NORGES HANDELSHØYSKOLE

Eksamen høstsemesteret 2009

Kurskode: INT010 Tittel: Anvendt metode

Dato: 11.11.2009 Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på tlf. 59670/41645914.

Hjelpemidler til eksamen:

Hjelpemidler tillatt: **JA**, alle trykte/egenskrevne

Kalkulator: **JA** *[I tråd med retningslinjer for bruk av kalkulator, jf. utfyllende bestemmelser til Forskrift om eksamen ved Norges Handelshøyskole (fulltidsstudiene).]*

Alle delspørsmål i oppgavesettet teller likt.

Oppgave 1

Det er kjent at endringer i oljeprisen påvirker økonomien i USA. En økonom ønsker å undersøke om prisen på et fat olje påvirker konsumprisindeksen (KPI). Økonomen samler inn to datasett; ett med 14 månedlige observasjoner av økningen av KPI i prosent, der pris på et fat olje er over $66, og ett med 9 månedlige observasjoner av økningen av KPI i prosent, der prisen på et fat olje er under $58. La gjennomsnittet og standardavvik være hhv. $\overline{X}_1 = 0{,}317\%$ og $S_1 = 0{,}12\%$ for utvalget med 14 observasjoner (populasjon 1), mens for utvalget med 9 observasjoner (populasjon 2) er $\overline{X}_2 = 0{,}210\%$ og $S_2 = 0{,}11\%$.

a) Hvilke forutsetninger må oppfylles for at vi skal kunne sammenligne om forventninger / varianser er like i disse to populasjonene?

b) Vis at hypotesen om like varianser blir akseptert med bruk av 5% signifikansnivå.

La μ_1 og μ_2 betegne forventningsverdiene til prosentøkningen av KPI for hhv. populasjon 1 og 2.

c) Test hypotesen H_0: $\mu_1 - \mu_2 = 0$ mot H_1: $\mu_1 - \mu_2 \neq 0$, med antagelsen om like varianser, på 5% signifikansnivå. Kommenter.

Oppgave 2

I en amerikansk studie ønsket man å undersøke hvordan pris (selgers prisforlangende) på bruktbiler avhenger av flere forklaringsvariable som bilens alder, kjørelengde, tilstand (dårlig, gjennomsnittlig, god) og type selger (privat eller forhandler). For en bestemt biltype (Chevrolet Camaro, anslått nybilpris $25000) ble det registrert n=31 ulike bruktbiler lagt ut for salg. For hver bil ble det registrert følgende data:

y: Pris (selgers prisforlangende i $)
x_1: Alder (i år)
x_2: Kjørelengde (i 1000 miles)
x_3: Indikatorvariabel for selger, 1 dersom bilforhandler, 0 ellers (privat)
x_4: Indikatorvariabel for tilstand, 1 dersom dårlig, 0 ellers
x_5: Indikatorvariabel for tilstand, 1 dersom gjennomsnittlig, 0 ellers

a) Formuler en lineær regresjonsmodell med y som avhengig variabel og $x_1,\ldots,x_5$ som forklaringsvariable. De to indikatorvariablene x_4 og x_5 representerer i fellesskap bilens tilstand. Hvilke kombinasjoner av verdier for x_4 og x_5 svarer til henholdsvis dårlig, gjennomsnittlig og god tilstand?

Utskrift fra Minitab er gitt under.

Regression Analysis: y versus x1; x2; x3; x4; x5

```
The regression equation is
y = 17357 - 1132 x1 - 33,2 x2 + 776 x3 - 3275 x4 - 2556 x5
```

Predictor	Coef	SE Coef	T	P
Constant	17357	1280	13,56	0,000
x1	-1131,9	274,8	-4,12	0,000
x2	-33,24	23,57	-1,41	0,171
x3	775,6	913,1	0,85	0,404
x4	-3275	1112	-2,95	0,007
x5	-2556,4	915,1	-2,79	0,010

S = 1543,99   R-Sq = 89,6%   R-Sq(adj) = 87,6%

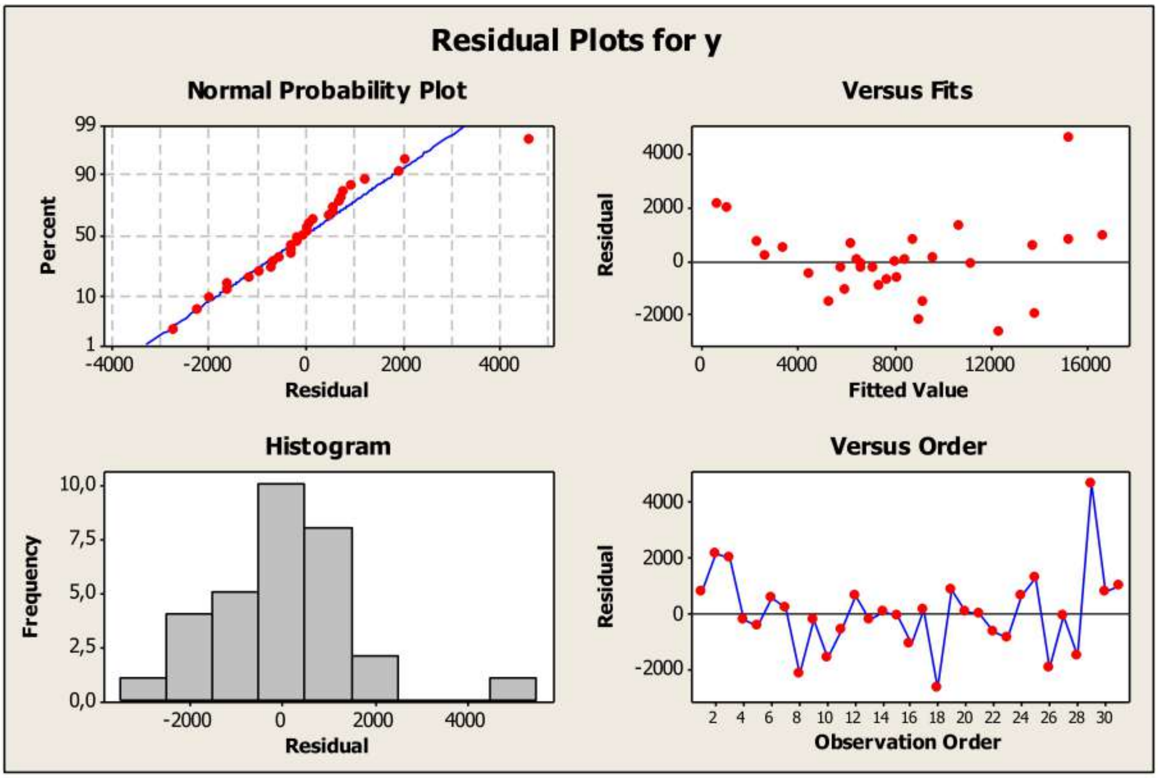
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	516005252	103201050	43,29	0,000
Residual Error	25	59597467	2383899		
Total	30	575602720			

Source	DF	Seq SS
x1	1	438119324
x2	1	29151570
x3	1	25641144
x4	1	4489888
x5	1	18603327

b) Tolk regresjonskoeffisienten knyttet til bilens alder. Hvilken pris predikeres for en helt ny bil solgt gjennom forhandler? Er den estimerte modellen rimelig ut fra disse resultatene?

c) Forklar kort hvorfor residualplottene under kan tyde på at modellantagelsene ikke er tilstrekkelig tilfredstilt.



I studien valgte man å transformere y til $y^* = \ln(y)$ og gjennomføre analysen med denne som avhengig variabel i regresjonen mot $x_1,\ldots,x_5$. Utskriften fra Minitab er gitt under. For resten av oppgaven vil vi benytte resultatene fra denne analysen.

Regression Analysis: ln y versus x1; x2; x3; x4; x5

```
The regression equation is
ln y = 10,0 - 0,178 x1 - 0,00376 x2 + 0,122 x3 - 0,217 x4 - 0,0651 x5


Predictor        Coef   SE Coef      T      P
Constant       9,9977    0,1007  99,32  0,000
x1           -0,17834   0,02161  -8,25  0,000
x2          -0,003759  0,001853  -2,03  0,053
x3            0,12247   0,07180   1,71  0,100
x4           -0,21749   0,08743  -2,49  0,020
x5           -0,06508   0,07195  -0,90  0,374


S = 0,121397   R-Sq = 95,8%   R-Sq(adj) = 95,0%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       5  8,4056  1,6811  114,07  0,000
Residual Error  25  0,3684  0,0147
Total           30  8,7741


Source  DF  Seq SS
x1       1  7,6234
x2       1  0,3823
x3       1  0,2873
x4       1  0,1006
x5       1  0,0121
```

d) Drøft tolkningen av β_2 med y vs ln(y) som venstresidevariabel.

e) Formuler en lineær regresjonsmodell for y* mot $x_1,\ldots,x_5$, og skriv opp den tilhørende estimerte modellen for y. Hvordan vil du nå anslå prisen på en ny bil solgt hos forhandler? Sammenlign med den tilsvarende predikerte prisen i b).

f) Obama har tilbud om å kjøpe en to år gammel Chevrolet Camaro med kjørelengde 21 tusen miles, som selges privat og som er i god stand. Prisen er satt til \$17500. Bruk den estimerte regresjonsligningen til å regne ut et estimat for forventet markedspris på en slik bil. Det er også beregnet et 95% konfidensintervall (CI) og et 95% prediksjonsintervall (PI) for logaritmen til prisen på en slik bil. Utskriften fra Minitab finner du under. Hvilket av intervallene CI eller PI er mest relevant for Obama dersom han ønsker å finne ut om pristilbudet på \$17500 er urimelig høyt?

```
Predicted Values for New Observations

New
Obs     Fit  SE Fit         95% CI              95% PI
 1   9,5621  0,0883  (9,3801; 9,7440)  (9,2529; 9,8713)
```

g) Finn et 95% konfidensintervall for parameteren β_5. Hvorfor er det ut fra dette intervallet rimelig å slutte at variabelen x_5 ikke har signifikant betydning i modellen?

Oppgave 3

Situasjonen er som i oppgave 2. I denne oppgaven ønsker vi å finne ut om det er signifikant forskjell på tilstanden til biler av merket Chevrolet Camaro som selges privat og som selges gjennom forhandler. Følgende kontingenstabell med beregnet forventet frekvens i parantes er satt opp basert på data fra oppgave 2:

Tilstand / Kategori	God	Gjennomsnittlig	Dårlig	SUM
Privat	0 (4,06)	4 (5,22)	14 (8,71)	18
Forhandler	7 (2,94)	5 (3,77)	1 (6,29)	13
SUM	7	9	15	31

a) Hvorfor bør den oppgitte kontingenstabellen ikke benyttes for en kjikvadrattest? Slå sammen god og gjennomsnittlig, og sett opp en ny kontingenstabell. Beregn også forventet frekvens for hver celle i den nye kontingenstabellen.

Kjikvadrattesten for uavhengighet kan også benyttes til å teste for **homogenitet,** dvs. vi ønsker å teste om kategoriene (her: forhandler og privat) har en lik andel av gode/gjennomsnittlige og dårlige biler til salgs. For fire celler vil vi altså teste

H_0: $p_{11} = p_{21}$ og $p_{12} = p_{22}$
H_1: minst en ulik

der p_{ij} er sannsynligheten for at en tilfeldig observasjon havner i rad i og kolonne j.

b) Gjennomfør en kjikvadrattest for å teste for homogenitet basert på kontingenstabellen du fant i a), dvs. test om det er signifikant forskjell på tilstanden av bilene som selges privat og som selges gjennom forhandler. Bruk 1% signifikansnivå.

Vi ønsker også å finne ut om det er forskjell i kjørelengde for biler solgt gjennom forhandler og privat. Det benyttes en ikke-parametrisk test, og det oppgis at testobservatorens verdi (rangsummen) er $T = 126{,}5$.

c) Beregn forventning og standardavvik til testobservatoren T, dvs. beregn $E(T)$ og σ_T. Test om kjørelengden på biler solgt gjennom forhandler er kortere enn biler solgt privat på 5% signifikansnivå?

Oppgave 4

Anta at totalkostnaden Y ved et prosjekt kan uttrykkes ved n kostnadskomponenter X_1, X_2, …, X_n, dvs. $Y = f\,(X_1,X_2,\ldots,X_n)$, der f er en kjent funksjon.

Anta at forventninger, varianser og kovarianser til de n kostnadskomponentene er gitt ved $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$ og $Cov(X_i,X_j) = \sigma_{ij}$.

Vi ønsker å beregne E(Y) og Var(Y), men det er vanskelig med mindre Y er en lineær funksjon av X_i'ene. I ingeniørlitteratur finner vi at

(1) $E(Y) \approx f(\mu_1, \mu_2, \ldots, \mu_n)$

(2) $Var(Y) \approx \sum_{i=i}^{n} \left(\frac{\partial f}{\partial X_i} \right)^2 \cdot \sigma_i^2 + 2 \sum_{i<j} \frac{\partial f}{\partial X_i} \cdot \frac{\partial f}{\partial X_j} \cdot \sigma_{ij}$

der de deriverte evalueres for forventningsverdiene.

a) Forklar at formlene gjelder eksakt dersom funksjonen f er lineær.

La $Y = f\,(X_1,X_2,X_3) = X_1 + X_2 \cdot X_3$.

b) Bruk formlene (1) og (2) til å finne et tilnærmet uttrykk for E(Y) og Var(Y).