Math 132

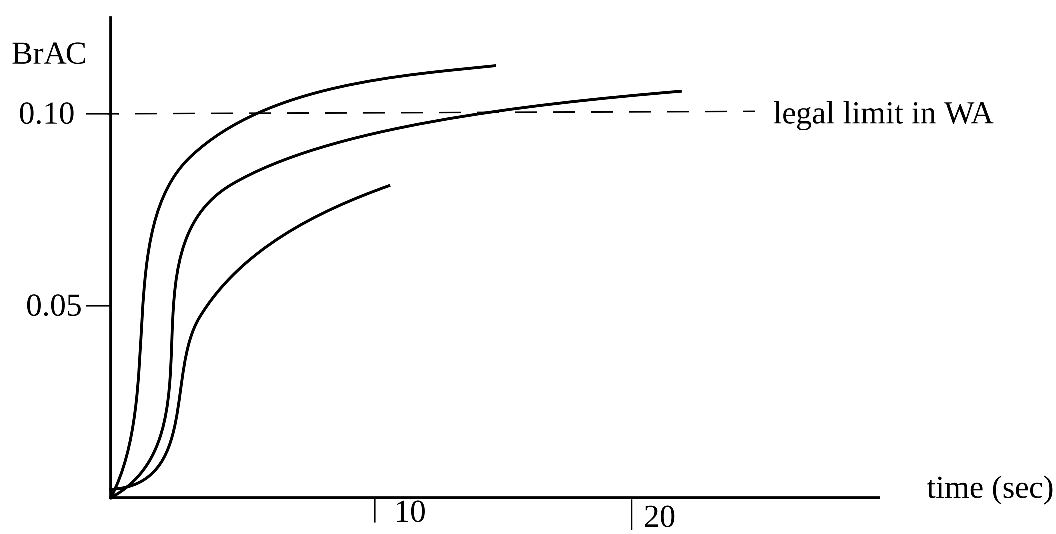
Alcohol Breath Testing and Least Squares Data Fitting

S. R. Lubkin

Application

Drunken driving kills over 15,000 people in the US every year. This is 38% of all traffic-related deaths. In the US there are 1.5 million DWI arrests every year, and each of these represents a blood-alcohol content (BAC) of a minimum of 0.08-0.10 (depending on the state), which is equivalent to a 170-pound man drinking 5-7 drinks in an hour.

One of the ways the State Patrol can measure blood alcohol content is by taking a sample of the driver's breath and measuring its alcohol content (BrAC), which is proportional to the blood alcohol level. (Alcohol has different solubilities in blood and in air.) The breath-analyzer gives a continuous reading of BrAC as the driver exhales, but the official reading is the maximum BrAC measured. Three typical BrAC profiles from the same person are graphed below.



These were each taken within minutes of each other. Notice that if the subject stops exhaling sooner, the official reading is lower. There are other factors which influence the BrAC profile and the maximum reading.

The variability in BrAC readings led a State Trooper to investigate mathematical models of alcohol exhalation[1]. His first step was to take BrAC profiles from theWashington State Alcohol Lab and find curves which would fit the data well. We will do the same task.

1. This is a true story. Only the data in this assignment is made up.

Mathematical Models

The trooper experimented with elementary functions that he knew and found two general functions that were simple yet could be made to fit the data pretty well:

$$\text{Model 1:} \quad B(t) = a(1 - e^{-ct})$$

$$\text{Model 2:} \quad B(t) = a(1 - e^{-ct}) + kt$$

The models contain two or three **parameters**, *a*, *c*, and *k*. A parameter is a number which is constant as the independent variable (time) changes, but may vary from “experiment” to “experiment”. So in the graph above, which has three curves, each curve will have different parameters, but these parameters will be constant along each individual curve.

Do this: To see what the general shape of these curves is, open up a Maple session, and plot the functions. You don’t need to know what the best parameter values are for this task, so just use any old value. A good value to start with for each parameter is 1. Try the following:

```
> model1:= a*(1 - exp(-c*t));
> model2:= a*(1 - exp(-c*t)) + k*t;
> p1:= subs(a=1, c=1, k=1, model1);
> plot(p1, t=0..10);
> p2:= subs(a=1, c=1, k=1, model2);
> plot(p2, t=0..10);
```

This choice of parameters looks OK for Model 1, but Model 2 doesn’t look much like a typical breath profile. Play with making c larger and k smaller in Model 2 till it looks more like the sample figure above. Remember to press return in Maple every time you change a command.

The task of data fitting is to determine the parameter values *a*, *c*, and *k* which will make all the measured data points $B_i(t_i)$ lie as close as possible to the curve *B(t)*. **“Close” means that the sum of the squares of the distances between each point $B_i(t_i)$ and the curve *B(t)* are as small as possible, within the constraint of the form of the function *B(t)*.**

We form the sum of squares *S* as

$$S = \sum_{i=1}^{n} (B_i(t_i) - B(t_i))^2$$

or, in the case of Model 1,

$$S = \sum_{i=1}^{n} (B_i(t_i) - a(1 - e^{-ct_i}))^2$$

We can then think of *S* as a function of the parameters *a*, *c*, and *k*. If we can minimize *S* by changing *a, c,* and *k,* we will have the version of Model 1 which fits the data best.

Methods of Solution

We’ll use the spreadsheet to organize our data and to form the sum of squares. Excel’s Solver function will minimize S for us. The Data Table will have the following columns:

- times of data points, t_i
- BrAC at those times, $B_i(t_i)$
- value of model function at those times, $B(t_i)$
- square of difference, $S_i(t_i)$

We will also construct a Parameter Table keeping track of our parameter values and the sum of squares. We can easily plot both the data and the fitted curve using the ChartWizard, and visually assess how good the fit is for different parameter values.

We will allow Solver to change the parameter values we are using until the least squares solution is found (the parameters which minimize S).

Implementation (step-by-step instructions)

1. Open the spreadsheet.
2. Copy the table headings and data below.

time (sec)	BrAC	B(t)	(BrAC-B(t))^2
1	.063		
2	.097		
3	.114		
4	.123		
5	.129		
6	.131		
7	.132		
8	.134		
9	.136		
10	.139		
11	.141		
12	.140		

3. Make a graph of the data using the ChartWizard and type “scatter” without the curve connecting the points. (By the way, Trendline can’t help you on this assignment. Why?)
4. Make a second table on the same spreadsheet, with the headings below, and initial guesses for

the three parameters.

a	c	k	S
1	1	0	

5. In the third column of the Data Table, evaluate Model 1 at the time indicated in the first column. Pull down the formula to all cells in the column, to evaluate *B(t)* at each time.
6. Change your graph to add the third column in the Data Table, or start a new graph. It is nicest if the actual data is in dots and the fitted function is in a line with no dots. Notice that the initial guess gives a terrible fit.
7. In the last column of the Data Table, form the square of the difference of the previous two columns.
8. In the Parameter Table, for the value of S, put the sum of the entries of the last column of the Data Table.
9. Select Solver from the Tools menu at the top of Excel[1]. In Solver's window, enter the address of the objective function *S* (the thing we want to optimize), tell it we want to *minimize* it, and enter the locations of the parameters that Solver is allowed to vary (of the Parameter Table). Remember, for Model 1, only *a* and *c* can be changed.
10. Click Solve. Watch as the parameter values in the Parameter Table change. Watch as the graph changes. Watch as the value of *S* decreases.
11. Repeat the process for Model 2. Be sure to tell Solver that it can change *a, c,* and *k*. Which model gives a lower least sum of squares *S*?

Assessment

The two models both can be made to fit the data closely. Model 2 gives a much better fit to most BrAC data, because the sum of squares is less. One major question at this point is, "What does a good fit to the data tell you?" If the data fitting doesn't help you accurately and reliably determine what the true breath alcohol content is (and hence the blood alcohol content), then it is of no help.

What the exercise in data fitting did actually do is lead to the construction of two mechanistic models using differential equations to describe the process of exhalation. The solutions of the two differential equation models turned out to be Models 1 and 2 above, but with the parameters *a*, *c*, and *k* representing various combinations of anatomical and chemical parameters - one of which was the BrAC. Use of the differential equation model and least squares data fitting gave a more reliable method for determining the true BrAC, no matter how the subject breathed into the machine.

---

1. If Solver isn't there under Tools, you need to go to Tools -> Add-Ins, scroll down, and click that Solver should be added. Then you can use Tools -> Solver. If Solver isn't in the Add-Ins, your installation of Excel will not work for this project, and you will either need to install Solver or find another computer.