IMT573 Lab 7: multiple regression

Your name:

November 18, 2020

Instructions

This is another computer lab where we ask you to work with multiple linear regression.

1. Please be clear! Ensure that the reader can understand why did you output whatever you did output.
2. Ensure that your result is readable even when no code is visible!

1 Diamonds price

1.1 Diamonds data

Our first task is to take a look at diamonds data in ggplot2 package. It contains the following variables:

carat diamond's mass, in carats (=0.2 g)

cut shape of the diamonds (consult Blue nile website for more explanations about diamonds cut, color and other characteristics).

color color of diamonds, D (colorless) is the best.

clarity transparency of diamonds, FL (flawless) is the best.

depth diamonds relative height (see brilliance.com for explanations)

table diamonds width, see above.

price in dollars

x, y, z dimension, mm

1. Load the data and perform basic sanity checks. How many diamonds do we have?
2. Inspect the variable *cut*. What kind of different cuts are there? How frequent are those?
   Hint: check out the function `table`

1.2 Regression analysis

Now it is time for regression analysis

1. Include two variables: the mass of diamonds (*carat*) and *cut*. Estimate model of a form

$$\text{price}_i = \beta_0 + \beta_1 \text{carat}_i + \beta_2 \text{cut}_i + \epsilon_i \tag{1}$$

2. What is the reference category for *cut*?
3. Interpret the following coefficients:
   (a) What is "carat" (correct value should be 7871)?
   (b) What is "cutIdeal" (correct value should be 1800)?
   (c) How much more expensive are ideal-cut diamonds compared to very good-cut diamonds (in average)?
   (d) What would be the (average) price for 1ct premium cut diamond?

2 Challenge (not graded)

If you have time and interest, then consider also doing the following tasks:

1. Plot the price versus mass and add the regression line to it.
2. Experiment with log scale for a) price, b) carat, c) both. Which plot does look the best?
3. Add the corresponding log-transform to your model. You can run log-transformed models just like `lm(log(price) ~log(carat))`
4. Add `cut` to this model and interpret the coefficients.

   Hint: see https://otoomet.bitbucket.io/machineLearning.pdf/ Section 4.1.6 Interactions and Feature Transformations.