UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST 102 — Videregående kurs i statistikk.
Eksamensdag:	Onsdag 31. mai 1995.
Tid for eksamen:	09.00 – 15.00.
Oppgavesettet er på 5 sider.	
Vedlegg:	2 tabeller.
Tillatte hjelpemidler:	Formelsamling for ST 101, ST 102, lommekalkulator.

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

Oppgave 1.

Anta at $Y_1, Y_2, \ldots, Y_n$ er uavhengige og normalfordelte med forventning μ og varians σ^2.

a) Anta at σ^2 er kjent. Vis at

$$\hat{\mu} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

er sannsynlighetsmaksimeringsestimatoren (SME) for μ.
Hvorfor kan vi også betegne $\hat{\mu}$ som en momentestimator?

b) Anta at både μ og σ^2 er ukjente. Finn simultan SME for (μ, σ^2).
Er denne estimatoren også en momentestimator?
Er SME forventningsrett i dette tilfellet?

c) Foreslå en testobservator for hypotesetestingsproblemet $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$ med signifikansnivå α når σ^2 er ukjent.

Argumenter for testobservatorens fordeling når $\mu = \mu_0$.

(Fortsettes side 2.)

Testen kan gjennomføres både ved hjelp av forkastningsgrenser og p-verdi (signifikanssannsynlighet). Angi for begge måter hvordan og diskuter kort sammenhengen.

d) Angi et 95% konfidensintervall for μ.
Forklar hva det betyr at dette er et konfidensintervall.
Utled at intervallet du oppga faktisk er et 95% konfidensintervall.

Diskuter kort sammenhengen mellom konfidensintervaller og hypotesetesting.

e) Benytt BLSS-utskriften under samt vedlagt tabell over Student's t-fordeling til å svare på følgende spørsmål.

1) Hva blir de forventningsrette estimatorene for μ og σ^2? Hva blir SME?
2) Kan man forkaste $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$ med nivå $\alpha = 0.001$?
3) Kan man forkaste $H_0 : \mu = 1$ mot $H_1 : \mu \neq 1$ med nivå $\alpha = 0.05$?
4) Hva blir et 99% konfidensintervall for μ?

```
  stat x

Statistics: x
Col    N    Mean      SD       Min        25%       50%       75%       Max
 1    25   1.623    1.486    -2.162     0.7543    1.648     2.662     4.018

  ht x

Hypothesis Test: one-sample, two-tail t.  Data: x.

Col    N     null     sample     sample      df     test     P     alpha  crit
             mean      mean       SE                stat   value          value
 1    25   0.0000    1.6232    0.29722     24.0     5.46   .0000   .0500  2.06
                                                                  .0100  2.80
                                                                  .0010  3.75
```

Oppgave 2.

Du skal i denne oppgaven undersøke om hvorvidt tvillingpar er eneggede eller toeggede har noen sammenheng med mors alder. Dataene du skal regne på er angitt i tabellen under. Dataene er konstruert, men de er i samsvar med hva man kunne observere hvis man velger ut $n = 200$ tvillingpar tilfeldig. Tabellen angir antall tvillingpar etter mors alder og om de er eneggede/toeggede.

(Fortsettes side 3.)

Mors alder	Eneggede	Toeggede
$\leq$ 25 år	30	30
26–34 år	40	80
$\geq$ 35 år	5	15

La X_{1j} = antall eneggede tvillingpar og X_{2j} = antall toeggede tvillingpar med mor i alderskategori nr. j, $j = 1, 2, 3$. La videre p_{1j} være sannsynligheten for at et tilfeldig utvalgt tvillingpar er enegget og har mor i alderskategori nr. j og la p_{2j} betegne den tilsvarende sannsynligheten for toeggede tvillingpar.

a) Foreslå en sannsynlighetsmodell for $(X_{11}, X_{12}, X_{13}, X_{21}, X_{22}, X_{23})$ og argumenter kort for dens gyldighet.

   Hvor mange parametre er det i modellen?

   Sannsynlighetsmaksimeringsestimatorene (SME) for p_{ij}-ene er gitt ved $\hat{p}_{ij} = X_{ij}/200$. (Dette skal du ikke vise.)
   Finn forventninger, varianser og kovarianser for SME.

Definer nå $p_{i\cdot}$ som sannsynligheten for at tvillingpar er enegget $(i = 1)$ eller toegget $(i = 2)$ og tilsvarende $p_{\cdot j}$ som sannsynligheten for at tvillingparets mor er i alderskategori nr. j.

Anta at morens alder en uavhengig av om tvillingparet er enegget eller toegget.

b) Hva blir da sammenhengen mellom $p_{ij}, p_{i\cdot}$ og $p_{\cdot j}$?

   Hvor mange parametre er det i modellen under uavhengighetsantagelsen?

   SME for $p_{i\cdot}$ blir nå $p^*_{i\cdot} = (X_{i1} + X_{i2} + X_{i3})/200$. Tilsvarende blir SME for $p_{\cdot j}$ lik $p^*_{\cdot j} = (X_{1j} + X_{2j})/200$. (Dette skal du ikke vise.)

   Hva blir SME for p_{ij} under uavhengighetsantagelsen?

Man kan teste nullhypotesen

H_0 : Hvorvidt tvillingparet er enegget eller toegget er uavhengig av mors alder

(mot alternativet at det er avhengighet) ved en kji-kvadratobservator oppgitt på generell form i Formelsamlingen for ST 102.

c) Hvordan blir kji-kvadratobservatoren seende ut i dette tilfellet?
   Hva er dens tilnærmede fordeling under nullhypotesen?
   Argumenter for at avhengighet rimeligvis skal resultere i en stor kji-kvadratobservator.
   Utfør testen og konkluder om dataene i tabellen tyder på avhengighet.

(Fortsettes side 4.)

d) Man kan også teste nullhypotesen i punkt c) ved en sannsynlighetskvotetest. Forklar ideen i en slik test og definer sannsynlighetskvoten Λ.
Utled et uttrykk for Λ i den konkrete situasjonen.
Beregn $2\ln\Lambda$ og sammenlign med kji-kvadratobservatoren fra punkt c). Kommenter.

e) 1) Man kan vise at dersom man hadde valgt ut et på forhånd bestemt antall tvillingmødre i hver alderskategori så kan likevel testene i punktene c) og d) benyttes til å undersøke om det er en sammenheng. Hvilken fordel kan en slik utvalgsplan ha?

2) Hva tyder resultatene i denne oppgaven på om sammenhengen mellom en tvillingmors alder og sannsynligheten for at tvillingen er toegget?

Oppgave 3.

I denne oppgaven behandles et spesialtilfelle av den multiple lineære regresjonsmodellen $Y_i = b_1 + b_2 x_{i2} + b_3 x_{i3} + \cdots + b_p x_{ip} + \varepsilon_i$ der x_{ij}-ene er gitte forklaringsvariable, $b_1, \ldots, b_p$ regresjonskoeffisienter og feilleddene ε_i uavhengige og normalfordelte med forventning $E\varepsilon_i = 0$ og varians $\mathrm{Var}\,\varepsilon_i = \sigma^2$. Vi lar altså $x_{i1} = 1$ for alle i.

Spesialtilfellet består i at x_{ij}-ene er sentrerte, dvs.

$$\sum_{i=1}^{n} x_{ij} = n\overline{x}_j = 0$$

og ukorrelerte, dvs.

$$\sum_{i=1}^{n} x_{ij}x_{ik} = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k) = 0$$

for $j, k = 2, 3, \ldots, p$ og $j \neq k$.

a) Vis at minste kvadratens estimatorene for $b_1, b_2, \ldots, b_p$ gis ved

$$\hat{b}_1 = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

og

$$\hat{b}_j = \frac{\sum_{i=1}^{n} x_{ij}Y_i}{S_j^2} \qquad j = 2, 3, \ldots, p$$

der $S_j^2 = \sum\limits_{i=1}^{n} x_{ij}^2$.

(Fortsettes side 5.)

b) Angi simultanfordelingen til $(Y_1, Y_2, \ldots, Y_n)$.
Forklar hvorfor minste kvadraters estimatorene $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_p$ også er SME?
Det er tilstrekkelig å vise dette under antagelse av at σ^2 er kjent.

c) Angi hva w_{ij} må være ($i = 1, \ldots, n$, $j = 1, \ldots, p$) for at vi kan skrive

$$\hat{b}_j = \sum_{i=1}^{n} w_{ij} Y_i$$

Vis at $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_p$ er normalfordelte med

$$\begin{aligned} E\hat{b}_j &= b_j \\ \mathrm{Var}(\hat{b}_1) &= \frac{\sigma^2}{n} \\ \mathrm{Var}(\hat{b}_j) &= \frac{\sigma^2}{S_j^2} \qquad \text{når } j = 2, 3, \ldots, p \end{aligned}$$

og

$$\mathrm{Cov}(\hat{b}_j, \hat{b}_k) = 0 \qquad \text{når } j \neq k.$$

d) De predikerte verdiene er definert ved

$$\widehat{Y}_i = \hat{b}_1 + \hat{b}_2 x_{i2} + \cdots + \hat{b}_p x_{ip}$$

Forklar hvorfor $\widehat{Y}_1, \widehat{Y}_2, \ldots, \widehat{Y}_n$ er normalfordelte med $E\widehat{Y}_i = EY_i$ og $\mathrm{Var}\,\widehat{Y}_i = \sigma^2 \left[\frac{1}{n} + \frac{x_{i2}^2}{S_2^2} + \frac{x_{i3}^2}{S_3^2} + \cdots + \frac{x_{ip}^2}{S_p^2} \right]$.

(Du skal ikke se på avhengighetene mellom $\widehat{Y}_1, \ldots, \widehat{Y}_n$).

e) Vis at $s^2 = \frac{1}{n-p} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$ er en forventningsrett estimator for σ^2.
Hint: Benytt uten bevis oppspaltningen

$$\sum_{i=1}^{n} (Y_i - EY_i)^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n} (\widehat{Y}_i - EY_i)^2$$

SLUTT