

Eksamen på Økonomistudiet vinter 2015-16

Multivariat analyse og Kategoriserede Data

Valgfag/kandidatfag

Fra 11/12 kl 12.00 til 30/12 før midnat

Besvarelsen med standardforside (uploades separat på Absalon) uploades som én pdf.fil (inkl. bilag) navngivet udelukkende med eksamensnummeret på Absalon, f.eks. 1234.pdf eller 55.pdf

Dette eksamenssæt består af 3 sider inklusive denne forside

Der er fokus på at undgå tilfælde af eksamenssnyd

I tilfælde af formodet eksamenssnyd, der bemærkes af fagenes eksamensadministration, af eksamenstilsynet eller af faglæreren, foretager studielederen en foreløbig undersøgelse af sagen. Dette foregår ved indhentning af udtalelse fra faglæreren, evt. fra eksamenstilsynet, og ved samtale med den studerende. Hvis studielederen finder formodningen om snyd bestyrket, indberetter han forholdet til rektor.

Den studerende skal under studiet og eksamenerne efterleve reglerne om videnskabelig redelighed. Videnskabelig uredelighed foreligger, når der ved forfalskning, plagiering, fortielse eller på anden måde vildledes om den pågældendes egne indsats eller resultater, eller når en anden studerende bistås hermed. Eksempelvis betragtes manglende kildeangivelser i skriftlige opgaver som fortielser. Forsøg på at snyde behandles på samme måde som gennemførte snyderier.

Rektor har følgende sanktionsmuligheder:

- Tildeling af advarsel
- Bortvisning fra eksamen
- Bortvisning fra universitetet for en begrænset periode eller permanent.

Det Samfundsvidenskabelige Fakultet
Studie- og eksamenskontoret
Oktober 2006

Multivariat analyse og Kategoriserede Data Eksamensopgave 2015

Husk et langt samlet pdf dokument, som indeholder et langt samlet SAS-program (som let kan afvikles!) som bilag sidst i pdf-filen. Eksamensopgaven skal besvares individuelt. Der anvendes plagieringskontrol.

Husk desuden alle formaningerne om at kalde biblioteket MULTKAT (store eller små bogstaver i datasæt-, fil- og biblioteksnavne er uden betydning) og om ikke at mellemlagre data på harddisken etc. Da lærer og censor selv kan afvikle SAS-programmet, er appendikser med SAS output helt overflødige.

Opgaverne i begge halvdele af kurset omhandler data fra 7. runde spørgeskemaundersøgelsen European Social Survey, der er indsamlet i en lang række lande i forsommeren 2014.

I får datasæt svarende til hver jeres kombination af tre regioner - en i Danmark, en i Tyskland og en i Sverige. Data er rensat, så fx observationer med manglende værdier er fjernet og relevante omregnede variable er tilføjet. Fordelingen udføres ud fra eksamensnummeret.

Opgave i Kategoriserede Data

Data er udleveret som rådata, dvs. at der ikke skal anvendes vægtvariable i en weight statement i Proc Catmod. Datasættene hedder `ess_kat_##`, hvor `##` er jeres eksamensnummer.

De variable er

`region_land`, hvori regionerne er defineret som

DK_DK01 Hovedstaden
DK_DK02 Sjælland
DK_DK03 Syddanmark
DK_DK04 Midtjylland
DE_DE1 Baden-Württemberg
DE_DE2 Bayern
DE_DE7 Hessen
DE_DE9 Niedersachsen
SE_SE110 Stockholms län
SE_SE224 Skåne län
SE_SE232 Västra Götalands län

`binge` (binge betyder drikke meget alkohol på en gang)

`drikker` (drikke alkohol i almindelighed)

`ryger`, `koen` (1=Mand, 2=Kvinde), `groent` (hvor ofte spises grøntsager)

For at få en oversigt, kan man køre:

```
proc freq data=multkat.ess_kat_##;  
tables region_land binge drikker ryger koen groent;  
run;
```

a) Der er en signifikant tredjeordens vekselvirkning mellem variablene `region_land`, `binge` og `drikker` i modellen med kun disse tre variable. Hvad betyder denne signifikante tredjeordens vekselvirkning?

b) Reducer modellen med alle de nævnte variable. Fortolk så meget som muligt ud fra uafhængigheder og betingede uafhængigheder. Tænk over at der er forskel på baggrundsvariable som `køn` (`koen`) og geografisk område (`region_land`) og de afledte variable om livsstil.

Opgave i Multivariat Analyse

Til opgaven anvendes korrelationsmatricen for 19 variable fra de samme tre geografiske områder i Danmark, Tyskland og Sverige. Da det udleverede datasæt er en korrelationsmatrix, skal I **IKKE** pille ved selve datasættet, men blot specificere det ved `data=multkat.ess_mult_##` i hvert procedurekald. Ellers opfattes datasættet som et rådatasæt med kun 22 observationer og ikke som en korrelationsmatrix estimeret ud fra mange hundrede observationer. Igen her betegner `##` jeres eksamensnummer.

En foreløbig faktoranalyse med alle variable udføres ved nedenstående procedurekald.

```
proc factor data=ess_mult_## method=ml rotate=promax;
var FCLCNTR FLTDPR FLTEEFF FLTLNL FLTSD
    ipadvnt ipbhprp ipcrtiv ipeqopt ipfrule ipgdtim iphlppl
    iplylfr ipmodst iprspot ipshabt ipstrgv ipsuces ipudrst;
run;
```

Ud fra outputtet det ses, hvad de variable omhandler. Alle variable er kodet, så en lav værdi svarer til "meget" mens en høj værdi svarer til "lidt". I praksis peger skalaerne for alle variable altså samme vej, så alle korrelationer er positive.

a) Gennemfør faktoranalyser først for "feel" variablene med et "F" som første bogstav i variabelnavnet og dernæst for de mange "important" variable med "IP" som de første bogstaver i variabelnavnet. Måske skal der bruges "Heywood". Husk fortolkninger!

b) Opstil en konfirmativ faktormodel for modellerne for hhv "feel" variablene og "important" variablene hver for sig. Da visse variable sikkert ikke kan beskrives med få faktorer, skal I vurdere, om antallet af variable skal reduceres for at opnå en brugbar model for de resterende variable. Husk fortolkninger!

c) Opstil en samlet stimodel for alle (på nær visse udeladte, jf spm b) udleverede variable på en gang. Husk fortolkninger!