Exercises: Simple Gaussian Linear Regression Part II

Valentina Zangirolami - valentina.zangirolami@unimib.it

November, 2023

(Thanks to Dr. Roberto Ascari for providing these exercises)
(Referring to the theoretical parts: 4, 5, 6, 7)

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% confidence 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 % confidence level), verify if the coefficients β_1 and β_2 are significant (you can use the following t-table for computing p-values).

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073

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 $\hat{\sigma} = 41.83$.

Exercise 3.1

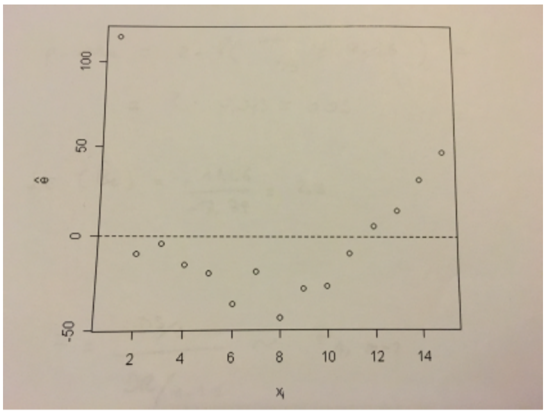
Complete the Table 1 writing the equations you should use.

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.

26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390

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**
Residual	3589.6		10.21	
Regression		-	-	-
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