Assignment for Applied Regression Analysis

朱强强

17064001

应用统计学1701

9.2

Correlation of Estimates						
Variable	**Label**	**Intercept**	**x1**	**x2**	**x3**	**x4**
Intercept	Intercept	1.0000	-0.9678	-0.9978	-0.9769	-0.9983
x1	x1	-0.9678	1.0000	0.9510	0.9861	0.9568
x2	x2	-0.9978	0.9510	1.0000	0.9624	0.9979
x3	x3	-0.9769	0.9861	0.9624	1.0000	0.9659
x4	x4	-0.9983	0.9568	0.9979	0.9659	1.0000

From the figure above we find that the matrix reveals the high correlations between these variables, so we consider that there is a significant multicollinearity.

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	Intercept	1	62.40537	70.07096	0.89	0.3991	0
x1	x1	1	1.55110	0.74477	2.08	0.0708	38.49621
x2	x2	1	0.51017	0.72379	0.70	0.5009	254.42317
x3	x3	1	0.10191	0.75471	0.14	0.8959	46.86839
x4	x4	1	-0.14406	0.70905	-0.20	0.8441	282.51286

The variance inflation factors are shown in the last column of the table above.

Collinearity Diagnostics							
			Proportion of Variation				
Number	**Eigenvalue**	**Condition Index**	**Intercept**	**x1**	**x2**	**x3**	**x4**
1	4.11970	1.00000	0.00000551	0.00036889	0.00001833	0.00021022	0.00003641
2	0.55389	2.72721	8.812348E-8	0.01004	0.00001265	0.00266	0.00010070
3	0.28870	3.77753	3.060952E-7	0.00057551	0.00031981	0.00159	0.00168
4	0.03764	10.46207	0.00012679	0.05745	0.00278	0.04569	0.00088373
5	0.00006614	249.57825	0.99987	0.93157	0.99687	0.94985	0.99730

The condition number of correlations between the regressors is shown in the third column of the table above. Because the last condition index surpasses 100, we believe that there is a sufficient evidence of multicollinearity in these data.

```
/*9.2*/
proc import datafile='E:\Applied Regression Analysis\Data Sets\data-ex-10-1_(Hald_Cement).XLS' out=data1 dbms=XLS replace;
    getnames=yes;
run;
proc print data=data1;
run;
proc reg data=data1;
    model y=x1-x4/corrb vif collin;
run;
```

9.2

The elements of the vector is

Eigenvalue
4.11970
0.55389
0.28870
0.03764
0.00006614

All elements are the sources of multicollinearity in these data.

9.4

Collinearity Diagnostics						
			Proportion of Variation			
Number	Eigenvalue	Condition Index	x1	x2	x3	x4
1	3.12306	1.00000	0.01027	0.00706	0.00796	0.01817
2	0.55350	2.37538	0.15559	0.00253	0.05983	0.03077
3	0.28834	3.29109	0.00806	0.07394	0.03747	0.47065
4	0.03510	9.43246	0.82609	0.91647	0.89474	0.48042

Because all condition indexes are less than 100, we believe that these is no evidence of multicollinearity in this data.

```
/*9.4*/
proc reg data=data1;
model y=x1-x4/noint collin;
run;
```

9.5

Collinearity Diagnostics								
			Proportion of Variation					
Number	Eigenvalue	Condition Index	Intercept	x1	x2	x3	x4
1	2.23570	1.00000	0	0.00263	0.00055897	0.00148	0.00047533
2	1.57607	1.19102	0	0.00427	0.00042729	0.00495	0.00045729
3	1.00000	1.49523	1.00000	0	0	0	0
4	0.18661	3.46134	0	0.06352	0.00208	0.04650	0.00072440
5	0.00162	37.10634	0	0.92958	0.99693	0.94707	0.99834

The regressors with centering is better.

```
/*9.5*/
data new_data1;
    set data1;
run;
proc standard data=new_data1 mean=0 out=new_data1;
    var x1-x4;
run;
proc reg data=new_data1;
    model y=x1-x4/collin;
run;
```

10.1

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-1.80837	7.90086	0.15251	0.05	0.8209
x2	0.00360	0.00069500	78.02809	26.80	<.0001
x7	0.19396	0.08823	14.06819	4.83	0.0378
x8	-0.00482	0.00128	41.40006	14.22	0.0009

The three procedures chose the same model, which is $y = \beta_0 + \beta_2 x_2 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon$.

```
/*10.1*/
proc import datafile='E:\Applied Regression Analysis\Data Sets\data-table-B1.XLS' out=data2 dbms=XLS replace;
    getnames=yes;
run;
proc print data=data2;
run;
proc reg ;
```

```
    model y=x1-x9/selection=forward;
run;
proc reg ;
    model y=x1-x9/selection=backward;
run;
proc reg ;
    model y=x1-x9/selection=stepwise;
run;
```

10.2

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	265.02306	44.17051	14.98	<.0001
Error	21	61.94123	2.94958		
Corrected Total	27	326.96429			

Root MSE	1.71743	R-Square	0.8106
Dependent Mean	6.96429	Adj R-Sq	0.7564
Coeff Var	24.66060		

```
/*10.2*/
proc reg data=data2;
    model y=x1 x2 x4 x7-x9/adjrsq mse cp;
run;
```

13.1

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.0709	2.1092	8.2848	0.0040
x	1	0.0177	0.00608	8.4904	0.0036

The model is $\hat{\pi} = \dfrac{1}{1 + e^{-0.67+0.0177x}}$

```
/*13.1*/
proc import datafile='E:\Applied Regression Analysis\Data Sets\data-prob-13-1.XLS' out=data3 dbms=XLS replace;
    getnames=yes;
run;
proc print data=data3;
run;
proc logistic ;
    model y=x/link=logit;
run;
```

13.4

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	2.0848	0.0804	672.3907	<.0001
x	1	-0.1357	0.00496	749.6913	<.0001

The model is $\hat{\pi} = \dfrac{1}{1+e^{2.0848-0.1357x}}$.

```
/*13.4*/
proc import datafile='E:\Applied Regression Analysis\Data Sets\data-prob-13-4.XLS' out=data4 dbms=XLS replace;
    getnames=yes;
run;
proc print data=data4;
run;
data new_data4;
input x f y ;
datalines;
5   100 1
5   400 0
7   378 0
7   122 1
9   353 0
9   147 1
11  324 0
11  176 1
13  289 0
13  211 1
15  256 0
15  244 1
17  223 0
17  277 1
19  190 0
```

```
19  310 1
21  157 0
21  343 1
23  128 0
23  372 1
25  391 1
25  109 0
;
run;
proc logistic ;
model y=x/link=logit scale=n aggregate ;
freq f;
run;
```