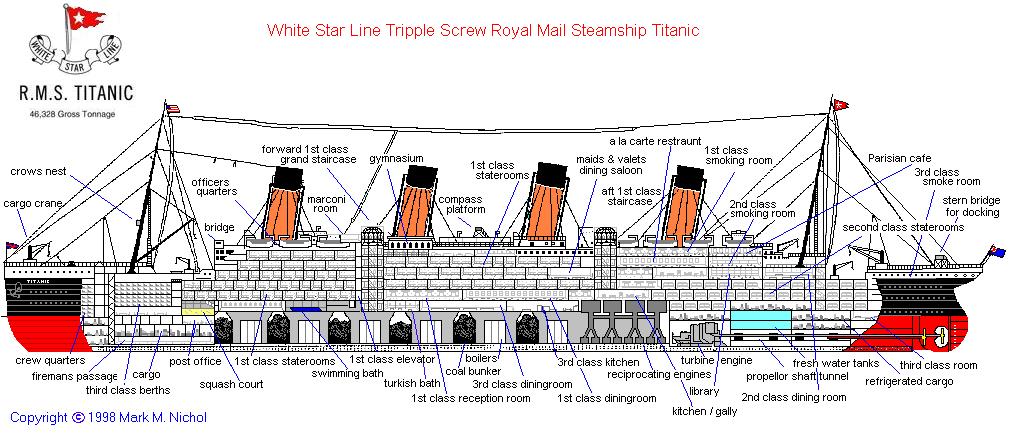
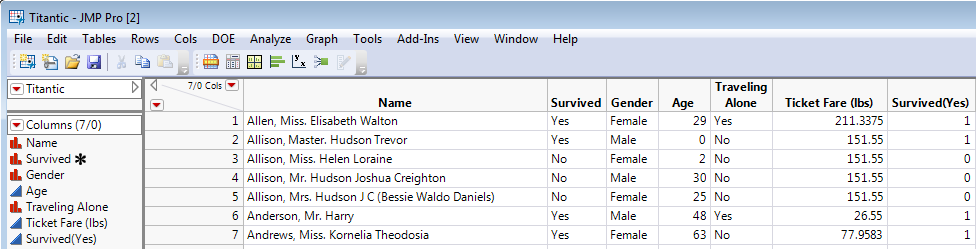
**18.1 – Introduction and Motivating Example**

In this section, we will introduce regression models in which the response variable is a binary categorical variable. The regression model we generally use for this type of situation is called a **logistic regression model**. Logistic regression is member of family models called Generalized Linear Models which we will also discuss in this section.

**Example 18.1 – RMS Titanic (Datafile: Titanic.JMP)**Consider the Titanic data from our course website. The response variable of interest is whether or not the person survived the Titanic disaster which is referred to as Survived in the dataset.

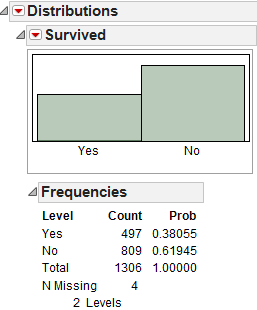


A snip-it of the dataset.



Goal: Understand the impact of the predictor variables (Gender, Age, Traveling Alone, and   
 Ticket Fare) on the response variable (Survived).

Before attempting to understand the conditional distribution of , let’s first consider the marginal distribution for survival. Because survived is a binary categorical variable, the summaries are fairly straight-forward – counts, percentages, and a bar graph for a visual representation.

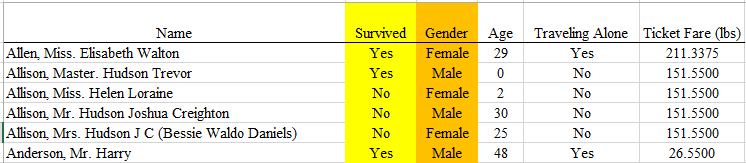


Give a brief description of what these summaries tell us about the marginal distribution for survival.

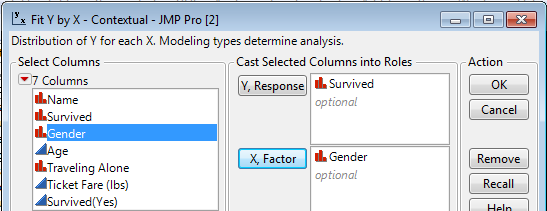
Definition of Odds for “Success”

Conditional Distribution of

Next, we will consider the impact of Gender on survival. In particular, the investigation will allow us investigate the impact from Gender on the likelihood of survival – or more simply the chance of survival.

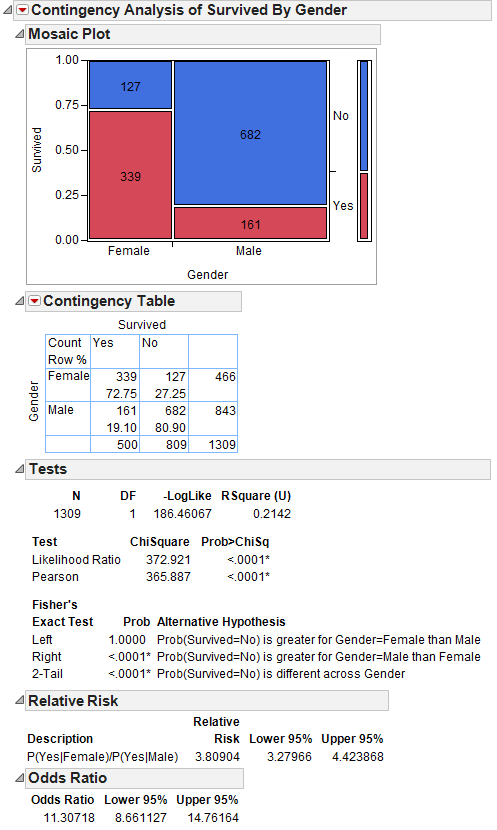


The **Fit Y by X** dialog box in JMP will be used initially as was done when the response was a continuous variable. To get started in JMP, select **Fit Y by X**, place Survived in the **Y, Response** box and Gender in the **X, Factor** box. Click OK.



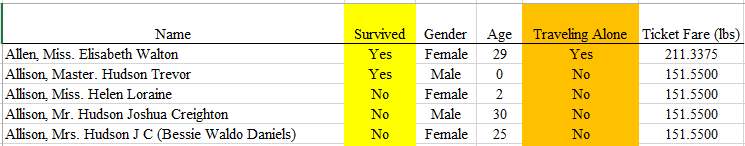
The following summaries are returned by JMP.

**Def’n:** Relative Risk (RR) and Odds Ratio (OR)

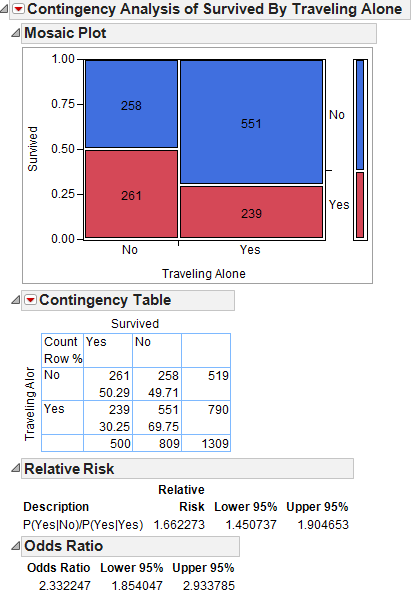


Conditional Distribution of Alone

Next, consider briefly the effect of Traveling Alone on the likelihood of survival.

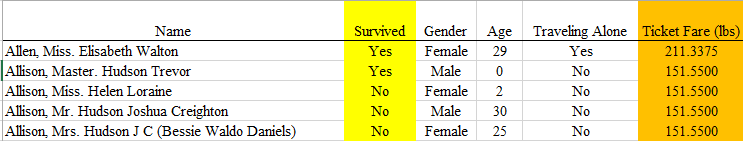


The following describes the conditional distribution of Survival | Traveling Alone

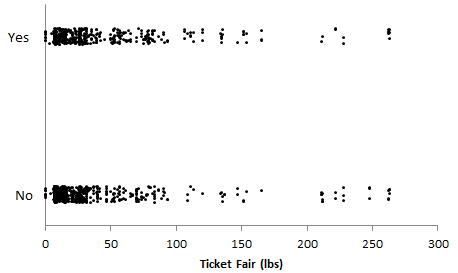


Which predictor, Gender or Traveling Alone, is better in explaining the likelihood of survival? Explain.

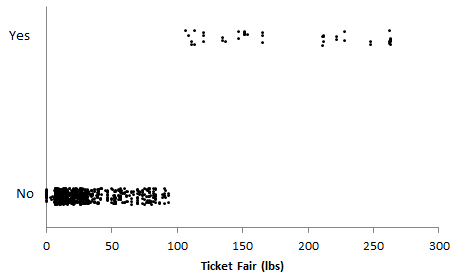
Conditional Distribution of (notice the predictor in numerical here)



I suppose a scatterplot type plot could be created and in doing so it might look something like the following.

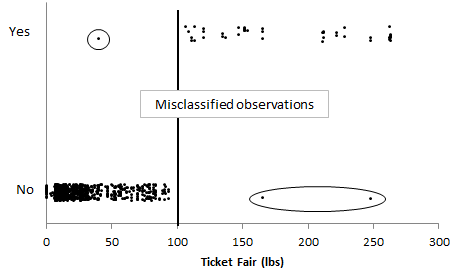


This scatterplot does not really provide much insight into the effect of Ticket Fare on the likelihood of survival. On the other hand, the plot below provides a great deal of information. Logistic regression can be used as a ***classification algorithm***.

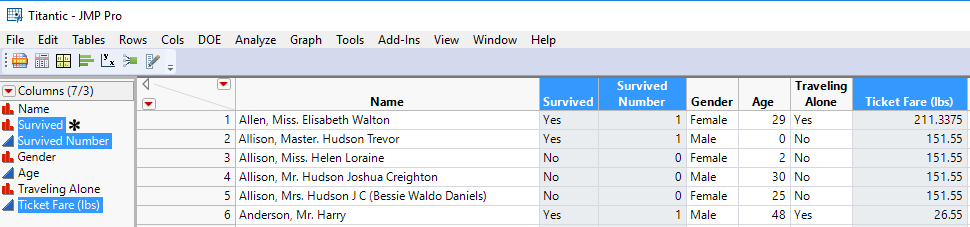


From the plot above, what can we say about the likelihood of survival condition on the Ticket Fare? Discuss.

Observations could be misclassified as well. In the following plot, the survival of 3 observations has been misclassified. Here, one individual who paid a low ticket Fare actually survived and two who paid a high ticket Fare did not survive. Our model would have made a mistake for these three observations. However, being wrong 3/1306 = 0.002 = 0.02% of the time is likely an acceptable misclassification rate.



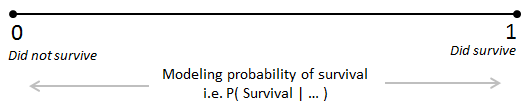
In order to make these types of plots, you will need to specify actual values for the Survived = Yes and Survived = No. The standard approach here would be to recode Survived = Yes to be a 1.0 and Survived = No to a zero as we did when creating terms from dichotomous nominal predictors .



Recoding Survived in the variable Survived Number above in JMP requires the use of a **Conditional > If** in the JMP Formula Calculator.

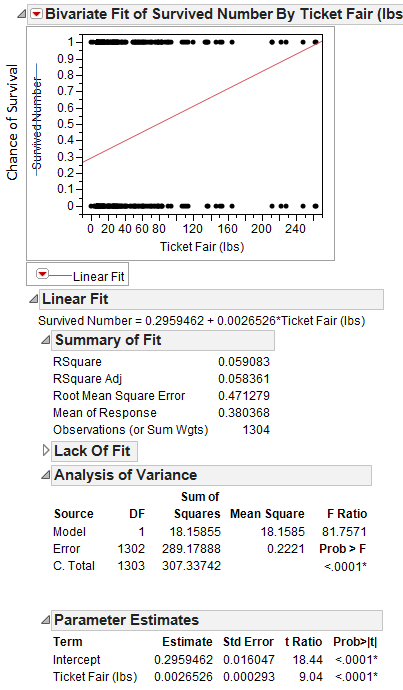
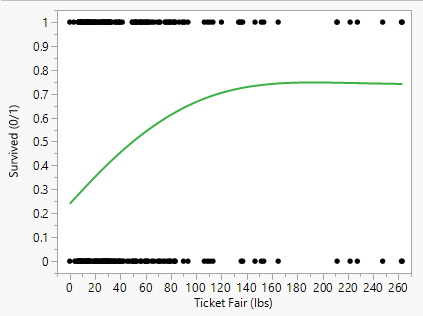


This type or recoding permits us to model the probability of survival. A smaller predicted probability translates to a smaller likelihood of survival and a larger predicted probability indicates an increase likelihood of survival. You must understand the coding being used by your software packages before attempting to interpret model outcomes. In particular, your software package may be modeling the probability of death in this situation instead of survival , which is perfectly acceptable if we want to focus on the negative outcome.



I suppose a naïve modeler could simply try to use the usual simple linear regression model. For example, our software packages (JMP/R) will certainly allow us to fit a model of the following form where Survived Number is treated as being numeric/continuous. As we will see, there are several reasons this model is inadequate.

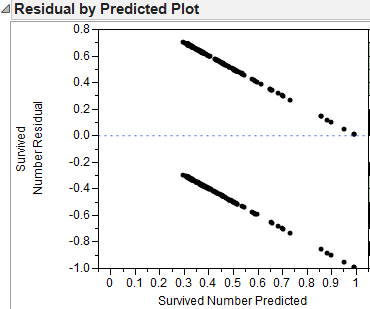
The outcome from JMP for this model are provide here. Note: This model is silly and should NOT be used!

Smoothing spline fit to scatterplot (better)

Residual Plot vs. )

Comments on residuals ()



As stated above, this model is inadequate and inappropriate for several reasons, the most straight-forward reasons are provided here.

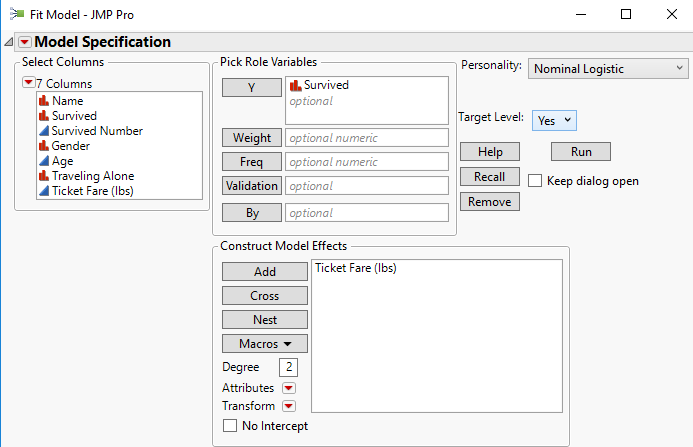
* A linear model permits predictions outside the range 0 to 1, which is not appropriate when modeling a chance of survival or more generally any probability. You can easily check this by saving the fitted/predicted values to the data table in JMP.
* The linear regression assumptions are violated, as shown in the plot of the residuals vs. the fitted values.

|  |  |
| --- | --- |
|  |  |

**18.2 - The Logistic Regression Model**

Because we are interested in estimating probabilities that are between 0 and 1, ideally we’d like to find a model where the range of possibilities from the predictions are constrained to be between 0 and 1. One such model is the ***logistic regression model***. The logistic model is a particular model in a broader family of models called **Generalized Linear Models (GLMs)**.

To fit a logistic regression model in JMP to predict survival rates based on age, make sure the response variable (*Survival*) is recognized as a categorical/nominal variable. Then, select **Analyze > Fit Model**, and enter the following.



JMP returns this output:

|  |  |
| --- | --- |
|  |  |

The simple (i.e. one predictor) logistic regression model has the following form.

where , i.e. the conditional probability of survival given ticket fare .

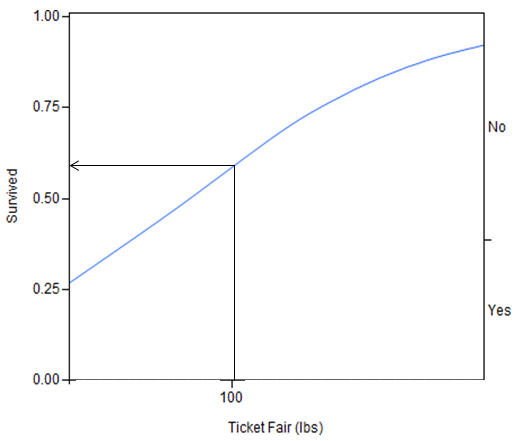
The right-hand side of this model looks familiar. The left-hand side is not something that we’ve seen before. The left-hand side ensures that in the end, the predicted probability will be between 0 and 1 – which is a fundamental requirement when modeling a probability. This function is called the *logit function (L)*, thus our model is modeling values for the logit . The logit is simply the natural logarithm of the odds for “success”, which is the odds for survival in this example.

In order to make predictions, we need to solve for in the following estimated logistic regression model.

Suppose one wants to predict the probability of survival given the individual paid 100 £ for their fare, i.e. . To obtain this predicted probability, we simple need to solve for in the equation below or use the general formula for the relationship between derived above.

Below we will solve for in the following.

Visually, we can see this value easily on the graph provided by JMP.



What would be the survival rate for a person who had paid just slightly more, say ticket fare = 101 £?

The most common approach to understanding the effect of a predictor/term like Ticket Fare is through the use of Odds Ratios (OR). Akin to what was done when investigating the conditional distributions above, e.g. and , the odds ratio is computed using the odds of survival at two specific values for the predictor. This is fairly straight-forward when the predictor is nominal and only has two levels. However the approach is pretty similar when the predictor is continuous and numeric as is the case here.

|  |  |
| --- | --- |
| Prediction at Ticket Fare = 100 £ | Prediction at Ticket Fare = 101 £ |
| Ratio of Odds for Survival | |

Questions

1. What is the interpretation of the odds of survival at Ticket Fare = 100 £.
2. What is the interpretation of the odds ratio in the context here?
3. What would it mean if the odds ratio were 1? Discuss.

It should be noted that a simple logistic regression model creates a ***constant odds ratio*** across the predictor variable. To convince ourselves of this, suppose we make a prediction for some other ticket Fare, say 240 £ and 241 £, i.e. a 1-unit increase in Ticket Fare. I have computed the odds and odds ratio here. It turns out, the odds ratio here will be exactly the same as the one computed above and this is true for any 1 unit increase in the predictor variable.

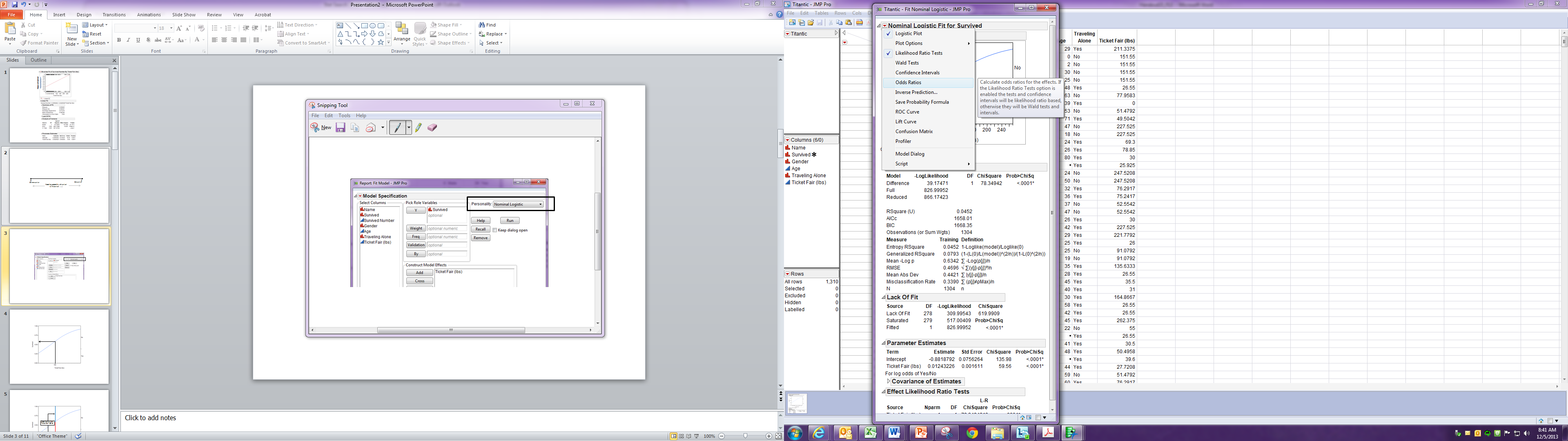
|  |  |
| --- | --- |
| Prediction at Ticket Fare = 240 £ | Prediction at Ticket Fare = 241 £ |
| Ratio of Odds for Survival | |

One can employ a little math and show that the odds ratio associated with a -unit change can be computed directly using the following.

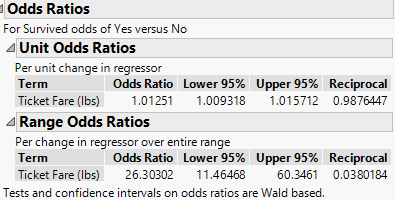
where *c* is the change in the units of predictor variable.

For example, in the above, the predictor variable was moved 1 unit, so and the estimated odds ratio is computed directly as

JMP automatically provides two odds ratios and these are obtained from the red drop-down arrow.



JMP provides two sets of Odds Ratios automatically. The Unit Odds Ratio is for a 1-unit increase in the predictor variable; whereas, the Range Odds Ratio is an odds ratio across the entire range of the predictor variable, i.e. using .



Questions

1. What is the interpretation of the unit odds ratio?

1. How do you think the reciprocal is computed?

The range odds ratio is automatically given and is the odds ratio across the entire range for Ticket Fare, from 0 (i.e. Jack Dawson) to 263 £, i.e. .

1. What is the interpretation of this odds ratio?

**18.3 – General Multiple Logistic Regression Model**

To review, logistic regression can be used to model a response that is a dichotomous nominal/categorical variable with levels generically called “success” and “failure”. For the RMS Titanic data we could view either survival or dying as a “success”, though in our analysis we have been viewing “survival” as a success. A nominal/categorical random variable with two levels in probability theory is called a *Bernoulli* random variable, i.e. where . The mean or expected value of a Bernoulli random variable is

In multiple logistic regression we are interested in understanding how relates to or can be modeled by a set of potential predictors , that is we are interested in developing a model for the *Conditional Probability* that given the **.**

or simply

However if we consider that the mean/expectation of a Bernoulli response variable is , we are really just interested in modeling the Conditional Expectation of the response given the potential predictors, i.e. the usual regression setting!

However because the conditional expectation must be between 0 and 1 the OLS regression model will not work! Thus we use the usual linear regression model for the logit function , i.e.

If we solve this equation for we obtain,

Thus this model is an example of a nonlinear regression model because the parameter estimates are imbedded in a nonlinear function . Model development is similar in that we need to decide which predictors/terms will give us the “best” model for the conditional probability of “success”.

**Odds Ratios**

The main way to interpret the parameter estimates in a logistic regression is by considering odds ratios associated with changes in the term/predictor values.

The logit function is the log of the odds for success ***given*** the predictor values . If we wish to compare populations defined by two sets of predictor values, say **,** in terms of their odds for success we take their ratio and we call this ratio the odds ratio (OR).

By exponentiating the difference in the logit functions we obtain the OR for the population defined the by the predictor values vs. the population defined by the predictor values .

For example, we could compare 50 year-old men traveling alone on the Titanic to 30 year-old women not traveling alone. Understanding the parameter estimates associated with individual terms/predictors in our logistic regression model makes use of this more general result.

**Interpreting Coefficients for Individual Continuous Terms/Predictors**

For example, consider the logistic regression of survival on the ticket price the passenger paid for the Titanic data. The simple logistic model would be,

Summary of the Logistic Regression

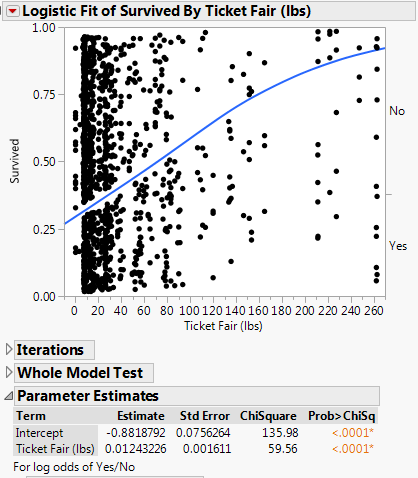
**Interpretation of the Coefficient for Ticket Fare** ()

Consider a -unit in the price paid for a ticket by a passenger. What is the OR associated with a -unit increase in the price paid? Thus we wish to contrast individuals with *Ticket Fare = x* to those who are *Ticket Fare = x + c*. 

Exponentiating both sides gives . Thus the multiplicative increase (or decrease if in odds associated with a *c* year increase in age is

This is how we interpret continuous terms in our model – we first pick an increment amount and then compute OR = .

of Survival on Ticket Fare



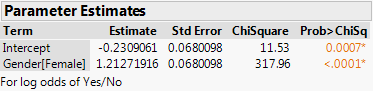
**Interpreting Coefficients Associated with Levels of a Nominal/Categorical Predictor**

For categorical/nominal predictors we can use contrast or dummy coded terms to represent all but one of the levels of the predictor. In JMP when performing logistic regression we only have contrast coding available! That is when coding Gender (M or F) in the Titanic data we would use contrast coding or dummy/indicator coding.

(Contrast) (Dummy/Indicator)

Consider the logistic regression of survival on Gender of the passenger only.

A summary of this simple model fit to the Titanic data is shown below.

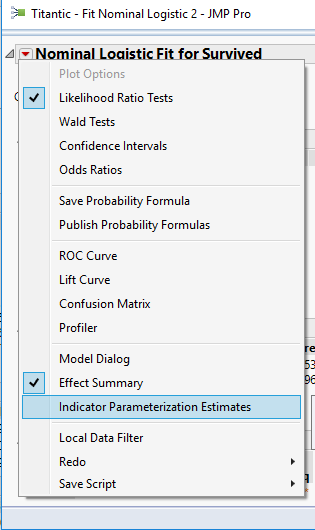


Let’s calculate the OR for survival associated with being a female passenger.

and thus the . For the model fit using the gender term for the Titanic data we have:

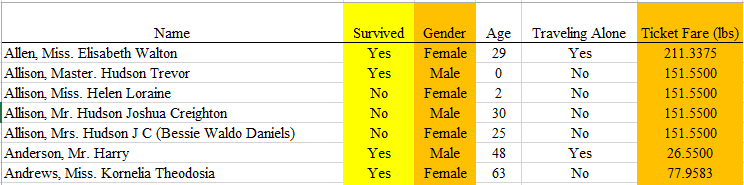
We still have to interpret each effect by holding all other terms/predictors constant, which has the same potential problems as OLS regression.

Using Dummy/Indicator Coding we have:





**Multiple Logistic Regression Model** -- C**onsidering two variables where one is categorical**



A multiple logistic regression model may improve our ability to successfully predict survival rate or probability. Such a model would allow us to consider two or more predictor variables simultaneously.

We can extend the simple logistic regression model given above

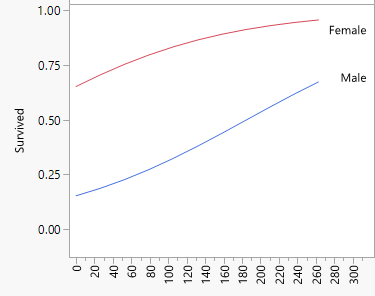
|  |  |
| --- | --- |
|  |  |

to include an effect due to Gender. Such a model will produce a predicted probability function akin to what was done in Section 11 (Factors and Terms).

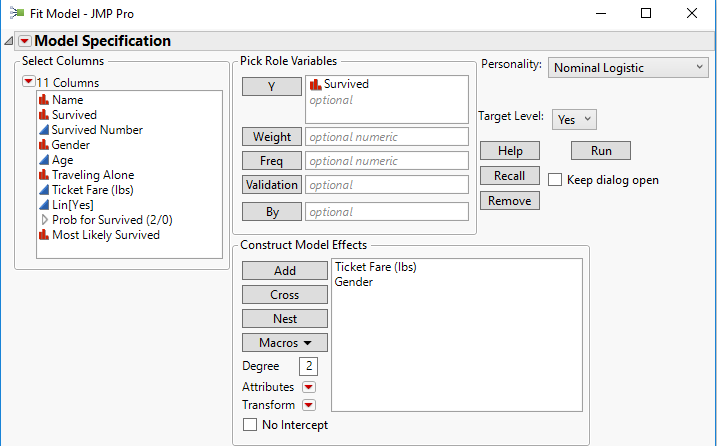
Using Graph Builder

|  |  |
| --- | --- |
|  |  |

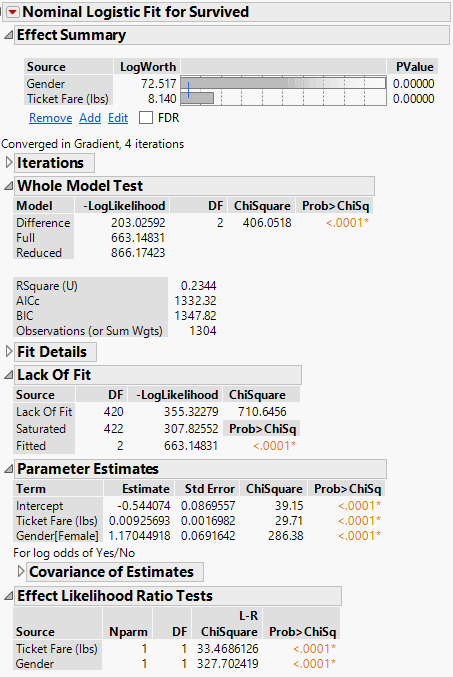
Using Interaction Plot in Profiler

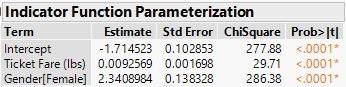


Getting this done in JMP

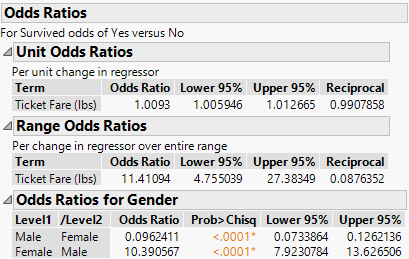


The following output is obtained after hitting Run.



Indicator Parameterization Estimates  


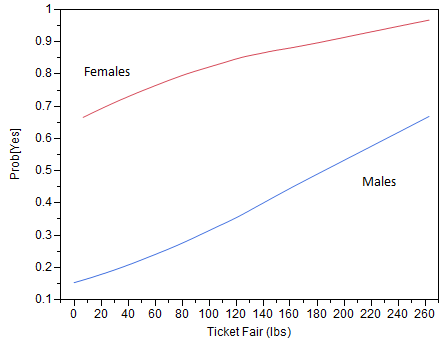
Once again, an investigation of the Odds Ratio is utilized to under the impact of each term in the model.



Unfortunately, JMP does not automatically produce the plot provided on bottom of page 384; however, this plot can be obtain by simply selecting **Save Predicted Formula**. JMP will compute all the predicted probabilities for each observation in your dataset.

|  |  |
| --- | --- |
| Saving out the predicted probabilities | The predicted probabilities and identification of the most likely outcome are provided by JMP. |

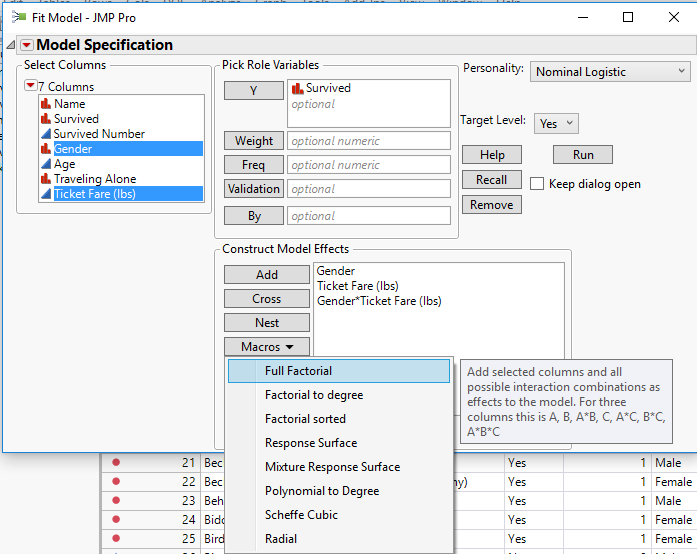
Once the predicted probabilities of survival are placed into your dataset, a plot similar to the following can be obtained fairly easily in JMP using the Interaction Profiler which we will later.



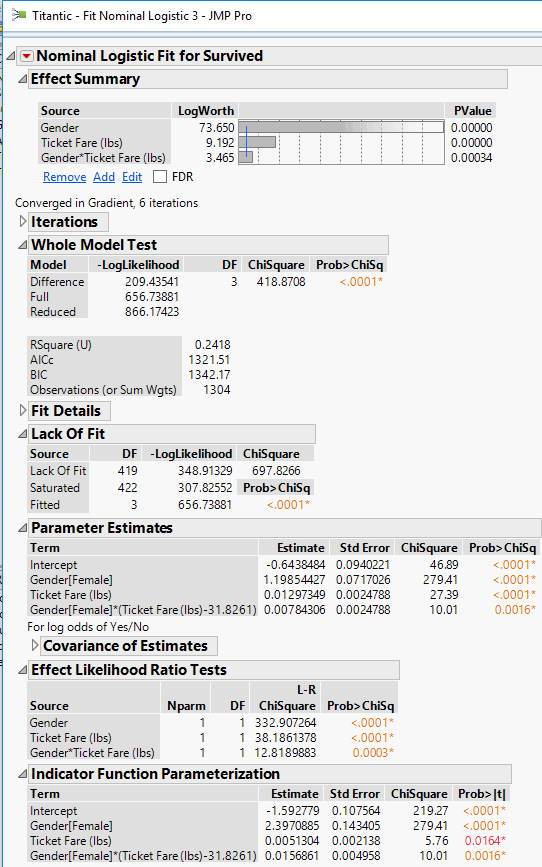
Comment: The model above did NOT include an interaction term between Ticket Fare and Gender. The plot above appears to indicate the estimated curves are not being *forced* in any way (remember a model without an interaction term forces the lines to be parallel, see Sections 10 & 11); however, the lack of interaction term is indeed adversely affecting our model though because the odds ratios – in a sense they are being “*forced to be parallel*.” Here that would mean that the odds for survival for female passengers are 10.39 times higher than that for men REGARDLESS of their ticket fare. This definitely may not be the case.

The logistic model used above could potentially be improved upon by including an interaction term between ticket fare and gender. This will alleviate the concern that the odds ratios are being restricted in some way.

Fitting this model in JMP.



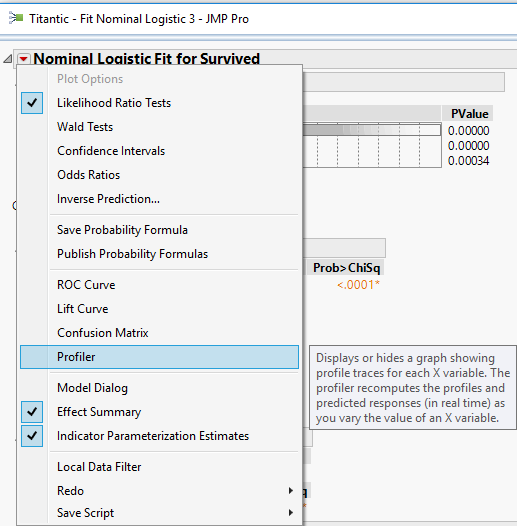
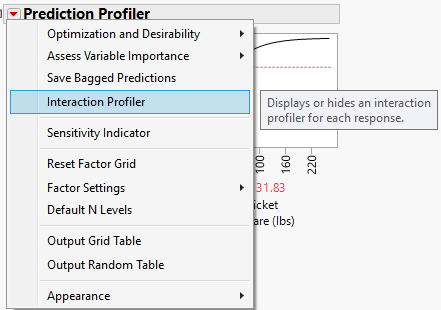
Model Summary from JMP



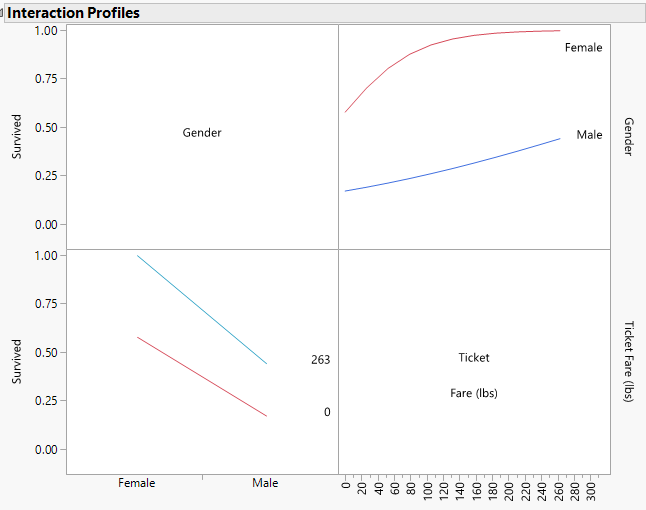
Comparing the two plots of the predicted survival rates

|  |  |
| --- | --- |
| Model with interaction term | Model without interaction term |

To obtain the probability profile plots shown above we can use Graph Builder as outlined above. An easier way is to use the **Profiler** in the **Nominal Logistic Fit** drop-down menu as shown below and then selecting **Interaction Profiler** in **Prediction Profiler** menu.

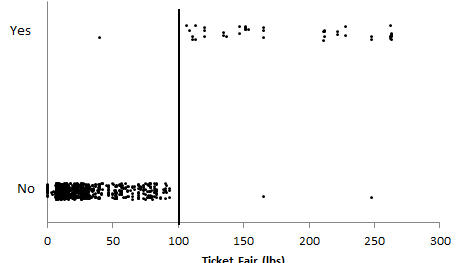
 

The resulting plot is shown below.

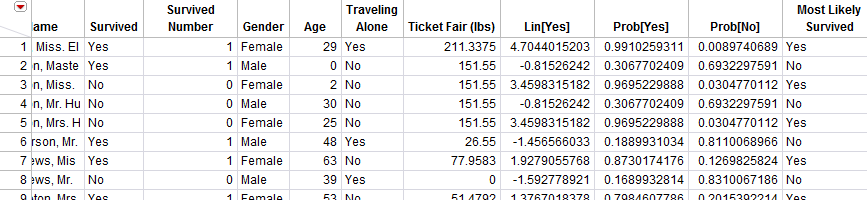


Concept of Misclassification Rate and Confusion Matrix

Consider once again the notation of observation being misclassified as was discussed briefly above.



When the **Save Predicted Formula** is selected, several quantities are placed into your dataset. These are shown here. The last variable is JMP attempt to make a decision regarding the survival of each individual. JMP computes for each observation and if this estimated probability is larger than 0.50, then JMP identifies the *Most Likely Survived* as Yes.

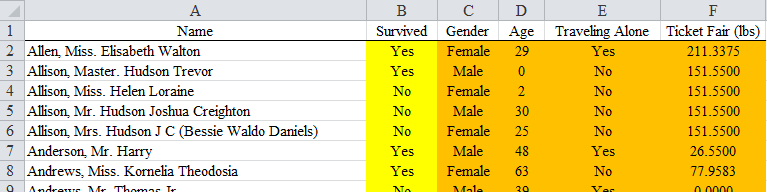


A summary of its ability to make these predictions is provided in a table which is commonly known as the Confusion Matrix. This can be obtained from the red-drop down menu in JMP.

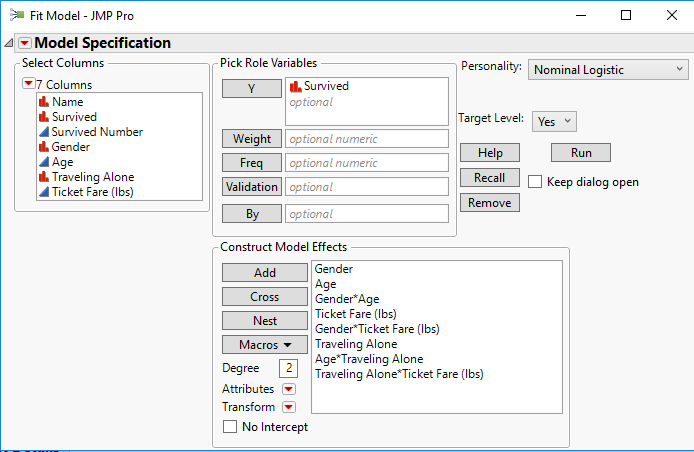
|  |  |
| --- | --- |
| The Misclassification Rate as computed in JMP | The Confusion Matrix is used to display how observations are being misclassified. |

How did JMP compute the misclassification rate?

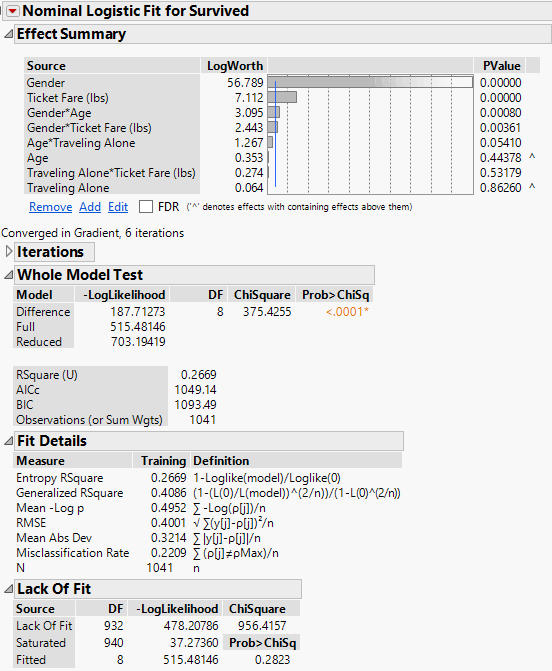
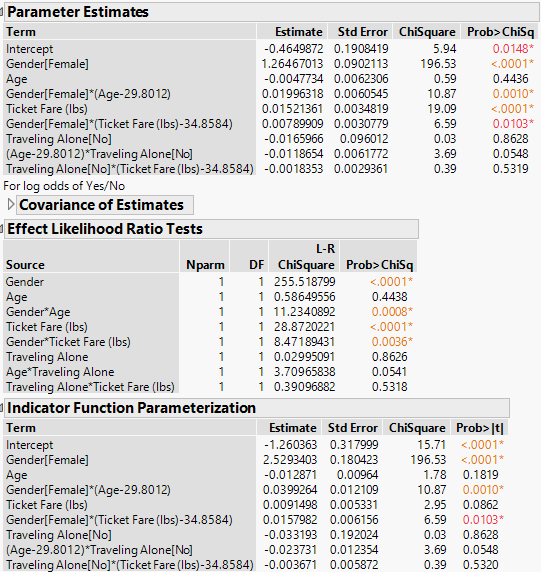
**Multiple Logistic Regression Model**: **Using all Variables**



Consider the following complete model. This model includes the interaction terms between each of the categorical and numerical predictor variables.

Fitting this full model in JMP  


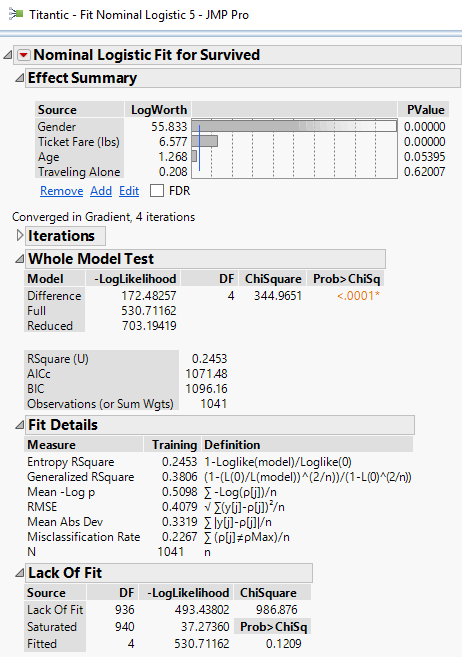
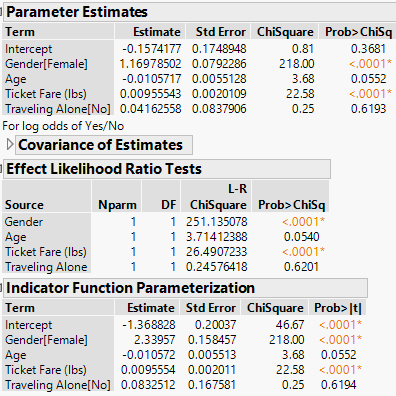
Summary measures, parameter estimates, and effect tests from this full model are given below.

Summary:

A much simpler model without all the interaction terms is provided here.

Notice the Entropy RSquare and Generalized Rsquare value dropped only slightly. Statistical tests exist to verify that this simpler model is sufficient in under the response. The “BIG F” test cannot be used, but analogous tests exist.

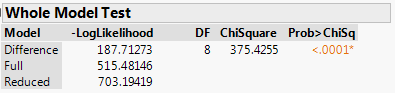
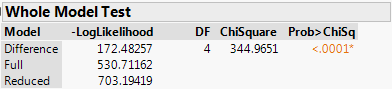
Summary:

**Big** Statistical tests exist to verify whether or not this simpler model is sufficient in order to sufficient understand the response variable. The “BIG F” test cannot be used, but rather we use the “BIG ” Test. More details on this test are found on page 394.

NH: Reduced model is sufficient.

AH: Full model is needed.

Full Model (with interactions) Reduced Model (no interactions)

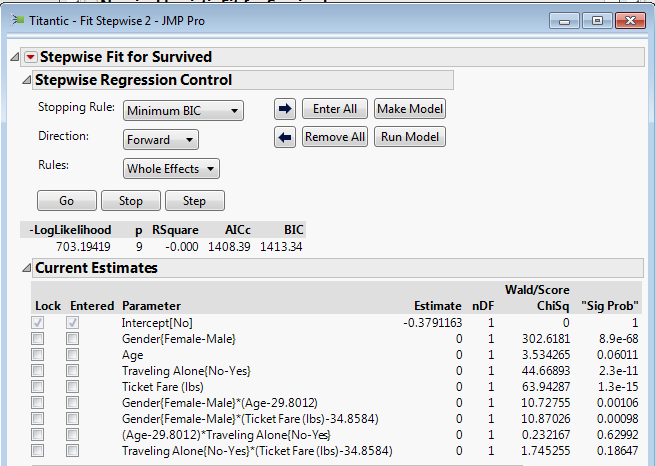
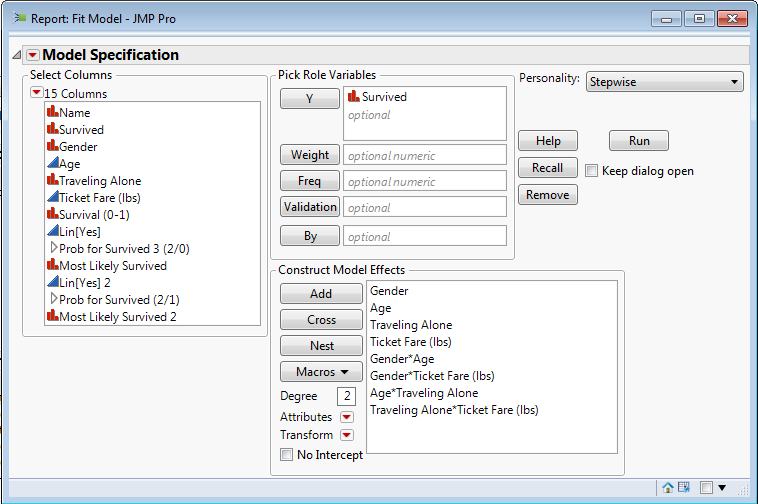
 

> pchisq(15.23,df=4,lower.tail=FALSE)

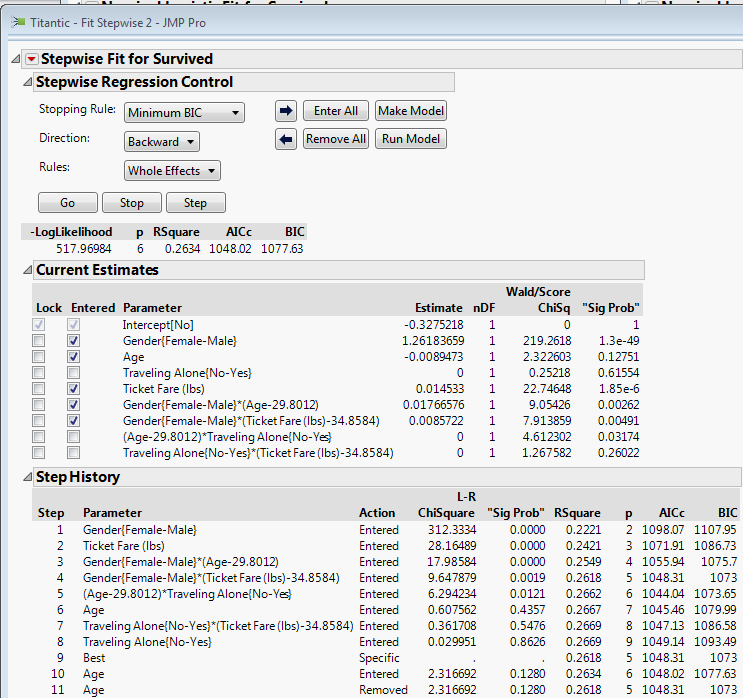
[1] 0.004247201

Thus the purely additive model (no interactions) is NOT supported. We can **use Stepwise Selection** methods to choose a reasonable sub-model starting with the full interaction model above defining the potential terms to consider.

**Stepwise Selection for Titanic Data**

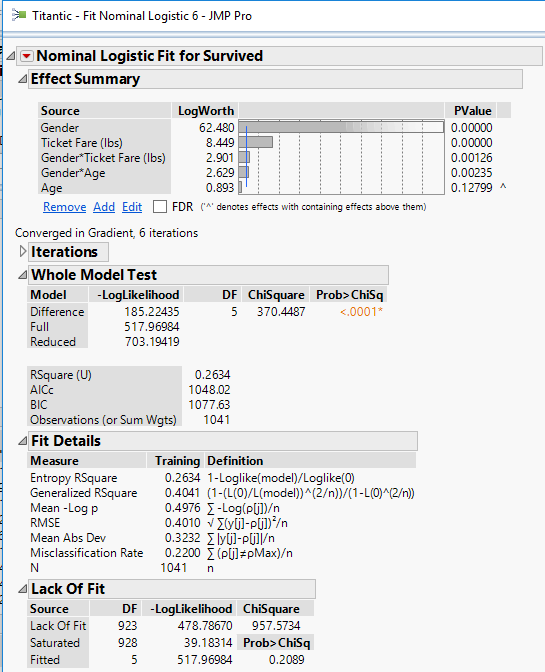
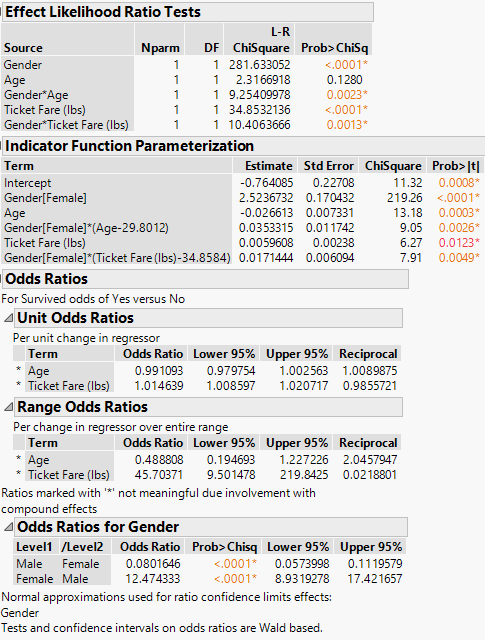


**Resulting Model**



Note: That both ticket fare and age are mean centered in the interaction terms with gender.

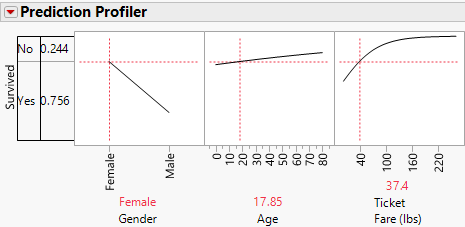
A summary of the stepwise selected model is shown below:

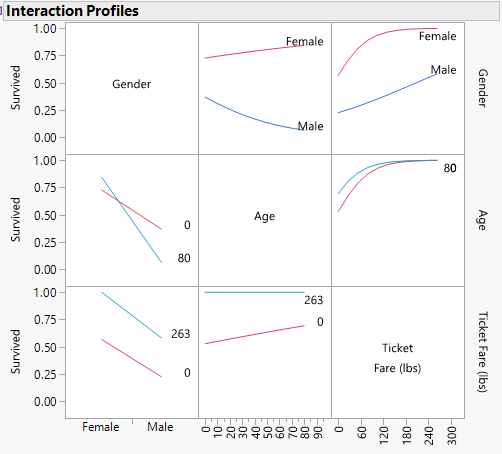
 

Notice the warnings about interpretation of the estimated OR’s for the continuous predictors that are involved in an interaction with the passenger’s gender. This is because predictors/terms involved in interactions **cannot** be interpreted marginally, i.e. without considering the level/value of the other term it is interacted with. The best way to understand a model involving interactions terms is through visualization. To do this in JMP we again select **Profiler** from the main model drop-down menu and then select **Interaction Profiler** from the **Prediction Profiler** drop-down menu as shown below.

Both the Profiler and Interaction Profiles for the model

are shown on the next page.

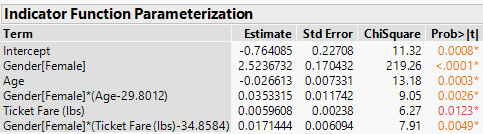




Comments:

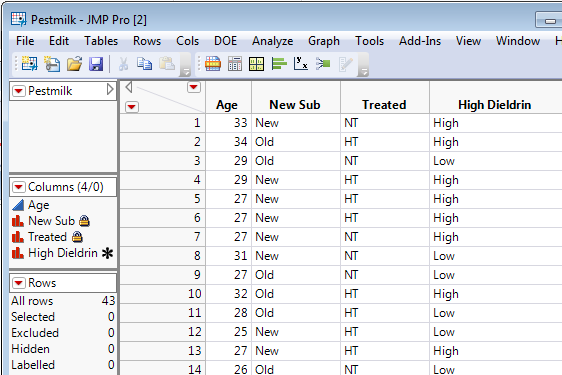
**Estimated Probabilities –** In the plots above the estimated probabilities of survival are shown. The two-step procedure below will allow for the computation of these probabilities for any predictor/term values we are interested in.

1. Find logit function value -
2. Compute



Examples:

**Example 18.2 – Dieldrin Levels in Breast Feeding Mothers in Western Australia   
(Datafile: Pestmilk.JMP)**  
This following is a study regarding breast feeding mothers in Western Australia. Earlier research discovered surprisingly high levels of pesticide levels in human breast milk.  The research conducted hopes to show that the levels had decreased as a result of stricter government regulations on the use of pesticides on food crops.   They found decreases for several types of pesticides; however, the levels of Dieldrin in human breast milk did not decrease and need to be further investigated.   By law new homes must be treated for termites in Australia.



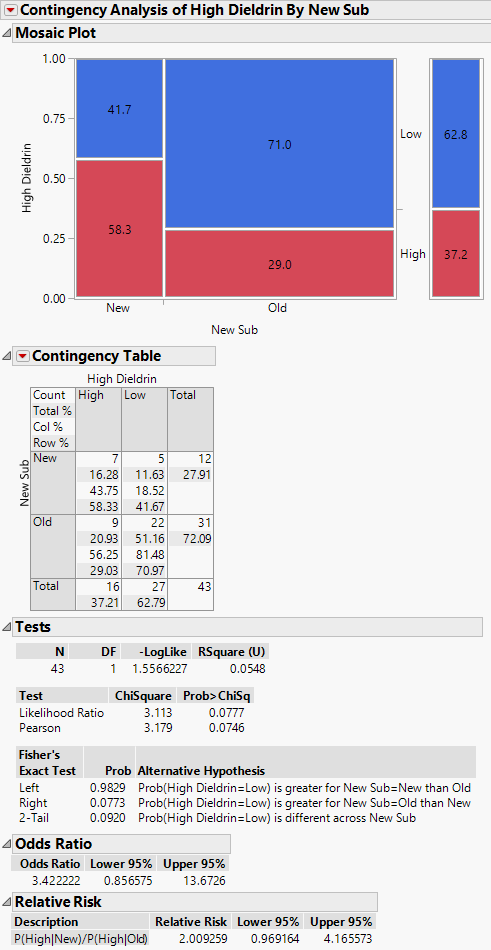
These data contain information on the mother's age in years, whether or not they lived in a new suburb, whether or not their house had been treated for termites within the past three years , and whether or not their breast milk contained above average (> .009 ppm) levels of the pesticide Dieldrin.

The variables in the dataset are:

1. Consider the conditional distribution of Dieldren | Suburb Type. How does the suburb type affect the likelihood of a having an increased risk of too much Dieldren?
2. Consider the conditional distribution of Dieldren | Treated. How does the treatment of home affect the likelihood of a having an increased risk of too much Dieldren?
3. Which predictor variable. Suburb Type or Treated, is more important in understanding the increased risk of having too much Dieldren?
4. Consider next, the conditional distribution of Dieldren | Age. Does age of mother appear to impact likelihood of having an increased risk of too much Dieldren ? Discuss.
5. Obtain the appropriate odds ratio for the logistic model in #4. Interpret this value. Use c = 5 yrs. as the increment.
6. Finally, consider the conditional distribution of Dieldren | (Suburb Type, Age, and Treated) by fitting an appropriate multiple logistic model. Explain/quantify the effect of each predictor. Are some predictors more useful than others? Discuss.

**Solution**

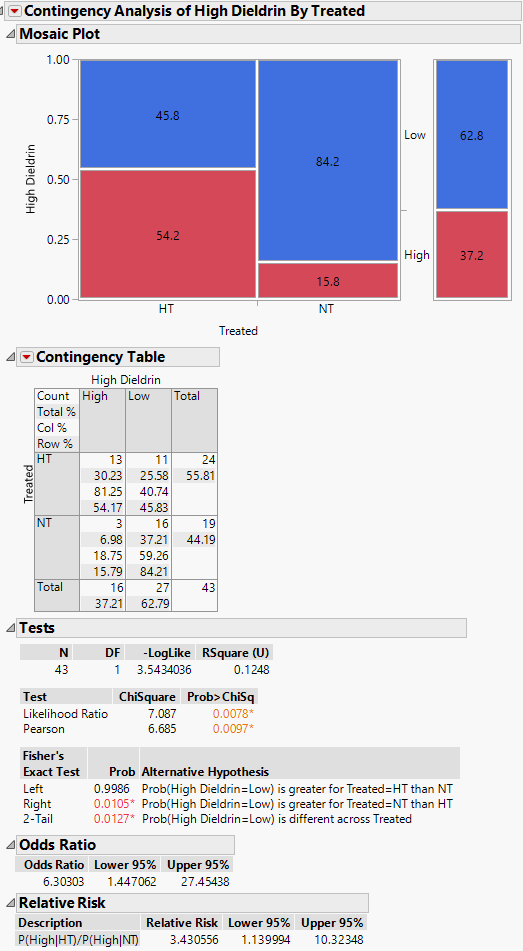
1. Consider the conditional distribution of Dieldren | Suburb Type. How does the suburb type   
 affect the likelihood of a having an increased risk of too much Dieldren?



Discussion

2. Consider the conditional distribution of Dieldren | House Treated. How does the treatment of   
 home affect the likelihood of a having an increased risk of too much Dieldren?

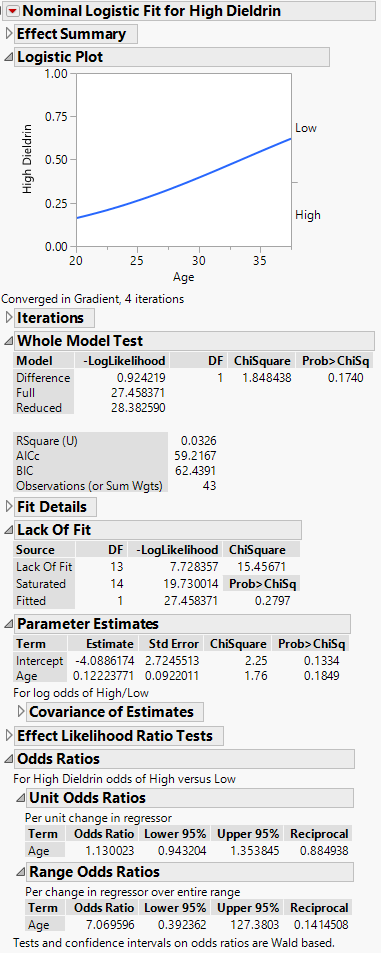
Discussion



3. Which predictor variable. Suburb Type or Treated, is more important in understanding the   
 increased risk of having too much Dieldren?

4. Consider next, the conditional distribution of Dieldren | Age. Does age of mother appear to   
 impact likelihood of having an increased risk of too much Dieldren ? Discuss.

5. Obtain the appropriate odds ratio for the logistic model in #4. Interpret this value. Use c = 5 yrs.   
 as the increment.

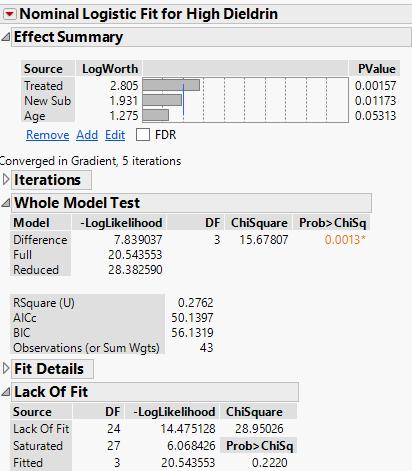
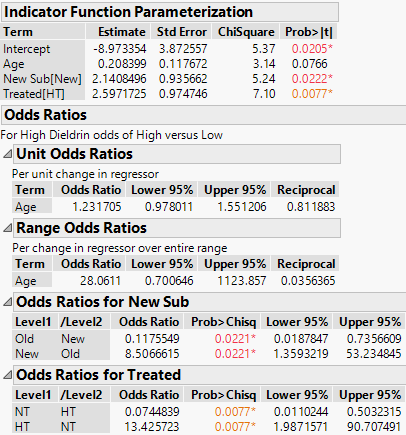


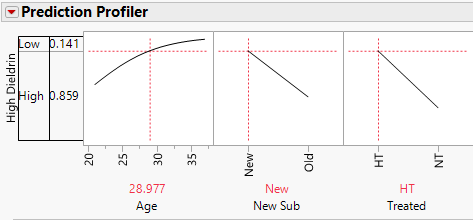
Discussion:

6. Finally, consider the conditional distribution of Dieldren | (Suburb Type, Age, and Treated) by fitting   
 an appropriate multiple logistic model. Explain/quantify the effect of each predictor. Are some   
 predictors more useful than others? Discuss.

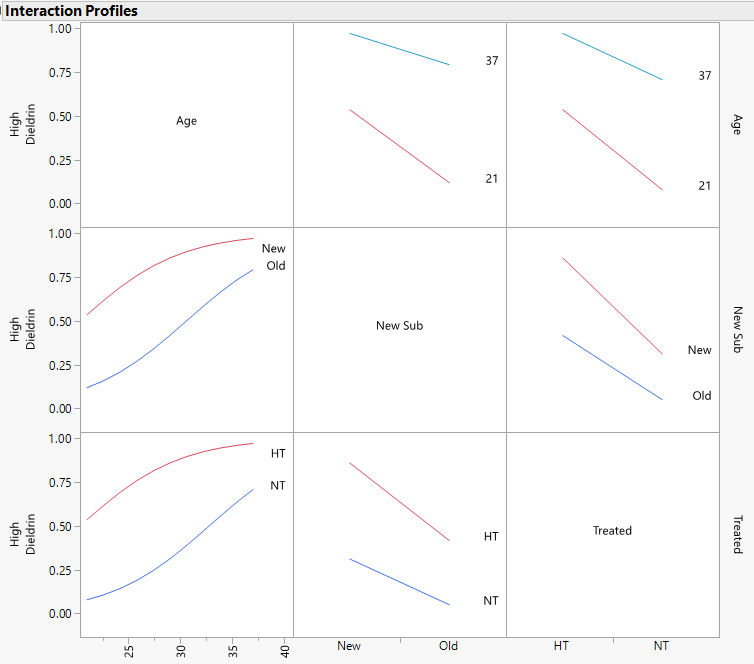
Assume

where



Comments:



Comments:

**18.4 – Generalized Linear Models (GLM)**

The logistic regression model that we have been studying is one type of a broader class of models called ***Generalized Linear Models***. Generalized linear models are an extension of regular linear models that allow: (i) the mean of a population to depend on a linear function of terms through a nonlinear ***link function***and (ii) the response probability distribution to be any member of a special class of distributions referred to as the ***exponential family***. The exponential family contains the normal distribution (OLS), the binomial distribution (logistic), and the Poisson distribution.

The link function is function that relates the mean of the response linearly to a set of terms based on the explanatory variables or predictors.

***OLS Regression***For a normally distributed response the link function is the identity function, thus,

or we typically write the model for the mean,

***Logistic Regression***  
For a binomial/Bernoulli response we know that

which we expressed as,

Recall: for a Bernoulli random variable.

Other members of the GLM family of models include:

Exponential, Gamma, Inverse Gaussian, Multinomial, and the Poisson. We will discuss the Poisson GLM model in the next section.

**18.5 - Poisson Regression**

“Recall” the Poisson distribution is given by

The response is a discrete random variable that represents the number of occurrences per time or space unit. In Poisson regression we seek a model for the mean of the response () as a function of terms based upon a set of predictors . For a Poisson random variable the mean and variance are both (i.e. and ), so traditional OLS will not be adequate because the constant error variance assumption would be violated.

For a Poisson distributed response the link function is so,

thus,

**Interpretation of Coefficients in the Poisson Regression Model**The interpretation of the coefficients in the Poisson regression model is as follows. Assume that we change one of the explanatory terms, for example, the first one, by one unit from *u* to *u+1* while holding all other terms fixed. This change affects the mean of the Poisson response by

To interpret consider the ratio of the mean function when term increases by 1 unit from to .

says the mean of the response gets a multiplicative increase by units per unit increase in the term .

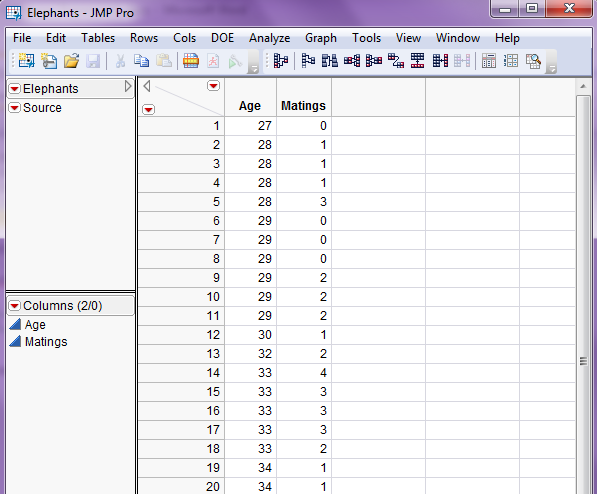
Wald Intervals and Tests for Poisson GLM Parameters

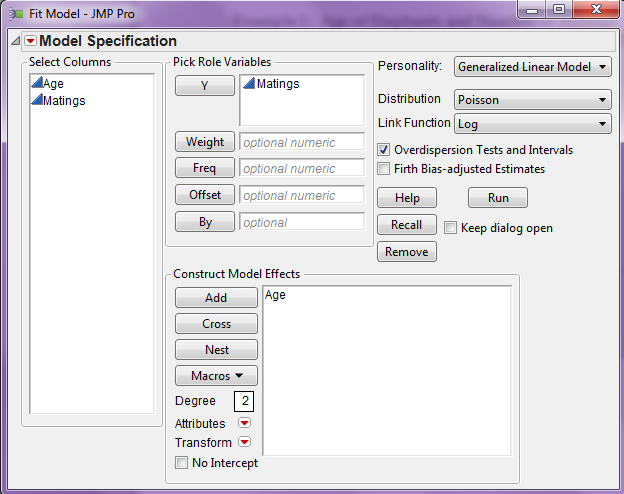
**95% CI for**  :

Therefore a CI for the multiplicative increase in the response is:

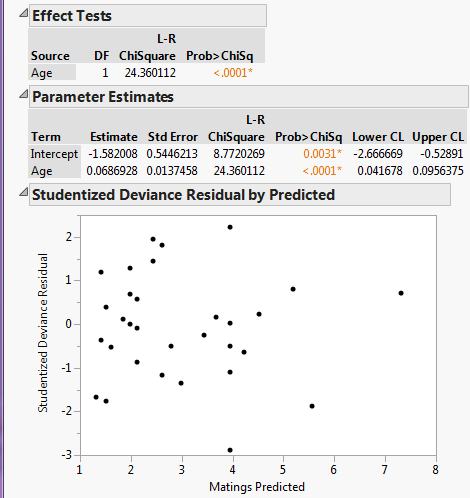
**95% CI for**  or  for a -unit increase.

**Example 18.3** - **Age of Elephants and Number of Matings** (**Data File: Elephants.JMP**)



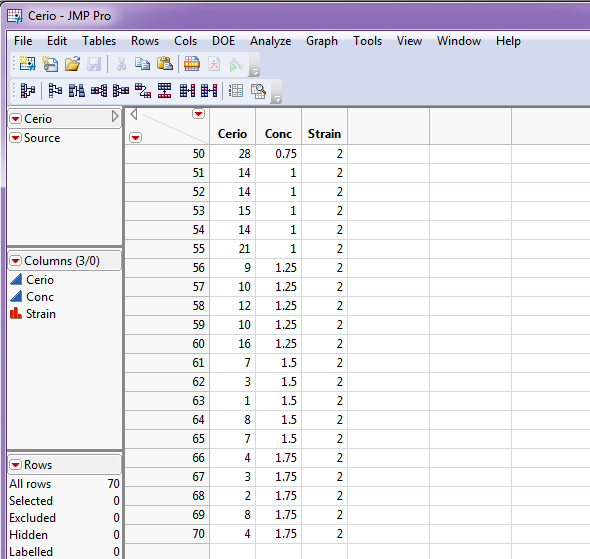
Select **Analyze > Fit Model** then change options in the dialog box as shown below:  


**Comments:**

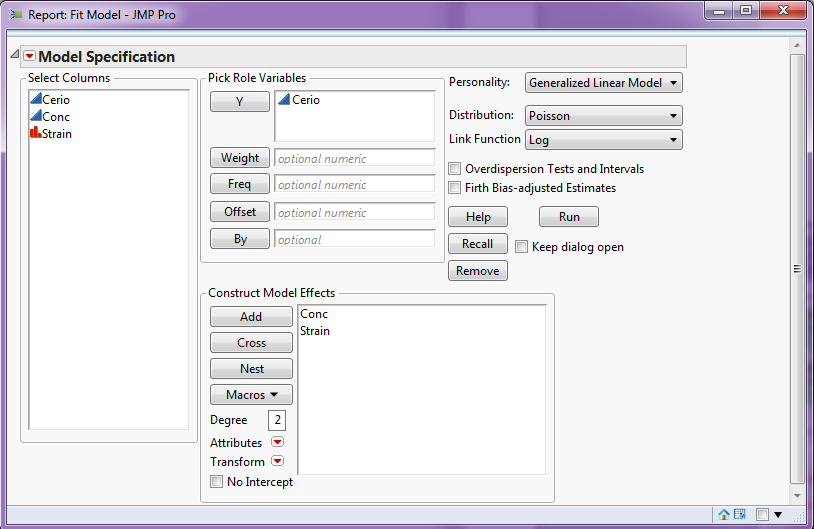


Interpretation of Coefficient for Age

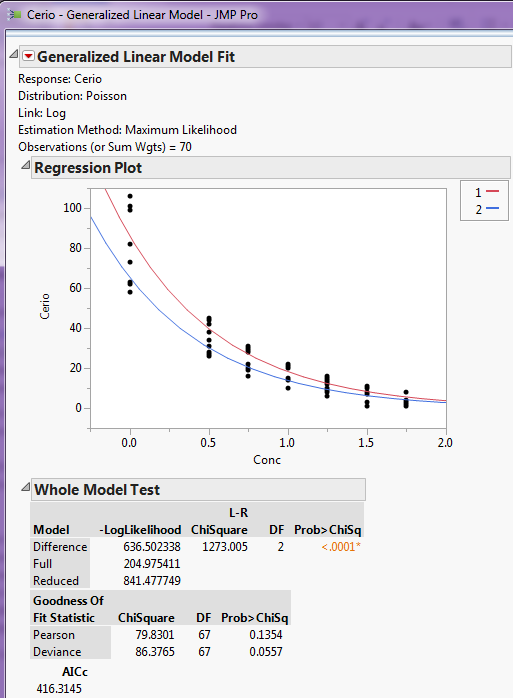
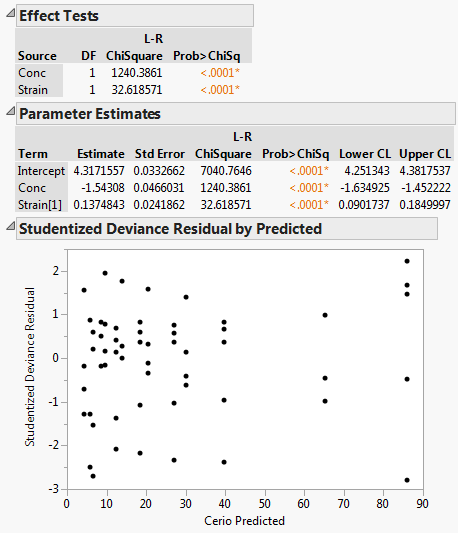
**Example 18.4 - Ceriodaphnia (Data File: Cerio.JMP)**

A portion of the data table is shown below:  


To fit a Poisson regression model for the number of ceriodaphnia as a function of the concentration and strain we again use **Analyze > Fit** Model to set up the model as shown below:



The results of the model fit are shown below:

We can clearly see from the regression plot that the ceriodaphnia count decreases with concentration and that strain 1 has higher count in general, with the largest difference at lower concentrations. Next we find the risk ratios associated with the concentration and strain.

Interpretation of Model Effects

Risk Ratio CI’s

**18.6 – General Inference for GLMs**

Below we present some general inferential tools for GLMs. As we have seen in the case of logistic/binomial regression and Poisson regression the parameters have special interpretations, but the mechanics of testing individual terms in our model or comparing full vs. reduced models is same regardless of the GLM model.

**Testing Individual Terms**

For testing the significance of a single term in our GLM adjusted for the other terms in the model, i.e.

we can use the large sample test for the significance of “slope” parameter ()

**General Chi-Square Test - (Big Test) for comparing nested models**  
Consider the comparing two rival models where the alternative hypothesis model

**General Chi-Square Statistic**

**=** (residual deviance of reduced model) – (residual deviance of full model)

******

If the full model is needed  is BIG and the associated p-value = will be small.

The deviance is like the RSS in an OLS model, however it differs depending on which GLM model we are using. The deviance for the models we are considering in this course are given below.

**Normal** *(usual OLS multiple regression)*

**Logistic**

**Poisson**

For any of these GLMs the deviance will always decrease when we add terms to our current model, so minimizing the deviance is NOT the criterion to use when model building. We need to consider the deviance decrease tempered against the number of parameters in our model using the General or Big- Test for comparing nested logistic or Poisson models. For the Normal model we use the Big F-test covered earlier.

**Residuals for GLMs**

Just as deviance (think RSS) is measured differently for GLMs, so are the residuals, which measure the discrepancy between the observed response values and the fitted values from the model. The deviance is an aggregation the of total variation represented in the residuals.

Logistic Model

* Deviance residuals

The deviance is simply the sum of the squared deviance residuals, i.e.

* Pearson residuals

Poisson Model

* Raw residuals (
* Pearson residuals
* Deviance residuals

As with scaled/studentized residuals in OLS regression, extreme values of these residuals indicate observations that are poorly fit, and that may possibly be classified as outliers. There are also measures for potential and influence, i.e. leverage and Cook’s distance for GLMs as well although we will not consider them here.