Chapter 21
Comparing Analyses

Chapter Table of Contents

<table>
<thead>
<tr>
<th>COMPARING ANALYSES OF DIFFERENT OBSERVATIONS</th>
<th>318</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracting Observations</td>
<td>318</td>
</tr>
<tr>
<td>Excluding Observations</td>
<td>322</td>
</tr>
<tr>
<td>COMPARING ANALYSES OF DIFFERENT VARIABLES</td>
<td>327</td>
</tr>
<tr>
<td>Deleting Variables</td>
<td>327</td>
</tr>
<tr>
<td>Transforming Variables</td>
<td>329</td>
</tr>
</tbody>
</table>
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.

![MINING Data](image)

→ Choose Edit:Observations:Find.

SAS OnlineDoc™: Version 7-1
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.
Chapter 21. Comparing Analyses

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).

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.
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.
Chapter 21. Comparing Analyses

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.

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

This creates a copy of the fit window.
Move the two fit windows side by side.

Choose **Edit:Windows:Freeze** in the fit window on the left.

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.
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 21.12. Two Windows, One Frozen

Figure 21.13. Edit: Observations Menu
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.

![Figure 21.15. Fit Window](image-url)
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.
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.

Figure 21.17. Comparing Two Models

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 (\( \text{DRILTI} \)). 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.