Probability and Statistics
USTH-B2

Exercices Chapter 5

1. Given the following data set (weight loss per week) $(9, 2, 5, 8, 4, 5)$:
   (a) Find the sample mean.
   (b) Find the sample median.
   (c) Find the sample mode.
   (d) Find the sample variance using the defining formula.
   (e) Find the sample standard deviation.
   (f) Find the first and third quartiles, Q_1 and Q_3.
2. Same exercice for the following data set $(21, 24, 15, 16, 24)$.
3. In a study of infant birth weight for different occupational groups this was recorded for some first-time mothers.
   The following table shows the results in grams (data specified in sorted order) for 20 births in total, 10 female births and 10 male births.

Females (x)	2474	2547	2830	3219	3429	3448	3677	3872	4001	4116
Males (y)	2844	2863	2963	3239	3379	3449	3582	3926	4151	4356

   Using classical computations or a software
   (a) What is the sample mean, variance and standard deviation of female births ? give some interpretations.
   (b) Same for male births. Compare with the female births.
   (c) Find the five quartiles for each distribution, using R and using a pen.
4. To compare the difficulty of 2 different courses at a university, the following grades distributions (given as number of pupils who achieved the grades) were registered :

	Course 1	Course 2	Total
Grade 12	20	14	34
Grade 10	14	14	28
Grade 7	16	27	43
Grade 4	20	22	42
Grade 2	12	27	39
Grade 0	16	17	33
Grade -3	10	22	32
Total	108	143	251

(a) What is the median of the 251 achieved grades ?

(b) What are the quartiles and the IQR ?

5. In a clinical trial of a cholesterol-lowering agent, 15 patients' cholesterol (in mMol/l) was measured before treatment and 3 weeks after starting treatment. Data are listed in the following table :

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before	9.1	8.0	7.7	10.0	9.6	7.9	9.0	7.1	8.3	9.6	8.2	9.2	7.3	8.5	9.5
After	8.2	6.4	6.6	8.5	8.0	5.8	7.8	7.2	6.7	9.8	7.1	7.7	6.0	6.6	8.4

(a) What is the median of the cholesterol measurements for the patients before treatment, and similarly after treatment ?

(b) Find the standard deviations of the cholesterol measurements of the patients before and after treatment.

(c) Find the sample covariance between cholesterol measurements of the patients before and after treatment.

(d) Find the sample correlation between cholesterol measurements of the patients before and after treatment.

(e) Compute the 15 differences (Dif = Before - After) and do various summary statistics and plotting on these : mean, variance, standard deviations, boxplot etc.

(f) Can we have some conclusions ?