Indian Statistical Institute
Bangalore Centre
B.Math Third Year 2017-2018
Second Semester

Mid-Semester Examination Date : 01.03.18

Statistics IV

Answer as much as you can. The maximum you can score is 60.
The notation used have their usual meaning unless stated otherwise.
Time :- 3 hours

1. Consider a 2×2 contingency table with variables X and Y.

(a) Define odds ratio (θ).

(b) In the following statement fill in the blanks with justification.

"$\theta = -$ when X and Y are independent, $\theta = -$ when there is a direct association and $\theta = -$ when there is a reverse association.

(c) For adults who sailed on the "Titanic" on its fateful voyage, the odds ratio between gender (female, male) and survival (yes, no) was 11.4. Consider the statement " The probability of survival for females was 11.4 times that for males".

(i) What is wrong with the interpretation ? Give the correct interpretation.

(ii) The odds of survival for females equaled 2.9. For each gender, find the proportion who survived.

(iii) Find a condition on n_{ij}'s (the entries in the two-way table) which would approximately imply the statement in " " above.

[1 + 2 x 3 + (4 + 3 + 4) = 18]

2. Suppose $X = (X_1, \cdots X_k)'$ follows multinomial distribution with parameters $(n, \pi_1, \cdots \pi_k)$. Let $\phi = (\phi_1, \cdots \phi_k)'$ where $\phi_i = \sqrt{\pi_i}$. Let

$$V = (V_1, \cdots V_k)',\ V_i = (X_i - n\pi_i)/\sqrt{n\pi_i}.$$

(a) Show that for any k-vector b the asymptotic distribution of $b'V$ is Normal with mean 0 and variance $b'(I_k - \phi\phi')b$.

(b) Suppose A is an idempotent matrix satisfying (i) $A\phi = 0$ and (ii) rank (A) = t. Find the asymptotic distribution of $V'AV$.

(c) Explain how you can test the hypothesis $H_0 : \pi_i = p_i,\ 1 \leq i \leq k$ against H_1 : not all π's equal p_i. Justify.

[8 + 6 + 4 =18]

3. Suppose X and Y are random variables with continuous distribution : their distribution functions being $F(x)$ and $G(y)$ respectively. $X_1, \cdots X_m$ and $Y_1, \cdots Y_n$ are independent random samples from the populations of X and Y respectively. Further, suppose $G(x) =$

$F(x-\triangle)$, where $\triangle$ is a unknown real-valued parameter. Let Q_i and R_j denote the ranks of X_i and Y_j, respectively, among the $N = m+n$ combined observations. Consider the following statistics.

$$W = \sum_{i=1}^{n} R_i \text{ and } U = \sum_{i=1}^{m}\sum_{j=1}^{n} \psi(Y_j - X_i),$$

where $\psi(t) = 1$ if $t > 0$ and 0 otherwise.

(a) Show that $W = U + n(n+1)/2$, provided there is no tie.

(b) Show that the distribution of W is symmetric about $n(N+1)/2$, provided $\triangle = 0$.

[4 + 3 = 7]

4. $X_1, X_2, \cdots X_n$, $n > 2$ are i.i.d continuous random variables and R_i is the rank of X_i. Let $X = (X_1, \cdots X_n)$.

   (a) Define anti-rank S_i of X_i.

   (b) Denote the sequences of ranks and anti-ranks of $|X_i|$'s by R^+ and S^+ respectively.

$$\text{Define } U_i = \begin{cases} 1 \text{ if } |X_{(i)}| \text{ corresponds to a positive} X_t \\ 0 \text{ otherwise} \end{cases} \text{ and } W^+ = \sum_{i=1}^{n} iU_i.$$

   Show that

   (i) $W^+ = \sum_{i=1}^{n} \psi(X_i)R_i^+$. Here the function ψ is as in Q3.

   (ii) U_i's are i.i.d. $B(1, 1/2)$ variables.

[2 + (3 + 6) = 11]

5. In order to study the relationship between dexterity (D) and aggression (A), n young adults were tested. Suppose the scores were (D_i, A_i), $i = 1, \cdots n$, the ranks of D_i and A_i were R_i and S_i respectively. Define

$$S_{DA} = \sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S}), \quad S_D = \sum_{i=1}^{n}(R_i - \bar{R})^2 \text{ and } S_A = \sum_{i=1}^{n}(S_i - \bar{S})^2,$$

$$r_{SP} = S_{DA}/\sqrt{S_D . S_A} \text{ and } U = \sum_{i=1}^{n} iR_i$$

   Show that if D and A are independent then $P[r_{SP} \geq a] = P[U \geq b]$, where $a = bc + d$, c and d are functions of n. Determine c and d. [6]

6. Suppose $X_1, \cdots X_n$ is a random samples from a continuous distribution with median θ. Show how you can obtain a distribution-free confidence interval for θ. [8]