INDIAN INSTITUTE OF SCIENCE EDUCATION & RESEARCH KOLKATA

Statistics - I (MA3205) – Mid-Semester Exam

Date: 20th February, 2019

Duration: 1.5 hours

Maximum points that you can score is 50. Good luck!

Question 1 (10 points)

Suppose that the third year students of IISER Kolkata are given two tests. The scores obtained by the ith student in the two tests are x_i and y_i. The final scorecard is based on the ranks of the students in each of the two tests. Denote the ranks of the ith student in the two tests by r_i and s_i. Assume that no two students have received the same score in any of the two tests. Show that the correlation coefficient, say R, between the data sets $\{r_1, r_2, \ldots, r_n\}$ and $\{s_1, s_2, \ldots, s_n\}$ is given by

$$R = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (r_i - s_i)^2.$$

[*Hint*: The ranks are a permutation of $\{1, 2, \ldots, n\}$.]

Question 2 (10 points)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with a density that is symmetric about 0. Suppose that $\mathbb{E}(|X_1|^k) = 2^k$ for $k = 1, 2$. Define

$$Y = \sum_{i=1}^{n} X_i, \quad \text{and} \quad Z = \sum_{i=1}^{n} \mathbb{1}(X_i > 0).$$

Find $\mathrm{Corr}(Y, Z)$.

Question 3 (10 points)

Consider the data set $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. At a data pre-processing stage, it is found that the x-observations have in fact been observed with additive measurement error. So, each x_i is actually $z_i + e_i$, where z_i is the true value of the explanatory variable and e_i is the unobserved measurement error. It is assumed that the e_i's have zero mean, variance s_e^2, and they are uncorrelated with the z_i's. Ideally, we would have liked to observe the z_i's and regress the y-variable on the z-variable. Given the present problematic situation, find the regression line of y on x. Compare it with the slope of the (unobserved) regression line of y on z. What happens to the slope of the former regression line if s_e^2 is large compared to s_z^2 (the variance of the unobserved z_i's)?

Please turn over

Anirvan Chakraborty 20/2/19

Question 4 (20 points)

Consider a data set $y_1, y_2, \ldots, y_n$. It is known that a median M for the data set $\{y_1, y_2, \ldots, y_n\}$ satisfies the relation $M = arg\min_a \sum_{i=1}^{n} |y_i - a|$. Let $f_1, f_2, \ldots, f_n$ be some positive integers.

(a) Find the minimizer of the function

$$a \mapsto \sum_{i=1}^{n} |x_i - a| f_i.$$

(b) Consider another data set $x_1, x_2, \ldots, x_n$, where each x_i is a non-zero rational number. Find the minimizer of the function

$$b \mapsto \sum_{i=1}^{n} |y_i - bx_i|.$$

[*Hint*: Try to express the function in part (b) in the form of the function in part (a).]

– END OF THE EXAM –

Anirva Chakraborty.
25/2/19.