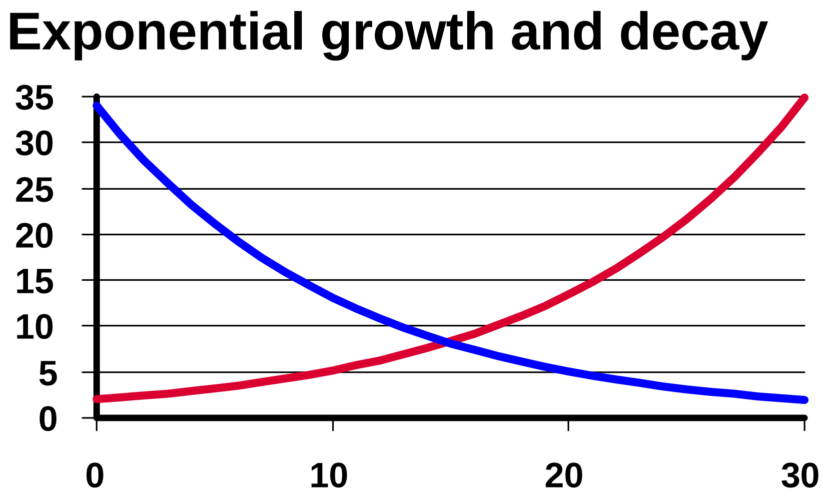
Chapter 8 (More on Assumptions for the Simple Linear Regression)

Chapter 8 covers the assumptions behind the SLR, and some alternative models that can be used. One of those alternatives involves the use of transformations to fit data that does not form a straight line when plotted.

Examples of curves: The exponential model (exponential growth and decay or mortality model)

$$Y_i = b_0 \exp^{b_1 X_i} e_i$$



Decreasing line: $b_0 = 34$, $b_1 = -0.953$, $e^{b1} = 0.909$

Increasing line: $b_0 = 2$, $b_1 = +0.953$, $e^{b1} = 1.1$

For this model the raw data should actually appear non-homogeneous

The equation is $Y_i = b_0 \exp^{b_1 X_i} e_i$, taking logs we get $\ln(Y_i) = \ln(b_0) + b_1 X_i + \ln(e_i)$. This is a simple linear regression where the dependent variable is $\ln(Y_i)$ and the independent variable is X_i. Once fitted the estimated slope is equal to b_1. The estimated intercept is $\ln(b_0)$ and the antilog must be taken to get back to the original equation ($Y_i = b_0 \exp^{b_1 X_i} e_i$).

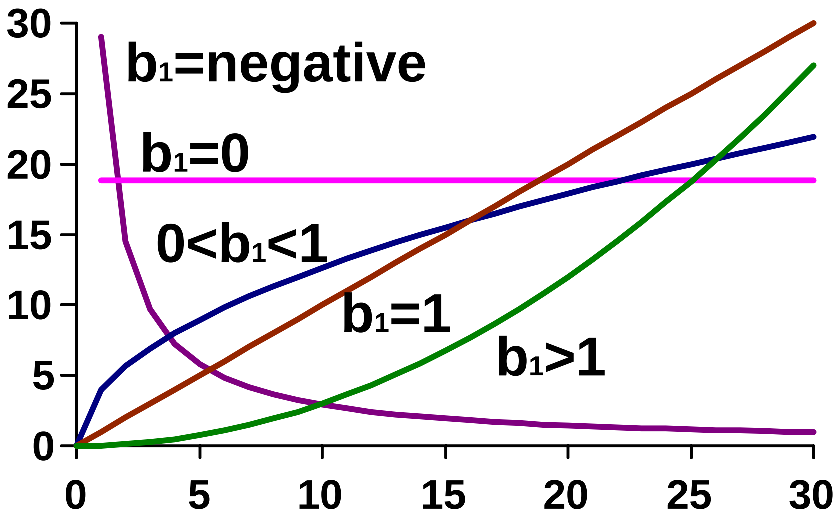
When fitted in SAS the model would be "MODEL LY = X;" where LY is the log of the Y variable and X is the unaltered X variable. These transformations must be done in the data step; they cannot be done in the proc step.

When the model is fitted on the log transformed data all of the usual assumptions apply. However, they apply to the log transformed version of the data, not to the raw data. Also, all tests of hypothesis and confidence intervals must be done on the log transformed data. Once all estimates of parameters, predicted values and confidence intervals are done they can be detransformed. Standard errors cannot be detransformed, so any estimates and confidence intervals must be done on the log transformed data.

In general, when the dependent variable, Y, is transformed the homogeneity is altered. If only the independent variable, X, is altered then the homogeneity is unchanged.

Another examples of curves: The power model

$$Y_i = b_0 X_i^{b_1} e_i$$



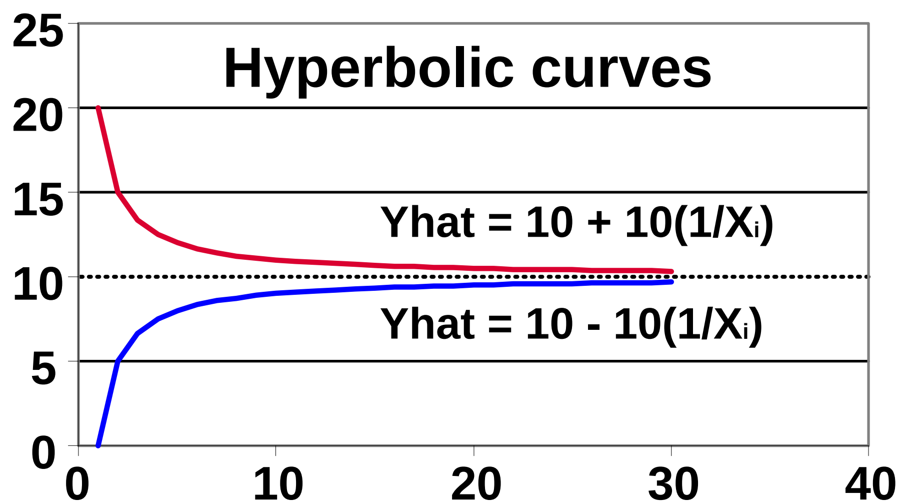
Values of b_0 and b_1: (29, -1), (19, 0), (4, 0.5), (1, 1), (0.03, 2)

For this model the raw data should appear non-homogeneous

The equation is $Y_i = b_0 X_i^{b_1} e_i$, taking logs we get $\ln(Y_i) = \ln(b_0) + b_1 \ln(X_i) + \ln(e_i)$. This is a simple linear regression where the dependent variable is $\ln(Y_i)$ and the independent variable is $\ln(X_i)$. Once fitted the estimated slope is equal to b_1. The estimated intercept is $\ln(b_0)$ and the antilog must be taken to get back to the original equation ($Y_i = b_0 \exp^{b_1 X_i} e_i$).

When fitted in SAS the model would be "MODEL LY = LX;" where LY is the log of the Y variable and LX is the log of the X variable. These transformations must be done in the data step; they cannot be done in the proc step.

Many other transformed models exist. While logarithmic transformations are common, inverse transformations and root transformations are also used. An example of an inverse transformation fitting a hyperbola is given below ($Y_i = b_0 + b_1 \frac{1}{X_i} + e_i$).



The example offered in your books is of the power model type above. In this example biologists have noted a possible relationship between the size of an island and the number of species of animals and plants on that island. The data in this example are for reptile and amphibian species on seven islands in the West Indies. Results for this analysis are given below.

Obs	AREA	SPECIES	Name
1	44218	100	Cuba
2	29371	108	Hispaniola
3	4244	45	Jamaica
4	3435	53	Puerto Rico
5	32	16	Monserrat
6	5	11	Saba
7	1	7	Redonda

According to the biologists the "relation" between species and islands is given by the formula $\text{Median}\{S|A\} = CA^{\gamma}$. If this model is fitted using logs we have $\text{Log}(\text{Median}\{S|A\}) = \log(C) + \gamma*\log(A)$. When estimated with SAS we get estimates of $\log(C) = 1.93651$ and $\gamma = 0.24968$. Lower and upper limits of the confidence interval on the parameter γ are 0.21856 and 0.28080, respectively.

This model is used for many other purposes. In instrument standardization, where two instruments are compared for readings on a range of standards this model is often used. If the two instruments match perfectly the model, $Y=b_oX^{b1}$ becomes $Y=X$ where $b_0=1$ and $b_1=1$. Note that this model always goes through the intercept, so when $X = 0$ then $Y = 0$.

This model is also used for fitting morphometric relationships in the biological sciences. These are relationships between the body parts of organism. For example, in various studies biologists will use total length or thoracic length to measure shrimp, or total length or fork length to measure fish. Subsequently, conversion formulas are needed to compare across studies. Paleontologists predict the size of dinosaurs from a few bones because the relationship between total size and selected bones, say the femur, is known for similar animals. A related use is the prediction of body mass from linear measurements. When comparing linear measurements with the equation $Y=b_oX^{b1}$, the power term (b_1) is expected to be 1, relating cm to cm. However, for body mass, relating gm to cm we expect the power term to be approximately 3, since gm must be related to cm^3.

```
          /*
          Biologists have noted a possible relationship between the size of an island
          and the number of species of animals and plants on that island.  The data in
          this example are for reptile and amphibian species on seven islands in the
          West Indies.
          */

          dm'log;clear;output;clear';
          options nodate nocenter nonumber ps=512 ls=99 nolabel;
          ODS HTML style=minimal rs=none
        ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\IsleSpecies01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\IsleSpecies01.html

          Title1 'Chapter 8: Correlation between the size of an island and the number of species'
        ! ;
          filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0801.csv';

          data Islands; infile input1 missover DSD dlm="," firstobs=2;
            input AREA SPECIES;
            label SPECIES = 'Number of reptile and amphibian species'
                  AREA = 'AREA (square miles)';
            LArea = log(area); *** the LOG(X) function gives the natural log (base e) ***;
            LSPECIES = log(SPECIES); *** the LOG10(X) function gives log base 10 ***;
          datalines;
NOTE: The infile INPUT1 is:
      File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\case0801.csv,
      RECFM=V,LRECL=256
NOTE: 7 records were read from the infile INPUT1.
      The minimum record length was 3.
      The maximum record length was 9.
NOTE: The data set WORK.ISLANDS has 7 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
          run;

          Title2 'Raw data listing';
          proc print data=Islands; run;
NOTE: There were 7 observations read from the data set WORK.ISLANDS.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

Chapter 8 : Correlation between the size of an island and the number of species
Raw data listing

Obs	AREA	SPECIES	LArea	LSPECIES
1	44218	100	10.6969	4.60517
2	29371	108	10.2878	4.68213
3	4244	45	8.3533	3.80666
4	3435	53	8.1418	3.97029
5	32	16	3.4657	2.77259
6	5	11	1.6094	2.39790
7	1	7	0.0000	1.94591

```
27         options ps=52 ls=111;
28         Title2 'Scatter plot of the taw data';
29         proc plot data=Islands; plot SPECIES * AREA; run;
30         Title2 'Scatter plot of the log transformed data';
NOTE: There were 7 observations read from the data set WORK.ISLANDS.
NOTE: The PROCEDURE PLOT printed page 2.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
31         proc plot data=Islands; plot LSPECIES * LAREA; run;
32         options ps=512 ls=99;
NOTE: There were 7 observations read from the data set WORK.ISLANDS.
NOTE: The PROCEDURE PLOT printed page 3.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

```
Chapter 8 : Correlation between the size of an island and the number of species
Scatter plot of the taw data

                     Plot of SPECIES*AREA.  Legend: A = 1 obs, B = 2 obs, etc.

       SPECIES |
               |
           120 +
               |
               |
               |
               |                                           A
               |
           100 +                                                                A
               |
               |
               |
               |
               |
            80 +
               |
               |
               |
               |
               |
            60 +
               |
               |       A
               |
               |        A
               |
            40 +
               |
               |
               |
               |
               |
            20 +
               |  A
               |
               |  A
               |  A
               |
             0 +
               |
               ---+-------------+-------------+-------------+-------------+-------------+--
                  0           10000         20000         30000         40000         50000
                                                   AREA
```

```
Chapter 8 : Correlation between the size of an island and the number of species
Scatter plot of the log transformed data

                    Plot of LSPECIES*LArea.  Legend: A = 1 obs, B = 2 obs, etc.
LSPECIES |
         |
     5.0 +
         |
         |
         |                                                                           A
         |                                                                              A
     4.5 +
         |
         |
         |
         |
     4.0 +                                                            A
         |
         |                                                             A
         |
         |
     3.5 +
         |
         |
         |
         |
     3.0 +
         |
         |                          A
         |
         |
     2.5 +
         |             A
         |
         |
         |
     2.0 +
         |  A
         |
         |
         |
     1.5 +
         |
         ---+-------------+-------------+-------------+-------------+-------------+----------
            0             2             4             6             8            10
                                                     LArea
```

```
34         Title2 'Regression without transformed values';
35         proc reg data=Islands;
36            model LSPECIES = LAREA / CLB;
37            output out=next1 r=resid p=yhat;
38         run;
39
40         Title3 'Listing of results from the regression output statement';
NOTE: The data set WORK.NEXT1 has 7 observations and 6 variables.
NOTE: The PROCEDURE REG printed page 4.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
41         proc print data=next1; run;
NOTE: There were 7 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PRINT printed page 5.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
Chapter 8 : Correlation between the size of an island and the number of species
Regression without transformed values

The REG Procedure
Model: MODEL1
Dependent Variable: LSPECIES

Number of Observations Read          7
Number of Observations Used          7
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6.99619	6.99619	425.30	<.0001
Error	5	0.08225	0.01645		
Corrected Total	6	7.07844			

Root MSE	0.12826	R-Square	0.9884
Dependent Mean	3.45438	Adj R-Sq	0.9861
Coeff Var	3.71289		

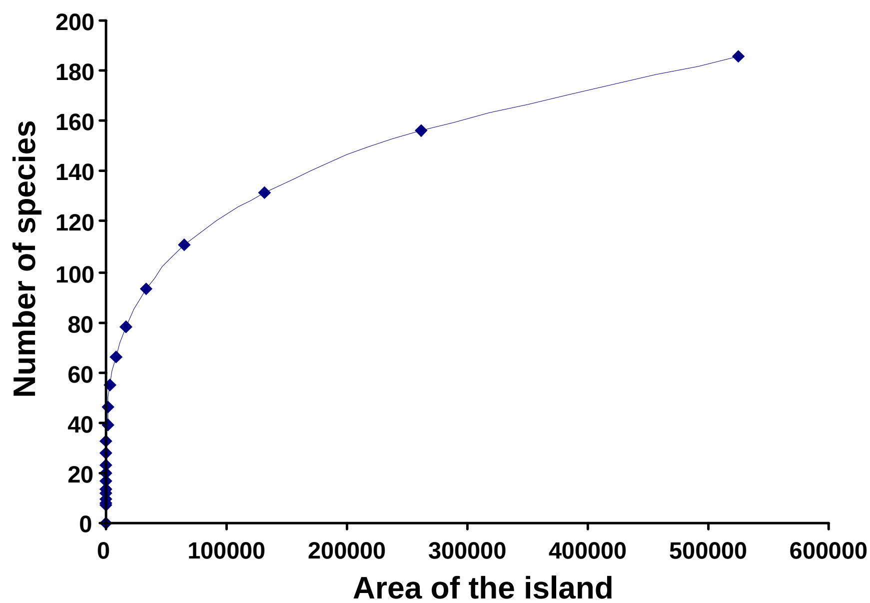
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	1.93651	0.08813	21.97	<.0001	1.70996	2.16306
LArea	1	0.24968	0.01211	20.62	<.0001	0.21856	0.28080

```
Chapter 8 : Correlation between the size of an island and the number of species
Regression without transformed values
Listing of results from the regression output statement
```

Obs	AREA	SPECIES	LArea	LSPECIES	yhat	resid
1	44218	100	10.6969	4.60517	4.60731	-0.00214
2	29371	108	10.2878	4.68213	4.50516	0.17698
3	4244	45	8.3533	3.80666	4.02215	-0.21549
4	3435	53	8.1418	3.97029	3.96935	0.00095
5	32	16	3.4657	2.77259	2.80183	-0.02924
6	5	11	1.6094	2.39790	2.33835	0.05954
7	1	7	0.0000	1.94591	1.93651	0.00940

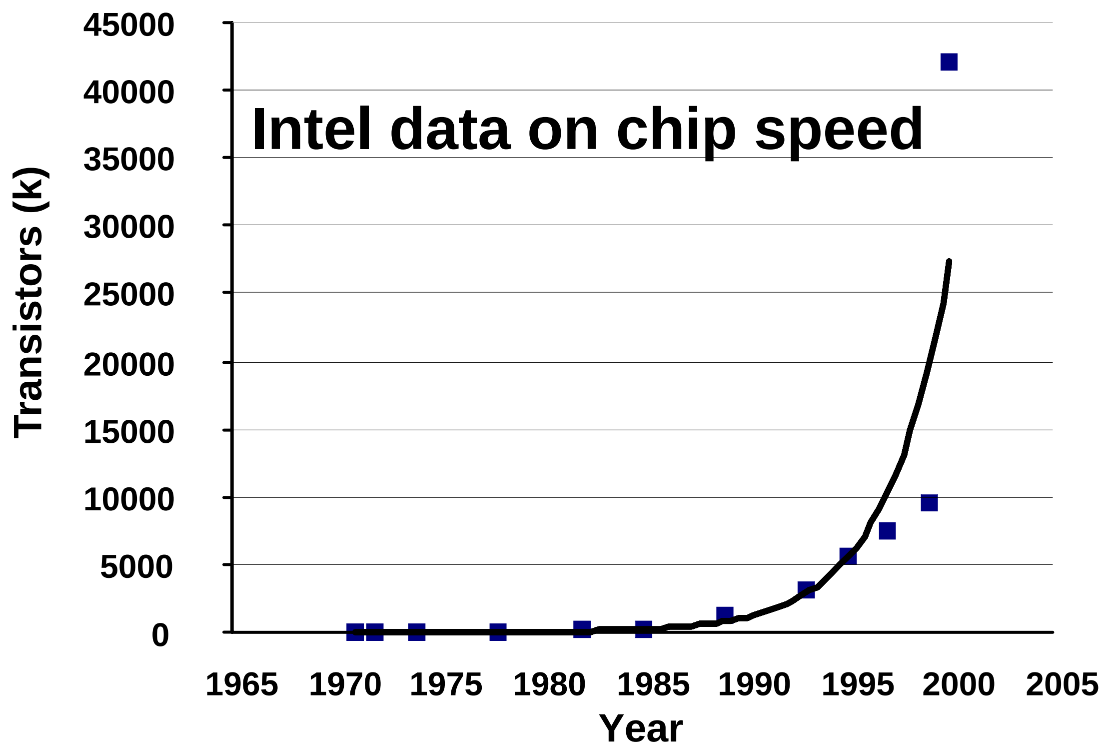
The original equation is then Median{S|A} = $CA^{\gamma} = \exp(1.93651)A^{0.24968} = 6.935*A^{0.250}$. When plotted this should form a curve



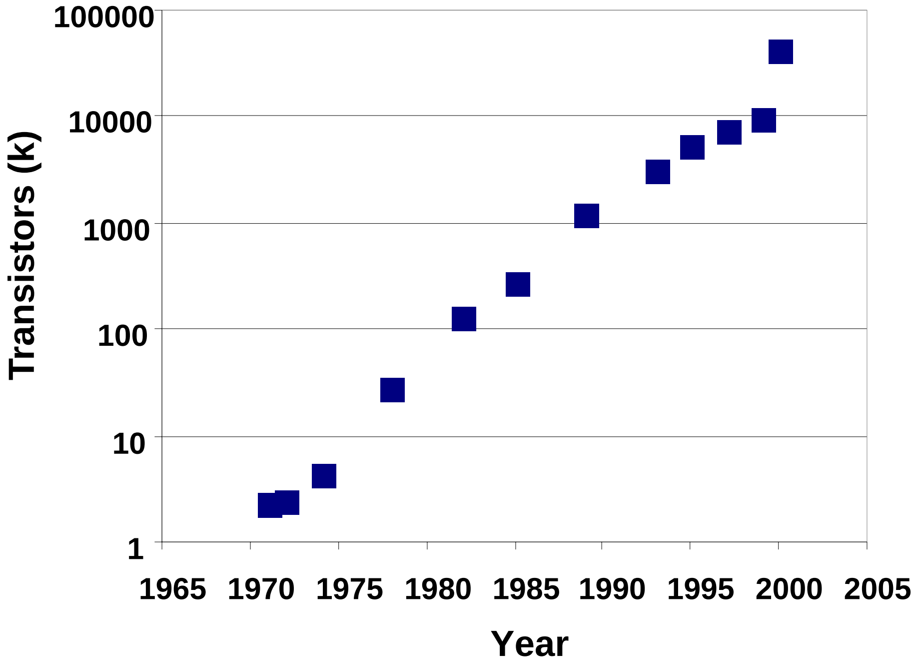
A brochure for the Intel Corporation, a major producer of computer processor chips, contains information about the history of it's chips. The table below was derived from that information.

Processor	Year of introduction	Number of Transistors (x1000)	Logarithm of No. of Transistors	Approx. MIPS (million instructions per second)
4004	1971	2.3	0.83	0.04
8080	1974	9	2.2	0.16
8086	1978	20	3	0.36
8088	1979	29	3.37	0.53
80286	1982	134	4.9	2.44
80386	1985	275	5.62	5
80486	1989	1200	7.09	21.82
Pentium (5)	1993	3000	8.01	54.55
Pentium Pro (6)	1995	5500	8.61	100

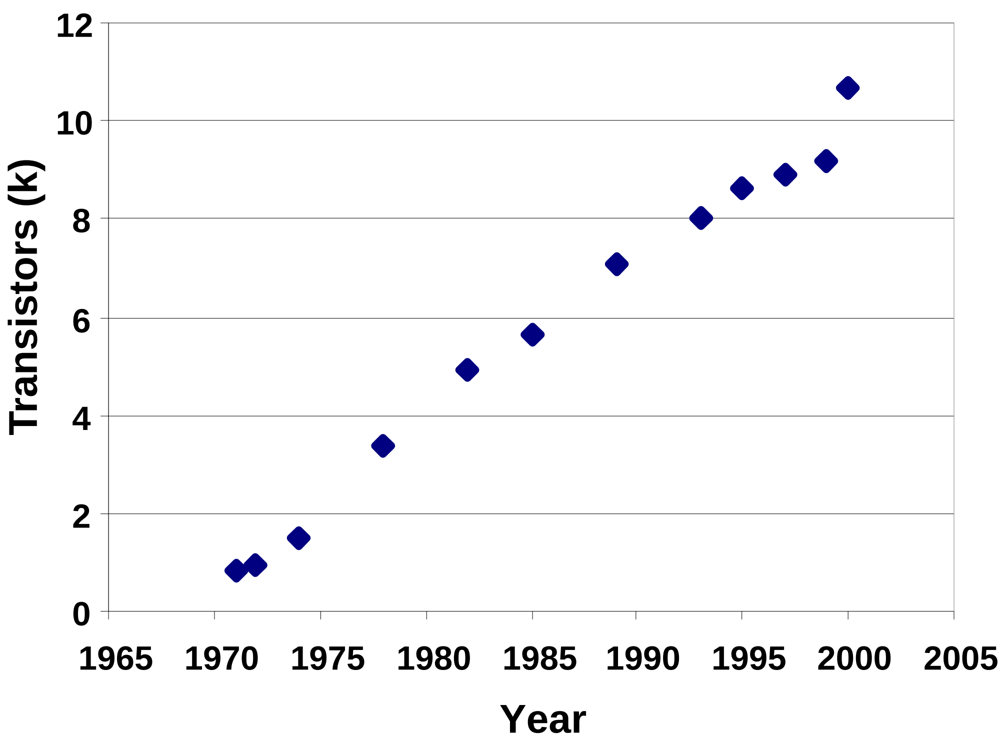
The first plot below is a plot of the data with an "exponential growth curve" graphed through the data. This appears to be a pretty good fit.



Below is a plot of the relationship between the independent variable "year of introduction" and the dependent variable plotted on a logarithmic scale (log 10 in this case) This fit also looks pretty good.



This same effect can be achieved by taking logarithms of the variable instead of plotting on a log scale. That graph, with natural logs, is given below.



This is the linear version of the exponential curve, and as you can see it is a simple linear regression. This data was regressed in SAS using PROC REG and the data is given in in separate pages. Answer the questions below from that computer output. Note that the intercept is meaningless in this particular example because there were no computers in the year "0000".

In 1965 Gordon Moore (a co-founder of Intel)) wrote an article that claimed that the rate of increase in computer capacity double every 18 months. On our logarithmic scale this translates into a rate of 0.462 logTRANS per year. This value was tested and rejected (P>F < 0.0001)

This doubling rate can be easily calculated. The curve is of the “exponential growth” type, $Y_i = b_0 \exp^{b_1 X_i} e_i$. For this model the value of b_0 is the “initial value” and the value of b_1 is the instantaneous rate of increase. The doubling time is calculated as

$$Y_i = b_0 e^{b_1 X_i} = (1.5254*10^{-275})e^{0.32150*\text{year}}$$

$$2b_0 = b_0 e^{b_1 (\text{doubling time})} = 3.0508E\text{-}275 = (1.5254*10^{-275})e^{0.32150(\text{doubling time})}.$$

$$2 = e^{0.32150(\text{doubling time})} \text{ and } \ln(2) = 0.32150*(\text{doubling time}) \text{ then}$$

$$\text{doubling time} = \ln(2)/0.32150 = 2.155978789$$

The annual rate is $e^{0.32150} = 1.379195006$, or about 38% per year, but this rate is “compounded” continously.

This model is used to describe the growth (positive b_1) of biological organisms and populations, growth in technology, and anything that grows proportionally, such as interest in the bank.

It is also used extensively in biology to describe various types of mortality (negative b_1) and types of decay for both biological material and radioactive material. Half lives are calculated in a fashion similar to the doubling time above.

```
1          ************************************************************;
2          *** A brochure from the Intel Corporation, a major producer ***;
3          *** of computer processor chips, contains information about ***;
4          *** the history of it's chips.  The information in this     ***;
5          *** example is from that brochure.                          ***;
6          ************************************************************;
7
8          dm'log;clear;output;clear';
9          options nodate nocenter nonumber ps=512 ls=99 nolabel;
10         ODS HTML style=minimal rs=none
10       ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\Intel01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\Intel01.html
11
12         OPTIONS LS=111 PS=256 NODATE NOCENTER NONUMBER;
13         DATA Intel; INFILE CARDS MISSOVER; LENGTH CHIP $ 16;
14          TITLE1 'Data from Intel Corporation';
15          TITLE2 'Increasing power of Intel computer processor chips over years';
16             INPUT CHIP $ 1-16 YEAR TRANS;
17               logTRANS = log(TRANS);
18               LABEL YEAR = 'Year of microchip introduction';
19               LABEL TRANS = 'Equivalent power of chip in 1000 transistors';
20               MIPS = trans * 0.01818181818;
21         CARDS;

NOTE: The data set WORK.INTEL has 12 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
34         ;
35         PROC PRINT DATA=Intel; TITLE3 'Raw data Listing'; RUN;
NOTE: There were 12 observations read from the data set WORK.INTEL.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Raw data Listing

Obs	CHIP	YEAR	TRANS	log TRANS	MIPS
1	4004	1971	2.3	0.8329	0.042
2	8008	1972	2.5	0.9163	0.045
3	8080	1974	4.5	1.5041	0.082
4	8086	1978	29.0	3.3673	0.527
5	80286	1982	134.0	4.8978	2.436
6	80386	1985	275.0	5.6168	5.000
7	80486	1989	1200.0	7.0901	21.818
8	Pentium (5)	1993	3100.0	8.0392	56.364
9	Pentium Pro (6)	1995	5500.0	8.6125	100.000
10	Pentium II	1997	7500.0	8.9227	136.364
11	Pentium III	1999	9500.0	9.1590	172.727
12	Pentium 4	2000	42000.0	10.6454	763.636

```
37     options ps=52 ls=111;
38     proc plot data=intel; plot trans * year; TITLE3 'Plot of the raw data'; run;
NOTE: There were 12 observations read from the data set WORK.INTEL.
NOTE: The PROCEDURE PLOT printed page 2.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
39         proc plot data=intel; plot logtrans * year;
       TITLE3 'Plot of the log transformed data'; run;
40
41         options ps=512 ls=111;
NOTE: There were 12 observations read from the data set WORK.INTEL.
NOTE: The PROCEDURE PLOT printed page 3.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

```
Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Plot of the raw data

                            Plot of TRANS*YEAR.  Legend: A = 1 obs, B = 2 obs, etc.

       TRANS |
             |
       50000 +
             |
             |
             |
             |
             |
             |                                                                                      A
       40000 +
             |
             |
             |
             |
             |
             |
       30000 +
             |
             |
             |
             |
             |
             |
       20000 +
             |
             |
             |
             |
             |
             |
       10000 +                                                                                   A
             |
             |                                                                              A
             |                                                                        A
             |
             |                                                                  A
             |                                                       A
           0 +     A  A    A          A           A       A
             |
             ---+-------------+-------------+-------------+-------------+-------------+-------------+--
              1970          1975          1980          1985          1990          1995          2000
                                                        YEAR
```

```
Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Plot of the log transformed data

                          Plot of logTRANS*YEAR.  Legend: A = 1 obs, B = 2 obs, etc.

      logTRANS |
               |
            12 +
               |
               |
               |
               |                                                                                       A
               |
            10 +
               |
               |
               |                                                                               A    A
               |                                                                         A
               |
             8 +                                                                   A
               |
               |
               |                                                        A
               |
               |
             6 +
               |                                            A
               |
               |                                     A
               |
               |
             4 +
               |
               |                         A
               |
               |
               |
             2 +
               |             A
               |
               |        A
               |     A
               |
             0 +
               |
               ---+-------------+-------------+-------------+-------------+-------------+-------------+-
                1970          1975          1980          1985          1990          1995          2000
                                                          YEAR
```

```
42         PROC REG DATA=Intel lineprinter;  ID YEAR;
43            TITLE3 'Computer Chip example using REG with CLM';
44            MODEL logTRANS = YEAR / CLI CLM CLB;
45               TEST YEAR=0.462;
46            output out=next r=resid;
47         RUN;
47       !      OPTIONS PS=45;
NOTE: The data set WORK.NEXT has 12 observations and 6 variables.
NOTE: The PROCEDURE REG printed pages 4-6.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.04 seconds
      cpu time            0.04 seconds
```

Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Computer Chip example using REG with CLM

The REG Procedure
Model: MODEL1
Dependent Variable: logTRANS

Number of Observations Read	12
Number of Observations Used	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	131.29985	131.29985	850.47	<.0001
Error	10	1.54386	0.15439		
Corrected Total	11	132.84371			

Root MSE	0.39292	R-Square	0.9884
Dependent Mean	5.80034	Adj R-Sq	0.9872
Coeff Var	6.77408		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-632.78864	21.89772	-28.90	<.0001	-681.57980	-583.99749
YEAR	1	0.32150	0.01102	29.16	<.0001	0.29694	0.34607

Output Statistics

Obs	YEAR	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean		95% CL Predict		Residual
1	1971	0.8329	0.8974	0.2028	0.4455	1.3493	-0.0878	1.8826	-0.0645
2	1972	0.9163	1.2189	0.1938	0.7872	1.6506	0.2427	2.1950	-0.3026
3	1974	1.5041	1.8619	0.1764	1.4689	2.2549	0.9023	2.8215	-0.3578
4	1978	3.3673	3.1479	0.1454	2.8240	3.4719	2.2144	4.0814	0.2194
5	1982	4.8978	4.4339	0.1227	4.1605	4.7074	3.5168	5.3511	0.4639
6	1985	5.6168	5.3985	0.1143	5.1439	5.6530	4.4867	6.3102	0.2183
7	1989	7.0901	6.6845	0.1174	6.4229	6.9461	5.7707	7.5982	0.4056
8	1993	8.0392	7.9705	0.1357	7.6682	8.2728	7.0443	8.8967	0.0687
9	1995	8.6125	8.6135	0.1489	8.2817	8.9453	7.6773	9.5497	-0.001002
10	1997	8.9227	9.2565	0.1640	8.8910	9.6220	8.3078	10.2052	-0.3339
11	1999	9.1590	9.8995	0.1806	9.4971	10.3020	8.9360	10.8631	-0.7405
12	2000	10.6454	10.2210	0.1893	9.7992	10.6429	9.2492	11.1928	0.4244

Sum of Residuals	-5.6503E-13
Sum of Squared Residuals	1.54386
Predicted Residual SS (PRESS)	2.30901

Test 1 Results for Dependent Variable logTRANS

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	25.07333	162.41	<.0001
Denominator	10	0.15439		

```
48         TITLE3 'Plot of residuals';
49         Proc plot; PLOT resid*YEAR / vref=0;
50         RUN;
NOTE: There were 12 observations read from the data set WORK.NEXT.
NOTE: The PROCEDURE PLOT printed page 7.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds
51         PROC UNIVARIATE DATA=NEXT NORMAL PLOT; VAR resid;
52         RUN;
NOTE: The PROCEDURE UNIVARIATE printed pages 8-10.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Plots of raw data & residuals

                          Plot of resid*YEAR.  Legend: A = 1 obs, B = 2 obs, etc.

      resid |
            |
        0.6 +
            |
            |
            |                                 A
        0.4 +                                                   A                            A
            |
            |
            |
        0.2 +                      A                  A
            |
            |
            |                                                             A
        0.0 +-------------------------------------------------------------------A---------------
            |     A
            |
            |
       -0.2 +
            |
            |        A
            |            A                                                           A
       -0.4 +
            |
            |
            |
       -0.6 +
            |
            |
            |                                                                             A
       -0.8 +
            |
            ---+------------+------------+------------+------------+------------+------------+--
             1970         1975         1980         1985         1990         1995         2000

                                                    YEAR
```

```
Data from Intel Corporation
Increasing power of Intel computer processor chips over years
Plots of raw data & residuals

The UNIVARIATE Procedure
Variable:  resid

                            Moments
N                          12    Sum Weights                 12
Mean                        0    Sum Observations             0
Std Deviation      0.37463418    Variance            0.14035077
Skewness           -0.5243539    Kurtosis            -0.4472994
Uncorrected SS     1.54385842    Corrected SS        1.54385842
Coeff Variation             .    Std Error Mean      0.10814757

              Basic Statistical Measures
    Location                    Variability
Mean     0.000000     Std Deviation            0.37463
Median   0.033830     Variance                 0.14035
Mode      .           Range                    1.20437
                      Interquartile Range      0.63072

           Tests for Location: Mu0=0
Test           -Statistic-    -----p Value------
Student's t    t         0    Pr > |t|    1.0000
Sign           M         0    Pr >= |M|   1.0000
Signed Rank    S         3    Pr >= |S|   0.8501

                   Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.939064    Pr < W      0.4860
Kolmogorov-Smirnov    D     0.136633    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.040599    Pr > W-Sq  >0.2500
Anderson-Darling      A-Sq  0.292949    Pr > A-Sq  >0.2500


             Extreme Observations
-------Lowest------        ------Highest-----
     Value      Obs             Value      Obs
-0.7404772       11          0.218314        6
-0.3578260        3          0.219373        4
-0.3338563       10          0.405601        7
-0.3026030        2          0.424396       12
-0.0644798        1          0.463898        5


   Stem Leaf                     #  Boxplot                      Normal Probability Plot
      4 126                      3     |          0.5+                                     *  *++++*
      2 22                       2  +-----+          |                                * *+++++
      0 7                        1  *--+--*          |                             +*++++
     -0 60                       2  |     |      -0.1+                        ++*+*
     -2 630                      3  +-----+          |                  *++*++*
     -4                                |             |             ++++++
     -6 4                        1     |         -0.7+    +++++*
        ----+----+----+----+                          +----+----+----+----+----+----+----+----+----+----+
    Multiply Stem.Leaf by 10**-1                           -2        -1         0        +1        +2
```