Statistics A – Summer term 2023

Excercise 4

1. Properties of the linear regression.
   (a) Show $\bar{y} = a + b\bar{x}$
   (b) Show $b = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$ and $a = \frac{\overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{xy}}{\overline{x^2} - \bar{x}^2}$
   (c) Show $\sum_{i=1}^{n} \epsilon_i = 0$
   (d) Show $\sum_{i=1}^{n} \epsilon_i x_i = 0$
   (e) Show $\sum_{i=1}^{N} (y_i - \overline{y})^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$
2. A housing agent sells last April 26 flats with the following data regarding size G in $[m^2]$ and rental M in [€]. We know $Cov(G, M) = 5760$ $\sigma_M = 223,61$ $\sigma_G = 21,9$ $\mu_M = 1100$ $\mu_G = 80$
   (a) Calculate and draw the regression line.
   (b) Calculate the average and marginal price per m^2.
   (c) Estimate the rental of flat of size 100 m^2.
3. Within a random sample of 100 data points (X, Y) we have $R = -0,93$. Decide which statement is right or wrong?
   (1) There is no dependence between X and Y since R is negative.
   (2) The data points scattering very near around the regression line.
   (3) The linear regression has positive slope.
   (4) The dependence of X and Y is inversely proportional.

4. In 2021 a brewery has the following output and costs:

Month	Output [Hektoliter]	Costs [€]
Jan	600	6500
Feb	680	8200
Mrz	720	7300
Apr	1010	8900
Mai	900	9900
Jun	990	10000
Jul	1270	10300
Aug	1440	12500
Sep	1380	11500
Okt	1010	9200
Nov	830	8200
Dez	1070	9300

(a) Calculate the Cost function via a linear regression.

(b) Interprete economically the parameters.

(c) Calculate the correlation between Output and costs?

(d) Estimate the total costs and average variable costs of an output of 1100 Hektoliter.