Using the F-test to Compare Two Models

When fitting data using nonlinear regression there are often times when one must choose between two models that both appear to fit the data well. After plotting the residuals of each model and looking at the r^2 values for each model, both models may appear to fit the data. In this case, an F-test can be conducted to see which model is statistically better[1]. It gives a definitive answer and does not rely on arbitrary interpretation of an r^2 value or residual plot. An F-test follows an F-distribution and can be used to compare statistical models.

The F-statistic is computed using one of two equations depending on the number of parameters in the models. If both models have the same number of parameters, the formula for the F statistic is $F=SS_1/SS_2$, where SS_1 is the residual sum of squares for the first model and SS_2 is the residual sum of squares for the second model. There are $N - V$ degrees of freedom, where N is the number of data points and V is the number of parameters being estimated (one degree of freedom is lost per parameter estimated). The resulting F statistic can then be compared to an F-table to extract the p-value. Alternatively, the p-value can be computed using MATLAB's built in function 'fcdf'. To do this simply type P=1-fdcf(F,df$_1$,df$_2$), where F is the computed F-statistic, and df$_1$ and df$_2$ are the degrees of freedom of each model equation. **If the p-value is large (greater than α) then the first model is statistically better than the second. If the p-value is small (less than 1-α) then the second model is statistically better than the first.**

If the models have different numbers of parameters, the formula becomes:

$$F = \frac{(SS_1 - SS_2)/(df_1 - df_2)}{SS_2/df_2}$$

The sum of squares for each model and the degrees of freedom for each model are calculated as before (note the models will have different degrees of freedom for this case). Additionally, the first model must be the one with fewer parameters (i.e. the simpler one). Once again, the F-statistic and degrees of freedom can be used to determine the p-value. Use df$_1$-df$_2$ and df$_2$ degrees of freedom when finding the p-value. **In this case a p-value less than α indicates that the more complex model (denominator of F-statistic) fits the data significantly better than the simpler model.**

[1] Note that the F-test tells you nothing about the physical significance of the model. It is ok to pick a model that does not fit the data as well if it has physical meaning. However, for illustration purposes the following assumes you do not know the physical laws governing the process.

Example

Say you are given the following data on bacteria growth:

Time (Hours)	Number of Bacteria (x10^3)
2.5	10.07
2.8	11.07
5.4	14.59
6.5	20.70
9.2	27.94
9.5	31.50
11.0	38.04
13.3	49.90
14.6	60.72
16.4	75.57

You want to determine if the bacteria growth is best modeled by an exponential model or a power law model. These models have the form:

$$N = Ae^{at} \text{ (exponential model)}$$
$$N = Bt^b \text{ (power law model)}$$

where N is the number of bacteria, t is the time passed, and A, a, B, and b are parameters to be calculated by nonlinear regression. MATLAB can be used to fit each model to the data. This process will not be discussed in detail here (see *Conducting a Nonlinear Fit Analysis in MATLAB*). Instead, we will focus on comparing the two nonlinear models.

First, let's look at the models plotted with the data points to get an idea of which model appears to be the better fit.
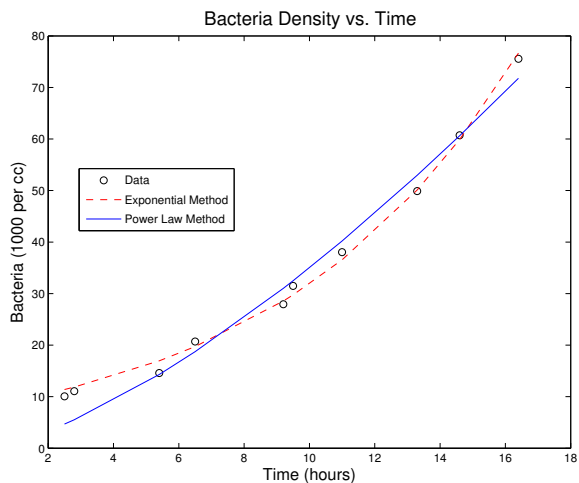

Figure 1: Bacteria density with nonlinear models

From this plot, it appears as if the exponential model fits the data better. But is this statistically significant? Let's look at some tests we can do to determine the better fit:

- Residual Plot analysis

  The residual plots for each of the two models are presented below. If you are unsure of what a residual plot is, see the BOSS section on Regression.
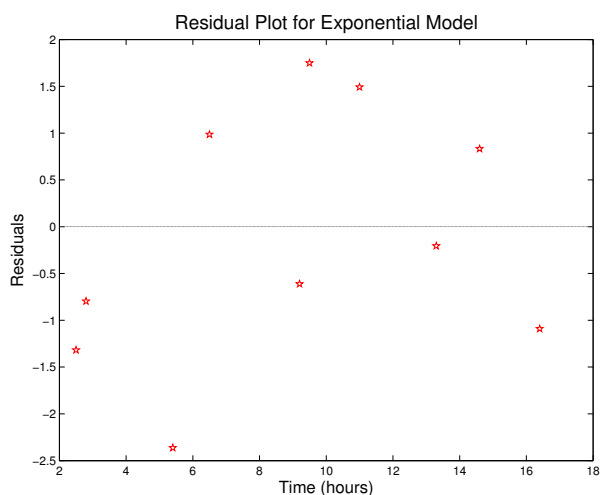

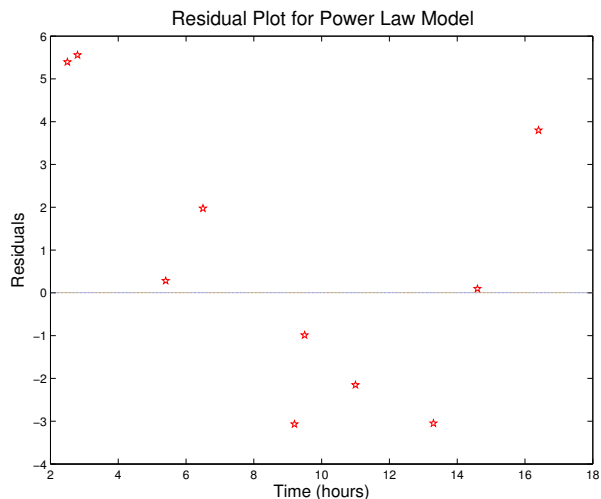Figure 2: Exponential Model Residual Plot

Figure 3: Power Law Model Residual Plot

The residuals for model 1 (the exponential model) are smaller in magnitude than the residuals for model 2 (power law). It also appears as if the residuals in model 2 follow a parabolic pattern. However, both

models appear to have relatively small, scattered residuals and it is difficult to conclude one model is definitively better than the other one.

- r^2 analysis

  The r^2 value can be calculated for each model based on the equation $r^2 = 1 - (SS_{residual}/SS_{Total})$. If you do not understand this equation see the *Fitting Curves to Data using Nonlinear Regression* document. Using this method, the following r^2 values were determined for each of the models:

$$\text{model 1 (exponential): } r^2 = 0.9963$$
$$\text{model 2 (power law): } r^2 = 0.9767$$

  Although the r^2 value for the exponential model is slightly greater than the r^2 value for the power law model, they are both indicative of a good fit. It cannot be concluded with certainty that one model is better than the other.

- F-statistic analysis

  The F-statistic can be computed by first noting that both equations have the same number of parameters. Thus, the form of the F-statistic equation is: $F = SS_1/SS_2$. The sum of squares of each model can be computed by summing the square of the residuals for each model. In this case $SS_1 = 16.5310$ and $SS_2 = 102.6796$. The degrees of freedom for each model is equal to the number of data points minus the number of parameters. There were ten data points and two parameters for each model so they both have eight degrees of freedom. Computing F gives: F = 0.1608. The corresponding P-value is 0.9909. The null hypothesis (that model 1 is simpler) is rejected if the F-statistic calculated from the data is greater than the critical value of the F-distribution for some desired false-rejection probability. Interpreting this result, it is clear that the F-statistic is not greater than the critical F, the null can not be rejected, and model 1 is statistically better than model 2.

  Let's look more in depth at how the F-test works. Notice in the simplest case, the F-statistic is just the sum of squares of the first model divided by the sum of squares of the second model. We expect that the model with the lower value for the sum of squares will fit the data better because this number reflects the total distance the model is from the data points and this was minimized during the regression procedure. Based on the sum of squares, we expect our result to indicate that model 1 is better than model 2 (a high p-value). This is best understood when compared to the F-distribution.
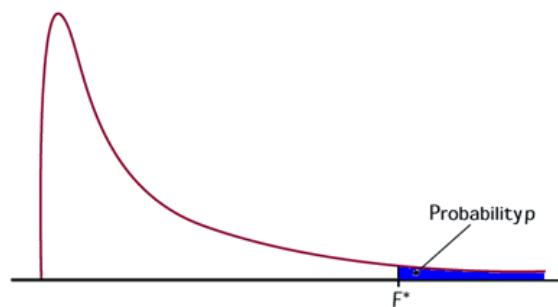

Figure 4: Example of F-distribution

The x-axis refers to the F-statistic, and the total area under the curve is equal to 1. Note the exact shape of the curve will vary depending on the degrees of freedom, but it will follow the same skewed trend. For an F-statistic of 0.1608, it is clear that the probability should be large. However, when reporting p-values small ones are often used. Notice that if we had reversed the sum of squares in the F-statistic and did the same computation we would have obtained a p-value of 1-0.9909=0.0091. This is the value often reported in published papers and is the value MATLAB gives if you don't subtract the result of the 'fcdf' command from 1. In short, always look at your F-statistic and p-value and make sure the result makes sense with your knowledge of the F-distribution. In general, low F-statistics should correspond to high p-values and

the model with the higher sum of squares should be shown to be the worse model (although the result might not necessarily be significant enough to produce a p-value less than the critical one).

In conclusion, we have seen that although two models may *appear* to fit a set of data well, one can be statistically better. This statistical difference is best identified by using an F-test since it is difficult to tell with residuals and r^2 analysis. Moreover, this result makes physical sense. It is expected that bacteria growth will follow an exponential distribution.

```matlab
1
2    %% Data from table
3    T=[2.5 2.8 5.4 6.5 9.2 9.5 11.0 13.3 14.6 16.4];
4    bacteria=[10.07 11.07 14.59 20.70 27.94 31.50 38.04 49.90 60.72 75.57];
5
6    % Plot the raw data and try to see if the data follows a known distribution
7    figure(1)
8    clf
9    plot(T,bacteria,'o')
10   ylabel('Bacteria (1000 per cc)')
11   xlabel('Time (hours)')
12   title('Bacteria Density vs. Time')
13   print -depsc BacteriaDensity
14
15   %% Define the model functions and fit
16   % Define Models and Initial Guesses
17   % Exponential model
18   model1=@(A,T) A(1).*exp(A(2).*T);
19   guess1=[1 1];
20   % Power law model
21   model2=@(B,T) B(1).*(T.^(B(2)));
22   guess2=[1 1 ];
23
24   % Determine values of the parameters using nlinfit
25   [alpha,R1,J1,CovB1,MSE1] = nlinfit(T,bacteria,model1, guess1);
26   [beta,R2,J2,CovB2,MSE2] = nlinfit(T,bacteria,model2, guess2);
27
28
29   %% Plot models with raw data
30   N1=(alpha(1))*exp(alpha(2)*T);
31   N2=(beta(1)).*(T.^(beta(2)));
32   x=linspace(2,18,1000);
33   figure(2)
34   clf
35   plot(T,bacteria,'ok')
36   hold on
37   plot(T,N1,'--r')
38   hold on
39   plot(T,N2,'-b')
40   ylabel('Bacteria (1000 per cc)','FontSize',14)
41   xlabel('Time (hours)','FontSize',14)
42   title('Bacteria Density vs. Time','FontSize',16)
43   legend('Data','Exponential Method','Power Law Method',0)
44   print -depsc BacteriaDensityPlusModels
45
46   %% Plot Residual Plots
47   figure(3)
48   clf
49   plot(x,0,T,R1,'rp')
50   xlabel('Time (hours)','FontSize',14)
51   ylabel('Residuals','FontSize',14)
52   title('Residual Plot for Exponential Model','FontSize',16)
53   print -depsc ResidualPlotModel1
54
55   figure(4)
```

```
56   clf
57   plot(x,0,T,R2,'rp')
58   xlabel('Time (hours)','FontSize',14)
59   ylabel('Residuals','FontSize',14)
60   title('Residual Plot for Power Law Model','FontSize',16)
61   print -depsc ResidualPlotModel2
62
63   %% Compare fits to see which is better
64   % Compute sum of squares for each model and the total sum of squares
65   ssmodel1=sum(R1.^2);
66   ssmodel2=sum(R2.^2);
67   sstotal=sum((bacteria-mean(bacteria)).^2);
68   % Calculate r^2 for each model
69   r2model1=1-ssmodel1./sstotal;
70   r2model2=1-ssmodel2./sstotal;
71   % Calculate the F statistic, The null hypothesis (that model 1 is simpler) is rejected if the F
72   % calculated from the data is greater than the critical value of the
73   % F-distribution for some desired false-rejection probability (e.g. 0.05)
74   F=ssmodel1./ssmodel2;
75   df=length(T)-2;
76   P=1-fcdf(F,df,df);
```