Tae Coding
Introduction to Data Science: CS61
Summer 2018
Class Exercise#5

Date Given: June 26, 2018   Due Date:
===

There are 4 problems in this assignment.

Please add the following table on the first page of your submission.

Problem Number	Answer
1	Answers:
2	Answers:
3	Answers:
4	Answers:

This assignment comprises 6 problems.

Problem#1

Linear Regression

The values of 2 variables X and Y are given below. Here X is the predictor variable and Y is the response variable.

X	25	30	35	40	45
Y	5	260	480	745	1100

Build your regression model using the following 3 methods.
  a) Closed form solution – using only the mean of 'x', 'y', 'x*y', 'x^2' variables.
  b) Closed form solution – using the correlation coefficient between 'x' and 'y' variables and the standard deviation of both variables.
  c) Closed form solution – using matrix (use Python for matrix manipulation)
  d) Python: Scikit-Learn

Make sure that your answers are the same using all the 3 methods.

Solve Problem#2 + #3 + #4 in Python.

Problem#2

17. You Explain It! Video Games and GPAs A student at Joliet
NW Junior College conducted a survey of 20 randomly selected
full-time students to determine the relation between the
number of hours of video game playing each week, x, and
grade-point average, y. She found that a linear relation exists
between the two variables. The least-squares regression line
that describes this relation is $\hat{y} = -0.0526x + 2.9342$.

(a) Predict the grade-point average of a student who plays
video games 8 hours per week.

(b) Interpret the slope.

(c) If appropriate, interpret the y-intercept.

(d) A student who plays video games 7 hours per week has a
grade-point average of 2.68. Is this student's grade-point
average above or below average among all students who
play video games 7 hours per week?

Problem#3

21. **Height versus Head Circumference** (Refer to Problem 23, Section 4.1) A pediatrician wants to determine the relation that exists between a child's height, x, and head circumference, y. She randomly selects 11 children from her practice, measures their heights and head circumferences and obtains the following data.

Height, x (inches)	Head Circumference, y (inches)	Height, x (inches)	Head Circumference, y (inches)
27.75	17.5	26.5	17.3
24.5	17.1	27	17.5
25.5	17.1	26.75	17.3
26	17.3	26.75	17.5
25	16.9	27.5	17.5
27.75	17.6		

Source: Denise Slucki, student at Joliet Junior College

(a) Find the least-squares regression line treating height as the explanatory variable and head circumference as the response variable.
(b) Interpret the slope and y-intercept, if appropriate.
(c) Use the regression equation to predict the head circumference of a child who is 25 inches tall.
(d) Compute the residual based on the observed head circumference of the 25-inch-tall child in the table. Is the head circumference of this child above average or below average?

Problem#4

23. **Weight of a Car versus Miles per Gallon** (Refer to Problem 25, Section 4.1) An engineer wants to determine how the weight of a car, x, affects gas mileage, y. The following data represent the weights of various domestic cars and their miles per gallon in the city for the 2008 model year.

Car	Weight (pounds), x	Miles per Gallon, y
Buick Lucerne	3,765	19
Cadillac DeVille	3,984	18
Chevrolet Malibu	3,530	21
Chrysler Sebring Sedan	3,175	22
Dodge Neon	2,580	27
Dodge Charger	3,730	18
Ford Focus	2,605	26
Lincoln LS	3,772	17
Mercury Sable	3,310	20
Pontiac G5	2,991	25
Saturn Ion	2,752	26

Source: www.roadandtrack.com

(a) Find the least-squares regression line treating weight as the explanatory variable and miles per gallon as the response variable.

(b) Interpret the slope and y-intercept, if appropriate.

(c) A Chevy Cobalt weighs 2,780 pounds and gets 22 miles per gallon. Is the miles per gallon of a Cobalt above average or below average for cars of this weight?

(d) Would it be reasonable to use the least-squares regression line to predict the miles per gallon of a Toyota Prius, a hybrid gas and electric car? Why or why not?