# Indian Statistical Institute, Bangalore <br> B. Math II, First Semester, 2021-22 <br> Final Examination, Statistics I <br> Maximum Score 100 <br> <br> Duration: 3 Hours 

 <br> <br> Duration: 3 Hours}
05.01.22

Students are allowed to consult the book Statistics by McClave and Sincich.
Values from the normal distribution qnorm $(0.9)=1.281552$, qnorm $(0.995)=2.575829$, qnorm $(0.95)=1.644854$
Instructor: Rituparna Sen rsen@isibang.ac.in ritupar.sen@gmail.com 9176620249

1. (10) Let $X_{1}, \cdots, X_{n}$ be a random sample from a distribution with $\operatorname{pdf} f(x \mid \lambda, \theta)=\lambda \exp (-\lambda(x-\theta))$, if $x>\theta$ where $\lambda \in \mathbb{R}^{+}$and $\theta \in \mathbb{R}$. Find the method of moments estimators of $\theta$ and $\lambda$.
2. $(5+5+10)$ In genetics, consider a single locus with two alleles denoted A and a with probabilities $p$ and $1-p$, respectively. Then the probabilities of genotypes AA, aa and Aa under random mating are $p^{2},(1-p)^{2}$ and $2 p(1-p)$ respectively.
We take a sample of size $n$ from such a population. In the sample the observed frequencies of the three genotypes AA, aa and Aa are $X_{1}, X_{2}, X_{3}$ respectively. The parameter of interest is $p$. Let $\hat{p}=\frac{1}{n}\left(X_{1}+\frac{X_{3}}{2}\right)$. Answer the following questions.
(a) What is the joint distribution of $X_{1}, X_{2}, X_{3}$ ?
(b) Show that $\hat{p}$ is unbiased for $p$.
(c) Show that the variance of $\hat{p}$ is $\frac{p(1-p)}{2 n}$
3. $(5+5)$ Answer any two of the following questions related to the class presentations.
(a) Consider the sample median, $T$, obtained from a sample of size $n$ from a given population. Describe the method of obtaining a bootstrap estimate of the variance of $T$.
(b) In multiple linear regression, why is it a problem if there is high correlation between the predictor variables?
(c) In what situation is the Mann-Whitney-Wilcoxon test used? Describe the testing procedure.
4. $(5+7+8)$ Consider the regression model

$$
y_{i}=\beta x_{i}+\epsilon_{i}, \quad 1 \leq i \leq n,
$$

where $\epsilon_{i}$ are iid with mean zero and variance $\sigma^{2}$ and $x_{i}$ are fixed.
Consider estimators of $\beta$ of the form $T \mathbf{a}=\sum_{i=1}^{n} a_{i} y_{i}$, for $\mathbf{a}=\left(\mathbf{a}_{\mathbf{1}}, \cdots, \mathbf{a}_{\mathbf{n}}\right)^{\mathbf{T}} \in \mathbb{R}^{\mathbf{n}}$.
(a) Find the least squares estimator $T$ of $\beta$.
(b) Find the mean and variance of $T_{\mathbf{a}}$ in terms of $a_{i}$ 's and the parameters $\beta$ and $\sigma$.
(c) Show that estimator $T$ minimizes the variance of $T_{\mathbf{a}}$ over all possible $\mathbf{a} \in \mathbb{R}^{\mathbf{n}}$.
5. $(10+10)$ Consider random variable $Y$ with triangular distribution whose probability density function (pdf) is given by

Assume that you can generate observations on uniform $[0,1]$.
(a) How would you draw observations on $Y$ using probability integral transform?
(b) If $U$ and $V$ are independent uniform $[0,1]$ then obtain the distribution of $(U+V) / 2$. Use this result to describe an alternative procedure to generate observations on $Y$.
6. $(4+4+4+4+4)$ A medical centre developed a new procedure for surgery of the knee, designed to reduce the recovery time after surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing knee surgery were randomly assigned to receive either the standard procedure or the new procedure. Summary statistics of the recovery times are given below.

| Type of <br> Procedure | Sample <br> Size | Mean recovery <br> Time(days) | Standard deviation of <br> Recovery Time |
| :---: | :---: | :---: | :---: |
| Standard | 110 | 217 | 34 |
| New | 100 | 186 | 29 |

Using $\alpha=0.05$ perform the appropriate test of hypothesis to determine if the mean recovery time is lower for the new procedure. In particular, answer the following questions.
(a) State the null and alternative hypotheses.
(b) State the test statistic and find its distribution under the null hypothesis.
(c) Is the distribution in the previous part exact or approximate? In the latter case, why does the approximation work?
(d) Find the critical region and compute the value of the test statistic.
(e) Is the null hypothesis rejected? What is the conclusion regarding the mean recovery time?

