Exercises: Simple Gaussian Linear Regression Part II

Luca Danese - l.danese1@campus.unimib.it

October 18, 2024

(Thanks to Valentina Zangirolami and Dr. Roberto Ascari for providing this material)

2 Computer repair data

A computer repair company is interested in knowing the relationship between the duration of interventions (measured in minutes) and the number of electronic components to be replaced or repaired. Therefore, a simple linear regression model was considered to explain the duration in minutes of interventions (y) as a function of the number of units (x) to be replaced.

A sample of 14 interventions provided the following data: $\bar{y} = 95.768$, $\bar{x} = 6$, $\sum_{i=1}^{14}(y_i - \bar{y})^2 = 31108.357$, and $\sum_{i=1}^{14}(x_i - \bar{x})^2 = 114$. The model provides a coefficient of determination $R^2 = 0.984$.

Exercise 2.1
Starting from the data, compute the maximum likelihood estimates of β_1 and β_2. Then, write the equation of the estimated linear regression model.

Exercise 2.2
Find the estimate for the variance σ^2 using the decomposition of the total sum of squares. Through a valid test, verify the goodness of fit at 5% significance level.

Exercise 2.3
Given the standard errors (S.E.) of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$, which correspond to $\sqrt{\hat{Var}(\hat{\beta}_1)} = 4.014$ and $\sqrt{\hat{Var}(\hat{\beta}_2)} = 0.604$. Through a valid test (at 5 % significance level), verify if the coefficients β_1 and β_2 are significant (you can use the following t-table for computing p-values).

Exercise 2.4
Given the ex. 2.2, is there any statistical test in the exercise 2.3 that might be unnecessary?

3 Bacteria mortality data

Suppose we want to analyze bacterial mortality (y) as a function of radiation exposure (x). The output of a linear regression of y as a function of x is partially summarized in the table below:

Table 1: Output of a linear regression.

Variable	**Coefficients**	**S.E.**	**T-value**	**P-value**
Constant	49.162	22.76		
Exposure (x)	-19.46		-7.79	$<$0.0001

where $n = 15$, $R^2 = 0.823$ and $s = 41.83$.

Exercise 3.1

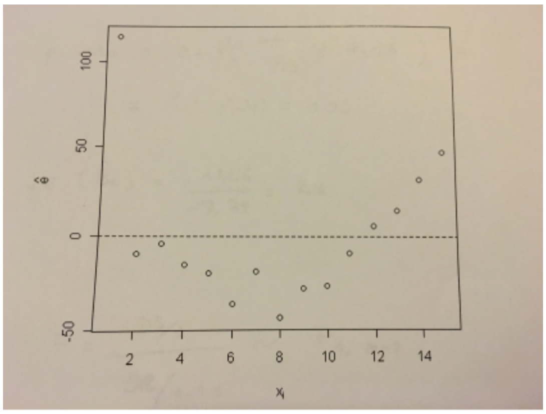
Complete the Table 1 writing the equations you should use.

Exercise 3.2

Through a valid statistical test, evaluate the goodness of fit of the model.

Exercise 3.3

Discuss about the hypothesis related to the model, looking the following residual plot.



4 Grades data

In 2011 among 62 adolescents, the variables x ”daily hours spent on average to video games” and y ”average report card grade” were observed. We proposed a gaussian simple linear model with y, as response variable, and obtained the following estimates: $\hat{\beta}_1 = 7.4$, $\hat{\beta}_2 = -0.48$, $SSE = 223$, $\frac{Var(\hat{\beta}_1)}{\sigma^2} = 3.43$ and $\frac{Var(\hat{\beta}_2)}{\sigma^2} = 0.07$.

Exercise 4.1

Compute the OLS estimates for β_1 and β_2 and provide an explanation.

Exercise 4.2

Through a valid test, use p-values to evaluate if the coefficients are significant (you can use the t-table below for computing p-values). Then, evaluate the goodness of fit using p-values.

5 Additional exercise

A linear regression model was estimated on 82 units. Complete the tables below and specify the hypothesis, the test statistic and p-value for inference.

Table 2: Analysis of variance.

Deviance	**Sum of squares**	**d.o.f.**	**F**	**p-value**
Regression	3589.6		10.21	
Residual		-	-	-
Total		-	-	-

where d.o.f. means degree of freedom.

Table 3: Output of a linear regression.

Variable	**Coefficients**	**S.E.**	**T-value**	**P-value**
Constant	12.7		0.82	
x_1	-19.3			0.002