Looking for Patterns and Applying the Method of Least Squares to Real Data

Note: You may notice differences between this Maple worksheet and the equivalent Mathematica notebook. These differences were introduced to preserve the content of these modules and were necessary because of major functional differences between Maple and Mathematica.

Introduction

OBJECTIVE: Minimize the sum of squared residuals to fit an arbitrary function to a set of data.

We look for patterns in real data sets including medical, socio-economic, meteorological, and pollution statistics. Does the least squares method apply only to finding lines of best fit, or does it apply to any type of curve that you wish to fit to a set of data points? We begin by analyzing data with standard fit functions and then utilize the calculus definition of best fit to find a nonstandard fit function.

Technology Guidelines

NOTE: If you have just finished a worksheet, **restart** *Maple* before executing a new worksheet.
TO OPEN SECTIONS,
Click on the **PLUS** sign at the left hand side of the screen *or* select **Expand All Sections** from the **View** drop down menu.

TO STOP AN EXECUTION
Click on **STOP** button from the toolbar.

ORDER OF EXECUTION
Execute commands in the order given. Do not skip any *Maple* Input lines within a given worksheet

Alternatively, you can execute the entire worksheet by selecting the **Execute Worksheet** command from the **Edit** drop down menu.

SAVING WORKSHEETS.

You can save anytime to any directory you choose, and it is wise to save often.
EXPERIENCING MAJOR PROBLEMS
Save if appropriate, and then shut down *Maple* and start it up again.

Time Series Data

Thirty-Two Years of Salaries: Men Versus Women

The following data represent the mean income for men and women in the United States from 1967 to 1998. The study includes people 15 years old and over. All income is in 1998 consumer price index adjusted dollars.

SOURCE: March Current Population Survey
PREPARED BY:
Income Statistics Branch/HHES Division, U.S. Bureau of the Census, U.S. Department of Commerce
Washington, D.C. 20233-8500, (301) 457-3242

First, we load in the **plots** and **stats** package.

```
> restart:
  with(plots):
  with(stats):
```

Warning, the name changecoords has been redefined

```
> menearnings := [27232, 28209, 29338, 29290, 29389, 31214,
      31587, 30495, 29784, 30168, 30635, 31180, 31039, 29916, 29419, 29180,
      29182, 30027, 30805, 31956, 32210, 32635, 33324, 31978, 31074, 30670,
      32143, 32887, 33126, 33553, 34794, 36315];
  womenearnings := [11500, 11631, 11997, 12195, 12412, 12930,
       12918, 12868, 12889, 13172, 13437, 13313, 13070, 13207, 13253, 13758,
      14148, 14805, 15174, 15729, 16301, 16703, 17119, 17085, 17027, 17070,
      17506, 17846, 18183, 18790, 19511, 20462];
  years := [1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976,
      1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988,
      1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998];
```

$$menearnings := [27232, 28209, 29338, 29290, 29389, 31214, 31587, 30495, 29784, 30168, 30635$$
$$31039, 29916, 29419, 29180, 29182, 30027, 30805, 31956, 32210, 32635, 33324, 31978, 31074, 3$$
$$32887, 33126, 33553, 34794, 36315]$$

$$womenearnings := [11500, 11631, 11997, 12195, 12412, 12930, 12918, 12868, 12889, 13172, 134$$
$$13070, 13207, 13253, 13758, 14148, 14805, 15174, 15729, 16301, 16703, 17119, 17085, 17027, 1$$
$$17846, 18183, 18790, 19511, 20462]$$

$$years := [1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1$$
$$1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 199$$

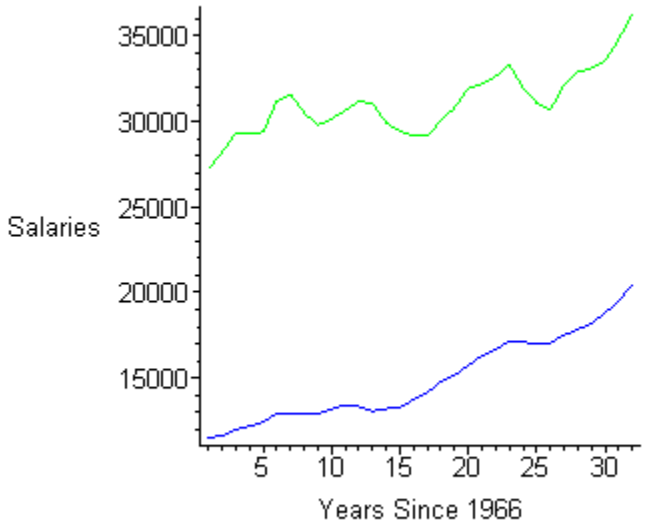
```
> data1:=[seq([i,menearnings[i]],i=1..nops(years))]:
  p1:=pointplot(data1, style=LINE, color=green,labels=["Years Since 1966", "Salaries"]):
  data2:=[seq([i,womenearnings[i]],i=1..nops(years))]:
  p2:=pointplot(data2,style=LINE, color=blue, labels=["Years Since 1966", "Salaries"]):
  data3:=[seq([years[i],menearnings[i],womenearnings[i]],i=1..nops(years))]:
```

```
> matrix([[`Year`,` Men's $`,` Women's $`],op(data3)]);
```

Year	*Men's $*	*Women's $*
1967	27232	11500
1968	28209	11631
1969	29338	11997
1970	29290	12195
1971	29389	12412
1972	31214	12930
1973	31587	12918
1974	30495	12868
1975	29784	12889
1976	30168	13172
1977	30635	13437
1978	31180	13313
1979	31039	13070
1980	29916	13207
1981	29419	13253
1982	29180	13758
1983	29182	14148
1984	30027	14805
1985	30805	15174
1986	31956	15729
1987	32210	16301
1988	32635	16703
1989	33324	17119
1990	31978	17085
1991	31074	17027
1992	30670	17070
1993	32143	17506
1994	32887	17846
1995	33126	18183
1996	33553	18790

As was probably expected, men's mean income exceeds women's mean income, but we can get a better perspective if we visualize these data through plots.

```
> print(plots[display]({p1,p2}));
```



Men's salaries are in green and women's are in blue.

Neither data set follows a straight line, but both have an upward trend. One way we can examine those trends is to compute the best-fit lines for each data set. We will do this using the **fit [leastsquares]( )** function in *Maple*.

```
> year:=[seq(i,i=1..nops(years))]:
  fit[leastsquare[[x,y]]]([year,menearnings]):
  fmen:=evalf(%);
  fit[leastsquare[[x,y]]]([year,womenearnings]):
  fwomen:=evalf(%);
```

$$fmen := y = 28386.89718 + 163.5857771\,x$$

$$fwomen := y = 10641.35484 + 260.3894795\,x$$

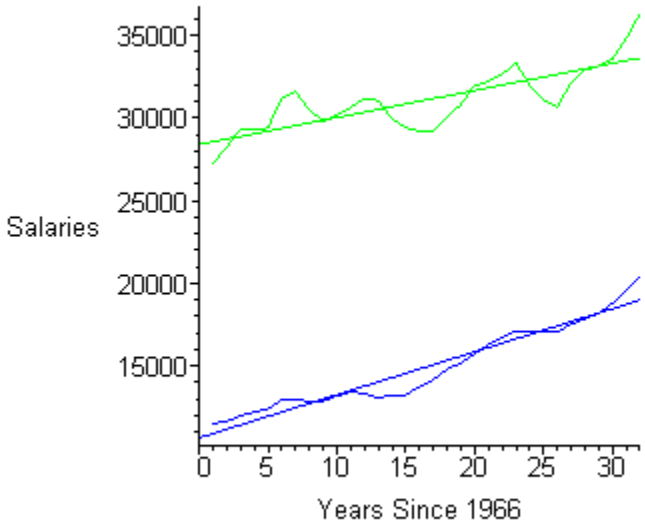
```
> print(`The average salary of males is approximately `, rhs(fmen));
  print(`The average salary of females is approximately `, rhs(fwomen));
  print(`where x represents the years since 1966`);
```

$$\text{The average salary of males is approximately }, 28386.89718 + 163.5857771\,x$$

$$\text{The average salary of females is approximately }, 10641.35484 + 260.3894795\,x$$

where x represents the years since 1966

```
> p3:=plot(rhs(fmen), x=0..nops(years),  color=green):
  p4:=plot(rhs(fwomen),x=0..nops(years),color=blue):
  print(plots[display]({p1,p2,p3,p4}, labels=["Years Since 1966", "Salaries"]));
```



What do these lines tell you about the rate at which salaries are increasing for men versus women? You could verify that these lines would intersect in about 183 years. Why could this be considered a worthless estimate?

You Try It: Part I

Lake Pollution

The following data sets represent weekly readings of aluminum pollutant levels in Lake Erie over a two-year period. What observations can you make about the amounts of aluminum pollutant? Does the amount of pollution follow a pattern? The first data set represents the data in the order that they were recorded. We then order those data to better analyze the distribution of the amounts of pollutant.

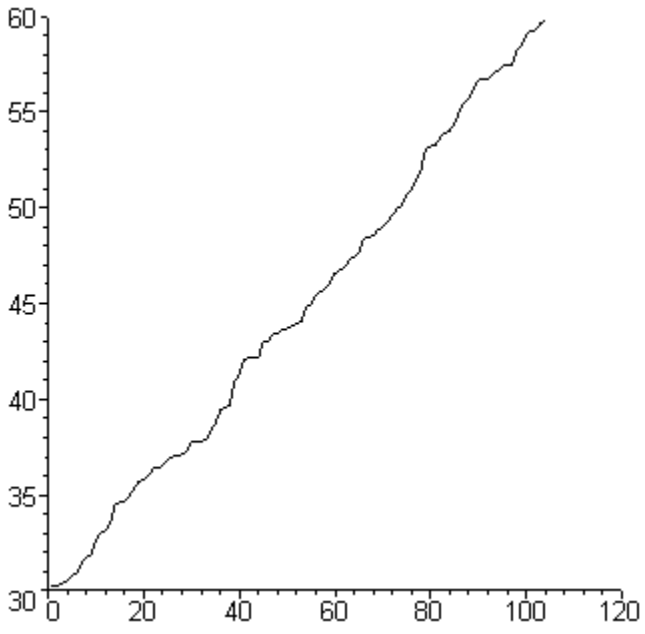
```
> pollutant := [59.7390, 43.6978, 45.4508, 55.6974, 43.6578, 56.6556, 49.9503,
      59.2501, 59.2778, 30.8085, 50.9860, 59.0104, 57.1471, 48.4696, 47.4461,
      37.0215, 52.0141, 53.8839, 35.8155, 43.8406, 58.4949, 56.7300, 45.6032,
      37.7449, 45.7960, 44.0860, 55.0693, 30.5343, 57.3854, 30.2706, 42.1799,
      44.9115, 40.9325, 57.4171, 30.2881, 46.1233, 36.4139, 46.7199, 36.3994,
      42.1318, 54.0219, 34.4747, 49.2455, 43.0440, 58.2225, 57.4492, 53.2136,
      30.9330, 48.8363, 32.6732, 53.6791, 31.7362, 35.7099, 39.6966, 48.3676,
      42.1876, 53.2427, 47.2459, 31.4635, 43.3965, 44.8305, 46.8731, 39.5781,
      37.0135, 37.8624, 37.1184, 36.5494, 41.2847, 48.9843, 50.6743, 48.5370,
      33.1112, 54.3590, 56.7471, 34.9079, 43.4007, 34.6556, 42.0995, 55.4403,
```

```
    42.9701, 56.2188, 39.4222, 33.5046, 38.6973, 59.6348, 38.3582, 46.5821,
    36.8290, 35.3832, 56.9938, 33.0390, 37.8167, 51.4268, 37.3356, 30.3771,
    37.7568, 43.9597, 50.1189, 47.6849, 53.0739, 34.6360, 36.0122, 31.8986,
    49.6269]:
sorted:=sort(pollutant);
```

```
sorted := [30.2706, 30.2881, 30.3771, 30.5343, 30.8085, 30.9330, 31.4635, 31.7362, 31.8986, 32.
33.0390, 33.1112, 33.5046, 34.4747, 34.6360, 34.6556, 34.9079, 35.3832, 35.7099, 35.8155, 36.0
36.4139, 36.5494, 36.8290, 37.0135, 37.0215, 37.1184, 37.3356, 37.7449, 37.7568, 37.8167, 37.8
38.6973, 39.4222, 39.5781, 39.6966, 40.9325, 41.2847, 42.0995, 42.1318, 42.1799, 42.1876, 42.9
43.3965, 43.4007, 43.6578, 43.6978, 43.8406, 43.9597, 44.0860, 44.8305, 44.9115, 45.4508, 45.6
46.1233, 46.5821, 46.7199, 46.8731, 47.2459, 47.4461, 47.6849, 48.3676, 48.4696, 48.5370, 48.8
49.2455, 49.6269, 49.9503, 50.1189, 50.6743, 50.9860, 51.4268, 52.0141, 53.0739, 53.2136, 53.2
53.8839, 54.0219, 54.3590, 55.0693, 55.4403, 55.6974, 56.2188, 56.6556, 56.7300, 56.7471, 56.9
57.3854, 57.4171, 57.4492, 58.2225, 58.4949, 59.0104, 59.2501, 59.2778, 59.6348, 59.7390]
```

Plot the ordered data, and see what the plot tells you about the pattern of the data.

```
> p1:=listplot(sorted, view=[0..120, 30..60]):
  print(p1);
```



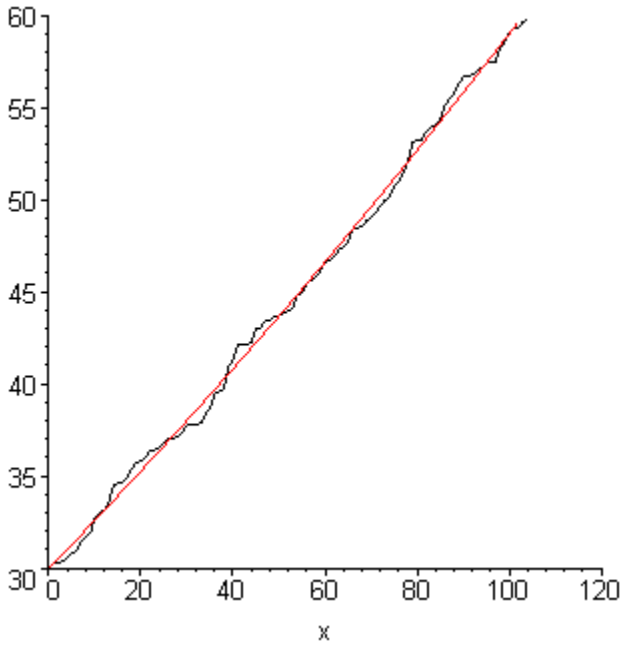
Look for fit functions for these data. Replace the equation in the **fit[leastsquare]** function with an appropriate expression.

```
> data1:=[[seq(i,i=1..nops(sorted))],sorted]:
```

```
> fit[leastsquare[[x,y],y=a+b*x+c*x^2, {a,b,c}]](data1):
  pollutantfit:=evalf(rhs(%));
  fitplot:=plot(pollutantfit,x=0..102, color=red):
```

```
print(plots[display]({fitplot,p1}));
```

$$pollutantfit := 29.98855475 + 0.2572804887\,x + 0.0003252390699\,x^2$$



The good fit for the linear model could imply that the pollutant distribution is somewhat uniform. What would uniform mean in this context?

Looking for Cause and Effect: Temperature as a Function of Latitude

Following are data showing 56 cities in the U. S., their January temperature (in degrees Fahrenheit) over a thirty-year period, and their latitudes.

```
> city := [MobileAL, MontgomeryAL, PhoenixAZ, LittleRockAR,
LosAngelesCA,SanFranciscoCA, DenverCO, NewHavenCT, WilmingtonDE, WashingtonI
JacksonvilleFL, KeyWestFL, MiamiFL, AtlantaGA, BoiseID, ChicagoIL,IndianapolisIN,
DesMoinesIA, WichitaKS, LouisvilleKY, NewOrleansLA, PortlandME, BaltimoreMD,
BostonMA, DetroitMI, MinneapolisMN, StLouisMO,HelenaMT, OmahaNE, ConcordNH,
AtlanticCityNJ, AlbuquerqueNM, AlbanyNY, NewYorkNY, CharlotteNC, RaleighNC,
BismarckND, CincinnatiOH, ClevelandOH, OklahomaCityOK, PortlandOR, HarrisburgP
PhiladelphiaPA, CharlestonSC, NashvilleTN, AmarilloTX, GalvestonTX,
HoustonTX,SaltLakeCityUT, BurlingtonVT, NorfolkVA, SeattleWA, SpokaneWA, Madis
MilwaukeeWI, CheyenneWY];

januarytemp:=[44, 38, 35, 31, 47, 42, 15, 22, 26, 30, 45, 65, 58, 37, 22, 19, 21, 11, 22, 27, 45, 1
25, 23, 21, 2, 24, 8, 13, 11, 27, 24, 14, 27, 34, 31, 0, 26, 21, 28, 33, 24, 24, 38, 31, 24, 49, 44, 18
32, 33, 19, 9, 13, 14];

latitude :=[31.2, 32.9, 33.6, 35.4, 34.3, 38.4, 40.7, 41.7, 40.5, 39.7, 31, 25, 26.3, 33.9, 43.7, 42.3
```

39.8, 41.8, 38.1, 39, 30.8, 44.2, 39.7, 42.7, 43.1, 45.9, 39.3, 47.1, 41.9, 43.5, 39.8, 35.1, 42.6, 40
35.9, 36.4, 47.1, 39.2, 42.3, 35.9, 45.6, 40.9, 40.9, 33.3, 36.7, 35.6, 29.4, 30.1, 41.1, 45, 37, 48.1
43.4, 43.3, 41.2];

city := [*MobileAL*, *MontgomeryAL*, *PhoenixAZ*, *LittleRockAR*, *LosAngelesCA*, *SanFranciscoCA*, *L*
NewHavenCT, *WilmingtonDE*, *WashingtonDC*, *JacksonvilleFL*, *KeyWestFL*, *MiamiFL*, *AtlantaG*
ChicagoIL, *IndianapolisIN*, *DesMoinesIA*, *WichitaKS*, *LouisvilleKY*, *NewOrleansLA*, *PortlandME*,
BostonMA, *DetroitMI*, *MinneapolisMN*, *StLouisMO*, *HelenaMT*, *OmahaNE*, *ConcordNH*, *Atlantic*
AlbuquerqueNM, *AlbanyNY*, *NewYorkNY*, *CharlotteNC*, *RaleighNC*, *BismarckND*, *CincinnatiOH*,
OklahomaCityOK, *PortlandOR*, *HarrisburgPA*, *PhiladelphiaPA*, *CharlestonSC*, *NashvilleTN*, *Amc*
GalvestonTX, *HoustonTX*, *SaltLakeCityUT*, *BurlingtonVT*, *NorfolkVA*, *SeattleWA*, *SpokaneWA*, *M*
MilwaukeeWI, *CheyenneWY*]

januarytemp := [44, 38, 35, 31, 47, 42, 15, 22, 26, 30, 45, 65, 58, 37, 22, 19, 21, 11, 22, 27, 45, 12
24, 8, 13, 11, 27, 24, 14, 27, 34, 31, 0, 26, 21, 28, 33, 24, 24, 38, 31, 24, 49, 44, 18, 7, 32, 33, 19, 9

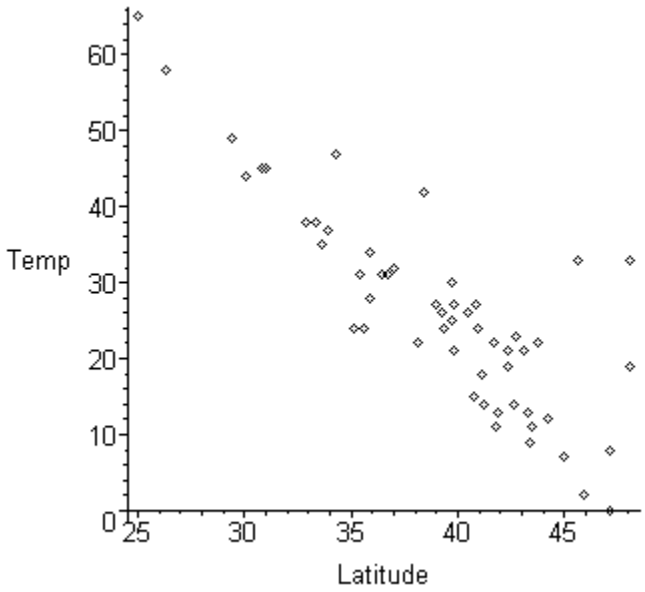
latitude := [31.2, 32.9, 33.6, 35.4, 34.3, 38.4, 40.7, 41.7, 40.5, 39.7, 31, 25, 26.3, 33.9, 43.7, 42.3,
38.1, 39, 30.8, 44.2, 39.7, 42.7, 43.1, 45.9, 39.3, 47.1, 41.9, 43.5, 39.8, 35.1, 42.6, 40.8, 35.9, 36.4
42.3, 35.9, 45.6, 40.9, 40.9, 33.3, 36.7, 35.6, 29.4, 30.1, 41.1, 45, 37, 48.1, 48.1, 43.4, 43.3, 41.2]

> **tabledata:=[seq([city[i], latitude[i], januarytemp[i]], i=1..nops(januarytemp))]:**
matrix([[`City `,`Latitude`, ` Jan Temp.`],op(tabledata)]);

City	Latitude	Jan Temp.
MobileAL	31.2	44
MontgomeryAL	32.9	38
PhoenixAZ	33.6	35
LittleRockAR	35.4	31
LosAngelesCA	34.3	47
SanFranciscoCA	38.4	42
DenverCO	40.7	15
NewHavenCT	41.7	22
WilmingtonDE	40.5	26
WashingtonDC	39.7	30
JacksonvilleFL	31	45
KeyWestFL	25	65
MiamiFL	26.3	58
AtlantaGA	33.9	37
BoiseID	43.7	22
ChicagoIL	42.3	19
IndianapolisIN	39.8	21
DesMoinesIA	41.8	11
WichitaKS	38.1	22
LouisvilleKY	39	27
NewOrleansLA	30.8	45
PortlandME	44.2	12
BaltimoreMD	39.7	25
BostonMA	42.7	23
DetroitMI	43.1	21
MinneapolisMN	45.9	2
StLouisMO	39.3	24
HelenaMT	47.1	8
OmahaNE	41.9	13
ConcordNH	43.5	11

First, do a scatterplot to see if there is a relationship between the January temperatures and latitude.

```
> scatterdata:=[seq([latitude[i], januarytemp[i]], i=2..nops(januarytemp))]:
  scatterplot:=pointplot(scatterdata, labels=["Latitude", "Temp"]):
  print(scatterplot);
```



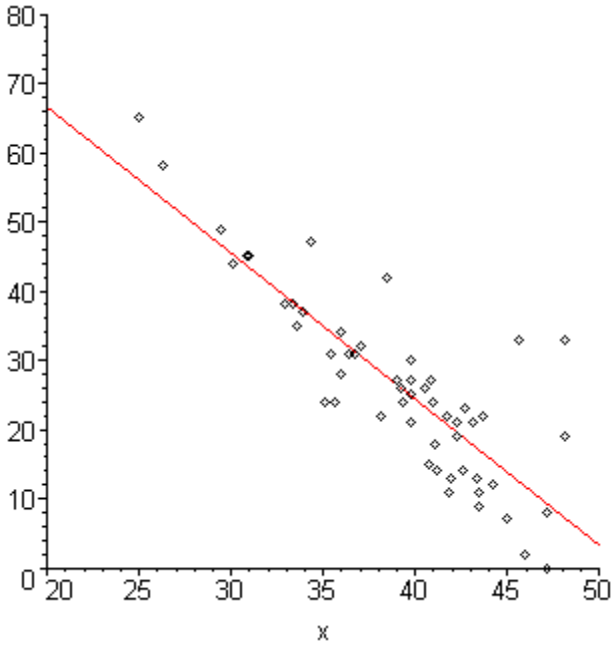
Since the scatterplot appears to have a somewhat linear pattern, that shows the average January temperature dropping as the latitude increases, we will investigate this relationship further. We will use *Maple's* **fit[leastsquare]( )** function to get results. Note that you choose a linear fit by specifying the (x,y) function. We can write the line of best fit as follows.

```
> fit[leastsquare[[x,y]]]([latitude,januarytemp]):
  ffit:=rhs(evalf(%));
```

$$ffit := 108.7277422 - 2.109587848\,x$$

Now we can plot this line with our scatterplot.

```
> yfit:=plot(ffit, x=0..50):
  print(plots[display]({yfit, scatterplot}, view=[20..50, 0..80]));
```



Can you identify cities that deviate from the pattern? Might you speculate on the reason for the deviation? Name one important factor that might explain at least a part of this deviation.

You Try It: Part II

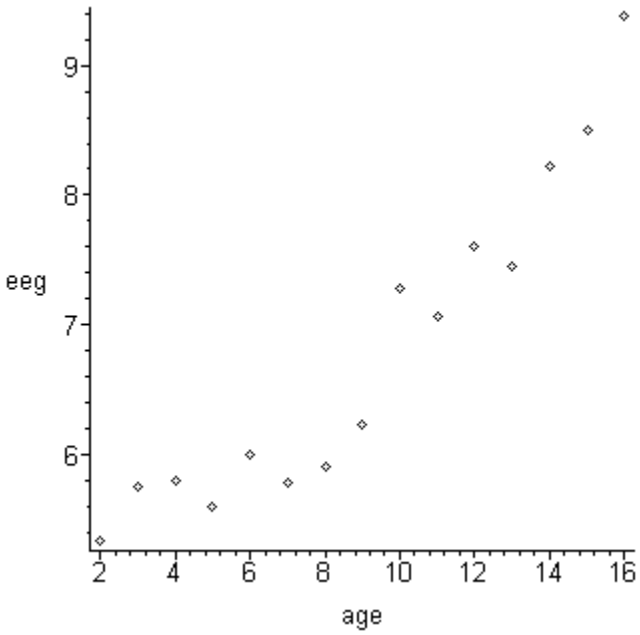
Here are two data sets you can explore. You can enter your own data sets and then re-execute the computations to explore relationships.

Age and EEG

The electroencephalogram (EEG) is a device used to measure brain waves. Neurologists have found that the peak EEG frequency in children increases with age. The data below represent the results of a study of children age 2 to16, and the EEG readings (in hertz) represent the average peak EEG frequencies for each age group. It would be reasonable to think of *age* as the independent variable and *eeg* as the dependent variable.

```
> age:=[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]:
  eeg:=[5.33, 5.75, 5.80, 5.60, 6.00, 5.78, 5.90, 6.23, 7.28, 7.06, 7.60, 7.45,8.23, 8.50, 9.38]
  tabledata:=[seq([age[i], eeg[i]], i=1..nops(age))]:
  matrix(tabledata);
  eegdata:=[seq([age[i],eeg[i]],i=1..nops(age))]:
  scatter:=pointplot(eegdata, labels=["age", "eeg"]):
  print(scatter);
```

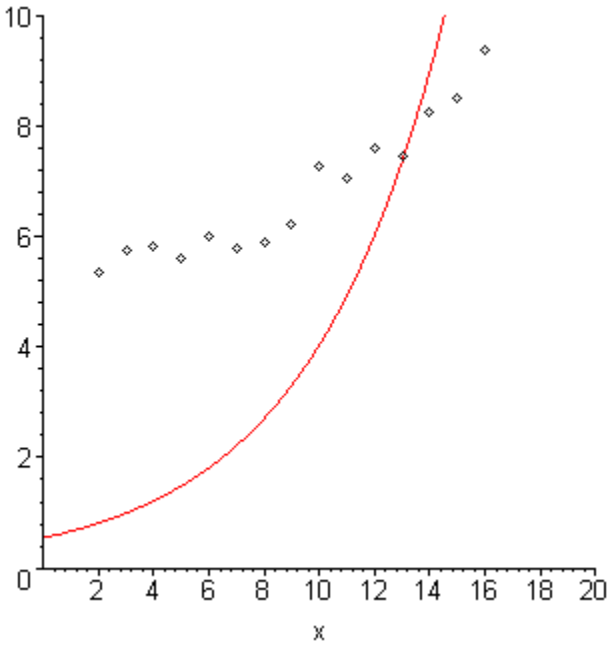
2	5.33
3	5.75
4	5.80
5	5.60
6	6.00
7	5.78
8	5.90
9	6.23
10	7.28
11	7.06
12	7.60
13	7.45
14	8.23
15	8.50
16	9.38




```
> fit[leastsquare[[x,y], y=a*exp(0.2*x), {a}]]([age,eeg]):
  agefit:=rhs(evalf(%));
```

$$agefit := 0.5461793856\, e^{(0.2\,x)}$$

```
> ageplot:=plot(agefit, x=0..20):
  print(plots[display]({ageplot, scatter}, view=[0..20, 0..10]));
```



Alter the function in **fit[leastsquare]( )** until you get a good fit.

Reliability of Construction Material

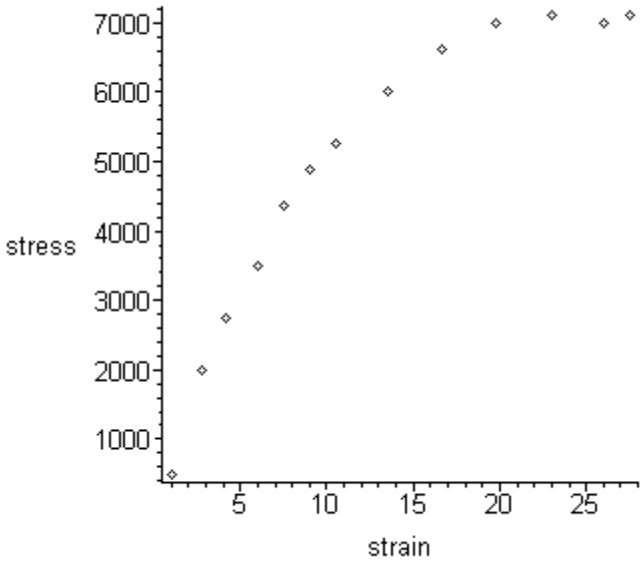
The Canadian Geotechnical Journal (Aug., 1985) reported on a study that was conducted to investigate the reliability of the use of fragmented Queenston Shale, a compaction shale, as a rockfill construction material. In particular, the researchers wanted to estimate the stress-strain relationship of the fragmented material. Their study yielded the following results, where axial strain (x) is given as percents and deviatoric stress (y) is given in kPa. You might want to consider a quadratic fit for these data.

```
> strain:=[1.0, 2.8, 4.2, 6.0, 7.5, 9.0, 10.5, 13.5, 16.7, 19.8, 23.0, 26.0, 27.5];
  stress:=[500, 2000, 2750, 3500, 4375, 4875, 5250, 6000, 6625, 7000, 7125, 7000, 7125];
```

$$strain := [1.0, 2.8, 4.2, 6.0, 7.5, 9.0, 10.5, 13.5, 16.7, 19.8, 23.0, 26.0, 27.5]$$

$$stress := [500, 2000, 2750, 3500, 4375, 4875, 5250, 6000, 6625, 7000, 7125, 7000, 7125]$$

```
> ss:=[seq([strain[i],stress[i]],i=1..nops(strain))]:
  scatter:=pointplot(ss, labels=["strain","stress"]):
  print(scatter);
```

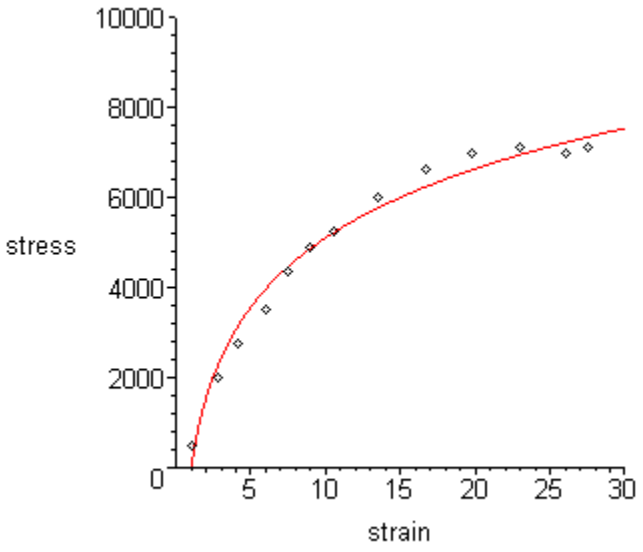


Alter the function in **fit[leastsquare]( )** until you get a good fit.

> **fit[leastsquare[[x,y], y=a*ln(x),{a}]]([strain,stress]):**
ssfit:=rhs(evalf(%));

$$ssfit := 2217.615053 \ln(x)$$

> **ssplot:=plot(ssfit, x=0..30):**
print(plots[display]({ssplot, scatter}, view=[0..30, 0..10000],labels=["strain","stress"]



Going Back to the Calculus to Determine the Curve of Best

Fit: Exponential Regression Coefficients

Data were collected concerning a specific genetic characteristic that researchers believed followed a particular exponential pattern. After observing a scatterplot of the data, they looked for a fit function of the form $y = a\,e^{(b\,x)} + c$. Because this is not a standard fit function, it is necessary to use the minimization techniques from multivariable calculus to determine a, b, and c from scratch. In all, there were 2723 data sets collected, and those data sets were placed into eight categories. In the set defined below, the first entry represents the frequency, the second, the categorical variable (x) and the third the observed genetic measurement (y).

```
> dataset := [[579, 1, 38.08], [1021, 2, 29.70], [607, 3, 25.42], [324, 4,23.15], [120, 5, 21.79], [46
  20.91], [17, 7, 19.37], [9, 8,19.36]];
```

$$dataset := [[579, 1, 38.08], [1021, 2, 29.70], [607, 3, 25.42], [324, 4, 23.15], [120, 5, 21.79], [4$$
$$[17, 7, 19.37], [9, 8, 19.36]]$$

```
> matrix([[`freq `,` count `,` observ`],op(dataset)]);
  y:=a*exp(b*x)+c;
```

$$\begin{bmatrix} freq & count & observ \\ 579 & 1 & 38.08 \\ 1021 & 2 & 29.70 \\ 607 & 3 & 25.42 \\ 324 & 4 & 23.15 \\ 120 & 5 & 21.79 \\ 46 & 6 & 20.91 \\ 17 & 7 & 19.37 \\ 9 & 8 & 19.36 \end{bmatrix}$$

$$y := a\,e^{(b\,x)} + c$$

We begin our process by writing the formula for the sum of the squares of the vertical distances between the curve of best fit and the observed data. This will be a function of a, b, and c, so we will find the first three partial derivatives and solve for the values of a, b, and c when we set those partial derivatives equal to 0.

```
> add(dataset[i,1]*(dataset[i,3]-a*exp(dataset[i,2]*b)-c)^2, i=1..nops(dataset)):
  g:=unapply(%, (a,b,c));
```

$$g := (a,b,c) \to 579\,(38.08 - a\,e^{b} - c)^2 + 1021\,(29.70 - a\,e^{(2b)} - c)^2 + 607\,(25.42 - a\,e^{(3b)} - c)^2 + 324\,(23.15 - a\,e^{(4b)} - c)^2 + 120\,(21.79 - a\,e^{(5b)} - c)^2 + 46\,(20.91 - a\,e^{(6b)} - c)^2 + 17\,(19.37 - a\,e^{(7b)} - c)^2 + 9\,(19.36 - a\,e^{(8b)} - c)^2$$

```
> g_a:=diff(g(a,b,c),a):
```

```
> g_b:=diff(g(a,b,c),b):
```

```
> g_c:=diff(g(a,b,c),c):
```

Now we will set up our equations to solve. Due to previous studies, we had an idea of a range of values for the parameters *a*, *b* and *c*. For that reason, we can use the command **solve** and give some initial starting values..

```
> solve({g_a=0, g_b=0, g_c=0, a<50, a>30, b<1, b>-1, c<30, c>10},{a,b,c}):
  sol:=%;
```

$$sol := \{a = 33.22210566,\ b = -0.6268550049,\ c = 20.29125400\}$$

```
> assign(sol);
```

```
> a;
  b;
  c;
```

$$33.22210566$$

$$-0.6268550049$$

$$20.29125400$$

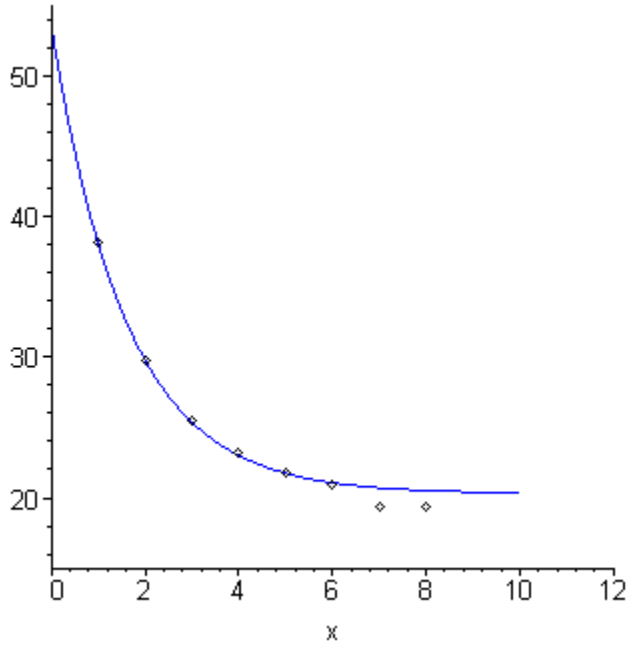
```
> yhat:=y;
```

$$yhat := 33.22210566\,e^{(-0.6268550049\,x)} + 20.29125400$$

Look at the results graphically.

```
> p1:=plot(yhat, x=0..10, view=[0..12, 15..55], color=blue):
  pts:=[seq([dataset[i,2],dataset[i,3]],i=1..nops(dataset))]:
  p2:=pointplot(pts):
```

```
print(plots[display]({p1,p2}));
```



This looks like a good fit. Compute the sum of the squares of the errors.

```
> yest:=[seq(eval(yhat, x=dataset[i,2]), i=1..nops(dataset))]:
  yact:=[seq(dataset[i,3],i=1..nops(dataset))]:
  wts:=[seq(dataset[i,1],i=1..nops(dataset))]:

> temp1:=(yact-yest):
  temp2:=[seq(temp1[i]^2,i=1..nops(dataset))]:
  temp3:=[seq(temp2[i]/yest[i], i=1..nops(dataset))]:
  sse1:=linalg[dotprod](wts, temp2):
  sse2:=linalg[dotprod](wts, temp3):
  print(`The sum of the squares of the vertical deviations for the 2723 data points is`, sse1);
  print(`The sum of the squares of the vertical deviations written as percentages for the 2723
  points is`, sse2);
```

The sum of the squares of the vertical deviations for the 2723 data points is, 59.96642967

The sum of the squares of the vertical deviations written as percentages for the 2723 data points i
2.740750961

```
> ?

>
```

This last entry is important for running a statistical test on how good a fit this model is. Because of our calculus work, we know that this is the best fit model for a function of this type.