

LAB SESSION 10 – MULTIPLE LINEAR REGRESSION

Analytics Primer

REGRESSION CAUTIONS

Multicollinearity

Example

- A real estate company is trying to model housing prices (in dollars) of their customers with the variables:
 - x_1 : Size of Home (square feet)
 - x_2 : Age of Home (years)
 - x_3 : Acreage of Land (acres)
 - x_4 : Number of Bedrooms
- Using a sample of 105 houses they derive the following model:

$$\hat{y} = 24,312 + 86.5x_1 - 324x_2 + 9,610x_3 + 3,617x_4$$

$$SSE = 27695831$$

$$SSR = 45963293$$

$$TSS = 73659124$$

Example

- A real estate company is trying to model housing prices (in dollars) of their customers with the variables:
 - x_1 : Size of Home (square feet)
 - x_2 : Age of Home (years)
 - x_3 : Acreage of Land (acres)
 - x_4 : Number of Bedrooms
 - x_5 : Number of Bathrooms
- Using a sample of 105 houses they derive the following model:

$$\hat{y} = 28,438 + 62.9x_1 - 336x_2 + 9,610x_3 - 2,102x_4 + 3,498x_5$$

$$s_{\hat{\beta}_4} = 4002$$

Example

1. Without examining the numbers, is there any potential for multicollinearity?
 - Yes! Number of bathrooms could easily be correlated with number of bedrooms, and square footage.

Example

2. Calculate the hypothesis test for significance on the variable x_4 . In the 4 variable model, the p-value was between (0.3, 0.4).

$$H_0: \beta_4 = 0$$

$$H_a: \beta_4 \neq 0$$

P-value > 0.5 → DON'T REJECT H_0

$$t = \frac{-2102 - 0}{4002} = -0.53$$

Example

3. What signs do you see that might signal multicollinearity?
 - Change in signs of coefficients
 - Shifts in values of coefficients

MORE EXAMPLES!

Example

- Use the following two models to answer the next two questions:

$$\hat{y} = 140 - 75x_1 + 212x_2 \quad SSR = 21,000 \quad TSS = 42,000$$

$$\hat{x}_1 = 90 + 67x_2 \quad SSR = 8,000 \quad TSS = 12,000$$

Example

1. What is the VIF for x_1 ?

$$R_1^2 = \frac{8000}{12000} = 0.67$$

$$VIF_1 = \frac{1}{1 - 0.67} = 3$$

Example

2. Does this model appear to have a problem with multicollinearity?
 - No! The VIF is less than 10.