31.05.24

Exam Seat No.

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

Date: 03-05-24 | FS MSC(P) Examination | Time: 11.30 a.m to 2:30 p.m

Day: Friday | **Sampling Theory (STA211IC11)**

Total Marks: 70

Q: 1 Choose the correct option and write in the answer sheet provided. (28)

1. An unbiased estimator of $V(\hat{Y}_{pps})$ is

(a) $\frac{1}{N(n-1)}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}-Y\right)^2$ (b) $\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}-Y\right)^2$

(c) $\frac{1}{n(N-1)}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}-Y\right)^2$ (d) $\frac{1}{N(N-1)}\sum_{i=1}^{n}\left(\frac{Y_i}{P_i}-Y\right)^2$

2. In random group method population is randomly divided in to groups which are
(a) mutually exclusive but not exhaustive. (b) not mutually but exhaustive
(c) mutually exclusive and exhaustive. (d) only exhaustive.

3. Under Simple Random Sampling $\hat{Y}_R$ equal to

(a) $\frac{\sum Y_i}{\sum X_i}$ (b) $\frac{\sum Y_i}{\sum X_i}.X$ (c) $\frac{\sum Y_i}{\sum X_i}.Y$ (d) $\frac{\sum X_i}{\sum Y_i}.X$

4. The bias of the estimator $\hat{Y}_{R_0}$ is
(a) $[N-1]S_{zx}$ (b) $-[1-n]S_{zx}$
(c) $-[N-n]S_{zx}$ (d) $-[N-1]S_{zx}$

5. Under SRS, the ratio estimator is more precise than the expansion estimator, when the variables X and Y have
(a) high negative correlation (b) low positive correlation
(c) high positive correlation (d) none of the above

6. Non-response in surveys mean
(a) non return of questionnaire by respondents. (b) non availability of respondents.
(c) refuse to give information by respondents. (d) all the above.

7. Whenever the groups are of same size in Random group method
(a) $V(\hat{Y}_{RHC}) > V(\hat{Y}_{PPS})$ (b) $V(\hat{Y}_{RHC}) = V(\hat{Y}_{PPS})$
(c) $V(\hat{Y}_{RHC}) < V(\hat{Y}_{PPS})$ (d) none of the above

8. Let $e_o = \frac{\hat{Y}-Y}{Y}$ then e_o satisfied
(a) $E(e_0) = 0$ (b) $E(e_0^2) = V(\hat{Y})/Y^2$
(c) both the above are correct. (d) both the above are wrong.

9. When the samples are drawn independently in the two phases of sampling, the approximate bias of the ratio estimator is
(a) $B(\hat{Y}_{RD}) = Y\left[\frac{V(\bar{X})}{\bar{X}^2} - \frac{Cov(\bar{X},\bar{Y})}{\bar{X}\bar{Y}}\right]$ (b) $B(\hat{Y}_{RD}) = Y\left[\frac{V(\bar{X})}{\bar{Y}^2} - \frac{Cov(\bar{X},\bar{Y})}{\bar{X}\bar{Y}}\right]$
(c) $B(\hat{Y}_{RD}) = Y\left[\frac{V(\bar{X})}{\bar{X}^2} + \frac{Cov(\bar{X},\bar{Y})}{\bar{X}\bar{Y}}\right]$ (d) none of the above.

10. Under srs in regression estimation
(a) $V(\hat{Y}_{SRS}) > MSE(\hat{Y}_{LR})$ (b) $MSE(\hat{Y}_R) > MSE(\hat{Y}_{LR})$
(c) both (a) and (b) are correct. (d) both (a) and (b) are wrong.

11. The combined ratio estimator in stratified sampling is given by
(a) $\hat{Y}_{RC} = \sum_{h=1}^{L}\left[\frac{\bar{Y}_h}{\bar{X}_h}\right]X_h$ (b) $\hat{Y}_{RC} = \sum_{h=1}^{L}\left[\frac{\bar{X}_h}{\bar{Y}_h}\right]X_h$
(c) $\hat{Y}_{RC} = \left[\frac{\sum_{h=1}^{L}\bar{Y}_h}{\sum_{h=1}^{L}\bar{X}_h}\right]Y$ (d) $\hat{Y}_{RC} = \left[\frac{\sum_{h=1}^{L}\bar{Y}_h}{\sum_{h=1}^{L}\bar{X}_h}\right]X$

P.T.O.

12. The ratio estimator $\hat{Y}_R$ less efficient than the expansion estimator $\hat{Y}$, if

(a) $\rho < \frac{1}{2}\left[\frac{C_y}{C_x}\right]$ (b) $\rho > \frac{1}{2}\left[\frac{C_x}{C_y}\right]$ (c) $\rho = \frac{1}{2}\left[\frac{C_x}{C_y}\right]$ (d) none of the above

13. Which of the following statement is true

(a) less the standard error, better it is (b) less the variance, better it is

(c) both (a) and (b) are correct (d) standard error is always unity

14. Sampling frame is a term used for

(a) a list of random numbers (b) a list of voters

(c) a list of sampling units of population (d) none of the above

Q2(a.) Explain Lahiri's method with suitable example. In which situation Lahiri method is more suitable than cumulative method? (04)

(b.) In pps sampling with out replacement, $\hat{Y}_{HT}$ is unbiased and its sampling variance is given by $V_{HT}(\hat{Y}_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)Y_i^2}{\pi_i} + \sum_{i}^{N} \sum_{i \neq j}^{N} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} Y_i Y_j$ where π_{ij} is the probability of inclusion of both the ith and jth unit in the sample.Prove it. (05)

OR

(b) The probability of selecting the i^{th} unit in the first effective draw is $\frac{X_i}{X}$ in Lahiri's method of pps sampling. Prove it. (05)

(c) Under ppswor, $\hat{Y}_{DR}$ is unbiased for the population total and unbiased estimator of $\hat{Y}_{DR}$ is $\frac{1}{n(n-1)}\sum_{i=1}^{n} (t_i - \bar{t})^2$ (05)

Q3(a) Explain Ratio estimator, Ratio type estimator & Almost unbiased ratio estimator. (04)

(b.) Derive bound for bias of ratio estimator. (05)

OR

(b.) When the second phase sample is a subsample of the first phase sample, the approximate bias of the ratio estimator is (05)

$$B(\hat{Y}_{RD}) = Y\left[\frac{V(\hat{X})}{X^2} - \frac{Cov(\hat{X},\hat{Y})}{XY} - \frac{Cov(\hat{X},\hat{X}_d)}{X^2} + \frac{Cov(\hat{Y},\hat{X}_d)}{YX}\right]$$

(c.) Prove that the product estimator $\hat{Y}_p$ is more efficient than $\hat{Y}$ if $\rho(\hat{X},\hat{Y}) < -\frac{1}{2}\frac{C(X)}{C(Y)}$ (05)

Q4. (a) (a) Explain Sampling and non sampling error. What are the sources of non sampling error?

(b) Under SRS $V(\hat{Y}_{SRS}) > MSE(\hat{Y}_{LR})$ and $MSE(\hat{Y}_R) > MSE(\hat{Y}_{LR})$. Prove it. (05)

(c) When the samples are drawn independently in two phases of sampling with the help of SRS the variance of the difference estimator is

$V(\hat{Y}_{DD}) = N^2\left[fS_Y^2 + \lambda^2(f+f')S_X^2 - 2\lambda f S_{XY}\right]$ Where $f = \frac{N-n}{Nn}$ and $f' = \frac{N-n'}{Nn'}$

Where n' and n are sample sizes corresponding to the first and second phases of sampling. (05)

OR

(c) When the second phase sample is a subsample of the first phase sample, the approximate mean square error is (05)

$$MSE(\hat{Y}_{RD}) = Y^2\left[\frac{V(\hat{Y})}{Y^2} + \frac{V(\hat{X})}{X^2} + \frac{V(\hat{X}_d)}{X^2} - 2\frac{Cov(\hat{X},\hat{Y})}{XY} - 2\frac{Cov(\hat{X},\hat{X}_d)}{X^2} + 2\frac{Cov(\hat{Y},\hat{X}_d)}{YX}\right]$$

---------X---------

Exam Seat No. _____

Date:3-12-2016	The Maharaja Sayajirao University of Baroda Subject-Statistics	Time:11.30-2.30 P.M.
Day: Saturday	M.Sc. Sem I (Dec 2016)(REGULAR & ATKT) STA2102-Linear Models(4 credits)	Marks:70

Supervisor's Sign _____

Marks Obtained _____

Examiner's signature _____

N.B.(i)This paper contains two parts.
(ii)Part-I has to be answered in the paper itself.
(iii) Part-II has to be given only when part-I is returned.
(iv) Part-I has to be returned to the invigilator during first 45 minutes of the commencement of the exam.

Part-I(28 marks)

1 Two statements are given below, for each of the statements choose the correct alternative and write it in the box: [2 x 5 = 10]

(a) Only S_1 is correct
(b) Only S_2 is correct
(c) Both S_1 and S_2 are correct
(d) Neither S_1 nor S_2 is correct

(i) S_1: The one-way classification is balanced if the numbers of observations under the different categories are same.
S_2: The random effects model are also termed as variance component model

(ii) S_1: In the usual notation if $|X'X| = 1$ the regressors are nonorthogonal.
S_2: If $|X'X| = 0$ there is an exact linear dependence among the regressors.

(iii) S_1: Ridge estimator shrinks the least square estimators toward the origin.
S_2: Ridge estimators are unbiased estimators.

(iv) S_1: If there is strong multi-collinearity then the variance of the least squares estimate of the regression coefficient β_j is very large.
S_2: If there is strong multi-collinearity between x_j and any subset of the other p-1 regressors, then the value of R_j^2 will be close to unity.

(v) S_1: In the plot of residuals against fitted values inward opening funnel indicates variance is a increasing function of y.
S_2: Variance inflation factor is given by $VIF = (1 - R_j^2)$

II. **Choose the correct alternative and write it in the box.[2 x 6=12]**

(i) In the usual notation a linear function of Y_1, Y_2----------Y_n is said to belong error if

(a) $E(d'Y) = 1$, $d' = (d_1, d_2$------$d_n)$
(b) $E(d'Y) > 1$, $d' = (d_1, d_2$------$d_n)$
(c) $E(d'Y) = 0$, $d' = (d_1, d_2$------$d_n)$
(d) $E(d'Y) > 0$, $d' = (d_1, d_2$------$d_n)$

(ii) In the regression model with p-1 predictor variables chosen from a set of P-1 possible predictor variables, which of the following indicates that bias is not a problem with a model?

(a) Mallow's $C_p \leq p$
(b) Mallow's $C_p \leq P$
(c) Mallow's $C_p > p$
(d) Mallow's $C_p > P$

(iii) A model in which both the types of a_{ij}'s - indicator variables and independent variables are present is called the

(a) Regression model
(b) Analysis of variance model
(c) Both (a) and (b)
(d) Analysis of covariance model

(iv) The full form of MINQUE is:

(a) Minimum quadratic unbiased estimator
(b) Minimum quadratic estimator
(c) Minimum norm quadratic unbiased estimator
(d) Minimum norm quadratic estimator

(v) In the usual notation Hat matrix is given by:

(a) $H = X(X'X)^{-1}$
(b) $H = (X'X)^{-1}$
(c) $H = X(X'X)^{-1}X'$
(d) $H = X(X'X)^{-1}$

(vi) Which one of the following is not a plausible remedy for near multicollinearity

(a) Use principal component analysis
(b) Drop one of the collinear variables
(c) Use a longer run of data
(d) Take logarithms of each of the variables

III. **Do as directed:[2 x 3=6]**

(i) State the properties of residuals.

onsider three independent random variables, X_1, X_2 and X_3, having common
riance σ^2 and expectations
(X_1) α α $E(X_2) = \mu_2$ and $E(X_3) = \mu_3$. Determine the condition

*****end of part-I*****

Exam Seat No.

Date: 8-11-2017	The Maharaja Sayajirao University of Baroda Subject-Statistics	Time: 11.30-2.30
Day: Wednesday	M.Sc. Sem I (Nov 2017)(REGULAR & ATKT) STA2102-Linear Models(4 credits)	Marks: 70

Part-II(42 marks)

1. (a) State and prove the gauss-markov theorem for linear model. Show that $\hat{\beta}$ is unbiased estimator of β, also obtain in the usual notation the dispersion matrix of $\hat{\beta}$. [4]
(b) In the usual notation explain the Gauss-Markov linear model with linear restrictions on parameters, also estimate the parameters. [4]
(c) Define (i) parametric function (ii) estimable function. [4]

OR

(c) Define (i) error space (ii) estimation space.

2. (a) Explain the method of determining simultaneous confidence intervals by using Cauchy-Schwartz inequality. [5]

(b) Explain how would you obtain power for F-test, what purpose does it serves explain? [6]

OR

(b) Explain Tukey's test.

3. (a) Explain the difference between regular residuals and PRESS residuals. [5]

(b) What is an outlier? Explain how would you test for an outlier. [5]

OR

(b) Explain normal probability plots.

4. Answer briefly: [9]

(a) Why Ridge estimator are also called shrinkage estimator?

(b) Explain the difference between ordinary least square estimates and ridge estimates.

(c) Explain the methods for choosing k in ridge estimator.

****

17-a
40

Exam Seat No._____

Date:8-11-2017	The Maharaja Sayajirao University of Baroda Subject-Statistics	Time:11.30-2.30 P.M.
Day: Wednesday	M.Sc. Sem I (Nov 2017)(REGULAR & ATKT) STA2102-Linear Models(4 credits)	Marks:70

Supervisor's Sign_____

Marks Obtained_____

Examiner's signature_____

N.B.(i)This paper contains two parts.
(ii)Part-I has to be answered in the paper itself.
(iii) Part-II has to be given only when part-I is returned.
(vi) Part-I has to be returned to the invigilator during first 45 minutes of the commencement of the exam.

Part-I(28 marks)

1. Choose the correct alternative and write it in the box**[2x9=18]**

(i)Which of the following assumptions are required to show the consistency, unbiasedness and efficiency of the OLS estimator?

(a) $E(u_t) = 0$

(b) $V(u_t) = \sigma^2$

(c) $Cov(u_t,u_{t-j}) = 0$ for all j

(d) $u_t \sim N(0,\sigma^2)$

(A) (b) and (d) only

(B) (a) and (c) only

(C) (a), (b) and (c) only

(D) (a) , (b) ,(c) and (d) only

(ii) What is the meaning of the term "hetroscedasticity"?

(A)The variance of the errors is not constant

(B) The variance of the dependent variable is not constant

(C) The errors are not linearly independent of each other

(D)The errors have non zero mean

(iii) What would be then consequences for the OLS estimator if heteroscedasticity is present in regression model but ignored?

(A) It will be biased
(B) It will be inconsistent
(C) It will be inefficient
(D) All of (A), (B) and (C) will be true

(iv) Which of the following are plausible approaches to dealing with a model that exhibits heteroscedasticity?

(i) Take logarithms of each of the variables
(ii) Use suitably modified standard errors
(iii) Use a generalized least squares procedure
(iv) Add lagged values of the variables to the regression equation

(A) (ii) and (iv) only
(B) (i) and (iii) only
(C) (i), (ii) and (iii) only
(D) (i), (ii), (iii) and (iv)

(v) Which one of the following is not a plausible remedy for near multicollinearity

(A) Use principal component analysis
(B) Drop one of the collinear variables
(C) Use a longer run of data
(D) Take logarithms of each of the variables

(vi) What will be the properties of OLS estimator in presence of multicollinearity

(A) It will be consistent, unbiased and efficient
(B) It will be consistent and unbiased but not efficient
(C) It will be consistent but not unbiased
(D) It will not be consistent

(vii) Multicollinearity refers to a situation in which

(A) Successive error terms derived from the application of regression analysis to time series data are correlated.
(B) There is a high degree of correlation between the independent variables included in a multiple regression model.
(C) The dependent variable is highly correlated with the independent variables in a regression analysis
(D) The application of a multiple regression model yields estimated that are nonlinear in form.

(viii) In the usual notation a model in which the coefficient a_{ij}'s take indicator variables is called

(A) Regression model

alysis of variance model

(C)Both (A) and (B)

(D)Analysis of covariance model

(x) In the usual notation a linear function of Y_1, Y_2----------Y_n is said to belong to error if

(A)$E(d'Y) = 1$, $d' = (d_1, d_2$------$d_n)$

(B)$E(d'Y) > 1$, $d' = (d_1, d_2$------$d_n)$

(C)$E(d'Y) = 0$, $d' = (d_1, d_2$------$d_n)$

(D)$E(d'Y) > 0$, $d' = (d_1, d_2$------$d_n)$

II. Fill in the blanks [2 x 5=10]

(i) In the usual notation Ridge estimator is given by_____.

(ii) In the usual notation VIF is given by_____

(iii) In the usual notation studentized residual is given by _____.

(iv) The vector space orthogonal to estimation space is called _____.

(v) Full form of MINQUE is_____.

********end of part-I********

63

Exam Seat No. _____

36 a
15

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

Date/Day
30/4/2024
Tuesday

FSMSC Previous Examination(May 2024)
STA2102C02-Linear Models(4 credits)

Time: 11.30 am to 2.30 pm

Total Marks: 70

Instructions: *(i)All the questions have to be answered in the answerbook only.*

Q-I **(a)** **Choose the correct alternative** **[16]**

1. In the theory of least square method of estimating the unknown parameters of the model we minimize the sum of
   (A) Squares of observations
   (B) observations
   (C) deviations of observations from their expected values
   (D) Square of the deviations of observations from their expected values.

2. In the usual notation VIF is defined as
   (A) $VIF=1/R_j^2$ (B) $VIF=1/(1-R_j^2)$
   (C) $VIF=R_j^2$ (D) $VIF=(1-R_j^2)$

3. Full form of VIF is:
   (A) Variance inflation factor
   (B) Varying inflation factor
   (C) Varied inflation factor
   (D) None of these

4. Full form of MINQUE is
   (A) Maximum norm quadratic estimator
   (B) Minimum norm quadratic unbiased estimator
   (C) Minimum norm quadratic estimator
   (D) Maximum norm quadratic unbiased estimator

5. Which of the property is/are satisfied by Ridge estimator?
   (A) Ridge estimator are biased estimator
   (B) Ridge estimator has smaller variance than the unbiased estimator.
   (C) Both (A) and (B).
   (D) None of them is correc..

6. The vector of residuals **e** in terms of hat matrix H can be represented as:
   (A) $\mathbf{e} = Hy$
   (B) $\mathbf{e} = yH$
   (C) $\mathbf{e} = (H-I)y$
   (D) $\mathbf{e} = (I-H)y$

36 b
15

7. In the usual notation PRESS residuals are given by
(A) $e_{(i)} = e_i/h_{ii}$ (B) $e_{(i)} = 1/h_{ii}$
(C) $e_{(i)} = e_i/(1-h_{ii})$ (D) $e_{(i)} = e_i/(1-h_{ii})^2$

8. One way classification fixed effect model is given by:

(A) $Y_{ij} = \mu + a_i + e_{ij}$
(B) $Y_{ij} = \mu + a_i + e_{ij}$, $a_i \sim N(0, \sigma_a^2)$
(C) $Y_{ij} = \mu + a_i + e_{ij}$, $e_{ij} \sim N(0, \sigma_e^2)$
(D) $Y_{ij} = \mu + a_i + e_{ij}$, $a_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma_e^2)$

(b) Do as Directed

1. State the properties of residuals [3]
2. State the form of Hat matrix. [2]
3. In the usual notation state the logistic growth model, Gompertz model and Weibull growth model. [3]
4. In the usual notations state the form of C_p mallows statistic. [2]
5. In the usual notations state the form of Ridge estimator. [2]

Q-II **Answer the following questions briefly(Attempt any two)** [12]

1. For the Gauss Markoff linear model for uncorrelated variables in the usual notation show that $\hat{\beta}_j$ is a unique linear unbiased estimator of β_j for j=1,2------p with $var(\hat{\beta}_j) = \sigma^2 c^{ij}$ and $cov(\hat{\beta}_j, \hat{\beta}_{j'}) = \sigma^2 c^{ij'}$
2. In the usual notation prove that $\sigma^2 = R_0^2/(n-r)$, where r = rank(X), n= no. of observations
3. Explain (i) fixed effect model (ii) random effect model and (iii) mixed effect model.

Q-III **Do as Directed**

1. Explain the method of determining simultaneous confidence intervals by using Cauchy- Schwartz inequality. [5]
2. Explain how would you obtain power for F-test, what purpose does it serves explain? [5]

OR

2. Explain Tukey's test. [5]

Q-IV **Attempt any one**

1. Explain one way classification random effect model [10]
2. Write a short note on outliers.

Q-V **Do as Directed(Each question carries 5 marks)**

1. How does Ridge estimator differs from ordinary least square estimators? Explain [10]
2. Write short note on C_p Mallow's statistic.

****

Exam Seat No. 130627

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

15/12/2023
Friday

FSM.Sc. Previous Examination(Nov 2023)

STA2102C02-Linear Models(4 credits)

Time: 11.30 am to 2.30 pm

Total Marks: 70

Instructions: *(i)All the questions have to be answered in the answerbook only.*

Q-I (a) Fill in the blanks [14]

1. In the usual notations Variance inflation factor is given by_____

2. In the usual notations Press residuals are represented as_____

3. Euclidean norm of matrix $A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$ is_____

4. Full form of MINQUE is_____.

5. In the usual notation if the value of standardized residual d_i is _____, it indicates outlier.

6. In the usual notation Ridge estimator is given by_____

7. The vector of residuals **e** in terms of hat matrix H can be represented as_____

Q-I (b) Choose the correct alternative [8]

1. In the usual notation a linear function l'y is said to belong to error if _____ irrespective of the values of θ_1, θ_2-----θ_m.
   (A) $E(l'y) > 0$
   (B) $E(l'y) < 0$
   (C) $E(l'y) = 0$
   (D) None of these

2. What is the meaning of the term "hetroscedasticity"?
   (A) The variance of the errors is not constant
   (B) The variance of the dependent variable is not constant

P.T.O.

99/10

(C) The errors are not linearly independent of each other
(D)The errors have non zero mean

3. In the plot of residuals against the fitted values outward opening funnel pattern indicates that the
   (A) variance is a decreasing function of y.
   (B) variance is an increasing function of y.
   (C) variance is constant.
   (D) Nothing can be said.

4. If there is strong multi-collinearity between x_j and any subset of the other (p-1) regressors, then the value of R_j^2 will be.
   (A) close to unity
   (B) close to zero
   (C) close to -1
   (D) can be anything.

Q-I (c) **Do as Directed** [6]

1. State the properties of residuals
2. State one way random effect model with assumptions
3. State the form of C_p statistics

Q-II **Answer briefly** [11]

1. If $l'\beta$ is any estimable linear function of the parameters β_1, β_2------- β_p, then in the usual notation prove that [7]
   (i) there exists a unique linear function $c'Y$ of the random variables Y_1, Y_2 ---------Y_n such that $c \in V(A')$ and $E(c'Y) = l'\beta$.
   (ii) $Var(c'Y)$ is less than the the variance of any other linear unbiased estimator of $l'\beta$.

2. Explain the Gauss Markoff set up for uncorrelated variables. [4]

OR

2. Define (i) error space (ii) estimation space [4]

Q-III **Do as Directed** [11]

1. State and prove the second fundamental theorem of least square theory. [6]

OR

1. Assuming the Gauss Markov linear set up for the observations and assuming that the observations are normally and independently distributed, derive the ANOVA test for the hypothesis that k independent and consistent conditions have assigned values: $H'\beta = O_0$. [6]
2. Explain briefly Scheffe's test [5]

Continue ...(3)

29.C
/40

Q-IV **Attempt any <u>two</u>** **[10]**

1. Explain briefly MINQUE theory for estimation of parameters in linear models
2. Explain normal probability plots and its uses
3. Explain Studentized residuals

Q-V **Attempt any <u>two</u>** **[10]**

1. Write a note on Mallow's C_p statistic
2. Explain the ill effects of multicollinearity
3. Explain principal component regression

****

40a/50

Exam Seat No. 150627

Date:2-12-2024	The Maharaja Sayajirao University of Baroda Subject-Statistics	Time:11.30A.M.-2.30 P.M.
Day: Monday	FSMSC I (Nov 2024) STA2102C02 -Linear Models (4 credits)	Marks:70

N.B. All the questions must be answered in the answer book only

Q-1 **Choose the correct alternative and write only answer in the answer book** [28]

1. What is the meaning of the term "hetroscedasticity"?
   (A)The variance of the errors is not constant
   (B) The variance of the dependent variable is not constant
   (C) The errors are not linearly independent of each other
   (D)The errors have nonzero mean

2. In the usual notation one way random effect model is given by
   (A) $Y_{ij} = \mu + a_i + e_{ij}$, $e_{ij} \sim N(0,\sigma^2_e)$
   (B) $Y_{ij} = \mu + a_i + e_{ij}$, $a_i \sim N(0,\sigma^2_a)$
   (C) $Y_{ij} = \mu + a_i + e_{ij}$, $a_i \sim N(0,\sigma^2_a)$, $e_{ij} \sim N(0,\sigma^2_e)$
   (D) None of these

3. Which one of the following is not a plausible remedy for near multicollinearity
   (A) Use principal component analysis
   (B) Drop one of the collinear variables
   (C) Use a longer run of data
   (D) Take logarithms of each of the variables

4. In the plot of residuals against the fitted values outward opening funnel pattern indicates that the
   (A) variance is a decreasing function of y.
   (B) variance is an increasing function of y.
   (C) variance is constant.
   (D) Nothing can be said.

PTO

5. In the usual notation a linear function l'y is said to belong to error if _____ irrespective of the values of Θ_1, Θ_2----- Θ_m.

(A) $E(l'y) > 0$
(B) $E(l'y) < 0$
(C) $E(l'y) = 0$
(D) None of these

6. Full form of MINQUE is_____.

(A) Minimum norm quadratic unbiased estimator
(B) Maximum norm quadratic unbiased estimator
(C) Minimax norm quadratic unbiased estimator
(D) Maximin norm quadratic unbiased estimator

7. In the usual notations Press residuals are represented as_____

(A) $e_{(i)} = e_i/h_{ii}$
(B) $e_{(i)} = e_i/(1-h_{ii})$
(C) $e_{(i)} = e_i^2/h_{ii}$
(D) $e_{(i)} = e_i^2/(1-h_{ii})$

8. In the usual notation which of the following is characteristic of standardized residual d_i:

(A) Mean of $d_i = 1$ and variance of $d_i = 0$

(B) Mean of $d_i = 1$ and variance of $d_i = 1$

(C) Mean of $d_i = 1$ and variance of $d_i = 0$

(D) Mean of $d_i = 0$ and variance of $d_i \sim 1$

9. In the regression model with p-1 predictor variables chosen from a set of P-1 possible predictor variables, which of the following indicates that bias is not a problem with a model?

(A) Mallow's $C_p \leq p$
(B) Mallow's $C_p \leq P$
(C) Mallow's $C_p > p$
(D) Mallow's $C_p > P$

40c/50

10. The vector of residuals **e** in terms of hat matrix H can be represented as:
   (A) $e = Hy$
   (B) $e = yH$
   (C) $e = (H-I)y$
   (D) $e = (I-H)y$

11. A linear manifold in a vector space V is any subset of vectors M closed under_____
   (A) Addition
   (B) Scalar multiplication
   (C) Division
   (D) Both (A) and (B)

12. In the usual notations normal equations for estimating the parameters are by least square method is _____
   (A) $(X'X)\hat{\beta} = X'Y$
   (B) $(X'X)^{-1}\hat{\beta} = X'Y$
   (C) $(X'X)\hat{\beta} = X'Y\,X$
   (D) $(X'X)\hat{\beta} = Y$

13. In the usual notation if the value of standardized residual d_i is _____, it indicates outlier.
   (A) Equal to 3
   (B) Less than 3
   (C) More than 3
   (D) None of these

14. Euclidean norm of matrix $B = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$ is_____
   (A) 6 (B) 8 (C) 10 (D) none of these

3/4 → PTO

Q-II **Attempt any TWO** **[10]**

1. State and prove the gauss-markov theorem for linear model. Show that $\hat{\beta}$ is unbiased estimator of β, also obtain in the usual notation the dispersion matrix of $\hat{\beta}$.

2. For the following model

$E(Y_i) = \delta + \beta (x_i - \bar{x})$, i=1,2,--------n

$D(Y) = \sigma^2 I$

Obtain the LS estimators of δ and β and show that cov($\hat{\delta}$, $\hat{\beta}$)= 0

3. Explain how you would reduce the Gauss Markov set up

$(Y, X\beta, \sigma^2 G)$ to standard form where G is a known positive definite symmetric matrix

Q-III **Attempt any TWO** **[10]**

1. Assuming the Gauss Markov linear set up for the observations and assuming that the observations are normally and independently distributed, derive the ANOVA test for the hypothesis that k independent and consistent conditions have assigned values: $H'\beta = \Theta_0$.

2. Write short note on Scheffe's test.

3. Explain the significance of power of F-test.

Q-IV **Attempt any TWO** **[12]**

1. Explain MINQUE theory

2. What do you mean by scaled residuals? State different types of scaled residuals. Explain any one of them.

3. Write a note on residual plots.

Q-V **Attempt any TWO** **[10]**

1. Explain the various methods for detecting multicollinearity.

2. Write short note on Principal component regression.

3. Explain the effects of multicollinearity.

****

Exam Seat No. 156626

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

09/12/2022 FSM.Sc. Previous Examination(Dec 2022) Time: 11.30 am

Friday STA2102C02-Linear Models(4 credits) to 2.30 pm

Total Marks: 70

Instructions: ***(i)All the questions have to be answered in the answerbook only.***

Q1. Choose the correct answer and write only answer in the answerbook[2 x 14=28]

(1) In the usual notation a linear function l'y is said to belong to error if _____ irrespective of the values of Θ_1, Θ_2-----Θ_m.

(A) $E(l'y) > 0$
(B) $E(l'y) < 0$
(C) $E(l'y) = 0$
(D) None of these

(2) In the usual notation a model in which the coefficient a_{ij}'s take indicator variables is called
(A) Regression model
(B) Analysis of variance model
(C) Both (A) and (B)
(D) Analysis of covariance model

(3) In the vector space theory estimation space and error space are _____to each other
(A) Orthogonal
(B) Perpendicular
(C) Can be either (A) or (B)
(D) None of these

(4) Full form of MINQUE is_____.
(A) Minimum norm quadratic unbiased estimator
(B) Maximum norm quadratic unbiased estimator
(C) Minimax norm quadratic unbiased estimator
(D) Maximin norm quadratic unbiased estimator

(5) A necessary and sufficient condition for the linear function l'ß of the parameters to be linearly estimable is_____
(A) Rank(A) = rank $\begin{pmatrix} A \\ l' \end{pmatrix}$
(B) Rank(A) > rank $\begin{pmatrix} A \\ l' \end{pmatrix}$
(C) Rank(A) < rank $\begin{pmatrix} A \\ l' \end{pmatrix}$
(D) None of these

(6) A linear manifold in a vector space V is any subset of vectors M closed under

(A) Addition
(B) Scalar multiplication
(C) Division
(D) Both (A) and (B)

(7) Which of the following is not a property of residuals?

(A) Mean of residuals is zero
(B) Approximate average variance is estimated by MSS_E
(C) Residuals are independent
(D) Only (B)

(8) In the usual notations Press residuals are represented as _____

(A) $e_{(i)} = e_i/h_{ii}$
(B) $e_{(i)} = e_i/(1-h_{ii})$
(C) $e_{(i)} = e_i^2/h_{ii}$
(D) $e_{(i)} = e_i^2/(1-h_{ii})$

(9) In the plot of residuals against fitted values inward opening funnel indicates

(A) Variance is a decreasing function of y
(B) Variance is a increasing function of y
(C) Nothing can be interpreted about variance.
(D) There are no such plots of residuals.

(10) In the usual notation if $|X'X| = 1$ indicates regressors are

(A) Orthogonal
(B) Nonorthogonal
(C) Nothing can be interpreted about regressors from the above
(D) There is exact linear dependence among the regressors.

(11) In the usual notations Normal probability plot is a plot of _____

(A) e_i versus cumulative probability $P_i = i$
(B) e_i versus cumulative probability $P_i = (i - ½)/n$
(C) $e_{[i]}$ versus cumulative probability $P_i = i$
(D) $e_{[i]}$ versus cumulative probability $P_i = (i - ½)/n$

(12) In the usual notations Standardized residuals are given by _____

(A) $d_i = e_i/\sqrt{MSSE}$
(B) $d_i = e_i^2/\sqrt{MSSE}$
(C) $d_i = e_i\sqrt{MSSE}$
(D) $d_i = e_i^2\sqrt{MSSE}$

(13) In the usual notation formula for C_p statistic is _____

(A) $C_p = SS_{res}(p) - n + 2p$
(B) $C_p = (SS_{res}(p) - n + 2p)/\hat{\sigma}^2$
(C) $C_p = SS_{res}(p) - (n + 2p)/\hat{\sigma}^2$
(D) $C_p = (SS_{res}(p))/\hat{\sigma}^2 - n + 2p$

50-C
50

(14) In the regression model with p-1 predictor variables chosen from a set of P-1 possible predictor variables, which of the following indicates that bias is not a problem with a model?

(A) Mallow's $C_p \leq p$

(B) Mallow's $C_p \leq P$

(C) Mallow's $C_p > p$

(D) Mallow's $C_p > P$

Q-II

(a) State and prove the gauss-markov theorem for linear model. Show that $\hat{\beta}$ is unbiased estimator of β, also obtain in the usual notation the dispersion matrix of $\hat{\beta}$. [6]

OR

(a) Explain the terms estimable parametric function', 'error function' and 'best linear unbiased estimator' in connection with the problem of linear estimation [6]

(b) Explain (i) fixed effect model (ii) random effect model and (iii) mixed effect model. [4]

Q-III

(a) In the usual notations prove that $R_0^2 \sim \sigma^2 \chi^2_{(n-r)}$ where r is the rank of matrix X. [5]

(b) In the usual notation explain the procedure of testing a single parametric function also obtain the confidence interval. [6]

OR

(b)Explain the use of OC curves in linear models also explain for what purpose power of F test is used? [6]

Q-IV

(a) Write short note on MINQUE theory. [7]

OR

(a) Write short note on outliers. [7]

(b) Explain PRESS residuals. [4]

Q-V Attempt any two: [10]

(a) Ridge regression

(b) Principal component regression

(c) Mallow's C_p statistic.

.....

Exam Seat No _____

Date 3-12-2016	The M.S. University of Baroda Subject-Statistics	Time.11.30-2.30 P.M.
Day Saturday	M.Sc. Sem I (Dec 2016)(REGULAR & ATKT) STA2102-Linear Models(4 credits)	Marks:70

Part-II(42 marks)

1.(a) For the Gauss Markoff linear model for uncorrelated variables in the usual notation show that $\hat{\beta}_j$ is a unique linear unbiased estimator of β_j for j=1,2------p with $var(\hat{\beta}_j) = \sigma^2 c^{jj}$ and $cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c^{ij}$ [6]

(b) Attempt any two: [2 x 3 = 6]

(i) Explain (i) fixed effect model (ii) random effect model and (iii) mixed effect model.

(ii) In the usual notation prove that $\sigma^2 = R_0^2/(n-r)$, where r = rank(X), n= no. of observations.

(iii) In the usual notation state and prove the necessary and sufficient condition for the linear function $l'\beta$ to be linearly estimable.

2. Attempt any two:[2 x 5=10]

(i) Write short note on power of F-test.

(ii) Explain how would you test multiple hypothesis.

(iii) Write short note on Tukey's test.

3. Explain in detail one way classification random effect model.[10]

OR

3. Write short note on MINQUE theory.[10]

4. (a) Explain by giving example what do you mean nonlinear regression models, how would you estimate parameters in case of non linear regression models? [5]

(b) Explain the various methods for detecting multicollinearity. [5]

OR

(b) Explain principal component regression.

****