

KRAV TIL SEMESTEROPPGÅVE I SVSOS3003: ANVENDT STATISTISK DATAANALYSE I SAMFUNNSVITENSKAP

A. FORMÅL

Semesteroppgåva skal vere eit sjølvstendig arbeid der ein demonstrerer korleis multipl regressjon kan nyttast til å analysere eit samfunnsvitskapeleg problem. Oppgåva skal ha eit opplegg omlag som ein tidsskriftartikkel, men bør gå lenger i å dokumentere data og analysar, gjerne ved hjelp av vedlegg. Det vert gitt karakter på semesteroppgåva. Den tel 60% i den endelege karakteren på kurset

C. FRAMLEGG TIL FRAMGANGSMÅTE

Det skal utarbeidast ei problemstilling med bakgrunn i sosiologisk eller statsvitskapeleg teori som bakgrunn for eit studium av kva som er årsak til variasjonen i ein høveleg valt avhengig variabel. Den avhengige variabelen og forklaringsvariablane kan anten hentast frå eit datasett som er klargjort og stilt til rådvelde for kurset (sjå eigen oversikt) eller dei kan hentast frå egne data - t.d. samla inn til bruk i ei hovudoppgåve - Data må likevel ha ein kvalitet som gjer at dei ulike analysekrava som vert stilt til oppgåva kan stettast. Bruk av egne data skal derfor godkjennast. Den avhengige variabelen skal anten ha målenivå intervallskala eller nominalskala. Dersom den har nominalskala nivå skal logit (logistisk) regressjon nyttast.

B. FORMALITETAR

Tittelside

Tittelsida skal minimum innehalde studentnummer og tittel som indikerer avhengig variabel.

Forord

Ikkje alle datasett er laga av SSB, men dei fleste. Ein føresetnad for å nytte **data frå SSB sine granskingar** er at følgjande vert teke med i ein fotnote eller i eit forord:

"(En del av) de data som er benyttet i denne publikasjonen er hentet fraundersøkelsen (årstall). Data i anonymisert form er stilt til disposisjon gjennom Norsk samfunnsvitenskapelig datatjeneste (NSD). Innsamling og tilrettelegging av data ble opprinnelig utført av Statistisk Sentralbyrå. Hverken Statistisk Sentralbyrå eller NSD er ansvarlige for analysen av dataene eller de tolkninger som er gjort her."

For kommune- eller fylkesdata eller andre typer data tilpassar ein formularet.

Litteraturreferansar

Referansar bør følgje ein akseptert standard.

Innlevering

Frist for innlevering er for våren 15 mai og for hausten 30 november. For sommarkurset sjå kursoversikt. Oppgåva skal leverast på instituttkontoret i 2 eksemplar.

Innbinding

er ikkje nødvendig, men den skal stiftast saman. Om de ønskjer å binde inn oppgåva skal det nyttast innbinding i A4 format.

D KRAVSPESIFIKASJONAR

Med utgangspunkt i den avhengige variabelen skal det gjerast ei kort teoretisk drøfting av moglege kausalsamband. Dette fører fram til ei modellformulering med framlegg til operasjonalisering av moglege årsaksvariablar henta innan datasettet. Dersom manglande opplysningar på ein eller fleire variablar fører til at eitt eller fleire case ikkje kan nyttast i

analysen, skal seleksjonsproblemet drøftast. Ved hjelp av regresjonsanalyse skal kausalmodellen estimerast og resultatata drøftast.

I konkretiseringa av ein første estimerbar regresjonsmodell skal følgjande element vere med:

- a) Med utgangspunkt i deskriptiv statistikk for variablane som skal inkluderast i modellen, skal fordelinga deira beskrivast og mogelege transformasjonar vurderast. Transformasjonar skal takast i bruk dersom dette kan forbetre analysen substansielt (dvs. det er teoretiske grunnar til å tru at det marginale sambandet mellom forklaringsvariabel og avhengig variabel er kurvelineært, jfr. pkt d) eller dersom det kan gjere testprosedyrane meir truverdige (residualen kjem nærmare opp mot normalfordelinga)
- b) Modellen skal innehalde minst ein kategorisk forklaringsvariabel med meir enn to kategoriar (MÅLENIVÅ: nominalsкала).
- c) Det skal gjennomførast ei drøfting av mogelege interaksjonar og minst eitt interaksjonsledd skal testast.
- d) Mogelege kurvelineære samanhengar skal drøftast og minst ein kurvelineære samanheng skal testast.

Med utgangspunkt i den estimerte modellen skal følgjande drøftast

- e) I OLS-regresjon skal det testast for heteroskedastisitet.
- f) I OLS regresjon skal effekten av autokorrelasjon vurderast dersom autokorrelasjon er aktuelt.
- g) I OLS-regresjon skal fordelinga av feilledet vurderast.
- h) I logit regresjon skal diskrimineringsproblem vurderast.
- i) Multikollinearitetsproblem skal vurderast
- j) Effekten av utliggarar og innflytelsesrike case vurderast og eventuelt illustrerast
- k) Modellspefifikasjonen skal vurderast.
- l) I drøftinga av resultatata skal det gjerast bruk av betinga effektplott (conditional effect plot).

Avslutningsvis skal substansen i resultatata frå regresjonsanalysen drøftast i høve til problemstillinga ein tok utgangspunkt i.

LEGG MERKE TIL AT

Punkta som er nemnt ovanfor kan ikkje brukast mekanisk som disposisjon for arbeidet. Det viktigase i arbeidet er modellspefifikasjonen. Diagnoseverktøyet er viktig for å forbetre spefifikasjonen (t.d. interaksjonsledd, kurveelement eller variabeltransformasjonar) og klargjere tolkingsproblem (t.d. fordelinga av restleddet, effekt på avhengig variabel når samanhengen er kurvelineær).