

Analyzing the Number-line Task

A Tutorial

John E. Opfer
Carnegie Mellon University

Introduction

The Number-line Task

The number-line task is a robust tool for characterizing representations of numerical value. Without demanding specific knowledge of measurement units (such as inches or centimeters), it taps into participants' mapping of spatial and numerical quantities across a wide range of values. The task has proven useful for characterizing subjects' representations across a wide range of ages (Siegler & Opfer, 2003). The purpose of this tutorial is to document procedures for analyzing performance on the number-line task.

The number-line task has two variants: the Number-to-Position (NP) Task and the Position-to-Number (PN) task. On the NP task (Fig. 1A), participants are shown a number and asked to estimate its position on the number line. On the PN task (Fig. 1B), participants are shown a position on a number line and asked to estimate the number that corresponds to it.

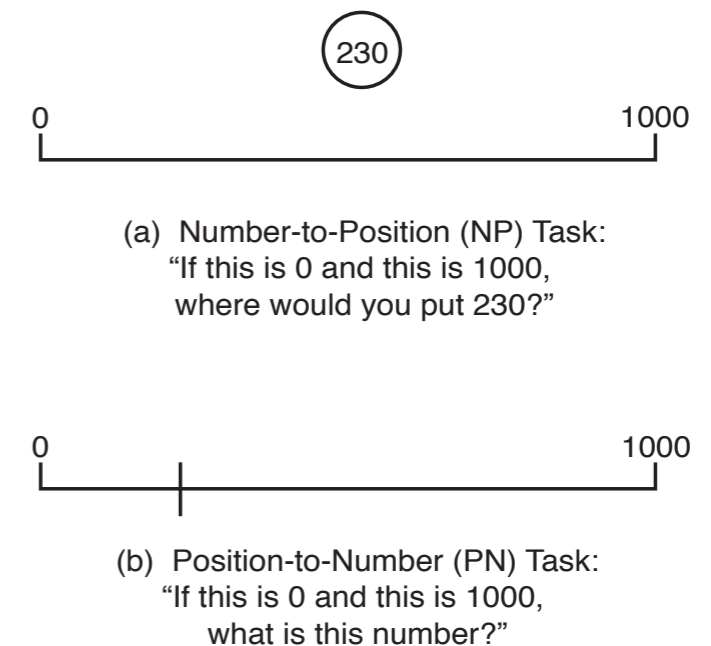
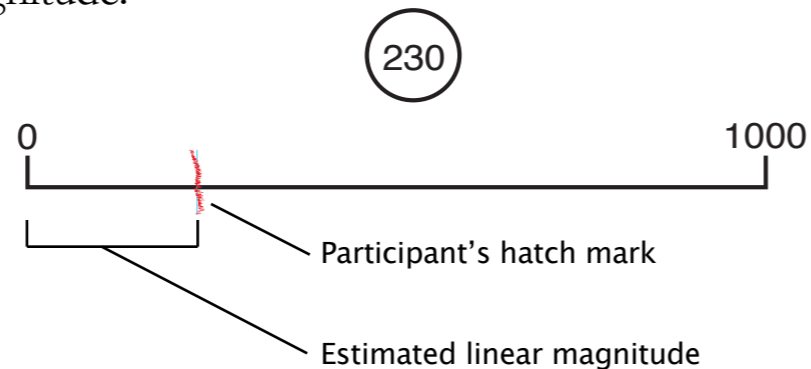


Figure 1. Two variants of the number-line task.

Recording the Data

Participants' performance on the number-line task is recorded in two ways, depending on the variant of the task.

On the NP task, participants provide a hatch mark on the number line to provide an estimate of linear magnitude.



To convert estimates of linear magnitude into a real number

- 1 Measure the distance from the left end point to the hatch mark (in linear units),
- 2 Divide that distance by the total length of the line, and
- 3 Multiply that value by the number given on the other endpoint.



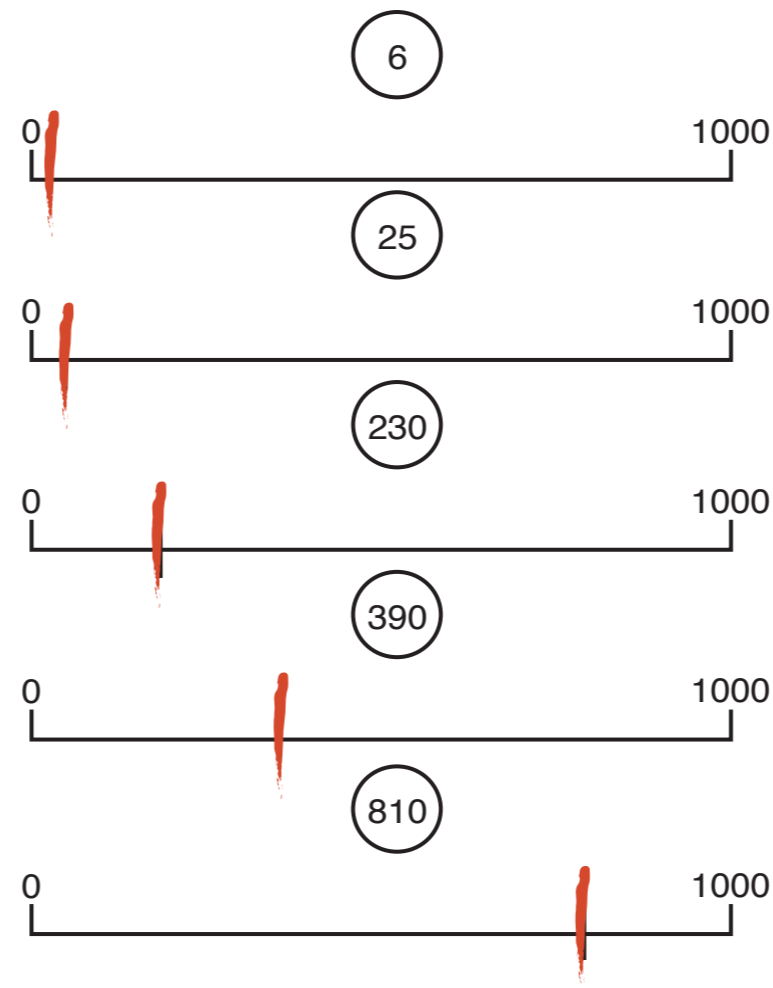
On the PN task, participants write out an integer; no conversion is needed. However, to generate stimuli on the PN task, one must convert numbers into linear magnitudes using the above procedure.

Obtaining Sufficient Data

By obtaining only a single estimate, it is not possible to discriminate among competing models of numeric representation.

Ideally, participants should be given a large range of quantities to estimate, with duplicate quantities. For example, in Siegler & Booth (2003), participants were given 24 different numerals (or linear magnitudes).

Thus, for each participant, there was a series of estimates that corresponded to a series of actual values.



Data Analysis

The Basics

The primary analysis of number-line performance involves the use of different regression models to characterize the relation between actual quantity (the numeral presented on the NP task or the linear magnitude presented on the PN task) and estimated values (participants' answers to the NP and PN problems).

This tutorial assumes basic understanding of regression and descriptive statistics and moderate facility in use of Microsoft Excel and SPSS. The user manuals of each program provides more detailed mathematical descriptions for advanced users.

Basic data analysis comprises three steps:

- 1 Graphing estimates against actual quantity,
- 2 Fitting estimates to actual quantities using different regression models, and
- 3 Comparing different regression models to identify the best fitting model.



Graphing Estimates

To graph estimates against actual quantity, produce a table in Excel or SPSS that includes a column (or variable) for actual quantities and a column(s) for estimates.

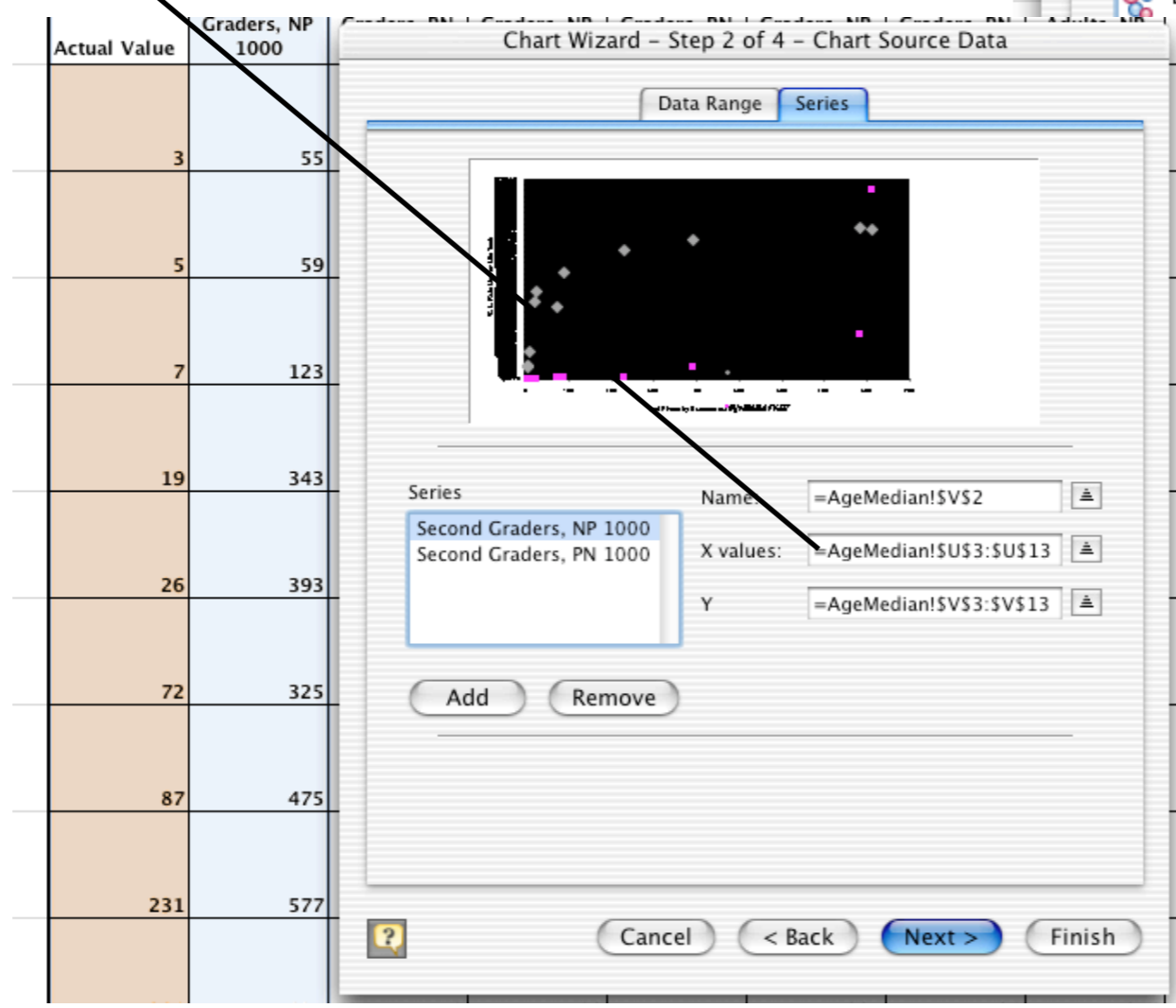
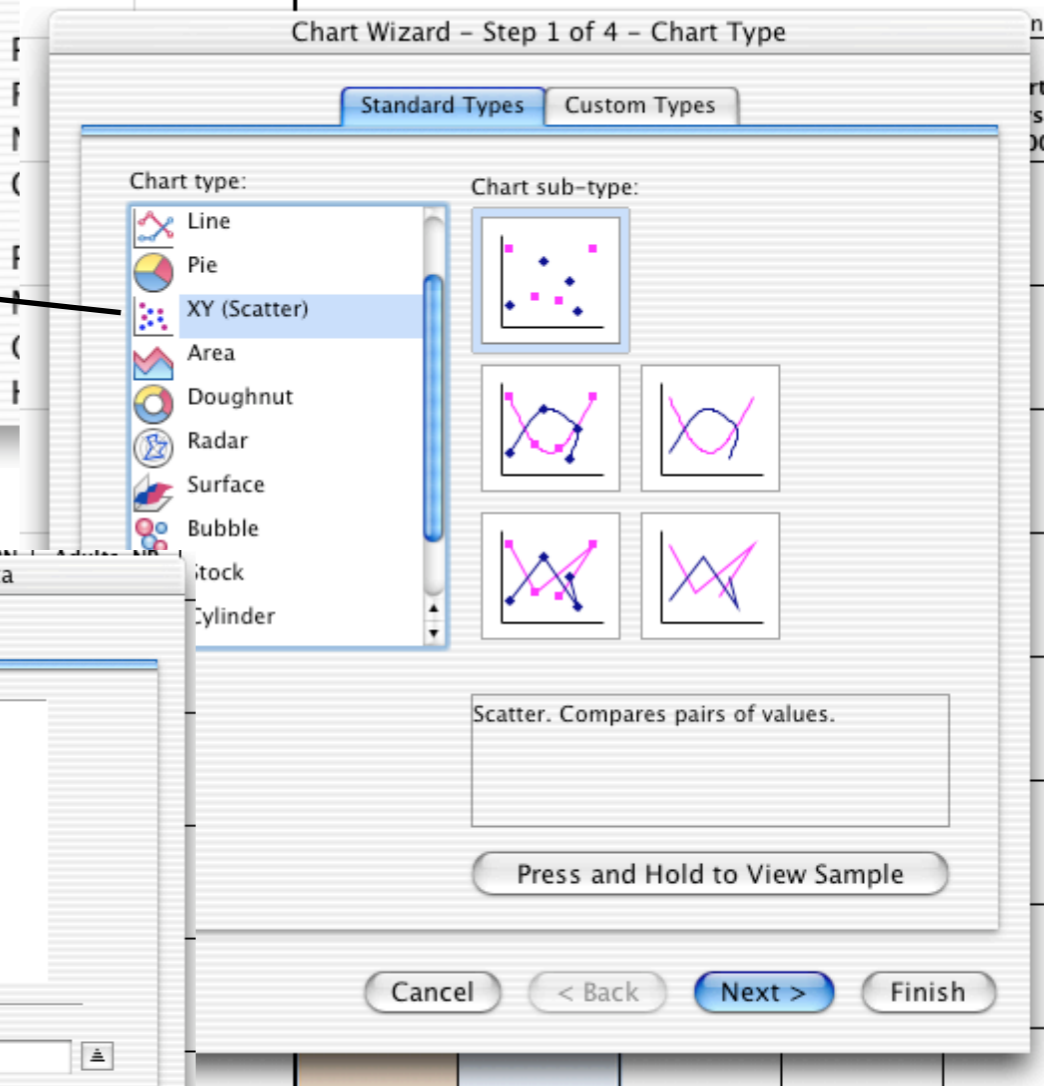
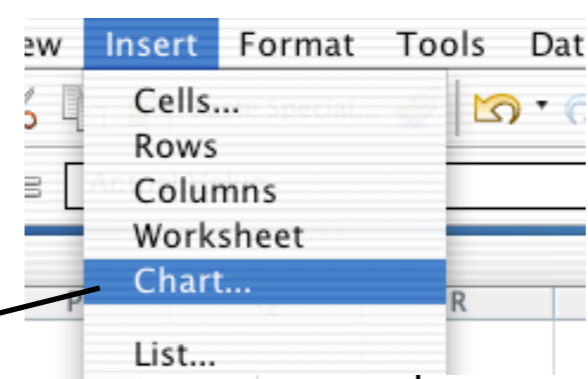
(Because estimates and actual values of zero cannot be modeled using certain regression models, it is best to add 1 to both actual and estimated quantities before beginning any analysis.)

Graphs may be produced for either groups' or individuals' estimates. In producing graphs for groups, there will be several estimates for each actual quantity. Use the median estimate so that wildly inaccurate individual estimates do not disproportionately affect the group data.

	Median + 1							
Actual Value	Second Graders, NP 1000	Second Graders, PN 1000	Fourth Graders, NP 1000	Fourth Graders, PN 1000	Sixth Graders, NP 1000	Sixth Graders, PN 1000	Adults, NP 1000	Adults, PN 1000
3	55	2	17	2	7	2	5	2
5	59	2	51	4.5	9	11	5	7.5
7	123	2	55	3	21	11	9	11
19	343	3	255	11	77	28.5	17	23.5
26	393	3	160	4	79	8.5	29	26
72	325	6	388	36	199	73.5	87	91
87	475	6	277	7.5	183	51	71	101
231	577	9.5	559	95.5	225	226	243	226
391	627	52	483	57.5	335	401	351	401
781	681	197	751	801	721	801	757	801
811	673	851	785	801	763	801	799	801

To create a graph in Excel

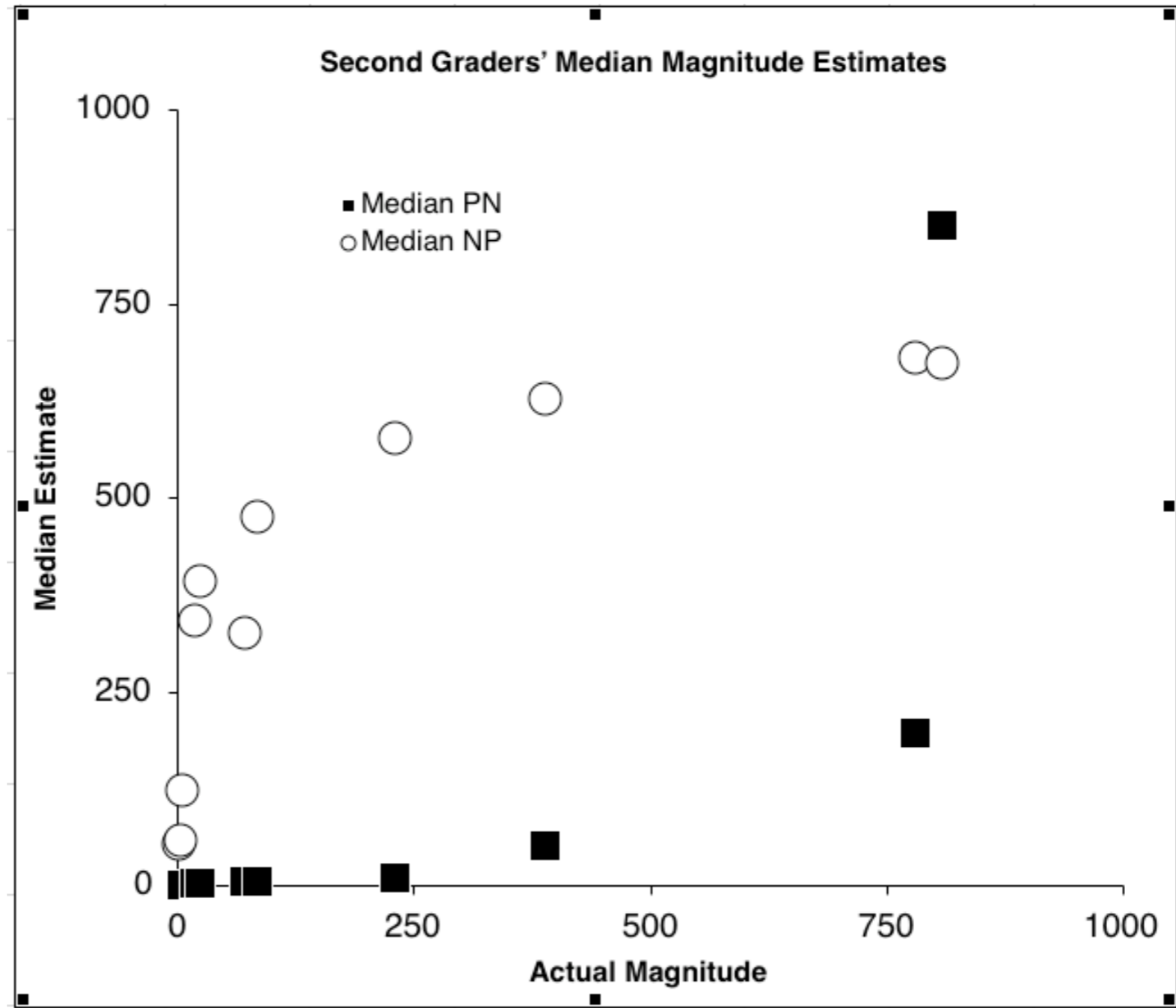
- 1 Select the data you wish to graph
- 2 Choose Insert > Chart... from the menu bar
- 3 Choose the XY (Scatter) chart.
- 4 Choose the series of actual quantities for the X values and the series of estimates for the Y values. Click Finish.



Actual Value	Graders, NP 1000
3	55
5	59
7	123
19	343
26	393
72	325
87	475
231	577

After the graph has been created, format it following the steps listed in the Excel user guide or in the on-line Help guide.

This graph will be used to obtain the fit of different regression functions in Excel.



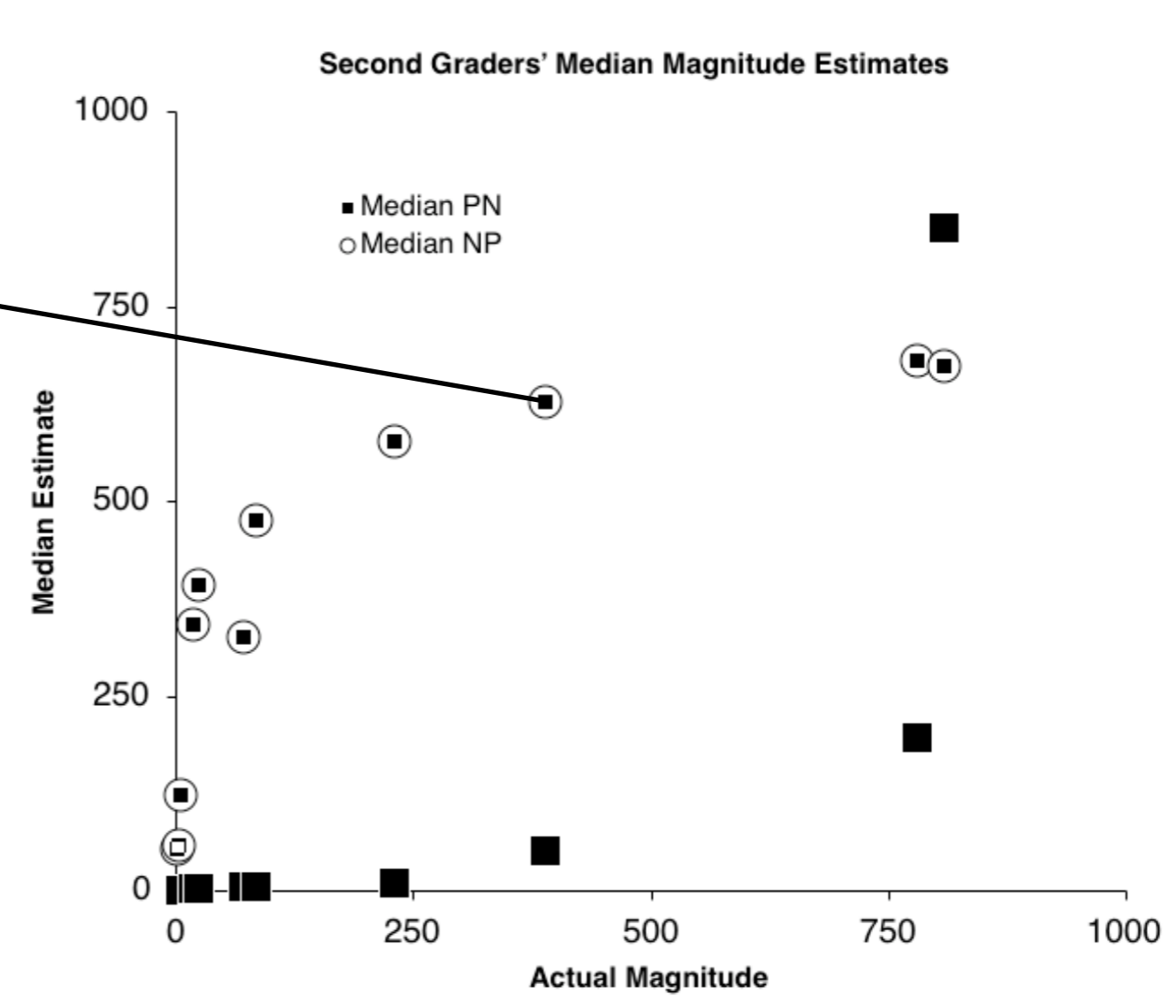
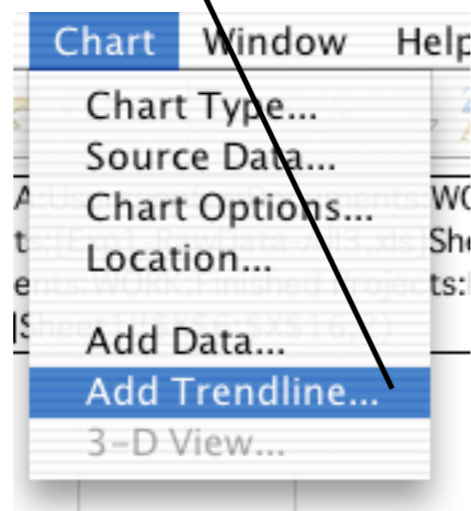
Fitting Estimates Using Regression Models

What is the mathematical relation between estimates and actual quantity? Regression models provide one class of relations, and Microsoft Excel provides an easy, fast way to find the best fitting regression model.

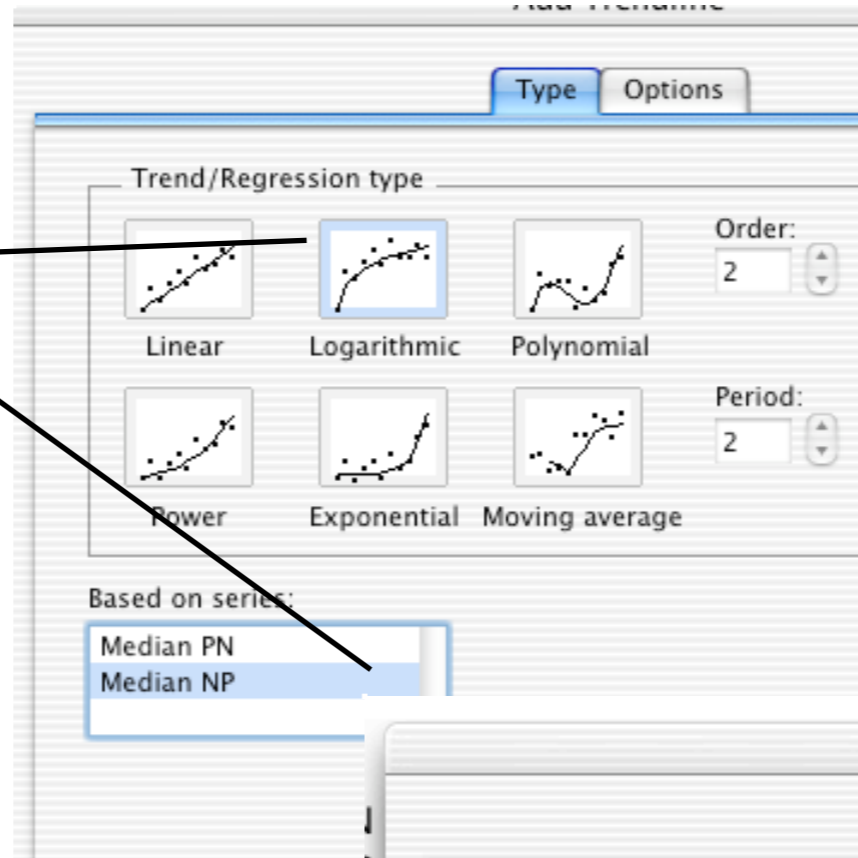
To generate regression lines

1 Select the data series on your chart that you wish to fit by single-clicking one element in the series.

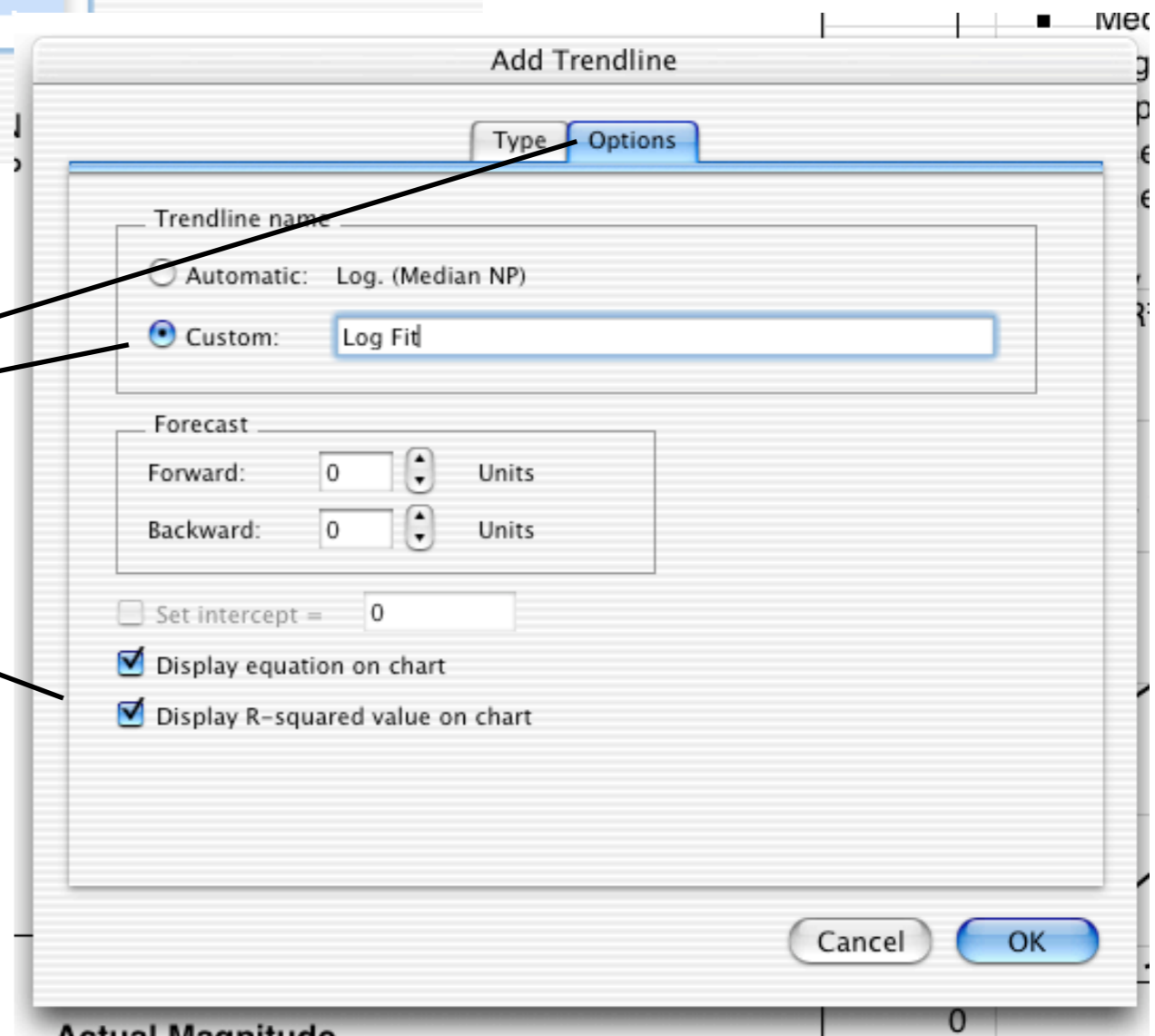
2 Select Chart > Add Trendline....



3 In the Add Trendline dialog box, choose the type of regression model you would like to use. Make sure that the correct series is highlighted.

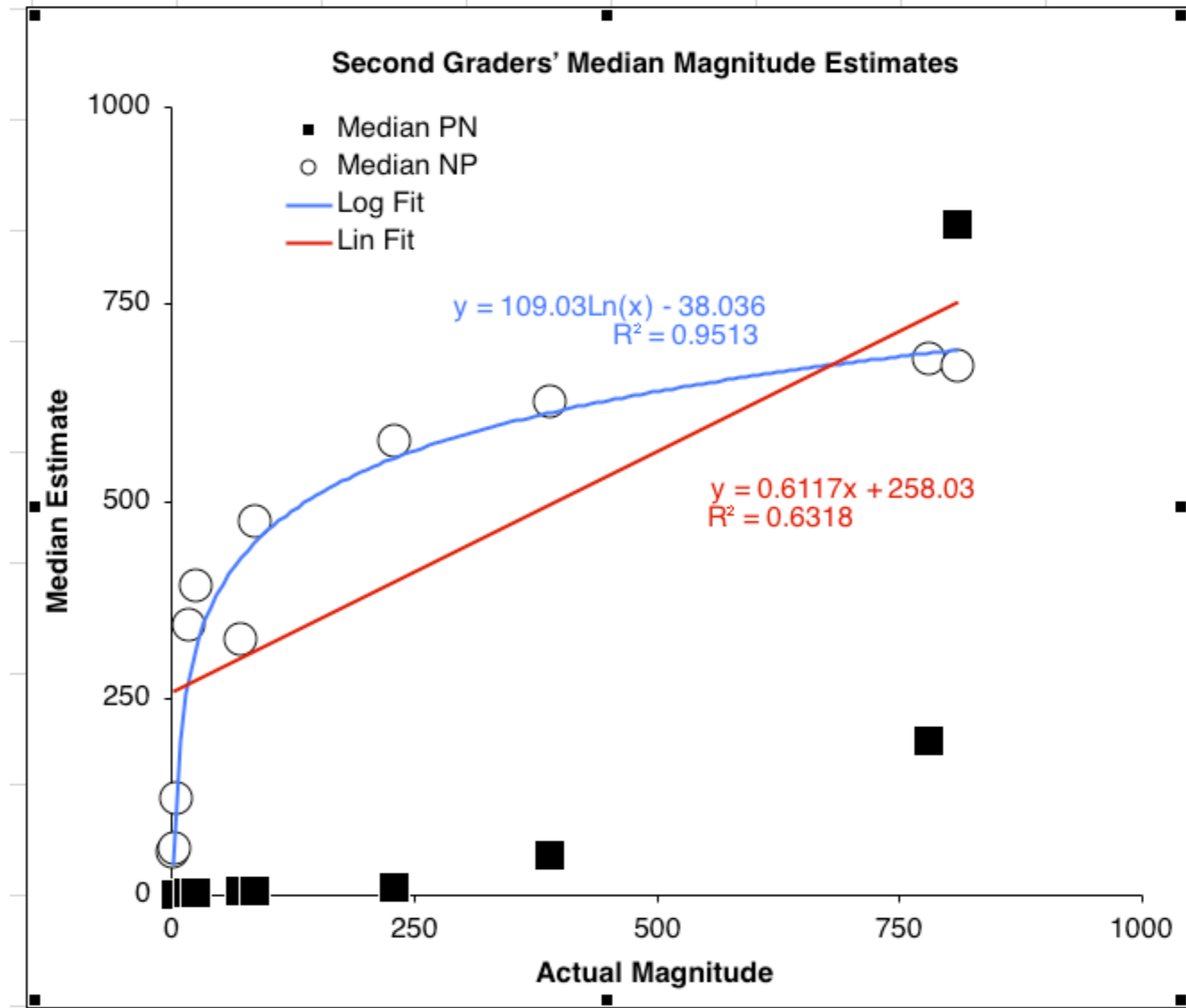


4 Select the Options tab to name your trendline and to display the equation and R-squared value on the chart.



5 Repeat steps 1 - 4 to produce additional trendlines.

The R-squared values on the finished chart indicate that the logarithmic regression fits second graders' Median NP estimates better than the linear regression line.



Comparing Regression Models

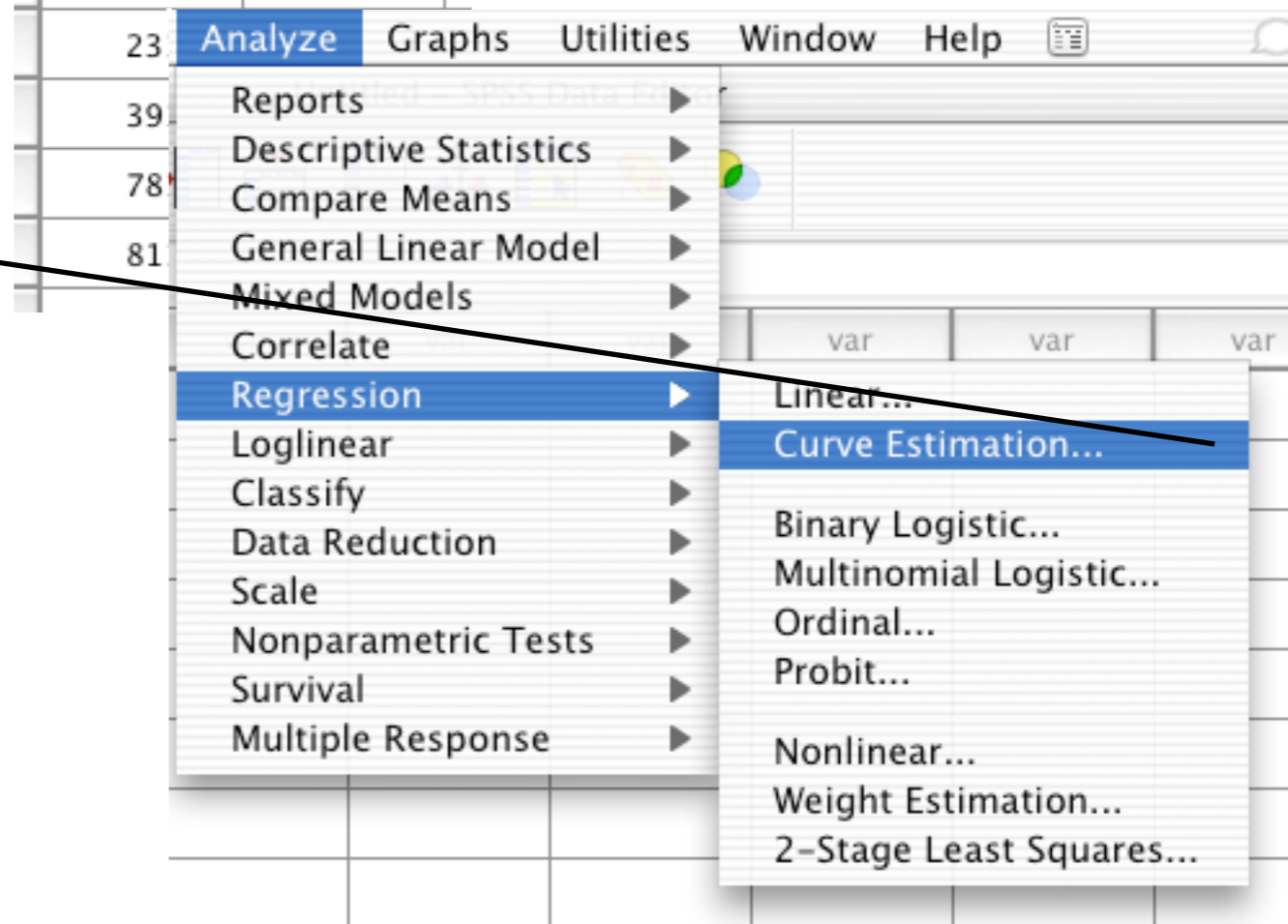
How reliable is the difference in fit between two regression lines? In this section, we'll use SPSS to address this question.

To compare the fit of two regression lines

1 Enter data in the SPSS Data Editor.

	actual	npest
	3.00	55.00
	5.00	59.00
	7.00	123.00
	19.00	343.00
	26.00	393.00
	72.00	325.00
	87.00	475.00

2 Choose Analyze > Regression > Curve Estimation...

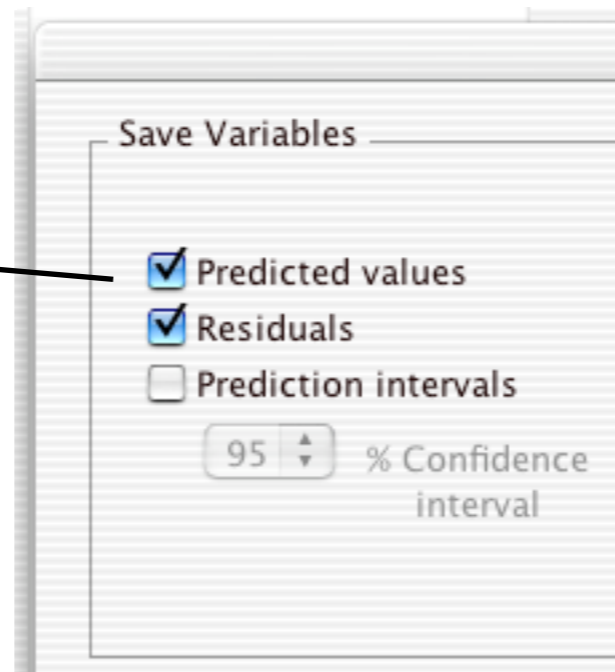
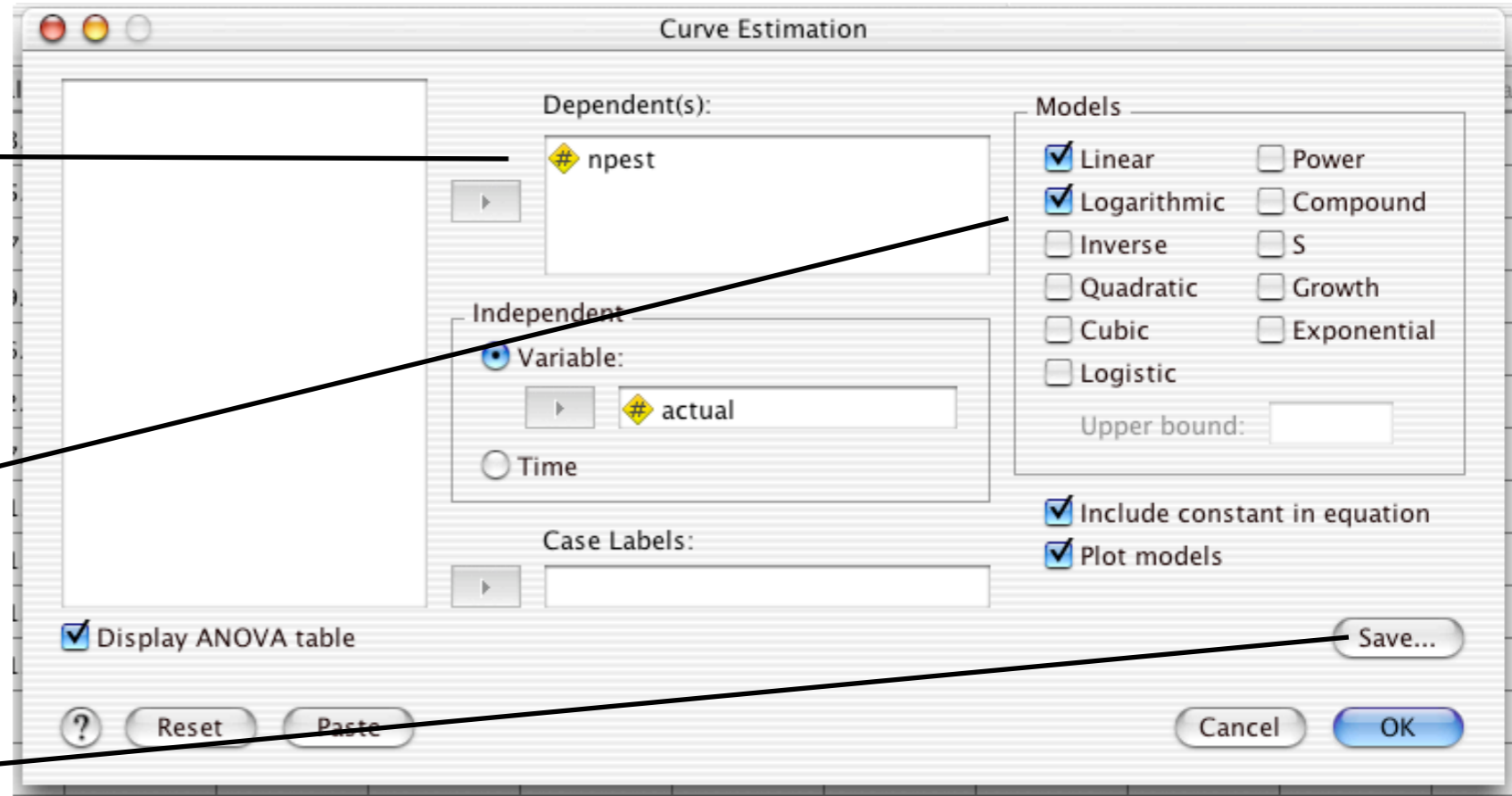


3 In the Curve Estimation pop-up window, choose the variable for actual values as your independent variable and the variable for estimates as your dependent variable.

4 Select the models you wish to compare, and check the “Include constant in equation” and “Plot models” boxes.

5 Choose Save...

6 Choose to Save Variables “Predicted values” and “Residuals”



The SPSS Data Editor should now display four new variable pictured to the right, with one variable (err_1) containing the residuals for the linear regression models and another variable (err_2) containing the residuals for the logarithmic regression model.

actual	npest	fit_1	err_1	fit_2	err_2
3.00	55.00	260.2557	-205.2557	57.44584	-2.44584
5.00	59.00	261.4792	-202.4792	115.9756	-56.9756
7.00	123.00	262.7027	-139.703	154.5282	-31.5282
19.00	343.00	270.0436	72.95640	268.9384	74.06161
26.00	393.00	274.3258	118.6742	304.8769	88.12311
72.00	325.00	302.4661	22.53393	421.5833	-96.5833
87.00	475.00	311.6422	163.3578	443.2665	31.73355
331.00	577.00	399.7335	177.2665	555.1537	21.84628
497.6127			129.3873	615.4554	11.54463
736.1931			-55.1931	694.7287	-13.7287
754.5455			-81.5455	699.0475	-26.0475

To compare the fit of the two models, perform a paired t-test on the absolute value of the residuals.

To perform t-test on residuals

1 Transform residuals for the two regression models into absolute values by choosing Transform > Compute...

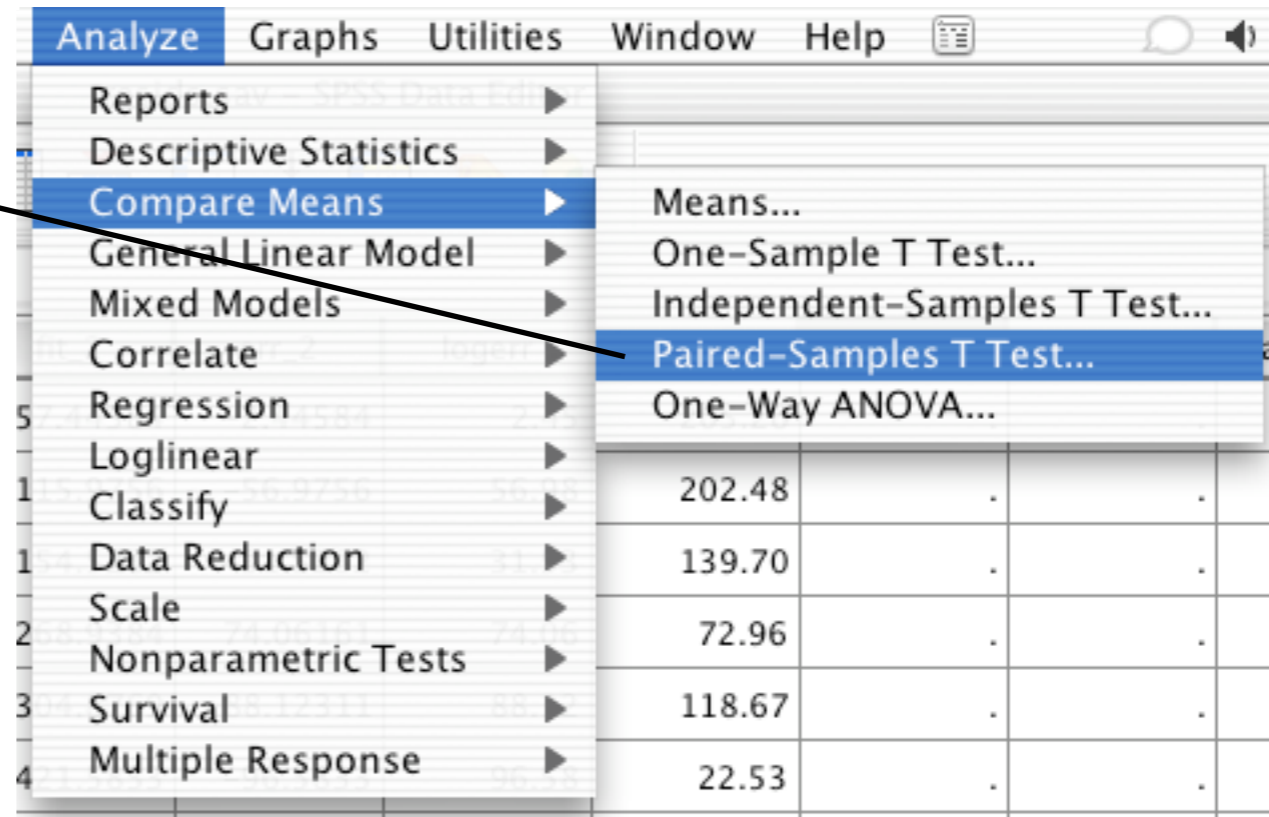
2 Create a new variable for the absolute value of the residuals in the Target Variable box.

3 Choose the ABS function and select the err_1 variable. Click OK.

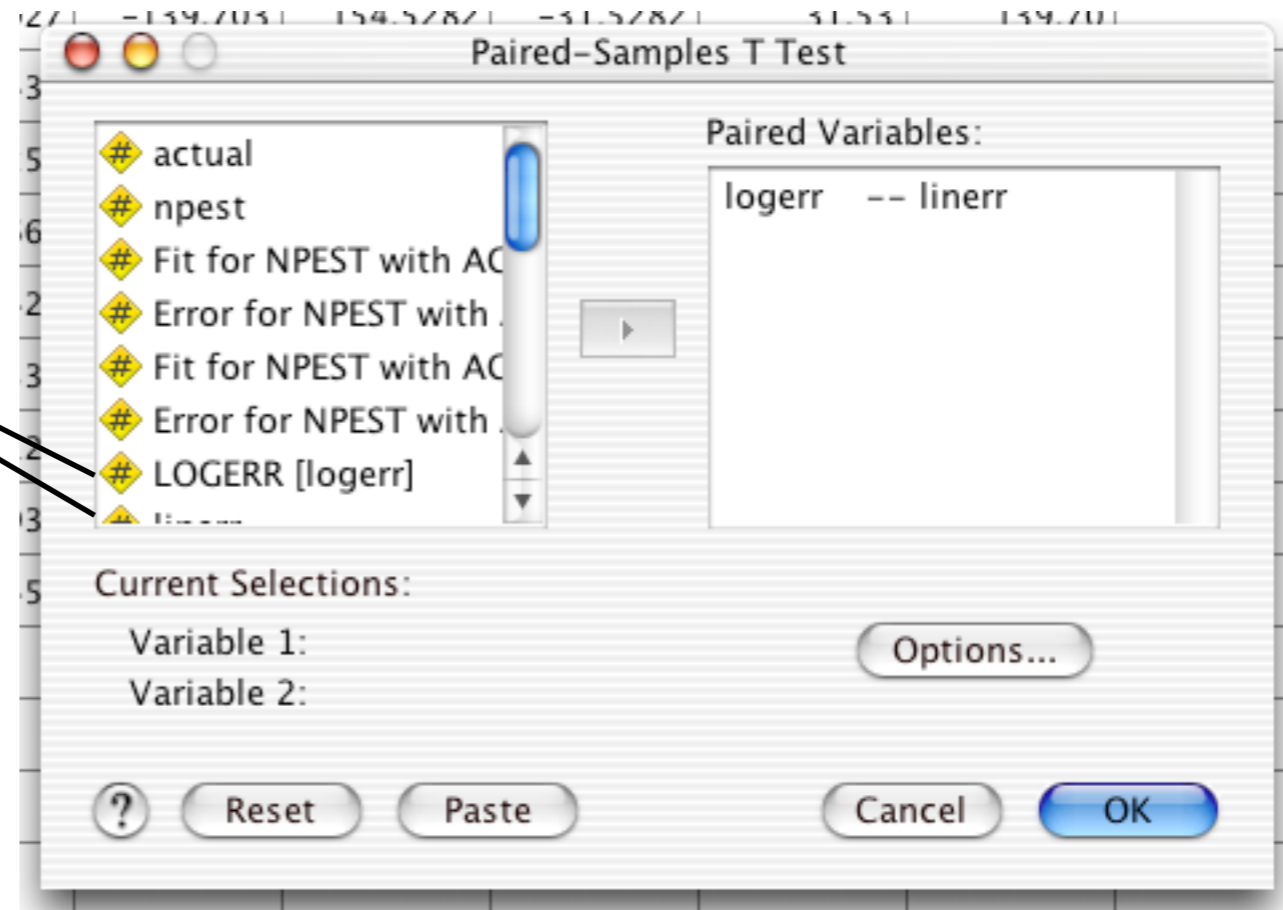
4 Repeat steps 1 - 3 for your next set of residuals.

The screenshot shows the SPSS 'Compute Variable' dialog box. The 'Target Variable' field contains 'LINERR'. The 'Numeric Expression' field contains 'ABS(err_1)'. Below the fields, there is a list of variables including 'actual', 'npest', 'Fit for NPEST with...', 'Error for NPEST wit...', 'LOGERR [logerr]', 'linerr', 'var00007', and 'var00008'. A 'Functions' list is open, showing 'ABS(numexpr)' selected. The background shows the 'Transform' menu with 'Compute...' selected.

5 Choose Analyze > Compare Means > Paired-Samples T Test...



6 Select the variables you created in steps 2 and 4 as Variable 1 and Variable 2. Click OK.



The SPSS Viewer should display your statistical output. In this data set, the residuals from the linear regression were reliably greater than the residuals from the logarithmic regression, $t(10) = 3.45, p < .01$.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	LOGERR	41.3289	11	32.43607	9.77984
	LINERR	124.396	11	60.73520	18.31235

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	LOGERR & LINERR	11	-.439	.177

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	LOGERR - LINERR	*****	80.42915	24.25030	*****	*****	-3.425	10	.006