

ACCESS - MATH

July 2001

Notes on Body Mass Index.

What is the Body Mass Index?

If you read newspapers and magazines it is likely that once or twice a year you run across an article about the body mass index (B.M.I.), and its use in determining health risk factors for overweight and underweight people. If you search the internet for "body mass index" you will find many sites which let you compute your B.M.I., and which tell you a little bit about it.

A person's B.M.I. is computed by dividing their weight by the square of their height, and then multiplying by a universal constant. If you measure weight in kilograms, and height in meters, this constant is the number one. Thus, the proponents of the B.M.I. index are claiming that for adults at equal risk levels (but different heights), weight should be proportional to the square of height. As we have discussed in class, if people were to scale equally in all directions ("self-similar") when they grew, volume and hence weight would scale as the cube of height. That particular power law seems a little high, since adults don't look like uniformly expanded versions of babies; we seem to get relatively stretched out when we grow taller. One might expect the best predictive power for weight as a function of height to be somewhere between 2 and 3, if one expected a power law at all. If there is a predictive power, and if it is much larger than 2, then one could argue that the body mass index might need to be modified to reflect this fact. (In fact, in some recent tables just such a correction has been made.)

What is our B.M.I. project?

We have all collected several heights and weights, and hopefully in aggregate we will have a good number of representative measurements, from baby-sized to adult. We will use this data, and see if it is consistent with a power law relating weights to heights. In particular, we will see if we get a power law with exponent near 2, as we might expect from the B.M.I.

I often do this experiment with my linear algebra classes, and we have gotten powers between 2.35 and 2.65. Also, last fall I found a national data base at the U.S. Center for Disease Control web site. It includes a wide variety of body measurements collected between 1976 and 1980. I will use national median heights and weights for boys and girls, age 2-19. By taking only the national medians I have taken a lot of the "noise" out of the data, compared to what ours will look like. From the national data a power law with power near 2.6 seems quite certain. However, I know of no mathematical model which explains this scaling.

How do you test for power laws?

When you studied logarithms in high school you might have wondered what they were good for. Well, here's one application. You will see others in the next four years.

Suppose we have a set of "n" data points, which you can think of as your height-weight data, but which could really be any set of paired data:

$$\{[h_1, w_1], [h_2, w_2], [h_3, w_3], [h_n, w_n]\}$$

We want to see if there is a power p and a proportionality constant C so that the formula

$$w = C h^p$$

effectively mirrors the real data. **Taking (natural) logarithms of the proposed power law yields**

$$\ln(w) = \ln(C) + p \ln(h)$$

If we define new variables, $y = \ln(w)$ and $x = \ln(h)$, and call $\ln(C) = b$, this is just the equation of a line

having slope p and y-intercept b :

$$y = p x + b$$

In other words, in order for there to be a power law for the original data, the ln-ln data should (approximately) satisfy the equation of a line. Furthermore, this process is reversible; if the ln-ln data lies on a line with slope p , then the original data satisfies a power law with exponent p . This last conclusion is true by the following computation, which takes the line equation and exponentiates it, recovering the original power law:

$$e^y = e^{(p x + b)}$$

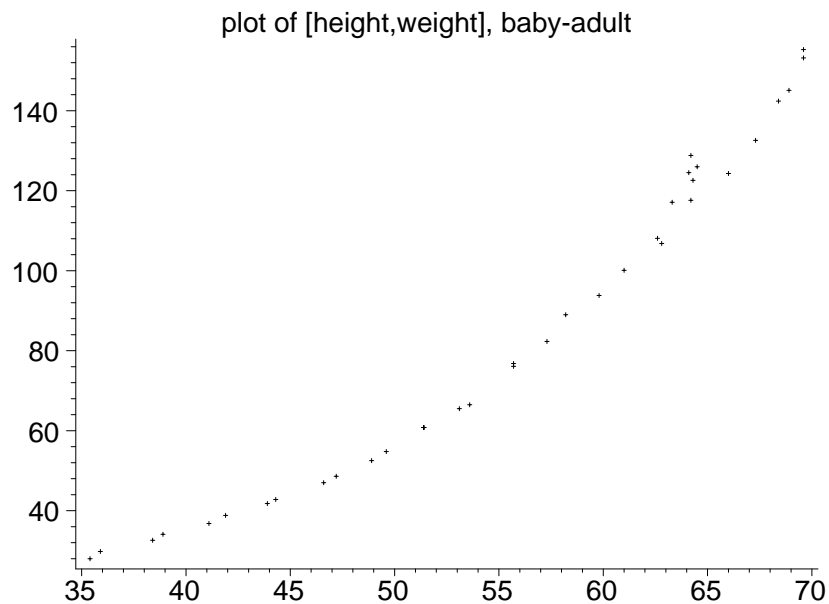
$$e^y = (e^x)^p e^b$$

$$w = C h^p$$

In practice, it is not too hard to see if data is well approximated by a line, so this trick with the logarithm is quite useful.

National data example:

```
[ > restart:
[ > with(plots):
Warning, the name changecoords has been redefined
[ > boyhw:=[[35.9,29.8],[38.9,34.1],[41.9,38.8],[44.3,42.8],
           [47.2,48.6],[49.6,54.8],[51.4,60.8],[53.6,66.5],
           [55.7,76.8],[57.3,82.3],[59.8,93.8],[62.8,106.8],
           [66.0,124.3],[67.3,132.6],[68.4,142.4],[68.9,145.1],
           [69.6,155.3],[69.6,153.2]]:
[ #boy heights (inches) weights (pounds): Ntl medians for ages 2-19
[ > girlhw:=[[35.4,28.0],[38.4,32.6],[41.1,36.8],[43.9,41.8],
            [46.6,47.0],[48.9,52.5],[51.4,60.8],[53.1,65.5],
            [55.7,76.1],[58.2,89.0],[61.0,100.1],[62.6,108.1],
            [63.3,117.1],[64.2,117.6],[64.3,122.6],[64.2,128.8],
            [64.1,124.5],[64.5,126]]:
[ #girl heights(inches) weights (pounds): Ntl medians for ages 2-19
[ > boys:=pointplot(boyhw):
girls:=pointplot(girlhw):
display({boys,girls},
        title='plot of [height,weight], baby-adult');
```



[And now for the ln-ln data:

[> with(linalg): #linear algebra package

Warning, the protected names norm and trace have been redefined and unprotected

```
[ > B:=evalm(boyhw): #evaluate matrix, this will
    #turn our list of points into a matrix, which will
    #be easiser to manipulate in Maple
    G:=evalm(girlhw):
    BG:=stackmatrix(B,G); #stack the matrices on top of eachother
```

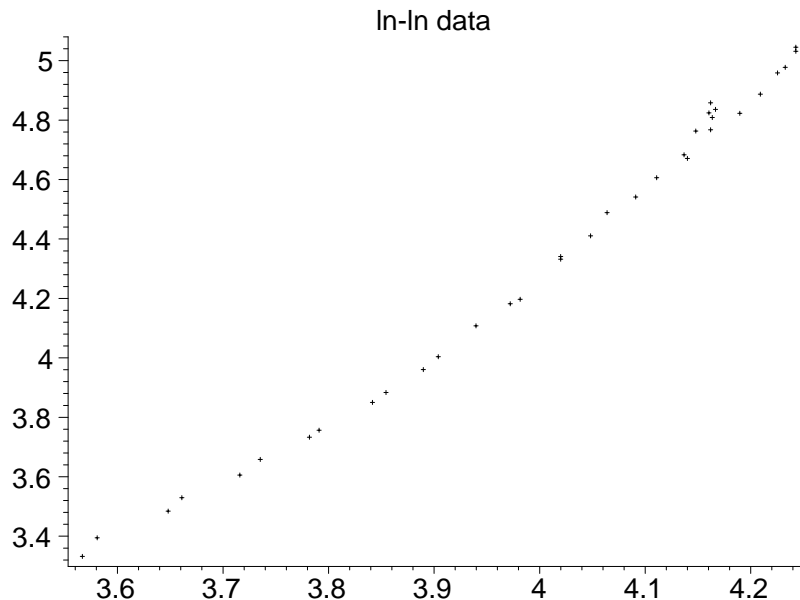
35.9	29.8
38.9	34.1
41.9	38.8
44.3	42.8
47.2	48.6
49.6	54.8
51.4	60.8
53.6	66.5
55.7	76.8
57.3	82.3
59.8	93.8
62.8	106.8
66.0	124.3
67.3	132.6
68.4	142.4
68.9	145.1
69.6	155.3
69.6	153.2
35.4	28.0
38.4	32.6
41.1	36.8
43.9	41.8
46.6	47.0
48.9	52.5
51.4	60.8
53.1	65.5

$BG :=$

```

[ > lnBGa:=map(ln,BG): #get ln-ln data
  lnBG:=map(evalf,lnBGa): #get floating points, otherwise
    #maple will try to work symbolically when you call
    #for least squares solution, and with many points will
    #get hung up.
[ > lnlnplot:=pointplot({seq([lnBG[i,1],lnBG[i,2]],i=1..36)}):
  display(lnlnplot,title='ln-ln data');

```



How do I find the best line fit to a collection of points?

From Calculus, there is a slope "m" and intercept "b" yielding a line which minimizes the sum of the squared vertical distances between your data points and the points on a line, If the "n" data points are

$$\{[x_1, y_1], [x_2, y_2], [x_3, y_3], [x_n, y_n]\}$$

then "m" and "b" satisfy

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix}$$

We could set up this system with several "do-loops", or we can look up the method of least squares in the help files. It turns out to exist in the statistics package. The format is kind of weird, but that's life with Maple.

```

[ > with(stats):
[

```

```

> Xs:=convert(col(lnBG,1),list): #convert the first column of
    #the ln-ln data into a list of the "x's" The least squares
    #command wants to have lists input, not matrix columns!?!?
Ys:=convert(col(lnBG,2),list):
> fit[leastsquare[[x,y]]]([Xs,Ys]); #this is Maple-ese
    #for find the least squares line fit, using variables
    #called x and y for the given values which we
    #called Xs's and Ys's when we defined them.

```

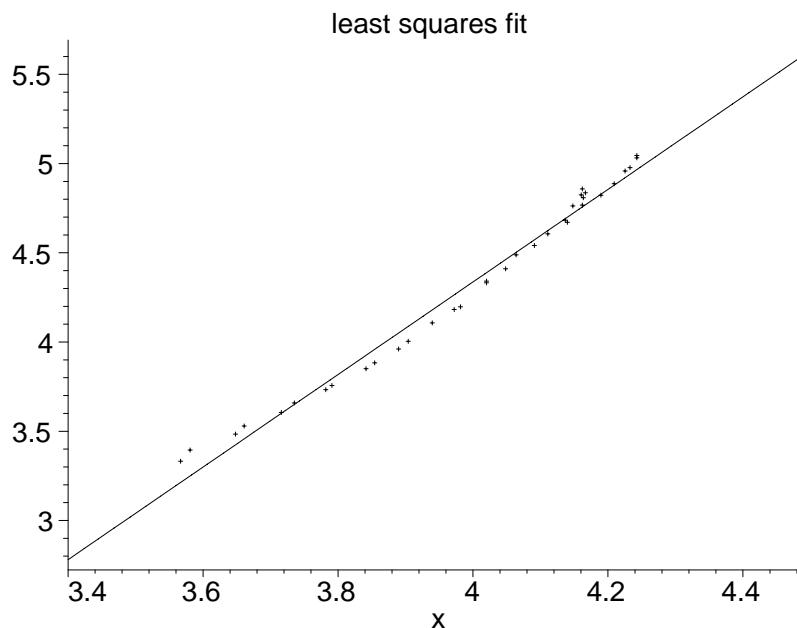
$$y = -6.038703303 + 2.593828004 x$$

[We can paste in the equation of the line and see how well we did.

```

> line:=plot(-6.038703600+2.593828078*x, x=3.4..4.5,
    color=black):
display({line,lnlnplot}, title='least squares fit');

```



You notice that until adolescence the boy and girl data are more or less indistinguishable. The "baby fat" of small children may explain why the 2-3 year olds are slightly above the line, and the peaking near adulthood for both the males and females is probably the effect of their respective hormones. But I would say this ln-ln data IS close to its least-squares line. But WHY?

[Finally, we can go back from the least squares line fit to a power law

```

> p:=2.593828078; #power
    C:=exp(-6.038703600); #proportionality constant
    p:=2.593828078
    C:=.002384648369
> powerplot:=plot(C*h^p,h=0..80,color=black):
display({powerplot,boys,girls},title=
    'power law approximation for height-weight data');

```

[>

