnt, vil den bedste forudsigelse være den, der er baseret på "Påbeg t-1".

## Multipel regressionsanalyse

Multipel regressionsanalyse har til formål at undersøge den lineære samvariation mellem to eller flere uafhængige variabler ($x_1$, $x_2$, $x_3$ ….) og en afhængig variabel y. Det vil sige af typen

$$y = a + b_1 x_1 + b_2 x_2 + \ldots.$$

De uafhængige variabler kan både være kvantitative og kvalitative, fx henholdsvis alder og køn.

Betragt følgende eksempel, der er en oversigt over samhørende værdier for pris på lejligheder, størrelse og kvalitet:

**Figur 22**: Ejerlejligheder: Pris, kvalitet og størrelse

| | A | B | C |
|---|---|---|---|
| 1 | Data vedrørende pris på ejerlejligheder | | |
| 2 | Y | | X1 |
| 3 | Pris 1000 kr. | Kvalitet | Størrelse kv.m |
| 4 | 753 | LAV | 52 |
| 5 | 897 | LAV | 66 |
| 6 | 1083 | MIDDEL | 69 |
| 7 | 1007 | MIDDEL | 74 |
| 8 | 1135 | MIDDEL | 79 |
| 9 | 1100 | MIDDEL | 83 |
| 10 | 1146 | MIDDEL | 88 |
| 11 | 903 | LAV | 92 |
| 12 | 1077 | MIDDEL | 97 |
| 13 | 1112 | MIDDEL | 101 |
| 14 | 1262 | MIDDEL | 105 |
| 15 | 1297 | HØJ | 106 |
| 16 | 1175 | MIDDEL | 108 |
| 17 | 1210 | HØJ | 116 |
| 18 | 1308 | HØJ | 130 |

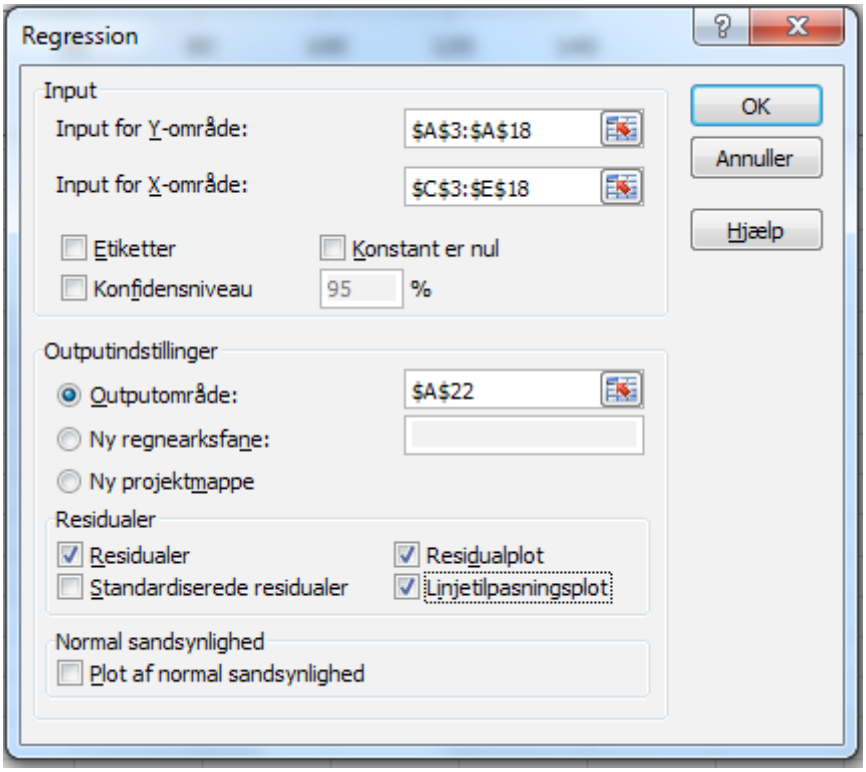
Kilde: Per Vejrup-Hansen, 2012.

Regnearket ændres til en tabel og ordnes efter kvalitet. Der tilføjes tre kolonner med koder for kvalitet således:

**Figur 23**: Ejerlejligheder: Pris, kvalitet, størrelse og dummyvariabler

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Data vedrørende pris på ejerlejligheder | | | | | |
| 2 | Y | Kolonne | X1 | X2 | X3 | Kolonne |
| 3 | Pris 1000 kr. | Kvalitet | Størrelse kv.m | LAV | MIDDEL | HØJ |
| 4 | 1083 | MIDDEL | 69 | 0 | 1 | 0 |
| 5 | 1007 | MIDDEL | 74 | 0 | 1 | 0 |
| 6 | 1135 | MIDDEL | 79 | 0 | 1 | 0 |
| 7 | 1100 | MIDDEL | 83 | 0 | 1 | 0 |
| 8 | 1146 | MIDDEL | 88 | 0 | 1 | 0 |
| 9 | 1077 | MIDDEL | 97 | 0 | 1 | 0 |
| 10 | 1112 | MIDDEL | 101 | 0 | 1 | 0 |
| 11 | 1262 | MIDDEL | 105 | 0 | 1 | 0 |
| 12 | 1175 | MIDDEL | 108 | 0 | 1 | 0 |
| 13 | 753 | LAV | 52 | 1 | 0 | 0 |
| 14 | 897 | LAV | 66 | 1 | 0 | 0 |
| 15 | 903 | LAV | 92 | 1 | 0 | 0 |
| 16 | 1297 | HØJ | 106 | 0 | 0 | 1 |
| 17 | 1210 | HØJ | 116 | 0 | 0 | 1 |
| 18 | 1308 | HØJ | 130 | 0 | 0 | 1 |

Den kvalitative variabel "Kvalitet" kan ikke anvendes direkte, men der skal dannes tre såkaldte "dummyvariabler". Dette gøres ved at udvide med tre kolonner for hhv. "LAV", "MIDDEL" og "HØJ" kvalitet, sådan at de tre kvalitetskategorier kodes med et "1"-tal, hvor kategorien forekommer i kolonnen, og med "O", hvor den ikke forekommer. I regressionsanalysen skal kun indgå to dummyvariabler, når der er tre kategorier. Den sidste kolonne opfatter Excel som referencekolonne, hvilket du kan se senere. Det vil sige, at dummyvariablerne LAV og MIDDEL indgår som uafhængige (forklarende) variabler – sammen med den kvantitative variabel "Størrelse". De tre forklarende variabler er markeret med X1, X2 og X3 over kolonnerne. De skal udgøre en sammenhængende blok (være nabokolonner). Celleområdet C3-E18 bliver således X-område i regressionen, idet overskrifter (etiketter medtages). Dialogboksen ser sådan ud (Vejrup-Hansen, 2012: 106f.).




De væsentligste oplysninger i den omfattende regressionsmatrix, som Excel producerer, er følgende:

**Figur 24:** Regressionskoefficienter

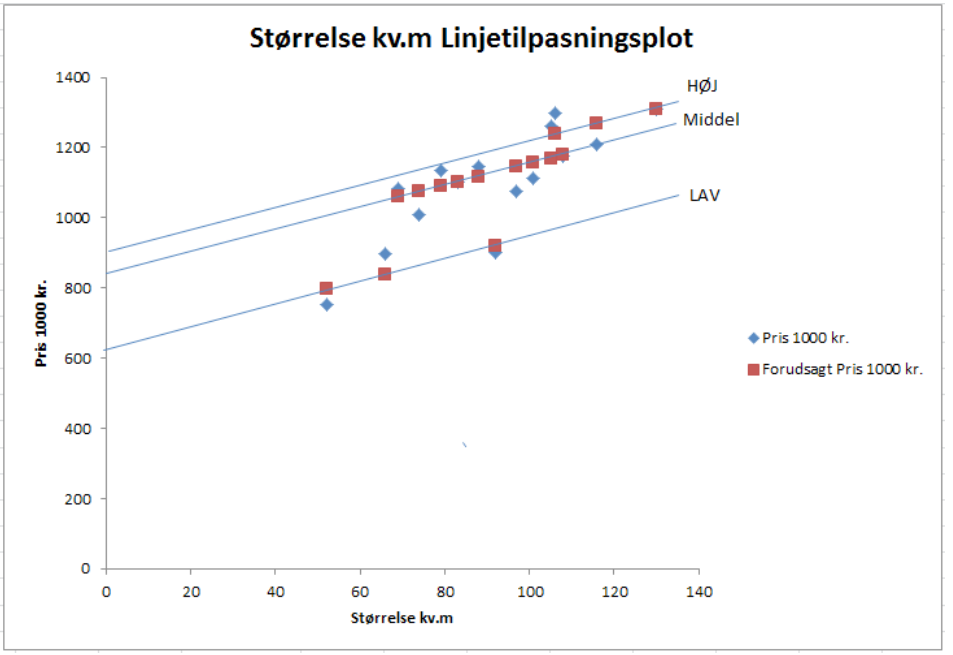
| | Koefficienter |
|---|---|
| Skæring | 914,12 |
| Størrelse | 3,05 |
| LAV | -276,43 |
| MIDDEL | -64,45 |

Resultatet kan skrives på følgende måde:

Pris = 914 +3,05 · Størrelse – 276 · LAV – 64 · MIDDEL

Denne regressionsodel angiver, at for en given m2-størrelse har en lejlighed af lav kvalitet en pris, der ligger 276.000 kroner under en lejlighed med høj kvalitet. En lejlighed med middelkvalitet ligger 64.00 kroner under en lejlighed med høj kvalitet. Konstantleddet på 912 (912.000) gælder for referencekategorien høj kvalitet, idet de to sidste led jo er lig med nul, når kvalitet er lig med høj. Konstantleddet angiver et niveau i det relevante område for størrelser af lejligheder. For lejligheder af lav kvalitet er konstanten lig med 912 minus 276, dvs. at koefficienten til en dummyvariabel angiver en niveauforskel. Koefficienten til Størrelse er 3,05— dvs. at prisen øges med godt 3.000 kroner pr. ekstra kvadratmeter. Det er et gennemsnit for alle lejlighedstyper under ét. Forskellen mellem lav, middel og høj kvalitet kommer kun til udtryk i en niveauforskel i den anvendte model. Regressionens output "Linjetilpasningsplot" (for variablen Størrelse) giver en udmærket illustration, jf. figuren nedenfor. De forventede værdier ifølge regressionsmodellen danner tre forskellige niveauer svarende til de tre kvalitetsvurderinger. I figuren er det fremhævet ved, at der "manuelt" er indtegnet tre rette linjer for høj, middel og lav. Linjerne er parallelle og har samme hældningskoefficient, dvs. koefficienten til Størrelse (Vejrup-Hansen 2012: 108).

**Figur 25**: Ejerlejligheder: Pris, kvalitet og størrelse: Tilpasningsplot



## Litteratur til fortsat læsning

**Ønsker du at vide mere om** statistik og anvendelsen af Excel til statistiske beregninger, kan du med fordel læse Vejrup-Hansen, Per (2012). *Statistik med Excel*. 2. udgave. Frederiksberg: Samfundslitteratur.