UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST102 — Videregående kurs i statistikk
Eksamensdag:	Onsdag 27. mai 1998.
Tid for eksamen:	09.00 – 15.00.

Oppgavesettet er på 3 sider.

Vedlegg:	1) 39 observasjoner om forbruk pr. capita. 2) Tabell over χ^2-fordelingen.
Tillatte hjelpemidler:	Formelsamlinger i ST101 og ST102, lommekalkulator.

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

Oppgave 1.

Hvordan forbruk avhenger av inntekt er et viktig spørsmål i økonomi. I vedlegg 1 er det gjengitt 39 observasjoner av sammenhengen mellom per capita konsum og inntekt i USA. Tallene dreier seg om tidsrommet 1946–1984 og er angitt i 1972 dollarverdi.

Vi skal i denne oppgaven og i den neste se hvordan en kan undersøke om samme sammenheng kan sies å gjelde for såvel de første 20 årene som de 19 siste. Et første skritt er derfor å tilpasse modellene

$$\begin{aligned} Y_{i1} &= A_1 + B_1 x_{i1} + \varepsilon_{i1} \qquad i = 1, \ldots, 20 \\ Y_{i2} &= A_2 + B_2 x_{i2} + \varepsilon_{i2} \qquad i = 1, \ldots, 19 \end{aligned}$$

der ε_{i1}, $i = 1, 2, \ldots, 20$ og ε_{i2}, $i = 1, 2, \ldots, 19$ antas å være uavhengige normalfordelte restledd med forventning 0 og $\mathrm{Var}(\varepsilon_{i1}) = \sigma_1^2$ og $\mathrm{Var}(\varepsilon_{i2}) = \sigma_2^2$.

Resultatet av tilpasning av de to modellene gir at minste kvadraters estimatene for A_1, B_1, A_2, B_2 blir

$$\widehat{A}_1 = 206.5, \quad \widehat{B}_1 = 0.84, \quad \widehat{A}_2 = -170.8 \quad \text{og} \quad \widehat{B}_2 = 0.95.$$

(Fortsettes side 2.)

Dessuten er kvadratsummene

$$\sum_{i=1}^{20}(y_{i1} - \widehat{A}_1 - \widehat{B}_1 x_{i1})^2 = 13552.2$$
$$\sum_{i=1}^{19}(y_{i2} - \widehat{A}_2 - \widehat{B}_2 x_{i2})^2 = 23148.9$$

a) Utled minste kvadraters estimatorer for A_1, B_1, A_2 og B_2. Forklar hvorfor disse også blir sannsynlighetsmaksimeringsestimatorer.

b) Det er hensiktsmessig å skrive modellene på formen

$$E(Y_{i1}) = \alpha_1 + \beta_1(x_{i1} - \overline{x}_1) \qquad i = 1, \ldots, 20$$
$$E(Y_{i2}) = \alpha_2 + \beta_2(x_{i2} - \overline{x}_2) \qquad i = 1, \ldots, 19$$

der $\overline{x}_1 = \frac{1}{20}\sum_{i=1}^{20} x_{i1}$ og $\overline{x}_2 = \frac{1}{19}\sum_{i=1}^{19} x_{i2}$.
Angi sammenhengen mellom de to parametersettene (A_1, B_1, A_2, B_2) og $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. Forklar hvorfor minste kvadraters estimatorene for (B_1, B_2) og (β_1, β_2) blir de samme.

c) Utled fordelingen til $\widehat{\beta}_1 - \widehat{\beta}_2$.

d) Utled en test med nivå α for $H_0\colon \sigma_1^2 = \sigma_2^2$ mot $H_1\colon \sigma_1^2 \neq \sigma_2^2$ og forklar hvordan man kan konstruere et $100(1-\alpha)\%$ konfidensintervall for σ_1^2/σ_2^2.

e) Beregn testen og konfidensintervallet du utledet i punkt d) på grunnlag av de oppgitte tallene. Velg $\alpha = 0.05$. Forklar resultatet og sammenhengen mellom de to størrelsene.

Tabell over 2.5 og 97.5 percentilene i Fisherfordelingen.

		Frihetsgrader teller			
		17	18	19	20
Frihetsgrader nevner	17	0.374 2.673	0.382 2.652	0.389 2.633	0.396 2.616
	18	0.377 2.617	0.385 2.596	0.392 2.576	0.401 2.560
	19	0.380 2.567	0.388 2.546	0.396 2.526	0.403 2.509
	20	0.382 2.523	0.391 2.501	0.399 2.482	0.406 2.464

Oppgave 2.

Vi skal i denne oppgaven betrakte samme situasjon som i oppgave 1, men i tillegg anta at $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

a) Angi en estimator for σ^2 basert på alle observasjonene.
Hva er forventningen og variansen? Beregn estimatet for σ^2 på grunnlag av tallene oppgitt i oppgave 1.

(Fortsettes side 3.)

b) Begrunn og utled en test med nivå α for nullhypotesen $H_0 : \beta_1 = \beta_2$ mot $H_1 : \beta_1 \neq \beta_2$ ved å ta utgangspunkt i fordelingen til $\widehat{\beta}_1 - \widehat{\beta}_2$ du utledet i punkt c) i oppgave 1.

c) Forklar hvordan en kan tilpasse en modell der $\beta_1 = \beta_2 = \beta$, det vil si der

$$\begin{aligned} Y_{i1} &= \alpha_1 + \beta(x_{i1} - \overline{x}_1) + \varepsilon_{i1} \qquad i = 1, \ldots, 20 \\ Y_{i2} &= \alpha_2 + \beta(x_{i2} - \overline{x}_2) + \varepsilon_{i2} \qquad i = 1, \ldots, 19 \end{aligned}$$

ved å sette opp en passende multippel regresjon.

d) Finn minste kvadraters estimatoren for β for modellen i punkt c).

Oppgave 3.

Betrakt den geometriske fordelingen med punktsannsynlighet

$$P(X = x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots$$

der $0 \leq p \leq 1$.

Vi skal anta vi observerer en geometrisk fordelt stokastisk variabel X.

a) Finn forventning og varians til X.

b) Utled en test med nivå α for $H_0 : p \geq \frac{1}{2}$ mot $H_1 : p < \frac{1}{2}$ basert på en observasjon av X.
Forklar spesielt betydningen kravene til feil av type I og type II har for å bestemme den kritiske verdien.

c) Beregn styrkefunksjonen for testen i punkt c) når $\alpha = 0.05$. Angi styrken i punktene $p = 0.1$, 0.25 og 0.4 og skisser styrkefunksjonen.

d) Finn p-verdien hvis den observerte verdien av X er $X = 7$.

e) Vis at X er en Markov (minimum variance unbiased) estimator for $\frac{1}{p}$.

Oppgave 4.

Tabellen nedenfor viser sammenhengen mellom yrkeserfaring og klasse for 786 kvinner basert på en spørreundersøkelse.

		Klasse: Arbeider	Klasse: Lavere funksjonær	Klasse: Høyere funksjonær
Yrkeserfaring	< 3 år	28	86	76
	3–10 år	68	178	162
	> 10 år	40	87	61

a) Formuler og begrunn en modell for observasjonene og en aktuell nullhypotese og alternativ. Hva innebærer nullhypotesen?

b) Test hypotesen du formulerte i punkt a) med nivå 0.05.

SLUTT