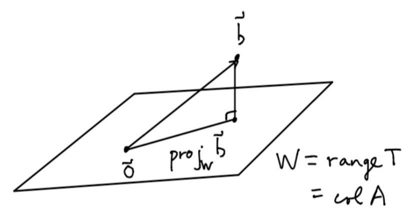
There's actually a smart way to find the least squares solutions that doesn't require an orthonormal basis for *Col A*. To understand it fully depends on concepts we talked about last week (and was one reason we spent a long time talking about orthogonal complements to subspaces). As a further result, it will turn out that one can compute projections onto a subspace with elementary matrix operations and without first constructing an orthonormal basis for the subspace !!! Consider the following chain of equivalent conditions on $\hat{\underline{\mathbf{x}}}$:

$$A\hat{\underline{\mathbf{x}}} = proj_{Col\,A}\,\underline{\boldsymbol{b}}$$

$$\Longleftrightarrow \quad \underline{\boldsymbol{z}} = \underline{\boldsymbol{b}} - A\hat{\underline{\mathbf{x}}} \in (Col\,A)^{\perp} = Nul\,A^T$$



$$\Longleftrightarrow \quad A^T(\underline{\boldsymbol{b}} - A\hat{\underline{\mathbf{x}}}) = \underline{\mathbf{0}}$$

$$\Longleftrightarrow \quad A^T\underline{\boldsymbol{b}} - A^TA\hat{\underline{\mathbf{x}}} = \underline{\mathbf{0}}$$

$$\Longleftrightarrow \quad A^TA\hat{\underline{\mathbf{x}}} = A^T\underline{\boldsymbol{b}}\,.$$

This last equation will always be consistent because projections exist. And if the columns of A are linearly independent the solutions to the top equation, and hence the final equation, will be unique. So the matrix A^TA will be invertible in that case. The final matrix equation is called the *normal equation* for least squares solutions.

<u>Exercise 2</u> Re-do Exercise 1 using the normal equation, i.e find the least squares solution $\hat{\boldsymbol{x}}$ to

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}.$$

$A x = \vec{b}$ inconsistent.

$A^T A \hat{x} = A^T \vec{b}$ instead

And then note that $A\,\hat{\boldsymbol{x}}$ is $proj_{Col\,A}\underline{\boldsymbol{b}}$, i.e. you found the projection of $[3\ 3\ 3]^T$ without ever finding and using an ortho-normal basis!!!

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

<u>Same</u> $\hat{x}$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ 9 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 12 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Wed $A\hat{x} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$ which is $\text{proj}_{ColA}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$ as claimed

Matlab assumes you want the least squares solution when you hand it an inconsistent system. This is because, as we'll discuss tomorrow, whenever an applied mathematician, engineer or scientist is using a finite-dimensional linear model for an actual experimental process, there is almost no chance that the actual data will fit the model exactly.

script:

```
% matlab assumes you want least squares solutions to inconsistent systems
A=[1 2; 0 1; 1 0]
b=[3;3;3]
aug=[A,b]
rref(aug)  %system is inconsistent (last column is pivot column)
x=linsolve(A,b)  %least squares solution
```

executes to produce:

```
A =

     1     2
     0     1
     1     0


b =

     3
     3
     3


aug =

     1     2  |  3
     0     1  |  3
     1     0  |  3


ans =

     1     0  |  0
     0     1  |  0
     0     0  |  1


x =

    2.0000
    1.0000
```

least squares soln

Exercise 3 In the case that $A^T A$ is invertible we may take the normal equation for finding the least squares solution to $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ and find $A\,\hat{\underline{\mathbf{x}}} = proj_{Col\,A}\underline{\mathbf{b}}$ directly:

$$A^T A\,\hat{\underline{\mathbf{x}}} = A^T\underline{\mathbf{b}}$$

$$\hat{\underline{\mathbf{x}}} = \left(A^T A\right)^{-1} A^T\underline{\mathbf{b}}$$

$$proj_{Col\,A}\underline{\mathbf{b}} = A\,\hat{\underline{\mathbf{x}}} = A\left(A^T A\right)^{-1} A^T\underline{\mathbf{b}}.$$

no orthogonal bases required

Verify for the third time that for $W = span\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}\right\}$, $proj_W\begin{bmatrix}3\\3\\3\end{bmatrix} = \begin{bmatrix}4\\1\\2\end{bmatrix}$ by "plug and chug".

if you want, you can check ex.

Wed Nov 21

- 6.6 Fitting data to "linear" models.

Announcements:

Warm-up Exercise: find the least-squares solution to

$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

(we'll use it in class today)

for inconsistent systems

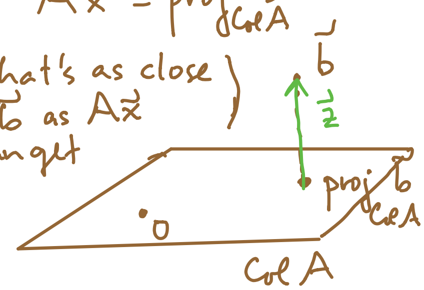
$A\vec{x} = \vec{b}$

but $\vec{b} \notin \text{Col}\,A$

solve instead in

$A\hat{x} = \text{proj}_{\text{Col}A}\vec{b}$

(that's as close to $\vec{b}$ as $A\vec{x}$ can get)



long way

① Find orthonormal (or just orthogonal) basis for Col A

② Use that basis to find $\text{proj}_{\text{Col}A}\vec{b}$

③ Solve consistent sys.

$A\hat{x} = \text{proj}_{\text{Col}A}\vec{b}$

short way: instead of $A\vec{x} = \vec{b}$

solve $A^TA\hat{x} = A^T\vec{b}$ normal eqtn

reason want $\vec{z} = \vec{b} - A\hat{x} \perp \text{Col}A$

$\Longleftrightarrow \quad A^T(\vec{b} - A\hat{x}) = \vec{0}$

$A^TA\hat{x} = A^T\vec{b}.$

Note $A\hat{x}$ will recover $\text{proj}_{\text{Col}A}\vec{b}$

$$\begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

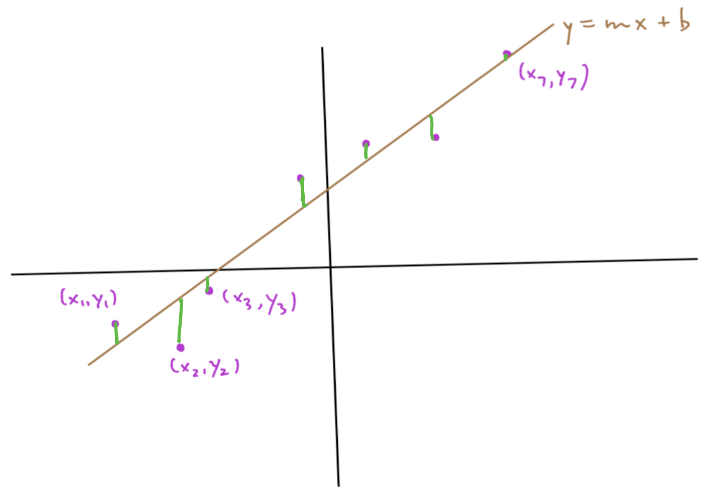
$$\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

Applications of least-squares to data fitting.

- Find the best line formula $y = m\,x + b$ to fit n data points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. We seek $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = m \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$



In matrix form, find $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}. \qquad A \begin{bmatrix} m \\ b \end{bmatrix} = \underline{\boldsymbol{y}}\,.$$

There is no exact solution unless all the data points are actually on a single line!

Least squares solution:

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \underline{\boldsymbol{y}}\,.$$

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \underline{\mathbf{y}}$$

As long as the columns of A are linearly independent (i.e.at least two different values for x_j) there is a unique solution $[m, b]^T$. Furthermore, you are actually solving

$$A \begin{bmatrix} m \\ b \end{bmatrix} = proj_W \underline{\mathbf{y}}$$

where

$$W = span\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\},$$

so

$$\left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - m \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\|^2$$

is as small as possible. In other words, you've minimized the sum of the *squared vertical deviations* from points on the line to the data points,

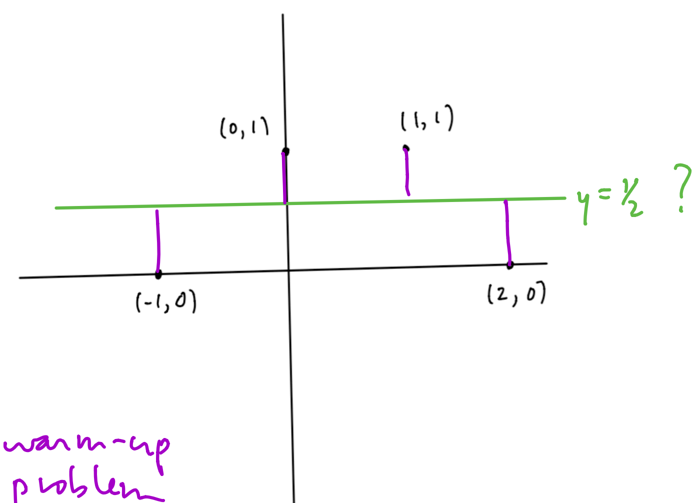
$$\sum_{i=1}^{n} (y_i - m x_i - b_i)^2 .$$

Exercise 1 Find the least squares line fit for the 4 data points $\{(-1, 0), (0, 1), (1, 1), (2, 0)\}$. Sketch.

want $m \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$

↑ x-vals ↑ corres y-vals

$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$



see warm-up problem

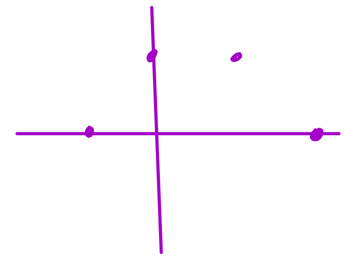
$\begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}$ so $y = 0x + \frac{1}{2} = \frac{1}{2}$

<u>Example 2</u> Find the best quadratic fit to the same four data points. This is still a "linear" model!! In other words, we're looking for the best quadratic function

$$p(x) = c_0 + c_1 x + c_2 x^2$$

to fit to the four data points

$$\{(-1, 0), (0, 1), (1, 1), (2, 0)\}.$$



We want to solve

$$c_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c_2 \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

e.g. $c_0 + c_1 x_1 + c_2 x_1^2 = y_1$

$c_0 + c_1 x_2 + c_2 x_2^2 = y_2$

$\vdots$

For our example this is the system

$$c_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

$= 1 \qquad = .5 \qquad = -.5$

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

with Matlab and the least squares normal equation (which matlab will apply automatically as well), we can let technology solve

$$A^T A \underline{\boldsymbol{c}} = A^T \underline{\boldsymbol{b}}$$

although this problem is small enough that one could also work it by hand.

This Matlab script

```
%in the following example the least square solution is
%actually an exact solution.
C=[1,-1,1; 1,0,0; 1,1,1; 1,2,4]
b2=[0;1;1;0]
c=linsolve(C,b2)   %least squares solution
c2=(transpose(C)*C)^(-1)*transpose(C)*b2  %also least squares solution
rref([C,b2])  %system was consistent
```

yields

```
C =

     1    -1     1
     1     0     0
     1     1     1
     1     2     4

b2 =

     0
     1
     1
     0
```

```
c =

    1.0000
    0.5000
   -0.5000
```

Matlab

```
c2 =

    1.0000
    0.5000
   -0.5000
```

magic matrix formula

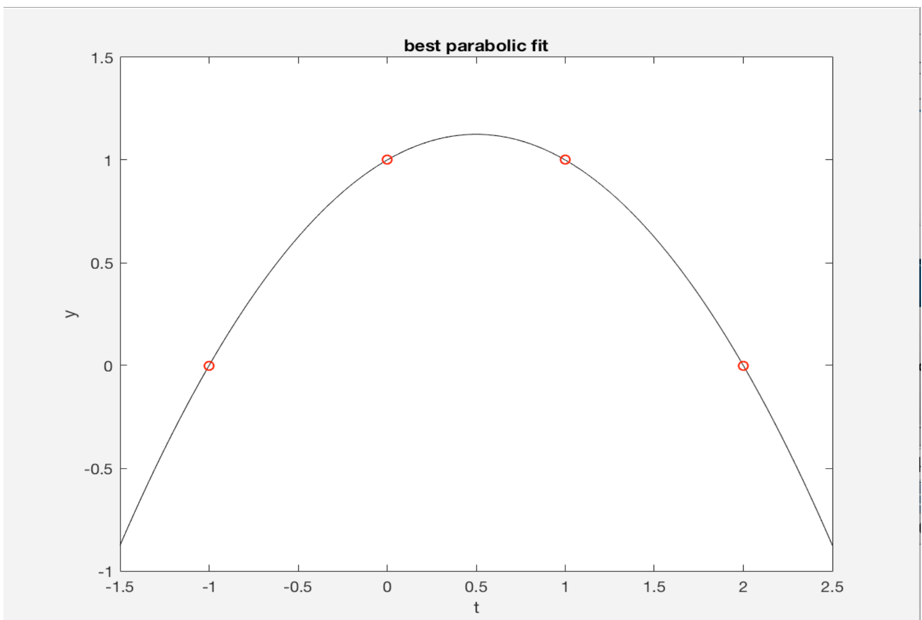
```
ans =

    1.0000         0         0    1.0000
         0    1.0000         0    0.5000
         0         0    1.0000   -0.5000
         0         0         0         0
```

For a plot, this script:

```
%plots...
t=linspace(-1.5,2.5,100) %left endpt, right endpt, numpoints
% "t" above is a vector 100 equally spaced numbers between -1.5 and 2.5
% the definition below is for an equally sized vector containing the
% parabolic approximation.  we use "t." to extract a scalar value from the
% vector
y=c(1)+c(2)*t+c(3)*t.*t
lucky1=plot(t,y,'black')
title('best parabolic fit')
xlabel('t')  %horizontal variable label
ylabel('y')   %vertical variable label
hold on   % the "hold" command lets us combine plots into one display
scatter([-1,0,1,2],[0,1,1,0],'red')
hold off
```

produces this display:



Math 2270-002
Week 13-14 homework,
due November 28.

6.5 Least square solutions
1, **3**, 5, **7**, **9**, 11, 15, **17**, 19

6.6 Linear models for data fitting
1, **7**, and exercise **w13.2** below about the human height-weight power law.

6.7 Inner product spaces
6.7.25 extended (Legendre polynomials): For functions in $C[-1, 1]$ Use Gram-Schmidt to find an orthogonal basis for $W = span\{1, t, t^2, t^3\}$, with respect to the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)\ \mathrm{d}t.$$

In the first part of the problem scale the orthogonal polynomials so that the coefficient of the leading power of t is 1. Then normalize the orthogonal basis to make it orthonormal. You can read more about Legendre polynomials at Wikipedia.

w13.1 In quiz 13 you found $proj_W\ \underline{\boldsymbol{b}}$, for $\underline{\boldsymbol{b}} = \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix}$ and $W = span\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -4 \end{bmatrix} \right\}$, by first finding an orthogonal basis for W and then using that basis to do the projection. Rework this projection problem by using the method of least squares algorithm from section 6.5, as we've also discussed in class.

Math 2270-002 Fall 2018

A Power Law For Human Heights and Weights

Body Mass Index

A person's BMI 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. If you measure height in inches and weight in pounds then the formula is

$$BMI = 703\frac{w}{h^2}$$

The graph of heights and weights for which BMI has a constant value B is the parabola

$$w = \frac{B}{703}\cdot h^2.$$

Thus, the assumption underlying BMI is that for adults at equal risk levels (but different heights), weight should be proportional to the square of height. This is a historical accident and at some point became a dogma. The BMI was popularized in the 1960's in the U.S., by proponents who were initially unware that they were repeating history. It is easy to deduce that if people were to scale equally in all directions when they grew, 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 length-wise when we grow taller. One would expect the best predictive power to be somewhere between 2 and 3. If the power is much larger than 2 then one could argue that the body mass index might need to be modified to reflect this fact.

It turns out a Belgian demographer, Adolphe Quetelet, also called the "Father of Statistics", originally proposed a power of p=2 for adults, based on his own data analysis during the early 1800's. In a footnote which history has forgotten, he said that a power of 2.5 is more appropriate if you want an approximation for people of all ages. He actually wrote that the square of the weight should scale like the fifth power of the height, because pre-calculators, fractional powers were harder for people to deal with. My recollection is that this footnote appears in the 1835 publication "Sur l'homme et le développement de ses facultés, ou Essai de physique sociale". I have read the footnote.

There is (or at least there was, 20 years ago) 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. The national data is shown on the next page; heights are given in inches and weights are in pounds.

w13.2) Find the power law

$$w = C\,h^p$$

predicted by this data, by finding a least squares line fit to the ln-ln data. (Combine the boy-girl data into one set.) We will discuss this further in class on Monday after Thanksgiving. Note that if such a power law holds, taking logarithms of both sides of the identity yields

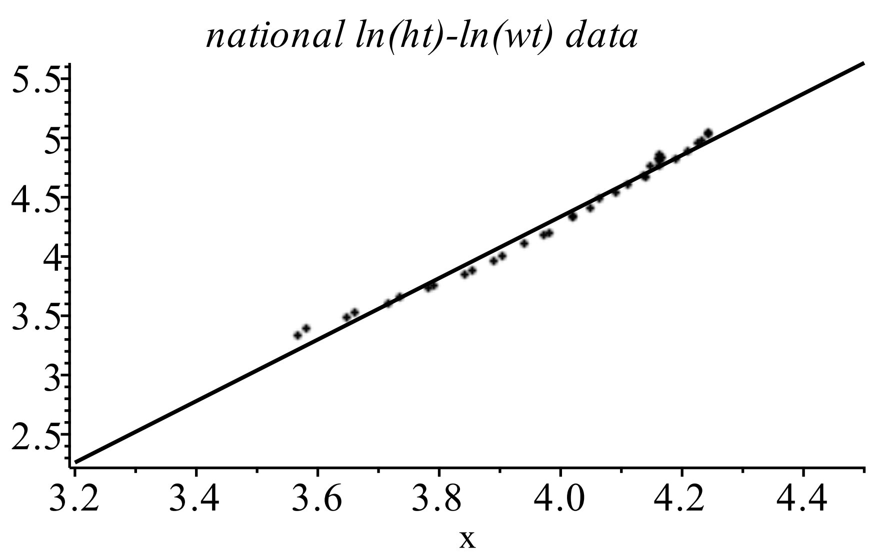
$$\ln(w) = \ln(C) + p\cdot\ln(h).$$

If we write $Y = \ln(w)$, $X = \ln(h)$ then this is the equation of a line in the $X - Y$ plane, where the slope is the original power p and the Y-intercept equals $\ln(C)$,

$$Y = Y_0 + p\,X$$

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

A graph of the best line fit to the national ln − ln data. It's a pretty good fit! (Infants are a little heavier than the line predicts, adolescent data is slightly below the line, and as adults mature they rise a bit above the line. The slope of the line will be the power in the approximate power law.



submission: I prefer that you use Matlab. In that case, submit a script to CANVAS which computes the least squares line fit; which recovers the power law; and which creates a graph of the log-log point scatterplot together with the least squares line (as above); and a separate plot which combines a scatter plot of the original height-weight data, together with the graph of the power law function. We will use an analogous script for a smaller problem in class on Monday. If you don't use Matlab please hand in hard copies of same results with the rest of your homework.