

NORGES HANDELSHØYSKOLE

Eksamen vårsemesteret 2009

Kurskode: INT010 Tittel: Anvendt metode

Dato: 25.05.2009 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

Anta at ett-års overgangssannsynligheten mellom det å være "i arbeid" og det å være "arbeidsløs" for unge menn i alderen 18-25 år kan uttrykkes ved overgangsmatrisen (**P**):

t \ t+1	i arbeid	arbeidsløs
	i arbeid	arbeidsløs
i arbeid	0,7	0,3
arbeidsløs	0,4	0,6

Merk at for tilstanden "i arbeid" menes de som er yrkesaktiv og de som studerer.

- Begrunn at **P** er en overgangsmatrise. Hva er sannsynligheten for at en ung mann som er arbeidsløs ett år skal være i arbeid året etter? Hva er sannsynligheten for at en ung mann som er arbeidsløs ett år skal være i arbeid to år etter?
- Anta at det i utgangspunktet er 80 prosent unge menn i arbeid, mens de resterende 20 prosentene er arbeidsløse. Hvordan fordeler de unge mennene seg i de to tilstandene etter ett år?
- Finn den stasjonære fordelingen til overgangsmatrisen **P**. (Hint: Om vektoren $\pi = [\pi_1, \pi_2]$ er den stasjonære fordeling for hhv. tilstand 1 (i arbeid) og tilstand 2 (arbeidsløs), da er $\pi_1 + \pi_2 = 1$.)

Oppgave 2

I en studie for å avdekke om det er lønnsforskjeller blant daglige ledere ansatt i selskaper lokalisert i ulike geografiske områder i London er det samlet inn data for 27 ledere på østsiden og 27 ledere på vestsiden, der lønn er oppgitt i tusen pund per år. Gjennomsnittslønn på østsiden (\bar{X}_1) er lik 60,9. På vestsiden er gjennomsnittslønn (\bar{X}_2) lik 70,2. Det oppgis at standardavvikene er $S_1 = 8,7$ og $S_2 = 7,6$.

La μ_1 og μ_2 betegne forventningsverdiene til lønn for daglige ledere på hhv. østsiden og vestsiden av London.

- Test hypotesen $H_0: \mu_1 - \mu_2 = 0$ mot $H_1: \mu_1 - \mu_2 < 0$, med antagelsen $\sigma_1 = \sigma_2$, på 5% signifikansnivå. Kommenter.

Et alternativ til testen i a) er å benytte en ikke-parametrisk test, Wilcoxon's test for to utvalg. Det oppgis at testobservatorens verdi (rangsummen) er $T = 537,5$.

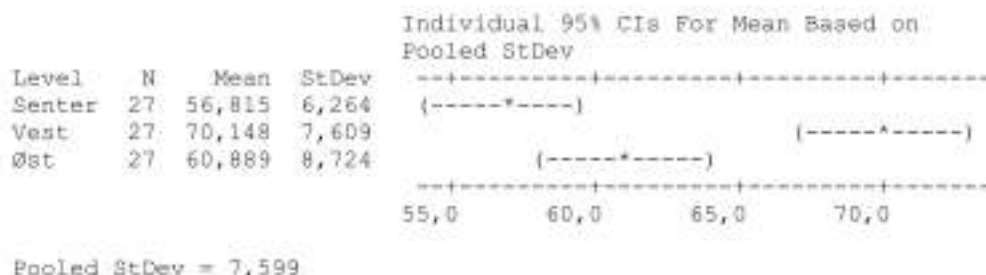
- Beregn forventning og standardavvik til testobservatoren T , dvs. beregn $E(T)$ og σ_T . Hva blir konklusjonen på hypotesetesten H_0 : Populasjonene 1 og 2 har lik lokalisering mot H_1 : Populasjon 1 ligger til venstre for populasjon 2 (dvs. tilsvarende hypotese som i a)) på 5% signifikansnivå? Regn ut p-verdien til testen.

Studien utvides ved å samle inn lønnsdata fra 27 ledere i sentrale London. Under er utskrift av en en-veis variansanalyse fra Minitab.

One-way ANOVA: Lønn versus Område

Source	DF	SS	MS	F	P
Område	2	2521,0	1260,5	21,83	0,000
Error	78	4504,1	57,7		
Total	80	7025,1			

S = 7,599 R-Sq = 35,89% R-Sq(adj) = 34,24%



- c) Sett opp nullhypotesen og alternativhypotesen for denne ANOVA. Hva blir konklusjonen av testen på 5% signifikansnivå? Kommenter.

Det er også gjennomført noen multiple sammenligninger tilhørende ANOVA'en over. Både Tukeys og Fishers metode er benyttet. Minitab-utskriften vises under.

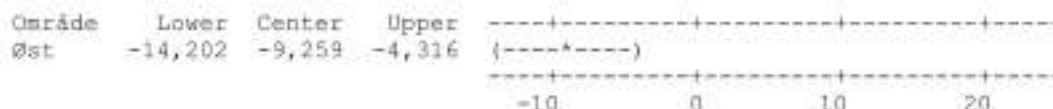
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Område

Individual confidence level = 98,07%

Område = Senter subtracted from:



Område = Vest subtracted from:



Fisher 95% Individual Confidence Intervals
All Pairwise Comparisons among Levels of Område

Simultaneous confidence level = 87,88%

Område = Senter subtracted from:

Område	Lower	Center	Upper
Vest	9,216	13,333	17,451
Øst	-0,043	4,074	8,192

Område = Vest subtracted from:

Område	Lower	Center	Upper
Øst	-13,377	-9,259	-5,142

d) Tolk kort resultatene fra disse parvise sammenligningene.

I tillegg til innsamlete lønnsdata har man kun samlet inn informasjon om ledere som arbeider i tre bestemte bransjer, dvs. bransje er en faktor med tre nivåer. Det er derfor gjennomført en to-veis variansanalyse, og utskrift fra Minitab er vist under.

Two-way ANOVA: Lønn versus Område; Bransje

Source	DF	SS	MS	F	P
Område	2	2520,99	1260,49	50,64	0,000
Bransje	2	2499,43	1249,72	50,21	0,000
Interaction	4	212,72	53,18	2,14	0,085
Error	72	1792,00	24,89		
Total	80	7025,14			

S = 4,989 R-Sq = 74,49% R-Sq(adj) = 71,66%

e) Hva er konklusjonene på de tre F-testene gjennomført i denne to-veis ANOVA om du benytter 5% signifikansnivå? Kommenter.

Oppgave 3

En analytiker i filmsektoren ønsker å estimere brutto inntekter generert av en film basert på ulike forklaringsvariabler. Analytikeren vurderer tre forklaringsvariabler; produksjonskostnader, markedsføringskostnader og en indikatorvariabel som er lik 1 om filmen er basert på en bok og 0 om ikke. Responsvariabelen og de to kostnadsvariablene er målt i millioner dollar. Datasettet er oppgitt i vedlegget.

Analytikeren starter med en enkel regresjonsmodell med kun produksjonskostnader som forklaringsvariabel. Minitab gir følgende utskrift:

Regression Analysis: Brutto inntekter versus Produksjonskostnader

The regression equation is
Brutto inntekter = 5,07 + 5,53 Produksjonskostnader

Predictor	Coef	SE Coef	T	P
Constant	5,071	5,949	0,85	0,405
Produksjonskostnader	5,5266	0,7493	7,38	0,000

S = 9,50616 R-Sq = 75,1% R-Sq(adj) = 73,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4916,3	4916,3	54,40	0,000
Residual Error	18	1626,6	90,4		
Total	19	6542,9			

Unusual Observations

Obs	Produksjonskostnader	Brutto inntekter	Fit	SE Fit	Residual	St. Resid
8	10,8	45,00	64,76	3,31	-19,76	-2,22R

R denotes an observation with a large standardized residual.

- a) Ta stilling til utsagnet: "Hvis et filmselskap øker produksjonskostnadene med én million dollar kan de forvente at bruttoinntekten øker med 5,53 millioner dollar."

Analytikeren ønsker å benytte den estimerte regresjonsmodellen til prediksjonsformål. Du får oppgitt at gjennomsnittet for de observerte produksjonskostnadene er 7,4 millioner dollar (dvs. $\bar{X} = 7,4$) og at standardavviket for produksjonskostnadene er 2,9 (dvs. $S_x = 2,9$).

- b) Beregn et 95% prediksjonsintervall for bruttoinntekter for en gitt film med verdi av produksjonskostnader lik 9,9 millioner dollar.
- c) Beregn et 95% konfidensintervall for forventningen til bruttoinntekter for mange filmer med verdi av produksjonskostnader lik 9,9 millioner dollar. Hvorfor er intervallene i b) og c) forskjellige?

Analytikeren inkluderer så de to andre forklaringsvariablene; markedsføringskostnader og indikatorvariabelen bok. Utskrift fra Minitab er gitt under.

Regression Analysis: Brutto inntekter versus Produksjonskostnader; Markedsføringskostnader; Bok

The regression equation is

Brutto inntekter = 7,84 + 2,85 Produksjonskostnader + 2,28 Markedsføringskostnader + 7,17 Bok

Predictor	Coef	SE Coef	T	P
Constant	7,836	2,333	3,36	0,004
Produksjonskostnader	2,8477	0,3923	7,26	0,000
Markedsføringskostnader	2,2782	0,2534	8,99	0,000
Bok	7,166	1,818	3,94	0,001

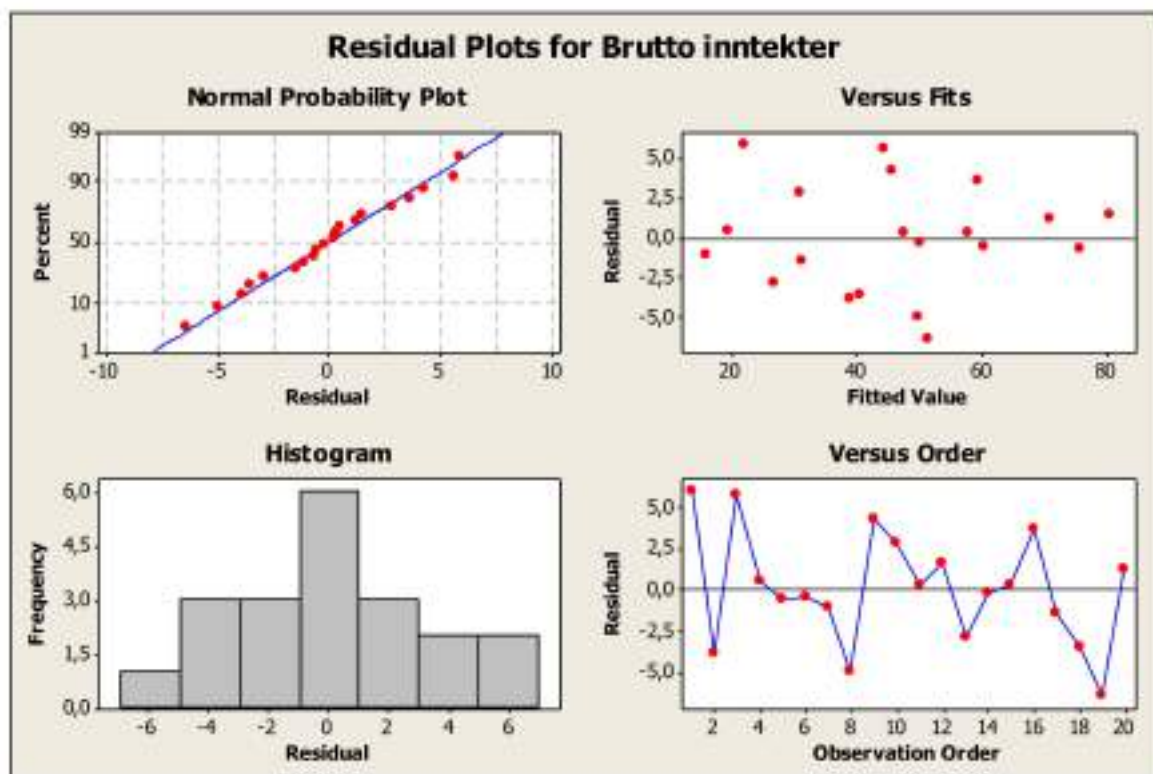
S = 3,68950 R-Sq = 96,7% R-Sq(adj) = 96,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	6325,2	2108,4	154,89	0,000
Residual Error	16	217,8	13,6		
Total	19	6542,9			

Source	DF	Seq SS
Produksjonskostnader	1	4916,3
Markedsføringskostnader	1	1197,3
Bok	1	211,5

- d) Drøft kort hva vi lærer av den estimerte regresjonsmodellen ovenfor (altså den utvidede modellen).
- e) F-verdien for regresjonen er 154,89. Formuler null og alternativhypotesen tilhørende denne F-verdien. Hva er antall frihetsgrader i testen? Hva er kritisk verdi på 5% nivå? Kan nullhypotesen forkastes?
- f) Analytikeren har studert noen diagnoseplott, se under. Forklar hvilken informasjon de ulike plottene gir. Eventuelt, hvilke konsekvenser har dette for tolkningen av regresjonen?



- g) Hvordan må analytikeren gå frem for å beregne et 95% prediksjonsintervall for en gitt film, dvs. vi har gitt verdiene av de tre forklaringsvariablene i den utvidede regresjonsmodellen? Du skal ikke beregne intervallet, kun kort beskrive fremgangsmåten.

Vedlegg til oppgave 3.

Datasett

Bruttoinntekter i mill. \$	Produksjonskostnader i mill. \$	Markedsføringskostnader i mill. \$	Basert på bok 1 – ja, 0 – nei
28	4,2	1,0	0
35	6,0	3,0	1
50	5,5	6,0	1
20	3,3	1,0	0
75	12,5	11,0	1
60	9,6	8,0	1
15	2,5	0,5	0
45	10,8	5,0	0
50	8,4	3,0	1
34	6,6	2,0	0
48	10,7	1,0	1
82	11,0	15,0	1
24	3,5	4,0	0
50	6,9	10,0	0
58	7,8	9,0	1
63	10,1	10,0	0
30	5,0	1,0	1
37	7,5	5,0	0
45	6,4	8,0	1
72	10,0	12,0	1