

Institutt for matematiske fag

Eksamensoppgåve i **TMA4255 Anvendt statistikk**

Fagleg kontakt under eksamen: Anna Marie Holand

Tlf: 951 38 038

Eksamensdato: August 2014

Eksamenstid (frå–til):

Hjelpemiddelkode/Tillatne hjelpemiddel: Alle trykte og handskrivne hjelpemiddel. Spesiell kalkulator.

Annan informasjon:

- I utskrifta frå MINITAB er komma brukt som desimalskilleteikn.
- Signifikansnivå 5% skal brukast om ikkje anna er spesifisert.
- Alle svar må grunngjevast.

Målform/språk: nynorsk

Sidetal: 9

Sidetal vedlegg: 0

Kontrollert av:

Dato

Sign

Oppg ve 1 Reikibehandling av Fibromyalgi

Fibromyalgi er ein kronisk smertetilstand med ukjend  rsak, som p virkar 2 – 4% av befolkninga. Pasientar med fibromyalgi brukar ofte komplement rande og alternativ medisin, slik som Reiki. Reiki er en form for energimedisin, der healeren kanaliserer "universell livsenergi" gjennom lett ber ring.

Ein studie vart gjennomf rd for   utforske nytten av Reiki som behandling mot smerte. Hundre pasientar med fibromyalgi vart rekrutterte. For   tene som placebo behandling inkluderte studien ein skodespelar som gav behandling, i tillegg til ein Reikimeister.

Smerte vart m lt ved   bruke ein visuell analog skala (VAS) (0 = inga smerte, 10 = vondaste smerte nokonsinne).

(Kjelde: Nassim Assefi, M.D., Andy Bogart, M.S., Jack Goldberg, Ph.D., and Dedra Buchwald, M.D. (2008). Reiki for the Treatment of Fibromyalgia: A Randomized Controlled Trial. *Journal of Alternative and Complementary Medicine*, 14(9): 1115-1122.)

- a) Dei 100 pasientane vart randomiserte i 4 grupper, kvar p  storleik $n = 25$, der dei fekk behandling anten med direkte ber ring eller utan ber ring (fr  avstand) fra anten ein Reikimeister (referert til som Reiki behandling) eller ein skodespelar (referert til som placebo behandling). VAS poengsk r f r behandling vart m lt for kvar av dei $n = 25$ pasientane i kvar av dei 4 gruppene. La gruppe A betegne pasientar som f r behandling fr  ein Reikimeister med direkte ber ring, gruppe B betegne pasientar som f r behandling fr  ein Reikimeister utan ber ring (fr  avstand), gruppe C betegne pasientar som f r behandling fr  ein skodespelar med direkte ber ring, gruppe D betegne pasientar som f r behandling fr  ein skodespelar utan ber ring (fr  avstand).

Dei deskriptive m la for dei ulike gruppene i VAS poengsk r f r behandling er gjevne i Tabell 1.

Smertegruppe	Utvalgsstørrelse	Gjennomsnitt	Standardavvik
Gruppe A	25	6.3	2.2
Gruppe B	25	6.4	2.6
Gruppe C	25	6.8	2.1
Gruppe D	25	6.1	2.4
Totalt	100	6.4	

Tabell 1: Deskriptive mål for gruppene A, B, C og D i Reiki behandlings datasettet.

For å kunne samanlikne VAS poengskår etter behandling på tvers av gruppene, bør VAS poengskårene vere lik på tvers av gruppene før behandling. Kan vi frå denne studien konkludere med at poengskårene før behandling er ulike mellom gruppene, A, B, C og D?

Skriv ned nullhypotesen og den alternative hypotesen, utfør ein hypotesetest basert på dei deskriptive måla ovanfor. Bruk signifikansnivå $\alpha = 0.05$. Spesifiser kva antakingar du gjer og konklusjonen av testen.

- b) Forskerane var interesserte i å samanlikne den direkte berørings behandlingen, gjeve anten av ein Reikimeister eller ein skodespelar.

La X_i betegne VAS poengskår etter behandling av Reikimeisteren med direkte berøring, $i = 1, \dots, 25$ og Y_j betegne VAS poengskår etter behandling av skodespelaren med direkte berøring, $j = 1, \dots, 25$.

Anta at X_i og Y_j er normalfordelt, høvevis $X_i \sim N(\mu_X, \sigma^2)$, $i = 1, \dots, 25$ og $Y_j \sim N(\mu_Y, \sigma^2)$, $j = 1, \dots, 25$.

Deskriptive mål for dette datasettet er $\bar{d} = \bar{x} - \bar{y} = -0.4$ og $s_p^2 = \frac{1}{48}(\sum_{i=1}^{25}(x_i - \bar{x})^2 + \sum_{i=1}^{25}(y_i - \bar{y})^2) = 1.23^2$, kor $\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i$ og $\bar{y} = \frac{1}{25} \sum_{i=1}^{25} y_i$.

Basert på desse dataene, har vi grunn til å tru at Reiki behandling har ein større effekt på smerte, målt som VAS poengskår, enn placebo behandling gjeve av ein skodespelar? Skriv ned nullhypotesen og den alternative hypotesen, vel ein testobservator og utfør ein hypotesetest. Bruk signifikansnivå $\alpha = 0.05$.

Spesifiser kva antakingar du gjer.

Oppg ve 2 Smaken av cheddarost

N r cheddarosten mognast skjer det ulike kjemiske prosessar som avgjer smaken p  sluttproduktet. Konsentrasjonen av fleire kjemikalier i denne kjemiske prosessen er knytt til smaken.

I ein observasjonsstudie av smaken p  cheddarostar, vart eit utval p  $n = 30$ modna cheddarost analysert for kjemisk sammensetning og utsatt for smakstestar. Den samla poengsk ren for testen vart funnen ved   kombinere poenga fr  fleire personar i eit smakspanel. F lgjande variablar vart m lt for kvar cheddarost.

- y : Den subjektive smakstest poengsk ret, gjennomsnittleg smakstest poengsk r varierte fr  0.7 til 57.2.
- x_1 : Den naturlege logaritmen p  konsentrasjon av eddiksyre i cheddarosten.
- x_2 : Den naturlege logaritmen p  konsentrasjon av hydrogensulfid i cheddarosten.
- x_3 : Konsentrasjon av melkesyre i cheddarosten.

F rst vart tre separate enkle regresjonsmodeller brukt til   studere tilh vet mellom smak og kvar av variablane,

$$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i, \quad (1)$$

$$y_i = \beta_0 + \beta_2 x_{2i} + \epsilon_i, \quad (2)$$

$$y_i = \beta_0 + \beta_3 x_{3i} + \epsilon_i, \quad (3)$$

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

En MINITAB utskrift fr  statistiske analyse finst in Figur 1, 2 og 3.

Predictor	Coef	SE Coef	T	P
Constant	-61,50	24,85	-2,48	0,020
X1	15,648	4,496	3,48	0,002

S = 13,8212 R-Sq = 30,2% R-Sq(adj) = 27,7%

Figur 1: Utskrift fra statistiske analysar av cheddarostdataene for modellen in ligning (1).

Predictor	Coef	SE Coef	T	P
Constant	-9,787	5,958	-1,64	0,112
X2	5,7761	0,9458	6,11	0,000

S = 10,8334 R-Sq = 57,1% R-Sq(adj) = 55,6%

Figur 2: Utskrift fra statistiske analysar av cheddarostdataene for modellen i ligning (2).

Predictor	Coef	SE Coef	T	P
Constant	-29,86	10,58	-2,82	0,009
X3	37,720	7,186	5,25	0,000

S = 11,7450 R-Sq = 49,6% R-Sq(adj) = 47,8%

Figur 3: Utskrift fra statistiske analysar av cheddarostdataene for modellen i ligning (3).

- a) Frå Figur 1 ser vi at variabelen eddiksyre, X_1 , er signifikant ved eit 5% signifikansnivå. Kva kan du konkludere med fra dei gjevne p-verdiane i Figur 1, 2 og 3 om dei tre kjemikalienes påvirkad på smaken? Kommenter verdiane av R^2 i Figur 1, 2 og 3. Grunngi svaret.

Finn eit 90% konfidensintervall for β_1 .

Kva er den predikerte poengskåret for smak for ein eddiksyreverd på $x_1^0 = 7$?

- b) Finn eit passende estimat for σ , og beregn er 90% konfidensintervall for σ i regresjonsmodellen i ligning (1) (Hint: bruk at SSE/σ^2 er kjikvadrat-fordelt).

Korleis kan vi bruke dette konfidensintervallet for å teste nullhypotesen $H_0 : \sigma = 1$? Skriv ned den alternative hypotesen som vert brukt, konklusjonen på denne testen og signifikansnivået.

Vidare vart regresjon av smak på alle tre kjemikalier utført.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i \quad (4)$$

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

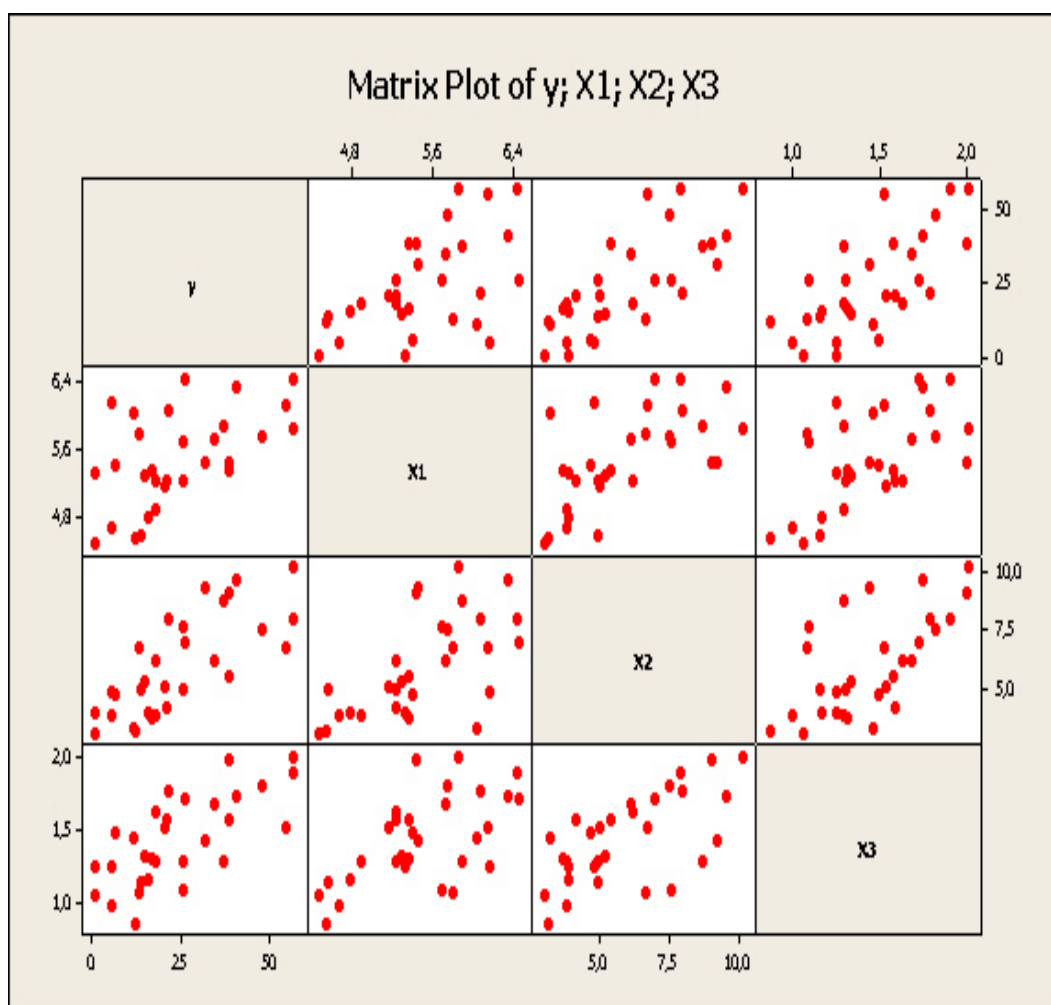
MINITAB utskriften frå statistiske analyse av denne tre-variabel modellen finst i Figur 4. Ein korrelasjonsmatrise og eit parvis spredningsplott finst i høvevis Figur 5 og Figur 6. Residualplott finst i Figur 7.

The regression equation is					
y = - 28,9 + 0,33 X1 + 3,91 X2 + 19,7 X3					
Predictor	Coef	SE Coef	T	P	
Constant	-28,88	19,74	-1,46	0,155	
X1	0,328	4,460	0,07	0,942	
X2	3,912	1,248	3,13	0,004	
X3	19,671	8,629	2,28	0,031	
S = 10,1307 R-Sq = 65,2% R-Sq(adj) = 61,2%					
Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	3	4994,5	1664,8	16,22	0,000
Residual Error	26	2668,4	102,6		
Total	29	7662,9			

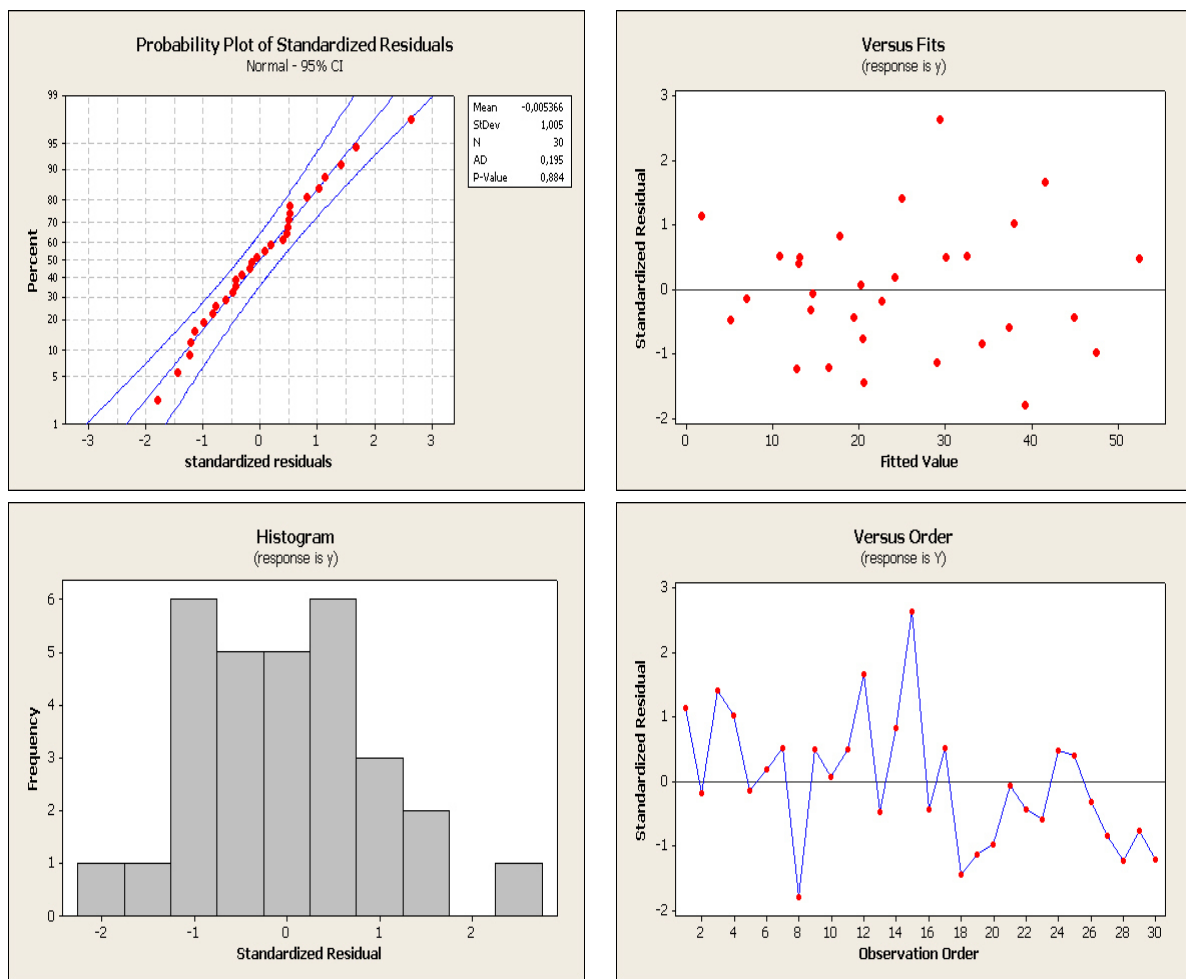
Figur 4: Utskrift fra statistiske analysar av cheddarostdataene for modellen i ligning (4).

Correlations: X1; X2; X3		
	X1	X2
X2	0,618	
X3	0,604	0,645

Figur 5: Pearson korrelasjon mellom variablene X_1 , X_2 og X_3 i cheddarostdatasettet.



Figur 6: Parvis spredningsplott av variablene i cheddarostdatasettet.



Figur 7: Residualplott (normalplott basert på standardiserte residualer i øvre venstre panel, standardiserte residualer mot tilpassede verdier i øvre høyre panel, histogram basert på standardiserte residualer i nedre venstre panel og standardiserte residualer mot rekkjefølgda på observasjonane i nedre høyre panel) for regresjonsmodellen i ligning (4) for cheddarostdatasettet.

- c) Forklar begrepet multikollinearitet. Kan dette vere et problem i regresjonsmodellen i ligning (4)? Kommenter korrelasjonsmatrisen i Figur 5 og det parvise spredningsplottet i Figur 6.

Frå tre-variabel regresjonmodellen i Figur 4 kan vi sjå at eddiksyre, X_1 , ikkje er signifikant, men i Figur 1 (ligning 1) er X_1 signifikant. Hva kan være årsaka til dette? Grunngi svaret.

Dette er ein observasjonsstudie. Vil det vere mogleg å designa eit eksperiment (design of experiment) for å utforske problemet vi studerer? Utdjup.

Oppgåve 3 Leikeplastklossar

Ein fabrikk produserer leikeplastklossar som kan vert saman sett. For å vere i stand til å setje saman klossane må dei ikkje ha defektar. Fabrikken tek prøver for å forsikre seg at klossane er av god kvalitet. Kvar dag vert det teke eit utval på $n = 250$ tilfeldige klossar og kvar kloss vert klassifisert som anten OK eller defekt. Resultatet etter to veker er gjevne under. Prosessen antas å vere i kontroll i løpet av desse to vekene utvala vart tekne.

Dag i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Mengd defekte	3	3	1	3	4	6	4	6	4	4	5	2	3	1

- a) Lag eit passande kontrolldiagram for å kontrollere sannsynligheten av defektar, p . Bruk alle dataene i tabellen ovanfor for å berekne kontrollgrensane (bruk 3σ grensar).

Kan vi anta at mengda defektar i kvart utval er tilnærma normalfordelt?

- b) Fabrikken er interessert i å oppdage om prosessen er ute av kontroll ved ein auke i sannsynligheten av defektar. Kor mange observasjonar, n , i kvar stikkprøve trengst for å oppdage ein endring fra $p = 0.2$ til $p_1 = 0.21$?

Oppg ve 4 Fedme og alkoholinntak

I ein studie s  p  samanhengen mellom fedme og alkoholinntak i Australia, vart eit utval p  $n_{\text{Lav}} = 165$ personar med lav grad av fedme, $n_{\text{Gjennomsnittleg}} = 161$ personar med gjennomsnittleg grad av fedme og $n_{\text{H g}} = 165$ personar med h g grad av fedme valt og deira alkoholinntak i mengd drikker per dag vart registrert. Alkoholinntaket vart klassifisert i fire grupper og f lgjande krysstabell vart observert.

Fedme	Alkoholinntak				Total
	0	1-2	3-5	6+	
Lav	45	45	41	34	165
Gjennomsnittleg	39	32	46	44	161
H�g	33	37	47	48	165
Total	117	114	134	126	491

- a) Basert p  desse dataene, kan vi konkludere med at Lav, Gjennomsnittleg og H g fedme populasjonar er ulike med omsyn p  alkoholinntak? Skriv ned nullhypotesen og den alternative hypotesen og utf r ein hypotesetest basert p  tabellen ovanfor. Bruk eit 5% signifikansniv . Det er oppgjeve at χ^2 -testobservatoren er lik 6.952. Du treng berre vise beregningen av eit av de 12 ledda i summen. Kva er konklusjonen fr  testen?