**Assignments**

**Please try to carry out the exercises on your own and make presentation. The presentation will be discussed on Friday morning (3rd March 2017) when we meet.**

1. To study the relationship between the length of a service call and number of components in the computer that must be repaired or replaced, a sample of records on service calls were taken and given in repair worksheet. Does the length of service call depends on number of components repaired/replaced? If yes then what will be the length of a service call (with 95% interval) when the number of components to be repaired/replaced is 5 and 12.
2. The data in worksheet ‘Stay’ represent the length of service in days (X) and cost per day in dollars (Y) for 32 patients on a home health care program. Develop a regression model and interpret. What will be the expected cost per day if length of service is 55 and 250 days?
3. The data in worksheet ‘advertisement’ appeared in the Wall Street Journal. The advertisement were selected by an annual survey conducted by Video Board Tests, Inc., a New York ad-testing company, based on interviews with 20,000 adults who were asked to name the most outstanding TV commercial they had seen, noticed, and liked. The retained impressions were based on a survey of 4,000 adults, in which regular product users were asked to cite a commercial they had seen for that product category in the past week. The datasheet provides the following information in three column: FIRM: Firm name, SPEND: TV advertising budget, 1983 ($ millions) and MILIMP: Millions of retained impressions per week. Does retained impressions influenced by advertisement spend?
4. One may wonder if people of similar heights tend to marry each other. For this purpose 96 pair of new married couple was selected and provided in worksheet couple file. Analyze the data relating to heights of husbands and wives and give your conclusion.
5. The data in worksheet ‘mpg’ contains information on 11 variables for n = 93 makes and models of automobiles. It is desired to construct a model to predict MPG Highway. 10 possible predictor variables was considered. They are 1. Horsepower, 2. Fuel tank, 3. Passengers, 4. Length, 5. Wheelbase, 6. Width, 7. U-Turn Space, 8. Rear seat, 9. Luggage and 10. Weight. Fit a parsimonious model, i.e., a model involving as few variables as possible, provided none of the omitted variables would add significantly to the predictive capability of the model.
6. The salary survey data given in worksheet ‘salary’ was developed from a salary survey of computer professional in a large organization. The objective of the survey was to identify and quantify those variable that determine salary differentials. The response variable is salary (S) and predictor are: (1) experience measured in years, (2) education (E ) coded as 1 for diploma, 2 for graduates and 3 for advance degree and (3) Management (M) which is coded as 1 for a person with management responsibility and 0 otherwise. Measure the effects of these three variables on salary.
7. A study was planned to identify the effect of advertisement in getting enquiries. The Response of the study is “number of inquiries resulting from advertisement”. The factors influencing the response was identified as Day of week (Monday through Friday) and Section of newspaper (news, business, and sports). The data given in the worksheet ‘paper’. Analyse the data and conclude.
8. The operations manager of a company that manufactures tires wants to determine whether there are any differences in the quality of workmanship among the three daily shifts. She randomly selects 528 tires and carefully inspects them. Each tire is either classified as perfect, satisfactory, or defective, and the shift that produced it is also recorded. The two categorical variables of interest are: shift and condition of the tire produced. The data summarized and presented below two-way table. Do these data provide sufficient evidence at the 5% significance level to infer that there are differences in quality among the three shifts?

|  |  |  |  |
| --- | --- | --- | --- |
| Shift | Perfect | Satisfactory | Defective |
| A | 106 | 124 | 15 |
| B | 67 | 85 | 10 |
| C | 37 | 72 | 12 |

1. An investment organization facing a problem in predicting earning per share (EPS) of various stock traded in the market. The predicted EPS regarded as valuable information for investment. A new methodology was proposed to remove the inconsistency. The new method requires past performance data, market share and company growth, etc.. Some of the components depend on analyst understanding of the market behaviour. To evaluate the effectiveness of proposed methodology, two analysts was assigned to project the EPS of 18 BSE 500 companies. Analyse the data and conclude whether the new method is valid or not.
2. A survey was conducted in restaurants to identify the variation in sales. Variables measured include: Sales (gross sales), Newcap (new capital invested), and Value (estimated market value of the business). All variables are measured in thousands of dollars. Management interested to know whether sales is dependent on ‘new capital invested’ or ‘market value of the businesses’
3. Sumeet, a Data Analyst became curious to find out the pizza delivery time at the front door of his house. He was interested to see how, by varying whether he ordered thick or thin crust, whether Coke was ordered with the pizza and whether garlic bread was ordered with the pizza, the response would be affected. He ordered the pizza from the same shop, being Domino's Pizza. To be consistent he ordered a Supreme pizza each time at approximately the same time of day. The response was measured from the time he closed the telephone to the time the pizza was delivered to the front door. He wrote each of the eight treatment (possible order combination) on a piece of paper twice, put them all into a hat, mixed them up, and took them out one at a time to allocate the order in which each treatment was done. The response and treatment for each pizza delivery the actual hour of delivery was recorded, along with whether the driver was male or female. Analyse the data and identify the factors affecting Pizza Delivery Time.

Pizza Experiment Data

Data Codes: Pizza Crust Thin/Thick 0/1

Coke Yes/No 0/1

Garlic Bread Yes/No 0/1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Order | Crust | Coke | G. Bread | Driver | Hour | Time to Deliver |
| 1 | 0 | 1 | 1 | M | 20.87 | 14 |
| 2 | 1 | 1 | 0 | M | 20.78 | 21 |
| 3 | 0 | 0 | 0 | M | 20.75 | 18 |
| 4 | 0 | 0 | 1 | F | 20.6 | 17 |
| 5 | 1 | 0 | 0 | M | 20.7 | 19 |
| 6 | 1 | 0 | 1 | M | 20.95 | 17 |
| 7 | 0 | 1 | 0 | F | 21.08 | 19 |
| 8 | 0 | 0 | 0 | M | 20.68 | 20 |
| 9 | 0 | 1 | 0 | F | 20.62 | 16 |
| 10 | 1 | 1 | 1 | M | 20.98 | 19 |
| 11 | 0 | 0 | 1 | M | 20.78 | 18 |
| 12 | 1 | 1 | 0 | M | 20.9 | 22 |
| 13 | 1 | 0 | 1 | M | 20.97 | 19 |
| 14 | 0 | 1 | 1 | F | 20.37 | 16 |
| 15 | 1 | 0 | 0 | M | 20.52 | 20 |
| 16 | 1 | 1 | 1 | M | 20.7 | 18 |

1. The measurement of body fat is a cumbersome and expensive procedure requiring the immersion of the person in the water. It is understood that amount of body fat (Y) depends on triceps skinfold thickness (X1), thigh circumference (X2), and mid-arm circumference (X3). The predictor variables are easy to obtain. It would, therefore, be very useful if a prediction model is developed with some predictor variables which can provide reliable estimates of the amount of body fat. Data on a sample of 20 healthy females 25 to 34 years old collected and given in datasheet body fat. Develop an appropriate model.