Chapter 39
Fit Analyses

Chapter Table of Contents

Chapter 39
Fit Analyses

Choosing **Analyze:Fit (Y X)** gives you access to a variety of techniques for fitting models to data. These provide methods for examining the relationship between a response (dependent) variable and a set of explanatory (independent) variables.

You can use least-squares methods for simple and multiple linear regression with various diagnostic capabilities when the response is normally distributed.

You can use generalized linear models to analyze the data when the response is from a distribution of the exponential family and a function can be used to link the response mean to a linear combination of the explanatory variables.

You can use spline and kernel smoothers for nonparametric regression when the model has one or two explanatory variables.

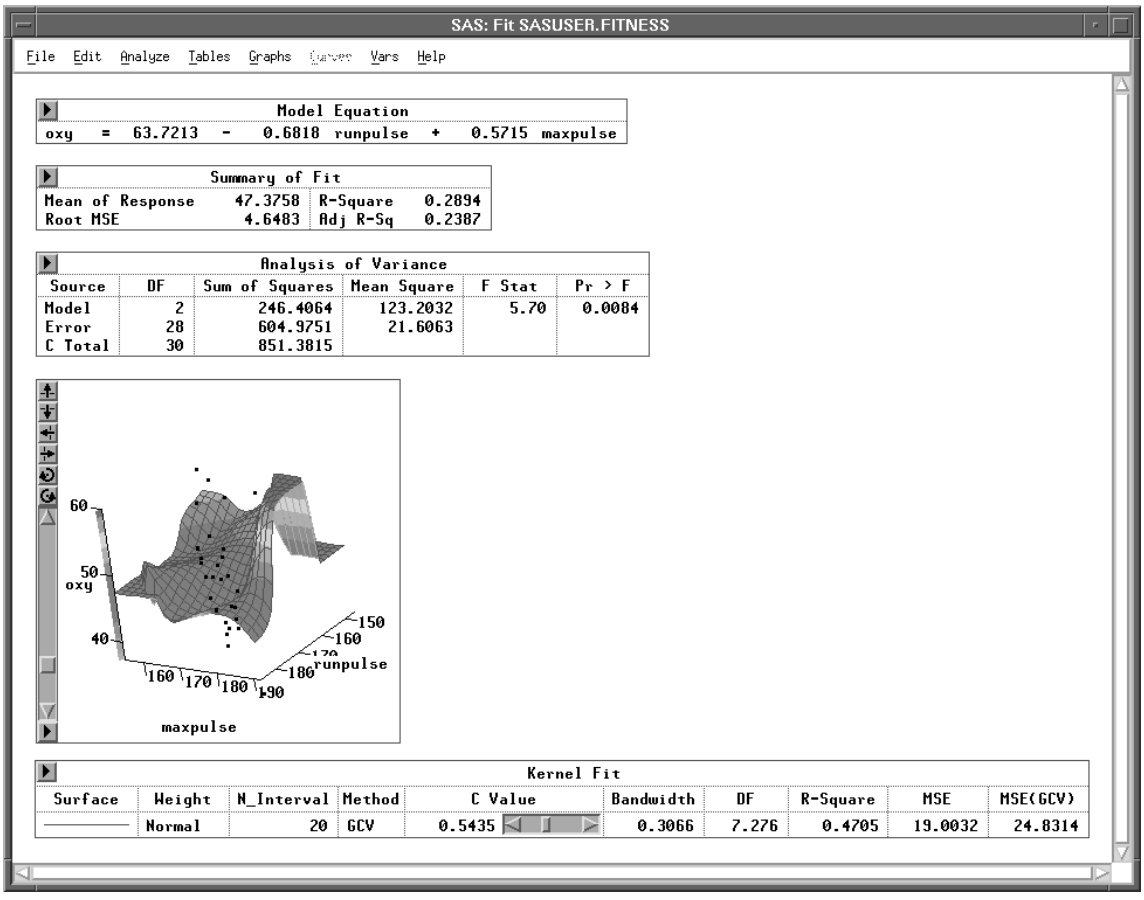


Figure 39.1. Fit Analysis

Statistical Models

The relationship between a response variable and a set of explanatory variables can be studied through a regression model

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

where y_i is the ith observed response value, $\mathbf{x}_i$ is the ith vector of explanatory values, and ε_i's are uncorrelated random variables with zero mean and a common variance.

If the form of the regression function f is known except for certain parameters, the model is called a *parametric regression model*. Furthermore, if the regression function is linear in the unknown parameters, the model is called a *linear model*.

In the case of linear models with the error term ε_i assumed to be normally distributed, you can use classical linear models to explore the relationship between the response variable and the explanatory variables.

A *nonparametric model* generally assumes only that f belongs to some infinite- dimensional collection of functions. For example, f may be assumed to be differentiable with a square-integrable second derivative.

When there is only one explanatory X variable, you can use nonparametric smoothing methods, such as smoothing splines, kernel estimators, and local polynomial smoothers. You can also request confidence ellipses and parametric fits (mean, linear regression, and polynomial curves) with a linear model. These are added to a scatter plot generated from **Y** by a single **X** and are described in the "Fit Curves" section.

When there are two explanatory variables in the model, you can create parametric and nonparametric (kernel and thin-plate smoothing spline) response surface plots. With more than two explanatory variables in the model, a parametric profile response surface plot with two selected explanatory variables can be created.

When the response y_i has a distribution from the exponential family (normal, inverse Gaussian, gamma, Poisson, binomial), and the mean μ_i of the response variable y_i is assumed to be related to a linear predictor through a monotone function g

$$g(\mu_i) = \mathbf{x}_i'\beta$$

where β is a vector of unknown parameters, you can explore the relationship by using generalized linear models.

Linear Models

SAS/INSIGHT fit analysis provides the traditional parametric regression analysis assuming that the regression function is linear in the unknown parameters. The relationship is expressed as an equation that predicts a response variable from a linear function of explanatory variables.

Besides the usual estimators and test statistics produced for a regression, a fit analysis can produce many diagnostic statistics. Collinearity diagnostics measure the strength of the linear relationship among explanatory variables and how this affects the stability of the estimates. Influence diagnostics measure how each individual observation contributes to determining the parameter estimates and the fitted values.

In matrix algebra notation, a linear model is written as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

where $\mathbf{y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ is the $n \times p$ design matrix (rows are observations and columns are explanatory variables), β is the $p \times 1$ vector of unknown parameters, and ϵ is the $n \times 1$ vector of unknown errors.

Each effect in the model generates one or more columns in a design matrix $\mathbf{X}$. The first column of $\mathbf{X}$ is usually a vector of 1's used to estimate the intercept term. In general, no-intercept models should be fit only when theoretical justification exists. Refer to the chapter on the GLM procedure in the *SAS/STAT User's Guide* for a description of the model parameterization.

The classical theory of linear models is based on some strict assumptions. Ideally, the response is measured with all the explanatory variables controlled in an experimentally determined environment. If the explanatory variables do not have experimentally fixed values but are stochastic, the conditional distribution of $\mathbf{y}$ given $\mathbf{X}$ must be normal in the appropriate form.

Less restrictive assumptions are as follows:

- The form of the model is correct (all important X variables have been included).
- Explanatory variables are measured without error.
- The expected value of the errors is 0.
- The variance of the errors (and thus the response variable) is constant across observations (denoted by σ^2).
- The errors are uncorrelated across observations.

If all the necessary assumptions are met, the least-squares estimates of β are the best linear unbiased estimates (BLUE); in other words, the estimates have minimum variance among the class of estimators that are unbiased and are linear functions of the responses. In addition, when the error term is assumed to be normally distributed, sampling distributions for the computed statistics can be derived. These sampling distributions form the basis for hypothesis tests on the parameters.

The method used to estimate the parameters is to minimize the sum of squares of the differences between the actual response values and the values predicted by the model. An estimator **b** for β is generated by solving the resulting normal equations

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}$$

yielding

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Let **H** be the projection matrix for the space spanned by **X**, sometimes called the hat matrix,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Then the predicted mean vector of the n observation responses is

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{H}\mathbf{y}$$

The sum of squares for error is

$$SSE = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i\mathbf{b})^2$$

where $\mathbf{x}_i$ is the ith row of the **X** matrix.

Assume that **X** is of full rank. The variance σ^2 of the error is estimated by the mean square error

$$s^2 = MSE = \frac{SSE}{n-p}$$

The parameter estimates are unbiased:

$$E(\mathbf{b}) = \beta$$

$$E(s^2) = \sigma^2.$$

The covariance matrix of the estimates is

$$\mathrm{Var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

The estimate of the covariance matrix, $\widehat{\mathrm{Var}}(\mathbf{b})$, is obtained by replacing σ^2 with its estimate, s^2, in the preceding formula:

$$\widehat{\mathrm{Var}}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}s^2$$

The correlations of the estimates,

$$\mathbf{S}^{-1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{S}^{-1/2}$$

are derived by scaling to one on the diagonal, where $\mathbf{S}$ = diag ($(\mathbf{X}'\mathbf{X})^{-1}$).

If the model is not full rank, the matrix $\mathbf{X}'\mathbf{X}$ is singular. A generalized (g2) inverse (Pringle and Raynor 1971), denoted as $(\mathbf{X}'\mathbf{X})^{-}$, is then used to solve the normal equations, as follows:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$$

However, this solution is not unique, and there are an infinite number of solutions using different generalized inverses. In SAS/INSIGHT software, the fit analysis chooses a basis of all variables that are linearly independent of previous variables and a zero solution for the remaining variables.

⊕ **Related Reading:** Multiple Regression, Chapter 14.

⊕ **Related Reading:** Analysis of Variance, Chapter 15.

Generalized Linear Models

Generalized linear models assume that the response y_i has a distribution from the exponential family (normal, inverse Gaussian, gamma, Poisson, binomial) and a function can be used to link the expected response mean and a linear function of the X effects. In SAS/INSIGHT software, a generalized linear model is written as

$$\mathbf{y} = \mu + \epsilon$$

$$\eta = g(\mu) = \eta_0 + \mathbf{X}\beta$$

where **y** is the $n \times 1$ vector of responses, μ is the $n \times 1$ expected response means, and ϵ is the $n \times 1$ vector of unknown errors.

The monotone function g links the response mean μ with a linear predictor η from the effects, and it is called the *link function*. The $n \times 1$ vector η_0 is the offset, **X** is the $n \times p$ design matrix, and β is the $p \times 1$ vector of unknown parameters. The design matrix is generated the same way as for linear models.

You specify the response distribution, the link function, and the offset variable in the fit method options dialog.

The Exponential Family of Distributions

The distribution of a random variable **Y** belongs to the exponential family if its probability (density) function can be written in the form

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

where θ is the natural or canonical parameter, ϕ is the dispersion parameter, and *a*, *b* and *c* are specific functions.

The mean and variance of **Y** are then given by (McCullagh and Nelder 1989)

$$E(y) = \mu = b'(\theta)$$

$$\mathrm{Var(y)} = \mathrm{a}(\phi)\mathrm{b}''(\theta)$$

The function $b''(\theta)$ can be expressed as a function of μ, $b''(\theta) = V(\mu)$, and it is called the *variance function*. Different choices of the function $b(\theta)$ generate different distributions in the exponential family. For a binomial distribution with *m* trials, the function $a(\phi) = \phi/m$. For other distributions in the exponential family, $a(\phi) = \phi$.

SAS/INSIGHT software includes normal, inverse Gaussian, gamma, Poisson, and binomial distributions for the response distribution. For these response distributions, the density functions $f(y)$, the variance functions $V(\mu)$, and the dispersion parameters ϕ with function $a(\phi)$ are

Normal

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) \quad \text{for } -\infty < y < \infty$$
$$V(\mu) = 1$$
$$a(\phi) = \phi = \sigma^2$$

Inverse Gaussian

$$f(y) = \frac{1}{\sqrt{2\pi y^3}\sigma} \exp\left(-\frac{1}{2\mu^2 y}\left(\frac{y-\mu}{\sigma}\right)^2\right) \quad \text{for } y > 0$$
$$V(\mu) = \mu^3$$
$$a(\phi) = \phi = \sigma^2$$

Gamma

$$f(y) = \frac{1}{y\Gamma(\nu)}\left(\frac{\nu y}{\mu}\right)^\nu \exp\left(-\frac{\nu y}{\mu}\right) \quad \text{for } y > 0$$
$$V(\mu) = \mu^2$$
$$a(\phi) = \phi = \nu^{-1}$$

Poisson

$$f(y) = \frac{\mu^y e^{-\mu}}{y!} \quad \text{for } y = 0, 1, 2, \ldots$$
$$V(\mu) = \mu$$
$$a(\phi) = \phi = 1$$

Binomial

$$f(y) = \binom{m}{r}\mu^r(1-\mu)^{m-r} \quad \text{for } y = r/m, r = 0, 1, 2, ..., m$$
$$V(\mu) = \mu(1-\mu)$$
$$a(\phi) = \phi/m = 1/m$$

Link Function

The link function links the response mean μ to the linear predictor η. SAS/INSIGHT software provides six types of link functions:

Identity $g(\mu) = \mu$

Log $g(\mu) = \log(\mu)$

Logit $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$

Probit $g(\mu) = \Phi^{-1}(\mu)$

Comp. Log-log $g(\mu) = \log(-\log(1-\mu))$

Power $g(\mu) = \mu^\lambda$ where λ is the value in the **Power** entry field.

For each response distribution in the exponential family, there exists a special link function, the canonical link, for which $\theta = \eta$. The canonical links expressed in terms of the mean parameter μ are

Normal	$g(\mu) = \mu$
Inverse Gaussian	$g(\mu) = \mu^{-2}$
Gamma	$g(\mu) = \mu^{-1}$
Poisson	$g(\mu) = \log(\mu)$
Binomial	$g(\mu) = \log(\frac{\mu}{1-\mu})$

† **Note:** Some links are not appropriate for all distributions. For example, logit, probit, and complementary log-log links are only appropriate for the binomial distribution.

The Likelihood Function and Maximum-Likelihood Estimation

The log-likelihood function

$$l(\theta, \phi; y) = \log f(y; \theta, \phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

can be expressed in terms of the mean μ and the dispersion parameter ϕ:

Normal	$l(\mu, \phi; y) = -\frac{1}{2}\log(\phi) - \frac{1}{2\phi}(y-\mu)^2 \quad \text{for } -\infty < y < \infty$
Inverse Gaussian	$l(\mu, \phi; y) = -\log(y^3\phi) - \frac{(y-\mu)^2}{2y\mu^2\phi} \quad \text{for } y > 0$
Gamma	$l(\mu, \phi; y) = -\log(y\Gamma(\frac{1}{\phi})) + \frac{1}{\phi}\log(\frac{y}{\mu\phi}) - \frac{y}{\mu\phi} \quad \text{for } y > 0$
Poisson	$l(\mu, \phi; y) = y\log(\mu) - \mu \quad \text{for } y = 0, 1, 2, \ldots$
Binomial	$l(\mu, \phi; y) = r\log(\mu) + (m-r)\log(1-\mu)$ for $y = r/m, r = 0, 1, 2, ..., m$

† **Note:** Some terms in the density function have been dropped in the log-likelihood function since they do not affect the estimation of the mean and scale parameters.

SAS/INSIGHT software uses a ridge stabilized Newton-Raphson algorithm to maximize the log-likelihood function *l*(μ , ϕ ; y) with respect to the regression parameters. On the *r*th iteration, the algorithm updates the parameter vector **b** by

$$\mathbf{b}_{(r)} = \mathbf{b}_{(r-1)} - \mathbf{H}^{-1}_{(r-1)}\mathbf{u}_{(r-1)}$$

where **H** is the Hessian matrix and **u** is the gradient vector, both evaluated at $\beta = \mathbf{b}_{(r-1)}$.

$$\mathbf{H} = (h_{jk}) = \left(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right)$$

$$\mathbf{u} = (u_j) = \left(\frac{\partial l}{\partial\beta_j}\right).$$

The Hessian matrix **H** can be expressed as

$$\mathbf{H} = -\mathbf{X}'\mathbf{W}_o\mathbf{X}$$

where **X** is the design matrix, $\mathbf{W}_o$ is a diagonal matrix with ith diagonal element

$$w_{oi} = w_{ei} + (y_i - \mu_i)\frac{V_i g_i'' + V_i' g_i'}{V_i^2 (g_i')^3 a_i(\phi)}$$

$$w_{ei} = E(w_{oi}) = \frac{1}{a_i(\phi) V_i (g_i')^2}$$

where g_i is the link function, V_i is the variance function, and the primes denote derivatives of g and V with respect to μ. All values are evaluated at the current mean estimate μ_i. $a_i(\phi) = \phi/w_i$, where w_i is the prior weight for the ith observation.

SAS/INSIGHT software uses either the full Hessian matrix **H** = - **X**' $\mathbf{W}_o$ **X** or the Fisher's scoring method in the maximum-likelihood estimation. In the Fisher's scoring method, $\mathbf{W}_o$ is replaced by its expected value $\mathbf{W}_e$ with ith element w_{ei}.

$$\mathbf{H} = \mathbf{X}'\mathbf{W}_e\mathbf{X}$$

The estimated variance-covariance matrix of the parameter estimates is

$$\hat{\mathbf{\Sigma}} = -\mathbf{H}^{-1}$$

where **H** is the Hessian matrix evaluated at the model parameter estimates.

The estimated correlation matrix of the parameter estimates is derived by scaling the estimated variance-covariance matrix to 1 on the diagonal.

† **Note:** A warning message appears when the specified model fails to converge. The output tables, graphs, and variables are based on the results from the last iteration.

Scale Parameter

A scale parameter is related to the dispersion parameter ϕ and is given by

Normal	$\sigma = \sqrt{\phi}$
Inverse Gaussian	$\sigma = \sqrt{\phi}$
Gamma	$\nu = 1/\phi$
Poisson	1
Binomial	1

The scale parameter is 1 for Poisson and binomial distributions. SAS/INSIGHT software provides different scale parameter estimates for normal, inverse Gaussian, and gamma distributions:

MLE	the maximum-likelihood estimate
Deviance	the mean deviance
Pearson	the mean Pearson χ^2
Constant	the value in the **Constant** entry field

When maximum-likelihood estimation is used, the Hessian **H** and the gradient **u** also include the term for the scale parameter.

† **Note:** You can request an exponential distribution for the response variable by specifying a gamma distribution with scale parameter set to 1.

Goodness of Fit

The log-likelihood can be expressed in terms of the mean parameter μ and the log-likelihood-ratio statistic is the scaled deviance

$$D^*(y;\hat{\mu}) = -2(l(\hat{\mu};y) - l(\hat{\mu}_{max};y))$$

where $l(\hat{\mu};y)$ is the log-likelihood under the model and $l(\hat{\mu}_{max};y)$ is the log-likelihood under the maximum achievable (saturated) model.

For generalized linear models, the scaled deviance can be expressed as

$$D^*(y;\hat{\mu}) = \frac{1}{\phi}D(y;\hat{\mu})$$

where $D(y;\hat{\mu})$ is the residual deviance for the model and is the sum of individual deviance contributions.

The forms of the individual deviance contributions, d_i, are

Normal	$(y-\hat{\mu})^2$
Inverse Gaussian	$(y-\hat{\mu})^2/(\hat{\mu}^2 y)$
Gamma	$-2\log(y/\hat{\mu})+2(y-\hat{\mu})/\hat{\mu}$
Poisson	$2y\log(y/\hat{\mu})-2(y-\hat{\mu})$
Binomial	$2(r\log(y/\hat{\mu})+(m-r)\log((1-y)/(1-\hat{\mu}))$ where *y=r/m*, *r* is the number of successes in *m* trials.

For a binomial distribution with m_i trials in the ith observation, the Pearson χ^2 statistic is

$$\chi^2=\sum_{i=1}^{n} m_i\frac{(y_i-\mu_i)^2}{V(\mu_i)}$$

For other distributions, the Pearson χ^2 statistic is

$$\chi^2=\sum_{i=1}^{n} \frac{(y_i-\mu_i)^2}{V(\mu_i)}$$

The scaled Pearson χ^2 statistic is χ^2 / ϕ. Either the mean deviance $D(y;\hat{\mu})/(n-p)$ or the mean Pearson χ^2 statistic $\chi^2/(n-p)$ can be used to estimate the dispersion parameter ϕ. The χ^2 approximation is usually quite accurate for the differences of deviances for nested models (McCullagh and Nelder 1989).

Quasi-Likelihood Functions

For binomial and Poisson distributions, the scale parameter has a value of 1. The variance of **Y** is $\mathrm{Var(y)}=\mu(1-\mu)/\mathrm{m}$ for the binomial distribution and $\mathrm{Var(y)}=\mu$ for the Poisson distribution. *Overdispersion* occurs when the variance of **Y** exceeds the $\mathrm{Var(y)}$ above. That is, the variance of **Y** is $\sigma^2 V(\mu)$, where σ>1.

With overdispersion, methods based on quasi-likelihood can be used to estimate the parameters β and σ. A quasi-likelihood function

$$Q(\mu;y)=\int_y^{\mu}\frac{y-t}{\sigma^2 V(t)}dt$$

is specified by its associated variance function.

SAS/INSIGHT software includes the quasi-likelihoods associated with the variance functions $V(\mu) = 1$, μ, μ^2, μ^3, and $\mu(1-\mu)$. The associated distributions (with the same variance function), the quasi-likelihoods $Q(\mu; y)$, the canonical links $g(\mu)$, and the scale parameters σ and ν for these variance functions are

$V(\mu) = 1$ **Normal**

$\sigma^2 Q(\mu; y) = -\frac{1}{2}(y-\mu)^2 \quad \text{for } -\infty < y < \infty$

$g(\mu) = \mu$

$\sigma = \sqrt{\phi}$

$V(\mu) = \mu$ **Poisson**

$\sigma^2 Q(\mu; y) = y \log(\mu) - \mu \quad \text{for } \mu > 0, y \geq 0$

$g(\mu) = \log \mu$

$\sigma = \sqrt{\phi}$

$V(\mu) = \mu^2$ **Gamma**

$\sigma^2 Q(\mu; y) = -y/\mu - \log(\mu) \quad \text{for } \mu > 0, y \geq 0$

$g(\mu) = \mu^{-1}$

$\nu = \phi^{-1}$

$V(\mu) = \mu^3$ **Inverse Gaussian**

$\sigma^2 Q(\mu; y) = -y/(2\mu^2) + 1/\mu \quad \text{for } \mu > 0, y \geq 0$

$g(\mu) = \mu^{-2}$

$\sigma = \sqrt{\phi}$

$V(\mu) = \mu(1-\mu)$ **Binomial**

$\sigma^2 Q(\mu; y) = r \log(\mu) + (m-r) \log(1-\mu)$

for $0 < \mu < 1$, $y = r/m$, $r = 0, 1, 2, \ldots, m$

$g(\mu) = \log(\frac{\mu}{1-\mu})$

$\sigma = \sqrt{\phi}$

SAS/INSIGHT software uses the mean deviance, the mean Pearson χ^2, or the value in the **Constant** entry field to estimate the dispersion parameter ϕ. The conventional estimate of ϕ is the mean Pearson χ^2 statistic.

Maximum quasi-likelihood estimation is similar to ordinary maximum-likelihood estimation and has the same parameter estimates as the distribution with the same variance function. These estimates are not affected by the dispersion parameter ϕ, but ϕ is used in the variance-covariance matrix of the parameter estimates. However, the likelihood-ratio based statistics, such as **Type I (LR)**, **Type III (LR)**, and **C.I.(LR) for Parameters** tables, are not produced in the analysis.

⊕ **Related Reading:** Logistic Regression, Chapter 16.

⊕ **Related Reading:** Poisson Regression, Chapter 17.

Nonparametric Smoothers

For a simple regression model with one or two explanatory variables,

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

a *smoother* $\widehat{f_\lambda}(\mathbf{x})$ is a function that summarizes the trend of **Y** as a function of **X**. It can enhance the visual perception of either a **Y**-by-**X** scatter plot or a rotating plot. The smoothing parameter λ controls the smoothness of the estimate.

With one explanatory variable in the model, $\widehat{f_\lambda}(\mathbf{x})$ is called a *scatter plot smoother*. SAS/INSIGHT software provides nonparametric curve estimates from smoothing spline, kernel, loess (nearest neighbors local polynomial), and fixed bandwidth local polynomial smoothers.

For smoothing spline, kernel, and fixed bandwidth local polynomial smoothers, SAS/INSIGHT software derives the smoothing parameter λ from a constant *c* that is independent of the units of **X**. For a loess smoother, the smoothing parameter λ is a positive constant α.

With two explanatory variables in the model, $\widehat{f_\lambda}(\mathbf{x})$ is called a *surface smoother*. SAS/INSIGHT software provides nonparametric surface estimates from thin-plate smoothing spline and kernel smoothers. The explanatory variables are scaled by their corresponding sample interquartile ranges. The smoothing parameter λ is derived from a constant *c* and both are independent of the units of **X**.

Similar to parametric regression, the R^2 value for an estimate is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{f_\lambda}(\mathbf{x}_i))^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

You can use the following methods to choose the λ value:

DF uses the λ value that makes the resulting smoothing estimate have the specified degrees of freedom (*df*).

GCV uses the λ value that minimizes the generalized cross validation (GCV) mean squared error.

C Value uses the λ value derived from the specified *c* value for nonparametric smoothers other than the loess smoother.

Alpha uses the specified α value for the loess estimator.

If you specify a **DF** value for a smoother, an iterative procedure is used to find the estimate with the specified *df*. You can choose a convergence criterion γ based on either the relative difference or the absolute difference. A smoother satisfying the following conditions is then created:

$$\frac{|df\,(\text{fitted}) - \text{df}(\text{specified})|}{\text{df}(\text{specified})} < \gamma \quad \text{for relative difference}$$

$$|df\,(\text{fitted}) - \text{df}(\text{specified})| < \gamma \quad \text{for absolute difference}$$

Smoother Degrees of Freedom

For a nonparametric smoother with a parameter λ, the fitted values can be written as

$$\hat{\mathbf{y}} = \mathbf{H}_\lambda \mathbf{y}$$

where $\mathbf{y}$ is the $n \times 1$ vector of observed responses y_i, $\hat{\mathbf{y}}$ is the $n \times 1$ vector of fitted values $\hat{y}_i = \widehat{f}_\lambda(x_i)$, and the smoother matrix $\mathbf{H}_\lambda$ is an $n \times n$ matrix that depends on the value of λ.

The degrees of freedom, or the effective number of parameters, of a smoother can be used to compare different smoothers and to describe the flexibility of the smoother. SAS/INSIGHT software defines the degrees of freedom of a smoother as

$$df_\lambda = \text{trace}(\mathbf{H}_\lambda)$$

which is the sum of the diagonal elements of $\mathbf{H}_\lambda$.

† **Note:** Two other popular definitions of degrees of freedom for a smoother are $\text{trace}(\mathbf{H}_\lambda \mathbf{H}'_\lambda)$ and $\text{trace}(2\mathbf{H}_\lambda - \mathbf{H}_\lambda \mathbf{H}'_\lambda)$ (Hastie and Tibshirani 1990).

Smoother Generalized Cross Validation

With the degrees of freedom of an estimate df_λ, the mean squared error is given as

$$\mathrm{MSE}(\lambda) = \frac{1}{\mathrm{n} - \mathrm{df}_\lambda} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \widehat{\mathrm{f}}_\lambda(\mathbf{x}_\mathrm{i}))^2$$

Cross-validation (CV) estimates the response at each x_i from the smoother that uses only the remaining $n - 1$ observations. The resulting cross validation mean squared error is

$$\mathrm{MSE}_{\mathrm{CV}}(\lambda) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \hat{\mathrm{f}}_{\lambda(\mathrm{i})}(\mathbf{x}_\mathrm{i}))^2$$

where $\hat{f}_{\lambda(i)}(\mathbf{x}_i)$ is the fitted value at x_i computed without the ith observation.

The cross validation mean squared error can also be written as

$$\mathrm{MSE}_{\mathrm{CV}}(\lambda) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\mathrm{y_i} - \widehat{\mathrm{f}}_\lambda(\mathbf{x}_\mathrm{i})}{1 - \mathrm{h}_{\lambda\mathrm{i}}} \right)^2$$

where $h_{\lambda i}$ is the ith diagonal element of the $\mathbf{H}_\lambda$ matrix (Hastie and Tibshirani 1990).

Generalized cross validation replaces $h_{\lambda i}$ by its average value, $\frac{1}{n} df_\lambda$. The generalized cross validation mean squared error is

$$\mathrm{MSE}_{\mathrm{GCV}}(\lambda) = \frac{1}{\mathrm{n}(1 - \mathrm{df}_\lambda/\mathrm{n})^2} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y_i} - \widehat{\mathrm{f}}_\lambda(\mathbf{x}_\mathrm{i}))^2$$

† **Note:** The function $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$ may have multiple minima, so the value estimated by SAS/INSIGHT software may be only a local minimum, not the global minimum.

Variables

To create a fit analysis, choose **Analyze:Fit (Y X)**. If you have already selected one or more variables, the first variable selected is the response or dependent variable, and it is assigned the **Y** variable role. The remaining variables are explanatory or independent variables, and they are assigned the **X** variable role. If you do not select any **X** effects, a model with only an intercept term (mean) is fit.

If you have not selected any variables, a variables dialog appears.

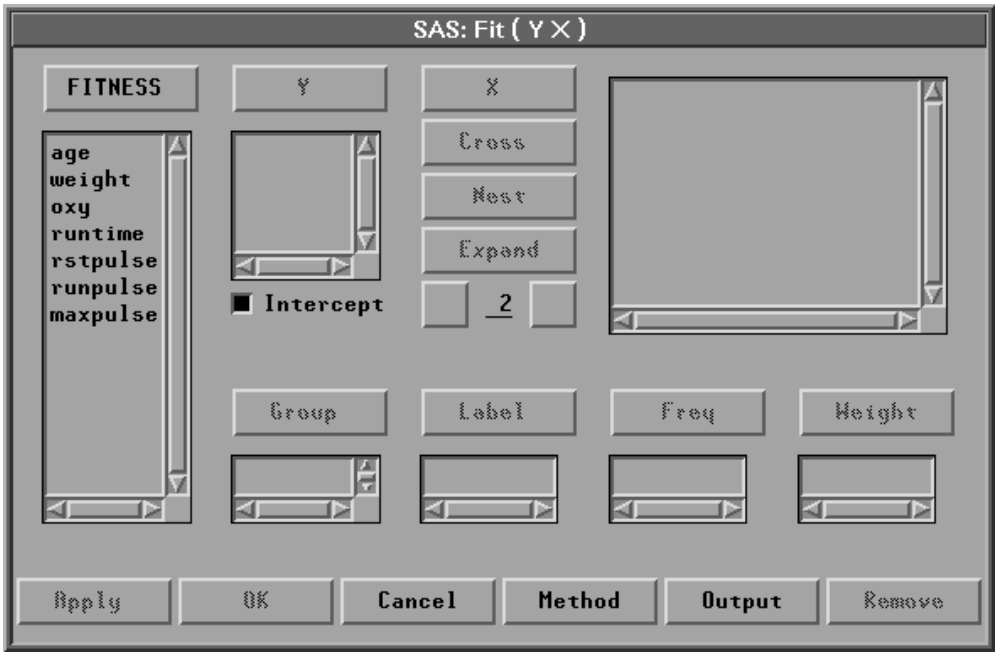


Figure 39.2. Fit Variables Dialog

In the dialog, select one **Y** variable for each fit analysis. Create **X** effects in the model by using the **X**, **Cross**, **Nest**, and **Expand** buttons. An *effect* is a variable or combination of variables that constitutes a term in the model. There are four ways to specify effects in SAS/INSIGHT software. In the following discussion, assume that **X1** and **X2** are interval variables and **A** and **B** are nominal variables.

You can use the **X** button to create regressor effects of the interval variables and main effects of the nominal variables. Select any variable, then click the **X** button. For example, selecting **A** and then clicking the **X** button adds **A** to the effects list.

You can use the **Cross** button to create crossed effects. These include polynomial effects of the interval variables, interactions of the nominal variables, and interaction effects of interval and nominal variables. Select two or more variables, then click the **Cross** button. For example, selecting **X1** and **X2** and then clicking the **Cross** button generates the crossed effect **X1*X2**.

You can use the **Nest** button to create nested effects. In a nested effect, a variable or crossed effect is nested within the effects of one or more nominal variables. Select a variable or crossed effect and one or more nominal variables, then click the **Nest**

button. For example, selecting **X1*X2**, **A**, and **B** and then clicking the **Nest** button generates the nested effect **X1*X2(A B)**.

You can use the **Expand** button and the associated entry field to create expanded effects. These include response surface effects for interval variables and factorial effects for nominal variables. The **Expand** button expands all possible effects to the degree of expansion specified in the entry field below the **Expand** button. The value **2** is the default degree of expansion. You can click the right button of the entry field to increase the expansion degree by 1 or the left button to decrease it by 1.

Choose the degree of expansion, then select variables or effects and click the **Expand** button. For example, with degree of expansion 2 and variables **A** and **B** selected, clicking the **Expand** button generates three effects

A B A*B

With degree of expansion 2 and variables **X1** and **X2** selected, clicking the **Expand** button generates five effects

X1 X2 X1*X1 X1*X2 X2*X2

Intercept is checked by default to include the intercept term in the model. As a general rule, no-intercept models should be fit only when theoretical justification exists.

You can select one or more **Group** variables if you have grouped data. This creates a fit analysis for each group.

You can select a **Label** variable to label observations in the plots.

You can select a **Freq** variable. If you select a **Freq** variable, each observation is assumed to represent n observations, where n is the value of the **Freq** variable.

You can select a **Weight** variable to assign relative weights for each observation in the analysis. The details of weighted analyses are explained in the "Weighted Analyses" section at the end of this chapter.

The fit variables dialog provides an **Apply** button. The **Apply** button displays the fit window without closing the fit variables dialog. This makes it easy to modify the model by adding or removing variables. Each time you modify the model using the **Apply** button, a *new* fit window is displayed so you can easily compare models. The **OK** button also displays a new fit window but closes the dialog.

Method

Observations with missing values for **Y**, **X**, **Weight**, or **Freq** variables are not used. Observations with nonpositive **Weight** or **Freq** values are not used. Only the integer part of **Freq** values is used.

To view or change the response distribution and link function, click the **Method** button in the variables dialog. This displays the dialog shown in Figure 39.3.

Figure 39.3. Fit Method Options Dialog

You can choose the response distribution and link function of the **Y** variables. If you choose a binomial distribution, specify either

- a **Y** variable with values 1 or 0 indicating success or failure
- a **Y** variable giving the number of successes in a certain number of trials, and a **Binomial** variable to give the corresponding number of trials

If you choose a power link function, specify the power value in the **Power** entry field.

If you select an **Offset** variable, it is treated as an **X** variable with coefficient fixed at 1.0.

You can choose the scale parameter for the response distribution. If you choose a **Constant** scale, specify the constant value in the **Constant** entry field.

With overdispersion in the model, you can specify the **Quasi-Likelihood** option to fit the generalized linear model using the quasi-likelihood functions.

If you choose a normal response distribution with a canonical link (identity for normal distributions), you can specify the **Exact Distribution** option to fit the linear model using the usual exact distributions for the test statistics.

You can specify the **Fisher's Scoring** option to use the Fisher's scoring method in the maximum-likelihood estimation for the regression parameters.

By default, SAS/INSIGHT software uses the **Normal** response distribution and **Canonical** link with the **Exact Distribution** option to perform a fit analysis for the linear model.

Output

To view or change the options associated with your fit analysis, click the **Output** button in the variables dialog. This displays the output options dialog shown in Figure 39.4.

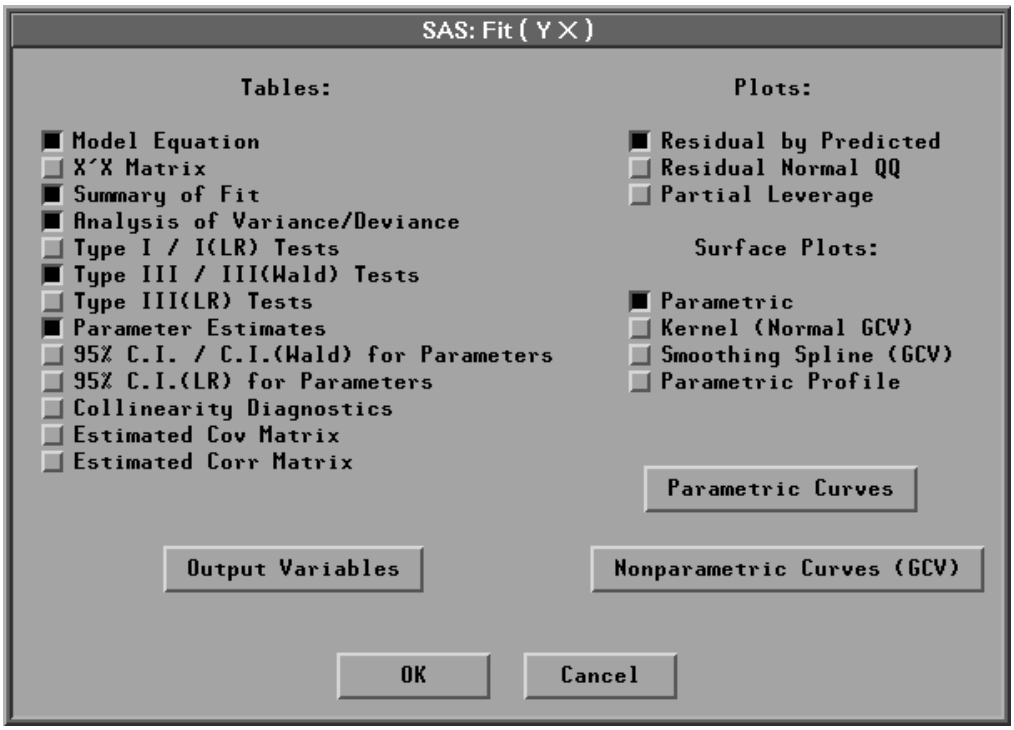


Figure 39.4. Fit Output Options Dialog

The options you set in this dialog determine the tables and graphs that appear in the fit window. Provided by default are tables of the model equation, summary of fit, analysis of variance or deviance, type III or type III (Wald) tests, and parameter estimates and a plot of residuals by predicted values.

When there are two explanatory variables in the model, a parametric response surface plot is created by default. You can also generate a nonparametric kernel or a thin-plate smoothing spline response surface plot. With more than two explanatory variables in the model, a parametric profile response surface plot with the first two explanatory variables can be created. The values of the remaining explanatory variables are set to their corresponding means in the plot. You can use the sliders to change these values of the remaining explanatory variables.

Click on the **Output Variables** button in the fit dialog to display the **Output Variables** dialog shown in Figure 39.5. The **Output Variables** dialog enables you to specify variables that can be saved in the data window. Output variables include predicted values and several influence diagnostic variables based on the model you fit.

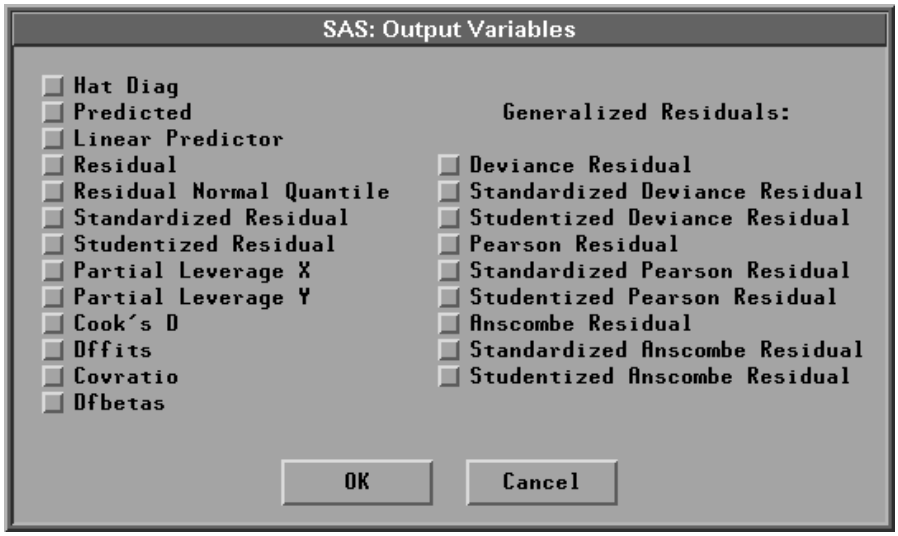


Figure 39.5. Output Variables Dialog

When there is only one explanatory variable in the model, a **Y**-by-**X** scatter plot is generated. The **Parametric Curves** and **Nonparametric Curves (GCV)** buttons display dialogs that enable you to fit parametric and nonparametric curves to this scatter plot.

Click on **Parametric Curves** to display the **Parametric Curves** dialog.

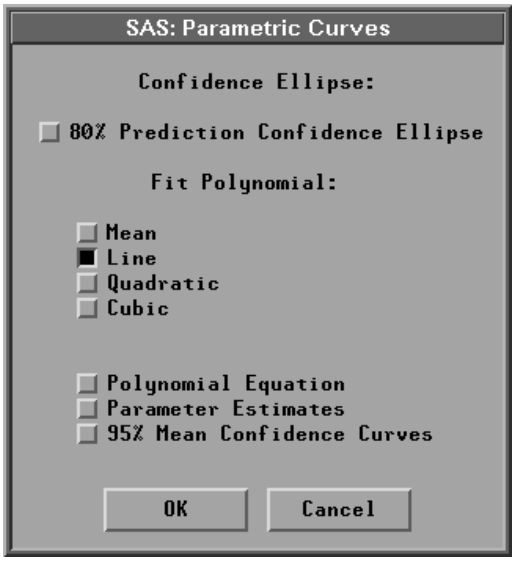


Figure 39.6. Parametric Curves Dialog

A regression line fit is provided by default. You can request an 80% prediction ellipse and other polynomial fits in the dialog. You can also request polynomial equation tables, parameter estimates tables, and 95% mean confidence curves for fitted polynomials.

The **Nonparametric Curves (GCV)** dialog in Figure 39.7 includes a smoothing spline, a kernel smoother, and a local polynomial smoother. You must specify the method, regression type, and weight function for a local polynomial fit.

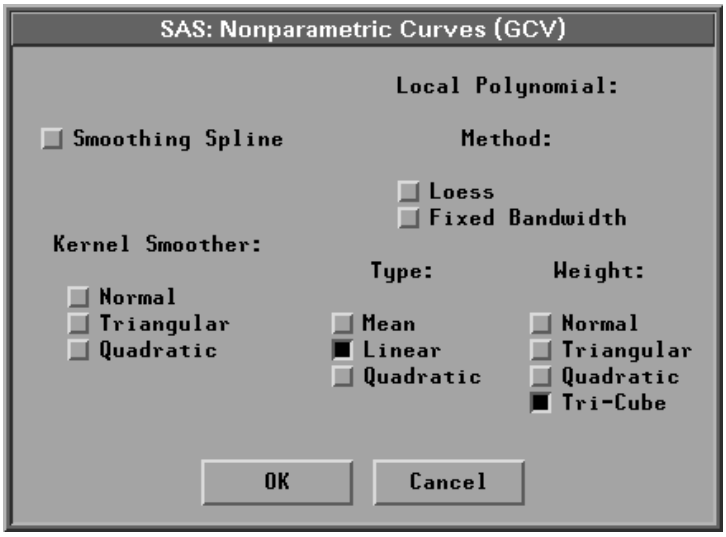


Figure 39.7. Nonparametric Curves Dialog

Tables

You can generate tables that present the results of a model fit and diagnostics for assessing how well the model fits the data. Set options in the output dialog as described in the "Output" section or choose from the **Tables** menu.

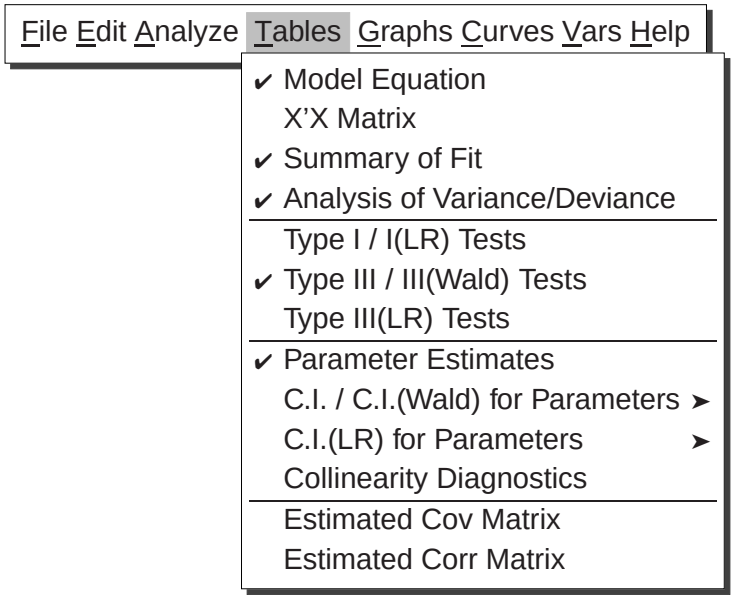


Figure 39.8. Tables Menu

Model Information

The first table in the fit analysis contains the model specification, the response distribution, and the link function, as illustrated in Figure 39.9.

When the model contains nominal variables in its effects, the levels of the nominal variables are displayed in the **Nominal Variable Information** table, as shown in Figure 39.9. The levels are determined from the formatted values of the nominal variables. An additional **Parameter Information** table, as illustrated in Figure 39.9, shows the variable indices for the parameters in the model equation, the X'X matrix, the estimated covariance matrix, and the estimated correlation matrix.

Model Equation

The model equation table gives the fitted equation for the model. Figure 39.9 shows an equation for a model with nominal variables, and Figure 39 shows an equation for a model without nominal variables

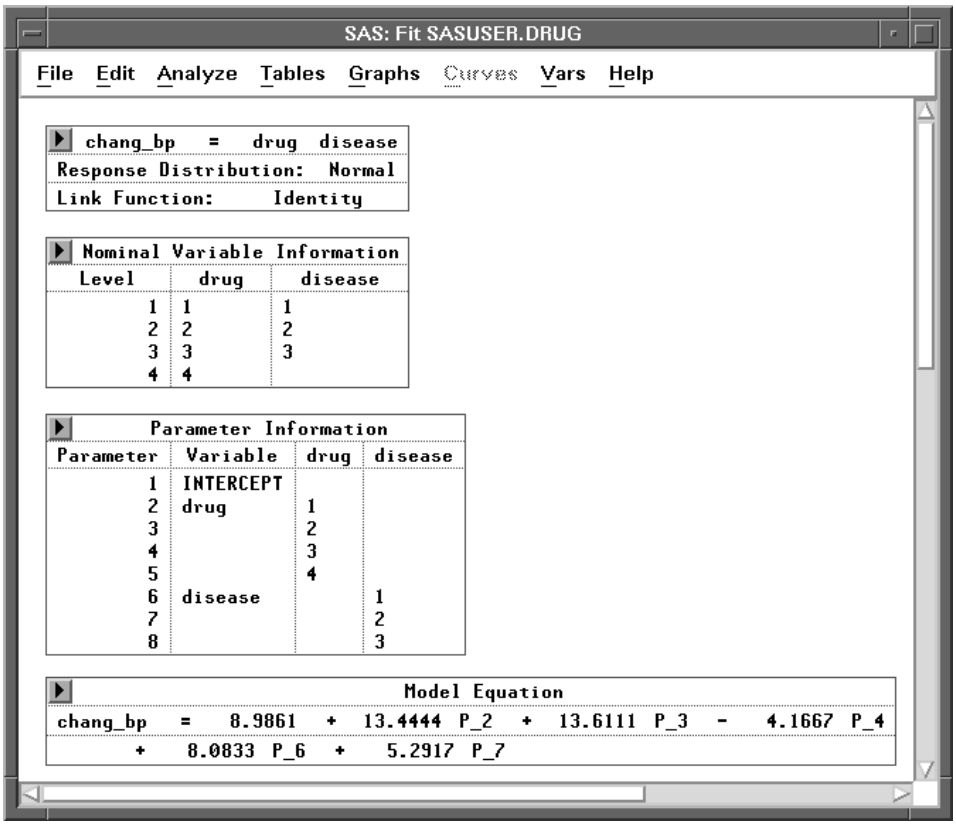


Figure 39.9. Model Information Tables

X'X Matrix

The X'X matrix table, as illustrated by Figure 39.10, contains the X'X crossproducts matrix for the model.

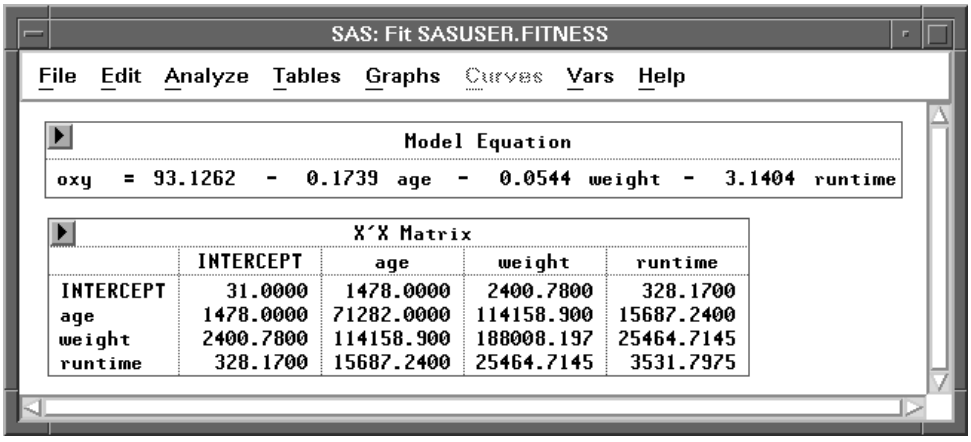


Figure 39.10. X'X Matrix for Linear Models

Summary of Fit for Linear Models

The **Summary of Fit** table for linear models, shown in Figure 39.11, includes the following:

Mean of Response is the sample mean, $\overline{y}$, of the response variable.

Root MSE is the estimate of the standard deviation of the error term. It is calculated as the square root of the mean square error.

R-Square R^2, with values between 0 and 1, indicates the proportion of the (corrected) total variation attributed to the fit.

Adj R-Sq An adjusted R^2 is a version of R^2 that has been adjusted for degrees of freedom.

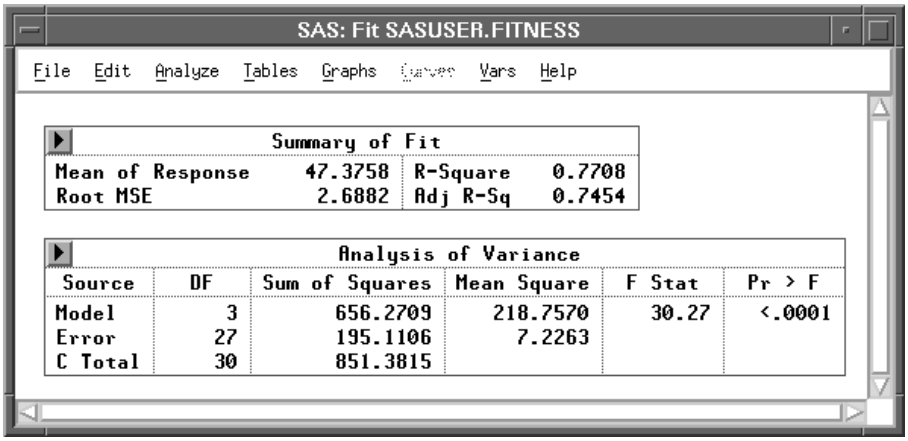


Figure 39.11. Summary of Fit, Analysis of Variance Tables for Linear Models

With an intercept term in the model, R^2 is defined as

$$R^2 = 1 - (SSE/CSS)$$

where CSS $= \sum_{i=1}^{n} (y_i - \overline{y})^2$ is the corrected sum of squares and SSE $= \sum_{i=1}^{n} (y_i - \hat{y})^2$ is the sum of squares for error.

The R^2 statistic is also the square of the multiple correlation, that is, the square of the correlation between the response variable and the predicted values.

The adjusted R^2 statistic, an alternative to R^2, is adjusted for the degrees of freedom of the sums of squares associated with R^2. It is calculated as

$$AdjR^2 = 1 - \frac{SSE/(n-p)}{CSS/(n-1)} = 1 - \frac{n-1}{n-p}(1-R^2)$$

Without an intercept term in the model, R^2 is defined as

$$R^2 = 1 - (SSE/TSS)$$

where $\mathrm{TSS} = \sum_{i=1}^{n} y_i^2$ is the uncorrected total sum of squares.

The adjusted R^2 statistic is then calculated as

$$AdjR^2 = 1 - \frac{SSE/(n-p)}{TSS/n} = 1 - \frac{n}{n-p}(1 - R^2)$$

† **Note:** Other definitions of R^2 exist for models with no intercept term. Care should be taken to ensure that this is the definition desired.

Summary of Fit for Generalized Linear Models

For generalized linear models, the **Summary of Fit** table, as illustrated by Figure 39.12, includes the following:

Mean of Response is the sample mean, $\overline{y}$, of the response variable.

SCALE is the constant scale parameter specified in the method dialog or a value of 1.0 for maximum-likelihood estimation for Poisson or binomial distributions.

SCALE (MLE) is the maximum-likelihood estimate of the scale parameter for normal, gamma, and inverse Gaussian distributions.

SCALE (Deviance) is the scale parameter estimated by the mean error deviance.

SCALE (Pearson) is the scale parameter estimated by the mean Pearson χ^2.

Deviance is the error deviance.

Deviance/DF is the mean error deviance, the error deviance divided by its associated degrees of freedom.

Pearson ChiSq is the Pearson χ^2 statistic.

Pearson ChiSq / DF is the mean Pearson χ^2, the Pearson χ^2 divided by its associated degrees of freedom.

When the scale parameter is a constant specified in the method dialog, or when the response has a Poisson or binomial distribution, the table also contains the scaled deviance and the scaled Pearson χ^2:

Scaled Dev is the error deviance divided by the dispersion parameter.

Scaled ChiSq is the Pearson χ^2 divided by the dispersion parameter.

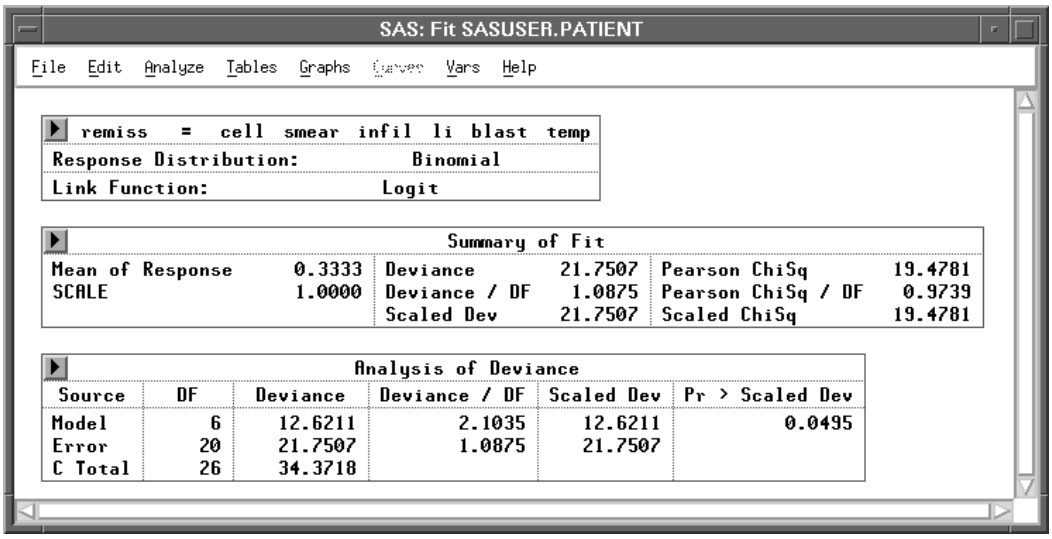


Figure 39.12. Summary of Fit and Analysis of Deviance Tables for Generalized Linear Models

Analysis of Variance for Linear Models

The **Analysis of Variance** table for linear models, shown in Figure 39.11, includes the following:

Source indicates the source of the variation. Sources include **Model** for the fitted regression and **Error** for the residual error. **C Total** is the sum of the **Model** and **Error** components, and it is the total variation after correcting for the mean. When the model does not have an intercept term, the uncorrected total variation (**U Total**) is displayed.

DF is the degrees of freedom associated with each source of variation.

Sum of Squares is the sum of squares for each source of variation.

Mean Square is the sum of squares divided by its associated degrees of freedom.

F Stat is the F statistic for testing the null hypothesis that all parameters are 0 except for the intercept. This is formed by dividing the mean square for model by the mean square for error.

Pr > F is the probability of obtaining a greater F statistic than that observed if the null hypothesis is true. This quantity is also called a p-value. A small p-value is evidence for rejecting the null hypothesis.

Analysis of Deviance for Generalized Linear Models

The **Analysis of Deviance** table for generalized linear models, as illustrated by Figure 39.12, includes the following:

Source indicates the source of the variation. Sources include **Model** for the fitted regression and **Error** for the residual error. **C Total** is the sum of the **Model** and **Error** components, and it is the total variation after correcting for the mean. When the model does not have an intercept term, the uncorrected total variation (**U Total**) is printed.

DF is the degrees of freedom associated with each source of variation.

Deviance is the deviance for each source of variation.

Deviance/DF is the deviance divided by its associated degrees of freedom.

When the scale parameter is a constant specified in the method dialog, or when the response has a Poisson or binomial distribution, the table also contains the following:

Scaled Dev is the deviance divided by the dispersion parameter.

Pr>Scaled Dev is the probability of obtaining a greater scaled deviance statistic than that observed if the null hypothesis is true. Under the null hypothesis, all parameters are 0 except for the intercept, and the scaled deviance has an approximate χ^2 distribution.

Type I Tests

Type I tests examine the sequential incremental improvement in the fit of the model as each effect is added. They can be computed by fitting the model in steps and recording the difference in error sum of squares (linear models) and log-likelihood statistics (generalized linear models). The **Type I Tests** table for linear models, as illustrated by Figure 39.13, includes the following:

Source is the name for each effect.

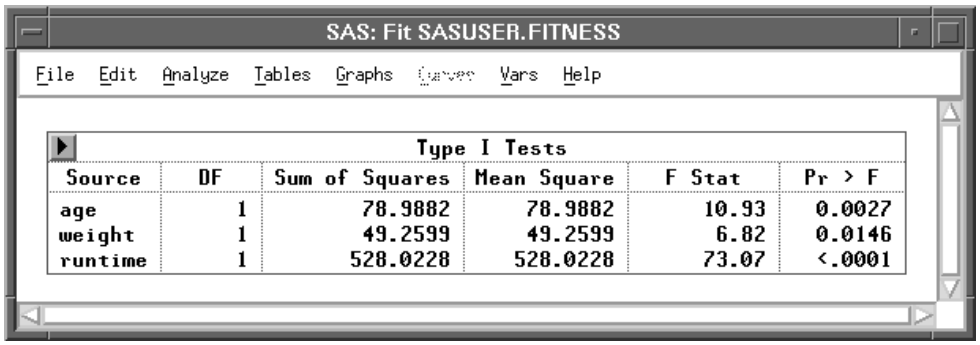
DF is the degrees of freedom associated with each effect.

Sum of Squares is the incremental error sum of squares for the model as each effect is added.

Mean Square is the sum of squares divided by its associated degrees of freedom.

F Stat is the F statistic for testing the null hypothesis that the parameters for the added effect are 0. This is formed by dividing the mean square for the effect by the mean square for error from the complete model.

Pr > F is the probability of obtaining a greater F statistic than that observed if the null hypothesis is true.



Type I Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
age	1	78.9882	78.9882	10.93	0.0027
weight	1	49.2599	49.2599	6.82	0.0146
runtime	1	528.0228	528.0228	73.07	<.0001

Figure 39.13. Type I Tests Table

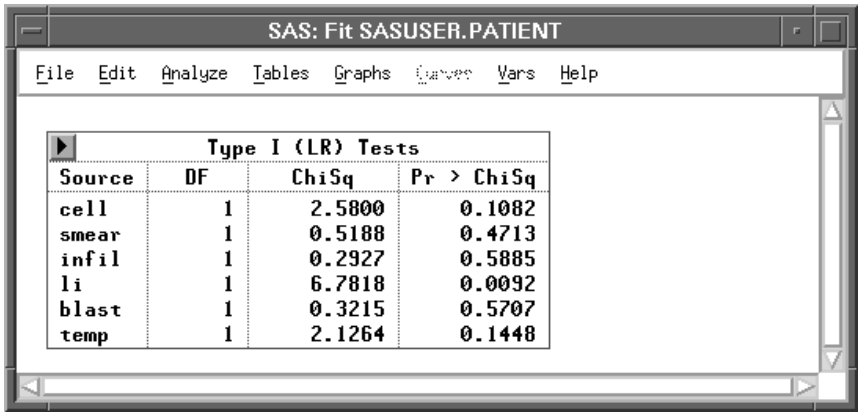
The **Type I (LR) Tests** table for generalized linear models, as illustrated by Figure 39.14, includes the following:

Source is the name for each effect.

DF is the degrees of freedom associated with each effect.

ChiSq is the χ^2 value for testing the null hypothesis that the parameters for the added effect are 0. This is evaluated as twice the incremental log-likelihood for the model as each effect is added, and it has an asymptotic χ^2 distribution under the null hypothesis.

Pr > ChiSq is the probability of obtaining a greater χ^2 statistic than that observed, if the null hypothesis is true.



Type I (LR) Tests

Source	DF	ChiSq	Pr > ChiSq
cell	1	2.5800	0.1082
smear	1	0.5188	0.4713
infil	1	0.2927	0.5885
li	1	6.7818	0.0092
blast	1	0.3215	0.5707
temp	1	2.1264	0.1448

Figure 39.14. Type I Likelihood Ratio Tests

Type III Tests

Type III tests examine the significance of each partial effect, that is, the significance of an effect with all the other effects in the model. They are computed by constructing a type III hypothesis matrix $\mathbf{L}$ and then computing statistics associated with the hypothesis $\mathbf{L}\beta = 0$. Refer to the chapter titled "The Four Types of Estimable Functions," in the *SAS/STAT User's Guide* for the construction of the matrix $\mathbf{L}$.

For linear models, the type III or partial sum of squares

$$(\mathbf{Lb})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{Lb})$$

is used to test the hypothesis $\mathbf{L}\beta = 0$.

The **Type III Tests** table for linear models, as illustrated by Figure 39.15, includes the following:

Source is the name for each effect.

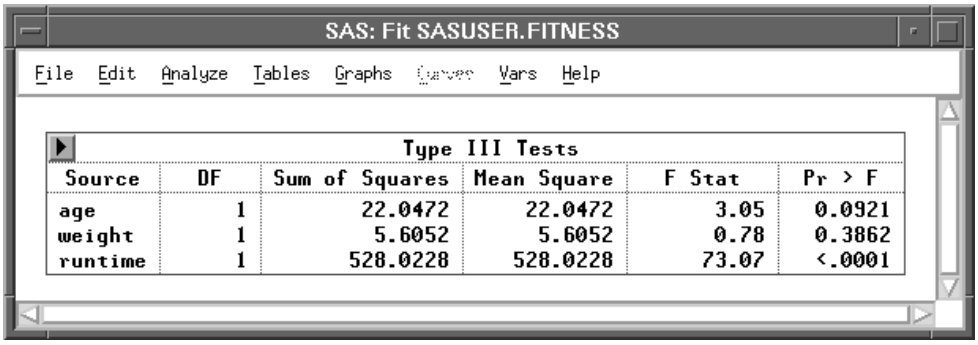
DF is the degrees of freedom associated with each effect.

Sum of Squares is the partial sum of squares for each effect in the model.

Mean Square is the sum of squares divided by its associated degrees of freedom.

F Stat is the F statistic for testing the null hypothesis that the linear combinations of parameters described previously for the hypothesis matrix $\mathbf{L}$ are 0. This is formed by dividing the mean square for the hypothesis matrix $\mathbf{L}$ by the mean square for error from the complete model.

Pr > F is the probability of obtaining a greater F statistic than that observed if the null hypothesis is true.



Type III Tests

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
age	1	22.0472	22.0472	3.05	0.0921
weight	1	5.6052	5.6052	0.78	0.3862
runtime	1	528.0228	528.0228	73.07	<.0001

Figure 39.15. Type III Tests Table for Linear Models

For generalized linear models, either the Wald statistic or the likelihood-ratio statistic can be used to test the hypothesis $\mathbf{L}\beta = 0$. For the linear model, the two tests are equivalent.

The Wald statistic is given by

$$(\mathbf{Lb})'(\mathbf{L}\widehat{\mathrm{Var}}(\mathbf{b})\mathbf{L}')^{-1}(\mathbf{Lb})$$

where $\widehat{\mathrm{Var}}(\mathbf{b})$ is the estimated covariance matrix of the parameters. The likelihood-ratio statistic is computed as twice the difference between the maximum log-likelihood achievable under the unconstrained model and the maximum log-likelihood for the model under the restriction or constraint $\mathbf{L}\beta = 0$. Both the Wald statistic and the likelihood-ratio statistic have an asymptotic χ^2 distribution.

The **Type III (Wald) Tests** and **Type III (LR) Tests** tables, as illustrated by Figure 39.16, include the following:

Source is the name for each effect.

DF is the degrees of freedom associated with each effect.

ChiSq is the Wald statistic for the Wald tests or the likelihood-ratio statistic for the LR tests of the null hypothesis that the parameters for the effect are 0. This has an asymptotic χ^2 distribution.

Pr > ChiSq is the probability of obtaining a greater χ^2 statistic than that observed, if the null hypothesis is true.

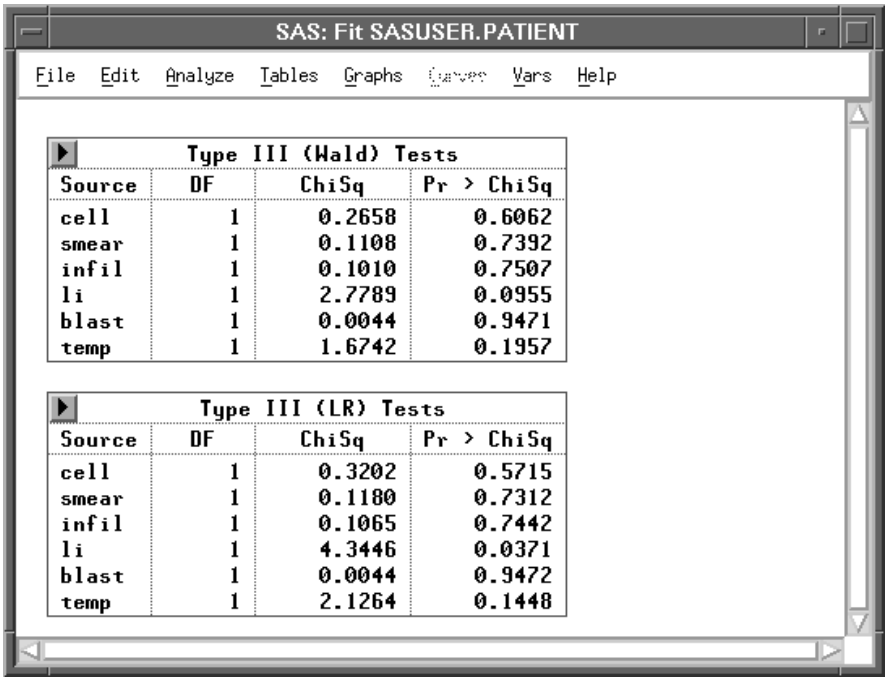


Figure 39.16. Type III Tests Tables for Generalized Linear Models

Parameter Estimates for Linear Models

The **Parameter Estimates** table for linear models, as illustrated by Figure 39.17, includes the following:

Variable names the variable associated with the estimated parameter. The name **INTERCEPT** represents the estimate of the intercept parameter.

DF is the degrees of freedom associated with each parameter estimate. There is one degree of freedom unless the model is not of full rank. In this case, any parameter whose definition is confounded with previous parameters in the model has its degrees of freedom set to 0.

Estimate is the parameter estimate.

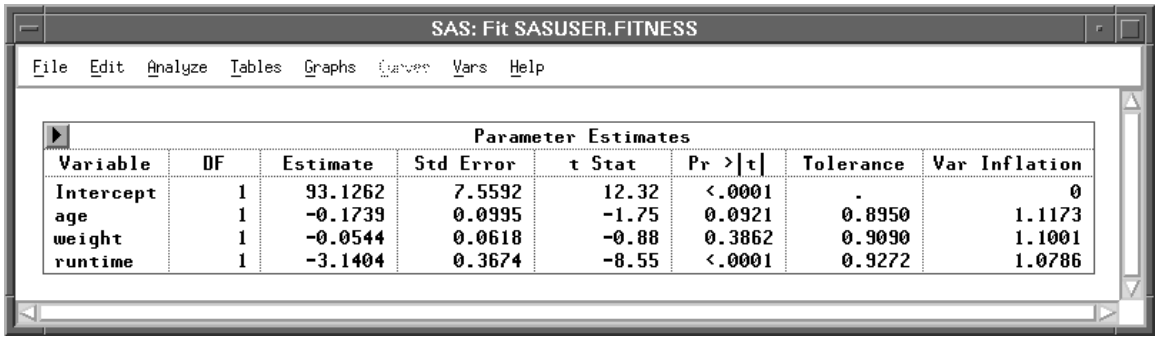
Std Error is the standard error, the estimate of the standard deviation of the parameter estimate.

t Stat is the t statistic for testing that the parameter is 0. This is computed as the parameter estimate divided by the standard error.

Pr > |t| is the probability of obtaining (by chance alone) a t statistic greater in absolute value than that observed given that the true parameter is 0. This is referred to as a two-sided p-value. A small p-value is evidence for concluding that the parameter is not 0.

Tolerance is the tolerance of the explanatory variable on the other variables.

Var Inflation is the variance inflation factor of the explanatory variable.



Parameter Estimates

Variable	DF	Estimate	Std Error	t Stat	Pr >|t|	Tolerance	Var Inflation
Intercept	1	93.1262	7.5592	12.32	<.0001	.	0
age	1	-0.1739	0.0995	-1.75	0.0921	0.8950	1.1173
weight	1	-0.0544	0.0618	-0.88	0.3862	0.9090	1.1001
runtime	1	-3.1404	0.3674	-8.55	<.0001	0.9272	1.0786

Figure 39.17. Parameter Estimates Table for Linear Models

The standard error of the *j*th parameter estimate b_j is computed using the equation

$$\mathrm{STDERR(b_j)} = \sqrt{(\mathbf{X'X})^{-1}_{\mathrm{jj}}\mathrm{s}^2}$$

where $(\mathbf{X'X})^{-1}_{jj}$ is the *j*th diagonal element of $(\mathbf{X'X})^{-1}$.

Under the hypothesis that β_j is 0, the ratio

$$t = \frac{b_j}{\mathrm{STDERR(b_j)}}$$

is distributed as Student's *t* with degrees of freedom equal to the degrees of freedom for the mean square error.

When an explanatory variable is nearly a linear combination of other explanatory variables in the model, the affected estimates are unstable and have high standard errors. This problem is called *collinearity* or *multicollinearity*. A fit analysis provides several methods for detecting collinearity.

Tolerances (*TOL*) and *variance inflation factors* (*VIF*) measure the strength of interrelationships among the explanatory variables in the model. Tolerance is $1 - R^2$ for the R^2 that results from the regression of the explanatory variable on the other explanatory variables in the model. Variance inflation factors are diagonal elements of $(\mathbf{X'X})^{-1}$ after **X'X** is scaled to correlation form. The variance inflation measures the inflation in the variance of the parameter estimate due to collinearity between the explanatory variable and other variables. These measures are related by *VIF* = 1 / *TOL*.

If all variables are orthogonal to each other, both tolerance and variance inflation are 1. If a variable is closely related to other variables, the tolerance goes to 0 and the variance inflation becomes large.

When the **X'X** matrix is singular, least-squares solutions for the parameters are not unique. An estimate is 0 if the variable is a linear combination of previous explanatory variables. The degrees of freedom for the zeroed estimates are reported as 0. The hypotheses that are not testable have *t* tests printed as missing.

Parameter Estimates for Generalized Linear Models

The **Parameter Estimates** table for generalized linear models, as illustrated by Figure 39.18, includes the following:

Variable names the variable associated with the estimated parameter. The name **INTERCEPT** represents the estimate of the intercept parameter.

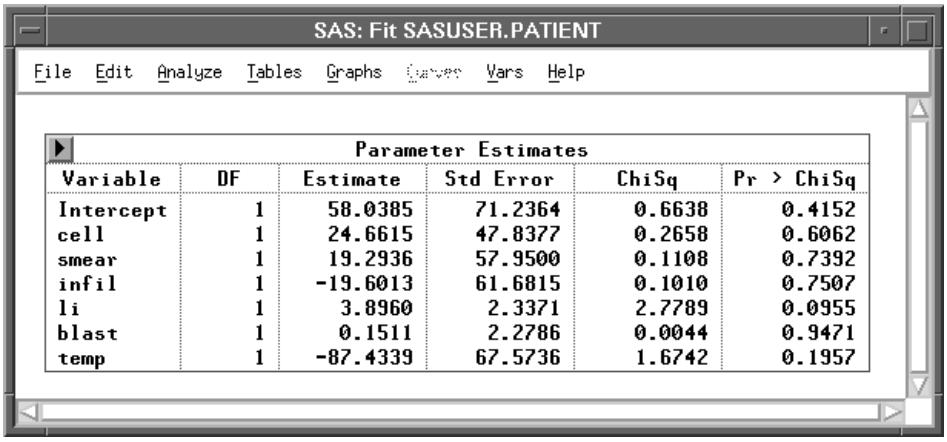
DF is the degrees of freedom associated with each parameter estimate. There is one degree of freedom unless the model is not full rank. In this case, any parameter that is confounded with previous parameters in the model has its degrees of freedom set to 0.

Estimate is the parameter estimate.

Std Error is the estimated standard deviation of the parameter estimate.

ChiSq is the χ^2 test statistic for testing that the parameter is 0. This is computed as the square of the ratio of the parameter estimate divided by the standard error.

Pr > ChiSq is the probability of obtaining an χ^2 statistic greater than that observed given that the true parameter is 0. A small p-value is evidence for concluding that the parameter is not 0.



Parameter Estimates

Variable	DF	Estimate	Std Error	ChiSq	Pr > ChiSq
Intercept	1	58.0385	71.2364	0.6638	0.4152
cell	1	24.6615	47.8377	0.2658	0.6062
smear	1	19.2936	57.9500	0.1108	0.7392
infil	1	-19.6013	61.6815	0.1010	0.7507
li	1	3.8960	2.3371	2.7789	0.0955
blast	1	0.1511	2.2786	0.0044	0.9471
temp	1	-87.4339	67.5736	1.6742	0.1957

Figure 39.18. Parameter Estimates Table for Generalized Linear Models

C.I. for Parameters

The **C.I. for Parameters** table gives a confidence interval for each parameter for each confidence coefficient specified. You choose the confidence interval for parameters either in the fit output options dialog or from the **Tables** menu, as shown in Figure 39.19.

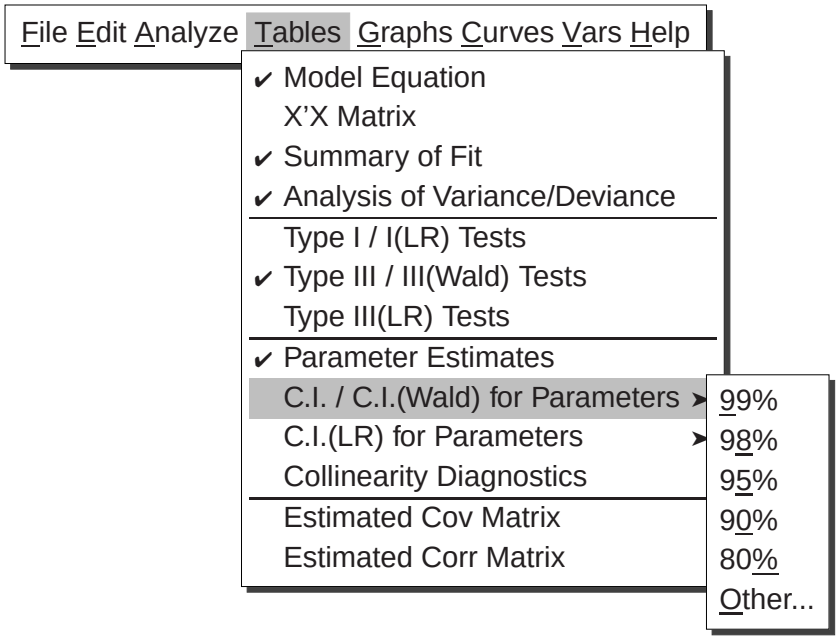


Figure 39.19. C.I. for Parameters Menu

Selecting **95% C.I. / C.I.(Wald) for Parameters** or **95% C.I.(LR) for Parameters** in the fit output options dialog produces a table with a 95% confidence interval for the parameters. This is the equivalent of choosing **Tables:C.I. / C.I.(Wald) for Parameters:95%** or **Tables:C.I.(LR) for Parameters:95%** from the **Tables** menu. You can also choose other confidence coefficients from the **Tables** menu. Figure 39.20 illustrates a 95% confidence intervals table for the parameters in a linear model.

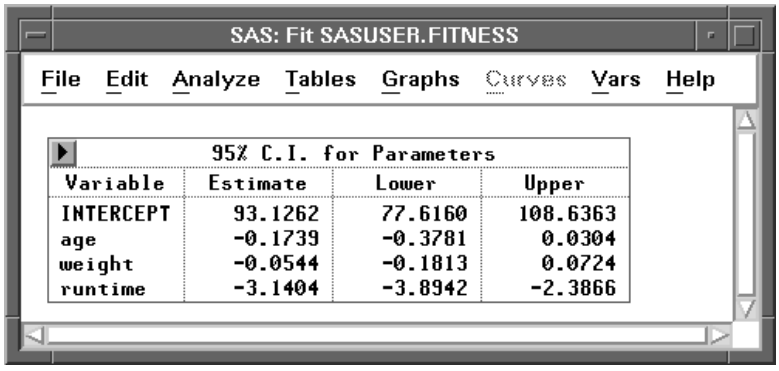


Figure 39.20. C.I. for Parameters Table

For linear models, a $100(1-\alpha)\%$ confidence interval has upper and lower limits

$$b_j \pm t_{(1-\alpha/2)} s_j$$

where $t_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ critical value of the Student's t statistic with degrees of freedom *n-p*, used in computing s_j, the estimated standard deviation of the parameter estimate b_j.

For generalized models, you can specify the confidence interval based on either a Wald type statistic or the likelihood function.

A $100(1-\alpha)\%$ Wald type confidence interval is constructed from

$$\left(\frac{\beta^j - b_j}{s_j}\right)^2 \leq \chi^2_{(1-\alpha),1}$$

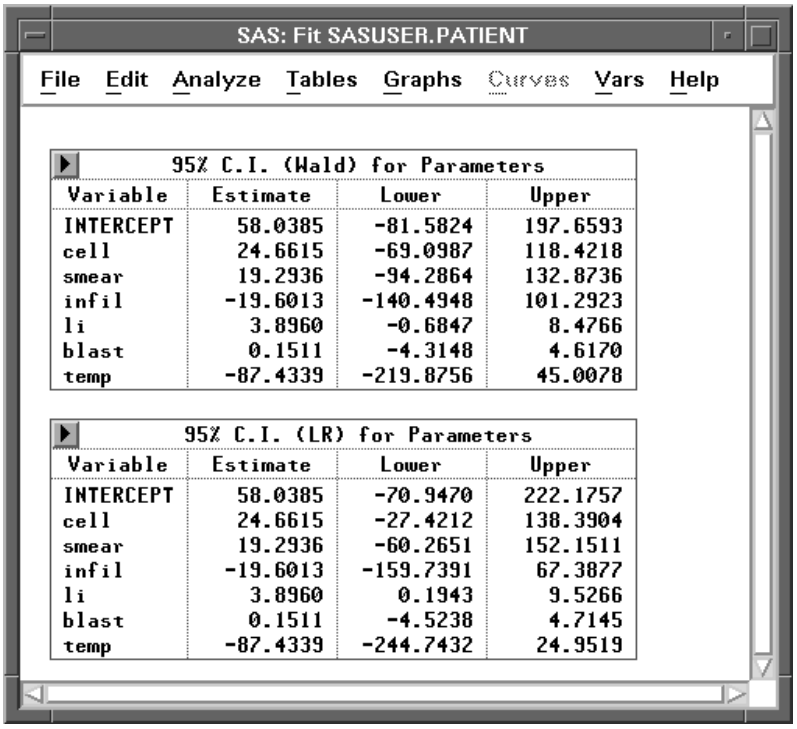
where $\chi^2_{(1-\alpha),1}$ is the $(1-\alpha)$ critical value of the χ^2 statistic with one degree of freedom, and s_j is the estimated standard deviation of the parameter estimate b_j.

Thus, $100(1-\alpha)\%$ upper and lower limits are

$$b_j \pm z_{(1-\alpha/2)} s_j$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ critical value of the standard normal statistic.

A table of 95% Wald type confidence intervals for the parameters is shown in Figure 39.21.



95% C.I. (Wald) for Parameters

Variable	Estimate	Lower	Upper
INTERCEPT	58.0385	-81.5824	197.6593
cell	24.6615	-69.0987	118.4218
smear	19.2936	-94.2864	132.8736
infil	-19.6013	-140.4948	101.2923
li	3.8960	-0.6847	8.4766
blast	0.1511	-4.3148	4.6170
temp	-87.4339	-219.8756	45.0078

95% C.I. (LR) for Parameters

Variable	Estimate	Lower	Upper
INTERCEPT	58.0385	-70.9470	222.1757
cell	24.6615	-27.4212	138.3904
smear	19.2936	-60.2651	152.1511
infil	-19.6013	-159.7391	67.3877
li	3.8960	0.1943	9.5266
blast	0.1511	-4.5238	4.7145
temp	-87.4339	-244.7432	24.9519

Figure 39.21. C.I. for Parameters Tables

The likelihood ratio test statistic for the null hypothesis

$$H_0\colon \beta_j = \beta_{j0}$$

where β_{j0} is a specified value, is

$$\lambda = -2(l(\widehat{\beta_0}) - l(\hat{\beta}))$$

where $l(\widehat{\beta_0})$ is the maximized log likelihood under H_0 and $l(\hat{\beta})$ is the maximized log likelihood over all β.

In large samples, the hypothesis is rejected at level α if the test statistic λ is greater than the $(1-\alpha)$ critical value of the chi-squared statistic with one degree of freedom.

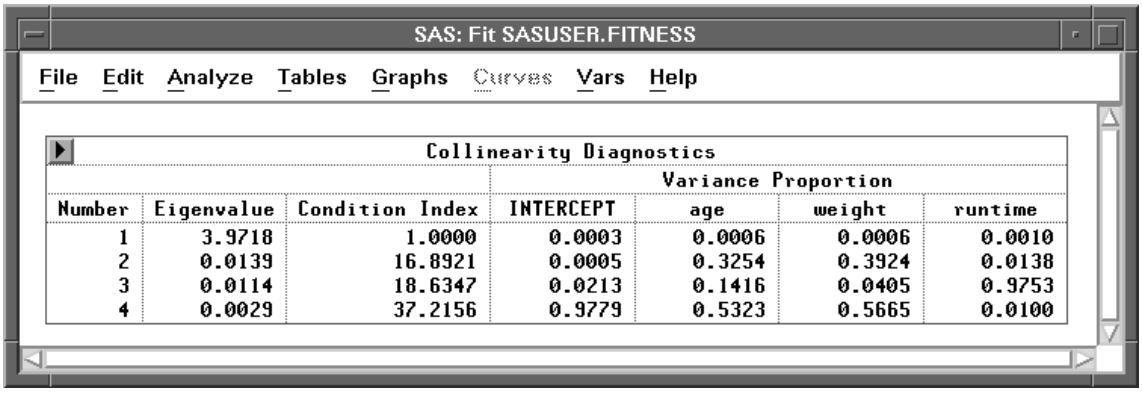
Thus a $100(1-\alpha)\%$ likelihood-based confidence interval is constructed using restricted maximization to find upper and lower limits satisfying

$$l(\widehat{\beta_0}) = l(\hat{\beta}) - \frac{1}{2}\chi^2_{(1-\alpha),1}$$

An iterative procedure is used to obtain these limits. A 95% likelihood-based confidence interval table for the parameters is illustrated in Figure 39.21.

Collinearity Diagnostics

The **Collinearity Diagnostics** table is illustrated by Figure 39.22.



Collinearity Diagnostics

			Variance Proportion			
Number	Eigenvalue	Condition Index	INTERCEPT	age	weight	runtime
1	3.9718	1.0000	0.0003	0.0006	0.0006	0.0010
2	0.0139	16.8921	0.0005	0.3254	0.3924	0.0138
3	0.0114	18.6347	0.0213	0.1416	0.0405	0.9753
4	0.0029	37.2156	0.9779	0.5323	0.5665	0.0100

Figure 39.22. Collinearity Diagnostics Table

Number is the eigenvalue number.

Eigenvalue gives the eigenvalues of the **X**'**X** matrix.

Condition Index is the square root of the ratio of the largest eigenvalue to the corresponding eigenvalue.

Variance Proportion is the proportion of the variance of each estimate accounted for by each component.

Detailed collinearity diagnostics use the eigenstructure of **X**'**X**, which can be written as

$\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}'$ where $\mathbf{V}$ is an orthogonal matrix whose columns are the eigenvectors of **X**'**X**, and $\mathbf{D}^2$ is a diagonal matrix of eigenvalues

$$d_1^2 \geq d_2^2 \geq \ldots \geq d_p^2$$

After scaling (**X**'**X**) to correlation form, Belsley, Kuh, and Welsch (1980) construct the condition indices as the square roots of the ratio of the largest eigenvalue to each individual eigenvalue, d_1/d_j, $j = 1, 2, \ldots, p$.

The *condition number* of the **X** matrix is defined as the largest condition index, d_1/d_p. When this number is large, the data are said to be *ill conditioned*. A condition index of 30 to 100 indicates moderate to strong collinearity.

For each variable, the proportion of the variance of its estimate accounted for by each component d_j can be evaluated. A collinearity problem occurs when a component associated with a high condition index contributes strongly to the variance of two or more variables. Thus, for a high condition index (>30), the corresponding row should be examined to see which variables have high values. Those would indicate near-linear dependence.

Estimated COV Matrix and Estimated CORR Matrix

The **Estimated COV Matrix** table contains the estimated variance-covariance matrix of the parameters. The **Estimated CORR Matrix** table contains the estimated correlation matrix of the parameters. Sample tables are shown in Figure 39.23.

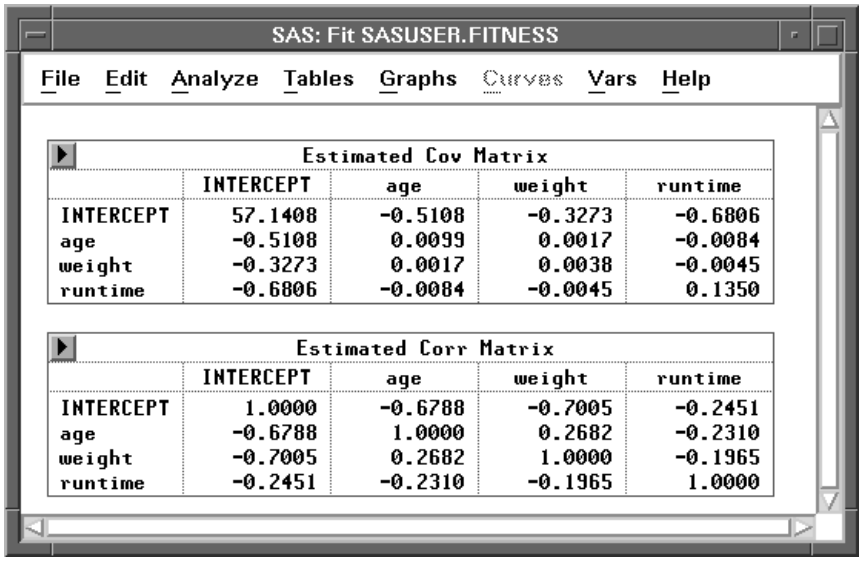


Figure 39.23. Estimated COV and CORR Matrices

Residual and Surface Plots

Residual plots provide visual displays for assessing how well the model fits the data, for evaluating the distribution of the residuals, and for identifying influential observations. Surface plots are three-dimensional displays of continuous response surfaces on the regular grids of the explanatory variables. They are much easier to comprehend than rotating plots.

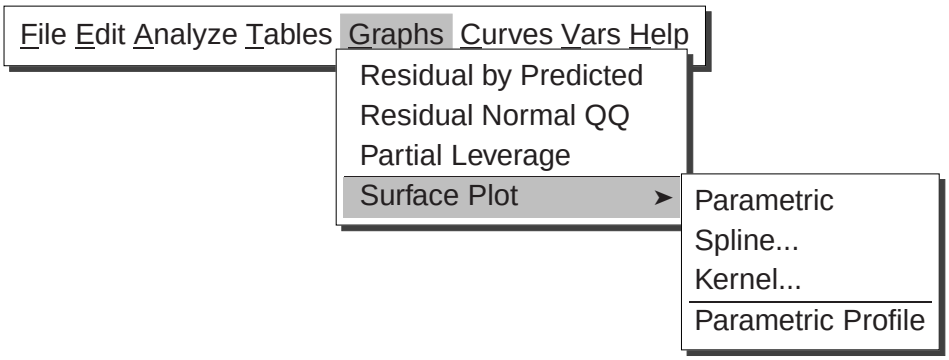


Figure 39.24. Graphs Menu

Residual-by-Predicted Plot

A residual-by-predicted plot is commonly used to diagnose nonlinearity or nonconstant error variance. It is also used to find outliers. A residual-by-predicted plot, as illustrated by the plot on the left in Figure 39.25, is a plot of residuals versus predicted response for each observation. See the "Predicted Values" and "Residuals" sections for a further explanation of the axis variables.

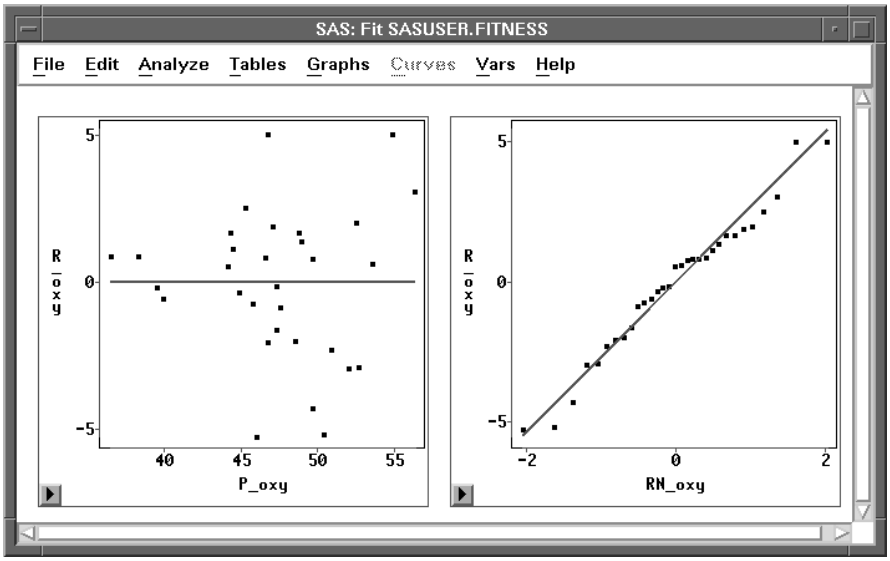


Figure 39.25. Residual-by-Predicted and Residual Normal QQ Plots

Residual Normal QQ Plot

A normal quantile-quantile plot of residuals is illustrated by the plot on the right in Figure 39.25. See the "Residual Normal Quantiles" section for an explanation of the **X** axis variable.

The empirical quantiles are plotted against the quantiles of a standard normal distribution. If the residuals are from a normal distribution with mean 0, the points tend to fall along the reference line that has an intercept of 0 and a slope equal to the estimated standard deviation.

Partial Leverage Plots

For linear models, the partial leverage plot for a selected explanatory variable can be obtained by plotting the residuals for the response variable against the residuals for the selected explanatory variable. The residuals for the response variable are calculated from a model having the selected explanatory variable omitted, and the residuals for the selected explanatory variable are calculated from a model where the selected explanatory variable is regressed on the remaining explanatory variables.

Let $\mathbf{X}_{[j]}$ be the $n \times (p-1)$ matrix formed from the design matrix $\mathbf{X}$ by removing the *j*th column, $\mathbf{X}_j$. Let $\mathbf{r}_{y[j]}$ be the partial leverage **Y** variable containing the residuals that result from regressing $\mathbf{y}$ on $\mathbf{X}_{[j]}$ and let $\mathbf{r}_{x[j]}$ be the partial leverage **X** variable containing the residuals that result from regressing $\mathbf{X}_j$ on $\mathbf{X}_{[j]}$. Then a partial leverage plot is a scatter plot of $\mathbf{r}_{y[j]}$ against $\mathbf{r}_{x[j]}$. Partial leverage plots for two explanatory variables are illustrated by Figure 39.26.

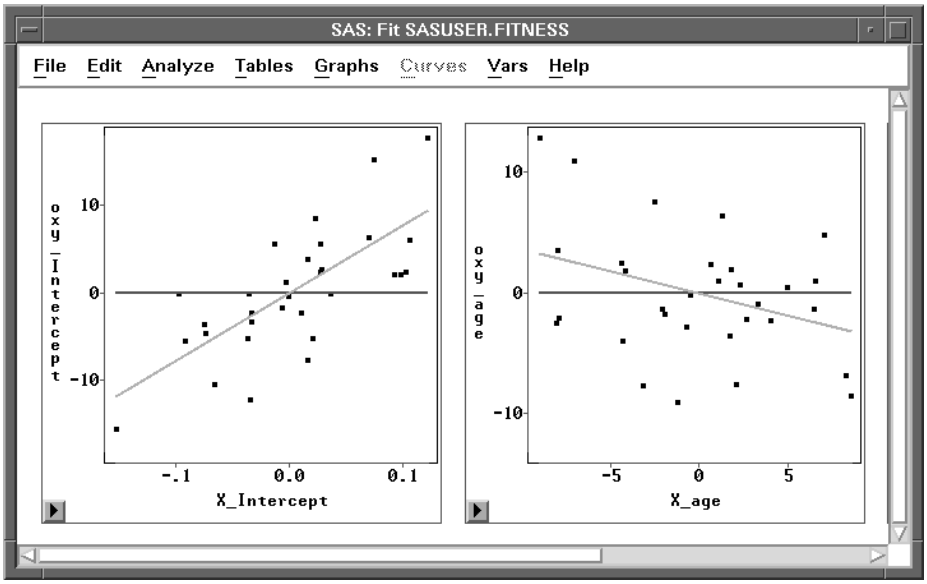


Figure 39.26. Partial Leverage Plots

In a partial leverage plot, the partial leverage **Y** variable $\mathbf{r}_{y[j]}$ can also be computed as

$$r_{y[j]i} = r_{x[j]i}b_j + (y_i - \hat{\mu}_i)$$

For generalized linear models, the partial leverage **Y** is also computed as

$$r_{y[j]i} = r_{x[j]i}b_j + (y_i - \hat{\mu}_i)g'(\hat{\mu}_i)$$

Two reference lines are also displayed in the plots. One is the horizontal line of **Y** = 0, and the other is the fitted regression of $\mathbf{r}_{y[j]}$ against $\mathbf{r}_{x[j]}$. The latter has an intercept of 0 and a slope equal to the parameter estimate associated with the explanatory variable in the model. The leverage plot shows the changes in the residuals for the model with and without the explanatory variable. For a given data point in the plot, its residual without the explanatory variable is the vertical distance between the point and the horizontal line; its residual with the explanatory variable is the vertical distance between the point and the fitted line.

Parametric Surface Plot

With two explanatory interval variables in the model, a parametric surface plot is a continuous surface plot of the predicted responses from the fitted parametric model on a set of regular grids of the explanatory variables. Figure 39.27 shows a response surface plot of **oxy** as a quadratic function of **age** and **weight**.

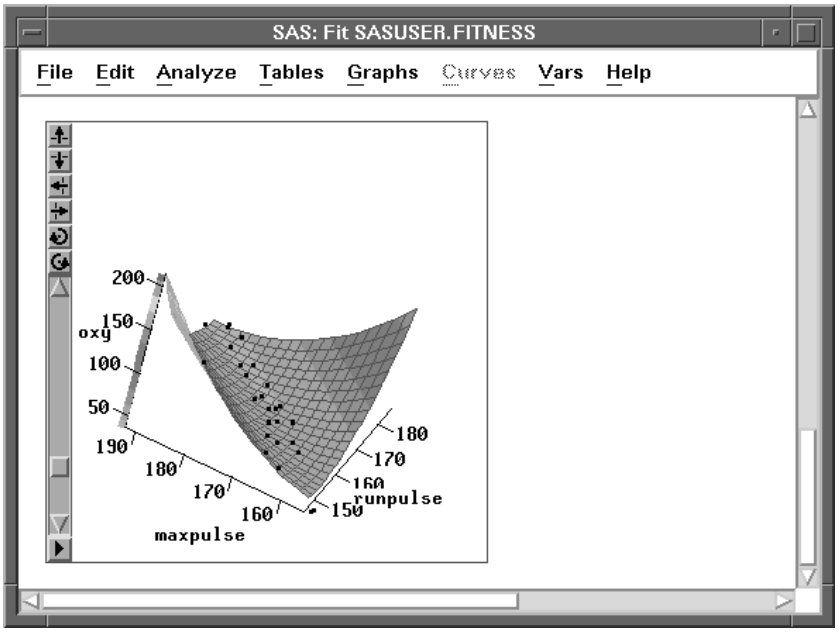


Figure 39.27. Parametric Surface Plot

The response surface is displayed with options **Drawing Modes:Smooth Color** and **Axes:Three Sections**.

Smoothing Spline Surface Plot

Two criteria can be used to select an estimator $\widehat{f_\lambda}$ for the function f:

- goodness of fit to the data
- smoothness of the fit

A standard measure of goodness of fit is the mean residual sum of squares

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{f_\lambda}(\mathbf{x}_i))^2$$

A measure of the smoothness of a fit is an integrated squared second derivative

$$J_2(f_\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}((\frac{\partial^2 f_\lambda}{\partial x_1^2})^2 + 2(\frac{\partial^2 f_\lambda}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 f_\lambda}{\partial x_2^2})^2)dx_1 dx_2$$

A single criterion that combines the two criteria is then given by

$$S(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{f_\lambda}(\mathbf{x}_i))^2 + \lambda J_2(f_\lambda)$$

where $\widehat{f_\lambda}$ belongs to the set of all continuously differentiable functions with square integrable second derivatives, and λ is a positive constant.

The estimator that results from minimizing $S(\lambda)$ is called a *thin-plate smoothing spline estimator*. Wahba and Wendelberger (1980) derived a closed form expression for the thin-plate smoothing spline estimator.

† **Note:** The computations for a thin-plate smoothing spline are time intensive, especially for large data sets.

The smoothing parameter λ controls the amount of smoothing; that is, it controls the trade-off between the goodness of fit to the data and the smoothness of the fit. You select a smoothing parameter λ by specifying a constant c in the formula

$$\lambda = c/100$$

The values of the explanatory variables are scaled by their corresponding interquartile ranges before the computations. This makes the computations independent of the units of X_1 and X_2.

After choosing **Graphs:Surface Plot:Spline** from the menu, you specify a smoothing parameter selection method in the **Spline Fit** dialog.

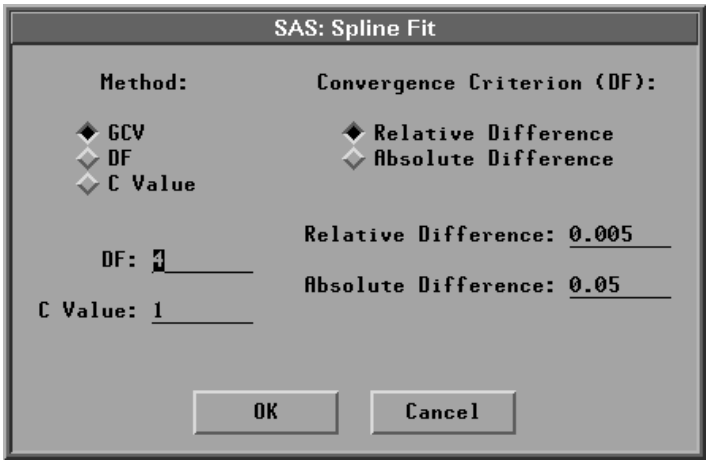


Figure 39.28. Spline Surface Fit Dialog

The default **Method:GCV** uses a *c* value that minimizes the generalized cross validation mean squared error $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$. Figure 39.29 displays smoothing spline estimates with *c* values of 0.0000831 (the GCV value) and 0.4127 (DF=6). Use the slider in the table to change the *c* value of the spline fit.

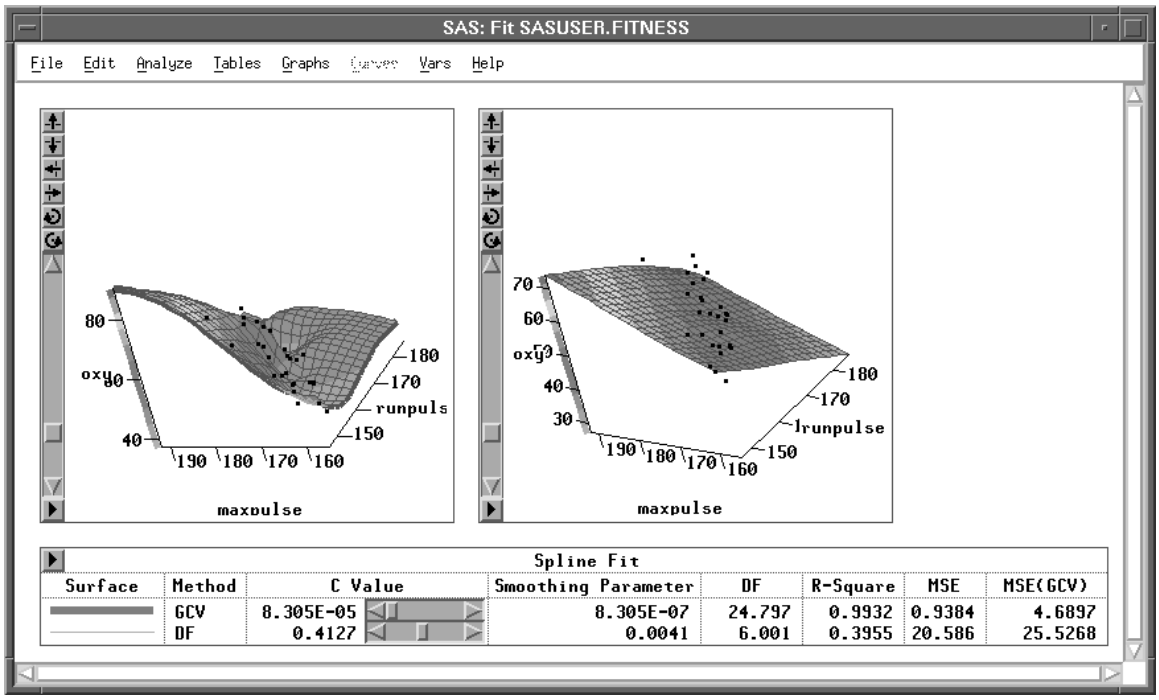


Figure 39.29. Smoothing Spline Surface Plot

Kernel Surface Plot

A *kernel estimator* uses an explicitly defined set of weights at each point **x** to produce the estimate at **x**. The kernel estimator of f has the form

$$\widehat{f_\lambda}(\mathbf{x}) = \sum_{i=1}^{n} W(\mathbf{x}, \mathbf{x}_i; \lambda, \mathbf{V_x}) y_i$$

where $\mathrm{W}(\mathbf{x}, \mathbf{x}_i; \lambda, \mathbf{V_x})$ is the weight function that depends on the smoothing parameter λ and the diagonal matrix $\mathbf{V_x}$ of the squares of the sample interquartile ranges.

The weights are derived from a single function that is independent of the design

$$W(\mathbf{x}, \mathbf{x}_i; \lambda, \mathbf{V_x}) = \frac{K_0((\mathbf{x} - \mathbf{x}_i)/\lambda, \mathbf{V_x})}{\sum_{j=1}^{n} K_0((\mathbf{x} - \mathbf{x}_j)/\lambda, \mathbf{V_x})}$$

where K_0 is a kernel function and λ is the bandwidth or smoothing parameter. The weights are nonnegative and sum to 1.

Symmetric probability density functions commonly used as kernel functions are

- $Normal$ $\quad K_0(\mathbf{t}, \mathbf{V}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\mathbf{t}'\mathbf{V}^{-1}\mathbf{t})$ for all t
- $Quadratic$ $\quad K_0(\mathbf{t}, \mathbf{V}) = \begin{cases} \frac{2}{\pi}(1 - \mathbf{t}'\mathbf{V}^{-1}\mathbf{t}) & \text{for } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} \leq 1 \\ 0 & \text{otherwise} \end{cases}$
- $Biweight$ $\quad K_0(\mathbf{t}, \mathbf{V}) = \begin{cases} \frac{3}{\pi}(1 - \mathbf{t}'\mathbf{V}^{-1}\mathbf{t})^2 & \text{for } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} \leq 1 \\ 0 & \text{otherwise} \end{cases}$
- $Triweight$ $\quad K_0(\mathbf{t}, \mathbf{V}) = \begin{cases} \frac{4}{\pi}(1 - \mathbf{t}'\mathbf{V}^{-1}\mathbf{t})^3 & \text{for } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} \leq 1 \\ 0 & \text{otherwise} \end{cases}$

You select a bandwidth λ for each kernel estimator by specifying *c* in the formula

$$\lambda = n^{-\frac{1}{6}} c$$

where n is the sample size. Both λ and *c* are independent of the units of **X**.

SAS/INSIGHT software divides the range of each explanatory variable into a number of evenly spaced intervals, then estimates the kernel fit on this grid. For a data point $\mathbf{x}_i$ that lies between two grid points, a linear interpolation is used to compute the predicted value. For $\mathbf{x}_i$ that lies inside a square of grid points, a pair of points that lie on the same vertical line as $\mathbf{x}_i$ and each lying between two grid points can be found. A linear interpolation of these two points is used to compute the predicted value.

After choosing **Graphs:Surface Plot:Kernel** from the menu, you specify a kernel and smoothing parameter selection method in the **Kernel Fit** dialog.

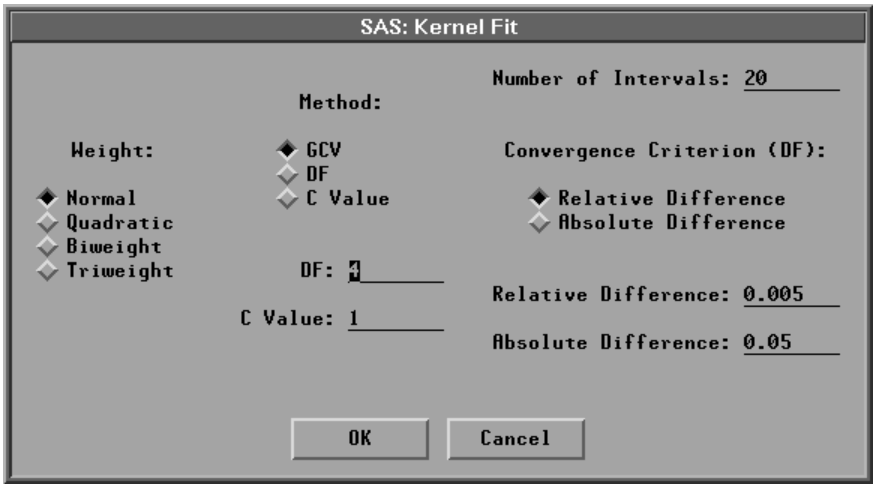


Figure 39.30. Kernel Surface Fit Dialog

By default, SAS/INSIGHT software divides the range of each explanatory variable into 20 evenly spaced intervals, uses a normal weight, and uses a *c* value that minimizes $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$. Figure 39.31 illustrates normal kernel estimates with *c* values of 0.5435 (the GCV value) and 1.0. Use the slider to change the *c* value of the kernel fit.

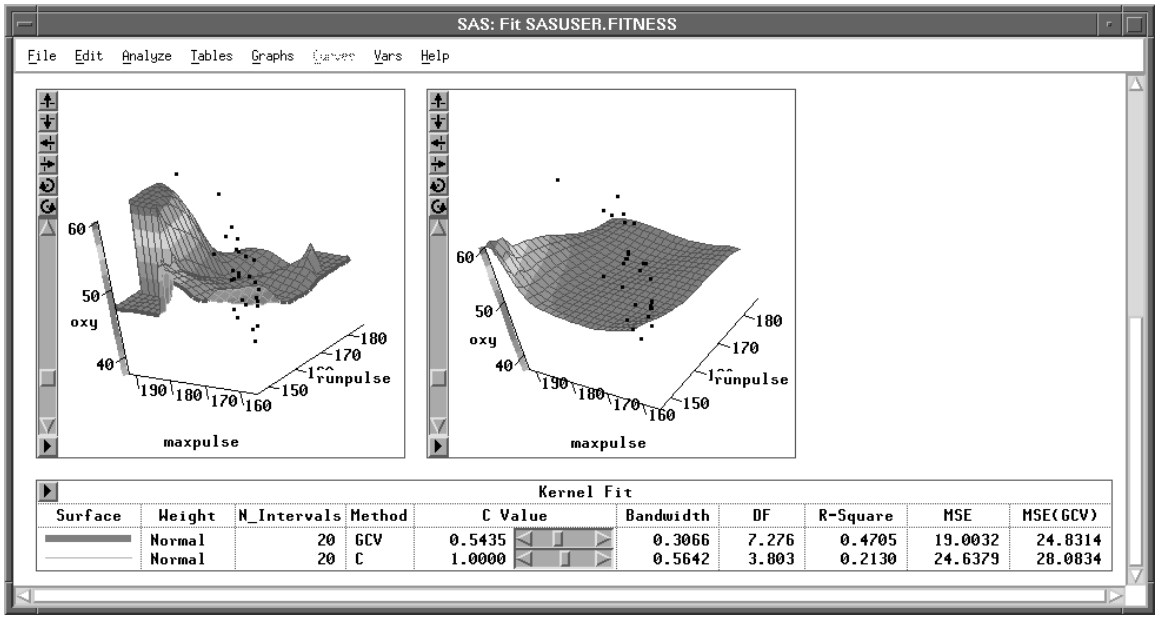


Figure 39.31. Kernel Surface Plot

Parametric Profile Surface Plot

With more than two explanatory interval variables in the model, a parametric profile surface plot is a continuous surface plot of the predicted responses from the fitted parametric model on a set of regular grids of a pair of explanatory variables. The values of the remaining explanatory variables are initially set at their means and can be adjusted with the sliders.

By default, the first two explanatory variables are used in the surface plot. You can also create profile surface plots for other explanatory variables by selecting the two variables before choosing **Graphs:Surface Plot:Parametric profile**. Figure 39.32 shows a parametric profile surface plot of **oxy** as a quadratic function of **runpulse** and **maxpulse** with **rstpulse** = 53.4516.

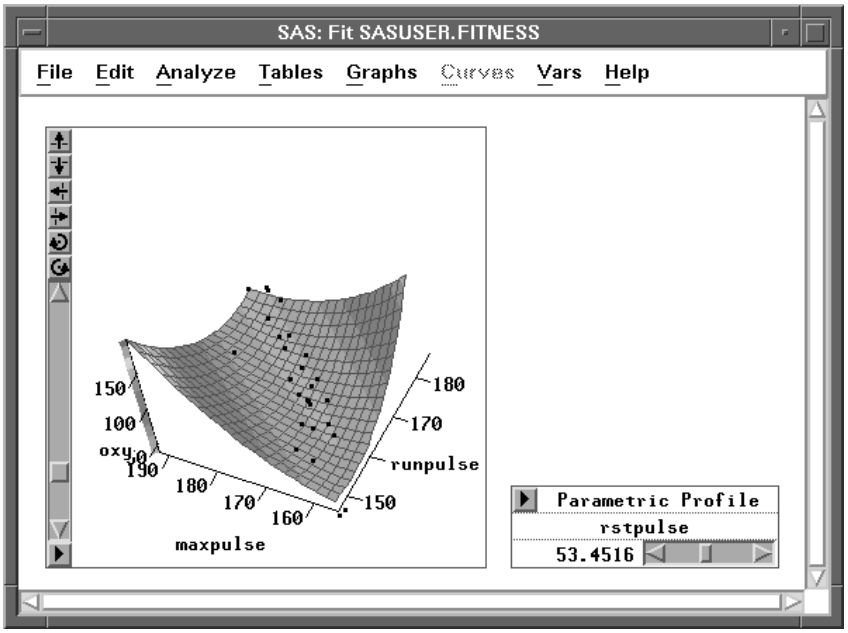


Figure 39.32. Parametric Profile Surface Plot

Fit Curves

When you are working with one explanatory variable, you can fit curves to the **Y**-by-**X** scatter plot generated when the analysis is first created. Use the output dialog (see Figure 39.4, Figure 39.6, and Figure 39.7) or the **Curves** menu in Figure 39.33 to fit curves to the scatter plot.

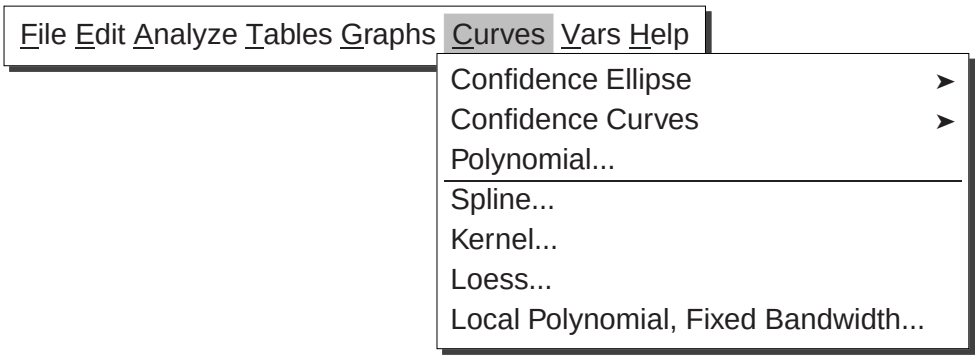


Figure 39.33. Curves Menu

There are two types of fitting techniques: parametric and nonparametric. Parametric techniques enable you to add confidence ellipses, fit regression polynomials, and add confidence curves of fitted polynomials to the **Y**-by-**X** scatter plot. Nonparametric techniques enable you to add spline, kernel, and local polynomial fits to the **Y**-by-**X** scatter plot.

Parametric Curves: Confidence Ellipses

SAS/INSIGHT software provides two types of confidence ellipses for each pair of **X** and **Y** variables assuming a bivariate normal distribution. One is a confidence ellipse for the population mean, and the other is a confidence ellipse for prediction.

Let $\overline{\mathbf{Z}}$ and $\mathbf{S}$ be the sample mean and the unbiased estimate of the covariance matrix of a random sample of size n from a bivariate normal distribution with mean μ and covariance matrix Σ.

The variable $\overline{\mathbf{Z}} - \mu$ is distributed as a bivariate normal variate with mean 0 and covariance $n^{-1}\mathbf{\Sigma}$, and it is independent of $\mathbf{S}$. The confidence ellipse for μ is based on Hotelling's T^2 statistic:

$$T^2 = n(\overline{\mathbf{Z}} - \mu)'\mathbf{S}^{-1}(\overline{\mathbf{Z}} - \mu)$$

A $100(1-\alpha)\%$ confidence ellipse for μ is defined by the equation

$$(\overline{\mathbf{Z}} - \mu)'\mathbf{S}^{-1}(\overline{\mathbf{Z}} - \mu) = \frac{2(n-1)}{n(n-2)}F_{2,n-2}(1-\alpha)$$

where $F_{2,n-2}(1-\alpha)$ is the $(1-\alpha)$ critical value of an F variate with degrees of freedom 2 and $n-2$.

A confidence ellipse for prediction is a confidence region for predicting a new observation in the population. It also approximates a region containing a specified percentage of the population.

Consider $\mathbf{Z}$ as a bivariate random variable for a new observation. The variable $\mathbf{Z} - \overline{\mathbf{Z}}$ is distributed as a bivariate normal variate with mean 0 and covariance $(1 + 1/n)\mathbf{\Sigma}$, and it is independent of $\mathbf{S}$.

A $100(1-\alpha)\%$ confidence ellipse for prediction is then given by the equation

$$(\mathbf{Z} - \overline{\mathbf{Z}})'\mathbf{S}^{-1}(\mathbf{Z} - \overline{\mathbf{Z}}) = \frac{2(n+1)(n-1)}{n(n-2)}F_{2,n-2}(1-\alpha)$$

The family of ellipses generated by different *F* critical values has a common center (the sample mean) and common major and minor axes.

The ellipses graphically indicate the correlation between two variables. When the variable axes are standardized (by dividing the variables by their respective standard deviations), the ratio of the two axis lengths (in Euclidean distances) reflects the magnitude of the correlation between the two variables. A ratio of 1 between the major and minor axes corresponds to a circular confidence contour and indicates that the variables are uncorrelated. A larger value of the ratio indicates a larger positive or negative correlation between the variables.

You can choose the level of the confidence region from the **Confidence Ellipse** menus, as illustrated by Figure 39.34.

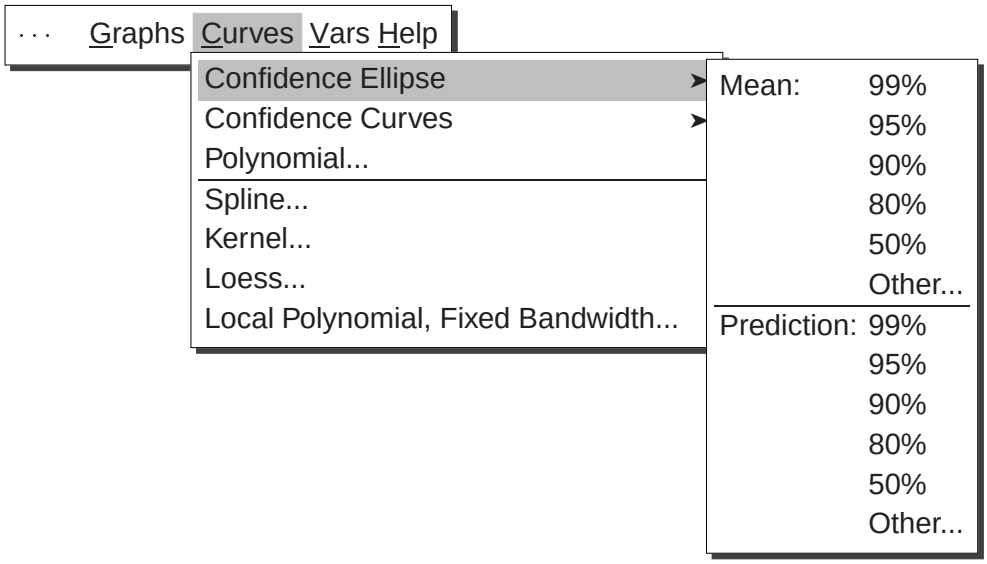


Figure 39.34. Confidence Ellipse Menu

A confidence ellipse for the population mean is displayed with dashed lines, and a confidence ellipse for prediction is displayed with dotted lines. Figure 39.35 displays a scatter plot with 50% and 80% confidence ellipses for prediction. Use the sliders in the **Confidence Ellipses** table to change the coefficient of the confidence ellipses.

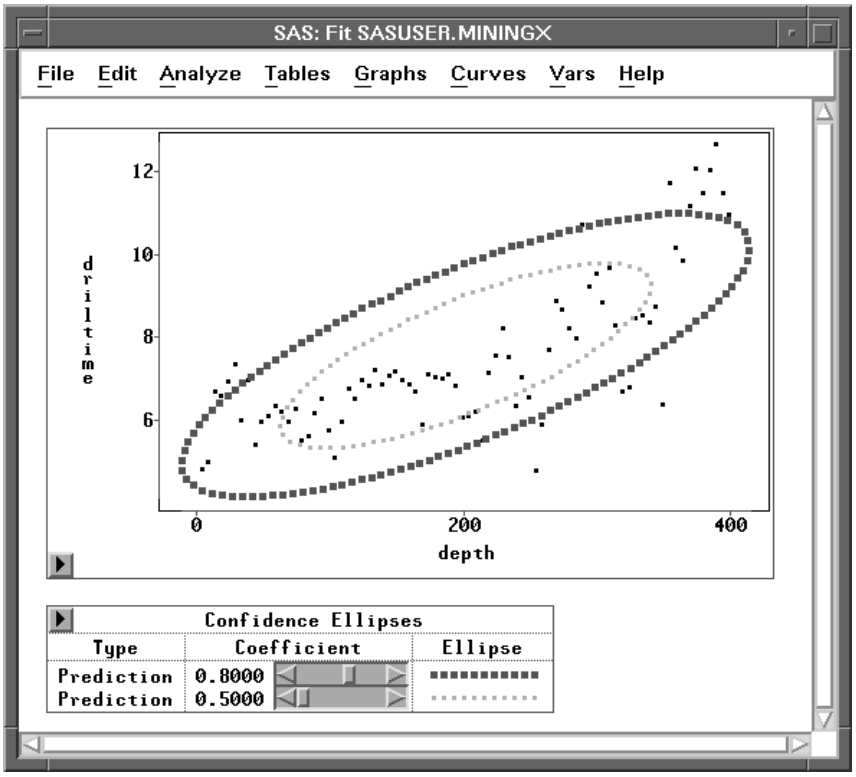


Figure 39.35. Confidence Ellipses for Prediction

Parametric Curves: Polynomial

Choose **Curves:Polynomial** from the menu to add a polynomial regression fit to the **Y**-by-**X** scatter plot. This displays the Polynomial Fit dialog in Figure 39.36.

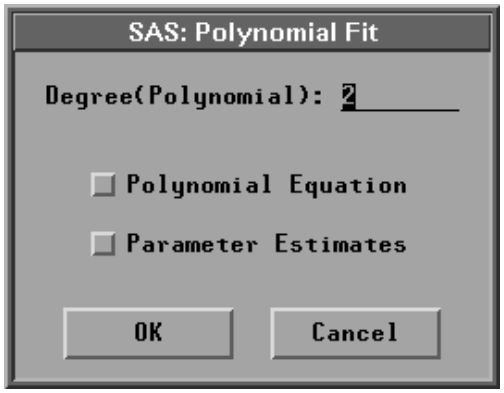


Figure 39.36. Polynomial Fit Dialog

In the **Polynomial Fit** dialog, you enter the degree for the polynomial fit. Select the **Polynomial Equation** or **Parameter Estimates** options to create a **Polynomial Equation** or **Parameter Estimates** table for the fitted curve.

Information about the polynomial fit is displayed in a table, as illustrated by Figure 39.37 The information includes the R^2 value and an F statistic and its associated p-value for testing the null hypothesis that all parameters are 0 except for the intercept. A parametric regression fit table includes the following:

Curve	is the curve in the **Y**-by-**X** scatter plot.
Degree(Polynomial)	is the degree for the polynomial fit.
Model DF	is the degrees of freedom for model.
Model Mean Square	is the mean square for model.
Error DF	is the degrees of freedom for error.
Error Mean Square	is the mean square for error.
R-Square	is the proportion of the (corrected) total variation attributed to the fit.
F Stat	is the F statistic for testing the null hypothesis that all parameters are zero except for the intercept. This is formed by dividing the mean square for model by the mean square for error.
Pr > F	is the probability under the null hypothesis of obtaining a greater F statistic than that observed.

Figure 39.37 displays a quadratic polynomial fit with **Polynomial Equation** and **Parameter Estimates** tables.

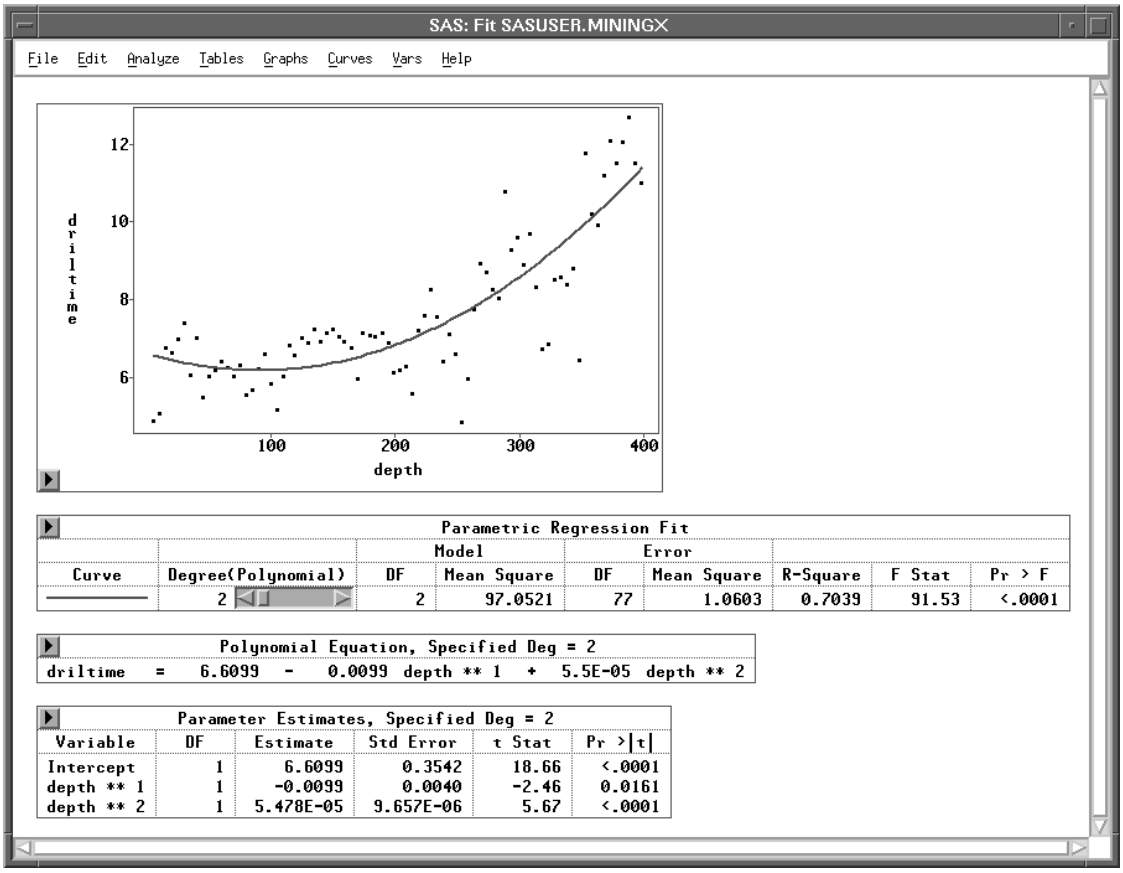


Figure 39.37. A Quadratic Polynomial Fit

You can use the **Degree(Polynomial)** slider in the **Parametric Regression Fit** table to change the degree of the polynomial curve fit. However, these will not change the **Polynomial Equation** and **Parameter Estimates** tables. You can produce a new **Polynomial Equation** or **Parameter Estimates** table by selecting the **Polynomial Equation** or **Parameter Estimates** option from the **Polynomial Fit** dialog.

Parametric Curves: Confidence Curves

You can add two types of confidence curves for the predicted values. One curve is for the mean value of the response, and the other is for the prediction of a new observation.

For the ith observation, a confidence interval that covers the expected value of the response with probability $1 - \alpha$ has upper and lower limits

$$\mathbf{x}_i\mathbf{b} \pm t_{(1-\alpha/2)}\sqrt{h_i}s$$

where $t_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ critical value of the Student's t statistic with degrees of freedom equal to the degrees of freedom for the mean squared error and h_i is the ith diagonal element of the hat matrix $\mathbf{H}$. The hat matrix $\mathbf{H}$ is described in the section "Output Variables" later in this chapter.

The $100(1 - \alpha)\%$ upper and lower limits for prediction are

$$\mathbf{x}_i\mathbf{b} \pm t_{(1-\alpha/2)}\sqrt{1 + h_i}s$$

You can generate confidence curves for a parametric regression fit by choosing the confidence coefficient from the **Curves:Confidence Curves** menu.

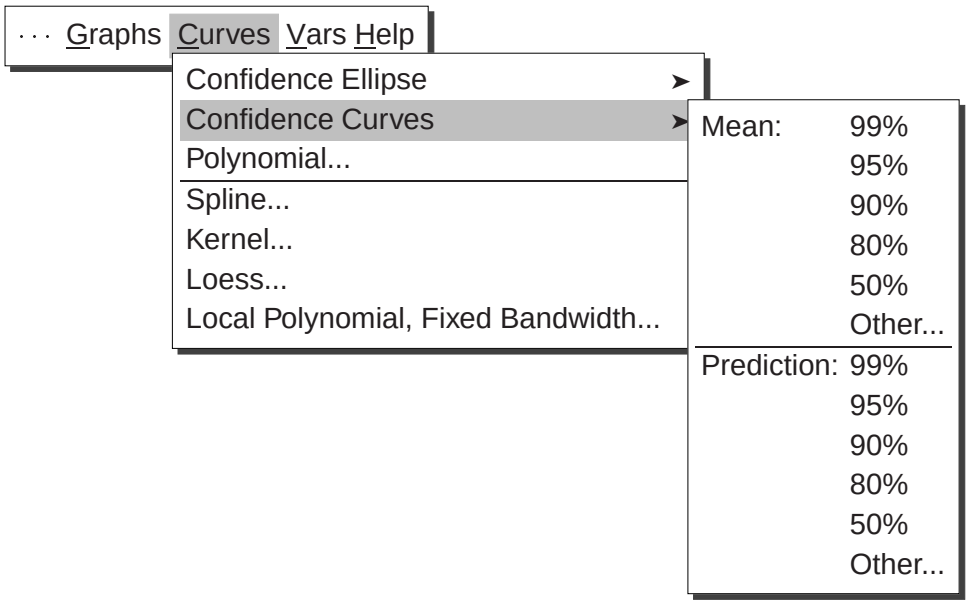


Figure 39.38. Confidence Curves Menu

Figure 39.39 displays a quadratic polynomial fit with 95% mean confidence curves for the response. Use the **Coefficient** slider to change the confidence coefficient.

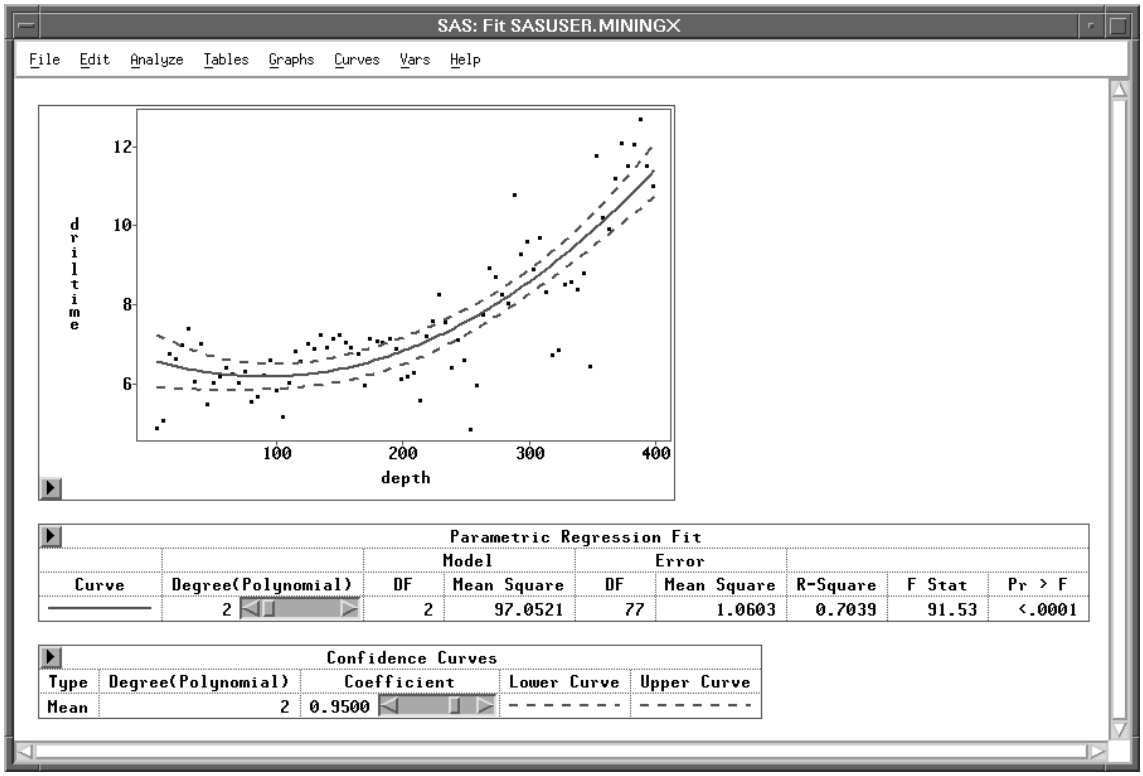


Figure 39.39. A Quadratic Polynomial Fit with 99% Mean Confidence Curves

Nonparametric Smoothing Spline

Two criteria can be used to select an estimator $\widehat{f_\lambda}$ for the function f:

- goodness of fit to the data
- smoothness of the fit

A standard measure of goodness of fit is the mean residual sum of squares

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{f_\lambda}(x_i))^2$$

A measure of the smoothness of a fit is the integrated squared second derivative

$$\int_{-\infty}^{\infty}(\widehat{f_\lambda}''(x))^2 dx$$

A single criterion that combines the two criteria is then given by

$$S(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{f_\lambda}(x_i))^2 + \lambda\int_{-\infty}^{\infty}(\widehat{f_\lambda}''(x))^2 dx$$

where $\widehat{f_\lambda}$ belongs to the set of all continuously differentiable functions with square integrable second derivatives, and λ is a positive constant.

The estimator that results from minimizing $S(\lambda)$ is called the *smoothing spline estimator*. This estimator fits a cubic polynomial in each interval between points. At each point x_i, the curve and its first two derivatives are continuous (Reinsch 1967).

The smoothing parameter λ controls the amount of smoothing; that is, it controls the trade-off between the goodness of fit to the data and the smoothness of the fit. You select a smoothing parameter λ by specifying a constant c in the formula

$$\lambda = (Q/10)^3 c$$

where Q is the interquartile range of the explanatory variable. This formulation makes c independent of the units of **X**.

After choosing **Curves:Spline**, you specify a smoothing parameter selection method in the **Spline Fit** dialog.

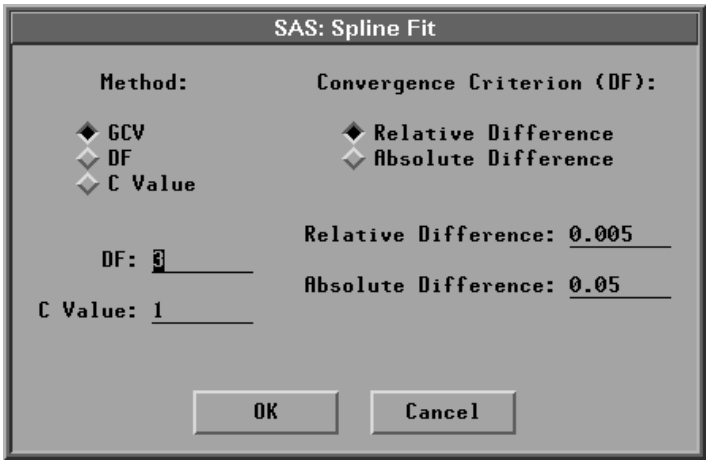


Figure 39.40. Spline Fit Dialog

The default **Method:GCV** uses a *c* value that minimizes the generalized cross validation mean squared error $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$. Figure 39.41 displays smoothing spline estimates with *c* values of 0.0017 (the GCV value) and 15.2219 (DF=3). Use the slider in the table to change the *c* value of the spline fit.

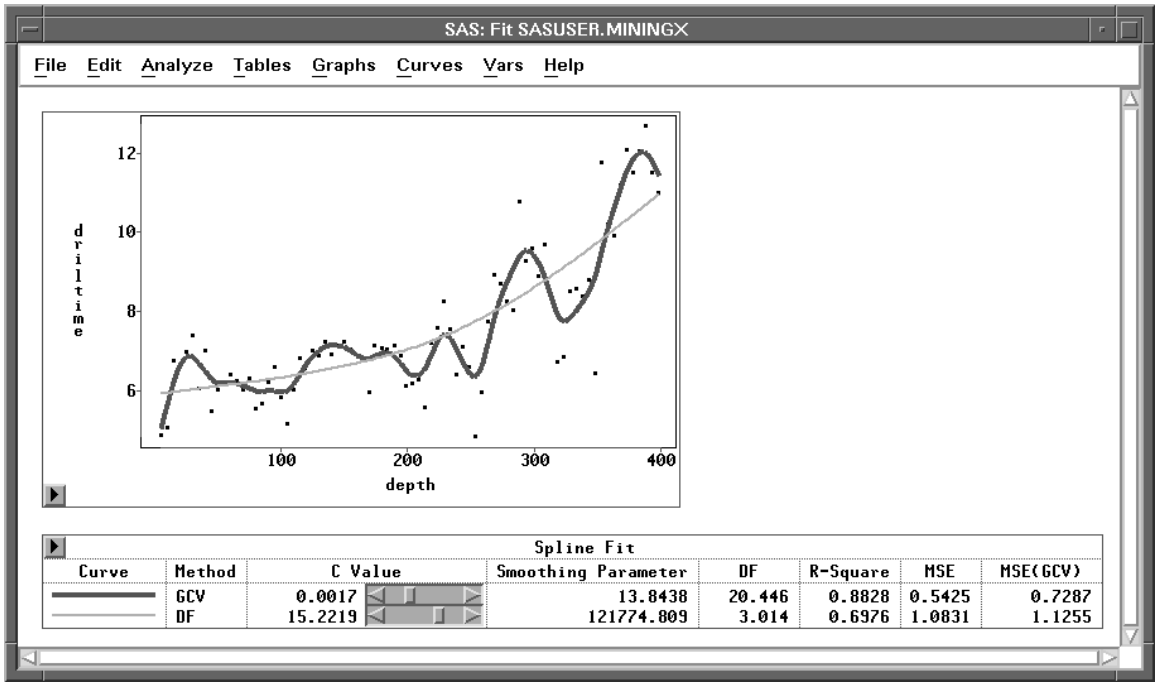


Figure 39.41. Smoothing Spline Estimates

Nonparametric Kernel Smoother

A *kernel estimator* uses an explicitly defined set of weights at each point *x* to produce the estimate at *x*. The kernel estimator of f has the form

$$\widehat{f_\lambda}(x) = \sum_{i=1}^{n} W(x, x_i; \lambda) y_i$$

where W$(x, x_i; \lambda)$ is the weight function that depends on the smoothing parameter λ.

The weights are derived from a single function that is independent of the design

$$W(x, x_i; \lambda) = \frac{K_0(\frac{x - x_i}{\lambda})}{\sum_{j=1}^{n} K_0(\frac{x - x_j}{\lambda})}$$

where K_0 is a kernel function and λ is the bandwidth or smoothing parameter. The weights are nonnegative and sum to 1.

Symmetric probability density functions commonly used as kernel functions are

- $Normal$ $\quad K_0(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$ $\quad$ for $-\infty < t < \infty$
- $Triangular$ $\quad K_0(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases}$
- $Quadratic$ $\quad K_0(t) = \begin{cases} \frac{3}{4}(1 - t^2) & \text{for } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases}$

You select a bandwidth λ for each kernel estimator by specifying *c* in the formula

$$\lambda = n^{-\frac{1}{5}} Q c$$

where *Q* is the sample interquartile range of the explanatory variable and n is the sample size. This formulation makes *c* independent of the units of **X**.

SAS/INSIGHT software divides the range of the explanatory variable into 128 evenly spaced intervals, then approximates the data on this grid and uses the fast Fourier transformation (Silverman 1986) to estimate the kernel fit on this grid. For a data point x_i that lies between two grid points, a linear interpolation is used to compute the predicted value. A small value of λ (relative to the width of the interval) may give unstable estimates of the kernel fit.

After choosing **Curves:Kernel**, you specify a kernel and smoothing parameter selection method in the **Kernel Fit** dialog.

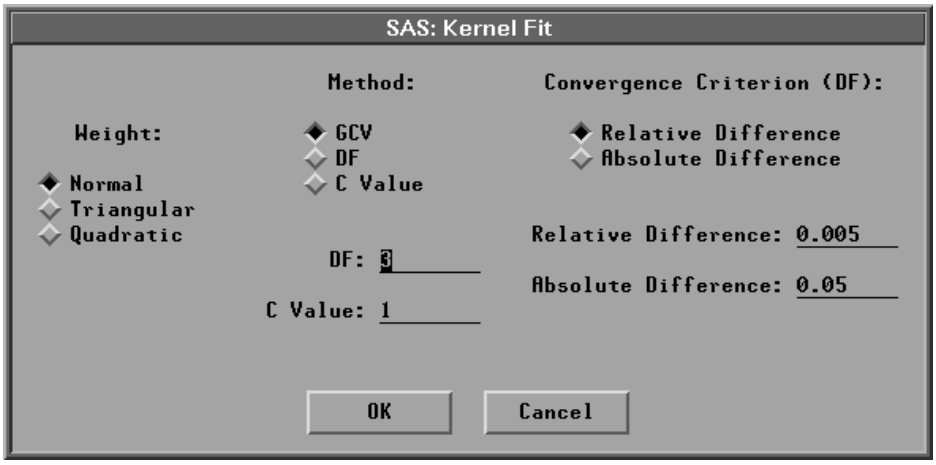


Figure 39.42. Kernel Fit Dialog

The default **Weight:Normal** uses a normal weight, and **Method:GCV** uses a *c* value that minimizes $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$. Figure 39.43 illustrates normal kernel estimates with *c* values of 0.0944 (the GCV value) and 0.7546 (DF=3). Use the slider to change the *c* value of the kernel fit.

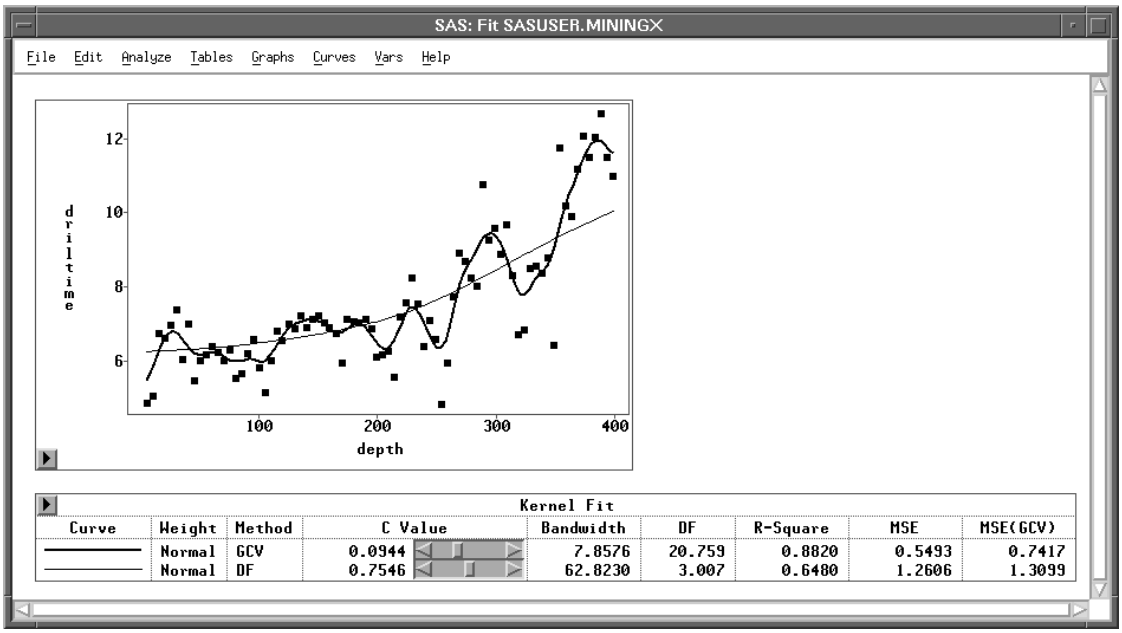


Figure 39.43. Kernel Estimates

Nonparametric Local Polynomial Smoother

The kernel estimator fits a local mean at each point *x* and thus cannot even estimate a line without bias (Cleveland, Cleveland, Devlin and Grosse 1988). An estimator based on locally-weighted regression lines or locally-weighted quadratic polynomials may give more satisfactory results.

A local polynomial smoother fits a locally-weighted regression at each point *x* to produce the estimate at *x*. Different types of regression and weight functions are used in the estimation.

SAS/INSIGHT software provides the following three types of regression:

- $Mean$ a locally-weighted mean
- $Linear$ a locally-weighted regression line
- $Quadratic$ a locally-weighted quadratic polynomial regression

The weights are derived from a single function that is independent of the design

$$W(x, x_i; \lambda_i) = K_0(\frac{x - x_i}{\lambda_i})$$

where K_0 is a weight function and λ_i is the local bandwidth at x_i.

SAS/INSIGHT software uses the following weight functions:

- $Normal$ $K_0(t) = \begin{cases} \exp(-t^2/2) & \text{for } |t| \leq 3.5 \\ 0 & \text{otherwise} \end{cases}$
- $Triangular$ $K_0(t) = \begin{cases} 1 - |t| & \text{for } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases}$
- $Quadratic$ $K_0(t) = \begin{cases} 1 - t^2 & \text{for } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases}$
- $Tri - Cube$ $K_0(t) = \begin{cases} (1 - |t|^3)^3 & \text{for } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases}$

† **Note:** The normal weight function is proportional to a truncated normal density function.

SAS/INSIGHT software provides two methods to compute the local bandwidth λ_i. The loess estimator (Cleveland 1979; Cleveland, Devlin and Grosse 1988) evaluates λ_i based on the furthest distance from k nearest neighbors. A fixed bandwidth local polynomial estimator uses a constant bandwidth λ at each x_i.

For a loess estimator, you select k nearest neighbors by specifying a positive constant α. For $\alpha \leq 1$, k is αn truncated to an integer, where n is the number of observations. For $\alpha > 1$, k is set to n.

The local bandwidth λ_i is then computed as

$$\lambda_i = \begin{cases} d_{(k)}(x_i) & \text{for } 0 < \alpha \leq 1 \\ \alpha d_{(n)}(x_i) & \text{for } \alpha > 1 \end{cases}$$

where $d_{(k)}(x_i)$ is the furthest distance from x_i to its k nearest neighbors.

† **Note:** For $\alpha \leq 1$, the local bandwidth λ_i is a function of k and thus a step function of α.

For a fixed bandwidth local polynomial estimator, you select a bandwidth λ by specifying *c* in the formula

$$\lambda = n^{-\frac{1}{5}} Q c$$

where Q is the sample interquartile range of the explanatory variable and n is the sample size. This formulation makes *c* independent of the units of **X**.

† **Note:** A fixed bandwidth local mean estimator is equivalent to a kernel smoother.

By default, SAS/INSIGHT software divides the range of the explanatory variable into 128 evenly spaced intervals, then it fits locally-weighted regressions on this grid. A small value of *c* or α may give the local polynomial fit to the data points near the grid points only and may not apply to the remaining points.

For a data point x_i that lies between two grid points $x_{i[j]} \leq x_i < x_{i[j+1]}$, the predicted value is the weighted average of the two predicted values at the two nearest grid points:

$$(1 - d_{ij})\hat{y}_{i[j]} + d_{ij}\hat{y}_{i[j+1]}$$

where $\hat{y}_{i[j]}$ and $\hat{y}_{i[j+1]}$ are the predicted values at the two nearest grid points and

$$d_{ij} = \frac{x_i - x_{i[j]}}{x_{i[j+1]} - x_{i[j]}}$$

A similar algorithm is used to compute the degrees of freedom of a local polynomial estimate, df_λ = trace($\mathbf{H}_\lambda$). The ith diagonal element of the matrix $\mathbf{H}_\lambda$ is

$$(1 - d_{ij})h_{i[j]} + d_{ij}h_{i[j+1]}$$

where $h_{i[j]}$ and $h_{i[j+1]}$ are the ith diagonal elements of the projection matrices of the two regression fits.

After choosing **Curves:Loess** from the menu, you specify a loess fit in the **Loess Fit** dialog.

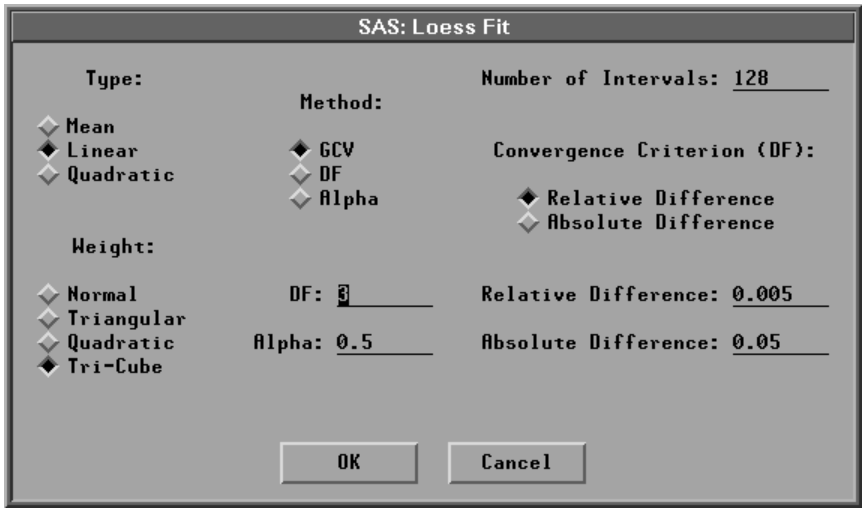


Figure 39.44. Loess Fit Dialog

In the dialog, you can specify the number of intervals, the regression type, the weight function, and the method for choosing the smoothing parameter. The default **Type:Linear** uses a linear regression, **Weight:Tri-Cube** uses a tri-cube weight function, and **Method:GCV** uses an α value that minimizes $\mathrm{MSE}_{\mathrm{GCV}}(\lambda)$.

Figure 39.45 illustrates loess estimates with **Type=Linear**, **Weight=Tri-Cube**, and α values of 0.0930 (the GCV value) and 0.7795 (DF=3). Use the slider to change the α value of the loess fit.

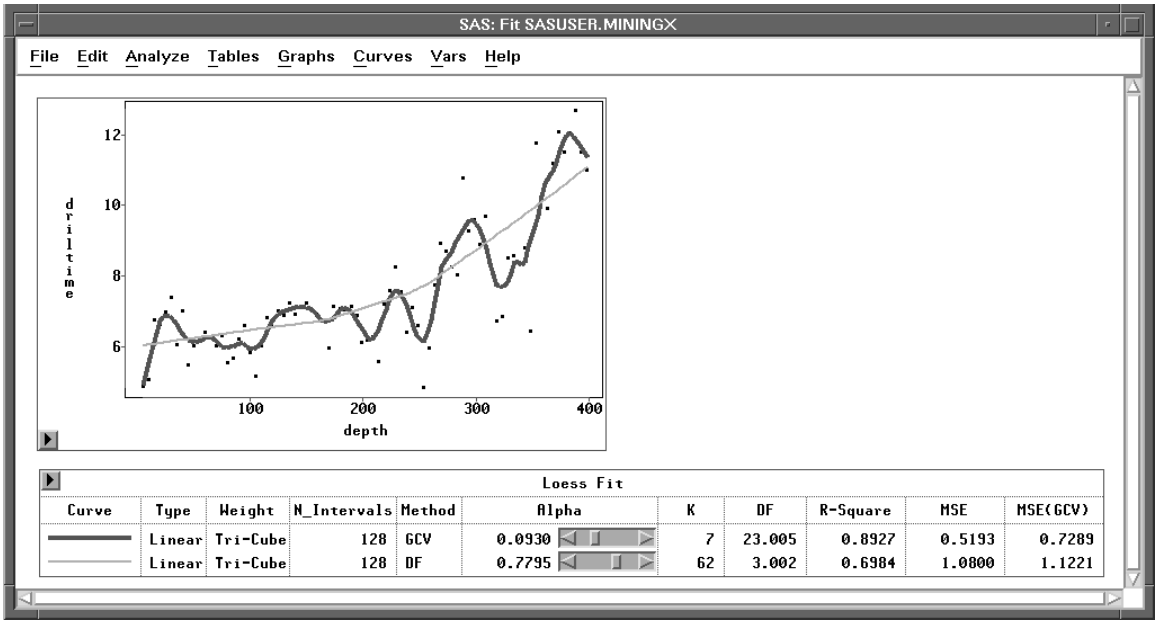


Figure 39.45. Loess Estimates

The loess degrees of freedom is a function of local bandwidth λ_i. For $\alpha \leq 1$, λ_i is a step function of α and thus the loess *df* is a step function of α. The convergence criterion applies only when the specified *df* is less than $df_{(\alpha=1)}$, the loess *df* for $\alpha = 1$. When the specified *df* is greater than $df_{(\alpha=1)}$, SAS/INSIGHT software uses the α value that has its *df* closest to the specified *df*.

Similarly, you can choose **Curves:Local Polynomial, Fixed Bandwidth** from the menu to specify a fixed bandwidth local polynomial fit.

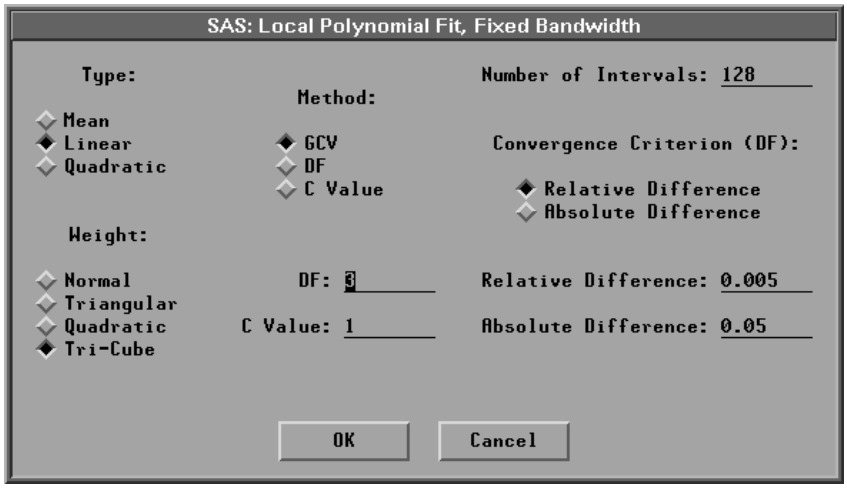


Figure 39.46. Fixed Bandwidth Local Polynomial Fit Dialog

Figure 39.47 illustrates fixed bandwidth local polynomial estimates with **Type=Linear**, **Weight=Tri-Cube**, and *c* values of 0.2026 (the GCV value) and 2.6505 (DF=3). Use the slider to change the *c* value of the local polynomial fit.

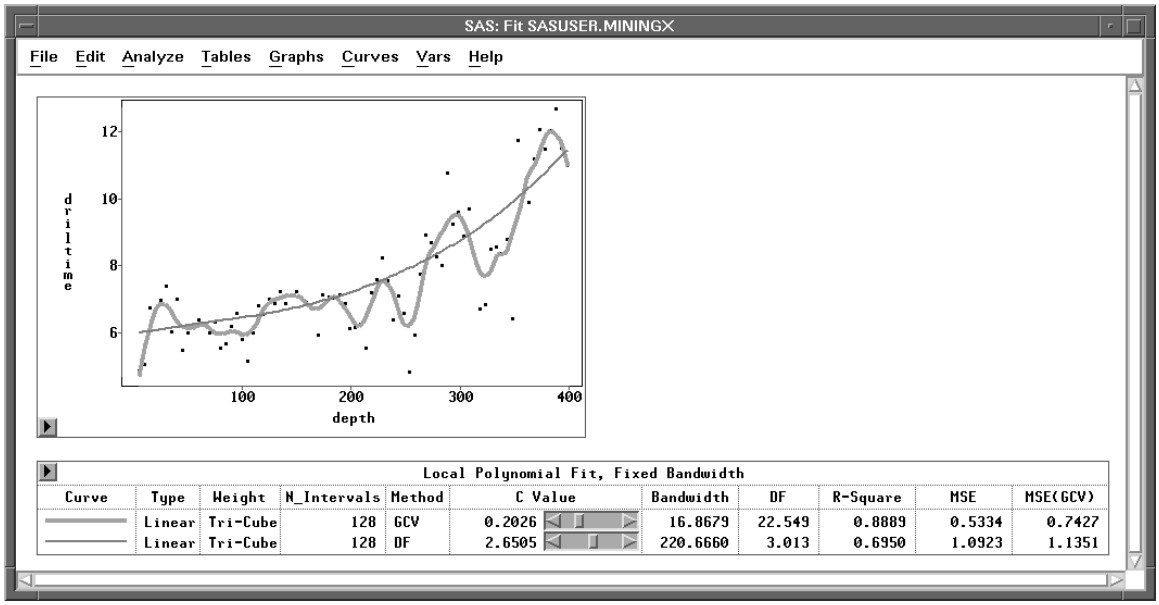


Figure 39.47. Fixed Bandwidth Local Polynomial Estimates

Output Variables

Output variables based on the model you fit can be saved in the data window. From the data window, you can store these variables in a SAS data set. This enables you, for example, to perform additional analyses using SAS/STAT software.

Axis variables in residual plots are automatically saved in the data window used to create the analysis. For example, when you create a residual-by-predicted plot, residual and predicted variables are always generated. These variables are deleted when you close the analysis window.

You can save variables permanently by using the fit output options dialog or the **Vars** menu shown in Figure 39.48. Such variables remain stored in the data window after you close the analysis window.

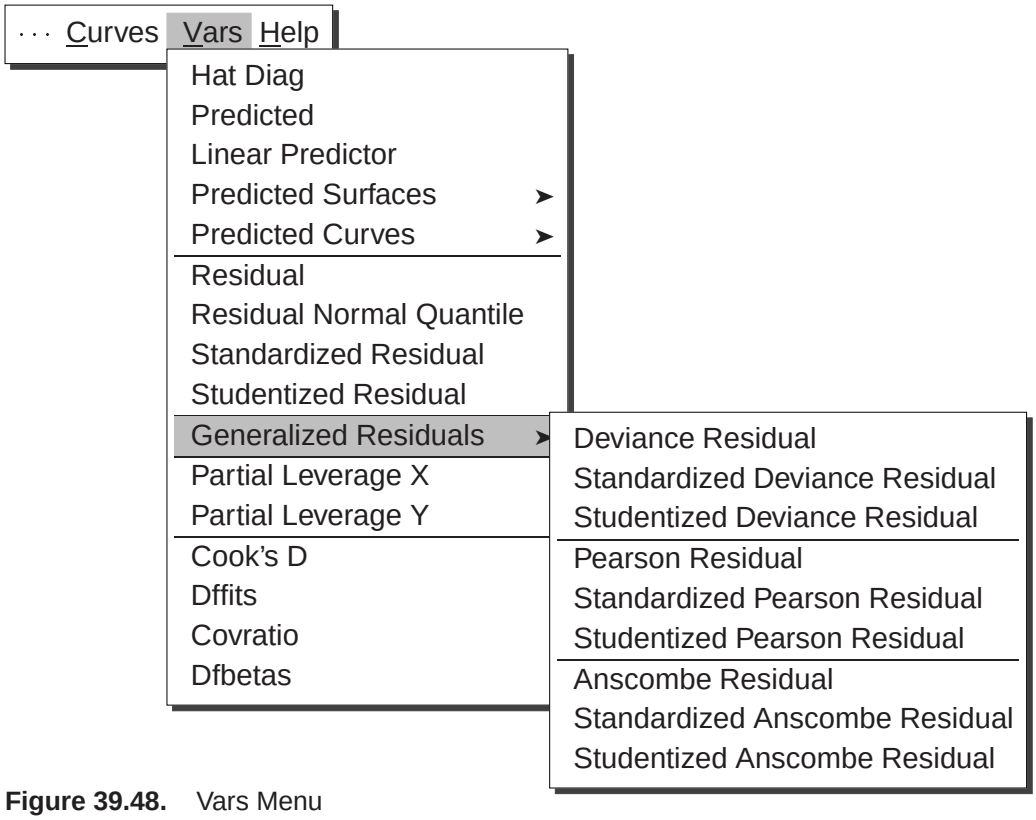


Figure 39.48. Vars Menu

SAS/INSIGHT software provides predicted and residual variables, a linear predictor, a residual normal quantile variable, partial leverage **X** and **Y** variables, and influence diagnostic variables.

Influence diagnostics are measures of the influence of each observation on the parameter estimates. These diagnostics include the hat diagonal values, standardized residuals, and studentized residuals. Cook's D, Dffits, Covratio, and Dfbetas also measure the effect of deleting observations.

Some influence diagnostics require a refit of the model after excluding each observation. For generalized linear models, numerical iterations are used for the fits, and

the process can be expensive. One-step methods are used to approximate these diagnostics after each fit. The process involves doing one iteration of the fit without the excluded observation, starting with the final parameter estimates and weights from the complete fit.

You can also create generalized residuals such as Pearson, deviance, and Anscombe residuals with generalized linear models. These residuals are applicable to the non-normal response distributions.

Generated variables use the naming conventions described later in this section. If a resulting variable name has more than 32 characters, only the first 32 characters are used. Generated variables also follow the same numbering convention as the analysis window when you create more than one fit analysis from the same data window. If the generated variable name is longer than 32 characters, the original variable name is truncated to the necessary length.

Hat Matrix Diagonal

Data points that are far from the centroid of the *X*-space are potentially influential. A measure of the distance between a data point, x_i, and the centroid of the *X*-space is the data point's associated diagonal element h_i in the hat matrix. Belsley, Kuh, and Welsch (1980) propose a cutoff of $2p/n$ for the diagonal elements of the hat matrix, where n is the number of observations used to fit the model, and p is the number of parameters in the model. Observations with h_i values above this cutoff should be investigated.

For linear models, the hat matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

can be used as a projection matrix. The hat matrix diagonal variable contains the diagonal elements of the hat matrix

$$h_i = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$$

For generalized linear models, an approximate projection matrix is given by

$$\mathbf{H} = \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{1/2}$$

where $\mathbf{W} = \mathbf{W}_o$ when the full Hessian is used and $\mathbf{W} = \mathbf{W}_e$ when Fisher's scoring method is used.

The values of h_i are stored in a variable named **H_yname**, where **yname** is the response variable name.

Predicted Values

After the model has been fit, the predicted values are calculated from the estimated regression equation.

For linear models, the predicted mean vector of the n observation responses is

$$\hat{\mu} = \mathbf{X}\mathbf{b} = \mathbf{H}\mathbf{y}$$

$$\hat{\mu_i} = \mathbf{x}_i\mathbf{b}$$

For generalized linear models,

$$\hat{\mu_i} = g^{-1}(\eta_{0i} + \mathbf{x}_i\mathbf{b})$$

where η_{0i} is the offset for the ith observation.

The predicted values are stored in variables named **P_yname** for each response variable, where **yname** is the response variable name.

Linear Predictor

The *linear predictor* values are the linear function values, $\mathbf{x}_i\mathbf{b}$, in the predicted values. The linear predictor values are stored in variables named **LP_yname** for each response variable, where **yname** is the response variable name.

Residuals

The *residuals* are calculated as actual response minus predicted value,

$$r_i = y_i - \hat{\mu_i}$$

The residuals are stored in variables named **R_yname** for each response variable, where **yname** is the response variable name.

Residual Normal Quantiles

The *normal quantile* of the ith ordered residual is computed as

$$\Phi^{-1}\left(\frac{i - 0.375}{n + 0.25}\right)$$

where Φ^{-1} is the inverse standard cumulative normal distribution.

If the residuals are normally distributed, the points on the residual normal quantile-quantile plot should lie approximately on a straight line with residual mean as the intercept and residual standard deviation as the slope.

The normal quantiles of the residuals are stored in variables named **RN_yname** for each response variable, where **yname** is the response variable name.

Predicted Surfaces

You can output predicted values from fitted kernel and thin-plate smoothing spline surfaces by choosing **Vars:Predicted Surfaces** from the menu.

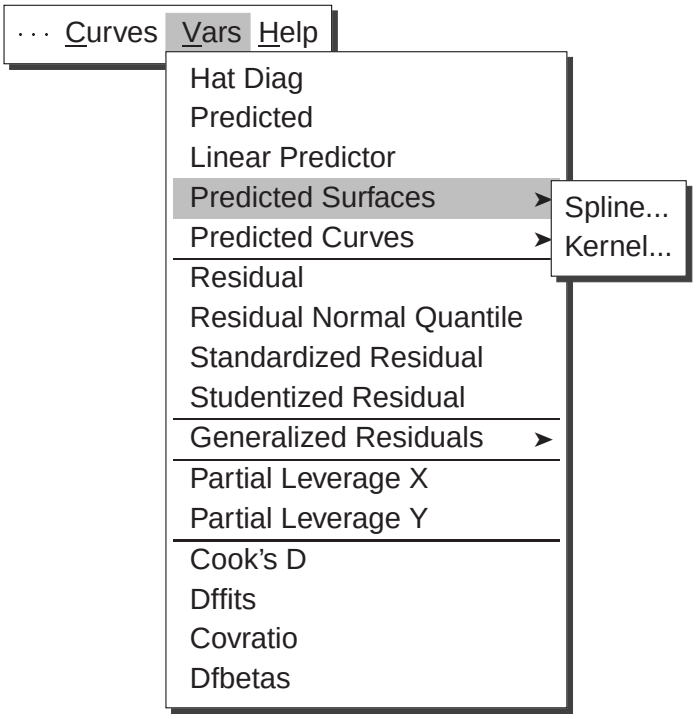


Figure 39.49. Predicted Surfaces Menu

For predicted values from a spline or kernel fit, you specify the surface fit in the dialogs, as shown in Figure 39.28 or Figure 39.30, respectively.

The predicted values for each response variable are stored in variables named **PS_yname** for spline and **PK_yname** for kernel, where **yname** is the response variable name.

Predicted Curves

You can output predicted values from fitted curves by choosing **Vars:Predicted Curves** from the menu.

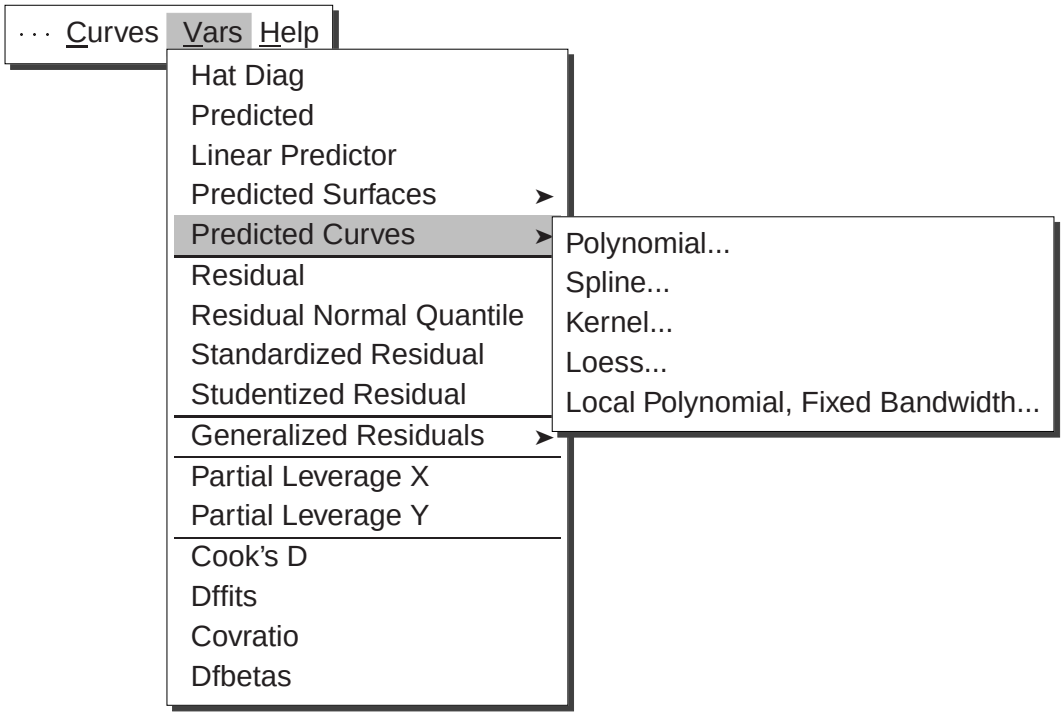


Figure 39.50. Predicted Curves Menu

After choosing **Vars:Predicted Curves:Polynomial** from the menu, you can specify the degree of polynomial in the **Polynomial Fit** dialog.

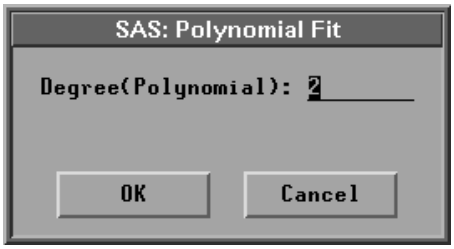


Figure 39.51. Predicted Polynomial Fit Dialog

For predicted values from a spline, kernel, loess, or fixed bandwidth local polynomial fit, you specify the curve fit in the dialogs, as shown in Figure 39.40, Figure 39.42, Figure 39.44, or Figure 39.46, respectively.

The predicted values for each response variable are stored in variables named **PP_yname** for polynomial, **PS_yname** for spline, **PK_yname** for kernel, and **PL_yname** for loess and fixed bandwidth local polynomial, where **yname** is the response variable name.

Standardized and Studentized Residuals

For linear models, the variance of the residual r_i is

$$\mathrm{Var(r_i)} = \sigma^2(1 - \mathrm{h_i})$$

and an estimate of the standard error of the residual is

$$\mathrm{STDERR(r_i)} = \mathrm{s}\sqrt{1 - \mathrm{h_i}}$$

Thus, the residuals can be modified to better detect unusual observations. The ratio of the residual to its standard error, called the *standardized residual*, is

$$r_{si} = \frac{r_i}{s\sqrt{1 - h_i}}$$

If the residual is standardized with an independent estimate of σ^2, the result has a Student's *t* distribution if the data satisfy the normality assumption. If you estimate σ^2 by $s^2_{(i)}$, the estimate of σ^2 obtained after deleting the ith observation, the result is a studentized residual:

$$r_{ti} = \frac{r_i}{s_{(i)}\sqrt{1 - h_i}}$$

Observations with $|r_{ti}| > 2$ may deserve investigation.

For generalized linear models, the standardized and studentized residuals are

$$r_{si} = \frac{r_i}{\sqrt{\hat{\phi}(1 - h_i)}}$$

$$r_{ti} = \frac{r_i}{\sqrt{\hat{\phi}_{(i)}(1 - h_i)}}$$

where $\hat{\phi}$ is the estimate of the dispersion parameter ϕ, and $\hat{\phi}_{(i)}$ is a one-step approximation of ϕ after excluding the ith observation.

The standardized residuals are stored in variables named **RS_yname** and the Studentized residuals are stored in variables named **RT_yname** for each response variable, where **yname** is the response variable name.

Deviance Residuals

The *deviance residual* is the measure of deviance contributed from each observation and is given by

$$r_{Di} = \mathrm{sign}(\mathrm{r_i})\sqrt{\mathrm{d_i}}$$

where d_i is the individual deviance contribution.

The deviance residuals can be used to check the model fit at each observation for generalized linear models. These residuals are stored in variables named **RD_yname** for each response variable, where **yname** is the response variable name.

The standardized and studentized deviance residuals are

$$r_{Dsi} = \frac{r_{Di}}{\sqrt{\hat{\phi}(1-h_i)}}$$

$$r_{Dti} = \frac{r_{Di}}{\sqrt{\hat{\phi}_{(i)}(1-h_i)}}$$

The standardized deviance residuals are stored in variables named **RDS_yname** and the studentized deviance residuals are stored in variables named **RDT_yname** for each response variable, where **yname** is the response variable name.

Pearson Residuals

The *Pearson residual* is the raw residual divided by the square root of the variance function $V(\mu)$.

The Pearson residual is the individual contribution to the Pearson χ^2 statistic. For a binomial distribution with m_i trials in the ith observation, it is defined as

$$r_{Pi} = \sqrt{m_i}\frac{r_i}{\sqrt{V(\hat{\mu}_i)}}$$

For other distributions, the Pearson residual is defined as

$$r_{Pi} = \frac{r_i}{\sqrt{V(\hat{\mu}_i)}}$$

The Pearson residuals can be used to check the model fit at each observation for generalized linear models. These residuals are stored in variables named **RP_yname** for each response variable, where **yname** is the response variable name.

The standardized and studentized Pearson residuals are

$$r_{Psi} = \frac{r_{Pi}}{\sqrt{\hat{\phi}(1-h_i)}}$$

$$r_{Pti} = \frac{r_{Pi}}{\sqrt{\hat{\phi}_{(i)}(1-h_i)}}$$

The standardized Pearson residuals are stored in variables named **RPS_yname** and the studentized Pearson residuals are stored in variables named **RPT_yname** for each response variable, where **yname** is the response variable name.

Anscombe Residuals

For nonnormal response distributions in generalized linear models, the distribution of the Pearson residuals is often skewed. Anscombe proposed a residual using a function $A(y)$ in place of *y* in the residual derivation (Anscombe 1953, McCullagh and Nelder 1989). The function $A(y)$ is chosen to make the distribution of $A(y)$ as normal as possible and is given by

$$A(\mu) = \int_{-\infty}^{\mu} V^{-1/3}(t)dt$$

where $V(t)$ is the variance function.

For a binomial distribution with m_i trials in the ith observation, the *Anscombe residual* is defined as

$$r_{Ai} = \sqrt{m_i}\frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}}$$

For other distributions, the Anscombe residual is defined as

$$r_{Ai} = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}}$$

where $A'(\mu)$ is the derivative of $A(\mu)$.

For the response distributions used in the fit analysis, Anscombe residuals are

Normal	$r_{Ai} = y_i - \hat{\mu}_i$
Inverse Gaussian	$r_{Ai} = (\log(y_i) - \log(\hat{\mu}_i))/{\hat{\mu}_i}^{1/2}$
Gamma	$r_{Ai} = 3((y_i/\hat{\mu}_i)^{1/3} - 1)$
Poisson	$r_{Ai} = \frac{3}{2}(y_i^{2/3}{\hat{\mu}_i}^{-1/6} - {\hat{\mu}_i}^{1/2})$
Binomial	$r_{Ai} = \sqrt{m_i}\left(B(y_i, \frac{2}{3}, \frac{2}{3}) - B(\hat{\mu}_i, \frac{2}{3}, \frac{2}{3})\right)(\hat{\mu}_i(1-\hat{\mu}_i))^{-1/6}$ where $B(z,a,b) = \int_0^z t^{a-1}(1-t)^{b-1}\,dt$

You can save Anscombe residuals to your data set by using the **Output Variables** dialog, as shown in Figure 39.5, or the **Vars** menu, as shown in Figure 39.48. These residuals are stored in variables named **RA_yname** for each response variable, where **yname** is the response variable name.

The standardized and studentized Anscombe residuals are

$$r_{Asi} = \frac{r_{Ai}}{\sqrt{\hat{\phi}(1-h_i)}}$$

$$r_{Ati} = \frac{r_{Ai}}{\sqrt{\hat{\phi}_{(i)}(1-h_i)}}$$

where $\hat{\phi}$ is the estimate of the dispersion parameter ϕ, and $\hat{\phi}_{(i)}$ is a one-step approximation of ϕ after excluding the ith observation.

The standardized Anscombe residuals are stored in variables named **RAS_yname** and the studentized Anscombe residuals are stored in variables named **RAT_yname** for each response variable, where **yname** is the response variable name.

Partial Leverage Variables

The *partial leverage output variables* are variables used in the partial leverage plots. For each interval **X** variable, the corresponding partial leverage **X** variable is named **X_xname**, where **xname** is the **X** variable name. For each pair of **Y** and **X** variables, the corresponding partial leverage **Y** variable is named **yname_xname**, where **yname** is the **Y** variable name and **xname** is the **X** variable name. Up to the first three characters of the response variable name are used to create the new variable name.

Cook's D

Cook's D measures the change in the parameter estimates caused by deleting each observation. For linear models,

$$D_i = \frac{1}{ps^2}(\mathbf{b} - \mathbf{b}_{(i)})'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \mathbf{b}_{(i)})$$

where $\mathbf{b}_{(i)}$ is the vector of parameter estimates obtained after deleting the ith observation.

Cook (1977) suggests comparing D_i to the F distribution with p and $n - p$ degrees of freedom.

For generalized linear models,

$$D_i = \frac{1}{p\hat{\phi}}(\mathbf{b} - \mathbf{b}_{(i)})'(\mathbf{X}'\mathbf{W}\mathbf{X})(\mathbf{b} - \mathbf{b}_{(i)})$$

where $\mathbf{W} = \mathbf{W}_o$ when the full Hessian is used and $\mathbf{W} = \mathbf{W}_e$ when Fisher's scoring method is used.

Cook's D statistics are stored in variables named **D_yname** for each response variable, where **yname** is the response variable name.

Dffits

The *Dffits statistic* is a scaled measure of the change in the predicted value for the ith observation. For linear models,

$$F_i = \frac{\hat{\mu}_i - \hat{\mu}_{(i)}}{s_{(i)}\sqrt{h_i}}$$

where $\hat{\mu}_{(i)}$ is the ith value predicted without using the ith observation.

Large absolute values of F_i indicate influential observations. A general cutoff to consider is 2; a recommended size-adjusted cutoff is $2\sqrt{p/n}$.

For generalized linear models,

$$F_i = \frac{\hat{\mu}_i - \hat{\mu}_{(i)}}{\sqrt{\hat{\phi}_{(i)}h_i}}$$

The Dffits statistics are stored in variables named **F_yname** for each response variable, where **yname** is the response variable name.

Covratio

Covratio measures the effect of observations on the covariance matrix of the parameter estimates. For linear models,

$$C_i = \frac{|s^2_{(i)}(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}|}{|s^2(\mathbf{X}'\mathbf{X})^{-1}|}$$

where $\mathbf{X}_{(i)}$ is the $\mathbf{X}$ matrix without the ith observation.

Values of C_i near 1 indicate that the observation has little effect on the precision of the estimates. Observations with $|C_i - 1| \geq 3p/n$ suggest a need for further investigation.

For generalized linear models,

$$C_i = \frac{|\hat{\phi}_{(i)}(\mathbf{X}'_{(i)}\mathbf{W}_{(i)}\mathbf{X}_{(i)})^{-1}|}{|\hat{\phi}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}|}$$

where $\mathbf{W}_{(i)}$ is the $\mathbf{W}$ matrix without the ith observation, $\mathbf{W} = \mathbf{W}_o$ when the full Hessian is used, and $\mathbf{W} = \mathbf{W}_e$ when Fisher's scoring method is used.

The Covratio statistics are stored in variables named **C_yname** for each response variable, where **yname** is the response variable name.

Dfbetas

Dfbetas is a normalized measure of the effect of observations on the estimated regression coefficients. For linear models,

$$B_{j,i} = \frac{b_j - b_{j(i)}}{s_{(i)}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}}$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{jj}$ is the *j*th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. Values of $B_{j,i} > 2$ indicate observations that are influential in estimating a given parameter. A recommended size-adjusted cutoff is $2/\sqrt{n}$.

For generalized linear models,

$$B_{j,i} = \frac{b_j - b_{j(i)}}{\sqrt{\hat{\phi}_{(i)}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}_{jj}}}$$

where $\mathbf{W} = \mathbf{W}_o$ when the full Hessian is used and $\mathbf{W} = \mathbf{W}_e$ when the Fisher's scoring method is used.

The dfbetas statistics are stored in variables named **Byname_xname** for each pair of response and explanatory variables, where **yname** is the response variable name and **xname** is the explanatory variable name. Up to the first two characters of the response variable name are used to create the new variable name.

Weighted Analyses

If the errors ε_i do not have a common variance in the regression model

$$y_i = f(\mathbf{x}_i) + \varepsilon_i$$

a weighted analysis may be appropriate. The observation weights are the values of the **Weight** variable you specified.

In parametric regression, the linear model is given by

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Let $\mathbf{W}$ be an $n \times n$ diagonal matrix consisting of weights $w_1 > 0, w_2 > 0, \ldots$, and $w_n > 0$ for the observations, and let $\mathbf{W}^{1/2}$ be an $n \times n$ diagonal matrix with diagonal elements $w_1^{1/2}, w_2^{1/2}, \ldots$, and $w_n^{1/2}$.

The weighted fit analysis is equivalent to the usual (unweighted) fit analysis of the transformed model

$$\mathbf{y}^* = \mathbf{X}^*\beta + \epsilon^*$$

where $\mathbf{y}^* = \mathbf{W}^{1/2}\mathbf{y}$, $\mathbf{X}^* = \mathbf{W}^{1/2}\mathbf{X}$, and $\epsilon^* = \mathbf{W}^{1/2}\epsilon$.

The estimate of β is then given by

$$\mathbf{b}_w = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$$

For nonparametric weighted regression, the minimizing criterion in spline estimation is given by

$$S(\lambda) = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i\{y_i - \widehat{f_\lambda}(x_i)\}^2 + \lambda \int_{-\infty}^{\infty} \{\widehat{f_\lambda}''(x)\}^2 dx$$

In kernel estimation, individual weights are

$$W(x, x_i; \lambda) = \frac{w_i K_0(\frac{x - x_i}{\lambda})}{\sum_{j=1}^{n} w_j K_0(\frac{x - x_j}{\lambda})}$$

For generalized linear models, the function $a_i(\phi) = \phi/(m_i w_i)$ for binomial distribution with m_i trials in the ith observation, $a_i(\phi) = \phi/w_i$ for other distributions. The function $a_i(\phi)$ is used to compute the likelihood function and the diagonal matrices $\mathbf{W}_o$ and $\mathbf{W}_e$.

The individual deviance contribution d_i is obtained by multiplying the weight w_i by the unweighted deviance contribution. The deviance is the sum of these weighted deviance contributions.

The Pearson χ^2 statistic is

$$\chi^2 = \sum_{i=1}^{n} w_i m_i (y_i - \mu_i)^2 / V(\mu_i)$$

for binomial distribution with m_i trials in the ith observation,

$$\chi^2 = \sum_{i=1}^{n} w_i (y_i - \mu_i)^2 / V(\mu_i)$$

for other distributions.

References

Anscombe, F.J. (1953), "Contribution to the Discussion of H. Hotelling's Paper," *Journal of the Royal Statistical Society*, Series B, 15, 229–230.

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

Cleveland, W.S. (1979), "Robust Locally-Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cleveland, W.S., Devlin, S.J., and Grosse, E. (1988), "Regression by Local Fitting: Methods, Properties, and Computational Algorithms," *Journal of Econometrics*, 37, 87–114.

Cleveland, W.S. and Grosse, E. (1991), "Computational Methods for Local Regression," *Journal of Statistics and Computing*, 1, 47–62.

Cleveland, W.S. (1993), *Visualizing Data*, Summit, New Jersey: Hobart Press.

Cook, R.D. (1977), "Detection of Influential Observations in Linear Regression," *Technometrics*, 19, 15–18.

Cook, R.D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.

Cox, D.R. and Snell, E.J. (1968), "A General Definition of Residuals," *Journal of the Royal Statistical Society*, Series B, 30, 248–275.

Dobson, A.J. (1990), *An Introduction to Generalized Linear Models*, New York: Chapman and Hall.

Eubank, R.L. (1988), *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker, Inc.

Hastie, Y.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*, New York: Chapman and Hall.

Hinkley, D.V., Reid, N., and Snell, E.J. (1991), *Statistical Theory and Modelling*, New York: Chapman and Hall.

Hoaglin, D.C. and Welsch, R.E. (1978), "The Hat Matrix in Regression and ANOVA," *The American Statistician*, 32, 17–22.

Kvalseth, T.O. (1985), "Cautionary Note About R^2," *The American Statistician*, 39, 279.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, Second Edition, London: Chapman and Hall.

Pringle, R.M. and Raynor, A.A. (1971), *Generalized Inverse Matrices with Applications to Statistics*, New York: Hafner Publishing Co.

Rawlings, J.O. (1988), *Applied Regression Analysis: A Research Tool*, Pacific Grove, CA: Wadsworth & Brooks.

Reinsch, C. (1967), "Smoothing by Spline Functions," *Numerische Mathematik*, 10, 177–183.

Silverman, B.W. (1982), "Kernel Density Estimation using the Fast Fourier Transform," *Applied Statistics*, 31, 93–99.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall.

Velleman, P.F. and Welsch, R.E. (1981), "Efficient Computing of Regression Diagnostics," *The American Statistician*, 35, 234–242.

Wahba, G. and Wendelberger, J.G. (1980), "Some New Mathematical Methods for Variational Objective Analysis using Splines and Cross Validation," *Monthly Weather Review*, 108, 1122–1143.

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