Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) II Year 2013-2014
Mid-semester Examination

Statistics I   Date 13.09.13

Answer as many questions as possible. The maximum you can score is 60.
The notation used have their usual meaning unless stated otherwise.

1. A coin was tossed three times and the results noted. This process was continued 100 times, i.e. there were 100 sets of three tosses. In 69 cases first toss showed head, in 49 cases heads obtained in second trial and in 53 cases third trial showed head. In 33 cases heads were obtained in both first and second tosses and in 21 cases both second and third trial resulted in head.

   Show that there were at most 15 occasions when tails occurred all three times. [4]

2. (a) Define (i) sample median and (ii) median of a continuous probability distribution. Suppose X is a continuous random variable with a density. Let M denote the median of X. Prove the following

   $$\text{for any real number C, } E(\mid X - C \mid) \geq E(\mid X - M \mid).$$

   (b) In an epidemiological study, the total organochlorines present in 37 milk samples were recorded. The data set is given below.

   27 43 52 53 54 56 61 63 64 65 68 70 72 75 83 95 96 97 101 105

   110 115 117 118 120 126 127 134 145 152 153 182 190 195 197 232 252

   (i) Prepare a frequency distribution table using class-intervals 0 to 40, 41 to 80 and so on.

   (ii) Find the median from the table. Find the three quartiles from the raw data. What is the reason for the difference in the two values of median? Your answer must be based only on this data set. [(2 + 5)+ (4 + (2 + 3 + 4)) = 20]

3. (a) When is an unimodal probability density function is said to be symmetric, positively skewed or negatively skewed ? Illustrate with graph.

   (b) Define a scale-free measure of skewness (say Sk) based only on three quartiles. What is its range and why ? Justify the use of Sk as a measure of skewness with the help of graphs.

   (c) Compute Sk for the data in Q. 2(b) and comment on the skewness of the data.

   [3 + (2 + 3 + 4) + 2 = 14]

4. In order to study the effects of insulin on reducing blood sugar level in rats, the following experiment was conducted. n rats were selected at random, each one was injected a different dose of insulin and the reduction in blood sugar level was noted.

(a) Suppose a linear model holds. Describe the model, stating clearly all the assumptions. Derive least square estimates of the coefficients. What is meant by residual sum of squares (R_0^2)? Show that

$$E[R_0^2] = (n-2)\sigma^2. \tag{1}$$

(b) Now suppose a quadratic model holds. Is equation (1) still true ? justify.

Suppose it is possible to extend the experiment with m more rats (assume $m \geq 2$). Suggest what doses of insulin may be given to these m rats (in terms of the earlier doses given), so that one can find an expression (say R_1^2) involving the reduction in blood sugar levels of the $n+m$ rats such that

$$E[R_1^2] = c\sigma^2, \text{ where c is known .}$$

Find R_1^2 and c.

[(3 + 4 + 1 + 6) + (2 + 7) = 23]

5. Suppose $U_1, U_2, \cdots U_n$ and $V_1, V_2, \cdots V_n$ are two sets of i.i.d. random variables. Suppose further every U_i is independent of every V_j. Finally, suppose the distributions of U_i and V_j are $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$, respectively, $1 \leq i, j \leq n$.

Consider the following statistics.

$$T_1 = \sqrt{n}\bar{U}/\sigma_1,\ T_2 = S_{uv}/(\sigma_1.\sqrt{S_{vv}}),\ T_3 = (S_{uu} - S_{uv}^2/S_{vv})/\sigma_1^2,$$
$$T_4 = \sqrt{n}\bar{V}/\sigma_2 \text{ and } T_5 = S_{vv}/\sigma_2^2,$$

where $S_{uu} = \sum_{i=1}^n (U_i - \bar{U})^2$, $S_{uv} = \sum_{i=1}^n (U_i - \bar{U})(V_i - \bar{V})$ and $S_{vv} = \sum_{i=1}^n (V_i - \bar{V})^2$.

Show the following.

(i) Each of these five statistics is independent of each of the others.

(ii) Three of them are distributed as normal variables and the other two as χ^2 variables. Derive the parameters of these distributions.

[10]