Paper 215-28

Teaching Statistical Methods Courses with Case Studies and JMP®

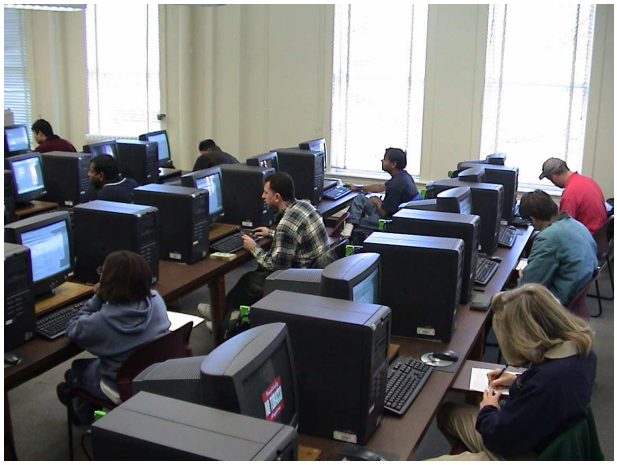
Mauromoustakos Andy and Carol Ojano-Dirain, U of Arkansas, Fayetteville, AR

ABSTRACT

The paper will focus on teaching statistical principles and analyses using cases studies and JMP (Version 5, 2002) as an analysis tool and a teaching/learning tool.

INTRODUCTION

At the University of Arkansas the Agricultural Statistics Laboratory is responsible in offering service graduate courses on Principles of Experimentation, Regression Applications and Experimental Design to graduate students from all Departments of the College of Agriculture Food and Life Sciences and others including the Biological and Health Science Departments. We have being using JMP as the primary tool for carrying out the analysis with very good success. Lately we have attempted with great deal of success according to the students to teach the material using Case Studies (Foster 2002). We will like to demonstrate and share our experiences since we strongly believe in this marriage to provide better understanding and facilitate a fun learning environment. We will present and discuss one such case study presented and discussed in the Applied Regression class during the fall semester of 2002. In fact this case study was one of the two problems that the students were asked to model during their final examination in the computer lab environment seen below.



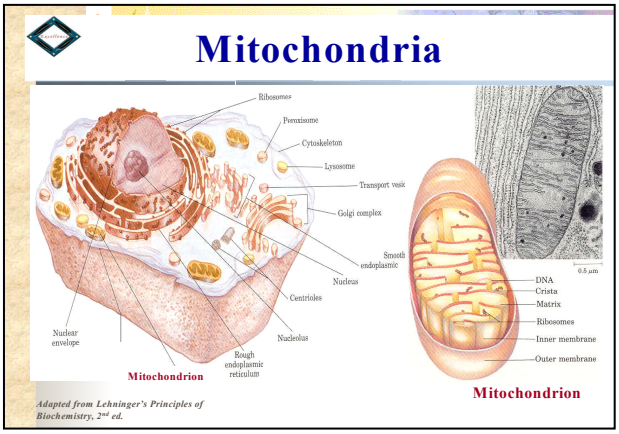
CASE STUDY (APPLIED REGRESSION)

Measurement of Succinate Ubiquinone Reductase in Liver Mitochondria of Broilers (CIILiverMito.JMP)

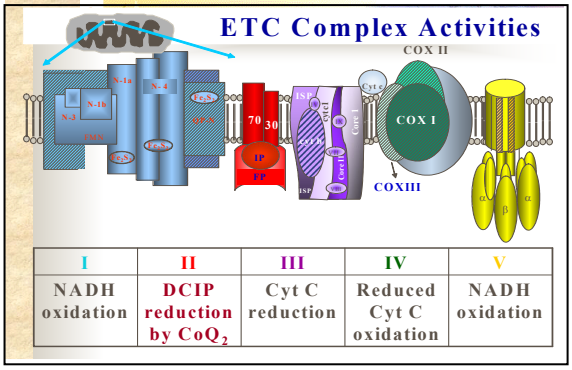
Mitochondria are specialized compartment present in all eukaryotic cells except in the red blood cells. They are responsible for creating more than 90% of the energy needed by the body to sustain life and support growth. When they fail, less and less energy is generated within the cell. Cell injury and even cell death follow. Mitochondrial dysfunction have been implicated with degenerative disorders and ageing, and in the pathogenesis of numerous human diseases including Alzheimer's, atherosclerosis, diabetes, lactic acidosis, developmental delays and many, many more. In our laboratory, we have initially documented that differences in mitochondrial function may contribute to the inherent differences in phenotypic expression of feed efficiency within a single genetic line of broilers.
The electron transport chain consists of major enzyme complexes, namely Complex I through IV and the enzyme complex for ATP synthesis, ATP synthase. Measuring the enzyme activity of the electron transport chain, either by polographic or spectrophotometric assays, is one way of assessing mitochondrial function. When we are measuring Complex activity in broilers, substrate concentration and amount of enzyme and mitochondria sample is usually adjusted until a linear decrease in absorbance is obtained. Activity is then measured for certain time and then time range for optimum enzymatic activity is selected. But for this case study, we are trying to look only on the activity of one of the enzyme complex, Complex II or Succinate Ubiquinone Reductase (SQR) at fix concentration of substrate, enzyme and sample, and for the whole 5-minute duration of the assay.
The data set (**CIILiverMito.JMP**) represents Complex II activity in broilers by spectrophotometric assay. CII activity was measured by following the decrease in absorbance at 600nm due to the reduction of dichloroindeophenol by Coenzyme Q (UQ-2) for 5 minutes with 10-second interval. Values are mean of 9 birds with two replicates per bird.
Our goal is to evaluate models that describe the relationship of Complex II activity through time and to determine activity rate of Succinate-UQ2 Reductase enzyme in isolated frozen liver mitochondria from broilers.

PROBLEM SUBJECT MATTER INTRODUCTION

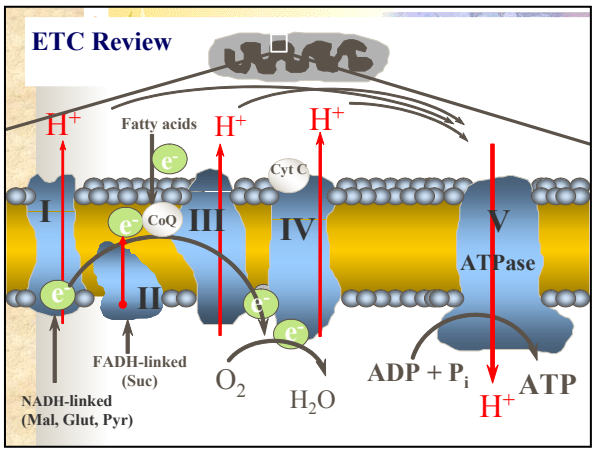
Mitochondria are specialized compartment present in all eukaryotic cells except in the red blood cells. They are responsible for creating more than 90% of the energy needed by the body to sustain life and support growth. When they fail, less and less energy is generated within the cell. Cell injury and even cell death follow.



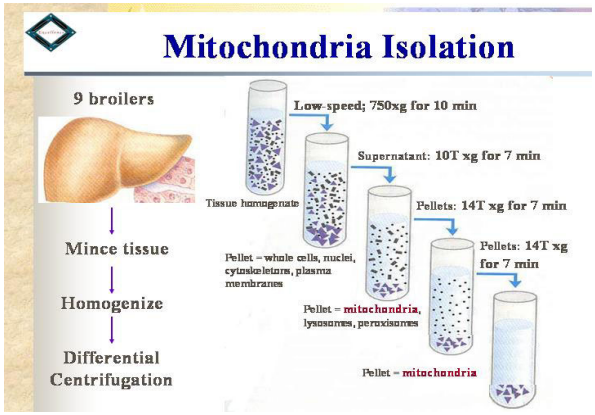
Mitochondrial dysfunction have been implicated in the pathogenesis of ageing and to numerous human diseases including Alzheimer's, atherosclerosis or heart disease, diabetes and liver diseases, lactic acidosis, developmental delays and many, many more.
The researchers we have initially documented that differences in mitochondrial function may contribute to the inherent differences in phenotypic expression of feed efficiency within a single genetic line of broilers.
Some ways to assess mitochondrial function include measuring RCR (index of ETC coupling), ADP:O (index of ATP synthesis coupled to cell respiration), respiration rates, generation of ROS, immunohistochemistry and measuring enzyme activity of the ETC Complexes. The advantage of enzyme complex activity includes "simplicity" of the assay in terms of equipment (spectrophotometer) and sample preparation and can be done on frozen samples. The first 4 assays require "live" mitochondria.



This picture below summarizes the flow of electrons and protons through the four complexes of the respiratory chain. Electrons from NADH and FADH-linked substrates reached UQ via complexes I and II, respectively. (Note that electrons from fatty acid B-oxidation can also enter the respiratory chain through UQ).Ubuquinol (reduced UQ) serves as mobile carrier of electrons and protons. It passes electrons to Complex III, which passes them to another mobile connecting link, Cyt C. Complex IV then transfers electron from reduced Cyt C. to O2 that is reduced to water. Electron flow through the complexes is accompanied by proton flow from the matrix to the inner membrane space.



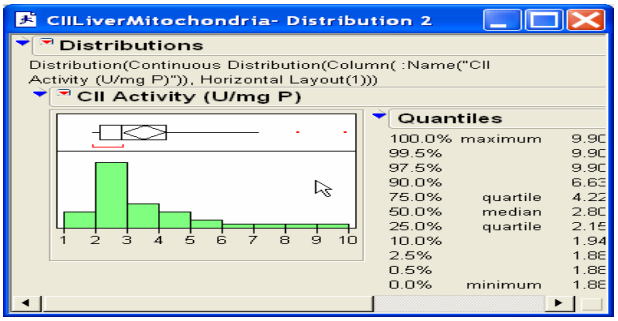
The liver was excised and immediately placed in a beaker containing PBS medium. The liver was minced and homogenized with Potter Elvejehm homogenizer. Mitochondria was isolated as shown in the following differential centrifugation.



The researchers measured Complex II activity by following the CoQ-catalyzed reduction or decrease in absorbance of dichloroindephenol or DCIP (a dye reagent and serve as electron acceptor). Set up the spectrophotometer, Abs at 600nm and temp. at 37C. in a 1-ml cuvette….mix gently and quickly and record absorbance. Literature suggested 5-minutes.

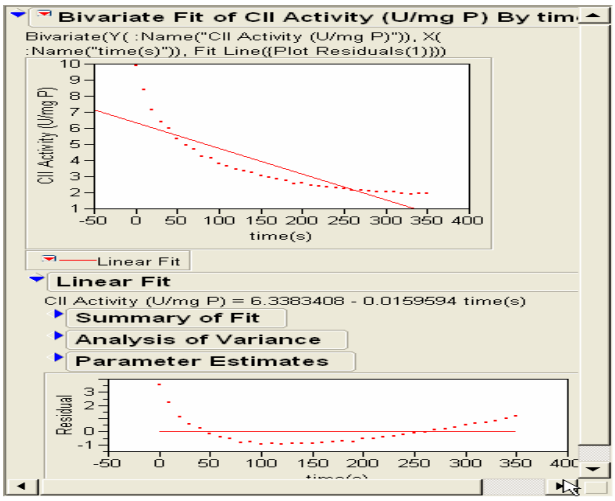
DATA EXPLORATION IN JMP5

As shown in the distribution of the data below, the highest activity is 9.9 and the lowest is 1.89 U/mg mitochondrial protein.

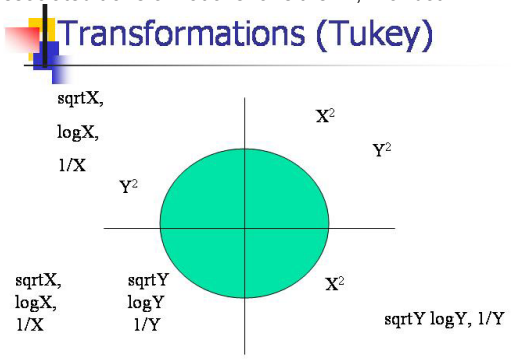


BIVARIATE PLOT AND FITS (FIT Y BY X)

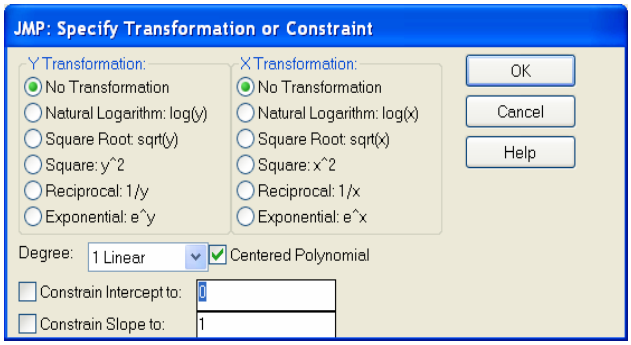
Below is plot of the data over time with a simple linear regression fit superimposed along with a plot of its residuals. As indicated, a linear regression is not adequate to describe the relationship between time and Complex II activity through time. The residuals also indicate curvature in the data. Our next step is to consider either transformation that linearize the data or higher order curvilinear or nonlinear models to model this relationship.



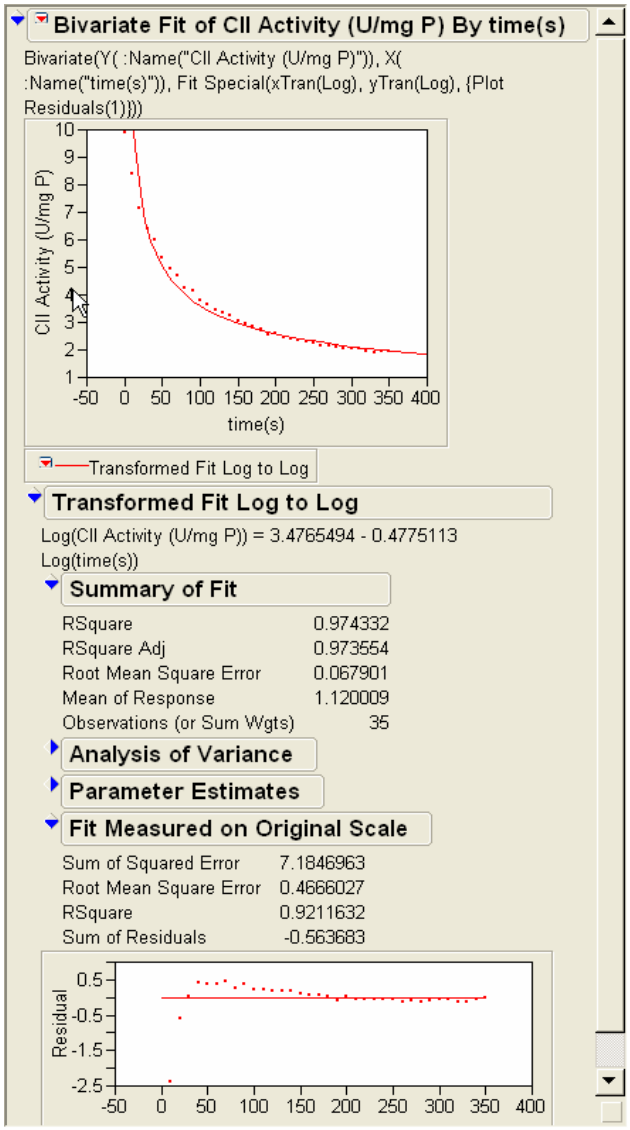
Tukey suggested the following: Match the curvature in your data to one of the four quadrants in the figure below and use the associated transformations for either X, Y or both.



Thus as a first attempt we tried to lower both the power of X and the power of Y by taking logarithms of both X and or Y since they are popular because of the interpretation of the slope in terms of percentage changes rather than absolute changes. We can do this in the Fit Y by X platform of JMP be selecting the Fit Special option and selecting the desired transformation as shown below.

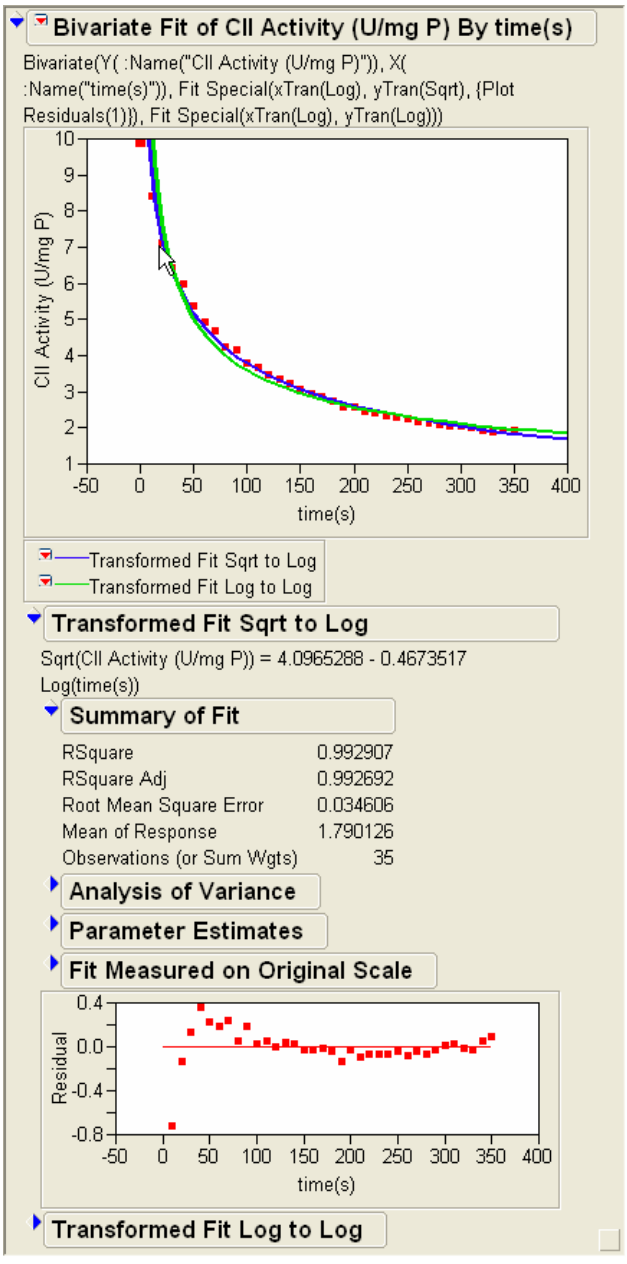


We first tried the log(y), log(X) transformation shown in output below because the "slope coefficients" ("*elasticities*") have nice interpretations but as you can see we did not do an "adequate enough job" as seen by the fit details and residuals plot.



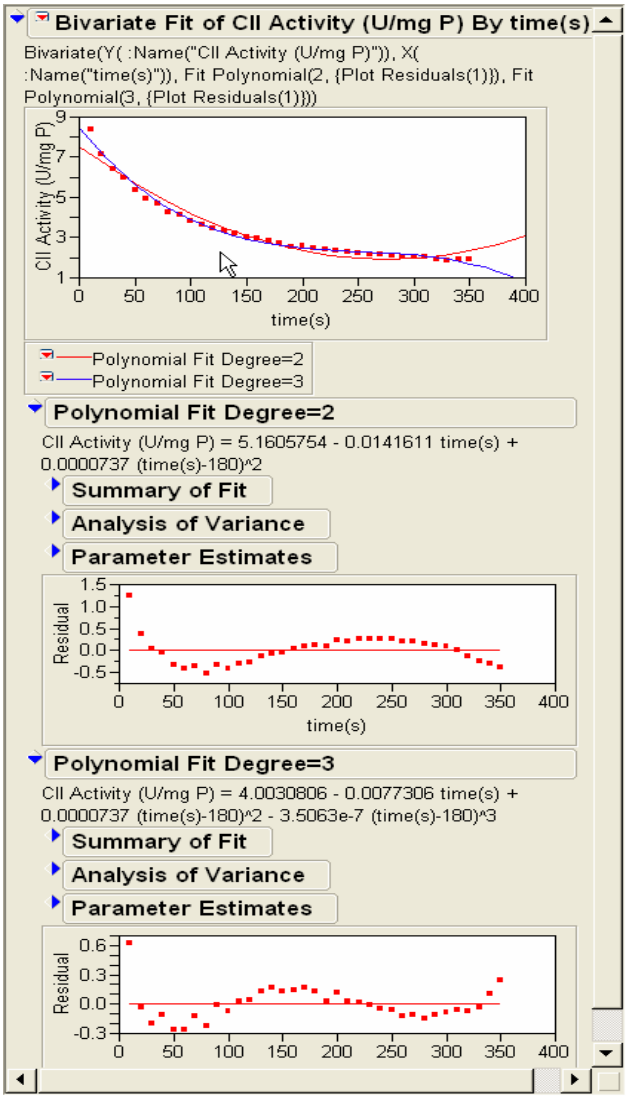
Other linear transformation of either X or Y or both such as sqrtX, sqrtY, 1/X and 1/Y were also tried and showed relatively good R^2 but obviously were not the best model for the data set in addition to the difficulty associated with explaining the results of these model in the context of the study or in terms of biochemical applications of such transformations.

We show below one of these (SQRT(y) vs Log(X), "blue line") "the most promising" that seems to do much better job of tracking the observations than the Log, Log such as shown below over imposed in the (Log(y) vs Log(X), "red line") fit . Again here you can see "the pattern of its residuals" below that exhibits similar structure of strong autocorrelation as before.



Since visual inspection of the data plot, might suggest that one might try a higher order polynomial model, we examined that possibility before we turn our attention to nonlinear models. The results do show that this approach will yield significantly worst fits and results as compared to what we have gotten thus far as sheen below.

Parameter estimates indicate that adding quadratic and/or cubic and quartic terms to the linear model significantly improve the model fit. But even the polynomial model of degree 2 does not seem to provide a very good description of the data relationships as it appears to capture the curvature at the beginning and end of the data curve. Although these concerns are minimized in the polynomial model of degree 3, the plot of residuals still indicate a strong curvature pattern, in addition to the complexity in interpreting polynomial models in biochemical terms, provide support for identifying more suitable model.



NONLINEAR PLATFORM

Having considered simple linear regression, linear transformations, and polynomial fits, my next step is to fit a non-linear model to my data. The Michaelis-Menten equation is a common model used in chemical kinetics. Since my data is a decay curve and measures enzyme activity through time and not enzyme activity at varying substrate concentration, we chose not to use Michaelis-Menten equation. Instead, we tried together models available in the literature and two of these are the one-phase exponential decay and a two-phase exponential decay. Note however, that in the case of radioactive decay, these models describe exactly what is going on physically. In the case of enzymatic measurements, these models maybe a simplification since the model considers that all receptors are either free or bound. Even though the model is simplified, it can predict experimental data very well. Even simple models can adequately predict the behavior of very complicated systems and can yield rate constants that have a physical meaning. Therefore, models that simplify true molecular or physiological mechanisms can be very useful, as long as they are not too simple.

In addition, the following table describes some of the terms and parameters that we will be using:

Variable	Remarks
X	Time, in seconds at 10-s interval
Y	Complex II activity, in U/mg mitochondrial protein 1U = micromoles/(time) per mg protein
Span	Difference between activity at $t_{initial}$ and plateau, same units as Y.
Plateau	Non-enzymatic activity, same unit as Y. It is the leveling-off or asymptotic value. Can be fixed at 0 or can be left as parameter.
K	Rate constant (sometimes also known as dissociation or association constant); expressed in units of inverse time. Equivalent to Km in the Michaelis-Menten equation, therefore, the lower the value, the better.
$t_{1/2}$	Half-life; the time it takes for Y to drop by 50%. It is equal to the natural logarithm of 2 divided by the rate constant: $T_{1/2} = \ln(2)/k = 0.693/k$.

First, let's use the simpler one-phase exponential decay.

I) One phase exponential decay model:

$$Y_i = a + be^{-k*Time_i} + e_i$$

Where:
a = plateau
b = span
k = rate constant
e_i = error

The first step is to add a new column formula and define the parameters and giving them initial values as a starting point for the nonlinear model parameters that are suggested by the plot of the data:

Span = 7.9 (difference of the highest observation minus the lowest observation.

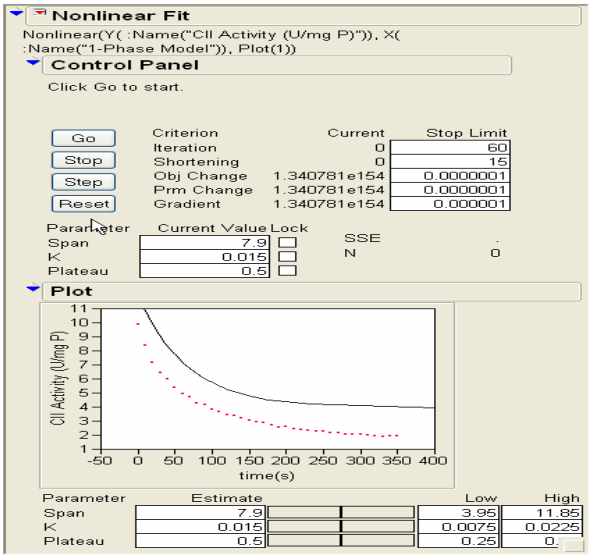
Plateau = 0.5 (Guess a number less than the lowest observation, since we are assuming that the complex activity will continue to decay through time.

K = 0.015 (Guess. Or can use the initial slope of the linear fit).

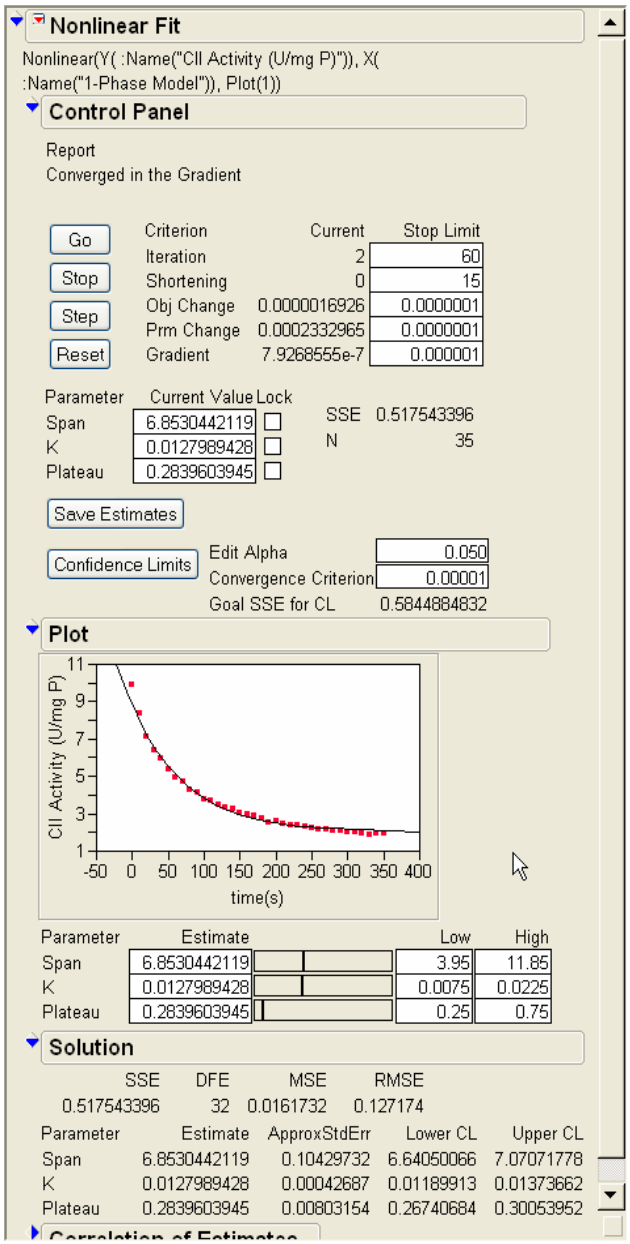
Next is to set up the formula for the model given as:

Y = Plateau + Span * (Exp(-K * time(s))

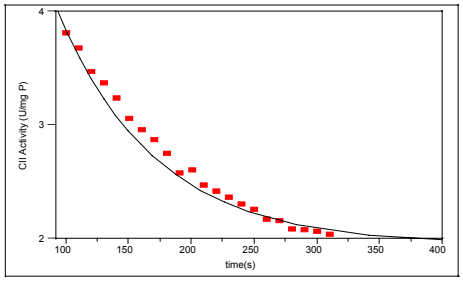
With the initial values shown in the initial Nonlinear Fir platform below



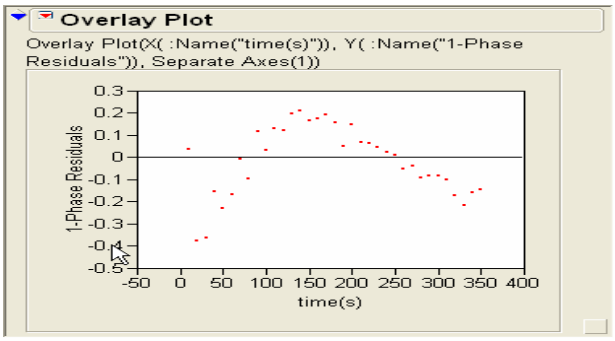
Below are the results when we select the GO button above and requesting additional iterations for asymptotic confidence intervals for the one-phase nonlinear parameter estimates:
NOTE:
* It took less than 5 iterations for the model to converge.
* The model estimate for Span, K and Plateau are 7.3, 0.014 and 0.28, respectively.
* In this model half-life is at 48 seconds, with a rate constant of 0.0144 (half-life=0.693/0.0144).



Since none of the parameter estimates have a confidence limit that contain zero, we do have some assurance that the parameters are different from zero. Span and Plateau however, are highly correlated. As we can see in the graph of the fitted model, the model seems to fit the data fairly well. But somewhere in the middle (t=130 to 185) the model appear to under predicts Complex II activity whereas at end of the curve (t=276 to 350) the model over predicts the data (see the plot below explored with the use of the magnifying tool).



In addition, like in linear model, another way of checking the quality of the fitted model is to check the residuals. Visibly from the residual plot plotted via the Overlay platform after saving the predicted values and creating a new variable for the residuals as the difference between the observed and predicted values the residuals of the fitted model still display a strong trend.



With that, let's consider using the other model reported in the literature, the two-phase exponential decay.

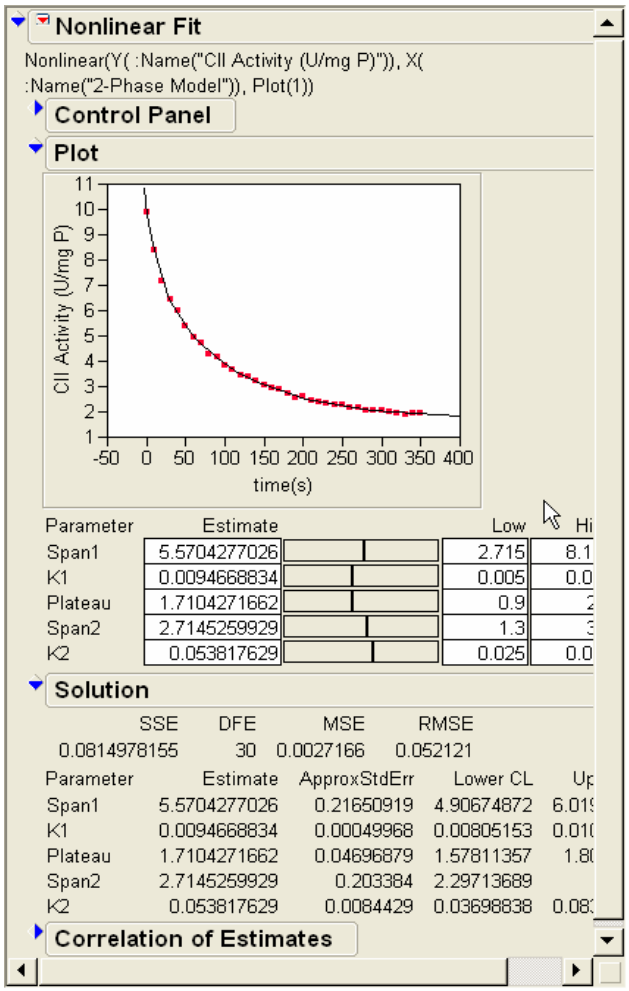
II) Two phase exponential decay model:

$$Y_i = a + b_1 e^{-k_1 * Time_i} + b_2 e^{-k_2 * Time_i} + e_i$$

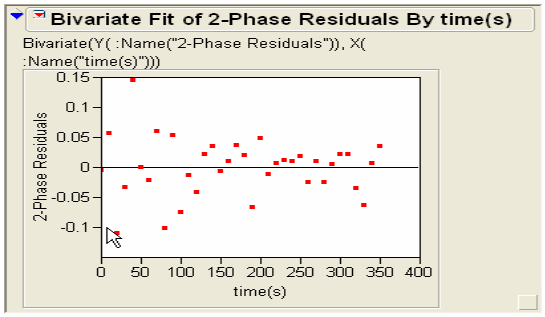
Where:
a = plateau
b_1 = Span 1 b_2 =Span 2
k_1 = rate constant 1
k_2 = rate constant 2
e_i = error

* Add new column (2-Phase Model)
* Set up initial value for parameters
  plateau = 1.8 (a number little bit lower than the lowest Complex II activity)
  Span 1 = 5.43 (it's usually a guess but one way to have initial value is to check the values of the original Complex II activity and look where the activity approximately slows down; e.g., between 70 and 80 seconds in this data set: [4.7 + 4.25]/2 = 4.47; then subtract this number from the highest observation: 9.9 – 4.47=5.43).
  Span 2 = 2.6 (guess, or [4.7 + 4.25]/2 = 4.47 – 1.9 (lowest observation))
  rate constant 1 = approximation, from graph appears lower than k2
  rate constant 2 = approximation, from graph appears higher than k1

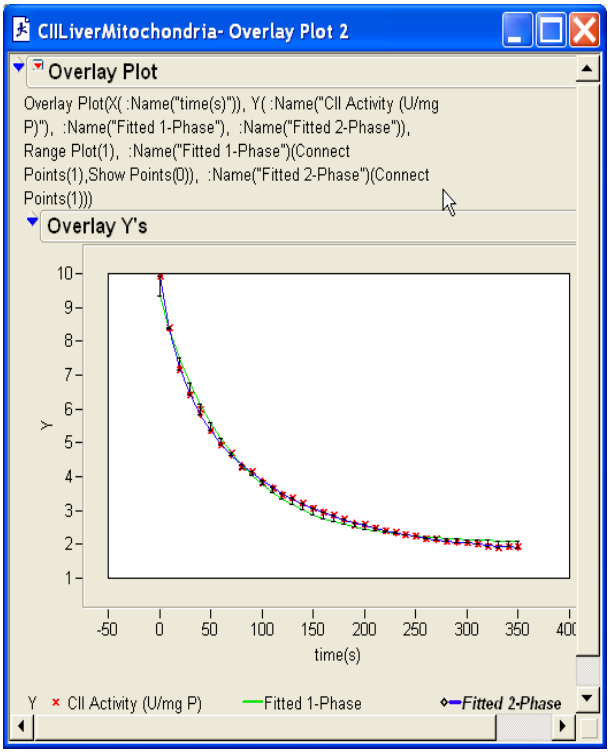
* Set up formula: **Ypred = Plateau + Span1 x $e^{-K1(time)}$ + Span2 x $e^{-K2(time)}$**
* Run model repeating the steps above and we get:



Therefore, the two half-lifes in this model are:
1). 0.693/k1 = 77 seconds
2). 0.693/k2 = 14 seconds (notice however that this doesn't start at time=0 instead, it starts at the end of k1, approximately at 168 seconds. [77 x 2] + 14 secs = 168 secs).
Neither of the parameters include zero, therefore we have some assurance that the parameters are different from zero. Not as apparent departure from the residual assumptions as indicated by the plot of residuals below.



Based on the above observations, the GOF measures ($RMSE_I$=.18 vs $RMSE_{II}$=.05), residuals and subsequent sensitivity analysis (not presented here) all indicate that the two-phase exponential decay model is better model in predicting Complex II activity in liver mitochondria from broilers. Below is an overlay plot of the two model, and indeed two-phase model seems to fit the relationships of the data better (as seen in the overlay graph below) than the one-phase model.



CONCLUSION

Based on all the analysis that was done on the data, we can conclude that Complex II activity in liver mitochondria of broilers follows a two-phase exponential decay model, when one substrate concentration is used and only one level of both enzyme and mitochondria sample were used. The model is: Ypred = 1.71 + 5.57 x e-0.009(time) + 2.71 x e –0.05(time). From this model we can also see that Y starts out equal to Span 1 + Span 2 + Plateau but decays to plateau with fast and slow components, and that the two rate constants differ 5-fold (0.009 vs 0.05). Some of the important implications of this relationship could include: valuable when studying inhibitors of this enzyme complex; in developing products (supplements or medicine) for disease associated with Complex II activity.
JMP and case studies make statistics courses "fun and exciting" and JMP easy learning curve allows for in class hands-on testing with very minimal prior experiences.

REFERENCES

Business Analysis Using Regression: A Casebook (Revised Edition 2002) by Dean P. Foster, Robert A. Stine, and Richard P. Waterman. Springer Verlag.

JMP® Statistics and Graphics Guide, Version 5 Copyright © 2002 by SAS Institute Inc., Cary, NC, USA.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Author Name: Andy Mauromoustakos
Company: U of Arkansas
Address: Agri Stat Lab, AGRX 101
City state ZIP: Fayetteville AR, 72701
Work Phone: (501) 575-5678
Fax: (501) 575-8643
Email: andym@uark.edu
Web:http://www.uark.edu/misc/andym/andy1.html