



# Fitting Regression Models Containing Categorical Factors

Presented by  
**Dr. Neil W. Polhemus**

# Virginia Piedmont



# Type of grass affects population of wild birds

Cool season grasses



Warm season grasses

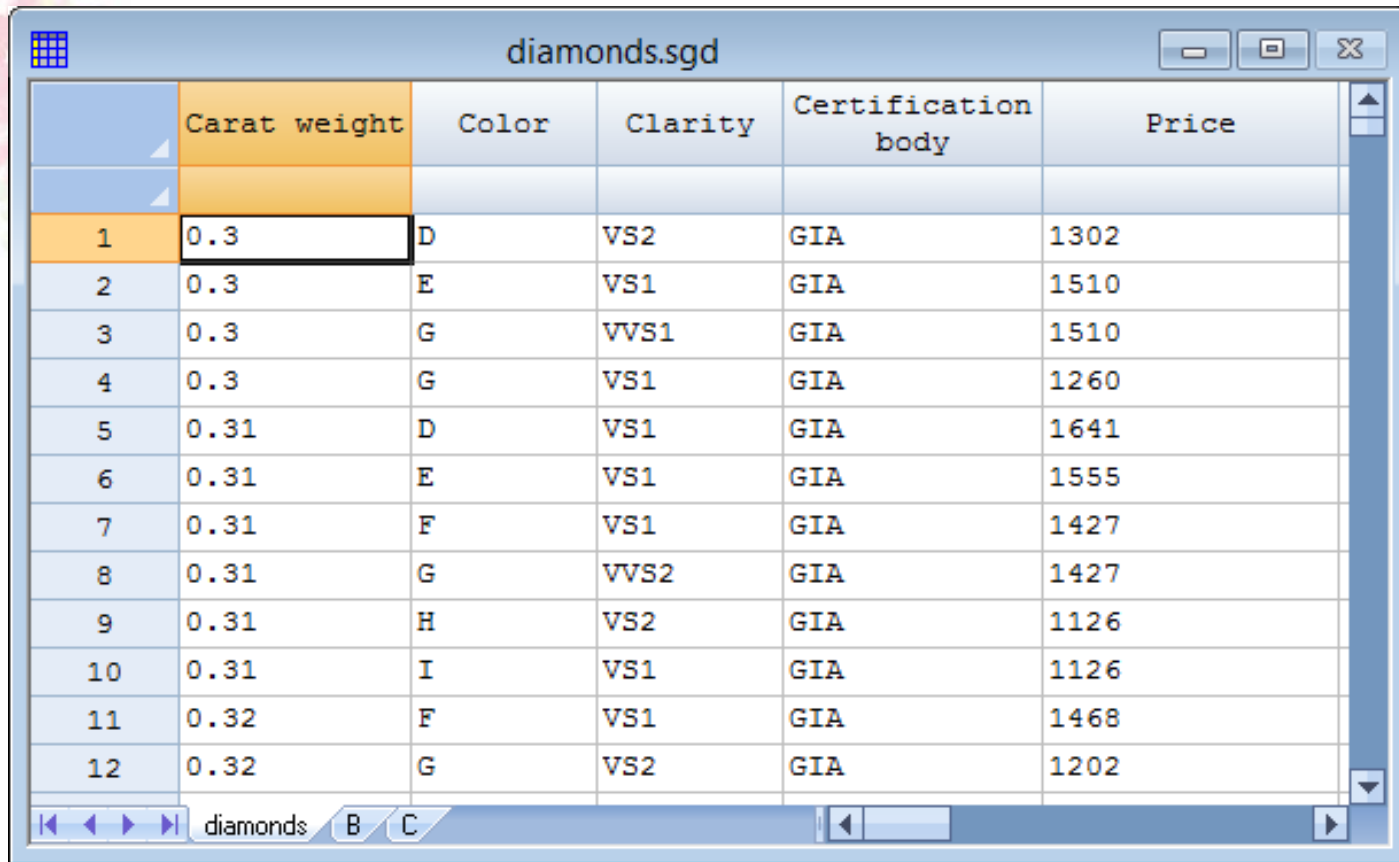


# Outline

1. Linear models with a single predictor
2. Linear models with multiple predictors
3. Logistic regression
4. Nonlinear regression models
5. Life data regression

# Example: Pricing of Diamonds

- The file *diamonds.sgd* contains information on 308 diamonds. (JSE Data Archive, Singfat Chu, National University of Singapore)

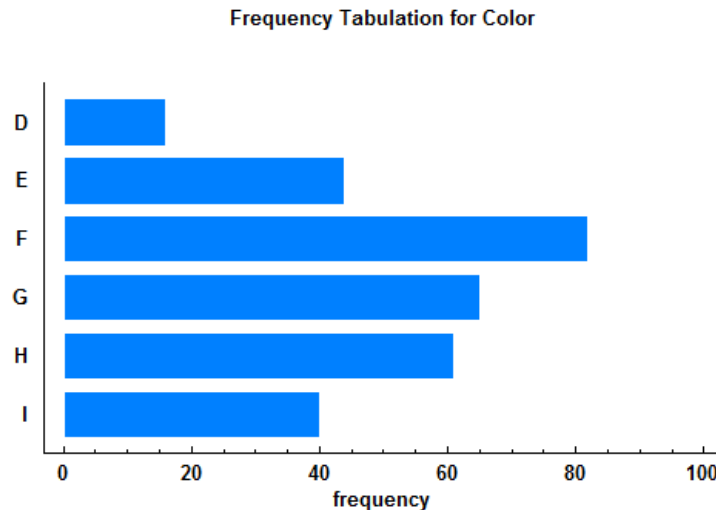


	Carat weight	Color	Clarity	Certification body	Price
1	0.3	D	VS2	GIA	1302
2	0.3	E	VS1	GIA	1510
3	0.3	G	VVS1	GIA	1510
4	0.3	G	VS1	GIA	1260
5	0.31	D	VS1	GIA	1641
6	0.31	E	VS1	GIA	1555
7	0.31	F	VS1	GIA	1427
8	0.31	G	VVS2	GIA	1427
9	0.31	H	VS2	GIA	1126
10	0.31	I	VS1	GIA	1126
11	0.32	F	VS1	GIA	1468
12	0.32	G	VS2	GIA	1202

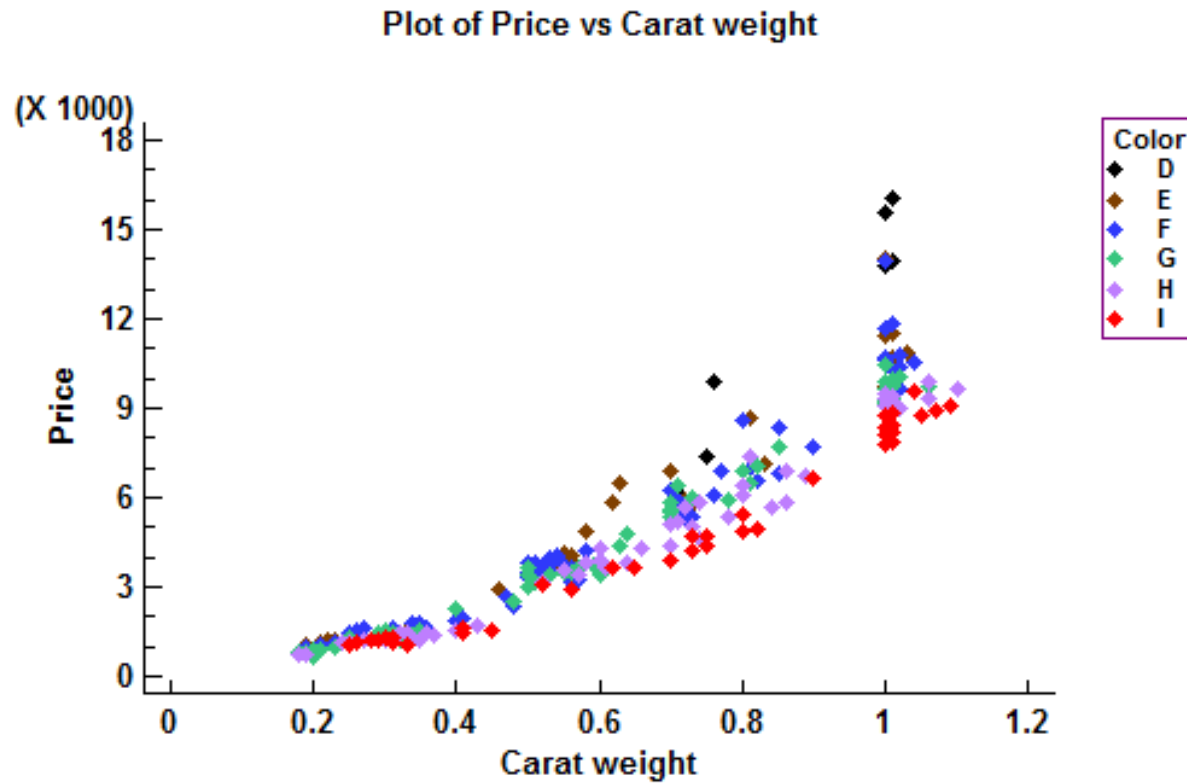


# Models with a single categorical predictor

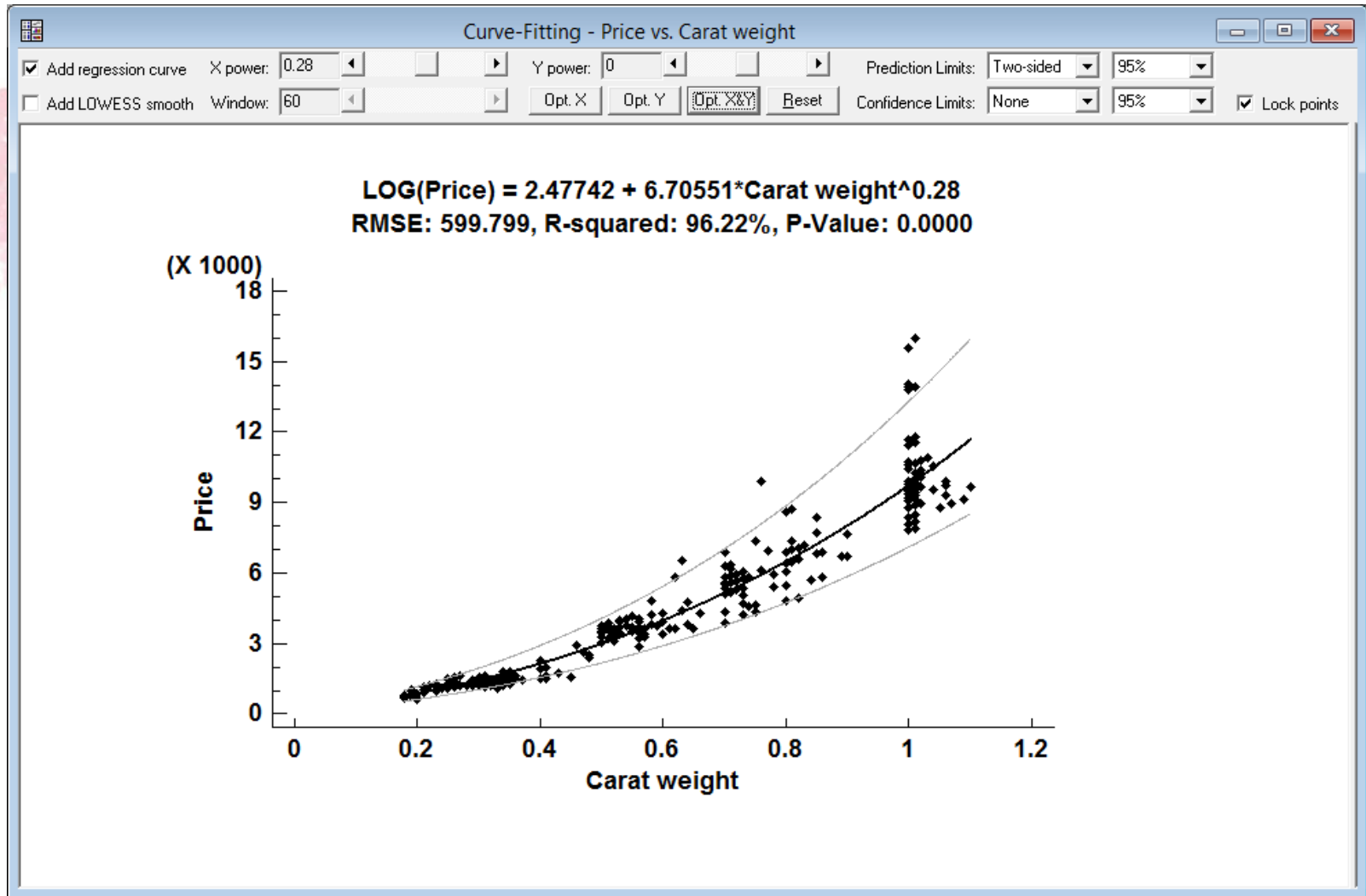
- Dependent variable:  $Y = \text{Price}$
- Independent variable:  $X = \text{Carat weight}$
- Categorical variable:  $C = \text{Color (6 levels)}$



# Coded Scatterplot

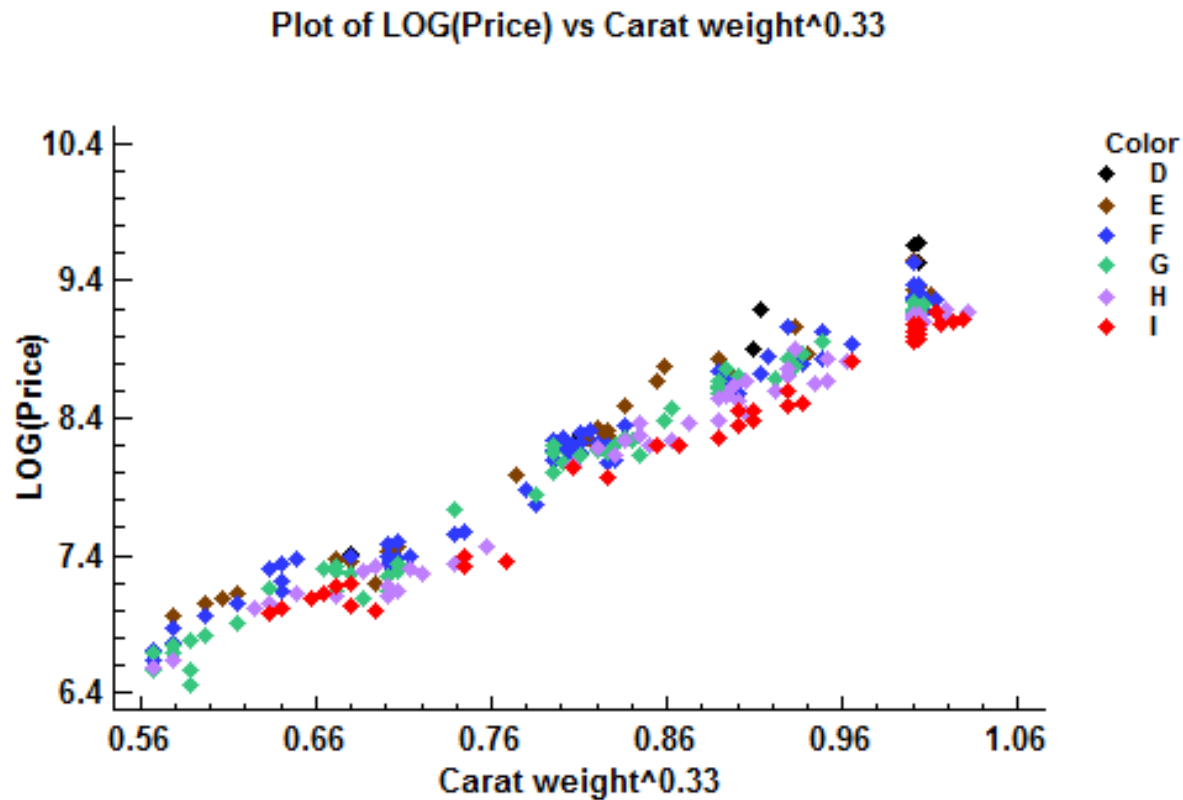


# Transformation





# Transformed Data



# Statistical Model

$$Y = \beta_0 + \beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3 + \beta_4 I_4 + \beta_5 I_5 \\ + \beta_6 X + \beta_7 I_1 X + \beta_8 I_2 X + \beta_9 I_3 X + \beta_{10} I_4 X + \beta_{11} I_5 X$$

where

$$Y = \text{LOG}(\text{Price})$$

$$X = \text{Carat weight}^{0.33}$$

Color	I1	I2	I3	I4	I5
D	0	0	0	0	0
E	1	0	0	0	0
F	0	1	0	0	0
G	0	0	1	0	0
H	0	0	0	1	0
I	0	0	0	0	1

# Model by Color

Color D:  $Y = \beta_0 + \beta_6 X$

Color E:  $Y = (\beta_0 + \beta_1) + (\beta_6 + \beta_7)X$

Color F:  $Y = (\beta_0 + \beta_2) + (\beta_6 + \beta_8)X$

Color G:  $Y = (\beta_0 + \beta_3) + (\beta_6 + \beta_9)X$

Color H:  $Y = (\beta_0 + \beta_4) + (\beta_6 + \beta_{10})X$

Color I:  $Y = (\beta_0 + \beta_5) + (\beta_6 + \beta_{11})X$

# Comparison of Regression Lines

Comparison of Regression Lines

Carat weight  
Color  
Clarity  
Certification body  
Price

Dependent Variable:  
LOG(Price)

Independent Variable:  
Carat weight<sup>0.33</sup>

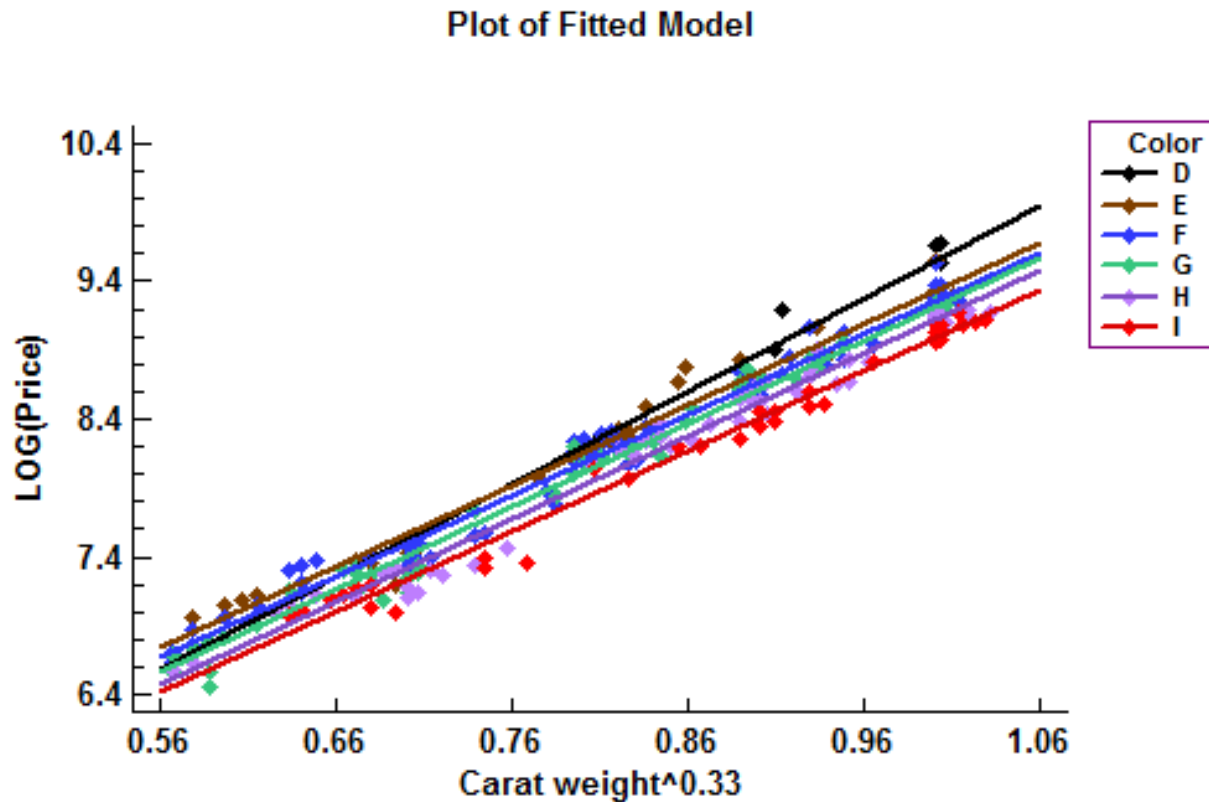
Level Codes:  
Color

(Select:)

☐ Sort column names

OK Cancel Delete Transform... Help

# Fitted Regression Model



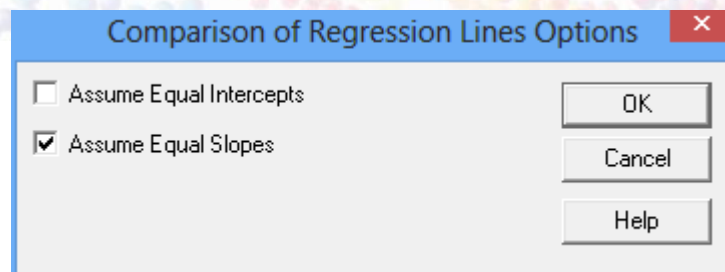
# Test for Differences in the Slopes and Intercepts

- Conditional sums of squares

**Further ANOVA for Variables in the Order Fitted**

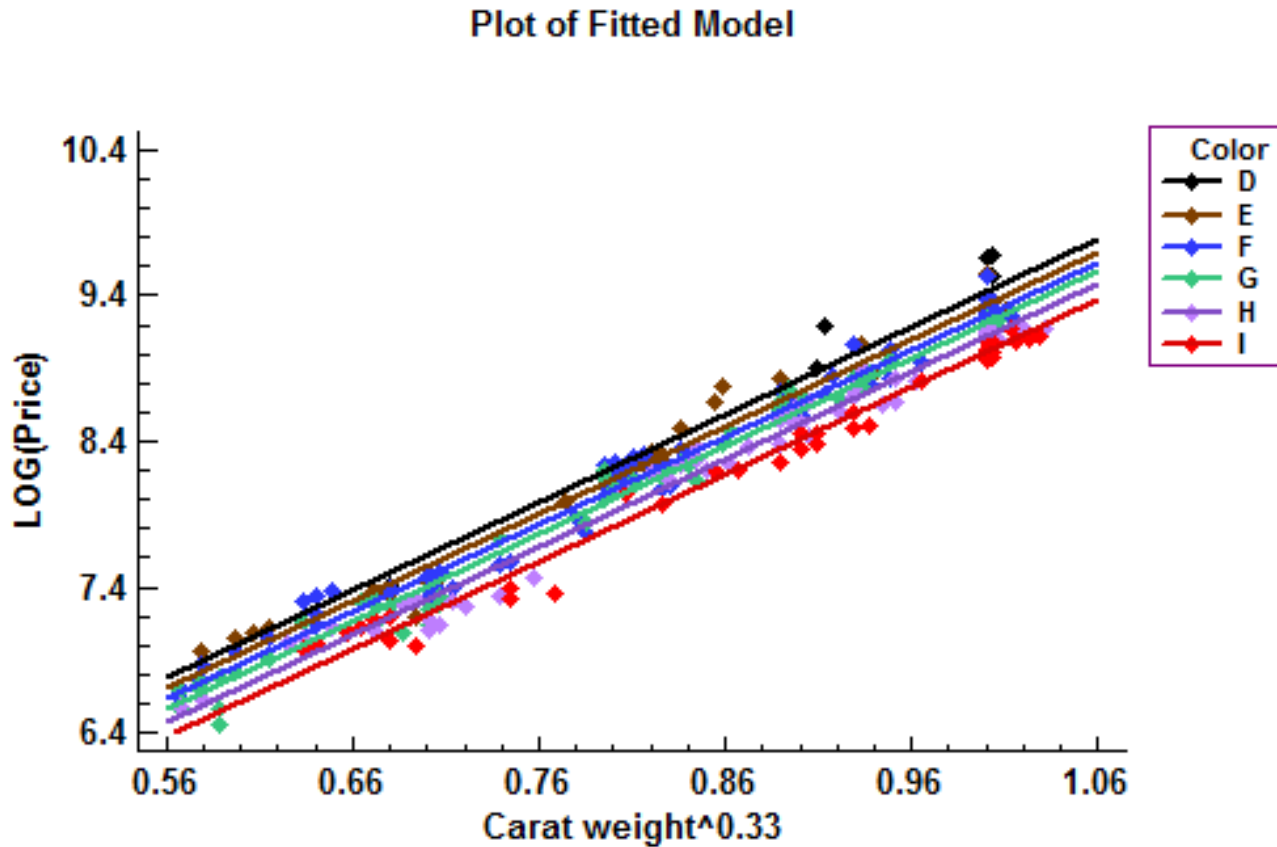
<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Carat weight^0.33	195.846	1	195.846	16521.44	0.0000
Intercepts	3.96947	5	0.793893	66.97	0.0000
Slopes	0.249657	5	0.0499314	4.21	0.0010
Model	200.065	11			

- Simplify the model using *Analysis Options*





# Parallel Regression Lines



# Fitted Model

## Comparison of Regression Lines - LOG(Price) versus Carat weight<sup>0.33</sup> by Color

Dependent variable: LOG(Price)

Independent variable: Carat weight<sup>0.33</sup>

Level codes: Color

Number of complete cases: 308

Number of regression lines: 6

### Multiple Regression Analysis

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	2.83694	0.147535	19.2289	0.0000
Carat weight <sup>0.33</sup>	6.7001	0.173817	38.5469	0.0000
Color=E	0.626838	0.18309	3.42366	0.0007
Color=F	0.536887	0.166205	3.23027	0.0014
Color=G	0.339188	0.167165	2.02906	0.0433
Color=H	0.277087	0.176293	1.57174	0.1171
Color=I	0.320909	0.187515	1.71137	0.0881
Carat weight <sup>0.33</sup> *Color=E	-0.844063	0.215945	-3.90869	0.0001
Carat weight <sup>0.33</sup> *Color=F	-0.821593	0.196678	-4.17735	0.0000
Carat weight <sup>0.33</sup> *Color=G	-0.661424	0.19826	-3.33614	0.0010
Carat weight <sup>0.33</sup> *Color=H	-0.700254	0.206253	-3.39512	0.0008
Carat weight <sup>0.33</sup> *Color=I	-0.871306	0.21686	-4.01783	0.0001

### Coefficients

Color	Intercept	Slope
D	2.83694	6.7001
E	3.46378	5.85604
F	3.37383	5.87851
G	3.17613	6.03868
H	3.11403	5.99985
I	3.15785	5.8288

### Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	200.065	11	18.1877	1534.31	0.0000
Residual	3.5088	296	0.011854		
Total (Corr.)	203.574	307			

R-Squared = 98.2764 percent

R-Squared (adjusted for d.f.) = 98.2123 percent

Standard Error of Est. = 0.108876

Mean absolute error = 0.0833828

Durbin-Watson statistic = 1.18873 (P=0.0000)

Lag 1 residual autocorrelation = 0.401379

# Multiple Predictors

- When dealing with multiple categorical and quantitative predictors, we can use either of 2 procedures:
  - **Multiple Regression** (have to type in expressions for each indicator variable)
  - **GLM: General Linear Model** (automatically generates the indicator variables)
- Be careful: the indicator variables are set up differently in GLM (as well as the DOE procedures.)

# Coding Comparison

- Comparison of Regression Lines

Color	I1	I2	I3	I4	I5
D	0	0	0	0	0
E	1	0	0	0	0
F	0	1	0	0	0
G	0	0	1	0	0
H	0	0	0	1	0
I	0	0	0	0	1

- GLM and DOE

Color	I1	I2	I3	I4	I5
D	1	0	0	0	0
E	0	1	0	0	0
F	0	0	1	0	0
G	0	0	0	1	0
H	0	0	0	0	1
I	-1	-1	-1	-1	-1

# Model Comparison

- Comparison of Regression Lines

Color	Intercept	Slope
D	$\beta_0$	$\beta_6$
E	$\beta_0 + \beta_1$	$\beta_6 + \beta_7$
F	$\beta_0 + \beta_2$	$\beta_6 + \beta_8$
G	$\beta_0 + \beta_3$	$\beta_6 + \beta_9$
H	$\beta_0 + \beta_4$	$\beta_6 + \beta_{10}$
I	$\beta_0 + \beta_5$	$\beta_6 + \beta_{11}$

- GLM and DOE

Color	Intercept	Slope
D	$\beta_0 + \beta_1$	$\beta_6 + \beta_7$
E	$\beta_0 + \beta_2$	$\beta_6 + \beta_8$
F	$\beta_0 + \beta_3$	$\beta_6 + \beta_9$
G	$\beta_0 + \beta_4$	$\beta_6 + \beta_{10}$
H	$\beta_0 + \beta_5$	$\beta_6 + \beta_{11}$
I	$\beta_0 - (\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5)$	$\beta_6 - (\beta_7 + \beta_8 + \beta_9 + \beta_{10} + \beta_{11})$

# Data Input Dialog Box

General Linear Models

Carat weight  
Color  
Clarity  
Certification body  
Price

Dependent Variables:  
LOG(Price)

Categorical Factors:  
Color  
Clarity  
Certification body

Quantitative Factors:  
 $\text{Carat weight}^{0.33}$

(Weights:)

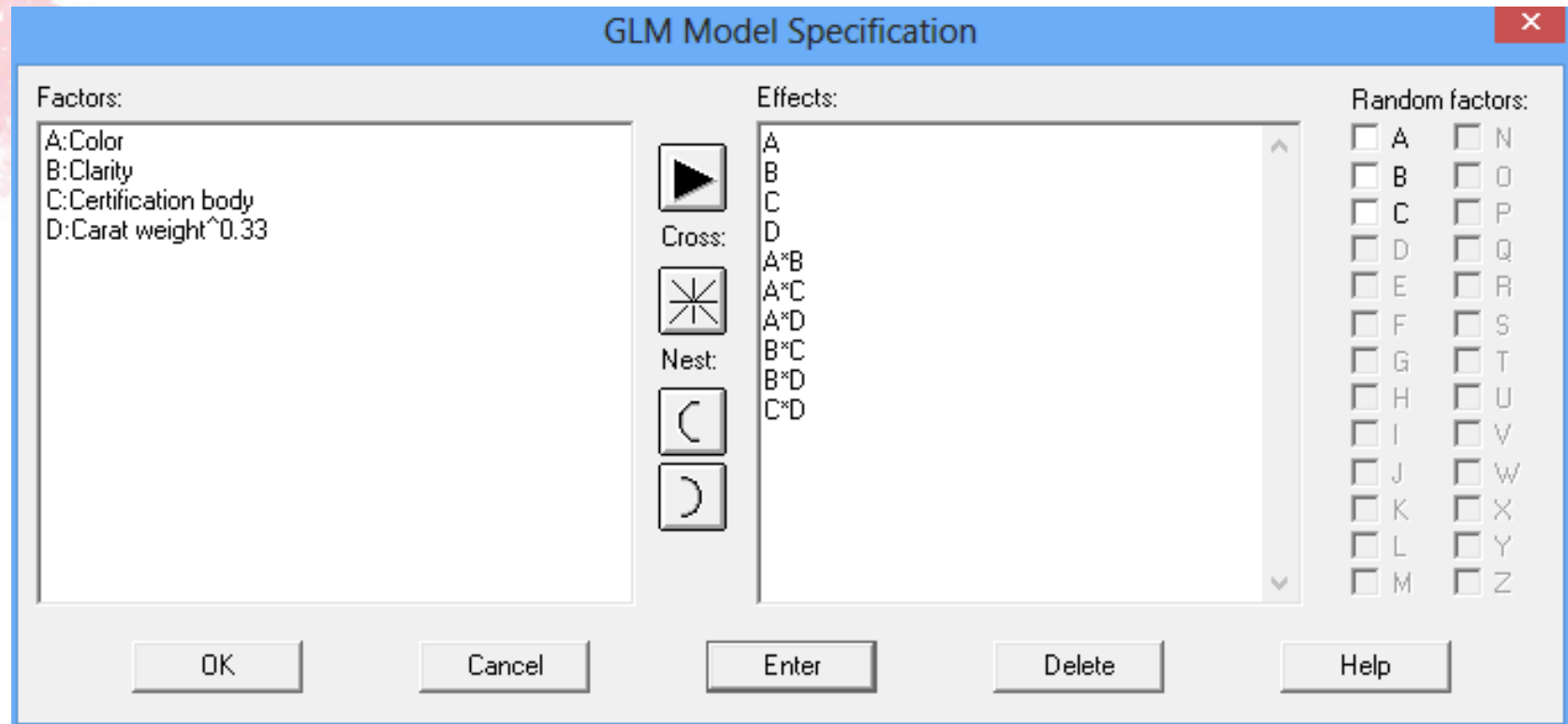
(Select:)

☐ Sort column names

OK Cancel Delete Transform... Help



# Model Specification



The image shows a 'GLM Model Specification' dialog box. It has three main sections: 'Factors:', 'Effects:', and 'Random factors:'. The 'Factors:' section contains a list of factors: A:Color, B:Clarity, C:Certification body, and D:Carat weight^0.33. The 'Effects:' section contains a list of effects: A, B, C, D, A\*B, A\*C, A\*D, B\*C, B\*D, and C\*D. The 'Random factors:' section contains a grid of checkboxes for factors A through Z. At the bottom, there are buttons for OK, Cancel, Enter, Delete, and Help. In the center, there are icons for 'Cross' (a triangle) and 'Nest' (a circle with a line), each with a corresponding label.

GLM Model Specification

Factors:

- A:Color
- B:Clarity
- C:Certification body
- D:Carat weight<sup>0.33</sup>

Effects:

- A
- B
- C
- D
- A\*B
- A\*C
- A\*D
- B\*C
- B\*D
- C\*D

Random factors:

<input type="checkbox"/> A	<input type="checkbox"/> N
<input type="checkbox"/> B	<input type="checkbox"/> O
<input type="checkbox"/> C	<input type="checkbox"/> P
<input type="checkbox"/> D	<input type="checkbox"/> Q
<input type="checkbox"/> E	<input type="checkbox"/> R
<input type="checkbox"/> F	<input type="checkbox"/> S
<input type="checkbox"/> G	<input type="checkbox"/> T
<input type="checkbox"/> H	<input type="checkbox"/> U
<input type="checkbox"/> I	<input type="checkbox"/> V
<input type="checkbox"/> J	<input type="checkbox"/> W
<input type="checkbox"/> K	<input type="checkbox"/> X
<input type="checkbox"/> L	<input type="checkbox"/> Y
<input type="checkbox"/> M	<input type="checkbox"/> Z

OK Cancel Enter Delete Help

# Results

## General Linear Models

Number of dependent variables: 1

Number of categorical factors: 3

A=Color

B=Clarity

C=Certification body

Number of quantitative factors: 1

D=Carat weight<sup>0.33</sup>

Number of observations: 308

### Analysis of Variance for LOG(Price)

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	202.896	61	3.32616	1206.95	0.0000
Residual	0.677934	246	0.00275583		
Total (Corr.)	203.574	307			

### Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Color	0.13531	5	0.027062	9.82	0.0000
Clarity	0.0865789	4	0.0216447	7.85	0.0000
Certification body	0.0585465	2	0.0292733	10.62	0.0000
Carat weight <sup>0.33</sup>	35.7245	1	35.7245	12963.25	0.0000
Color*Clarity	0.130609	20	0.00653043	2.37	0.0011
Color*Certification body	0.033009	10	0.0033009	1.20	0.2929
Color*Carat weight <sup>0.33</sup>	0.0286298	5	0.00572597	2.08	0.0688
Clarity*Certification body	0.0173235	8	0.00216543	0.79	0.6156
Clarity*Carat weight <sup>0.33</sup>	0.0320343	4	0.00800858	2.91	0.0224
Certification body*Carat weight <sup>0.33</sup>	0.0489169	2	0.0244584	8.88	0.0002
Residual	0.677934	246	0.00275583		
Total (corrected)	203.574	307			

# Simplified Model

Type III Sums of Squares

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Color	4.40806	5	0.881612	312.82	0.0000
Clarity	0.203955	4	0.0509888	18.09	0.0000
Certification body	0.0697858	2	0.0348929	12.38	0.0000
Carat weight <sup>0.33</sup>	68.6518	1	68.6518	24359.17	0.0000
Color*Clarity	0.141277	20	0.00706386	2.51	0.0005
Clarity*Carat weight <sup>0.33</sup>	0.0669007	4	0.0167252	5.93	0.0001
Certification body*Carat weight <sup>0.33</sup>	0.052897	2	0.0264485	9.38	0.0001
Residual	0.758126	269	0.00281831		
Total (corrected)	203.574	307			

# Predicting New Observations

C:\Data\webinar\diamonds.sgd

	Carat weight	Color	Clarity	Certification body	Price
302	1.01	H	VS1	HRD	9153
303	1.01	I	VVS1	HRD	8873
304	1.01	I	VS1	HRD	8175
305	1.02	F	VVS2	HRD	10796
306	1.06	H	VVS2	HRD	9890
307	1.02	H	VS2	HRD	8959
308	1.09	I	VVS2	HRD	9107
309	.75	E	VVS2	HRD	

**Model Results for LOG(price)**

	<i>Fitted</i>	<i>Std. Error</i>	<i>Lower 95.0% CL</i>	<i>Upper 95.0% CL</i>	<i>Lower 95.0% CL</i>	<i>Upper 95.0% CL</i>
<i>Row</i>	<i>Value</i>	<i>for Forecast</i>	<i>for Forecast</i>	<i>for Forecast</i>	<i>for Mean</i>	<i>for Mean</i>
309	8.85939	0.0564242	8.7483	8.97048	8.82176	8.89703

STATGRAPHICS Expression Evaluator

Expression:  
EXP(8.85939)

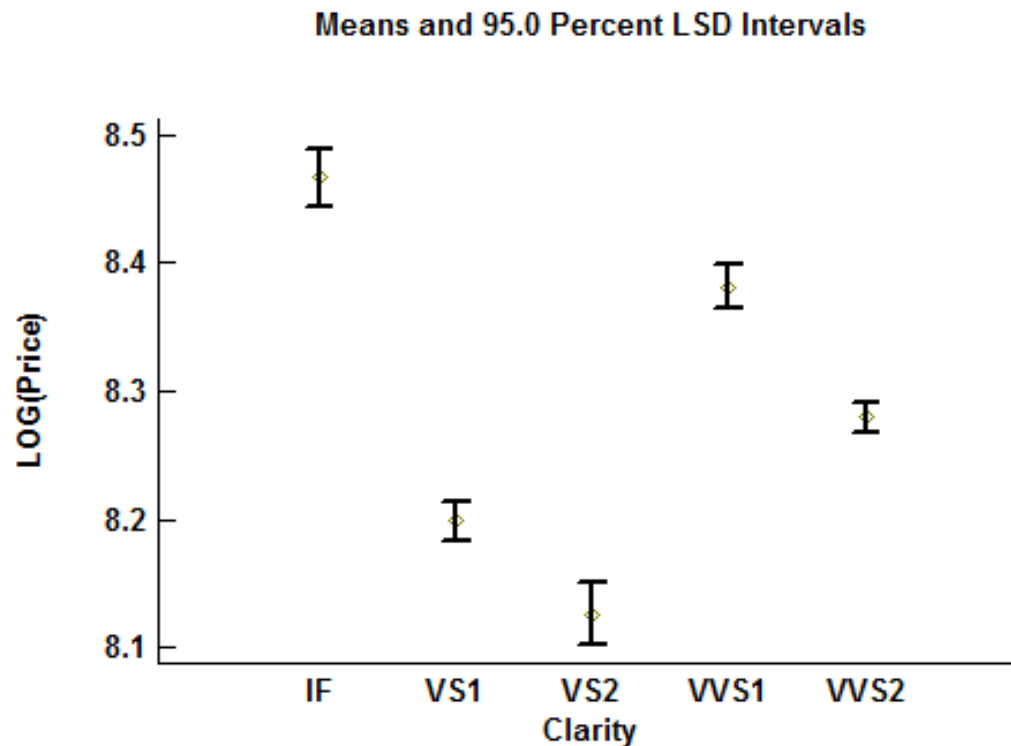
Variables:  
Carat weight  
Color  
Clarity  
Certification body  
Price

Operators:  
ABS(?)  
ACOS(?)  
ACOSG(?)  
ACOSH(?)  
ACOSR(?)  
ASIN(?)  
ASING(?)  
ASINH(?)  
ASINR(?)  
ATAN(?)

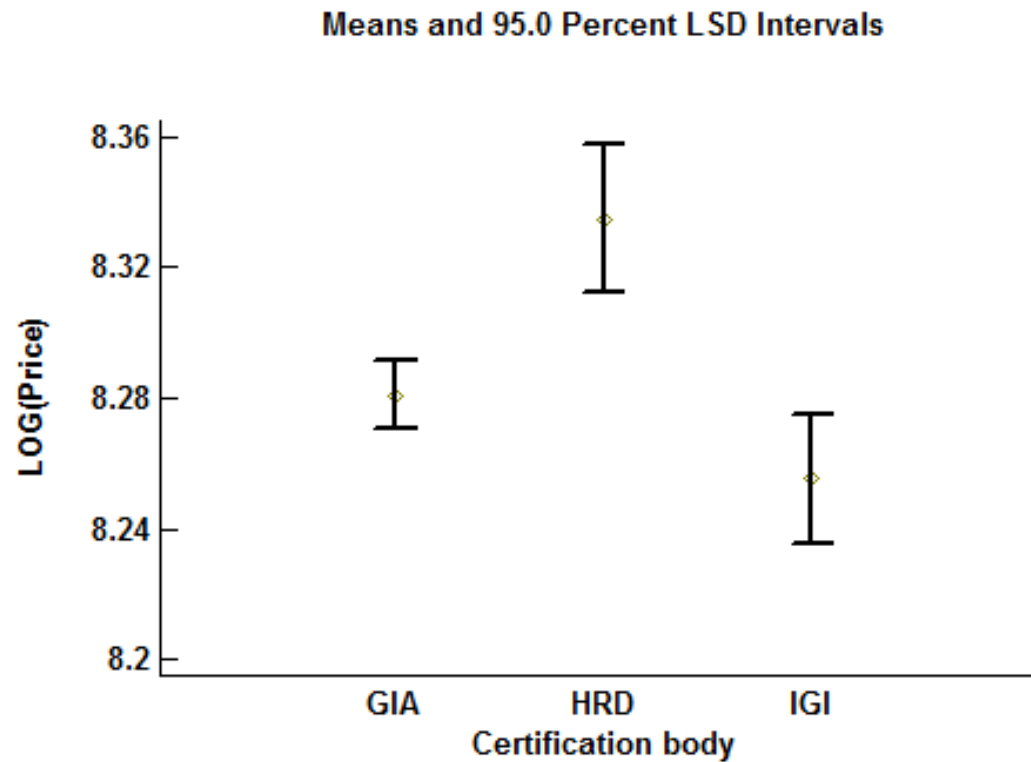
Result:  
7040.18692045

Clear Cancel Help

# Least Squares Means

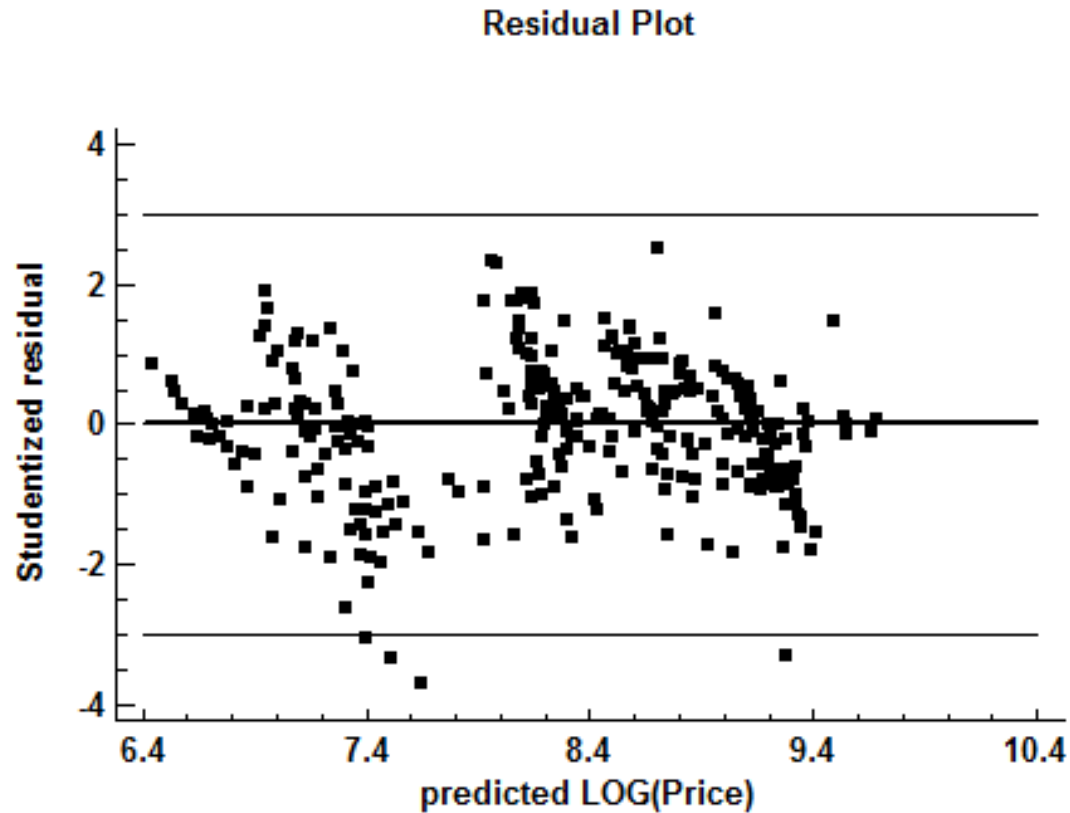


# Least Squares Means





# Residual Plot



# Logistic Regression



Can we predict how well Barry Bonds would do when he came to bat?

# Bonds Data from 2001

C:\Data\webinar\barrybonds.sgd															
	Game number	At-bat	Home game	On first	On second	On third	Outs	Inning	Runs in inning	Walk	Intentional walk	Result	ERA	Giants score	Opposing team score
		within game	1=TRUE						after first pitch to Bonds			>0 if got on base	opposing pitcher	before at-bat	before at-bat
1	1	1	1	1	0	0	1	1	0	0	0	0	4.20	0	0
2	1	2	1	0	1	0	1	3	0	1	1	5	4.20	0	0
3	1	3	1	0	0	0	1	5	1	0	0	4	4.20	1	0
4	1	4	1	1	0	1	2	6	0	1	0	5	5.49	3	1
5	1	5	1	0	0	0	2	8	0	0	0	0	4.00	3	2
6	2	1	1	0	0	0	2	1	0	0	0	0	5.75	0	0
7	2	2	1	1	1	0	2	3	0	0	0	0	5.75	0	3
8	2	3	1	0	1	0	2	5	0	0	0	0	5.75	1	4
9	2	4	1	1	0	0	1	7	3	0	0	2	4.00	4	5
10	2	5	1	1	0	0	2	8	0	1	0	5	5.21	7	6
11	3	1	1	0	0	0	2	1	0	0	0	0	4.30	0	1
12	3	2	1	1	0	0	2	3	3	0	0	1	4.30	1	1
13	3	3	1	1	0	0	2	5	0	0	0	0	4.30	4	2

n = 648 at-bats

Source: JSE Data Archive, Jerome P. Reiter, Duke University

# Model

- $Y = 1$  if Bonds reached base and 0 otherwise
- Predictors:
  - ERA of opposing pitcher
  - Runs already scored that inning
  - Opposing team's score
  - Inning
  - # of outs when he came to bat
  - Whether a runner was on first base
  - Whether a runner was on second base
  - Whether a runner was on third base
  - Whether it was a home game

# Data Input Dialog Box

The screenshot shows the 'Logistic Regression' dialog box in Statgraphics. On the left is a list of available variables: Game number, At-bat, Home game, On first, On second, On third, Outs, Inning, Runs in inning, Walk, Intentional walk, Result, ERA, Giants score, and Opposing team score. The 'Dependent Variable' is set to 'Result>0'. The 'Sample Sizes' field is empty. Under 'Quantitative Factors', 'ERA', 'Giants score', 'Opposing team score', and 'Inning' are listed. Under 'Categorical Factors', 'Outs', 'On first', 'On second', 'On third', and 'Home game' are listed. The '(Select:)' field is empty. At the bottom left, there is a checkbox for 'Sort column names' which is unchecked. At the bottom are buttons for 'OK', 'Cancel', 'Delete', 'Transform...', and 'Help'.

Logistic Regression

Game number  
At-bat  
Home game  
On first  
On second  
On third  
Outs  
Inning  
Runs in inning  
Walk  
Intentional walk  
Result  
ERA  
Giants score  
Opposing team score

Dependent Variable:  
▶ Result>0

(Sample Sizes:)  
▶

Quantitative Factors:  
▶ ERA  
Giants score  
Opposing team score  
Inning

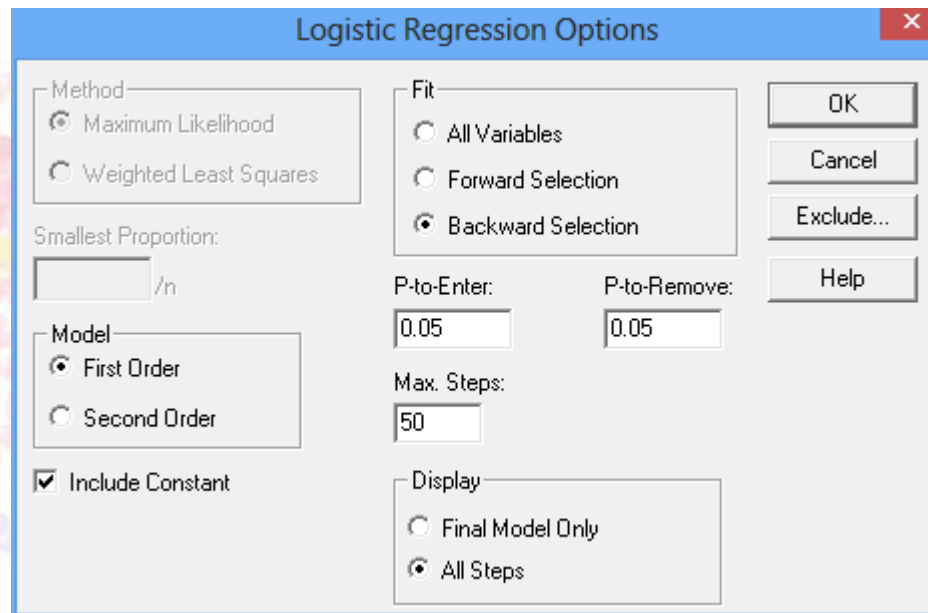
Categorical Factors:  
▶ Outs  
On first  
On second  
On third  
Home game

(Select:)  
▶

☐ Sort column names

OK Cancel Delete Transform... Help

# Analysis Options



The image shows a screenshot of the 'Logistic Regression Options' dialog box. The dialog has a blue title bar with a close button. It contains several sections: 'Method' with radio buttons for 'Maximum Likelihood' (selected) and 'Weighted Least Squares'; 'Smallest Proportion' with a text box and '/n'; 'Model' with radio buttons for 'First Order' (selected) and 'Second Order', and a checked 'Include Constant' checkbox; 'Fit' with radio buttons for 'All Variables', 'Forward Selection', and 'Backward Selection' (selected); 'P-to-Enter' and 'P-to-Remove' text boxes both set to '0.05'; 'Max. Steps' text box set to '50'; and 'Display' with radio buttons for 'Final Model Only' and 'All Steps' (selected). On the right side, there are four buttons: 'OK', 'Cancel', 'Exclude...', and 'Help'.

Logistic Regression Options

Method

- ☒ Maximum Likelihood
- ☐ Weighted Least Squares

Smallest Proportion:

/n

Model

- ☒ First Order
- ☐ Second Order

☒ Include Constant

Fit

- ☐ All Variables
- ☐ Forward Selection
- ☒ Backward Selection

P-to-Enter:

P-to-Remove:

Max. Steps:

Display

- ☐ Final Model Only
- ☒ All Steps

OK

Cancel

Exclude...

Help



# Results

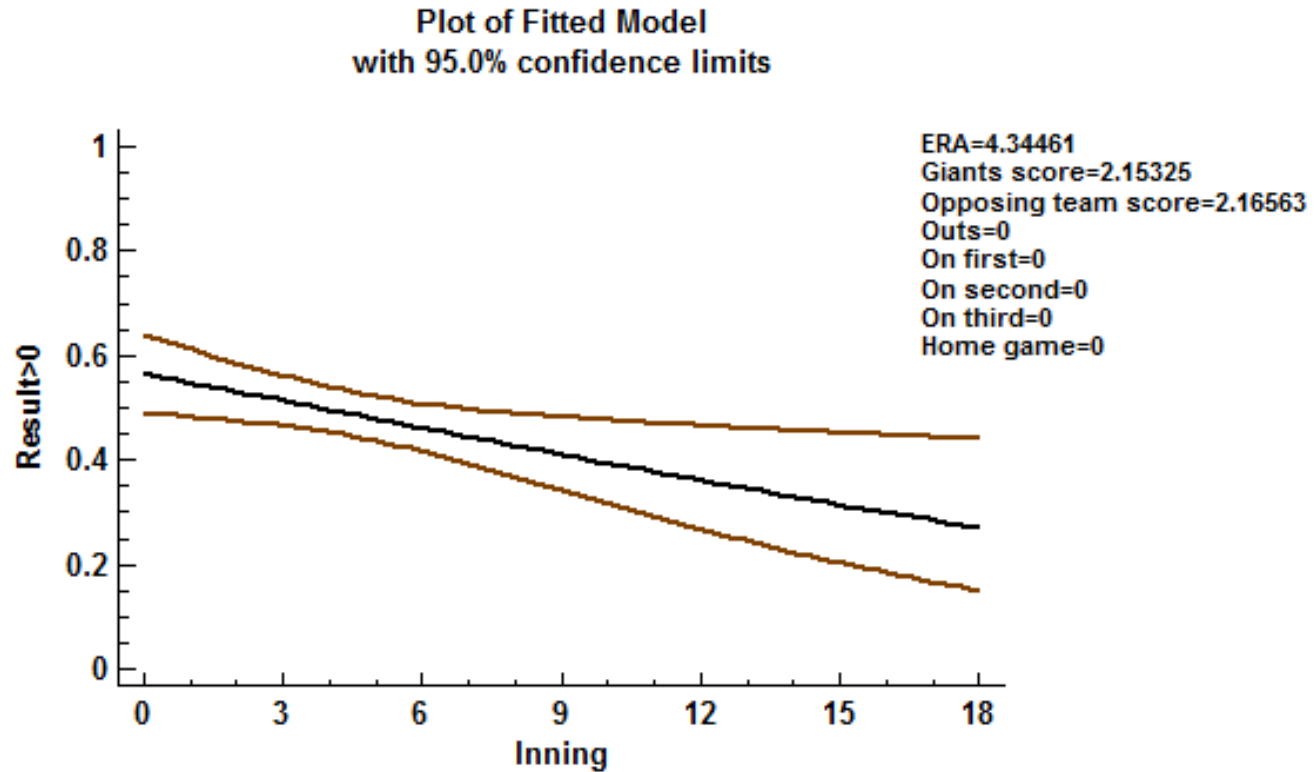
## Likelihood Ratio Tests

<i>Factor</i>	<i>Chi-Square</i>	<i>Df</i>	<i>P-Value</i>
Inning	6.34437	1	0.0118
On second	12.3916	1	0.0004

## Estimated Regression Model (Maximum Likelihood)

		<i>Standard</i>	<i>Estimated</i>
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Odds Ratio</i>
CONSTANT	1.02656	0.250474	
Inning	-0.0695187	0.0277701	0.932843
On second=0	-0.762319	0.221567	0.466583

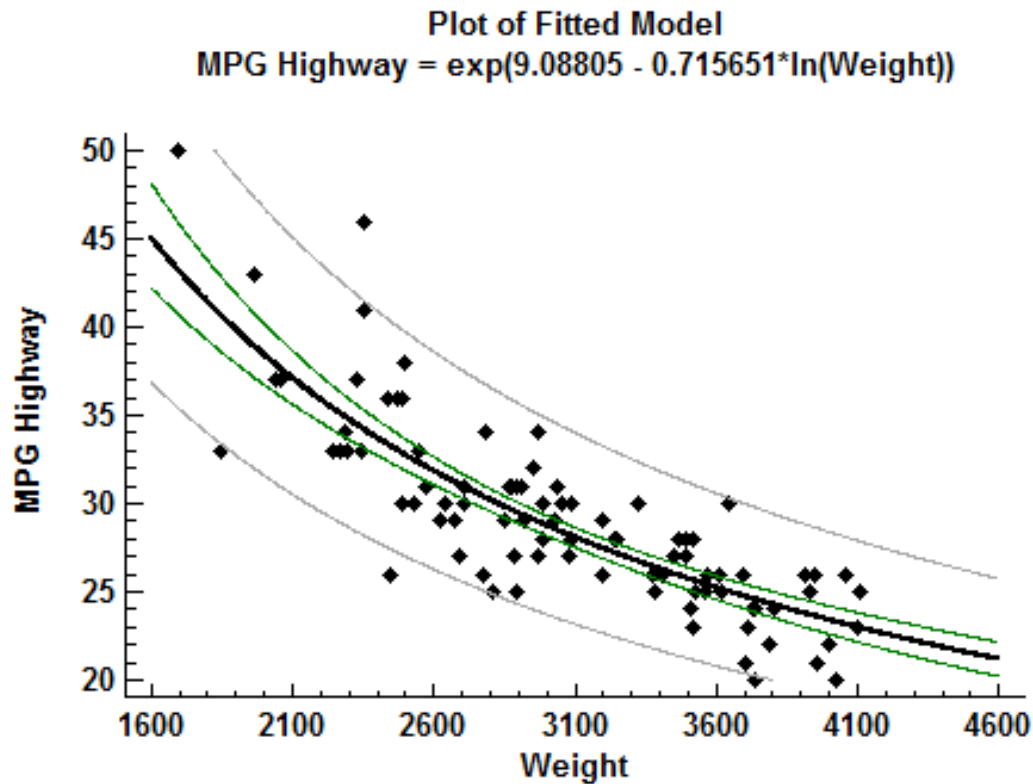
# Plot of Probability of Reaching Base



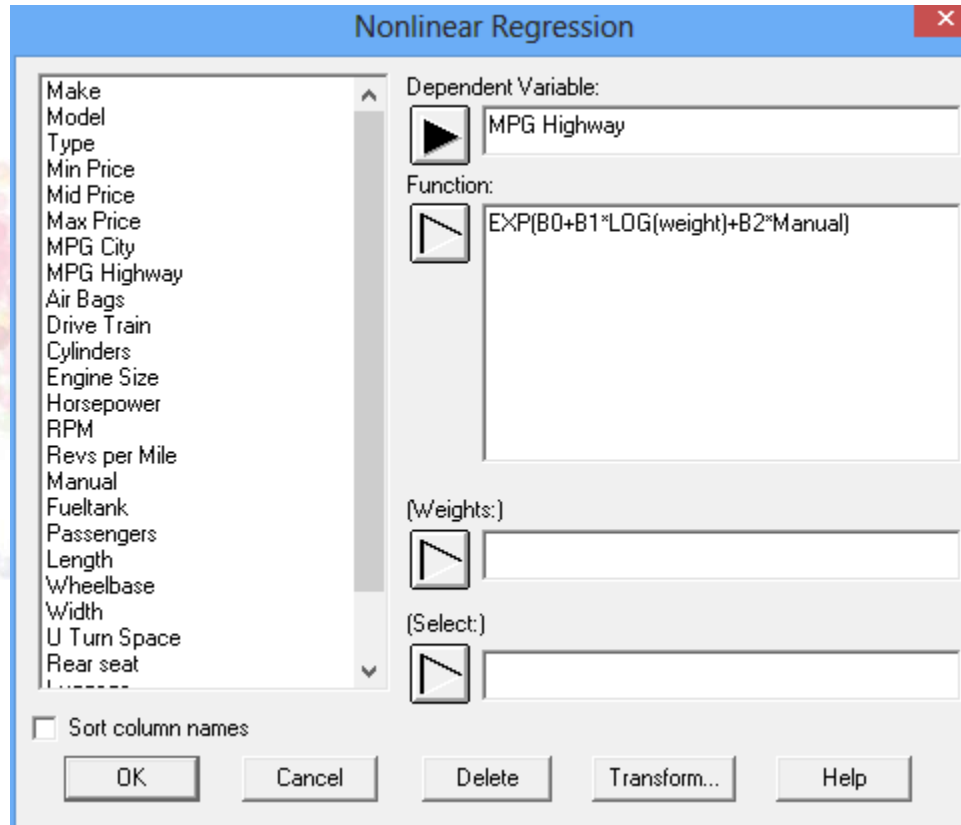
# Nonlinear Models

- Data file: 93cars.sgd
  - Y: MPG Highway
  - X: weight
  - C: manual
- The relationship between Y and X is nonlinear.

# Multiplicative Model



# Nonlinear Regression



The image shows a 'Nonlinear Regression' dialog box from the Statgraphics software. On the left is a list of variables: Make, Model, Type, Min Price, Mid Price, Max Price, MPG City, MPG Highway, Air Bags, Drive Train, Cylinders, Engine Size, Horsepower, RPM, Revs per Mile, Manual, Fuel tank, Passengers, Length, Wheelbase, Width, U Turn Space, and Rear seat. The 'Dependent Variable' is set to 'MPG Highway'. The 'Function' is set to  $\text{EXP}(B0+B1*\text{LOG}(\text{weight})+B2*\text{Manual})$ . There are empty fields for '(Weights:)' and '(Select:)', each with a selection icon. At the bottom, there is a checkbox for 'Sort column names' and buttons for 'OK', 'Cancel', 'Delete', 'Transform...', and 'Help'.

Nonlinear Regression

Dependent Variable: MPG Highway

Function:  $\text{EXP}(B0+B1*\text{LOG}(\text{weight})+B2*\text{Manual})$

(Weights:)

(Select:)

☐ Sort column names

OK Cancel Delete Transform... Help

# Results

## Nonlinear Regression - MPG Highway

Dependent variable: MPG Highway (miles per gallon in highway driving)

Independent variables:

weight (pounds)

manual (0=no, 1=yes)

Function to be estimated:  $\text{EXP}(B0+B1*\text{LOG}(\text{weight})+B2*\text{manual})$

Initial parameter estimates:

B0 = 0.1

B1 = 0.1

B2 = 0.1

Number of observations: 93

Estimation method: Marquardt

Estimation stopped due to convergence of parameter estimates.

Number of iterations: 10

Number of function calls: 45

$$\begin{aligned}\text{MPG Highway} = & \text{EXP}(9.66687 \\ & -0.783745 * \text{LOG}(\text{weight}) \\ & -0.0440352 * \text{manual})\end{aligned}$$

### Estimation Results

			Asymptotic	95.0%
		Asymptotic	Confidence	Interval
Parameter	Estimate	Standard Error	Lower	Upper
B0	9.66687	0.494256	8.68495	10.6488
B1	-0.783745	0.060318	-0.903577	-0.663912
B2	-0.0440352	0.0292223	-0.10209	0.0140201

### Analysis of Variance

Source	Sum of Squares	Df	Mean Square
Model	80506.6	3	26835.5
Residual	786.421	90	8.73801
Total	81293.0	93	
Total (Corr.)	2615.31	92	

R-Squared = 69.9301 percent

R-Squared (adjusted for d.f.) = 69.2619 percent

Standard Error of Est. = 2.95601

Mean absolute error = 2.19093

Durbin-Watson statistic = 1.65999

Lag 1 residual autocorrelation = 0.168988

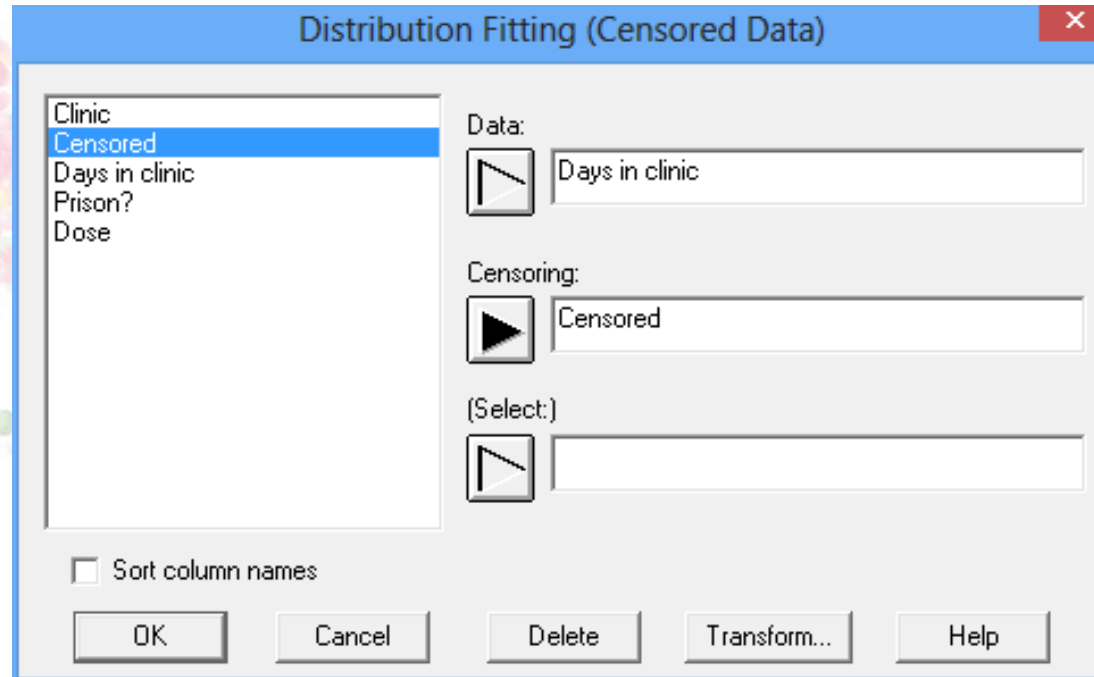
# Life Data Regression

- Data file: methadone.sgd

	Clinic	Censored	Days in clinic	Prison?	Dose
	1 or 2	1=still in clinic		1=spent time in prison	
1	1	0	428	0	50
2	1	0	275	1	55
3	1	0	262	0	55
4	1	0	183	0	30
5	1	0	259	1	65
6	1	0	714	0	55
7	1	0	438	1	65
8	1	1	796	1	60
9	1	0	892	0	50
10	1	0	393	1	65
11	1	1	161	1	80
12	1	0	836	1	60
13	1	0	523	0	55
14	1	0	612	0	70

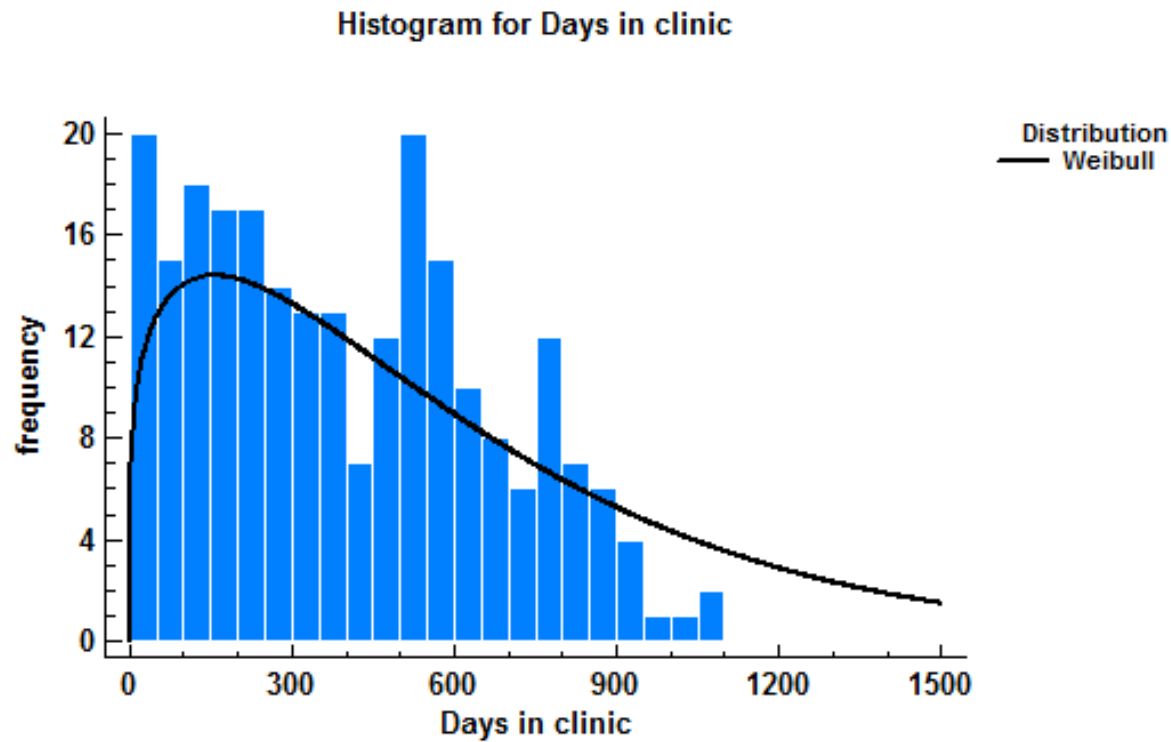
Caplehorn, J. (1991). Methadone dosage and retention of patients in maintenance treatment. *Medical Journal of Australia*.

# Distribution Fitting





# Distribution Fitting



# Data Input Dialog Box

Life Data Regression

Clinic  
Censored  
Days in clinic  
Prison?  
Dose

Dependent Variable:  
Days in clinic

(Censored:)  
Censored

Quantitative Factors:  
Dose

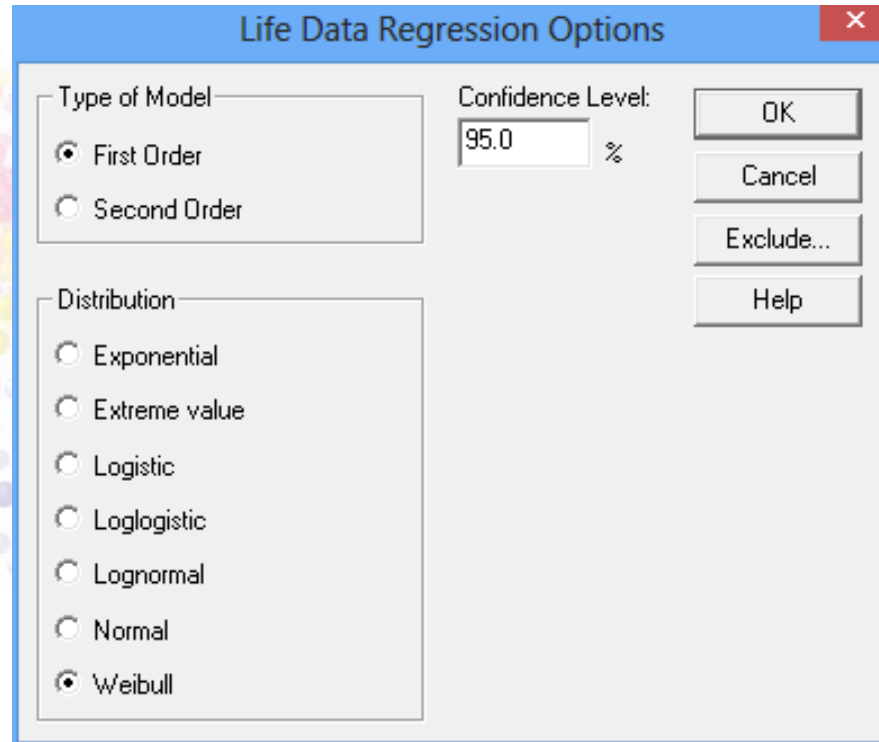
Categorical Factors:  
Clinic  
Prison?

(Select:)

☐ Sort column names

OK Cancel Delete Transform... Help

# Analysis Options



The image shows a software dialog box titled "Life Data Regression Options". It features a blue title bar with a close button (X) in the top right corner. The dialog is divided into several sections. On the left, there is a "Type of Model" section with two radio buttons: "First Order" (which is selected) and "Second Order". Below this is a "Distribution" section with seven radio buttons: "Exponential", "Extreme value", "Logistic", "Loglogistic", "Lognormal", "Normal", and "Weibull" (which is selected). To the right of these sections is a "Confidence Level:" label followed by a text input field containing "95.0" and a percentage symbol (%). On the far right, there are four buttons stacked vertically: "OK", "Cancel", "Exclude...", and "Help". The background of the slide features a decorative pattern of colorful dots (pink, yellow, green, blue) arranged in a curved, trail-like shape.

Life Data Regression Options

Type of Model

- ☒ First Order
- ☐ Second Order

Confidence Level: 95.0 %

Distribution

- ☐ Exponential
- ☐ Extreme value
- ☐ Logistic
- ☐ Loglogistic
- ☐ Lognormal
- ☐ Normal
- ☒ Weibull

OK

Cancel

Exclude...

Help

# Results

## Life Data Regression - Days in clinic

Dependent variable: Days in clinic

Censoring: Censored (1=still in clinic)

Factors:

Dose

Clinic (1 or 2)

Prison? (1=spent time in prison)

Number of uncensored values: 150

Number of right-censored values: 88

### Estimated Regression Model - Weibull

		<i>Standard</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Conf. Limit</i>	<i>Conf. Limit</i>
CONSTANT	5.29378	0.309429	4.68731	5.90025
Dose	0.0244232	0.00459046	0.015426	0.0334203
Clinic=1	-0.709329	0.157267	-1.01757	-0.401091
Prison?=0	0.229509	0.120806	-0.00726626	0.466285
SIGMA	0.729943	0.0493229	0.6394	0.833308

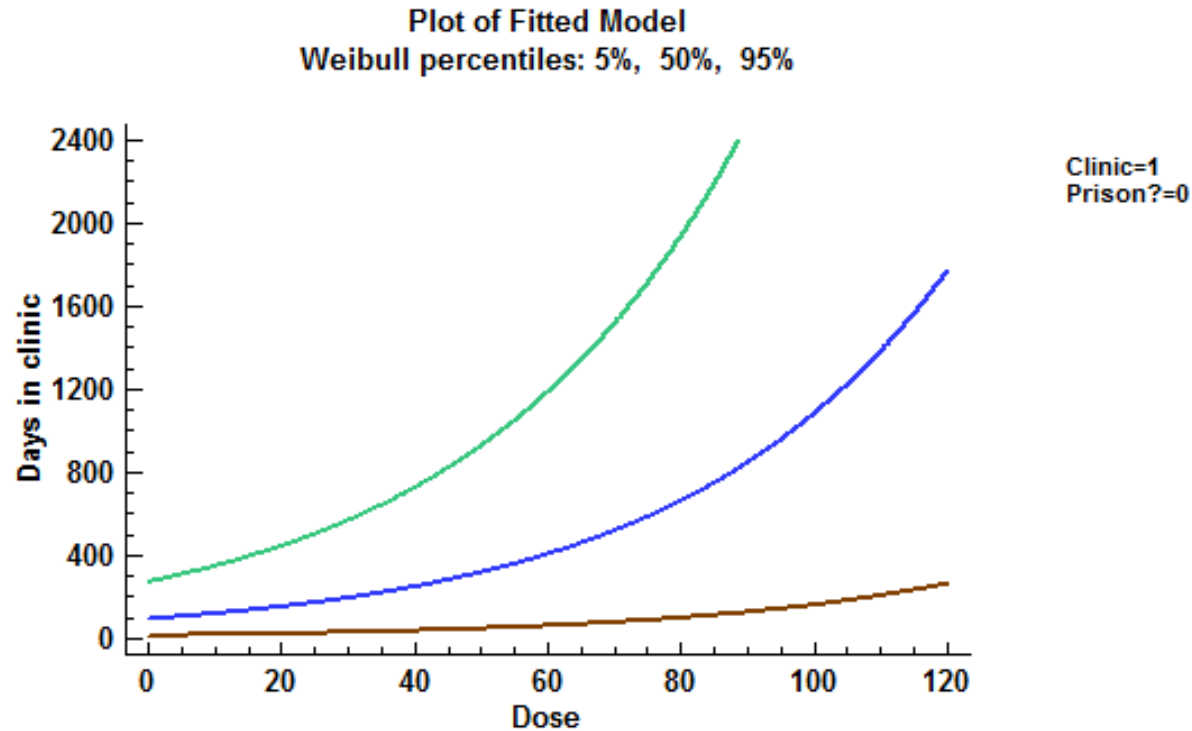
Log likelihood = -1084.48

### Likelihood Ratio Tests

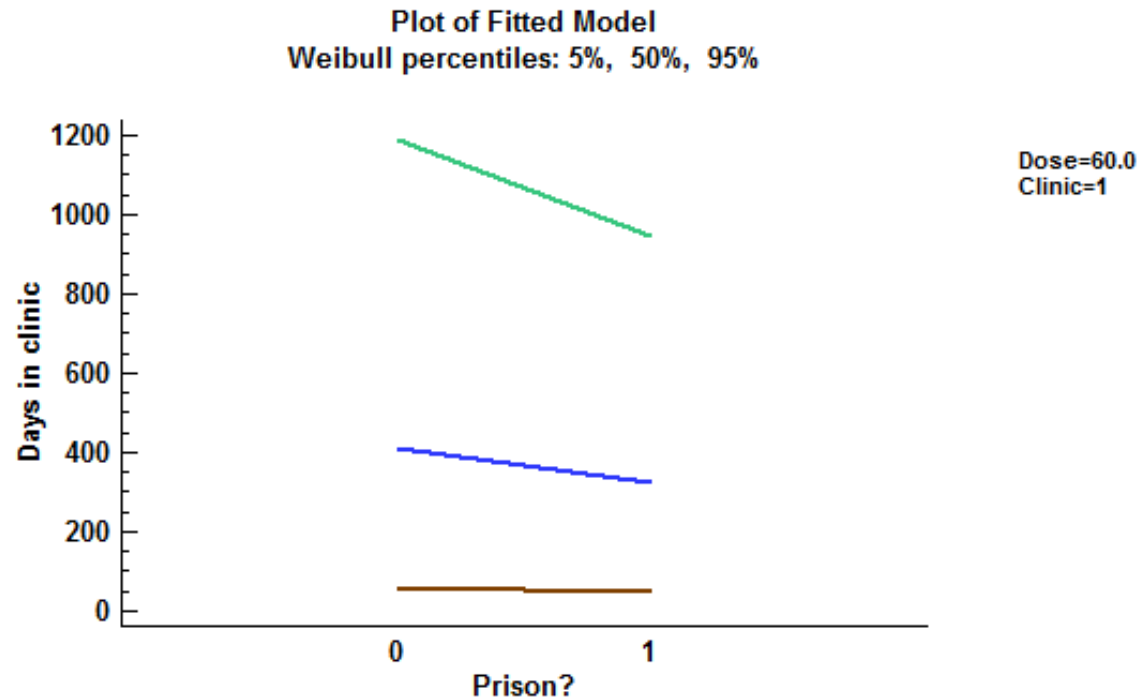
<i>Factor</i>	<i>Chi-Square</i>	<i>Df</i>	<i>P-Value</i>
Dose	28.5474	1	0.0000
Clinic	24.9152	1	0.0000
Prison?	3.54961	1	0.0596

Days in clinic =  $\exp(5.29378 + 0.0244232 * \text{Dose} - 0.709329 * \text{Clinic}=1 + 0.229509 * \text{Prison}?=0)$

# Percentile Plot



# Percentile Plot



# More Information

- Video, slides and sample data may be found at [www.statgraphics.com/webinars](http://www.statgraphics.com/webinars).