

Projekt nr. 2:

Norske busselskabers effektivitet.

Datamaterialet til denne opgave er stillet til rådighed af Torben Holvad, Jens Leth Hougaard, Dorte Kronborg og Hans Kurt Kvist. Datasættet indeholder driftsoplysninger vedrørende 1991 for 157 offentligt støttede busselskaber i Norge. Datasættet findes i form af tekstfilen 03-2-dat.txt som kan hentes fra kursets hjemmeside. Denne fil indeholder 157 linier, en for hvert selskab. Forklaringen på hvad de 22 oplysninger i hver linie betyder findes på filen var.pdf, som er en (ikke for køn) scannet udgave af en fotokopi (også vedhæftet den trykte version af opgaveteksten).

Datasættet har været brugt som eksempel i flere artikler om efficiensmål, benchmarking m.v. inden for økonomi. En af dem (af ovennævnte forfattere) findes på nettet som

<http://www.mes.cbs.dk/~stdk/bus1102ny.pdf>

Opgaven går ud på at undersøge, om disse økonomiske metoders konklusioner bekræftes eller afkræftes — eller overhovedet kan efterprøves — ved hjælp af almindelige statistiske metoder. Primært skal man altså diskutere og udføre de statistiske analyser, man anser for relevante til sammenligning af busselskabernes effektivitet i udnyttelsen af deres ressourcer. I anden omgang skal man så diskutere disse konklusioners overensstemmelse og sammenlignelighed med de konklusioner, der findes i den nævnte artikel.

Tekniske detaljer.

Ved optælling vil man se, at variabellisten kun indholder 21 variabelnavne. Det skyldes at den første oplysning i hver linie, som er en slags løbenummer (med “huller”, som formodentlig skyldes at visse selskaber er røget ud af en eller anden grund) ikke er nævnt.

Datafilen 03-2-dat.txt ser ret kaotisk ud, fordi den er genereret som tabulator-separeret tekstfil fra et Excel regneark, og derfor formatteret på en måde, som forskyder tallenes placering i søjlerne når de er for lange til at stå i én søjle. Men der er faktisk (som man opdager, hvis man prøver at indlæse data i en statistikpakke) præcis 22 tal i hver linie. For en ordens skyld er også Excel regnearket lagt ud under navnet 03-2-dat.xls.

For en af de 22 variable (PERS_KM) forekommer der manglende værdier. Disse er på datafilen kodet som NA (not available).

Rapporten afleveres fredag d. 19. december.