

1 Opgave Brugte biler

1.1 Data

Opgavens datagrundlag er de annoncerede priser på brugte biler i "Søndagsavisen" på to søndage i oktober 1999. Til datasættet udvalgt bilmodellerne VW Golf og Ford Escort.

I datasættet optræder følgende variable:

- FABR angiver fabrikatet
- MODEL angiver modellen
- TYP angiver typen
- AARL angiver fabrikationsåret
- ALDER angiver alderen i hele år beregnet ud fra fabrikationsåret
- KM angiver kilometerstanden
- PRIS angiver den annoncerede pris
- LPRIS angiver logaritmen til den annoncerede pris

Data ligger i datasættet `SASDATA.BRUGTBIL`, i SAS-biblioteket `SASDATA` i kursusbiblioteket

1.2 Opgave

Vurder relationen mellem pris og alder (i år og/eller i kilometer) for de to bilmærker.

De, der har haft opgaven før, ved godt, at det er smart at kigge på logaritmen til prisen ; det gider vi ikke at gøre et stort nummer ud af i denne opgave, som er en træning i GLM. Men man bør dog lige overbevise sig om, at en lineær regressionsmodel log pris mod alder både giver bedre mening, (fast procentisk afskrivning), og passer bedre til data (for det ene af bilmærkernes vedkommende ser man tydeligt at hvis man tegner regressionen pris mod alder, ligger observationer *over linien* til at begynde med, derefter *under linien* for slutteligt at ligge *over linien* igen - og så er modellen jo ikke god).

SAS-opgaven går ud på at gennemføre og dokumentere (i passende omfang) analysen ved at starte med den store model (med bilmærkeafhængigt intercept, - koefficient til alder og evt til kmstand), gennemføre (og kommentere) tests for modelreduktion.

Lærerens yndlingsmodel for dette datasæt er en model, der har forskellig "intercept" for hvert bilmærke, fælles afskrivningsprocent - og (pga materialets sparsomhed og multikollineariteten mellem alder og kilometer) ingen korrektion for kilometerstanden.

En god besvarelse vil netop undervejs i modelreduktionen have forholdt sig til multikollineariteten mellem alder og kilometer; dels bare ved at tegne dem op mod hinanden, dels ved at kigge på $(\mathbf{X}'\mathbf{X})^{-1}$ (eller, hvad der er det samme, kovariansen mellem estimatorer, specielt for alder og km). (Man kan da også kigge på VIF, selv om vi ikke har lært om den, men kun set om den på SAS's hjælpesider.

I anden del af opgaven skal man for et af bilmærkerne (fx Ford) og en bestemt alder (fx 7 år) bestemme den fittede værdi (på log-skalaen - regnestykket skrives i besvarelsen og usikkerheden (spredningen) på denne værdi - regnestykket skrives i besvarelsen. Hvis man har lyst, kan man også prøve at flytte fitted værdi og spredning op på krone -skalaen

Besvarelsen afleveres på papir - der bliver ikke givet ordens- eller stil-karakter. Besvarelsen skal dokumentere at man har fornemmelse for data-situationen, og at man kan bruge sit SAS til en sådan generel lineær model (den kaldes undertiden en kovariansanalyse) og at man har en vis forståelse for, hvad det er, SAS gør.