

Løsningsforslag INT010 høst 2011

Oppgave 1

- a) H_0 : Forventet pris på forsikring er lik i de fem byene.
 H_1 : Minst en by har forventning forskjellig fra de andre.
Vi kan ikke forkaste hypotesen om lik forventning.
- b) Analysen forteller oss at det er klart signifikante forskjeller i pris mellom selskapene.
- c) Tukeys test forteller oss at selskap 2 og 4 (Gjensidige og Tryg) er signifikant dyrere enn selskap 1, 3, 5 og 6 (If, DnB Nor, Storebrand og Nemi). Det er ikke signifikante prisforskjeller mellom selskapene innenfor hver av disse to gruppene.
- d) Testen har 0,5 prosent sannsynlighet for feilaktig å påstå at det er signifikant prisforskjell mellom to bestemte selskaper. Testen har 5 prosent sannsynlighet for feilaktig å påstå signifikant forskjell mellom minst to selskaper blant alle de som sammenlignes.
- e) Toveisanalysen forteller oss at det både er signifikante prisforskjeller mellom byer og mellom selskaper. Prisforskjellen mellom byene trer tydeligere fram enn i enveisanalysen fordi vi får kontrollert for systematisk prisvariasjon som skyldes forskjeller mellom selskapene.
- f) For å undersøke samspillseffekter måtte vi ha samlet inn minst to forsikringspriser for hver kombinasjon av selskap og by. Tolkningen av en samspillseffekt ville være at det var stedsspesifikke prisforskjeller mellom selskapene. For eksempel kunne Gjensidige generelt være et dyrt selskap, og Oslo generelt en dyr by, men Gjensidige kunne likevel være billigere i Oslo enn det man ellers ville forvente ut fra prisene i Oslo og prisene til Gjensidige.

Oppgave 2

- a) Gjennomsnitt lik 0,7308 forteller oss at ca 73 % av de omsatte praksisene er gruppepraksiser.
- b) Vi ser at både listestørrelse og salgsmåned har signifikant forklaringskraft og at dummyen for gruppepraksis er nær ved å være signifikant på 5 %-nivå.

Predikert verdi av en praksis øker med ca 570 kroner per pasient på listen, alt annet likt.

Den gjennomsnittlige prisveksten for en praksis med gitte karakteristika har vært nærmere 4300 kroner per måned over de årene vi har data for. Tar vi utgangspunkt i gjennomsnittsverdien for omsatte praksiser som vi finner i den deskriptive tabellen representerer dette en årlig prisvekst i størrelsesorden 9 prosent utover den alminnelige prisstigningen som alt er hensyntatt ved at tallene er deflatert med konsumprisindeksen.

Gruppepraksiser omsettes i gjennomsnitt for 120 000 kroner mer enn solopraksiser.

“Bokstavlig tolket” predikerer konstantleddet verdien for en solopraksis omsatt i desember 2005. At konstantleddet er negativt er det imidlertid ingen grunn til å legge vekt på da listestørrelse null ligger godt utenfor det intervallet vi har data for. Den negative verdien forteller oss bare at vi ikke kan predikere lineært ut av observasjonsområdet.

Justert R^2 er 53 %. Det må sies å være relativt bra for en modell med bare tre variabler, men neste halvparten av variasjonen i verdien må altså forklares med andre faktorer enn listestørrelse, omsetningstidspunkt og hvorvidt det er gruppe- eller solopraksis.

Standardavviket til feilleddet (S) kvantifiserer at det gjennomsnittlige avviket fra predikert verdi i utvalget er i nesten 200 000 kroner. Det er konsistent med at det er viktige verdidrivere som modellen ikke fanger opp.

Vi ser fra Durbin-Watsonobservatoren at det ikke er vesentlige problemer med autokorrelasjon siden verdien ligger relativt nær 2. Dette bekreftes også av at det ikke er noen påfallende mønstre i residualplottet. De øvrige residualplottene gir heller ikke indikasjoner på at det er store avvik fra normalfordeling i feilleddene eller at det er alvorlige problemer med heteroskedastisitet.

[Alle disse poengene trenger naturligvis ikke være med for å få full uttelling, men kandidaten må vise at vedkommende forstår den økonomiske fortolkningen av utskriften og den statistiske fortolkningen av residualplottene.]

- c) $T = 4267/1808 = 2,36$.
- d) 95 % konfidensintervall for $b^{\text{listestørrelse}} = 570,83 \pm 2,01 \cdot 81,81 \approx 571 \pm 164$ der 2,01 er en kritisk grense hentet fra T-tabell med 48 frihetsgrader. Vi ser altså at det er relativ stor usikkerhet omkring verdiøkningen per pasient på listen.
- e) $\text{Var}(\text{verdi} | \text{listestørrelse} \text{ salgsmnd gruppepraksis}) = S^2 = 198\,485^2$ (Ca 39,4 milliarder!)
- f) Fra kolonne (2) ser vi at linearitet i listestørrelse er en svært god tilnærming i intervallet vi har data for. Fra kolonne (3) ser vi at verdien som betales for medfølgende inventar ikke predikerer verdien av opparbeidet praksis, utover at praksiser der det av en eller annen grunn ikke medfølger inventar prises klart lavere enn andre praksiser med ellers like karakteristika. Fra kolonne (5) ser vi at variabelen inventar_null, drøftet ovenfor, mister mye av sin forklaringskraft når det kontrolleres for kommunens befolkningsgrunnlag. Det må altså være i små kommuner at praksiser omsettes uten medfølgende inventar. (Det som her fanges opp er at utkantkommuner som sliter med å rekruttere leger ofte stiller legekantor til disposisjon for legene.) Fra kolonne (7) ser vi at praksiser i små kommuner innenfor Stor-Oslo-regionen omsettes for høyere priser enn praksiser i kommuner med samme antall innbyggere som ikke ligger i Stor-Oslo. Estimaten er marginalt signifikant. I kolonne (9) kjøres regresjonen på nytt med to ikke-signifikante variabler inne. Til tross for manglende signifikans øker justert R^2 . Det virker derfor rimelig å anta at dette er reelle verdidrivere, men at utvalget vårt er for lite til å få presise estimater. Vi ser imidlertid at de estimatene som modellen gir, er økonomisk relevante. Kroner 95 tusen ekstra for en gruppepraksis, 63 000 i fratrekk for en praksis der det ikke medfølger inventar og 94 000 ekstra for en praksis i “Oslo-nære” områder.
- g) Predikert verdi er

$$\text{verdi} = -259570 + 508 \cdot 600 + 4493 \cdot 64 + 2,194 \cdot 260392 - 3,123 \cdot 67803 = \underline{692\,333}$$

- h) Prediksjonsintervallet er mest relevant. Det fanger med 95 % sikkerhet opp enkeltstående nye verdier av Y-variabelen for gitte verdier av X-variablene. Dette er videre enn CI siden det omfatter både usikkerhet i estimeringen av EY og at realisert Y-verdier varierer omkring EY.
- i) Vi antar feilleddet er normalfordelt rundt regresjonslinjen, og det estimerte standardavviket i modell 6 er 179593. 90-prosentpersentilen i denne fordelingen ligger derfor på $1,2816 \cdot 179593 = 229\,879$. (1,2816 er 90-prosentpersentilen for en standard normalfordelt variabel.)
- Vårt estimat for verdien av praksisen er da $692\,333 + 229\,879 \approx \underline{922\,000}$.

Oppgave 3

- a) $e^{0,20+0,5 \cdot 0,20 \cdot 0,20} = e^{0,20+0,02} = e^{0,22} \approx 1,25$.
- b) $1,2 = e^{\mu + \frac{1}{2}\mu^2} \Leftrightarrow \ln 1,2 = \mu + \frac{1}{2}\mu^2 \Leftrightarrow \mu^2 + 2\mu - \ln 1,44 = 0 \Leftrightarrow \underline{\mu \approx 0,168}$
- c) $P(Y > 1) = P(\ln Y > \ln 1) = P(X > 0) = 1 - P(X < 0) = 1 - G[(0 - 0,2)/0,2] = 1 - G(-1) = G(1) \approx \underline{0,84}$.
Fortolkningen er at det er 84 % sannsynlighet for positiv avkastning.
- d) $Y_{(2)} = Y_1 \cdot Y_2 \Leftrightarrow \ln Y_{(2)} = \ln Y_1 + \ln Y_2 = X_1 + X_2$ som er normalfordelt $N(2\mu, 2\sigma^2)$. Da er $Y_{(2)}$ lognormalt fordelt med forventning $e^{2\mu + \sigma^2}$.
- e) $P(Y_{(2)} > 1) = P(\ln Y_{(2)} > \ln 1) = P(X_1 + X_2 > 0) = 1 - P(X_1 + X_2 < 0) = 1 - G[(0 - 0,4)/(\sqrt{2} \cdot 0,2)] = 1 - G(-\sqrt{2}) = G(\sqrt{2}) = G(1,41) \approx \underline{0,92}$.