

Analys av diamanterpriser med multipel linjär regression

Oskar Söderlund*

Juni 2016

Sammanfattning

I det här examensarbetet undersöker vi vilka variabler som påverkar priset på diamanter. Målet med arbetet är att hitta en linjär modell, med hjälp av multipel linjär regression, som på bästa sätt förklarar variationen av priset på diamanterna. Det visar sig att sambandet mellan diamanternas pris och förklarande variabler ej kan uttryckas linjärt och vi använder oss istället av en multiplikativ modell härledd från en logaritmttransformation av responsvariabeln. Grundmodellen bestod av sju förklarande variabler som vi tror kan ha en påverkan på priset. Efter vi upptäckt att flertalet variabler haft stark korrelation har vi slagit samman dessa tre variabler till ett gemensamt medelvärde. Vi har även behövt använda en kvadratisk term av diamanternas carat för att få en bättre modellanpassning. Undersökning av modellens residualer plottade mot det predikterade värdet visade på att små och stora diamanter ej hade ett kontinuerligt samband vilket resulterade i en uppdelning av observationerna och totalt fyra stycken modeller testades för att finna de två mest lämpliga. De resulterande modellerna pekade mot att vikten var den variabel som mest förklarade prisets variation. Klarheten var mer inflytelserik än färgen för små diamanter och vice versa för stora. Diamanternas certifiering visade sig ha betydelse på priset även fast det inte troddes ha det.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: Oskar_soderlund@hotmail.com. Handledare: Jan-Olov Persson, Gudrun Brattström.