

QUEEN MARY UNIVERSITY OF LONDON

MTH5120

Statistical Modelling I

Exercise Sheet 10

1.

We use the Bridge.txt dataset available on QMPlus, where information from 45 bridge projects are compiled. The response and predictor variables are as follows:

- Y : Time is the design time in person-days;
- X_1 : DArea is the deck area of bridge (000 sq ft);
- X_2 : CCost is the construction cost (\$000);
- X_3 : Dwgs is the number of structural drawings;
- X_4 : Length is the length of bridge (ft);
- X_5 : Spans is the number of spans.

Take the logarithm transformation of all the variables.

- (a) Use RStudio to find the best reduced model using the AIC procedure and state which is the best reduced model;
- (b) For the best reduced model, comment on the values of VIF.
- (c) For the best reduced model, comment on the significance of the exogenous variables;
- (d) For the best reduced model, comment on the significance of the overall regression and of the individual explanatory variables in the presence of the other variables (based on Anova table).

2. When fitting the model

$$E[Y_i] = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$$

to a set of $n = 25$ observations, the following results were obtained using the general linear model notation:

$$\mathbf{X}^t \mathbf{X} = \begin{pmatrix} 25 & 219 & 10232 \\ 219 & 3055 & 133899 \\ 10232 & 133899 & 6725688 \end{pmatrix}, \quad \mathbf{X}^t \mathbf{Y} = \begin{pmatrix} 559.60 \\ 7375.44 \\ 337071.69 \end{pmatrix}$$
$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{pmatrix} 0.1132 & -0.0044 & -0.00008 \\ -0.0044 & 0.0027 & -0.00004 \\ -0.00008 & -0.00004 & 0.000001 \end{pmatrix}$$

Also $\mathbf{Y}^t \mathbf{Y} = 18310.63$ and $\bar{Y} = 22.384$.

- (a) Compute the *AIC* and the *VIF*

(b) In the same way, run a two dimensional model:

$$E[Y_i] = \beta + \beta_1 x_{1,i}$$

to the same set of 25 observations and we have the following results:

$$\mathbf{X}^t \mathbf{X} = \begin{pmatrix} 25 & 219 \\ 219 & 3055 \end{pmatrix}, \quad \mathbf{X}^t \mathbf{Y} = \begin{pmatrix} 559.60 \\ 7375.44 \end{pmatrix}$$
$$(\mathbf{X}^t \mathbf{X})^{-1} = \begin{pmatrix} 0.1075 & -0.0077 \\ -0.0077 & 0.00087 \end{pmatrix}$$

Compute the *AIC* and the *VIF*

(c) Which is the best model across the two models, the one with two explanatory variables or the one with one explanatory variable.