Linear Regression Models: Tutorial Sheet

1. Describe how to use to 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
∅	200	x1, x2, x3	74
		x1, x2, x4	75
x1	150	x1, x2, x5	79
x2	145	x1, x3, x4	72
x3	135	x1, x3, x5	85
x4	136	x1, x4, x5	95
x5	139	x2, x3, x4	83
		x2, x3, x5	82
x1, x2	97	x2, x4, x5	78
x1, x3	81	x3, x4, x5	85
x1, x4	94		
x1, x5	88	x1, x2, x3, x4	93
x2, x3	87	x1, x2, x3, x5	120
x2, x4	108	x1, x2, x4, x5	104
x2, x5	87	x1, x3, x4, x5	101
x3, x4	105	x2, x3, x4, x5	89
x3, x5	82		
x4, x5	86	x1, x2, x3, x4, x5	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.

13. In the context of regression models, explain what is meant by Heteroscedascity and Homoscedascity. Support your answers with sketches.

14. Explain the term "Influence" in the context of linear regression models. Support your answer with sketches.

15. Explain the term "Cook's Distance" in the context of linear regression models.

16. The Durbin Watson Test was carried out to test for Autocorrelation. Briefly describe autocorrelation. You may support your answer with sketches.

17. State your conclusion to the following procedure.

```
> durbinWatsonTest(myModel)

lag Autocorrelation D-W Statistic p-value
1     -0.08428163      2.143578   0.806

Alternative hypothesis: rho != 0
```

18. In certain circumstances, Robust Regression may be used in preference to Ordinary Least Squares Regression. Describe what these circumstances might be.

19. State one difference between Ordinary Least Squares and Robust regression techniques, in terms of computing regression equations.

20. Explain the process of Huber Weighting, stating the algorithm used to compute weightings.

21. Suppose that Huber Weighting, with a tuning constant of $k = 13.45$ was applied to the observations tabulated below. What would be the outcome of the procedure for each case.

Observation (i)	Residual (e_i)
11	-9.07
14	14.54
18	22.91