MAASAI MARA UNIVERSITY
REGULAR UNIVERSITY EXAMINATIONS
2023/2024 ACADEMIC YEAR
FIRST YEAR SECOND SEMESTER
SCHOOL OF PURE APPLIED AND HEALTH SCIENCES
MASTER OF SCIENCE IN APPLIED STATISTICS
STA 8214 NON-PARAMETRIC METHODS

Instructions to candidates:

Answer any three questions.

All Symbols have their usual meaning

DATE: TIME:

Question 1

(a) The residual sum of squares (RSS) is given by

$$RSS = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}'\boldsymbol{\beta}$$

(i) Differentiating the RSS with respect to $\boldsymbol{\beta}$, show that the ridge estimator

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y}$$

(**5 Marks**)

(ii) Given $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$ where $\mathbf{H} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$ is called the hat matrix. Use this relationship to show that

$$\boldsymbol{\beta}_R = \mathbf{C}_\lambda\hat{\boldsymbol{\beta}}$$

where $\mathbf{C}_\lambda = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-\mathbf{1}}\mathbf{X}'\mathbf{X}$ (**2 Marks**)

(iii) Given that

$$Var(\hat{\boldsymbol{\beta}}_R) = \sigma^2 = [(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-\mathbf{1}}]\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})]^{-\mathbf{1}}$$

and

$$Bias(\hat{\boldsymbol{\beta}}_R) = \lambda\boldsymbol{\beta}(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-\mathbf{1}}$$

, show that

$$MSE(\boldsymbol{\beta}_R) = \sigma^2\sum_{i=1}^{p}\frac{c_i}{(c_i+\lambda)^2} + \lambda^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-\mathbf{2}}\beta$$

where $C_i's$ are the eigenvalues of $\mathbf{X}'\mathbf{X}$ (**5 Marks**)

(b) (i) Give the difference between Least Absolute Selection and Shrinkage Operator estimator(LASSO) introduced by Tibshirani (1996) and ridge estimator (**2 Marks**)

(ii) Write three problems of finding LASSO estimator (**6 Marks**)

Question 2

(a) Show that when we calculate the Kolmogorov- Smirnov statistics $D_n = \sup_{-\infty<x<\infty}|F_n(x) - F_o(x)|$ we get $D_n = 2n$ maximum positive quantities (**10 Marks**)

(b) The following table shows the number of cycles to failure of 22 ball bearings

17.88	28.92	33.00	41.52	52.12	45.60	48.48	51.84	51.96	54.12	56.56
67.40	68.64	68.88	84.12	93.12	98.64	105.12	105.84	127.92	128.04	173.40

This data is used to test the hypothesis

$$H_o : F_o(x) = 1 - e^{-\lambda x},\ x > 0\ , \lambda > 0$$

where λ is unknown
(i) Show that the maximum likelihood estimator of λ is $\hat{\lambda} = 0.0138$ (**4 Marks**)
(ii)Use Kolmogorov- Smirnov one sample test at $\alpha = 0.05$ to check if the given data has the distribution given by $F_o(x)$ and mean $\frac{1}{\hat{\lambda}} = 72.3873$ (**6 Marks**)

Question 3
The following table shows students performance during the first week of class. Use $\alpha = 0.05$ for your desired level of confidence

90	72	90
64	95	89
74	88	100
77	57	35
100	64	95
65	80	84
90	100	76

(a) You are required to compute

(i) the mean and standard deviation (**4 Marks**)

(ii) the Kurtosis (**4 Marks**)

(iii) the standard error of Kurtosis (**3 Marks**)

(b) Use the Kurtosis and the standard error of the Kurtosis to find a z-score (**2 Marks**)

(c) (i) Use the mean and standard deviation to find the skewness (**3 Marks**)

(ii) Find the standard error of the skewness (**2 Marks**)

(iii) Use the skewness and the standard error of the skewness to find a z-score (**2 Marks**)

Question 4

(a) Let $X_{(r)}$ and $X_{(s)}$ be two order statistics and assuming continuity so that $F(\epsilon_p) = p$ where ϵ_p is the quartile order $p(0 < p < 1)$ of the continuous function F. Show that

$$P_r X_{(r)} \leq \epsilon_p \leq X_{(s)} = \sum_{i=r}^{b-1} \binom{n}{i} p^i (1-p)^{n-i}$$

(**5 Marks**)

(b) (i) Define Tolerance intervals (**2 Marks**)

(ii) Let $(X_{(r)}, X_{(s)})$ for $r < s$ be a random interval, write the probability of, i.e

$$P_r[F(X_{(s)}) - F(X_{(r)}) > 1 - r]$$

using $U_{(s)}$ and $U_{(r)}$ which are the r-th and s-th order statistics from the U(0,1) (**2 Marks**)

(iii) Write also the joint density function

$$f_{U_{(r)},U_{(s)}}(x, y)$$

of the r-th and s-th order statistics (**3 Marks**)

(c) Prove the following properties of the empirical Distribution function, $F_n(x)$
(i)is unbiased estimator of f(x) (**3 Marks**)
(ii) is a weakly consistent estimator of the unknown distribution function $F(x)$, that is, $Var(F_n\infty)$ $\to 0$ as $n \to \infty$ with probability 1 (**5 Marks**)