NATIONAL HEIGHT-WEIGHT DATA

AND THE BMI INDEX

There is a database at the U.S. Center for Disease Control, of national body data collected between 1976 and 1980. From this data I have extracted the median heights and weights for boys and girls, age 2-19. Here is the national data; heights are given in inches and weights are in pounds.

```
> with(linalg):with(plots):
> title:=matrix(1,5,[‘age‘,‘boy height‘,‘weight‘,‘girl
  height‘,‘weight‘]):
> A1:=transpose(matrix([[35.9,38.9,41.9,44.3,47.2,49.6,51.4,53.6,55.
  7,
                57.3,59.8,62.8,66.0,67.3,68.4,68.9,69.6,69.6],

  [29.8,34.1,38.8,42.8,48.6,54.8,60.8,66.5,76.8,82.3,93.8,
              106.8,124.3,132.6,142.1,145.1,155.3,153.2]])):
     #boy heights, weights: medians for ages 2-19
> A2:=transpose(matrix([[35.4,38.4,41.1,43.9,46.6,48.9,51.4,53.1,55.
  7,
              58.2,61.0,62.6,63.3,64.2,64.3,64.2,64.1,64.5],

  [28.0,32.6,36.8,41.8,47.0,52.5,60.8,65.5,76.1,89.0,100.1,
             108.1,117.1,117.6,122.6,128.8,124.5,126.0]])):
     #girl data
> ages:=matrix(18,1,[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,
  17,18,19]):

> BG:=augment(ages,A1,A2):   #boys-girls combined
> A:=stackmatrix(title,BG);
```

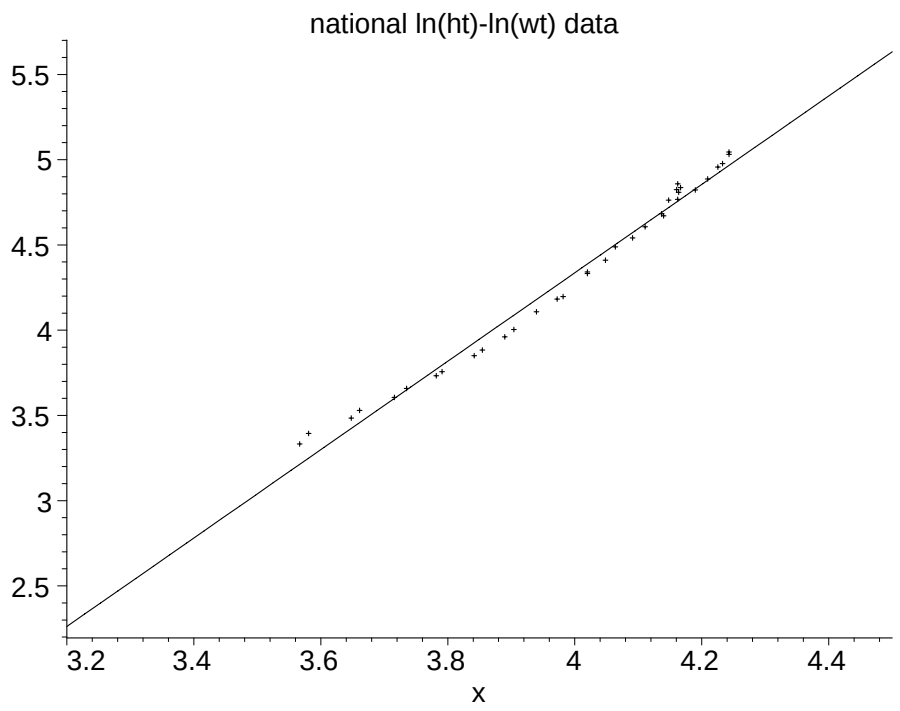
$A :=$

age	*boy height*	*weight*	*girl height*	*weight*
2	35.9	29.8	35.4	28.0
3	38.9	34.1	38.4	32.6
4	41.9	38.8	41.1	36.8
5	44.3	42.8	43.9	41.8
6	47.2	48.6	46.6	47.0
7	49.6	54.8	48.9	52.5
8	51.4	60.8	51.4	60.8
9	53.6	66.5	53.1	65.5
10	55.7	76.8	55.7	76.1
11	57.3	82.3	58.2	89.0
12	59.8	93.8	61.0	100.1
13	62.8	106.8	62.6	108.1
14	66.0	124.3	63.3	117.1
15	67.3	132.6	64.2	117.6
16	68.4	142.1	64.3	122.6
17	68.9	145.1	64.2	128.8
18	69.6	155.3	64.1	124.5
19	69.6	153.2	64.5	126.0

When we do the ln-ln analysis of this data, we find that the least squares line fit is

```
> ln(w)=2.593488078*ln(x) -6.037404653
```

Here is a picture of the least squares line, and the ln-ln data:

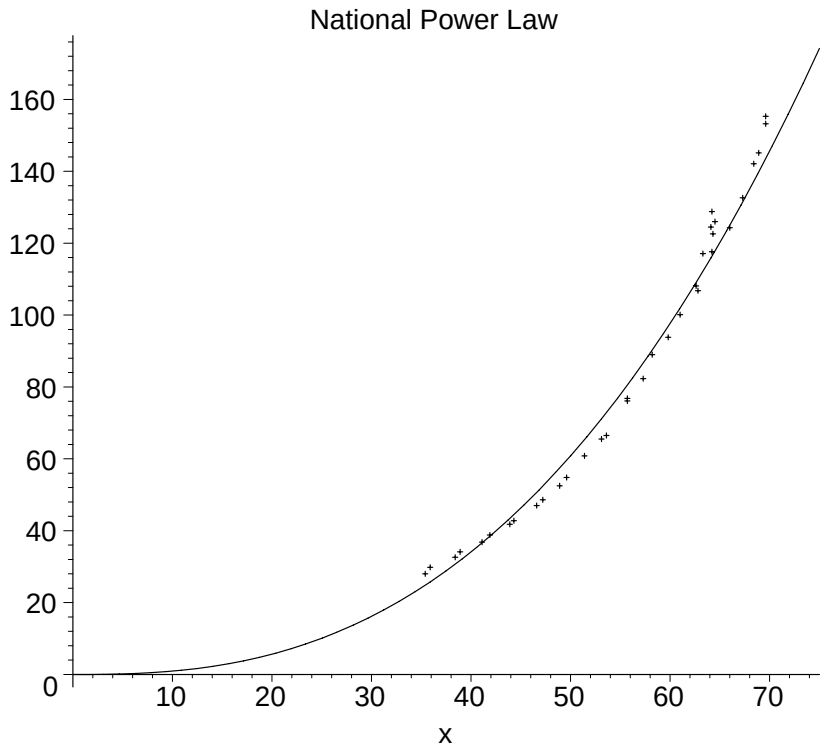


You notice that until adolescence the boy and girl data are more or less indistinguishable. The "baby fat" of small children may explain why the 2-year olds are slightly above the line, and the peaking near adulthood for both the males and females is quite likely to be the effect of their respective hormones. But this data IS very close to a linear fit. One might wonder what sort of biological advantage this sort of scaling is a consequence of. I do.

Going back to the power law:

```
> p:=2.593488078:
  C:=exp(-6.037404653):
  f:=x->C*x^p:
```

And here's a picture of the experimental power law, graphed with the actual heights and weights



Correlation coefficient: In statistics one measures for a possible linear relationship between variables by using the sample correlation coefficient. In class, and in the text (page 189) we discussed the fact that the correlation coefficient is really the cosine of the angle between the normalized vector of x-values and the normalized vector of y-values. (The normalization was to subtract off the average values of each data set so that the averages were both zero.) A correlation coefficient near 1 implies high positive correlation. If you compute the correlation coefficient for our ln-ln data:

$$costheta := (x, y) \rightarrow \frac{\text{dotprod}(x, y)}{\text{norm}(x, 2)\,\text{norm}(y, 2)}$$

```
> costheta(xvect-xvectav,yvect-yvectav);
     #this is the sample correlation coefficient for our
     #ln-ln data.
```

.9924868726

```
> evalf(180/Pi)*arccos(%);
     #the angle between our normalized deviation vectors,
     #in degrees
```

7.027814524