MAS programmes - Statistical Analysis (Autumn Term)

Exercises 1

The following exercises should be tackled 'by-hand', to test your understanding (recollection) of some of the basic statistical concepts outlined in Lecture 1. They should also be attempted during the first Practical Computing Session, as a means of getting started with the R software. *To aid hand calculations, useful quantities from the relevant distributions are given as hints.*

1. 25 grapefruit were exposed to sunlight in such a way that only half of each (uncut) fruit was exposed, the other half remaining in shade. The response variable recorded was the percentage of solids in each grapefruit half, and interest was in whether or not the exposure to sunlight affects the percentage of solids in the exposed half of the grapefruit.

	% solids	
Grapefruit	Shaded	Exposed
1	8.59	8.49
2	8.59	8.59
3	8.09	7.84
4	8.54	7.89
5	8.09	8.19
6	8.49	7.84
7	7.89	7.89
8	8.59	7.89
9	8.54	7.79
10	7.99	7.84
11	7.89	7.79
12	8.09	7.84
13	7.89	7.89
14	8.54	8.07
15	7.84	7.97
16	7.49	7.57
17	7.89	7.92
18	7.79	7.97
19	7.84	8.17
20	8.89	8.67
21	8.54	8.07
22	8.04	7.97
23	8.59	8.62
24	8.19	7.92
25	8.59	7.97

Carry out an appropriate test of the hypothesis that the percentage of solids is the same in both the shaded and exposed grapefruit halves.

[*Hint*: $t_{24;0.025} = 2.06$ (is the 0.975-quantile of the t_{24} distribution)].

2. In 1898 Hermon Bumpus published data consisting of measurements taken from 136 house sparrows that had suffered badly during a severe storm. About half of the birds saved subsequently died, and Bumpus decided to analyse his data looking for evidence relating to Darwin's theory of natural selection. From the analysis of all his data, Bumpus concluded that the ability to survive was generally less for those birds that were most different from the norm.

We are concerned in this example with a subset of the data, namely the 49 female sparrows, and will consider the variable weight (measured in grams) and how it varies between the sample of birds who survived and those who died.

Weights of female sparrows in grams

Survived					Died				
25.3	22.6	25.1	23.2	24.4	26.3	25.8	26.0	23.2	26.5
25.1	24.6	24.0	24.2	24.9	24.2	26.9	27.7	23.9	26.1
24.1	24.0	26.0	24.9	25.5	24.6	23.6	26.0	25.0	24.8
23.4	25.9	24.2	24.2	27.4	22.8	24.8	24.6	30.5	24.8
24.0					23.9	24.7	26.9	22.6	26.1
					24.8	26.2	26.1		

(a) Calculate the 95% confidence interval for the *variance* of the weight of female sparrows that did not survive. Is the variance of the weight of the survivors compatible with this interval?

[*Hint*: $\chi^2_{27;0.025} = 43.19$ and $\chi^2_{27;0.975} = 14.57$].

(b) Using a suitable hypothesis test, test the hypothesis that the variance of the weight is the same for the two groups, female sparrows that did not survive and surviving female sparrows.

[*Hint*: $F_{27,20;0.025} = 2.38$].

(c) Use a t-test to test the hypothesis that the population mean weight of female sparrows does not depend on whether or not the sparrow survived.

[*Hint*: $t_{47;0.025} = 2.01$].

Recall that for a two-sample t-test, we use

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad \text{where} \quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

Note: when using R, you should also investigate graphically the assumptions made in each of the tests.