

NORGES HANDELSHØYSKOLE

Eksamen høst 2011

Kurskode: INT010 Tittel: Anvendt metode

Dato: 2.11.2011 Kl. 09.00-12.00

Faglærer går ikke rundt i eksamenslokalene, men kan kontaktes av eksamensvakten på
tlf. 59248/924 27 289.

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

I Aftenpostens nettutgave 26.4.2011 er det en artikkel som undersøker hvor mye det koster å forsikre en ganske vanlig familiebil i de fem største byene i Norge. Avisen har bedt seks forsikringsselskaper gi tilbud på et spesifikt forsikringstilfelle i hver by. Dataene er gjengitt under.

Bilforsikring Oslo Kristiansand Stavanger Bergen Trondheim						
			DnB NOR			
Oslo	8 222	12 110	10 577	12 133	9 291	7 810
Kristiansand	6 783	11 115	8 348	11 383	7 718	7 123
Stavanger	5 494	11 368	7 132	9 371	8 920	6 855
Bergen	6 391	10 255	6 929	9 838	7 292	5 943
Trondheim	7 854	11 945	8 550	11 195	7 451	7 029

Aftenposten omtalte tallene som følger:

Tabellen viser hva de enkelte selskapene forlanger for eksempelet vi har satt opp. Om vi legger sammen og finner gjennomsnittsprisen på byene, får vi følgende resultat:

Bergen: 7775

Stavanger: 7857

Kristiansand: 8745

Trondheim: 9004

Oslo: 10023

Det er altså Bergen som er den billigste byen å forsikre bilen sin i, over 1000 kroner i snitt billigere enn Kristiansand og drøye to tusen billigere enn Oslo.

Artikkelen forstetter med at representanter for selskapenes forklarer hvorfor prisene varierer mellom ulike byer. Forskjellene forklares generelt med variasjoner i skadefrekvens, noe som igjen kan relateres til bakenforliggende faktorer som trafikkmengde, veikvalitet og klima.

Du har et studentengasjement i NAF og blir bedt om å analysere tallene nærmere. Du starter med en enveis variansanalyse og får følgende resultater:

One-way ANOVA: Pris versus By

Source	DF	SS	MS	F	P
By	4	20475551	5118888	1,30	0,298
Error	25	98588734	3943549		
Total	29	119064284			

S = 1986 R-Sq = 17,20% R-Sq(adj) = 3,95%

a) Hvilken hypotese er det som testes og hva forteller analysene?

Artikkelen i Aftenposten har hovedfokus på prisforskjellen mellom byer. Du blir også bedt om å undersøke om det er systematiske prisforskjeller mellom forsikringsselskapene. Du gjør en ny enveis variansanalyse, denne gang med de ulike selskapene som faktor. Du gjør dessuten Tukeys test for parvise sammenligninger.

One-way ANOVA: Pris versus Selskap

Source	DF	SS	MS	F	P
Selskap	5	93089432	18617886	17,20	0,000
Error	24	25974852	1082285		
Total	29	119064284			

S = 1040 R-Sq = 78,18% R-Sq(adj) = 73,64%



Pooled StDev = 1040

Grouping Information Using Tukey Method

Selskap	N	Mean	Grouping
2	5	11359	A
4	5	10784	A
3	5	8307	B
5	5	7734	B
6	5	6952	B
1	5	6949	B

Means that do not share a letter are significantly different.

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

Individual confidence level = 99,50%

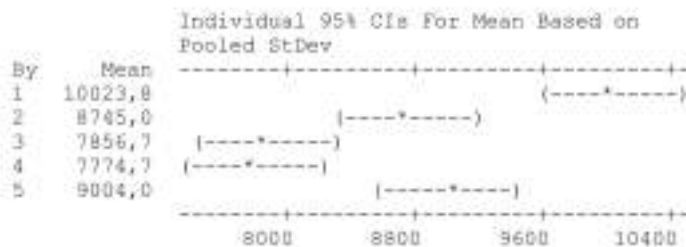
- Hva forteller variansanalysen oss? (Øvre del av utskriften)
- Hva forteller Tukeys test? (Nedre del av utskriften)
- Forklar opplysningene som gis i de tre nederste linjene av utskriften om konfidensnivåer.

Du avslutter analysen din med å gjøre en toveis variansanalyse der du åpner for at det både kan være prisforskjeller mellom byer og mellom forsikringsselskaper.

Two-way ANOVA: Pris versus By; Selskap

Source	DF	SS	MS	F	P
By	4	20475551	5118888	18,62	0,000
Selskap	5	93089432	18617886	67,71	0,000
Error	20	5499301	274965		
Total	29	119064284			

S = 524,4 R-Sq = 95,38% R-Sq(adj) = 93,30%



- e) Endrer dette noen av dine tidligere konklusjoner? Hvordan er det mulig at konklusjonene kan endre seg ved å gå fra enveis- til toveis variansanalyse?
- f) Hvordan måtte datainnsamlingen være lagt opp for at du skulle kunne undersøke om det er samspillseffekter mellom byer og forsikringselskaper, og hva ville være tolkningen av en eventuell samspillseffekt?

Oppgave 2

Fastlegepraksiser omsettes ikke fritt og man får derfor ikke testet ut den reelle betalingsvilligheten i markedet gjennom budrunder. Overdragelse foregår ved at kommunen først ansetter en ny fastlege når en liste blir ledig. Deretter forhandler den nye og gamle legen om verdien av praksisen. De materielle verdiene er vanligvis små og lette å verdsette, men de immaterielle verdiene er større og vanskeligere å verdsette. De immaterielle verdier kalles "verdien av opparbeidet praksis" og ligger i å overta en praksis som allerede er i drift med legesekretær ansatt, husleieavtale og inventar på plass, journalsystem med oversikt over pasientenes diagnoser og faste medisiner, gjennomarbeidede kvalitetssikringsrutiner, mm. Disse elementene gjør at den nye legen er operativ fra første dag, og jo høyere kvaliteten er, desto mer effektivt kan den nye legen arbeide. Dette har åpenbart stor økonomisk verdi.

På hjemmesiden til Den norske legeforening ligger det oversikt over prisen på en del praksisovertagelser i årene 2006 til 2011. Foruten den avtalte verdien av opparbeidet praksis inneholder datasettet følgende informasjon: Pris for inventar, salgsdato, kommune, listestørrelse (antall pasienter) og om praksisen er del av en gruppepraksis eller ikke. Dette er få variabler, men Legeforeningen ønsker å utforske hvor godt de kan predikere verdivariabelen for de omsatte praksisene. Målsettingen er å utvikle en modell som kan være til hjelp ved framtidige verdifastsettelse.

Med utgangspunkt i informasjonen på Legeforeningens hjemmeside lager du følgende datasett:

	Gjennomsnitt	Standardavvik	Min	Median	Max	Antall obs.
Verdi	575 321	291 108	27 313	564 599	1 245 823	52
Salgsmd	35,94	15,77	9	32	64	52
Listestørrelse	1110,7	345,1	106	1200	1700	52
Listestørrelse ²	1 350 421	708 582	11 236	1 440 000	2 890 000	52
Gruppepraksis	0,7308	0,4479	0	1	1	52
Inventar	105 614	118 292	0	71 708	546 351	52
Inventar_null	0,3269	0,4737	0	0	1	52
Befolkning	138 587	209 735	4852	28 781	599 230	52
(Befolk/1000) ²	62 348	129 042	23	828	359 076	52
StorOslo	0,2308	0,4254	0	0	1	52

Alle kroneverdier er deflatert med konsumprisindeksen til juni 2011-priser. Salgsmånedene er nummerert slik at januar 2006 er måneden nummer 1. *Gruppepraksis* er en dummyvariabel som markerer om praksisen inngår i et kontorfelleskap med andre leger. *Inventar_null* er en dummyvariabel for at prisen på inventaret er satt til null eller ikke fylt ut. *Befolkning* er kommunens innbyggertall i 2011 og hentet fra statistikkbanken i SSB, Tabell 05212. $(Befolk/1000)^2$ er befolkningstallet i hele tusen kvadrert. *StorOslo* er en dummyvariabel som markerer at kommunen ligger innenfor Stor-Oslo slik denne regionen er definert i SSB Notater 2009/24. Dummyvariabelen er satt til null for Oslo selv. Det er altså en markør for de mindre kommunene rundt Oslo.

Du kjører en første regresjon som vist under:

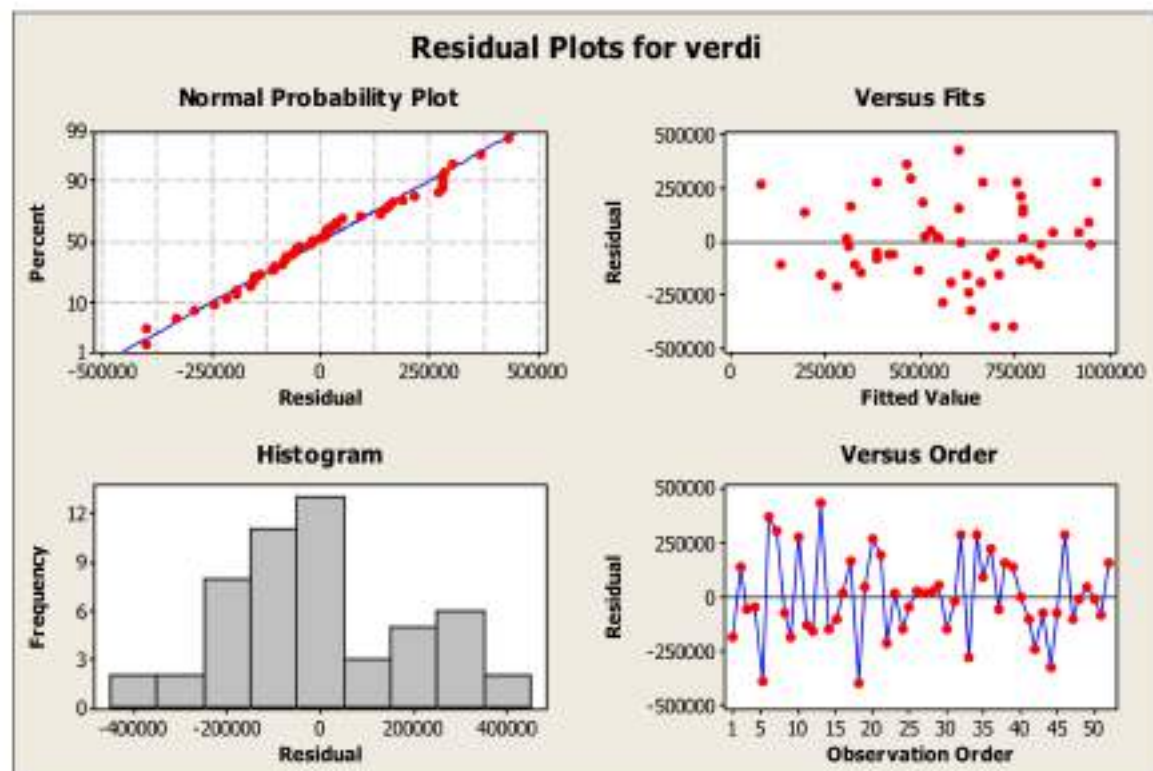
The regression equation is

verdi = - 300582 + 571 listestørrelse + 4267 salgsmnd + 121144 gruppepraksis

Predictor	Coef	SE Coef	T	P
Constant	-300582	118048	-2,55	0,014
Listestørrelse	570,83	81,81	6,98	0,000
salgsmnd	4267	1808		0,022
gruppepraksis	121144	63795	1,90	0,064

S = 198485 R-Sq = 56,2% R-Sq(adj) = 53,5%

Durbin-Watson statistic = 2,38712



- Gruppepraksis er en dummyvariabel. Hva forteller det oss at den har gjennomsnittsverdi 0,7308?
- Kommenter regresjonsutskriften og diagnoseplottene.
- Regn ut T-verdien til "salgsmnd" (Den svarte firkanten i utskriften.)
- Regn ut et 95 prosent konfidensintervall for koeffisienten til listestørrelse.
- Hva er estimert $\text{Var}(\text{verdi} | \text{listestørrelse}, \text{salgsmnd}, \text{gruppepraksis})$?
(Altså; den estimerte variansen til venstresidevariabelen betinget på høyresidevariablene i regresjonen.)

I tabellen på neste side er det rapportert en serie regresjoner der du utforsker spesifikasjonen nærmere med tanke på å komme fram til en best mulig modell.

Verdien av opparbeidet praksis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listestørrelse	571*** (81.8)	647* (346.0)	544*** (78.8)	544*** (78.6)	514*** (73.4)	508*** (74.0)	515*** (72.3)	524*** (72.9)
Listestørrelse ²		-0.038 (0.170)						
Salgsmd	4267** (1808)	4293** (1830)	4836*** (1752)	4580** (1725)	4306*** (1599)	4493*** (1607)	4156** (1581)	3933.5** (1601)
Gruppepraksis	121144* (63795)	122541* (64727)	143897** (61572)	138118** (61100)	95844 (58241)			95293 (57566)
Inventar			-0.263 (0.295)					
Inventar_null			-182775** (74634)	-139962** (57103)	-91041 (55140)			-63467 (57780)
Befolkning					1.744*** (0.639)	2.194*** (0.620)	2.501*** (0.629)	2.075*** (0.673)
(Befolk/1000) ²					-2.455** (1.028)	-3.123*** (1.006)	-3.514*** (1.006)	-2.891*** (1.060)
StorOslo							113228* (62670.0)	94135 (65516)
Konstantledd	-300582** (118047)	-335728* (195284)	-219709* (118948)	-248203** (114348)	-278734** (106258)	-259570** (98785)	-300070*** (99057)	-326346*** (110126)
N	52	52	52	52	52	52	52	52
adj. R ²	0.535	0.526	0.577	0.579	0.640	0.619	0.637	0.648
S	198.485	200.476	189.284	188.877	174.769	179.593	175.417	172.738

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

N is the number of observations. S in the last row is the estimated standard error of the residual.

f) Kommenter kort hva vi lærer ved å gå fra kolonne (1) til (8).

g) Bruk modell (6) til å predikere verdien av en praksis med følgende karakteristika:

Values of Predictors for New Observations

New Obs	listestørrelse	salgsmd	befolkning	befolkning2
1	600	64,0	260392	67803

Dette er en praksis i Bergen med 600 pasienter som selges i april 2011.

Du bruker Minitab til å sjekke prediksjonen din og får følgende output:

New Obs	Fit	SE Fit	95% CI	95% PI
1	██████	97880	(495296; 889115)	(280736; 1103674)

h) Er det CI eller PI som er mest relevant for å bedømme usikkerheten omkring prediksjonen i spørsmål g)? Begrunn svaret.

i) Anta at praksisen som skal verdsettes ligger på 90-prosentpersentilen mht. verdi utover det man vil forvente basert på de karakteristikaene som inngår i regresjonsmodellen. Hvordan vil du da verdsette praksisen?

Oppgave 3

I en studie av kursendringer på børsnoterte aksjer benyttes ofte den såkalte lognormale fordelingen som modell. At forholdet, Y , mellom aksjekursen ved slutten og begynnelsen av en periode er lognormalt fordelt, betyr at logaritmisk avkastning, $X = \ln Y$, er normalfordelt. Anta derfor at X er fordelt $N(\mu, \sigma^2)$ og at $Y = e^X$.

Det kan vises at $EY = Ee^X = e^{\mu + \frac{1}{2}\sigma^2}$.

- Regn ut EY når $\mu=0,20$ og $\sigma=0,20$.
- Regn også ut μ gitt at $EY=1,2$ og $\sigma=\mu$.
- Regn ut $P(Y>1)$ når $\mu=0,20$ og $\sigma=0,20$. Hva forteller denne sannsynligheten oss?

Avkastningen over to perioder kan uttrykkes som $Y_{(2)}=Y_1 \cdot Y_2$

- Hva er fordelingen til $Y_{(2)}$ hvis Y_1 og Y_2 er uavhengige og identisk, lognormalt fordelte med parametre μ og σ som ovenfor.
- Regn ut $P(Y_{(2)}>1)$ når $\mu=0,20$ og $\sigma=0,20$.