Simple Linear Regression

Problem 1: The values of y and their corresponding values of y are shown in the table below

x	0	1	2	3	4
y	2	3	5	4	6

a) Compute SS_{xx}, SS_{yy}, and SS_{xy}.
b) Find the least squares regression line by choosing appropriate dependent and independent variables based on your answer in part a.
c) Interpret the meaning of the values of a and b calculated in part c.
d) Plot the scatter diagram and the regression line. (use a graphing paper)
e) Calculate r and r^2 and explain what they mean.
f) Predict y for x = 7.
h) Compute the standard deviation of errors.
i) Construct a 95% confidence interval for B.

13.25 An auto manufacturing company wanted to investigate how the price of one of its car models depreciates with age. The research department at the company took a sample of eight cars of this model and collected the following information on the ages (in years) and prices (in hundreds of dollars) of these cars.

Age	8	3	6	9	2	5	6	3
Price	45	210	100	33	267	134	109	235

a. Construct a scatter diagram for these data. Does the scatter diagram exhibit a linear relationship between ages and prices of cars?
b. Find the regression line with price as a dependent variable and age as an independent variable.
c. Give a brief interpretation of the values of a and b calculated in part b.
d. Plot the regression line on the scatter diagram of part a and show the errors by drawing vertical lines between scatter points and the regression line.
e. Predict the price of a 7-year-old car of this model.
f. Estimate the price of an 18-year-old car of this model. Comment on this finding.

13.26 The following table gives information on the amount of sugar (in grams) and the calorie count in one serving of a sample of 13 varieties of Kellogg's cereal.

Sugar (grams)	4	15	12	11	8	6	7	2	7	14	20	3	13
Calories	120	200	140	110	120	80	190	100	120	190	190	110	120

Source: kelloggs.com.

a. Construct a scatter diagram for these data. Does the scatter diagram exhibit a linear relationship between the amount of sugar and the number of calories per serving?
b. Find the predictive regression equation of the number of calories on the amount of sugar.
c. Give a brief interpretation of the values of a and b calculated in part b.
d. Plot the predictive regression line on the scatter diagram of part a and show the errors by drawing vertical lines between scatter points and the predictive regression line.
e. Calculate the predicted calorie count for a cereal with 16 grams of sugar per serving.
f. Estimate the calorie count for a cereal with 52 grams of sugar per serving. Comment on this finding.

13.27 A diabetic is interested in determining how the amount of aerobic exercise impacts his blood sugar. When his blood sugar reaches 170 mg/dL, he goes out for a run at a pace of 10 minutes per mile. On different days, he runs different distances and measures his blood sugar after completing his run. Note: The preferred blood sugar level is in the range of 80 to 120 mg/dL. Levels that are too low or too high are extremely dangerous. The data generated are given in the following table.

Distance (miles)	2	2	2.5	2.5	3	3	3.5	3.5	4	4	4.5	4.5
Blood sugar (mg/dL)	136	146	131	125	120	116	104	95	85	94	83	75

a. Construct a scatter diagram for these data. Does the scatter diagram exhibit a linear relationship between distance run and blood sugar level?
b. Find the predictive regression equation of blood sugar level on the distance run.
c. Give a brief interpretation of the values of a and b calculated in part b.
d. Plot the predictive regression line on the scatter diagram of part a and show the errors by drawing vertical lines between scatter points and the predictive regression line.
e. Calculate the predicted blood sugar level count after a run of 3.1 miles (5 kilometers).
f. Estimate the blood sugar level after a 10-mile run. Comment on this finding.

13.28 While browsing through the magazine rack at a bookstore, a statistician decides to examine the relationship between the price of a magazine and the percentage of the magazine space that contains advertisements. The data are given in the following table.

Percentage containing ads	37	43	58	49	70	28	65	32
Price ($)	5.50	6.95	4.95	5.75	3.95	8.25	5.50	6.75

a. Construct a scatter diagram for these data. Does the scatter diagram exhibit a linear relationship between the percentage of a magazine's space containing ads and the price of the magazine?
b. Find the estimated regression equation of price on the percentage containing ads.
c. Give a brief interpretation of the values of a and b calculated in part b.
d. Plot the estimated regression line on the scatter diagram of part a, and show the errors by drawing vertical lines between scatter points and the predictive regression line.
e. Predict the price of a magazine with 50% of its space containing ads.
f. Estimate the price of a magazine with 99% of its space containing ads. Comment on this finding.

13.30 The following table gives the total 2008 payroll (on the opening day of the season, rounded to the nearest million dollars) and the number of runs scored during the 2008 season by each of the American League baseball teams.

Team	Total Payroll (millions of dollars)	Runs Scored
Baltimore Orioles	67	782
Boston Red Sox	123	845
Chicago White Sox	96	811
Cleveland Indians	82	805
Detroit Tigers	115	821
Kansas City Royals	71	691
Los Angeles Angels	114	765
Minnesota Twins	65	829
New York Yankees	201	789
Oakland Athletics	62	646
Seattle Mariners	99	671
Tampa Bay Rays	63	774
Texas Rangers	69	901
Toronto Blue Jays	81	714

Source: ESPN.com.

a. Find the least squares regression line with total payroll as the independent variable and runs scored as the dependent variable.
b. Is the equation of the regression line obtained in part a the population regression line? Why or why not? Do the values of the y-intercept and the slope of the regression line give A and B or a and b?
c. Give a brief interpretation of the values of the y-intercept and the slope obtained in part a.
d. Predict the number of runs scored by a team with a total payroll of \$84 million.

13.80 The owner of a small factory that produces working gloves is concerned about the high cost of air conditioning in the summer but is afraid that keeping the temperature in the factory too high will lower productivity. During the summer, he experiments with temperature settings from 68°F to 81°F and measures each day's productivity. The following table gives the temperature and the number of pairs of gloves (in hundreds) produced on each of the 8 randomly selected days.

Temperature (°F)	72	71	78	75	81	77	68	76
Pairs of gloves	37	37	32	36	33	35	39	34

a. Do the pairs of gloves produced depend on temperature, or does temperature depend on pairs of gloves produced? Do you expect a positive or a negative relationship between these two variables?
b. Taking temperature as an independent variable and pairs of gloves produced as a dependent variable, compute SS_{xx}, SS_{yy}, and SS_{xy}.
c. Find the least squares regression line.
d. Interpret the meaning of the values of a and b calculated in part c.
e. Plot the scatter diagram and the regression line.
f. Calculate r and r^2, and explain what they mean.
g. Compute the standard deviation of errors.
h. Predict the number of pairs of gloves produced when $x = 74$.
i. Construct a 99% confidence interval for B.
j. Test at the 5% significance level whether B is negative.
k. Using $\alpha = .01$ can you conclude that ρ is negative?

13.81 The following table gives information on the limited tread warranties (in thousands of miles) and the prices of 12 randomly selected tires at a national tire retailer as of July 2009.

Warranty (thousands of miles)	60	70	75	50	80	55	65	65	70	65	60	65
Price per tire ($)	95	70	94	90	121	70	84	80	92	79	66	95

a. Taking warranty length as an independent variable and price per tire as a dependent variable, compute SS_{xx}, SS_{yy}, and SS_{xy}.
b. Find the regression of price per tire on warranty length.
c. Briefly explain the meaning of the values of a and b calculated in part b.
d. Calculate r and r^2 and explain what they mean.
e. Plot the scatter diagram and the regression line.
f. Predict the price of a tire with a warranty length of 73,000 miles.
g. Compute the standard deviation of errors.
h. Construct a 95% confidence interval for B.
i. Test at the 5% significance level if B is positive.
j. Using $\alpha = .025$, can you conclude that the linear correlation coefficient is positive?

13.83 The following table gives information on the incomes (in thousands of dollars) and charitable contributions (in hundreds of dollars) for the last year for a random sample of 10 households.

Income	Charitable Contributions
76	15
57	4
140	42
97	33
75	5
107	32
65	10
77	18
102	28
53	4

a. With income as an independent variable and charitable contributions as a dependent variable, compute SS_{xx}, SS_{yy}, and SS_{xy}.
b. Find the regression of charitable contributions on income.
c. Briefly explain the meaning of the values of a and b.
d. Calculate r and r^2 and briefly explain what they mean.
e. Compute the standard deviation of errors.
f. Construct a 99% confidence interval for B.
g. Test at the 1% significance level whether B is positive.
h. Using the 1% significance level, can you conclude that the linear correlation coefficient is different from zero?

15.5 Use the following time-series data to answer the given questions.

Time Period	Value	Time Period	Value
1	27	6	66
2	31	7	71
3	58	8	86
4	63	9	101
5	59	10	97

a. Develop forecasts for periods 5 through 10 using 4-month moving averages.

b. Calculate MAD, MSE and MAPE from the forecast errors.

15.6 Following are time-series data for eight different periods. Use exponential smoothing to forecast the values for periods 3 through 8. Use the value for the first period as the forecast for the second period. Compute forecasts using two different values of alpha, $\alpha = .1$ and $\alpha = .8$. Compute the errors for each forecast and compare the errors produced by using the two different exponential smoothing constants.

Time Period	Value	Time Period	Value
1	211	5	242
2	228	6	227
3	236	7	217
4	241	8	203

15.8 The U.S. Census Bureau publishes data on factory orders for all manufacturing, durable goods, and nondurable goods industries. Shown here are factory orders in the United States over a 13-year period ($ billion).

a. Use these data to develop forecasts for the years 6 through 13 using a 5-year moving average.

Year	Factory Orders ($ billion)
1	2,512.7
2	2,739.2
3	2,874.9
4	2,934.1
5	2,865.7
6	2,978.5
7	3,092.4
8	3,356.8
9	3,607.6
10	3,749.3
11	3,952.0
12	3,949.0
13	4,137.0

b. Calculate MAD, MSE and MAPE.

Questions from Multiple Regression will be similar to the lab problem.