```
1          ********************************************************;
2          *** Brain size of selected mammals.                  ***;
3          *** Biologists are interested in the effect of brain  ***;
4          *** size on variable slike gestation and litter size  ***;
5          ********************************************************;
6
7          dm'log;clear;output;clear';
8          options nodate nocenter nonumber ps=512 ls=99 nolabel;
9          ODS HTML style=minimal rs=none
9        ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\BrainSize01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\BrainSize01.html
10
11         Title1 'Chapter 9 : Relationship of brain size to selected variables';
12         filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0902.csv';
13
14         data BrainSize; length species $ 24; infile input1 missover DSD dlm="," firstobs=2;
15            input SPECIES $ BRAIN BODY GESTATION LITTER;
16              label Species = 'Mammal species'
17                    Brain = 'Brain size (gms)'
18                    Body = 'Body size (gms)'
19                    Gestation = 'Gestation period (days)'
20                    Gestation = 'Litter size';
21            LBrain = log(brain);
22            LBody = log(body);
23         datalines;
NOTE: The infile INPUT1 is:
      File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0902.csv,
      RECFM=V,LRECL=256
NOTE: 96 records were read from the infile INPUT1.
      The minimum record length was 19.
      The maximum record length was 36.
NOTE: The data set WORK.BRAINSIZE has 96 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
24         run;
25         proc sort data=brainsize; by brain; run;
NOTE: There were 96 observations read from the data set WORK.BRAINSIZE.
NOTE: The data set WORK.BRAINSIZE has 96 observations and 7 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds
26         PROC PRINT DATA=BrainSize; TITLE2 'Raw data Listing'; RUN;
NOTE: There were 96 observations read from the data set WORK.BRAINSIZE.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.20 seconds
      cpu time            0.10 seconds
```

Chapter 9 : Relationship of brain size to selected variables
Raw data Listing

Obs	species	BRAIN	BODY	GESTATION	LITTER	LBrain	LBody
1	House mouse	0.45	0.02	19	5.0	-0.79851	-3.72970
2	Deer mouse III	0.52	0.02	24	5.0	-0.65393	-4.07454
3	Deer mouse II	0.63	0.03	23	5.0	-0.46204	-3.64966
4	Hampster I	0.67	0.04	21	4.6	-0.40048	-3.32424
5	Deer mouse I	0.68	0.03	23	3.7	-0.38566	-3.61192
6	Deer mouse IV	0.69	0.02	24	5.0	-0.37106	-3.72970
7	Rat I	0.72	0.05	23	7.3	-0.32850	-2.99573
8	Pygmy gerbil	1.04	0.07	21	4.0	0.03922	-2.73337
9	Hampster II	1.12	0.13	16	6.3	0.11333	-2.04022
10	Elephant shrew I	1.14	0.05	51	1.5	0.13103	-3.01593
11	Hopping mouse	1.18	0.15	27	5.6	0.16551	-1.89712
12	Elephant shrew II	1.37	0.06	46	1.5	0.31481	-2.74887
13	Flying squirrel	1.89	0.05	40	3.1	0.63658	-2.95651
14	Rat II	2.38	0.34	21	8.0	0.86710	-1.07881
15	Tree shrew	3.15	0.15	46	3.0	1.14740	-1.89712
16	Hedgehog	3.50	0.93	34	4.6	1.25276	-0.07257
17	Guinea pig	4.28	0.97	67	2.6	1.45395	-0.03046
18	Chinchilla	5.25	0.43	110	2.0	1.65823	-0.84397
19	Tree squirrel	6.23	0.33	38	3.0	1.82938	-1.10866

20	Gentle lemur	7.80	0.22	145	2.0	2.05412	-1.51413
21	Aardvark	9.60	2.20	31	5.0	2.26176	0.78846
22	Bush baby	9.90	0.70	135	1.0	2.29253	-0.35667
23	Acouchis	9.90	0.78	98	1.2	2.29253	-0.24846
24	Long-nose armadillo	12.00	3.70	120	4.0	2.48491	1.30833
25	Slow loris	12.80	1.20	90	1.2	2.54945	0.18232
26	Jack rabbit	13.30	2.90	41	2.5	2.58776	1.06471
27	Quokka	17.50	3.50	26	1.0	2.86220	1.25276
28	Agoutis	20.30	2.80	104	1.3	3.01062	1.02962
29	Hyrax	20.50	3.80	225	2.4	3.02042	1.33500
30	Lemur	22.00	2.10	135	1.0	3.09104	0.74194
31	Nutria	23.00	5.00	132	5.5	3.13549	1.60944
32	Porcupine III	24.00	6.60	113	1.0	3.17805	1.88707
33	Domestic cat	28.40	2.50	63	4.0	3.34639	0.91629
34	Bat-eared fox	28.50	3.20	65	4.0	3.34990	1.16315
35	Linkajou	31.20	2.00	77	1.1	3.44042	0.69315
36	Porcupine I	37.00	11.00	112	1.2	3.61092	2.39790
37	Porcupine II	37.00	14.00	112	1.2	3.61092	2.63906
38	Gray fox	37.30	3.80	63	3.7	3.61899	1.33500
39	Canadian beaver	40.00	20.00	128	2.9	3.68888	2.99573
40	Raccoon	41.60	5.30	63	3.5	3.72810	1.66771
41	Beaver	45.00	25.00	128	4.0	3.80666	3.21888
42	Red fox	48.00	6.00	52	4.0	3.87120	1.79176
43	Badger	53.00	6.00	60	2.2	3.97029	1.79176
44	Howler monkey	54.00	7.70	139	1.0	3.98898	2.04122
45	Leaf monkey	65.50	5.80	168	1.0	4.18205	1.75786
46	Vervet guenon	67.00	4.60	195	1.0	4.20469	1.52606
47	Dog	70.20	8.50	63	4.0	4.25135	2.14007
48	Ring-tail monkey	73.00	3.70	180	1.0	4.29046	1.30833
49	Lynx	75.00	12.00	60	2.5	4.31749	2.48491
50	Capybara	76.00	30.00	123	3.0	4.33073	3.40120
51	Rhesus monkey I	84.60	6.00	175	1.0	4.43793	1.79176
52	Duikers	93.00	13.00	120	1.0	4.53260	2.56495
53	White-handed gibbon	102.00	5.50	210	1.0	4.62497	1.70475
54	Domestic goat	106.00	30.00	151	2.0	4.66344	3.40120
55	Rhesus monkey II	107.00	8.70	165	1.1	4.67283	2.16332
56	Spider monkey II	109.00	7.70	140	1.0	4.69135	2.04122
57	Spider monkey I	114.00	9.10	140	1.0	4.73620	2.20827
58	Barking deer	124.00	16.00	183	1.1	4.82028	2.77259
59	Domestic sheep	125.00	49.00	150	2.4	4.82831	3.89182
60	Leopard	157.00	46.00	92	2.5	5.05625	3.82864
61	Western baboon	179.00	32.00	180	1.0	5.18739	3.46574
62	Domestic pig	180.00	190.00	115	8.0	5.19296	5.24702
63	Hamadryas baboon	183.00	21.00	180	1.0	5.20949	3.04452
64	Wild boar	185.00	150.00	120	4.0	5.22036	5.01064
65	Vicuna	198.00	45.00	300	1.1	5.28827	3.80666
66	Black buck antilope	200.00	39.00	180	1.0	5.29832	3.66356
67	Barbary sheep	210.00	66.00	158	1.2	5.34711	4.18965
68	Axis deer	219.00	89.00	218	1.0	5.38907	4.48864
69	Fallow deer	223.00	80.00	240	1.0	5.40717	4.38203
70	Llama	225.00	93.00	330	1.0	5.41610	4.53260
71	Tapir	250.00	230.00	390	1.0	5.52146	5.43808
72	Lion	260.00	180.00	108	3.0	5.56068	5.19296
73	Pygmy hippopotamus	260.00	150.00	205	1.0	5.56068	5.01064
74	Caribou	288.00	110.00	225	1.0	5.66296	4.70048
75	Tiger	302.00	210.00	104	3.0	5.71043	5.34711
76	Yak	334.00	250.00	255	1.0	5.81114	5.52146
77	Orangutan	343.00	37.00	270	1.0	5.83773	3.61092
78	Fur seal	355.00	250.00	254	1.0	5.87212	5.52146
79	Chimpanzee	360.00	45.00	230	1.0	5.88610	3.80666
80	Sea lion	363.00	100.00	343	1.0	5.89440	4.60517
81	Elk	365.00	120.00	235	1.0	5.89990	4.78749
82	Sambar	383.00	120.00	246	1.1	5.94803	4.78749
83	Grizzly bear	400.00	250.00	219	2.3	5.99146	5.52146
84	Gorilla	406.00	140.00	265	1.0	6.00635	4.94164
85	Red deer	435.00	200.00	255	1.0	6.07535	5.29832
86	Harp seal	442.00	110.00	240	1.0	6.09131	4.70048
87	Cattle	456.00	520.00	280	1.0	6.12249	6.25383
88	Eland	480.00	560.00	255	1.0	6.17379	6.32794
89	Beaked whale	500.00	250.00	240	1.8	6.21461	5.52146
90	Porpoise	537.00	56.00	270	1.0	6.28600	4.02535
91	Weddell seal	550.00	400.00	310	1.0	6.30992	5.99146
92	Hippopotamus	590.00	1400.00	240	1.0	6.38012	7.24423
93	Horse	712.00	480.00	330	1.0	6.56808	6.17379
94	Human being	1300.00	65.00	270	1.0	7.17012	4.17439
95	Dolphin	1600.00	160.00	360	1.0	7.37776	5.07517
96	African elephant	4480.00	2800.00	655	1.0	8.40738	7.93737

From the plots note the following:

```
plot Brain * Gestation = Species;
```

Nonhomogeneous variance -

```
plot Brain * Litter = Species;
```

Nonhomogeneous variance -

```
plot Brain * Body = Species;
```

Much scatter and probable nonhomogeneous variance

```
plot LBrain * LBody = Species;
```

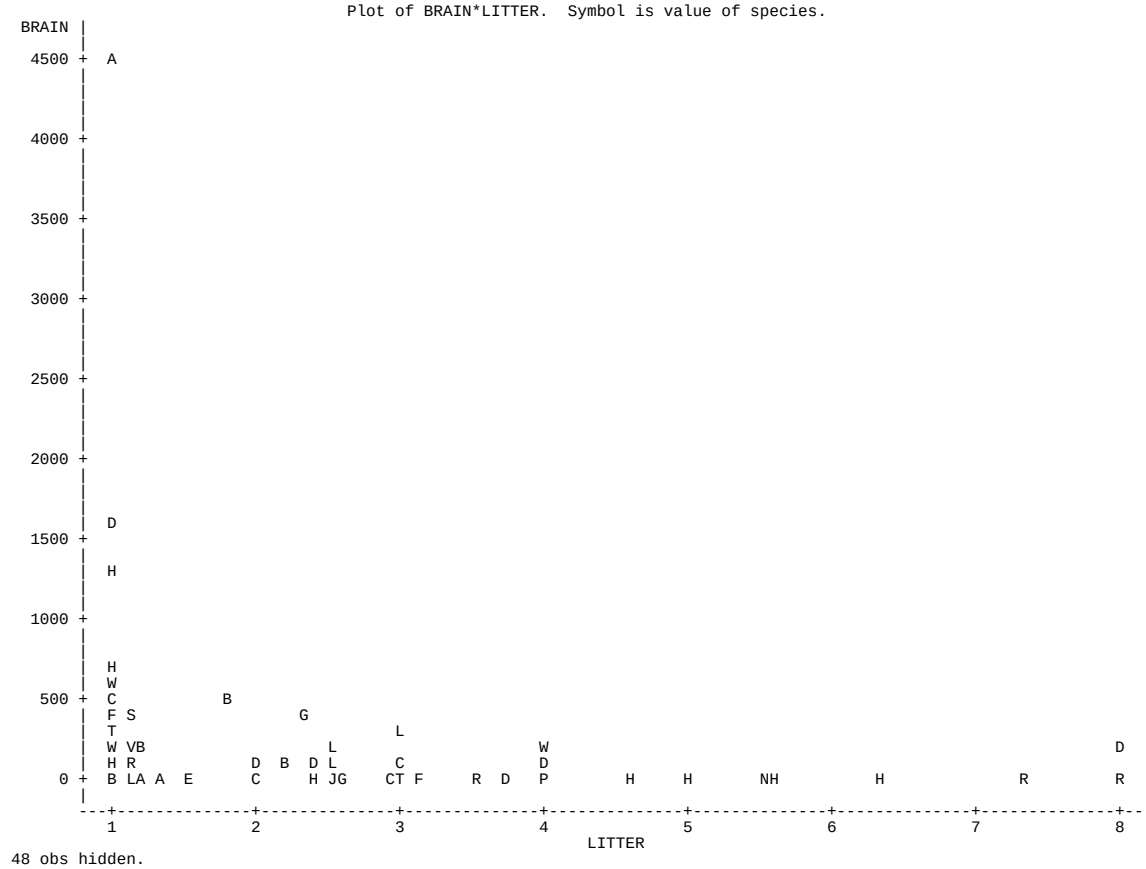
Looks pretty good

```
28         options ps=65 ls=132;
29         proc plot data=BrainSize;  TITLE2 'Plot of the raw data';
30            plot Brain * Gestation = Species;
31            plot Brain * Litter = Species;
32            plot Brain * Body = Species;
33            plot LBrain * LBody = Species;
34         RUN;
34       !      OPTIONS PS=256;
35
NOTE: There were 96 observations read from the data set WORK.BRAINSIZE.
NOTE: The PROCEDURE PLOT printed pages 2-5.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.13 seconds
      cpu time            0.01 seconds
```

```
Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data

                                Plot of BRAIN*GESTATION.  Symbol is value of species.

      BRAIN |
            |
       4500 +                                                                                                A
            |
            |
            |
            |
       4000 +
            |
            |
            |
            |
       3500 +
            |
            |
            |
            |
       3000 +
            |
            |
            |
            |
       2500 +
            |
            |
            |
            |
       2000 +
            |
            |
            |                                                  D
       1500 +
            |                                      H
            |
            |
       1000 +
            |
            |                                                H
            |                                    H           W
        500 +                                   B E  PC
            |                                 G CHSF G          S
            |                 L             P  C   Y  O                T
            |             L  DW     B  W     A  F        V    L
            |          B       C  H D LRR V  W
          0 +    HHDHFEEDG L SAACLCB G           H
            |
            ---+-------------+-------------+-------------+-------------+-------------+-------------+-------------+-------------+--
               0             100           200           300           400           500           600           700
                                                         GESTATION
NOTE: 36 obs hidden.
```

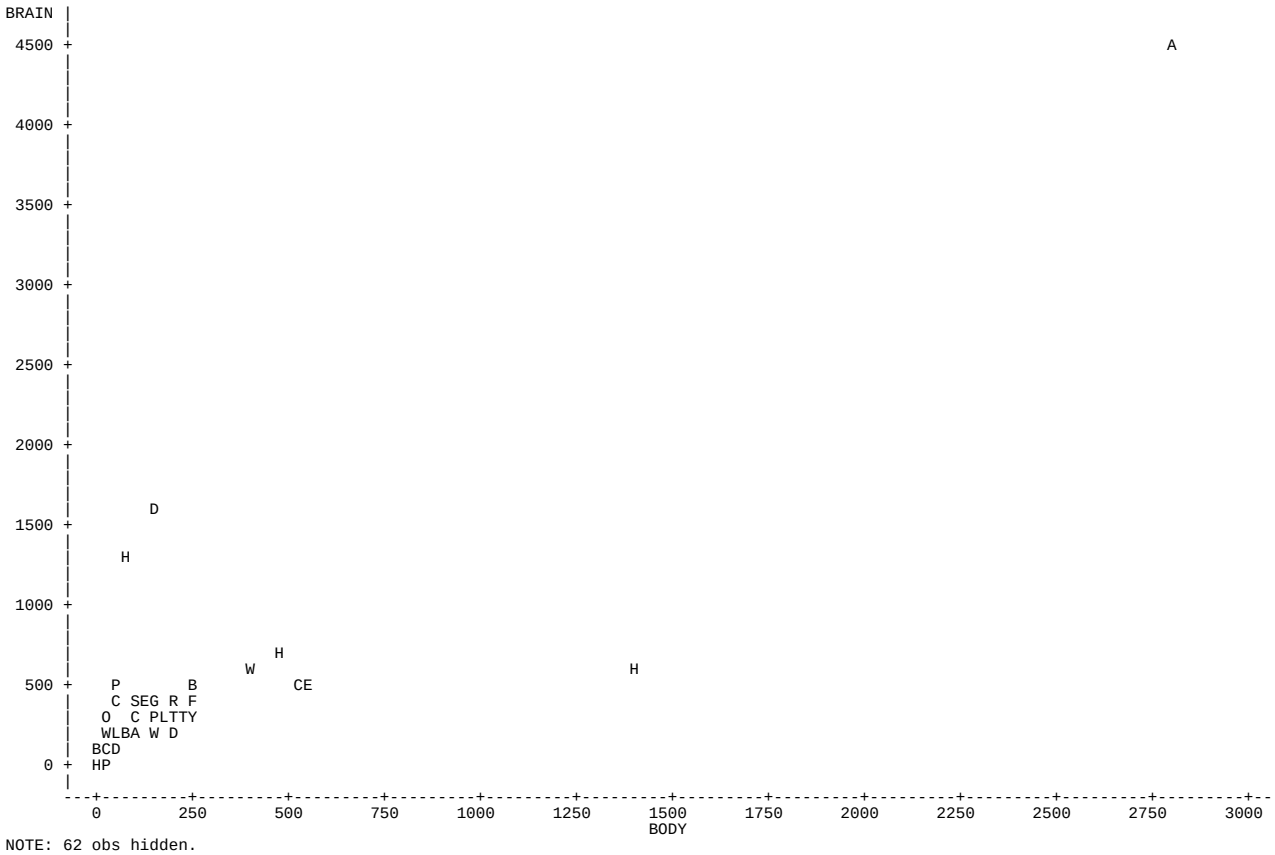
Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data

Plot of BRAIN*LITTER. Symbol is value of species.



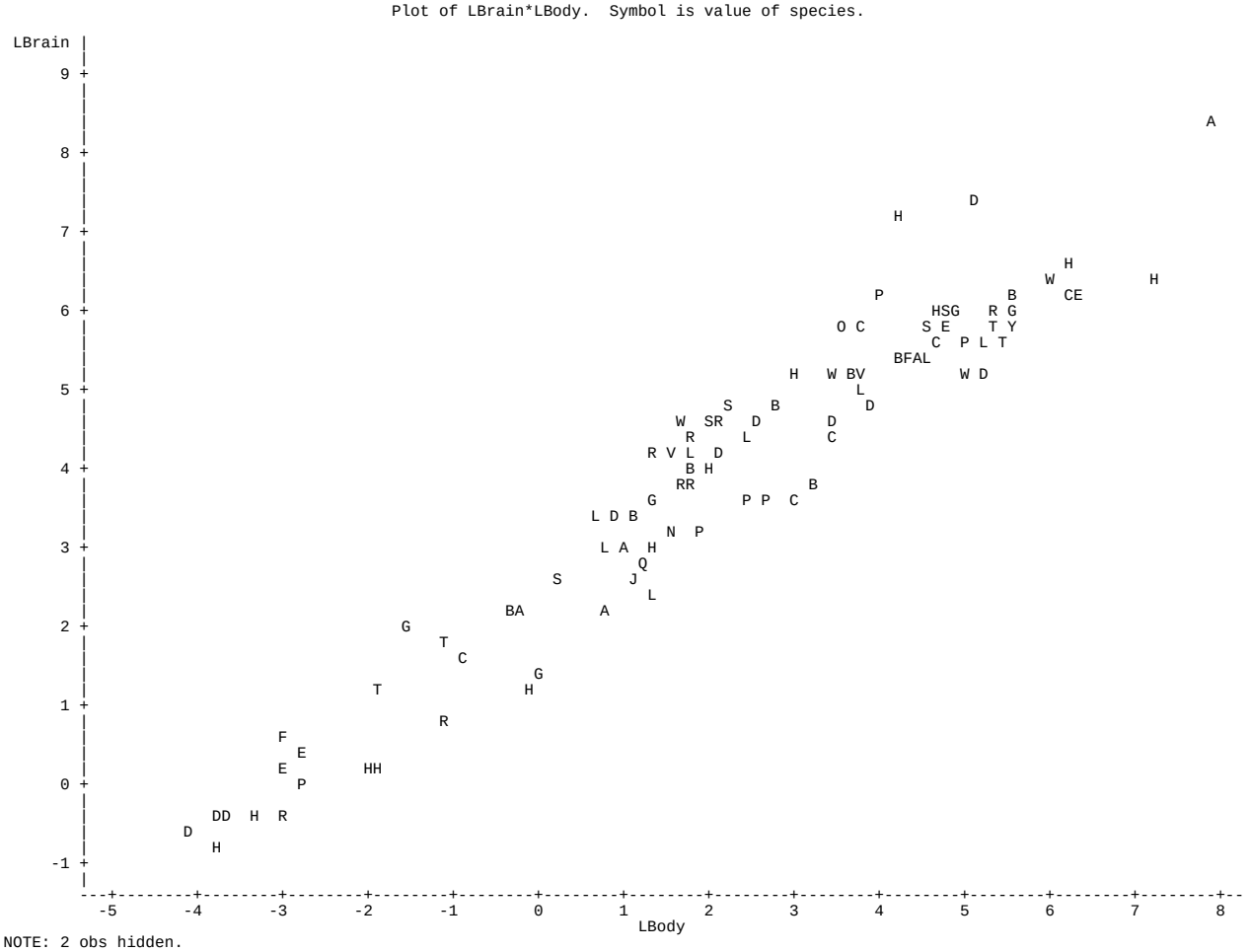
NOTE: 48 obs hidden.

Plot of BRAIN*BODY. Symbol is value of species.



NOTE: 62 obs hidden.

```
Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data
```




```
NOTE: 2 obs hidden.

36         proc corr data=BrainSize; var BRAIN BODY GESTATION LITTER LBrain LBody; run;
NOTE: The PROCEDURE CORR printed page 6.
NOTE: PROCEDURE CORR used (Total process time):
      real time           0.07 seconds
      cpu time            0.04 seconds
37
```

CORRELATIONS: A property called “multicolinearity” can cause considerable problems with multiple regression. It can be displayed as strong pairwise correlation among independent variables or as correlations among linear combinations of variables. Pairwise correlations among variable is a first check for possible multicolinearity. Correlations en excess of 0.9 are often indicators of problems.

```
Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data

The CORR Procedure

   6  Variables:    BRAIN     BODY      GESTATION LITTER    LBrain    LBody

                                SimpleStatistics
Variable            N          Mean       Std Dev           Sum       Minimum       Maximum
BRAIN              96     218.97698     506.96801         21022       0.45000          4480
BODY               96     108.32775     329.41140         10399       0.01700          2800
GESTATION          96     151.26042     108.06609         14521      16.00000     655.00000
LITTER             96       2.31042       1.74703     221.80000       1.00000       8.00000
LBrain             96       3.86458       2.17113     370.99922      -0.79851       8.40738
LBody              96       2.13050       2.91107     204.52768      -4.07454       7.93737
```

Pearson Correlation Coefficients, N = 96
Prob > |r| under H0: Rho=0

	BRAIN	BODY	GESTATION	LITTER	LBrain	LBody
BRAIN	1.00000	0.86403	0.72819	-0.25456	0.51980	0.48346
		<.0001	<.0001	0.0123	<.0001	<.0001
BODY	0.86403	1.00000	0.62467	-0.18274	0.42200	0.46914
	<.0001		<.0001	0.0747	<.0001	<.0001
GESTATION	0.72819	0.62467	1.00000	-0.60375	0.80626	0.77150
	<.0001	<.0001		<.0001	<.0001	<.0001
LITTER	-0.25456	-0.18274	-0.60375	1.00000	-0.61199	-0.51370
	0.0123	0.0747	<.0001		<.0001	<.0001
LBrain	0.51980	0.42200	0.80626	-0.61199	1.00000	0.96429
	<.0001	<.0001	<.0001	<.0001		<.0001
LBody	0.48346	0.46914	0.77150	-0.51370	0.96429	1.00000
	<.0001	<.0001	<.0001	<.0001	<.0001	

```
38         Title2 'Initial fit of the raw data';
39         PROC REG DATA=BrainSize lineprinter;
40            TITLE3 'Simple Regression with REG';
41            MODEL Brain = Body Gestation Litter / vif collin stb;
42            output out=next1 r=resid p=YHat;
43         RUN;
43       !      OPTIONS PS=45;
44
NOTE: The data set WORK.NEXT1 has 96 observations and 9 variables.
NOTE: The PROCEDURE REG printed page 7.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.14 seconds
      cpu time            0.07 seconds
```

Chapter 9 : Relationship of brain size to selected variables
Initial fit of the raw data
Simple Regression with REG

The REG Procedure
Model: MODEL1
Dependent Variable: BRAIN

Number of Observations Read	96
Number of Observations Used	96

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	19777226	6592409	130.73	<.0001
Error	92	4639348	50428		
Corrected Total	95	24416574			

Root MSE	224.56111	R-Square	0.8100
Dependent Mean	218.97698	Adj R-Sq	0.8038
Coeff Var	102.55010		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	1	-225.29213	83.05875	-2.71	0.0080	0	0
BODY	1	0.98588	0.09428	10.46	<.0001	0.64059	1.81717
GESTATION	1	1.80874	0.35445	5.10	<.0001	0.38555	2.76402
LITTER	1	27.64864	17.41429	1.59	0.1158	0.09528	1.74367

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variation: Intercept	BODY	GESTATION	LITTER
1	2.61903	1.00000	0.00986	0.02470	0.01429	0.01758
2	0.98271	1.63251	0.00543	0.28738	0.00774	0.06285
3	0.35660	2.71006	0.01262	0.38206	0.11314	0.19407
4	0.04165	7.92937	0.97209	0.30586	0.86482	0.72550

```
45         Title2 'Fit of the transformed data';
46         PROC REG DATA=BrainSize lineprinter;
47            TITLE3 'Simple Regression with REG';
48            MODEL LBrain = LBody Gestation Litter / vif collin stb;
49               test LBody = 1;
50            output out=next2 r=resid p=YHat;
51         RUN;

51       !      OPTIONS PS=45;
52
53         options ps=52 ls=111;
NOTE: The data set WORK.NEXT2 has 96 observations and 9 variables.
NOTE: The PROCEDURE REG printed pages 8-10.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.15 seconds
      cpu time            0.08 seconds
```

Multiple regression – Most of the interpretations and diagnostics developed from simple linear regression also apply to multiple regression. New diagnostics include the standardized partial regression coefficient, the variance inflation factor and the condition index.

Note the regression coefficients and their standard errors and the t-tests of the coefficients against an hypothesized value of zero. The interpretation is the same.

The R^2 is now called the coefficient of multiple determination. It's square root is the correlation between Y_i and the predicted value.

Note the test of H_0: $\beta_1 - 1$ done with the test statement. Multiple tests are also possible.

Confidence intervals on predicted means and individual points (not included here) work the same as with simple linear regression.

Standardized partial regression coefficient – These are regression coefficients from "standardized" data. ($Z = (Y_i - \bar{Y})/s$). These can be used to interpret the strength of the contribution of each variable to the regression.

Variance inflation factor and the condition index – These are used to evaluate multicolinearity. Variance inflation values should be about 1. They should not be over 10 and should not have a mean greater than 2. The last value (bottom) condition index number should not be over 40.

Residual plots in multiple regression are usually plotted on the predicted value. For this analysis the predicted value will be dominated by "LBody".

```
Chapter 9 : Relationship of brain size to selected variables
Fit of the transformed data
Simple Regression with REG

The REG Procedure
Model: MODEL1
Dependent Variable: LBrain

Number of Observations Read          96
Number of Observations Used          96
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	425.79663	141.93221	593.13	<.0001
Error	92	22.01497	0.23929		
Corrected Total	95	447.81160			

Root MSE	0.48918	R-Square	0.9508
Dependent Mean	3.86458	Adj R-Sq	0.9492
Coeff Var	12.65795		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	1	2.67415	0.16206	16.50	<.0001	0	0
LBody	1	0.61874	0.02722	22.73	<.0001	0.82962	2.49265
GESTATION	1	0.00171	0.00078916	2.16	0.0330	0.08501	2.88736
LITTER	1	-0.16714	0.03620	-4.62	<.0001	-0.13449	1.58777

Collinearity Diagnostics

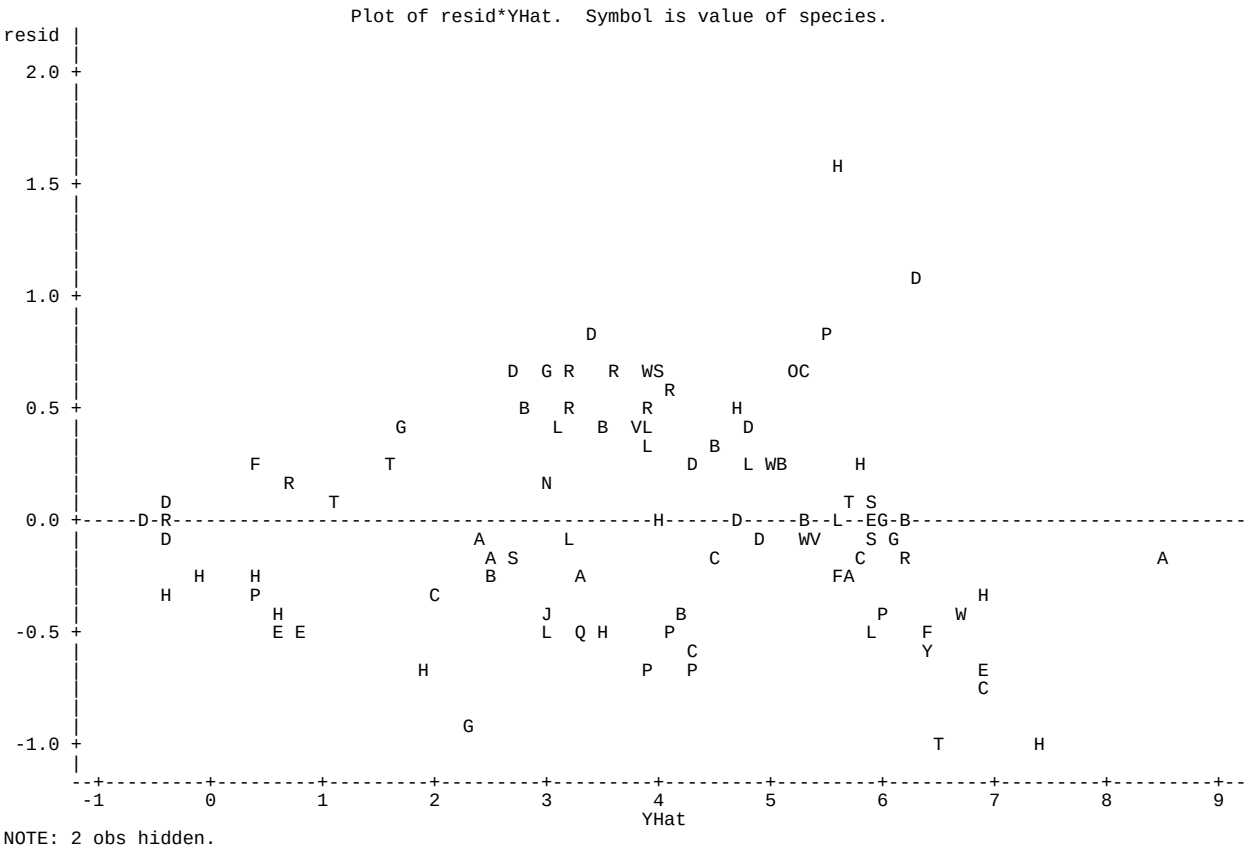
Number	Eigenvalue	Condition Index	Proportion of Variation: Intercept	LBody	GESTATION	LITTER
1	2.89305	1.00000	0.01018	0.01961	0.01194	0.01372
2	0.90920	1.78381	0.00653	0.09691	0.01174	0.12694
3	0.14524	4.46305	0.08496	0.74490	0.21299	0.29486
4	0.05251	7.42294	0.89833	0.13859	0.76333	0.56447

Test 1 Results for Dependent Variable LBrain

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	46.94597	196.19	<.0001
Denominator	92	0.23929		

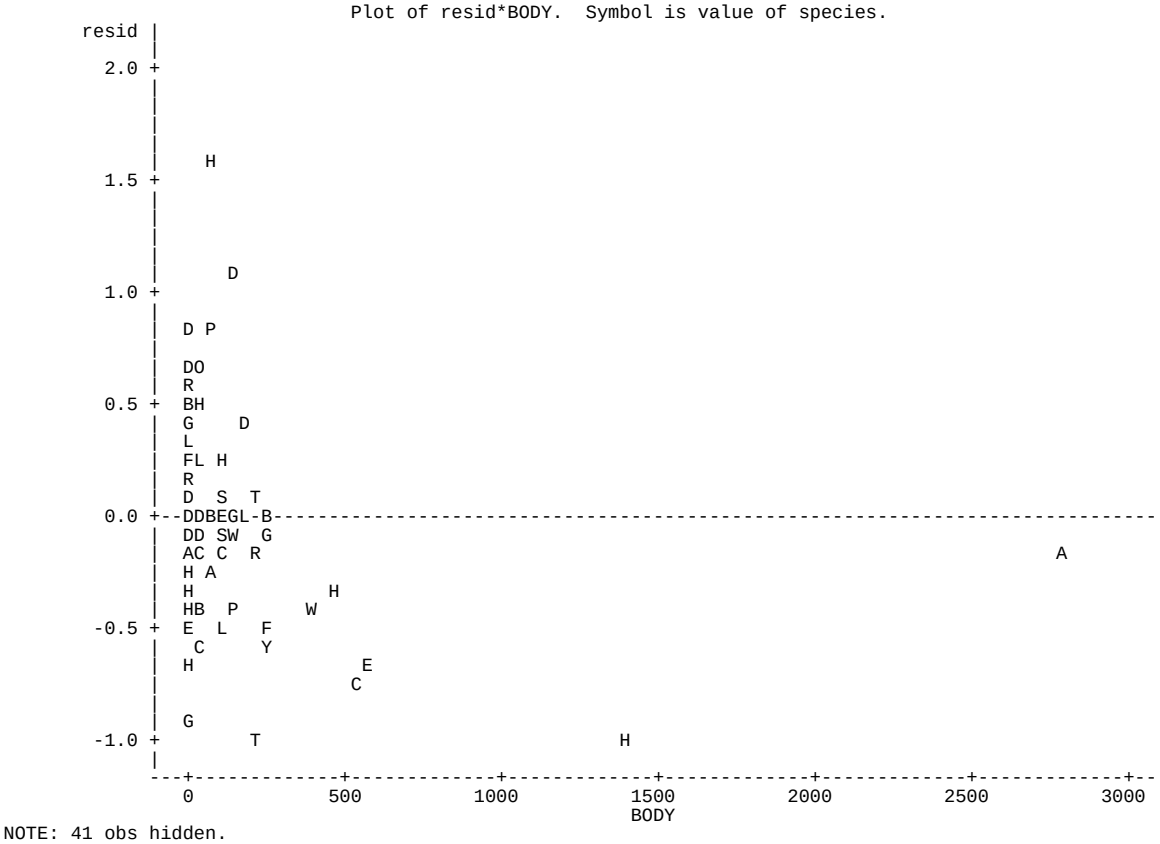
```
54         proc plot data=next2;  TITLE2 'Plot of the raw data';
55            plot resid * YHat = Species / vref=0;
56            plot resid * body = Species / vref=0;
57            plot resid * Gestation = Species / vref=0;
58            plot resid * Litter = Species / vref=0;
59         RUN;

59       !      OPTIONS PS=256;
60
NOTE: There were 96 observations read from the data set WORK.NEXT2.
NOTE: The PROCEDURE PLOT printed pages 11-14.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.08 seconds
      cpu time            0.03 seconds
```

Chapter 9 : Relationship of brain size to selected variables



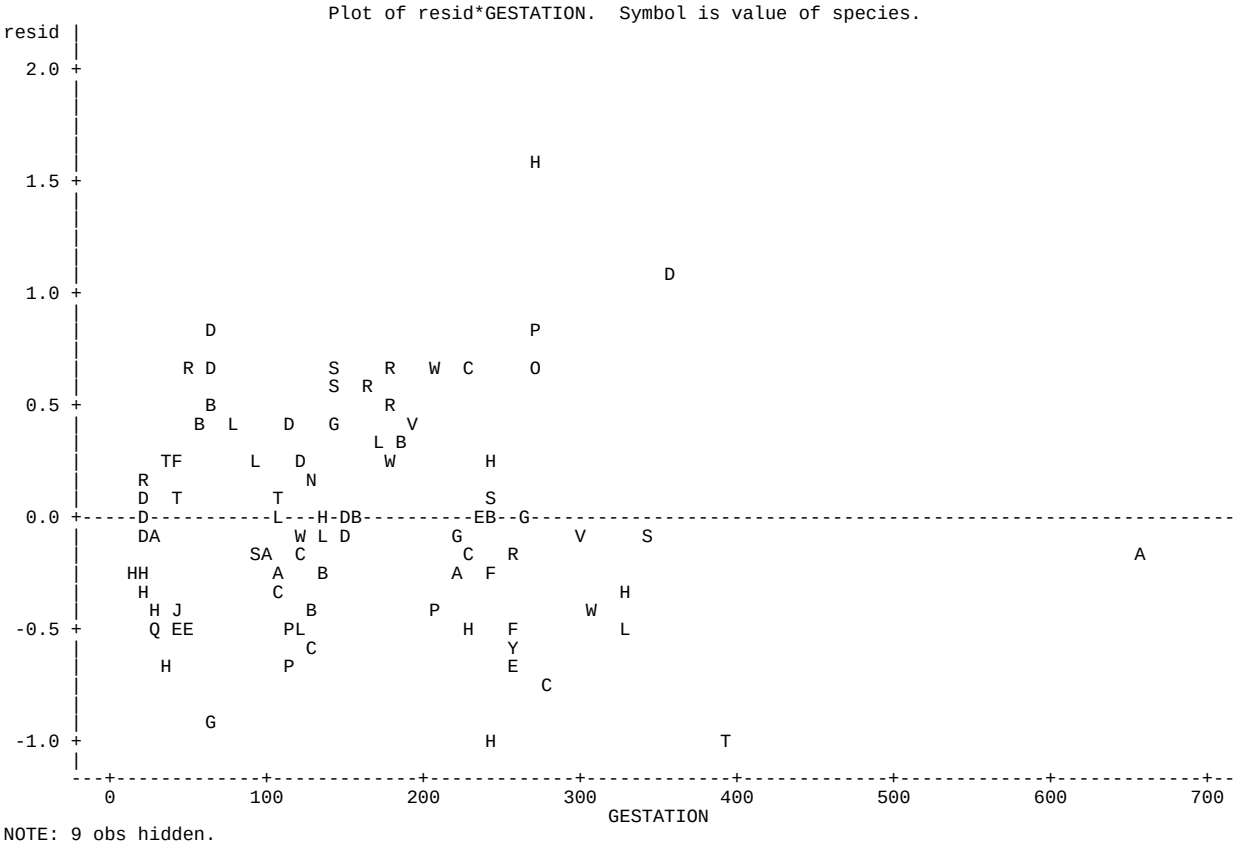
NOTE: 2 obs hidden.

Chapter 9 : Relationship of brain size to selected variables

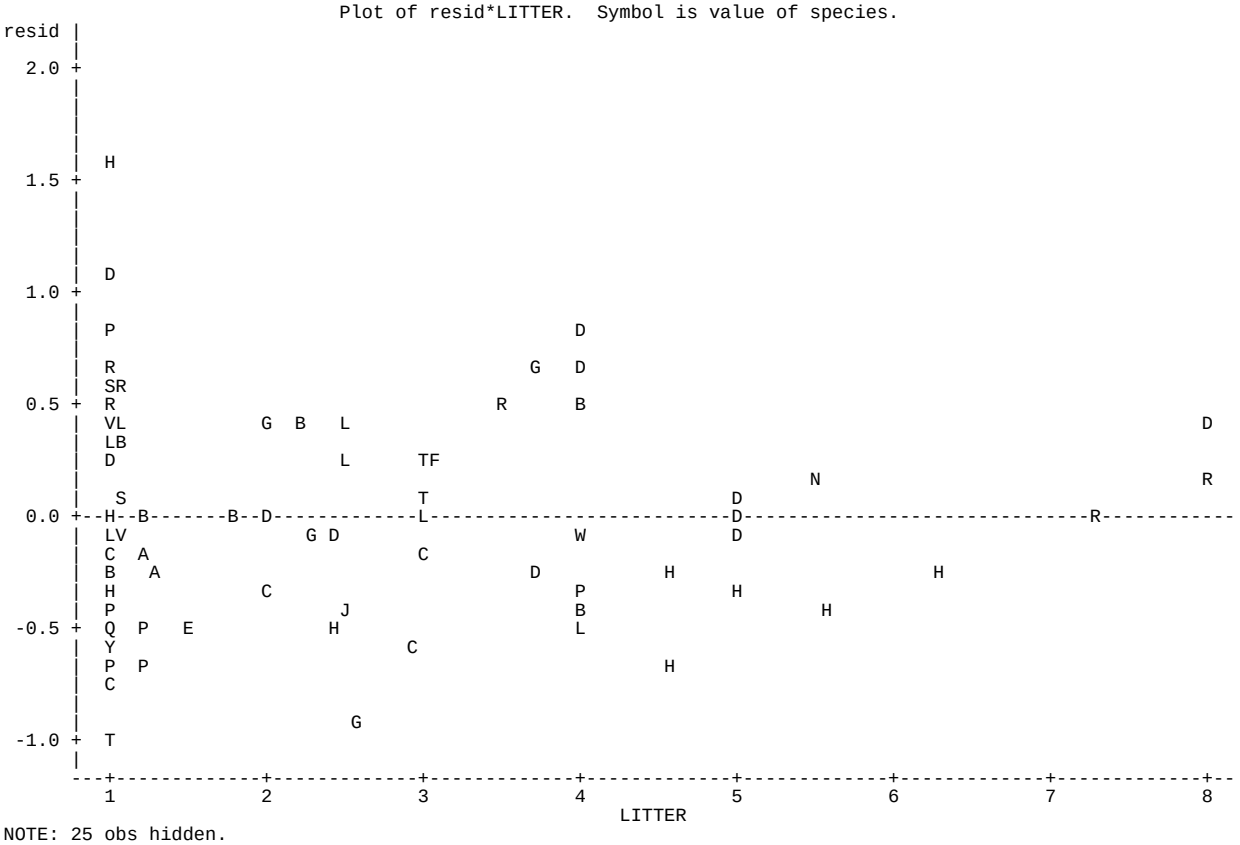


NOTE: 41 obs hidden.

Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data



NOTE: 9 obs hidden.



NOTE: 25 obs hidden.

```
61         PROC UNIVARIATE DATA=NEXT2 NORMAL PLOT; VAR resid;
62         RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 14.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.09 seconds
      cpu time            0.03 seconds
63         ODS HTML close;
64         quit;


Chapter 9 : Relationship of brain size to selected variables
Plot of the raw data

The UNIVARIATE Procedure
Variable:  resid

Moments
N                          96    Sum Weights                96
Mean                        0    Sum Observations            0
Std Deviation      0.48139016    Variance           0.23173649
Skewness           0.40284286    Kurtosis            0.3355239
Uncorrected SS     22.0149663    Corrected SS       22.0149663
Coeff Variation             .    Std Error Mean     0.04913168

Basic Statistical Measures
    Location                     Variability
Mean      0.00000     Std Deviation            0.48139
Median   -0.04334     Variance                 0.23174
Mode        .         Range                    2.63821
                      Interquartile Range      0.70032

Tests for Location: Mu0=0
Test            -Statistic-    -----p Value------
Student's t     t         0    Pr > |t|    1.0000
Sign            M        -5    Pr >= |M|   0.3584
Signed Rank     S       -81    Pr >= |S|   0.7690

Tests for Normality
Test                    --Statistic---    -----p Value------
Shapiro-Wilk            W     0.982393    Pr < W      0.2258
Kolmogorov-Smirnov      D     0.073231    Pr > D     >0.1500
Cramer-von Mises        W-Sq  0.069917    Pr > W-Sq  >0.2500
Anderson-Darling        A-Sq  0.431494    Pr > A-Sq  >0.2500

Quantiles (Definition 5)
Quantile        Estimate
100% Max       1.6190788
99%            1.6190788
95%            0.7044858
90%            0.6353334
75% Q3         0.3458140
50% Median    -0.0433379
25% Q1        -0.3545094
10%           -0.5477752
5%            -0.6895717
1%            -1.0191271
0% Min        -1.0191271
```

Extreme Observations

------Lowest------		------Highest-----	
Value	Obs	Value	Obs
-1.019127	92	0.704486	53
-1.016438	71	0.813986	47
-0.881229	17	0.827172	90
-0.732271	87	1.115645	95
-0.689572	32	1.619079	94