**ØF-notat nr. 17/2007**

# Det lille kvantitative metodeheftet

**av**

**Vegard Johansen**

# Østlandsforskning

Østlandsforskning er et forskningsinstitutt som ble etablert i 1984 med fylkeskommunene og høgskolestyrene/de regionale høgskolesentra i fylkene Oppland, Hedmark og Buskerud som stiftere i samarbeid med Kommunaldepartementet.

Østlandsforskning er lokalisert i høgskolemiljøet på Lillehammer og har i tillegg kontorer i Hamar. Instituttet driver anvendt, tverrfaglig og problemorientert forskning og utvikling.

Østlandsforskning er orientert mot en bred og sammensatt gruppe brukere. Den faglige virksomheten er konsentrert om to områder:

Næringsliv og regional utvikling
Velferd, organisasjon og kommunikasjon

Østlandsforsknings viktigste oppdragsgivere er departement, fylkeskommuner, kommuner, statlige etater, råd og utvalg, Norges forskningsråd, næringslivet og bransjeorganisasjoner.

Østlandsforskning har samarbeidsavtaler med Høgskolen i Lillehammer, Høgskolen i Hedmark og Norsk institutt for naturforskning. Denne kunnskapsressursen utnyttes til beste for alle parter.

ØF -notat nr. 17/2007

# Det lille kvantitative metodeheftet

av

Vegard Johansen

østlandsforskning

**Tittel:** Det lille kvantitative metodeheftet

**Forfatter:** Vegard Johansen

**ØF-notat nr.:** 17/2007

**ISSN nr.:** 0808-4653

**Prosjektnummer:** 10123

**Prosjektnavn:** NFR Basisbevilgning

**Oppdragsgiver:** Grunnbevilgning fra NFR

**Prosjektleder:** Vegard Johansen

**Referat:** Dette notatet er basert på undervisningsnotater i forbindelse med undervisning i samfunnsvitenskapelig metode ved Høgskolen i Lillehammer og Norges Teknisk - Naturvitenskapelige Universitet. I notatet gjennomgås noen grunnleggende måter å analysere kvantitative data på ved statistikkprogrammet SPSS.

Notatet er beregnet for studenter, forskere og andre som har liten eller ingen erfaring med bruk av kvantitativ metode. I heftet redegjøres det kortfattet for forskningsprosessen og deretter fokuseres det på kvantitativ metode. Veiledningen er laget til versjon 14.0 av SPSS.

**Emneord:** Forskningsprosess, statistiske begreper, survey, SPSS, beskrivende analyse, univariat analyse, bivariat analyse, estimering, feilmargin, hypotesetesting, kjikvadrattest, t-test, korrelasjon, multivariat analyse, omkoding, konstruksjon av sammensatt variabel

**Dato:** Oktober 2007

**Antall sider:** 49

**Pris:** Kr 120,-

**Utgiver:** Østlandsforskning
Serviceboks
2626 Lillehammer

Telefon 61 26 57 00
Telefax 61 25 41 65
e-mail: post@ostforsk.no
http://www.ostforsk.no

# Forord

Dette notatet er basert på mine undervisningsnotater i forbindelse med undervisning i samfunnsvitenskapelig metode ved Høgskolen i Lillehammer og Norges Teknisk - Naturvitenskapelige Universitet. På bakgrunn av hyggelige tilbakemeldinger fra studenter og kolleger ved Østlandsforskning fikk jeg lyst til å gjøre mine notater tilgjengelig for enda flere.

Hensikten med dette notatet er å gjøre det praktiske analysearbeidet lettere for studenter og andre forskere som har liten eller ingen erfaring med bruk av kvantitativ metode. I notatet gjennomgås noen grunnleggende måter å analysere kvantitative data på ved statistikkprogrammet SPSS.

I heftet benyttes datasett fra tre prosjekter gjennomført av Østlandsforskning.

*Evaluering av Jysla Løye i TastaSkolen. Et prosjekt om fysisk aktivitet i skolen*
Prosjektleder: Hans Olav Bråtå
Oppdragsgiver: Stavanger kommune

*Mot mer makt for kvinner i lokalpolitikken*
Prosjektleder: Ingrid Guldvik
Oppdragsgiver: KS – Kommunesektorens interesse- og arbeidsgiverorganisasjon

*Ung i Gjøvik*
Prosjektleder: Vegard Johansen
Oppdragsgiver: Gjøvik kommune

Lillehammer, oktober 2007

Torhild Andersen
forskningsleder

Vegard Johansen
prosjektleder

# Innhold

**Forord** ..... **5**
**Innhold** ..... **7**
**Figurer og tabeller** ..... **8**
**1 Sentrale begreper** ..... **9**
1.1 Forskningsprosessen ..... 9
1.2 Innsamling av "survey-data" ..... 10
1.3 Populasjon, sannsynlighetsutvelging og utvalg ..... 12
1.4 Enhet, variabel, verdi, målenivå og årsakssammenhenger ..... 12
**2 Introduksjon til SPSS** ..... **15**
2.1 Meny og ikoner i SPSS ..... 15
2.2 Registrering av svar i SPSS ..... 16
**3 Beskrivende analyse** ..... **21**
3.1 Univariat analyse ..... 21
3.2 Bivariat analyse ..... 24
**4 Estimering** ..... **31**
4.1 Estimering av populasjonsgjennomsnitt ..... 32
4.2 Estimering av feilmargin ..... 34
**5 Hypotesetesting** ..... **35**
5.1 Kjikvadrattesten ..... 35
5.2 t-test for to uavhengige utvalg ..... 38
5.3 Signifikanstest av korrelasjon ..... 41
5.4 Vurdering av slutningsstatistikk ..... 42

# Figurer og tabeller

**Figur 1.** Forskningsprosessens faser (kvantitativ tilnærming) ..... 10
**Tabell 1.** Huskeliste når vi utvikler spørreskjema. ..... 11
**Figur 2.** Populasjon - nettoutvalg. ..... 12
**Figur 3.** Fire målenivåer. ..... 13
**Figur 4.** Årsaksmodell ..... 13
**Figur 5.** Meny og alternativer ..... 15
**Figur 6.** Ikoner i SPSS ..... 16
**Figur 7.** Datamatrisen ..... 16
**Figur 8**. Variabelmatrisen ..... 17
**Figur 9.** Resultatvinduet ..... 17
**Figur 10.** Punching av variabelinformasjon ..... 17
**Tabell 2.** Punching av variabelinformasjon ..... 18
**Figur 11.** Punching av svar ..... 18
**Tabell 3.** Kodebok eller allerede utfylte verdier i spørreskjema ..... 19
**Figur 12.** Gjennomføring av univariat analyse av kategorisk variabel ..... 21
**Figur 13.** Resultater av univariat analyse av kategorisk variabel ..... 22
**Tabell 4.** Presentasjon av variabelen Ordførers kjønn. ..... 22
**Figur 14.** Univariat analyse av kategorisk variabel ..... 23
**Figur 15.** Resultater (output) av univariat analyse av kontinuerlig variabel ..... 24
**Tabell 5.** Presentasjon av variabelen Kvinneandel i kommunestyret. ..... 24
**Figur 16.** Gjennomføring av bivariat analyse. Avhengig variabel ”medlem idrettslag” . 25
**Figur 17.** Resultater av bivariat analyse. Avhengig variabel ”medlem idrettslag” ..... 26
**Tabell 6.** Presentasjon av sammenhengen mellom kjønn og medlemskap i idrettslag. 26
**Figur 18.** Gjennomføring av bivariat gjennomsnittsanalyse av variabel ”skoletrivsel” .. 27
**Figur 19.** Resultater av gjennomsnittsanalyse av variabel ”skoletrivsel” ..... 28
**Figur 20.** Gjennomføring av Pearson korrelasjonsanalyse ..... 29
**Figur 21.** Resultater av Pearson korrelasjonsanalyse ..... 29
**Figur 22.** Slutte om fordelingen i populasjonen basert på utvalgsfordelingen ..... 31
**Figur 23.** Illustrasjon av konfidensintervall ..... 32
**Tabell 7.** Tradisjonell beregning av konfidensintervall ..... 32
**Figur 24.** Gjennomføre beregning av 95 prosent konfidensintervall ..... 33
**Figur 25.** Tolkning av resultater for beregning av 95 prosent konfidensintervall ..... 33
**Tabell 8.** Huskeregler angående variasjoner i feilmargin ..... 34
**Figur 26.** Gjennomføring av moderne utregning av kjikvadrat ..... 37
**Figur 27.** Resultater (output) av kjikvadrat-test ..... 38
**Figur 28.** Gjennomføring av moderne utregning av t-test for to uavhengige utvalg ..... 40
**Figur 29.** Resultater moderne utregning av t-test ..... 40
**Figur 30.** Mulige feilslutninger ved slutningsstatistikk. ..... 42

# 1 Sentrale begreper

I kapittel 1 gjennomgås følgende:

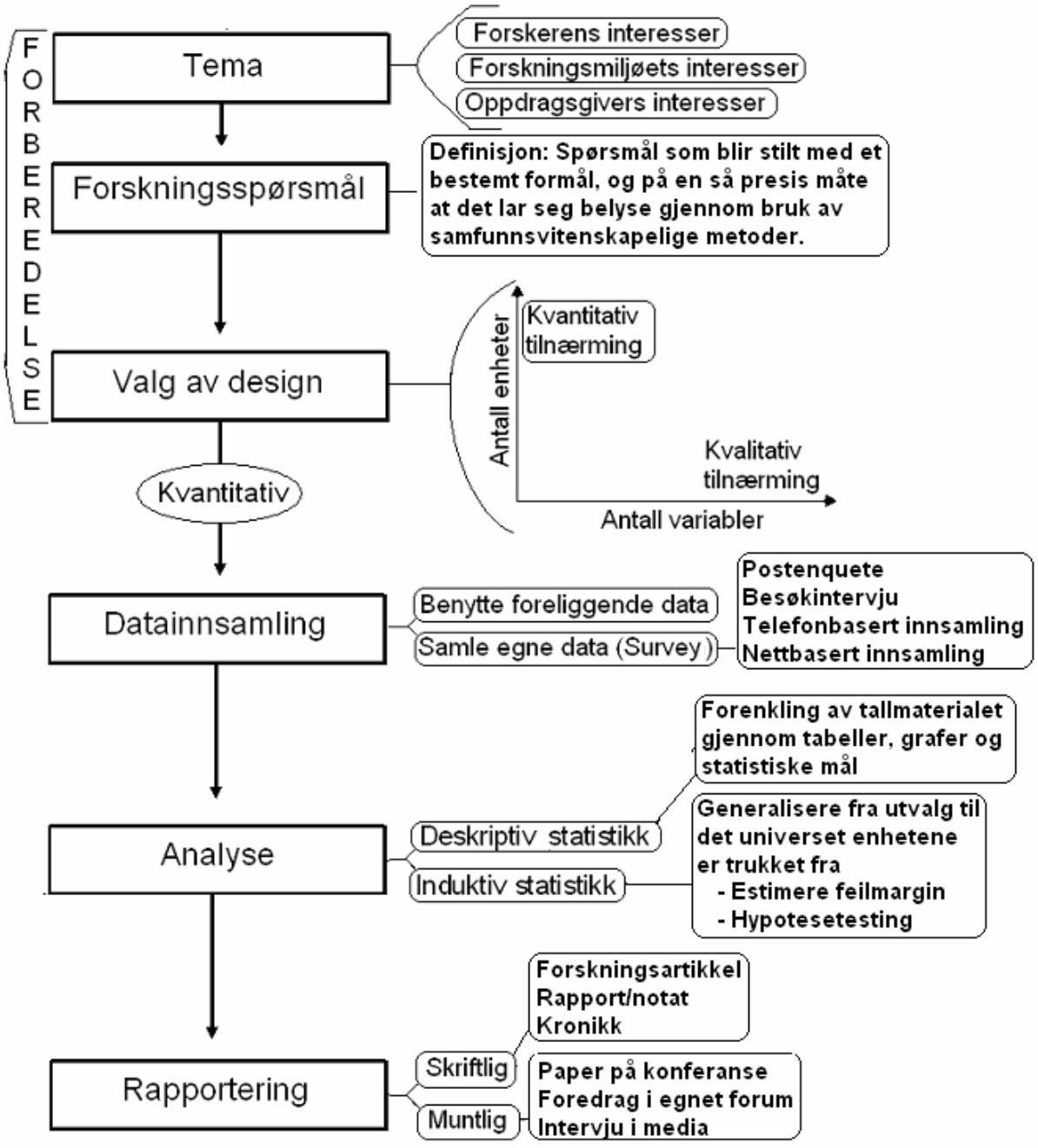
- Forskningsprosessen
- Tips for innsamling av survey-data
- Sentrale begreper ved gjennomføring av kvantitativ analyse

## 1.1 Forskningsprosessen

Forskningsprosessen kan deles inn i fire faser (se illustrasjon på neste side).

- **Forberedelse:** Her utvikles/formuleres forskningsspørsmålet ut fra et gitt tema, og så finner man ut hvordan man best kan besvare dette spørsmålet (valg av tilnærming).
- **Datainnsamling:** Finne relevante og pålitelige data som gjenspeiler virkeligheten som undersøkes, enten ved å samle inn egne data via survey eller å bruke tilgjengelige data.
- **Analyse:** Data bearbeides ved beskrivende statistikk og/eller slutningsstatistikk.
- **Rapportering:** Kan foregå både muntlig og skriftlig.

**Figur 1.** *Forskningsprosessens faser (kvantitativ tilnærming)*




## 1.2 Innsamling av "survey-data"

En survey er en systematisk og strukturert utspørring av et, ofte stort, utvalg personer om et hvilket som helst tema. En survey-undersøkelse har fire faser:

Forberedelse

a. Valg av tema for undersøkelsen og utarbeidelse av spørreskjema
b. Prestudie
c. Hvis populasjonen er stor trekkes et bruttoutvalg

Datainnsamling

a. Valg av innsamlingsteknikk; telefon, post, besøkintervju eller nettbasert
b. De som svarer på undersøkelsen utgjør nettoutvalget
c. Enhetenes svar "punches"/registreres i statistikkprogram

Dataanalyse

a. Beskrivende statistikk - analyser bestemmes av variablenes målenivå
b. Generaliserende statistikk - analyser bestemmes av variablenes målenivå

Rapportering

a. Skriftlig
b. Muntlig

Spørreskjemaer er prekodede skjema med faste spørsmål og (ofte) faste svaralternativer. Denne standardiseringen innebærer at vi kan samle inn mye data på kort tid, undersøke likheter og variasjoner i respondenters svar, kartlegge fenomener, generalisere resultater og undersøke sammenhenger mellom fenomener.

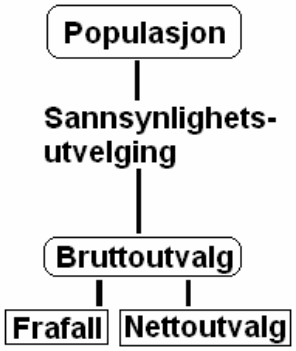
**Tabell 1.** *Huskeliste når vi utvikler spørreskjema.*

| | |
|---|---|
| Struktur | - Introdusere undersøkelsen på første side ved å fortelle om hensikten med undersøkelsen, hvem som gjennomfører den, eventuell godkjennelse av NSD eller etiske komiteer, svarfrist og kontaktopplysninger<br>- Spørreskjemaet må ikke bli for langt, da kan det tenkes at vi bare får med de respondentene som er mest interessert i temaet<br>- Rekkefølgen på spørsmålene kan pirre respondentens interesse, tema legges tydelig opp, og sensitive spørsmål tas inn underveis |
| Spørsmål | - Forholde oss til tidligere forskning for å velge ut relevante spørsmål<br>- Gi svar på forskningsspørsmålene<br>- Entydige, presist og enkelt formulert<br>- Ikke ledende |
| Svaralternativer | - Gjensidig utelukkende og mest mulig uttømmende<br>- Både positive og negative svarmuligheter<br>- Alternativet "vet ikke" ved vanskelige spørsmål |
| Prestudie | - Teste skjemaet på noen utvalgte respondenter |
| Punching | - Utforme skjemaet slik at det er lett å punche |

## 1.3 Populasjon, sannsynlighetsutvelging og utvalg

Populasjon, sannsynlighetsutvalg og nettoutvalg er tre sentrale begreper.

**Figur 2.** *Populasjon - nettoutvalg.*




**Populasjon:** En gruppe av objekter med spesifikke egenskaper definert slik at vi entydig kan avgjøre om et objekt tilhører gruppen eller ikke.

**Sannsynlighetsutvelging:** Ved store populasjoner må vi ofte trekke utvalg. For å øke sannsynligheten for å få et representativt utvalg foretas *sannsynlighetsutvelging.*[1]

**Nettoutvalg:** Det er sjelden at alle respondenter svarer på en survey. Vi skiller mellom *bruttoutvalget* (alle som blir forespurt) og *nettoutvalget* (de som deltar).

Bruttoutvalg – Nettoutvalg = *Frafall*.

Fremgangsmåten ved en spørreundersøkelse kan være slik:

1. Populasjonen defineres
2. Fra populasjonen trekkes et bruttoutvalg ved sannsynlighetsutvelging.
3. Innsamlingsmetode velges og innsamlingen gjennomføres
4. De som svarer på undersøkelsen utgjør nettoutvalget
5. Vi undersøker om nettoutvalget gjenspeiler populasjonen
    a. *Svarprosent (netto/brutto):* Avhenger ofte av flere faktorer som innsamlingsmetode, typen populasjon, tema for undersøkelsen m.m.
    b. Frafallsanalyse: Reduserer usikkerheten ved frafallet og innebærer å sammenligne utvalgsfordelingen med fordelingen i populasjonen på sentrale områder.

## 1.4 Enhet, variabel, verdi, målenivå og årsakssammenhenger

**Enheter:** Forskningsspørsmålene angir hvem vi skal vite noe om og i kvantitative undersøkelser kaller vi de som undersøkes for enheter. Enhetene er oftest mennesker, men kan også være dyr, ting, hendelser, land m.m. Hvis datainnsamlingen foregår ved hjelp av spørreskjema kaller vi enhetene for respondenter.

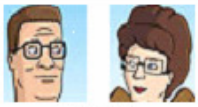
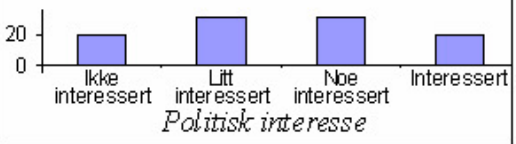
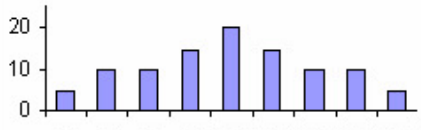
[1] Tre metoder er. 1. Enkel tilfeldig trekking (ETT): Fra en populasjon på N enheter trekkes et utvalg på n enheter tilfeldig, der hver enhet har lik sannsynlighet for å komme med i utvalget. 2. Stratifisert utvelging (to trinn): Populasjonen deles inn i grupper (strata) før utvalget trekkes ved ETT, og hensikten er å sikre at utvalget er representativt på stratifiseringsvariablene. 3. Klyngeutvelging (to eller flere trinn): Først trekkes det ut et utvalg av klynger, og deretter trekkes det utvalg av enheter fra hver klynge ved ETT.

**Variabel:** Fenomener ved enhetene beskrives ved at vi operasjonaliserer variabler. Operasjonalisering vil si å knytte teoretiske begreper til empiriske indikatorer; vi gjør generelle fenomener konkrete slik at de kan måles eller klassifiseres. Denne prosessen resulterer i at vi beskriver fenomenet med en eller flere variabler. Hvis enheten er mennesker er kjønn, utdanning, yrke og inntekt eksempler på typiske variabler.

**Verdier:** Variabler varierer med ulike verdier/kategorier som kan skilles logisk. Hvis vår variabel er kjønn kan vi skille mellom verdiene mann og kvinne.

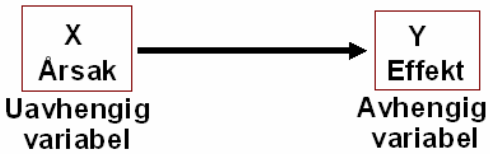
**Målenivå:** Variabler med ulike egenskaper deles inn i et hierarki med fire målenivåer.

**Figur 3.** *Fire målenivåer.*

| Målenivå | Karakteristikk | Eksempel |
|---|---|---|
| Nominalnivå | Gjensidig utelukkende kategorier |  *Kjønn* |
| Ordinalnivå | Gjensidig utelukkende kategorier<br>Kategoriene kan rangeres |   *Politisk interesse* |
| Intervallnivå | Gjensidig utelukkende kategorier<br>Kategoriene kan rangeres<br>Lik avstand mellom kategoriene |   *Intelligens* |
| Forholdstallsnivå | Gjensidig utelukkende kategorier<br>Kategoriene kan rangeres<br>Lik avstand mellom kategoriene<br>Absolutt nullpunkt |   *Inntekt* |

**Årsakssammenhenger:** Noen ganger ønsker vi å undersøke sammenhenger mellom variabler, og ofte er det en retning på disse sammenhengene. D.v.s. at ett fenomen forklares ved ett eller flere andre fenomener. Vi snakker ikke om lovmessighet, men ulik grad av sannsynlighet for at dersom X inntreffer så inntreffer også Y.

**Figur 4.** *Årsaksmodell*




Når vi ønsker å "forklare" variasjonen i et fenomen er det en rekke krav som bør innfris:

- X skal komme før Y
- Samvariasjon mellom X og Y
- Kontroll for andre variabler ($X_1$, $X_2$, $X_n$)

# 2 Introduksjon til SPSS

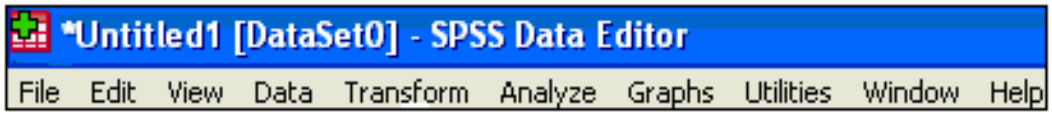
I kapittel 2 gjennomgås følgende:

- Meny og ikoner i SPSS
- Registrering av svar i SPSS

## 2.1 Meny og ikoner i SPSS

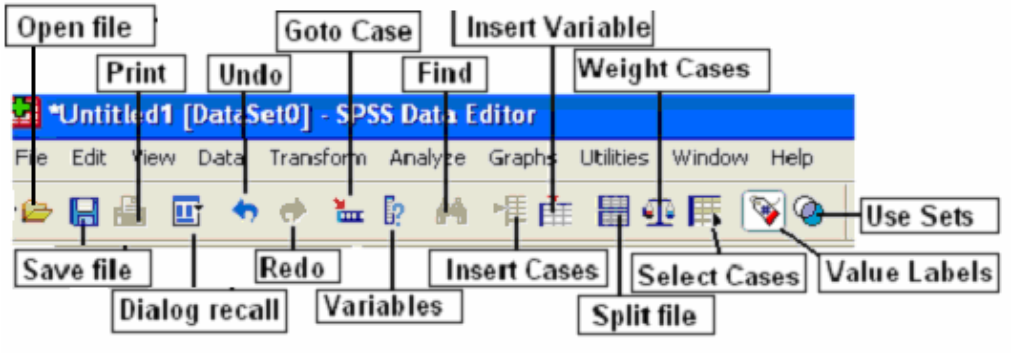
Øverst i SPSS-vinduet er det en menylinje med 10 forskjellige alternativer:

**Figur 5.** *Meny og alternativer*




File har alternativer for å lagre, skrive ut resultater og hente inn nye filer
Edit gir deg muligheten til å kopiere (f.eks. kopiere tall eller datamatrise)
View gir deg mulighet til å tilpasse utseendet av SPSS
Data lar deg endre datamatrisen systematisk
Transform har kommandoer for å endre verdier og variabler
Analyze er menyen for å kjøre statistisk analyse av datamaterialet
Graphs gir tilgang til grafer og grafisk analyse
Utilities gir informasjon fra datamatrisen og SPSS
Window gir muligheten til å holde orden på de ulike vinduene i SPSS
Help gir tilgang til hjelpemenyen i SPSS

**Figur 6.** *Ikoner i SPSS*

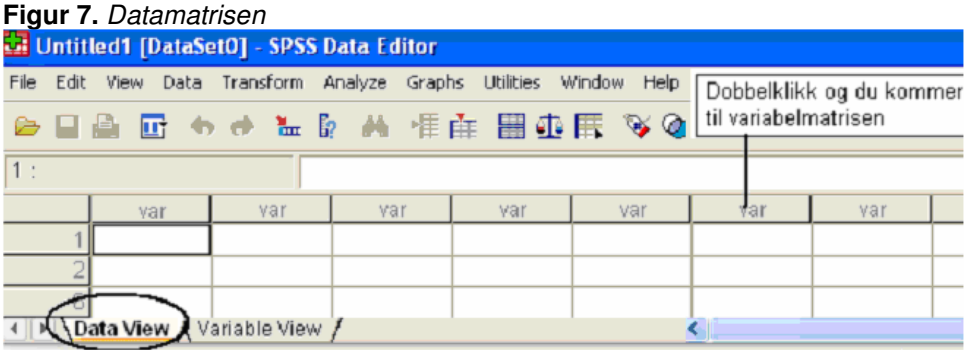



Open file: Åpner filer avhengig av hvilket vindu som er aktivt
Save file: Lagrer filer avhengig av hvilket vindu som er aktivt
Print: Skriver ut vinduet som er aktivt (brukes oftest på resultater)
Dialog recall: De siste "dialogboksene" huskes av SPSS og kan fås tilbake
Undo: Her kan du angre på operasjoner en har utført, såfremt de ikke er lagret
Redo: Her kan du gjøre om igjen operasjoner som du har angret på
Goto Case: Her kommer det opp en boks og du kan velge hvilken enhet du vil se på
Variables: Her er en boks der du kan finne informasjon om samtlige variabler i datafilen
Find: Her kan du søke etter verdier i de ulike ved å plassere musa i kolonnen du ser på
Insert Cases: Ved å klikke på en bestemt enhet i datavinduet setter du inn en ny enhet
Insert Variables: Ved å klikke på en enhet i datavinduet setter du inn en ny variabel
Value Labels: I datavinduet kan innholdet vises som eller verditeksten som er definert

## 2.2 Registrering av svar i SPSS

I SPSS er det tre vinduer vi kan møte[2]:

**Figur 7.** *Datamatrisen*

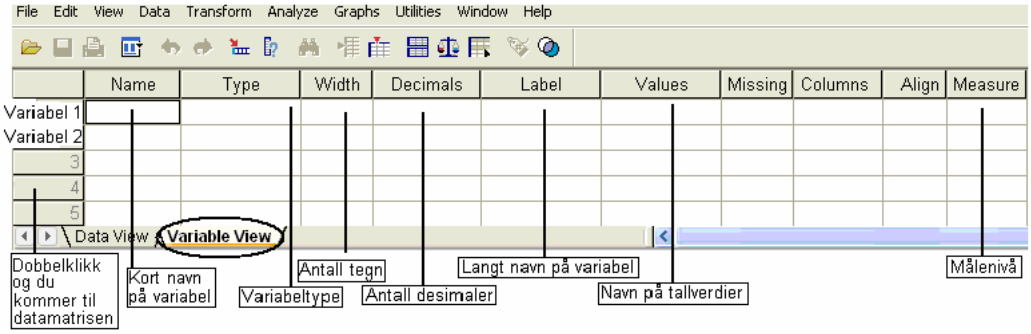



*Data view* er utformet som et ruteark med vertikale *kolonner* og horisontale *rader*:

a. Hver kolonne representerer en variabel
b. Hver rad representerer en enhet

[2] I tillegg har vi Syntax-vinduet der man må skrive inn kommandoer. Dette vinduet har sine fordeler når man skal ha store, presise operasjoner, men bruk av syntax-vinduet skal vi ikke si mye mer om.

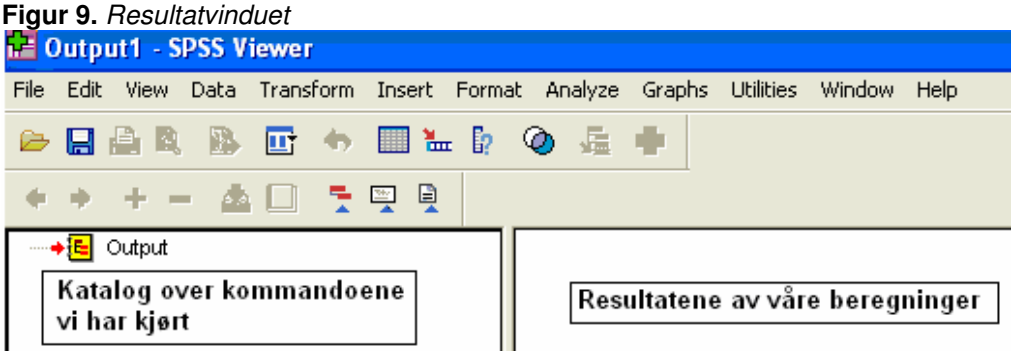
**Figur 8**. *Variabelmatrisen*




*Variable View* gir oss opplysninger om variablene som er med i datasettet, herunder:

a. informasjon om variablenes verdier
b. informasjon om variablenes målenivå

**Figur 9.** *Resultatvinduet*

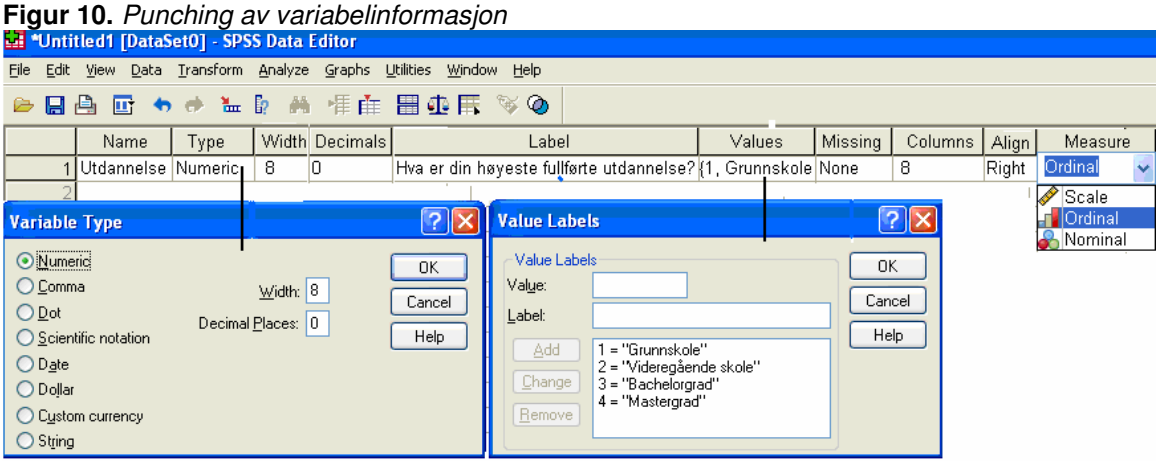



*Output* gir oss resultatene av de beregningene SPSS har utført. Outputvinduet kommer automatisk opp når vi kjører en beregning.

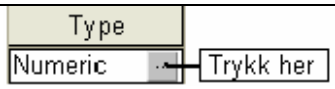
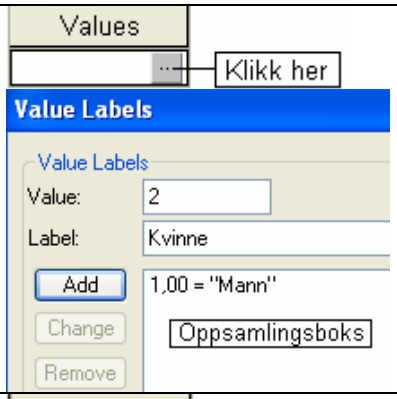
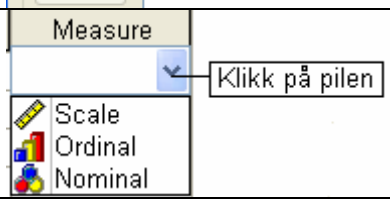
I SPSS bruker vi *Variable view* og *Data view* når vi skal registrere svar.

*Variable view* brukes til å registrere informasjon om variablene våre.

**Figur 10.** *Punching av variabelinformasjon*

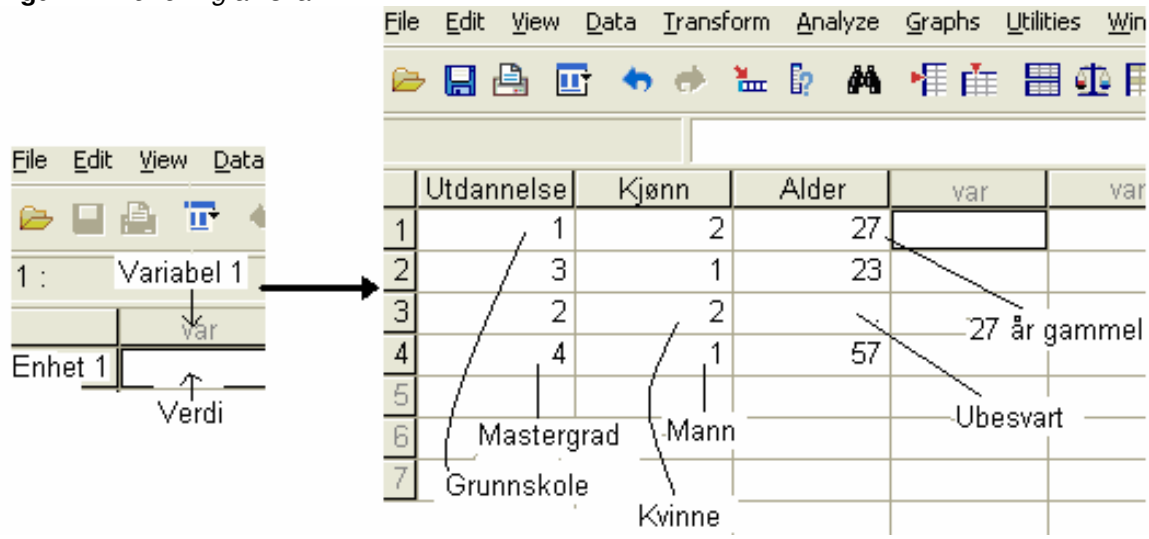


**Tabell 2.** *Punching av variabelinformasjon*

| Hva? | Symbol | Forklaring |
|---|---|---|
| Name: |  | Kort navn på variabel i ett ord |
| Type: |  | 1. Avgjøre hvilken type data vi skal registrere<br>- Numeric = talldata<br>- String = tekstdata<br>2. Width: Hvor mange tegn vi skal ha plass til<br>3. Decimal Places: Antall desimaler |
| Label: |  | Registrere spørsmålet vi har stilt |
| Values: |  | Om vi har talldata kobler vi verdinavn til tall gjennom boksen *Value Labels*. Det er tre rom:<br>*Value:* Her skriver vi inn verdien<br>*Value Label*: Her skriver vi inn det verdinavnet vi vil ha på verdien<br>*Oppsamlingsboks:* Når vi har fylt ut *Value* og *Label* trykker vi på *Add*, og det nye verdinavnet legges inn i oppsamlingsboksen<br>Eksempel: I variabelen Kjønn er verdi 1 = Mann og verdi 2 = Kvinne. Vi setter inn 2 i *Value*, Kvinne i *Label*, og trykker *Add* |
| Measure |  | Klikk på pilen for Measure og vi får da tre alternativer til målenivå:<br>*Scale:* Om variabel er på forholdstallsnivå<br>*Ordinal:* Om variabel er på ordinalnivå<br>*Nominal:* Om variabel er på nominalnivå |

*Data view* brukes til å registrere respondentenes svar.

**Figur 11.** *Punching av svar*



I figuren viser vi noen eksempler ut fra tre variabler: Utdannelse, Kjønn og Alder. For å være effektive puncher vi verdiene til hver enkelt enhet (vertikalt). Når vi puncher benyttes enten en *kodebok* eller at man allerede har skrevet inn *verdier i spørreskjemaet*. Se eksempler nedenfor.

**Tabell 3.** *Kodebok eller allerede utfylte verdier i spørreskjema*

Kodebok

| **Utdannelse** |
|---|
| 1 = Grunnskole |
| 2 = Videregående skole |
| 3 = Bachelorgrad |
| 4 = Mastergrad |

Avkrysset spørreskjema

*Hva er din høyeste fullførte utdannelse?*

| | |
|---|---|
| 1 | Grunnskole |
| X | Videregående skole |
| 3 | Bachelorgrad |
| 4 | Mastergrad |

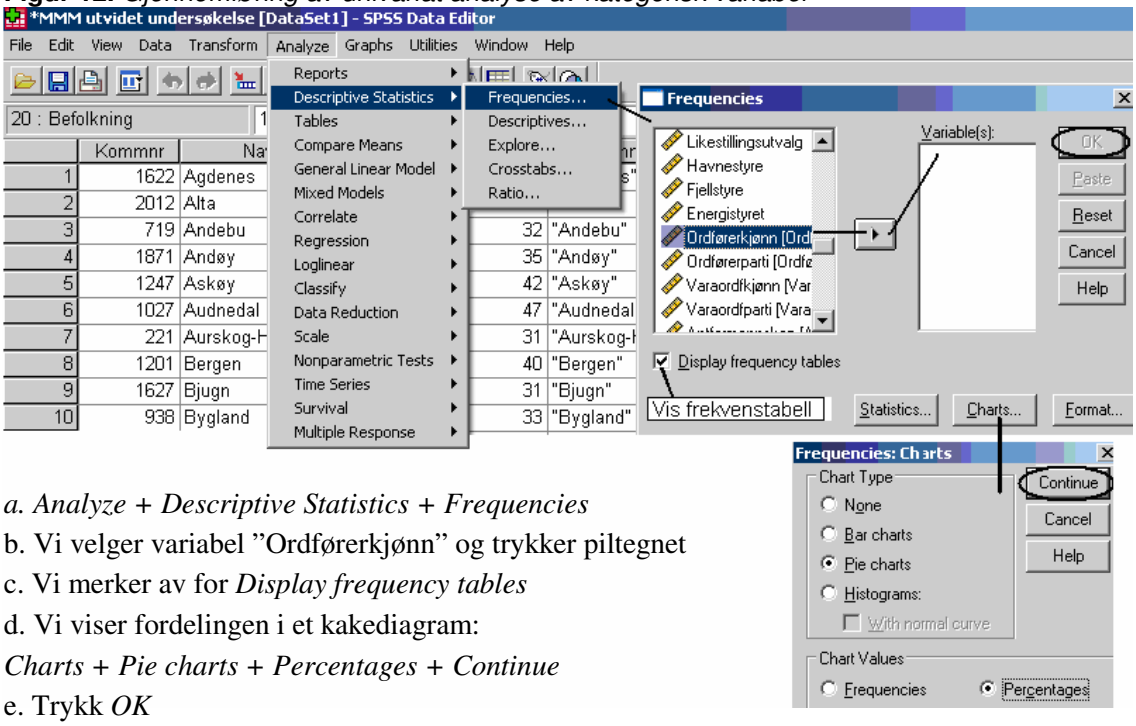
# 3 Beskrivende analyse

Vi ser i kapittel 3 på fordelingsanalyse for en variabel (univariat) og to variabler (bivariat).

## 3.1 Univariat analyse

Variablenes målenivå bestemmer hvordan vi analyserer dem. Kategoriske variabler, d.v.s. variabler på nominal- eller ordinalnivå, analyseres ut fra prosentuerte tabeller eller grafer. Gjennomføringen illustreres i figuren under (eks. kjønn på ordfører):

**Figur 12.** *Gjennomføring av univariat analyse av kategorisk variabel*




*a. Analyze + Descriptive Statistics + Frequencies*
b. Vi velger variabel "Ordførerkjønn" og trykker piltegnet
c. Vi merker av for *Display frequency tables*
d. Vi viser fordelingen i et kakediagram:
*Charts + Pie charts + Percentages + Continue*
e. Trykk *OK*

Den andre delen handler om å tolke Output-feltet - resultatene som vises i figuren på neste side.

**Figur 13.** *Resultater av univariat analyse av kategorisk variabel*

**Statistics**

Ordførerkjønn

| | | | |
|---|---|---|---|
| N | Valid | 433 | Antall svar |
| | Missing | 0 | Antall "hull" i datamatrisen |

**Ordførerkjønn**

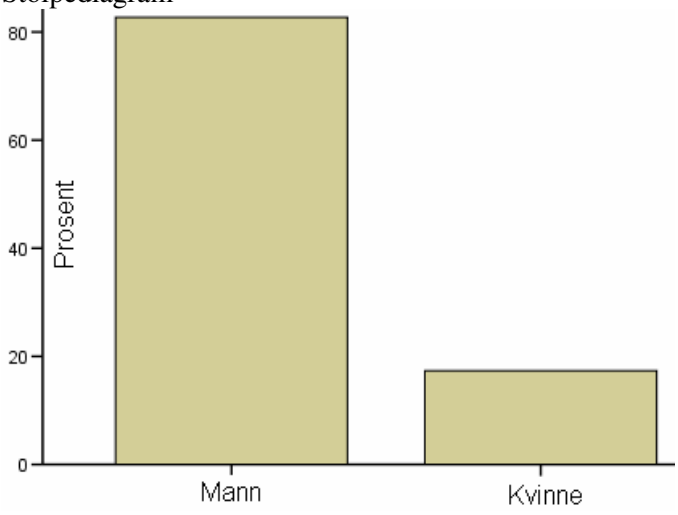
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | Mann | 358 | 82,7 | 82,7 | 82,7 |
| | Kvinne | 75 | 17,3 | 17,3 | 100,0 |
| | Total | 433 | 100,0 | 100,0 | |
| | | Antallet | Prosent | Prosent valide svar | Kummulativ prosent |

*Frequency* = Faktisk observerte antallet enheter med ulike verdier
*Percent* = Frekvensen delt på det totale antallet enheter
*Valid Percent* = Frekvensen delt på enhetene med oppgitt verdi
*Cumulative Percent* = Summen av prosentandelene fra den første verdien til den aktuelle verdi-en

Dersom vi skal bruke resultatene som vises i Output-vinduet i en rapport bør vi redigere den først. Det vi kan gjøre er å kopiere resultatene fra SPSS til andre program (eks. Word/Excel) og så lage grafer/tabeller. I figuren under presenteres resultatene fra vår analyse av variabelen *Ordførerkjønn*.

**Tabell 4.** *Presentasjon av variabelen Ordførers kjønn.*

Stolpediagram




N = 433

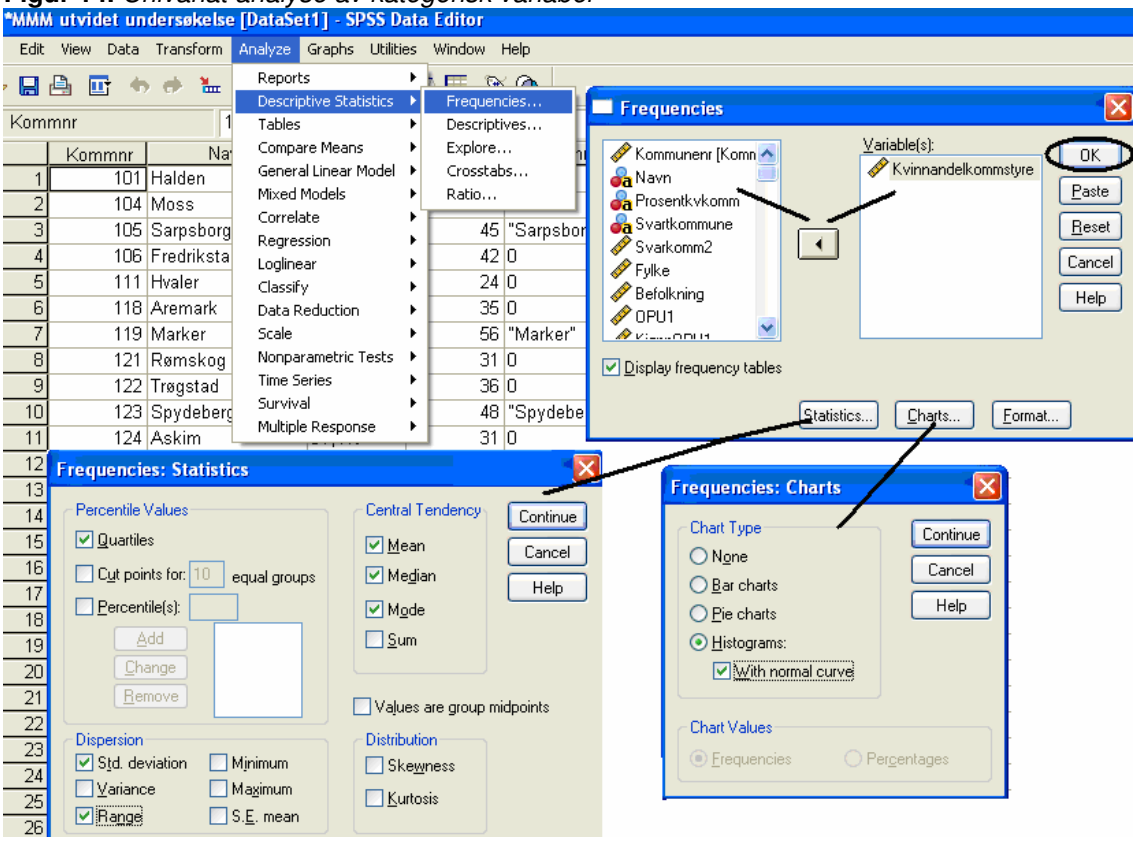
Tabell

| | Antall | Prosent |
|---|---|---|
| Mann | 358 | 82,7 |
| Kvinne | 75 | 17,3 |
| Totalt | 433 | 100 |

Kort konklusjon
Nesten 5 av 6 ordførere er menn, mens 1 av 6 ordførere er kvinner.

Når vi har kontinuerlige variabler, d.v.s. variabler på forholdstalls-, intervall- eller ordinalnivå med mange verdier, blir det for mange verdier til å få en god oversikt over fordelingen ved å se på hver enkelt verdis prosentandel. Vi benytter statistiske mål på fordelingens tyngdepunkt (gjennomsnitt, median modus) og spredning (variasjonsbredde, kvartiler, standardavvik). I vårt eksempel brukes en variabel som viser til andel kvinner i kommunestyret:

**Figur 14.** *Univariat analyse av kategorisk variabel*




*a. Analyze + Descriptive Statistics + Frequencies*

b. Vi velger variabelen "Kvinnandelkommstyre" og trykker på piltegnet

c. Vi merker ikke av for *Display frequency tables*

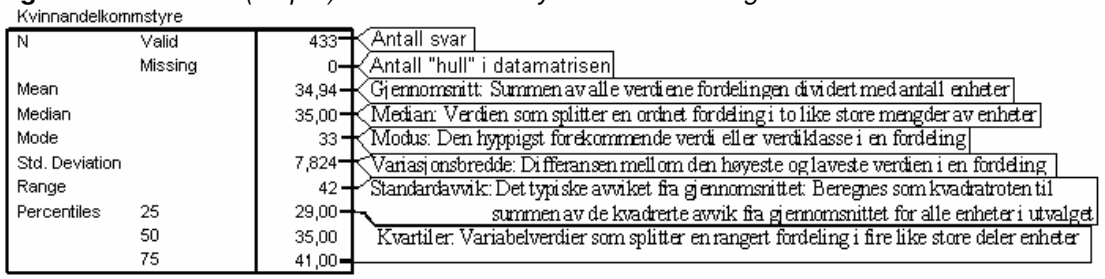
*d.* Vi velger statistiske mål: Trykk på *Statistics + Quartiles + Mean + Median + Mode + Std. Deviation + Range + Continue*

e. Vi viser fordelingen i et histogram: *Charts + Histograms + With normal curve + Continue*

f. Trykk *OK*

Den andre delen handler om å tolke Output-feltet, d.v.s. resultatene som vises på neste side

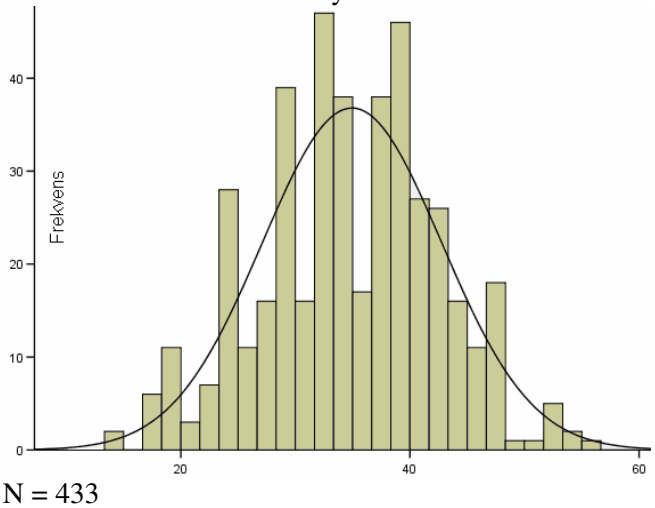
**Figur 15.** *Resultater (output) av univariat analyse av kontinuerlig variabel*




Dersom vi ønsker å bruke tabellen/figuren som vises i Output-vinduet bør vi redigere den først. Vi kan presentere resultatene av variabelen *Kvinneandel i kommunestyret* på ulike måter.

**Tabell 5.** *Presentasjon av variabelen Kvinneandel i kommunestyret.*

*Histogram*
Andel kvinner i kommunestyret




N = 433

*Tabell.*
*Mål på s*predning og sentraltendens for andel kvinner i kommunestyret

| **Mål** | **Resultat** |
|---|---|
| Sentraltendens | |
| Gjennomsnitt | 34,9 |
| Median | 35 |
| Modus | 33 |
| Spredningsmål | |
| Variasjonsbredde | 42 |
| Standardavvik | 7,8 |
| Kvartil 1 | <0-29> |
| Kvartil 2 | <30-35> |
| Kvartil 3 | <36-40> |
| Kvartil 4 | <41- 71> |

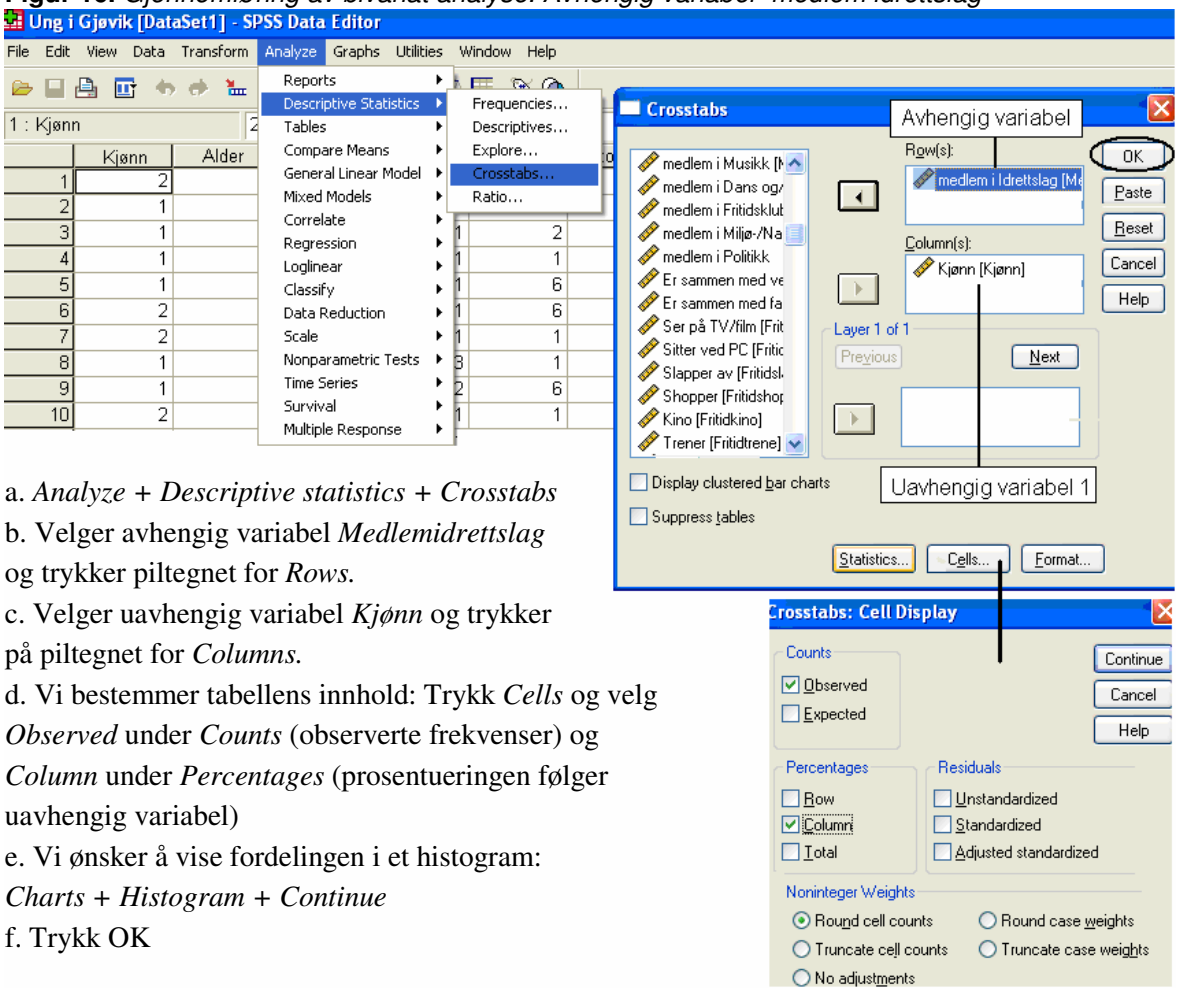
Kort konklusjon

Variasjonen mellom kommunene er stor – 42 prosentpoeng – men halvparten av kommunene har mellom 29 og 41 prosent kvinnelige representanter. Det typiske avviket fra gjennomsnittet er lite, under 8 prosentpoeng. Sentraltendensmålene ligger svært nært hverandre.

## 3.2 Bivariat analyse

Valg av bivariat analyse avhenger av variablenes målenivå. Når begge variabler er på nominalnivå eller ordinalnivå med få verdier benyttes prosentuerte krysstabeller/grafer. Eks. Medlemskap i idrettslag (ja/nei) i forhold til kjønn (mann/kvinne).

**Figur 16.** *Gjennomføring av bivariat analyse. Avhengig variabel "medlem idrettslag"*




a. *Analyze* + *Descriptive statistics* + *Crosstabs*
b. Velger avhengig variabel *Medlemidrettslag* og trykker piltegnet for *Rows.*
c. Velger uavhengig variabel *Kjønn* og trykker på piltegnet for *Columns.*
d. Vi bestemmer tabellens innhold: Trykk *Cells* og velg *Observed* under *Counts* (observerte frekvenser) og *Column* under *Percentages* (prosentueringen følger uavhengig variabel)
e. Vi ønsker å vise fordelingen i et histogram: *Charts* + *Histogram* + *Continue*
f. Trykk OK

Den andre delen handler om å tolke Output-feltet (neste side).

- Den øverste tabellen forteller hvor mange enheter som har valide verdier, det vil si at de inkluderes i analysen, og hvor mange som ikke har valide svar.
- Den neste tabellen tolkes vertikalt: 59 prosent av guttene og 48 prosent av jentene er medlem av idrettslag, totalt er 52 prosent medlem i idrettslag.

**Figur 17.** *Resultater av bivariat analyse. Avhengig variabel "medlem idrettslag"*

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| Er du medlem i idrettslag * Kjønn | 1288 | 99,2% | 10 | ,8% | 1298 | 100,0% |

Antallet som er inkludert i analysen

Antallet som ikke er inkludert i analysen

Antallet og andelen i hele utvalget

Andelen av utvalget som er inkludert i analysen

Andelen av utvalget som ikke er inkludert i analysen

**Er du medlem i idrettslag * Kjønn Crosstabulation**

Uavhengig variabel

| | | | Kjønn | | Total |
|---|---|---|---|---|---|
| | | | Gutt | Jente | |
| Er du medlem i idrettslag | Nei | Count | 227 | 387 | 614 |
| | | % within Kjønn | 41,2% | 52,5% | 47,7% |
| | Ja | Count | 324 | 350 | 674 |
| | | % within Kjønn | 58,8% | 47,5% | 52,3% |
| Total | | Count | 551 | 737 | 1288 |
| | | % within Kjønn | 100,0% | 100,0% | 100,0% |

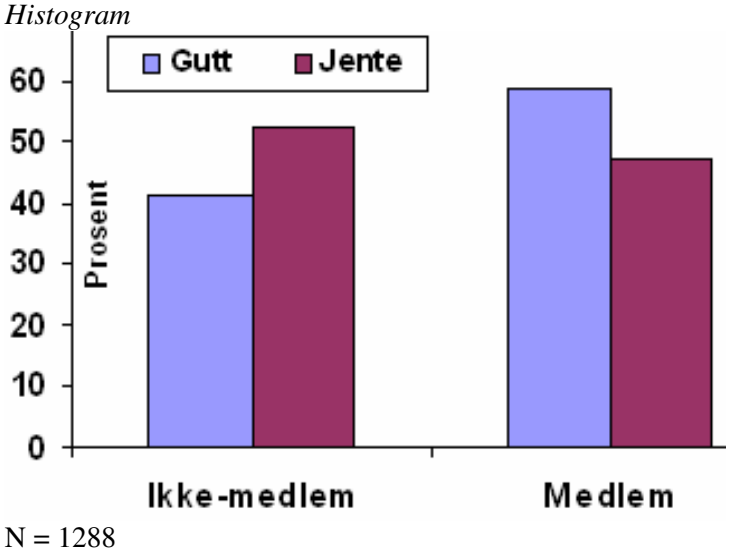
Marginalfordelingen

Avhengig variabel

Dersom vi ønsker å bruke tabellen som vises i Output-vinduet bør vi redigere den først. Vi kan presentere resultatene av den bivariate analysen på ulike måter.

**Tabell 6.** *Presentasjon av sammenhengen mellom kjønn og medlemskap i idrettslag.*

*Histogram*




N = 1288

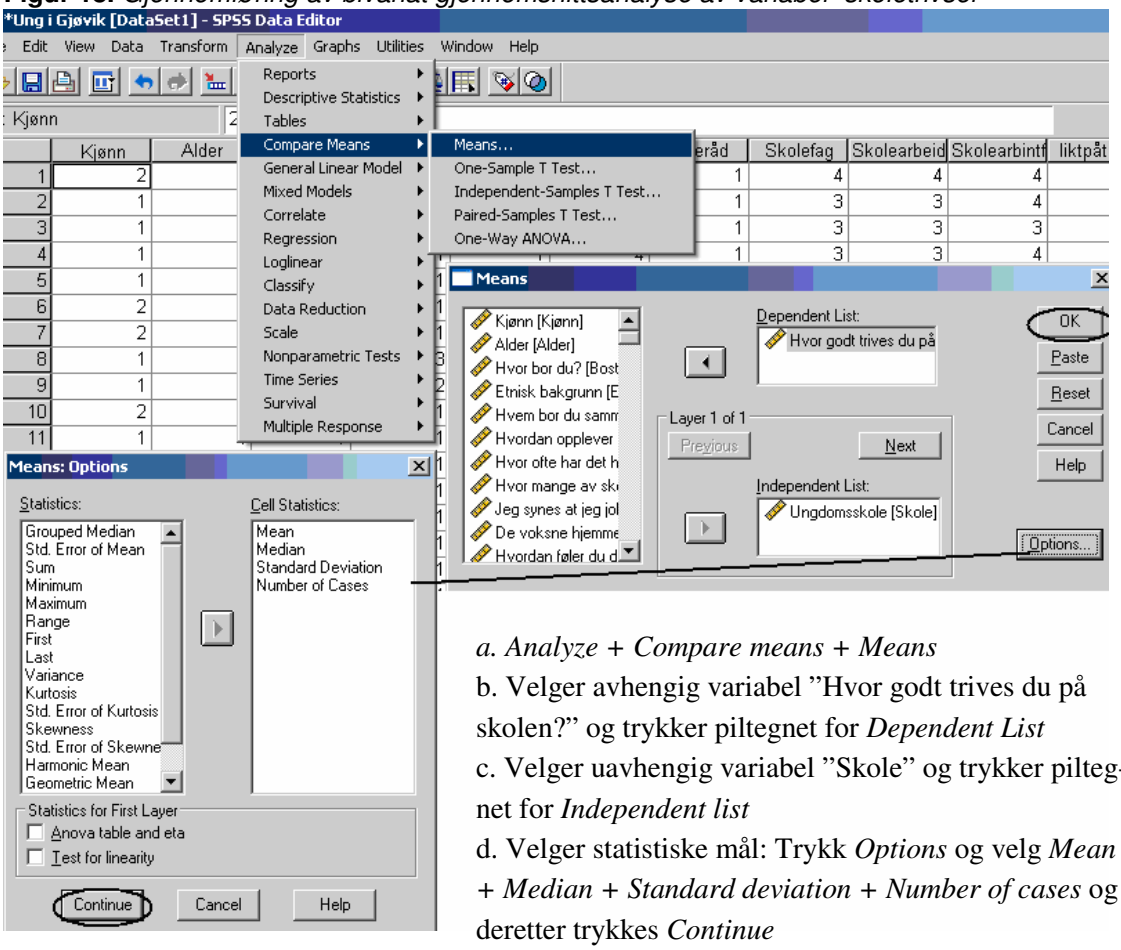
*Tabell*

| | Gutt | Jente | Totalt |
|---|---|---|---|
| Ikke-medlem | 41 | 53 | 47 |
| Medlem | 59 | 47 | 52 |
| Sum | 100 | 100 | 100 |
| N | 551 | 737 | 1288 |

Kort konklusjon: Gutter i utvalget er oftere medlem av idrettslag komparert med jentene.

Andre ganger er avhengig variabel på forholdstalls-, intervall- eller ordinalnivå med mange verdier, mens den andre variabelen er på ordinalnivå med få verdier eller nominalnivå. Ved slike tilfeller kan vi sammenligne statistiske mål som gjennomsnitt, median og standardavvik. Analysen går ut på å studere forskjeller mellom verdier på uavhengig variabel, og gjennomføres slik:

**Figur 18.** *Gjennomføring av bivariat gjennomsnittsanalyse av variabel "skoletrivsel"*




*a. Analyze + Compare means + Means*

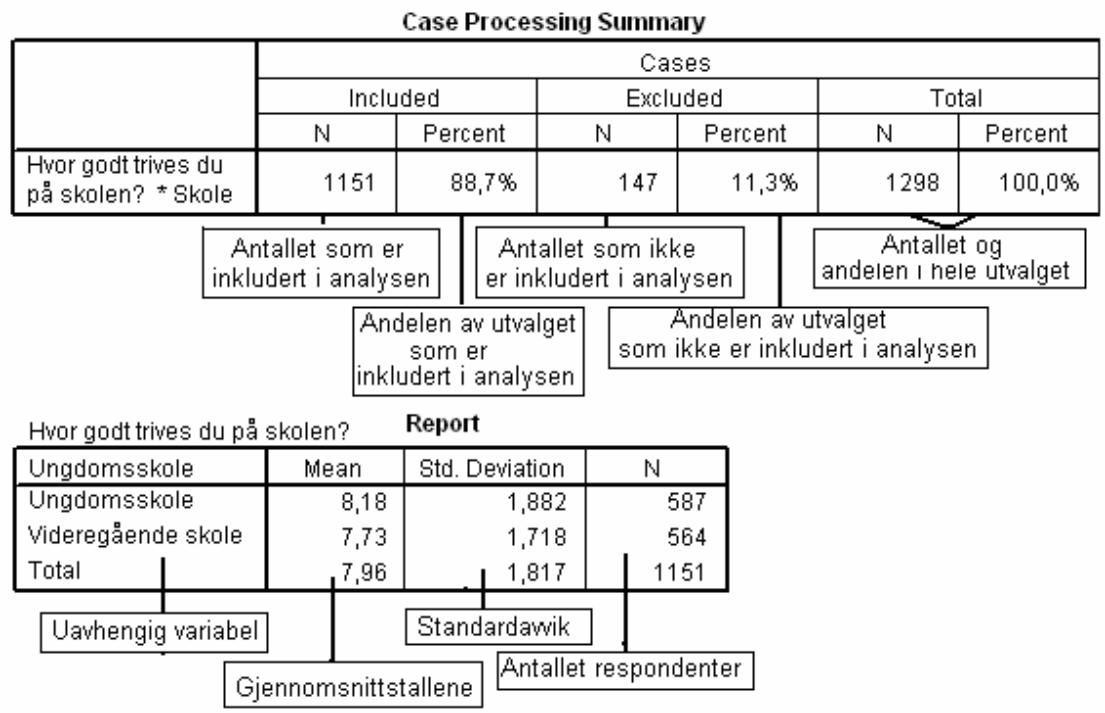
b. Velger avhengig variabel "Hvor godt trives du på skolen?" og trykker piltegnet for *Dependent List*

c. Velger uavhengig variabel "Skole" og trykker piltegnet for *Independent list*

d. Velger statistiske mål: Trykk *Options* og velg *Mean + Median + Standard deviation + Number of cases* og deretter trykkes *Continue*

Den andre delen handler om å tolke Output-feltet (neste side).

**Figur 19.** *Resultater av gjennomsnittsanalyse av variabel "skoletrivsel"*



**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Included | | Excluded | | Total | |
| | N | Percent | N | Percent | N | Percent |
| Hvor godt trives du på skolen? * Skole | 1151 | 88,7% | 147 | 11,3% | 1298 | 100,0% |

Antallet som er inkludert i analysen
Antallet som ikke er inkludert i analysen
Antallet og andelen i hele utvalget
Andelen av utvalget som er inkludert i analysen
Andelen av utvalget som ikke er inkludert i analysen

**Report**

Hvor godt trives du på skolen?

| Ungdomsskole | Mean | Std. Deviation | N |
|---|---|---|---|
| Ungdomsskole | 8,18 | 1,882 | 587 |
| Videregående skole | 7,73 | 1,718 | 564 |
| Total | 7,96 | 1,817 | 1151 |

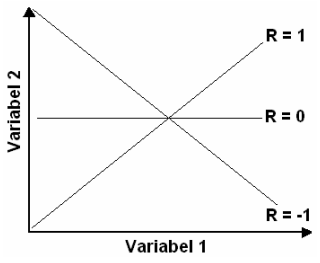
Uavhengig variabel
Standardavvik
Gjennomsnittstallene
Antallet respondenter

Den øverste tabellen forteller hvor mange enheter som har valide verdier, det vil si at de inkluderes i analysen, og hvor mange som ikke har valide svar. Tabellen under er den vi konkluderer ut fra:

a. Gjennomsnittet for elever på ungdomsskolen i utvalget er 8,2 mens gjennomsnittet er 7,7 for elever på videregående skole
b. Standardavvikene viser videre at svarene til elevene på ungdomsskolen er noe mer spredt enn svarene til elevene på videregående skole
c. Den siste delen N viser til antallet respondenter

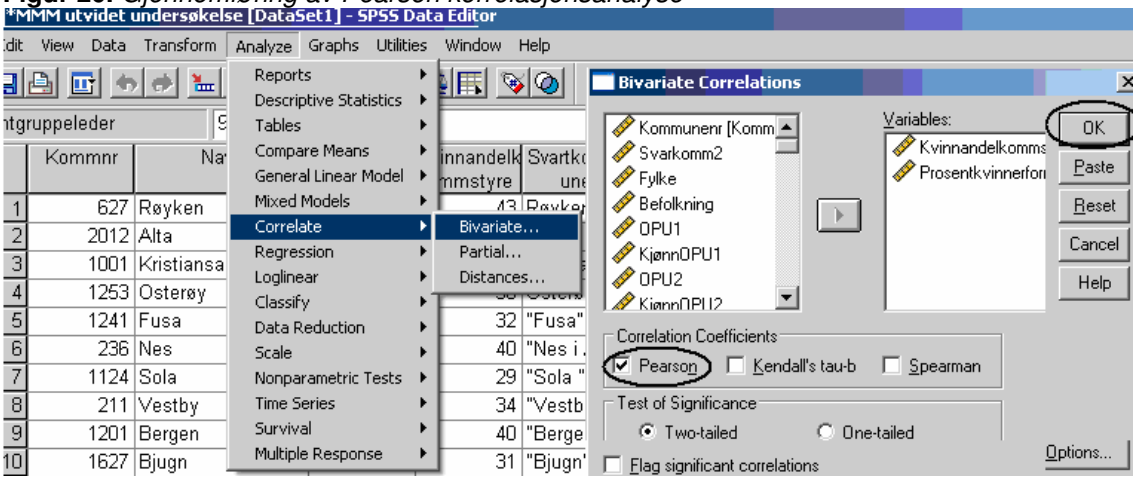
Når begge variabler er på forholdstalls-, intervall- eller ordinalnivå med mange verdier, kan vi foreta Pearson korrelasjonsanalyse.

Korrelasjonen, ved bruk av Pearsons korrelasjonsmål r, angir både hvilken retning samvariasjon mellom variabler går og hvor sterk sammenhengen er:
Ved r = -1 er det en perfekt negativ korrelasjon
Ved r = 0 er det ingen korrelasjon
Ved r = 1 er det en perfekt positiv korrelasjon.




Korrelasjonstesten gjennomføres slik. I eksempelet på neste side ser vi på andelen *kvinner i kommunestyret* og andelen *kvinner i formannskapet.*

**Figur 20.** *Gjennomføring av Pearson korrelasjonsanalyse*




*a. Analyze + Correlate + Bivariate*

*b.* Vi merker av for "Kvinnandelkommstyre" og "Prosentkvinnerformannskap" og trykker på pilen.

c. Vi haker av for Pearson som korrelasjonskoeffisient

d. Trykk OK

Den andre delen handler om å tolke Output-feltet.

**Figur 21.** *Resultater av Pearson korrelasjonsanalyse*

**Correlations**

| | | Kvinnandel kommstyre | Prosentkvinne rformannskap |
|---|---|---|---|
| Kvinnandelkommstyre | Pearson Correlation | 1 | ,332 |
| | Sig. (2-tailed) | | ,000 |
| | N | 433 | 330 |
| Prosentkvinnerforman nskap | Pearson Correlation | ,332 | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 330 | 330 |

Korrelasjonskoeffisienten r

Signifikansnivå

Antall enheter

I den første ruten og siste ruten ser vi korrelasjonen for hver variabel satt opp mot seg selv, og korrelasjonen er naturligvis 1. De rutene som er interessante er når variablene krysser hverandre. Vi ser i rute 2 (og 3) at Pearson Correlation er 0,33, signifikansnivået er lavere enn 0,000 og antallet enheter (N) er lik 330. Vi konkluderer med at korrelasjonen mellom andel kvinner i kommunestyret og andel kvinner i formannskapet er svak og positiv (+ 0,33).

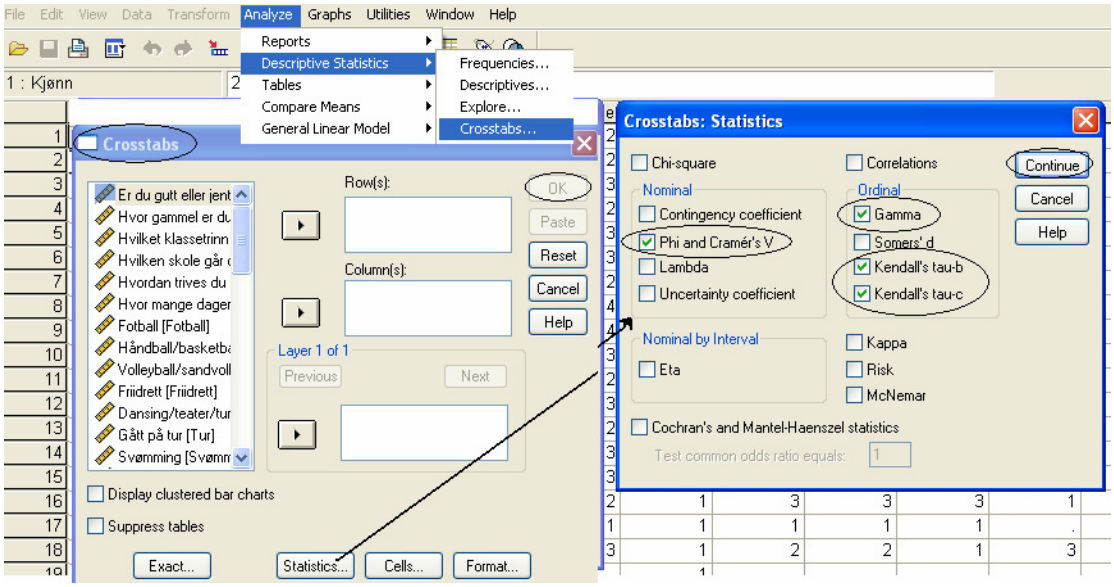
**Boks 1. Andre korrelasjonsmål**

Det finnes også en rekke andre korrelasjonsmål som måler styrken i sammenhengen mellom to variabler. Når en avgjør hva slags korrelasjonsmål en ønsker å bruke må en se på målenivået og det er den variabelen med det laveste målenivået som bestemmer hva slags mål vi kan bruke. Vi skal ikke se eksplisitt på disse målene, men gi en kort oversikt over dem.

a. Nominalnivå: Phi og Cramer's V

b. Ordinalnivå: Gamma, Kendalls's tau-b og Kendall's tau-c

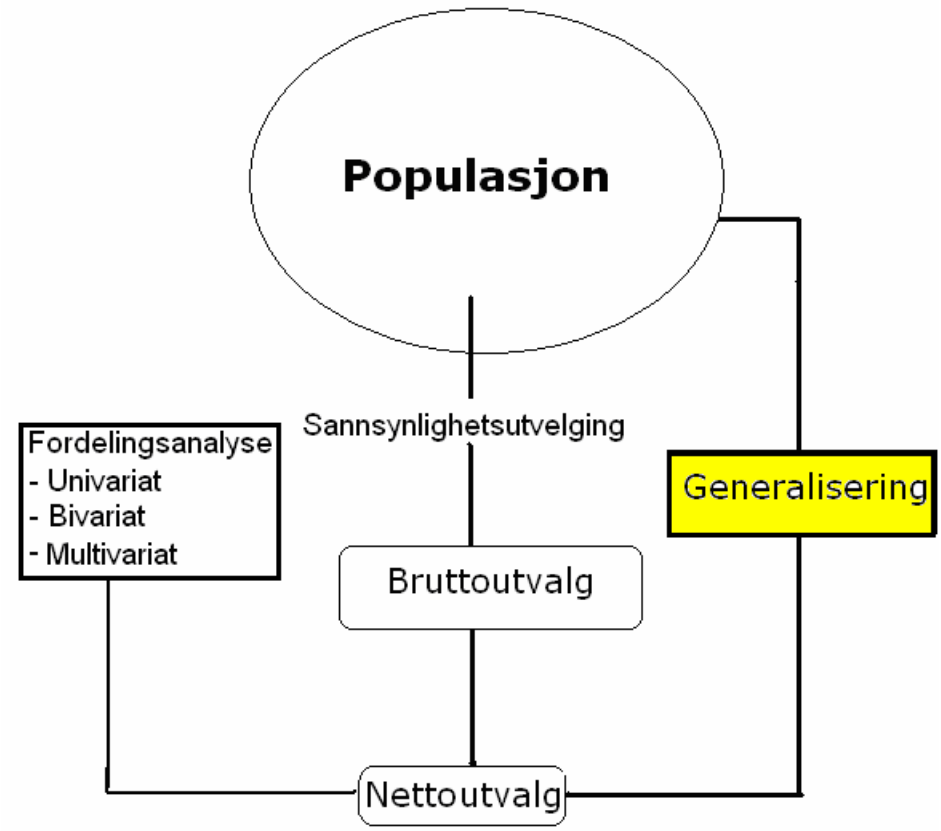
c. Kontinuerlig variabel: Pearson r (som vi har gjennomgått)



# 4 Estimering

Når vi bruker et utvalg av populasjonen som grunnlag for våre analyser er det knyttet en viss usikkerhet til resultatene når vi generaliserer fra utvalget til populasjonen. Vi kjenner jo ikke fordelingen i populasjonen, men vi kan generalisere resultater fra utvalget til populasjonen.

**Figur 22.** *Slutte om fordelingen i populasjonen basert på utvalgsfordelingen*




Generalisering

a. Estimere: Anslå en parameter i populasjonen ved hjelp av en estimator i utvalget.
b. Hypotesetesting: Undersøke om forskjeller mellom utvalg kan generaliseres til å gjelde for populasjoner.

## 4.1 Estimering av populasjonsgjennomsnitt

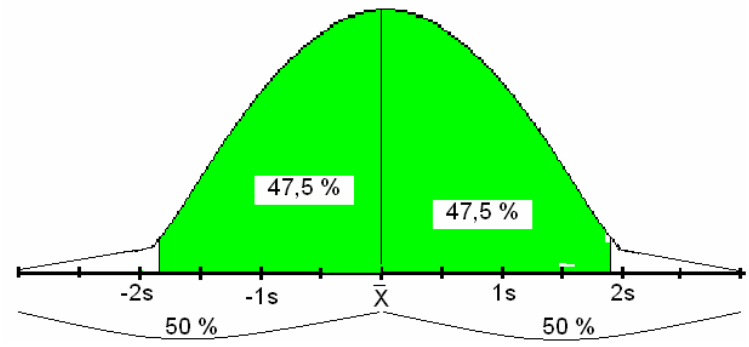
Når vi har kontinuerlige variabler (forholdstall, intervall og ordinalnivå med mange verdier) beregnes utvalgsgjennomsnittet og standardavvik. Vi kan bruke variabelens utvalgsgjennomsnitt og standardavvik til å estimere populasjonsgjennomsnittet. Tre forhold spiller inn på hvor presist dette estimatet blir:

a. Utvalgets størrelse: Flere enheter gir sikrere estimat
b. Spredningen: Dess større standardavvik, dess større estimatområde
c. Valg av sikkerhetsintervall: Dess sikrere vi vil være, dess større estimatområde

I slutningsstatistikk er det en konvensjon å benytte seg av en sikkerhetsmargin på 95 prosent når vi beregner populasjonsgjennomsnitt. Det er da 2,5 prosent sannsynlighet for at populasjons-gjennomsnittet er større estimatet og 2,5 prosent sannsynlighet for at populasjonsgjennomsnittet er mindre. Sikkerhetsmarginen kalles konfidensintervall.

Ved 1,96 standardfeil vil 95 prosent av de mulige utvalgsgjennomsnittene vi kunne trukket, befinne seg. Det er m.a.o. 95 prosent sannsynlig at vi har rett når vi sier at populasjonsgjennom-snittet befinner seg innenfor et område på +/- 1,96 standardfeil.

**Figur 23.** *Illustrasjon av konfidensintervall*



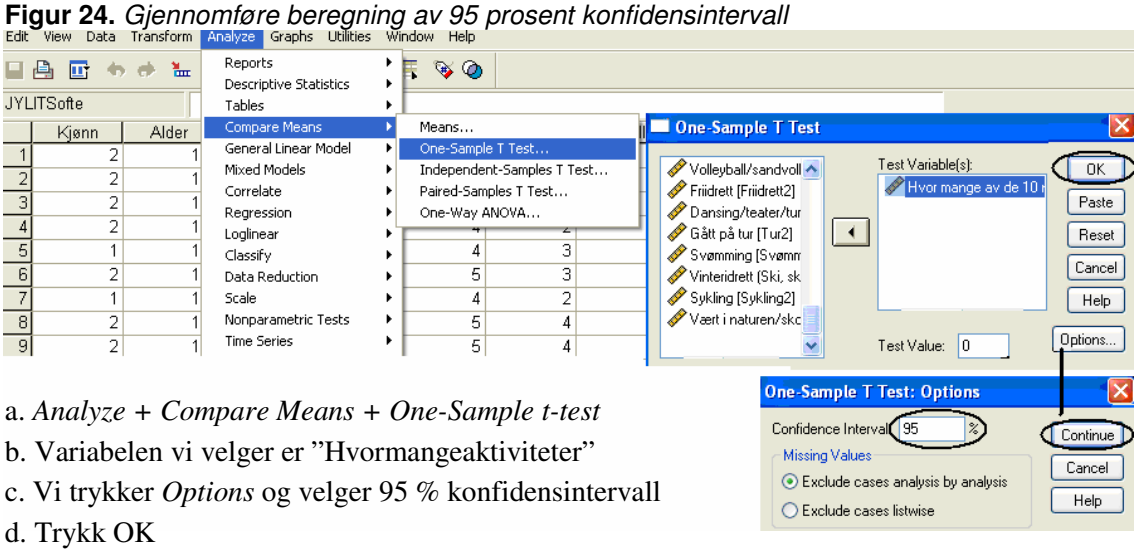

Vi skal nå vise tradisjonell og moderne utregning av konfidensintervall.

**Tabell 7.** *Tradisjonell beregning av konfidensintervall*

| | |
|---|---|
| Beregne standardfeilen (Samplingsfordelingens estimerte standardavvik) | $\text{Standardfeil til gjennomsnittet} = \frac{\text{Standardavvik}}{\sqrt{\text{Antall enheter}}}$ |
| Eks: Snitthøyden for jenter i et utvalg på 200 er 168 cm. Standardavvik på 10 cm.<br>$1\,SF = \frac{\text{Standardavvik}}{\sqrt{\text{Antall enheter}}} = \frac{10}{\sqrt{200}} = 0{,}7$<br>95 % sikkerhetsmarginen = 0,7 * (+/- 1,96 standardfeil) = +/- 1,4 cm | |
| Konklusjon: Det er 95 prosent sikkert når vi konkluderer med at gjennomsnittshøyden blant jenter i populasjonen er i intervallet 166,6 til 169,4 cm. | |

Vi kan også beregne et 95 prosent konfidensintervall i SPSS. I eksempelet vårt har et utvalg av elever i en norsk storby krysset av for hvilke aktiviteter som de har hatt i gymtimene. I alt kan man krysse av for 10 aktiviteter. Vi skal beregne gjennomsnittet for populasjonen av elever på ungdomsskolene i storbyen i forhold til antall aktiviteter.

**Figur 24.** *Gjennomføre beregning av 95 prosent konfidensintervall*




a. *Analyze* + *Compare Means* + *One-Sample t-test*
b. Variabelen vi velger er "Hvormangeaktiviteter"
c. Vi trykker *Options* og velger 95 % konfidensintervall
d. Trykk OK

Den andre delen handler om å tolke output-feltet.

**Figur 25.** *Tolkning av resultater for beregning av 95 prosent konfidensintervall*

**One-Sample Statistics**

| | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| Hvor mange av de 10 nevnte aktivitetene har elevene deltatt i? | 551 | 3,97 | 2,582 | ,110 |

**Gjennomsnitt** **Standardavvik**

**One-Sample Test**

| | Test Value = 0 | | | | | |
|---|---|---|---|---|---|---|
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference Lower | Upper |
| Hvor mange av de 10 nevnte aktivitetene har elevene deltatt i? | 36,085 | 550 | ,000 | 3,969 | 3,75 | 4,19 |

**Laveste populasjonsgjennomsnitt** **Høyeste populasjonsgjennomsnitt**

I første tabell er det to viktige resultater: Gjennomsnittet = 3,97 og Standardavvik = 2,58.
I andre tabell ses konfidensintervallet: Gjennomsnittet i populasjonen kan bli så lavt som 3,75 og så høyt som 4,19.

**Konklusjonen:** Med 95 prosent sannsynlighet har populasjonen av elevene i storbyen gjennomført mellom 3,75 og 4,2 av de 10 utvalgte aktiviteter.

## 4.2 Estimering av feilmargin

Ved variabler på nominal eller ordinalnivå med få verdier estimeres konfidensintervall for prosentfordelinger i utvalg. Vi foretar kun tradisjonell beregning av feilmarginer.

**Eks:** I en landsdekkende undersøkelse, med et utvalg på 700, sier 9 prosent av ungdom mellom 13 og 15 år at de mobbes. Hva er feilmarginen med 95 prosent sikkerhet?
Først beregnes standardfeilen til prosentandelen.

$$SF(p) = \sqrt{\frac{p(100-p)}{n}} \qquad \underset{\text{Mobbing}}{SF(p)} = \sqrt{\frac{p(100-p)}{n}} = \sqrt{\frac{9(100-9)}{700}} = \sqrt{1{,}17} = 1{,}08$$

Så beregnes feilmarginen, og den er gitt ved: +/- 1.96 SF * 1.08 = 2,1 prosent.

**Tabell 8.** *Huskeregler angående variasjoner i feilmargin*

Feilmarginen blir mindre med økende utvalgsstørrelse

| Antall | Potensiell feilmargin ved 95 % sikkerhet |
|---|---|
| 500 respondenter | +/- 4,4 |
| 1000 respondenter | +/- 3,1 |
| 10000 respondenter | +/- 1 |

Feilmarginen varierer med resultatet

| | Feilmargin ved 95 % sikkerhet | | |
|---|---|---|---|
| Antall | 10 prosent | 50 prosent | 90 prosent |
| 500 respondenter | +/- 2,6 | +/- 4,4 | +/- 2,6 |
| 1000 respondenter | +/- 1,9 | +/- 3,1 | +/- 1,9 |

Feilmarginen varierer med størrelsen på populasjonen

| | Potensiell feilmargin ved 95 % sikkerhet | | |
|---|---|---|---|
| Antall | 2000 i pop | 10000 i pop | Uendelig pop |
| 500 respondenter | +/- 3,8 | +/- 4,2 | +/- 4,4 |
| 1000 respondenter | +/- 2,2 | +/- 2,9 | +/- 3,1 |

# 5 Hypotesetesting

Når vi har påpekt en sammenheng mellom to variabler, er det interessant å undersøke om denne sammenhengen er signifikant eller ikke. Forskjeller mellom utvalg må være av en viss størrelse for at vi skal konkludere med at det også er forskjell mellom de respektive populasjonene.

Prinsippet bak hypotesetesting er å formulere en nullhypotese ($H_0$) som sier at det ikke er differanse mellom populasjonene og en alternativ hypotese ($H_1$) om at det er forskjell mellom populasjonene. Så beregner vi, ved bruk av en testobservator, om det er sannsynlig at $H_0$ bør beholdes eller forkastes.

## 5.1 Kjikvadrattesten

Når avhengig variabel er på nominalnivå, eller på ordinalnivå med få verdier, benyttes kjikvadrattesten for å teste om resultater i utvalget kan generaliseres til populasjonen.

**Tradisjonell utregning**
Det antas å være en sammenheng mellom kjønn og medlemskap i idrettslag.
1. Vi formulerer hypoteser:
   $H_1$ = Det er en sammenheng mellom kjønn og medlemskap i idrettslag
   $H_0$ = Det er ikke en sammenheng mellom kjønn og medlemskap i idrettslag
2. Vi bestemmer oss for å benytte oss kjikvadrattesten.
3. Vi velger et signifikansnivå på 0.05, d.v.s. 5 prosent usikkerhet ved resultat
4. Beregne tabellens kjikvadratverdi

**Observert fordeling (O)**

| Medlem i idrettslag | Kjønn | | Total |
|---|---|---|---|
| | Gutt | Jente | |
| Ja | 324 | 350 | 674 |
| Nei | 168 | 281 | 449 |
| Total | 492 | 631 | 1123 |

**Estimert fordeling (E)**

| | | Kjønn | | Total |
|---|---|---|---|---|
| Medlem i Idrettslag | | Gutt | Jente | |
| | Ja | | | 674 |
| | Nei | | | 449 |
| Total | | 492 | 631 | 1123 |

| Rute | O | E | O-E | $(O-E)^2$ | $(O-E)^2/E$ |
|---|---|---|---|---|---|
| 1 | 324 | 295 | 29 | 841 | 2,85 |
| 2 | 350 | 379 | -29 | 841 | 2,22 |
| 3 | 168 | 197 | -29 | 841 | 4,27 |
| 4 | 281 | 252 | -29 | 841 | 3,34 |
| Sum | 1123 | 1123 | 0 | | $X^2 = 12,68$ |

$$E = \frac{\text{n (kolonne) x n (rekke)}}{\text{n (hele tabellen)}}$$

**Beregninger**

Gutt medlem = (674 x 492)/1123 = 295,3
Jente medlem = (674 x 631)/1123 = 378,7
Gutt ikke medlem = (449 x 492)/1123 = 196,7
Jente ikke medlem = (449 x 631)/1123 = 252,3

5. Beregning av kritisk verdi (signifikansnivået (0.05) og antallet frihetsgrader (1).[3]

| Antall frihetsgrader | 5 prosent | 1 prosent |
|---|---|---|
| 1 | **3,84** | 6,64 |
| 2 | 5,99 | 9,21 |


6. Sammenligning: Tabellens kjikvadratverdi (12,7) er høyere enn kritisk verdi (3,84).

**Konklusjon:** Med 95 prosent sikkerhet kan vi slutte at vi også vil finne kjønnsforskjeller når det gjelder medlemskap i idrettslag i populasjonene av gutter og jenter.

**Moderne utregning**

Vi antar at det er en sammenheng mellom kjønn og medlemskap i dans-/teatergruppe.

1. Vi formulerer hypoteser:

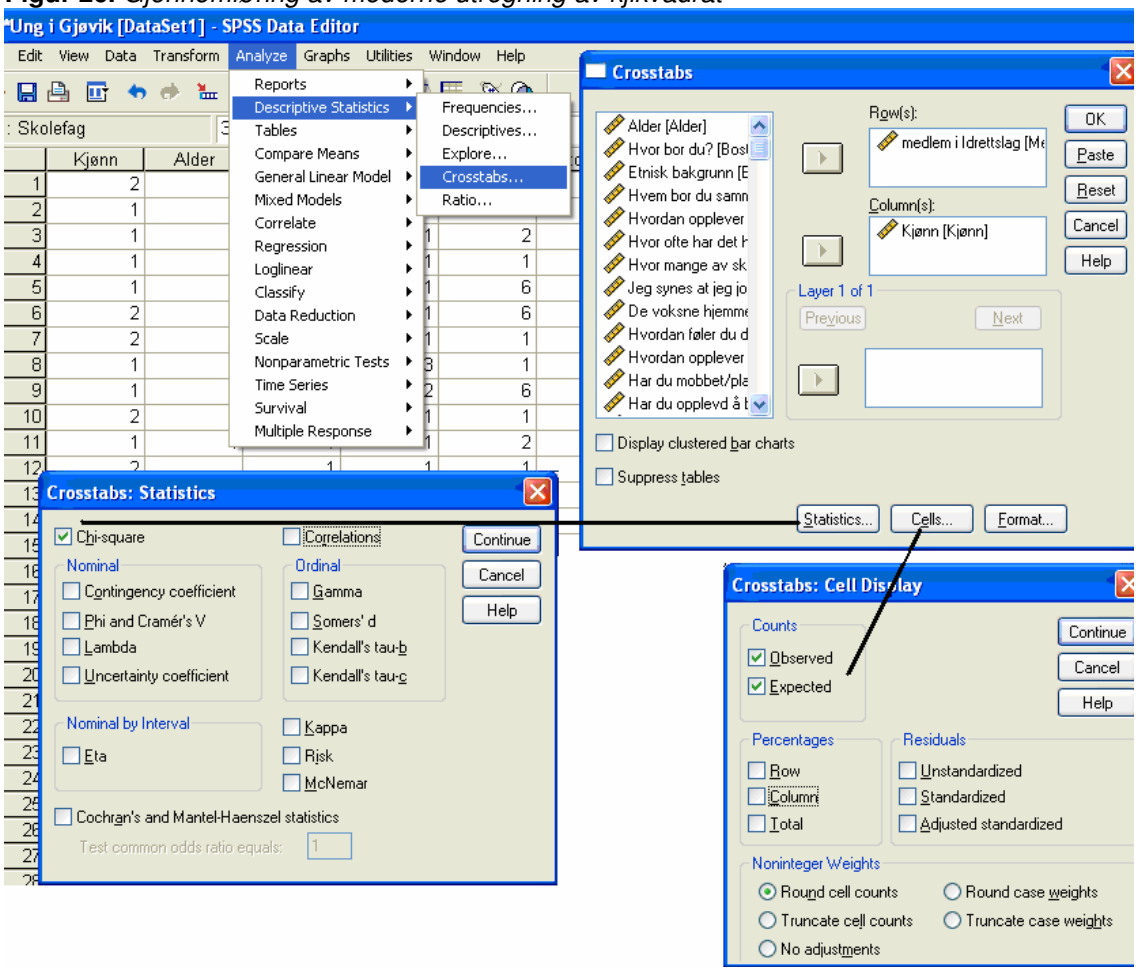
   $H_1$ = Det er en sammenheng mellom kjønn og medlemskap i dans-/teatergruppe.

   $H_0$ = Det er ikke en sammenheng mellom kjønn og medlemskap i d-/t-gruppe.

2. Vi bestemmer oss for kjikvadrattesten og velger et signifikansnivå på 0.05.
3. Finne utvalgsobservatoren

[3] Formel: (m-1)x(n-1), der m er antallet verdier på den første variabelen (kjønn) og n er antallet verdier på den andre variabelen (medlemskap i idrettslag).

**Figur 26.** *Gjennomføring av moderne utregning av kjikvadrat*




a. *Analyze* + *Descriptive statistics* + *Crosstabs.*

b. Vi velger avhengig variabel ”Medlemdans” og trykker piltegnet for *Rows.*

c. Vi velger uavhengig variabel ”Kjønn” og trykker piltegnet for *Columns.*

d. Vi bestemmer ”celleinnholdet”: Trykk *Cells* og velg *Observed og Expected* under *Counts* (observerte og estimerte frekvenser) og *Column.*

e. Vi velger kjikvadrattesten: Trykk på *Statistics* og merk av for *Chi-square.*

Den andre delen handler om å tolke Output-feltet, d.v.s. resultatene på neste side.

**Figur 27.** *Resultater (output) av kjikvadrat-test*

**Meddans2 * Kjønn Crosstabulation**

| | | | Kjønn | | Total |
|---|---|---|---|---|---|
| | | | Gutt | Jente | |
| Meddans2 | Nei | Count | 529 | 645 | 1174 |
| | | Expected Count | 502,2 | 671,8 | 1174,0 |
| | Ja | Count | 22 | 92 | 114 |
| | | Expected Count | 48,8 | 65,2 | 114,0 |
| Total | | Count | 551 | 737 | 1288 |
| | | Expected Count | 551,0 | 737,0 | 1288,0 |

Count = O (Observert fordeling)
Expected count = E (Estimert fordeling)

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 28,171[b] | 1 | ,000 | | |
| Continuity Correction[a] | 27,129 | 1 | ,000 | | |
| Likelihood Ratio | 30,725 | 1 | ,000 | | |
| Fisher's Exact Test | | | | ,000 | ,000 |
| Linear-by-Linear Association | 28,149 | 1 | ,000 | | |
| N of Valid Cases | 1288 | | | | |

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 48,77.

Antall frihetsgrader

Kjikvadratverdien

p-verdien

Value: Den faktiske kjikvadratverdien = 28,2
DF: Antallet frihetsgrader = 1
Asymp. Sig.: p-verdien som SPSS har regnet ut for tabellen er lavere enn 0,000

**Konklusjon:** Med 95 prosent sikkerhet kan vi slutte at vi vil finne forskjeller i forhold til medlemskap i dans-/teatergruppe i populasjonene av gutter og jenter

## 5.2 t-test for to uavhengige utvalg

Når vi skal undersøke forskjeller mellom grupper og avhengig variabel er kontinuerlig, kan vi benytte oss av t-testen. Hensikten med t-tester er å undersøke om eventuelle forskjeller mellom grupper i utvalget er så store at vi også kan hevde at disse gruppene har ulike gjennomsnitt i populasjonen. Spørsmålet vi da må stille i en t-test er om denne forskjellen skyldes tilfeldige

avvik mellom utvalgene, eller om forskjellen er så stor at vi også må anta at det er forskjeller mellom populasjonene av de ulike gruppene.

t-fordelingen er en symmetrisk statistisk sannsynlighetsfordeling som er litt flatere enn normal-fordelingen i små utvalg og i store utvalg identisk med normalfordelingen.t-testen finnes i en rekke varianter for ett og to utvalg: I toutvalgstester testes forskjellen mellom to gjennomsnitt, og vi kan både se på avhengige (før-etter test av det samme utvalget) og uavhengige utvalg.

Vi kan se på et eksempel av t-test for to uavhengige utvalg.

**Tradisjonell utregning**

Vi antar at det er en sammenheng mellom kjønn og tid som brukes til trening.

1. Vi formulerer følgende hypoteser:
   $H_1$ = Det er en sammenheng mellom kjønn og tid som brukes til trening.
   $H_0$ = Det er ikke en sammenheng mellom kjønn og tid som brukes til trening.
2. Avhengig variabel er på forholdstallsnivå og er operasjonalisert som antall minutter man trener hver dag. Vi kan da benytte oss a t-test for uavhengige utvalg.
3. Vi bestemmer oss for et signifikansnivå på 0.05.
4. Finne utvalgsobservatoren

En standardfeil for differansen i gjennomsnittlig treningstid er:

| | Antall | Gjennomsnitt | Standardavvik |
|---|---|---|---|
| Menn | 64 | 47,2 minutter | 24,1 minutter |
| Kvinner | 71 | 32,7 minutter | 20,0 minutter |
| Totalt | 135 | 39,6 minutter | 23,1 minutter |

$$SF(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$SF(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{24{,}1*24{,}1}{64} \quad \frac{20{,}0*20{,}0}{71}}$$

$$= \sqrt{9{,}08 + 5{,}63} = 3{,}84$$

Når vi har mange enheter (over 120) benytter vi oss +/- 1,96 SF for finne kritisk verdi.
Utregning: +/- 1,96 * 3,84 minutter = 7,5 min

5. Sammenligning: Kritisk verdi er 7,5 minutter, mens gjennomsnittlig forskjell i utvalget av kvinner og menns daglige trening er 14,5 min.

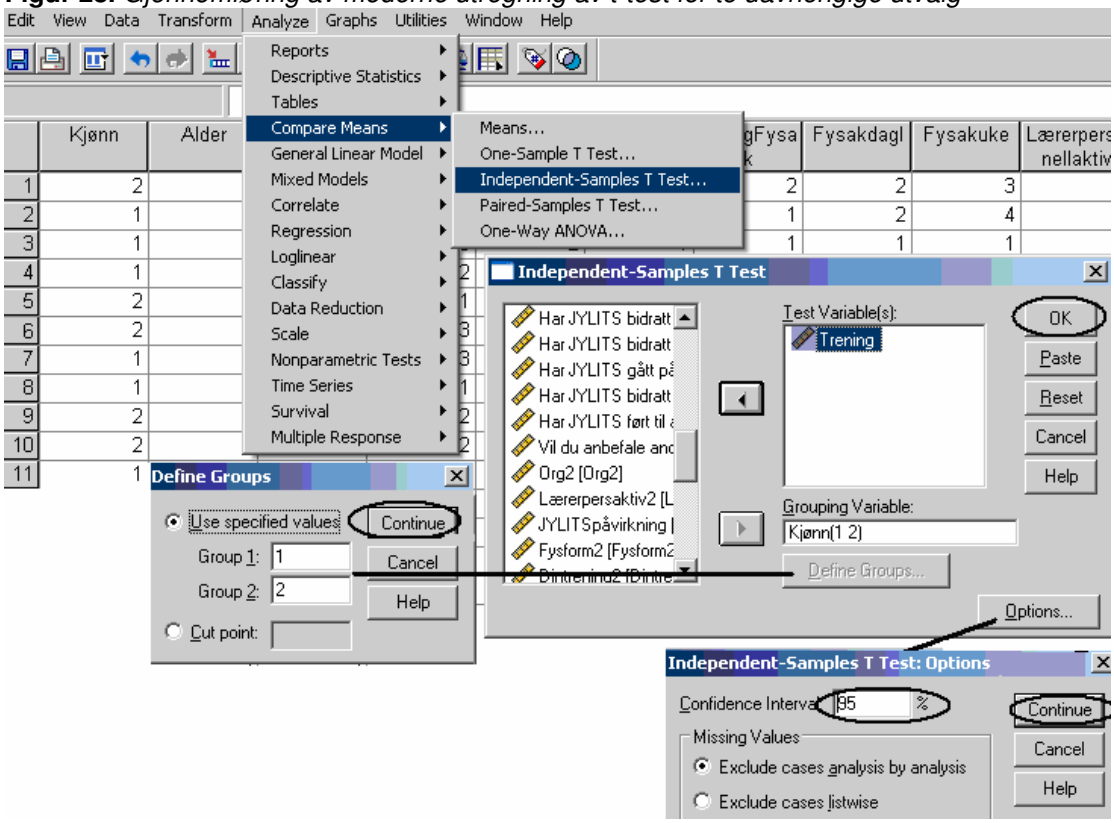
6. Konklusjon: Forskjellen mellom kvinner og menns treningsmengde er større enn kritisk verdi, og vi kan på denne bakgrunn forkaste $H_0$ med 95 prosent sikkerhet.

**Moderne utregning**

Vi gjennomfører den samme beregningen i SPSS.

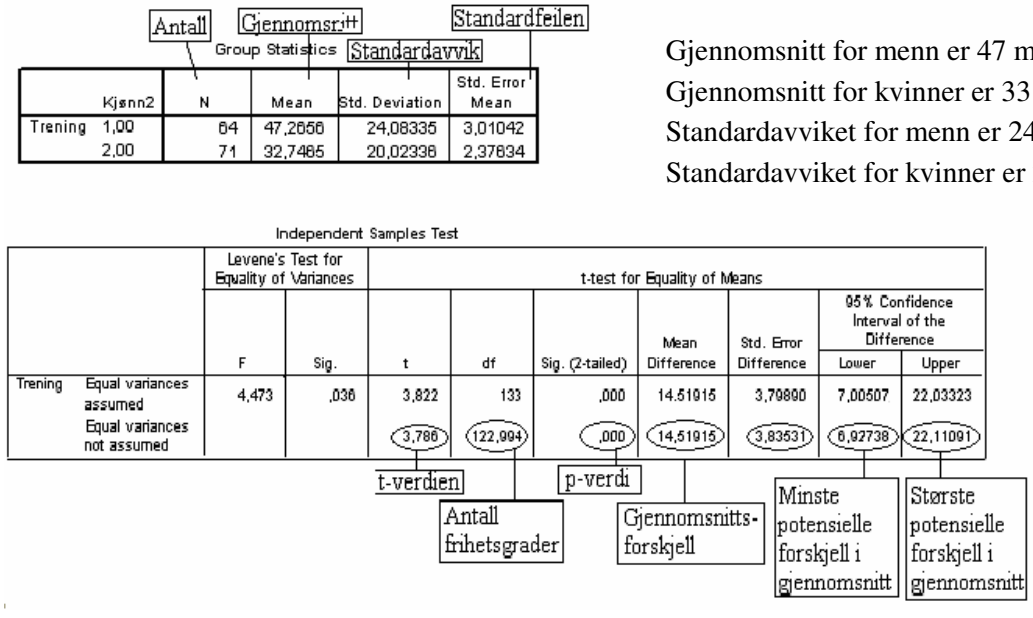
a. *Analyze + Compare Means + Independent Samples t-test*
b. Valg av avhengig variabel: "Trening" og trykk pilen
c. Valg av uavhengig variabel: "Kjønn" og trykk pilen
d. Deretter defineres gruppene: 1 = mann & 2 = kvinne
e. Vi velger konfidensintervall: Trykk *Options og* merk av for 95 % konfidensintervall

**Figur 28.** *Gjennomføring av moderne utregning av t-test for to uavhengige utvalg*




Den andre delen handler om å tolke Output-feltet, d.v.s. resultatene i figuren under

**Figur 29.** *Resultater moderne utregning av t-test*




Gjennomsnitt for menn er 47 min
Gjennomsnitt for kvinner er 33 min
Standardavviket for menn er 24 min
Standardavviket for kvinner er 20 min

Levene's Test viser om det er sannsynlig at den avhengige variabelen har lik varians i de to populasjonene av menn og kvinner. Sannsynligheten er så liten (sign<0,05) at det er tryggest å anta at variansen er ulik. Vi bruker t-testen som ikke forutsetter lik varians.

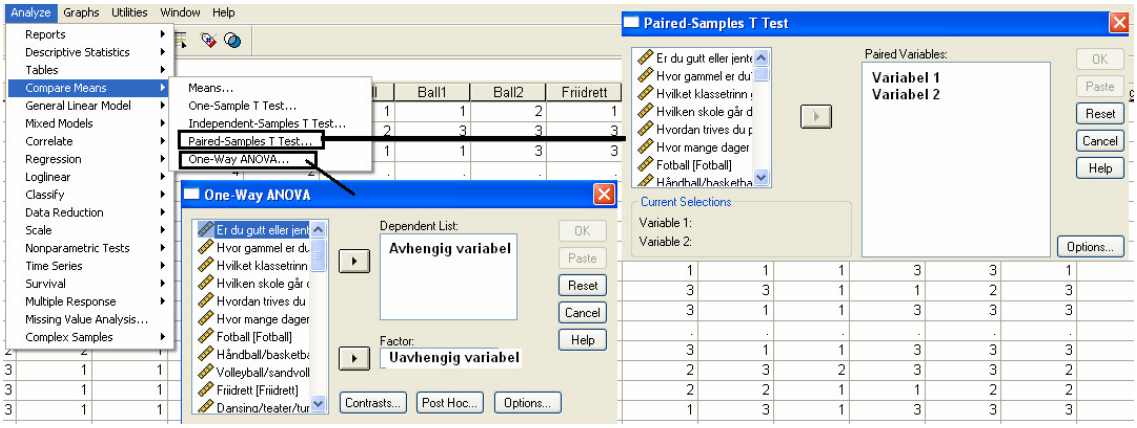
Vi får følgende resultater:

- t: t-verdien er 3,8
- df: Antallet frihetsgrader er 123
- p-verdien er lavere enn 0,000
- Gjennomsnittsforskjellen er 14,5 minutter

**Konklusjon:** Med 95 prosent sikkerhet slutter vi at vi vil finne kjønnsforskjeller i forhold til antall minutter trening i populasjonene av menn og kvinner

**Boks 2. Andre tester av gjennomsnitt**

Det finnes to andre tester av gjennomsnitt som vi skal nevne, men ikke gjennomføre.




a. t-testen for parede utvalg brukes hvis vi ønsker å sammenligne to variabler for det samme utvalget. Vi tester $H_0$ om at gjennomsnittsverdien for disse forskjellene er null
b. En-veis-Anova brukes når vi skal teste forskjeller mellom 3 eller flere ulike grupper. Vi beregner gjennomsnitt for hver enkelt gruppe, og beregner også spredningen innad i gruppen og mellom gruppene

## 5.3 Signifikanstest av korrelasjon

Når begge variabler er på forholdstall, intervall eller ordinalnivå med mange verdier kan vi foreta en signifikanstest ved Pearson korrelasjonsanalyse. Vi tester om korrelasjonen er så sterk at den sannsynliggjør at vi vil finne en korrelasjon mellom de samme variablene i populasjonen.

1. Formulere hypoteser:

$H_1$ = Korrelasjonen mellom kvinneandel i kommunestyret og kvinneandel i formannskapet er signifikant

$H_0$ = Korrelasjonen mellom kvinneandel i kommunestyret og kvinneandel i formannskapet er ikke signifikant

2. Resultater (gjennomføring i SPSS se kap. 3.2)

**Correlations**

| | | Kvinnandel kommstyre | Prosentkvinne rformannskap |
|---|---|---|---|
| Kvinnandelkommstyre | Pearson Correlation | 1 | ,332 |
| | Sig. (2-tailed) | | ,000 |
| | N | 433 | 330 |
| Prosentkvinnerforman nskap | Pearson Correlation | ,332 | 1 |
| | Sig. (2-tailed) | ,000 | |
| | N | 330 | 330 |

Korrelasjonskoeffisienten r

Signifikansnivå

Antall enheter

3. Bestemmer oss for et signifikansnivå på 0.05

**Tradisjonell utregning**

Standardfeilen for korrelasjonen beregnes slik:

r = korrelasjonen

n = antallet enheter

$$SF(r) = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - (0{,}33 * 0{,}33)}{330 - 2}} = 0{,}05$$

Nullhypotesens gyldighetsområde er gitt ved: 0,052 * 1,96 = 0,101.
Korrelasjonen 0,33 er større enn gyldighetsområdet til $H_0$.
Vi konkluderer med at $H_0$ forkastes.

**Moderne beregning**
Sig. (2-tailed) verdien er lavere enn 0,000 (se figuren over).
Vi kan forkaste $H_0$ med 95 prosent sikkerhet.

## 5.4 Vurdering av slutningsstatistikk

Ved hypotesetesting er det nullhypotesen vi tester. Testingen går ut på å beregne sannsynligheten for å få det observerte resultat dersom nullhypotesen er sann; d.v.s. sannsynligheten for å forkaste en riktig nullhypotese. Det betyr at slutningsstatistikk innebærer muligheter for å gjøre feil:

**Figur 30.** *Mulige feilslutninger ved slutningsstatistikk.*

| | $H_0$ er sann | $H_0$ er falsk |
|---|---|---|
| $H_0$ forkastes | Feil (type I) | Korrekt |
| $H_0$ beholdes | Korrekt | Feil (type II) |

Feil av type I vil være å forkaste en sann nullhypotese: Vi kan redusere muligheten for å forkaste $H_0$ ved å bruke et strengere signifikansnivå (eks. 0,01)

Feil av type II vil være å beholde nullhypotesen selv om den er gal. Vi kan øke muligheten for å forkaste $H_0$ ved å bruke et mindre strengt signifikansnivå (eks. 0,1)

Ergo: Reduserer vi faren for å begå type I-feil øker vi faren for å begå feil av type II

**Boks 3. Omkoding av variabel og konstruksjon av sammensatte variabler**

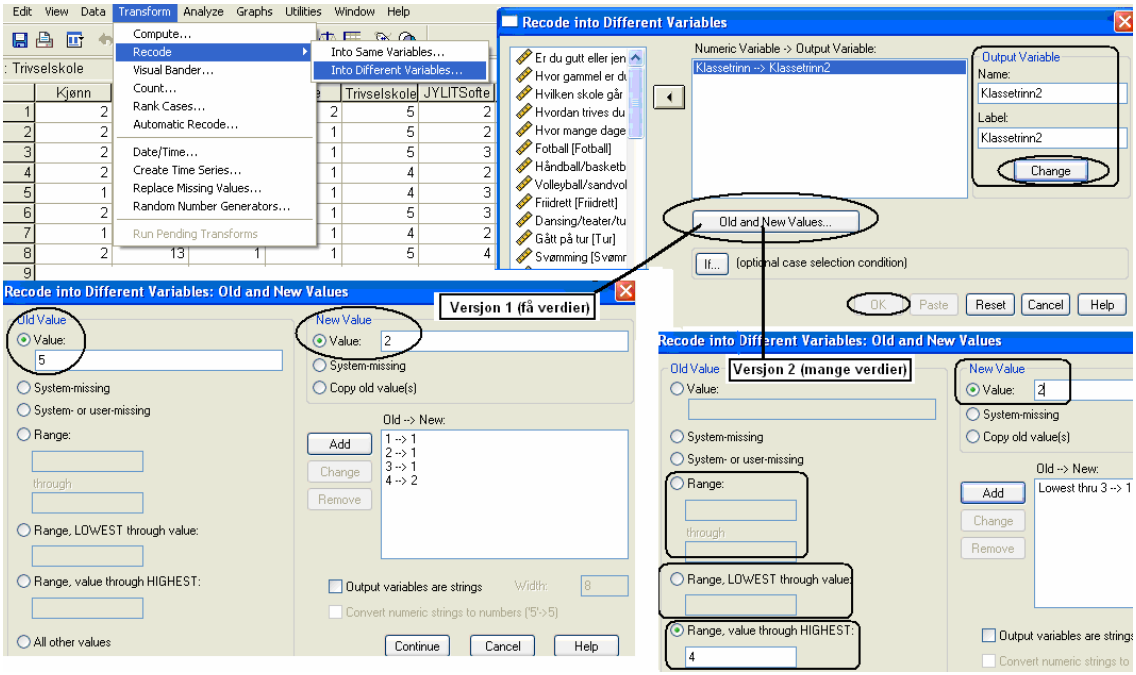
Hvis variabler har mange verdier og vi ønsker å benytte dem i krysstabeller, må vi kode variablene om til færre kategorier.

Vi omkoder variabelen ”Klassetrinn” fra seks til to verdier:

Før: 1 = 8. klasse, 2 = 9. klasse, 3 = 10. klasse, 4 = 1VG, 5 = 2VG og 6 = 3VG

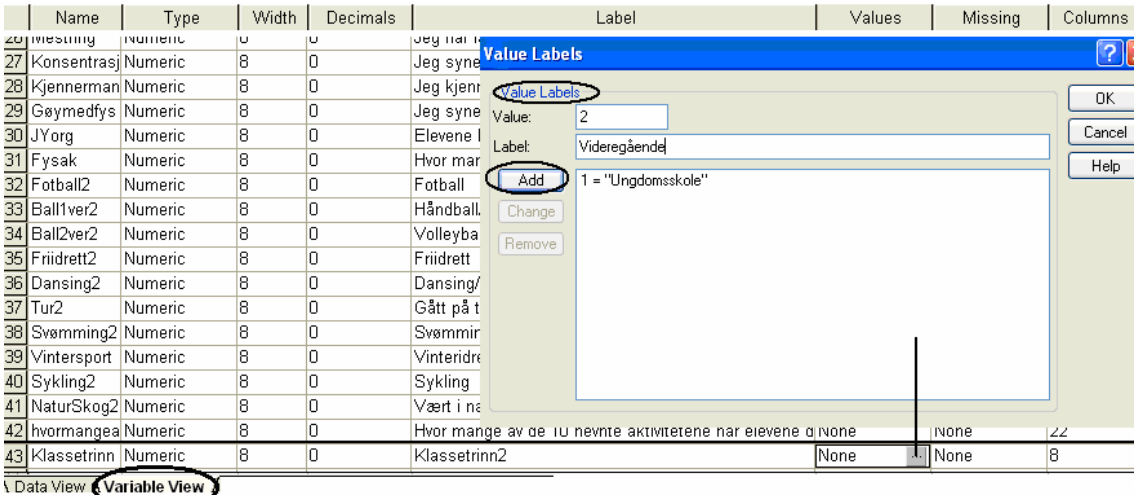
Etter: 1 = ungdomsskole (verdiene 1, 2 og 3) og 2 = videregående (verdiene 4, 5 og 6)

Konstruksjon av den nye variabelen




- Trykk på *Transform* + *Recode* + *Into Different Variables* (lager ny variabel)
- Velger variabelen ”Klassetrinn” og trykker på piltegnet - variabelen havner i oppsamlingsboksen som *Numeric Variable*
    - Vi skriver navn for den nye variabelen: Name = ”Klassetrinn2” og Label = ”Klassetrinn2”
    - Vi trykker *Change* og da navn i oppsamlingsboks som *Output Variable*
- Så trykker vi oss inn på Old and New Values og får opp et nytt vindu
    - Til venstre skal vi skrive de gamle verdiene (*Old Value*) og til høyre skriver vi de nye verdiene (*New Value*)
        - Ved få verdier bruker vi tekstfeltet for de gamle verdiene:
          1 = 1 + Add, 2 = 1 + Add og 3 = 1 + Add
          4 = 2 + Add, 5 = 2 + Add og 6 = 2 + Add
        - Ved mange verdier bruker vi Range-funksjonen:
          Merk av *Range, LOWEST through value* 3 = 1 + Add
          Merk av *Range value through, HIGHEST* 4 = 2 + Add
- Trykk *Continue*
- Kommer tilbake til *Recode into Different Variables* og trykker *OK*

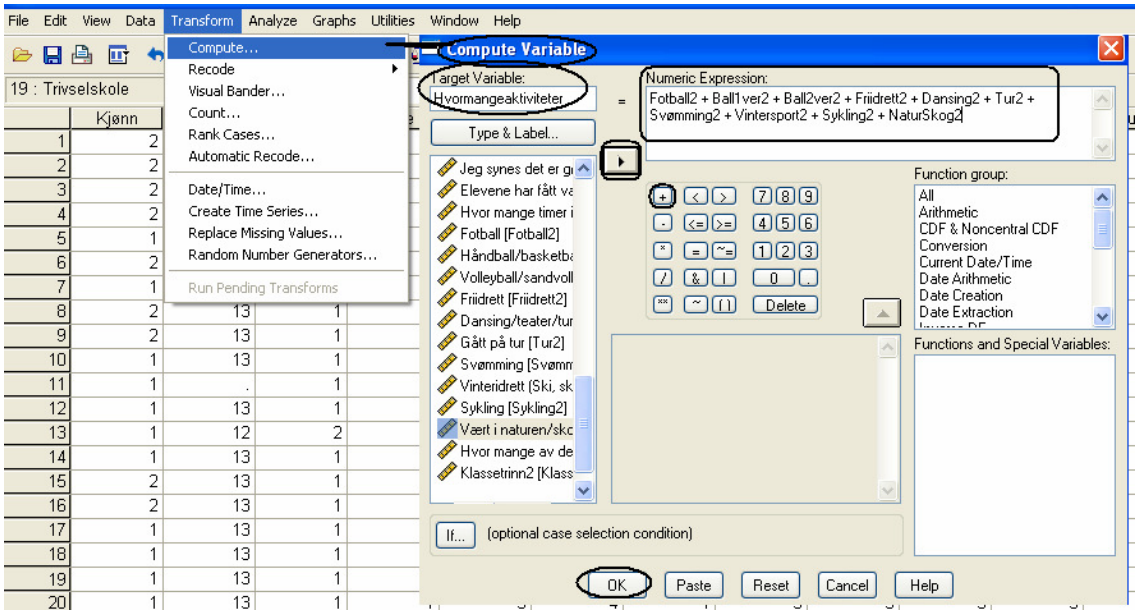
Det neste skrittet er teksting av nye verdier




- Vi går til *Variable View* og finner den nye omkodede variabelen "Klassetrinn2".
- Vi går til *Values*, velger vinduet *Value Labels* og foretar registrering (se kap 2.2)

Vi har vist omkoding. Andre ganger ønsker vi å slå sammen variabler for å gjennomføre analyser. Vi skal lage variabelen "Hvormangeaktiviteter" som vi så på tidligere. Vårt utgangspunkt er 10 aktiviteter der elevene har svart Ja = 1 og Nei = 0.

Konstruksjon av den nye variabelen




Vi velger *Transform* + *Compute*

- Under *Target Variable* skriver vi navn på variabelen "Hvormangeaktiviteter"
- Vi merker av for *Fotball2* trykker på piltegnet og *Fotball 2* kommer i boksen *Numeric Expression.* Vi trykker på tegnet for "+" og tar for oss neste variabel.
- Når vi har lagt inn alle variablene trykker vi på *OK*

Vi går til *Variable View* og finner den nye sammenslåtte variabelen nederst

## Boks 4. Multivariat analyse

Det er sjelden nok at vi bare analyserer sammenhenger mellom to variabler. Vi skal nå redegjøre for hvordan vi gjør analyser der vi kontrollerer for flere uavhengige variabler.

- I forhold til den beskrivende analysen skal vi se på trivariat analyse
- I forhold til slutningsstatistikken skal vi se på multivariat regresjon

### Trivariat analyse

I trivariat analyse kontrolleres effekten for uavhengig variabel med en annen uavhengig variabel. Eks: Avhengig variabel = Dans med to verdier (1 = artig og 2 = Kjedelig)

Kjønn og Dans

| | Gutt | Jente |
|---|---|---|
| Artig | 58 | 84 |
| Kjedelig | 42 | 16 |
| Sum | 100 | 100 |
| N | 106 | 110 |

Mestring og Dans

| | Lav | Høy |
|---|---|---|
| Artig | 58 | 78 |
| Kjedelig | 42 | 22 |
| Sum | 100 | 100 |
| N | 69 | 147 |

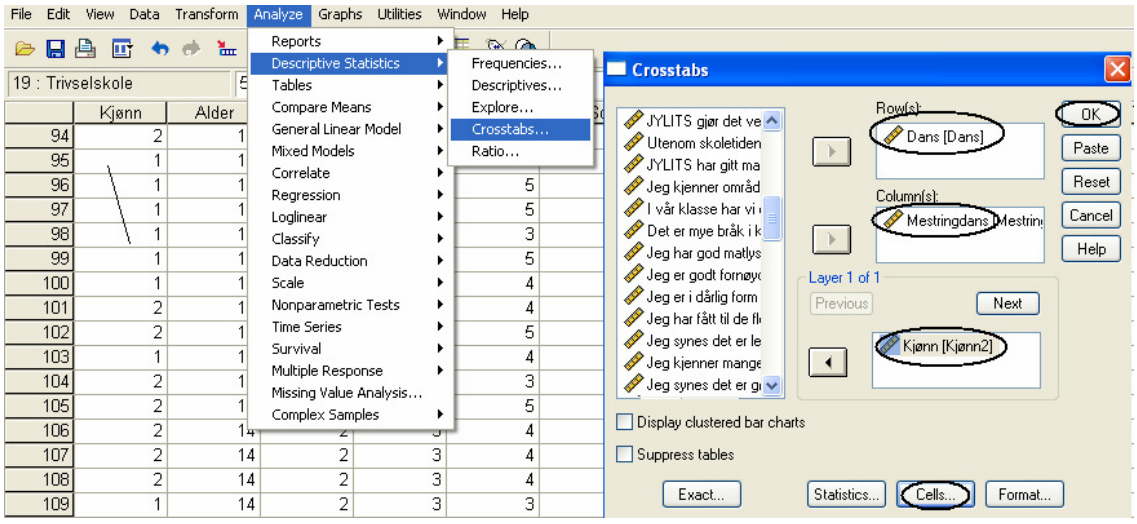
Kjønn og Mestring

| | Gutt | Jente |
|---|---|---|
| Lav mestring | 44 | 20 |
| Høy mestring | 56 | 80 |
| Sum | 100 | 100 |
| N | 106 | 110 |

- Vi har funnet en sammenheng mellom Kjønn og Dans.
- Vi har funnet en sammenheng mellom opplevelse av Mestring og Dans.
- I tillegg har vi funnet en sammenheng mellom Kjønn og Mestring

Kan den positive sammenhengen mellom mestring og holdning til dans egentlig skyldes at kjønn påvirker begge to? Vi gjennomfører analysen i SPSS.

- *Analyze + Descriptive Statistics + Crosstabs*
- Vi legger Dans som avhengig variabel (*Row*), vi legger Mestringsdans som uavhengig variabel 1 (*Column*) og Kjønn er uavhengig variabel 2 (*Layer 1 of 1*)



Vi får ut følgende resultat i Output-vinduet:

Dans * Mestringdans * Kjønn Crosstabulation

| Kjønn | | | Mestringdans: Lav grad mestring | Mestringdans: Høy grad av mestring | Total |
|---|---|---|---|---|---|
| Gutt | Dans | Artig | 46,8% | 67,8% | 58,5% |
| | | Kjedelig | 53,2% | 32,2% | 41,5% |
| | Total | | 100,0% | 100,0% | 100,0% |
| Jente | Dans | Artig | 81,8% | 84,1% | 83,6% |
| | | Kjedelig | 18,2% | 15,9% | 16,4% |
| | Total | | 100,0% | 100,0% | 100,0% |

Vi konsentrerer oss i det videre arbeidet om en av verdiene i den avhengige variabelen - når avhengig variabel er dikotom (2 verdier) spiller det ingen rolle hvilken verdi vi velger for analysen – og velger ut verdien ”Artig”. Vi får følgende tabell:

| Kjønn | Jente | | Gutt | |
|---|---|---|---|---|
| Mestring | Lav | Høy | Lav | Høy |
| Dans er ”Artig” | 82 | 84 | 47 | 68 |

Tabellen omskrives og vi beregner effekter

| | Jente | Gutt | Delsammenhenger |
|---|---|---|---|
| Høy mestring | 84 | 68 | 16 |
| Lav mestring | 82 | 47 | 35 |
| Delsammenhenger | 2 | 19 | |

a. Gjennomsnittlig effekt av ”Kjønn” kontrollert for ”Mestring” = (16 + 35)/2 = 25,5 %.

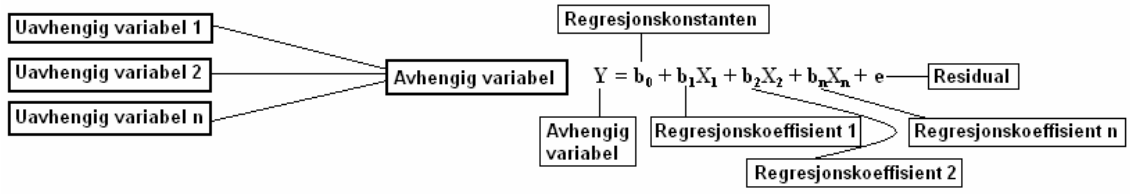
b. Gjennomsnittlig effekt av ”Mestring” kontrollert for ”Kjønn” = (2 + 19)/2 = 10,5 %.

c. Delsammenhengene er ulike

- Det betyr at effekten av ”Kjønn” varierer etter hvilken verdi av ”Mestring” vi holder konstant: Effekten av ”Kjønn” kontrollert for ”Mestring” er større for gutter enn jenter.
- Det betyr at effekten av ”Mestring” varierer etter hvilken verdi ”Kjønn” vi holder konstant: Effekten av ”Mestring” kontrollert for ”Kjønn” er større for ”Lav” enn ”Høy”.

**Regresjonsanalyse**

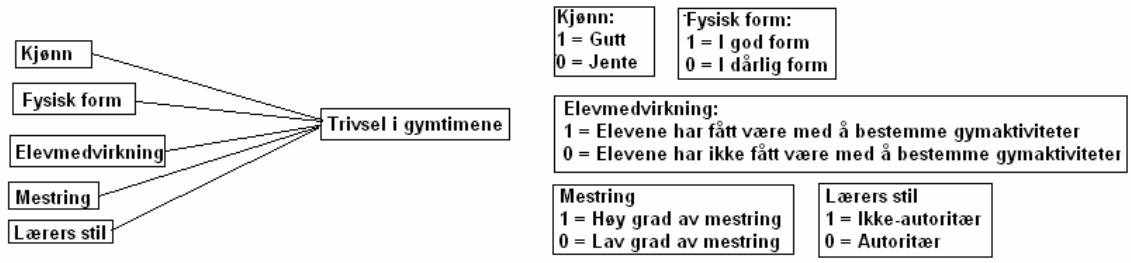
Regresjonsanalysen er basert på tanken om at sammenhengen mellom en avhengig variabel Y og et sett uavhengige variabler (X), kan framstilles i form av en lineær funksjon.



I funksjonen finner vi følgende:

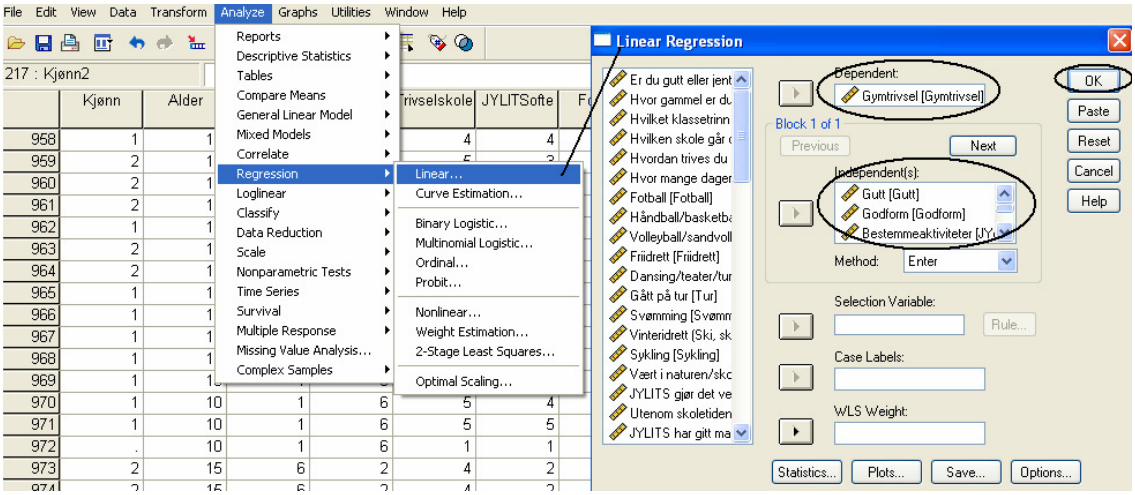
- Variasjonen i avhengig variabel (Y) forklares av uavhengige variabler ($X_1 \ldots X_n$).
- Regresjonskonstanten ($b_0$) er den predikerte verdi av Y når x-variablene er 0.
- Regresjonskoffesienten ($b_1$) viser endringen i Y når X endres med en måleenhet, kontrollert for de andre variablene i modellen.
- Residualene (e) er forskjellen mellom observerte og predikerte verdier av Y.

Vi skal gjennomføre en lineær multivariat regresjonsanalyse i SPSS. Vår avhengige variabel er en sumskårevariabel, sammenslått variabel, av ulike påstander som går på trivsel i gymtimene. Avhengig variabel "gymtrivsel" har laveste verdi 1 (stor grad av mistrivsel) og høyeste verdi 10 (stor grad av trivsel). Vi lager følgende modell:




Vi gjennomfører analysen

- *Analyze + Regression + Linear*
- Avhengig variabel "Gymtrivsel" legges under *Dependent Variable*
- Uavhengige variabler legges i boksen for *Independent*
- Trykk *OK*



Vi får ut følgende resultater som skal tolkes:

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | ,432[a] | ,187 | ,179 | 1,735 |

a. Predictors: (Constant), Ikke-autoritær, Godform, Bestemmeaktiviteter, Gutt, Mestretaktiviteter

**Den justerte kvadrerte korrelasjonskoeffisienten**

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 350,077 | 5 | 70,015 | 23,271 | ,000[a] |
| | Residual | 1522,391 | 506 | 3,009 | | |
| | Total | 1872,469 | 511 | | | |

a. Predictors: (Constant), Ikke-autoritær, Godform, Bestemmeaktiviteter, Gutt, Mestretaktiviteter

b. Dependent Variable: Gymtrivsel

**df viser til frihetsgrader som her er antallet uavhengige variabler**

**Kvadratsummene viser her til hvor mye regresjonen vår forklarer av den totale variansen i avhengig variabel: 350/1872 = 0,187 (som er lik R square i tabellen over)**

**Regresjonskoeffisientene**

**t-verdien**

**signifikansnivået**

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | 4,660 | ,166 | | 28,124 | ,000 |
| | Gutt | ,392 | ,157 | ,103 | 2,495 | ,013 |
| | Godform | ,460 | ,164 | ,118 | 2,811 | ,005 |
| | Bestemmeaktiviteter | ,707 | ,161 | ,177 | 4,398 | ,000 |
| | Mestretaktiviteter | ,656 | ,101 | ,276 | 6,506 | ,000 |
| | Ikke-autoritær | ,378 | ,179 | ,086 | 2,117 | ,035 |

a. Dependent Variable: Gymtrivsel

*Adjusted $R^2$:* Forteller oss at de uavhengige variablene forklarer rundt 18 prosent av variasjonen i avhengig variabel (Y). Det betyr m.a.o. at andre faktorer enn de vi har inkludert i analysen forklarer 82 prosent av variasjonen i Y.

Regresjonskonstanten ($b_0$) er gitt ved verdi 4,7: Det vil si at en elev som er jente, i dårlig form, som ikke får være med å bestemme gymaktiviteter, som har lav mestring av aktivitetene og som har en autoritær lærer har en skåre på 4,7 trivselspoeng i gym

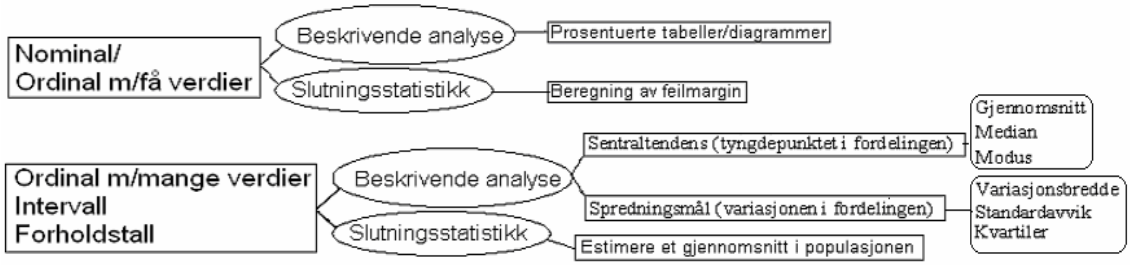
Regresjonskoeffisienter forteller oss endringer når vi går fra verdi 0 til 1. Det betyr:

- at gutter trives bedre i gymtimene enn jenter
- at de som er i god form trives bedre enn de som er i dårlig form
- at de som får være med å bestemme aktiviteter trives bedre enn de som ikke får det
- at de som mestrer aktiviteter trives bedre enn de som ikke mestrer aktivitetene
- at de som har en ikke-autoritær lærer trives bedre enn de som har en autoritær

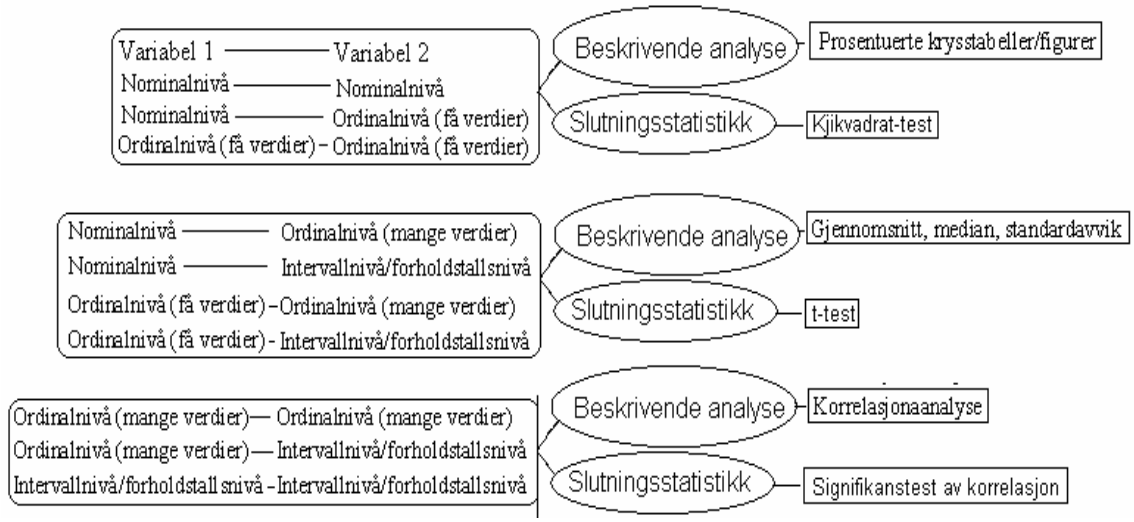
Signifikansverdiene er under signifikansnivået 0.05: Kjønn, fysisk form, medvirkning, mestring og lærerstil har signifikant betydning for trivsel i gymtimene

**Boks 5. Oversikt univariat og bivariat analyse**

Univariat analyse




Bivariat analyse



**Det lille kvantitative metodeheftet**

Dette notatet er basert på mine undervisningsnotater i forbindelse med undervisning i samfunnsvitenskapelig metode ved Høgskolen i Lille-hammer og Norges Teknisk - Naturvitenskapelige Universitet. I notatet gjennomgås noen grunnleggende måter å analysere kvantitative data på ved statistikkprogrammet SPSS.

Notatet er beregnet for studenter, forskere og andre som har liten eller ingen erfaring med bruk av kvantitativ metode. I heftet redegjøres det kortfattet for forskningsprosessen og deretter fokuseres det på kvantitativ metode. Veiledningen er laget til versjon 14.0 av SPSS.

**Notat nr.: 17/2007**
**ISSN nr: 0808-4653**