Business Analytics Yue Summer 2025

Case Problems #3, Due 7/16/2025 or 7/17/2025

1. Solve one of the following two case studies:

CASE PROBLEM 3: SELECTING A POINT-AND-SHOOT DIGITAL CAMERA

Consumer Reports tested 166 different point-and-shoot digital cameras. Based upon factors such as the number of megapixels, weight (g), image quality, and ease of use, they developed an overall score for each camera tested. The overall score ranges from 0 to 100, with higher scores indicating better overall test results. Selecting a camera with many options can be a difficult process, and price is certainly a key issue for most consumers. By spending more, will a consumer really get a superior camera? And, do cameras that have more megapixels, a factor often considered to be a good measure of picture quality, cost more than cameras with fewer megapixels? Table 14.13 shows the brand, average retail price ($), number of megapixels, weight (g), and the overall score for 13 Canon and 15 Nikon subcompact cameras tested by *Consumer Reports* (*Consumer Reports* website).

Managerial Report

1. Develop numerical summaries of the data.
2. Using overall score as the dependent variable, develop three scatter diagrams, one using price as the independent variable, one using the number of megapixels as the independent variable, and one using weight as the independent variable. Which of the three independent variables appears to be the best predictor of overall score?

Observation	Brand	Price ($)	Megapixels	Weight (g)	Score
1	Canon	330	10	198	66
2	Canon	200	12	142	66
3	Canon	300	12	198	65
4	Canon	200	10	170	62
5	Canon	180	12	142	62
6	Canon	200	12	198	61
7	Canon	200	14	142	60
8	Canon	130	10	198	60
9	Canon	130	12	142	59
10	Canon	110	16	142	55
11	Canon	90	14	142	52
12	Canon	100	10	170	51
13	Canon	90	12	198	46
14	Nikon	270	16	142	65
15	Nikon	300	16	198	63
16	Nikon	200	14	170	61
17	Nikon	400	14	198	59
18	Nikon	120	14	142	57
19	Nikon	170	16	170	56
20	Nikon	150	12	142	56
21	Nikon	230	14	170	55
22	Nikon	180	12	170	53
23	Nikon	130	12	170	53
24	Nikon	80	12	198	52
25	Nikon	80	14	198	50
26	Nikon	100	12	113	46
27	Nikon	110	12	142	45
28	Nikon	130	14	113	42

3. Using simple linear regression, develop an estimated regression equation that could be used to predict the overall score given the price of the camera. For this estimated regression equation, perform an analysis of the residuals and discuss your findings and conclusions.
4. Analyze the data using only the observations for the Canon cameras. Discuss the appropriateness of using simple linear regression and make any recommendations regarding the prediction of overall score using just the price of the camera.

CASE PROBLEM 4: FINDING THE BEST CAR VALUE

When trying to decide what car to buy, real value is not necessarily determined by how much you spend on the initial purchase. Instead, cars that are reliable and don't cost much to own often represent the best values. But, no matter how reliable or inexpensive a car may cost to own, it must also perform well.

To measure value, *Consumer Reports* developed a statistic referred to as a value score. The value score is based upon five-year owner costs, overall road-test scores, and predicted reliability ratings. Five-year owner costs are based on the expenses incurred in the first five years of ownership, including depreciation, fuel, maintenance and repairs, and so on. Using a national average of 19,300 kilometers per year, an average cost per kilometer driven is used as the measure of five-year owner costs. Road-test scores are the results of more than 50 tests and evaluations and are based upon a 100-point scale, with higher scores indicating better performance, comfort, convenience, and fuel economy. The highest road-test score obtained in the tests conducted by *Consumer Reports* was a 99 for a Lexus LS 460L. Predicted-reliability ratings (1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, and 5 = Excellent) are based on data from *Consumer Reports'* Annual Auto Survey.

A car with a value score of 1.0 is considered to be "average-value." A car with a value score of 2.0 is considered to be twice as good a value as a car with a value score of 1.0; a car with a value score of .5 is considered half as good as average; and so on. The data for 20 family sedans, including the price ($) of each car tested, follow.

Car	Price ($)	Cost/Kilometer	Road-Test Score	Predicted Reliability	Value Score
Nissan Altima 2.5 S (4-cyl.)	23,970	.37	91	4	1.75
Kia Optima LX (2.4)	21,885	.36	81	4	1.73
Subaru Legacy 2.5i Premium	23,830	.37	83	4	1.73
Ford Fusion Hybrid	32,360	.39	84	5	1.70
Honda Accord LX-P (4-cyl.)	23,730	.35	80	4	1.62
Mazda6 i Sport (4-cyl.)	22,035	.36	73	4	1.60
Hyundai Sonata GLS (2.4)	21,800	.35	89	3	1.58
Ford Fusion SE (4-cyl.)	23,625	.35	76	4	1.55
Chevrolet Malibu LT (4-cyl.)	24,115	.35	74	3	1.48
Kia Optima SX (2.0T)	29,050	.45	84	4	1.43
Ford Fusion SEL (V6)	28,400	.42	80	4	1.42
Nissan Altima 3.5 SR (V6)	30,335	.43	93	4	1.42
Hyundai Sonata Limited (2.0T)	28,090	.41	89	3	1.39
Honda Accord EX-L (V6)	28,695	.42	90	3	1.36
Mazda6 s Grand Touring (V6)	30,790	.46	81	4	1.34
Ford Fusion SEL (V6, AWD)	30,055	.44	75	4	1.32
Subaru Legacy 3.6R Limited	30,094	.44	88	3	1.29
Chevrolet Malibu LTZ (V6)	28,045	.42	83	3	1.20
Chrysler 200 Limited (V6)	27,825	.44	52	5	1.20
Chevrolet Impala LT (3.6)	28,995	.42	63	3	1.05

Managerial Report

1. Develop numerical summaries of the data.
2. Use regression analysis to develop an estimated regression equation that could be used to predict the value score given the price of the car.
3. Use regression analysis to develop an estimated regression equation that could be used to predict the value score given the five-year owner costs (cost/kilometer).
4. Use regression analysis to develop an estimated regression equation that could be used to predict the value score given the road-test score.
5. Use regression analysis to develop an estimated regression equation that could be used to predict the value score given the predicted-reliability.
6. What conclusions can you derive from your analysis?

2. (a) What are assumptions for the error terms of regression analysis and explain why they are important.

   (b) Explain what is Simpson's Paradox.

   (c) Use examples in regression analysis to demonstrate what is Simpson's Paradox and provide suggestions for conducting regression analysis.