

Regression on the Metric Cars Data

```
/****** carsreg1.sas *****/
options linesize=79 noovp formdlim='- ' nodate;
title 'Regression with Metric Cars Data: Part One';

data auto;
  infile 'mcars4.data' firstobs=2 ; /* Skipping the header on line 1 */
  input id country $ Lper100k 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;
/* Product terms to check interactions */
  Wc1 = weight*c1; Wc2 = weight*c2;
  Lc1 = length*c1; Lc2 = length*c2;
  label country = 'Country of Origin'
  Lper100K = 'Fuel Consumption in L/100k'
  weight = 'Weight in kg'
  length = 'Length in cm'
  c1 = 'USminusJapan'
  c2 = 'EUminusJapan' ;

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

proc glm;
  title2 'One-factor ANOVA ignoring covariates';
  class country;
  model Lper100k = country;
  means country;
  lsmeans country / pdiff adjust=bon;

proc reg simple; /* Simple says display means etc. */
  title2 'With both Weight and Length: Parallel Planes Model';
  model Lper100k = 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 glm;
  title2 'Parallel planes with proc glm and LSMEANS';
  class country;
  model Lper100k = weight length country;
  lsmeans country / pdiff adjust=bon;

proc reg;
  title2 'Test Parallel Planes with Proc Reg';
  model Lper100k = weight length c1 c2 Wc1 Wc2 Lc1 Lc2;
  Parallel: test Wc1=Wc2=Lc1=Lc2=0;
```

```

proc glm;
  title2 'Test Parallel Planes with Proc GLM';
  class country;
  model Lper100k = weight length country
              weight*country length*country;
  contrast 'Parallel' weight*country 1 -1 0,
              weight*country 0 1 -1,
              length*country 1 -1 0,
              length*country 0 1 -1 / E;
  /* Over-parameterized model is tricky to deal with. */

proc reg noprint; /* No printed output this time
                  Just generate residuals */
  model Lper100k = weight length c1 c2;
  output out=auto2 rstudent=delstud;
  /* New SAS dataset auto2 has original vars plus deleted
     studentized residuals. */

proc univariate normal plot;
  var delstud;

proc plot;
  plot delstud * (weight length); /* Y then X */
  plot delstud * weight = country;

```

Regression with Metric Cars Data: Part One
 One-factor ANOVA ignoring covariates

2

The GLM Procedure

Class Level Information

Class	Levels	Values
country	3	Europ Japan US
Number of Observations Read		100
Number of Observations Used		100

Regression with Metric Cars Data: Part One 3
 One-factor ANOVA ignoring covaiates

The GLM Procedure

Dependent Variable: Lper100k Fuel Consumption in L/100k

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	129.103708	64.551854	6.63	0.0020
Error	97	943.807892	9.729978		
Corrected Total	99	1072.911600			

R-Square	Coeff Var	Root MSE	Lper100k Mean
0.120330	25.40553	3.119291	12.27800

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	2	129.1037082	64.5518541	6.63	0.0020

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	2	129.1037082	64.5518541	6.63	0.0020

Regression with Metric Cars Data: Part One 4
 One-factor ANOVA ignoring covaiates

The GLM Procedure

Level of country	N	-----Lper100k----- Mean	Std Dev
Europ	14	10.1785714	3.61900508
Japan	13	10.6846154	2.36320621
US	73	12.9643836	3.13255176

Regression with Metric Cars Data: Part One 5
 One-factor ANOVA ignoring covaiates

The GLM Procedure

Least Squares Means

Adjustment for Multiple Comparisons: Bonferroni

country	Lper100k LSMEAN	LSMEAN Number
Europ	10.1785714	1
Japan	10.6846154	2
US	12.9643836	3

Least Squares Means for effect country
 Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Lper100k

i/j	1	2	3
1		1.0000	0.0086
2	1.0000		0.0511
3	0.0086	0.0511	

Regression with Metric Cars Data: Part One 6
 With both Weight and Length: Parallel Planes Model

The REG Procedure

Number of Observations Read 100
 Number of Observations Used 100

Descriptive Statistics

Variable	Sum	Mean	Uncorrected SS	Variance	Standard Deviation
Intercept	100.00000	1.00000	100.00000	0	0
weight	141321	1413.21000	212686149	131009	361.95176
length	484.92000	4.84920	2381.41200	0.30240	0.54991
c1	73.00000	0.73000	73.00000	0.19909	0.44620
c2	14.00000	0.14000	14.00000	0.12162	0.34874
Lper100k	1227.80000	12.27800	16148	10.83749	3.29203

Descriptive Statistics

Variable	Label
Intercept	Intercept
weight	Weight in kg
length	Length in cm
c1	USminusJapan
c2	EUminusJapan
Lper100k	Fuel Consumption in L/100k

Regression with Metric Cars Data: Part One 7
 With both Weight and Length: Parallel Planes Model

The REG Procedure

Model: MODEL1

Dependent Variable: Lper100k Fuel Consumption in L/100k

Number of Observations Read 100
 Number of Observations Used 100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	797.30574	199.32644	68.71	<.0001
Error	95	275.60586	2.90111		
Corrected Total	99	1072.91160			

Root MSE	1.70327	R-Square	0.7431
Dependent Mean	12.27800	Adj R-Sq	0.7323
Coeff Var	13.87250		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-5.28270	2.92605	-1.81
weight	Weight in kg	1	0.00546	0.00147	3.71
length	Length in cm	1	2.34597	0.98033	2.39
c1	USminusJapan	1	-1.99424	0.58499	-3.41
c2	EUminusJapan	1	-0.50652	0.66016	-0.77

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	0.0742
weight	Weight in kg	1	0.0004
length	Length in cm	1	0.0187
c1	USminusJapan	1	0.0010
c2	EUminusJapan	1	0.4448

```
> # Estimated fuel consumption for a Japanese car of average weight and length.
> -5.28270 + 0.00546*1413.21 + 2.34597*4.8492
[1] 13.80950
```

Regression with Metric Cars Data: Part One 8
 With both Weight and Length: Parallel Planes Model

The REG Procedure
 Model: MODEL1

Test country Results for Dependent Variable Lper100k

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	20.01754	6.90	0.0016
Denominator	95	2.90111		

Regression with Metric Cars Data: Part One 9
 With both Weight and Length: Parallel Planes Model

The REG Procedure
 Model: MODEL1

Test wgt_len Results for Dependent Variable Lper100k

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	334.10102	115.16	<.0001
Denominator	95	2.90111		

Regression with Metric Cars Data: Part One
 Parallel planes with proc glm and LSMEANS

10

The GLM Procedure

Class Level Information

Class	Levels	Values
country	3	Europ Japan US
Number of Observations Read		100
Number of Observations Used		100

Regression with Metric Cars Data: Part One
 Parallel planes with proc glm and LSMEANS

11

The GLM Procedure

Dependent Variable: Lper100k Fuel Consumption in L/100k

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	797.305741	199.326435	68.71	<.0001
Error	95	275.605859	2.901114		
Corrected Total	99	1072.911600			

R-Square	Coeff Var	Root MSE	Lper100k Mean
0.743123	13.87250	1.703266	12.27800

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight	1	746.6769100	746.6769100	257.38	<.0001
length	1	10.5937459	10.5937459	3.65	0.0590
country	2	40.0350853	20.0175427	6.90	0.0016

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight	1	39.86971491	39.86971491	13.74	0.0004
length	1	16.61365020	16.61365020	5.73	0.0187
country	2	40.03508534	20.01754267	6.90	0.0016

Regression with Metric Cars Data: Part One
 Parallel planes with proc glm and LSMEANS

12

The GLM Procedure
 Least Squares Means
 Adjustment for Multiple Comparisons: Bonferroni

country	Lper100k LSMEAN	LSMEAN Number
Europ	13.2981897	1
Japan	13.8047068	2
US	11.8104679	3

Least Squares Means for effect country
 Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Lper100k

i/j	1	2	3
1		1.0000	0.0338
2	1.0000		0.0029
3	0.0338	0.0029	

Regression with Metric Cars Data: Part One
 Test Parallel Planes with Proc Reg

13

The REG Procedure
 Model: MODEL1
 Dependent Variable: Lper100k Fuel Consumption in L/100k

Number of Observations Read 100
 Number of Observations Used 100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	817.63884	102.20486	36.43	<.0001
Error	91	255.27276	2.80520		
Corrected Total	99	1072.91160			

Root MSE 1.67487 R-Square 0.7621
 Dependent Mean 12.27800 Adj R-Sq 0.7412
 Coeff Var 13.64124

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-10.07335	16.83584	-0.60
weight	Weight in kg	1	0.01692	0.01024	1.65
length	Length in cm	1	0.64989	5.94151	0.11
c1	USminusJapan	1	5.01091	17.14931	0.29
c2	EUminusJapan	1	-1.83239	18.58103	-0.10
Wc1		1	-0.01123	0.01036	-1.08
Wc2		1	-0.01035	0.01171	-0.88
Lc1		1	1.18552	6.03345	0.20
Lc2		1	2.83971	6.70494	0.42

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	0.5511
weight	Weight in kg	1	0.1019
length	Length in cm	1	0.9131
c1	USminusJapan	1	0.7708
c2	EUminusJapan	1	0.9217
Wc1		1	0.2810
Wc2		1	0.3793
Lc1		1	0.8447
Lc2		1	0.6729

Regression with Metric Cars Data: Part One 14
 Test Parallel Planes with Proc Reg

The REG Procedure
 Model: MODEL1

Test Parallel Results for Dependent Variable Lper100k

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	5.08328	1.81	0.1333
Denominator	91	2.80520		

Regression with Metric Cars Data: Part One 15
 Test Parallel Planes with Proc GLM

The GLM Procedure

Class Level Information

Class	Levels	Values
country	3	Europ Japan US

Number of Observations Read	100
Number of Observations Used	100

Regression with Metric Cars Data: Part One 16

Test Parallel Planes with Proc GLM

The GLM Procedure

Coefficients for Contrast Parallel

		Row 1	Row 2	Row 3	Row 4
Intercept		0	0	0	0
weight		0	0	0	0
length		0	0	0	0
country	Europ	0	0	0	0
country	Japan	0	0	0	0
country	US	0	0	0	0
weight*country	Europ	1	0	0	0
weight*country	Japan	-1	1	0	0
weight*country	US	0	-1	0	0
length*country	Europ	0	0	1	0
length*country	Japan	0	0	-1	1
length*country	US	0	0	0	-1

Regression with Metric Cars Data: Part One
Test Parallel Planes with Proc GLM

17

The GLM Procedure

Dependent Variable: Lper100k Fuel Consumption in L/100k

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	817.638842	102.204855	36.43	<.0001
Error	91	255.272758	2.805195		
Corrected Total	99	1072.911600			

R-Square Coeff Var Root MSE Lper100k Mean
0.762075 13.64124 1.674872 12.27800

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight	1	746.6769100	746.6769100	266.18	<.0001
length	1	10.5937459	10.5937459	3.78	0.0551
country	2	40.0350853	20.0175427	7.14	0.0013
weight*country	2	19.4775459	9.7387729	3.47	0.0352
length*country	2	0.8555546	0.4277773	0.15	0.8588

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight	1	17.12279082	17.12279082	6.10	0.0154
length	1	2.17433967	2.17433967	0.78	0.3810
country	2	1.96673461	0.98336731	0.35	0.7052
weight*country	2	3.33166710	1.66583355	0.59	0.5543
length*country	2	0.85555461	0.42777730	0.15	0.8588

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Parallel	4	20.33310050	5.08327512	1.81	0.1333

Regression with Metric Cars Data: Part One 18
 Test Parallel Planes with Proc GLM

The UNIVARIATE Procedure
 Variable: delstud (Studentized Residual without Current Obs)

Moments

N	100	Sum Weights	100
Mean	-0.0042162	Sum Observations	-0.4216169
Std Deviation	1.02669295	Variance	1.05409842
Skewness	-0.663718	Kurtosis	0.45971263
Uncorrected SS	104.357521	Corrected SS	104.355743
Coeff Variation	-24351.324	Std Error Mean	0.1026693

Basic Statistical Measures

Location		Variability	
Mean	-0.00422	Std Deviation	1.02669
Median	0.00873	Variance	1.05410
Mode	-2.59615	Range	4.62121
		Interquartile Range	1.34603

Note: The mode displayed is the smallest of 8 modes with a count of 4.

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----
Student's t	t -0.04107	Pr > t 0.9673
Sign	M 1	Pr >= M 0.9204
Signed Rank	S 143	Pr >= S 0.6253

Tests for Normality

Test	--Statistic---	-----p Value-----
Shapiro-Wilk	W 0.956707	Pr < W 0.0024
Kolmogorov-Smirnov	D 0.103727	Pr > D <0.0100
Cramer-von Mises	W-Sq 0.150833	Pr > W-Sq 0.0232
Anderson-Darling	A-Sq 1.067598	Pr > A-Sq 0.0084

Quantiles (Definition 5)

Quantile	Estimate
100% Max	1.84071359
99%	1.80067907
95%	1.52813039
90%	1.34985827
75% Q3	0.78864798
50% Median	0.00873159

25% Q1 -0.55737986
 10% -1.51338017
 5% -2.14163966

Regression with Metric Cars Data: Part One
 Test Parallel Planes with Proc GLM

19

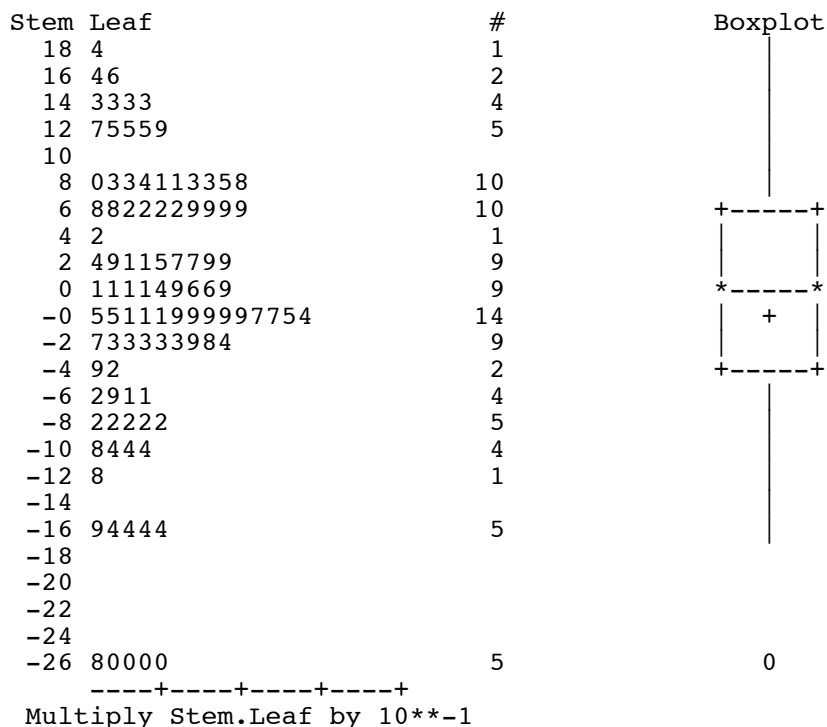
The UNIVARIATE Procedure
 Variable: delstud (Studentized Residual without Current Obs)

Quantiles (Definition 5)

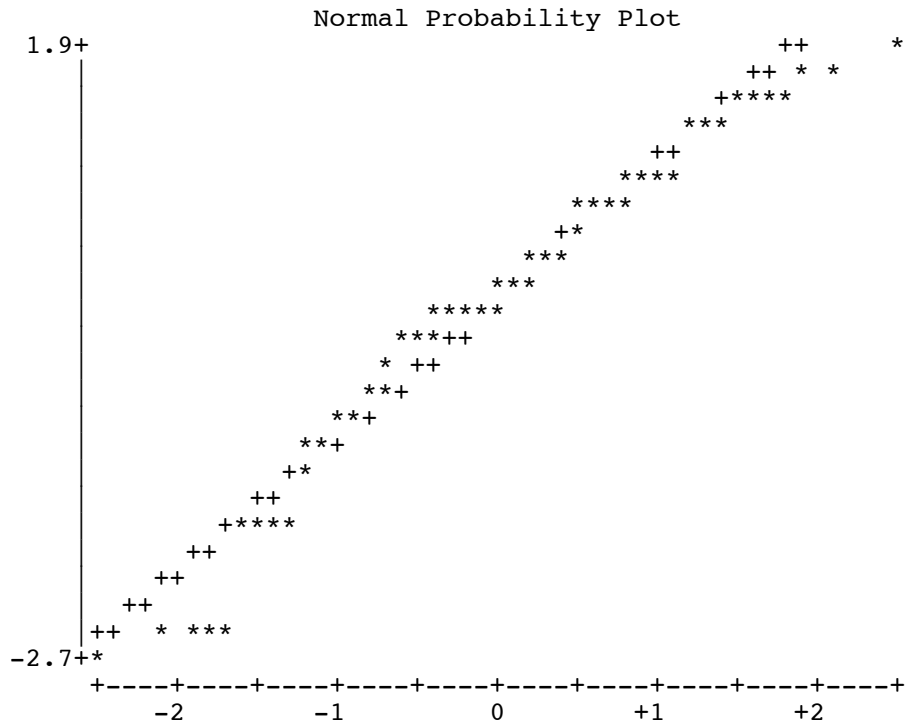
Quantile	Estimate
1%	-2.68832211
0% Min	-2.78049372

Extreme Observations

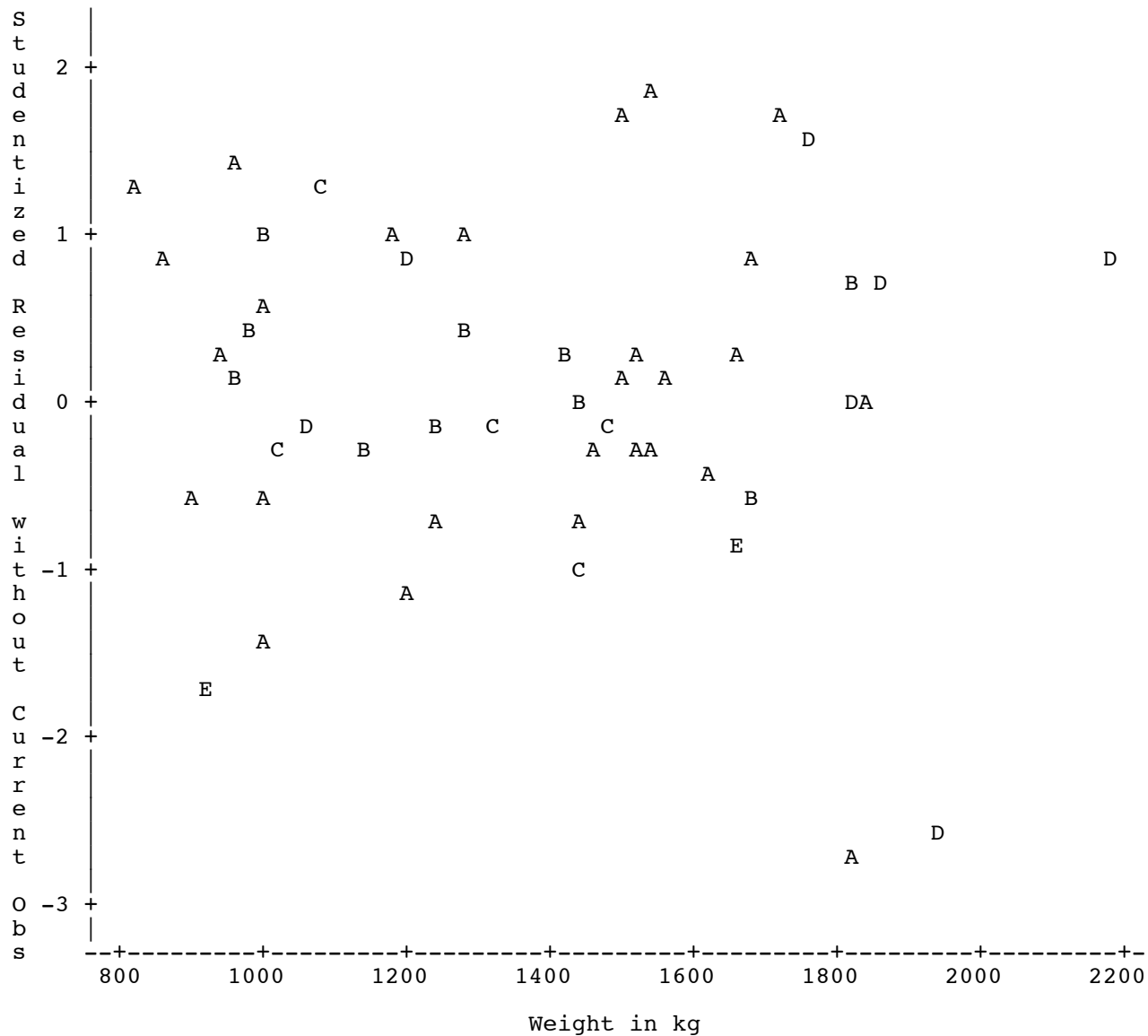
-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-2.78049	44	1.52813	86
-2.59615	80	1.52813	92
-2.59615	73	1.73885	43
-2.59615	65	1.76064	75
-2.59615	13	1.84071	9



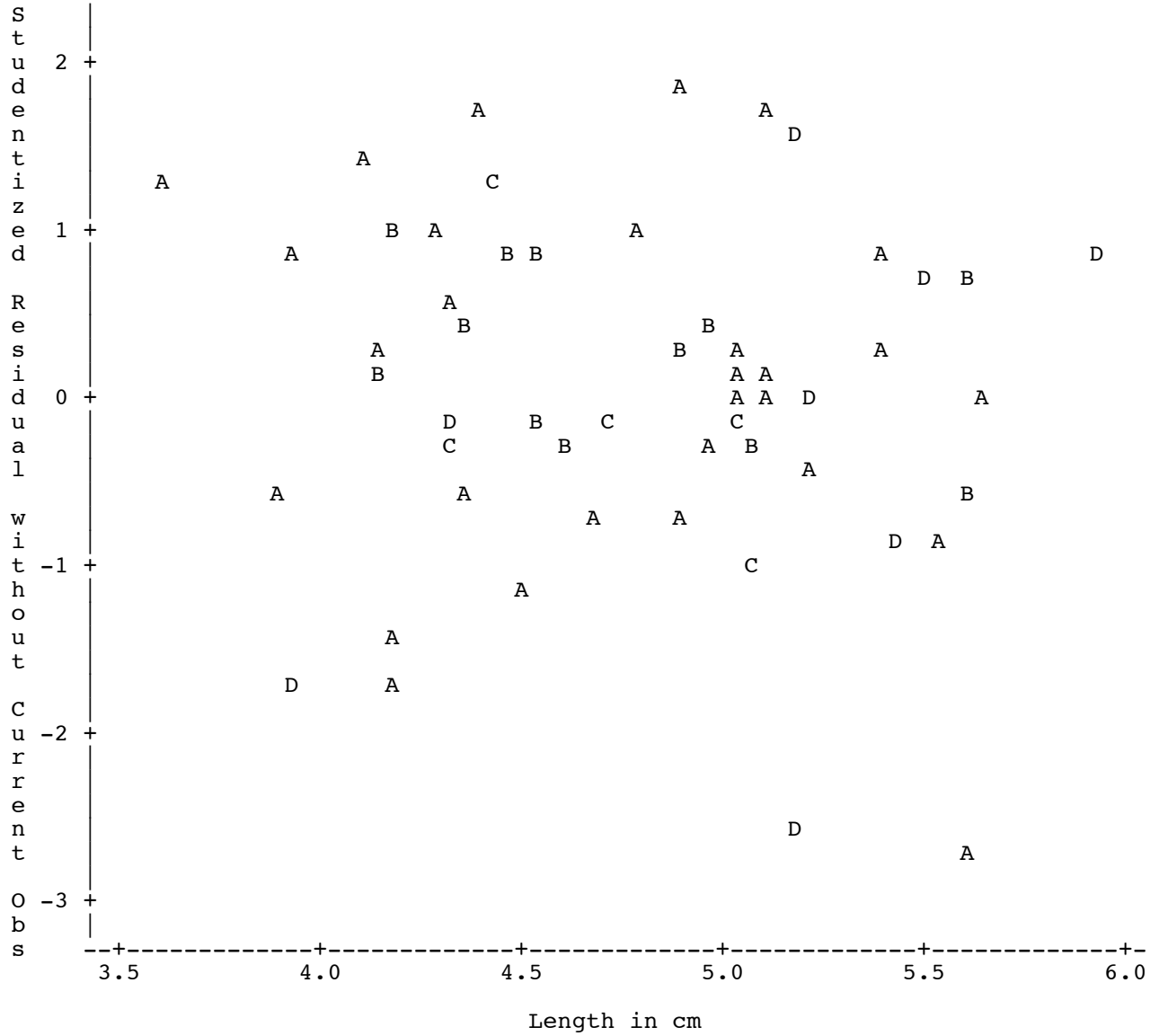
The UNIVARIATE Procedure
Variable: delstud (Studentized Residual without Current Obs)



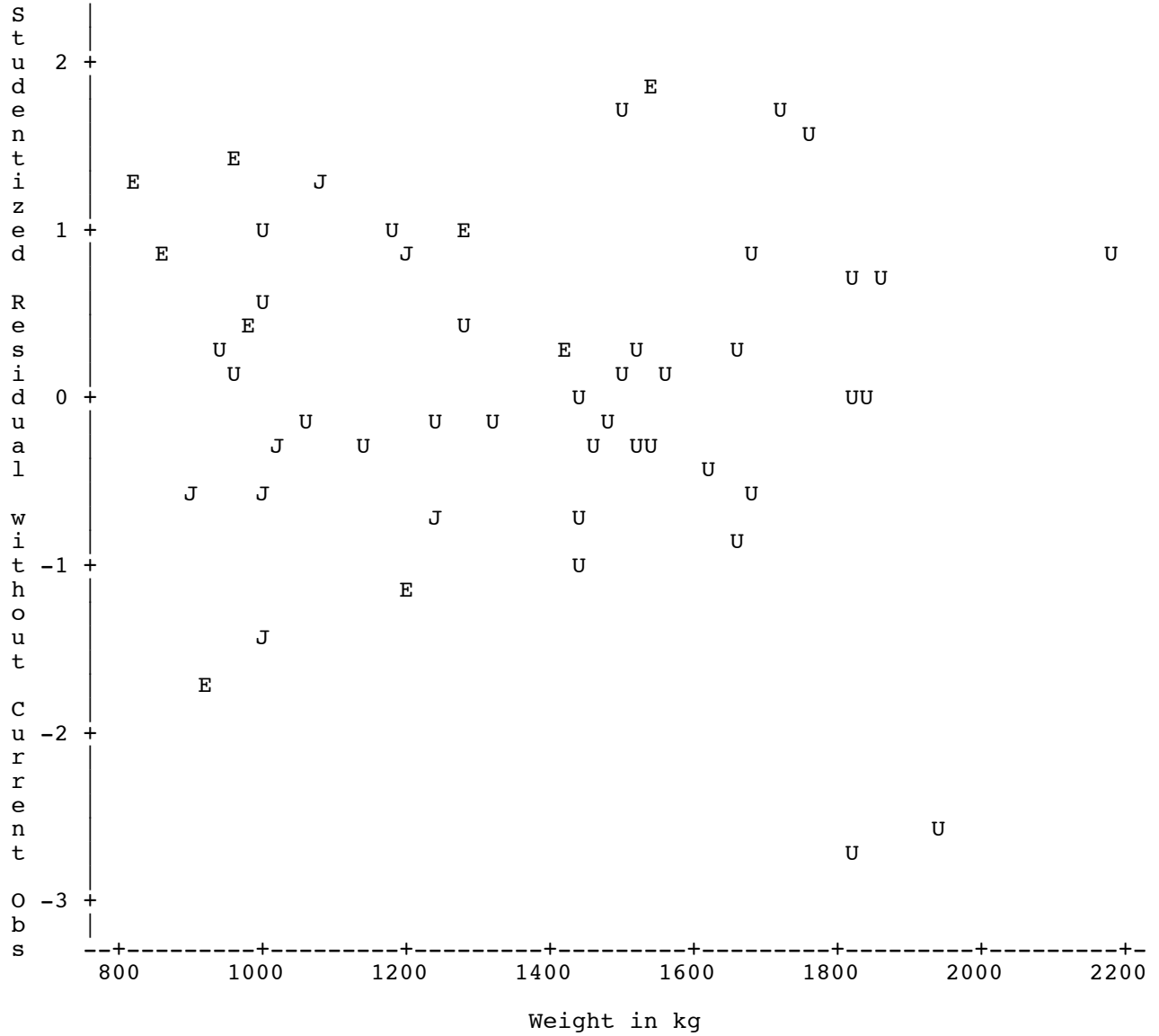
Plot of delstud*weight. Legend: A = 1 obs, B = 2 obs, etc.



Plot of delstud*length. Legend: A = 1 obs, B = 2 obs, etc.



Plot of delstud*weight. Symbol is value of country.



NOTE: 49 obs hidden.