MSc Introductory Material - Block B

Statistical Methods Exercises

1. Do some detective work on the following data, which is of mean January temperatures in $^{\circ}\mathrm{C}$ for 60 U.S. Metropolitan Areas.

   Is there anything suspicious about the numbers?

   $$\begin{array}{r|l} & \text{Unit} = 0.1 \\ 1 & 2 \quad \text{represents } 1.2 \end{array}$$

LO	−111
−6	6
−5	00
−4	44444
−3	383
−2	7727722
−1	611116166
−0	5500
+0	055055
1	616
2	7
3	38
4	444
5	55
6	
7	2272
8	8
9	4
HI	127,116,194,122,127

   Note that the top line here indicates a data point outside the main range of the table, with the value -11.1 .

2. A chemical theory suggests that the temperature at which a certain reaction occurs is $180^{\circ}\mathrm{C}$. The results of 15 determinations were:
   183.2, 179.4, 175.4, 179.0, 180.6, 178.1, 178.1, 175.7, 179.3, 176.9, 178.3, 179.1$, 180.2, 174.2, 179.4.

   a. What conclusions about the theory can be drawn from these data?
   b. Investigate the view that the experimental procedure used is unsatisfactory since it determines reaction temperature with a standard deviation exceeding $4^{\circ}\mathrm{C}$.

3. Two types A and B of solid fuel boilers are under investigation for the dust content of their flue gases. For this purpose 8 type A boilers and 6 type B boilers were used under identical fuelling and extraction conditions. Over the period of the investigation the following were the quantities in gm of dust deposited in similar traps in each of the 14 flues.

   A: $64, 75, 39, 47, 49, 81, 72, 61$
   B: $31, 57, 63, 40, 54, 49$.
   Discuss the evidence for the following claims:

   a. Deposits are equally variable for the two types of boiler.

b. The mean deposits for the two types are the same.

Estimate the difference in the average amounts of dust deposited by the two types of boiler.

4. Using R find the p-value for the Normal standardized test statistic values z where the number in brackets indicates the form of the alternative hypothesis as follows:(1: reject for large positive z; 2: reject for large negative z; 3: reject for large $|z|$).

   1.83 (1), -1.62 (2), 1.45 (3), 2.02 (1), 0.36 (2), 2.35 (3), 3.54 (3), -1.763 (2), -1.894 (3), 1.046 (1).

5. Using R, complete the following in a similar manner where the numbers in square brackets represent degrees of freedom.

   a. Find the p-value for the following t values. (1: reject for large positive $t, 2$: reject for large negative $t, 3$: reject for large $|t|$).
   $1.83[5](1), -1.62[8](2), 1.45[2](3), 2.02[12](1), 0.36[36](2), 2.35[19](3)$, 3.54 [42] (3), -1.763 [74] (2), -1.894 [7] (3), 1.046 [200] (1).

   b. Find the p-value for the following χ^2 values c. Here [] denote the degrees of freedom for the χ^2 distribution. (1: reject for large $c, 2$: reject for small $c, 3$: reject for large or small c).
   1.83 [5] (1), 2.62 [8] (2), 1.45 [2] (3), 22.02 [12] (1), 20.36 [36] (2), 32.35 [19] (3), 53.54 [42] (3), 91.76 [74] (2), 11.89 [7] (3), 121.04 [180] (1).

   c. Find the p-value for the following F values f. Here [,] denote the degrees of freedom for the F distribution. (1: reject for large $f, 3$: reject for large or small f.)
   $10.83[5, 9](1), 0.32[8, 6](3), 3.45[12, 2](3), 12.02[2, 12](1), 0.36[10, 36]$ (3), 2.35 [19,18] (3), 3.54 [42,24] (3), 1.763 [74, 83] (3), 2.894 [7,17] (3), 1.446 [200,200] (1).

6. In comparing two methods of chlorinating sewage, eight pairs of batches of sewage were treated. Each pair was taken on a different day, the two batches on any one day being taken close together in time, and the two treatments were randomly assigned to the two batches in each pair. Treatment A involved an initial period of rapid mixing while treatment B did not. The results, in log coliform density per ml, were as follows.

Day	1	2	3	4	5	6	7	8
	A2.8	B3.1	B3.4	A3.0	B2.7	B2.9	B3.5	A2.6
	B3.2	A3.1	A2.9	B3.5	A2.4	A3.0	A3.2	B2.8

   Assess the difference in treatments.

7. The toxicity of a drug may be measured by the proportion p of mice in a standard laboratory population that would die after injection with a dose of given strength. Out of 30 mice randomly selected from the population 8 died after injection. Find, using R, confidence limits for p

   a. approximately, from Normal approximation;
   b. exactly.

8. For the data set D5 for discrete data would a Poisson model be appropriate for the number of particles emitted in a $7\frac{1}{2}$ second period?

9. The table below classifies 801 women by their attitudes to the proposition that ‘most men are better suited emotionally to politics than most women’ and by length of education.

	Years' education		
	≤ 8	$9-12$	≥ 13
Agree	83	255	76
Disagree	38	207	142

(U.S. General Social Survey 1975)

   Test for association between attitude to the proposition and length of education.

10. The data below give the number of pages, in hundreds, and the price, to the nearest pound, of a random sample of 10 text books produced by one publisher.

Book:	i	1	2	3	4	5	6	7	8	9	10
Pages:	a_i	2.4	2.8	2.9	3.7	3.9	4.3	5.3	6.2	6.9	7.1
Price:	b_i	10	11	15	19	14	17	15	18	25	23

$$\sum_{i=1}^{10} a_i = 45.5 \sum_{i=1}^{10} a_i^2 = 233.95 \sum_{i=1}^{10} b_i = 167 \sum_{i=1}^{10} b_i^2 = 2995 \sum_{i=1}^{10} a_i b_i = 823.2$$

Plot the data. Find the least squares regression line to predict price on the basis of number of pages. Plot this on your graph. It is alleged that the publisher increases the price of a book by 30p for each extra 10 pages beyond the basic set-up costs. Do the sample data provide evidence against this allegation?