Indian Statistical Institute
Bangalore Centre
B.Math (Hons.) II Year 2018-2019
Semestral Examination

Statistics I Date 19.11.18

Answer as many questions as possible. The maximum you can score is 120.
The notation used have their usual meaning unless stated otherwise.
State clearly the assumptions you make and the results you use.
No numerical computation is required. It is enough to present you answer as an algebraic expression.

1. In a town invaded by smallpox, $n_{1.}$ inhabitants were attacked, $n_{.1}$ had been vaccinated and n_{11} were vaccinated but attacked. The population size of the town was n.

   (a) What is the percentage of the population who were not vaccinated and not attacked ?

   (b) Explain how you can find out whether the data support the hypothesis that the vaccination was effective. [4 + 8 = 12]

2. Consider n distinct real numbers $x_1, x_2, \cdots x_n$. For a real number A, define $D(A) = \sum_{i=1}^{n} \|x_i - A\|$. Show that $D(A)$ is minimum when $A =$ the median of $x_1, x_2, \cdots x_n$. [10]

3. Suppose X is a random variable following standard normal distribution.

   (a) For a fixed positive real number l, show that $Prob.(a \leq X \leq a+l)$ is maximum if $a = -l/2$.

   (b) Let $\mathcal{I}$ denote the class of intervals on the real line such that $Prob(X \in I) = p$, for every I in $\mathcal{I}$. Find the interval $I_0 \in \mathcal{I}$ with minimum length.

   (c) Suppose $X_1, X_2, \cdots X_n$ are i.i.d. random variables following $N(\mu, 1)$. Two persons A and B were asked to obtain a 80 % confidence interval for μ using the observed values of X_i's. A found an interval $[\bar{X} - a, \bar{X} + b], a \neq b$, while B found the interval $[\bar{X} - b, \bar{X} + b]$. Which one would you recommend and why ? [9 + 4 + 7 = 20]

4. (a) Write down the probability density of non-central χ^2 distribution with k degrees of freedom and non-centrality parameter λ, in terms of probability density of central χ^2 distribution.

   Suppose X is a random variable following $N(\mu, 1)$. Write down the probability density of $Y = X^2$. [Derivation not needed].

   (b) Suppose X and Y are independent χ^2 variables. Suppose their degrees of freedom are k and m and non-centrality parameters 0 and λ respectively. Derive the probability density of $X + Y$. [(4 + 3) + 8 = 15]

5. Discharging warm water from a nuclear power plant is suspected to be harmful to a species of zoo-plankton. It is known from the past record that the average count of them in mid-July is 1.2 per liter. In an area close to the nuclear plant 5 one-liter water samples are collected from the river; the counts of zoo-plankton are found to be $0, 0, 1, 2, 1$. Formulate a test, with a type I error probability not more than 8%, to verify whether the data support the hypothesis that warm water is really harmful to zoo-plankton. [8]

Notation :

$$S_{qq} = \sum_{i=1}^{n}(Q_i - \bar{Q})^2 \text{ and } S_{pq} = \sum_{i=1}^{n}(P_i - \bar{P})(Q_i - \bar{Q}). \tag{1}$$

6. Suppose $X_1, \cdots X_n$ are i.i.d. $N(\mu, 1)$ variables and $A_1, \cdots A_n$ are constants such that $a = (A_1, \cdots A_n) \neq 0$. Show the following.

   (a) $\bar{X}$ follows $N(\mu, 1/n)$.

   (b) S_{xx} follows $\chi^2(n-1)$.

   (c) S_{xx} and $\bar{X}$ are independent.

   (d) Suppose $\mu = 0$. Then

   (i) S_{ax}^2/S_{aa} follows $\chi^2(1)$ and

   (ii) $S_{xx} - S_{ax}^2/S_{aa}$ follows $\chi^2(n-2)$.

   (iii) Now suppose $Y_1, \cdots Y_n$ are i.i.d. $N(0, \sigma_y^2)$ variables. Then, $S_{xx} - S_{xy}^2/S_{yy}$ follows $\chi^2(n-2)$.

   [Here $S_{xx}, S_{ax}, S_{aa}, S_{xy}, S_{yy}$ are as in (1).] [3 + 6 + 5 + (4 + 8 + 10) = 36]

7. 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 certain dose of insulin and the reduction in blood sugar level was noted. Suppose X_i denote the ith dose of insulin and Y_{ij} denote the reduction in the blood sugar level of the jth rat among the ones receiving X_i. Consider the model

$$Y_{ij} = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_{ij},\ j = 1, \cdots n_i, i = 1, \cdots k.$$

   where ϵ_{ij}'s are i.i.d $N(0, \sigma^2)$ variables.

   Define $SS_W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{i,j} - \bar{Y}_i)^2$ and $SS_B = \sum_{i=1}^{k} n_i(\bar{Y}_i - \bar{Y})^2$, where $\bar{Y}_i$ and $\bar{Y}$ have their usual meaning.

   (a) Show that SS_W and SS_B are independent.

   (b) Find the distribution of SS_W. [7 + 6 = 13]

8. Suppose (X, Y) follows bivariate normal distribution, with parameters $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ and ρ.

   Define random variables U and V as follows. $U = Y - q - c(X - p)$ and $V = X - p$.

   Find values of p, q and c in terms of the parameters of the distribution such that (i) $E(U) = E(V) = 0$ and (ii) U, V are independent. What would be the distribution of U if these conditions are satisfied ?

   Using the above results find the conditional distribution of Y, given $X = x$. [7 + 3 + 9 =19]