Introduction to Statistical Learning and Applications
Linear Regression Exercises

Razan MHANNA

February 20, 2024

Exercise 1

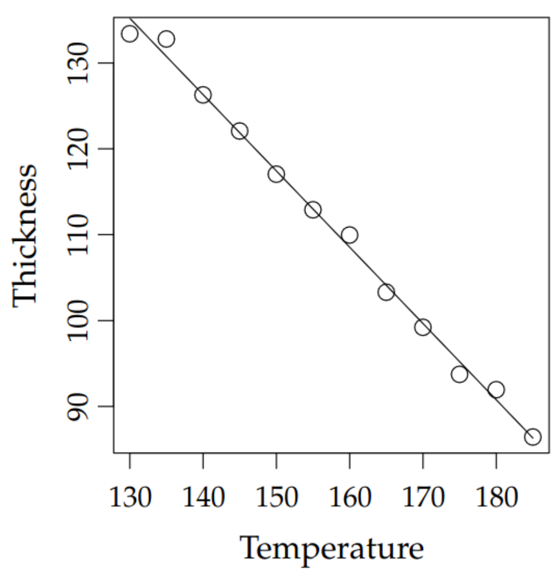
Given the following data points, we want to estimate the relation $y_t = a_0 + a_1 x_1 + \epsilon_t$, t=1 ... n.

x_1	54	53	59	66	63	62	65	60	59	65	70	65
y_1	20	19	21	21	23	20	25	24	28	27	31	33

1)Calculate the estimations of a_0 an a_1. Give the serie of residuals.
2) Calculate the standard error of of a_0 an a_1.
3) Calculate the coeffficient of correlation r_{xy}, is it significantly different than 0(do a hypothesis test)
4) Test the null hypothesis H_0: a_1=0 against $a_1 \neq$0. Why this hyppthesis is important to test? Give a confidence interval for a_1 with 95
5) Test the null hypothesis H_0: a_1=0.5 against $a_1$0.5 . Then the hypothesis H_0: a_1=0.5 against a_1 ¿ 0.5.
6) Calculate SST; SSR and SSE. Deduce R. Conclude.

Exercise 2

On a machine that folds plastic film the temperature may be varied in the range of 130-185 °C. For obtaining, if possible, a model for the influence of temperature on the folding thickness, n = 12 related set of values of temperature and the fold thickness were measured that is illustrated in the following figure:



a) Determine by looking at the figure, which of the following sets of estimates for the parameters in

the usual regression model is correct:
1)$\hat{\beta}_0 = 0, \hat{\beta}_1 = -0.9, \hat{\sigma} = 36$
2)$\hat{\beta}_0 = 0, \hat{\beta}_1 = 0.9, \hat{\sigma} = 3.6$
3)$\hat{\beta}_0 = 252, \hat{\beta}_1 = -0.9, \hat{\sigma} = 3.6$
4)$\hat{\beta}_0 = -252, \hat{\beta}_1 = -0.9, \hat{\sigma} = 36$
5)$\hat{\beta}_0 = 252, \hat{\beta}_1 = -0.9, \hat{\sigma} = 36$

b) What is the only possible correct answer:
1) The proportion of explained variation is 50% and the correlation is 0.98
2) The proportion of explained variation is 0% and the correlation is 0.98
3) The proportion of explained variation is 96% and the correlation is 1
4) The proportion of explained variation is 96% and the correlation is 0.98
5) The proportion of explained variation is 96% and the correlation is 0.98

Exercise 3
We measure the annual income R_t and the annual savings E_t over a period of 12 years, for a given socio-professional category. The results are obtained (in thousands of euros):
$\overline{R}=\frac{1}{12}\sum_{i=1}^{12} R_i = 19.7$
$\overline{E}=\frac{1}{12}\sum_{i=1}^{12} E_i = 6.1$
$\sum_{i=1}^{12} \mathrm{R}_i^2 = 4827$
$\sum_{i=1}^{12} \mathrm{E}_i^2 = 456$
$\sum_{i=1}^{12} R_i E_i = 1480$
The model equation $E_t = aR_t + b + \epsilon_t$
1) Calculate $\hat{a}$ and $\hat{b}$.
2) Test the hypothesis H_0: a=0.
3) Calculate the coefficient of determination. Conclude.

Exercise 4
Having $y_t = 1.251x_t - 32.82 + e_t$, n=20, R=0.23, $\hat{\sigma_\epsilon}$=10.66.
1) Calculate SST, SSR, SSE, the law of student and the standard error of $\hat{a_1}$.
2] Is the coefficient of the variable x significantly >1 ?

Exercise 5
A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device, and a study is therefore performed in which the life time is determined as a function of temperature. The following data is found:

Temperature in Celcius (t)	10	20	30	40	50	60	70	80	90
Lite time in hours(y)	420	365	285	220	176	117	69	34	5

1) Calculate the 95% confidence interval for the slope in the usual linear regression model, which expresses the life time as a linear function of the temperature.
2)Can a relation between temperature and life time be documented on level 5%?