List 11

Exercise 1. One company investigated the relationship between the company's weekly income and television advertising expenditure and radio advertising expenditure. The table shows the data (in EUR thousand):

TV	12	10	15	9	12	18	14	8	9	20
Radio	15	17	6	6	4	9	1	4	3	13
Income	70	77	81	70	66	82	68	61	65	85

a) On the basis of graphs of points with coordinates (x, y) from the data, check what the relationship between the examined features may be. Note: plots for two independent variables should be made separately.
b) Determine the correlation coefficients between the company's weekly income and television advertising expenditure and between the company's weekly income and radio advertising expenditure.
c) Calculate the linear regression equation of the company's weekly income depending television and radio advertising expenditures.
d) Check the significance of multiple linear regression.
e) Determine the company's weekly income without any advertising and with television advertising expenditure equals to 16.000 euro and radio advertising expenditure equals to 6.000 euro.
f) Determine and interpret the coefficient of multiple determination.

Take a significance level equal to 0.05.

Exercise 2. Health researcher are interested in factors that influence heart diseases. The scientist examined 498 towns and gather data on the percentage of people in each town who smoke, the percentage of people in each town who do sports, and the percentage of people in each town who have heart disease.

a) Load the file heart.data.csv.
b) On the basis of graphs of points with coordinates (x, y) from the data, check what the relationship between the examined features may be. Note: plots for two independent variables should be made separately.
c) Determine the correlation coefficients between the percentage of people who do sport and the percentage of people who have heart diseases and between the percentage of people who smoke and the percentage of people who have heart diseases.
d) Calculate the linear regression equation of the percentage of people who have heart diseases depending the percentage of people who do sport and smoke.
e) Check the significance of multiple linear regression.
f) Predict the percentage of people who have heart diseases when the percentage of people who do sport is equal to 25% and the percentage of people who smoke is 56%.
g) Determine and interpret the coefficient of multiple determination.

Take a significance level equal to 0.05.

Exercise 3. The dependence of infant height (in cm) on several factors was examined: age (in days), height at birth (in cm), and weight at birth (in kg). The following data was obtained:

Height	57.5	52.8	61.3	67.0	53.6	62.7	56.3	68
Age	77	68	73	86	66	81	75	93
Birth height	45.4	46.8	46.3	49.0	43.8	45.9	56.1	56.1
Birth weight	2.7	2.2	4.3	5.1	3.3	4.5	2.6	4.4

a) On the basis of graphs of points with coordinates (x, y) from the data, check what the relationship between the examined features may be. Note: plots for three independent variables should be made separately.
b) Determine the correlation coefficients between the height and age, birth height and birth weight separately.
c) Calculate the linear regression equation of the infant height depending on age, birth height and birth weight.
d) Check the significance of multiple linear regression.
e) Determine the height depending on 75 days old person with birth height equals to 47.5 cm and birth weight equals to 4 kg.
f) Determine and interpret the coefficient of multiple determination.

Take a significance level equal to 0.05.

Exercise 4. The behavior of the newly produced car while driving was investigated. For this purpose, the distance travelled (in m) was measured depending on the driving time (in s). The results are presented in the table below:

x_i	5	7	11	14	20	24	30
y_i	45	53	82	143	332	635	901

a) On the basis of the graph of points with coordinates (x, y) from the table, check what type of relationship between the examined features can be expected.
b) Determine the correlation coefficient between driving time and distance travelled by car.
c) Calculate the polynomial regression equation for the distance travelled depending on driving time.
d) Decide on the degree of the regression polynomial and check the significance of regression coefficients.
e) Add a polynomial regression functions to an existing plot.
f) Determine the distance travelled after 17 and 50 seconds.
g) Determine and interpret the coefficient of determination.

Take a significance level equal to 0.05.

Exercise 5. It was decided to test the processor temperature [^{o}C] depending on the computer operation time [in h]. The results are presented in the table:

x_i	1	4	7	8	10	13	18	22	25	28
y_i	35	45	59	77	62	50	39	47	54	71

a) On the basis of the graph of points with coordinates (x, y) from the table, check what type of relationship between the examined features can be expected.
b) Determine the correlation coefficient between the processor temperature and the computer operation time.
c) Calculate the polynomial regression equation for the processor temperature depending on computer operation time.
d) Decide on the degree of the regression polynomial and check the significance of regression coefficients.
e) Add a polynomial regression functions to an existing plot.
f) Determine the temperature after 15 and 48 hours.
g) Determine and interpret the coefficient of determination.

Take a significance level equal to 0.05.

Exercise 6. The speed of the motorcycle (in km/h) at a certain turn of the racetrack was analyzed depending on the percentage of the distance covered (in %) by a motorcyclist at that turn. The results are presented in the table:

x_i	5	15	30	40	55	65	80	90
y_i	130	98	78	65	58	73	104	147

a) On the basis of the graph of points with coordinates (x, y) from the table, check what type of relationship between the examined features can be expected.
b) Determine the correlation coefficient between the speed of the motorcycle and the percentage of the distance covered at that turn.
c) Calculate the polynomial regression equation for the speed of the motorcycle depending on the percentage of the distance covered at that turn.
d) Decide on the degree of the regression polynomial and check the significance of regression coefficients.
e) Add a polynomial regression functions to an existing plot.
f) Determine the speed of the motorcycle after completing 25% of that turn.
g) Determine and interpret the coefficient of determination.

Take a significance level equal to 0.05.