**PSYC3101: SPSS-ØVING i mediator- og moderatoranalyse med bruk av regresjon**

***Innledning***

Før dere setter i gang med øvingen, må dere gjøre dere kjent med datafilen **alkohol.sav**, som finnes i øvingsmappen ”Øving – mediering og moderering”. Denne inneholder kun tre variabler hentet fra en undersøkelse gjennomført blant norske 18-åringer (N = 581). Tema for undersøkelsen er bruk av alkohol og forventninger til alkoholbruk. Datafilen inneholder følgende variabler:

**1**. **Kjønn**:

Kodingen er slik:

Jente = 0, Gutt = 1

**2. Alkoholkonsum (ALCONS\_18)**

Dette er en sumskåre av fire spørsmål hvor ofte man for tiden drikker fire ulike typer alkohol; lettvin, vin, øl og brennevin. Alle fire er målt på en skala fra 0 (aldri), 1 (mindre enn 1 gang i måneden), 2 (1-3 ganger i måneden), 3 (ukentlig).

Denne sumskåren har variabelnavnet **Alcons\_18**, der lavest mulig skåre = 0 og høyest mulig oppnåelig skåre er 12. Høy skåre = stort forbruk av alkohol.

**3. Alkoholforventninger (SUMPOS\_18)**

Dette er en sumskåre av fire spørsmål om alkoholforventinger (vist nedenunder)**.** Jo høyere skåre, jo mer positive forventinger. Kun sumskårer er med i datafilen.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Uenig* | *Delvis uenig* | *Delvis enig* | *Enig* |
| 1. Folk blir mer vennlige når de har drukket litt alkohol | 0 | 1 | 2 | 3 |
| 2. Alkohol gjør det lettere å ha de moro sammen med andre | 0 | 1 | 2 | 3 |
| 3. Det blir mer gøy på fest når det blir servert alkohol der | 0 | 1 | 2 | 3 |
| 4. Folk blir i bedre humør når de drikker litt alkohol | 0 | 1 | 2 | 3 |

***OPPGAVER DERE SKAL LØSE***

**Oppgave 1: Studer histogramfordelinger over alkoholbruk ved 18 års alder (ALCONS\_18), samt positive alkoholforventinger ved samme alder (SUMPOS\_18).**

For å få frem histogram:

Graphs

Legacy Dialogs

Histogram

Hvordan vil du beskrive histogrammene? – gi en kort kommentar til hvordan alkoholbruk og alkoholforventinger er blant ungdommene.

**Oppgave 2: Testing av mediatoreffekt: Går effekten av kjønn på alkoholbruk gjennom positive** **alkoholforventinger?**

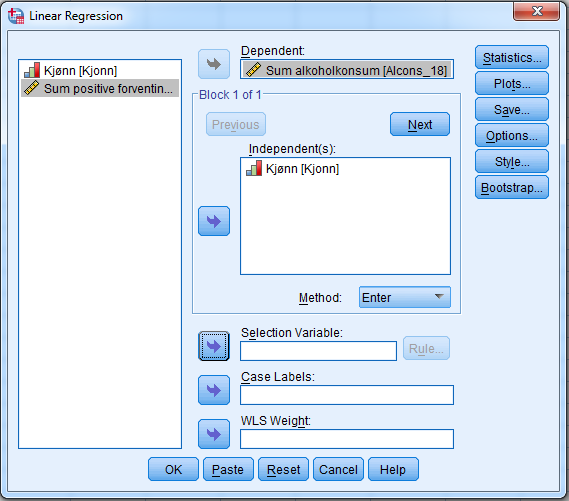
I denne deloppgaven skal dere undersøke om det er noen *kjønnsforskjeller* i alkoholbruk og om en eventuell effekt av kjønn på alkoholbruk er *mediert* gjennom alkoholforventinger. Dere kan starte med å illustrere den antatte mediatoreffekten av alkoholforventinger et *stidiagram* (se forelesningsnotatene hvis dere er usikre)

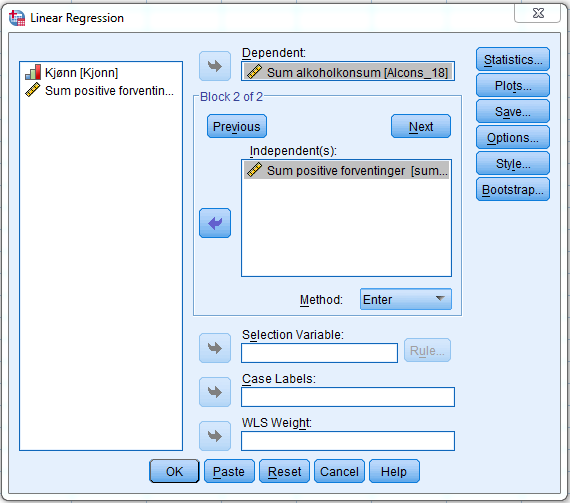
Deretter kan dere benytte regresjonsanalyse til å teste de fire trinnene/stegene i en mediatoranalyse som foreslått av Baron & Kenny (se forelesningsnotatene om dere er usikre):

* Steg 1: undersøke om kjønn (X) er relatert til alkoholbruk (Y). Hvis ja, forklar i hvilken retning kjønnsforskjellen går (hvem drikker mest av gutter og jenter og hvor stor er forskjellen?)
* Steg 2: undersøk om det er kjønnsforskjeller i alkoholforventinger (M). Hvis ja, beskriv effekten.
* Steg 3 og 4. Bruk hierarkisk regresjonsanalyse (se beskrivelse på neste side). Vurder på dette grunnlag om det er en mediatoreffekt til stede, og om vi i så fall har full eller delvis mediering.

*Bruk av hierarkisk regresjonsanalyse for å undersøke om det er en mediatoreffekt av alkoholforventinger.*

Velg vanlig lineær regresjonsanalyse (Analyze -> Regression -> Linear). Avhengig variabel er alkoholbruk. I blokk 1 tas kun kjønn med som uavhengig variabel.. Klikk deretter på ”Next”, for å komme til blokk 2, her tas alkoholforventinger også med som uavhengig variabel.



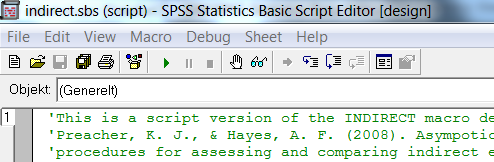


Tolk utskriften av analysen. Er det grunn til å tro at det er en mediatoreffekt av positive alkoholforventninger tilstede? Forklar.

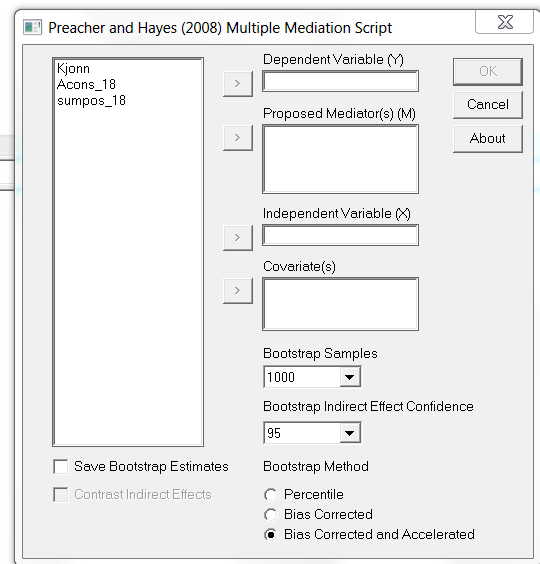
Lag til slutt et *stidiagram*, der du skriver inn tall fra regresjonsanalysene som beskriver relasjonen mellom de tre variablene. Se forelesningsnotater for eksempel på oppsett av en slik figur.

**Oppgave 3**

1. Lag en tabell der du beregner både den direkte, indirekte og totale effekten av kjønn på alkoholbruk. Bruk fremgangsmåte som beskrevet i forelesningsnotater.
2. Signifikanstesting av den indirekte effekten av kjønn på alkoholkonsum, mediert av alkoholforventninger med bruk av Preacher & Hayes fremgangsmåte:
   1. Dobbeltklikk/åpne den såkalte Makroen ”indirect.sbs” som ligger på Fronterrommet til PSYC3101 under mappen ”Øving – mediering og moderering” og bruk denne til å estimere den indirekte effekten.
   2. Trykk på symbolet som ser ut som en ”Play” knapp.



* 1. Klikk så over variablene i de boksene der du mener de hører hjemme og klikk OK.



* 1. Studer utskriften og svar på spørsmålene på neste side:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Koeffisient | Signifikant forskjellig fra null på 5%-nivå? | Hvilken ”sti” tilsvarer dette i Baron og Kenny’s modell (a, b, c, c’, a\*b) |
| Effekt av kjønn på mediator |  |  |  |
| Effekt av mediator på alkoholkonsum |  |  |  |
| Total effekt av kjønn på alkoholkonsum |  |  |  |
| Direkte effekt av kjønn på alkoholkonsum (kontrollert for effekt av mediator) |  |  |  |
| Indirekte effekt av kjønn på alkoholkonsum |  |  |  |

**Oppgave 4. Moderatoreffekt (interaksjonseffekt)**

Vi har funnet en sammenheng mellom positive alkoholforventinger og alkoholbruk, og at alkoholforventinger har en mediatoreffekt mht relasjonene mellom kjønn og alkoholbruk. Imidlertid kan vi forvente at kjønn spiller en annen rolle i dette tilfellet; relasjonen mellom positive alkoholforventinger og alkoholbruk kan være forskjellige for jenter og gutter. Sagt mer teknisk: effekten av positive alkoholforventinger (M) på alkoholbruk (Y) kan være forskjellig for ulike nivå på variabel X (Kjønn). Dette kalles en interaksjonseffekt, eller en moderatoreffekt.

Vi kan teste denne effekten gjennom å undersøke sammenhengen mellom positive alkoholforventninger (M) og alkoholbruk (X) separat for gutter og jenter (X), og så undersøke om regresjonskoeffisientene for gutter og jenter er ulike, men en bedre måte er nok å inkludere et såkalt interaksjonsledd i regresjonsanalysen.

Metode 1: Uten sentrering

Modell uten interaksjonseffekt ser slik ut:

= a + b1X + b2M

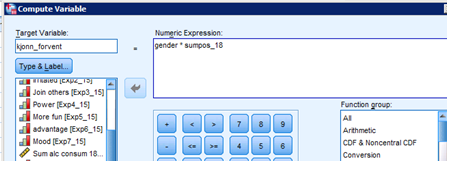
Modell med interaksjonsledd vil se slik ut:

= a + b1X + b2M + b3X\*M

Koeffisienten b3 til interaksjonsleddet (X\*M) vil her si hvorvidt effekten av alkoholforventinger på alkoholbruk er sterkere eller svakere for gutter enn for jenter. Hvis b3 blir forskjellig fra null og/eller R2 øker når vi tar med interaksjonsleddet i modellen, kan vi si at vi har en interaksjonseffekt. Blir b3 positiv, er effekten sterkere for gutter, blir den negativ er det en sterkere for jenter. Hvorfor tror du det blir slik? (tips: finn ut hvem er kodet 0 og 1 på kjønn).

**Slik gjør du det i praksis:**

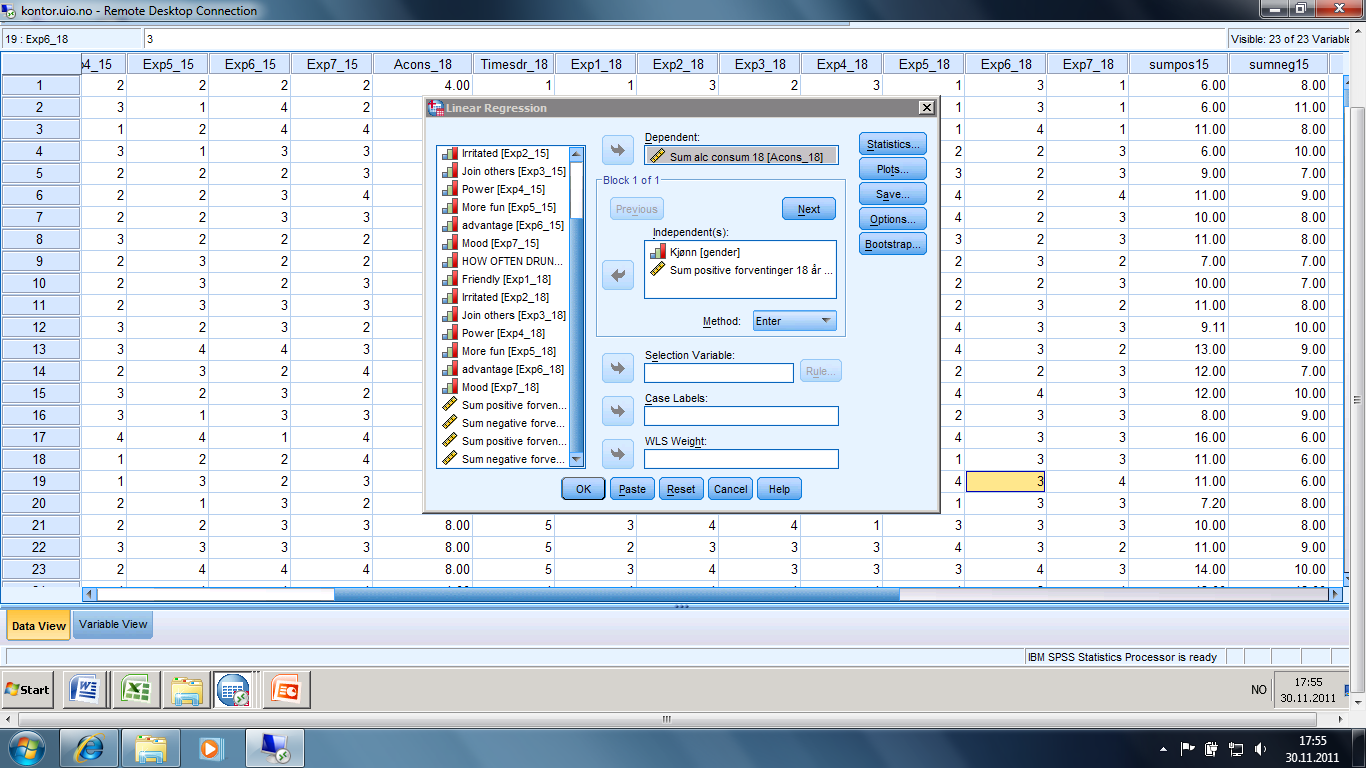
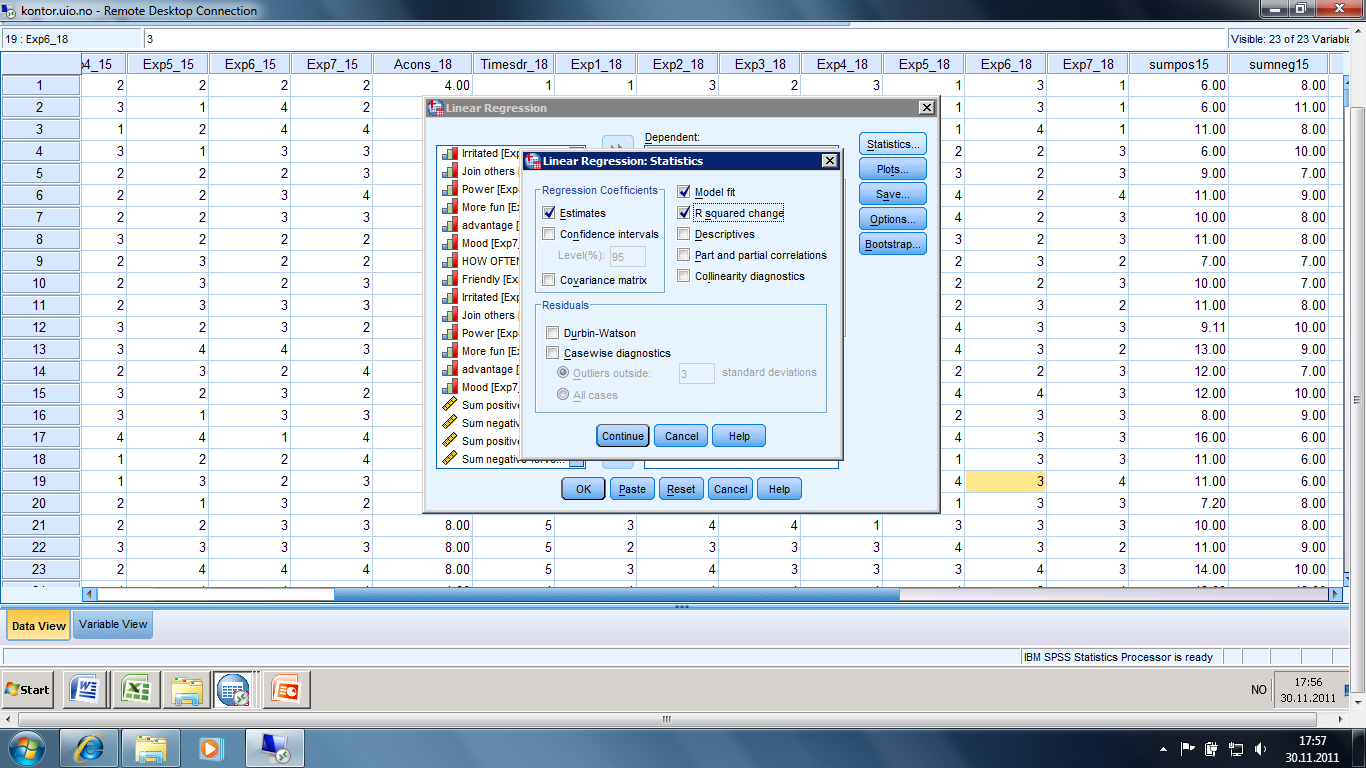
Bruk Transform - COMPUTE-kommandoen og lag en ny variabel der du multipliserer kjønn med positive alkoholforventinger:



COMPUTE KJONN\_FORVENT = GENDER\*SUMPOS\_18.

Benytt deretter hierarkisk regresjonsanalyse. I blokk 1 blir kjønn *og* positive alkoholforventinger uavhengige variabler, i blokk 2 tar du med interaksjonsleddet mellom de to som uavhengig variabel.

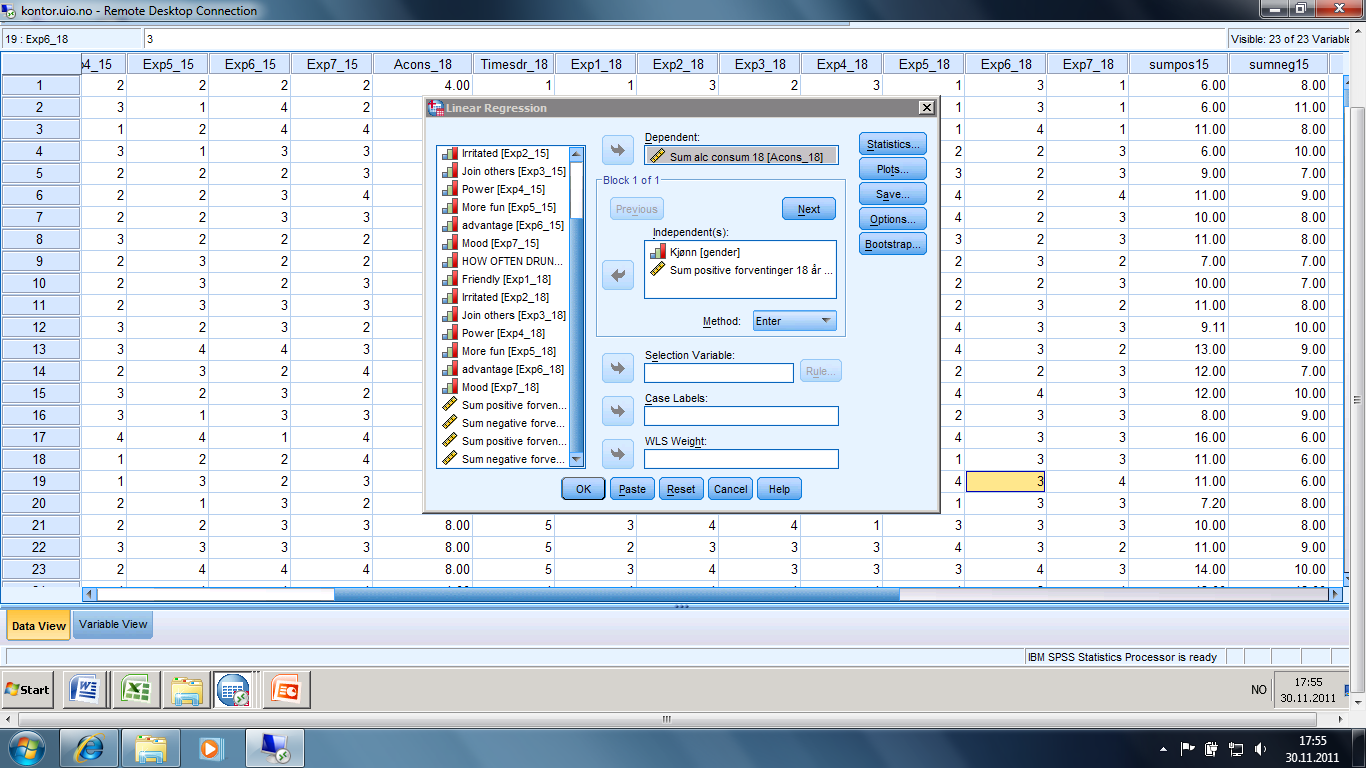
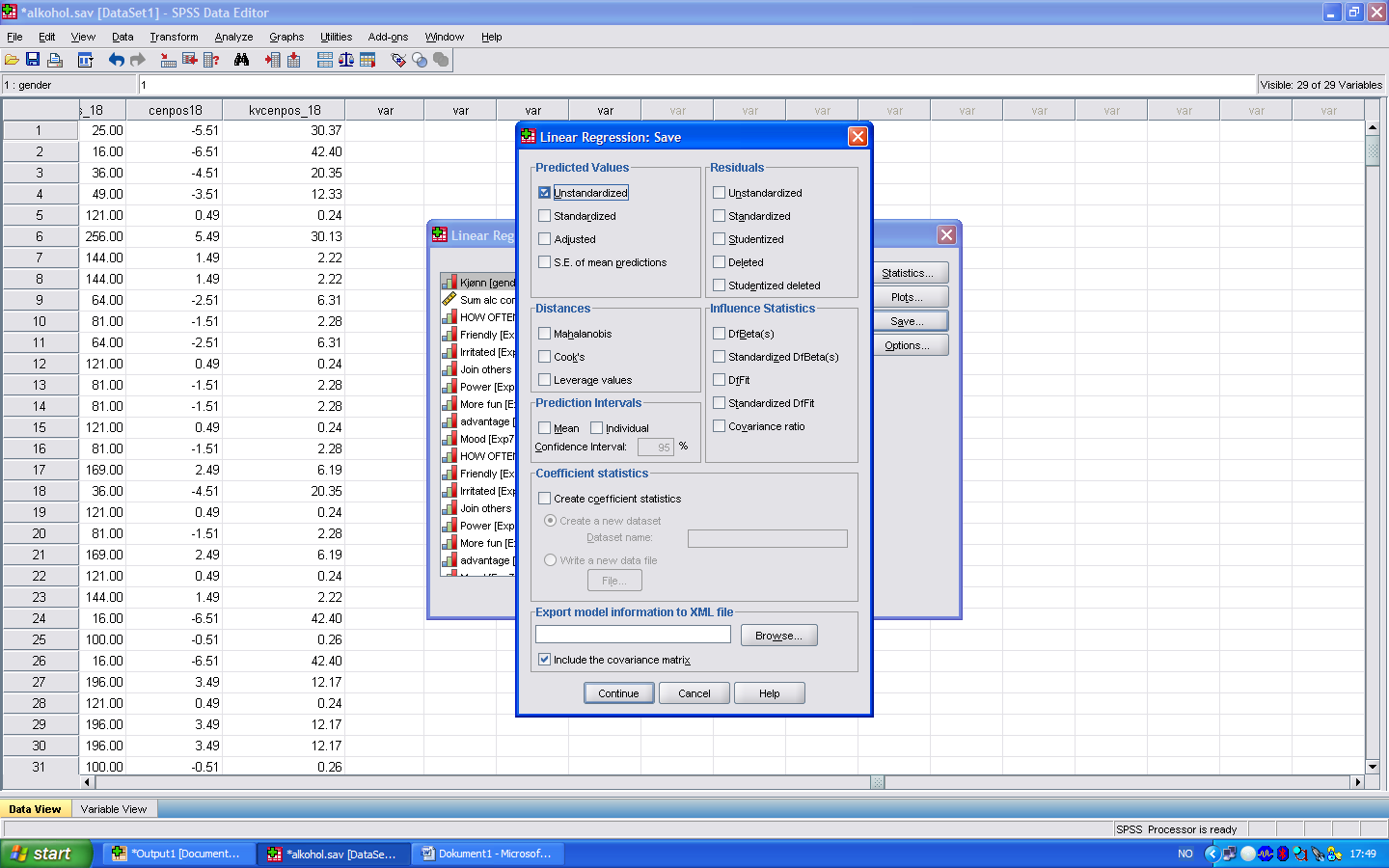
I tillegg må du bestille en test for om andel forklart varians (R2) øker signifikant fra modell 1 til modell 2. Dette gjør du gjennom å klikke på knappen ”Statistics” og velg deretter ”R squared change”.

Klikk Continue og deretter OK.

Tolk resultatene du får ut – ser det ut til å være en signifikant interaksjonseffekt? Tips: se om R2 øker signifikant. Prøv også å finne ut i hvilken retning forskjellen går.

**Grafisk framstilling av interaksjonseffekt:**

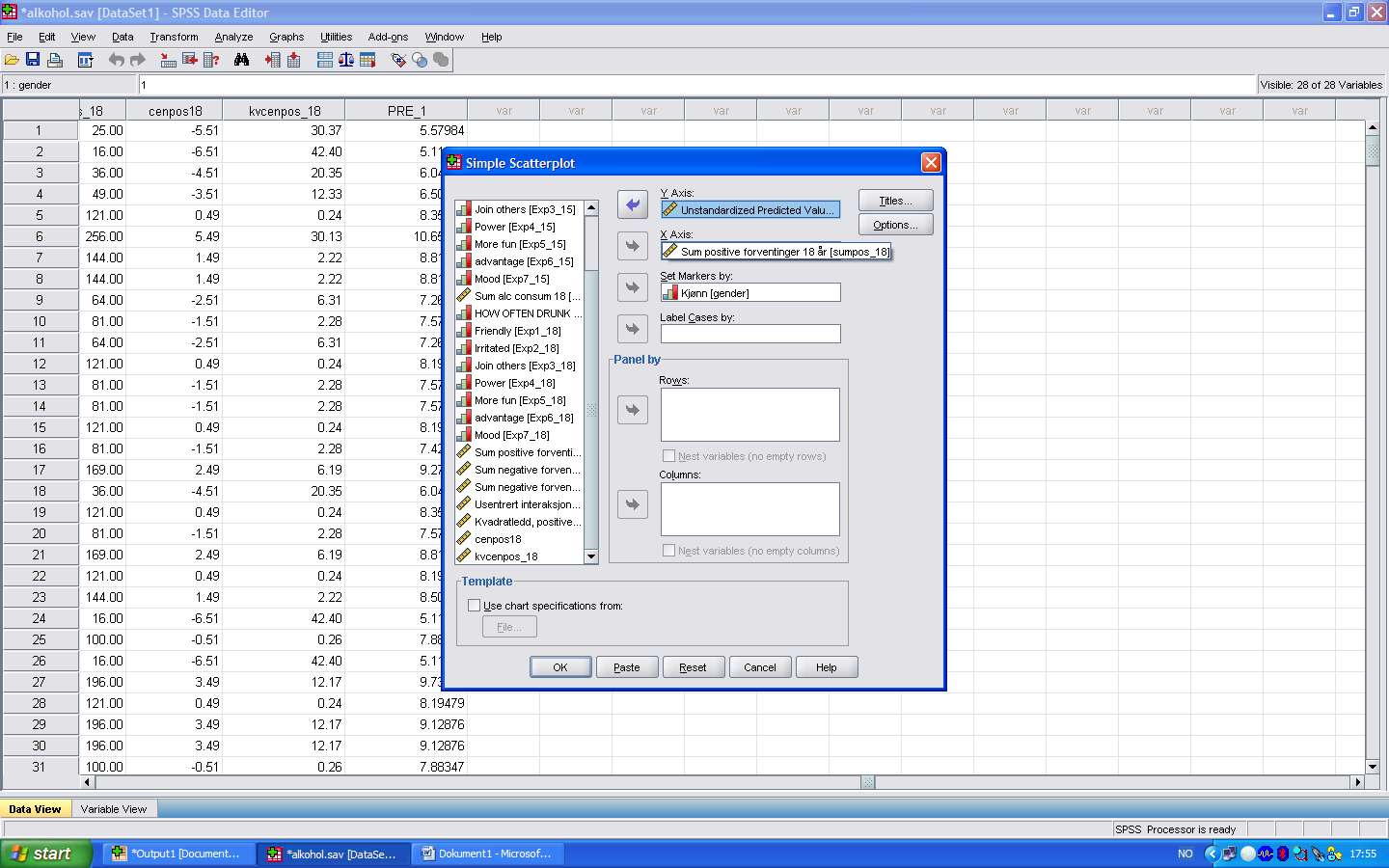
Gjør samme analyse på nytt, men lagre predikert verdi på alkoholbruk. Klikk på knappen ”Save”… og velg deretter ”Unstandardised” under Predicted values. 

Klikk Continue og OK, og du får nå en ny variabel i datafilen, **PRE\_1**.

Deretter kan vi lage et scatterplot/spredningsdiagram med separate linjer for gutter og jenter, noe som kan lette tolkningen av interaksjonseffekten.

Velg **Graphs…**, **Legacy Dialogs** og deretter **Scatter/dot...**

Velg deretter **Simple Scatter** og klikk på **Define:**



Variabelen PRE\_1 (predikert alkoholbruk) settes på Y-axis.

Positive alkoholforventinger ved 18 års alder på X-axis.

I Set markers by velger du Kjønn.

Gjør denne grafen det lettere å tolke interaksjonseffekten?

**Oppgave 5: Ikke-linearitet/Kurvelinearitet**

Vi har mistanke om at effekten av positive alkoholforventinger (M) på alkoholbruk (Y) er kurvelineær. Det kan f eks tenkes at det å gå fra høye forventinger til svært høye forventninger ikke skaper like stor endring i alkoholbruk som f eks å gå fra lave forventinger til moderate forventinger. Sagt mer teknisk: effekten av M på Y er forskjellig for ulike verdier av M.

En kurvelineær sammenheng kan modelleres på mange måter, men i vårt tilfelle er vi interessert i å teste ut om det er slik at effekten av alkoholforventinger på alkoholbruk ”flater ut” etter hvert, og da kan en modell med et kvadratledd være grei å teste ut (det finns flere muligheter, men kvadratledd duger i vårt tilfelle).

Prinsippet blir da å undersøke om modellen med kvadratledd har signifikant høyere R2 enn modellen uten kvadratledd av alkoholforventinger.

Modell uten kvadratledd, dvs. ren lineær effekt:

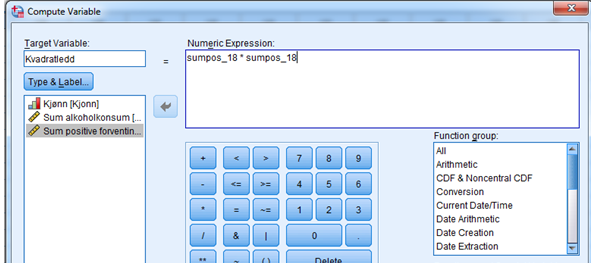
= a + b1M

Modell med kvadratledd, dvs. med kurvlineær effekt:

= a + b1M + b2M2

Et kvadratledd lager vi gjennom å multiplisere den uavhengige variabelen vi tror har en kurvlineær effekt med seg selv dvs. Kvadratledd = (sumpos\_18)2, eller sumpos\_18 \*sumpos\_18

Bruk Transform – Compute-kommandoen til å lage en ny variabel:



Bruk samme fremgangsmåte som i oppgave 4 for å teste ut om kvadratleddet bidrar til å øke modellens forklaringskraft (dvs. bruk hierarkisk regresjon og se om R2 øker når kvadratleddet inkluderes i modellen).

**Grafisk framstilling av ikke-linearitet/kurvelinearitet**

Hvis du finner tegn på at vi har en ikke-lineær effekt av alkoholforventinger på alkoholbruk, er det vanskelig å tolke denne med bare å se på utskriften fra regresjonsanalysen. Det blir mye lettere å illustrere denne grafisk, gjennom å lage et scatterplot mellom alkoholforventinger og predikert alkoholbruk ut i fra modellen som har med kvadratleddet.

Bruk samme fremgangsmåte som når dere illustrerte interaksjonseffekten grafisk i oppgave 4, dvs. lagre ”unstandardized predicted value” når dere estimerer regresjonsmodellen med kvadratleddet inkludert.

**Til slutt:**

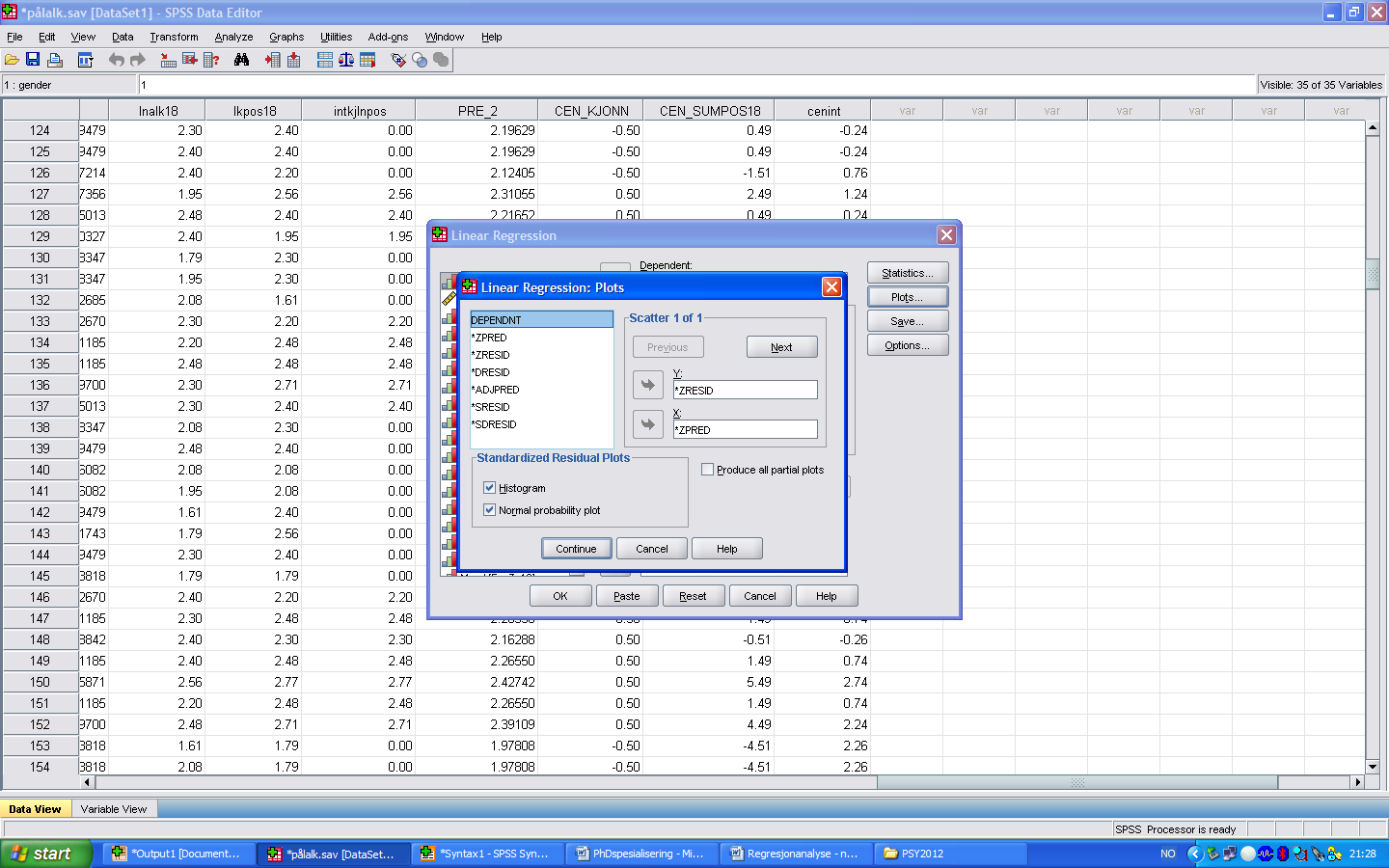
Estimer en regresjonsmodell med disse variablene:

* Kjønn, alkoholforventninger, interaksjonsleddet mellom kjønn og alkoholforventinger og kvadratleddet for alkoholforventinger.

Lagre ”Unstandardized predicted value» for denne modellen/analysen. Benytt til slutt samme fremgangsmåte som grafisk framstilling i oppgave 4 for å lage et scatterplot med predikert alkoholbruk ut i fra denne modellen som Y-aksen og alkoholforventinger som X-aksen, med separate linjer for gutter og jenter.

**Oppgave 6: Sjekk om forutsetningene om normalfordeling og homoskedastisitet for residualene er oppfylt**

Forutsetningen om normalfordelte residualer kan lett undersøkes gjennom å klikke på Plots, og deretter krysse av disse to. Histogrammet du får ut er også egnet til å se om det er noen utliggere i *modellen* du har estimert.



Du kan undersøke om residualene er likt fordelt for ulike predikerte verdier av alkoholbruk gjennom å bestille dette diagrammet (plot), mellom standardiserte residualer og standardiserte predikerte verdier av den avhengige variabelen.

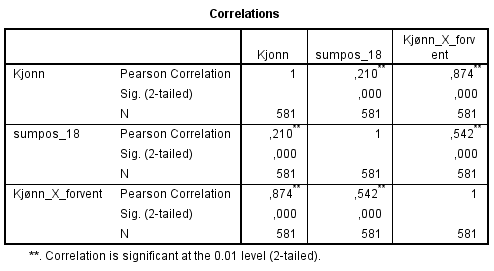
Ser det ut til at forutsetningene om residualene er oppfylt?

**Frivillig tilleggsoppgave: Om sentrering av variabler i regresjonsanalysen**

Når vi lager en ny variabel ut i fra variabler som allerede er inkludert i modellen, slik som med interaksjonsleddet og kvadratleddet, så blir denne nye variabelen naturlig nok svært høyt korrelert med de variablene den er laget ut i fra. **Dette har ingen konsekvenser for modellens forklaringskraft, dvs. R2**, men dette kan skape et problem som kalles multikollinearitet.

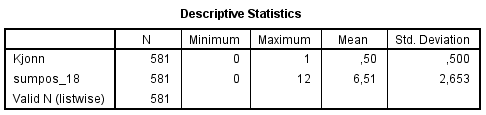
Multikollinearitet er først og fremst et problem for signifikanstestene av effekten den enkelte uavhengige variabel i modellen, samt at det kan vanskeliggjøre tolkingen av den enkelte effekt. Dette skyldes at de uavhengige variablene får lite unik varians, dvs. varians som den ikke deler med de andre uavhengige variablene i modellen. Dermed øker standardfeilen til den enkelte regresjonskoeffisient og dette gjør det vanskelig å få signifikante effekter.

Undersøk korrelasjonen mellom kjønn, positive alkoholforventinger og Interaksjonsleddet mellom kjønn og alkoholforventinger. Korrelasjonstabellen skal bli slik; som du ser er korrelasjonene mellom kjønn og interaksjonsleddet svært høy:



En måte å redusere dette problemet på, er og gjennomsnitts-sentrere de uavhengige variablene som man lager interaksjonsledd ut i fra eller variabler som man lager kvadratledd ut i fra. Sentrering betyr at man gjør om de opprinnelige variablene til avviksskårer fra gjennomsnittsverdi, altså lages den nye sentrerte variabelen gjennom å trekke fra gjennomsnittsverdien fra (rå)skåre på den opprinnelige variabelen.

I vårt tilfelle er gjennomsnittskåren på kjønn og positive forventinger ved 18 års alder:



De nye sentrerte skårene for kjønn og positive forventinger blir dermed slik:

COMPUTE CEN\_KJONN = GENDER - 0.50.

COMPUTE CEN\_ SUMPOS18 = SUMPOS\_18 - 6.51.

**Oppgave:**

1. Lag sentrerte skårer for kjønn og positive alkoholforventinger ved 18 års alder.
2. Korreler de sentrerte skårerne med de opprinnelige skårene.
3. Lag interaksjonsledd mellom de sentrerte skårene til kjønn og positive alkoholforventinger ved 18 års alder.
4. Korreler interaksjonsleddet med både de sentrerte og opprinnelige skårene til kjønn og positive alkoholforventinger. Hva finner du og hvorfor tror du dette skjer?
5. Estimer modellen uten og med interaksjonsledd på nytt, men bruk nå de sentrerte skårene, samt interaksjonsleddet som er laget ut i fra de sentrerte skårene. Hva finner du?