Freund & Wilson (1997) : Prediction of the distance of a nebulae (megaparsecs) from the recession velocity (km/sec) (Case 7.1.1)

Obs	Velocity	Distance	Vel*Vel	Dist*Dist	Vel*Dist	Predicted	Residual
1	170	0.032	28900	0.00102	5.44	0.0971	0.1779
2	290	0.034	84100	0.00116	9.86	0.1451	0.1299
3	-130	0.214	16900	0.0458	-27.82	0.2206	-0.0066
4	-70	0.263	4900	0.06917	-18.41	0.303	-0.04
5	-185	0.275	34225	0.07563	-50.875	0.3579	0.5421
6	-220	0.275	48400	0.07563	-60.5	0.605	0.295
7	200	0.45	40000	0.2025	90	0.6325	-0.6005
8	290	0.5	84100	0.25	145	0.6737	-0.2237
9	270	0.5	72900	0.25	135	0.6737	-0.0437
10	200	0.63	40000	0.3969	126	0.7698	-0.2698
11	300	0.8	90000	0.64	240	0.7972	-0.7632
12	-30	0.9	900	0.81	-27	0.7972	-0.2972
13	650	0.9	422500	0.81	585	0.811	-0.011
14	150	0.9	22500	0.81	135	1.0169	0.0831
15	500	0.9	250000	0.81	450	1.0856	-0.1856
16	920	1	846400	1	920	1.0856	0.0144
17	450	1.1	202500	1.21	495	1.0856	0.3144
18	500	1.1	250000	1.21	550	1.0856	0.9144
19	500	1.4	250000	1.96	700	1.2915	-0.3915
20	960	1.7	921600	2.89	1632	1.4974	0.5026
21	500	2	250000	4	1000	1.5661	0.4339
22	850	2	722500	4	1700	1.6622	-0.6622
23	800	2	640000	4	1600	1.7171	-0.0171
24	1090	2	1188100	4	2180	1.8956	0.1044
25	1200	.	1440000	.	.	2.0466	
Sum	**8955**	**21.873**	**6511425**	**29.5178**	**12513.7**		**0**
Mean	**373.125**	**0.91137**	**271309.4**	**1.23**	**521.404**		

Intermediate Calculations

Sum X =	**8955**	**Sum Y =**	**21.873**
Sum X^2 =	**6511425**	**Sum Y^2 =**	**29.51780**
Mean X=	**373.125**	**Mean Y=**	**0.91137**
Sum XY =	**12513.695**	**n =**	**24**

Correction factors and Corrected values (Sums of squares and cross-products)

CF for X =	C_{xx} = 3341334.375	Corrected SS X =	S_{xx} = 3170090.625
CF for Y =	C_{yy} = 19.93450537	Corrected SS Y =	S_{yy} = 9.58328967
CF for XY =	C_{xy} = 8161.363	Corrected SS XY =	S_{xy} = 4352.331899

Derivation of the equations (solution)

Given our original equation, $Y_i = b_0 + b_1X_i + e_i$, we can determine that the residual is defined as $e_i = Y_i - (b_0 + b_1X_i)$. We wish to minimize the deviations of the observations from the

regression line, so we minimize $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$. Derivatives with respect to the two unknown values (b_0 and b_1),

$$\frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial b_0} = 2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)(-1) \quad \text{and} \quad \frac{\partial\left(\sum_{i=1}^{n} e_i^2\right)}{\partial b_1} = 2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)(-X_i)$$

are solved for zero, yielding two equations with two unknowns. These are called the normal equations, $nb_0 + b_1\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i$ and $b_0\sum_{i=1}^{n} X_i + b_1\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i Y_i$. These two equations can be solved for the two unknows, yielding the solution to the regression line parameter estimates. These equations are $b_1 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \frac{\Sigma X_i Y_i - (X_i)(Y_i)/n}{\Sigma X_i^2 - \Sigma(X_i)^2/n}$ and $b_0 = \bar{Y}_i - b_1\bar{X}_i$.

Model Parameter Estimates

Slope = b_1 = 4352.331899 / 3170090.625 = 0.001372936
Intercept = b_0 = 0.91137 – 0.001372936 * 373.125 = 0.399098213
Regression Line $Y_i = b_0 + b_1 * X_i + e_i$
Y_i = 0.399098213 + 0.001372936 * X_i + e_i

ANOVA Table

SSTotal = 9.58328967
SSRegression = 4352.331899^2 / 3170090.625 = 5.975473638
SSError = 9.58328967 – 5.975473638 = 3.607816032

Source	**df**	**SS**	**MS**	**F**
Regression	1	5.975473638	5.975473638	36.43767278
Error	22	3.607816032	0.163991638	
Total	23	9.583289670		

Standard error of b_1 : where $t_{(0.05/2,\ 45\ df)}$ = 2.073875294 ;

$$S_{b_1} = \sqrt{MSE/\sum X_i^2} = 0.163991638 / 3170090.625 = 0.0002274443$$

P(0.0013729 – 2.073875*0.00022744 ≤ β_1 ≤ 0.0013729 + 2.073875*0.00022744) = 0.95
P(0.000901245 ≤ β_1 ≤ 0.001844627) = 0.95

Testing b_1 against a specified value : H_0: β_1 = 0.0015 versus H_1: $\beta_1 \neq 0.0015$

$$t = \frac{b_1 - \beta_{1|H_0}}{S_{b_1}} = (0.0013729 - 0.0015) / 0.0002274443 = -0.558659419$$

Note that t^2 = F = 0.312100 ; This test would be done in SAS as an F statement

The variance of a linear combination is given by the sum of the variances plus twice the covariances.

e.g. for **A = aX + bY + cZ**

then **$Var(A) = a^2\sigma^2_X + b^2\sigma^2_Y + c^2\sigma^2_Z + 2(ab\sigma_{XY} + ac\sigma_{XZ} + bc\sigma_{YZ})$**

where the covariances are assumed equal to zero if the variables are independent

For the linear combination $\hat{Y}_i = b_0 + b_1 X_i$, the standard error of $\hat{Y}_i$ is as follows.

Standard error of the regression line ($\hat{Y}_i$): $S_{\hat{Y}_{y|x}} = \sqrt{MSE\left(\frac{1}{n} + \frac{\left(X_i - \bar{X}.\right)^2}{\sum\left(X_i - \bar{X}.\right)^2}\right)}$

The calculation above DOES NOT assume that the covariances of the regression coefficients are independent. However, for the variance of individual points the linear combination is $Y_i = b_0 + b_1 X_i + e_i = \hat{Y}_i + e_i$. For this linear combination the terms for the predicted value and residuals are assumed independent (i.e. $\hat{Y}_i$ is independent of e_i).

$$S_{Y_{y|x}} = \sqrt{MSE\left(\frac{1}{n} + \frac{\left(X_i - \bar{X}.\right)^2}{S_{xx}}\right) + MSE} = \sqrt{MSE\left(1 + \frac{1}{n} + \frac{\left(X_i - \bar{X}.\right)^2}{S_{xx}}\right)}$$

Standard error of b_0 is the same as the standard error of the regression line where $X_i = 0$

$$S_{Y_{y|x}} = \sqrt{MSE\left(\frac{1}{n} + \frac{\left(X_i - \bar{X}.\right)^2}{\sum\left(X_i - \bar{X}.\right)^2}\right)} = \sqrt{0.163991638\left(\frac{1}{24} + \frac{\left(0 - 373.125\right)^2}{3170090.625}\right)} =$$

SQRT(0.163991638 (0.041666667 + (0 - 373.125)^2 / 3170090.625)) = 0.118469734

Confidence interval on b_0 where b_0 = 0.399098213 and $t_{(0.05/2,\ 45\ df)}$ = 2.073875294

P(0.399098 – 2.073875*0.1184697 ≤ β_0 ≤ 0.399098 + 2.073875*0.1184697) = 0.95

P(0.153406758 ≤ β_0 ≤ 0.644789668) = 0.95

Estimate the predicted value and standard error of an individual observation (e.g. the distance to a nebula that is moving at 1200 km/sec)

Y_i = 0.399098213 + 0.001372936 * X_i = 0.399098213 + 0.001372936(1200) = 2.046622

$$S_{Y_{y|x}} = \sqrt{MSE\left(1 + \frac{1}{n} + \frac{\left(X_i - \bar{X}.\right)^2}{S_{xx}}\right)} = \sqrt{0.163991638\left(1 + \frac{1}{24} + \frac{\left(1200 - 373.125\right)^2}{3170090.625}\right)}$$

SQRT(0.163991638 (0.041666667 + (12000 - 373.125)^2 / 3170090.625)) = 0.205432598

P(2.046622 – 2.073875*0.2054326 ≤ $\mu_{x=1200}$ ≤ 2.046622 + 2.073875*0.2054326) = 0.95

P(799.1094964 ≤ $\mu_{x=10}$ ≤ 1313.372185) = 0.95

Calculate the coefficient of Determination and correlation

R^2 = 5.975473638 / 9.583289670 = 0.623530525 or 62.35305248 %

r = 0.78963949

```
1          /*
2          The Big Bang
3          Edwin Hubble discovered a relationship between a nebula's distance
4          from earth and the velocity with which it was moving away from earth.
5          The data in this analysis is for the first 24 nebulae Hubble measured.
6          */
7
8          dm'log;clear;output;clear';
9          options nodate nocenter nonumber ps=512 ls=132;
10         ODS HTML style=minimal rs=none
body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\Hubble01.html' ;
NOTE: Writing HTML Body file:
C:\Geaghan\Current\EXST3201\Fall2005\SAS\Hubble01.html
11
12         Title1 'Chapter 6 : Hubble example';
13         filename input
'C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0701.csv';
14
15         data Hubble; infile input missover DSD dlm="," firstobs=2;
16           input VELOCITY DISTANCE;
17           label velocity = 'Velocity in km/sec'
18                 distance = 'Distance in megaparsecs';
19         datalines;
NOTE: The infile INPUT is:
      File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0701.csv,
      RECFM=V,LRECL=256
NOTE: 24 records were read from the infile INPUT.
      The minimum record length was 5.
      The maximum record length was 23.
NOTE: The data set WORK.HUBBLE has 24 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.15 seconds
      cpu time            0.00 seconds
20         run;
21
22         data madeup; VELOCITY=1200; DISTANCE=.; output; run;
NOTE: The data set WORK.MADEUP has 1 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
23
24         data hubble; set hubble madeup;
25
NOTE: There were 24 observations read from the data set WORK.HUBBLE.
NOTE: There were 1 observations read from the data set WORK.MADEUP.
NOTE: The data set WORK.HUBBLE has 25 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
26         proc sort data=Hubble; by velocity; run;
NOTE: There were 25 observations read from the data set WORK.HUBBLE.
NOTE: The data set WORK.HUBBLE has 25 observations and 2 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.07 seconds
      cpu time            0.00 seconds
27         Title2 'Raw data listing';
28         proc print data=Hubble; run;
NOTE: There were 25 observations read from the data set WORK.HUBBLE.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.13 seconds
```

Chapter 6 : Hubble example
Raw data listing

Obs	VELOCITY	DISTANCE
1	-220	0.27500
2	-185	0.27500
3	-130	0.21400
4	-70	0.26300
5	-30	0.90000
6	150	0.90000
7	170	0.03200
8	200	0.45000
9	200	0.63000
10	270	0.50000
11	290	0.03400
12	290	0.50000
13	300	0.80000
14	450	1.10000
15	500	0.90000
16	500	1.10000
17	500	1.40000
18	500	2.00000
19	650	0.90000
20	800	2.00000
21	850	2.00000
22	920	1.00000
23	960	1.70000
24	1090	2.00000
25	1200	.

```
29
30         Title2 'Regression with labels and simple output';
31         proc reg data=Hubble;
32            model distance = velocity / p;
33         run;
```

Chapter 6 : Hubble example
Regression with labels and simple output

The REG Procedure
Model: MODEL1
Dependent Variable: DISTANCE Distance in megaparsecs

Number of Observations Read	25
Number of Observations Used	24
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.97547	5.97547	36.44	<.0001
Error	22	3.60782	0.16399		
Corrected Total	23	9.58329			

Root MSE	0.40496	R-Square	0.6235
Dependent Mean	0.91137	Adj R-Sq	0.6064
Coeff Var	44.43383		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.39910	0.11847	3.37	0.0028
VELOCITY	Velocity in km/sec	1	0.00137	0.00022744	6.04	<.0001

Chapter 6 : Hubble example
Regression with labels and simple output

The REG Procedure
Model: MODEL1
Dependent Variable: DISTANCE Distance in megaparsecs

OutputStatistics

Obs	Dependent Variable	Predicted Value	Residual
1	0.2750	0.0971	0.1779
2	0.2750	0.1451	0.1299
3	0.2140	0.2206	-0.006617
4	0.2630	0.3030	-0.0400
5	0.9000	0.3579	0.5421
6	0.9000	0.6050	0.2950
7	0.0320	0.6325	-0.6005
8	0.4500	0.6737	-0.2237
9	0.6300	0.6737	-0.0437
10	0.5000	0.7698	-0.2698
11	0.0340	0.7972	-0.7632
12	0.5000	0.7972	-0.2972
13	0.8000	0.8110	-0.0110
14	1.1000	1.0169	0.0831
15	0.9000	1.0856	-0.1856
16	1.1000	1.0856	0.0144
17	1.4000	1.0856	0.3144
18	2.0000	1.0856	0.9144
19	0.9000	1.2915	-0.3915
20	2.0000	1.4974	0.5026
21	2.0000	1.5661	0.4339
22	1.0000	1.6622	-0.6622
23	1.7000	1.7171	-0.0171
24	2.0000	1.8956	0.1044
25	.	2.0466	.

Sum of Residuals	0
Sum of Squared Residuals	3.60782
Predicted Residual SS (PRESS)	4.20513

```
35         options nolabel;
36         Title2 'Regression without labels and with output statement';
NOTE: The PROCEDURE REG printed pages 2-3.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.28 seconds
      cpu time            0.04 seconds
37         proc reg data=Hubble;
38            model distance = velocity / clb;
39            test velocity = 0.0015;
40            output out=next1 r=resid p=yhat lclm=lclm uclm=uclm lcl=lcli
ucl=ucli stdp=stdp stdi=stdi;
41         run;
```

Chapter 6 : Hubble example
Regression without labels and with output statement

The REG Procedure
Model: MODEL1
Dependent Variable: DISTANCE

Number of Observations Read	25
Number of Observations Used	24
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.97547	5.97547	36.44	<.0001
Error	22	3.60782	0.16399		
Corrected Total	23	9.58329			

Root MSE	0.40496	R-Square	0.6235
Dependent Mean	0.91137	Adj R-Sq	0.6064
Coeff Var	44.43383		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	0.39910	0.11847	3.37	0.0028	0.15341	0.64479
VELOCITY	1	0.00137	0.00022744	6.04	<.0001	0.00090125	0.00184

Model: MODEL1

Test 1 Results for Dependent Variable DISTANCE

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.05118	0.31	0.5820
Denominator	22	0.16399		

```
42
43         Title3 'Listing of results from the regression output statement';
NOTE: The data set WORK.NEXT1 has 25 observations and 10 variables.
NOTE: The PROCEDURE REG printed pages 4-5.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
44         proc print data=next1; run;
NOTE: There were 25 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PRINT printed page 6.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
```

Chapter 6 : Hubble example
Regression without labels and with output statement
Listing of results from the regression output statement

Obs	VELOCITY	DISTANCE	yhat	stdp	lclm	uclm	lcli	ucli	resid	stdi
1	-220	0.27500	0.09705	0.15821	-0.23106	0.42517	-0.80460	0.99871	0.17795	0.43477
2	-185	0.27500	0.14511	0.15148	-0.16905	0.45926	-0.75156	1.04177	0.12989	0.43236
3	-130	0.21400	0.22062	0.14117	-0.07214	0.51338	-0.66878	1.11001	-0.00662	0.42886
4	-70	0.26300	0.30299	0.13035	0.03267	0.57332	-0.57928	1.18526	-0.03999	0.42542
5	-30	0.90000	0.35791	0.12345	0.10189	0.61393	-0.52008	1.23590	0.54209	0.42336
6	150	0.90000	0.60504	0.09700	0.40388	0.80620	-0.25855	1.46863	0.29496	0.41641
7	170	0.03200	0.63250	0.09470	0.43611	0.82889	-0.22999	1.49499	-0.60050	0.41588
8	200	0.45000	0.67369	0.09156	0.48380	0.86357	-0.18735	1.53472	-0.22369	0.41518
9	200	0.63000	0.67369	0.09156	0.48380	0.86357	-0.18735	1.53472	-0.04369	0.41518
10	270	0.50000	0.76979	0.08593	0.59159	0.94799	-0.08874	1.62832	-0.26979	0.41397
11	290	0.03400	0.79725	0.08480	0.62139	0.97311	-0.06080	1.65530	-0.76325	0.41374
12	290	0.50000	0.79725	0.08480	0.62139	0.97311	-0.06080	1.65530	-0.29725	0.41374
13	300	0.80000	0.81098	0.08432	0.63611	0.98584	-0.04687	1.66882	-0.01098	0.41364
14	450	1.10000	1.01692	0.08449	0.84170	1.19214	0.15900	1.87484	0.08308	0.41368
15	500	0.90000	1.08557	0.08755	0.90399	1.26714	0.22633	1.94480	-0.18557	0.41432
16	500	1.10000	1.08557	0.08755	0.90399	1.26714	0.22633	1.94480	0.01443	0.41432
17	500	1.40000	1.08557	0.08755	0.90399	1.26714	0.22633	1.94480	0.31443	0.41432
18	500	2.00000	1.08557	0.08755	0.90399	1.26714	0.22633	1.94480	0.91443	0.41432
19	650	0.90000	1.29151	0.10392	1.07600	1.50702	0.42446	2.15855	-0.39151	0.41808
20	800	2.00000	1.49745	0.12751	1.23300	1.76189	0.61696	2.37793	0.50255	0.42456
21	850	2.00000	1.56609	0.13637	1.28328	1.84891	0.67992	2.45227	0.43391	0.42730
22	920	1.00000	1.66220	0.14935	1.35247	1.97192	0.76707	2.55732	-0.66220	0.43162
23	960	1.70000	1.71712	0.15700	1.39151	2.04272	0.81637	2.61786	-0.01712	0.43433
24	1090	2.00000	1.89560	0.18281	1.51648	2.27471	0.97416	2.81704	0.10440	0.44431
25	1200	.	2.04662	0.20543	1.62058	2.47266	1.10490	2.98834	.	0.45409

```
46         Title2 'Regression forced through the origin';
47         proc reg data=Hubble;
48            model distance = velocity / noint;
49         run;
NOTE: The PROCEDURE REG printed page 7.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
Chapter 6 : Hubble example
Regression forced through the origin

The REG Procedure
Model: MODEL1
Dependent Variable: DISTANCE

Number of Observations Read                          25
Number of Observations Used                          24
Number of Observations with Missing Values            1
```

NOTE: No intercept in model. R-Square is redefined.

```
                            Analysis of Variance
                                   Sum of           Mean
Source                     DF        Squares         Square    F Value    Pr > F
Model                       1       24.04889       24.04889     101.14    <.0001
Error                      23        5.46890        0.23778
Uncorrected Total          24       29.51780

Root MSE              0.48763    R-Square     0.8147
Dependent Mean        0.91137    Adj R-Sq     0.8067
Coeff Var            53.50434

                              Parameter Estimates
                Parameter     Standard
Variable    DF    Estimate        Error    t Value    Pr > |t|      95% Confidence Limits
VELOCITY     1     0.00192   0.00019109      10.06      <.0001     0.00153        0.00232
```

One megaparsec = 979.8 billion years, estimate = 0.00192*979.8 = 1.88 billion years with a range of 1.50 to 2.27 billion years.

```
51         Title2 'Regression with PROC MIXED';
52         proc mixed data=Hubble;
53            model distance = velocity;
54         run;
NOTE: 1 observation is not included because of missing values.
NOTE: The PROCEDURE MIXED printed page 8.
NOTE: PROCEDURE MIXED used (Total process time):
      real time           0.10 seconds
      cpu time            0.01 seconds

Chapter 6 : Hubble example
Regression with PROC MIXED

The Mixed Procedure
                  Model Information
Data Set                      WORK.HUBBLE
Dependent Variable            DISTANCE
Covariance Structure          Diagonal
Estimation Method             REML
Residual Variance Method      Profile
Fixed Effects SE Method       Model-Based
Degrees of Freedom Method     Residual

            Dimensions
Covariance Parameters              1
Columns in X                       2
Columns in Z                       0
Subjects                           1
Max Obs Per Subject               25
```

```
          Number of Observations
Number of Observations Read              25
Number of Observations Used              24
Number of Observations Not Used           1

Covariance Parameter
      Estimates
Cov Parm     Estimate
Residual       0.1640

           FitStatistics
-2 Res Log Likelihood            40.8
AIC (smaller is better)          42.8
AICC (smaller is better)         43.0
BIC (smaller is better)          43.9

        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
VELOCITY        1      22      36.44    <.0001



56         Title2 'Regression with PROC GLM';
57         proc glm data=Hubble;
58            model distance = velocity;
59         run;
NOTE: The PROCEDURE GLM printed pages 9-10.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.20 seconds
      cpu time            0.01 seconds


Chapter 6 : Hubble example
Regression with PROC GLM

The GLM Procedure

Number of Observations Read          25
Number of Observations Used          24


Dependent Variable: DISTANCE
                                     Sum of
Source                       DF         Squares     Mean Square    F Value    Pr > F
Model                         1      5.97547364      5.97547364      36.44    <.0001
Error                        22      3.60781603      0.16399164
Corrected Total              23      9.58328967

R-Square     Coeff Var      Root MSE    DISTANCE Mean
0.623531      44.43383      0.404959         0.911375

Source                       DF       Type I SS     Mean Square    F Value    Pr > F
VELOCITY                      1      5.97547364      5.97547364      36.44    <.0001

Source                       DF     Type III SS     Mean Square    F Value    Pr > F
VELOCITY                      1      5.97547364      5.97547364      36.44    <.0001

                                  Standard
Parameter         Estimate           Error    t Value    Pr > |t|
Intercept     0.3990982130      0.11846973       3.37      0.0028
VELOCITY      0.0013729361      0.00022744       6.04      <.0001
```

```
61          GOPTIONS DEVICE=CGMLT97L ctitle=black ctext=black
62                   ftext='TimesRoman' ftitle='TimesRoman';
63
64          GOPTIONS GSFNAME=OUT1;
65          FILENAME OUT1 'C:\Geaghan\Current\EXST3201\Fall2005\SAS\Hubble01.CGM';
66          PROC GPLOT DATA=Hubble;
67             TITLE1 'Simple Linear Regression Example';
68             TITLE2 'Hubbles nebula distance data';
69           PLOT distance*velocity=1 distance*velocity=2 distance*velocity=3
70                / overlay HAXIS=AXIS1 VAXIS=AXIS2;
71           AXIS1 LABEL=('Velocity in km/sec') MINOR=(N=1)
72                 color=black ORDER=-600 TO 1400 BY 200;
73           AXIS2 LABEL=(ANGLE=90 'Distance in megaparsecs') MINOR=(N=5)
74                 color=black ORDER=0 TO 3 BY 0.5;
75           SYMBOL1 color=red   V=None I=RLcli99  L=1 MODE=INCLUDE;
76           SYMBOL2 color=green V=None I=RLclm99  L=1 MODE=INCLUDE;
77           SYMBOL3 color=blue  V=dot  I=None     L=1 MODE=INCLUDE; run;
NOTE: Regression equation :  DISTANCE =  0.399098 + 0.001373*VELOCITY.
NOTE: 1 observation(s) contained a MISSING value for the DISTANCE * VELOCITY
request.
NOTE: Regression equation :  DISTANCE =  0.399098 + 0.001373*VELOCITY.
NOTE: 1 observation(s) contained a MISSING value for the DISTANCE * VELOCITY
request.
NOTE: 1 observation(s) contained a MISSING value for the DISTANCE * VELOCITY
request.
NOTE:  51 RECORDS WRITTEN TO
C:\Geaghan\Current\EXST3201\Fall2005\SAS\Hubble01.CGM
```