

Multipel OLS ekstra eksempel

Mads Hove

9/19/2022

OLS-regression kræver, at man holder tungen lige i munden når man foretager tolkning. Jeg lægger derfor en ekstra tolkning fra min hånd op her, som eksempel på, hvordan man kan gribe det an.

Eksemplet er fra en undersøgelse jeg har lavet om sammenhængen mellem politikeres kampagneindsats på Facebook (målt ved kr. brugt på Facebook-annoncer) og valgresultat (målt ved andel personlige stemmer i partiet i storkredsen). Synes I det er spændende og/eller vil læse mere ligger eksamensbesvarelsen på GitHub: https://github.com/madshove-thy/4V-eksamen/blob/main/4V_eksamensbesvarelse_Simon_Mads.pdf

Hypotesen lyder: *Større kampagneindsats fra kandidaten på Facebook øger kandidatens valgresultat*. Testen er foretaget ved multipel OLS regression og opsummeret i nedenstående tabel. Tallene i parentes angiver standardfejl, mens de andre er ustandardiserede koefficienter.

Inden du læser videre kan du jo passende selv aflæse tabellen og sige tolkningen højt for dig selv. Husk at læs noten under tabellen før du tolker (her finder du information om skaleringen). Vær også opmærksom på, at koefficienterne er ustandardiseret, at standardfejl står i parentes og signifikansniveau er udtrykt ved stjernerne. Find tabellen her:

Tabel 3: Testresultater hypotese 1

| | Valgresultat | |
|-------------------------|---------------------|---------------------|
| | Model I | Model II |
| | (1) | (2) |
| Kampagneindsats | 0.012*** (0.003) | 0.009* (0.004) |
| Partileder | | 0.156 (0.122) |
| Minister | | 0.045 (0.055) |
| Antal år i FT | | -0.0002 (0.003) |
| Konstant | 0.243*** (0.023) | 0.240*** (0.032) |
| N | 123 | 123 |
| R ² | 0.134 | 0.154 |
| Adjusted R ² | 0.127 | 0.125 |
| Residual Std. Error | 0.222 (df = 121) | 0.222 (df = 118) |
| F Statistic | 18.690*** | 5.376*** |

Note: *p < .05; **p < .01; ***p < .001. Den afhængige variabel er intervalskaleret fra 0 til 1, hvor 1 angiver at politikerens modtager alle personlige stemmer i partiet i storkredsen. Kampagneindsats er intervalskaleret og angiver stigning i 10.000 kr. brugt på Facebook-annoncer. Estimer er fra ustandardiserede OLS regressioner.

Tabellen indeholder model I og II, hvor model I er den bivariate sammenhæng mellem kampagneindsats og valgresultat, mens model II kontrollerer for, hvorvidt politikerens partileder (skaleret 0/1), minister (skaleret 0/1) samt hvor mange år politikerens har siddet i Folketinget (antal år). I modellerne vil højere koefficienter for kampagneindsats samt signifikans indikere, at kampagneindsats korrelerer med et bedre valgresultat, hvorfor vi vil finde opbakning til hypotesen, hvis koefficienten på kampagneindsats er positiv samt statistisk signifikant.

Model I viser, hvordan kampagneindsats er positivt korreleret med valgresultat, hvor en forøgelse i kampagneindsats på 10.000 kr. øger andelen af personlige stemmer inden for partiet i storkredsen med 1,2 procentpoint. Model II viser, hvordan denne sammenhæng svækkes en smule når vi holder hvorvidt politikerens partileder, minister og erfaring konstant, hvor sammenhængen er på 0,9 procentpoint. Begge koefficienter er ligeledes statistisk signifikante på et 95 procents konfidensinterval.

Læg mærke til, hvordan denne tolkning kun er mulig, fordi vi i noten kan se, hvordan den uafhængige og afhængige variabel er skaleret (hvilket man nok også vil skrive i brødteksten inden man præsenterer tabellen). Således ved vi, at ét trin op på kampagneindsats (X) er lig et øget forbrug på 10.000 kr., og at valgresultat (Y) er intervalskaleret mellem 0 og 1, hvorfor koefficienten for kampagneindsats på 0,012 er det samme som en stigning på 1,2 procentpoint i valgresultat.

Da jeg i det her tilfælde er interesseret i sammenhængen mellem kampagneindsats og valgresultat vil jeg ikke bruge lang tid på at kommentere på kontrolvariablene - man vil trods alt tidligere (i operationalisering) have forklaret, hvorfor de indgår i regressionen. Jeg vil dog ofre en kort paragraf til at sige, at hvad angår kontrolvariablene så fejler de alle i at opnå statistisk signifikans, da p-værdien ikke når under 0,05.

Hvad angår modellens forklaringskraft kan vi se, at model I kan forklare 12,7 procent af variationen i valgresultat, mens model II kan forklare 12,5 procent, hvorfor inddragelse af kontrolvariable ikke øger forklaringskraften af politikerens valgresultat ift. en situation, hvor vi bivariat ser på kampagneindsats og valgresultat. F-testen viser desuden at modellen som samlet hele er signifikant.

Husk også at gør tolkningen substantiel. Er en sammenhæng ml. kampagneindsats og valgresultat på 0,9 procentpoint stor, moderat eller lille? I det her tilfælde vil der være flere meningsfulde veje til den substantielle tolkning. Eksempelvis kunne man opstille et (ikke urealistisk) scenarie, hvor en politiker anvender 100.000 kr. mere på Facebook-annoncer end en anden politiker (NB: fra samme parti). Hvis modellen (og sammenhængen) er rigtig, vil dette betyde, at en politiker relativt kan øge sit valgresultat med 9 procentpoint af alle personlige stemmer i partiet i storkredsen, hvilket må siges at være en meget kraftig effekt.

En anden måde at lave en substantiel tolkning på er ved at sige, hvor mange stemmer 0,9 procentpoint af de personlige stemmer i partiet i storkredsen er udtryk for. Dette vil naturligvis variere for parti til parti og fra storkreds til storkreds, men i en situation, hvor mange politikere vælges med en margin der hedder 200-1000 stemmer, da vil 10.000 kr. Facebook-annoncer for nogle politikere være forskellen på at blive valgt eller ikke at blive valgt. Og da det for den enkelte politiker jo er hele målet med at stille op til valg, så vil det også vidne substantielt om en meget betydelig effekt.

Nu håber jeg nogen af jer sidder tilbage og tænker: “lyder det ikke næsten for godt til at være sandt” eller “kunne man ikke forestille sig risiko for endogenitet/simultanitet?”. For jo, virkeligheden er helt sikkert ikke som modellen estimerer - for jeg kan fx ikke udelukke, at årsagen til sammenhængen faktisk kunne være, at de politikere, som bruger mange kr. på Facebook-annoncer generelt er mere populære - eller at det kun er politikere med gode valgresultater, der faktisk har råd til at køre mange Facebook annoncer (omvendt kausalitet). Den udfordring samler vi op på, når vi skal snakke om målefejl.