

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

- Eksamen i: ST 301 — Statistiske metoder og anvendelser.
- Eksamensdag: Torsdag 20. juni 1996.
- Tid for eksamen: 09.00 – 14.00.
- Oppgavesettet er på 5 sider.
- Vedlegg: Tabeller av  $t$ -,  $F$ -,  $\chi^2$ - og normalfordeling.
- Tillatte hjelpemidler: Alle trykne og skrevne samt kalkulator.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

### Oppgave 1.

Vi vil i denne oppgaven studere høyt blodtrykk for menn over 40 år. Totalt 433 menn har blitt undersøkt, og i tillegg til om personen har høyt blodtrykk (målt som en enten/eller variabel), har en også registrert om personen røyker, om personen har unormal fedme og om personen snorker. Følgende data ble observert (fra Norton og Dunn(1995), "Snoring as a risk factor for disease: an epidemiological survey", Br. Med. J):

Røyker	Fedme	Snorker	Antall	Høyt blodtrykk
0	0	0	60	5
1	0	0	17	2
0	1	0	8	1
1	1	0	2	0
0	0	1	187	35
1	0	1	85	13
0	1	1	51	15
1	1	1	23	8

(Fortsettes side 2.)

For de tre første kolonnene indikerer 0 nei, mens 1 indikerer ja. Siste kolonne gir antall som har høyt blodtrykk av det totale antallet (4. kolonne).

- a) Som tittelen på artikkelen til Norton og Dunn indikerer, er en her interessert i å undersøke om snorking har betydning for høyt blodtrykk. Hvorfor er det fornuftig å også registrere om personene røyker og/eller har unormal fedme?
- b) Hva er den naturlige responsvariabelen å se på her? Hva slags type regresjonsmetode vil du så bruke?

En regresjonsanalyse basert på de 3 forklaringsvariablene gir følgende resultat:

	Value	Std. Error
(Intercept)	-2.3777	0.3801
røyking	-0.0678	0.2781
fedme	0.6953	0.2851
snorking	0.8719	0.3975

Devians = 1.6184

- c) Forklar hva deviansen betyr. Hvor mange frihetsgrader har vi her?
- d) Utifra tabellen ovenfor, forklar hvorfor det virker naturlig å kutte ut variabelen *røyking*.

En regresjonsanalyse basert på bare *fedme* og *snorking* gir

	Value	Std. Error
(Intercept)	-2.3920	0.3757
fedme	0.6954	0.2851
snorking	0.8654	0.3966

Devians = 1.6781

- e) Gjør en test basert på devianser for å undersøke om røyking har betydning for høyt blodtrykk.

Tabellen nedenfor gir resultatene når foruten *fedme* og *snorking* også interaksjonsleddet mellom disse variablene er tatt med (interaksjonsleddet er definert som produktet av *fedme* og *snorking* kolonnene).

(Fortsettes side 3.)

	Value	Std. Error
(Intercept)	-2.3026	0.3964
fedme	0.1054	1.1222
snorking	0.7621	0.4271
fedme:snorking	0.6387	1.1609

Devians = 1.3439

- f) Beregn sannsynligheten for høyt blodtrykk for en mann for alle kombinasjonene (røyker, fedme), (røyker, ikke fedme), (ikke-røyker, fedme) og (ikke-røyker, ikke fedme).
- g) Av de modeller vi har sett på, hvilken modell vil du foretrekke? Tolk koeffisienten tilhørende snorking for denne modellen.

## Oppgave 2.

Vi vil i denne oppgaven se på et forsøk der en prosess er kjørt med to ulike katalysatorer, kalt A og B. Videre kjøres prosessen med to ulike temperaturnivåer og to konsentrasjonsnivåer for katalysatoren. Eksperimentet er en del av en større undersøkelse i beplantning, og responsen er avkastning. Dataene er som følger:

	Temp.	Kons.	Kata.	Avka.
1	160	20	A	60
2	180	20	A	72
3	160	40	A	54
4	180	40	A	68
5	160	20	B	52
6	180	20	B	83
7	160	40	B	45
8	180	40	B	80

Her står Temp. for temperatur, Kons. for konsentrasjon, Kata. for katalysator-type og Avka. for avkastning.

- a) Forsøket er *designt*. Forklar hva vi mener med det. Hva slag type design er det snakk om her?
- b) Det ble bestemt å kjøre de 8 eksperimentene i en tilfeldig rekkefølge. Forklar hvorfor dette er en fornuftig fremgangsmåte.

(Fortsettes side 4.)

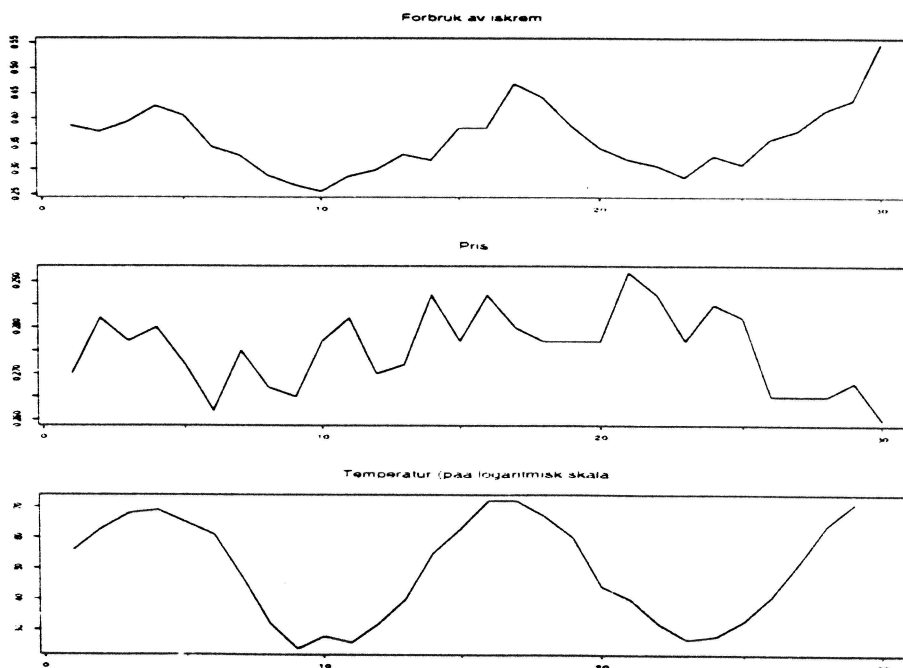
Vi vil betrakte en modell hvor kun 1. ordens leddene (hovedeffektene) er tatt med. Nedenfor er deler av en anova tabell basert på dataene ovenfor:

	d.f.	Sum of squares	Mean sum of squares	F-statistic	P-value
Temp.		1058.0			
Kons.		50.0			
Kata.		4.5			
Residuals		205.0			
Total	7	1317.5			

- c) Forklar hvordan man finner de manglende tall i tabellen og fyll ut. Kommenter resultatene.
- d) Diskuter om det her har noen hensikt å innføre interaksjonsledd for disse data.

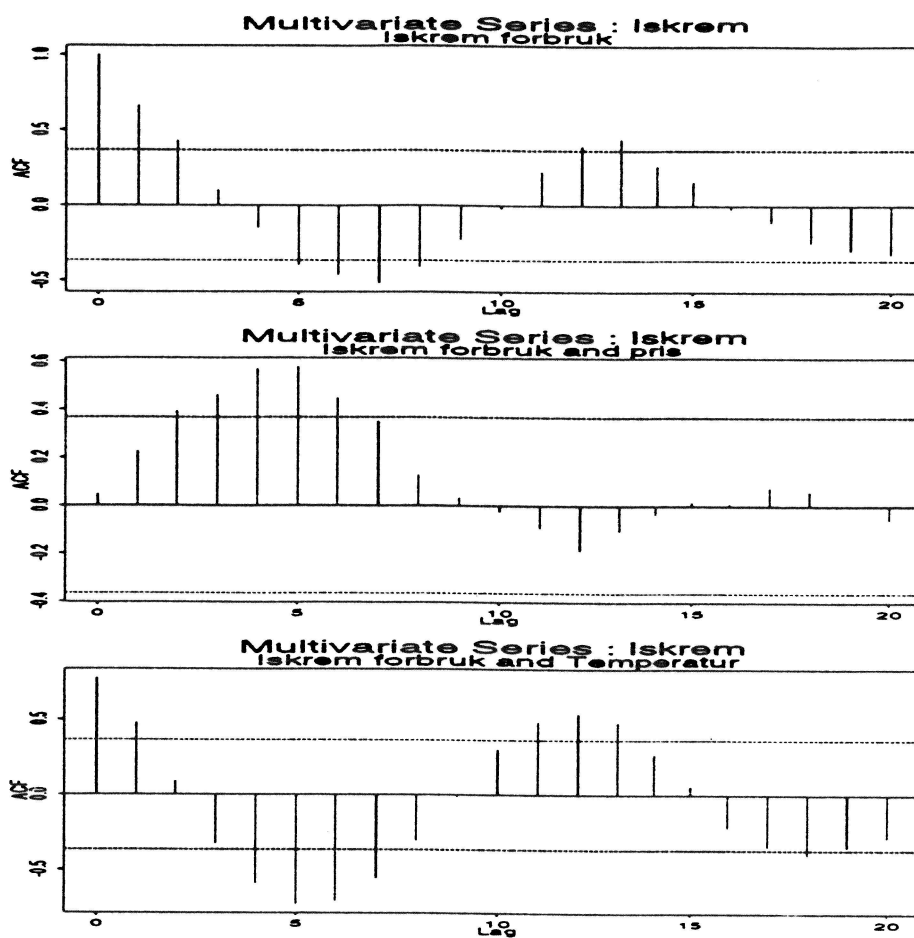
### Oppgave 3.

En undersøkelse ble i perioden 1951-1953 utført for å se på hvordan forbruket av iskrem avhenger av pris og temperatur. Disse tre variablene ble målt over 30 fire-ukersperioder og er plottet nedenfor.



Videre har vi nedenfor plott av autokorrelasjonsfunksjonen for iskremforbruket (øverst) samt krysskorrelasjonsfunksjonene mellom iskremforbruk og pris (midten) og mellom iskremforbruk og temperatur (nederst).

(Fortsettes side 5.)



Basert på dette, gi en beskrivelse av hvordan du vil gå frem for å analysere slike data. Gi gjerne et forslag til en modell.

SLUTT