



Fagleg kontakt under eksamen:
Mette Langaas (988 47 649)

NYNORSK

EKSAMEN I TMA4255 ANVENDT STATISTIKK

Fredag 25. mai 2012
Tid: 9:00–13:00

Tal på studiepoeng: 7.5.

Tillatte hjelpemiddel: Alle trykte og handskrivne hjelpemiddel. Spesiell kalkulator.

Sensurfrist: 18. juni 2012.

Eksamensresultata blir annonserte frå <http://studweb.ntnu.no/>.

Merk deg følgjande:

- Signifikansnivå 5% skal brukast om ikke anna er spesifisert.
- Alle svar må grunngjevast.

Oppg ve 1 Vitamin C

I ein medisinsk studie p  marsvin vart det nytta to ulike kjelder til vitamin C-inntak. Desse var appelsinjuice (tilskot 1) og syntetisk askorbinsyre (tilskot 2). Responsm let som vart brukt var lengda til odontoblastceller i fortennene til marsvina. Forskarane hadde som m l   finne ut om det var ein forskjell i dette responsm let for dei to tilskota.

$X_1, X_2, \dots, X_{n_1}$ angir odontoblastlengdene til eit tilfeldig utval av n_1 marsvin som fekk tilskot 1 og $Y_1, Y_2, \dots, Y_{n_2}$ angir odontoblastlengdene til eit tilfeldig utval av n_2 marsvin som fekk tilskot 2. Vi g r ut fr  at $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $E(Y_j) = \eta$ og $\text{Var}(Y_j) = \tau^2$ for $i = 1, 2, \dots, n_1$ og $j = 1, 2, \dots, n_2$, og at dei to tilfeldige utvalga er uavhengige.

Totalt fekk $n_1 = 10$ marsvin tilskot 1 og $n_2 = 10$ marsvin tilskot 2. Datasettet finn du (sortert) i tabellen under. Lengdene er i mikrometer (10^{-6} meter).

Tilskot	Observasjonar									
Tilskot 1: Appelsinjuice	8.2	9.4	9.6	9.7	10.0	14.5	15.2	16.1	17.6	21.5
Tilskot 2: Askorbinsyre	4.2	5.2	5.8	6.4	7.0	7.3	10.1	11.2	11.3	11.5

Deskriptive m l for datasettet er $\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = 13.18$, $\bar{y} = \frac{1}{10} \sum_{j=1}^{10} y_j = 8.00$, $s_x = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2} = 4.44$, og $s_y = \sqrt{\frac{1}{9} \sum_{j=1}^{10} (y_j - \bar{y})^2} = 2.77$.

- Skriv ned nullhypotesen og den alternative hypotesen som forskarane  nskjer   teste. Kva for f resetnader m  gjelde for at du skal kunne bruke ein to-utvals t -test? Er det fornuftig   anta at variansane σ^2 og τ^2 er like? Her kan du bruke signifikansniv  $\alpha = 0.02$ dersom du vel   utf re ein test for   unders ke om variansane er ulike. Utf r ein t -test. Vil du konkludere med at det er forskjell i lengda til odontoblastcellene i ein populasjon som bruker tilskot 1 og ein populasjon som bruker tilskot 2?
- Eit alternativ til ein to-utvals t -test er Wilcoxon rang-sum-test (ogs  kalt Mann-Whitney test). Kva for f resetnader byggjer Wilcoxon rang-sum-testen p ? N r b r du bruke Wilcoxon rang-sum-test i staden for to-utvals t -test? Utf r Wilcoxon rang-sum-testen for dataa i tabellen over. Vil du konkludere med at det er forskjell i lengda til odontoblastcellene i ein populasjon som bruker tilskot 1 og ein populasjon som bruker tilskot 2?

Oppg ve 2 Kjemisk produkt

Vi skal studere ein kjemisk prosess og lage ein statistisk modell som relaterer mengda kjemisk produkt til temperatur og trykk for den kjemiske prosessen. Vi bruker f lgjande notasjon.

- y , **produkt**. Mengda produkt fr  den kjemiske prosessen.
- x_1 , **temperatur**. Temperaturen som vart brukt under prosessen.
- x_2 , **trykk**. Trykket som vart brukt under prosessen.

Den kjemiske prosessen vart k yrd for $n = 21$ ulike kombinasjonar av temperatur og trykk, og mengda produkt vart m lt.

Ein multipel line r regresjonsmodell kan brukast til   analysere dataa. La (y_i, x_{1i}, x_{2i}) vere m lingane gjort ved i te k yring av prosessen.

$$\text{Modell A} \quad y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i$$

Her er ε_i u.i.f. $N(0, \sigma^2)$ for $i = 1, \dots, n$.

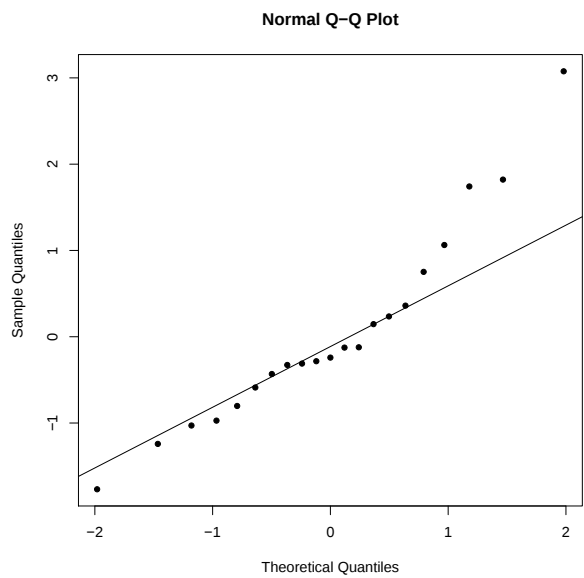
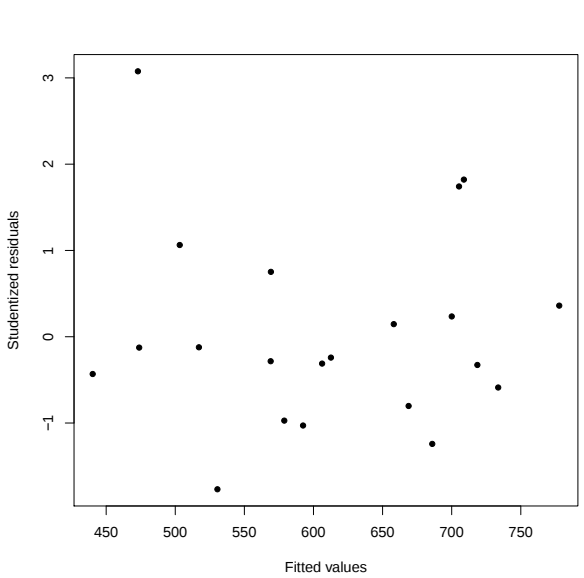
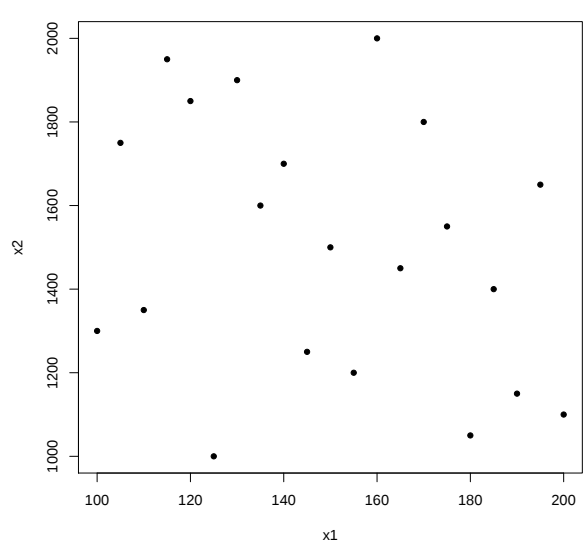
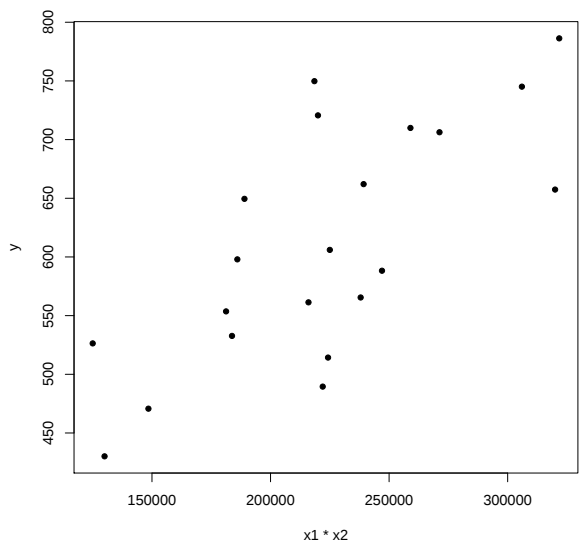
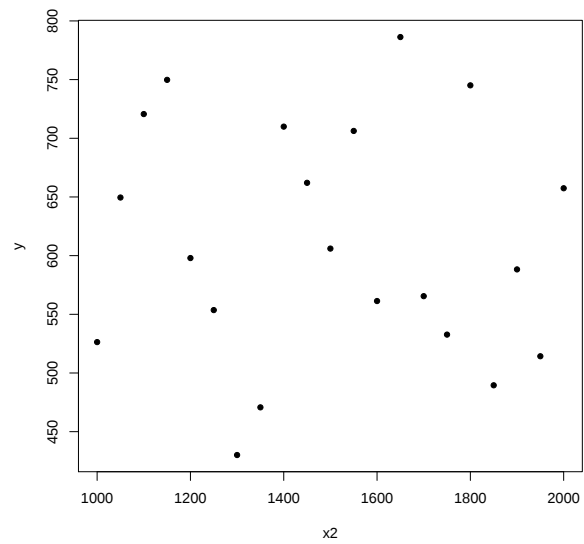
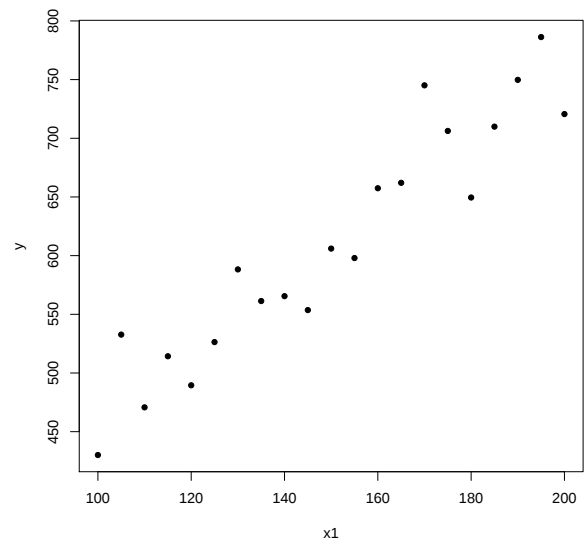
Spreiingsplott (scatter plots) finn du p  side 4 ( vre og midtre radene). Modell A vart tilpassa til dataa, og resultatet finn du p  side 5. Plott av studentiserte residual er p  side 4 (nedre rada).

a) Skriv ned den tilpassa regresjonsmodellen.

Kva er estimatet for σ^2 ?

Det er gitt ein p -verdi i rada for x_2 i resultata. Forklar med ord kva denne p -verdien betyr.

Basert p  plotta og regresjonsresultata, vil du seie at modell A er ein god modell for dataa? Du m  spesifisere kva for eigenskapar til plotta og regresjonsresultata du bruker for   komme fram til svaret ditt.



Regression analysis					
Predictor	Coef	SE Coef	T	P	
Constant	181.8	156.6	1.16	0.262	
x1	2.146	1.018	2.11	0.050	
x2	-0.0440	0.1043	-0.42	0.678	
x1*x2	0.0007774	0.0006942	1.12	0.278	
S = 27.2502 R-Sq = 93.8% R-Sq(adj) = 92.7%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	3	190830	63610	85.66	1.825e-10
Residual Error	17	12624	743		
Total	20	203454			

Gå ut frå at konstantleddet, β_0 , er med i regresjonsmodellen. Vi skal no sjå på ulike regresjonsmodellar, der vi har med ulike kombinasjonar av forklaringsvariablane x_1 , x_2 og x_1x_2 . Resultat frå tilpasninga av modellane er presenterte i tabellen under. Kvar rad i tabellen svarer til ein modell. Talet på forklaringsvariablar inkludert i kvar modell finn du i kolonnen med namn **par** (ikkje medrekna konstantleddet). I kolonnen med namn MSE finn du gjennomsnittlig kvadratavvik (mean squared error) til regresjonen.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	par	MSE	R2	R2.adj	C_p
1	151	3.067			1	1179	89.0	88.4	13.2
2	641		-0.020		1	10667	0.4	-4.9	255.9
3	315			0.001	1	5093	52.4	49.9	113.3
4	15	3.263	0.071		2	753	93.3	92.6	3.3
5	117	2.563		0.0005	2	709	93.7	93.0	2.2
6	505		-0.258	0.002	2	885	92.2	91.3	?
7	182	2.146	-0.044	0.001	3	743	93.8	92.7	4.0

- b) I kolonnen heilt til høgre i tabellen finn du Mallows C_p . Eitt av tala i denne kolonnen er erstatta med eit spørjeteikn. Skriv ned definisjonen av C_p og rekn ut den manglande C_p -verdien.
 Forklar korleis du kan bruke C_p til å samanlikne dei ulike regresjonsmodellane.
 Kva for ein av desse 7 regresjonsmodellane meiner du er den «beste» for dette datasettet?

Oppg ve 3 Behandling av tennisolboge

Uttrykket *tennisolboge* blir brukt om ein betennelsestilstand i olbogen, som er  rsak til smerter. Tilstanden er vanleg hos personar som driv med racketsport, men alle aktivitetar som involverer gjenteke vriding av handleddet (som ved   bruke eit skrujern) kan vere  rsak til denne tilstanden. Tilstanden kan  g skuldast konstant bruk av tastatur eller mus.

I ein randomisert klinisk studie  nskte ein   samanlikne tre ulike behandlingsm tar for tennisolboge, A: fysioterapi, B: injeksjon med steroid og C: vent-og-sj  (pasientane i vent-og-sj -gruppa fekk ikkje noka behandling, men vart bedne om   belaste olbogen s  lite som mogleg).

Vi skal sj  p  korttidseffekt av behandling ved   sj  p  m lingar etter 6 veker. Alle deltakarane i studien hadde berre ein arm med tilstanden tennisolboge. Fleire responsm l vart registrerte i studien, og vi vil i det f lgjande sj  p  to ulike responsm l.

- a) Behandlinga vart rekna som ein suksess dersom pasienten vurderte at ho/han var vorte mykje betre eller heilt bra (ein standardisert betringsskala vart brukt). Talet p  suksessar og fiaskoar etter 6 vekers behandling finn du i tabellen under.

Behandling	Fiasko	Suksess	Totalt
A (fysioterapi)	22	41	63
B (injeksjon)	14	51	65
C (vent-og-sj�)	44	16	60
Totalt	80	108	188

Vil vil unders ke om suksessraten er ulik for behandlingane. Skriv ned nullhypotesen og den alternative hypotesen og utf r ein hypotesetest p  grunnlag av informasjonen i tabellen over. For   forenkle utrekningsbyrda kan det opplysast at χ^2 -testobservatoren blir 36.6, og du treng berre   vise korleis du reknar ut eitt av dei 6 ledda i summen.

Kva for f resetnader m  gjelde for at du skal kunne bruke denne testen?

Kva er konklusjonen din basert p  denne testen?

Vi skal no sj  p  eit responsm l som blir kalla *smertefri gripestyrke*. Dette vart m lt med eit digitalt dynamometer og normalisert til gripestyrken i den friske armen. Ein smertefri gripestyrke p  100 betyr at den affiserte armen og den ikkje-affiserte armen fungerer like godt. Deskriptive m l for kvar av behandlingsgruppene er presenterte i tabellen p  toppen av side 7.

Behandling	Utvalgsstorleik	Gjennomsnitt	Standardavvik
A (fysioterapi)	63	70.2	25.4
B (injeksjon)	65	83.6	22.9
C (vent-og-sjå)	60	51.8	23.0
Totalt	188	69.0	

- b) Vi ønskjer å undersøke om den forventta smertefrie gripestyrken varierer mellom behandlingsgruppene. Skriv ned nullhypotesen og den alternative hypotesen og utfør ein hypotesetest basert på dei deskriptive måla i tabellen over.

Kva for føresetnader må gjelde for at du skal kunne bruke denne testen?

Kva er konklusjonen din basert på denne testen?

- c) Vi ønskjer å parvist samanlikne den forventta smertefrie gripestyrken for alle behandlingsgruppene. Gå no ut frå at utvalgsstorleikane for alle behandlingane er $n_A = n_B = n_C = 63$ og at gjennomsnitt og empirisk standardavvik for kvar behandlingsgruppe er som gitt i tabellen over.

Utfør samanlikningane ved å konstruere simultane konfidensintervall for den forventta differansen i smertefri gripestyrke mellom alle par av behandlingar ved å bruke Tukeys metode. Bruk eit overordna konfidensnivå på 95% for alle samanlikningane.

Kva for føresetnader må gjelde?

Finn det felles individuelle konfidensnivået som er brukt for kvar av samanlikningane.

- d) La μ_A angi den forventta smertefrie gripestyrken i ein populasjon der fysioterapi blir brukt for behandling av tennisolboge, og la μ_C angi den forventta smertefrie gripestyrken i ein populasjon der vent-og-sjå blir brukt. Definer den relative differansen mellom desse to forventningsverdiane som

$$\gamma = \frac{\mu_A - \mu_C}{\mu_C}.$$

Dette kan fortolkast som den forventta relative auken i smertefri gripestyrke når fysioterapi blir brukt i staden for vent-og-sjå-behandling. Foreslå ein estimator, $\hat{\gamma}$, for γ basert på to uavhengige utval av storleik n_A og n_C frå fysioterapi og vent-og-sjå-behandlingsgruppene.

Bruk tilnærma metodar til å finne forventningsverdi og varians til denne estimatoren, $E(\hat{\gamma})$ og $\text{Var}(\hat{\gamma})$.

Bruk utvalgsstoleikane, gjennomsnitta og dei empiriske standardavvik for fysioterapi og vent-og-sjå-behandlingsgruppene som er presenterte i tabellen på toppen av denne sida til å rekne ut numerisk verdi for $\hat{\gamma}$ og estimere numeriske verdier for $E(\hat{\gamma})$ og $\text{Var}(\hat{\gamma})$.

Dersom du fekk vite at dei to utvala ikkje var uavhengige, korleis ville det påverke approksimasjonen du har funne for $E(\hat{\gamma})$ og $\text{Var}(\hat{\gamma})$?