

## Appendiks 5

### Forutsetninger for lineær regresjonsanalyse

Det er flere krav til årsaksslutninger i regresjonsanalyse. En naturlig forutsetning er tidsrekkefølge og i andre rekke spiller variabeltype inn. Ringdal (2007) skiller mellom tre typer variabler. Eksogene variabler kan bare være x-variabler, det finnes også de som kan være både x- og y-variabler og endogene variabler som helst bør være y-variabler. Eksempler på den første typen er bakgrunnsvariabler som alder og kjønn. Den neste typen er variabler som kan beskrive oppnådde statuser som utdanning og variabler knyttet til yrkesaktivitet, som lønn. Den tredje typen, endogene variabler, måler verdier og holdninger. I utgangspunktet bør slike variabler helst opptre som avhengige variabler i regresjonsanalyser, fordi det er meget vanskelig å fastslå eventuelle årsaksrekkefølger mellom dem. Hvis verdier og holdninger skal benyttes som x-variabler, bør det være basert på sterk teori, og aller helst bør analysen være basert på tidsdata med minst to måletidspunkt. Det tredje kravet til en årsaksslutning betyr at den observerte sammenheng mellom X og Y ikke skal forsvinne når en kontrollerer for en annen variabel. I denne analysen brukes flere endogene variabler som x-variabler. Dette er basert på et sterkt teorigrunnlag. Innenfor rammene av denne oppgaven har det dessverre ikke vært mulig å bruke tidsdata.

Den første forutsetningen for regresjonsanalysen er at X er korrekt spesifisert. Disse punktene må være oppfylt for denne forutsetningen (Ringdal 2007:380).

- Alle relevante X må være tatt med og alle irrelevante fjernet fra modellen. Alle X-variablene oppfattes som faste, det vil si uten målefeil.
- Sammenhengen mellom X-variablene og Y er lineære
- Modellen er additiv, det betyr at det ikke er samspill mellom X-variablene.

Deretter følger fire forutsetninger om restleddet eller residualene, og en om sammenhengen mellom X-variablene i multippel regresjon (Ringdal 2007:380).

1. Residualene har et gjennomsnitt på 0 i populasjonen.
2. Residualene har lik varians med alle X, homoskedastisitet.
3. Residualene er ukorrelerte med hverandre og med X-variablene.
4. X-variablene må ikke være perfekt korrelerte, verken parvis eller gruppevis.

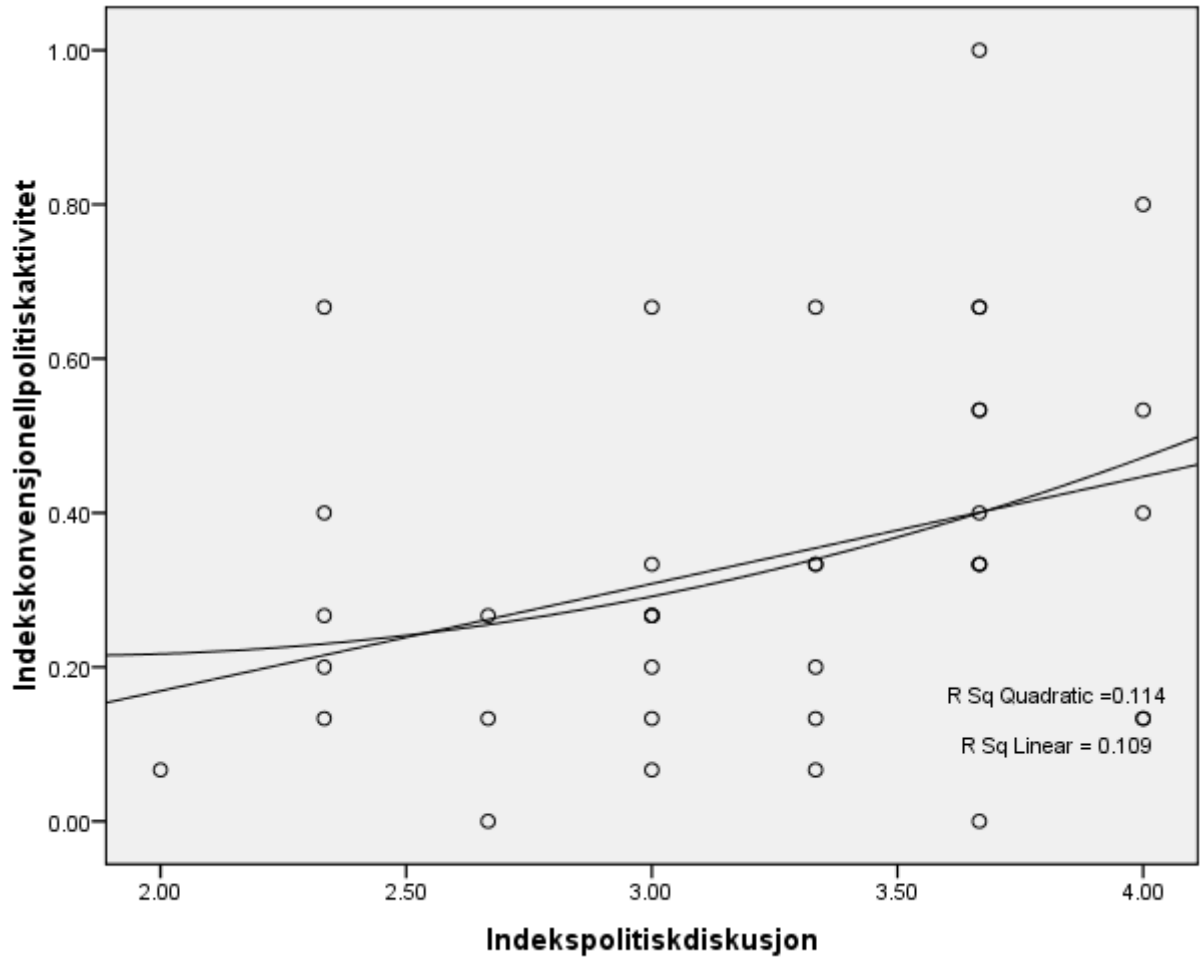
Hvis ikke forutsetningene er oppfylte, kan vi ikke stole på at resultatene er riktige. Det er ikke mulig å teste alle forutsetningene for regresjon, som eksempelvis om alle de relevante X-

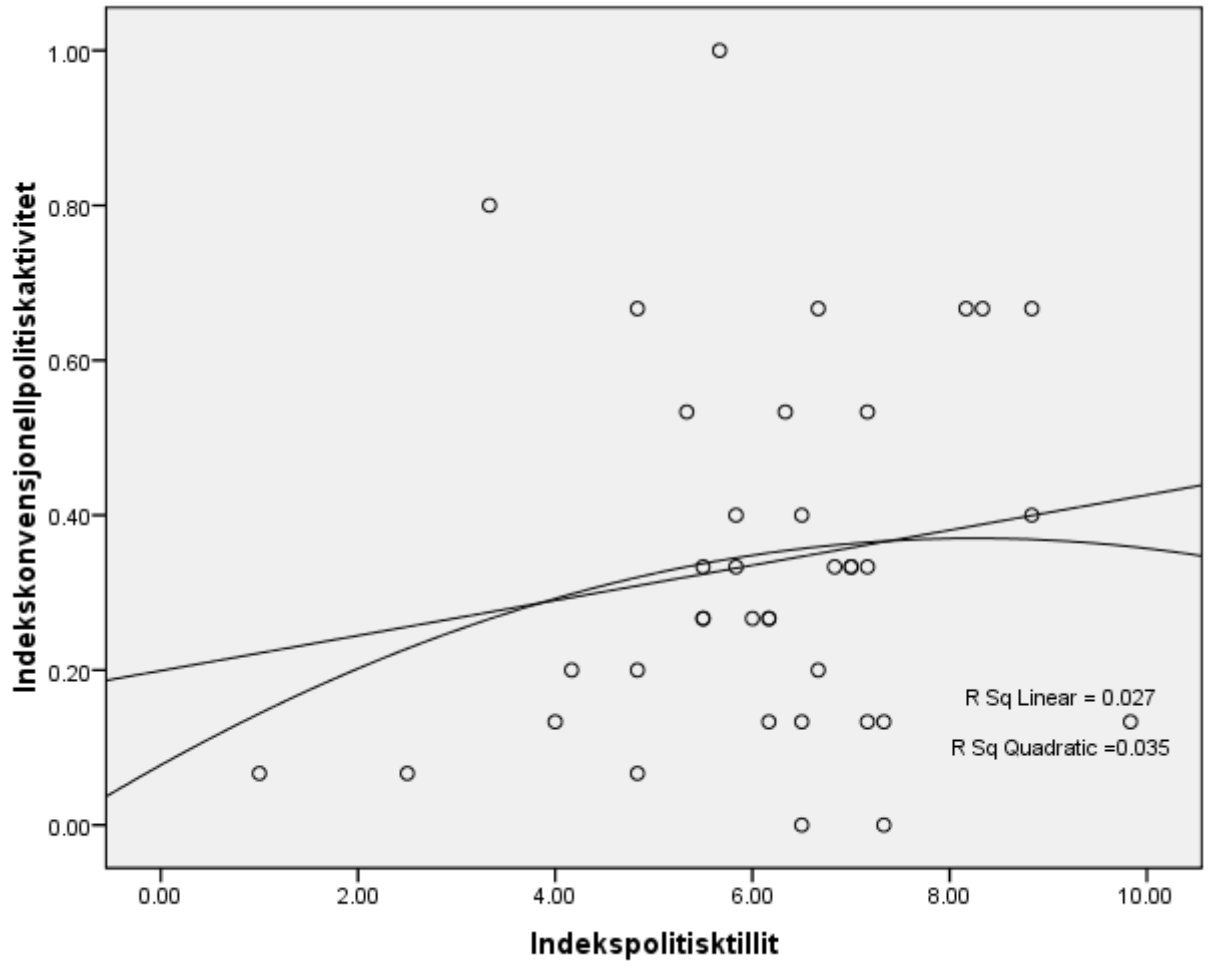
variablene er tatt med, om vi har målefeil på X-ene eller om forventet verdi til feilledet er 0. Jeg vil her teste linearitetskravet og kravet om normalfordelte residualer. Multikollinearitet er testet for i kapittel 7. Kravet om fravær av heteroskedastisitet er ikke relevant for denne analysen, da det er viktig i sammenhengen gyldig statistisk generalisering av resultatene fra utvalget til populasjonen. Jeg har undersøkt et ikke-sannsynlighetsutvalg fra en sjelden populasjon, det er derfor ikke aktuelt å generalisere. Autokorrelasjon er heller ikke aktuelt da jeg ikke bruker tidsseriedata.

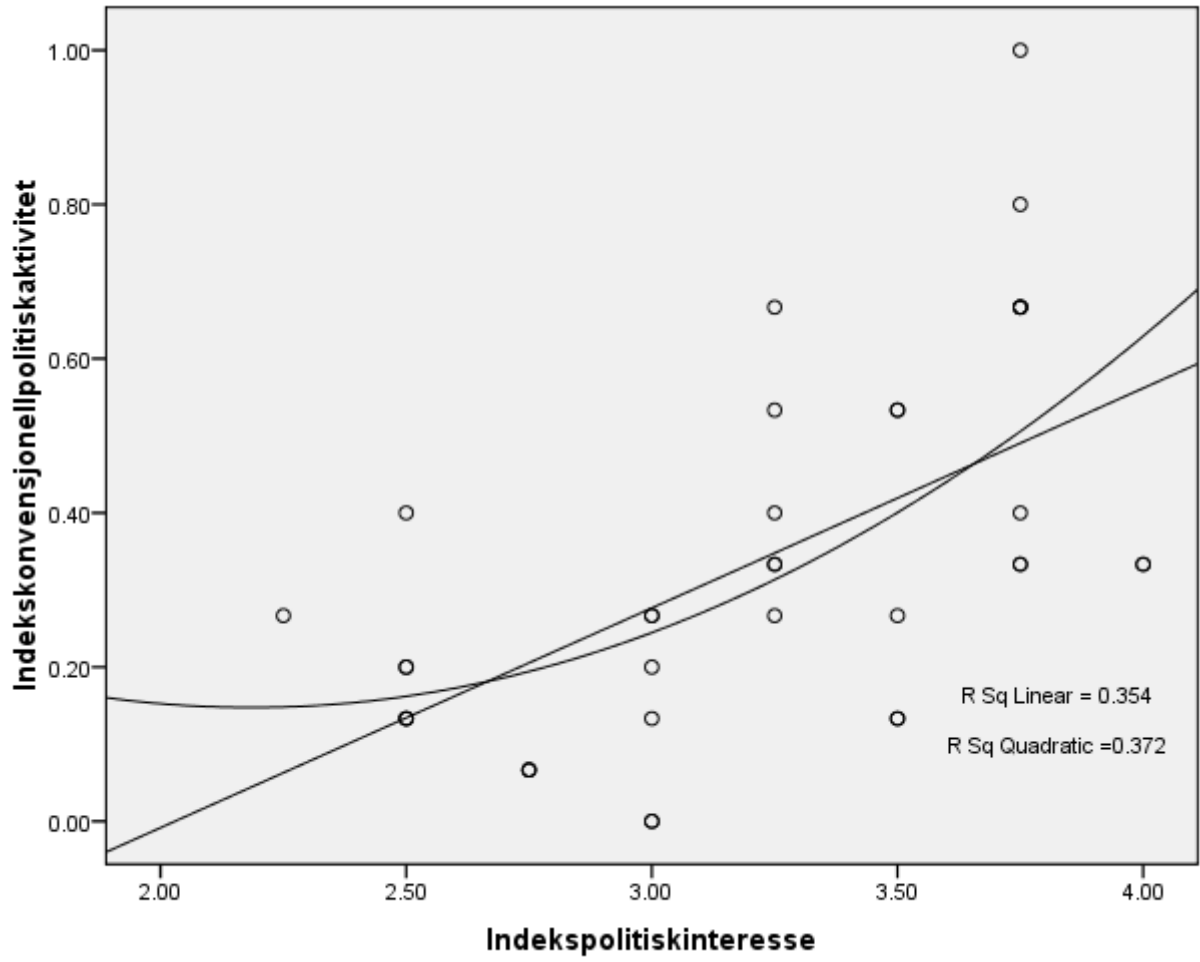
### **Testing av linearitetsforutsetningen**

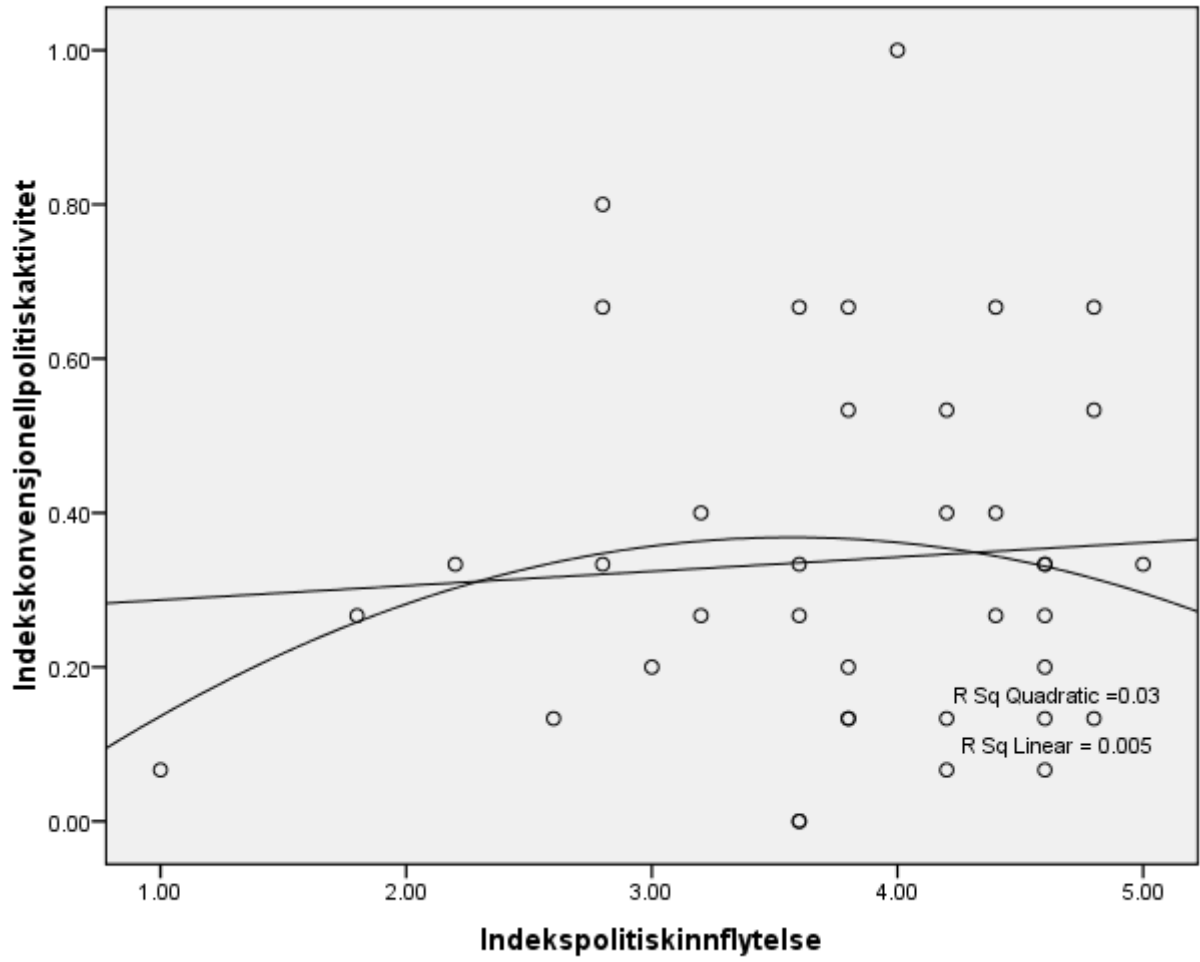
Forutsetningen om at sammenhengen mellom X og Y er lineær, er grunnlaget for lineær regresjon, men økningshastigheten i Y kan være sterkere eller svakere enn økningshastigheten i X. Eikemo og Clausen (2007:65) påpeker at dette ikke er uvanlig, og at linearitetsforutsetningen ikke er så alvorlig å bryte. Det betraktes som mer "bekvem" å ha lineære modeller fordi de er lettere å tolke.

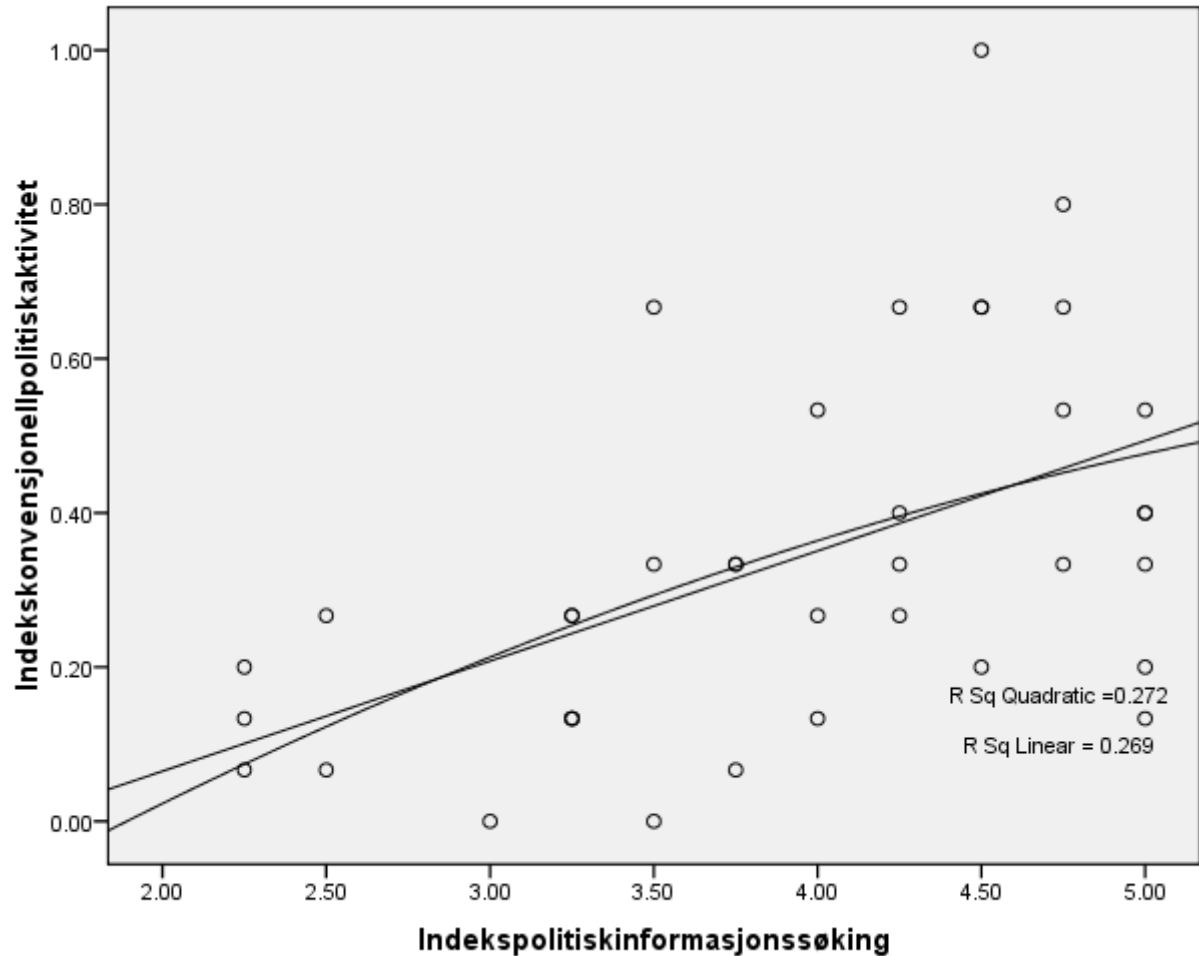
For å teste linearitetsforutsetningen ser jeg nærmere på et prikkdiagram over den observerte sammenhengen mellom de uavhengige variablene og Y. Jeg går inn på grahps, scatter, simple scatter og setter inn variablene. Jeg dobbeltklikker så på plottene i outlook-vinduet og trykker deretter på høyre museknapp på den samme grafen i det nye vinduet som kommer opp. Jeg velger da "add fit line at total" og merker av for quadratic og linear og trykker apply. Da kan jeg se teste både om det er en lineær eller en mulig kurvelineær sammenheng (Eikemo og Claussen 2007:65). Nedenfor følger prikkdiagram for de uavhengige holdningsvariablene.







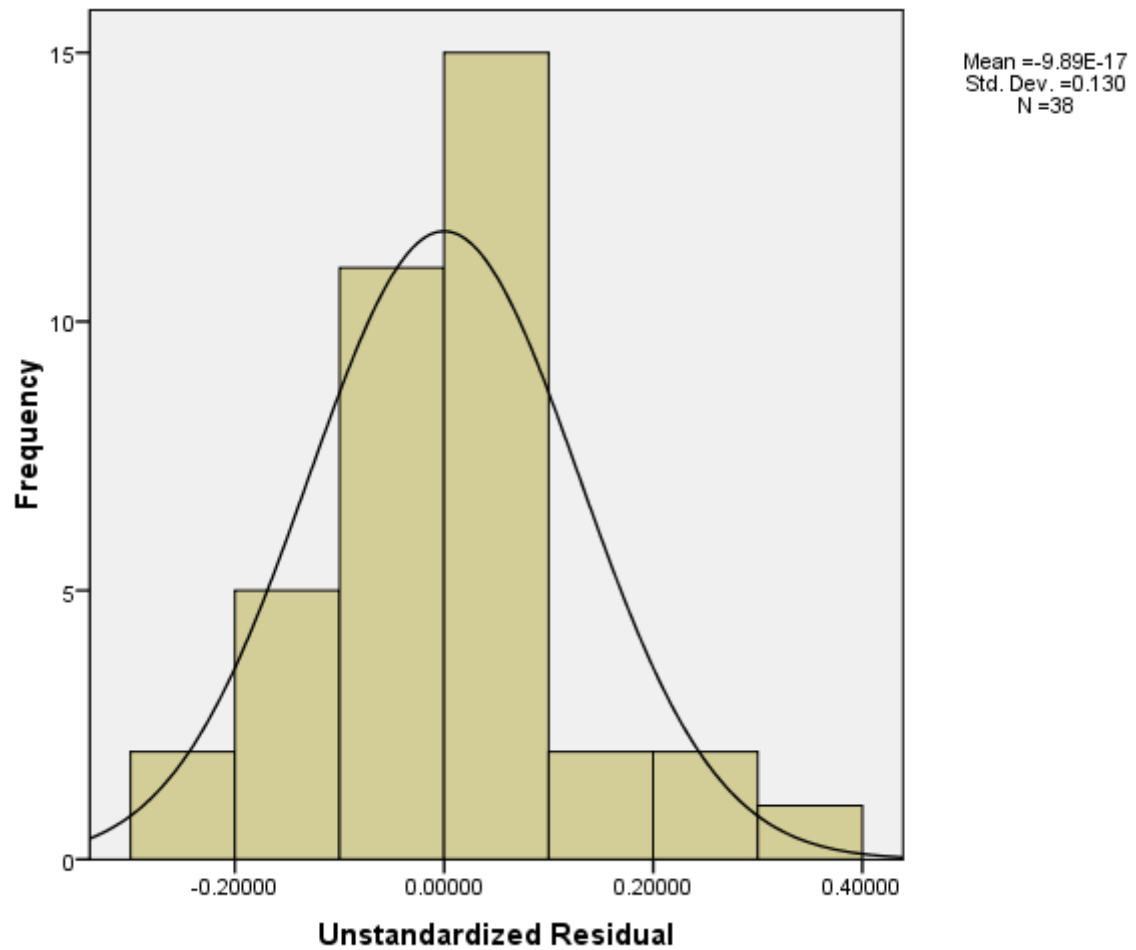




Kort oppsummert er det en lineær modell en god tilnærming for dette materialet. Når variasjonsbredden i det fenomenet jeg studerer er begrenset, gir den rette linjen en god nok approksimasjon innen for det intervallet jeg i praksis har data for (Skog 2004:237).

### Kravet om normalfordelte residualer

Denne forutsetningen gjelder kun dersom utvalgsstørrelsen er svært liten, hvilket er tilfellet for denne undersøkelsen. Jeg bruker data med få observasjoner, derfor er dette viktig å teste. Normalfordelingen er et mål som alltid vil oppfylles helt. Men grafen nedenfor viser at residualene er tilnærmet normalfordelte. Halen til høyre er imidlertid litt lengre enn det venstre.



### Oppsummering

Denne modellen bryter ikke alvorlig med noen av forutsetningene for lineær regresjon og resultatene er dermed pålitelige til å si noe om dette utvalget.