MULTILINEAR REGRESSION

1. An experimenter wants to study a system which depends on one single natural variable ξ_1. For this variable, he defines $m = 6$ different levels $\xi_{1,1} = 10, \ldots, \xi_{6,1} = 60$ and for each of the $\xi_{1,i}$ he obtains $n_i = 3$ observations $y_{i,1}, \ldots, y_{i,3}$ for the response y. The data are in file `reglin.dat`. The experimenter wants to fit these data with the following linear model $y = a_0 + a_1\xi_1 + \varepsilon$.

   - What are the value of k and p? What do you choose for x_1?
   - Compute matrix X and vector $\mathbf{y}$.
   - Compute matrices C and H.
   - Compute $\hat{\mathbf{a}} = (\hat{a}_0, \hat{a}_1)$.
   - Compute SSE, SSR, SST, $MSE = \widehat{\sigma^2}$
   - Is there a linear relationship between the response y and at least one regression variable?
   - Test $H_0 : a_j = 0$ versus $H_1 : a_j \neq 0$ for $j = 0, 1$.
   - Compute confidence intervals $\{(a_j)_L, (a_j)_U\}$ for $j = 0, 1$.
   - Compute R^2, R_a^2, the residual vectors $\mathbf{d}$, $\mathbf{r}$, and the Cook distance vector D.
   - Compute $SSLOF$ and $SSPE = \widehat{\sigma^2}$. Is the regression model $y = a_0 + a_1\xi_1 + \varepsilon$ valid?
   - Plot data, the regression curve and the two confidence interval curves.

2. Now the experimenter wants to fit the same data with the following model $y = a_0 + a_1 \ln(\xi_1)/\xi_1 + \varepsilon$

   - Compute matrix X and vector $\mathbf{y}$.
   - Compute matrices C and H.
   - Compute $\hat{\mathbf{a}} = (\hat{a}_0, \hat{a}_1)$.
   - Compute SSE, SSR, SST, $MSE = \widehat{\sigma^2}$
   - Is there a linear relationship between the response y and at least one regression variable?
   - Test $H_0 : a_j = 0$ versus $H_1 : a_j \neq 0$ for $j = 0, 1$.
   - Compute confidence intervals $\{(a_j)_L, (a_j)_U\}$ for $j = 0, 1$.
   - Compute R^2, R_a^2, the residual vectors $\mathbf{d}$, $\mathbf{r}$, and the Cook distance vector D.

- Compute $SSLOF$ and $SSPE = \widehat{\sigma^2}$. Is the regression model $y = a_0 + a_1 \ln(\xi_1)/\xi_1 + \varepsilon$ valid?
- Plot data, the regression curve and the two confidence interval curves.

3. An experimenter wants to study a system which depends on two single natural variables ξ_1 and ξ_2. For these variables, he defines $m = 9$ different levels $\{(10, 30), (15, 30), \ldots, (20, 40)\}$, and for each of them he obtains $n_i = 2$ observations for the response y. The data are in file `regmlin.dat`. The experimenter wants to fit these data with the following model $y = a_0 + a_1\xi_1 + a_2\xi_2 + a_3\xi_1^2 + a_4\xi_2^2 + a_5\xi_1\xi_2 + \varepsilon$.
   - What are the value of k and p? What do you choose for $x_1, \ldots, x_5$?
   - Compute matrix X and vector $\mathbf{y}$.
   - Compute matrices C and H.
   - Compute $\hat{\mathbf{a}}$.
   - Compute SSE, SSR, SST, $MSE = \widehat{\sigma^2}$
   - Is there a linear relationship between the response y and at least one regression variable?
   - Test $H_0 : a_j = 0$ versus $H_1 : a_j \neq 0$ for $j = 0, \ldots, 5$.
   - Compute confidence intervals $\{(a_j)_L, (a_j)_U\}$ for $j = 0, \ldots, 5$.
   - Compute R^2, R_a^2, the residual vectors $\mathbf{d}$, $\mathbf{r}$, and the Cook distance vector D.
   - Compute $SSLOF$ and $SSPE = \widehat{\sigma^2}$. Is the quadratic model valid?