BA Homework #4

Business Analytics - Spring 2020, NYU

INSTRUCTIONS & SUBMISSION

Complete the problems below and submit via NYUClasses. Submit a PDF file with your answers, graphs, and R code.

Problem 1: Bicycling World Problem

1. *Bicycling World*, a magazine devoted to cycling, reviews hundreds of bicycles throughout the year. Its Road-Race category contains reviews of bicycles used by riders primarily interested in racing. One of the most important factors in selecting a bicycle for racing is its weight. The following data show the weight (pounds) and price ($) for ten racing bicycles reviewed by the magazine:

Model	Weight (lbs)	Price ($)
Fierro 7B	17.9	2200
HX 5000	16.2	6350
Durbin Ultralight	15.0	8470
Schmidt	16.0	6300
WSilton Advanced	17.3	4100
bicyclette vélo	13.2	8700
Supremo Team	16.3	6100
XTC Racer	17.2	2680
D'Onofrio Pro	17.7	3500
Americana #6	14.2	8100

a. Develop a scatter chart with weight as the independent variable. What does the scatter chart indicate about the relationship between the weight and price of these bicycles?

b. Use the data to develop an estimated regression equation that could be used to estimate the price for a bicycle, given its weight. What is the estimated regression model?

c. Test whether each of the regression parameters β_0 and β_1 is equal to zero at a 0.05 level of significance. What are the correct interpretations of the estimated regression parameters? Are these interpretations reasonable?

d. How much of the variation in the prices of the bicycles in the sample does the regression model you estimated in part b explain?

Problem 2: Assembly Line Problem

2. In a manufacturing process the assembly line speed (feet per minute) was thought to affect the number of defective parts found during the inspection process. To test this theory, managers devised a situation in which the same batch of parts was inspected visually at a variety of line speeds. They collected the following data:

Line Speed (ft/min)	Number of Defective Parts Found
20	21
20	19
40	15
30	16
60	14
40	17

a. Develop a scatter chart with line speed as the independent variable. What does the scatter chart indicate about the relationship between line speed and the number of defective parts found?

b. Use the data to develop an estimated regression equation that could be used to predict the number of defective parts found, given the line speed. What is the estimated regression model?

c. Test whether each of the regression parameters β_0 and β_1 is equal to zero at a 0.01 level of significance. What are the correct interpretations of the estimated regression parameters? Are these interpretations reasonable?

d. How much of the variation in the number of defective parts found for the sample data does the model you estimated in part b explain?

Problem 3: Jensen Tire & Auto Problem

3. Jensen Tire & Auto is deciding whether to purchase a maintenance contract for its new computer wheel alignment and balancing machine. Managers feel that maintenance expense should be related to usage, and they collected the following information on weekly usage (hours) and annual maintenance expense (in hundreds of dollars).

Weekly Usage (hours)	Annual Maintenance Expense ($100s)
13	17.0
10	22.0
20	30.0
28	37.0
32	47.0
17	30.5
24	32.5
31	39.0
40	51.5
38	40.0

a. Develop a scatter chart with weekly usage hours as the independent variable. What does the scatter chart indicate about the relationship between weekly usage and annual maintenance expense?

b. Use the data to develop an estimated regression equation that could be used to predict the annual maintenance expense for a given number of hours of weekly usage. What is the estimated regression model?

Problem 4: Toyota Problem

The Toyota Camry is one of the best-selling cars in North America. The cost of a previously owned Camry depends on many factors, including the model year, mileage, and condition. To investigate the relationship between the car's mileage and the sales price for Camrys, the following data show the mileage and sale price for 19 sales (PriceHub Web site, February 24, 2012).

Miles (1000s)	Price ($1,000s)
22	16.2
29	16.0
36	13.8
47	11.5
63	12.5
77	12.9
73	11.2
87	13.0
92	11.8
101	10.8
110	8.3
28	12.5
59	11.1
68	15.0
68	12.2
91	13.0
42	15.6
65	12.7
110	8.3

a. Develop a scatter chart for these data with miles as the independent variable. What does the scatter chart indicate about the relationship between price and miles?

b. Develop an estimated regression equation showing how price is related to miles. What is the estimated regression model?

c. Test whether each of the regression parameters β_0 and β_1 is equal to zero at a 0.01 level of significance. What are the correct interpretations of the estimated regression parameters? Are these interpretations reasonable?

d. How much of the variation in the sample values of price does the model estimated in part b explain?

e. For the model estimated in part b, calculate the predicted price and residual for each automobile in the data. Identify the two automobiles that were the biggest bargains.

f. Suppose that you are considering purchasing a previously owned Camry that has been driven 60,000 miles. Use the estimated regression equation developed in part b to predict the price for this car. Is this the price you would offer the seller?

Problem 5: Dodger Stadium Attendance

The Dodger Stadium, with a capacity for 56,000, is the largest ballpark is the world. The data consists of the following:

month	Month of year in which game took place
day	Day of month in which game took place
attend	Total attendance recorded
day_of_week	Day of week in which game took place
opponent	Visitor team played
temp	Temperature recorded
skies	Overcast (clear/cloudy) during the game
day_night	Binary of when game was played (night vs day)
cap	Binary for promotion: cap
shirt	Binary for promotion: shirt
fireworks	Binary for promotion: fireworks
bobblehead	Binary for promotion: bobblehead

Link to dataset:
https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/dodgers.csv

Questions:

1. Complete an exploratory data analysis and answer the following:
    a. How many times did promotions take place during the year (cap vs shirts vs bobblehead vs fireworks)?
    b. How does attendance vary *with and without* promotions
    c. What patterns exist with programming of games (weather, time, month, day, etc)?
    d. Which opposing teams bring is attendance above average?.
2. Answer the following questions using predictive modeling techniques:
    a. Will the bobblehead promotions increase attendance?
    b. Are bobblehead promotions better than all other promotions?
    c. Giving your predictions, how many bobblehead should we ordered for the summer time (Jun - Aug)