Problem set 3: Multiple regression
Data analysis part, ØKA201

Bjørnar Karlsen Kivedal

Excercise 1

Use the data set `HousePrice.gdt` for this excercise.

a)

Generate a correlation matrix between all of the variables in the data set and generate six scatter plots: between `price` and `size`, `bedrooms`, `bathromms`, `lotsize`, `age` and `monthsold`. Which variable is most highly correlated with the price?

b)

Estimate the model

$$price_i = \beta_0 + \beta_1 bathrooms_i + u_i$$

Interpret the estimated coefficients.

c)

Estimate the model

$$price_i = \beta_0 + \beta_1 bathrooms_i + \beta_2 size_i + u_i$$

Interpret the estimated slope coefficients. Compare $\hat{\beta}_1$ with what you estimated in b). What is the difference and what can be the reason for this?

d)

Estimate the model

$$price_i = \beta_0 + \beta_1 size_i + u_i$$

Interpret the estimated slope coefficient

e)

Estimate the model

$$price_i = \beta_0 + \beta_1 size_i + \beta_2 bathrooms_i + \beta_3 bedrooms_i + \beta_4 lotsize_i + \beta_5 age_i + \beta_6 monthsold_i + u_i$$

(i) Interpret the estimated slope coefficients (some makes less sense than others).

(ii) Do any of the regressors have coefficients with an unexpected sign?

(iii) Which of the variables have statistically signinficant effects on sales price at a 5% significance level?

(iv) Are the regressors jointly statistically significant at at 5% significance level? Explain

(v) Are the regressors other than size jointly statistically significant at at 5% significance level? Explain

f)

Which of the models you have estimated would you recommend the realtor to use in order to estimate the sales price of a house?

Excercise 2

a)

Assume that we have the estimated model

$$\hat{Y}_i = \underset{(6)}{10} + \underset{(40)}{72}\, X_i + \underset{(5)}{16} Z_i$$

where standard errors are in parentheses.

(i) Obtain the test statistic for $\hat{\beta}_2$, the coefficient of Z_i

(ii) Is Z_i statistically significant at a 5% significance level?

(iii) Provide a 95% confidence interval for β_2

b)

Assume that we have the estimated model

$$\hat{Y}_i = \underset{(28)}{110} + \underset{(8)}{18} X_i + \underset{(10)}{24}\, Z_i$$

where standard errors are in parentheses.

(i) Obtain the test statistic for $\hat{\beta}_2$, the coefficient of Z_i

(ii) Is Z_i statistically significant at a 5% significance level?

(iii) Provide a 95% confidence interval for β_2

c)

Assume that we have the estimated model

$$\hat{Y}_i = \underset{[2.5]}{10} + \underset{[3]}{72} X_i + \underset{[5]}{16} Z_i$$

where test statistics are in brackets.

(i) Is Z_i statistically significant at a 5% significance level?

(ii) Obtain the standard error for $\hat{\beta}_2$, the coefficient of Z_i

(iii) Provide a 95% confidence interval for β_2

d)

Assume that we have the estimated model

$$\hat{Y}_i = \underset{[3.5]}{110} + \underset{[12]}{18}\, X_i + \underset{[10]}{24}\, Z_i$$

where test statistics are in brackets.

(i) Is Z_i statistically significant at a 5% significance level?

(ii) Obtain the standard error for $\hat{\beta}_2$, the coefficient of Z_i

(iii) Provide a 95% confidence interval for β_2