

INDIAN STATISTICAL INSTITUTE
SQC & OR Unit, Hyderabad

MS in Quality Management Science: 2015 -17
III SEMESTER: FINAL EXAMINATION

Subject: Applied Regression Analysis (ARA)

Date: 7 November 2016

Duration: 3 Hours

Max. Marks: 100

INSTRUCTIONS

*This paper contains questions for 120 marks. Answer as many as you can but the maximum you can score is 100 marks. You will also be given soft copy of this question paper for using data for analysis. You are free to use appropriate statistical software for analysis but the answers will have to be given in the answer sheet given to you. You need to also submit the soft copy of your analysis if any, only in a **word file format** opened in your name before the end of the examination without fail.*

- Q1. Answer any **five** of the following. Your answer should be brief (Max. 20 lines) and you may use examples wherever relevant.
- a. Different variable selection methods in a multiple regression set up highlighting criteria used in variable selection in each method. Also, mention the salient merits and demerits of each method.
 - b. Methods to handle multi collinearity among regressors. Also, mention the salient merits and demerits of each method.
 - c. Methods to handle non-linearity in modelling and mention the main significant difference of these methods over traditional linearization approaches.
 - d. Practical significance of validation in regression modelling and list commonly adopted validation methods with significance of each method.
 - e. Measures for checking the model adequacy in a multiple regression set up and salient merits and/or demerits of each measure.
 - f. Role of different classification and tree based methods in modelling and main criterion used in each method.
 - g. How outliers influence model adequacy and list different methods to handle outliers in the data while modelling. Specify the decision criteria used in each method.

[5X10 = 50 Marks]

- Q2. Following table contains the data on Exports in Tons of Production made by a company during the 2007-12.
- a) Fit an appropriate forecasting model to the data and give the quarter wise forecast along with the suitable prediction interval for the next two years.
 - b) Explain the criteria you have adopted in building and selecting the final model.

- c) Also find the AIC and BIC estimates for each of the models you have tried in questions a) & b) above and explain how you would use these criteria in your final model selection.

[10+5+10 = 25 Marks]

Year	Quarter	Period	Exports in Tons
2007	Apr	1	362
	Jul	2	385
	Oct	3	432
	Jan	4	341
2008	Apr	5	382
	Jul	6	409
	Oct	7	498
	Jan	8	387
2009	Apr	9	473
	Jul	10	513
	Oct	11	582
	Jan	12	474
2010	Apr	13	544
	Jul	14	582
	Oct	15	681
	Jan	16	557
2011	Apr	17	628
	Jul	18	707
	Oct	19	773
	Jan	20	592
2012	Apr	21	627
	Jul	22	725
	Oct	23	854
	Jan	24	661

Q3. The following is the data on Cost of Healthcare which was collected by the Department of Health and Social Services of the State of New Mexico and cover 52 of the 60 licensed facilities in New Mexico in 1988. The variables in the data are the characteristics which describe the facilities size, volume of usage, expenditures and revenue. The location of the facility is indicated whether it is in the rural or non-rural area.

- i) Using the data given, build a logistic regression model to understand whether the rural facilities are different from non-rural facilities. List the model adequacy

measures that need to be used in logistic modelling and verify the model adequacy of the developed model.

- ii) Also, build an appropriate model for patient care revenue using hospital characteristics. Comment on the adequacy of the model developed by using appropriate diagnostic measures.
- iii) Estimate the AIC and BIC for models tried and comment on the adequacy of the model selected in both cases using these criteria.

[15+15+15 = 45 Marks]

Variable	Description
RURAL	Rural home (1) and Non-rural home (0)
BED	Number of Beds in home
MCDAYS	Annual medical in-patient days (hundreds)
TDAYS	Annual total patient days (hundreds)
PCREV	Annual total patient care revenue (\$100)
NSAL	Annual nursing salaries (\$100)
FEXP	Annual facilities expenditures (\$100)
NETREV	PCREV-NSAL-FEXP

RURAL	BED	MCDAYS	TDAYS	PCREV	NSAL	FEXP	NETREV
0	244	128	385	23521	5230	5334	12957
1	59	155	203	9160	2459	493	6208
0	120	281	392	21900	6304	6115	9481
0	120	291	419	22354	6590	6346	9418
0	120	238	363	17421	5362	6225	5834
1	65	180	234	10531	3622	449	6460
1	120	306	372	22147	4406	4998	12743
1	90	214	305	14025	4173	966	8886
0	96	155	169	8812	1955	1260	5597
1	120	133	188	11729	3224	6442	2063
0	62	148	192	8896	2409	1236	5251
1	120	274	426	20987	2066	3360	15561
0	116	154	321	17655	5946	4231	7478
1	59	120	164	7085	1925	1280	3880
1	80	261	284	13089	4166	1123	7800
1	120	338	375	21453	5257	5206	10990
1	80	77	133	7790	1988	4443	1359
1	100	204	318	18309	4156	4585	9568
1	60	97	213	8872	1914	1675	5283
1	110	178	280	17881	5173	5686	7022

0	120	232	336	17004	4630	907	11467
0	135	316	442	23829	7489	3351	12989
1	59	163	191	9424	2051	1756	5617
0	60	96	202	12474	3803	2123	6548
1	25	74	83	4078	2008	4531	-2461
1	221	514	776	36029	1288	2543	32198
1	64	91	214	8782	4729	4446	-393
0	62	146	204	8951	2367	1064	5520
1	108	255	366	17446	5933	2987	8526
1	62	144	220	6164	2782	411	2971
0	90	151	286	2853	4651	4197	-5995
0	146	100	375	21334	6857	1198	13279
1	62	174	189	8082	2143	1209	4730
1	30	54	88	3948	3025	137	786
0	79	213	278	11649	2905	1279	7465
1	44	127	158	7850	1498	1273	5079
0	120	208	423	29035	6236	3524	19275
1	100	255	300	17532	3547	2561	11424
1	49	110	177	8197	2810	3874	1513
1	123	208	336	22555	6059	6402	10094
1	82	114	136	8459	1995	1911	4553
1	58	166	205	10412	2245	1122	7045
1	110	228	323	16661	4029	3893	8739
1	62	183	222	12406	2784	2212	7410
1	86	62	200	11312	3720	2959	4633
1	102	326	355	14499	3866	3006	7627
0	135	157	471	24274	7485	1344	15445
1	78	154	203	9327	3672	1242	4413
1	83	224	390	12362	3995	1484	6883
0	60	48	213	10644	2820	1154	6670
1	54	119	144	7556	2088	245	5223
0	120	217	327	20182	4432	6274	9476
