Project

Guidelines

1. Collaboration and (email) discussion are strictly discouraged. You may use any preexisting published resources for your paper, but be sure to cite all references your paper. To reiterate, the paper has to be your original work. Feel free to ask any questions that may arise your STOR455 instructor.
2. Late project will not be accepted.

1 Diamond pricing

You have been hired as a consultant by a large firm that sells diamonds. The firm has asked you to develop a model to predict the price of a diamond based on characteristics of the diamond.

Some background information on diamonds:
The weight of a diamond stone is indicated in terms of carat units. One carat is equivalent to 0.2 grams. All other things being equal, larger diamond stones command higher prices in view of their rarity. Being products of nature, diamonds have “birth marks" or inclusions only visible under a jeweler's magnifying glass (called a loupe) or a microscope. Diamonds with no inclusions under a loupe with a 10-power magnification are labeled IF (“internally flawless"). Lesser diamonds are categorized in descending order as “very very slightly imperfect" VVS1 or VVS2 and “very slightly imperfect" VS1 or VS2. The most prized diamonds display color purity. They are not contaminated with yellow or brown tones. Top color purity attracts a grade of D. Subsequent degrees of color purity are rated E, F, G, ... all the way down the alphabet ladder. The cut (or faceting) of a raw diamond stone relies on the experience and the craftsmanship of the diamond cutter. The optimal cut should neither be too deep nor too shallow for it will impede the trajectory of light and thereby the brilliance or “fire" of a diamond stone. All of the diamonds considered here are round cut.

To assist shoppers, independent certification bodies assay diamond stones and provide a certificate listing a stone's caratage, grade of clarity, color and cut. For the stones you will consider, three certification bodies were used: New York based Gemmological Institute of America (GIA), Antwerp based International Gemmological Institute (IGI) and Hoge Raad Voor Diamant (HRD). Their reputations could be a factor in the pricing of the diamond stones.

For more information on the factors that influence the price of a diamond, see the file “background.txt".

2 Your project

Data on 200 round-cut diamonds were collected including price (in US $), weight (in carats), color (either D, E, F, G, H or I) and clarity (either IF, VVS1, VVS2, VS1 or VS2), and certification body. The firm has asked you to create a sensible pricing model for diamond stones based on these characteristics. In addition to determining a sensible pricing model, the firm would like you to address the following issues:

1. Is the model that you have created useful? In particular, does the model have predictive power? Are the estimates of coefficients in your model sensible?
2. All other things being equal, what is the average price difference between a grade D diamond and one graded E?
3. All other things being equal are there price differences among the three certification bodies?

As part of your consultant duties, you need to write a short paper for summarizing your results for the diamond company. In your paper, describe the data, noting all important features and fully address the issues described above. The employees of the firm have had only a basic course in statistics so they don't understand technical terms like VIF, Cp, PRESSp, etc. This means that you need to briefly describe technical terms in the paper. For example, “We use the forward subset selection method to select a model. This is an automatic model search strategy used to select the predictors in a regression model."

3 Paper Format

- Your paper should include the following sections: introduction, analysis and results, conclusions, references (if any), appendices. The paper must be typed, doublespaced, and at least 10 point font. The body of the report (introduction through conclusions sections) must be no longer than 3 pages of text total (tables and figures not included in that total). To make it easier for me to gauge report length, put your tables and figures at the end of the paper.
- If you wish, you can include an appendix containing any results that you don't refer to directly in your report but that you deem are useful. Note: if a result is important, put it in the report not the appendix. I probably won't spend much time reading your appendices. I certainly won't give a better grade to a person just because he or she turns in a long appendix!
- In a second appendix, you should include your final SAS code that you used to analyze the data.