# 2016

BOOKLET NO. TEST CODE: **STA**

*Forenoon*

**No. of Questions: 10** **Time: 2 hours**

- **Answer as many questions as you can. All questions carry equal weight.**
- **Do not feel discouraged if you are not able to answer all the questions.**
- **Partial credit may be given for partial answer.**
- **Full credit will be given for complete and rigorous arguments.**

*Write your Name, Registration Number, Test Code, Booklet No. etc., in the appropriate places on the answer-booklet.*

**ALL ROUGH WORK MUST BE DONE ON THIS BOOKLET AND/OR ON THE ANSWER-BOOKLET. YOU ARE NOT ALLOWED TO USE CALCULATORS.**

**STOP! WAIT FOR THE SIGNAL TO START.**

1. Find all $2 \times 2$ matrices $A$ with real entries such that $A^2 = -I$, where $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

2. Find all continuously differentiable functions $f$ from the real line to the real line satisfying

$$(f(x))^2 = \int_0^x [f(t)^2 + f'(t)^2]dt + 2016,$$

for all real $x$.

3. Suppose $X$ and $Y$ are two random variables with finite variances such that

   (a) $E(X) = E(Y)$, and

   (b) for some $\beta \neq 0$ and for all $x, y$

$$E(Y|X = x) = \beta x \text{ and } E(X|Y = y) = \frac{y}{\beta}.$$

   Show that $P(X = Y) = 1$.

4. Prove that

$$\sum_{i=0}^{n} \frac{e^{-n} n^i}{i!} \to \frac{1}{2} \text{ as } n \to \infty.$$

5. Consider an irreducible and aperiodic Markov chain with a finite or countably infinite state space. Let the transition matrix $P$ be symmetric. Find $\lim_{n\to\infty} p_{ij}^{(n)}$ for all $i, j$, where $p_{ij}^{(n)}$ is the $(i, j)$-th element of $P^n$.

6. Let $p \in \left(0, \frac{1}{2}\right)$ be unknown. There are two coins with probabilities of head $p$ and $1 - p$. One of the two coins is picked at random. This coin is tossed 10 times independently. Let $X_i = 1$ if the $i$-th toss results in a head, and $X_i = 0$ otherwise, for $i = 1, ..., 10$.

   (a) Are $X_1, ..., X_{10}$ identically distributed? Are they independent? Justify both your answers.

   (b) Find the maximum likelihood estimator of $p$ based on $X_1, ..., X_n$.

7. There are 30 multiple choice questions (with 5 possible answers for each, exactly one being correct) in a certain examination. A student knows the answers to $k$ questions and answers them correctly. For the remaining $30-k$ questions, the student guesses randomly among the 5 choices. Let $X$ be the total number of correct answers given by the student. The integer $k$ is unknown to the examiner. Find the maximum likelihood estimator of $k$ based on $X$.

8. Let $\beta_1, \beta_2, \beta_3$ be the true interior angles of a triangle. Suppose that $Y_1, Y_2, Y_3$ are independent measurements of $\beta_1, \beta_2, \beta_3$, respectively. We assume $Y_i$ is normally distributed with mean $\beta_i$ and variance $\sigma^2$ for $i = 1, 2, 3$, where $\sigma > 0$ is unknown. Obtain the best linear unbiased estimators (BLUEs) of $\beta_1, \beta_2, \beta_3$ based on these measurements.

9. In a large survey, an approximate 95% confidence interval (using normal approximation) for the proportion of literate people in some state turned out to be (42.2%, 61.8%). A subsequent concern was whether more than 50% of the people in that state are literate. Formulate this as a hypothesis testing problem. Based on the given confidence interval, perform an appropriate test of this hypothesis at a significance level of 5%. You may use the relations

$$\int_{-\infty}^{1.645} \phi(x)dx = 0.95 \text{ and } \int_{-\infty}^{1.96} \phi(x)dx = 0.975,$$

where $\phi(x)$ is the standard normal density.

10. For a bivariate data set $(x_i, y_i)$, $i = 1, ..., 50$, a statistician tries to fit the model

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

where $x_i$'s are assumed to be fixed, and $\epsilon_i$'s are independently and identically distributed as normal with mean 0 and variance $\sigma^2$. Here the real numbers $\alpha, \beta$ and $\sigma^2 > 0$ are unknown. The model is fitted using maximum likelihood estimation, and the residuals $r_i = y_i - \hat{y}_i$ are plotted against $x_i$'s, where $\hat{y}_i$'s are the fitted values. These plots are shown below for four different data sets. If you think some of these plots is/are actually impossible, then identify them with justification. Suggest, with justification, how you would modify your model (if necessary) for the other plot(s).