

Institutt:	IKBM	
Eksamen i:	Stat100	Statistikk
Tid:	Torsdag 10. desember 2009	09.00-12.30
Emneansvarlig:	Trygve Almøy:	

Tillatte hjelpemiddel: C3: alle typer kalkulatorer, alle andre hjelpemiddel

Opgaveteksten er på: 4 sider

Alle deloppgaver teller likt.

Alle steder der du utfører statistiske tester, skal du skrive opp både nullhypotesen og den alternative hypotesen, samtidig må du oppgi signifikansnivået som du tester på (dersom det ikke er spesifisert i oppgaveteksten).

Oppgave 1

Et idrettslag ville prøve ut forskjellige intensive treningsmetoder for langdistanse, den ene fokuserte på intervalltrening (korte intervaller med maksimal innsats), den andre mer på langkjøring.

De gikk fram på følgende måte: 10 løpere ble valgt ut, alle løp et testløp over 10 km.

Deretter ble 5 av løperne trukket ut til intervall, mens de 5 andre drev langkjøring. Etter 8 uker løp samtlige en ny 10 km (kalt etter trening).

Løper.	1	2	3	4	5	6	7	8	9	10
Treningsmetode	L	I	I	L	L	I	I	I	L	L
Tid testløp	46:40	45:30	42:30	41:40	44:50	41:50	39:50	39:10	42:50	39:50
Tid etter trening	46:00	45:00	42:10	41:10	44:40	41:00	39:10	38:10	41:50	40:00

Tabell 1: Her står I for intervall, mens L står for langkjøring. Tid er gitt i minutter og sekunder slik at for eksempel 46:40 betyr 46 minutter og 40 sekunder

- Kan du påvise at langkjøringsopplegget hadde positiv effekt? Svar ved hypotesetest. Sett opp modellen du bruker.
- Dersom vi tester om intervalltrening har positiv effekt, får vi en P-verdi på 0,002. Vis hvordan denne er regnet ut.
- Kan du påvise at intervalltreningen og trening ved langkjøring har forskjellig effekt? Svar ved et konfidensintervall. Sett opp modellen du bruker

I et større forsøk ble 100 personer trukket tilfeldig ut. Alle løp en 5 kilometer. Deretter ble det for hver person ved loddtrekning avgjort om vedkommende skulle prøve intervall eller langkjøring. Det ble trukket ut 50 til intervall og 40 av disse hadde framgang, mens det ble trukket ut 50 langkjøring, her hadde 35 framgang.

- Lag et 95 % konfidensintervall for andelen som har framgang ved intervalltrening. Forklar treneren hva dette konfidensintervallet betyr. Bruk konfidensintervallet til å teste om intervalltrening har effekt

- e) Kan du på dette grunnlaget påstå at det er avhengighet mellom treningsmetode og framgang?

Oppgave 2

Limnanthes Alba er en plante med frøolje som har kjemiske egenskaper som gjør den svært verdifull. Oljen kommer fra blomsten og flere blomster pr plante jo mer lønnsom blir den. Problemet er at vi kjenner ikke hvilke dyrkningsforhold som gir størst antall blomster. Forskere prøvde å forklare blomsterantallet ved forskjellig lysintensitet. (målt i $\text{m mol/m}^2/\text{sekund}$) En plantet ut 12 planter der en hadde en plante i hvert sitt forsøksrom (der lysintensiteten ble kontrollert) og talte opp antall blomster pr. plante.

Lysintensitet	150	150	300	300	450	450	600	600	750	750	900	900
Antall blomster	62	77	55	54	50	61	39	45	31	44	36	41

La Y_i være antall blomster og x_i lysintensitet for plante i , der $i = 1, 2, \dots, 12$. Data ble analysert etter følgende modell: $Y_i = \alpha + \beta x_i + \epsilon_i$, der ϵ_i -ene er antatt å være uavhengige, normalfordelte med forventning 0 og standardavvik σ .

Etter denne oppgaven finner du noen figurer samt en Minitabutskrift.

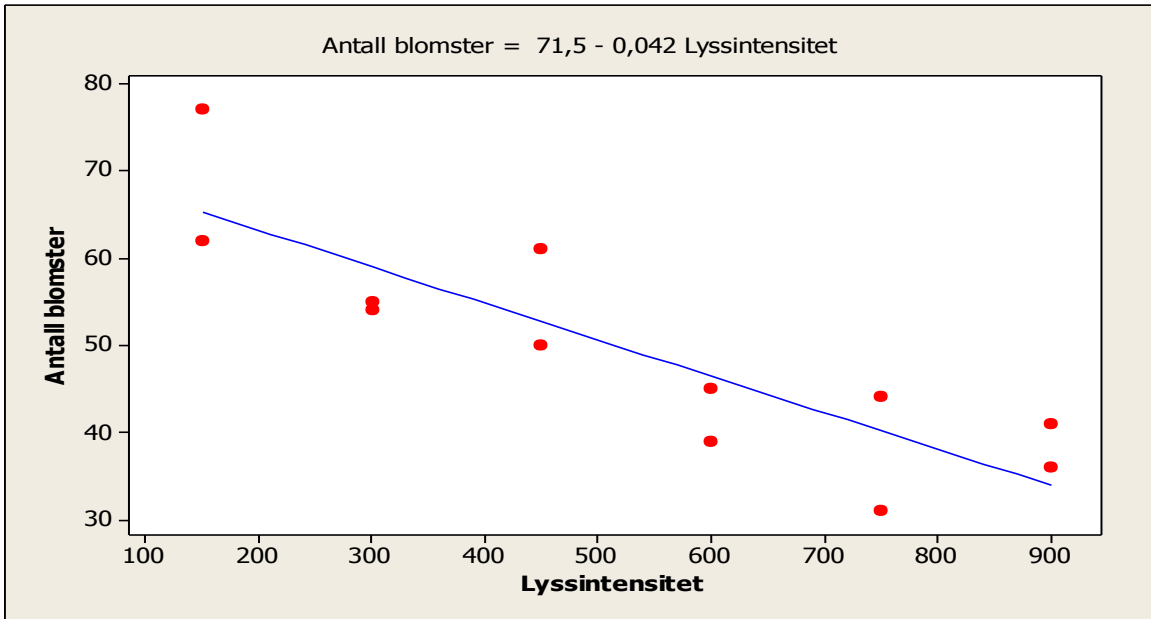
- Synes du at denne modellen passer til data?
Beregn residualet for de to første plantene i tabellen.
Forklar kort hvordan figur 2 kan brukes til modellsjekk.
- Gi en forklaring (tolkning) av hva de tre parametrene i modellen sier.
Estimer de tre ukjente parametrene.
- Finn R^2 og gi en konkret forklaring av hva denne betyr.
- Sett opp en test på om økt lysstyrke gir redusert antall blomster.
Hva blir konklusjonen på testen?
- Dersom en ønsker å ha en lysintensitet på 200, finn i så fall et 95 % konfidensintervall (KI) for forventet blomsterantall.
Et prediksjonsintervall (PI) for antallet blomster på en enkelt plante utsatt for denne lysintensiteten er (46,20; 80,14). Forklar forskjell på dette og konfidensintervallet.
Hva skjer med bredden til intervallene (KI og PI) når antallet observasjoner blir svært stort?

Regression Analysis: Antall blomster versus Lyssintensitet

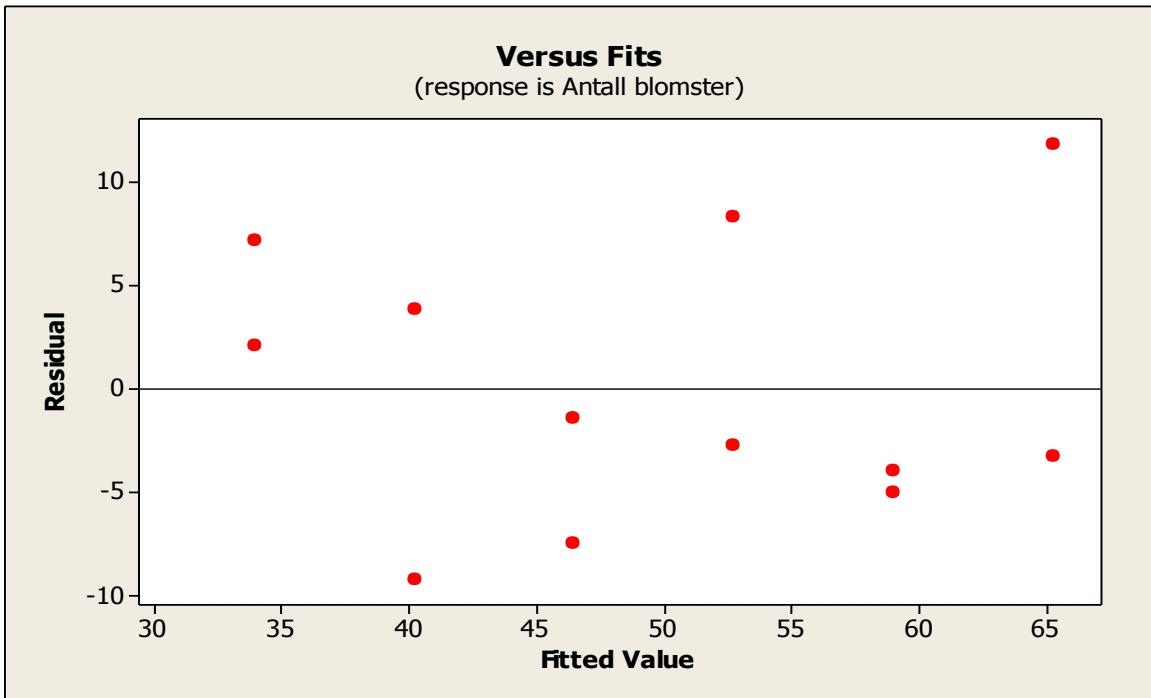
Predictor	Coef	SE Coef
Constant	71,5	4,543
Lyssintensitet	-0,042	0,0078

Analysis of Variance

Source	SS
Regression	1376,6
Residual Error	476,3



Figur 1: Antall blomster mot lysintensitet



Figur 2: Residualer mot tilpassede verdier

Oppgave 3

Mango er en frukt som er en utmerket kilde til C-vitamin, men det er et problem at innholdet av dette kan variere mellom hvert tre. For å undersøke C-vitamininnholdet nærmere ble det gjort en undersøkelse i Tanzania, der ble 4 steder (hvert med 10 trær) valgt ut. (alle målinger er i milligram pr. 100 gram). En statistisk analyse ga blant annet:

Source	SS (sum of squares)
Sted	70,266
Error	210,66

Sted	Mean
1	28,73
2	29,20
3	27,01
4	25,90

- a) Sett opp modellen du vil bruke for å analysere dette forsøket. Forklar **alle** symboler som inngår i denne modellen. Estimer de 5 parametre i modellen.
- b) Skriv opp hypotese og alternativ for å teste om det er forskjell mellom de 4 stedene i Tanzania. Utfør testen på 5 % signifikansnivå. Trekk konklusjon.
- c) De to første stedene i Tanzania er høyland, de to siste er lavland. Gir denne undersøkelsen som resultat at du er sikker på at det er forskjell mellom høyland og lavland?

Senere ble også 2 steder i Kenya valgt ut (kalt sted A og sted B). På hvert sted ble 10 mango trær tilfeldig trukket ut også her ble det målt C-vitamininnhold (milligram pr. 100 gram) i en frukt fra hvert tre. Dette ga gjennomsnitt og standardavvik for sted A på 25 og 2,4, mens det ga et gjennomsnitt og standardavvik på sted B på 28 og 2,6 (alt målt i milligram pr. 100 gram).

- d) Er det høyere C- vitamin innhold i Tanzania enn i Kenya, svar ved å utføre en test på 1 % signifikansnivå. Hvilke antagelser må du gjøre for å besvare dette?