

Ekstaoppgave i STK1110 høsten 2007

Oppgave E6

Denne oppgaven omhandler boligmarkedet i Oslo. Våren 2000 ble det fra Aftenpostens boligannonser registrert følgende opplysninger om 100 leiligheter:

- Prisantydning/verditakst (i 1000 kr)
- Leilighetens areal i kvadratmeter
- Antall rom i leiligheten
- Månedlig husleie (i kroner)
- Om det er balkong tilknyttet leiligheten (1=ja, 0=nei)
- Om det er garasje tilknyttet leiligheten (1=ja, 0=nei)
- x -koordinat (i km, retning øst)
- y -koordinat (i km, retning nord)

x - og y -koordinatene angir hvor i Oslo leilighetene ligger. Origo i koordinatsystemet er plassert på Bygdøy, mens det nordøstlige hjørnet for de leilighetene som er registrert ligger på Grefsen.

Dataene er gitt i filen "leiligheter.txt" på kurssets hjemmeside. Hver linje i filen gir opplysninger om en leilighet, mens kolonnene i filen svarer til variablene beskrevet ovenfor.

Dataene skal analyseres med multippel lineær regresjon (bruk MATLAB), der responsvariabelen er prisantydning/verditakst (unøyaktig referert til som pris) og de andre variablene er forklaringsvariable.

a) Tilpass en multippel lineær regresjonsmodell med alle forklaringsvariablene. Forklar hva de estimerte regresjonsparameterne forteller deg om hvordan prisen avhenger av forklaringsvariablene.

b) Undersøk hvilke av forklaringsvariablene som har signifikant betydning for prisen. Tilpass en ny modell der du fjerner de forklaringsvariablene som ikke er signifikante, og gi en fortolkning av den modellen du nå kommer fram til.

Du bør fjerne én forklaringsvariabel om gangen. Først fjerner du den forklaringsvariabelen som har størst P -verdi (forutsatt at P -verdien er større enn 5%) og tilpasser en regresjonsmodell uten denne forklaringsvariabelen. Så fjerner du den forklaringsvariabelen som nå har størst P -verdi (forutsatt at P -verdien er større enn 5%) og tilpasser en regresjonsmodell uten denne forklaringsvariabelen. Slik fjerner du én og én forklaringsvariabel og tilpasser nye regresjonsmodeller helt til alle forklaringsvariablene i modellen er signifikante (har P -verdi mindre enn 5%).

c) Bruk ulike plott av residualene til å kontrollere om forutsetningene for den multiple lineære regresjonsmodellen er (rimelig godt) oppfylt for den modellen du kom fram til i punkt b.