**PSYC3101: ØVINGSOPPGAVE I REGRESJONSANALYSE**

Å få folk i Norge til utføre mer miljøvennlige handlinger er en sentral miljøpolitisk målsetning. Vi har i den forbindelse foretatt en spørreundersøkelse blant 219 nordmenn over 18 år for å finne ut hvilke faktorer som kan ha innflytelse på ens miljøvaner. Ut i fra forskjellige psykologiske atferdsmodeller har vi en hypotese om at hvorvidt en person utfører miljøvennlige handlinger (heretter kalt *miljøatferd*) avhenger av personens *holdning* til miljøvern og i hvilken grad personen føler seg *utsatt for å få en helseskade* (risikovurdering) som følge av miljøtrusler. I tillegg antar vi at visse sosiodemografiske variabler som kjønn, alder og utdanning har innflytelse på ens tilbøyelighet til å "beskytte miljøet".

På grunnlag av dette kan vi sette opp en regresjonsmodell:

Predikert miljøatferd = a + b1 holdning + b2 risikovurdering+ b3 alder+ b4 kjønn+ b5 utdanning

Datafilen heter «**Datafil - øving** **regresjon.sav**»- Åpne filen og kikk litt på variablene før øvingen starter.

**🡺Variablene som er samlet inn i spørreundersøkelsen er beskrevet på side 7. Kikk på denne før du besvarer disse spørsmålene:**

**Ta stilling til følgende påstander:**

|  |  |  |
| --- | --- | --- |
|  | Riktig | Galt |
| Kvinner er kodet med verdien 0 og menn har verdien 1…………………... |  |  |
| Variabelen KJONN er en dummy-kodet variabel………………………... |  |  |
| En høy skåre på HOLDNING betyr at man er lite miljøbevisst/ har en negativ holdning til miljøvern……………………………………………... |  |  |
| En lav skåre på RISIKO betyr at man oppfatter risikoen for å få en helseskade som følge av miljøtrusler som lav……………………………... |  |  |
| Jo høyere skåre man har på MILJØATFERD, jo mer miljøvennlige handlinger rapporteres……………………………………………………. |  |  |

**1. Studer den avhengige variabelen**

Lag en *histogramfordeling* over den avhengige variabelen miljøvaner, MILJØATFERD. Det gjør du på følgende måte:

Klikk på **Graphs - Legacy Dialogs -** Velg **Histogram...** - Klikk variabelen MILJØATFERD over i boksen det står **Variable**.Kryss også av for **”Display normal curve”**. Klikk **OK**.

Studer histogramfordelingen. Hva forteller denne deg? (stikkord – spredning/varians)

**2. Bivariat regresjon: miljøatferd og holdninger**

Før vi estimerer hele regresjonsmodellen (med alle forklaringsvariablene), vil vi undersøke den *bivariate* sammenhengen mellom miljøholdninger og miljøatferd. Vi estimerer derfor en regresjonsmodell med kun variabelen HOLDNING som *uavhengig* variabel. Bruk MILJØATFERD som *avhengig* variabel. Dette gjøres på følgende måte:

1. Velg **Analyze** på menylinjen.
2. Klikk deretter på **Regression** og velg **Linear...**
3. Klikk variabelen MILJØATFERD over i boksen **Dependent:**
4. Klikk variabelen HOLDNING over i boksen **Independent(s):**
5. Klikk på **OK**-knappen.

**Studer resultatet og ta stilling til disse påstandene:**

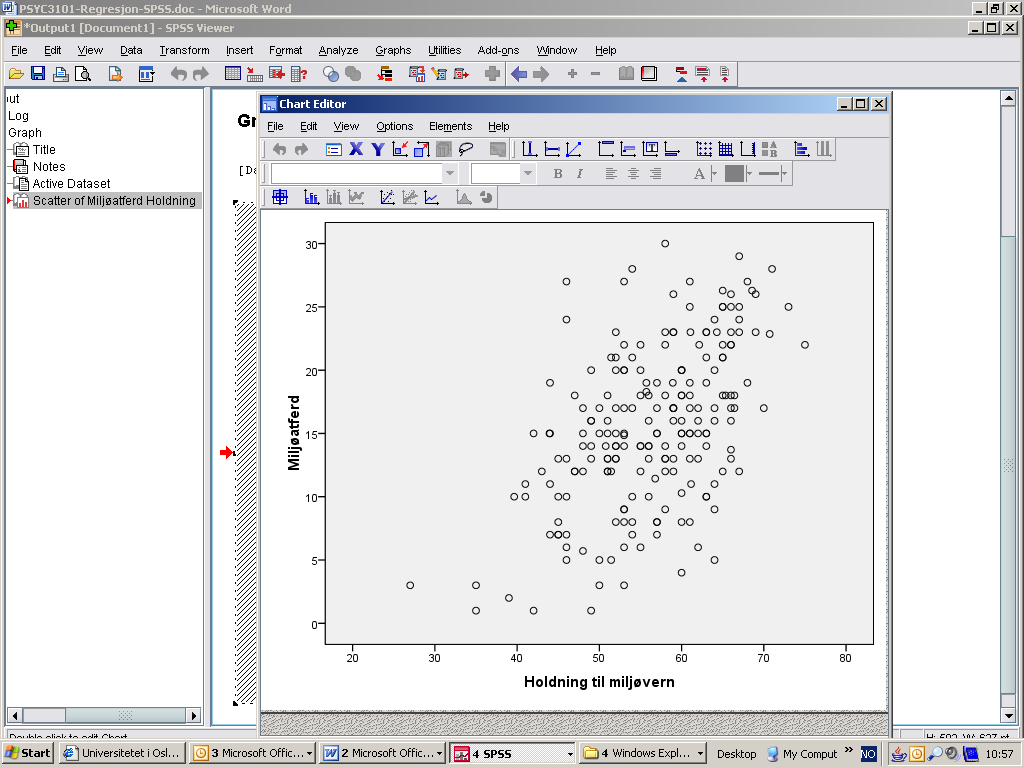
|  |  |  |
| --- | --- | --- |
|  | Riktig | Galt |
| 52,1 % av den totale variasjonen i miljøatferd kan forklares ut i fra skåre på miljøholdninger |  |  |
| 27,1 % av den totale variasjonen i miljøatferd kan forklares ut i fra skåre på miljøholdninger |  |  |
| Regresjonskoeffisienten (b) til HOLDNING er lik 0.412 |  |  |
| Regresjonskoeffisienten (b) er signifikant forskjellig fra 0 |  |  |
| Når skåre på miljøholdninger øker, synker skåre på miljøvaner |  |  |
| Når skåre på miljøholdninger øker med ett standardavvik, øker skåre på miljøatferd i gjennomsnitt med 0.521 standardavvik |  |  |
| Resultatet tyder på at ”positive” holdninger til miljøet har sammenheng med at man utfører mer miljøvennlige handlinger |  |  |

**3. Spredningsdiagram med regresjonslinje.**

Samvariasjonen mellom miljøholdninger og miljøatferd kan også presenteres som et spredningsdiagram (scatterplot) med regresjonslinje. Akkurat denne funksjonen er litt lite brukervennlig i SPSS, så det kan hende at dere vil slite litt med dette. Uansett, benytt følgende måte:

1. Velg **Graphs** – **Legacy Dialogs** på menylinjen. Klikk deretter på **Scatter/Dot**...
2. Du får nå opp en boks med flere valgmuligheter. Klikk på det bildet der det står **Simple** **Scatter** og klikk deretter på **Define**.
3. Klikk variabelen HOLDNING over til der det står **X-axis:**, og MILJØATFERD over til der det står **Y-axis:,** deretter **OK**.
4. Du får nå et scatterplot mellom de to variablene, men uten regresjonslinje. For at regresjonslinjen skal vises, må du først dobbeltklikke på scatterplottet i Output-vinduet.
5. Når du dobbeltklikker på grafen, kommer den opp i et nytt vindu. For å få frem regresjonslinjen, klikker du på denne knappen:

(Knappen heter: *Add Fit Line at Total*).



1. Lukk vinduet Chart Editor. 🗷

Hva kalles den vertikale avstanden mellom en observert verdi (punktene i spredningsdiagrammet) og regresjonslinjen?

**4. Multippel regresjonsanalyse**

Du ønsker nå å estimere hele regresjonsmodellen, dvs. ha med alle de uavhengige variablene i regresjonsmodellen (miljøholdning, risikovurdering, alder, utdanning og kjønn). Bruk samme fremgangsmåte som i oppgave 2, men klikk variablene RISIKO, ALDER, UTDANN og KJONN over i **Independent(s)** boksen i tillegg til HOLDNING. Klikk på **OK**-knappen og kjør i vei.

1. Tolk resultatet av utskriften: Hvilke av de uavhengige variablene har en signifikant effekt (på 5 % -nivå) på miljøvaner?

|  |  |  |
| --- | --- | --- |
|  | Signifikant effekt på miljøvaner? | |
|  | Ja | Nei |
| HOLDNING (Holdning til miljø) |  |  |
| ALDER |  |  |
| KJONN |  |  |
| UTDANN (utdanning) |  |  |
| RISIKO (Opplevd risiko for helseskade) |  |  |

1. Tolk regresjonskoeffisienten til de uavhengige variablene:

|  |  |  |
| --- | --- | --- |
|  | Riktig | Galt |
| Menn utfører i gjennomsnitt mer miljøvennlige handlinger enn kvinner |  |  |
| En høy skåre på HOLDNING (holdning til miljøvern) har sammenheng med høy grad av miljøvennlig atferd |  |  |
| De som opplever høy grad av risiko for helseskade er tilbøyelige til å utføre mer miljøvennlig atferd enn de med lav grad av opplevd risiko |  |  |

1. Hvilken av de uavhengige variablene vil du si har sterkest effekt på miljøatferd (les: hvilken variabel ser ut til å forklare mest av variasjonen i miljøvaner)?
2. Hva er skjedd med R2 etter at du inkluderte de andre uavhengige variablene i modellen? Hvor mye av den totale variasjonen i miljøatferd forklares av de uavhengige variablene? Hvorfor tror du at det er skjedd en endring?
3. Hva er Justert R2 (Adjusted R2) i modellen – og hva er forskjellen mellom Justert R2 og R2?
4. Sammenlign regresjonskoeffisienten til miljøholdninger (HOLDNING) med den regresjonskoeffisienten du fikk i den bivariate analysen. Hva er skjedd og hvorfor tror du dette er skjedd?

**5. Undersøk om det kan være multikollinearitetsproblemer**

Dette kan undersøkes på flere måter, men den enkleste er rett og slett å se på hvor høyt korrelerte de uavhengige variablene er med hverandre. Dette gjør du gjennom å:

1. Velg **Analyze** på menylinjen.
2. Velg **Correlate** og deretter **Bivariate**.
3. Klikk de uavhengige variablene over i boksen for **Variables** og klikk **OK**.

|  |  |  |
| --- | --- | --- |
|  | Ja | Nei |
| Har du grunn til å tro at vi har et kolinearitetsproblem? |  |  |

En litt mer avansert måte, er å estimere den såkalte **Toleransen** til hver uavhengig variabel. Denne bør være høyere enn 0.25 (enkelte setter grensen høyere, ved .40). Toleransen finner du gjennom å bestille regresjonsanalysen på nytt, men du må i tillegg be om at Toleransen for hver variabel estimeres. Benytt samme fremgangsmåte som i oppgave 2, men før du klikker på OK-knappen, klikker du på **”Statistics”** og krysser av for ”**Collinearity Diagnostics”**.

|  |  |  |
| --- | --- | --- |
|  | Ja | Nei |
| Har noen av de uavhengige variablene en Toleranse som er mindre enn .25? |  |  |

**6. Undersøk relasjonen mellom residualene og predikerte verdier.**

Estimer modellen på nytt, men klikk av for noen flere valgmuligheter i tillegg:

1. Velg **Analyze** på menylinjen.
2. Klikk deretter på **Regression** og velg **Linear…**
3. Klikk variabelen MILJØATFERD over i boksen **Dependent:**
4. Klikk de uavhengige variablene over i boksen **Independen(s):**
5. Klikk på **Save** og kryss deretter av for **Unstandarized** under *Predicted Values* og *Residuals*.
6. Klikk på **Continue**.
7. Klikk på **OK**-knappen.
8. Gå deretter tilbake til ***datafilen***. Du har nå fått to nye variabler i datafila, pred\_1 som er predikert verdi for miljøatferd og variabelen res\_1 er residualene. Hva utrykker disse variablene?

|  |  |  |
| --- | --- | --- |
|  | Riktig | Galt |
| Pred\_1 (predikert verdi) viser hvilke skåre på miljøatferd vi tror personen har ut i fra skåre på de uavhengige variablene i regresjonsmodellen |  |  |
| Res\_1 (residualen) uttrykker forskjell mellom faktisk skåre og predikert skåre på miljøvaner |  |  |

1. En forutsetning for OLS-estimering er at residualleddet er noenlunde normalfordelt og har et gjennomsnitt på null. Dette kan du finne ut gjennom å se på et *histogram* over residualen. Gjør dette på samme måte som under oppgave 1 (side 1), men klikk ***res\_1*** over i boksen **Variable**. Kryss også av for **”Display normal curve”**.

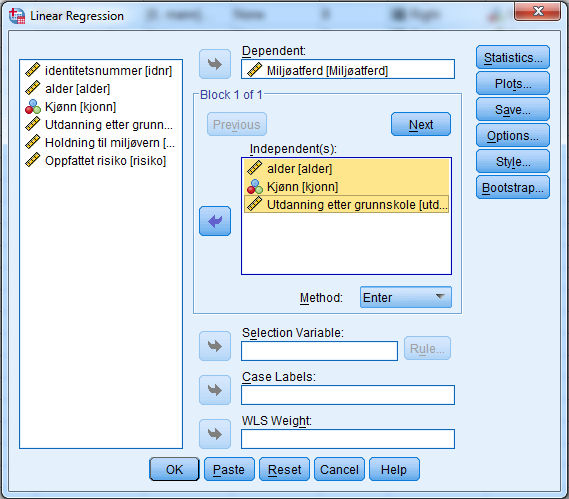
Synes du forutsetningen om et normalfordelt residualledd (restledd) er oppfylt?

## **7. Blokkvis regresjon (hierarkisk regresjon)**

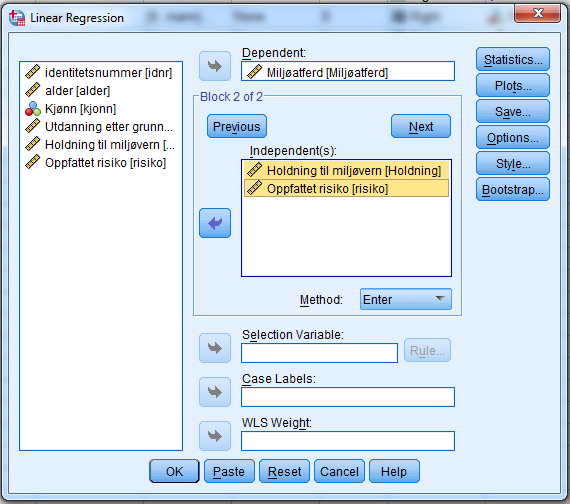
I denne analysen slipper vi inn de uavhengige variabler i grupper, eller ”blokker” som det også kalles.

Vi vil fortsatt undersøke om alder, kjønn, utdanning, holdninger og risikovurdering har sammenheng med miljøatferd, men nå vil vi først kun bruke de sosiodemografiske variablene som uavhengige variabler (Kjønn, alder, utdanning). Dette blir da *blokk 1*. Så kan vi i neste omgang i tillegg inkludere de uavhengige variabler av mer ”psykologisk art” i modellen (holdninger og risikovurdering) – dette blir *blokk 2*. Da må vi undersøke om det å inkludere Blokk 2 i tillegg til Blokk 1 forklarer mer varians i Miljøatferd det kun Blokk 1 gjør.

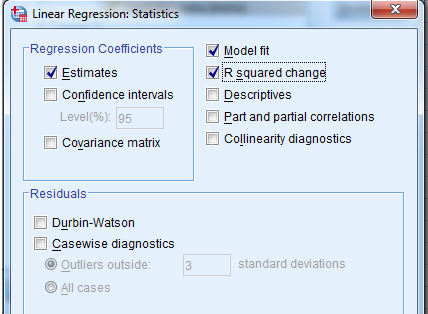
1. Klikk over *Alder*, *Kjønn* og *Utdanning* til Block 1 of 1:



2. Klikk på ”**Next**”. 3. Deretter klikker du *Holdning* og *Risiko* over i Block 2 of 2. Klikk deretter på ”**Statistics**”.



3. Velg ”**R squared change**”.



Kikk på resultatutskriften av den blokkvise analysen.

1. Øker R2 når blokk 2 inkluderes, og hvor stor er i så fall økningen?
2. Er økningen i R2 signifikant forskjellig fra null?
3. Hva betyr dette resultatet?

**OVERSIKT OVER DE ULIKE VARIABLENE I DATAFILA:**

|  |  |
| --- | --- |
| ALDER | Alder |
| KJONN | Kjønn  1 = kvinne  0 = mann |
| UTDANN | Antall år med utdanning etter grunnskole |
| RISIKO | Sumskåre av respondentenes sannsynlighetsvurdering for helseskade som følge av **miljørisiko**. Omfatter hull i ozonlaget, drivhuseffekten, kjemiske tilsetningsstoffer i mat, smittefarlig mat, industriforurensning, trafikkforurensning, radioaktiv stråling fra kjernekraftverk. Laveste skåre = 7, høyeste skåre = 35.  *Høy skåre = Oppfatter sannsynligheten for helseskade som høy* |
| HOLDNING | Sumskåre environmental concern scale (holdning til miljøvern). Består av 15 holdningsspørsmål. Laveste skåre = 15, høyeste = 75.  *Høy skåre = Positiv holdning til miljøvern/ høy grad av miljøbevissthet* |
| MILJØATFERD | Sumskåre av respondentens miljøvaner. Inkluderer unngått å kjøpe produkter fra firma som ikke tar hensyn til miljøet, kjøpt produkter som kan resirkuleres, skiftet fra ett merke til et annet, levert papir/glass til resirkulasjon, redusert bilbruk, brukt biologisk nedbrytbare såper, lest varedeklarasjon på produktet, kjøpt økomatvarer. Laveste skåre = 0, høyeste skåre = 32.  *Høy skåre = Høy grad av miljøatferd* |