Indian Statistical Institute

Bangalore Centre

B.Math Second year

Semestral Examination

Statistical Methods I .04

Answer as much as you can. The maximum you can score is 100

Time :- 3 1/2 hours

1. The caloric intake (in the units of 1000 calories) of 204 15 and 16 year old females is shown in the following frequency distribution table. By mistake a part of the table has been ommitted. Find out

   (a) the relative frequency and the number of females with caloric intake between 2000 and 2500,

   (b) the mean and median caloric intake for this group of females

   and (c) the proportion of females taking 2000 or more calories.

   [4 + 7 + 5 = 16]

2. (a) Consider two independent random variable X_1 and X_2, X_i following Gamma $(\alpha, p_i), i = 1, 2$. Find the probality distributions of

   (i) $X_1 + X_2$, (ii)X_1/X_2 and (iii) $X_1/(X_1 + X_2)$.

   (b) Suppose X_i follows $\chi^2(k_i), i = 1, 2$ and that they are independent. Derive the probality density of $(X_1/k_1)/(X_2/k_2)$.

   [(6 + 6 + 8) + 5 = 25]

3. Suppose $X_1, X_2, \cdots X_n$ are i.i.d. random variables following $N(\mu, \sigma^2)$.

   (a) Assume $\mu = 0, \sigma = 1$. Derive the distributions of X_1^2 and $\sum_{i=1}^n X_i^2$.

   (b) Show that $\bar{X}$ and $s^2 = (1/(n-1)) \sum_{i=1}^n (Xi - \bar{X})^2$ are independent.

   (c) Derive the distribution of s^2.

   [(5+3) + 9 + 5 = 22]

4. (a) Suppose the number of misprints for every page of a book is found and the data is given in the form of a frequency distribution table. Show how to obtain a 95 confidence interval for the average number of misprints per page of books printed by the same press.

   (b) Obtain the confidence interval numerically from the table below.

   [State your assumptions clearly.]

   [7 + 7 = 14]

5. The aim of a study is to investigate whether a certain drug affects human pulse rate. n patients were chosen at random. For each of them the dose (in c.c.) of the drug given and the pulse rate were noted.

(a) Suppose a linear model was to fitted. Derive the expressions for (i) least square estimate of the slope and (ii) the variance of this estimate.

(b) Derive a test procedure for testing whether the slope is zero.

(c) From the data described above, is it possible to test whether the linear model is not adequate and a higher degree polynomial model would give a better fit ? If not, suggest a modification to the experiment. Explain (without proof) how the above test can be performed with data from the modified experiment.

[(6 + 7) + 6 + 8 = 27]

6. The weekely repair cost (in units of Rs 1000) of a certain machine has observed to follow the probability density function given below.

$$f(x) = \begin{cases} 3(1-x)^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

It is desirable that actual cost of repairs does not exceed the budget B in 9 out of 10 times. How much money should be allotted in the budget ?

[10]

Indian Statistical Institute

Bangalore Centre

B.Math Second year

Back Paper Examination

Statistical Methods I .04

Answer as much as you can. The maximum you can score is 100

Time :- 3 hours

1. (a) Define sample median of a given data set. Consider a data on the yields of rice from villages of South Karnataka, given in the form of a frequency distribution table. Show how you obtain an approximate value of sample median. Justify your answer staying clearly the assumption made.

   (b) Suppose the median yield of rice from a sample of n_1 villages of north Karnataka is M_1 and the same from n_2 villages of south Karnataka is M_2. Show that the median yield of rice from the whole of Karnataka lies between M_1 and M_2.

   [7+5=12]

2. (a) Define absolute mean deviation about A ($Md(A)$). Show that $Md(A)$ is minimum when $A =$ the median.

   (b) Define root mean square deviation about A ($sd(A)$). Show that $sd(A)$ is minimum when A is the sample mean.

   [(2+7) +(2+3)=14]

3. (a) Consider a random variable taking values $1, 2, \cdots k$. Let $F_1 = n, F_2, \cdots f_k$ denote the cumulative frequencies of 'greater than type". Show that

$$\bar{x} = (1/n) \sum_{i=1}^{k} F_i.$$

   (b) Suppose X is a continuous random variable with probability density $f(x)$. Let $\mu = E(X)$ and $M =$ mode of the distribution. Suppose $f(x)$ is symmetric about a. Show that $a = \mu$.

   [7 + 7 = 14]

4. Consider a pair of random variables X, Y with joint density

$$f(x, y) = c.exp[(-1/2)Q(x, y)],$$

   where $Q(x, y) = g^2(x - a)^2 + 2gh(x - a)(y - b) + h^2(y - b)^2$.

   (a) Find the marginal density of X.

(b) Find (i) $E(X)$ and (ii) $Cov(X,Y)$.

(c) Derive the conditional dendity of Y given $X = x$.

(d) Show that if $Cov(X,Y) = 0$, then X and Y are independent.

(e) Suppose X follows univariate normal distribution and Y is another random variable. Is the statement in (d) true ? Justify.

[6 + (6+6) + 4 + 4 + 5 = 31]

5. Suppose $X_1, X_2, \cdots X_n$ are i.i.d. random variables following $N(\mu, \sigma^2)$.
   Derive the distribution of $s^2 = (1/(n-1)) \sum_{i=1}^{n} (Xi - \bar{X})^2$.

[12]

6. To get an idea about the proportion (p) of defective items produced in a factory, 50 random samples of size 10 were selected and the numbers of defective items $(x_1, \cdots x_10)$ noted.

   (a) Obtain (i) an estimate and (ii) a 95 confidence interval for p.

   (b) Show how you can test whether p exceeds the desirable level of $1/10$. [State clearly the results you use.]

[(3+6) + 8 = 17]