

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i STK1000 — Innføring i anvendt statistikk

Eksamensdag: Tirsdag, 1. desember 2009

Tid for eksamen: 09.00 – 12.00

Oppgavesettet er på 4 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Lærebok: Moore & McCabe «Introduction to the practice of statistics», ordliste for bruk i STK1000, godkjent kalkulator

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

### Oppgave 1

Følgende oppgave er basert på artikkelen «The antihypertensive effects of fish oil» av Knapp & FitzGerald publisert i *New England Journal of Medicine* 1989.

Forskere rekrutterte 14 frivillige menn med høyt blodtrykk, og valgte tilfeldig 7 av dem til å gå på en diett basert på fiskeolje, mens de resterende 7 gikk på en diett som brukte helt vanlig matolje. Reduksjonen i diastolisk blodtrykk (målt i torr) er gitt i Tabell 1.

	Reduksjon i torr						
Diett med fiskeolje	8	12	10	14	2	0	0
Diett med vanlig olje	-6	0	1	2	-3	-4	2

Tabell 1: Reduksjon i blodtrykk

a) Hvorfor randomiserte forskerene de 14 frivillige mennene mellom de to gruppene?

Er individene et tilfeldig utvalg fra en bestemt befolkning (dvs, er det en SRS)? Hva slags konsekvenser har dette?

Finn gjennomsnitt og standardavvik for observert reduksjon til de to diettene og kommentér resultatet.

b) Beregn et 95% konfidensintervall for forventet reduksjon med fiskeolje og ett for forventet reduksjon med vanlig olje og gjør rede for antagelsene som ligger til grunn for beregningene.

Bruk disse to konfidensintervallene til å teste om forventet reduksjon av de to diettene er null mot tosidige alternativ.

c) Bruk en passende statistisk test for å se om gruppene har ulik forventet reduksjon i blodtrykk og gjør rede for antagelsen som ligger til grunn for denne testen.

(Fortsettes på side 2.)

## Oppgave 2

Følgende oppgave er basert på artikkelen «*Light intensity effects on meadowfoam Groth and Flowering*» av Seddigh & Jolliff publisert i *Crop Science* 1994.

*Limnanthes alba* er en liten blomsterplante som gror blant annet i fuktige enger, og har frøolje med særegne kjemiske egenskaper som gjør den interessant å dyrke industrielt. Frøoljen kommer fra blomstene på planten, og jo flere blomster hver plante har jo mer lønnsom er den å dyrke. Undersøkelsen vi skal se på prøver å finne det dyrkingsmiljøet som gir flest blomster per plante.

Jordbruksforskere plantet ti og ti planter i isolerte forsøksrom som var identiske – bortsett fra at man varierte lysintensiteten og når man startet med lysbehandling på plantene. En faktor som pleier å være viktig for antall blomster er tidspunkt man starter lysbehandling i forhold til *fotoperiodisk floral induksjon* (FFI), som er da man øker lysperioden fra 8 til 16 timer per dag for å få plantene til å blomstre. I forsøket startet man lysbehandling enten ved FFI eller 24 dager før FFI. Det ble planlagt to gjentak for hvert mulig lysregime. Man hadde seks forskjellige lysintensiteter, så forsøket brukte i alt 240 frø. Disse ble tilfeldig tilordnet til hver behandlingsgruppe.

Isteden for å analysere antall blomster i hver behandlingsgruppe, analyserte forskerne *gjennomsnittlig antall blomster* på gruppene av ti og ti planter. Resultatet av forsøket er oppgitt i Tabell 2.

a) Forklar hvorfor man kan anse hvert av gjennomsnittene i Tabell 2 som uavhengige tilfeldige variabler.

Hvorfor er fordelingen til gjennomsnittlig antall blomster per plante nærmere normalfordelingen enn fordelingen til antall blomster på hver plante?

Kommentér om det er rimelig å kunne fastsette de lysforholdene som *forårsaker* at plantene har flest blomster ved å gjennomføre dette forsøket.

b) Anta at standardavviket til antall blomster per plante i en bestemt gruppe er  $\sigma$ . Hva er standardavviket til gjennomsnittlig antall blomster per plante i denne gruppen?

Hvor mange blomster måtte man ha hatt per gruppe for å halvere standardavviket til gjennomsnittlig antall blomster per plante?

		Lysintensitet ( $\mu\text{mol}/\text{m}^2/\text{sec}$ )					
		150	300	450	600	750	900
Starttid	Ved FFI	62.3	55.3	49.6	39.4	31.3	36.8
		77.4	54.2	61.9	45.7	44.9	41.9
	24 dager før FFI	77.8	69.1	57.0	62.9	60.3	52.6
		75.6	78.0	71.1	52.2	45.6	44.4

Tabell 2: Gjennomsnittlig antall blomster per plante for 12 behandlinger

La  $Y_1, Y_2, \dots, Y_{24}$  være gjennomsnittlig antall blomster per plante for hver av de 24 plantegruppene. Vi skal se på disse som responsvariabler i en multiplert lineær regresjonsmodell med to prediktorvariabler. Vi lar  $X_{1,1}, X_{1,2}, \dots, X_{1,24}$  være lysintensiteten (målt i  $\mu\text{-mol}/\text{m}^2/\text{sekund}$ ) og

(Fortsettes på side 3.)

$X_{2,1}, X_{2,2}, \dots, X_{2,24}$  angir starttiden for lysbehandlingen, der

$$X_{2,i} = \begin{cases} 0, & \text{hvis forsøket startes ved FFI} \\ 1, & \text{hvis forsøket startes 24 dager før FFI} \end{cases} .$$

Under følger deler av utskriften Minitab gir for en regresjonstilpassing av datasettet.

Regression Analysis: Antall blomster versus Lysintensitet; Starttid

The regression equation is

Antall blomster = 71,3 - 0,0405 Lysintensitet + 12,2 Starttid

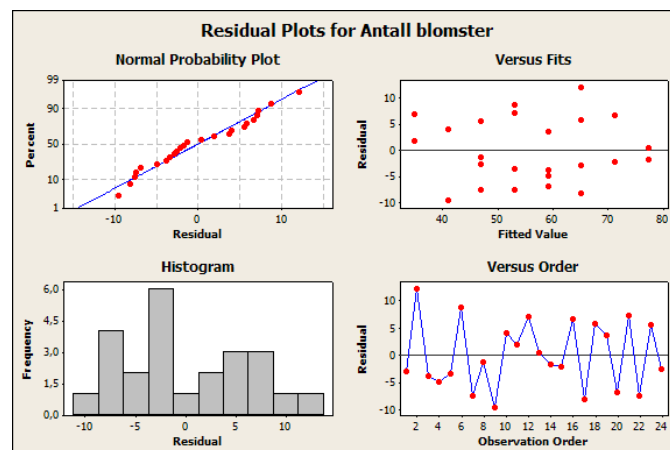
Predictor	Coef	SE Coef	T	P
Constant	71,306	3,274	21,78	0,000
Lysintensitet	-0,040471	0,005132	-7,89	0,000
Starttid	12,158	2,630	4,62	0,000

S = 6,44107    R-Sq = 79,9%    R-Sq(adj) = 78,0%

c) Formulér regresjonsmodellen Minitab har tilpasset og skriv ned antagelsene som ligger til grunn for denne modellen.

Finn estimater av alle modellens fire parametre.

Figur 1 viser Minitabs «four in one» residualplott, mens Figur 2 viser et kryssplott for alle observasjonene, hvor man har skilt hver tidskategori med et eget plottesymbol og har tegnet en regresjonslinjetilpassing for hver tidskategori.

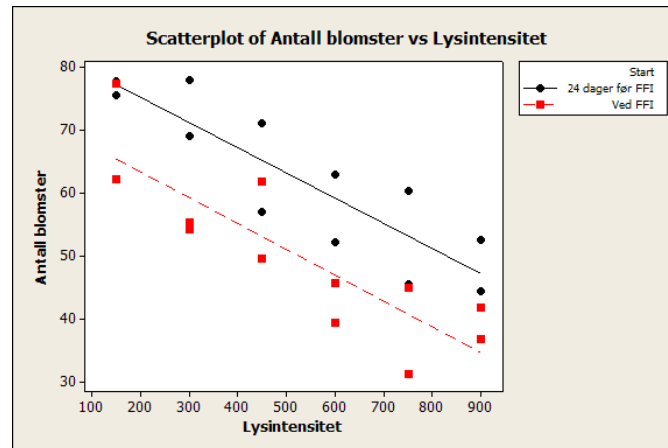


Figur 1: Regresjonsdiagnostikkplott

d) Diskutér modellantagelsene på bakgrunn av Figur 1 og Figur 2, og Minitabutskriften over.

Forklar hvordan en  $R^2$ -verdi på 79.9% kan tolkes.

(Fortsettes på side 4.)



Figur 2: Kryssplott for observasjonene med regresjonslinje for hver tidskategori

e) Hvilke hypotesetester ligger til grunn for  $P$ -verdiene i regresjonsutskriften?

Kommentér hva det vil si at alle  $P$ -verdiene i regresjonsanalysen er tilnærmet null.

Hvor mange frihetsgrader brukes i beregningene av disse  $P$ -verdiene?

Oppgi et 95% konfidensintervall for regresjonskoeffisienten til  $X_{2,i}$ .

f) Forklar med utgangspunkt i den tilpassede regresjonsmodellen hvordan antall blomster på *Limnanthes alba* forandres med de forskjellige lysregimene.

Hva slags lysregime virker rimelig å prøve hvis man ønsket å utføre et nytt forsøk for å finne enda bedre produksjonsmåter?

SLUTT