was: Logistic Cost Curve by the Levenberg-Marquardt Method by J.M. Redwood

Procedure *mnlfit*

Problem

A method was needed for estimating the cash flows of engineering development projects undertaken by a certain company. One such project took 13 months to complete and the cumulative cost returns were collected throughout the life of the project. The accounts were closed 2 months after completion of the project, when the last bills were brought to account.

The cumulative costs were collected at the end of each month up to the final fixed price of £1,000,000. The data therefore comprises 16 pairs of (end) of month numbers and the cumulative costs in £k.

The cumulative logistic distribution function often fits data from growth situations that are limited by a finite resource. In this case, the costs are limited by the fixed price for the job. They grow slowly at first as just a few, then more and more people on the development team become involved. They then increase more rapidly as parts are bought in and manufacturing, assembly and test proceed, and then taper off as the manufacturing and development teams reduce with final evaluation and delivery to the customer, followed by settlement of the last bills from suppliers.

Data

The data recorded was

```
> cost := [[1,2],[2,11],[3,36],[4,87],[5,138],[6,234],[7,352],[8,489],
  [9,643],[10,750],[11,854],[12,924],[13,948],[14,975],[15,995],[16,1000]]:
  n := nops(cost);
```

$$n := 16$$

The shape of the plot is typical of the cumulative expenditure for a fixed price project, and the sigmoidal form suggests that the logistic equation should fit the data. (Note the familiar slow start because engineers and draughtsmen were still involved with a different project!)

The (cumulative) logistic distribution function is given by $f(t) = \dfrac{a_1}{1 + \mathbf{e}^{(a_2 - a_3 t)}}$

where a_1, a_2, a_3 are constants and *t* is the independent variable.

Fitting the Logistic Equation to the Data

Starting values for $p = [a_1, a_2, a_3]$ are chosen from previous experience, and all weights are set to unity.

```
> p := vector(3,[1000,1,1]); w := vector(n,1):
```

$$p := [1000, 1, 1]$$

Define logistic function:

```
> f := a[1]/(1+exp(a[2]-a[3]*t[1]));
```

$$f := \frac{a_1}{1 + \mathbf{e}^{(a_2 - a_3 t_1)}}$$

Run fitting procedure to get parameter estimation and statistics:

```
> costv := linalg[col](cost,2):
  month := matrix(n,1,linalg[col](cost,1)):
  f1 := mnlfit(f,t,month,costv,w,a,p,10^(-20),B,6);
```

No of Iterations	Std Deviation of Residuals
16	8.57642

Analysis of Covariance

Source	Σ(Squares)	DF	Mean Square
Regression	2393263	2	$0.119663\ 10^7$
Residuals	956	13	73.5550
Total	2394219	15	

Tests of Covariance

R^2	F	Prob of F by chance for Normal data
1.00	16300.	$0.811\ 10^{-22}$

Final values of parameters

Parameter	Value	Standard Error
1.	1006.72	5.30992
2.	4.78432	0.0799611
3.	0.592026	0.0107814

Matrix of Covariances

$$\begin{bmatrix} 0.383322 & -0.00280745 & -0.000486808 \\ -0.00280745 & 0.0000869251 & 0.0000113418 \\ -0.000486808 & 0.0000113418 & 0.158030\ 10^{-5} \end{bmatrix}$$

Matrix of Correlation Coefficients

$$\begin{bmatrix} 1.00000 & -0.486359 & -0.625469 \\ -0.486359 & 1.00000 & 0.967696 \\ -0.625469 & 0.967696 & 1.00000 \end{bmatrix}$$

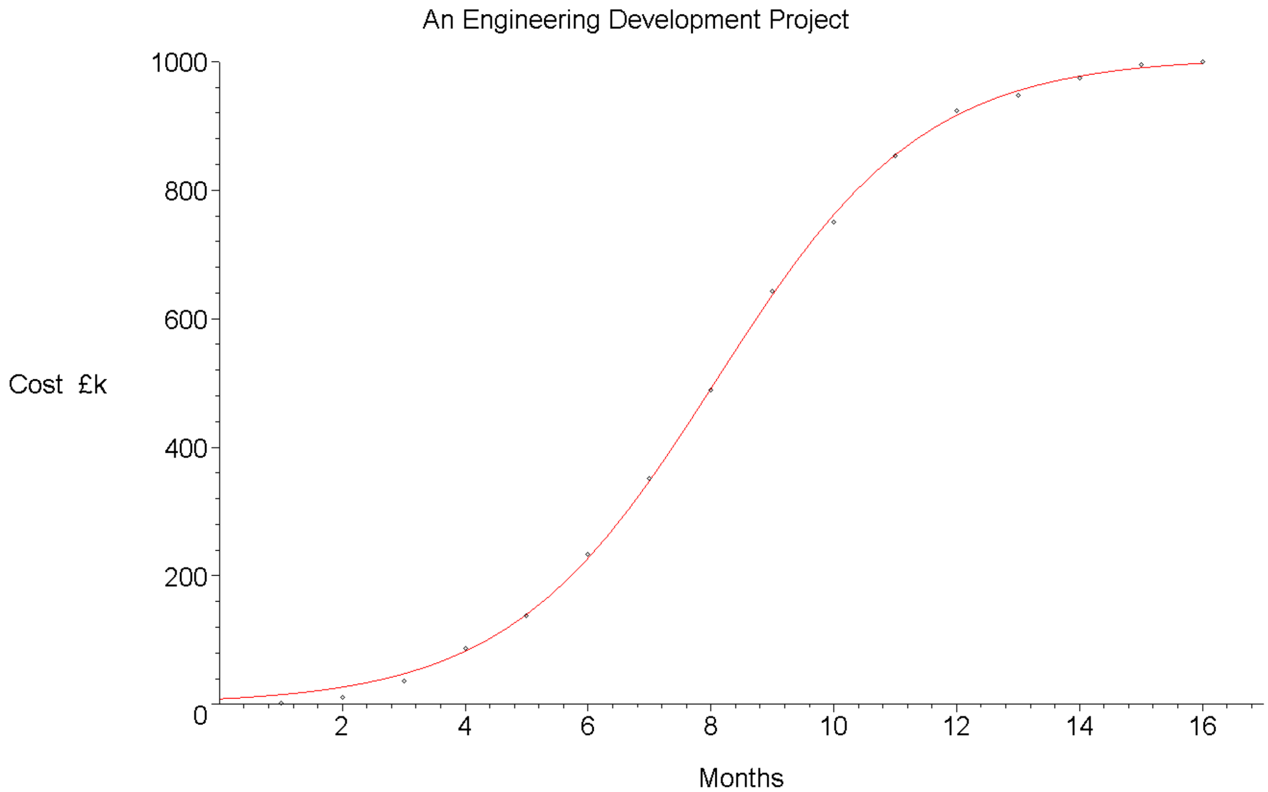
Table of residuals

$x\ obs_1$	y obs	y calc	difference	wt*difference / sd of residuals
1.	2.	14.9866	-12.9866	-1.51422
2.	11.	26.7687	-15.7687	-1.83861
3.	36.	47.3711	-11.3711	-1.32586
4.	87.	82.4950	4.5050	0.525277
5.	138.	139.865	-1.865	-0.217457
6.	234.	227.320	6.680	0.778880
7.	352.	347.535	4.465	0.520613
8.	489.	491.253	-2.253	-0.262697
9.	643.	636.974	6.026	0.702624
10.	750.	762.020	-12.020	-1.40152
11.	854.	854.858	-0.858	-0.100042
12.	924.	916.638	7.362	0.858400
13.	948.	954.811	-6.811	-0.794154
14.	975.	977.327	-2.327	-0.271325
15.	995.	990.245	4.755	0.554427
16.	1000.	997.539	2.461	0.286950

$$f1 := \frac{1006.72}{1 + \mathbf{e}^{(4.78432 - 0.592026\, t_1)}}$$

The last expressions gives the desired logistic function that fits the data.
Plot it against the data from which it was obtained.

```
> F1:=unapply(subs(t[1]=t,f1),t):
  aa := pointplot(cost,view=[0..17,0..1000],labels=[`Months`,`Cost
  £k`],title=`An Engineering Development Project`): bb := plot(F1(x),x=0..n):
  display(aa,bb);
```



By eye, the fit appears very good and this is supported by the relatively small *Std Deviation of Residuals* statistic.
However, one should plot the residuals to check that there is no clear trend:
Systematic error for small time, even if R^2 = 1.0 ...

```
> pts := zip((x,y)->[x,y],[seq(j,j=1..n)],[seq(cost[i,2]-F1(i),i=1..n)]):
  pointplot(pts,labels=["Month","Actual - Est Cost     £k"]);
```