

HW4 (Due February 26th)

You may work in groups of two. Both names must appear on the turned-in assignment. Note that you may not work with someone who was in your group previously in this class.

1) Consider the following dataset:

x_1	y
Republican	10
Republican	8
Democrat	5
Independent	3
Democrat	8
Independent	5
Democrat	10
Independent	2

- (a) Create columns to use as indicator variables in a model which incorporates the qualitative variable x_1 . Provide the dataset and the regression model of the response variable with respect to these predictor variables. (using α/β notation).
- (b) Obtain estimates of the coefficients for the model in (a) using R.
- (c) How does someone being a Republican (as opposed to Democrat) impact the response variable (y)?
- (d) How does someone being a Democrat (as opposed to Independent) impact the response variable (y)?
- (e) Obtain the SSE when 8-fold cross validation is used.
- (f) Obtain the MAE when 2-fold cross validation is used.

2) Consider the following two models:

I. $y = 2 + 3x_1 + 5x_2 - 10x_3 + \varepsilon$

II. $y = 2 - 3x_1 + 4x_2 - 8x_3 + 8x_1x_2 + \varepsilon$

- (a) According to Model I, compute the amount that the prediction of y will change by when x_1 increases by 1 unit and $x_2 = 10$ and $x_3 = 20$ do not change.
- (b) According to Model I, compute the amount that the prediction of y will change by when x_1 increases by 1 unit and $x_2 = 20$ and $x_3 = 30$ do not change.
- (c) According to Model I, compute the amount that the prediction of y will change by when x_1 increases by 1 unit and x_2 and x_3 stay the same (regardless of their value).
- (d) According to Model II, compute the amount that the prediction of y will change by when x_1 increases by 1 unit and $x_2 = 10$ and $x_3 = 20$ do not change.
- (e) According to Model II, compute the amount that the prediction of y will change by when x_1 increases by 1 unit and $x_2 = 20$ and $x_3 = 30$ do not change.

HW4 (Due February 26th)

3) Consider the following dataset:

x_1	x_2	y
50000	300	8
360123	457	16
230000	193	10.5
114095	645	10

With the goal being to have predictor variables which are the same scale, create new predictor variables using, log, and 3 other power transformations. Comment on which transformation(s) you think would be a good to use for which particular variable, if not both.