Problem Sheet Unit 1:

DATA Statistics

Q1.   There are two local factories that produce radios. Each radio produced at factory A is defective with probability 0.05, whereas each one produced at factory B is defective with probability .01. Suppose you purchase two radios that were produced at the same factory, which is equally likely to have been either factory A or factory B. If the first radio that you check is defective, what is the conditional probability that the other one is also defective?

Q2.   Suppose that an insurance company classifies people into one of three classes — good risks, average risks, and bad risks. Their records indicate that the probabilities that good, average, and bad risk persons will be involved in an accident over a 1-year span are, respectively, .05, .15, and .30. If 20 percent of the population are "good risks," 50 percent are "average risks," and 30 percent are "bad risks," what proportion of people have accidents in a fixed year? If policy holder A had no accidents in 1987, what is the probability that he or she is a good (average) risk?

Q3:   The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint.

3.4 2.5 4.8 2.9 3.6

2.8 3.3 5.6 3.7 2.8

4.4 4.0 5.2 3.0 4.8

Assume that the measurements are a simple random sample.
        (a) What is the sample size for the above sample?
        (b) Calculate the sample mean for these data.
        (c) Calculate the sample median.
        (d) Plot the data by way of a dot plot.
        (e) Compute the 20% trimmed mean for the above data set.
        (f) Is the sample mean for these data more or less descriptive?
        (g) Calculate the sample variance as well as standard deviation in tensile strength for both samples.

Q4:   Twenty adult males between the ages of 30 and 40 participated in a study to evaluate the effect of a specific health regimen involving diet and exercise on the blood cholesterol. Ten were randomly selected to be a control group, and ten others were assigned to take part in the regimen as the treatment group for a period of 6 months. The following data show the reduction in cholesterol experienced for the time period for the 20 subjects:

Control group: 7 3 −4 14 2

            5 22 −7 9 5

Treatment group: −6 5 9 4 4

            12 37 5 3 3

        (a) Do a dot plot of the data for both groups on the same graph.

(b) Compute the mean, median, and 10% trimmed mean for both groups.
(c) Calculate the sample variance as well as standard deviation in tensile strength for both samples.

(d) Explain why the difference in means suggests one conclusion about the effect of the regimen, while the difference in medians or trimmed means suggests a different conclusion.

Q5:   The following data set is related to that in Exercise 1.24. It gives the percentages of the families that are in the upper income level, for the same individual schools in the same order as

72.2 31.9 26.5 29.1 27.3 8.6 22.3 26.5
20.4 12.8 25.1 19.2 24.1 58.2 68.1 89.2
55.1 9.4 14.5 13.9 20.7 17.9 8.5 55.4
38.1 54.2 21.5 26.2 59.1 43.3

(a) Calculate the sample mean.
(b) Calculate the sample median.
(c) Construct a relative frequency histogram of the data.
(d) Compute Quartile and plot it also.

Q1:   If S = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} and A = {0, 2, 4, 6, 8}, B = {1, 3, 5, 7, 9}, C = {2, 3, 4, 5}, and D = {1, 6, 7}, list the elements of the sets corresponding to the following events:
   (a) A ∪ C;
   (b) A ∩ B;
   (c) C';
   (d) (C' ∩ D) ∪ B;
   (e) (S ∩ C)';
   (f) A ∩ C ∩ D'.

Q2. In an experiment to study the relationship of hypertension and smoking habits, the following data are collected for 180 individuals:

	Nonsmokers	Moderate Smokers	Heavy Smokers
H	21	36	30
NH	48	26	19

where H and NH in the table stand for Hypertension and Non-hypertension, respectively. If one of these individuals is selected at random, find the probability that the person is
   (a) experiencing hypertension, given that the person is a heavy smoker;
   (b) a non-smoker, given that the person is experiencing no hypertension.

Q3. The probability that an automobile being filled with gasoline also needs an oil change is 0.25; the prob- ability that it needs a new oil filter is 0.40; and the probability that both the oil and the filter need changing is 0.14.
   (a) If the oil has to be changed, what is the probability that a new oil filter is needed?
   (b) If a new oil filter is needed, what is the probability that the oil has to be changed?

Q4. In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective.
   (a) Now, suppose that a finished product is randomly selected. What is the probability that it is defective?
   (b) if a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

Q5. A paint-store chain produces and sells latex and semigloss paint. Based on long-range sales, the probability that a customer will purchase latex paint is 0.75. Of those that purchase latex paint, 60% also purchase rollers. But only 30% of semigloss paint buyers purchase rollers. A randomly selected buyer purchases a roller and a can of paint. What is the probability that the paint is latex?

Q6. A certain form of cancer is known to be found in women over 60 with probability 0.07. A blood test exists for the detection of the disease, but the test is not infallible. In fact, it is known that 10% of the time the test gives a false negative (i.e., the test incorrectly gives a negative result) and 5% of the time the test gives a false positive (i.e., incorrectly gives a positive result). If a woman over 60 is

known to have taken the test and received a favourable (i.e., negative) result, what is the probability that she has the disease?

Q7. A construction company employs two sales engineers. Engineer 1 does the work of estimating cost for 70% of jobs bid by the company. Engineer 2 does the work for 30% of jobs bid by the company. It is known that the error rate for engineer 1 is such that 0.02 is the probability of an error when he does the work, whereas the probability of an error in the work of engineer 2 is 0.04. Suppose a bid arrives and a serious error occurs in estimating cost. Which engineer would you guess did the work? Explain and show all work.