

Linear Regression Models: Tutorial Sheet

1. Describe how to use the Akaike Information Criterion for model selection.
2. Compare and contrast three types of variable selection procedure.
3. Explain what variable selection procedures are used for.
4. Model Selection Question x_1, x_2, x_3 and x_4 .

Suppose we have 5 predictor variables. Use **Forward Selection** and **Backward Selection** to choose the optimal set of predictor variables, based on the AIC measure.

Variables	AIC	Variables	AIC
\emptyset	200	x_1, x_2, x_3	74
		x_1, x_2, x_4	75
x_1	150	x_1, x_2, x_5	79
x_2	145	x_1, x_3, x_4	72
x_3	135	x_1, x_3, x_5	85
x_4	136	x_1, x_4, x_5	95
x_5	139	x_2, x_3, x_4	83
		x_2, x_3, x_5	82
x_1, x_2	97	x_2, x_4, x_5	78
x_1, x_3	81	x_3, x_4, x_5	85
x_1, x_4	94		
x_1, x_5	88	x_1, x_2, x_3, x_4	93
x_2, x_3	87	x_1, x_2, x_3, x_5	120
x_2, x_4	108	x_1, x_2, x_4, x_5	104
x_2, x_5	87	x_1, x_3, x_4, x_5	101
x_3, x_4	105	x_2, x_3, x_4, x_5	89
x_3, x_5	82		
x_4, x_5	86	x_1, x_2, x_3, x_4, x_5	100

5. What is Multicollinearity? Describe the implications of Multicollinearity?
6. Contrast the uses of Training Data, Validation Data and Testing Data, when creating a predictive model.
7. What is meant by overfitting, in the context of predictive models?
8. Describe how you would use the Variance Inflation Factor to make an assessment about multicollinearity.
9. Describe the process of model validation, with reference to training, validation and testing phases.

10. State two ways of methodically diagnosing the severity of multi-collinearity. How are these techniques related? How are they used to make decisions about the data?
11. State two ways in which a multiple regression technique could be affected by severe multicollinearity?
12. Explain what variable selection procedures are used for.