UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	STK1110 — Statistiske metoder og dataanalyse 1.
Eksamensdag:	Tirsdag 5. desember 2006.
Tid for eksamen:	15.30 – 18.30.
Oppgavesettet er på 3 sider.	
Vedlegg:	Tabell for normalfordeling, tabell for t-fordeling.
Tillatte hjelpemidler:	Godkjent lommeregner og Formelsamling for STK1100 og STK1110.

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

Oppgave 1.

a) Forklar hva som menes med at en estimator er forventningsrett og at den er konsistent.

I resten av oppgaven er $X \sim \text{bin}(n, p)$ og $\hat{p} = \frac{X}{n}$.

b) Forklar hvorfor $\hat{p}$ er forventningsrett og konsistent.

c) La $\theta = p(1-p)$. En rimelig estimator for θ er $\hat{\theta} = \hat{p}(1-\hat{p})$. Er den forventningsrett? Hvis svaret er nei, foreslå en modifikasjon av $\hat{\theta}$ som er det.

(Fortsettes side 2.)

Oppgave 2.

Anta at $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte med tetthet

$$f(x|\sigma) = \frac{1}{2\sigma}\exp(-|x|/\sigma), \qquad -\infty < x < \infty$$

a) Bestem sannsynlighetsmaksimeringsestimatoren, $\hat{\sigma}$, for σ. Finn $E(\hat{\sigma})$.

b) Vis at sannsynlighetsmaksimeringsestimatoren $\hat{\sigma}$ er tilnærmet $N(\sigma, \sigma^2/n)$ når antallet observasjoner, n, er stort.

c) Forklar hvorfor både intervaller av formen

$$(\hat{\sigma} - z_{1-\alpha/2}\ \hat{\sigma}/\sqrt{n}\,, \hat{\sigma} + z_{1-\alpha/2}\ \hat{\sigma}/\sqrt{n}\,)$$

og

$$\left(\frac{\hat{\sigma}}{(z_{1-\frac{\alpha}{2}}/\sqrt{n}) + 1}, \frac{\hat{\sigma}}{-(z_{1-\frac{\alpha}{2}}/\sqrt{n}) + 1}\right)$$

er konfidensintervaller med tilnærmet konfidenskoeffisient $1 - \alpha$ når antallet observasjoner er stort. Her er $z_{1-\frac{\alpha}{2}}$, $1 - \alpha/2$ kvantilen i standardnormalfordelingen $N(0, 1)$.

Oppgave 3.

Dataene nedenfor er et utsnitt av 17 årlige målinger av prosentandel snøinnhold, x_i, 1. april og gjennomsnittlig vannføring om våren, y_i, målt i tommer i en elv i USA.

i	x_i	y_i
1	23.1	10.5
2	32.8	16.7
⋮	⋮	⋮
16	21.1	10.5
17	27.6	16.1

Her er $\sum_{i=1}^{17} x_i = 511.5$, $\sum_{i=1}^{17} y_i = 267.1$, $\sum_{i=1}^{17} x_i^2 = 16628.7$, $\sum_{i=1}^{17} y_i^2 = 4549.43$ og $\sum_{i=1}^{17} x_i y_i = 8653.45$.

(Fortsettes side 3.)

La vannføring være responsvariabel. Vi antar følgende lineære regresjonsmodell

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad i = 1, \ldots, 17$$

der $\epsilon_1, \ldots, \epsilon_{17}$ er uavhengige $N(0, \sigma^2)$-fordelte variable.

a) Beregn minste kvadraters estimatene $\hat{\beta}_0$ og $\hat{\beta}_1$ for dataene ovenfor og skisser den tilpassede regresjonslinja.

b) Du skal også tilpasse en modell uten konstantledd, dvs.

$$Y_i = \gamma x_i + \eta_i$$

der $\eta_1, \ldots, \eta_{17}$ er uavhengige $N(0, \tau^2)$-fordelte variable.

Finn minste kvadraters estimatoren $\hat{\gamma}$. Beregn estimatet for γ i dataene ovenfor og skisser også denne regresjonslinja i samme diagram som regresjonslinja fra punkt a).

c) Forklar hvorfor estimatoren $\hat{\beta}_0$ er normalfordelt. Hvordan kan nullhypotesen

$$H_0 : \beta_0 = 0 \qquad \text{mot} \qquad H_A : \beta_0 \neq 0$$

testes? Angi en øvre og nedre grense for p-verdien. Kommenter resultatet av testen i lys av resultatene du fant i punkt a) og b).
Her kan du bruke at kvadratsummen av residualene $RSS = 45.56$ og uttrykket for $\text{Var}(\hat{\beta}_0)$ fra formelsamlingen.

Oppgave 4.

En logistisk fordelt variabel har kumulativ fordelingsfunksjon

$$F(x) = \frac{1}{1 + e^{-x}}, \qquad -\infty < x < \infty .$$

Forklar hvordan man beregner et $Q - Q$ plott for sammenligning av den kumulative fordelingsfunksjonen Φ til en $N(0, 1)$-fordelt variabel og F gitt ovenfor. Angi verdiene som svarer til kvartilene og 0.1 og 0.9 kvantilene, dvs. verdiene for $p = 0.1, 0.25, 0.5, 0.75$ og 0.9. Skisser plottet og kommenter utseendet.

SLUTT