Indian Statistical Institute
Bangalore Centre
B.Math Third Year 2014-2015
Second Semester

Mid-Semester Examination Date : 4.3.15

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 study with three categorical variables X, Y and Z. When are X and Y said to be conditionally independent, given $Z = k$?

   Derive the form of the log-linear model for the number of observed units in the (i, j, k)th cell in the case when X and Y are conditionally independent, given Z.

   [2 + 5 = 7]

2. (a) Define odds ratio for a 2×2 contingency table.

   (b) 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 + (3 + 3 + 4) = 11]

3. 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 that each π_i is a function of $\theta_1, \cdots \theta_q, q < k$. State clearly the result you use without proof.

   [8 + 6 + 4 =18]

4. 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 $G(x) = F(x - \Delta)$, where Δ 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 $\Delta = 0$.

(c) To see whether the "phenolic" contents differ significantly in the wood wastes of two species of trees, the percentages of phenolic contents were obtained from 6 randomly chosen pieces of woods of species A and 9 of species B. Explain how a distribution-free test can be conducted. State clearly your assumptions.

[4 + 3 + 4 = 11]

5. The finalists in a skating contest are rated by two judges A and B on a 10-point scale. Let X_i and Y_i denote the scores of the ith candidate given by judges A and B respectively, $i = 1, \cdots n$. The organizers want to see whether the ratings of judges A,B are closely related. Suppose R_i denote the rank of X_i and $U = \sum_{i=1}^{n} iR_i$. Prove (a) and (b) and justify the procedure described in (c).

(a) $E(U) = n(n+1)^2/4 = M$, say.

(b) The distribution of U is symmetric about M.

(c) If U is 'considerably larger' than M, then conclude that the rating on one judge is consistent with that of another.

[3 + 2 + 6 = 11]

6. Suppose $X_1, \cdots X_n$ is a random samples from a continuous distribution with median θ. Let $S(\theta)$ denote the number of positive members among $X_1 - \theta, \cdots X_n - \theta$.

(a) How is $S(\theta)$ distributed ?

(b) Explain a test procedure for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$.

(c) Show how you can obtain a confidence interval for θ, which does not depend on the distribution of X_i's.

(d) Describe an estimator of θ based on $S(\theta)$. Explain why you view this as a 'reasonable estimator'.

[1 + 1 + 5 + 6 = 13]