

Tutorial Sheet 1

1. Numeric Transformations, such as logarithmic transformation, are often used in statistical analysis as an approach for dealing with non-normal data.
 - (i.) (1 Mark) Describe the purpose of Tukey's Ladder (referencing direction and relative strength).
 - (ii.) (2 Marks) Give two examples of a transformation for various types of skewed data (i.e. an example for both types of skewness).
 - (iii.) (1 Mark) Discuss the limitations of numeric transformations.
2. The typing speeds for one group of 12 Engineering students were recorded both at the beginning of year 1 of their studies. The results (in words per minute) are given below:

149	146	112	142	168	153
137	161	156	165	170	159

Use the Dixon Q-test to determine if the lowest value (118) is an outlier. You may assume a significance level of 5%.

- (i.) (1 Mark) State the Null and Alternative Hypothesis for this test.
- (ii.) (2 Marks) Compute the test statistic
- (iii.) (1 Mark) State the appropriate critical value.
- (iv.) (1 Mark) What is your conclusion to this procedure.

- Outliers**
- (i.) (3 Marks) Provide a brief description for three tests from the family of Grubb's Outliers Tests. Include in your description a statement of the null and alternative hypothesis for each test
 - (ii.) (2 Marks) Describe any required assumptions for tests, and the limitations of these tests.

Model Selection 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	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

3. Model Selection Question

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

\emptyset	200	x1,x2,x3	74
		x1,x2,x4	75
x1	150	x1,x2,x5	78
x2	170	x1,x3,x4	72
x3	135	x1,x3,x5	82
x4	130	x1,x4,x5	70
x5	140	x2,x3,x4	80
		x2,x3,x5	82
x1,x2	90	x2,x4,x5	78
x1,x3	81	x3,x4,x5	75
x1,x4	84		
x1,x5	78	x1,x2,x3,x4	83
x2,x3	87	x1,x2,x3,x5	130
x2,x4	78	x1,x2,x4,x5	104
x2,x5	87	x1,x3,x4,x5	101
x3,x4	85	x2,x3,x4,x5	89
x3,x5	88		
x4,x5	86	x1,x2,x3,x4,x5	100