

Indian Statistical Institute, Bangalore
B. Math I, Second Semester, 2022-23
Mid-semester Examination

Intro to Statistics and Computation with Data
Maximum Score 50

23.02.23

Duration: 1.5 Hours

1. (3+9+6) The length of stay in a hospital after a particular treatment is of interest to patients and hospitals. A random sample of 50 patients who received the treatment was selected. The length of stay, in number of days, is recorded in the table below.

Length of stay	5	6	7	8	9	12	21
Number of patients	4	13	14	11	6	1	1

- (a) Is the distribution symmetric, skewed to the right or skewed to the left?
- (b) Find the median, first and third quartiles.
- (c) Using the $1.5 \times \text{IQR}$ rule, identify outliers, if any.
2. (2) Researchers are interested in studying whether regular walking reduces cholesterol. A sample of 100 patients with high cholesterol is selected. The patients are advised by the doctor to walk regularly. They are asked to keep a daily record of the number of minutes walked each day. After six months, their cholesterol is tested again and the walking record obtained. Identify and explain the type(s) of bias that can affect this study.

3. (10+2+3) Let $X_1, X_2 \dots X_n$ be a random sample from Beta distribution with pdf given by

$$f(x | a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\beta(a, b)} \quad \text{if } x \in (0, 1),$$

where $a, b \in \mathbb{R}^+$ are both unknown.

- (a) Obtain the method of moments (MoM) estimator for a .
- (b) Is this the only MoM estimator, or is it possible to obtain other estimators based on the method of moments?
- (c) Is the estimator consistent? Justify your answer.
4. (15) In a study of the relationship between birth order and college success, an investigator found that 140 in a sample of 200 college graduates were firstborn or only children. In a sample of 120 non-graduates of comparable age and socioeconomic background, the number of firstborn or only children was 66. Obtain a 95% confidence interval for the difference between the proportions of firstborn or only children in the two populations from which these samples were drawn. [Use $q_{\text{norm}}(0.975)=1.96$].