

Eksamen på Økonomistudiet vinter 2013-14

## **Multivariat analyse og Kategoriserede Data**

Valgfag/kandidatfag

Fra 9/12 kl 15.00 til 29/12 før midnat)

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

Dette eksamenssæt består af 6 sider incl 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 fortelser. 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 2013

Husk et langt samlet pdf dokument, som indeholder et langt samlet SAS-program (som let kan afvikles!) som bilag. 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.

Eksamensopgaven skal besvares individuelt. Tolv siders almindelig tekst er det absolutte maksimum, idet dog tabeller og figurer ikke tæller med i dette sideloft. Da lærer og censor selv kan afvikle SAS-programmet, er appendikser med SAS output helt overflødige.

Opgaverne i de to halvdele af kurset omhandler begge data fra spørgeskemaundersøgelser i USA indsamlet af Pew Research Center, der finansieres af <http://www.pewtrusts.org/>. Dette forskningscenter stiller mange interessante datamaterialer til rådighed for akademisk arbejde. Til opgaven her er data bearbejdet og gjort repræsentative ved en fornuftig vægtning af de oprindelige data ud fra de oplyste vægte i data. Det er gjort på en sådan måde, at I hver får unikke personlige datasæt med det oprindelige antal observationer. Konklusioner ud fra de datasæt, der anvendes i denne eksamensopgave, svarer altså til, hvad man ville få ud fra de oprindelige komplette datasæt, men mindre forskelle kan optræde. I kan altså ikke bruge disse eksamensdata i andre sammenhænge, så hvis emnerne interesserer jer, skal I gå til de originale datamaterialer fra Pew's hjemmesider.

Benyt datasættene svarende til jeres eksamensnummer: Eksamensnummer 7 skal eksempelvis bruge Dating\_7 og Middleclass\_7, mens eksamensnummer 21 skal bruge Dating\_21 og Middleclass\_21.

## Opgave i Kategoriserede Data

Data er udleveret som rådata, dvs. at der ikke skal anvendes vægtvariable. Datasættene hedder Dating\_##, hvor ## er eksamensnummeret. Der er måske enkelte nuller i den udleverede tabel, men de er uden betydning for de relevante vekselvirkninger. Så nuller er IKKE en del af opgaven!

Til programmet skal anvendes et format, der kaldes ved linien

```
options fmtsearch=(multkat.dating_64);
```

eller

```
options fmtsearch=(multkat.dating_32);
```

der skal ind i begyndelsen af alle SAS sessioner.

Dette format, dvs for et 64 bits maskine filen

```
dating_64.sas7bcats
```

og for en 32 bits maskine filen

dating\_32.sas7bcat

skal downloades og lægges i den mappe, der tilordnes som biblioteket multkat. Variablene er kodet med 1 for ja og 1 for nej, men formatet bevirker, at der skrives Yes og No.

Hvis formatet ikke virker, kan I indlede hvert sas-program med

```
option nofmterr;
```

Denne option slår brugen af formatet fra, og da de to værdier er 1=Yes og 2=No, er formatet ikke så vigtigt.

Data omhandler synspunkter på emnet "online dating" og søgen efter en partner.

**a) Der er mindst én signifikant (eller næsten signifikant) tredjeordens vekselvirkning mellem de fire variable date9a date9b date9c date9d i modellen med kun disse fire variable. Hvad betyder de signifikante tredjeordens vekselvirkninger?**

Betragt nu modellen med alle variable svarende til procedurekaldet nedenfor. Notationen @3 betyder, at der blandt de utroligt mange vekselvirkninger i den lange liste med lodrette streger højst medtages tredjeordens virkninger.

```
proc catmod data=dating_##;  
model date9a*date9b*date9c*date9d*sex*mar4*date1a=_response_/ pred=freq;  
loglin date9a|date9b|date9c|date9d|sex|mar4|date1a @3  
response out=ny1;  
run;  
quit;
```

**b) Reducer denne model yderligere og fortolk så meget som muligt ud fra uafhængigheder og betingede uafhængigheder. Mange enkeltvist signifikante vekselvirkninger kan med lidt held testes væk sammen med de insignifikante. Brug evt. et lavere signifikansniveau fx. 0.1% , da antal observationer, n = 2252, er stort.**

Variable	Label
date1a	DATE1a. Have YOU, personally, ever used an online dating site such as Match.com, eHarmony, or OK Cupid?
date9a	DATE9a. Many people have different opinions about online dating. Do you agree or disagree with this statement: Online dating is a good way to meet people
date9b	DATE9b. Many people have different opinions about online dating. Do you agree or disagree with this statement: Online dating allows people to find a better match for themselves because they can get to know a lot more people

Variable	Label
date9c	DATE9c. Many people have different opinions about online dating. Do you agree or disagree with this statement: People who use online dating sites are desperate
date9d	DATE9d. Many people have different opinions about online dating. Do you agree or disagree with this statement: Online dating keeps people from settling down because they always have options for people to date
mar4	MAR4. Would you say you are currently looking for a romantic partner, or that you are not currently looking for a partner?
sex	SEX. Respondent's sex

## Opgave i Multivariat Analyse

Til opgaven stilles korrelationsmatricen for 20 variable i en stikprøve på 2508 individer til rådighed i datasættet Middleclass\_##, hvor ## er eksamensnummeret. Da det udleverede datasæt er en korrelationsmatrix, skal I **IKKE** pille ved selve datasættet, men blot specificere det ved `data=multkat.middleclass_##` i hvert procedurekald. Ellers opfattes datasættet som et rådatasæt med kun 23 observationer og ikke som en korrelationsmatrix estimeret ud fra 2508 observationer. De anvendte labels er ret lange, så husk at labels kan slås til og fra med

```
options nolabel;
options label;
```

De variable er listet i tabellen nedenfor. De omhandler amerikaneres opfattelse af finanskrisens årsager, klasseinddelingen i USA, samt til hvordan de to præsidentkandidater i 2012, Barack Obama og Mitt Romney, vil behandle de forskellige samfundsklasser. Data er indsamlet i august 2012, efter det stod klart at Barack Obamas republikanske udfordrer netop blev Mitt Romney.

De variable er inddelt i visse grupper af parallelle spørgsmål, men spørgsmålene kan også grupperes efter lower, middle og upper class, og der kan også grupperes efter politisk tilhørsforhold. Man kan forestille sig, at almene holdninger til samfundsforhold påvirker, hvorledes respondenterne opfatter de to præsidentkandidaters synspunkter.

En faktoranalyse med alle 20 variable og 8 faktorer udføres ved nedenstående procedurekald.

```
proc factor data=middleclass_## method=ml heywood nfactor=8
r=promax flag=0.5 fuzz=0.2;
var q16a q16aa q16ab q16ac q16b q16c
q21aa q21ab q21ac q21ad q21ae q21af q21ag
q29a q29b q29c q29d
q8ah q8bh q8ch;
run;
```

Alle variablene er kodet, så en høj værdi betyder, at de deler det nævnte synspunkt i lav grad, mens en lav værdi betyder, at de er meget enige.

a) Gennemfør faktoranalyser først for visse relevante delgrupper af variable og dernæst for hele dette datamateriale. Husk fortolkninger!

b) Opstil en konfirmativ faktormodel for næsten alle de udleverede variable, da enkelte variable måske falder uden for rammen. Husk fortolkninger!

c) Opstil en stimodel for næsten alle de udleverede variable, da enkelte variable måske falder uden for rammen. Husk fortolkninger!

Variable	Label
q16a	Q.16a (Next,) Thinking about what Barack Obama might do for some groups if he were elected for a second term, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE], or not? / Middle class people
q16aa	Q.16aa (Next,) Thinking about what Mitt Romney might do for some groups if he were elected president, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE ITEMS IN SAME ORDER AS Q.16], or not? / Middle class people
q16ab	Q.16ab Thinking about what Mitt Romney might do for some groups if he were elected president, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE ITEMS IN SAME ORDER AS Q.16], or not? / Wealthy people
q16ac	Q.16ac Thinking about what Mitt Romney might do for some groups if he were elected president, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE ITEMS IN SAME ORDER AS Q.16], or not? / Poor people
q16b	Q.16b Thinking about what Barack Obama might do for some groups if he were elected for a second term, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE], or not? / Wealthy people
q16c	Q.16c Thinking about what Barack Obama might do for some groups if he were elected for a second term, do you think his policies would help [INSERT FIRST ITEM; RANDOMIZE], or not? / Poor people
q21aa	Q.21aa How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / The Obama administration
q21ab	Q.21ab How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / The Bush administration
q21ac	Q.21ac How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / Foreign competition
q21ad	Q.21ad How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / Banks and financial institutions
q21ae	Q.21ae How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame

<b>Variable</b>	<b>Label</b>
	[ITEM]: a lot, a little or not at all? / Large corporations
q21af	Q.21af How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / Congress
q21ag	Q.21ag How much do you blame: [INSERT FIRST ITEM, READ AND RANDOMIZE] for the difficulties the middle class has faced in the past ten years. Do you blame [ITEM]: a lot, a little or not at all? / Middle Class people themselves
q29a	Q.29a Compared to the average person, do you think that rich people generally are more likely or less likely to be / honest
q29b	Q.29b Compared to the average person, do you think that rich people generally are more likely or less likely to be / intelligent
q29c	Q.29c Compared to the average person, do you think that rich people generally are more likely or less likely to be / greedy
q29d	Q.29d Compared to the average person, do you think that rich people generally are more likely or less likely to be / hard-working
q8ah	Q.8a As I read off some different groups, please tell me if you think they are paying their fair share in federal taxes, paying too much or paying too little? / lower-income people
q8bh	Q.8b As I read off some different groups, please tell me if you think they are paying their fair share in federal taxes, paying too much or paying too little? / Middle-income people
q8ch	Q.8c As I read off some different groups, please tell me if you think they are paying their fair share in federal taxes, paying too much or paying too little? / Upper-income people