Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Non-Linear Models

We now relax some assumptions seen in all of the models presented so far in these sessions. The first assumption is that the parameters seen on right hand side of the regression equation are in the form of a linear combination of regression coefficients multiplied by explanatory variables :

$$\sum_{i=0}^{k} \beta_i x_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

We have seen that such linear combinations can include a wealth of complex explanatory variables like products of explanatory variables and squares [and powers] of measured measured explanatory variables. The x_i can be [in principle] any function of a set of explanatory variables. Nevertheless, the parameters β_i only appear in the linear combination form noted above. These models are called 'Linear Models'.

An additional set of assumptions needs to be mentioned. If the model has an explicit 'error' term, for a model to be considered 'Linear', the error term ϵ must add to the linear combination term as :

$$\sum_{i=0}^{k} \beta_i x_i + \epsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon$$

Also, with conditional models, to be 'Linear', the subject specific term u must add to the expression :

$$\sum_{i=0}^{k} \beta_i x_i + u + \epsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u + \epsilon$$

or

$$\sum_{i=0}^{k} \beta_i x_i + u = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

if there is no explicit 'error' term [like with Logistic Regression, for example].

This additive feature extends to models that have more than one subject specific term or the 'multi-level' type models.

There are huge collection of models available that do not require this linear combination form. These models are referred to as 'Non-Linear Models' or 'Non-Linear Regressions'. This naming can be a bit confusing since we have already devoted considerable attention to nonlinearity. The exposition so far has considered nonlinearity of measured explanatory variables but within the linear combination requirement noted; within the world of Linear Models.

The examples which follow should help to clarify these matters.

Michelis-Menten Model :

We will begin with a model usually called a Michelis-Menten model. In its simplest form, there is a

single measured positive explanatory variable x and the model has two parameters β_0 and β_1 and the right hand side of the equation has :

$$\frac{\beta_0 x}{\beta_1 + x}$$

This function can also be written as $\dfrac{\beta_0}{1+\dfrac{\beta_1}{x}}$ so that, for large x, the function is tending to β_0 : a horizontal asymptote.

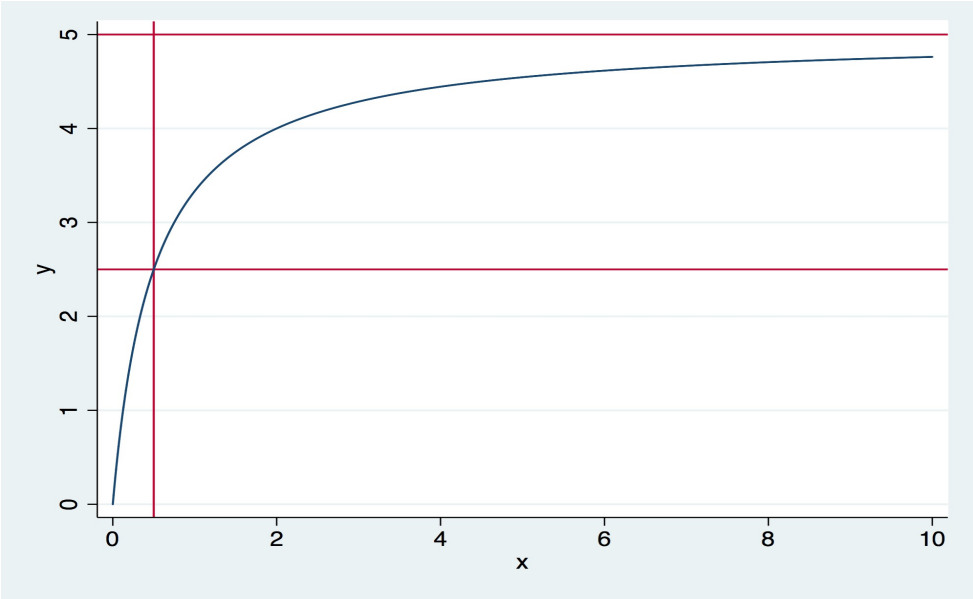
Now notice that when x is equal to β_1 , the function equals $\dfrac{\beta_0}{2}$.

$\dfrac{\beta_0}{2} = \dfrac{\beta_0 x}{\beta_1 + x}$ if and only if $\dfrac{1}{2} = \dfrac{x}{\beta_1 + x}$ if and only if $\beta_1 + x = 2x$ if and only if $x = \beta_1$

So β_0 and β_1 will both be positive. This function ranges from 0 to β_0 .

With a bit of calculus, it can be shown that the function is always concave down. It is the arc of a hyperbola.

Here is an example with $\beta_0 = 5$ and $\beta_1 = 0.5$:



We now consider this model with an additive error term :

$$\frac{\beta_0 x}{\beta_1 + x} + \epsilon$$

This is the form of the Michelis-Menten model that can be fit in Stata or in R.

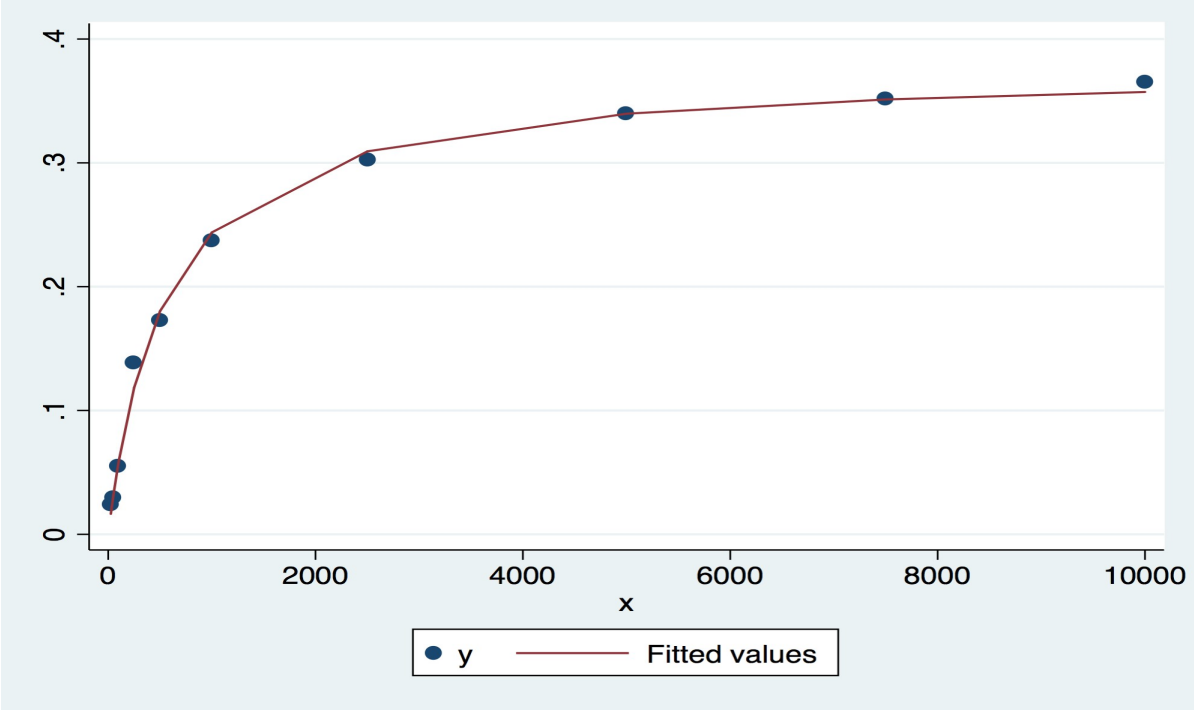
We can use our knowledge for comparing two groups by using an indicator variable G and then

considering a model like :

$$\frac{(\beta_0+\beta_2 G)x}{(\beta_1+\beta_3 G)+x}+\epsilon$$

An example from Marasovic[2017] :

```
. list x y

     +---------------+
     |     x       y |
     |---------------|
  1. |    25   .0243 |
  2. |    50   .0292 |
  3. |   100   .0546 |
  4. |   250   .1388 |
  5. |   500   .1726 |
     |---------------|
  6. |  1000   .2374 |
  7. |  2500   .3023 |
  8. |  5000   .3395 |
  9. |  7500   .3515 |
 10. | 10000   .3652 |
     +---------------+

. nl (y={b0}*x/({b1}+x))
(obs = 10)

Iteration 0:  residual SS =  .1672244
Iteration 1:  residual SS =  .0780477
Iteration 2:  residual SS =  .0262549
Iteration 3:  residual SS =  .0051114
Iteration 4:  residual SS =  .0009641
Iteration 5:  residual SS =  .0007095
Iteration 6:  residual SS =  .0007066
Iteration 7:  residual SS =  .0007066
Iteration 8:  residual SS =  .0007066
Iteration 9:  residual SS =  .0007066


      Source |       SS             df       MS
-------------+------------------------------    Number of obs =         10
       Model |  .57270148           2  .286350739    R-squared     =     0.9988
    Residual |  .00070662           8  .000088328    Adj R-squared =     0.9985
-------------+------------------------------    Root MSE      =   .0093983
       Total |   .5734081          10   .05734081    Res. dev.     =  -67.19721

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   .3767161   .0066251    56.86   0.000     .3614386    .3919936
         /b1 |   544.9243   40.90001    13.32   0.000     450.6087    639.2399
------------------------------------------------------------------------------


. predict yh
(option yhat assumed; fitted values)

. twoway (scatter y x)(line yh x)
```



Fitting a line or a parabola or even a lowess smoother is unsuccessful [not shown here]. The fit with the Michelis-Menten model [above] is very fine, indeed.

It is worth emphasizing that this model has an additive error term ϵ . In the output, we get an estimate of the standard deviation of ϵ which is the `Root MSE = .0093983`

We can fit two Michelis-Menten models using the Puromycin data from Bates & Watts[1988]

```
. nl (v=({b0}+{b2}*t)*c/(({b1}+{b3}*t)+c))
(obs = 23)

Iteration 0:  residual SS =   44201.1
Iteration 1:  residual SS =   12451.8
Iteration 2:  residual SS =  3575.968
Iteration 3:  residual SS =  2139.057
Iteration 4:  residual SS =   2056.78
Iteration 5:  residual SS =   2055.08
Iteration 6:  residual SS =  2055.054
Iteration 7:  residual SS =  2055.053
Iteration 8:  residual SS =  2055.053
Iteration 9:  residual SS =  2055.053
Iteration 10:  residual SS =  2055.053


      Source |       SS             df       MS
-------------+------------------------------------    Number of obs =         23
       Model |  417561.95            4  104390.487    R-squared     =     0.9951
    Residual |  2055.0531           19  108.160691    Adj R-squared =     0.9941
-------------+------------------------------------    Root MSE      =   10.40003
       Total |     419617           23  18244.2174    Res. dev.     =   168.6001

------------------------------------------------------------------------------
           v |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |     160.28   6.896012    23.24   0.000     145.8465    174.7136
```

```
         /b2 |    52.4037   9.551015     5.49   0.000      32.4132     72.3942
         /b1 |   .0477082   .0082812     5.76   0.000     .0303755    .0650408
         /b3 |   .0164131    .011429     1.44   0.167     -.007508    .0403342
------------------------------------------------------------------------------

. nl (v=({b0}+{b2}*t)*c/({b1}+c))
(obs = 23)

Iteration 0:  residual SS =   44201.1
Iteration 1:  residual SS =   12573.7
Iteration 2:  residual SS =  3767.496
Iteration 3:  residual SS =  2328.525
Iteration 4:  residual SS =  2242.811
Iteration 5:  residual SS =   2240.92
Iteration 6:  residual SS =  2240.892
Iteration 7:  residual SS =  2240.891
Iteration 8:  residual SS =  2240.891
Iteration 9:  residual SS =  2240.891


      Source |       SS             df       MS
-------------+------------------------------    Number of obs =         23
       Model |  417376.11            3   139125.37    R-squared     =     0.9947
    Residual |  2240.8915           20  112.044574    Adj R-squared =     0.9939
-------------+------------------------------    Root MSE      =   10.58511
       Total |     419617           23  18244.2174    Res. dev.     =   170.5913

------------------------------------------------------------------------------
           v |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         /b0 |   166.6041   5.807425    28.69   0.000       154.49    178.7182
         /b2 |   42.02597   6.272138     6.70   0.000     28.94252    55.10942
         /b1 |   .0579718   .0059102     9.81   0.000     .0456434    .0703002
------------------------------------------------------------------------------
```

In addition to the t-tests, one could compare the two [nested] models with an F-test.

Many years ago and before these non-linear least squares methods were available, two [now obsolete] methods were used. One was called the 'double reciprocal' method and sometimes noted as the 'Lineweaver-Burk' method. The other was called the 'Eadie-Hofstee' method.

For clarity, here is the rationale for the double reciprocal method :

$y=\frac{ax}{(b+x)}$ then $1/y=\frac{(b+x)}{(ax)}=\frac{1}{a}+(\frac{b}{a})*1/x$

So 1/y is a line in 1/x. Sounds good, right? Can we then use linear least squares?

The crucial point is that the assumptions for the nonlinear least squares and this transformed least squares are not the same. In Stata, nl assumes an additive error term that has constant variance.
A regression of 1/y versus 1/x assumes an additive error term with constant variance but for 1/y now.
The two sets of assumptions are very different.

If we fit the model :

$\frac{1}{y}=\alpha_0+\alpha_1*\frac{1}{x}+\epsilon$ there is no simple relationship between the estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ and the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

The rationale for the 'Eadie-Hofstee' method is :

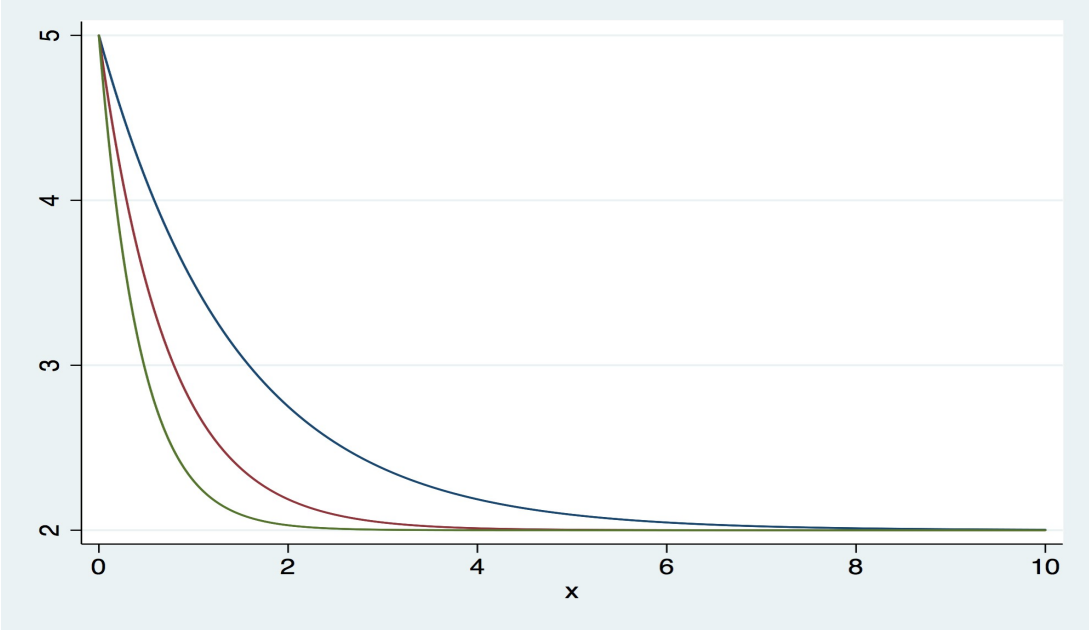
$\frac{a}{y}=\frac{(b+x)}{x}=\frac{b}{x}+1$ then $a=b*\frac{y}{x}+y$ rearranging $y=a-b*\frac{y}{x}$

So y versus y/x is a line. Notice, now that if we tried to model this, we would have the so-called explanatory variable y/x being a function the response y. More troubles with any attempt to fit.

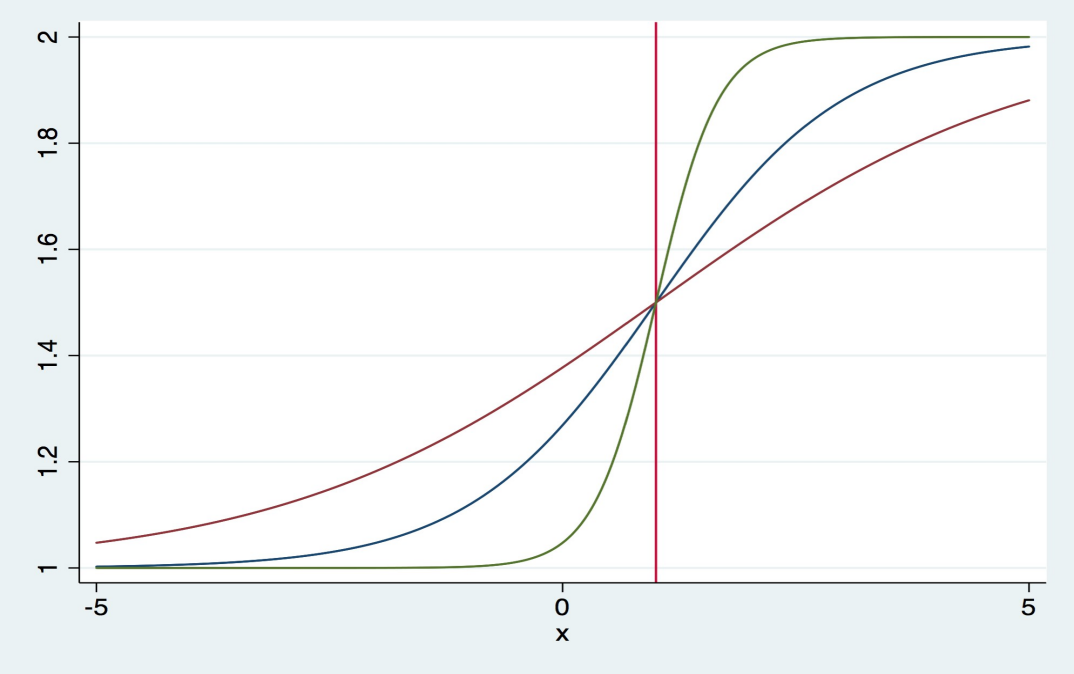
There is a [now obsolete] literature on so-called transformed Linear Models. Be careful out there !

Stata's nl command has a number of models pre-built.

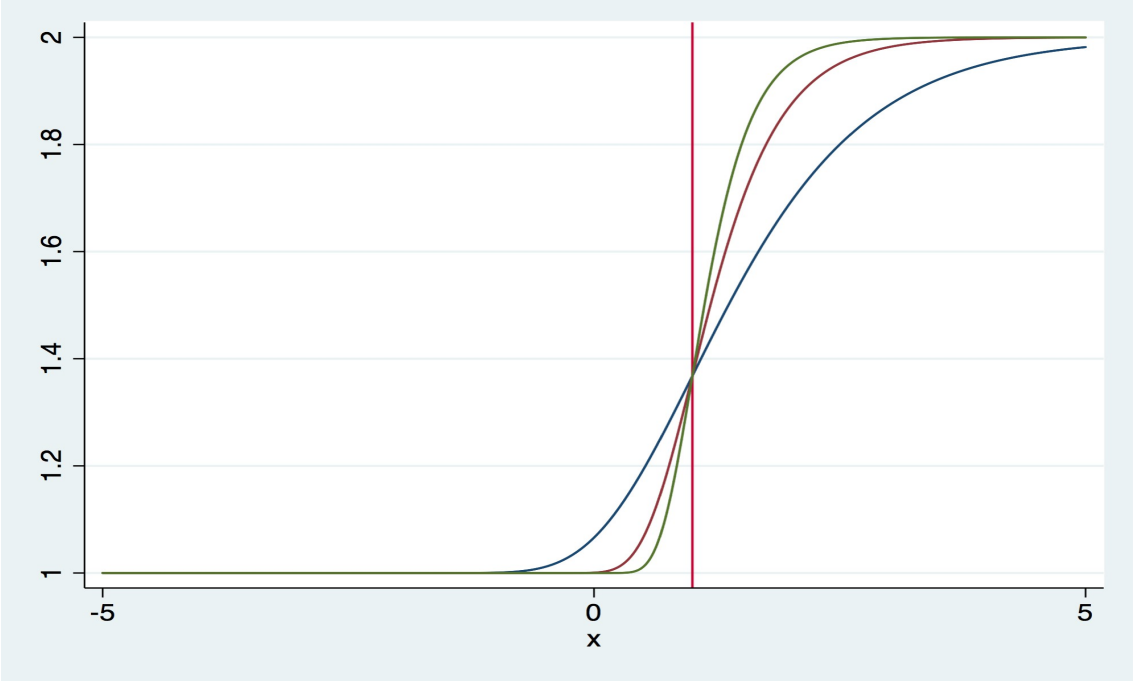
$y=\beta_0+\beta_1*\beta_2^{x}+\epsilon$ called exp3 : shown below for $\beta_0=2$ $\beta_1=3$ and $\beta_2=0.5,\ 0.25,\ 0.1$ and $\epsilon=0$



$y=\beta_0+\frac{\beta_1}{1+e^{-\beta_2*(x-\beta_3)}}+\epsilon$ called log4 : shown below for
$\beta_0=1$ $\beta_1=1$ $\beta_2=0.5,\ 1,\ 3$ $\beta_3=1$ and $\epsilon=0$



$y=\beta_0+\beta_1 e^{-e^{(-\beta_2*(x-\beta_3))}}+\epsilon$ called gom4 : shown below for
$\beta_0=1$ $\beta_1=1$ $\beta_2=1,\ 2,\ 3$ $\beta_3=1$ and $\epsilon=0$



See 'help nl' in Stata for lots of details.
Bates & Watts[1988] offers 18 different nonlinear models and various ways to extend these models.
The illustrations in Bates & Watts are recommended for study.

Further, there are a number of additional concepts in Bates & Watts.

The Nonlinear Models literature contains many more very different models.
There are multi-level nonlinear models available in Stata and in R.

Meddings, Scott and Fick [1989]