

Homework 2- Masters Group 9

Group Members:

- A) Oluwaseyitan Awojobi
- B) Mohamed Elgendy
- C) Ramya Koya
- D) Karthik Sagar Tadi
- E) Vijay Tulluri

SOLUTION

5. Model 1.

- a. The R^2 value is 0.7619 while the adjusted R^2 value is 0.6862.
- b. Based on the t value and $p > |t|$, the NW value is significant with a p value of $<.0001$

Number of Model Parameters	15
Number of Observations	59

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	172186	12299	10.06	<.0001
Error	44	53807	1222.879841		
Corrected Total	58	225993			

Model Fit Statistics

R-Square	0.7619	Adj R-Sq	0.6862
AIC	432.1213	BIC	444.1162
SBC	463.2844	C(p)	15.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	1400.0	281.6	4.97	<.0001
Education	1	-11.0467	8.9980	-1.23	0.2261
HCPot	1	-0.6712	0.4556	-1.47	0.1478
HHSiz	1	-38.0416	40.2752	-0.94	0.3501
JanTemp	1	-1.4411	0.7638	-1.89	0.0658
JulyTemp	1	-2.9503	1.9351	-1.52	0.1345
NOxPot	1	1.1785	0.9144	1.29	0.2042
NW	1	5.3027	0.9064	5.85	<.0001
PopDensity	1	0.00472	0.00434	1.09	0.2826
Rain	1	0.9695	0.5851	1.66	0.1046
RelHum	1	0.1360	1.1508	0.12	0.9065
S02Pot	1	0.0846	0.1357	0.62	0.5362
WC	1	-1.4923	1.2325	-1.21	0.2324
income	1	-0.00043	0.00129	-0.33	0.7435
pop	1	3.402E-6	4.116E-6	0.83	0.4129

6. Model 2.

- a. In this model, we are removing data that is considered insignificant to the overall decision making process which is the factors considered relevant to the mortality rate. Insignificant data have a tendency of reducing the precision of the standard errors of valid predictor variables.
- b. The R2 value is 0.4180 while the adjusted R2 is 0.4078.
- c. The NW value is significant with a P-value of <.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	94473	94473	40.94	<.0001
Error	57	131520	2307.364768		
Corrected Total	58	225993			

Model Fit Statistics			
R-Square	0.4180	Adj R-Sq	0.4078
AIC	458.8531	BIC	460.9910
SBC	463.0082	C(p)	2.0000

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	887.9	10.4124	85.27	<.0001
NW	1	4.4855	0.7010	6.40	<.0001

7.

- a. 7 Principal components were selected.
- b. This seems like a reasonable number. Given that the Principal components selected gives an 86.5% explanation for the variation which is pretty high. Since the goal of the PCA is to remove as much dimensionality as possible while still keeping as much variation, the selected principal components are reasonable.

```

75
76
77
78 *-----*
79 Summary of Exported Principal Components
80 *-----*
81
82 Remark:The number of inputs is used as the maximum number of principal components
83
84 Total number of input variables: 14
85 Maximum number cutoff of principal components: 14
86 Cumulative proportional eigenvalue cutoff: 0.85
87 Proportional eigenvalue increment cutoff: 0.001
88 Number of the selected principal components: 7
89 Total variation explained by the selected principal components: 0.8654705662
90
91
92 *-----*
93 * Score Output
94 *-----*
95
96
97 *-----*
98 * Report Output
99 *-----*
100

```

8. Model 3

- a. The R2 is 0.6390 and adjusted R-Square is 0.5895
- b. The significant variables in respect to the t value statistics and associated probabilities ($Pr > |t|$) are PC_1, PC_2, PC_3, PC_6.

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	144410	20630	12.90	<.0001
Error	51	81582	1599.651016		
Corrected Total	58	225993			

Model Fit Statistics			
R-Square	0.6390	Adj R-Sq	0.5895
AIC	442.6779	BIC	447.1385
SBC	459.2982	C(p)	8.0000

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	941.2	5.2070	180.75	<.0001
PC_1	1	-11.3761	2.6159	-4.35	<.0001
PC_2	1	22.9959	3.3944	6.77	<.0001
PC_3	1	-10.1128	3.8837	-2.60	0.0120
PC_4	1	-5.6194	4.5557	-1.23	0.2230
PC_5	1	6.8720	5.1380	1.34	0.1870
PC_6	1	21.4491	5.4783	3.92	0.0003
PC_7	1	1.5443	6.9471	0.22	0.8250

9. Model 4

- R² value is 0.6152 and adjusted R² value is 0.5867
- Significant values based on the t value statistic and associated probabilities are PC_1, PC_2, PC_4 and PC_6

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	139036	34759	21.59	<.0001
Error	54	86957	1610.308208		
Corrected Total	58	225993			

Model Fit Statistics

R-Square	0.6152	Adj R-Sq	0.5867
AIC	440.4420	BIC	443.3508
SBC	450.8297	C(p)	5.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	941.2	5.2243	180.15	<.0001
PC_1	1	-11.3761	2.6246	-4.33	<.0001
PC_2	1	22.9959	3.4057	6.75	<.0001
PC_3	1	-10.1128	3.8966	-2.60	0.0121
PC_6	1	21.4491	5.4965	3.90	0.0003

10. SUMMARY

Model 1: This is significant with a P- value of < .001. All of the variables however, except one (NW) are insignificant in relation to its P value and T statistics. It also has a low F value of 10.06.

Model 2: This model considers just 1 independent variable. It is considered significant with a P- value of <.0001 and a considerably high F-statistics of 40. 94. It however has the lowest R2 and adjusted R2 values.

Model 3: This model has a low F-value and above average R-squared. It however has a combination of relevant principal components which are able to predicts effectively causes of mortality.

Model 4: This has a low F –value and an above average R-Squared. Al variables here are statistically significant.

The most preferred Model would be Model 4. This is because all variables (Principal components) in this model are statistically significant. It also explains a reasonable amount of variation given its R-squared value.