Testing Model Assumptions: Tutorial Sheet

1. Numeric Transformations, such as logarithmic transformation, are often used in statistical analysis as an approach for dealing with non-normal data.

   (a) Discuss the importance of numeric transformations, such as logarithmic transformation, in statistics.

   (b) Describe the process of transformations

   (c) Describe the purpose of Tukey's Ladder (referencing direction and relative strength).

   (d) Give two examples of a transformation for various types of skewed data (i.e. an example for both types of skewness).

   (e) 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	118	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.) State the Null and Alternative Hypothesis for this test.

   (ii.) Compute the test statistic

   (iii.) State the appropriate critical value.

   (iv.) What is your conclusion to this procedure.

3. **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.

4. Use the Dixon Q-test to determine if there is an outlier present in this sample data. You may assume a significance level of 5%.

$$131, 136, 103, 117, 123, 127, 122, 132, 135$$

   (i) (1 Mark) State the null and alternative hypotheses 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?

5. Suppose that the results of an experimental procedure resulted in the collection of datasets X. Consider the following inference procedure performed on data set X.

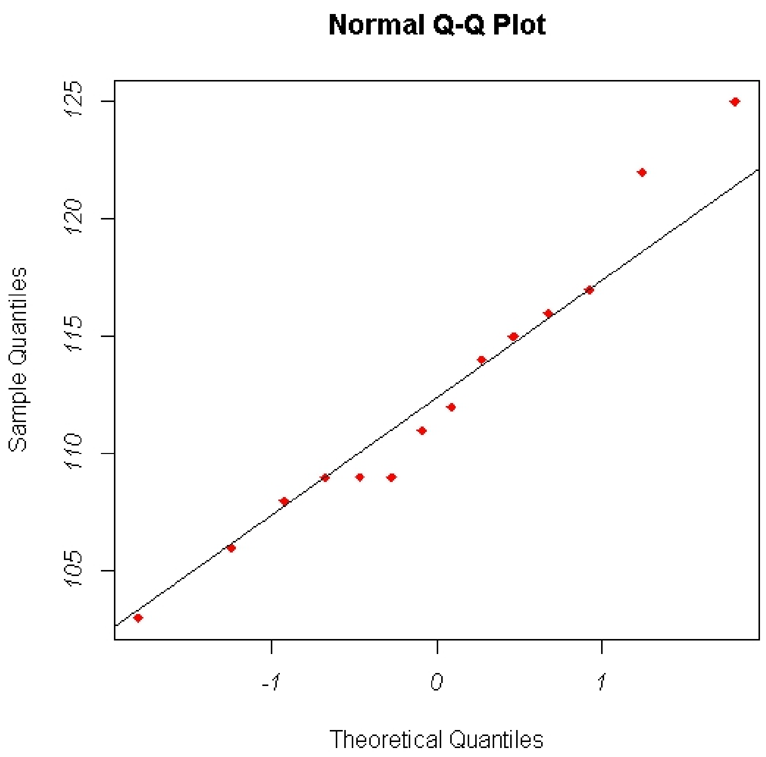
```
> shapiro.test(X)

Shapiro-Wilk normality test

data:  X
W = 0.9619, p-value = 0.6671
```

   (i) (1 Mark) Describe the purpose of this procedure.
   (ii) (1 Mark) What is the null and alternative hypothesis?
   (iii) (1 Mark) What is your conclusion about this procedure?

6. A graphical procedure was carried out to assess whether or not this assumption of normality is valid for data set `Y`. Consider the Q-Q plot in the figure below.



   (iv) (1 Mark) Provide a brief description on how to interpret this plot.
   (iv) (1 Mark) What is your conclusion for this procedure? Justify your answer.

7. The following statistical procedure is based on this dataset.

6.98	8.49	7.97	6.64
8.80	8.48	5.94	6.94
6.89	7.47	7.32	4.01

```
> grubbs.test(x, two.sided=T)

Grubbs test for one outlier

data:  x
G = 2.4093, U = 0.4243, p-value = 0.05069
alternative hypothesis: lowest value 4.01 is an outlier
```

   i. (1 Mark) Describe what is the purpose of this procedure. State the null and alternative hypothesis.
   ii. (1 Mark) Write the conclusion that follows from it.
   iii. (1 Mark) State any relevant assumptions for this procedure.

8. 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, any required assumptions and the limitations of these tests.

9. Showing your working, use the Dixon Q Test to test the hypothesis that the maximum value of the following data set is an outlier.

$$20, 20, 22, 23, 24, 25, 26, 27, 29, 38$$

Formulas and Tables

Critical Values for Dixon Q Test

N	$\alpha = 0.10$ *Confidence*= 0.90	$\alpha = 0.05$ *Confidence*= 0.95	$\alpha = 0.01$ *Confidence*= 0.99
3	0.941	0.97	0.994
4	0.765	0.829	0.926
5	0.642	0.71	0.821
6	0.56	0.625	0.74
7	0.507	0.568	0.68
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568
11	0.392	0.444	0.542
12	0.376	0.426	0.522
13	0.361	0.41	0.503
14	0.349	0.396	0.488
15	0.338	0.384	0.475
16	0.329	0.374	0.463