

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i STK4900/9900 — Statistiske metoder og anvendelser

Eksamensdag: 11. juni 2008

Tid for eksamen: 09.00 – 12.00

Oppgavesettet er på 2 sider.

Vedlegg: Tabell over t-fordeling, F-fordeling og kjikvadratfordeling.

Tillatte hjelpemidler: Alle trykte og skrevne samt godkjent lommeregner

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

Oppgave 1

Som en del av en lønnsomhetsstudie for melkebruk ble samlet arbeidsinnsats i årsverk registrert for 50 bruk, og disse brukene ble kategorisert etter hvor mange melkekyr de hadde. De kategoriene som ble brukt, var:

Kategori 1	Kategori 2	Kategori 3	Kategori 4
< 10 kyr	10 – 20 kyr	21 – 31 kyr	> 31 kyr

a) På grunnlag av dette ble det gjort en enveis variansanalyse med arbeidsinnsats som avhengig variabel og de 4 kategoriene som klasser. Variansanalysetabellen ble:

.	Df	SS	MS	F	P
Kategori	3	5.505	1.835	4.89	?
Rest	46	17.267	0.3754		
Total	49	22.772			

Forklar hvordan frihetsgradene er funnet, hvordan middelkvadratsummene er funnet og hvordan F er funnet. Hva kan du si om P-verdien ut fra disse tallene?

b) Anta at en vil analysere det samme datamaterialet ved en enkel, lineær regresjonsmodell, der x -variabelen er 1 for kategori 1, 2 for kategori 2, 3 for kategori 3 og 4 for kategori 4. Sett opp modellen, og forklar hva de forskjellige leddene betyr. Hvor mange frihetsgrader får resten i denne analysen? Forklar prinsippet for testing av en hypotese om at stigningsforholdet er null.

c) Hva er forskjellen mellom analysen i a) og analysen i b)? Se på en situasjon der analysen i a) gir signifikans, mens regresjonsanalysen i b) ikke gir signifikans. Hvordan vil du tolke dette resultatet? Tegn gjerne en figur.

(Fortsettes på side 2.)

Oppgave 2

Det var meldt opp $n = 32$ studenter på STK4900/9900 i år. Frammøtet på de praktiske øvelsene i løpet av kurset var som følger:

w = uke											
d = dag	1	2	3	4	5		6	7	8	9	10
dw = dag innen uke	1	2	3	4	5		1	2	3	4	5
antall studenter	30	24	23	21	23		22	21	17	22	16

a) Vi er interessert i å studere tidsutviklingen av dette frammøtet. Forklar hvorfor det er rimelig å bruke en logistisk regresjonsmodell.

b) En slik modell med d som regresjonsvariabel ga følgende delvise utskrift:

Coefficients :	Estimate	Std.error	z - value
(Intercept)	1.65944	0.28845	5.753
d	-0.15447	0.04379	-3.527

Null deviance: 21.5932 on 9 degrees of freedom

Residual deviance: 8.5831 on 8 degrees of freedom.

Bruk disse tallene til å teste på to måter om frammøtet på øvelsen avhenger av tid. Gi en konklusjon.

c) En multipel logistisk regresjon med dw og w som regresjonsvariable ga:

Coefficients	Estimate	Std.error
(Intercept)	2.47268	0.50039
dw	-0.21029	0.08783
w	-0.68704	0.24789

Null deviance: 21.5932 on 9 degrees of freedom

Residual deviance: 8.0406 on 7 degrees of freedom.

Bruk utskriftstabellen til å teste om det er noen effekt av uke.

d) Modellen under b) kan sees på som nøstet i modellen under c). Hvorfor? Bruk dette og de oppgitte deviansverdiene til å teste om det er noen effekt av uke. Hvorfor gir dette en konklusjon som er forskjellig fra den du fant i c)? Etter din mening, er det en netto effekt av uke i disse dataene? I hvilken forstand? Begrunn svaret.

SLUTT