1. Exercise 1:
   Let consider the following observations:

observation	country	% agricultural population	calories per day per person
1	Switzerland	4	3432
2	France	5.7	3273
3	Sweden	4.9	3049
4	United States	3	3642
5	Russia	14.8	3394
6	China	69.6	2628
7	India	63.8	2204
8	Brazil	26.2	2643
9	Peru	38.3	2192
10	Algeria	24.7	2687
11	Zaire	65.7	2159

   (a) Make a drawing of those observations

   (b) Estimate the parameters of the model:

$$Y_i = \beta_0 + \beta_1.x_i + \varepsilon_i$$

   (c) Explain the different outputs of the lm functions used in the R software and precise the tests that are performed and how they are performed.

   (d) Construct a confidence interval at 95% for the regression curve, at a point x_0.

   (e) Draw the points, the regression curve and the curves associated to the confident interval on the same graphic.

2. Exercise 2:
   Let consider the observations given by the file test1.Rdata

   (a) Perform the multiple regression model and analyse all the outputs that are available.

   (b) What are the variables that are really connected to the response variable?

3. Exercise 3:
   Let consider the observations given by the file test2.Rdata

   (a) Perform the multiple regression model and analyse all the outputs that are available.

   (b) Is this model interesting?

   (c) Perform a non-linear model and analyse it.

   (d) What are the variables that are really connected to the response variable?

4. Exercise 4:
   Three teams take turns on a production line (morning, afternoon, night). They use four work stations denoted A,B,C and D. During one month, we count the number of defective parts according to the team and the work station. Here is the observations :

	A	B	C	D	mean
morning	26	13	35	6	20
afternoon	18	17	31	2	17
night	31	24	33	4	23
mean	25	18	33	4	20

We would like to analyse this table with a analysis of variance with two factors.

(a) For the moment, we analyse the variables team and work station separately.

i. Can we say that the performance are not the same with respect to the team?

ii. Can we say that the difficulties associated to each work station are not the same?

(b) Now we take into account both variables in the same time. At first we consider the additive model $Y_{i,j} = \mu + a_i + b_j + \varepsilon_{i,j}$ where the $\varepsilon_{i,j}$ are i.i.d. random variables whose distribution is $\mathcal{N}(0, \sigma^2)$.

i. How to validate the use of the additive model?

ii. We accept the additive model. How to estimate $\mathbb{E}[Y_{i,j}]$ in this model?

iii. Is the variable team influent?

iv. Is the variable work station influent?