

# Chapter 21

## Comparing Analyses

### Chapter Table of Contents

---

<b>COMPARING ANALYSES OF DIFFERENT OBSERVATIONS</b> . . . . .	314
Extracting Observations . . . . .	314
Excluding Observations . . . . .	318
<b>COMPARING ANALYSES OF DIFFERENT VARIABLES</b> . . . . .	323
Deleting Variables . . . . .	323
Transforming Variables . . . . .	325



# Chapter 21

## Comparing Analyses

You can compare analyses that use different observations or variables. For example, you can exclude certain observations from a model and see how that affects the fit. You can delete and transform variables to create and compare different models.

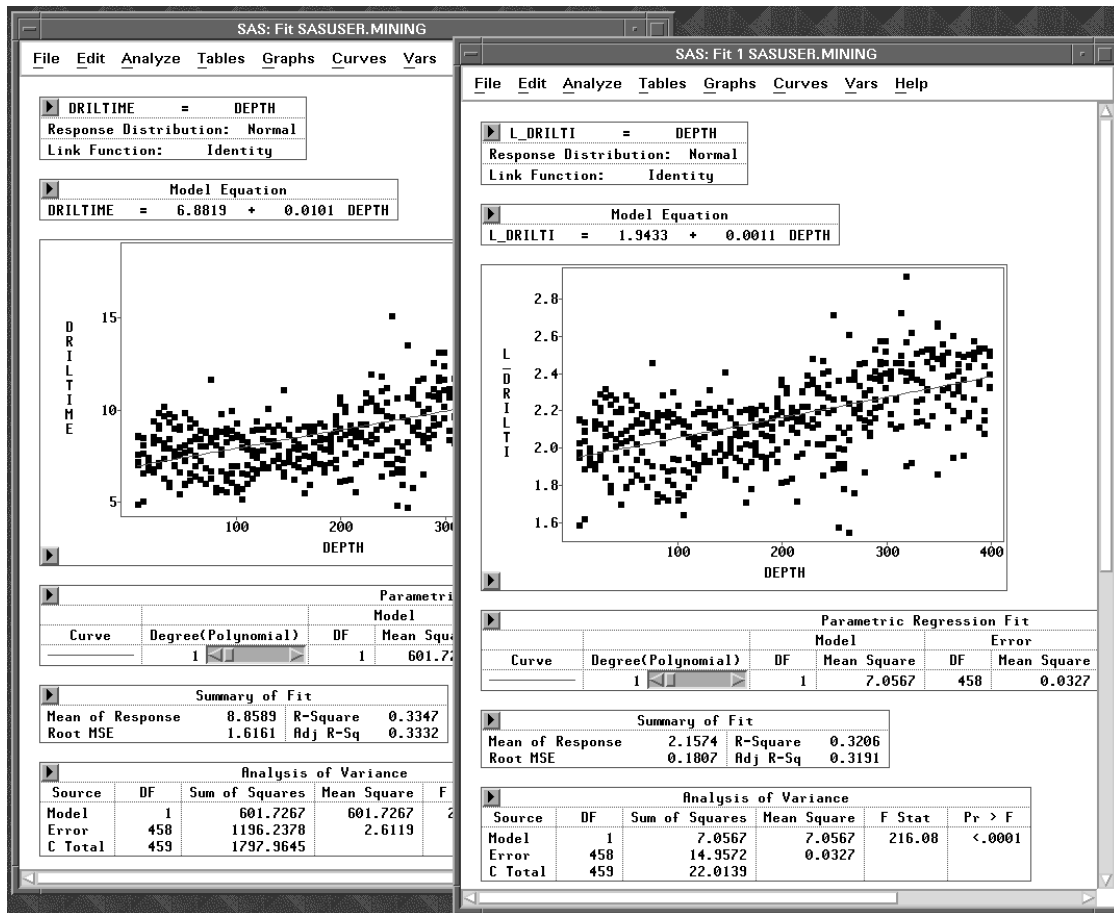


Figure 21.1. Comparing two Regression Analyses

## Comparing Analyses of Different Observations

There are two ways to compare analyses that use different observations. You can *extract* observations or you can *exclude* them.

### Extracting Observations

You can compare analyses made with different observations by *extracting* a subset, that is, by creating a new data set that contains a subset of observations from the original data set. Then you can request separate analyses for each data set.

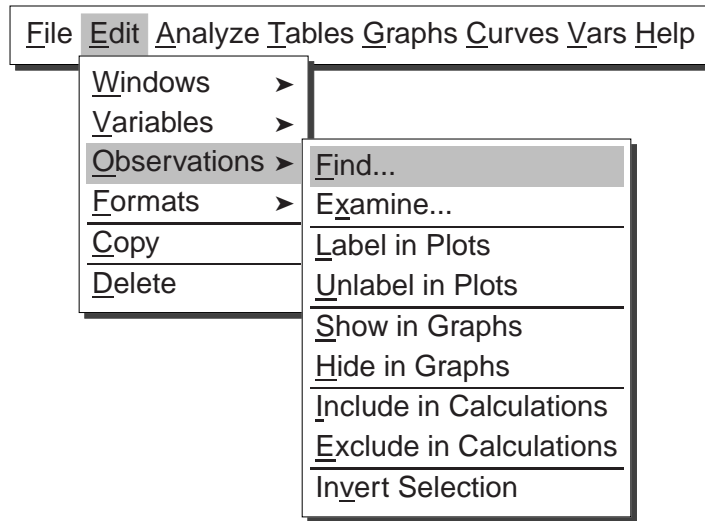
Consider the **MINING** data. This data set contains results of an experiment to examine drilling times (**DRILTIME**) for different drilling methods (**METHOD**). As it turned out, the experimenters encountered difficulties due to changing rock types after a depth of about 200 feet. It might be worthwhile to compare the distribution of **DRILTIME** for depths greater than 200 feet to the distribution of **DRILTIME** for the entire data set. To compare the two distributions, you need to select the observations where **DEPTH** is greater than 200 feet and extract them into a new data window.

⇒ Open the **MINING** data set.

	4	Int	Int	Nom	Int				
480		DEPTH	DRILTIME	METHOD	REP				
■	1	5	7.61	Wet	1				
■	2	5	8.68	Wet	2				
■	3	5	8.61	Wet	3				
■	4	5	7.25	Dry	1				
■	5	5	7.07	Dry	2				
■	6	5	4.90	Dry	3				
■	7	10	8.16	Wet	1				
■	8	10	8.13	Wet	2				
■	9	10	7.71	Wet	3				
■	10	10	8.55	Dry	1				
■	11	10	6.62	Dry	2				
■	12	10	5.07	Dry	3				

Figure 21.2. **MINING** Data

⇒ Choose **Edit:Observations:Find**.



**Figure 21.3.** Finding Observations

This displays the **Find Observations** dialog.



**Figure 21.4.** Find Observations Dialog

⇒ Select > in the **Test** list and **200** in the **Value** list.



Figure 21.5. Finding **DEPTH** > 200

⇒ Click the **OK** button.

This selects all observations where **DEPTH** is greater than 200 feet. To see the selected observations, either choose **Find Next** from the data pop-up menu or scroll down using the vertical scroll bar on the right (as indicated by the arrow in the figure).

	Int	Int	Nom	Int
480	DEPTH	DRILTIME	METHOD	REP
238	200	10.42	Dry	1
239	200	10.76	Dry	2
240	200	6.15	Dry	3
241	205	9.97	Wet	1
242	205	8.71	Wet	2
243	205	10.19	Wet	3
244	205	7.67	Dry	1
245	205	10.24	Dry	2
246	205	6.19	Dry	3
247	210	8.19	Wet	1
248	210	8.80	Wet	2
249	210	8.95	Wet	3

Figure 21.6. Observations where **DEPTH** > 200

⇒ Choose **Extract** from the data pop-up menu.

A new data set containing observations where **DEPTH** is greater than 200 feet appears, as shown in Figure 21.7. The new data window is named automatically by adding a subscript to the original name. You may have to scroll to the top of the data window to duplicate the next figure.

4	Int	Int	Nom	Int					
240	DEPTH	DRILTIME	METHOD	REP					
1	205	9.97	Wet	1					
2	205	8.71	Wet	2					
3	205	10.19	Wet	3					
4	205	7.67	Dry	1					
5	205	10.24	Dry	2					
6	205	6.19	Dry	3					
7	210	8.19	Wet	1					
8	210	8.80	Wet	2					
9	210	8.95	Wet	3					
10	210	8.32	Dry	1					
11	210	9.22	Dry	2					
12	210	6.29	Dry	3					

Figure 21.7. MINING1 Data

Now create distribution analyses for both data sets.

⇒ Select **DRILTIME** in the **MINING** data window.

⇒ Choose **Analyze:Distribution (Y)**.

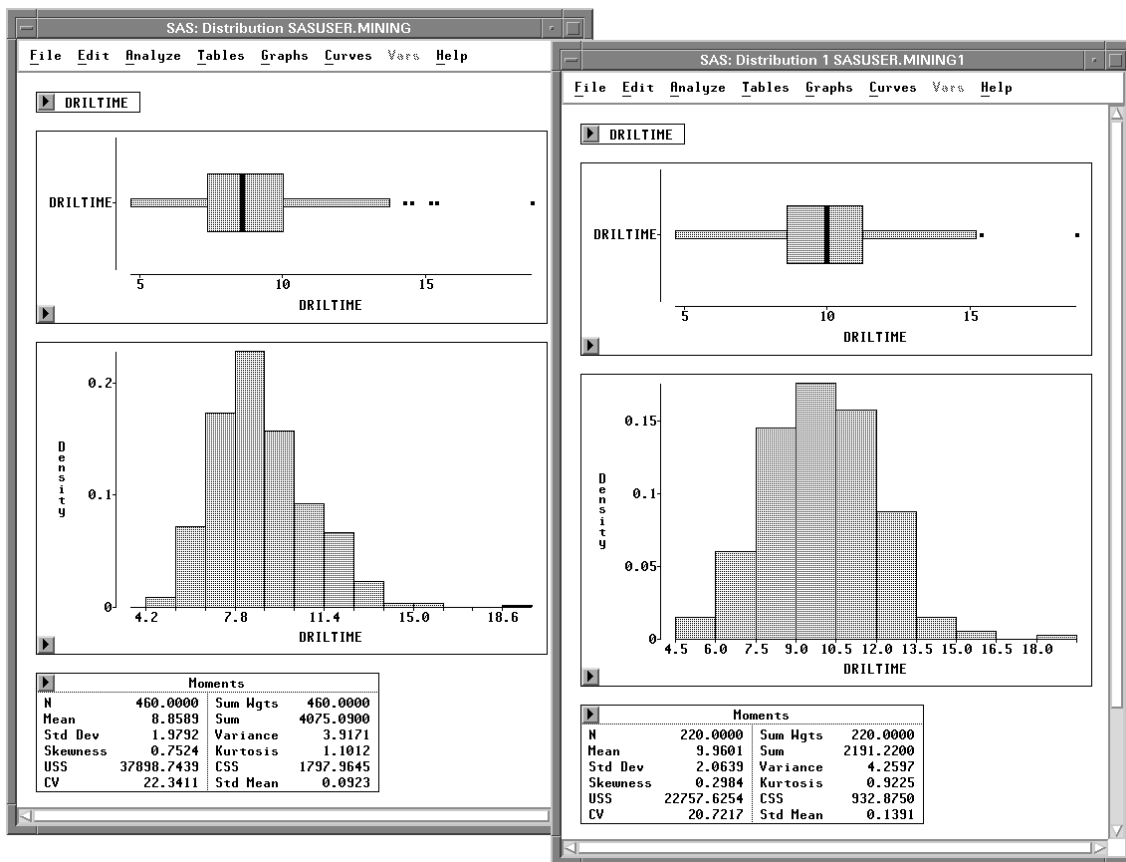
A distribution analysis using all the observations appears on your display.

⇒ Select **DRILTIME** in the **MINING1** data window.

⇒ Choose **Analyze:Distribution (Y)**.

A distribution analysis using the subset of observations appears on your display.

⇒ Move the two analysis windows side-by-side to compare the distributions.



**Figure 21.8.** Comparing Two Distribution Analyses

The mean drilling time at depths greater than 200 feet was 9.9601, while the mean overall was only 8.8589. The drills may have found harder rock at greater depths. You may want to create an additional analysis to compare depths greater than 200 feet with depths less than or equal to 200 feet.

⇒ Choose **File:End** to delete **MINING1** and the two analysis windows.

† **Note:** Sometimes you will want to compare analyses that use different subsets of observations based on the values of some variable. If this is the case, you can assign the variable the **Group** role, as described in Chapter 22, “Analyzing by Groups.”

⊕ **Related Reading:** Distributions, Chapter 38.

## Excluding Observations

Another way to compare analyses using different observations is to *exclude* observations, that is, to remove them from calculations in the analysis. The observations still appear in graphs. To illustrate this technique, consider a simple linear regression model with **DRILTIME** as the response variable and **DEPTH** as the explanatory variable.

⇒ Select **DRILTIME**, then **DEPTH**, then choose **Analyze:Fit (Y X)**.

This displays a fit window.



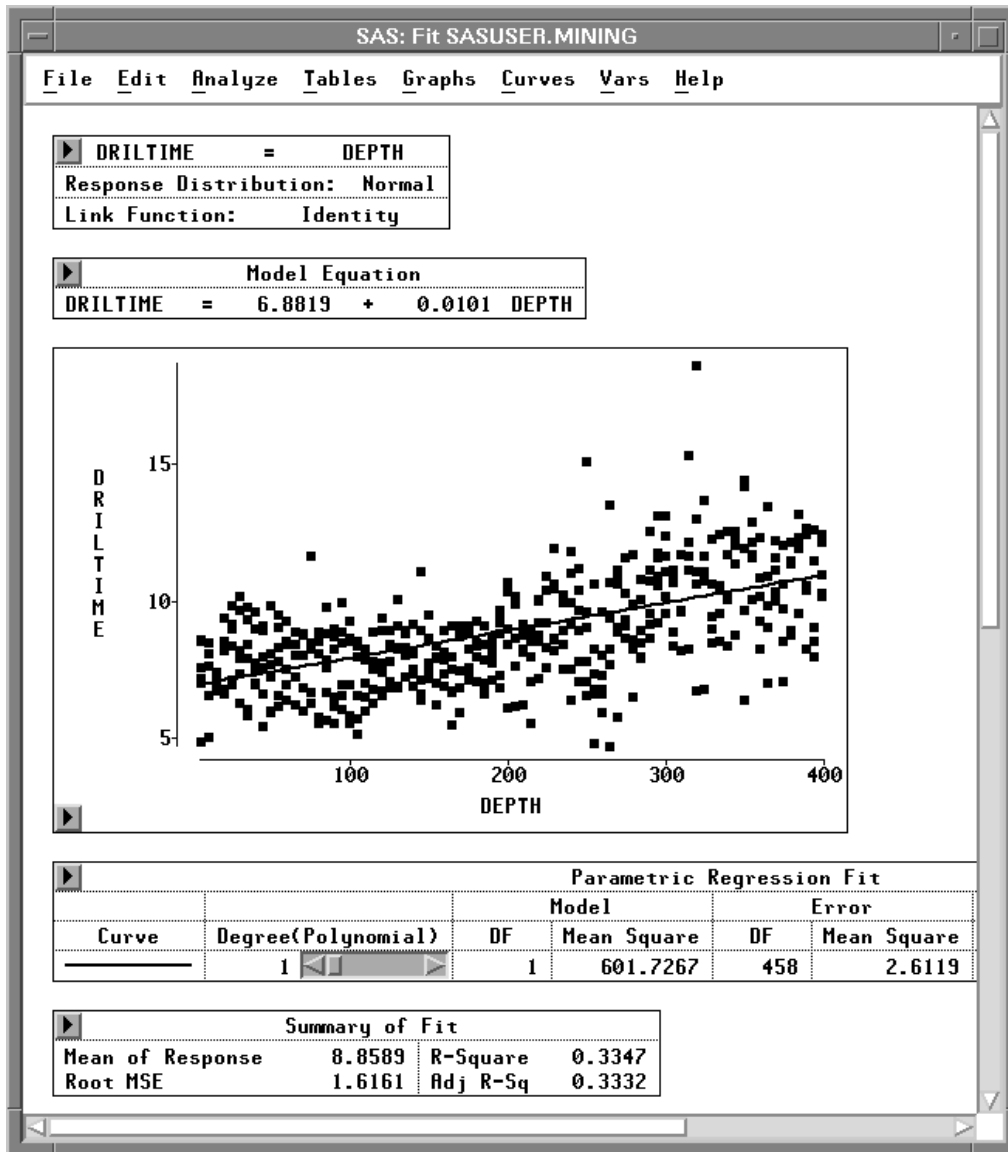
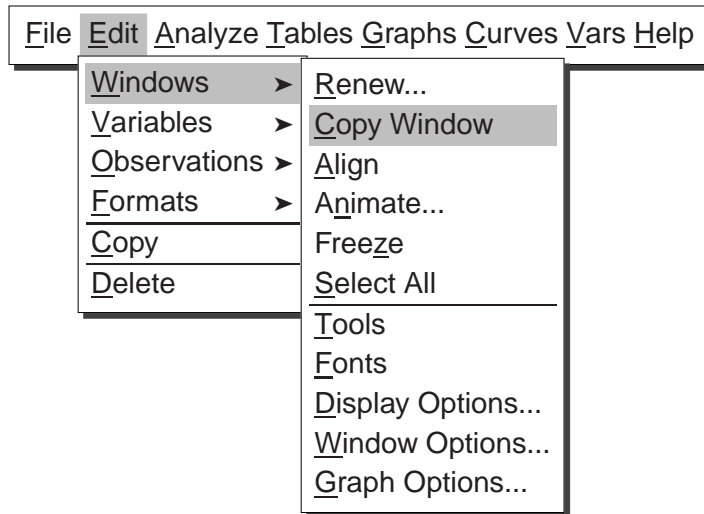


Figure 21.9. Fit Window

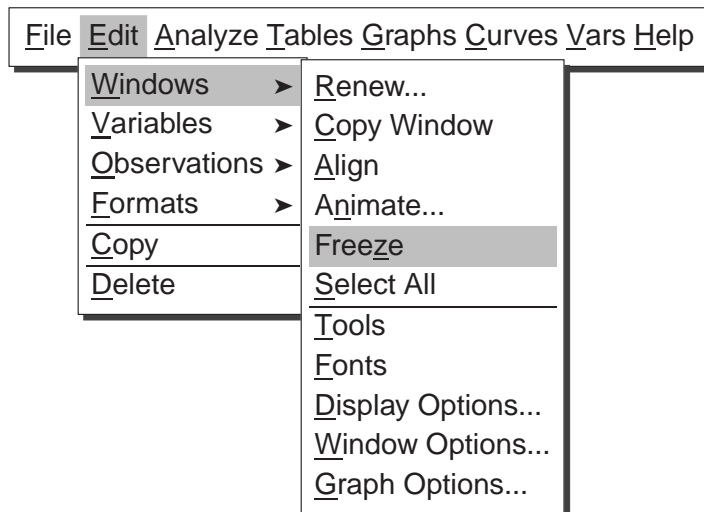
⇒ Choose **Edit:Windows:Copy Window** in the fit window.

This creates a copy of the fit window.



**Figure 21.10.** Edit:Windows Menu

- ⇒ Move the two fit windows side by side.
- ⇒ Choose **Edit:Windows:Freeze** in the fit window on the left.



**Figure 21.11.** Edit:Windows Menu

This freezes the window, as indicated by the frost in the corners of the window. *Freezing* a window converts the window to a static image that ignores any changes to the data. Normally, all SAS/INSIGHT windows are linked to their data, and any changes to the data are automatically reflected in all analyses. By freezing a window, you can compare windows using different observations without creating additional data sets.

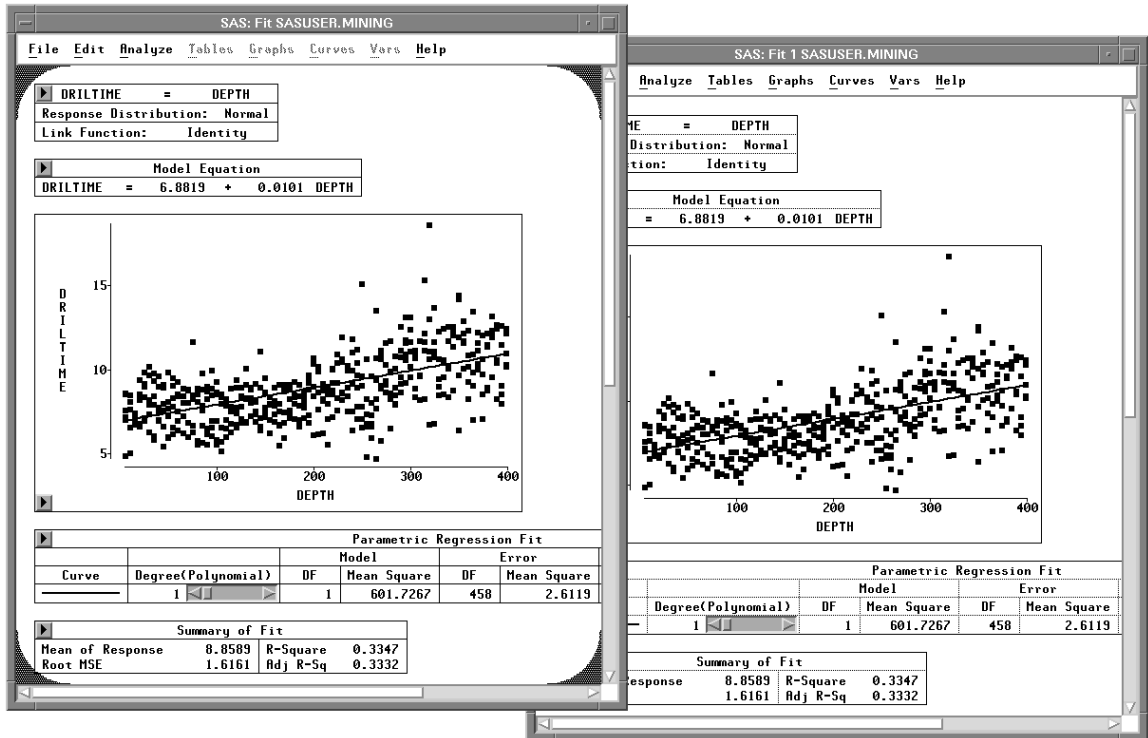


Figure 21.12. Two Windows, One Frozen

Now exclude a few observations from the window on the right.

- ⇒ Select the three observations with the largest values of **DRILTIME** in the scatter plot.
- ⇒ Choose **Edit:Observations:Exclude in Calculations**.

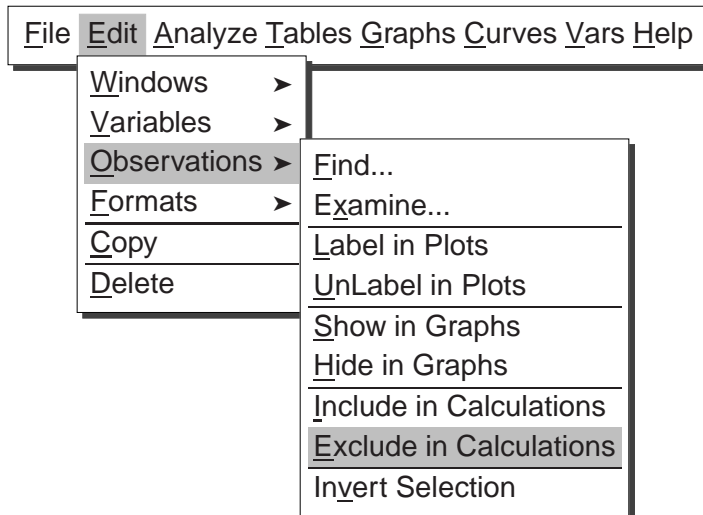
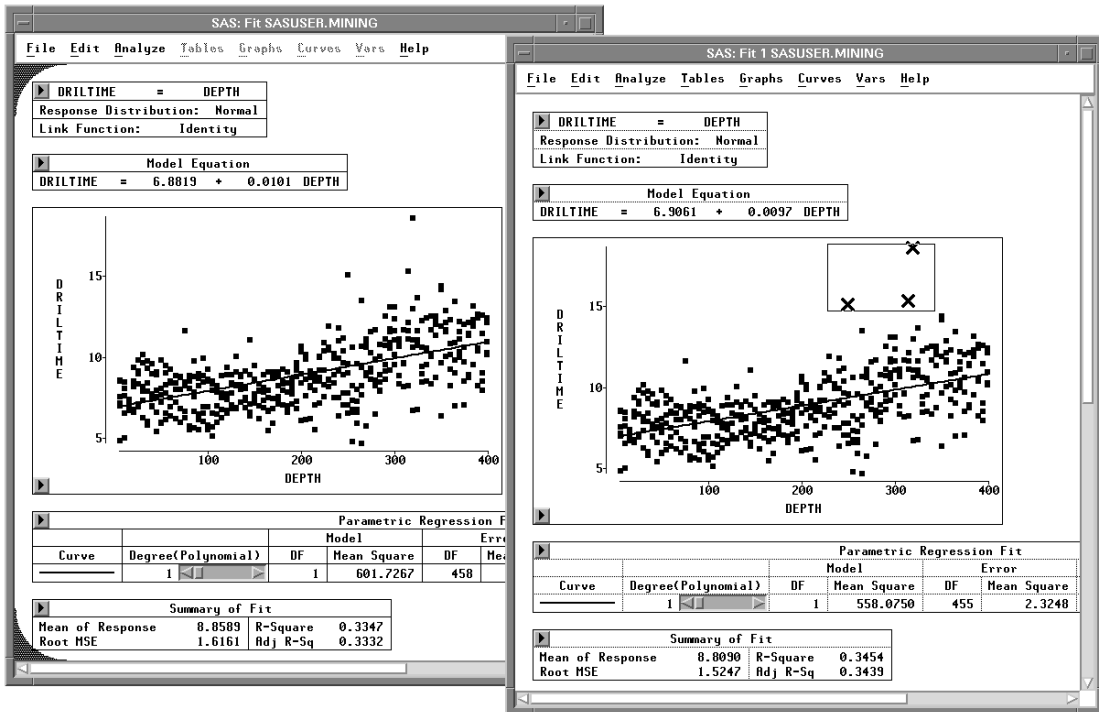


Figure 21.13. Edit: Observations Menu

## Part 2. Introduction

This recalculates the fit analysis without the selected observations. Normally, both windows would be recalculated, but since the window on the left is frozen, it does not change. Now you can compare the two fit windows.



**Figure 21.14.** Comparing Two Fit Windows

To thaw a frozen window, follow these steps.

⇒ **Choose Edit:Windows:Freeze again.**

This recalculates the frozen window and restores its dynamic behavior.

⇒ **Close all analysis windows before proceeding to the next section.**

## Comparing Analyses of Different Variables

You have already seen one easy way to compare analyses using different variables. The **Apply** button, discussed in Chapter 14, “Multiple Regression,” and Chapter 16, “Logistic Regression,” enables you to create models quickly with different effects.

In this section, you will see two additional ways to compare analyses using different variables. In any analysis, you can *delete* variables or you can *transform* them.

### Deleting Variables

You can delete any effect in a fit analysis. To see this, do the following:

⇒ **Select DRILTIME, then DEPTH, then METHOD in the data window.**

⇒ **Choose Analyze:Fit (Y X).**

A fit window appears, as shown in Figure 21.15.

The screenshot shows the SAS Fit window for the model  $DRILTIME = DEPTH \text{ METHOD}$ . The window includes the following sections:

- Model Equation:**  $DRILTIME = 7.5113 + 0.0098 \text{ DEPTH} - 1.1873 \text{ P}_3$
- Summary of Fit:**

Mean of Response	8.8589	R-Square	0.4243
Root MSE	1.5049	Adj R-Sq	0.4218
- Analysis of Variance:**

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	2	762.9420	381.4710	168.43	<.0001
Error	457	1035.0226	2.2648		
C Total	459	1797.9645			
- Type III Tests:**

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
DEPTH	1	562.7783	562.7783	248.49	<.0001
METHOD	1	161.2152	161.2152	71.18	<.0001
- Parameter Estimates:**

Variable	METHOD	DF	Estimate	Std Error	t Stat	Pr >  t
Intercept		1	7.5113	0.1587	47.32	<.0001
DEPTH		1	0.0098	0.0006	15.76	<.0001
METHOD	Dry	1	-1.1873	0.1407	-8.44	<.0001
	Het	0	0	.	.	.

Figure 21.15. Fit Window

⇒ Choose **Edit:Windows:Copy Window**.

Now you have two identical fit windows.

⇒ Select **METHOD** in one of the fit windows.

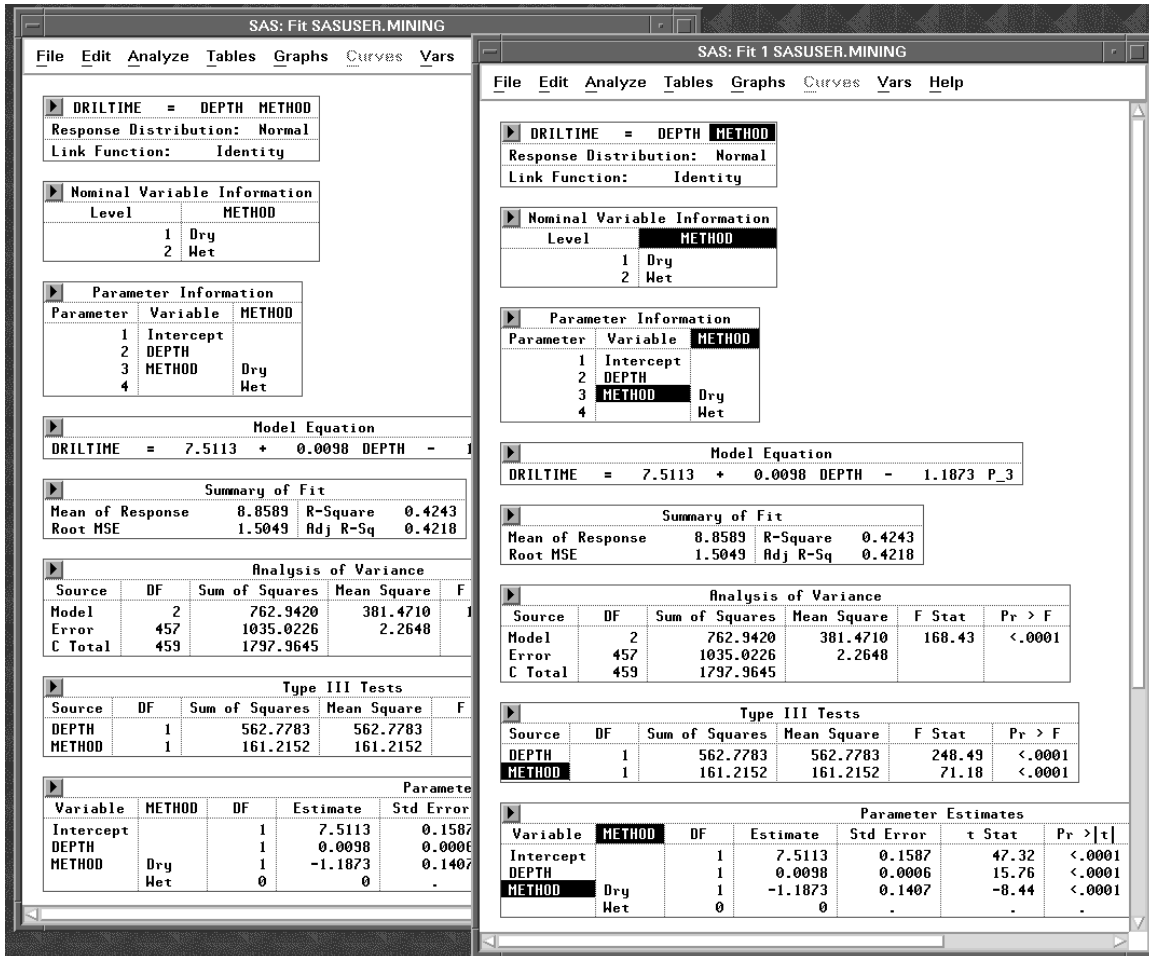


Figure 21.16. Two Fit Windows, **METHOD** Selected in One

⇒ Choose **Edit>Delete**.

This recalculates the fit window without the effect you deleted. Now you have two fit windows for two different models.

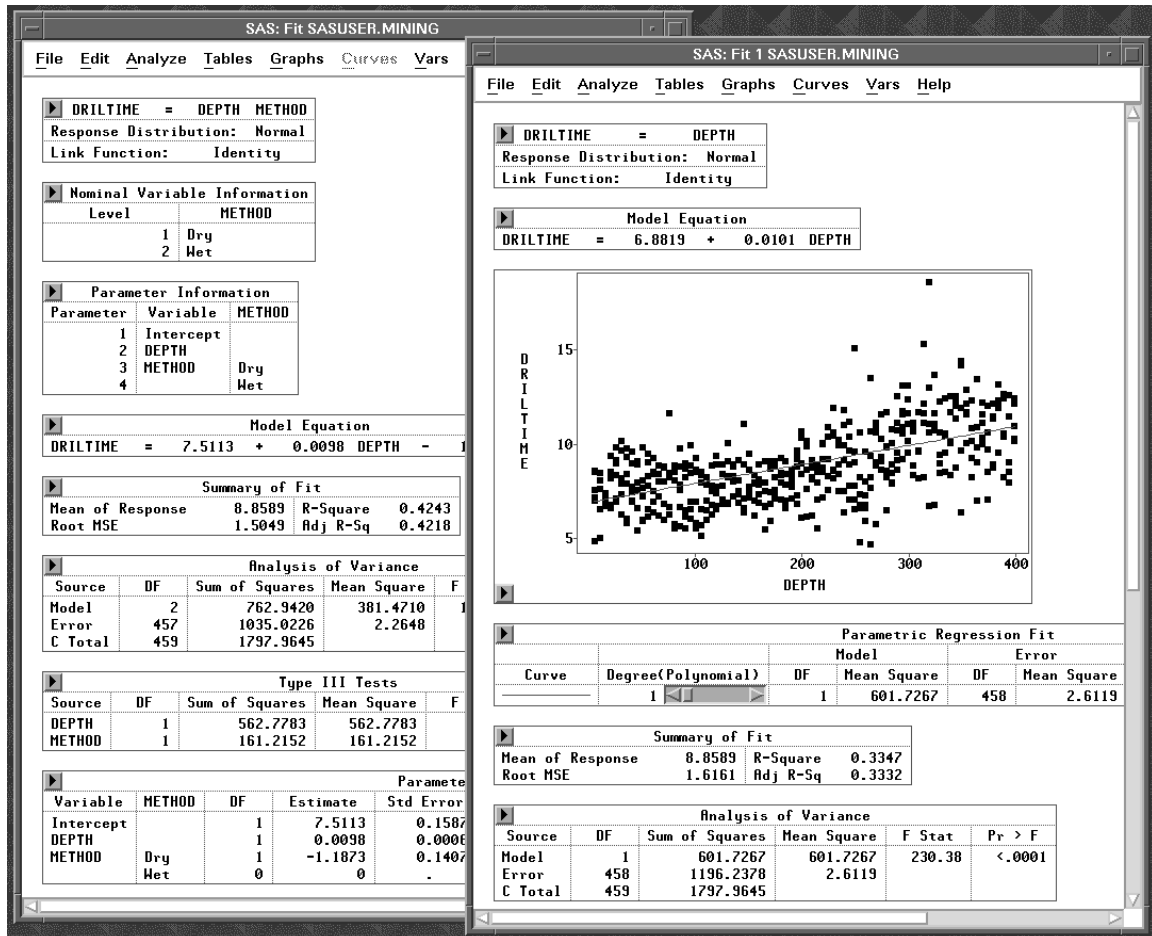


Figure 21.17. Comparing Two Models

Deleting **METHOD** caused the adjusted R-square value to drop from 0.4218 to 0.3332. It was expected that different drilling methods might produce different drilling times.

## Transforming Variables

You can compare analyses by transforming variables in any window.

⇒ Create identical fit windows for **DRILTIME = DEPTH**.

Either delete **METHOD** from the first window or choose **Edit:Windows:Copy Window** in the second window.

⇒ Select **DRILTIME** in one of the fit windows.

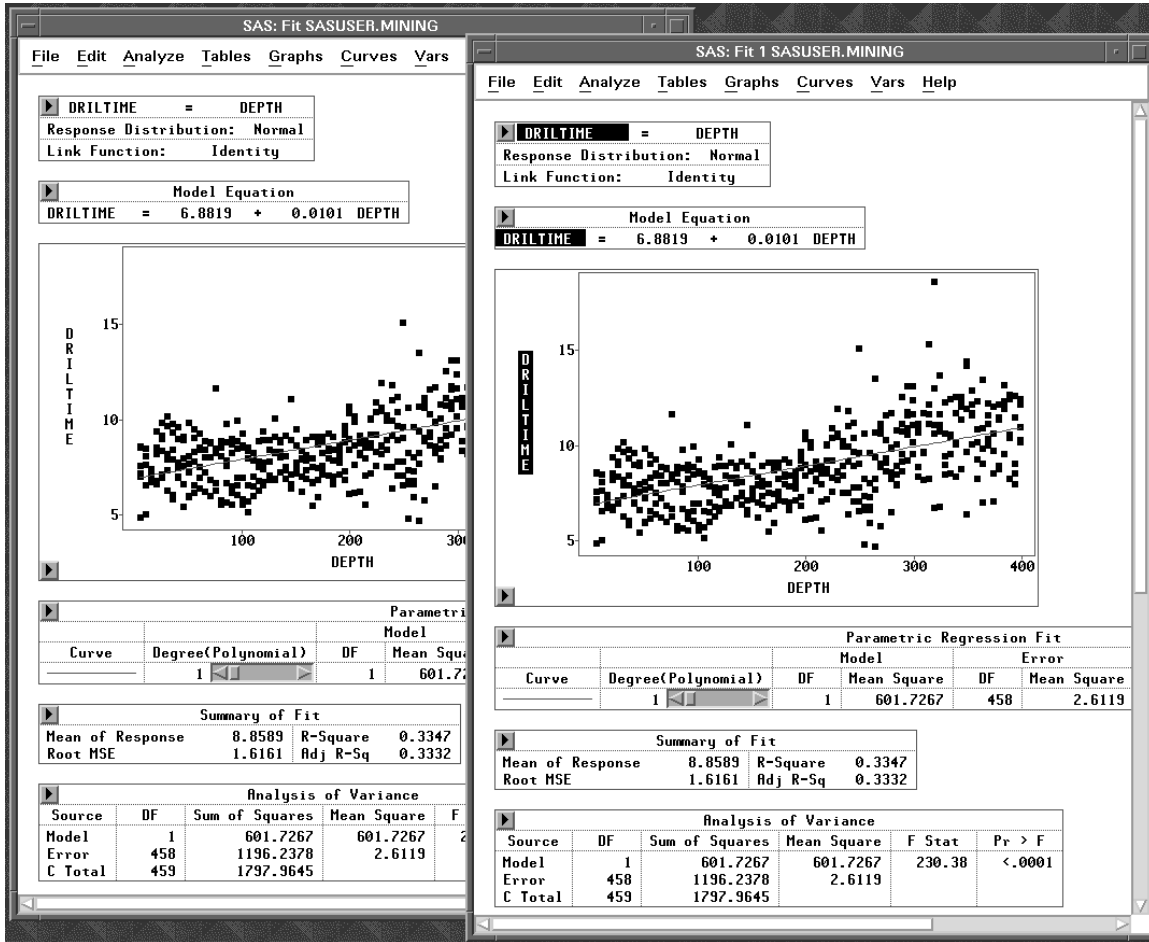


Figure 21.18. Two Fit Windows, DRILTIME Selected

⇒ Choose Edit:Variables:log( Y ).

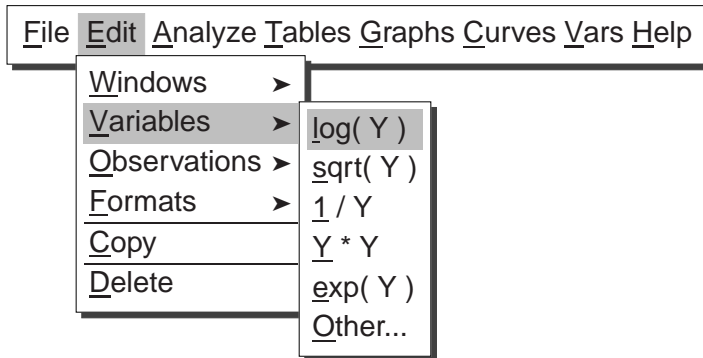


Figure 21.19. Edit:Variables Menu

This recalculates the fit window using the log of the response variable (L\_DRILTI). Now you have two fit windows for two different models.



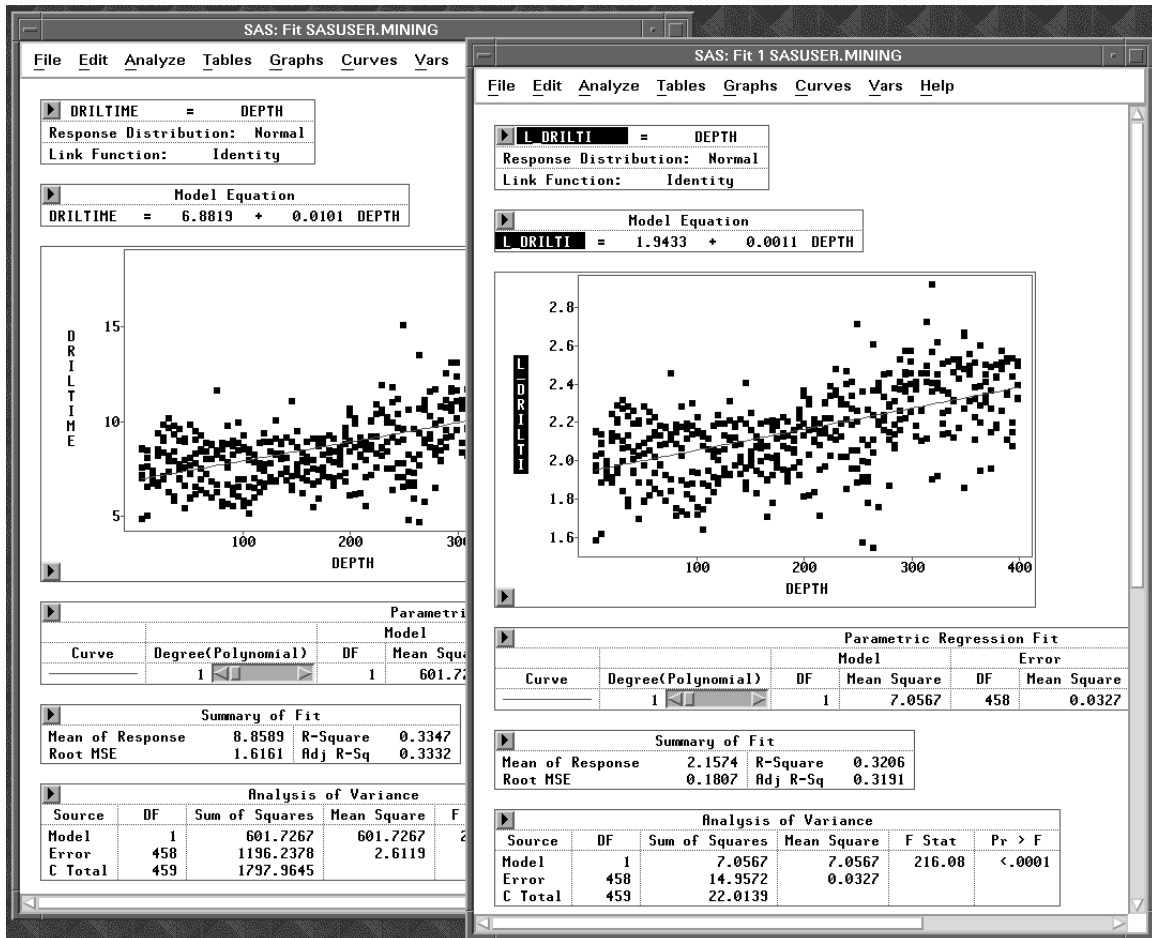


Figure 21.20. Comparing Two Fit Analyses

In this case, the log transform did not improve the fit. To undo the log transform, you can choose **Edit:Windows:Renew**.

In this chapter you have seen how to compare analysis windows that use different observations by extracting and excluding. You have also compared analyses using different variables by deleting and transforming. In the next chapter, you will see how to compare analyses using **Group** variables.

- ⊕ **Related Reading:** Transformations, Chapter 20.
- ⊕ **Related Reading:** Linear Models, Chapter 39.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/INSIGHT User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 752 pp.

**SAS/INSIGHT User's Guide, Version 8**

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-490-X

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**U.S. Government Restricted Rights Notice.** Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.