

Regression with Metric Cars Data

Here is the file `mcars2.dat`

Ident	Cntry	kpl	weight	length
1	US	5.04	2178.0	591.82
2	Japan	10.08	1026.0	431.80
3	US	9.24	1188.0	426.72
4	US	7.98	1444.5	510.54
5	US	7.98	1485.0	502.92
6	US	7.98	1485.0	502.92
7	Europ	9.66	972.0	436.88
8	US	7.56	1665.0	543.56
.
.
.

and so on

```
/****** mcarsreg1.sas *****/
options linesize=79 noovp formdlim='-';
title 'Regression with Metric Cars Data: Part One';

data auto;
  infile 'mcars2.dat' firstobs=2 ;      /* Skipping the header on line 1 */
  input id country $ kpl weight length;
/* Indicator dummy vars: Ref category is Japanese */
  if country = 'US' then c1=1;
  else if country = 'Japan' then c1=0;
  else if country = 'Europ' then c1=0;
  if country = 'Europ' then c2=1;
  else if country = 'US' then c2=0;
  else if country = 'Japan' then c2=0;
  label country = 'Country of Origin'
        kpl = 'Kilometers per Litre'
        weight = 'Weight in kg'
        length = 'Length in cm';

proc freq;
  title2 'Check Dummy variable creation';
  tables country*(c1 c2) / norow nocol nopercnt;

proc reg;
  title2 'With both Weight and Length';
  model kpl = weight length c1 c2;
  country: test c1 = c2 = 0;      /* Country controlling for wgt, length */
  wgt_len: test weight=length=0; /* wgt, length controlling for Country */

proc reg;
  title2 'With Weight but not Length';
  model kpl = weight c1 c2;
  country: test c1 = c2 = 0;
  USvsEURO: test c1=c2;        /* US vs European controlling for weight */

proc plot;
  plot kpl*weight=country;
```

Here is mcarsreg1.lst

Regression with Metric Cars Data: Part One 1
Check Dummy variable creation
15:35 Sunday, January 28, 2007

The FREQ Procedure

Table of country by c1

country(Country of Origin)			
	c1		
Frequency	0	1	Total
Europ	14	0	14
Japan	13	0	13
US	0	73	73
Total	27	73	100

Table of country by c2

country(Country of Origin)			
	c2		
Frequency	0	1	Total
Europ	0	14	14
Japan	13	0	13
US	73	0	73
Total	86	14	100

The REG Procedure
 Model: MODEL1
 Dependent Variable: kpl Kilometers per Litre

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	461.08779	115.27195	44.73	<.0001
Error	95	244.80150	2.57686		
Corrected Total	99	705.88930			

Root MSE	1.60526	R-Square	0.6532
Dependent Mean	8.79480	Adj R-Sq	0.6386
Coeff Var	18.25237		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	24.49010	2.75828	8.88
weight	Weight in kg	1	-0.00207	0.00139	-1.49
length	Length in cm	1	-0.02874	0.00925	-3.11
c1		1	1.36286	0.55139	2.47
c2		1	1.24528	0.62222	2.00

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
weight	Weight in kg	1	0.1389
length	Length in cm	1	0.0025
c1		1	0.0152
c2		1	0.0482

Regression with Metric Cars Data: Part One 3
 With both Weight and Length
 15:35 Sunday, January 28, 2007

The REG Procedure
 Model: MODEL1

Test country Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	8.50611	3.30	0.0411
Denominator	95	2.57686		

Regression with Metric Cars Data: Part One 4
 With both Weight and Length
 15:35 Sunday, January 28, 2007

The REG Procedure
 Model: MODEL1

Test wgt_len Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	169.74773	65.87	<.0001
Denominator	95	2.57686		

Regression with Metric Cars Data: Part One 5
 With Weight but not Length
 15:35 Sunday, January 28, 2007

The REG Procedure
 Model: MODEL1

Dependent Variable: kpl Kilometers per Litre

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	436.21151	145.40384	51.76	<.0001
Error	96	269.67779	2.80914		
Corrected Total	99	705.88930			

Root MSE	1.67605	R-Square	0.6180
Dependent Mean	8.79480	Adj R-Sq	0.6060
Coeff Var	19.05728		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	16.22634	0.76312	21.26
weight	Weight in kg	1	-0.00604	0.00057080	-10.58
c1		1	1.23615	0.57413	2.15
c2		1	1.45959	0.64566	2.26

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
weight	Weight in kg	1	<.0001
c1		1	0.0338
c2		1	0.0260

Regression with Metric Cars Data: Part One 6
 With Weight but not Length
 15:35 Sunday, January 28, 2007

The REG Procedure
 Model: MODEL1

Test country Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	8.61683	3.07	0.0511
Denominator	96	2.80914		

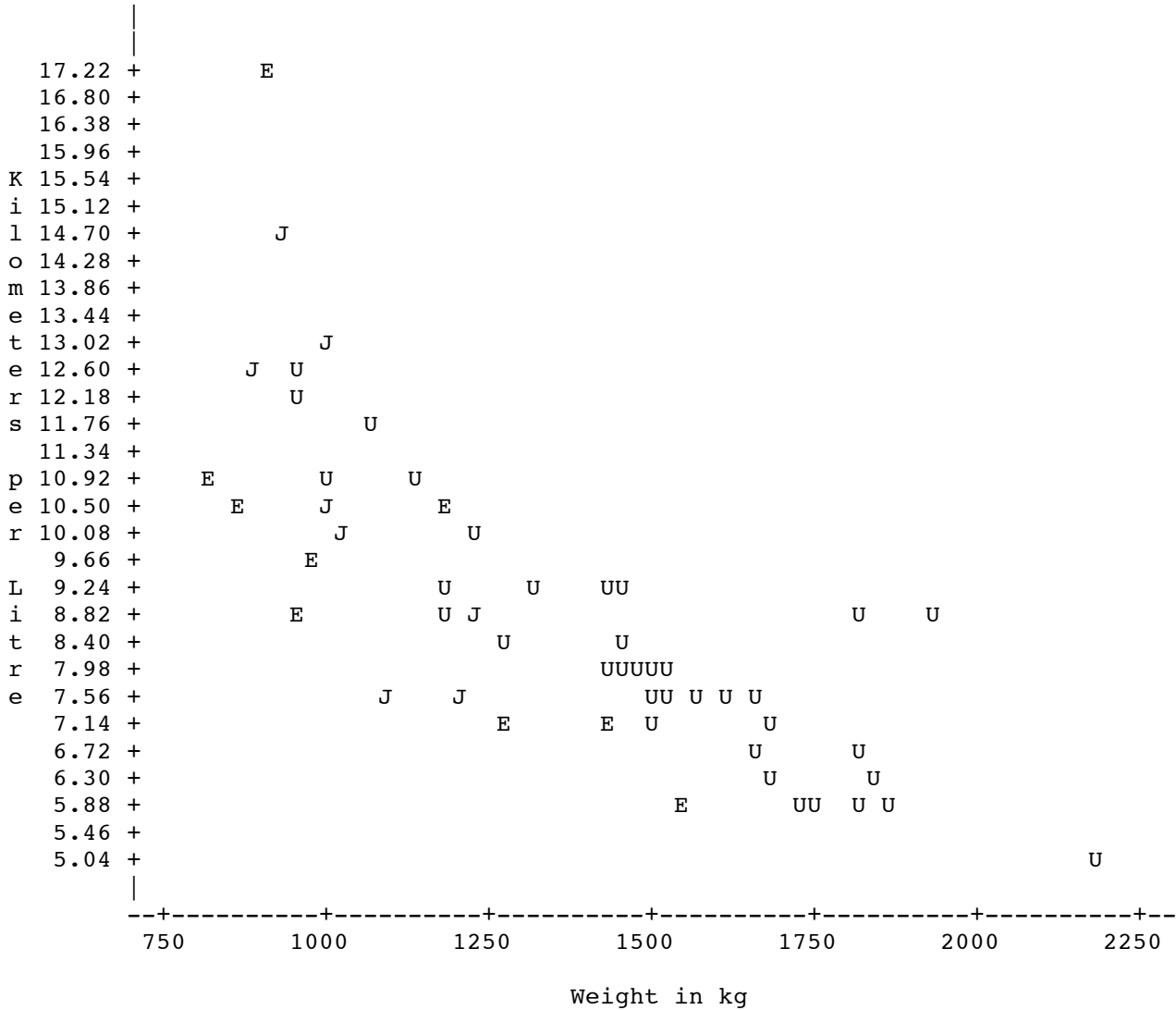
Regression with Metric Cars Data: Part One 7
 With Weight but not Length
 15:35 Sunday, January 28, 2007

The REG Procedure
 Model: MODEL1

Test USvsEURO Results for Dependent Variable kpl

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.45529	0.16	0.6881
Denominator	96	2.80914		

Plot of $kpl \cdot weight$. Symbol is value of country.



NOTE: 47 obs hidden.

```
/dos/brunner/429s07/regress > head switch.data ; cat switch.sas switch.lst
```

```
      x1      x2      x3      y
1      5.76    5.12    5.48   27.60
2      2.76    4.27    3.88   20.16
3     12.41   12.11   11.96   53.48
4      9.41   10.23    9.16   41.61
5      5.19    4.49    4.43   23.60
6      8.31    8.10    9.95   46.42
7     21.39   19.98   19.28   80.40
8      6.15    2.61    2.53   15.21
9     11.52   12.47   10.25   44.26
```

```
/****** switch.sas *****/
```

```
options linesize=79 pagesize=100 noovp formdlim='-';
title 'Adding an independent variable can change the conclusions';
```

```
data ohno;
  infile 'switch.data' firstobs=2 ;      /* Skipping the header on line 1 */
  input ident x1 x2 x3 y;
```

```
proc reg;
  title2 'Just x1 and x2';
  model y = x1 x2;
```

```
proc reg corr;
  title2 'Include x3 as well as x1 and x2';
  model y = x1 x2 x3;
```

```
Adding an independent variable can change the conclusions      1
      Just x1 and x2
                                23:51 Wednesday, January 31, 2007
```

```
      The REG Procedure
      Model: MODEL1
      Dependent Variable: y
```

```
      Analysis of Variance
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	39759	19879	644.69	<.0001
Error	97	2991.05857	30.83566		
Corrected Total	99	42750			

Root MSE	5.55299	R-Square	0.9300
Dependent Mean	47.03240	Adj R-Sq	0.9286
Coeff Var	11.80673		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	6.07321	1.27156	4.78	<.0001
x1	1	1.72679	0.42206	4.09	<.0001
x2	1	2.12112	0.43425	4.88	<.0001

Adding an independent variable can change the conclusions 2
 Include x3 as well as x1 and x2
 23:51 Wednesday, January 31, 2007

The REG Procedure

Correlation

Variable	x1	x2	x3	y
x1	1.0000	0.9688	0.9636	0.9554
x2	0.9688	1.0000	0.9727	0.9581
x3	0.9636	0.9727	1.0000	0.9977
y	0.9554	0.9581	0.9977	1.0000

Adding an independent variable can change the conclusions 3
 Include x3 as well as x1 and x2
 23:51 Wednesday, January 31, 2007

The REG Procedure

Model: MODEL1

Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	42676	14225	18523.4	<.0001
Error	96	73.72506	0.76797		
Corrected Total	99	42750			

Root MSE	0.87634	R-Square	0.9983
Dependent Mean	47.03240	Adj R-Sq	0.9982
Coeff Var	1.86327		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.09545	0.20130	25.31	<.0001
x1	1	0.08843	0.07172	1.23	0.2206
x2	1	-0.97442	0.08496	-11.47	<.0001
x3	1	4.87014	0.07902	61.63	<.0001

Reproducing the first set of parameter estimates, for comparison ...

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	6.07321	1.27156	4.78	<.0001
x1	1	1.72679	0.42206	4.09	<.0001
x2	1	2.12112	0.43425	4.88	<.0001