

Inge Dyrhol

# Folketalsutvikling og kommunale prioriteringar



**HØGSKULEN I VOLDA**



**MØREFORSKING  
VOLDA**

2007

<b>Forfattar</b>	Inge Dyrhol
<b>Ansvarleg utgjevar</b>	Høgskulen i Volda
<b>ISSN</b>	0805-8075
<b>Sats</b>	Inge Dyrhol
<b>Distribusjon</b>	<a href="http://www.hivolda.no/fou">http://www.hivolda.no/fou</a>

© Forfattar/Høgskulen i Volda

Føresegnene i åndsverklova gjeld for materialet i denne publikasjonen. Materialet er publisert for at du skal kunne lese det på skjermen eller framstille eksemplar til privat bruk. Utan særskild avtale med forfattar/Høgskulen i Volda er all anna eksemplarframstilling og tilgjengeleggjering berre tillate så langt det har heimel i lov eller avtale med Kopinor, interesseorgan for rettshavarar til åndsverk.

**Notatserien** er for ulike slag publikasjonar av mindre omfang, t.d. forprosjektnotat, foredrag, artikkelutkast o.a. Eit hovudføre mål med serien er å stimulere til publisering og fagleg debatt i miljøet. Spreinga går i hovudsak til fagmiljøet i Volda, til eksterne fagmiljø og personar som forfattar(ar) ønskjer kommentarar frå. Kvar forfattar er ansvarleg for sitt arbeid.

*Dette notatet inneheld ein analyse av den statistiske samanhengen mellom folketalsutviklinga i kommunar og korleis kommunar prioriterer mellom sektorar. Lesarar utan eigentleg interesse for multippel lineær regresjonsanalyse får orsake at eg og har dokumentert ein repetisjon og vidareføring av kunnskapar om det emnet, vonleg på ein måte som og andre kan ha ein viss nytte av<sup>1</sup>.*

Multippel lineær regresjonsanalyse kan nyttast til fleire føremål. Grovt sett kan føremåla klassifiserast i predikasjon og forklaring. Ein kan ikkje oppnå forklarande kraft utan predikativ kraft (Hair Jr. mfl 2005:189-190). Ein predikativ bruk kan difor vere uttrykk for ein resignasjon i høve til ein forklarande bruk. I dette skulle også ligge at ein analyse som eigentleg er ute etter forklaring ikkje er verdlaus om ein berre eit stykkje på veg maktar meir enn brukbar predikasjon. Ein kan også hevde at ein analyse som ikkje ein gong gjev god predikasjon kan vere nyttig i dei tilfelle der rimelege hypotesar om samanhengar ikkje blir støtta, – det aukar også innsikta.

Når ein analyse har predikativ kraft kan den vere nyttig i avgjerdssituasjonar. Eg kan tenkje meg at dette kan vere tilfelle i visse medisinske avgjerdssituasjonar. Dersom analysen predikerer at ein viss ”dose” behandling har positiv effekt, så kan det vere aktuelt å gje behandlinga sjølv om ein ikkje heilt forstår kvifor behandlinga verkar. Det er kanskje viktigare å kjenne til ev skadelege bieffektar. Innan økonomi kan det vere liknande situasjonar. Eit eksempel kan vere når Norges Bank fastset styringsrenta. Rett nok har Norges Bank ikkje vike attende for forsøk på å forklare verknaden på inflasjonen, men eg har like mykje sans for ein tidlegare sentralbanksjef i USA som i 2003 sa: ”Sentralbanken veit ikkje korleis økonomien fungerer ...”.

For kommunar er positiv folketalsutvikling noko ein ikkje kan ta avgjerd om, men som ein har eit sterkt ønskje om. Derimot kan kommunar ta avgjerd om kor stor prosent av budsjettet sitt dei vil bruke på ymse sektorar. Fullstendig fridom er det ikkje tale om på grunn av lover m.m. Prosenttal har også ei samla grense på 100%. Likevel kan det vere interessant å sjå kva fordeling mellom sektorar som kjenneteiknar kommunar som har vekst i folketallet. Dersom kommunar med positiv folketalsutvikling viser seg å bruke mykje på sektor  $x$  så kan det vere rimeleg å rå andre kommunar til å gjere det same, – sjølv om ein ikkje veit så mykje om kva som er årsak og kva som er verknad, eller om prioriteringa av sektoren verkar direkte eller indirekte.

Tiebout-hypotesen (Tiebout 1956) seier at folk røyster med føtene sine og, dvs. flyttar dit lokalpolitikarane legg tilhøva best til rette. Forskinga gjev eit samla inntrykk av at denne

---

<sup>1</sup> Eg meiner det er for vanleg å akseptere kunnskap på nokså reint puggenivå når det gjeld statistiske metodar ein brukar, ikkje minst blant folk som elles er sterkt kritiske til den såkalla puggeskulen. Men eg har mykje att å lære, og vonar dette ikkje pregar det som er skrive i nemnande grad. Eg er og medviten om punkt som byggjer på ”pugg”. For andre kan det sikkert verke litt tilfeldig kva som er gått litt nøyare innpå og kva som ikkje er det.

hypotesen ikkje blir særleg støtta verken internasjonalt eller i Norge (Berg og Rattsø 2005:43-44). Eg skal ikkje eksplisitt studere flytting, men endring i folketal som i denne samanheng er den viktige målvariabelen for kommunane.

Eg har studert samanhengen mellom endring i folketal og kor stor prosent av brutto driftsutgifter kommunane prioriterer å bruke på ymse sektorar. Prosentar er relative storleikar. Her følgjer nokre kommentarar om val av Prosent brutto driftsutgifter pr. sektor.

I mine data frå 424 kommunar (nesten alle), og som er henta frå KOSTRA, er Pearson's  $r$  -0,339 mellom Endring folketal 2006 i prosent og Brutto driftsutgifter 2006 totalt pr. innb. Signifikansen er på 0,001 nivå. Eg berre nemner dette innleiingsvis. Det er eit kjent resultat at små kommunar, som også gjerne er høvesvis tungvinte å drive, har høge Brutto driftsutgifter pr. innb. Slike kommunar har og gjerne nedgang i folketalet. Eit syn er at kommunane kjem gunstigare ut i inntektssystemet enn vekstkommunar. Ein kan ikkje trekkje den konklusjon at høge Brutto driftsutgifter pr. innb. har negativ innverknad på folketalsutviklinga. Ein rimelegare hypotese er at (graden av) høgare Brutto driftsutgifter pr. innb. ikkje er tilstrekkeleg til å sikre positiv utvikling i folketalet. Samvariasjonen som påvist ovanfor gjer det ikkje mindre interessant å sjå Prosent brutto driftsutgifter pr. sektor.

Brutto driftsutgifter pr. innb. pr. sektor kunne og vere interessant å trekkje inn. Desse variablane er ikkje i same grad som Prosent brutto driftsutgifter pr. sektor avgjerdsvariablar sett frå den einskilde kommunen si side. T.d. storleiken av frie inntekter pr. innb., som kommunen har svært avgrensa innverknad på, spelar ei vesentleg rolle for Brutto driftsutgifter pr. innb. pr. sektor. Desse absolutte tala pr. sektor kan også alle vere høgare i ein kommune enn i ein annan, i motsetnad til prosenttal. Ein multipel lineær regresjonsanalyse med Endring folketal 2006 i prosent og Brutto driftsutgifter 2006 pr. innb. pr. sektor gjev justert  $R^2$  på 0,452<sup>2</sup>. Ein multipel lineær regresjonsanalyse med Endring folketal 2006 i prosent og Prosent brutto driftsutgifter 2006 pr. sektor gjev justert  $R^2$  på 0,480. Justert  $R^2$  aukar til 0,522 om ein tek med begge sett uavhengige variablar, og til 0,547 om ein også tek med settet Prosent endring brutto driftsutgifter 2006 pr. sektor. Bruk av siste settet åleine gjev justert  $R^2$  lik 0,162. For setta Prosent brutto driftsutgifter 2006 pr. sektor og Prosent endring brutto driftsutgifter 2006 pr. sektor vert justert  $R^2$  lik 0,477. Alle F-verdiane<sup>3</sup> er signifikante på 0,001 nivå.

---

<sup>2</sup>  $R$ ,  $R^2$  og justert  $R^2$  er forklart noko nøyare i Appendiks 1.

<sup>3</sup> F-testen testar nullhypotesen at alle koeffisientane til dei uavhengige variablane, dvs med unntak av konstanten, er null. Dersom ein analyse ikkje går igjennom denne testen, så har ein så langt "ingen modell i det heile". Utan signifikant F-verdi er ikkje  $R^2$  signifikant større enn null (Hair Jr. mfl 2005:215). (Kvar einskildkoeffisient inngår når  $R$  vert rekna ut). Men sjølv med grei F-test kan analysen ha mange svake sider i høve til føresetnadene som ligg bak multipel regresjonsanalyse.

Som nemnt gjev ein multippel lineær regresjonsanalyse med Endring folketal 2006 i prosent og Prosent brutto driftsutgifter 2006 pr. sektor ein justert  $R^2$  på 0,480. I tillegg å bruke Brutto driftsutgifter 2006 totalt pr. innb. som uavhengig variabel gjev berre små endringar også i regresjonskoeffisientane for Prosent brutto driftsutgifter 2006 pr. sektor. Dette tyder på at kommunane si *prioritering* mellom ymse sektorar verkeleg har sjølvstendig ”forklaringskraft” overfor folketalsutviklinga.

Etter ei samla vurdering av funna framanfor og den avgjerdsretta intensjonen med dette arbeidet går eg berre vidare med Prosent brutto driftsutgifter pr. sektor som uavhengige variablar. Fordeling mellom sektorar er sjeldan noko som har oppstått brått. I dei fleste tilfelle er prosentane eit resultat av ei utvikling over lang tid. Eg vel difor å ta utgangspunkt i tal frå 2006, same året som endring folketal i prosent er frå. Eg vil i drøftinga likevel trekkje inn data frå 2005.

I KOSTRA er den kommunale aktiviteten delt i femten sektorar: Adm., styring og fellesutgifter, Barnehagar, Barnevern, Bustader, Brann og ulukkesvern, Grunnskule, Kyrkje, Kommunehelse, Kultur, Fys. planl., kulturminne, natur og nærmiljø, Næring, Pleie og omsorg, Samferdsel, Sosialteneste, VAR. Denne inndelinga er ikkje heilt uttømande. Summen av Prosent brutto driftsutgifter pr. sektor blir difor ikkje fullt 100%, noko som og har ei meir teknisk årsak.

## Drøfting i høve føresetnader for multippel lineær regresjon

Partielle regresjonsplott<sup>4</sup> viser ingen klare ikkje-lineære trekk.

Residualen (restleddet) oppfyller i stor grad føresetnaden om normalfordeling<sup>5</sup>.

*Heteroskadastisitet* (Hair Jr. mfl 2005:83-84) mellom standardisert avhengig variabel og standardiserte residualar er ikkje noko problem<sup>6</sup>.

### Korrelasjon mellom forklaringsvariablane

*Kolinearitet* er korrelasjon mellom to uavhengige variablar, *multikolinearitet* mellom tre eller fleire. Det er ikkje uvanleg å kalle begge delar multikolinearitet, jf. Hair Jr. mfl (2005:186), som skriv:

The impact of multicollinearity is to reduce any single independent variable's predictive power by the extent to which it is associated with the other independent variables (: 186).

og

The simplest and most obvious means of identifying collinearity in an examination of the correlation matrix for the independent variables. The presence of high correlations (generally .90 and higher) is the first indication of substantial collinearity. Lack of any high correlation values, however, does not ensure a lack of collinearity. Collinearity may be due to the combined effect of two or more other independent variables (termed *multicollinearity*) (:227)

M.a. Johannessen, Kristoffersen og Tufte (2004:317) brukar 0,7 som grense for kolinearitetsproblem. Det er grunn til å streke under at all samvariasjon mellom uavhengige variablar svekker klårleiken i ein multippel lineær regresjonsanalyse.

Mellom Prosent brutto driftsutgifter 2006 for ymse sektorar er det ingen korrelasjonskoeffisientar over 0,6. Den høgste er 0,589 mellom Prosent brutto driftsutgifter 2006 Barnehagar og Prosent brutto driftsutgifter 2006 Grunnskule.

---

<sup>4</sup> Dette er diagram som viser samanhengen mellom den avhengige variable og kvar einskild uavhengige variabel. I *partiell* ligg at effektane av dei andre uavhengige variable er haldne konstante. Den grafiske framstillinga er eit visuelt hjelpemiddel for m.a. å oppdage ikkje-lineære samanhengar (Hair Jr. mfl 2005:173), som ikkje er foreinlege med lineær regresjon.

<sup>5</sup> Føresetnaden om normalfordeling har slik bakgrunn: "Seldom will our predictions be perfect. We assume that random error will occur, but we assume that this error is an estimate of the true random error in the population ..., not just the error in prediction for our sample" (Hair Jr. mfl 2005:174).

<sup>6</sup> Det motsette, homoskadastisitet tyder "lik varians". I vårt tilfelle tyder ikkje eit plott på at standardiserte residualar fordelar seg ulikt på ulike nivå av standardisert avhengig variabel. Plottet er gjort i SPSS ved *Regression > Linear > Plots* og standardisert avhengig variabel *ZPRED* som x-variabel og standardiserte residualar *ZRESID* som y-variabel (Wenstrøm og Bagøien).

Men som nemnt kan det likevel vere problem med multikolaritet. Dei to mest vanlege måtane å vurdere dette på er ved å rekne ut *toleransen* og den inverse av denne som vert kalla *VIF* (Variance Inflation Factor). Toleransen er:

... the amount of variability of the selected independent variable not explained by the other independent variables (Hair Jr. mfl 2005:227).

Ein tek difor for seg kvar uavhengige variabel for seg og reknar ut ein  $R^2$  som viser i kva grad den utvalde variabelen let seg forklare ved hjelp av dei andre uavhengige variablane. Toleransen til den utvalde variabelen er då  $1 - R^2$ . Dess høgare toleranse, dess mindre multikolaritet. Hair Jr. mfl (2005:230) skriv at ein minste toleranse på 0,1 er ei mykje brukt grense. Dette tilsvarar ei øvre grense for VIF på 10. Ingen av Prosent brutto driftsutgifter 2006 for ymse sektorar bryt desse grensene for toleranse og VIF. Slike grenser er ikkje absolute, men ingenting tilseier at å gå vidare i denne analysen er utanfor vanleg praksis for det som blir akseptert når det gjeld korrelasjon mellom forklaringsvariablane.

## Avgrensa modell

Den endelege modellen er stegvis arbeidd fram i samsvar med Hair Jr. mfl (2005:211-212). Det er først køyrt ein regresjonsanalyse berre med den uavhengige variabelen som har høgast korrelasjon med den avhengige, i dette tilfelle  $r = 0,599$  for Prosent brutto driftsutgifter barnehagar 06.

Ein analyse med Endring folketal 2006 i prosent og Prosent brutto driftsutgifter 2006 pr. sektor gjev som nemnt justert  $R^2$  på 0,480 når alle sektorar er med. I den analysen er Adm., styring og fellesutgifter, Barnevern, Bustader, Brann og ulukkesvern, Næring, Pleie og omsorg, Samferdsel og Sosialteneste ikkje signifikante<sup>7</sup> på nivå 0,1. Desse vert sett bort frå. Partielle korrelasjonskoeffisientar<sup>8</sup> er rekna ut, og Prosent brutto driftsutgifter 2006 for dei andre sektorane vert arbeidd inn ein om gongen etter storleik. Under vegg i denne prosessen oppstår det ikkje insignifikans for nokon av dei uavhengige variablane.

Den avgrensa modellen oppnår ein justert  $R^2$  på 0,484 (F-verdien er signifikant på 0,001 nivå). Koeffisientane er som i tabell 1.

---

<sup>7</sup> Her er nytta ein såkalla *t*-test som har som nullhypotese at regresjonskoeffisienten er null (Montgomery, Peck og Vining 2006:84). Når ein regresjonskoeffisient vert sagt ikkje å vere signifikant, blir sannsynlegheita for at ein verdi ulik null berre er statistiske tilfeldig rekna som for stor. Ei god utlegging om dette finst i Johannesen, Kristoffersen og Tuft (2004:352-354).

<sup>8</sup> Merk at partielle korrelasjonskoeffisientar er noko anna enn regresjonskoeffisientar. Ein partiell korrelasjonskoeffisient "measures the strength of the relationship between the ... dependent variable and a single independent variable when the effects of the other independent variables in the model are held constant. For example,  $r_{Y, X_2, X_1}$  measure the variation in  $Y$  associated with  $X_2$  when the effect of  $X_1$  on both  $X_2$  and  $Y$  is held constant" (Hair Jr. mfl 2005:173).

Prosent brutto driftsutgifter 2006	Ustandardisert <i>b</i> -koeffisient	Standardisert $\beta$ -koeffisient	Signifikansnivå
Barnehagar	0,235	0,387	0,001
Grunnskule	0,074	0,178	0,001
Kyrkje	-0,511	-0,117	0,01
Kommunehelse	-0,200	-0,191	0,001
Kultur	0,157	0,141	0,001
Fys. planl., kulturminne, natur og nærmiljø	0,324	0,180	0,001
VAR	0,059	0,073	0,05

Tabell 1: Regresjonskoeffisientar med Endring folketal 2006 i prosent som uavhengig variabel

## Drøfting

I Appendiks 2 er ustandardiserte og standardiserte koeffisientar gjort ein del greie for.

Det er liten grunn til å tru at kommunar tenkjer på Kyrkje og Kommunehelse som verkemiddel i arbeidet for å betre folketalsutviklinga. Høg Prosent brutto driftsutgifter 2006 for desse sektorane er venteleg heller eit resultat av negativ folketalsutvikling. For Kyrkje kan også sterke historiske røtar for sokneinndeling forsterke verknaden.

Proseduren ovanfor er og gjennomført med Prosent brutto driftsutgifter 2005 for sektorane, men framleis Endring folketal 2006 i prosent som avhengig variabel. Prosent brutto driftsutgifter for 2005 kan ikkje kome av Endring folketal 2006 i prosent, slik at analysen kan tenkjast å bidra med nyttig innsikt. Kyrkje fell ut som insignifikant, men Barnevern kjem til. Men det er heller ingen grunn til å tru at kommunar heller tenkjer på Barnevern som verkemiddel for å betre folketalsutviklinga, eller at folk i nemnande grad flyttar til ein kommune på grunn av godt Barnevern. Koeffisientane blir elles ikkje så ulike dei i tabell 1. Fys. Planl., kulturminne, natur og nærmiljø og Grunnskule får auka både sin ustandardiserte og standardiserte koeffisient, medan det omvendte gjeld for Barnehagar og Kultur.

Forklaring av variasjonen i Endring folketal 2006 i prosent får i mine data både for 2005 og 2006 for Prosent brutto driftsutgifter størst bidrag, når vi ser bort frå Kyrkje og Kommunehelse, frå Barnehagar, Fys.plal., kulturminne, natur og nærmiljø, Grunnskule og Kultur. Bidraget frå Barnehagar er om lag det doble av bidraget frå dei andre. Toleransane er: Barnehagar 0,249, Fys.plal., kulturminne, natur og nærmiljø 0,716, Grunnskule 0,195 og Kultur 0,506. Dette svekkjer først og fremst oppfatninga av Grunnskule som ein viktig variabel, sidan låg toleranse bør føre til mindre vekt på den standardiserte koeffisienten.

At justert  $R_2 = 0,451$  for analysen med data frå 2005 for uavhengige variablane, og altså lågare enn for analysen med data for uavhengige variablar frå 2006, støttar isolert sett tidlegare forskning i at kommunale tenester ikkje spelar så stor rolle for flytting og dermed gjerne folketalsutvikling. På den andre sida er det interessant at den forklarte delen av variasjonen slett ikkje er liten verken med data for 2005 eller 2006.



## Appendiks 1

### $R$ , $R^2$ og justert $R^2$

$R$ , den multiple korrelasjonskoeffisienten, vert rekna ut slik:  $R = \sqrt{\sum_{j=1}^n r_{jy} \left( \sum_{i=1}^n \frac{r_{iy}}{r_{ij}} \right)}$ . Her viser  $y$

til den avhengige variabelen,  $r$  er korrelasjonskoeffisient og  $j$  og  $i$  tel opp dei uavhengige variablane ( $x$ -ane). Vi ser at gitt korrelasjonskoeffisientane blant  $x$ -ane blir  $R$  stor dersom alle  $x$ -ane enkeltvis er sterkt korrelerte med  $y$ , og gitt korrelasjonskoeffisientane mellom kvar av  $x$ -ane og  $y$  blir  $R$  større dess svakare korrelasjon det er mellom  $x$ -ane innbyrdes. Kvar ny  $x$  vil tilføre meir når den er svakt korrelert med dei andre.  $R$  vil vere mellom 0 og 1.

$R^2$ , den kvadrerte multiple korrelasjonskoeffisienten, er: "Measure of the proportion of the variance of the dependent variable about its mean that is explained by the independent ... variables. The coefficient can vary between 0 and 1. If the regression model is properly applied and estimated, the researcher can assume that the higher the value of  $R^2$ , the greater the explanatory power of the regression equation, and therefore the better the prediction of the dependent variable" (Hair Jr. mfl 2005:170). Her er å merke at "explanatory" ikkje kan tolkast på same måte som i skiljet mellom "Prediction with Multiple Regression" og "Explanation with Multiple Regression" (:190). På norsk vert det ofte sagt at stor  $R^2$  viser *forklaringskraft*. Det er viktig å ha klart for seg at  $R^2$  "forklarer" varians.

Justert  $R^2 = 1 - (1 - R^2) \frac{k - 1}{k - n - 1}$ , der  $k$  er talet på observasjonssett. Ta som utgangspunkt: Om ingen uavhengige variable, dvs  $n = 0$ , vil justert  $R^2$  vere lik  $R^2$ . Når  $n$  aukar vil justert  $R^2$  alt anna likt bli mindre. Behovet for og bruken av justert  $R^2$  kan skildrast slik: "Although the addition of independent variables will always cause the coefficient of determination (innsett her:  $R$ ) to rise (innsett her: "In general,  $R^2$  never decreases when a regressor is added to the model" (Montgomery, Peck og Vining 2006:83)), the adjusted coefficient of determination may fall if the added independent variables have little explanatory power or if the *degrees of freedom* become too small. This statistic is quite useful for comparison between equations with different numbers of variables, differing sample sizes, or both" (Hair Jr. mfl 2005:170). Fridomsgrad i samband med multiple lineære regresjonsanalysar er talet på observasjonar minus talet på ukjende parameter (Sverdrup 1964:228). Talet på ukjende parameter er lik talet på uavhengige variablar pluss ein (Hair Jr. mfl 2005:171,197). Det er koeffisientane og restleddet som er parametrane (Sverdrup 1964:214). Av formelen ser vi at auka fridomsgrad gjev auka justert  $R^2$ . Dersom vi har to observasjonar og ein uavhengig variabel vert fridomsgraden null. To likningar og to ukjende gjev nøyaktig løysing. Dersom det er fleire observasjonar vil vi få "prediction errors (actual minus predicted dependent values)", og i estimeringa av den tilfeldige feilen er der  $n - 2$  fridomsgrader, dvs ein fridomsgrad for kvar observasjon utover to. Stor fridomsgrad "indicates the prediction is fairly robust with regard to being representative of the overall sample of respondents", noko som gjev betre grunnlag

for generalisering, men dette kan altså gå ut over ”predictive accuracy” (:171). Målet er å oppnå høgast mogleg ”predictive accuracy with the most degrees of freedom”, og dette endar med eit skjøn (Hair Jr. mfl 2005:197). Justert  $R^2$  er i denne samanhengen eit mykje brukt hjelpemiddel som motverkar freistinga til å trekkje inn stadig nye uavhengige variablar for å få auka forklaringskraft.

---

## Appendiks 2

### Ustandardiserte og standardiserte koeffisientar

I vår gransking er observasjonane av alle uavhengige variablar i same målestokk, nemleg prosent. Generelt kan målestokken vere ulik. Brukar ein observasjonane som dei er, gjev den multiple lineære regresjonsanalysen ustandardiserte regresjonskoeffisientar, og som gjerne vert symboliserte med  $b$ -ar. Ustandardiserte regresjonskoeffisientar blir slik:  $b_j = \sum_{i=1}^n \frac{r_{iy}}{r_{ij}} \cdot \frac{S_y}{S_j}$ , der  $r$  er korrelasjonskoeffisientar og  $S$  standardavvik.

Når ikkje data for alle uavhengige variablar har same målestokk, viser dei ustandardiserte regresjonskoeffisientane *openbert* ikkje kor viktige kvar av dei uavhengige variablane er. uavhengige variablar har same målestokk. Hair Jr. mfl (2005:174) skriv om ustandardiserte koeffisientar:

In the multiple predictor model (e.g.,  $Y = b_0 + b_1X_1 + b_2X_2$ ), the regression coefficients are partial coefficients because each takes into account not only the relationships between  $Y$  and  $X_1$  and between  $Y$  and  $X_2$ , but also between  $X_1$  and  $X_2$ .

Hair Jr. mfl (2005:225) skriv:

In order to use the regression coefficients for explanatory purposes, we must first ensure that all of the independent variables are on comparable scale. Yes even then, differences in variability from variable to variable can affect the size of the regression coefficient. What is needed is a way to make all independent variables comparable in both scale and variability. We can achieve both these objectives and resolve the problem in explanation by using regression coefficients called the beta coefficients.

og

If we still want the original regression coefficients for predictive purposes, ...  
... , ... multiple regression gives us not only the regression coefficients, but also coefficients resulting from the analysis of standardized data termed beta ( $\beta$ ) coefficients.

Standardisering inneber (vanlegvis) ein transformasjon av dei originale variablane til nye variablar med gjennomsnitt/forventning 0 og standardavvik 1 (Hodges og Lehmann 1964:161 og Hair Jr. mfl 2005:225). Variabelen  $X$  vert standardisert  $X^* = \frac{X - E(X)}{SD(X)}$ . Merk at den

avhengige variabelen også blir standardisert, og restleddet får gjennomsnitt 0 og fell bort. Koeffisientane vert no standard regresjonskoeffisientar, også kalla  $\beta$ (beta)-vektar eller  $\beta$ (beta)-koeffisientar.  $\beta$ -koeffisientane er rekna ut slik:  $\beta_j = \sum_{i=1}^n \frac{r_{iy}}{r_{ij}}$ , der  $r$ -ane er

korrelasjonskoeffisientar (Montgomery, Peck og Vining 2006:106-107 og Cooley og Lohnes 1962:34). Midtbø (2007:102) skriv:

Betakoeffisienten uttrykker den gjennomsnittlige endringa i standardavviket til den avhengige variabelen som følge at én enhets endring i standardavviket til forklaringsvariabelen når andre forklaringsvariabler holdes konstant.

Cooley og Lohnes (1962:32) skriv: “These coefficients have the technical name standard partial regression coefficients”, og: “For a given set of predictors, comparison of the absolute values of the beta weights indicates the relative contributions of the corresponding variables to the prediction of the criterion”.

Begge slags regresjonskoeffisientar har arbeidd inn i seg uttrykk for samvariasjon med andre uavhengige variablar, og er difor partielle (Montgomery, Peck og Vining 2006:108). I si forklaring av partielle deriverte skriv Selmer (1961:231): ”Vi må betrakte  $z$  (innsett her: den avhengige variable) som ein funksjon av én variabel av gangen, og spør etter den momentane veksthastighet av  $z$  når for eksempel bare  $x$  varierer og  $y$  holdes konstant (innsett her:  $x$  og  $y$  uavhengige variable), men for øvrig vilkårlig”. Montgomery, Peck og Vining (2006:84) skriv: ”... a **partial or marginal** test because the regression coefficient  $\beta_j$  (innsett her: ustandardisert regresjonskoeffisient) depends on all of the other variables  $x_i$  ( $i \neq j$ ) that are in the model. Thus, this is a test of the **contribution** on  $x_j$  **given the other regressors in the model**”.

I Hair Jr. mfl (2005:175) står om standardiserte  $\beta$ -koeffisientar at dei ”enable the researcher to compare directly the relative effect of each independent variable on the dependent variable”, men vidare (:226):

Although the beta coefficients represent an objective measure of importance that can be directly compared, two cautions must be observed in their use:

- First, they should be used as a *guide to the relative importance of individual independent variables only when collinearity is minimal*. ... collinearity can distort the contributions of any independent variable even if beta coefficients are used.
- Second, the beta values can be *interpreted only in the context of the other variables in the equation*.

In summary, beta coefficients should be used only as a guide to the relative importance of the independent variables included in the equation, and only for those variables with minimal multicollinearity.

## Referansar

Berg, Erling og Jørn Rattsø 2005: *Evaluering av kommuneoverføringer som regionalpolitisk virkemiddel*, SØF-rapport 02/05, side 43-44.

Cooley, William W. og Paul R. Lohnes 1962: *Multivariate Procedures for the Behavioral Sciences*, John Wiley & Sons, Inc.

Hair Jr., Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson og Ronald. L. Tatham 2005: *Multivariate Data Analysis*, 6. utg., Pearson Prencice Hall.

Hodges, J. L. Jr. og E. L. Lehmann 1964: *Basic Concepts of Probability and Statistics*, Holden-Day, Inc.

Johannesen, Arbjørn, Line Kristoffersen og Per Arne Tufta 2004: *Forskningsmetode for økonomisk-administrative fag*, Abstrakt forlag.

Midtbø, Tor 2007: *Regresjonsanalyse for samfunnsvitere*, Universitetsforlaget.

Montgomery, Douglas C., Elisabeth A. Peck og G. Geoffrey Vining 2006: *Introduction to linear regression analysis*, 4. utg. Wiley.

Selmer, Ernst 1961: *Differensial og integralregning*, Universitetsforlaget.

Sverdrup, Erling 1964: *Lov og tilfeldighet I*, Universitetsforlaget.

Tiebout, C. M. 1956: A pure theory of local expenditure, *Journal of Political Economy*, 64:416-424.

Wenstøp, Fred og Espen Bagøie: *Innføring i SPSS 11.0, Del IV, Multippel regresjon*, [http://web.bi.no/itavd/itdoc.nsf/bf25ab0f47ba5dd785256499006b15a4/5497dee80d0ed7e9c1256e520037b158/\\$FILE/Innf%C3%B8ring%20i%20SPSS%2011.0%20Del%204.doc](http://web.bi.no/itavd/itdoc.nsf/bf25ab0f47ba5dd785256499006b15a4/5497dee80d0ed7e9c1256e520037b158/$FILE/Innf%C3%B8ring%20i%20SPSS%2011.0%20Del%204.doc)