

# Spørgeskemaundersøgelser og beregning af Chi-i-anden

Af

**Jon Urskov Pedersen og Inger Steensgaard Jensen**

Et tværfagligt undervisningsmateriale i matematik og samfundsfag fra



EGÅ GYMNASIUM

Materialet er udarbejdet med støtte fra Undervisningsministeriet, og kan frit kopieres og viderebearbejdes med angivelse af kilde.

## Indhold

Forord: .....	3
Forløb 1. Vælgeradfærd i det 21. århundrede. ....	4
Vælgeradfærd i det 21. Århundrede .....	6
Samfundsvidenskabelig metode.....	7
Induktivt og hypotetisk deduktivt .....	8
Problemformulering og problemstillinger.....	8
Chi-i-anden testen og samfundsfag.....	9
Surveybanken – en introduktion .....	12
Note om $\chi^2$ -test.....	18
Hvad kan en $\chi^2$ -test bruges til? .....	18
Hvornår kan en $\chi^2$ -test bruges? .....	19
Hvilken matematik ligger bag bestemmelsen af $p$ ? .....	20
Sådan beregnes $p$ i excel: .....	21
Sådan beregnes $p$ på TI-89eren:.....	21
Eksempel 1: Er der en sammenhæng mellem ens uddannelse og hvordan man stemmer værdipolitisk? .....	22
Eksempel 2: Er der sammenhæng mellem køn og hvorvidt man synes, at det er vigtigt at have en kernefamilie? .....	25
Opgaver: .....	27
Forløb 2. Sociologisk Projekt opgave.....	28
Vejledning til udarbejdelse af spørgeskemaer i Lectio.....	30
Litteraturliste:.....	37

## Forord:

Følgende materiale er tiltænkt elever i gymnasiet som har samfundsfag på enten A eller B- niveau og samtidig matematik på enten A eller B-niveau. Materialet lægger op til at berøre en række faglige mål i både matematik og samfundsfag:

### Faglige mål matematik:

- anvende simple statistiske eller sandsynlighedsteoretiske modeller til beskrivelse af et givet datamateriale eller fænomener fra andre fagområder...
- anvende simple funktionsudtryk i modellering af givne data, kunne foretage simuleringer og fremskrivninger og forholde sig reflekterende til idealiseringer og rækkevidde af modellerne
- demonstrere viden om matematikanvendelse inden for udvalgte områder, herunder viden om anvendelse i behandling af en mere kompleks problemstilling

### Faglige mål samfundsfag:

- formulere præcise faglige problemstillinger, herunder operationaliserbare hypoteser
- anvende viden om samfundsvidenskabelig metode til kritisk at vurdere undersøgelser og til at gennemføre mindre empiriske undersøgelser
- formidle og tydeliggøre faglige sammenhænge ved hjælp af beregninger, tabeller, diagrammer, modeller og begrebskemaer.

I materialet gennemgås der to forløb. Dels et forløb om vælgeradfærd, hvor der vil være en teoretisk og metodisk optakt. Forløbet vil dernæst bestå af en vejledning til brugen af surveybanken fra Aalborg Universitet og efterfølgende gennemgå chi- i- anden testen.

Til det andet forløb vil der være en kort introduktion. Forløbet er udarbejdet som et projektforsøg, der har en anden tilgang til en spørgeskemaundersøgelse, idet det anvender et spørgeskema udarbejdet af eleverne selv i Lectio. Der er mange måder at tilrettelægge de enkelte forløb på, dette er blot eksempler som der kan bruges mere eller mindre fra.

## Forløb 1. Vælgeradfærd i det 21. århundrede.

Nedenstående materiale kan anvendes på to måder. Den første er i form af et kort forløb, hvor eleverne bliver introduceret til surveybanken, og hvordan man arbejder i samfundsfag.

Lektionsplanen vil se ud som følger:

### Forslag til Lektionsplan:

Lektios Nr:	Fag(emne):	Lektie
Lektion 1	Samfundsfag, Samfundsvidenskabelig metode – hvordan undersøger vi ? Opstilling af problemformuleringer og hypoteser	Noter s -
Lektion 2	Matematik: Beregning af Chi-i-anden	Noter
Lektion 3	Samfundsfag/matematik: Behandling af egne hypoteser	Opgaver

En anden måde at anvende materialet på, er at lade det indgå som et AT-forløb, og dermed lade eleverne arbejde langt mere med at udarbejde problemformuleringer og opstille hypoteser. Eleverne vil få en langt mere uddybende forståelse for det teoretiske, metodiske og ikke mindst det, at regne med chi-i-anden. Et forslag til et AT-forløb med litteraturhenvisninger kunne forløbe som følger:

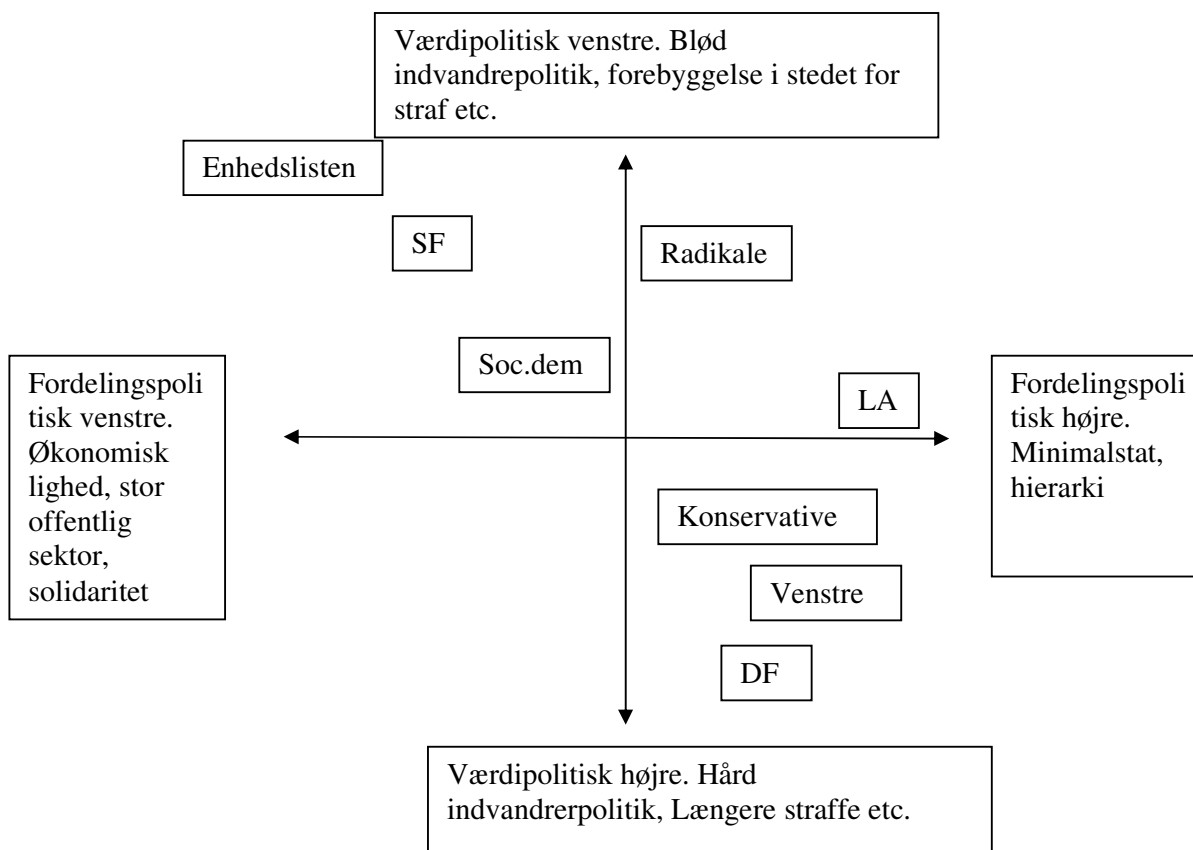
### Forslag til Lektionsplan:

Lektios Nr:	Fag(emne):	Lektie
Lektion 1	Samfundsfag, Samfundsvidenskabelig metode – hvordan undersøger vi ? Anvendelse af surveybanken: <a href="http://www.surveybank.aau.dk">www.surveybank.aau.dk</a>	Fold dig ud: 13-22 Nyttige begreber s49, 61-62(om krydstabel) samt 51-52(om hypoteser)

<b>Lektion 2</b>	Matematik Chi-i-anden test	Hypotesetest Fold dig ud: 89-91
<b>Lektion 3</b>	Samfundsfag Opstilling af hypoteser Vælgeradfærd	Hvorfor mener de det? S 102-109 Demokrati, magt og politik i Danmark: s 67-71
<b>Lektion 4.</b>	Samfundsfag	Note bearbejdning af data fra surveybanken.
<b>Lektion 5</b>	Matematik Test af hypoteser	Eksempelregning. Se note om chi-i-anden test.
<b>Lektion 6</b>	Samfundsfag Vælgeradfærd og medier	Politik og samfund i forandring: s 54-61
<b>Lektion 7</b>	Samfundsfag. Udarbejd en problemformulering der kan testes vha. flere hypoteser. Undersøg hypoteser vha. surveybanken	Fold dig ud: 52-57
<b>Lektion 8</b>	Matematik. Test hypoteserne matematisk	Se Lectio
<b>Lektion 9</b>	Matematik/Samfundsfag	På tværs af Samfundsfag s.9-15
<b>Lektion 10</b>	Matematik/Samfundsfag	Se Lectio
<b>Lektion 11</b>	Matematik/Samfundsfag	Se Lectio
<b>Lektion 12</b>	Fremlæggelse	

## Vælgeradfærd i det 21. Århundrede

I det senmoderne samfund ser vi en tendens til, at de klassiske sociale klasser eller sociale grupper er under opløsning. De politiske partier i Danmark var førhen typisk kendetegnede ved, at de var knyttet til en specifik klasse, og at disse klasser kæmpede mod hinanden. I dag inddeler vi ikke befolkningen i klasser, og vi snakker slet ikke om, at klasserne kæmper mod hinanden. I Danmark har vi fået en universel velfærdsstat, som ingen partier i Folketinget ønsker at ændre fundamentalt på eller forsøger at ændre fundamentalt på. I politik snakker man om, at den fordelingspolitiske ideologiske skillelinje har fået mindre betydning, og den værdipolitiske ideologiske skillelinje nu spiller en langt større rolle. Traditionelt opdelte man partierne på baggrund af en fordelingspolitisk akse, hvor emner som økonomisk ulighed/lighed solidaritet/hierarki, reform/tradition stod over for hinanden. I dag er det andre spørgsmål, der i højere grad er i centrum, nemlig emner som indvandreproblematik, kriminalitet, miljø etc. Dermed ikke sagt at partierne ikke er uenige fordelingspolitisk, men værdipolitikken har større karakter af politiske issues.



Dette skift i dansk politik og dette opbrud med klasserne, har medført, at vælgerne i langt højere grad står overfor individuelle valg, når der skal stemmes til et Folketingsvalg. Der er ikke så mange kernevælgere, dvs. vælgere som stemmer på det samme parti hver gang. Vi er i stedet en nation af marginalvælgere, der fra valg til valg tager stilling til hvilket parti vi ønsker at stemme på. Man taler om at vælgerne i højere grad stemmer på baggrund af de politiske issues, der er aktuelle. Det vil altså sige, politiske emner som eksempelvis indvandrere og klima. Derfor drejer det sig som for partierne at få valgkampene drejet over på de issues, hvor de står stærkest, hvor man kan sige de er issueowners. Mange af disse issues er at finde på den værdipolitiske akse.

En anden diskussion indenfor vælgeradfærd er de konkrete vælgerforskelle. Er det stadig muligt at kategorisere vælgerne, nu hvor arbejderne ikke længere nødvendigvis stemmer på socialdemokraterne eller hvor bønderne stemmer på Venstre? Ifølge valgforsker Jørgen Goul Andersen, er det faktisk stadig muligt, at se en tendens til at arbejderne er en smule mere venstreorienteret mens funktionærerne er mere højreorienterede. På den værdipolitiske akse ses det, at højtuddannede i højere grad er mere venstreorienterede, mens lavtuddannede befinder sig på den værdipolitiske højrefløj.

Når vælgerne skal sætte deres kryds afhænger partivalget i dag i høj grad også af identitetsdannelse. Arbejdet og klassesilhørsforholdet skaber ikke længere vores identitet, det er os der selv skal skabe identiteten. Partivalget er dermed med en identitetsskabende proces, og de politiske holdninger er en måde at udtrykke sine personlige holdninger i forhold til forskellige politiske issues. Ideologierne bruges til at gøre forskel på partierne og danner grundlag for vælgerens refleksioner. (læs evt. mere i ”Hvorfor mener de det” s 102-113)

### **Samfundsvidenskabelig metode**

I ovenstående har vi præsenteret en række teorier, som i høj grad bygger på empirisk materiale. Man kunne undersøge ovenstående ændrede vilkår ved hjælp af en samfundsvidenskabelig metode. Hvilken metode der skal anvendes, afhænger af hvilken problemstilling der stilles op. Formålet med at undersøge en problemstilling er, at vi skal have indsamlet informationer, der kan belyse den. Der

er mulighed for at behandle allerede indsamlede og behandlede informationer/data, og der er mulighed for selv at udarbejde data. Informationer/data kan antage mange forskellige former. Det kan dels være samfundsvidenskabelige teorier, som vi præsenterede i starten, men det kan også være interviews af nøglepersoner. I dette tilfælde arbejder vi med spørgeskemaundersøgelser. Dels som andre har udarbejdet, og dels som vi selv har udarbejdet i det sidste forløb.

### **Induktivt og hypotetisk deduktivt**

I det samfundsvidenskabelige arbejde med en konkret problemstilling kan selve undersøgelsen antage to forskellige metodiske retninger, nemlig den induktive og den hypotetisk deduktive. Ved den induktive metode tager man udgangspunkt i empirien, og udarbejder ud fra den noget generelt. Man kunne forestille sig at Jørgen Goul Andersens resultater og dermed teori, er blevet til ved at anvende induktiv metode, idet han i mange år har arbejdet med vælgeradfærd og har udført mange interviews og lavet mange spørgeskemaundersøgelser. Den hypotetisk deduktive metode udarbejder man på baggrund af nogle teorier, og nogle antagelser/ hypoteser, og går ud og tester disse. Det er denne metode vi anvender i dette eksempel. Det er dog vigtigt at understrege, at der er yderst sjældent forskere udelukkende vil arbejde induktivt eller deduktivt i sin forskning.

### **Problemformulering og problemstillinger**

Hvis vi tager udgangspunkt i dette konkrete tilfælde, så har vi på baggrund af noget teori arbejdet med, at vælgere har skiftet adfærd, og at der knytter sig andre faktorer til partivalg end den klassiske klasseinddeling, idet denne ikke længere er eksisterende. På baggrund af disse teorier kan vi opstille en problemformulering der lyder:

#### **Eksisterer der nogen sammenhæng mellem uddannelse og valg af parti?**

Problemformuleringen er relativ kort og i forhold til vælgeradfærd ville det være mere relevant at undersøge andre sammenhænge som, holdning til miljø, indvandrere og valg af parti. Men for eksemplets skyld vælger vi ovenstående problemformulering. Det er nu muligt at opstille problemstillinger eller hypoteser, der giver svar på denne problemformulering.

**Hypotese 1:** Personer med en længerevarende uddannelse vil i højere grad stemme på den værdipolitiske venstrefløj end personer med kortere uddannelse.

**Hypotese 2:** Dansk Folkepartis vælgere er hovedsageligt vælgere med kort eller slet ingen uddannelse.



**Hypotese 3:** De Radikale og SF har hovedsageligt vælgere med længerevarende uddannelse.

Det er muligt at opstille mange flere hypoteser. Men ovenstående hypoteser kunne begrundes teoretisk, og man kunne komme ind på en diskussion omkring identitetsdannelsen bag de enkelte partivalg. Det vil jeg undlade her, men henviser til "Hvorfor mener de det" side 102-113. Til gengæld vil vi nu prøve at teste den første af de ovenstående hypoteser ved hjælp af en kvantitativ metode, dvs. en samfundsvidenskabelig undersøgelse, der baserer sig på store talmængder, svar og facts, modsat den kvalitative der ønsker at opnå fyldige svar. (Se endvidere: "Fold dig ud" s. 13-29). Vores kvantitative undersøgelse vil bearbejde data fra Aalborg Universitet.

**Opgave 1:** Udarbejd en problemformulering der tager udgangspunkt i vælgeradfærd og i det faktum at vælgerne har ændret adfærd. Brug gerne den værdipolitiske akse til at komme med forslag.

**Opgave 2:** Opstil 2-3 hypoteser der svarer på ovenstående problemformulering

### Chi-i-anden testen og samfundsfag

Når vi skal bearbejde vores hypoteser, må vi være bevidste om, hvordan de kan undersøges i det konkrete talmateriale. I det konkrete tilfælde, er det muligt at finde spørgsmål, der dels går på partivalg og på hvilken uddannelse den enkelte har. Man kan starte med at se på tabellen hvor mange, der har en bestemt uddannelse og hvor mange der har stemt på et bestemt parti. Dette giver dog ikke svar på vores hypoteser. Vi er derfor nødt til at indsætte dem i en krydstabel, så vi kan se hvor mange, der har stemt på et bestemt parti og har en bestemt uddannelse. I den forbindelse er det vigtigt, at være bevidst om hvordan man læser en krydstabel. Idet tallene oftest med fordel omregnes til procent, er det vigtigt at se på hvilket led de giver 100%:

Stemte du ved folketingsvalget d. 13. november 2007?	A: Socialdemokraterne	F
Hvilken uddannelse har du ud over skoleuddannelsen?		
Grundlæggende erhvervsuddannelse	7.0	
Afsluttet erhvervsuddannelse	30.6	
Videregående uddannelse af kort varighed	10.0	
Videregående uddannelse af mellemlang varighed	22.3	
Videregående uddannelse af lang varighed	7.0	
Ingen	18.8	
Andet noter:	4.4	
Ved ikke	0.0	
Total	100.0	
N=	229	

Vi ser altså her hvor mange af socialdemokraternes vælgere der har en kort mellemlang, lang uddannelse osv. Vi ser IKKE hvor mange af dem med en kort uddannelse der har stemt på socialdemokraterne. Når man læser en tabel er det derfor altid vigtigt at bemærke på hvilket led den giver 100%.

Når man følger den efterfølgende vejledning og udarbejder en krydstabel, vil man hurtigt opdage at der tilsyneladende er forskel på, hvordan stemmerne fordeler sig, dels indenfor de forskellige partier og i forhold til hvilken uddannelse man har. Betyder det så at man kan bekræfte alle tre hypoteser? Nej! For forskellene kan skyldes tilfældigheder. Derfor er det vigtigt at inddrage en matematisk analyse, der kan undersøge om forskellene skyldes en tilfældighed, og her kommer Chi-i-anden testen ind, som bliver gennemgået i detaljer i afsnittet om Chi-i-anden testen. Det er vigtigt at bemærke, at når man matematisk arbejder med en chi-i-anden test opstiller man en  $H_0$  hypotese, hvilket er en hypotese, der siger at der ingen sammenhæng er. Det er vigtigt ikke at blande de matematiske hypoteser sammen med de overordnede hypoteser man har stillet op til besvarelse af en problemformulering. Der kan dog forekomme tilfælde hvor  $H_0$  hypotesen vil lægge sig op ad de i forvejen opstillede hypoteser. Vi kunne eksempelvis have opstillet en hypotese som følger:

**Hypotese 1a:** Der er ingen sammenhæng mellem uddannelse og valg af parti.

## Surveybanken – en introduktion

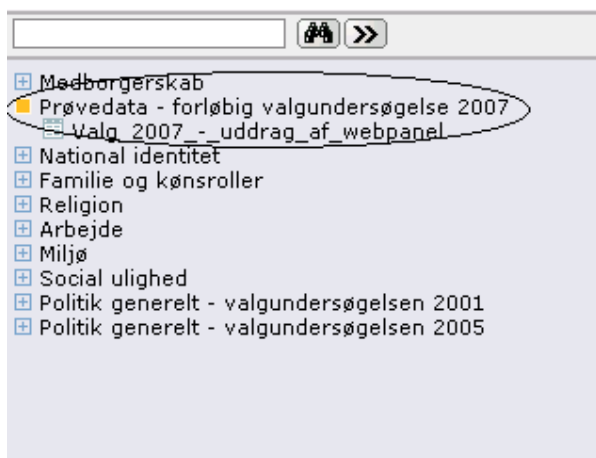
I det følgende vil der være en guide til brugen af surveybanken på Aalborg Universitet

Gå ind på [Ålborg Universitets surveybank](#)

Klik på start databank.

Du kan nu i venstre side af menuen se alle de undersøgelser, som Aalborg Universitet har lagt op til frit forbrug. Mange af undersøgelserne kan bruges i mange forskellige sammenhænge og i mange aspekter af samfundsfag. Det skal dog bemærkes at flere af dem kan være op til 10 år gamle. Fælles for dem er, at det er samme metode som nedenstående guide viser der gør sig gældende når man skal behandle data. Desuden vil det være muligt at finde både bøger og artikler der knytter sig til alle artiklerne.

Klik på krydset Prøvedata - forløbige valgundersøgelse 2007(Dette vil sandsynligvis indenfor kort tid skifte navn til ”Politik generelt – valgundersøgelsen 2007”)

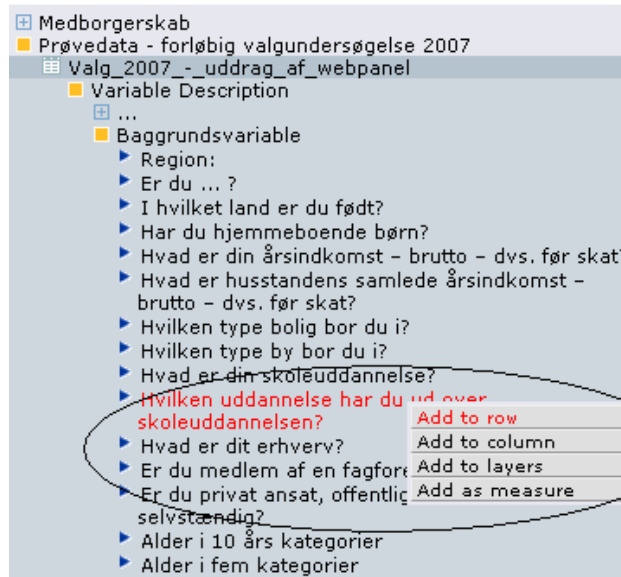


Klik nu på Valg:\_ 2007 uddrag af webpanel og vælg derefter ”Tabulation” i midten af bjælken i toppen

Klik på "Variable description"

Klik på "Baggrundsvariable"

Klik på "uddannelse" og vælg "Add to Row". Denne mulighed kan forsvinde hvis du eventuelt kommer til at dobbeltklikke. Du skal i så fald klikke på tabulation i den øverste bjælke

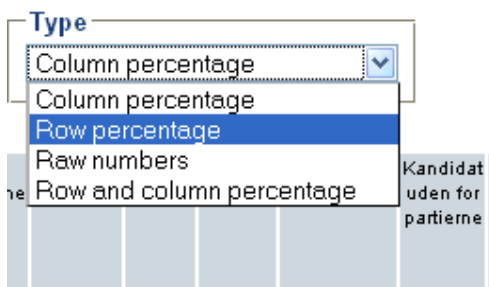


Du vil nu kunne se hvilken uddannelse de adspurgte i undersøgelsen har.

Luk evt. baggrundsvariable først for at få overskuelighed og klik nu på "Partivalg"

Klik på stemte du ved Folketingsvalget d. 13 november 2007?" og vælg "Add to column"

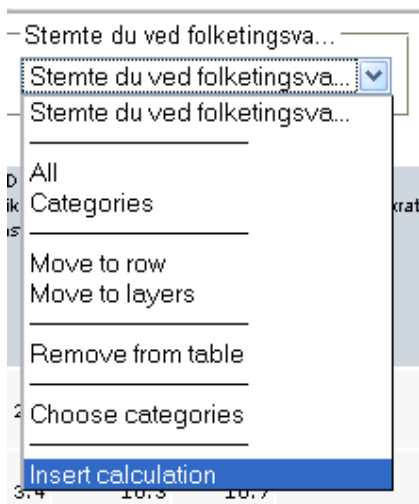
Det er nu muligt at lave en krydstabel over sammenhæng mellem uddannelse og partivalg. Som den er indskrevet nu, viser den hvilken uddannelse det enkelte partis vælgere har. Prøv nu at klikke på "Type" og ændr "Column percentage" til "row percentage".



Nu ser du, hvilket parti der er valgt indenfor de forskellige uddannelser.

Som tabellen ser ud nu, kan den se ret uoverskuelig ud. Det er derfor muligt at slå nogle kategorier sammen og dermed forenkle tabellen.

Klik på ”stemte du ved Folketingsvalget” og vælg ”Insert Calculation”



Lav nu nogle kategorier som er mere enkle. I dette tilfælde kobler vi partierne sammen i værdipolitisk højre og værdipolitisk venstre. Bemærk at I det gennemgåede eksempel anvender vi ikke Socialdemokraterne og Liberal Alliance, da de ligger på den værdipolitiske midte.

-Select type

Add ▼

---

Select categories

Deselect all Select all

---

All

- A: Socialdemokraterne
- B: Det Radikale Venstre
- C: De Det Konservative Folkeparti
- F: Socialistisk Folkeparti
- K: Kristendemokraterne
- O: Dansk Folkeparti
- V: Venstre
- Y: Ny Alliance
- Ø: Enhedslisten
- Kandidatudenforpartierne

---

Constant  

Decimals 1

Label Værdipolitisk Venstre

---

Herefter klikker du ok og vælger nu "choose catagories" som vist.

Stemte du ved folketingsva...

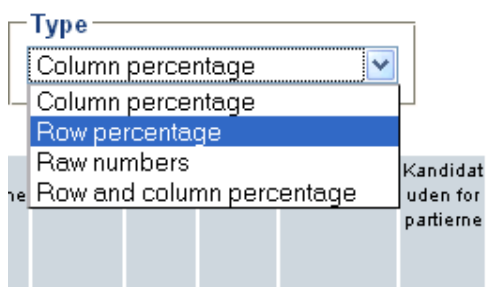
Stemte du ved folketingsva... ▼

- Stemte du ved folketingsva...
- All
- Categories
- Move to row
- Move to layers
- Remove from table
- Choose categories
- Insert calculation

Du kan nu vælge de kategorier, du vil have vist og får dermed det enkle overblik

Det sidste skridt er nu at overføre dine data til Excel, hvorfra man kan udregne Chi-i-anden testen.

Du overfører dine data til Excel, ved at klikke på det grønne Excel ikon i højre hjørne



Det er dog vigtigt at bemærke at inden data overføres til Excel er det en fordel at ændre tallene til Raw numbers, for at kunne regne med dem i chi- i- anden

Dette var blot et eksempel på, hvordan man kunne undersøge vælgeradfærd, med udgangspunkt i partivalg og uddannelse. Der er naturligvis mange andre variable der er interessante at sætte sammen og undersøge.

Nu er det i Excel at vi undersøger tallene og ser om der er en sammenhæng mellem eksempelvis uddannelse og partivalg. Vi vil derfor gennemgå chi-i-anden testen og se på nogle eksempler på hvordan den udregnes og hvordan man behandler data fra surveybanken her.





## Note om $\chi^2$ -test

$\chi$  er et græsk bogstav, der hedder chi og udtales ki. Derfor hedder testen ”chi-i-anden-test”.  $\chi^2$ -testen bruges ofte til at afgøre, hvorvidt der er en sammenhæng mellem forskellige kategorier i en krydstabel. Testen kan kun bruges til at afsløre en sammenhæng, den kan ikke forklare hvorfor der er en sammenhæng.

## Hvad kan en $\chi^2$ -test bruges til?

En  $\chi^2$ -test kan bruges til at afgøre, om et givet observationssæt er uafhængigt. Dette gøres ved at sammenligne den observerede fordeling med en fordeling, hvor man har beregnet hvordan tallene ville se ud, hvis der ikke var afhængighed. Disse opstilles i to tabeller – O-tabellen og E-tabellen:

Observerede data:

O-tabel	Kategori 1	Kategori 2	...	Sum
Kategori A	$O_{11}$	$O_{12}$		$R_1$
Kategori B	$O_{21}$	$O_{22}$		$R_2$
...				
Sum	$K_1$	$K_2$		$N$

En enkelt indgang refereres til som  $O_{ij}$ , hvor  $i$  er rækkenummeret og  $j$  er kolonnennummeret.  $O_{12}$  er altså antallet af personer, der tilhører kategori A og 2.  $R_i$  er det samlede antal personer i række  $i$ ,  $n$  er stikprøvens størrelse, dvs. alle adspurgte personer.

Forventede data:

E-tabel	Kategori 1	Kategori 2	...	Sum
Kategori A	$E_{11}$	$E_{12}$		
Kategori B	$E_{21}$	$E_{22}$		
...				
Sum				$N$

Hvor det forventede antal hvis der ikke var afhængighed beregnes som  $E_{ij} = \frac{R_i \cdot K_j}{n}$ .

Selve størrelsen  $\chi^2$  beregnes således for  $r$  rækker og  $k$  kolonner:

$$\chi^2 = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \dots + \frac{(O_{rk} - E_{rk})^2}{E_{rk}} = \sum_{i,j=1}^{r,k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Overordnet set er  $\chi^2$  et mål for hvor meget observationerne og forventningerne afviger fra hinanden. Jo større  $\chi^2$ -værdi, desto større forskel.

På baggrund af  $\chi^2$ -teststørrelsen, har man udviklet en test, der kan afgøre, om der er signifikant forskel på det observerede og det forventede datasæt – hvis der ikke er, kan det observerede datasæt altså siges at være uafhængigt. Man siger, at man tester hypotesen  $H_0$ : at der er uafhængighed i det observerede datasæt.

$\chi^2$ -testen returnerer en talværdi, der benævnes med  $p$ . Tallet  $p$  er sandsynligheden for, at der ikke er en sammenhæng i det observerede datasæt.

Hvis  $\chi^2$ -testen returnerer en talværdi,  $p$ , der er mindre end 0,05 betyder det, at der er forskel på det observerede og det beregnede datasæt på signifikansniveau 5 %. Dvs. det observerede datasæt må siges at indeholde en afhængighed – og man siger at hypotesen  $H_0$  forkastes.

Sandsynligheden for, at der er en afhængighed er så  $1 - p$ .

Dvs.  $\chi^2$ -testen kan bruges til at undersøge om der er afhængighed imellem to faktorer eller ej.

Hvis der afsløres afhængighed i det observerede datasæt opstilles en såkaldt Q-tabel for at finde ud af, hvor afhængigheden ligger. Dette gøres ved at beregne hvert led i summen ovenfor og opstille det på tabelform således:

<b>Q-tabel</b>	Kategori 1	Kategori 2	...	Sum
Kategori A	$\frac{(O_{11} - E_{11})^2}{E_{11}}$	$\frac{(O_{12} - E_{12})^2}{E_{12}}$		
Kategori B				
...			$\frac{(O_{st} - E_{st})^2}{E_{st}}$	
Sum				$\chi^2$

Den tabelindgang med den højeste talværdi er så der, hvor den største afvigelse mellem O og E-tabellen ligger.

### Hvornår kan en $\chi^2$ -test bruges?

Der er følgende tommelfingerregler ved brug af testen:

- Alle forventede værdier (E-tabellen) skal være mindst 1.
- Højst 20% af værdierne er mindre end 5.

## Hvilken matematik ligger bag bestemmelsen af $p$ ?

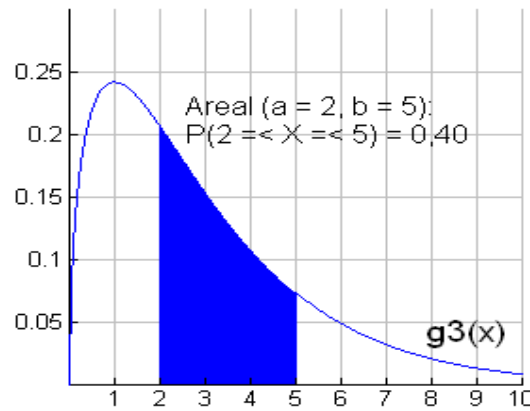
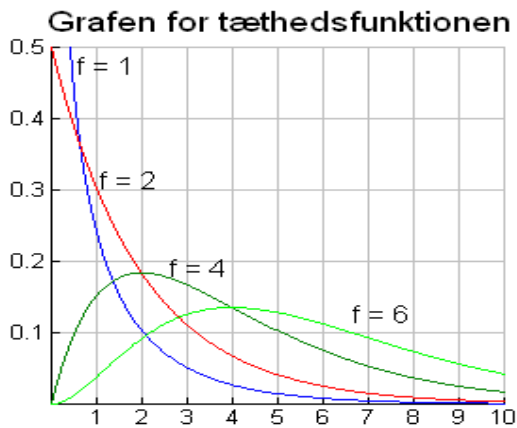
Lad  $X$  være en stokastisk variabel, der antager værdier i  $[0, \infty[$ . Vi siger, at  $X$  er  $\chi^2$ -fordelt med  $f$  frihedsgrader ( $f$  er fordelingsens parameter) hvis,

$$P(a \leq X \leq b) = \int_a^b g_f(x) dx \quad , \text{ hvor } 0 \leq a \leq b$$

og hvor tæthedsfunktionen er:

$$g_f(x) = \frac{x^{0,5 \cdot f - 1} \cdot e^{-0,5 \cdot x}}{\int_0^\infty x^{0,5 \cdot f - 1} \cdot e^{-0,5 \cdot x} dx} \quad , x \geq 0$$

Nævneren i tæthedsfunktionen er en konstant, der afhænger af  $f$  og som sikrer, at arealet under grafen for  $g_f(x)$  er 1. I vores tilfælde med krydstabeller, med  $r$  rækker og  $k$  kolonner i O-tabellen, er antallet af frihedsgrader  $f = (r - 1) \cdot (k - 1)$ .



$P(a \leq X \leq b)$  er altså lig med arealet mellem  $x$ -aksen og grafen for  $g_f(x)$  fra  $x=a$  til  $x=b$  (se figuren ovenfor til højre). Vi siger, at  $P(X \leq x)$  beregnes som det bestemte integral af  $g_f(x)$  med nedre grænse 0 og øvre grænse  $x$ . Bemærk, at  $P(0 \leq X < \infty) = 1$ , da arealet under hele grafen er 1. Dette svarer til, at man ikke kan undgå at få fat i en af personerne i stikprøven når man vælger en ud!  $\chi^2$  beregnes som nævnt ovenfor ved formlen:

$$\chi^2 = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \dots + \frac{(O_{rk} - E_{rk})^2}{E_{rk}} = \sum_{i,j=1}^{r,k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Denne størrelse er  $\chi^2$ -fordelt med  $f = (r - 1) \cdot (k - 1)$  frihedsgrader. P-værdien kan nu beregnes som:

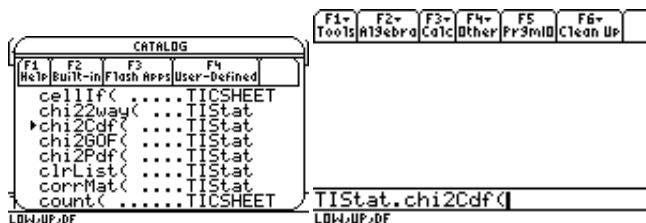
$$p = P(X \geq \chi^2) = 1 - \int_0^{\chi^2} g_f(x) dx$$

## Sådan beregnes $p$ i excel:

1. O-tabellen skrives ind i Excel.
2. E-tabellen beregnes i Excel
3. Skriv `=chitest(` i en celle.
4. Marker de celler, der indeholder O-tabellens datamateriale – dvs. rækken og kolonnen med ”sum” skal ikke markeres (f.eks. A2 til B3).
5. Lav et semikolon
6. Marker de celler, der indeholder E-tabellens datamateriale (f.eks. A8 til B9).
7. Afslut med )
8. Nu skulle der stå `=chitest(A2:B3;A8:B9)` i cellen.
9. Tryk enter. Excel returnerer en værdi for  $p$  Afgør hvorvidt hypotesen er be- eller afkræftet – og med hvilken sandsynlighed.
10. Beregn Q-tabellen hvis hypotesen forkastes og kommenter på den.

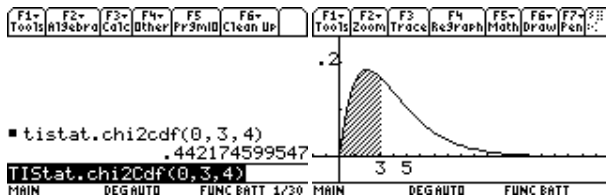
## Sådan beregnes $p$ på TI-89eren:

Vi vil nu se på, hvordan man bestemmer sandsynligheder i en  $\chi^2$ -fordeling. Hertil benyttes funktionen ”tistat.chi2cdf”, der har tre parametre: nedre værdi, øvre værdi og antal frihedsgrader. Funktionen findes ved følgende tastninger i hovedskærmen: ”CATALOG”, ”F3”, pil til ”chi2Cdf(”, ”ENTER”. Nederst på skærmen står de tre parametre, der kræves: low, up, df.



## Eksempel:

Lad  $X$  være en stokastisk variabel, der er  $\chi^2$ -fordelt med 4 frihedsgrader. Vi bestemmer sandsynligheden for at  $X$  er mindre end eller lig med 3, altså  $P(X \leq 3)$  (nedre værdi er 0, da fordelingen starter ved denne værdi):



Vi har således, at  $P(X \leq 3) = 0,4422$ . Dette svarer til det skraverede areal under tæthedsfunktionen på figuren til højre. (Denne figur fås ved følgende tastninger i hovedskærmen: "Y=" (grøn F1), "CATALOG", "F3", pil til "chi2Pdf(", "ENTER", skriv "x,4)", "ENTER", "GRAPH" (grøn F3) (måske skal koordinatsystemet justeres i "WINDOW"), "F5", "7", "0" (lower limit), "ENTER", "3" (upper limit), "ENTER" (værdierne på akserne skrives vha. "F7", "7").

I tilfældet med krydstabellerne er vi typisk interesseret i sandsynligheden for at den stokastiske variabel antager en værdi større end  $\chi^2$ . Lad os sige, at vi i den givne situation har beregnet  $\chi^2$  til at være 3. Så er  $p$  givet ved:

$$p = 1 - P(X \leq 3) = 1 - 0,442175 = 0,557825 \approx 55,8\%$$

Da  $p > 5\%$  er der uafhængighed.

### Eksempel 1: Er der en sammenhæng mellem ens uddannelse og hvordan man stemmer værdipolitisk?

Her tester vi hypotesen fra forrige afsnit. Eksemplet er bygget op så afsnit a er fælles lige meget hvilken metode der benyttes til at beregne  $p$ . I afsnit b beskrives hvordan  $p$  beregnes i "hånden", i afsnit c hvordan det gøres i Excel og i afsnit d hvordan det gøres på TI-89eren.

#### a) Opstilling af E-tabellen:

Først opstilles tabellen over de observerede data - **O-tabellen**:

Uddannelse/stemt ved valget 13/11-07	Værdipolitisk højre	Værdipolitisk venstre	N
Grundlæggende erhvervsuddannelse	24	15	39
Afsluttet erhvervsuddannelse	47	37	84
Videregående udd. (kort)	18	23	41
Videregående udd. (mellemlang)	39	62	101
Videregående udd. (lang)	30	34	64
Ingen	37	41	78
Total	195	212	407

Det ses at der er 2 kolonner og 6 rækker – dvs. antallet af frihedsgrader er  $f = (2 - 1) \cdot (6 - 1) = 5$ . Følgende hypotese ønskes testet:

$H_0$ : Hvordan man stemte værdipolitisk til folketingsvalget i 2007 har ikke noget at gøre med ens uddannelsesmæssige baggrund (der er uafhængighed).

For at undersøge hypotesen opstilles **E-tabellen**:

Uddannelse/stemt ved valget 13/11-07	Værdipolitisk højre	Værdipolitisk venstre	N
Grundlæggende erhvervsuddannelse	19	20	39
Afsluttet erhvervsuddannelse	40	44	84
Videregående udd. (kort)	20	21	41
Videregående udd. (mellemlang)	48	53	101
Videregående udd. (lang)	31	33	64
Ingen	37	41	78
Total	195	212	407

Hvor indgængenes værdier er beregnes ud fra formlen:  $E_{ij} = \frac{R_i \cdot K_j}{n}$ . For eksempel er antallet af personer, der har en mellemlang videregående uddannelse og som har stemt værdipolitisk højre beregnet som:

$$E_{41} = \frac{R_4 \cdot K_1}{n} = \frac{101 \cdot 195}{407} = 18,69 \approx 19$$

**b) Beregning af  $p$  ved hjælp af tæthedsfunktionen og integralregning:**

$\chi^2$  beregnes som:

$$\chi^2 = \sum_{i,j=1}^{r,k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 8,87532$$

Da der er 5 frihedsgrader er:

$$p = P(x \geq 8,87532) = 1 - \int_0^{8,87532} g_5(x) dx$$

Hvor

$$g_5(x) = \frac{x^{0,5 \cdot 5 - 1} \cdot e^{-0,5 \cdot x}}{\int_0^{\infty} x^{0,5 \cdot 5 - 1} \cdot e^{-0,5 \cdot x} dx}$$

Dette giver  $p = 1 - 0,885858 = 0,114142 = 11,4\%$ . Da dette er større end 5 % er hypotesen  $H_0$  bekræftet – Det betyder faktisk at vi ikke kan bekræfte vores hypotese 1. Betyder det så at der ingen

sammenhæng er ? Det er ikke sikkert. Man kunne eventuelt slå nogle kategorier sammen eksempelvis grundlæggende + afsluttet og se om det giver en forskel.

For at grave dybere ned er vi nødt til at undersøge hvor forskellen er. Det gøres ved at beregne **Q-tabellen**. Hver enkelt indgang er en udregning af:

$$Q_{ij} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Uddannelse/stemt ved valget 13/11-07	Værdipolitisk højre	Værdipolitisk venstre	N
Grundlæggende erhvervsuddannelse	1,512	1,390	
Afsluttet erhvervsuddannelse	1,134	1,043	
Videregående udd. (kort)	0,138	0,127	
Videregående udd. (mellemlang)	1,822	1,676	
Videregående udd. (lang)	0,014	0,013	
Ingen	0,004	0,003	
Total			8,87532

Af tabellen ses det, at det især er de med en erhvervsmæssig eller en mellemlang videregående uddannelse, der fordeler sig skævt på de to kategorier.

#### c) Beregning af $p$ vha. Excel:

Ved at følge beskrivelsen ovenfor fås  $p = 0,114141849 \approx 11,4\%$  og hypotesen er bekræftet Nu ville man jo have forventet den samme talværdi som i b) – forskellen ligger i at de 8,87532 som vi bruger i de to andre tilfælde er afrundet. Havde vi taget alle decimaler med havde vi fået samme resultat som excel.

#### d) Beregning af $p$ vha. TI-89eren:

Da der er 5 frihedsgrader skal vi bruge kommandoen `tstat.chi2Cdf(0, 8.87532, 5)`. Det giver 0,885858237334.  $p$  beregnes da som:

$$p = 1 - 0,885858 = 0,11414 \approx 11,41\%$$

Da  $p$  igen er over 5 % accepteres hypotesen( $h_0$ ).



**Eksempel 2: Er der sammenhæng mellem køn og hvorvidt man synes, at det er vigtigt at have en kernefamilie?**

Nedenstående er resultatet af en spørgeskemaundersøgelse fra det senere beskrevne forløb, hvor der er blevet svaret på om det er vigtigt at have en kernefamilie.

O-tabellen:

Svar/Køn	Kvinde	Mand	Total
I høj grad	58	16	74
I mindre grad	13	7	20
I nogen grad	37	37	74
Slet ikke	4	2	6
Ved ikke	4	2	6
Total	116	64	180

E-tabellen:

Svar/Køn	Kvinde	Mand	Total
I høj grad	48	26	74
I mindre grad	13	7	20
I nogen grad	48	26	74
Slet ikke	4	2	6
Ved ikke	4	2	6
Total	116	64	180

Q-tabellen:

Svar/Køn	Kvinde	Mand	Total
I høj grad	2,229	4,041	
I mindre grad	0,0010	0,002	
I nogen grad	2,396	4,342	
Slet ikke	0,005	0,008	
Ved ikke	0,005	0,008	
Total			13,04

I excel giver chi-i-anden testen  $p = 0,011097 = 1,1097\%$ , dvs. hypotesen må forkastes på 95%-signifikansniveau. ( $H_0$ ) og vi kan derfor bekræfte at der sandsynligvis er en signifikant sammenhæng mellem tallene.

## **Opgaver:**

### **Opgave 1:**

Slå nogle kategorierne grundlæggende + afsluttet uddannelse sammen i eksempel 1 og undersøg om det nu er muligt at bekræfte vores hypotese

### **Opgave 2.**

Test Hypotese 2 og 3. Undersøg hvorvidt det også her er nødvendigt at slå flere kategorier sammen for at få bekræftet hypoteserne.

### **Opgave 3.**

Undersøg dine egne hypoteser ved hjælp af Chi-i-anden testen.

## Forløb 2. Sociologisk Projekt opgave

Følgende forløb er et eksempelforløb over, hvordan en spørgeskemaundersøgelse og chi-i-anden test kan anvendes i en sociologisk projekt opgave. Opgaven udarbejdes i grupper og i dette tilfælde u som en synopsis. Udgangspunktet for elevernes projektarbejde er, at de har gennemgået følgende kernestof(eller lignende):

Sociologi og Modernitet; 46-59

Sociologi og Modernitet; 65-71, 156-172

Sociologi af Poul Brejnød 29- 34

Ungdomssociologi s. 80-83

Selve projekt opgaven ser ud som følger

1. I fællesskab(klassen) skal I udarbejde et spørgeskema, der undersøger de ændrede vilkår blandt unge i det senmoderne samfund. I skal anvende eleverne på XX gymnasium som jeres case.
2. I læsegrupper skal I udarbejde en synopsis, hvor I på baggrund af ovenstående statistiske materiale, egne artikler samt sociologiske teorier opstiller hypoteser, undersøger og diskuterer de ændrede vilkår blandt unge i det senmoderne samfund.
3. Produktet skal være en skriftlig synopsis som fremlægges på klassen. Produktet er både en PowerPoint der præsenterer den samt selve synopsisen som afleveres.

Præcisering:

I grupperne skal I undersøge, hvorvidt nedenstående teoretikers teorier afspejles i unges adfærd og holdninger til det senmoderne samfund.

Gruppe 1 +2: Jürgen Habermas

Gruppe 2+3 Anthony Giddens

Gruppe 4+5 Ulrich Beck

## Vejledning til udarbejdelse af spørgeskemaer i Lectio.

For at kunne analysere data i Lectio er det nødvendigt, at få datoen sorteret inde i excel. I det følgende gennemgås, hvordan man opretter et spørgeskema i Lectio, og hvordan man bearbejder data i Excel, så der kan udarbejdes en chi-i-anden test på det.

[Karakterer](#) | [Spørgeskema](#) | [Dokumenter](#) |

Start med at logge ind på lectio og klikke på spørgeskema

Klik herinde på ”opret nyt spørgeskema” og igen på ”opret spørgeskema”



Nu giver det næsten sig selv. Men giv spørgeskemaet en titel, angiv hvem der skal svare, hvor lang tid det skal løbe etc. og gå så ellers i gang med at indtaste dit spørgeskema. Men HER er det vigtigt at være opmærksom på, at hvis der senere skal være mulighed for at krydstabulere SKAL der vælges radioknapper og ingen af de andre. Flueben som i spørgeskemaet giver mulighed for flere svar, kan ikke bruges i en krydstabulering. Sørg derfor ALTID for at eleverne får formuleret spørgsmålene så radioknapperne kan vælges. Det kan betyde at der skal formuleres et par ekstra spørgsmål.

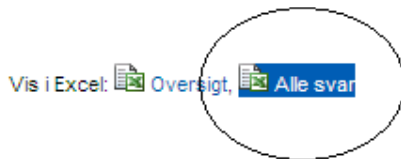
## Spørgsmål

↑ Flyt op ↓ Flyt Ned		+ Indsæt spørgsmål ✕ Slet spørgsmål
Afsnit nr. 1		Valgmuligheder
Spørgsmålstype	Overskrift/tekst	Ikke aktuelt ved pågældende type
Overskrift	Overskrift/tekst	
Spørgsmål/Tekst	Radioknapper	

Når undersøgelsen skal hentes ind klikkes der på ”vis” som på nedenstående billede:

Egne spørgeskemaer				
Titel	Svarfrist	Frigives	Udløber	Vis
2U alkohol				Vis
Evaluering af undervisningen	31/5-2008 11:30			Vis
Samfundsfaglig undersøgelse 3.U	12/3-2009 09:30	12/3-2009 08:00	31/3-2009 12:00	Vis

Man får nu en grafisk oversigt over svarene. For at behandle svarene skal de overføres til Excel, hvilket de bliver ved at klikke på Excelikonet:



Nu er Lectio spørgeskemasvarene hentet ind i Excel. Og der er mange måder at gøre det følgende på. Jeg vælger ikke den hurtigste, da det kræver man er godt hjemme i Excel. Jeg vælger en metode der skaber et godt overblik, og der er sikkert mange andre måder at gøre det på, men denne virker.

Første trin er at finde de to variable man ønsker at undersøge. I dette spørgeskema ønsker vi at undersøge om der er sammenhæng mellem køn og om hvor vigtigt det er at have en kernefamilie.

Det næste man gør, er, at afmærke svarene og få dem flyttet op ved siden af hinanden.

Personlige præferencer - Familien	I hvor høj grad er det at få en kernefamilie vigtigt for dig i fremtiden?	Respondent #1	Radioknapper/Et svar	I høj grad
Personlige præferencer - Familien	I hvor høj grad er det at få en kernefamilie vigtigt for dig i fremtiden?	Respondent #2	Radioknapper/Et svar	I høj grad
Personlige præferencer - Familien	I hvor høj grad er det at få en kernefamilie vigtigt for dig i fremtiden?	Respondent #3	Radioknapper/Et svar	I høj grad
Personlige præferencer - Familien	I hvor høj grad er det at få en kernefamilie vigtigt for dig i fremtiden?	Respondent #4	Radioknapper/Et svar	nogen grad

Når du afmærker, kan du holde ctrl+shift nede mens du trykker på piletasten, så får du hurtigt afmærket alle dine respondenter

Kopier dem, enten ved at klikke på kopiikonet eller ved at klikke på ctrl+c.





Sæt nu svarene ind ved siden af den variable du vil krydstabulere med, så at der ud fra respondent 1 står eksempelvis kvinde og i høj grad:

Overskrift	Spørgsmål/Tekst	Respondent	Type	Svar	Svar 1
Køn	Køn?	Respondent #1	Radioknapper/Et svar	Kvinde	I høj grad
Køn	Køn?	Respondent #2	Radioknapper/Et svar	Kvinde	I høj grad
Køn	Køn?	Respondent #3	Radioknapper/Et svar	Kvinde	I høj grad

Dette betyder, at du nu har sorteret så to svar fra den samme respondent står ud for hinanden.

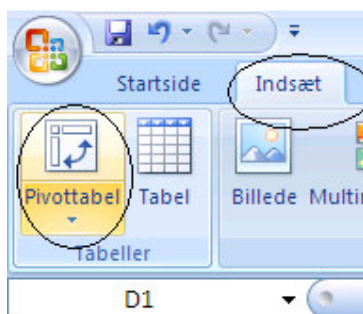
Slet type og ryk svarene sammen med respondenten, dette gør du ved at klikke på type, så det afmærkes og klippe kolonnen ud.

Nu afmærker du respondent og de to svar. Husk kun at afmærke ned til næste svar. Igen brug ctrl og page down til at gøre det hurtigere:

	Køn?	Respondent #183	Kvinde	I mindre grad
	Køn?	Respondent #184	Kvinde	I nogen grad
	Køn?	Respondent #185	Kvinde	I høj grad
	Årgang?	Respondent #1	1.g	
	Årgang?	Respondent #2	1.g	
	Årgang?	Respondent #3	2.g	

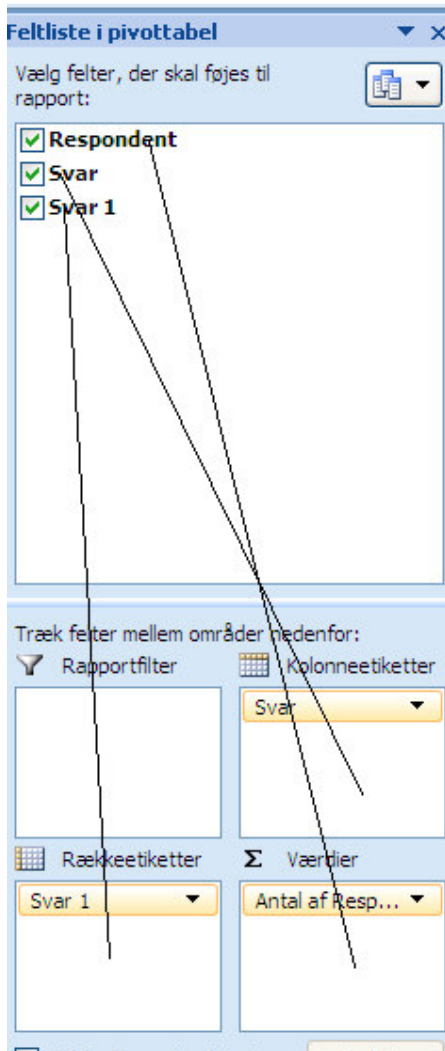
Kopier de afmærkede respondenter ud ved at klikke på ctrl+c og indsæt dem i et nyt regneark.

Klik nu på indsæt og derefter på pivottabel:





Klik ok og indsæt svar 1 + svar i henholdsvis rækkeetiketter/kolonneetiketter eller omvendt. Indsæt derefter respondenterne under sum tegnet:



Du har nu lavet en krydstabel og kan derfra udarbejde en chi-i-anden test.

## Litteraturliste:

Andersen , Lars ”Anvendelse af statistik”, [http://www.fals.info/files/11/anvendelse\\_af\\_statistik.pdf](http://www.fals.info/files/11/anvendelse_af_statistik.pdf).

Brejnrød, Poul: Sociologi Gyldendal 2007

Carstensen, Jens m. fl., ”Mat A2 stx”, Systime, 1. Udgave, 2006.

Christoffersen, Mogens nygaard: Ungdomssociologi Columbus 2001

Hansen, Lise mfl: Hvorfor mener de det? Columbus 2005

Henriksen, Per mfl: Fold dig ud Columbus 2007

Jensen, Inger Steensgaard Note Egå Gymnasium.

Jensen, Pia Møller Note. Egå Gymnasium.

Petersen Jakob Bøje, Note Frederiksborg Gymnasium og HF.