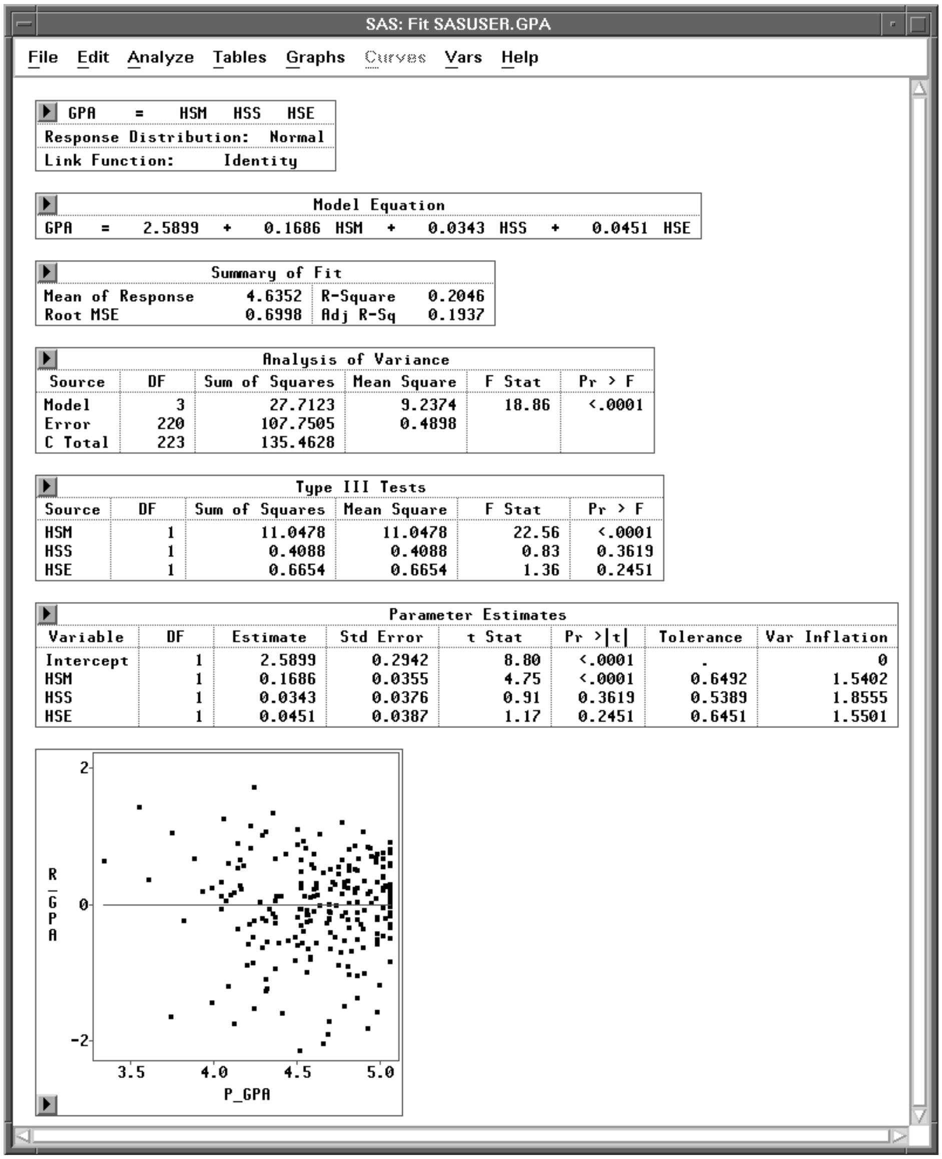
Chapter 14
Multiple Regression

Chapter Table of Contents

Chapter 14
Multiple Regression

You can create multiple regression models quickly using the fit variables dialog. You can use diagnostic plots to assess the validity of the models and identify potential outliers and influential observations. You can save residuals and other output variables from your models for future analysis.



GPA = HSM HSS HSE	
Response Distribution:	Normal
Link Function:	Identity

Model Equation
GPA = 2.5899 + 0.1686 HSM + 0.0343 HSS + 0.0451 HSE

Summary of Fit			
Mean of Response	4.6352	R-Square	0.2046
Root MSE	0.6998	Adj R-Sq	0.1937

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	3	27.7123	9.2374	18.86	<.0001
Error	220	107.7505	0.4898		
C Total	223	135.4628			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
HSM	1	11.0478	11.0478	22.56	<.0001
HSS	1	0.4088	0.4088	0.83	0.3619
HSE	1	0.6654	0.6654	1.36	0.2451

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	2.5899	0.2942	8.80	<.0001	.	0
HSM	1	0.1686	0.0355	4.75	<.0001	0.6492	1.5402
HSS	1	0.0343	0.0376	0.91	0.3619	0.5389	1.8555
HSE	1	0.0451	0.0387	1.17	0.2451	0.6451	1.5501

Figure 14.1. Multiple Regression Analysis

Creating the Analysis

The **GPA** data set contains data collected to determine which applicants at a large midwestern university were likely to succeed in its computer science program. The variable **GPA** is the measure of success of students in the computer science program, and it is the response variable. A *response variable* measures the outcome to be explained or predicted.

Several other variables are also included in the study as possible explanatory variables or predictors of **GPA**. An *explanatory variable* may explain variation in the response variable. Explanatory variables for this example include average high school grades in mathematics (**HSM**), English (**HSE**), and science (**HSS**) (Moore and McCabe 1989).

To begin the regression analysis, follow these steps.

⟹ **Open the GPA data set.**

⟹ **Choose Analyze:Fit (Y X).**

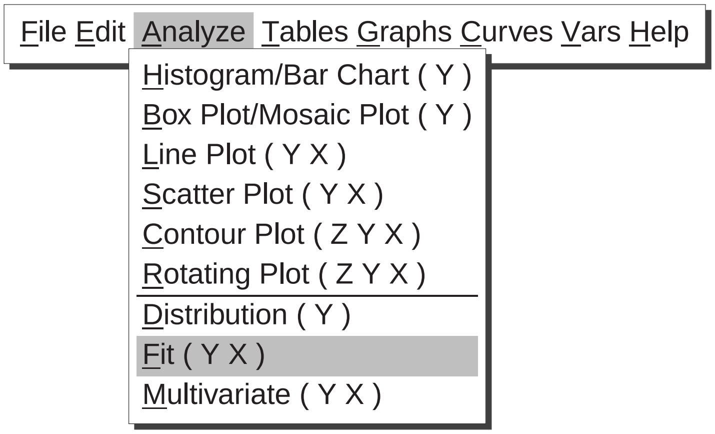


Figure 14.2. Analyze Menu

The fit variables dialog appears, as shown in Figure 14.3. This dialog differs from all other variables dialogs because it can remain visible even after you create the fit window. This makes it convenient to add and remove variables from the model. To make the variables dialog stay on the display, click on the **Apply** button when you are finished specifying the model. Each time you modify the model and use the **Apply** button, a new fit window appears so you can easily compare models. Clicking on **OK** also displays a new fit window but closes the dialog.

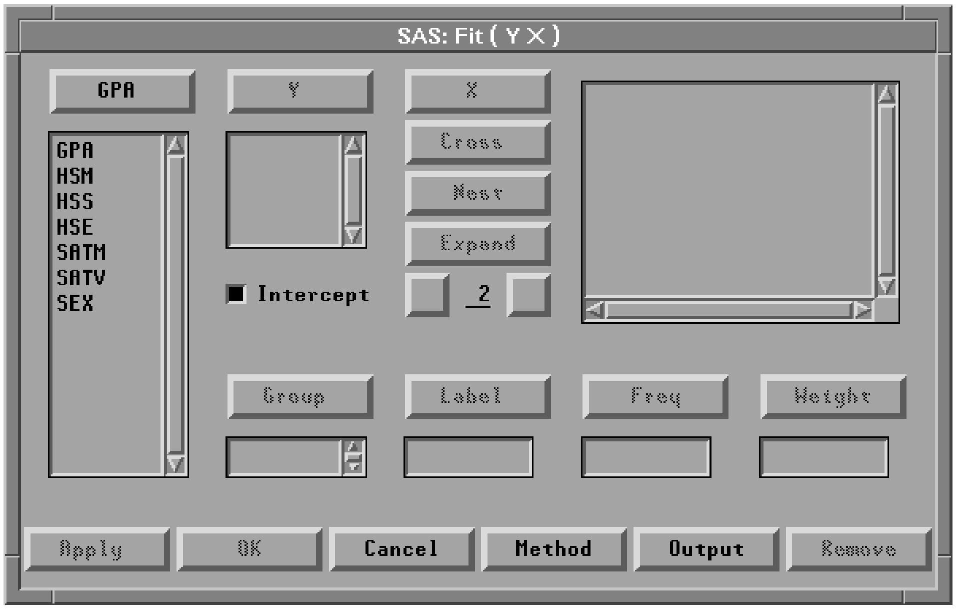


Figure 14.3. Fit Variables Dialog

⟹ **Select the variable GPA in the list on the left, then click the Y button.**
GPA appears in the **Y** variables list.

⟹ **Select the variables HSM, HSS, and HSE, then click the X button.**
HSM, **HSS**, and **HSE** appear in the **X** variables list.

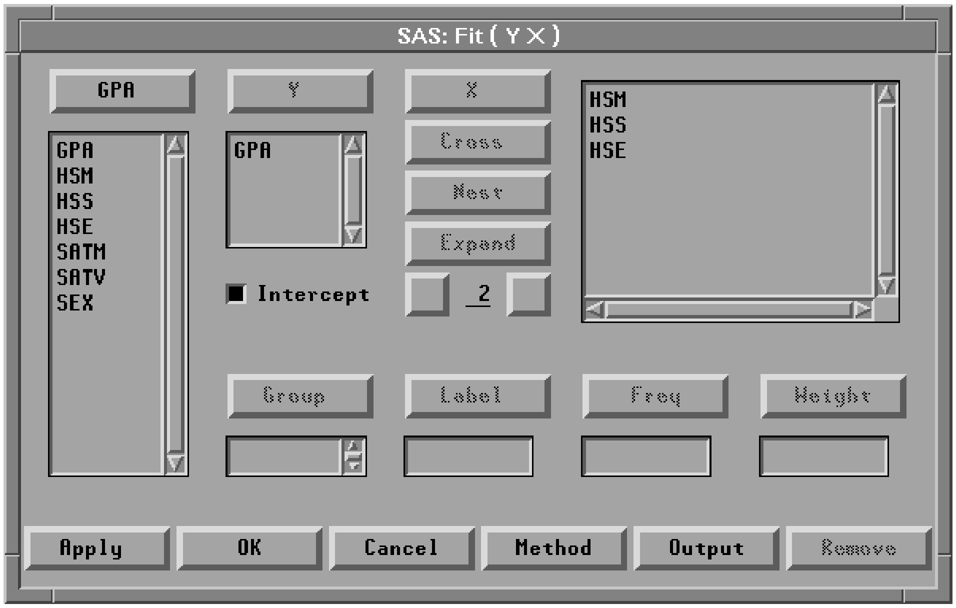
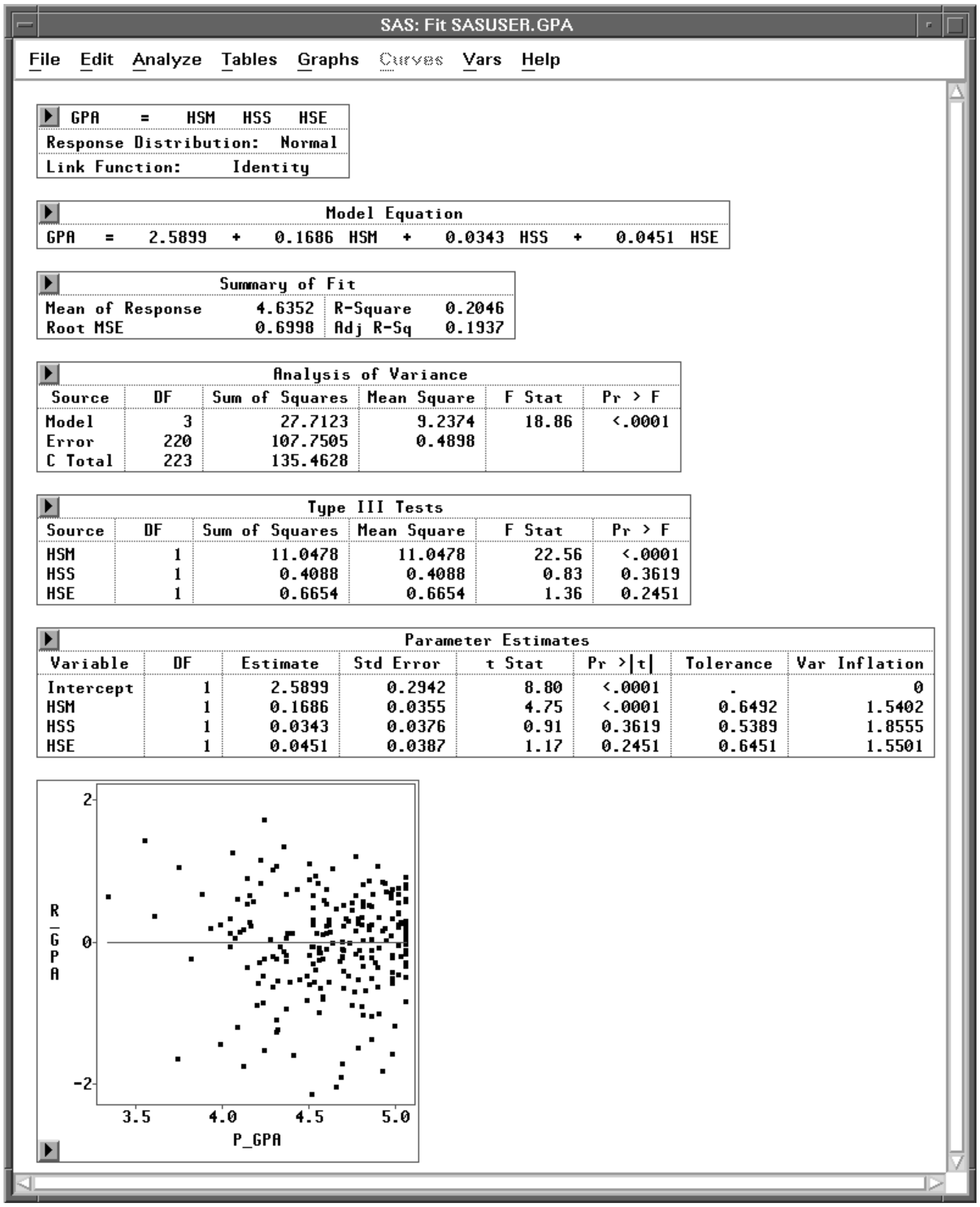


Figure 14.4. Variable Roles Assigned

⟹ **Click the Apply button.**
A fit window appears, as shown in Figure 14.5.



GPA = HSM HSS HSE	
Response Distribution:	Normal
Link Function:	Identity

Model Equation

GPA = 2.5899 + 0.1686 HSM + 0.0343 HSS + 0.0451 HSE

Summary of Fit

Mean of Response	4.6352	R-Square	0.2046
Root MSE	0.6998	Adj R-Sq	0.1937

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	3	27.7123	9.2374	18.86	<.0001
Error	220	107.7505	0.4898		
C Total	223	135.4628			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
HSM	1	11.0478	11.0478	22.56	<.0001
HSS	1	0.4088	0.4088	0.83	0.3619
HSE	1	0.6654	0.6654	1.36	0.2451

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	2.5899	0.2942	8.80	<.0001	.	0
HSM	1	0.1686	0.0355	4.75	<.0001	0.6492	1.5402
HSS	1	0.0343	0.0376	0.91	0.3619	0.5389	1.8555
HSE	1	0.0451	0.0387	1.17	0.2451	0.6451	1.5501

Figure 14.5. Fit Window

This window shows the results of a regression analysis of **GPA** on **HSM, HSS**, and **HSE**. The regression model for the *i*th observation can be written as

$$\mathsf{GPA}_i = \beta_0 + \beta_1 \mathsf{HSM}_i + \beta_2 \mathsf{HSS}_i + \beta_3 \mathsf{HSE}_i + \epsilon_i$$

where GPA_i is the value of GPA; β_0 to β_3 are the regression coefficients (parameters); $\mathsf{HSM}_i, \mathsf{HSS}_i$, and HSE_i are the values of the explanatory variables; and ϵ_i is the random error term. The ϵ_i's are assumed to be uncorrelated, with mean 0 and variance σ^2.

By default, the fit window displays tables for model information, **Model Equation**, **Summary of Fit**, **Analysis of Variance**, **Type III Tests**, and **Parameter Estimates**, and a residual-by-predicted plot, as illustrated in Figure 14.5. You can display other tables and graphs by clicking on the **Output** button on the fit variables dialog or by choosing menus as described in the section "Adding Tables and Graphs" later in this chapter.

Model Information

Model information is contained in the first two tables in the fit analysis. The first table displays the model specification, the response distribution, and the link function. The **Model Equation** table writes out the fitted model using the estimated regression coefficients $\hat{\beta}_0$ to $\hat{\beta}_3$:

$$\hat{\text{GPA}} = 2.5899 + 0.1686 \text{ HSM} + 0.0343 \text{ HSS} + 0.0451 \text{ HSE}$$

Summary of Fit

The **Summary of Fit** table contains summary statistics including **Root MSE** and **R-Square**. The **Root MSE** value is **0.6998** and is the square root of the mean square error given in the **Analysis of Variance** table. **Root MSE** is an estimate of σ in the preceding regression model.

The **R-Square** value is **0.2046**, which means that 20% of the variation in **GPA** scores is explained by the fitted model. The **Summary of Fit** table also contains an adjusted R-square value, **Adj R-Sq**. Because **Adj R-Sq** is adjusted for the number of parameters in the model, it is more comparable over models involving different numbers of parameters than **R-Square**.

Analysis of Variance

The **Analysis of Variance** table summarizes information about the sources of variation in the data. **Sum of Squares** represents variation present in the data. These values are calculated by summing squared deviations. In multiple regression, there are three sources of variation: **Model**, **Error**, and **C Total**. **C Total** is the total sum of squares corrected for the mean, and it is the sum of **Model** and **Error**. Degrees of Freedom, **DF**, are associated with each sum of squares and are related in the same way. **Mean Square** is the **Sum of Squares** divided by its associated **DF** (Moore and McCabe 1989).

If the data are normally distributed, the ratio of the **Mean Square** for the **Model** to the **Mean Square** for **Error** is an *F statistic*. This *F* statistic tests the null hypothesis that *none* of the explanatory variables has any effect (that is, that the regression coefficients β_1, β_2, and β_3 are all zero). In this case the computed *F* statistic (labeled **F Stat**) is 18.8606. You can use the *p*-value (labeled **Pr > F**) to determine whether to reject the null hypothesis. The *p*-value, also referred to as the *probability* value or *observed* significance level, is the probability of obtaining, by chance alone, an *F* statistic greater than the computed *F* statistic when the null hypothesis is true. The smaller the *p*-value, the stronger the evidence against the null hypothesis.

In this example, the *p*-value is so small that you can clearly reject the null hypothesis and conclude that at least one of the explanatory variables has an effect on **GPA**.

Type III Tests

The **Type III Tests** table presents the Type III sums of squares associated with the estimated coefficients in the model. Type III sums of squares are commonly called partial sums of squares (for a complete discussion, refer to the chapter titled "The Four Types of Estimable Functions" in the *SAS/STAT User's Guide*). The Type III sum of squares for a particular variable is the increase in the model sum of squares due to adding the variable to a model that already contains all the other variables in the model. Type III sums of squares, therefore, do not depend on the order in which the explanatory variables are specified in the model. Furthermore, they do not yield an additive partitioning of the **Model** sum of squares unless the explanatory variables are uncorrelated (which they are not for this example).

F tests are formed from this table as explained previously in the "Analysis of Variance" section. Note that when **DF = 1**, the Type III *F* statistic for a given parameter estimate is equal to the square of the *t* statistic for the same parameter estimate. For example, the **T Stat** value for **HSM** given in the **Parameter Estimates** table is **4.7494**. The corresponding **F Stat** value in the **Type III Tests** table is **22.5569**, which is **4.7494** squared.

Parameter Estimates

The **Parameter Estimates** table, as shown in Figure 14.5, displays the parameter estimates and the corresponding degrees of freedom, standard deviation, *t* statistic, and *p*-values. Using the parameter estimates, you can also write out the fitted model:

$$\mathrm{G\hat{P}A} = 2.5899 + 0.1686\mathrm{HSM} + 0.0343\mathrm{HSS} + 0.0451\mathrm{HSE}.$$

The *t* statistic is used to test the null hypothesis that a parameter is 0 in the model. In this example, only the coefficient for **HSM** appears to be statistically significant ($p \leq 0.0001$). The coefficients for **HSS** and **HSE** are not significant, partly because of the relatively high correlations among the three explanatory variables. Once **HSM** is included in the model, adding **HSS** and **HSE** does not substantially improve the model fit. Thus, their corresponding parameters are not statistically significant.

Two other statistics, tolerance and variance inflation, also appear in the **Parameter Estimates** table. These measure the strength of interrelationships among the explanatory variables in the model. Tolerances close to 0 and large variance inflation factor values indicate strong linear association or collinearity among the explanatory variables (Rawlings 1988, p. 277). For the **GPA** data, these statistics signal no problems of collinearity, even for **HSE** and **HSS**, which are the two most highly correlated variables in the data set.

Residuals-by-Predicted Plot

SAS/INSIGHT software provides many diagnostic tools to help you decide if your regression model fits well. These tools are based on the *residuals* from the fitted model. The residual for the *i*th observation is the observed value minus the predicted value:

$$\mathrm{GPA}_i - \mathrm{G\hat{P}A}_i .$$

The plot of the residuals versus the predicted values is a classical diagnostic tool used in regression analysis. The plot is useful for discovering poorly specified models or heterogeneity of variance (Myers 1986, pp. 138–139). The plot of **R_GPA** versus **P_GPA** in Figure 14.5 indicates no such problems. The observations are randomly scattered above and below the zero line, and no observations appear to be outliers.

Adding Tables and Graphs

The menus at the top of the fit window enable you to add tables and graphs to the fit window and output variables to the data window. When there is only one **X** variable, you can also fit curves as described in Chapter 13, “Fitting Curves.”

Following are some examples of tables and graphs you can add to a fit window.

Collinearity Diagnostics Table

⟹ **Choose Tables:Collinearity Diagnostics.**

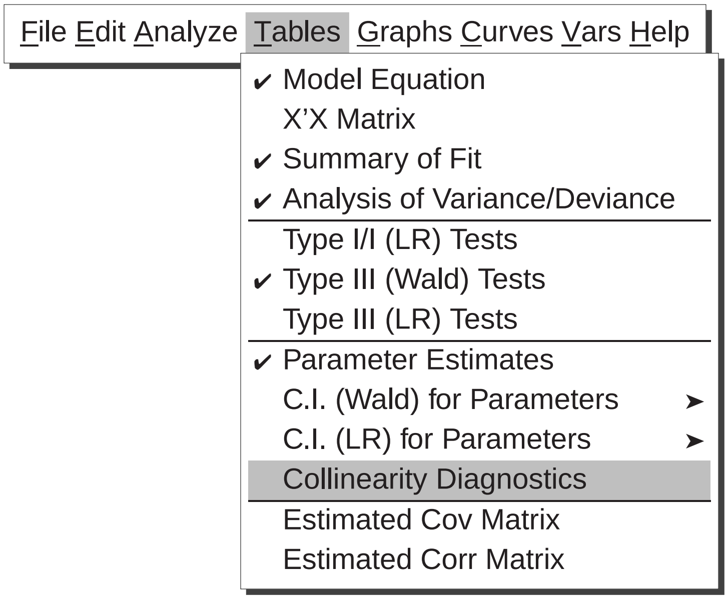
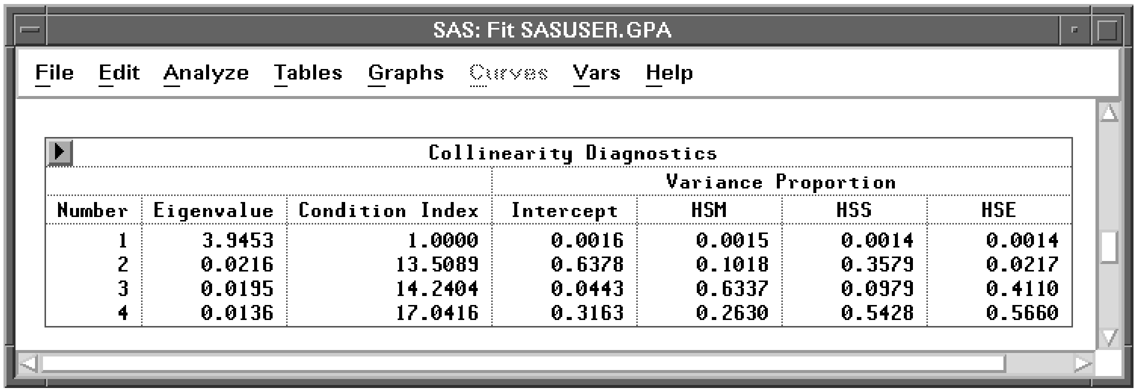


Figure 14.6. Tables Menu

This displays the table shown in Figure 14.7.



Collinearity Diagnostics

			Variance Proportion			
Number	Eigenvalue	Condition Index	Intercept	HSM	HSS	HSE
1	3.9453	1.0000	0.0016	0.0015	0.0014	0.0014
2	0.0216	13.5089	0.6378	0.1018	0.3579	0.0217
3	0.0195	14.2404	0.0443	0.6337	0.0979	0.4110
4	0.0136	17.0416	0.3163	0.2630	0.5428	0.5660

Figure 14.7. Collinearity Diagnostics Table

When an explanatory variable is nearly a linear combination of other explanatory variables in the model, the estimates of the coefficients in the regression model are unstable and have high standard errors. This problem is called *collinearity*. The **Collinearity Diagnostics** table is calculated using the eigenstructure of the $X'X$ matrix. See Chapter 13, "Fitting Curves," for a complete explanation.

A collinearity problem exists when a component associated with a high condition index contributes strongly to the variance of two or more variables. The highest condition number in this table is **17.0416**. Belsley, Kuh, and Welsch (1980) propose that a condition index of 30 to 100 indicates moderate to strong collinearity.

Partial Leverage Plots

Another diagnostic tool available in the fit window is partial leverage plots. When there is more than one explanatory variable in a model, the relationship of the residuals to one explanatory variable can be obscured by the effects of other explanatory variables. Partial leverage plots attempt to reveal these relationships (Rawlings 1988, pp. 265–266).

$\Longrightarrow$ **Choose Graphs:Partial Leverage.**

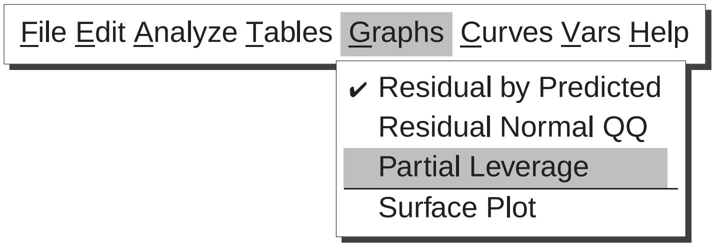


Figure 14.8. Graphs Menu

This displays the graphs shown in Figure 14.9.

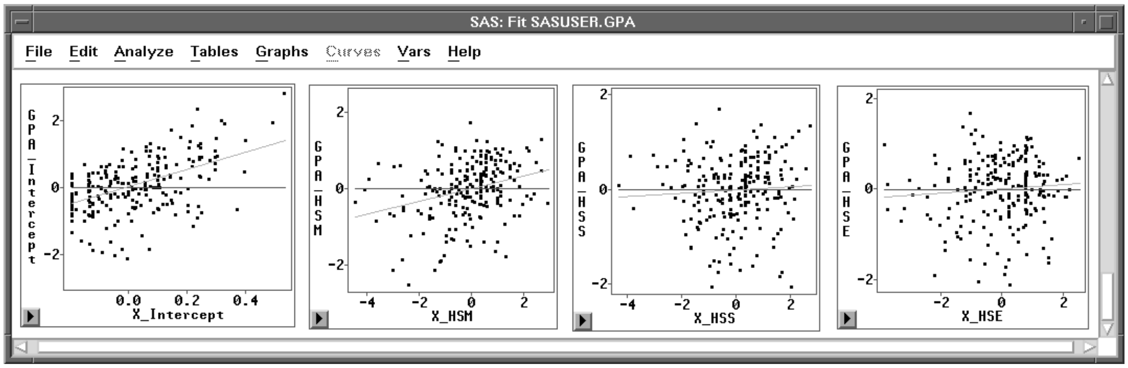


Figure 14.9. Partial Leverage Plots

In each plot in Figure 14.9, the x-axis represents the residuals of the explanatory variable from a model that regresses that explanatory variable on the remaining explanatory variables. The y-axis represents the residuals of the response variable calculated with the explanatory variable omitted.

Two reference lines appear in each plot. One is the horizontal line Y=0, and the other is the fitted regression line with slope equal to the parameter estimate of the corresponding explanatory variable from the original regression model. The latter line shows the effect of the variable when it is added to the model last. An explanatory variable having little or no effect results in a line close to the horizontal line Y=0.

Examine the slopes of the lines in the partial leverage plots. The slopes for the plots representing **HSS** and **HSE** are nearly 0. This is not surprising since the coefficients for the parameter estimates of these two explanatory variables are nearly 0. You will examine the effect of removing these two variables from the model in the section “Modifying the Model” later in this chapter.

Curvilinear relationships not already included in the model may also be evident in a partial leverage plot (Rawlings 1988). No curvilinearity is evident in any of these plots.

Residual-by-Hat Diagonal Plot

The fit window contains additional diagnostic tools for examining the effect of observations. One such tool is the residual-by-hat diagonal plot. *Hat diagonal* refers to the diagonal elements of the hat matrix (Rawlings 1988). Hat diagonal measures the leverage of each observation on the predicted value for that observation.

Choosing **Fit (Y X)** does not automatically generate the residual-by-hat diagonal plot, but you can easily add it to the fit window. First, add the hat diagonal variable to the data window.

⟹ **Choose Vars:Hat Diag.**

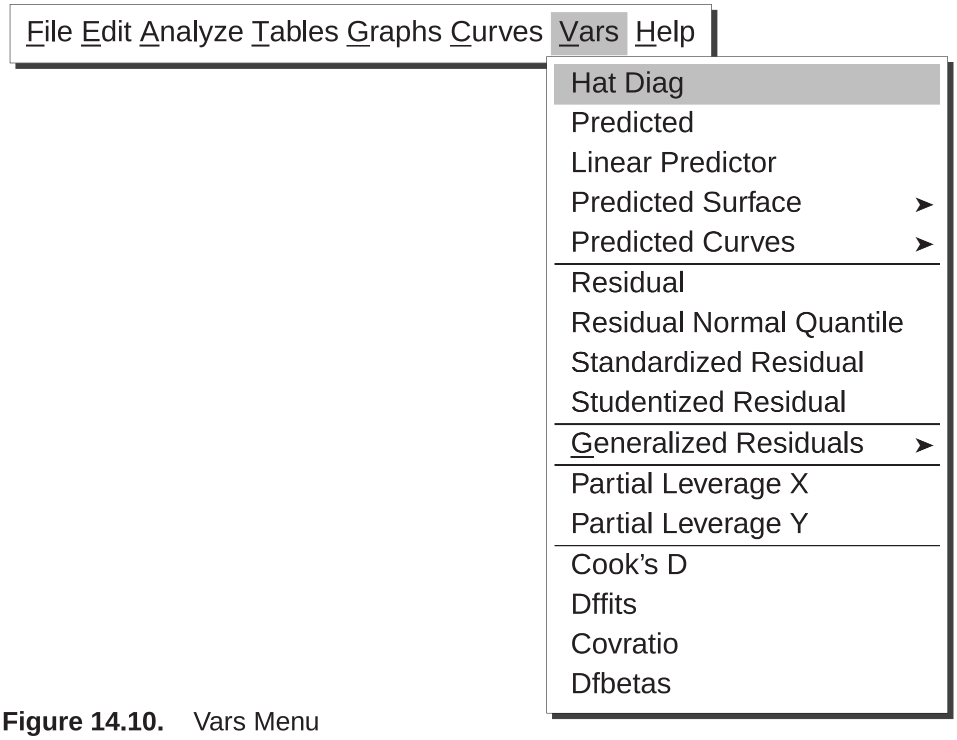
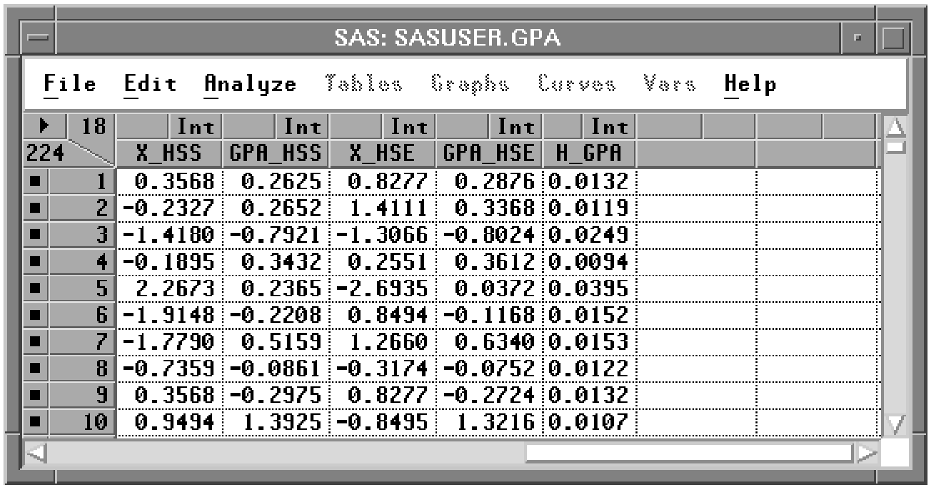


Figure 14.10. Vars Menu

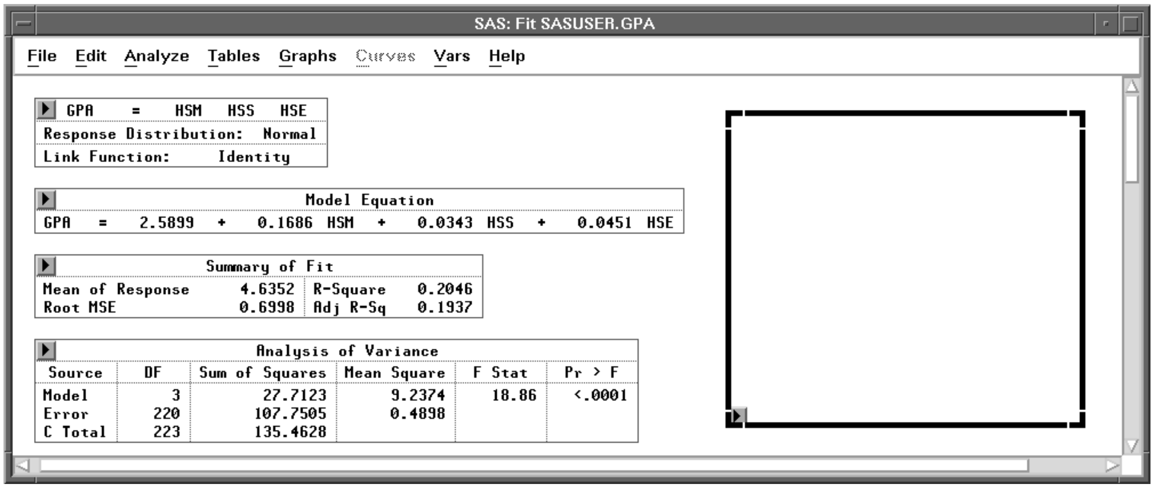
This adds the variable **H_GPA** to the data window, as shown in Figure 14.11. (The residual variable, **R_GPA**, is added when a residual-by-predicted plot is created.)



18	Int	Int	Int	Int	Int
224	X_HSS	GPA_HSS	X_HSE	GPA_HSE	H_GPA
1	0.3568	0.2625	0.8277	0.2876	0.0132
2	-0.2327	0.2652	1.4111	0.3368	0.0119
3	-1.4180	-0.7921	-1.3066	-0.8024	0.0249
4	-0.1895	0.3432	0.2551	0.3612	0.0094
5	2.2673	0.2365	-2.6935	0.0372	0.0395
6	-1.9148	-0.2208	0.8494	-0.1168	0.0152
7	-1.7790	0.5159	1.2660	0.6340	0.0153
8	-0.7359	-0.0861	-0.3174	-0.0752	0.0122
9	0.3568	-0.2975	0.8277	-0.2724	0.0132
10	0.9494	1.3925	-0.8495	1.3216	0.0107

Figure 14.11. GPA Data Window with H_GPA Added

⟹ **Drag a rectangle in the fit window to select an area for the new plot.**



```
GPA    =    HSM   HSS   HSE
Response Distribution:  Normal
Link Function:      Identity

                        Model Equation
GPA   =    2.5899   +    0.1686  HSM   +    0.0343  HSS   +    0.0451  HSE

             Summary of Fit
Mean of Response      4.6352  R-Square    0.2046
Root MSE              0.6998  Adj R-Sq    0.1937

                      Analysis of Variance
Source      DF   Sum of Squares  Mean Square   F Stat    Pr > F
Model        3          27.7123       9.2374    18.86    <.0001
Error      220         107.7505       0.4898
C Total    223         135.4628
```

Figure 14.12. Selecting an Area

⟹ **Choose Analyze:Scatter Plot (Y X).**

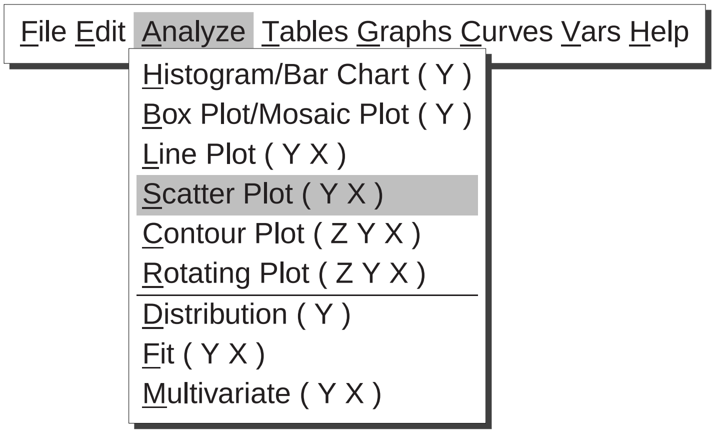


Figure 14.13. Analyze Menu

This displays the scatter plot variables dialog.

⟹ **Assign R_GPA the Y role and H_GPA the X role, then click on OK.**

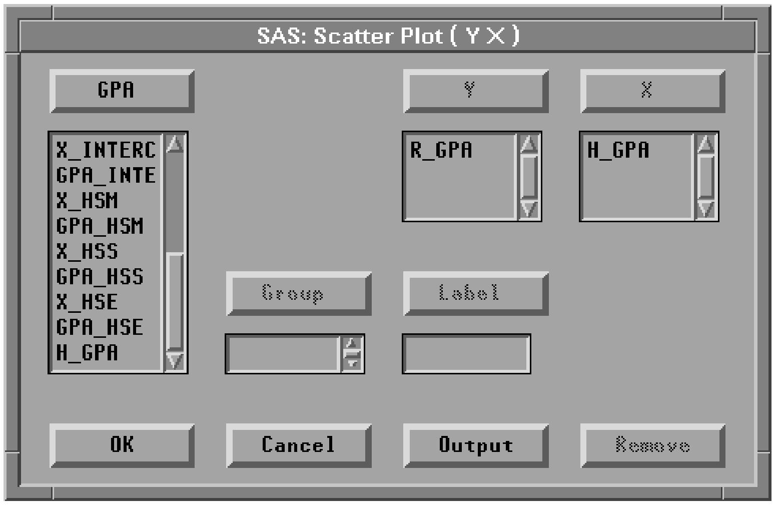
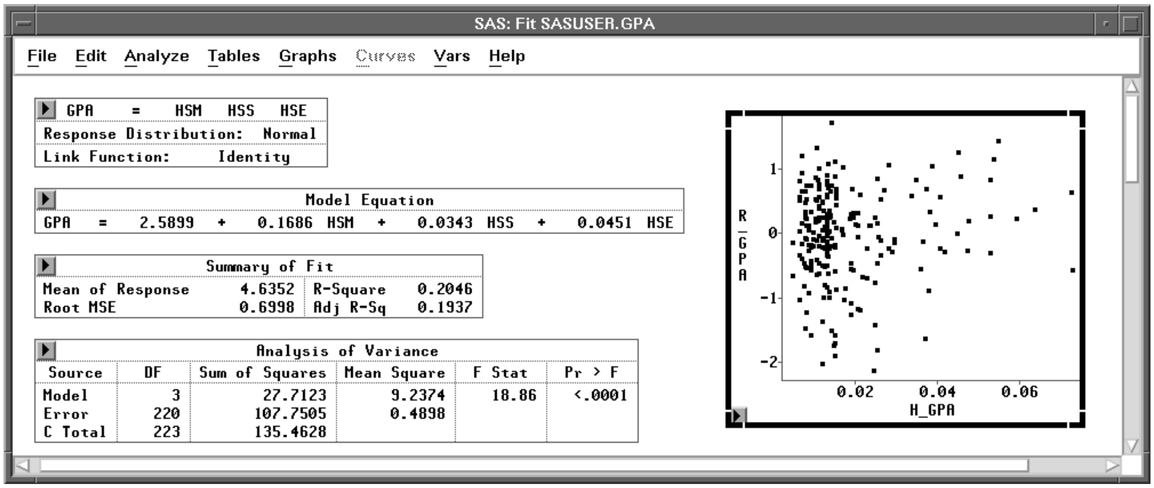


Figure 14.14. Scatter Plot Variables Dialog

The plot appears in the fit window in the area you selected.



```
GPA    =    HSM   HSS   HSE
Response Distribution:  Normal
Link Function:      Identity
```

```
                          Model Equation
GPA   =    2.5899   +    0.1686  HSM   +    0.0343  HSS   +    0.0451  HSE
```

Summary of Fit			
Mean of Response	4.6352	R-Square	0.2046
Root MSE	0.6998	Adj R-Sq	0.1937

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	3	27.7123	9.2374	18.86	<.0001
Error	220	107.7505	0.4898		
C Total	223	135.4628			

Figure 14.15. Residual by Hat Diagonal Plot

Belsley, Kuh, and Welsch (1980) propose a cutoff of $2p/n$ for the hat diagonal values, where *n* is the number of observations used to fit the model and *p* is the number of parameters in the model. Observations with values above this cutoff should be investigated. For this example, **H_GPA** values over 0.036 should be investigated. About 15% of the observations have values above this cutoff.

There are other measures you can use to determine the influence of observations. These include Cook's D, Dffits, Covratio, and Dfbetas. Each of these measures examines some effect of deleting the *i*th observation.

⟹ **Choose Vars:Dffits.**
A new variable, **F_GPA**, that contains the Dffits values is added to the data window.

Large absolute values of Dffits indicate influential observations. A general cutoff to consider is 2. It is, thus, useful in this example to identify those observations where **H_GPA** exceeds 0.036 and the absolute value of **F_GPA** is greater than 2. One way to accomplish this is by examining the **H_GPA** by **F_GPA** scatter plot.

⟹ **Choose Analyze:Scatter Plot (Y X).**
This displays the scatter plot variables dialog.

⟹ **Assign H_GPA the Y role and F_GPA the X role, then click on OK.**
This displays the **H_GPA** by **F_GPA** scatter plot.

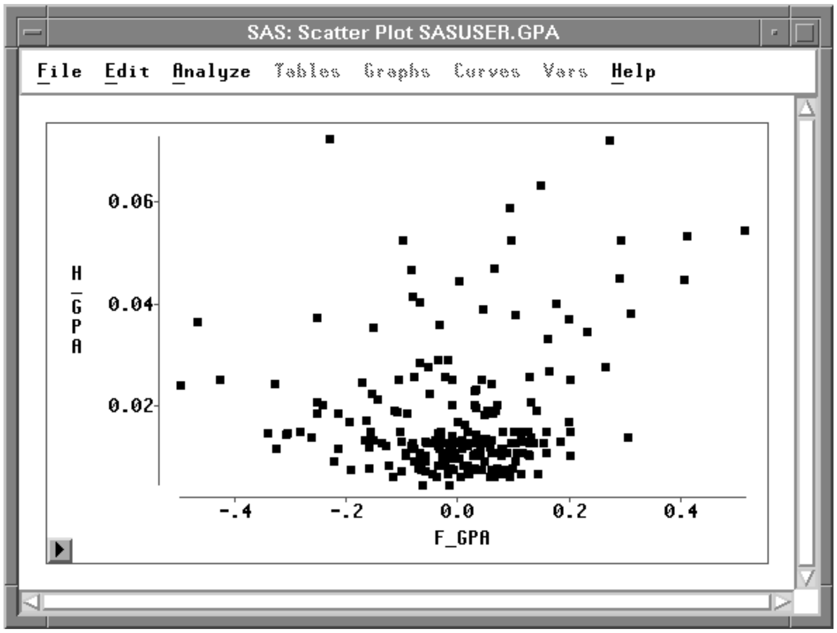


Figure 14.16. **H_GPA** by **F_GPA** Scatter Plot

None of the observations identified as potential influential observations (**H_GPA > 0.036**) are, in fact, influential for this model using the criterion $|F\_GPA| > 2$.

Modifying the Model

It may be possible to simplify the model without losing explanatory power. The change in the adjusted R-square value is one indicator of whether you are losing explanatory power by removing a variable. The estimate for **HSS** has the largest *p*-value, **0.3619**. Remove **HSS** from the model and see what effect this has on the adjusted R-square value.

From the fit variables dialog, follow these steps to request a new model with **HSS** removed. Remember, if you click **Apply** in the variables dialog, the dialog stays on the display so you can easily modify the regression model. You may need to rearrange the windows on your display if the fit variables dialog is not visible.

⟹ **Select HSS in the X variables list, then click the Remove button.**
This removes **HSS** from the model.

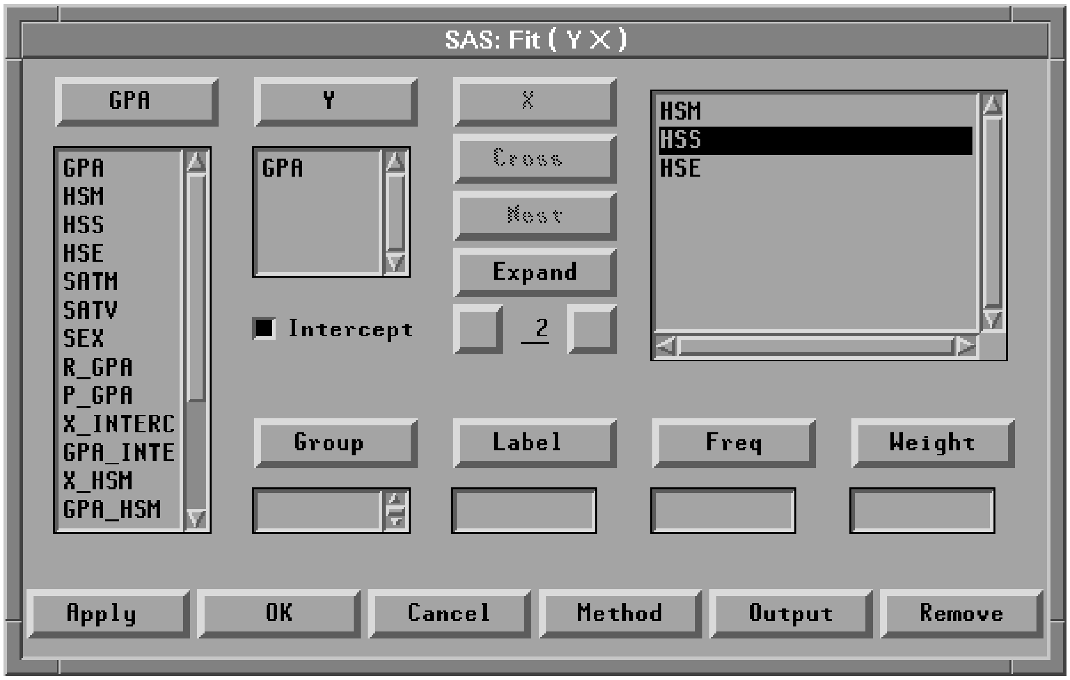
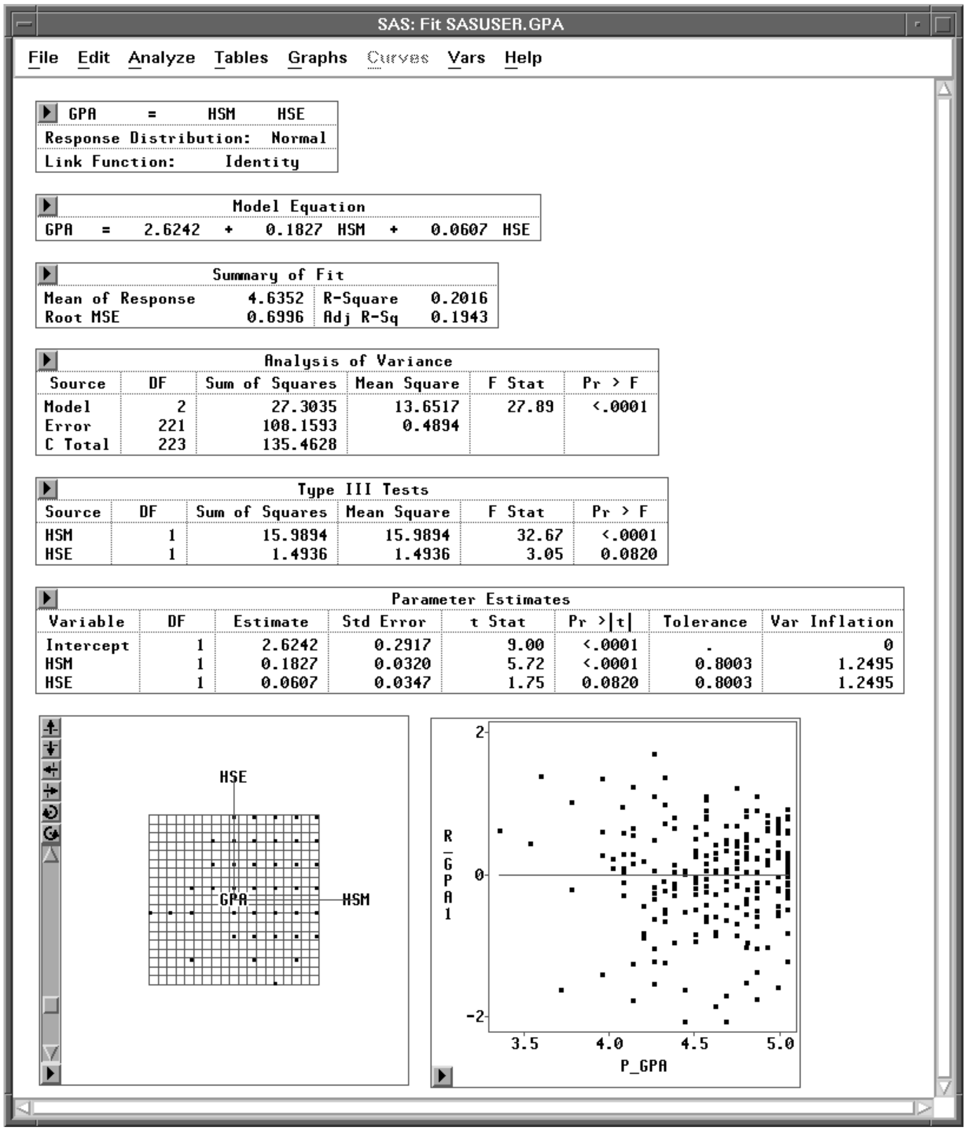


Figure 14.17. Removing the Variable **HSS**

⟹ **Click the Apply button.**
A new fit window appears, as shown in Figure 14.18.



GPA	=	HSM	HSE
Response Distribution:		Normal	
Link Function:		Identity	

Model Equation

GPA = 2.6242 + 0.1827 HSM + 0.0607 HSE

Summary of Fit			
Mean of Response	4.6352	R-Square	0.2016
Root MSE	0.6996	Adj R-Sq	0.1943

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	2	27.3035	13.6517	27.89	<.0001
Error	221	108.1593	0.4894		
C Total	223	135.4628			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
HSM	1	15.9894	15.9894	32.67	<.0001
HSE	1	1.4936	1.4936	3.05	0.0820

Parameter Estimates

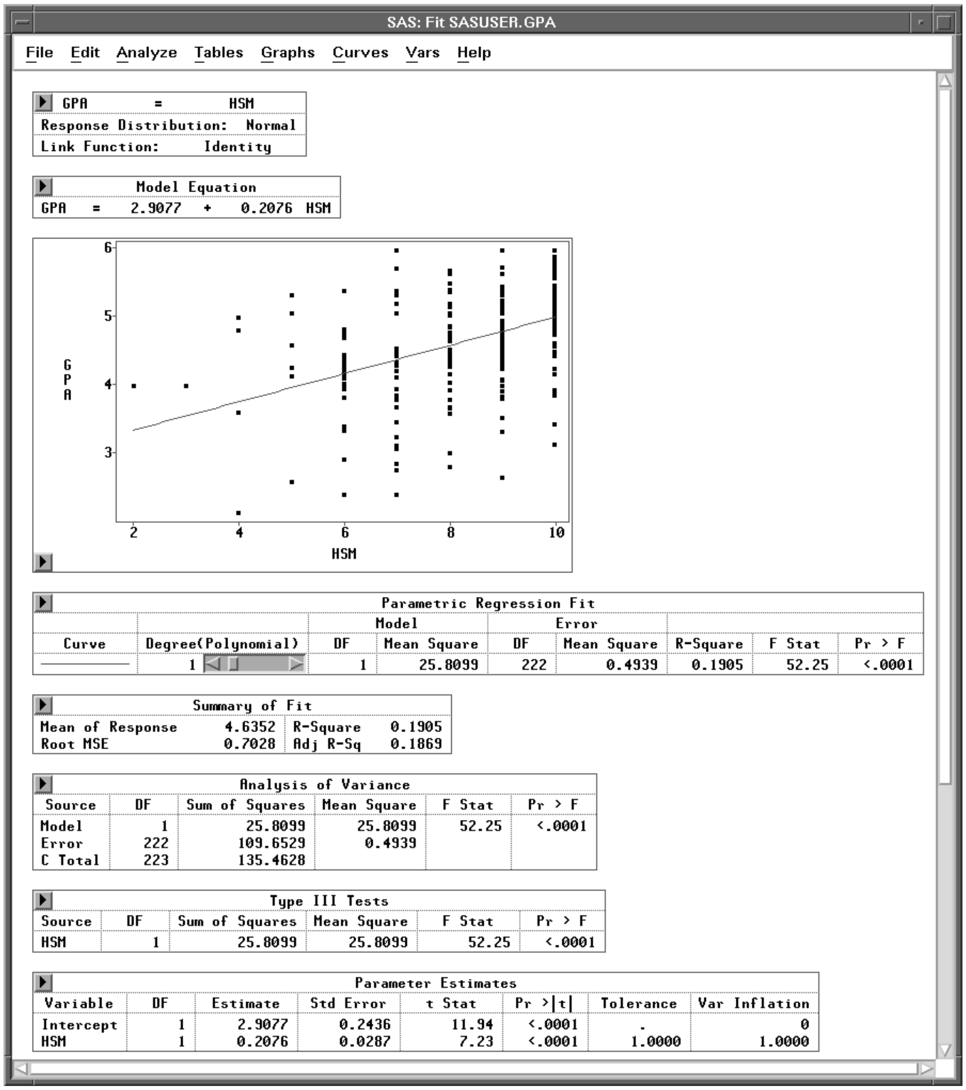
Variable	DF	Estimate	Std Error	t Stat	Pr > |t|	Tolerance	Var Inflation
Intercept	1	2.6242	0.2917	9.00	<.0001	.	0
HSM	1	0.1827	0.0320	5.72	<.0001	0.8003	1.2495
HSE	1	0.0607	0.0347	1.75	0.0820	0.8003	1.2495

Figure 14.18. Fit Window with HSM and HSE as Explanatory Variables

Reposition the two fit windows so you can compare the two models. Notice that the adjusted R-square value has actually increased slightly from 0.1937 to 0.1943. Little explanatory power is lost by removing **HSS**. Notice that within this model the *p*-value for **HSE** is a modest 0.0820. You can remove **HSE** from the new fit window without creating a third fit window.

⟹ **Select HSE in the second fit window.**

⟹ **Choose Edit:Delete in the second fit window.**
This recomputes the second fit using only **HSM** as an explanatory variable.



GPA	=	HSM
Response Distribution:		Normal
Link Function:		Identity

Model Equation
GPA = 2.9077 + 0.2076 HSM

Parametric Regression Fit

		Model		Error				
Curve	Degree(Polynomial)	DF	Mean Square	DF	Mean Square	R-Square	F Stat	Pr > F
————	1	1	25.8099	222	0.4939	0.1905	52.25	<.0001

Summary of Fit

Mean of Response	4.6352	R-Square	0.1905
Root MSE	0.7028	Adj R-Sq	0.1869

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	1	25.8099	25.8099	52.25	<.0001
Error	222	109.6529	0.4939		
C Total	223	135.4628			

Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
HSM	1	25.8099	25.8099	52.25	<.0001

Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	2.9077	0.2436	11.94	<.0001	.	0
HSM	1	0.2076	0.0287	7.23	<.0001	1.0000	1.0000

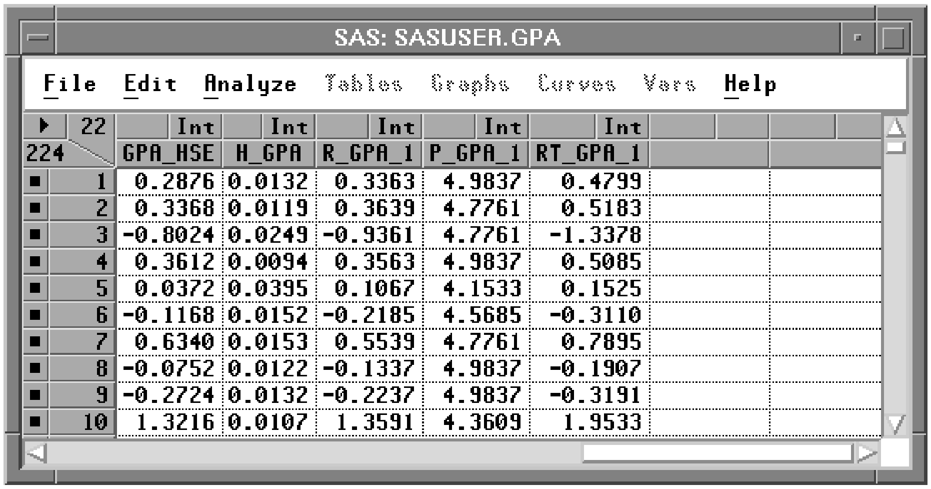
Figure 14.19. Fit Window with HSM as Explanatory Variable

The adjusted R-square value drops only slightly to **0.1869**. Removing **HSE** from the model also appears to have little effect. So, of the three explanatory variables you considered, only **HSM** appears to have strong explanatory power.

Saving the Residuals

One of the assumptions made in carrying out hypothesis tests in regression analysis is that the errors are normally distributed (Myers 1986). You can use residuals to check assumptions about errors. For this example, the *studentized* residuals are used because they are somewhat better than ordinary residuals for assessing normality, especially in the presence of outliers (Weisberg 1985). You can create a distribution window to check the normality of the residuals, as described in Chapter 12, "Examining Distributions."

⟹ **Choose Vars:Studentized Residual.**
A variable called **RT_GPA_1** is placed in the data window, as shown in Figure 14.20.



▸ 22	Int	Int	Int	Int	Int
224	GPA_HSE	H_GPA	R_GPA_1	P_GPA_1	RT_GPA_1
■ 1	0.2876	0.0132	0.3363	4.9837	0.4799
■ 2	0.3368	0.0119	0.3639	4.7761	0.5183
■ 3	-0.8024	0.0249	-0.9361	4.7761	-1.3378
■ 4	0.3612	0.0094	0.3563	4.9837	0.5085
■ 5	0.0372	0.0395	0.1067	4.1533	0.1525
■ 6	-0.1168	0.0152	-0.2185	4.5685	-0.3110
■ 7	0.6340	0.0153	0.5539	4.7761	0.7895
■ 8	-0.0752	0.0122	-0.1337	4.9837	-0.1907
■ 9	-0.2724	0.0132	-0.2237	4.9837	-0.3191
■ 10	1.3216	0.0107	1.3591	4.3609	1.9533

Figure 14.20. GPA Data Window with RT_GPA_1 Added

Notice the names of the last three variables. The number you see at the end of the variable names corresponds to the number of the fit window that generated the variables. See Chapter 39, "Fit Analyses," for detailed information about how generated variables are named.

⊕ **Related Reading:** Linear Models, Residuals, Chapter 39.

References

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), *Regression Diagnostics*, New York: John Wiley and Sons, Inc.

Freedman, D., Pisani, R., and Purves, R. (1978), *Statistics*, New York: W.W. Norton & Company, Inc.

Moore, D.S. and McCabe, G.P. (1989), *Introduction to the Practice of Statistics*, New York: W.H. Freeman and Company.

Myers, R.H. (1986), *Classical and Modern Regression with Applications*, Boston, MA: Duxbury Press.

Rawlings, J.O. (1988), *Applied Regression Analysis: A Research Tool*, Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software.

Weisberg, S. (1985), *Applied Linear Regression, Second Edition*, New York: John Wiley and Sons, Inc.

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

ISBN 1–58025–490–X

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

1st printing, October 1999

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