

Exercises: Simple Gaussian Linear Regression Part I

Valentina Zangirolami - valentina.zangirolami@unimib.it

November 23, 2023

(Referring to the theoretical parts: 4, 5, 6, 7)

(The results obtained in the previous practical part - simple linear regression - can be useful)

1 Mother and Daughter heights data

Let consider a sample of data with $n = 11$ observations (Table 1) with two variables:

- **mother's height** x (independent variable);
- **daughter's height** y (dependent variable).

Table 1: Mother and Daughter heights data: data are expressed in centimeters.

	1	2	3	4	5	6	7	8	9	10	11
x	153.7	156.7	173.5	157.0	161.8	140.7	179.8	150.9	154.4	162.3	166.6
y	163.1	159.5	169.4	158.0	164.3	150.0	170.3	158.9	161.5	160.8	160.6

We would like to find out if there exists a relationship between these two variables.

Exercise 1.5

Starting from the data (in Table 1), write the equation of the gaussian simple linear regression model together with the associated assumptions. Explain the difference from simple linear regression (make a comparison between the assumptions of ex. 1.1 and this case).

Exercise 1.6

Let consider the following system of hypothesis

$$\begin{cases} H_0: \beta_2 = 1 \\ H_1: \beta_2 \neq 1 \end{cases}$$

Compute the t-test and the p-value (use the t-table that you can find below).

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

Exercise 1.7

Let consider the following system of hypothesis

$$\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{cases}$$

Compute the F-test and the p-value (you can find the F-table here <https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/F-Tables/>). In this case, does an equivalent test exists? Specify the hypothesis, the test statistic and compute the p-value.

Exercise 1.8

Provide the confidence intervals for β_r , $r = 1, 2$ at level $1 - \alpha = 0.95$.

Exercise 1.9

During the theoretical lectures, you exploited the relationship between R^2 and the t-test to prove the equivalence among two statistical tests in the case of simple linear regression. Provide the formula and verify that it holds with the data. (In the exercise 1.4, you have already computed the R^2 and the components of its decomposition.)