UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1110 — Statistiske metoder og dataanalyse 1.
Eksamensdag:	Onsdag 3. desember 2008.
Tid for eksamen:	9.00 – 12.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	Tabell for normal, χ^2- og t-fordelingen.
Tillatte hjelpemidler:	Godkjent lommeregner og Formelsamling for STK1100 og STK1110.

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

Oppgave 1.

For å bestemme usikkerheten til et måleinstrument for hastighet måler man flere ganger hastigheten til en gjenstand som beveges med en kjent hastighet. Målingene nedenfor angir resultatet fra 5 slike forsøk der gjenstanden hadde en hastighet på 100 km per time

$$102.93,\ 99.27,\ 102.67,\ 99.82,\ 98.24.$$

En rimelig modell er her at målingene er realisasjoner av uavhengige normalfordelte variable, $X_1, \ldots, X_5$, som alle er normalfordelt $N(\mu_0, \sigma^2)$, der $\mu_0 = 100$ er kjent.

a) Vis at fordelingen til $\frac{1}{\sigma^2}\sum_{i=1}^{5}(X_i - \mu_0)^2$ er χ^2-fordelt med 5 frihetsgrader. Finn en forventningsrett estimator for σ^2. Hva er variansen?

b) Bruk resultater fra punkt a) til å utlede en test med nivå α for nullhypotesen

$$H_0 : \sigma^2 = \sigma_0^2 \text{ mot alternativet } H_A : \sigma^2 > \sigma_0^2$$

der σ_0^2 er kjent.

(Fortsettes side 2.)

c) Angi en øvre og nedre grense for p-verdien til testen basert på observasjonene i begynnelsen av oppgaven når $\sigma_0^2 = 2$. Forklar omhyggelig hva p-verdier er og hvordan resultatet du fant skal tolkes.

Oppgave 2.

Følgende tabell viser et utsnitt av vektene, y_i, til 12 barn med en spesiell type spiseforstyrrelse samt deres høyde, x_i, og alder, z_i.

Table 1: Vekt i kg, y_i, høyden i meter, x_i, og alder i år, z_i, for 12 barn med spiseforstyrrelser.

Vekt, y_i	Høyde, x_i	Alder, z_i
23.9	1.44	8
26.5	1.50	10
19.8	1.24	6
$\vdots$		$\vdots$
28.3	1.55	12
25.4	1.45	9

Her er vekt respons, og høyde og alder er uavhengige variable. Betrakt først situasjonen der sammenhengen mellom vekt og høyde beskrives med en enkel lineær regresjonsmodell:

$$Y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \ldots, 12.$$

a) Forklar hvilke forutsetninger en slik modell bygger på. Hvilke egenskaper til estimatorene for parametrene β_0 og β_1 kan utledes av det som i læreboka kalles "standard statistisk modell", det vil si uten å anta at feilleddene $e_1, \ldots, e_{12}$ er normalfordelte?

b) Finn et estimat for β_1 og bestem et 95% konfidensintervall for β_1. Her kan du bruke at $\sum_{i=1}^{12} y_i = 280.869$, $\sum_{i=1}^{12} x_i = 16.078$, $\sum_{i=1}^{12} x_i^2 = 21.873$ og $\sum_{i=1}^{12} x_i y_i = 381.526$. Residual kvadratsum er $RSS_1 = \sum_{i=1}^{12}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 41.645$, der $\hat{\beta}_0$ og $\hat{\beta}_1$ er estimater for parametrene β_0 og β_1.

En lineær regresjonsmodell uten konstantledd der høyde og alder er uavhengige variable kan formuleres som

$$Y_i = \alpha_1 x_i + \alpha_2 z_i + e_i, i = 1, \ldots, 12.$$

(Fortsettes side 3.)

c) Forklar hvordan modellen kan uttrykkes på matriseform,

$$\mathbf{Y} = \mathbf{X}\alpha + \mathbf{e},$$

der $\mathbf{X}$ er en matrise av kjente tall og $\mathbf{Y}$ og $\mathbf{e}$ er vektorer av tilfeldige variable.

d) Finn et estimat for α_2. Her kan du i tillegg til de størrelsene som er oppgitt i punkt b), bruke at $\sum_{i=1}^{12} z_i^2 = 976, \sum_{i=1}^{12} x_i z_i = 144.247$ og $\sum_{i=1}^{12} z_i y_i = 2534.908$. Du kan også trenge

$$\begin{pmatrix} 21.873 & 144.247 \\ 144.247 & 976.0 \end{pmatrix}^{-1} = \begin{pmatrix} 1.804 & -0.267 \\ -0.267 & 0.040 \end{pmatrix}.$$

Residual kvadratsum er $RSS_2 = \sum_{i=1}^{12}(y_i - \hat{\alpha}_1 x_i - \hat{\alpha}_2 z_i)^2 = 28.3$, der $\hat{\alpha}_1$ og $\hat{\alpha}_2$ er estimater for koeffesientene α_1 and α_2. Hva blir den estimerte standardfeilen til estimatoren for α_2?

Oppgave 3.

Tabellen nedenfor viser antall ulykker per måned på en bestemt veistrekning i løpet av en 20-måneders periode. I 9 av månedene var det altså ingen ulykker, i 7 av månedene var det en ulykke osv.

Table 2: Antall ulykker per måned i en periode på 20 måneder.

Antall ulykker	Antall måneder
0	9
1	7
2	3
3	1

En rimelig modell i denne situasjonen er at antall ulykker antas å være realisasjoner av uavhengige Poisson fordelte variable $X_1, \ldots, X_{20}$ med samme punktsannsynlighet

$$P(X = x) = \frac{\lambda^x}{x!}\exp(-\lambda), \quad x = 0, 1, \ldots$$

a) Finn moment- og sannsynlighetsmaksimeringsestimatoren til λ. Beregn estimatene ved å bruke observasjonene fra tabellen i begynnelsen av oppgaven.

(Fortsettes side 4.)

b) Utled et tilnærmet 95% konfidensintervall for λ. Beregn intervallet på grunnlag av observasjonene i tabellen.

Se nå på en situasjon der en bare bruker opplysninger om det har funnet sted ulykker i hver av de 20 månedene eller ikke. La $Y_i = 1$ hvis $X_i = 0$ og $Y_i = 0$ ellers. Da er $Y = \sum_{i=1}^{20} Y_i$ antallet ulykkesfrie måneder.

c) Begrunn at antallet måneder der det ikke skjer ulykker, det vil si Y, er binomisk fordelt og angi punktsannsynligheten til Y. Bestem moment- og sannsynlighetsmaksimeringsestimatoren til λ i dette tilfellet. Hva blir estimatene?

d) Utled og beregn et tilnærmet 95% konfidensintervall for λ på grunnlag av antall ulykkesfrie måneder.

SLUTT