Tutorial n^o 3
Using the R software

Exercise 1. *Eat apples !*
We want to compare the vitamin C values of five apples varieties, denoted by V_1, V_2, V_3, V_4 et V_5. For every variety the vitamin C value was measured, units $mg/(100g)$, in five randomly chosen apples. We got a dataset which is summed up in the following table :

V_1	V_2	V_3	V_4	V_5
$93,6$	$95,3$	$94,5$	$98,8$	$94,6$
$95,3$	$96,9$	$97,0$	$98,2$	$97,8$
$96,0$	$95,8$	$97,8$	$97,8$	$98,0$
$93,7$	$97,3$	$97,0$	$97,2$	$95,0$
$96,2$	$97,7$	$98,3$	$97,9$	$98,9$

What can we deduce from the data ? To answer to this question, use the analysis of variance.

This exercise is taken from the exercise book of François Husson and Jérôme Pagès : Statistiques générales pour utilisateurs*, PUR.*

Exercise 2. *Tasting beers*
Eight marks of acidity were collected for four white beers. These marks were rounded to the integer and summed up in the following table.

	beer 1	beer 2	beer 3	beer 4
mark 1	5	0	5	0
mark 2	5	1	6	0
mark 3	5	2	6	1
mark 4	6	2	7	1
mark 5	7	3	8	2
mark 6	7	4	9	3
mark 7	8	6	10	4
mark 8	10	6	10	4

1. After having computed the summary statistics for the marks of any the four beers, use boxplots to display them.
2. We want to know whether or not these beers taste the same using the acid marks. What is an appropriate statistical tool ? Explain why.
3. Write down the corresponding statistical model.
4. State the analysis of variance table.
5. What is the percent of the variation of the acid mark explained by the beer factor ?

6. What test can you use to compare all the beers ? Write down the test procedure. Decide wether or not these beers taste the same using the acid marks with a type I error of 1 %.

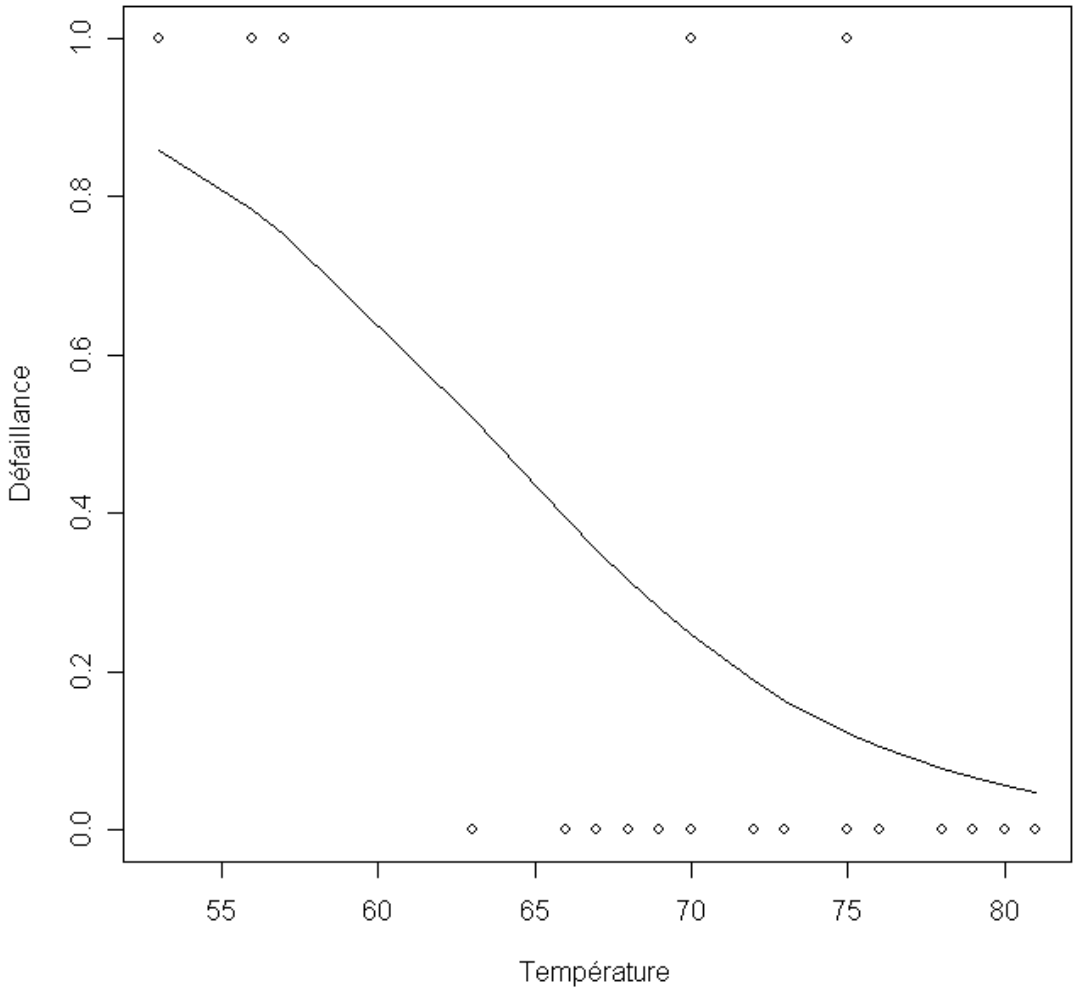
This exercise is taken from the exercise book of A.C. Davison : Statistical Models*, Cambridge University Press.*

Exercise 3. Défaillance d'un élément mécanique

The space shuttle Challenger exploded shortly after its launch on 28 January 1986, with a loss of seven lives. The accident is likely to have been caused by a leakage of gas from one of the fuel-tanks. Rubber insulating rings, so-called 'O-rings', were not pliable enough after the overnight low temperature of 31°F, and did not plug the joint between the fuel in the tanks and the intense heat outside.
We summed up in the following table the dataset from Dalal et *al.*, 1986. These data were collected during the 24 previous spaceflights of an American shuttle : the $Temperature$ variable is the temperature when the shuttle launches, in °F, and the variable $Damaged\ O-ring$ is 0 if none of the « O-ring » was damaged during the launch of the shuttle and 1 if any of them was damaged.

$Temperature$	$Damaged$	$Temperature$	$Damaged$
53	1	70	1
56	1	70	1
57	1	72	0
63	0	73	0
66	0	75	0
67	0	75	1
67	0	76	0
67	0	76	0
68	0	78	0
69	0	79	0
70	0	80	0
70	1	81	0

1. What is the explained variable ? What are its values ? What is the explaining variable ? What are its values ? What statistical model can we use ?
2. Display using a scatter plot the data from the previous table. Can you see any link between the temperature and the vulnerability of the « O-rings » ?
3. Using a logistic regression, state whether or not the temperature variable accounts for the vulnerability of the « O-rings ». Display the predicted probabilities on the previous scatter plot. Compare you graphic with the following one.
4. What was the predicted probability that at least one of the « O-rings » would be damaged the 28 January 1986 ? Even though we built this model using values ranging far away from the temperature of 31°F would you have proceeded to the launch ?


Défaillance
0.0
0.2
0.4
0.6
0.8
1.0
55
60
65
70
75
80
Température