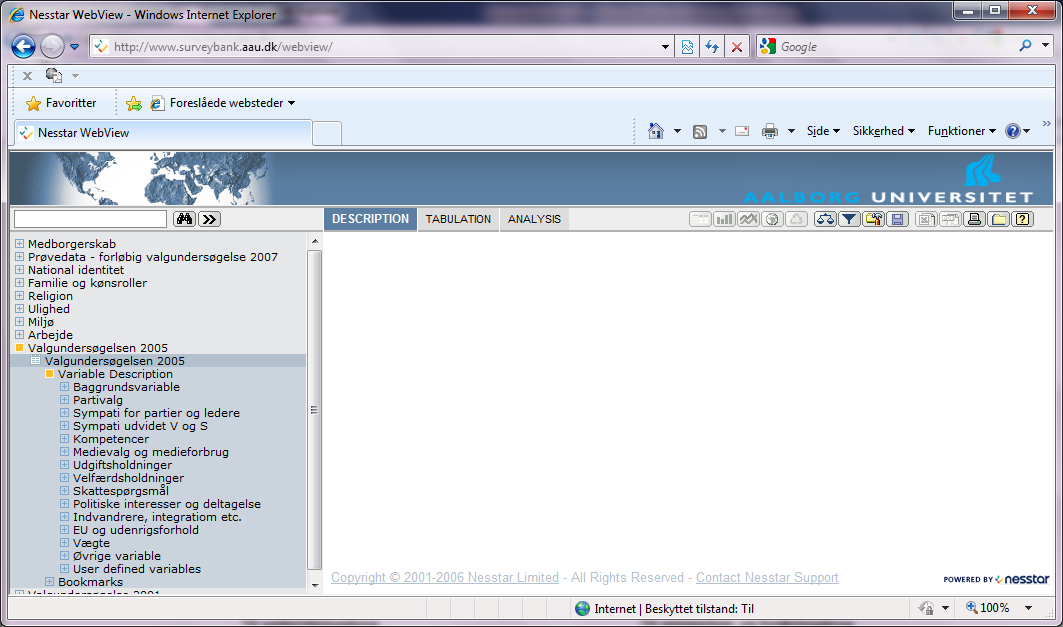
**Surveybank.dk   
Hent data og arbejd med krydstabeller og Chi2-test i EXCEL**

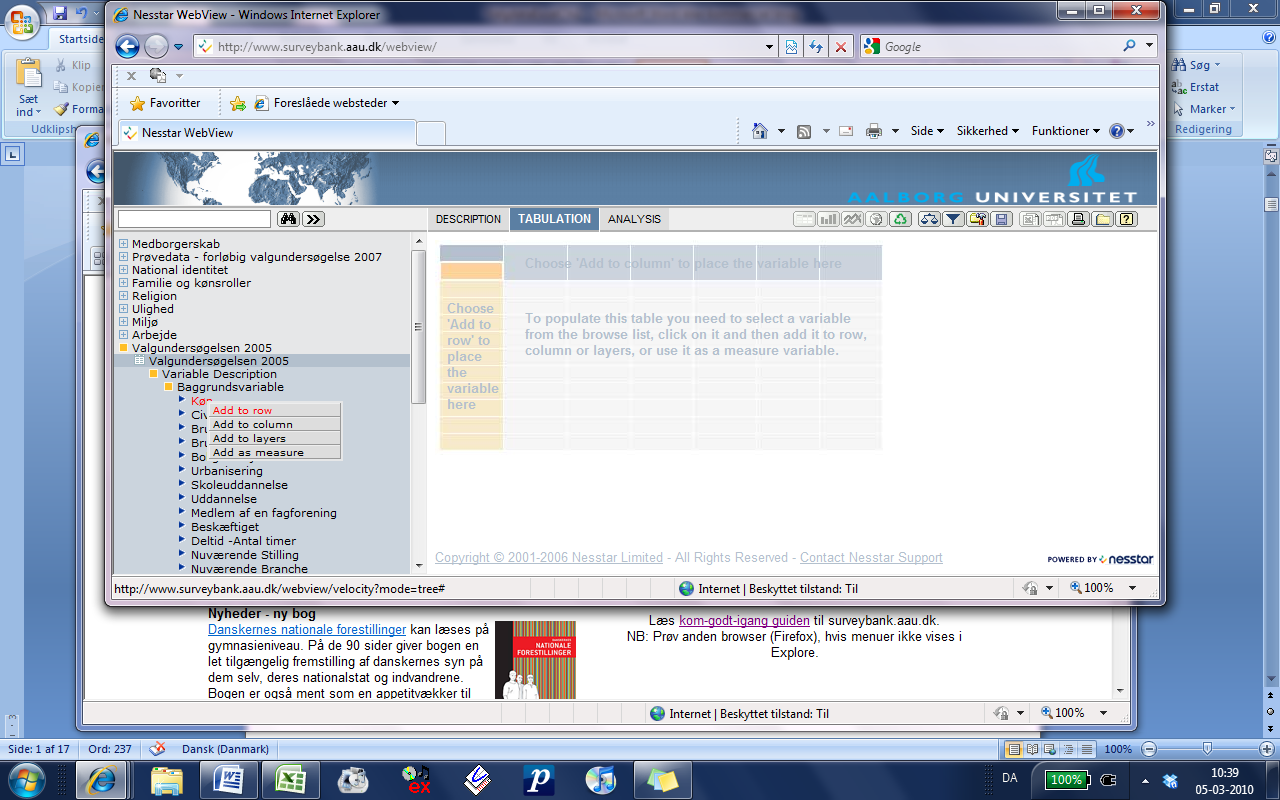
I det følgende skal vi se på hvordan vi kan hente data og arbejde med data. I noten arbejdes med materialet i Excel fordi vi får data exporteret til EXCEL. I kan sagtens bruge andre programmer, men EXCEL er et rigtig godt program at lave dette regne- og præsentationsarbejde. Der er tre faser i vores arbejde:

1. Find data og exporter dem til Excel.
2. Udregninger og testresultat.
3. Tolkning og præsentation af resultat.

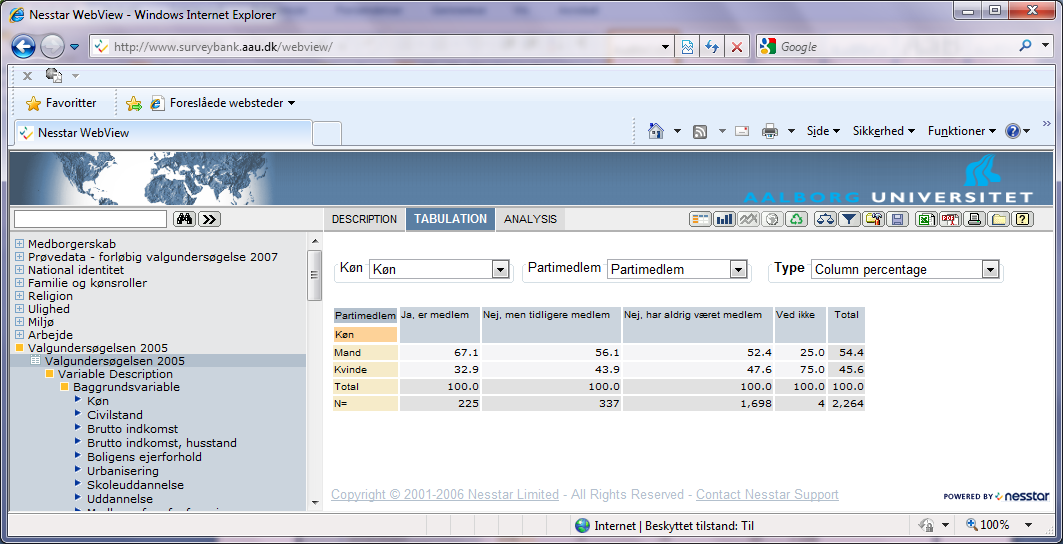
**Eksempel: Sammenhæng mellem køn og partimedlemsskab.   
Er det i højere grad mænd, der melder sig ind i partier?**

**Første del: Find data**

* Åben hjemmesiden <http://www.surveybank.aau.dk/> og tryk på
* Tryk i menuen til venstre på **Valgundersøgelse 2005 > Valgundersøgelse 2005 > Variable Discription**
* Tryk på fanen Tabulation
* Vælg **Baggrundvariable** og vælg **Køn** > **Add to row**



* Vælg **Partivalg > Partimedlem** som den anden variabel og vælg **Add to column**. Du får nu følgende tabel:

****

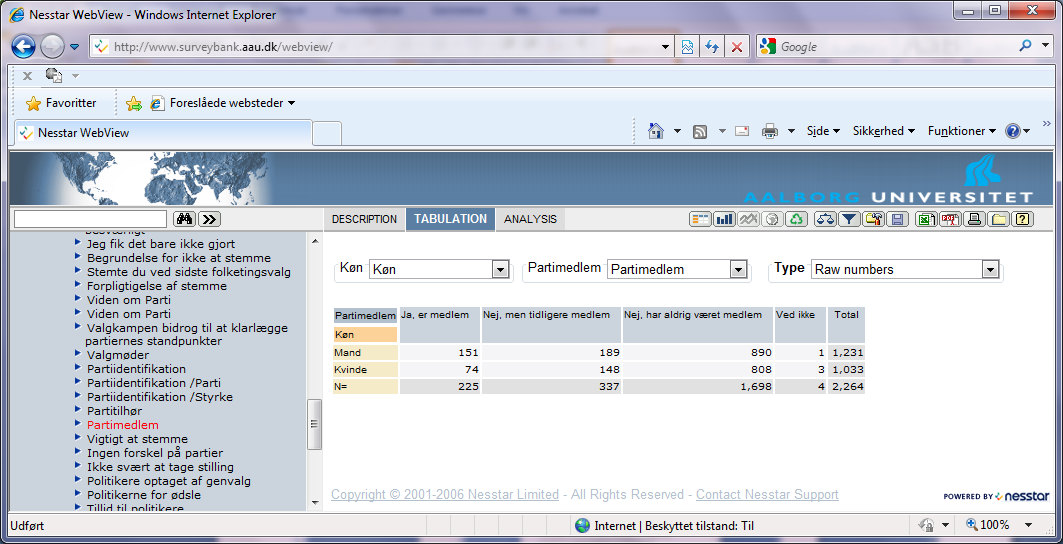
Data angives som udgangspunkt i % summeret sammen i hver kolonne. I yderkolonnerne ser vi hvor mange svar det er beregnet ud fra. Stikprøven er på 2264 personer.

* Vi vil gerne have antal så derfor vælge vi i menuen **Type > Raw Numbers**

Eksporter til Excel

Slet alt

Grafisk fremstilling

****

Skift mellem %-visning og antal

* Eksporter data til Excel og åben arket i Excel.

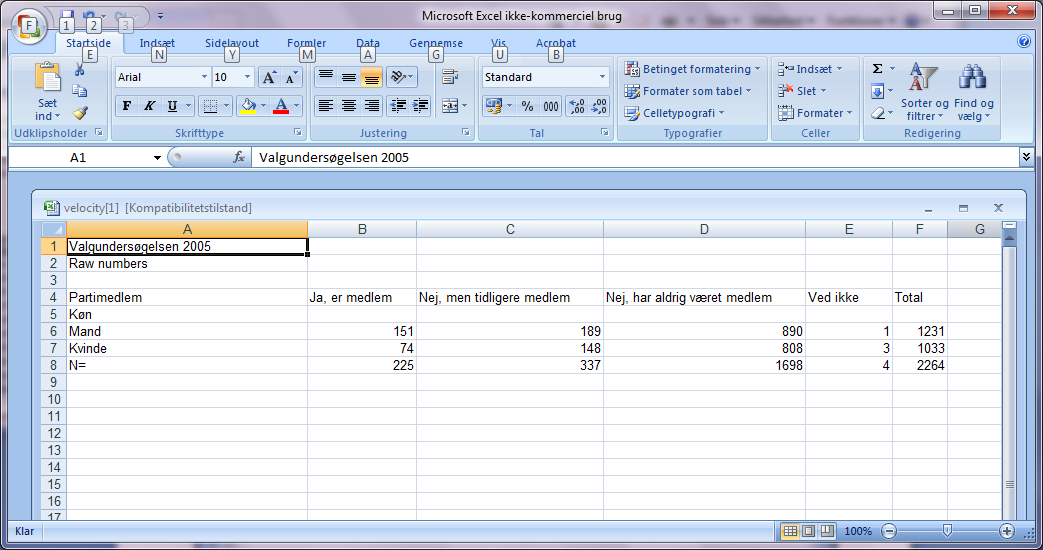
**Testtype:** Uafhængighedstest mellem **køn** og **partimedlemskab**

**H0:** Der er uafhængighed mellem **køn** og **partimedlemskab**

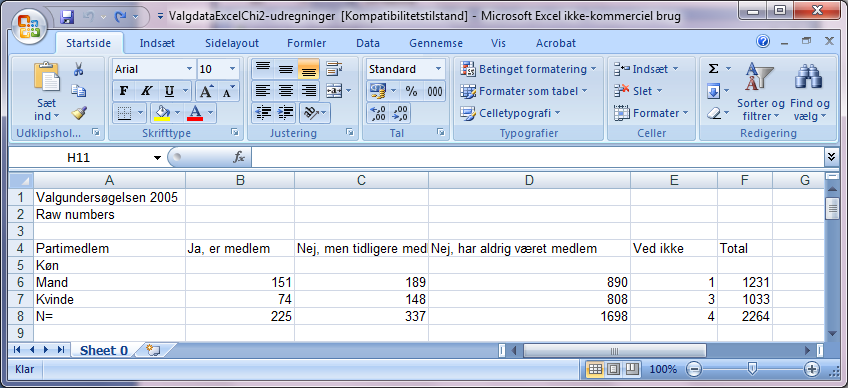
**Signifikansniveau:** 5%

**Anden del: Udregninger og testresultat**

I det følgende skal vi se hvordan vi kan bruge Excel’s formelfunktion til at udregne O-tabel og Q-tabel for os. Det datasæt vi får er følgende:

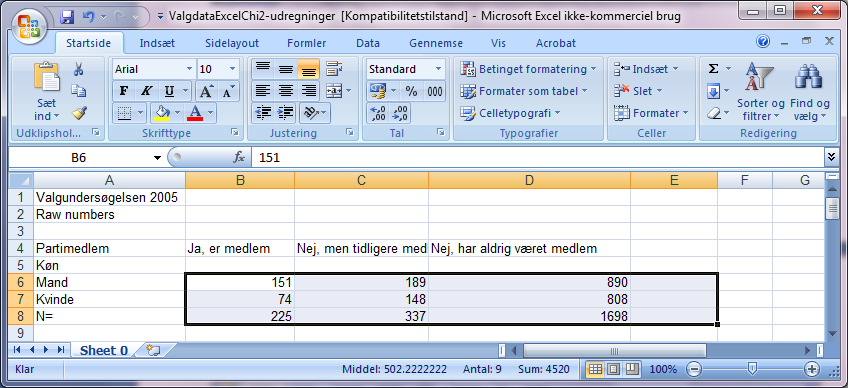
****

* Gem denne fil under et passende navn.

****

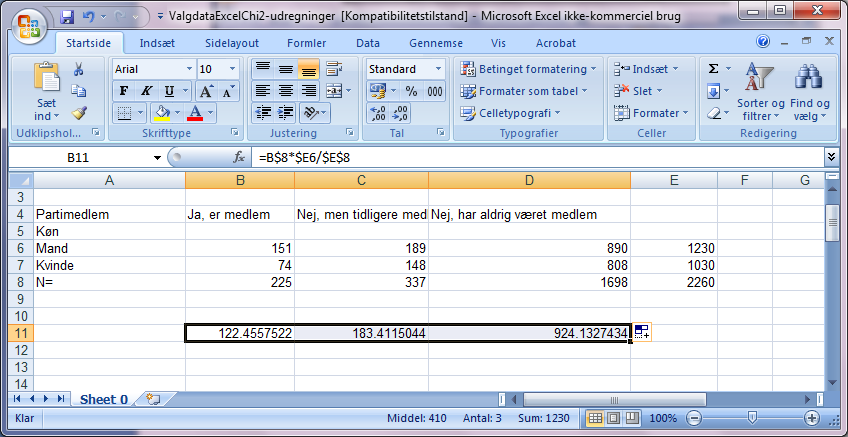
* Når vi laver chi2-test må ingen af cellerne i O-skemaet og E-skemaet være under 5. Har vi data hvor nogle af cellerne er lavere kan vi vælge at slå nogle kategorier sammen eller at stryge dem helt. Her vælger vi at stryge søjlen **Ved ikke** fordi tallene her er helt ubetydelige. Samtidig kan man sige, at hvis de udspurgte ikke ved om de har været medlemmer så er det rimeligt nok ikke at tælle dem med i nogen af kategorierne. Vi stryger derfor søjlen **Ved ikke** og **Total**, som vi jo så skal regne om.
* Afmærk blokken som vist nedenfor og tryk på **Sum-knappen**:

Sum-knap

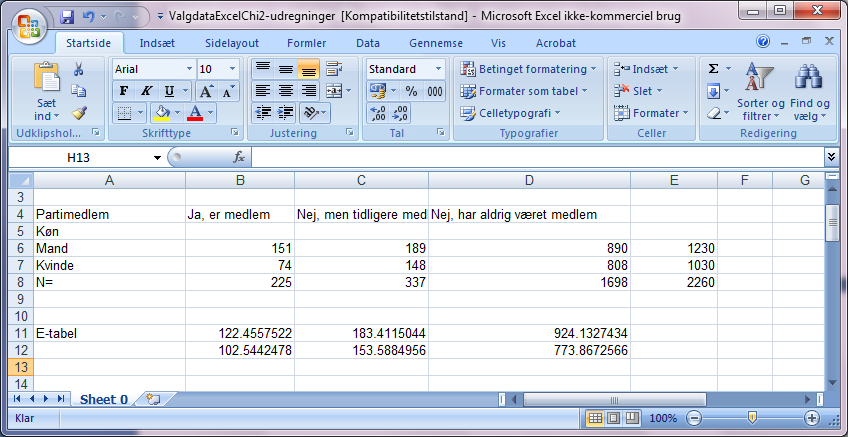
****

Nu skal vi prøve at regne E-værdierne ud i Excel

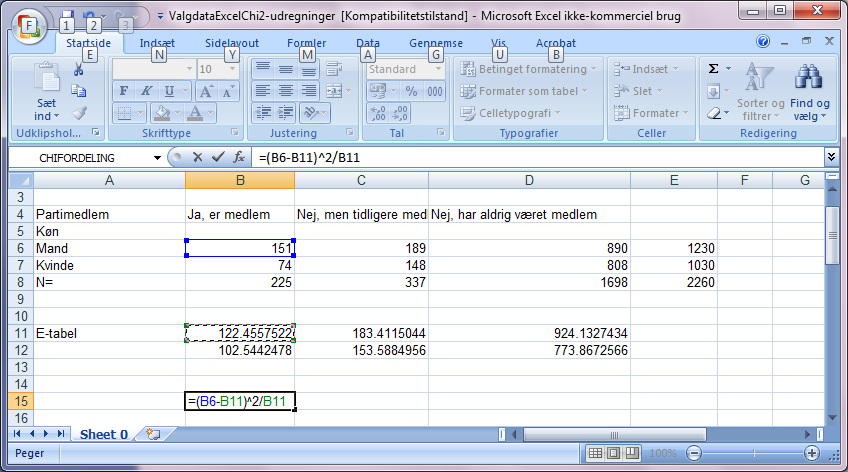
* Klik på feltet B11 og skriv den formel der skal regne q-værdien i første celle: =B8\*E6/E8 og tryk **enter.**
* Vi vil nu rette denne formel lidt til så den kan kopieres. Hvis vi kopierer den vandret vil den næste celle få formlen =C8\*F6/F8 og det er forkert. Det er rigtig nok at det første felt skal være C8, men de to andre skal ikke skifte bogstav. Det sikrer man ved at skrive $ foran. Retter vi formlen i det første felt til =B8\*$E6/$E8, så passer det når vi kopierer vandret. Den anden celle bliver C8\*E6/E8 og det er rigtigt, men det fungerer ikke, når vi kopierer nedad. Cellen lige under den første celle bliver =B9\*E7/E9 og her skulle de to 9-taller stadigvæk være 8-taller. Det sikrer vi igen ved at sætte $ foran, så den endelige formel for feltet B11 bliver =B$8\*$E6/$E$8  
  (Du kan huske det på at det felt du dividerer med – her E8 - skal have $-tegn foran begge to. Alle de andre steder hvor E eller 8 optræder, skal der $-tegn foran).
* Kopier nu formlen ved først at trække den hen:



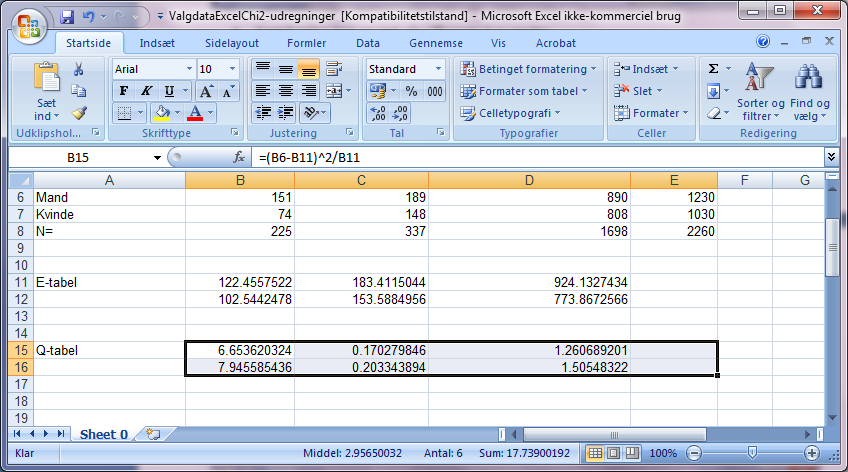
* Derefter trækkes helle denne linje ned:

****

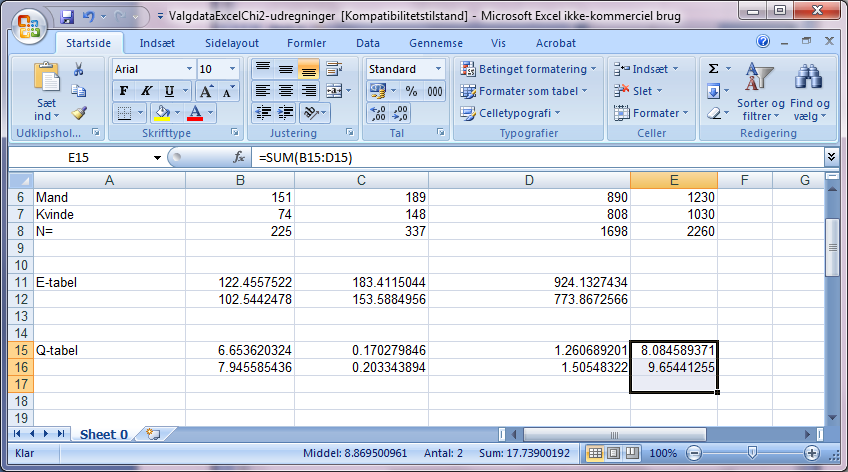
* Stil dig nu i feltet B15 og skriv =(B6-B11)^2/B11 og tryk **enter**. Her skal vi ikke indsætte $-tegn fordi her er ikke nogle felter i formlen der skal ”blive stående” når formlen kopieres ned.

****

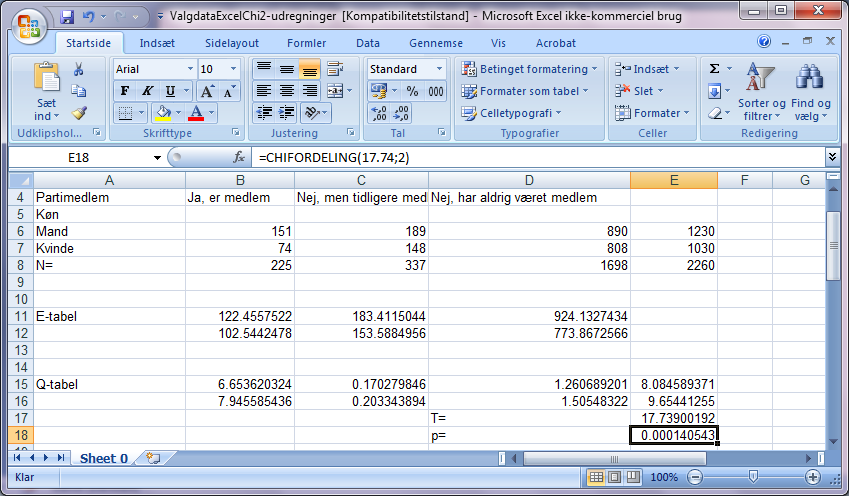
* Træk formlen hen og ned, så vi får vores 6 q-værdier.
* Afmærk blokken som vist nedenfor og tryk på sum-knappen for at få udregnet summen af q-værdierne i hver af de to rækker:



* Afmærk blokken som vist nedenfor og tryk på sum-knappen for at få udregnet den samlede sum af q-værdierne:

****

Vi får den samlede T-værdi T=17.74 i en Chi2-fordeling med f=2\*1=2 frihedsgrader



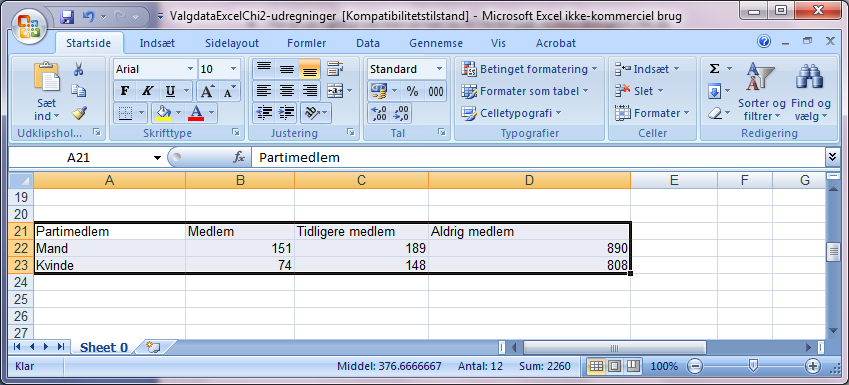
* For at udregne p-værdien skriver du fx i feltet E18 **=Chifordeling(17.74 ; 2)**   
  Bemærk at der er et semikolon mellem T-værdien og f-værdien. Resultatet er en p = 0.00014 = 0.014%.

**Tredje del: Tolkning og præsentation af resultatet.**

Vores konklusion er at der er en markant forskel mellem kvinder og mænd i forhold til partimedlemsskab. Den markante forskel er at **mænd er overrepræsenteret i partiforeninger og kvinder er underrepræsenteret.** Sagt på jævnt dansk: **Den er flere mænd der organiserer sig i partiforeninger end kvinder**.

En vigtig del er at præsenterer resultatet med en grafisk fremstilling, der tydeliggør den konklussion, vi har fundet. Der er mulighed for at lave grafiske fremstillinger af datasættet ud fra hjemmesiden ved at benytte knappen **Grafisk fremstilling** (se s 3). Her vil vi dog kun arbejde med grafisk fremstilling i Excel, fordi det giver os mulighed for at redigere i det materiale, vi vil anskueliggøre.

* Kopier det relevante datasæt til fx A21 som vist nedenfor og ret det til så overskriften står lige over kolonnen. Ret samtidig overskrifterne, hvis de er for lange. Her retter vi til **medlem** – **tidligere medlem** og **aldrig medlem**:



* Prøv at vælge **Grupperede søjler** eller hvad den kan hedde i din udgave af programmet. Er det en god grafik? Noget der forvirrer er at man åbenbart har spurgt færre kvinder, og det gør det mindre tydeligt.
* Prøv at lave samme afmærkning som ovenfor og vælg en der viser resultatet i %. Hos mig hedder den **100% stablet søjle**:

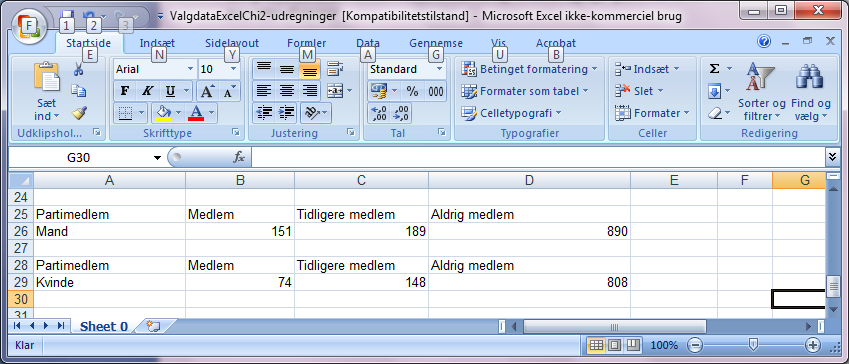
Er den god? Næ. Den viser tydeligt at der er flere mænd med i alle kategorier. Vi vil hellere have % udregnet på den anden led, så vi så på kvinder for sig.

* Lav grafen igen og se om du ikke kan finde en knap der hedder **Skift kolonne/række**

Denne grafiske fremstilling viser det fint

Lagkager er også gode til at vise forskellen i to grupper:

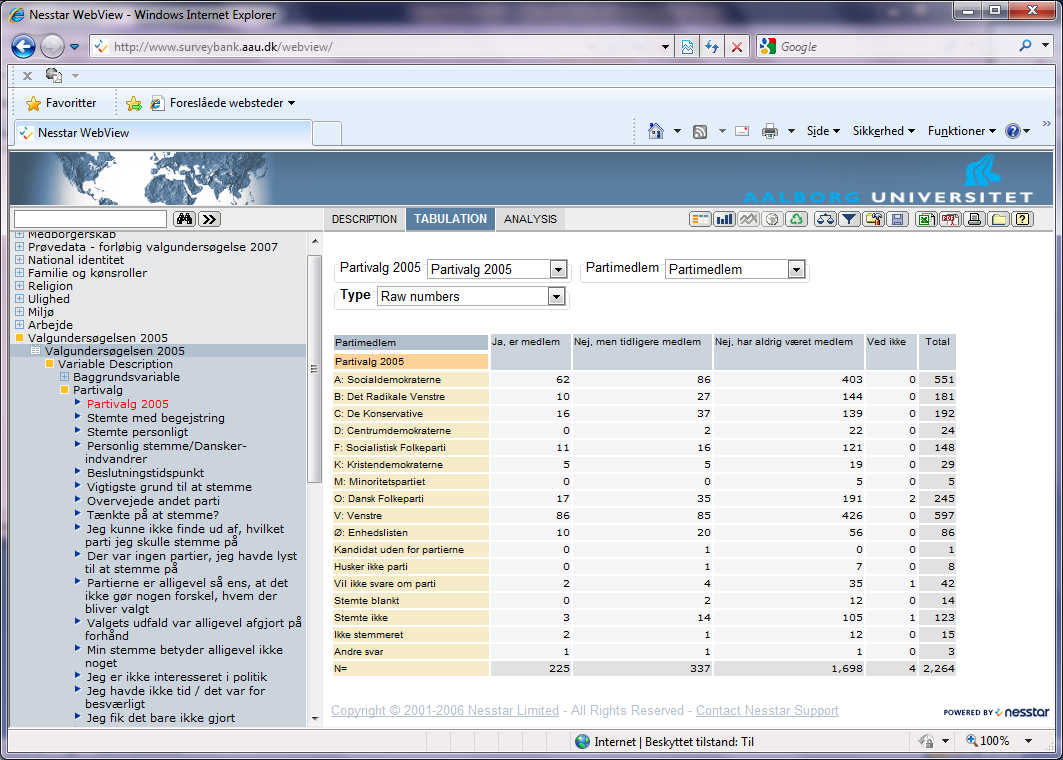
* Stil data op som nedenfor, så mænd og kvinder står for sig. Afmærk hvert skema og vælg ”Lagkagen”. Det er en udmærket grafisk illustration af vores resultat:

****

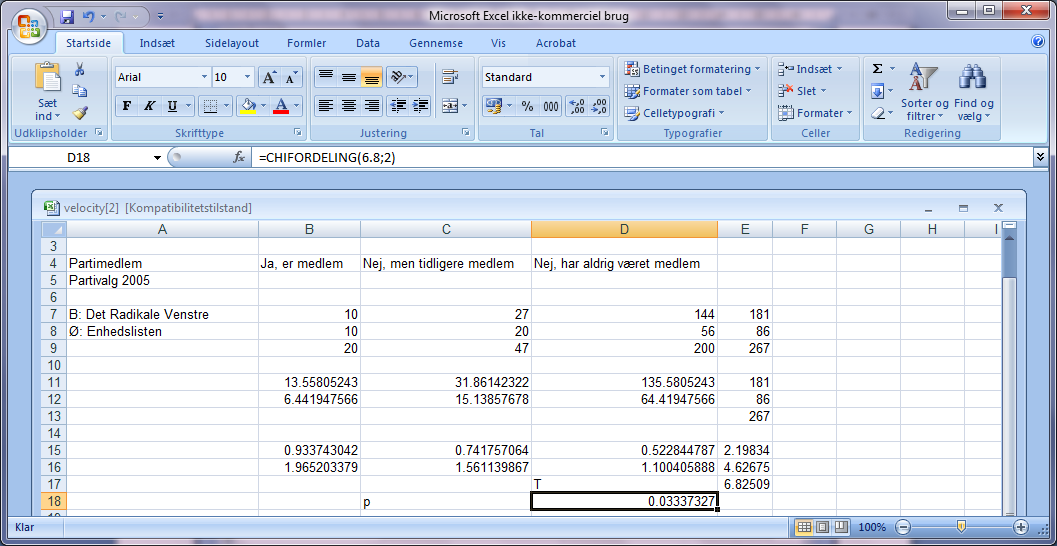
|  |  |
| --- | --- |
|  |  |

**Hele datasæt eller uddrag?**

Ofte vil vi nøjes med at arbejde med uddrag af datasæt. Hvis vi fx ser på sammenhængen mellem medlemskab og partivalg får vi hele datasættet:



Her kunne vi vælge at sammenligne to partier (hvis det var relevant). Det betyder at vi bare trækker et par rækker af vores skema ud og arbejder på dem. Vil vi fx se på Enhedslisten og Radikale venstre får vi følgende data-sæt:

****

Vi kan anskueliggøre forskellen mellem de to partier ved to ”lagkager” (undgå 3D-figurerne … de er sværere at aflæse)

Vi kunne også vælge at vise hele materialet. Nedenfor har jeg sorteret datasættet så dem med den største medlemsandel står først. Derved sætter man fokus på hvem der har den største medlemsprocent i forhold til stemmer. Samtidig kan vi også benytte grafikken til at se andre ting fx spørgsmlet om hvem der har haft den største medlemstilbagegang:

Fx kan du se at Vesnstre er det eneste parti (i 2005) der har flere nuværede medlemmer end tidligere medlemmer blandt dem dem der stemmer på dem. De konservative har mere end dobbelt så mange forhåndværende medlemmer, der stadig stemmer konservativt.