BodyFat Selection Methods

Jack Tubbs

October 2023

Contents

Theory

1 Checking Model Assumptions

The validity of the regression assumptions are examined. Various techniques have been proposed for verifying whether or not the assumptions are valid.

2 Checks for Normality

2.1 Visual plots for the Residuals

If the proposed model assumption hold then the residuals satisfy $\hat{e}_i \sim N(0, \sigma^2(1 - h_{ii}))$. A simple plots of the residuals versus the predicted values $\hat{y}_i$ or the j^{th} independent variable x_{ji} often reveals if this assumption has been satisfied. Draper and Smith has an extensive discussion concerning plots of this type. Essentially, you should not "see" any patterns in the scatterplots. Likewise, the amount of variability should be somewhat constant across the plots.

2.2 QQ and PP Plots

The QQ or PP plots can be used to assess normality of the residuals. That is, let $z_{(1)} \leq z_{(2)} \leq \ldots \leq z_{(n)}$ represent the ordered values of n independent and identically distributed $N(0, 1)$ random variables. It can be shown that the expected value of $z_{(i)}$ is

$$E(z_{(i)}) \approx \gamma_i = \Phi^{-1}[(i - 3/8)/(n + 1/4)]$$

where Φ is the cdf for the standard normal given by

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-1/2t^2} dt.$$

The QQ plot consists of a scatterplot of $(z_{(i)}, \gamma_i)$. If the data are normal then the resulting scatterplot should lie close to the line ($\gamma_i = z_{(i)}$). The PP plot is similar to the QQ plot using the ordered pairs $(\Phi(z_{(i)}), [i/n])$. If either plot differs greatly from the diagonal line, then the normality assumption likely does not hold. Note, one should only check these plots when the model is appropriate as the residuals for inappropriate models often appear to be non-normally distributed.

3 Residual Analysis

3.1 Standardized Residuals

From the previous chapter we have that $\hat{e}_i = y_i - \hat{y}_i \sim N(0, \sigma^2(1 - h_{ii}))$. Since σ^2 is unknown it can be estimated with either

$$\hat{\sigma}^2 = MS_E$$

or

$$\hat{\sigma}^2_{(i)} = \frac{(n - p)\hat{\sigma}^2 - \hat{e}_i^2/(1 - h_{ii})}{(n - p - 1)}$$

where $\hat{\sigma}^2_{(i)}$ is the mean square for the error whenever the i^{th} observation has been omitted from the regression model. SAS provides two standardized residuals. They are:

1. Internally Studentized Residual (STUDENT) is given by

$$s_i = \frac{\hat{e}_i}{\hat{\sigma}(1-h_{ii})^{1/2}}.$$

2. Externally Studentized Residual (RSTUDENT) is given by

$$s_{(i)} = \frac{\hat{e}_i}{\hat{\sigma}_{(i)}(1-h_{ii})^{1/2}}.$$

3. Their properties are;

   - $s_i \sim t(df = n-p)$.
   - $s_{(i)} \sim t(df = n-p-1)$.

3.2 Leverages

The hat matrix H has the following properties;

1. $SS(\beta) = \hat{\beta}'X'y = y'Hy = y'H^2y = \hat{y}'\hat{y}$.
2. $\sum_{i=1}^{n} var(\hat{y}_i)/n = tr[\sigma^2 H]/n = \sigma^2 p/n$.
3. $H\mathbf{j} = \mathbf{j}$ whenever the y intercept is included in the model. In which case, the sum of every row and every column of H equals 1.
4. $0 \leq h_{ij} \leq 1$ and $\sum_{i=1}^{n} h_{ii} = p = rank(X)$. Since, the average of the diagonal elements for H is p/n the i^{th} observation is said to be a leverage point if $h_{ii} \geq 2p/n$.
5. Since $\hat{y} = Hy$ we have

$$\hat{y}_i = h_{ii}y_i + \sum_{i \neq j} h_{ij}y_j.$$

   This indicates the importance that y_i has upon $\hat{y}_i$ is given by the magnitude of h_{ii}.

3.3 Detection of Influential Observations

If one suspects that the i^{th} observation has an unusual influence upon the prediction equation $\hat{y}$ one can recompute the regression model with the i^{th} observation omitted from the calculation. Suppose that the i^{th} observation is omitted then resultant regression estimate becomes

$$\hat{\beta}_{(i)} = (X'_{(i)}X_{(i)})^{-1}X'_{(i)}y_{(i)}$$

from which we have a new predicted value for y given by $\hat{y}_{(i)} = X\hat{\beta}_{(i)}$.

Cook's Distance

Cook's distance is a measure of how far the original "line" ($\hat{y}$) is from the "new line" ($\hat{y}_{(i)}$) when the i^{th} observation is omitted from the calculation. Cook's distance is

$$\begin{aligned} D_i &= (\hat{y} - \hat{y}_{(i)})'(\hat{y} - \hat{y}_{(i)})/(p\hat{\sigma}^2) \\ &= (\hat{\beta} - \hat{\beta}_{(i)})'X'X(\hat{\beta} - \hat{\beta}_{(i)})/(p\hat{\sigma}^2) \\ &= [\frac{\hat{e}_i}{\hat{\sigma}(1-h_{ii})^{1/2}}]^2[\frac{h_{ii}}{p(1-h_{ii})}]. \end{aligned}$$

DFFITS

A related measure to Cook's distance is the DFFITS statistic given by

$$DFFITS_i^2 = (\hat{\beta} - \hat{\beta}_{(i)})'X'X(\hat{\beta} - \hat{\beta}_{(i)})/(\hat{\sigma}^2_{(i)}).$$

COVRATIO

Another measure of influence is the COVRATIO which is the ratio of the determinant of the covariance matrix for the estimate $\hat{\beta}$, given by $det[\hat{\sigma}^2(X'X)^{-1}]$ when the i^{th} observation has been removed for the computation of the estimate $\hat{\beta}_{(i)}$. That is,

$$COVRATIO = det[\hat{\sigma}^2_{(i)}(X'_{(i)}X_{(i)})^{-1}]/det[\hat{\sigma}^2(X'X)^{-1}].$$

This statistic should be close to one whenever the observation has little influence upon the estimation of β. If the statistics is much different from one then the observation is said to be influential.

Belsley, Kuh, and Welsch suggest that observations with

$$\mid COVRATIO - 1 \mid \geq \frac{3p}{n}$$

where p is the number of parameters in the model and n is the number of observations used to fit the model, are worth investigation.

DFBETAS

The DFBETAS statistics are the scaled measures of the change in each parameter estimate and are calculated by deleting the i^{th} observation:

$$DFBETAS_j = \frac{\hat{\beta}_j - \hat{\beta}_{(i)j}}{\hat{\sigma}_{(i)}\sqrt{(X'X)^{-1}_{jj}}}$$

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

In general, large values of DFBETAS indicate observations that are influential in estimating a given parameter. Belsley, Kuh, and Welsch recommend 2 as a general cutoff value to indicate influential observations and as a size-adjusted cutoff.

4 Multicollinearity

In this section the problem of having linear dependence among the "independent" variables is discussed whenever there is a linear dependency among the independent variables. That is, let

$$X = (x_0, x_1, x_2, \ldots, x_{p-1})$$

where x_i is the $n \times 1$ vector of responses for the i^{th} variable and $x_0 = j$. The independent variables are said to have linear dependence whenever

$$\sum_{i=0}^{p-1} t_i x_i = 0,$$

for $t_i \neq 0$. If the above condition holds then $(X'X)^{-1}$ does not exist. Actually, linear dependency seldom holds when one is using observed data. Instead, one has "near" linear dependency which implies that $(X'X)^{-1}$ is ill conditioned. This ill-conditionality can affect the resultant estimates since they are based upon $(X'X)^{-1}$. Yet, a more severe problem occurs when computing the variance of the estimates, $\hat{\beta}$, $\hat{y}$ and the estimated residuals $\hat{e}$. This is equivalent to the problem of dividing a number by a very small positive value. Montogomery and Peck list four primary reasons for having a mulicollinearity problem within the regression model. They are:

1. The data collection method employed.
2. Constraints on the model or in the population.
3. Model specification.
4. An over-defined model.

4.1 Detecting Multicollinearity

There are two ways of detecting multicollinearity in these notes[1]. The first is a statistical solution using the R-square for each of the independent variables and the second method makes use of a result from matrix theory.

Tolerances and Variance Inflation Factors

When there is a linear dependency among the independent variables the R-square for the variable x_j using the remaining independent variables will be close to 1, $R_j^2 \simeq 1$. In which case, the j^{th} variable may not be needed in the model. When the variables are centered and scaled before computing the least squares estimates, it follows that

$$TOL_j = C_{(jj)} = (1 - R_j^2), \quad j = 1, 2, \ldots, p-1,$$

and

$$VIF_j = TOL_j^{-1},$$

where $C = (X'X)^{-1} = C_{(ij)}$. Some authors suggest that if $VIF_j > 10$ then collinearity may be present.

[1] Neither of these procedures suggest which independent variable needs to be removed from the model. Rather, they indicate that there is a problem with the independent variables in the model.

Eigenvalues and Condition Numbers

If the independent variables are centered and scaled, it follows that

$$tr[X'X] = \sum_{j=0}^{p-1} \lambda_j = p,$$

where λ_j is the j^{th} eigenvalue of the positive semi-definite matrix $X'X$. Since the eigenvalues are real numbers, they can be ordered so that $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{(p-1)} \geq 0$. The matrix $X'X$ is said to be ill conditioned whenever one of the $\lambda's$ becomes small. The condition number is defined as

$$CN_j = \sqrt{\lambda_0/\lambda_j}.$$

If $CN_j > 30$, collinearity may be present.

5 Principle Component Analysis

Principle Component Analysis is concern with explaining the variance-covariance structure of a set of p-variables through a set of a few ($q < p$) linear combinations of the original p variables. This method is used to aid one in (1) dimension reduction and (2) data interpretation.

Before considering an example using this method it is necessary to review some matrix algebra concerning eigenvalues and eigenvectors.

5.1 Eigenvalues and Eigenvectors

Suppose that A is an $n \times n$ matrix. Is there a transformation of vectors $\vec{x}$ that transforms A into a constant multiple of itself? That is, does there exist a vector $\vec{x}$ satisfying,

$$A\vec{x} = \lambda\vec{x}$$

for some a constant λ? If so, then

$$(A - \lambda I_n)\vec{x} = 0,$$

or

$$\mid A - \lambda I_n \mid = 0$$

for $\vec{x} \neq 0$. It can be shown that

$$\sum_{i=0}^{n} a_i \lambda^i = 0.$$

This last equation is called the characteristic equation for a $n \times n$ matrix A. The matrix A could possibly have as many as n different values λ that satisfy the characteristic equation. These solutions are called the characteristic values or the eigenvalues for the matrix A. Suppose that λ_1 is a solution and

$$A\vec{x}_1 = \lambda_1\vec{x}_1, \vec{x}_1 \neq 0,$$

then $\vec{x}_1$ is said to a characteristic vector or eigenvector of A corresponding to the eigenvalue λ_1. Note: The eigenvalues may or may not be real numbers.

5.2 Properties of Eigenvalues–Eigenvectors

1. The $n \times n$ matrix A has at least one eigenvalue equal to zero if and only if A is singular.
2. The eigenvalues for A, $C^{-1}AC$ and CAC^{-1} have the same set of eigenvalues for any nonsingular matrix C.
3. The matrices A and A' have the same set of eigenvalues but need not have the same eigenvectors.
4. Let A be a nonsingular matrix with eigenvalue λ, then $1/\lambda$ is an eigenvalue of A^{-1}.
5. The eigenvectors are not unique, for if $\vec{x}_1$ is an eigenvector corresponding to λ_1 then $c\vec{x}_1$ is also n eigenvector, since $Ac\vec{x}_1 = \lambda_1 c\vec{x}_1$, for any nonzero value c.
6. Let A be an $n \times n$ real matrix, then there exists a nonsingular, complex matrix Q such that $Q'AQ = T$, where T is a complex, upper triangular matrix, and the eigenvalues of A are the diagonal elements of the matrix T.
7. Suppose that A is a symmetric matrix;
   (a) Then the eigenvalues of A are real numbers.
   (b) For each eigenvalue there exists a real eigenvector (each element is a real number).
   (c) Let λ_1 and λ_2 be eigenvalues of A with corresponding eigenvectors $\vec{x}_1$ and $\vec{x}_2$, then $\vec{x}_1$ and $\vec{x}_2$ are orthogonal vectors, that is, $\vec{x}_1'\vec{x}_2 = 0$.
   (d) There exists an orthogonal matrix P ($P'P = PP' = I_n$) such that $P'AP = D$, where D is a diagonal matrix whose diagonal elements are the eigenvalues of the matrix A.

5.3 Principle Components

Principal component analysis was originated by Pearson (1901) and later developed by Hotelling (1933). The application of principal components is discussed by Rao (1964); Cooley and Lohnes (1971); Gnanadesikan (1977). Excellent statistical treatments of principal components are found in Kshirsagar (1972); Morrison (1976); Mardia, Kent, and Bibby (1979).

If you have a data set that contains p numeric variables, you can compute p principal components. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix. The eigenvectors are usually taken with unit length. The principal components are sorted by descending order of the eigenvalues, which are equal to the variances of the components.

Principal components have a variety of useful properties (Rao 1964; Kshirsagar 1972):

- The eigenvectors are orthogonal, so the principal components represent jointly perpendicular directions through the space of the original variables.
- The principal component scores are jointly uncorrelated. Note that this property is quite distinct from the previous one.
- The first principal component has the largest variance of any unit-length linear combination of the observed variables. The jth principal component has the largest variance of any unit-length linear combination orthogonal to the first j – 1 principal components. The last principal component has the smallest variance of any linear combination of the original variables.

- The scores on the first j principal components have the highest possible generalized variance of any set of unit-length linear combinations of the original variables.
- The first j principal components provide a least squares solution to the model

$$Y = X\beta + \epsilon$$

where Y is an $n \times p$ matrix of the centered observed variables; X is the $n \times j$ matrix of scores on the first j principal components; β is the matrix of eigenvectors; ϵ is an matrix of residuals; and you want to minimize trace, the sum of all the squared elements in . In other words, the first j principal components are the best linear predictors of the original variables among all possible sets of j variables, although any nonsingular linear transformation of the first j principal components would provide equally good prediction. The same result is obtained if you want to minimize the determinant or the Euclidean (Schur, Frobenius) norm of rather than the trace.

- In geometric terms, the j-dimensional linear subspace that is spanned by the first j principal components provides the best possible fit to the data points as measured by the sum of squared perpendicular distances from each data point to the subspace. This contrasts with the geometric interpretation of least squares regression, which minimizes the sum of squared vertical distances. For example, suppose you have two variables. Then, the first principal component minimizes the sum of squared perpendicular distances from the points to the first principal axis. This contrasts with least squares, which would minimize the sum of squared vertical distances from the points to the fitted line.

Principal component analysis can also be used for exploring polynomial relationships and for multivariate outlier detection (Gnanadesikan 1977), and it is related to factor analysis, correspondence analysis, allometry, and biased regression techniques (Mardia, Kent, and Bibby 1979).

Suppose that one has a p-dimensional vector given by $X = \begin{pmatrix} x_1 & x_2 & \dots & x_p \end{pmatrix}$ has covariance matrix Σ. The idea is to find a new q-dimensional vector $Y = \begin{pmatrix} y_1 & y_2 & \dots & y_p \end{pmatrix}$ where

$$y_i = \sum_{j=1}^{p} a_{ij} x_j,$$

for $i = 1, 2, \dots, q$ and $var(y_i) = a_i' \Sigma a_i$, $cov(y_i, y_k) = a_i' \Sigma a_k = 0$, $var(y_1) \geq var(y_2) \geq \dots var(y_q)$ for $a_i' = \begin{pmatrix} a_{1i} & a_{2i} & \dots & a_{pi} \end{pmatrix}$.
This problem has the following solution;

1. Suppose that the matrix Σ has associated real eigenvalue–eigenvectors given by (λ_i, a_i) where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, then the i^{th} principle component is given by

$$y_i = a_i' X = a_{i1} x_1 + a_{i2} x_2 + \dots + a_{ip} x_p,$$

and $var(y_i) = \lambda_i$ for $i = 1, 2, \dots, p$, $cov(y_i, y_k) = a_i' \Sigma a_k = 0$ for $i \neq k$. Note, the eigenvalues λ_i are unique, however, the eigenvectors (and hence the vectors $\vec{y_i}$) are not.

2. The total variance for the p dimensions is $tr[\Sigma] = \sum_{i=1}^{p} \lambda_i$. Hence, the proportion of variance explained by the k^{th} principle component is $\lambda_k / \sum_{i=1}^{p} \lambda_i$.

3. If the matrix X is centered and scaled so that Σ is the correlation matrix, then $\sum_{i=1}^{p} \lambda_i = p$.

6 Model Selection

The methods presented in this section are used to determine appropriate subset models in the multiple regression problem. Various statistics can be used, including; R^2, the adjusted R^2, s^2 the residual mean square (s^2), and Mallow's C_p statistic.

Mallow's statistics – C_p

The statistics is

$$C_p = RSS_p/s^2 - (n - 2p),$$

where RSS_p is the residual sum of squares from the model containing $p =$ rank(X) parameters (including β_0), and $s^2 = MS_E(r)$ is the residual mean square from the full model containing r predictor variables [assumed to be the most reliable estimate of σ^2]. Note, when $p = r + 1$, $C_p = p$. The idea of using this statistics is to find the smallest value of p such that $C_p \approx p$.

6.1 Subset Selection Methods

Statistical selection procedures include;

Forward Selection

The forward procedure starts with a single variable model (often selected with the highest R^2) and then adds additional variables that satisfy an entry criteria. The process continues until no other variables satisfy the entry criteria.

Backward Selection

The backward procedure starts with the complete (full) model and then eliminates variables that satisfy an exit criteria. The procedure is as follows;

1. Compute the regression model using all the predictor variables.
2. The partial F-value is calculated for every predictor variable using the type II sum of squares.
3. The lowest partial F-value, F_L, is compared with a preselected significance level, F_0.
   (a) If $F_L < F_0$, remove the variable corresponding to F_L, say X_L then recompute the model using the reduced model.
   (b) If $F_L > F_0$, adopt the regression model as calculated.

Stepwise Selection

Stepwise is a forward selection method that selects the best single predictor (say, the variable with the largest R^2), X_1 and fit the equation $\hat{y} = f(X_1)$. If this model is not significant, stop and conclude that $\hat{y} = \bar{y}$. If the model is significant, select the next predictor variable, say X_2 based upon the one with the largest partial F-value and the equation is given by $\hat{y} = f(X_1, X_2)$. This model checked for improvement in the R^2 and partial F-values for both variables in the equation. (This differs from the forward procedure in that a first

variable may be excluded from the model at this step whereas in forward selection once a variable enters the model it remains). These partial F-values are used to determine whether or not a variable remains in the model or is excluded. The procedure continues until no new variables satisfy the entry criteria.

7 Shrinkage Methods

The following material is given in Seber and Lee in chapter 12. Shrinkage methods were first proposed by James and Stein (1961). Suppose that $Z \sim N_p(\mu, \sigma^2\ I_p)$ when $p > 2$. The obvious estimate of μ is Z, which is the minimum variance unbiased estimate. However, this estimate is unsatisfactory since the expected length of Z, given by $||\ Z\ ||^2$ tends to be too large since,

$$\begin{aligned} E(||\ Z\ ||^2) &= \sum_{i-1}^{p} E[Z_i]^2 \\ &= \sum_{i-1}^{p} E[\sigma^2 + \mu_i^2] \\ &= p\sigma^2 + ||\ \mu\ ||^2 \\ &> ||\ \mu\ ||^2\ . \end{aligned}$$

Thus, some of the elements of the estimate are too large. This suggests "shrinking" some of the elements of Z, and considering an estimate of the form $\tilde{\mu} = cZ$, where $0 < c < 1$.

This estimate is biased, but it is possible to choose c so that $\tilde{\mu}$ has a smaller mean-square error than Z as an estimate of μ. Consider

$$\begin{aligned} E(||\ \tilde{\mu} - \mu\ ||^2) &= \sum_{i=1}^{p} E[(cZ_i - \mu_i)^2] \\ &= \sum_{i=1}^{p} E[(cZ_i - c\mu + c\mu - \mu_i)^2] \\ &= \sum_{i=1}^{p} E[(c(Z_i - \mu_i) - (1-c)\mu_i)^2] \\ &= \sum_{i=1}^{p} [c^2\sigma^2 + (1-c)^2\mu_i^2] \\ &= c^2 p\sigma^2 + (1-c)^2\ ||\ \mu\ ||^2\ . \end{aligned}$$

which is minimized by choosing $c = ||\ \mu\ ||^2\ /(p\sigma^2 + ||\ \mu\ ||^2)$. Thus the optimal estimate can be written as

$$\tilde{\mu} = (1 - \frac{p\sigma^2}{p\sigma^2 + ||\ \mu\ ||^2})Z.$$

However, this estimate is not a practical estimate since it requires knowledge of $||\ \mu\ ||$. Which in turns suggest the estimate

$$\tilde{\mu} = (1 - \frac{p\sigma^2}{|| Z ||^2})Z.$$

One can do better than this as James and Stein showed that of all estimates of the form

$$\tilde{\mu} = (1 - \frac{b}{|| Z ||^2})Z,$$

are best, in the sense of minimum mean-square error. The choice for b then becomes $b = (p-2)\sigma^2$ provided that $p > 2$.

In the regression case with orthonormal predictor variables (i.e., $X'X = I_p$) and known variance σ^2 and $\beta_0 = 0$, one can apply the James-Stein estimate procedure directly by setting $Z = \hat{\beta}$ and $\mu = \beta$ to obtain the "shrinkage" estimator

$$\tilde{\beta} = (1 - \frac{(p-2)\sigma^2}{|| \hat{\beta} ||^2})\hat{\beta}$$

which has the smallest MSE (mean-square error) of any estimator of the form $(1 - c)\hat{\beta}$. If σ^2 is unknown then it can be replaced by the usual estimate s^2 of σ^2 where the optimality property no longer holds.

7.1 Ridge Regression

Ridge regression is a popular method for detecting multicollinearity within a regression model. It was first proposed by Hoerl and Kennard (1970) and it was one of the first biased estimation procedures. The idea is fairly simple. Since the matrix $X'X$ is ill-condition or nearly singular one can add positive constants to the diagonal matrix and insure that the resulting matrix is not ill-conditioned. That is, consider the biased normal equations given by

$$(X'X + kI_n)\beta = X'y.$$

With a resulting biased estimate for β given by

$$\tilde{\beta}(k) = (X'X + kI_n)^{-1}X'y,$$

where k is called the shrinkage parameter. Since, $E(\tilde{\beta}) \neq \beta$ some do not want to use such a procedure. However in spite of the fact that $\tilde{\beta}$ is biased, it does have the effect of reducing the variance in the estimator. It can be shown that,

$$var(\hat{\beta}_j) = \sigma^2 1/\lambda_j,$$

where λ_j is the j^{th} eigenvalue of $X'X$. So when $X'X$ is ill-conditioned some of the $\lambda_j' s$ are very small, hence $var(\hat{\beta}_j)$ is very large. However,

$$var(\tilde{\beta}_j) = \sigma^2\lambda_j/(\lambda_j + k)^2.$$

Consider the example where $\sigma^2 = 1$, $\lambda_1 = 2.985, \lambda_2 = 0.01$, and $\lambda_3 = 0.005$, the usual least squares estimation gives,

$$\sum_{j=1}^{3} var(\hat{\beta}_j) = \sigma^2 \sum_{i=1}^{3} 1/\lambda_j = .3350 + 100 + 200 = 300.3350.$$

However, if $k = 0.10$ we have,

$$\sum_{j=1}^{3} var(\hat{\beta}_j) = \sigma^2 \sum_{i=1}^{3} \lambda_j/(\lambda_j + k)^2 \approx 2.3.$$

This process of reducing the total variance is very desirable and has led people to proposed similar estimation procedures called shrinkage estimators. In this class, we are interested using this procedure as a way of identifying multicollinearity and the variables which may contribute to this problem.

7.2 LASSO

The LASSO procedure is similar to ridge regression in that the estimator is a "shrinkage estimate" whereby one trades off the unbiased property for an estimate that is much more precise (smaller dispersion or mean square error). That is, the least squares estimate is

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\{\|y - X\beta\|^2\}.$$

Whereas the ridge estimate is

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}}\{\|y - X\beta\|^2 + \lambda \|\beta\|^2\}$$

where $\|\beta\| = \sum_{i=1}^{p} \beta_i^2$ is the usual l_2 or Euclidean norm. The LASSO estimate is

$$\hat{\beta}_{lasso} = \underset{\beta}{\operatorname{argmin}}\{\|y - X\beta\|^2 + \lambda \|\beta\|\}$$

where $\|\beta\| = \sum_{i=1}^{p} |\beta_i|$ is the l_1 norm.

Hastie, Tibshirani and Friedman book entitled "The Elements of Statistical Learning: Data Mining, Inference, and Prediction" have a very understandable discussion of the above methods. I have reproduced some of their material here.

> In this section we discuss and compare the three approaches discussed so far for restricting the linear regression model: subset selection, ridge regression and the lasso. In the case of an orthonormal input matrix X the three procedures have explicit solutions. Each method applies a simple transformation to the least squares estimate $\hat{\beta}_j$. Ridge regression does a "proportional shrinkage". Lasso translates each coefficient by a constant factor λ, truncating at zero. This is called "soft thresholding," and is used in the context of wavelet-based smoothing in Section 5.9. Best-subset selection drops all variables with coefficients smaller than the Mth largest; this is a form of "hard thresholding."

7.3 Elastic Net Selection

This section applies to the following procedures: REGSELECT.

The METHOD=ELASTICNET option specifies the elastic net method proposed by Zou and Hastie (2005), which bridges the LASSO method and ridge regression. The elastic net method strikes a balance between having a parsimonious model and borrowing strength from correlated regressors, by solving the

least squares regression problem with constraints on both the sum of the absolute coefficients and the sum of the squared coefficients.

More specifically, the elastic net coefficients $\beta = (\beta_1, \beta_2, \ldots, \beta_m)$ are the solution to the constrained optimization problem

$$\min\|\mathbf{y} - \mathbf{X}\beta\|^2 \quad subject\ to\ \sum_{j=1}^{m} \mid \beta_j \mid \ \leq t_1, \sum_{j=1}^{m} \beta_j^2 \leq t_2$$

This can be written as the equivalent Lagrangian form

$$\min\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_1 \sum_{j=1}^{m} \mid \beta_j \mid + \lambda_2 \sum_{j=1}^{m} \beta_j^2$$

Elastic net can be treated as a convex combination of LASSO and ridge penalty; pure LASSO and pure ridge are two limiting cases. If t_1 is set to a very large value or, equivalently, if λ_1 is set to 0, then the elastic net method reduces to ridge regression. If t_2 is set to a very large value or, equivalently, if λ_2 is set to 0, then the elastic net method reduces to LASSO. If t_1 and t_2 are both large or, equivalently, if λ_1 and λ_2 are both set to 0, then the elastic net method reduces to ordinary least squares regression.

The elastic net method can overcome the limitations of LASSO in the following three scenarios:

- If you have more parameters than observations ($m > n$), the LASSO method selects at most n variables before it saturates, because of the nature of the convex optimization problem. This can be a defect for a variable selection method. By contrast, the elastic net method can select more than n variables in this case because of the ridge regression regularization.
- If there is a group of variables that have high pairwise correlations, then whereas LASSO tends to select only one variable from that group, the elastic net method can select more than one variable.
- If you have more observations than parameters ($n > m$), and there are high correlations between predictors, then it has been empirically observed that the prediction performance of LASSO is dominated by ridge regression. In this case, the elastic net method can achieve better prediction performance by using ridge regression regularization.

SAS

7.4 Code

```
libname LDATA '/home/jacktubbs/my_shared_file_links/jacktubbs/myfolders/SAS Data Sets/';
options center nodate pagesize=100 ls=80;
title1 'Body Fat Data';

/*
Density determined from underwater weighing
 Percent body fat from Siri's (1956) equation
 Age (years)
 Weight (lbs)
 Height (inches)
 Neck circumference (cm)
 Chest circumference (cm)
 Abdomen 2 circumference (cm)
 Hip circumference (cm)
 Thigh circumference (cm)
 Knee circumference (cm)
 Ankle circumference (cm)
 Biceps (extended) circumference (cm)
 Forearm circumference (cm)
 Wrist circumference (cm)
*/



data bodyfat; set ldata.bodyfat; run;

title2 'Scatterplot of Entire Data';
proc sgscatter data=bodyfat;
  matrix per_fat  age wt ht neck chest abdomen hip
        /diagonal=(histogram normal);run;
proc sgscatter data=bodyfat;
  matrix per_fat  thigh knee ankle biceps forearm  wrist
        /diagonal=(histogram normal);run;


/*
Use per_fat or density as the dependent variable with a subset of the data
*/

title2 'Simple Random Sampling of 40 values';
proc surveyselect data=ldata.bodyfat
   method=srs n=40 out=new_bfat seed = 54321;
run;

proc sgplot data=new_bfat;
```

```
histogram per_fat;
density per_fat;
density per_fat/ type=kernel;
run;

proc sgscatter data=new_bfat;
matrix per_fat thigh knee ankle abdomen hip chest biceps forearm wrist age
         /diagonal=(histogram normal);
run;


title2 'Multiple Regression';
  proc reg data=new_bfat  plots = (diagnostics partial);
model per_fat=thigh knee ankle abdomen hip chest biceps forearm wrist age
        /partial   ss1 ss2;
run;

title2 'Multiple Regression';
title3 'Ridge and Check for Collinearity';
proc reg data=new_bfat plots(only)=ridge(unpack VIFaxis=log)
outest=b ridge=0 to 0.02 by .002 plots = (diagnostics partial);
model per_fat=thigh knee ankle abdomen hip chest biceps forearm wrist age
        /vif tol collin ;
run;

title2 'Forward Regression';
proc reg data=new_bfat;*   plots = (diagnostics partial);
model per_fat=thigh knee ankle abdomen hip chest biceps forearm wrist age
        /selection=forward details=summary;
run;

title2 'Stepwise Regression';
proc reg data=new_bfat;*   plots = (diagnostics partial);
model per_fat=thigh knee ankle abdomen hip chest biceps forearm wrist age
        /selection=cp best=5 details=summary;
run;

title2 'Stepwise Regression';
proc glmselect data=new_bfat   plots = all;
model per_fat=thigh knee ankle abdomen hip chest biceps forearm wrist age
        /selection=stepwise choose=cp details=summary;
run;

title2 'Lasso Regression';
title3 'Standardize the data';

proc stdize data=new_bfat out=data1;
var per_fat thigh knee ankle abdomen biceps forearm age wrist;  run;

proc glmselect data=data1   plots = all;
```

```
model per_fat=thigh knee ankle abdomen biceps forearm age wrist
      / selection=lasso choose=cp details=summary;
run;

title2 'Correlation and PCR';
proc corr data=new_bfat pearson spearman kendall;
var per_fat thigh knee ankle abdomen biceps forearm age wrist;
run;

proc princomp data=new_bfat  out=PCmeasures standard
              plots=patternprofile;
var thigh knee ankle abdomen biceps forearm age wrist;
run;

proc reg data=PCmeasures plots = (diagnostics partial);
model per_fat=Prin1 -- Prin2 / b ss1;
run;

proc reg data=PCmeasures plots(only)= none;
model per_fat=Prin1 -- Prin6 / selection=cp best=10;
run;

quit;

ods graphics off;
```

Output

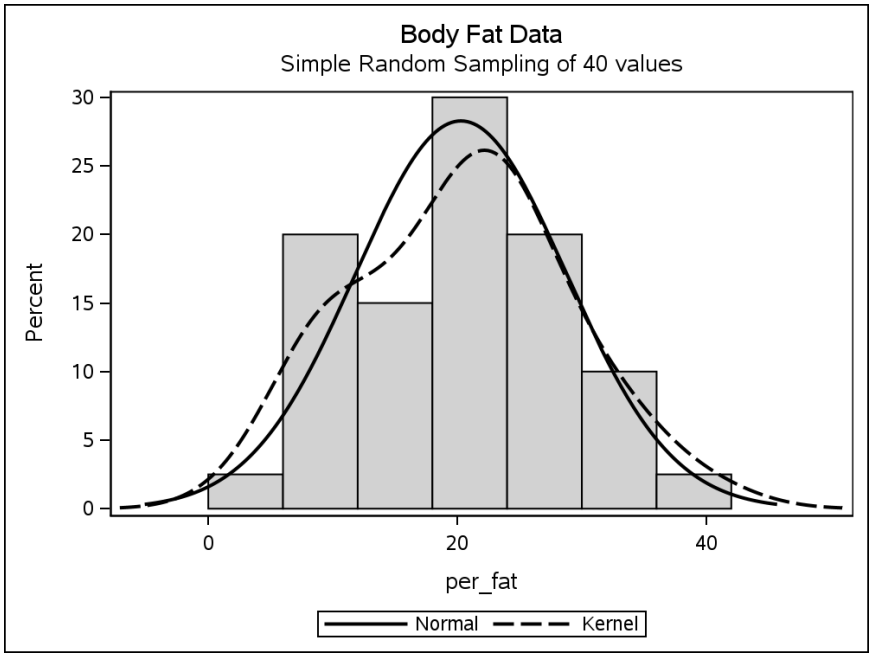
Body Fat Data

Simple Random Sampling of 40 values

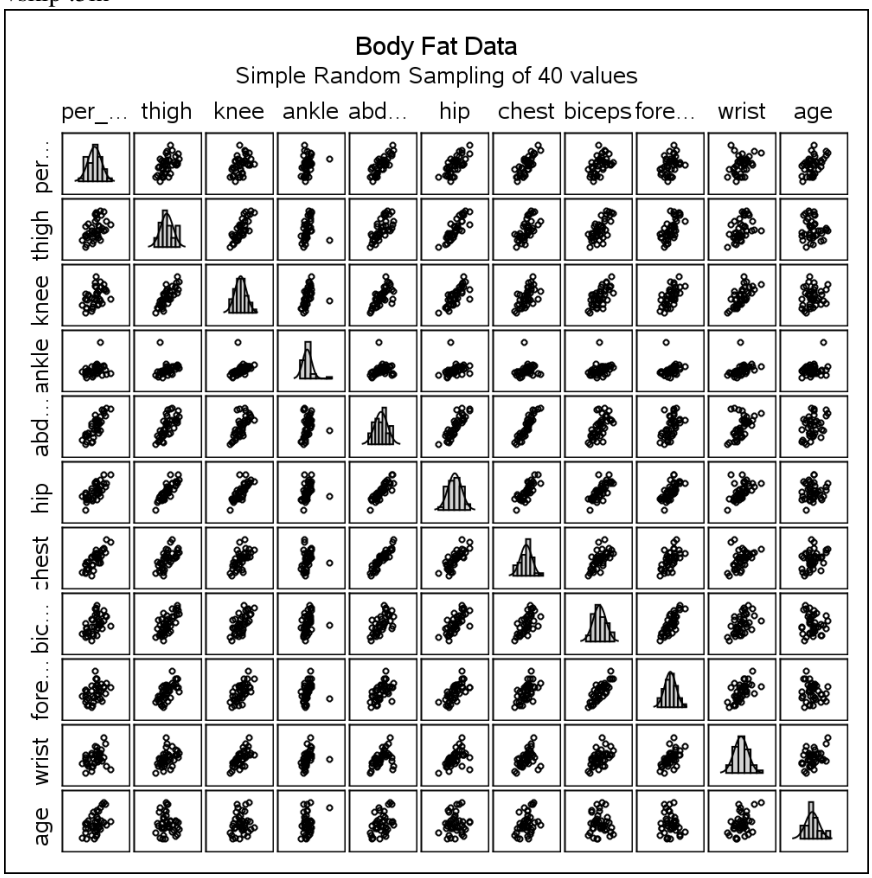
The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	BODYFAT
Random Number Seed	54321
Sample Size	40
Selection Probability	0.15873
Sampling Weight	6.3
Output Data Set	NEW_BFAT


Body Fat Data
Simple Random Sampling of 40 values
Percent
30
25
20
15
10
5
0
0
20
40
per_fat
Normal
Kernel

vskip .5in


Body Fat Data
Simple Random Sampling of 40 values
per_...
thigh
knee
ankle
abd...
hip
chest
biceps
fore...
wrist
age

Body Fat Data

Multiple Regression

The REG Procedure

Number of Observations Read	40
Number of Observations Used	40

Analysis of Variance					
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	10	2205.82714	220.58271	10.88	<.0001
Error	29	588.19661	20.28264		
Corrected Total	39	2794.02375			

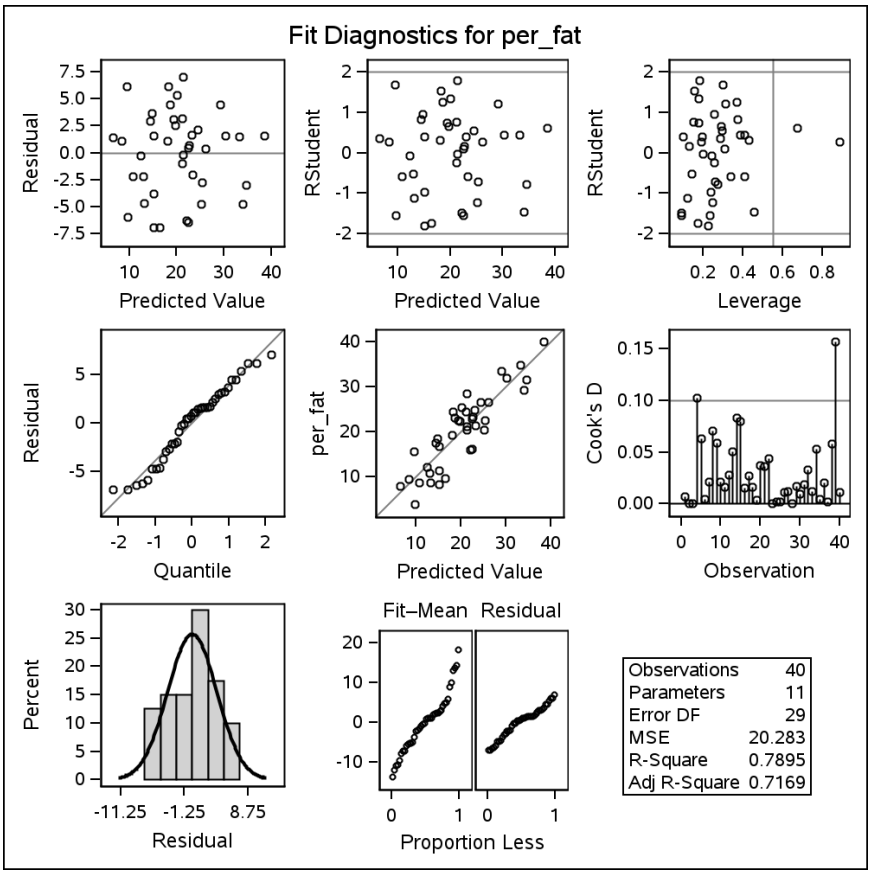
Root MSE	4.50363	*R-Square*	0.7895
Dependent Mean	20.28750	*Adj R-Sq*	0.7169
Coeff Var	22.19902		

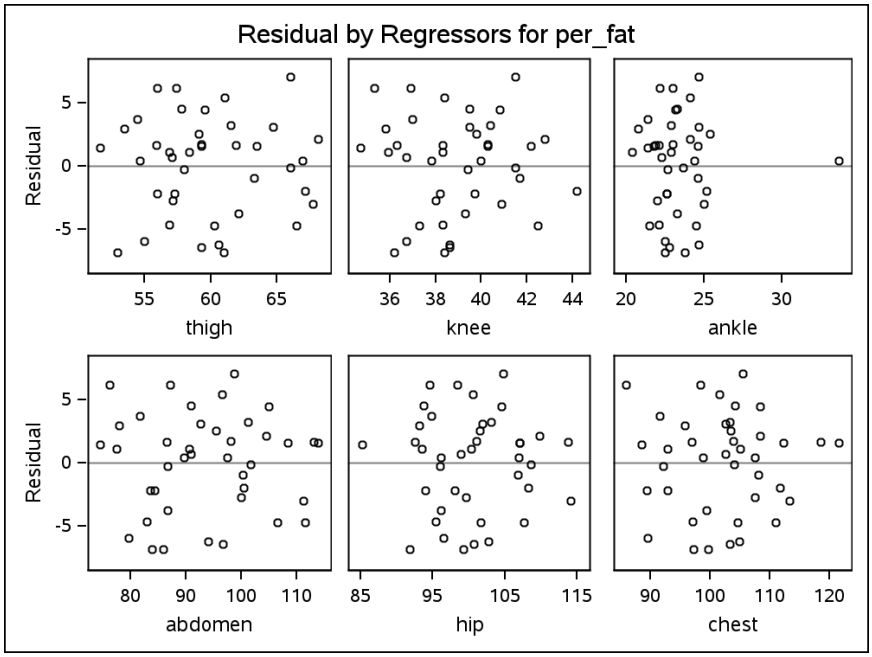
Parameter Estimates							
Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*	*Type I SS*	*Type II SS*
Intercept	1	−43.13906	21.34025	−2.02	0.0525	16463	82.88323
thigh	1	−0.12706	0.40602	−0.31	0.7566	627.00385	1.98614
knee	1	−0.99785	0.71784	−1.39	0.1751	24.45059	39.19263
ankle	1	0.41394	0.39613	1.04	0.3047	36.70898	22.14782
abdomen	1	0.25444	0.26469	0.96	0.3444	1395.08396	18.74172
hip	1	0.53385	0.36426	1.47	0.1535	16.43055	43.56506
chest	1	0.30955	0.26296	1.18	0.2487	30.17664	28.10717
biceps	1	0.37224	0.51926	0.72	0.4792	0.41822	10.42338
forearm	1	−0.34412	0.98429	−0.35	0.7292	24.56715	2.47911
wrist	1	−1.05353	1.52025	−0.69	0.4938	1.25932	9.74063
age	1	0.17767	0.11347	1.57	0.1282	49.72788	49.72788

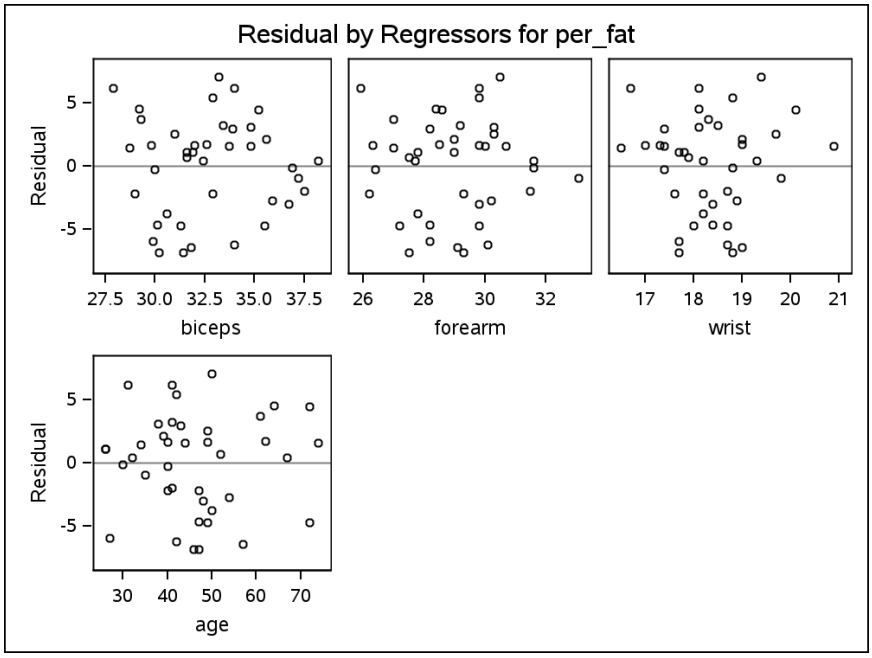
Body Fat Data

Multiple Regression

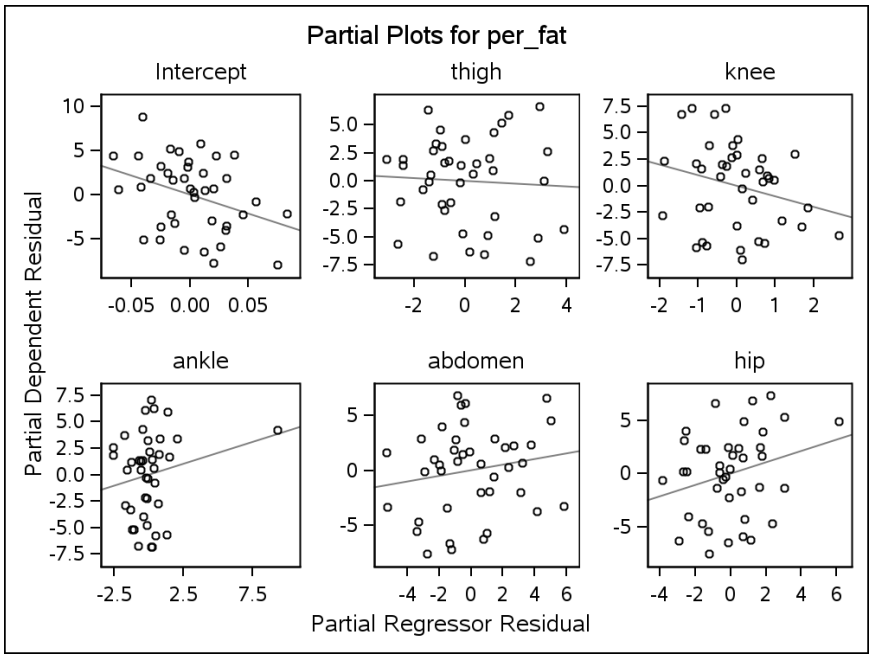
The REG Procedure

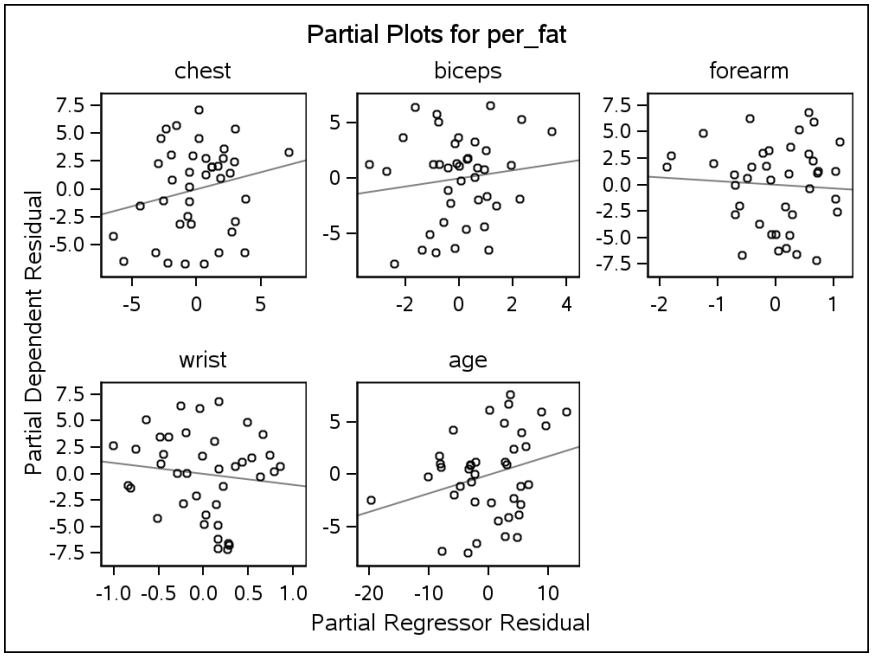



Residual by Regressors for per_fat
Residual
5
0
-5
55
60
65
thigh
36
38
40
42
44
knee
20
25
30
ankle
Residual
5
0
-5
80
90
100
110
abdomen
85
95
105
115
hip
90
100
110
120
chest


Residual by Regressors for per_fat
Residual
5
0
-5
27.5
30.0
32.5
35.0
37.5
biceps
26
28
30
32
forearm
17
18
19
20
21
wrist
Residual
5
0
-5
30
40
50
60
70
age

Partial Regression Residual Plot





Body Fat Data

Multiple Regression

Ridge and Check for Collinearity

The REG Procedure

Model: MODEL1

Dependent Variable: per_fat

Number of Observations Read	40
Number of Observations Used	40

Note	No intercept in model. R-Square is redefined.

Analysis of Variance					
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	10	18586	1858.62502	83.09	<.0001
Error	30	671.07984	22.36933		
Uncorrected Total	40	19257			

Root MSE	4.72962	*R-Square*	0.9652
Dependent Mean	20.28750	*Adj R-Sq*	0.9535
Coeff Var	23.31299		

Parameter Estimates							
Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*	*Tolerance*	*Variance Inflation*
thigh	1	−0.10148	0.42619	−0.24	0.8134	0.00085634	1167.76649
knee	1	−1.18396	0.74763	−1.58	0.1238	0.00065739	1521.17306
ankle	1	0.36278	0.41516	0.87	0.3891	0.00590	169.50730
abdomen	1	0.58386	0.21905	2.67	0.0123	0.00131	762.80739
hip	1	0.21511	0.34485	0.62	0.5375	0.00046341	2157.91462
chest	1	0.06110	0.24413	0.25	0.8041	0.00089360	1119.06866

Parameter Estimates							
Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*	*Tolerance*	*Variance Inflation*
biceps	1	0.60734	0.53146	1.14	0.2622	0.00184	544.14715
forearm	1	−0.40968	1.03312	−0.40	0.6945	0.00062297	1605.21291
wrist	1	−1.83595	1.54393	−1.19	0.2437	0.00069468	1439.51948
age	1	0.15516	0.11859	1.31	0.2007	0.01737	57.57421

			Collinearity Diagnostics						
Number	*Eigenvalue*	*Condition Index*	*Proportion of Variation*						
			thigh	*knee*	*ankle*	*abdomen*	*hip*	*chest*	*b*
1	9.92059	1.00	0.0867	0.0667	0.05956	0.01326	0.00470	0.00906	0.0
2	0.06033	12.82378	0.00047295	0.00011748	0.00019889	9.451604E−8	0.00012200	0.00003164	0.0
3	0.00981	31.79265	0.056839	0.050118	0.28833	0.04441	0.066810	0.00851	0.03
4	0.00375	51.41663	0.009473	0.00571	0.63631	0.06300	0.00007925	0.00372	0.0
5	0.00225	66.37126	0.09671	0.04462	0.01868	0.00444	0.00647	0.01202	0.3
6	0.00133	86.24990	0.07840	0.01721	0.01308	0.00162	0.00141	0.20594	0.4
7	0.00068974	119.92959	0.29187	0.00355	0.01292	0.53277	0.00973	0.28836	0.0
8	0.00051282	139.08671	0.17688	0.68267	0.00010658	0.00020234	0.16883	0.18193	0.0
9	0.00039296	158.88867	0.06369	0.00802	0.00567	0.30400	0.38438	0.18188	0.1
10	0.00033100	173.12335	0.29130	0.23759	0.02464	0.05398	0.42831	0.11760	0.0

Body Fat Data

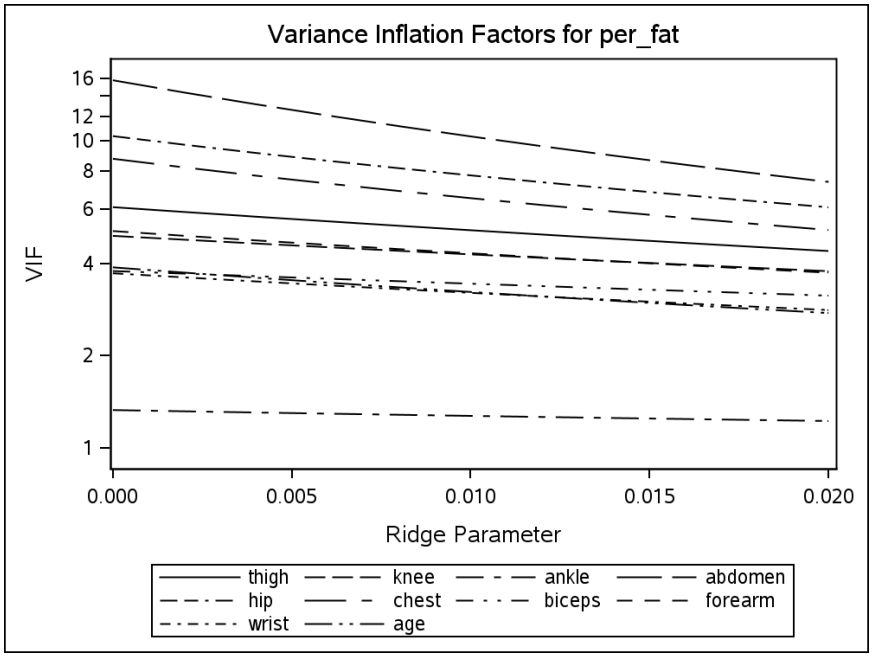
Multiple Regression

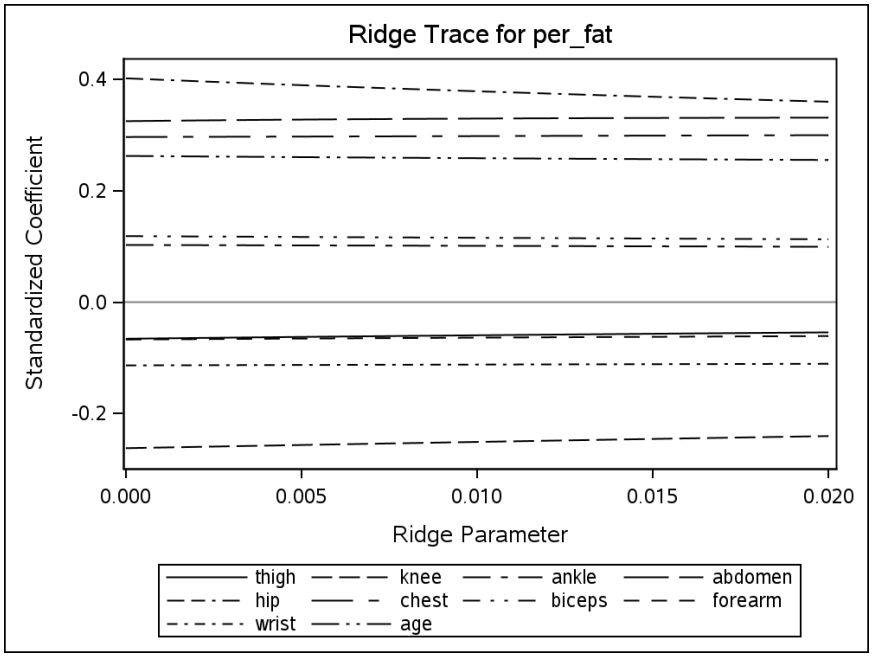
Ridge and Check for Collinearity

The REG Procedure

Model: MODEL1

Dependent Variable: per_fat




Ridge Trace for per_fat
Standardized Coefficient
0.4
0.2
0.0
-0.2
0.000
0.005
0.010
0.015
0.020
Ridge Parameter
thigh
knee
ankle
abdomen
hip
chest
biceps
forearm
wrist
age

Body Fat Data

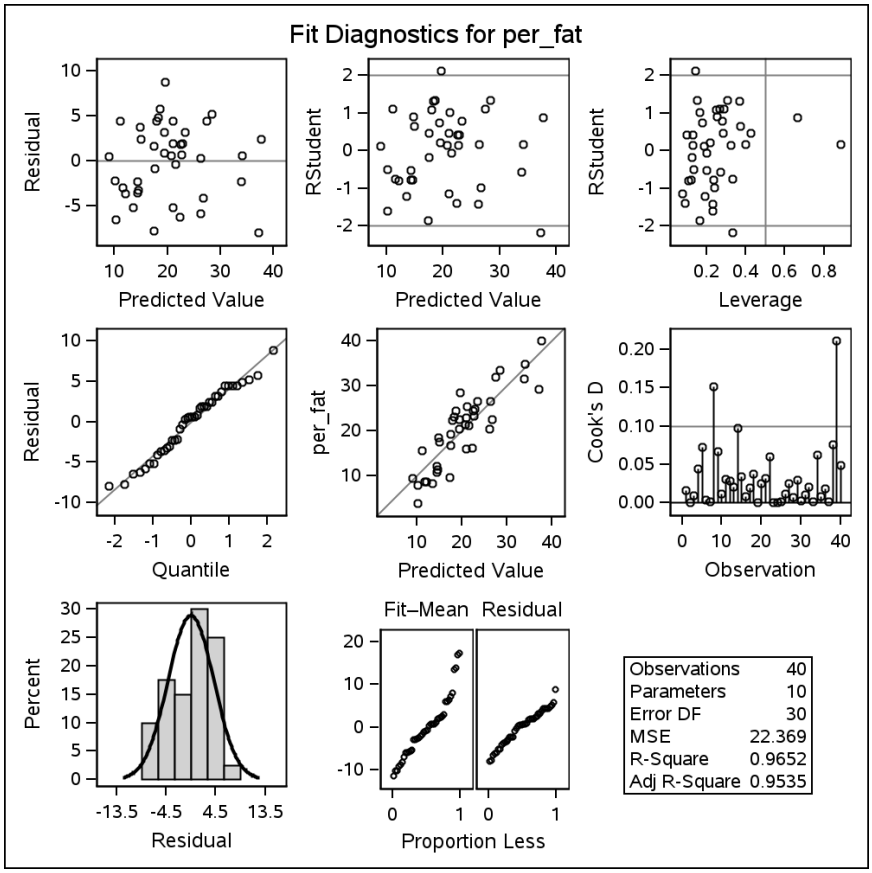
Multiple Regression

Ridge and Check for Collinearity

The REG Procedure

Model: MODEL1

Dependent Variable: per_fat

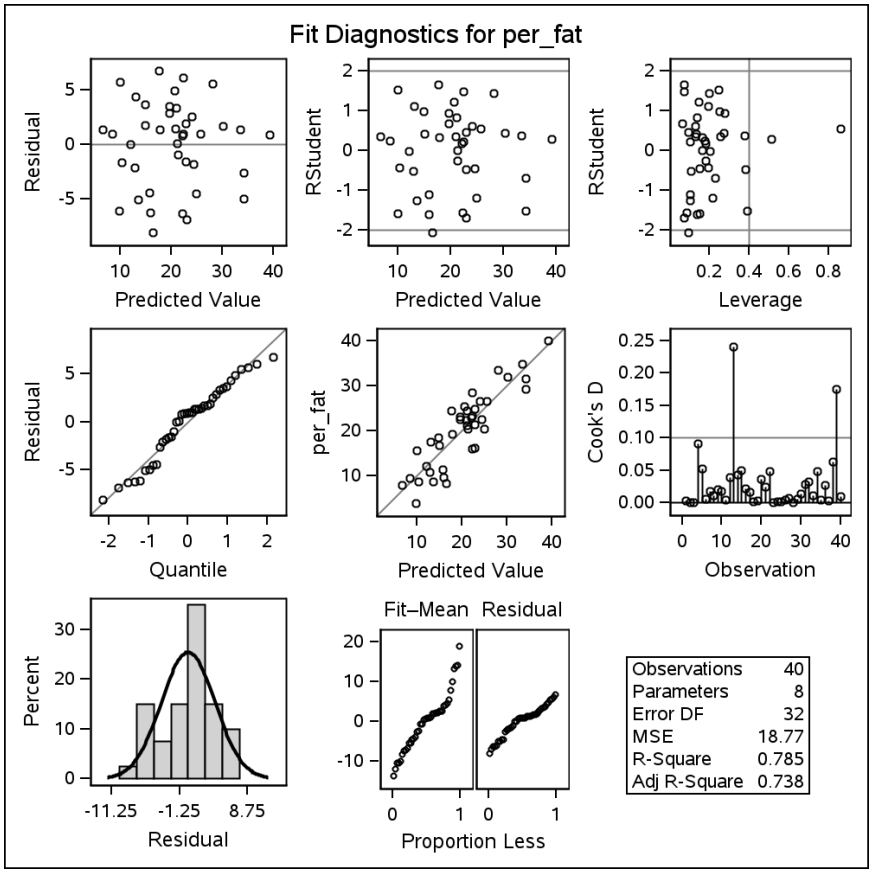


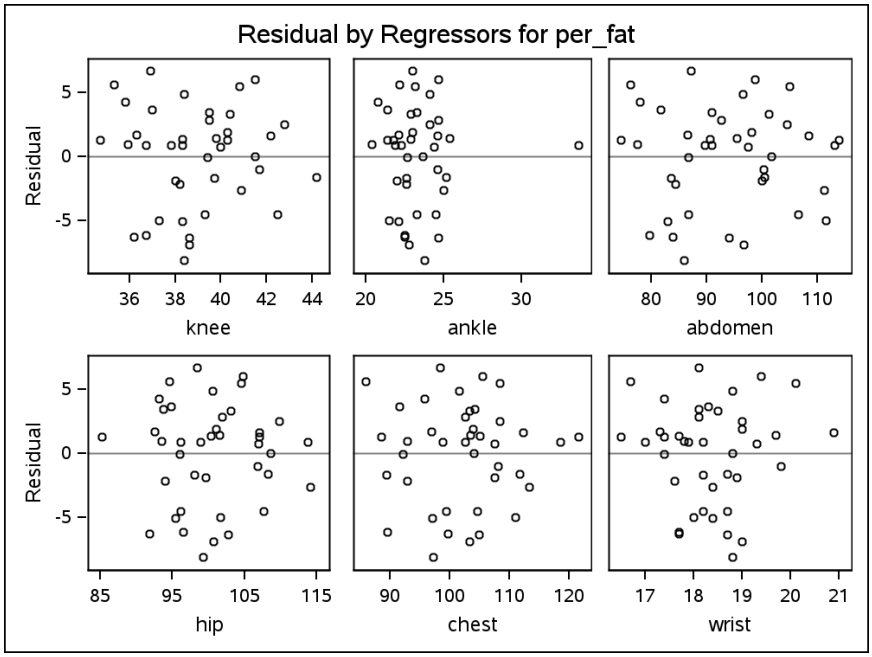
Body Fat Data

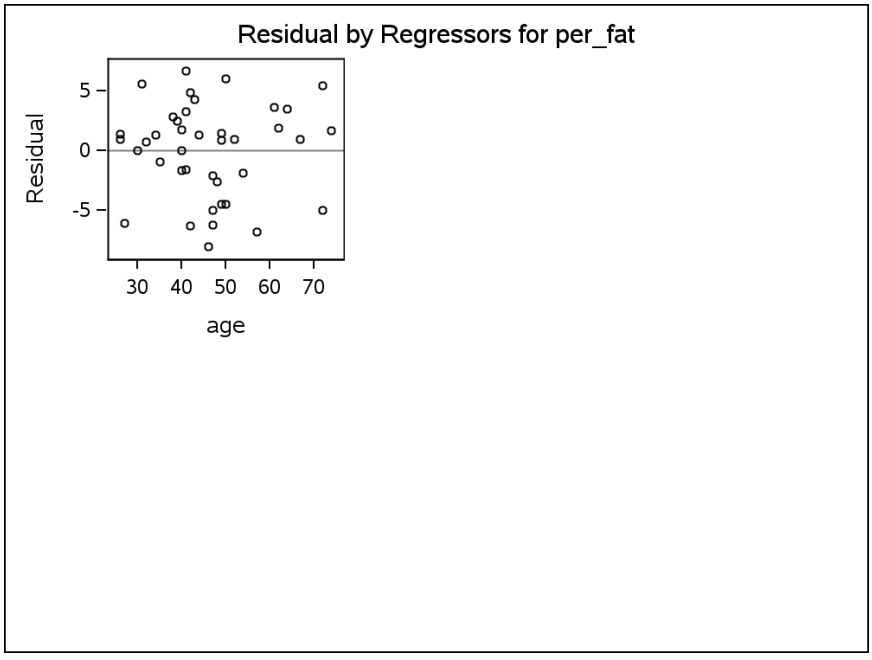
Forward Regression

Number of Observations Read	40
Number of Observations Used	40

Summary of Forward Selection							
Step	*Variable Entered*	*Number Vars In*	*Partial R-Square*	*Model R-Square*	*C(p)*	*F Value*	*Pr > F*
1	abdomen	1	0.6874	0.6874	7.0683	83.54	<.0001
2	knee	2	0.0349	0.7223	4.2578	4.65	0.0376
3	ankle	3	0.0213	0.7436	3.3243	2.99	0.0924
4	chest	4	0.0110	0.7545	3.8119	1.57	0.2192
5	age	5	0.0119	0.7664	4.1756	1.73	0.1973
6	hip	6	0.0138	0.7802	4.2766	2.07	0.1597
7	wrist	7	0.0048	0.7850	5.6143	0.72	0.4039



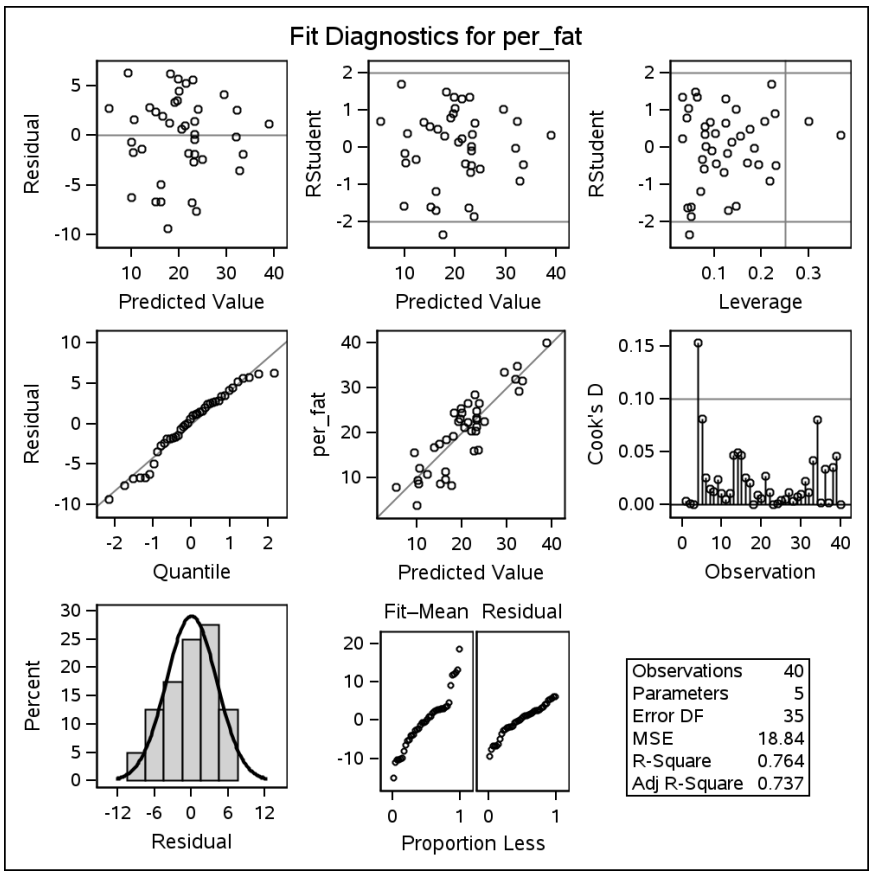

Residual by Regressors for per_fat
Residual
5
0
-5
36 38 40 42 44
knee
20 25 30
ankle
80 90 100 110
abdomen
85 95 105 115
hip
90 100 110 120
chest
17 18 19 20 21
wrist

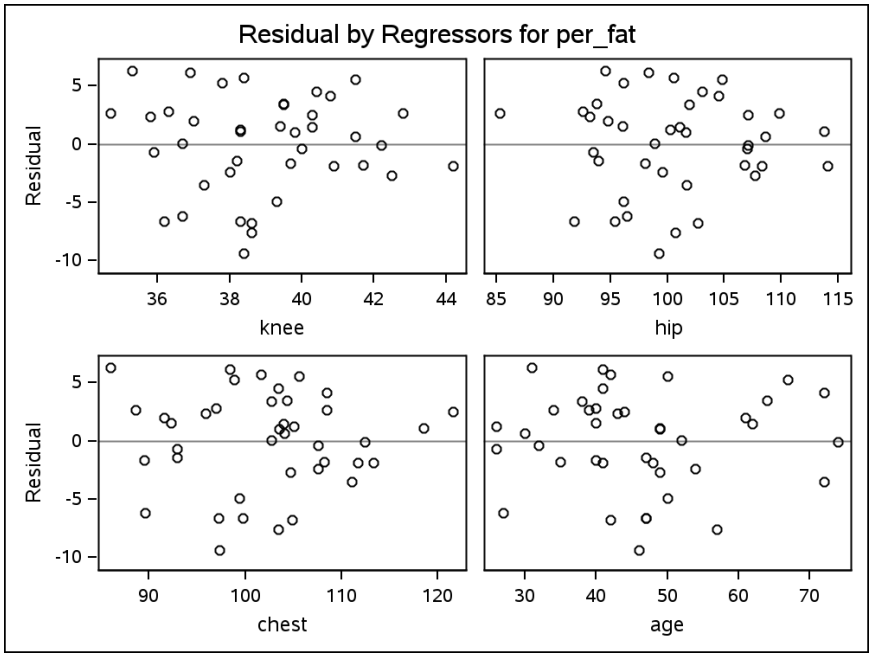

Residual by Regressors for per_fat
Residual
5
0
-5
30 40 50 60 70
age

C(p) Selection Method

Number of Observations Read	40
Number of Observations Used	40

Number in Model	*C(p)*	*R-Square*	*Variables in Model*
4	2.5108	0.7640	knee hip chest age
4	3.3045	0.7582	knee abdomen hip age
3	3.3243	0.7436	knee ankle abdomen
5	3.4923	0.7714	knee abdomen hip chest age
5	3.5779	0.7708	knee hip chest wrist age





Body Fat Data

Stepwise Regression

The GLMSELECT Procedure

Data Set	WORK.NEW_BFAT
Dependent Variable	per_fat
Selection Method	Stepwise
Select Criterion	SBC
Stop Criterion	SBC
Choose Criterion	C(p)
Effect Hierarchy Enforced	None

Number of Observations Read	40
Number of Observations Used	40

Dimensions	
Number of Effects	11
Number of Parameters	11

Body Fat Data

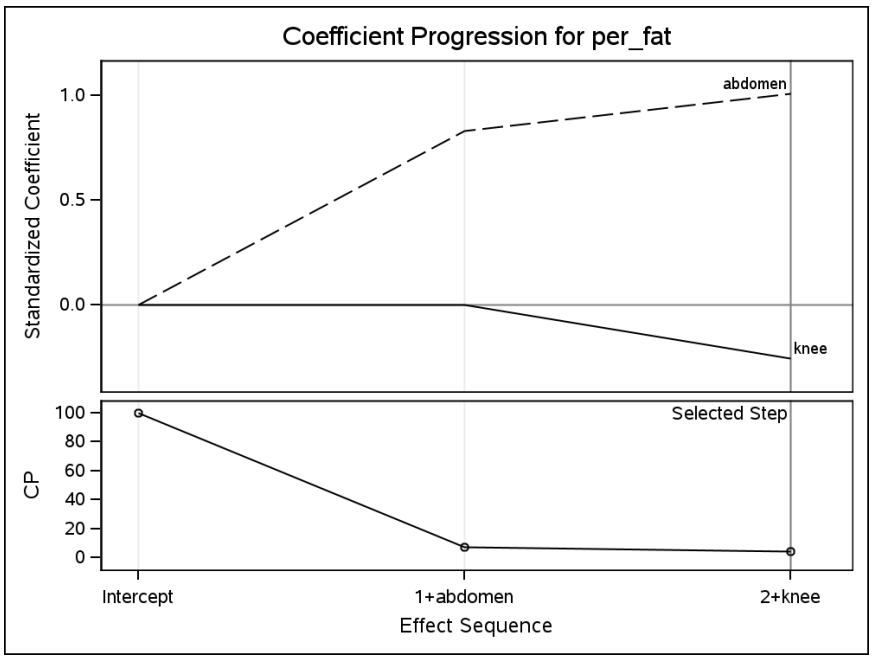
Stepwise Regression

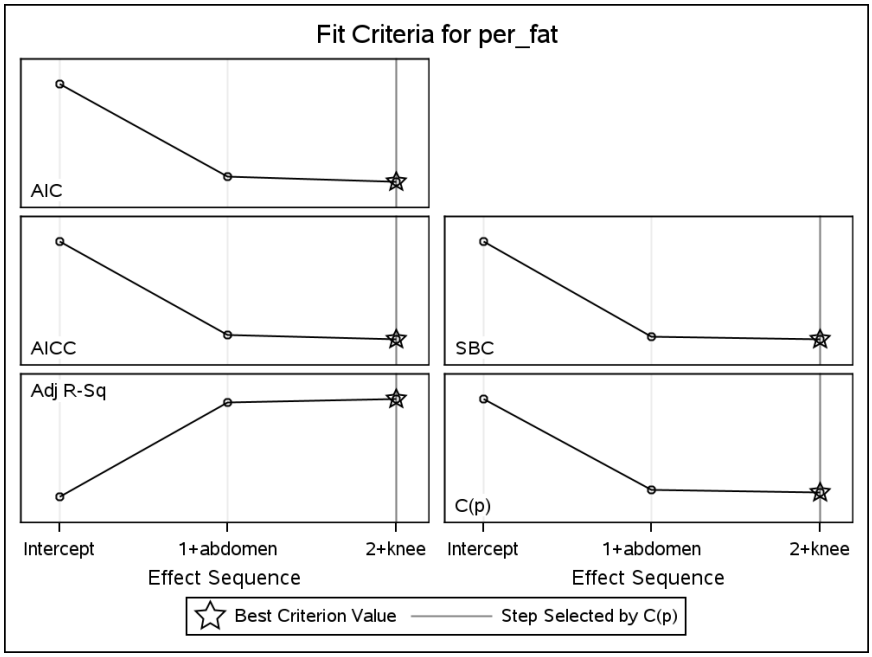
The GLMSELECT Procedure

Stepwise Selection Summary					
Step	*Effect Entered*	*Effect Removed*	*Number Effects In*	*CP*	*SBC*
0	*Intercept*		1	99.7544	173.5432
1	*abdomen*		2	7.0683	130.7247
2	*knee*		3	4.2578*	129.6760*
** Optimal Value of Criterion*					

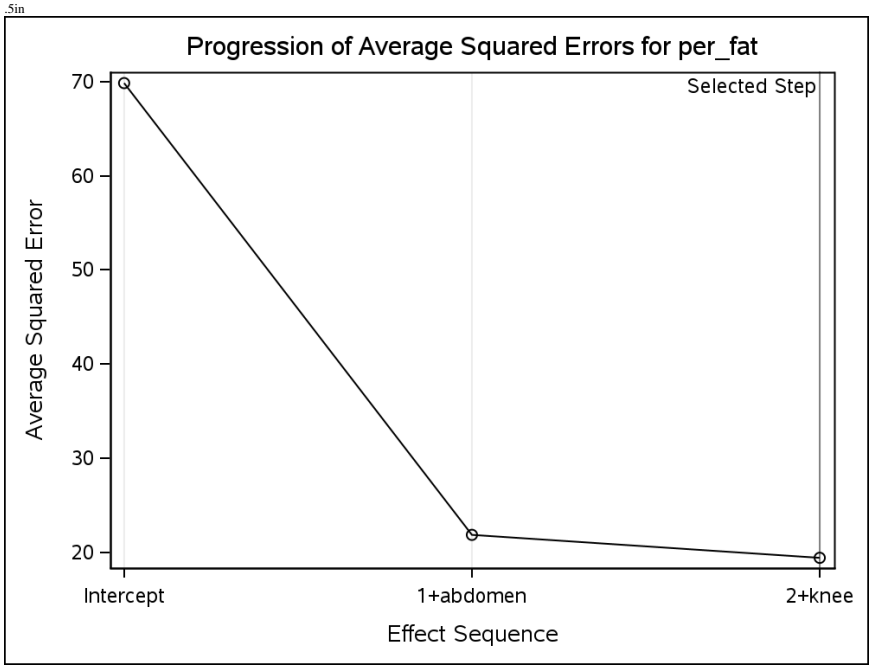
Selection stopped at a local minimum of the SBC criterion.

Stop Details				
Candidate For	*Effect*	*Candidate SBC*		*Compare SBC*
Entry	ankle	130.1737	>	129.6760
Removal	knee	130.7247	>	129.6760




Fit Criteria for per_fat
AIC
AICC
SBC
Adj R-Sq
C(p)
Intercept
1+abdomen
2+knee
Effect Sequence
Best Criterion Value
Step Selected by C(p)

.5in


Progression of Average Squared Errors for per_fat
Selected Step
Average Squared Error
70
60
50
40
30
20
Intercept
1+abdomen
2+knee
Effect Sequence

Body Fat Data

Stepwise Regression

The GLMSELECT Procedure

Selected Model

Effects:	Intercept knee abdomen

Analysis of Variance				
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*
Model	2	2018.05481	1009.02740	48.11
Error	37	775.96894	20.97213	
Corrected Total	39	2794.02375		

Root MSE	4.57953
Dependent Mean	20.28750
R-Square	0.7223
Adj R-Sq	0.7073
AIC	166.60932
AICC	167.75218
BIC	126.87841
C(p)	4.25778
SBC	129.67596

Parameter Estimates				
Parameter	*DF*	*Estimate*	*Standard Error*	*t Value*
Intercept	1	−15.275023	13.314855	−1.15
knee	1	−0.983639	0.456035	−2.16
abdomen	1	0.788591	0.093850	8.40

Body Fat Data

Lasso Regression

Standardize the data

The GLMSELECT Procedure

Data Set	WORK.DATA1
Dependent Variable	per_fat
Selection Method	LASSO
Stop Criterion	SBC
Choose Criterion	C(p)
Effect Hierarchy Enforced	None

Number of Observations Read	40
Number of Observations Used	40

Dimensions	
Number of Effects	9
Number of Parameters	9

Body Fat Data

Lasso Regression

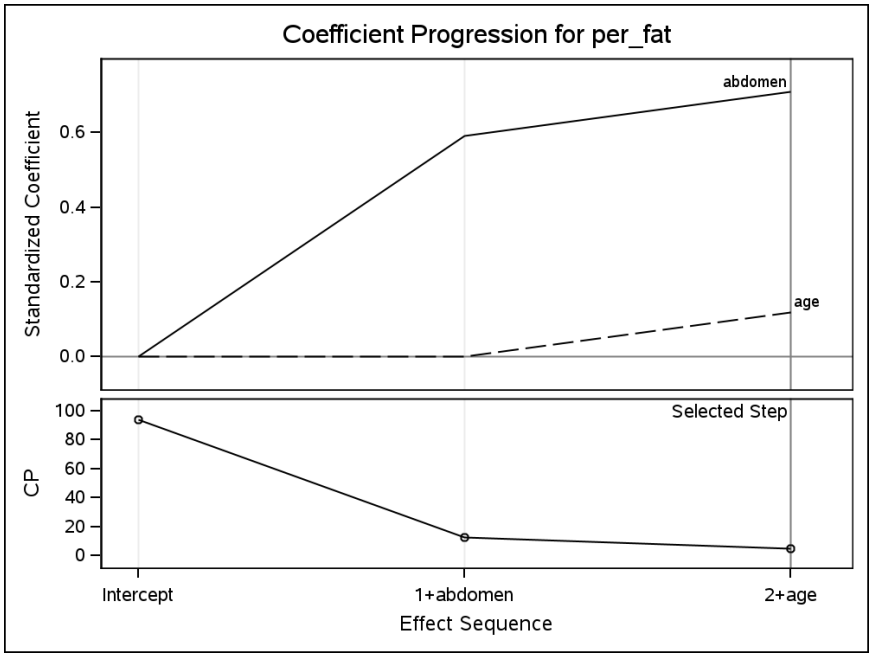
Standardize the data

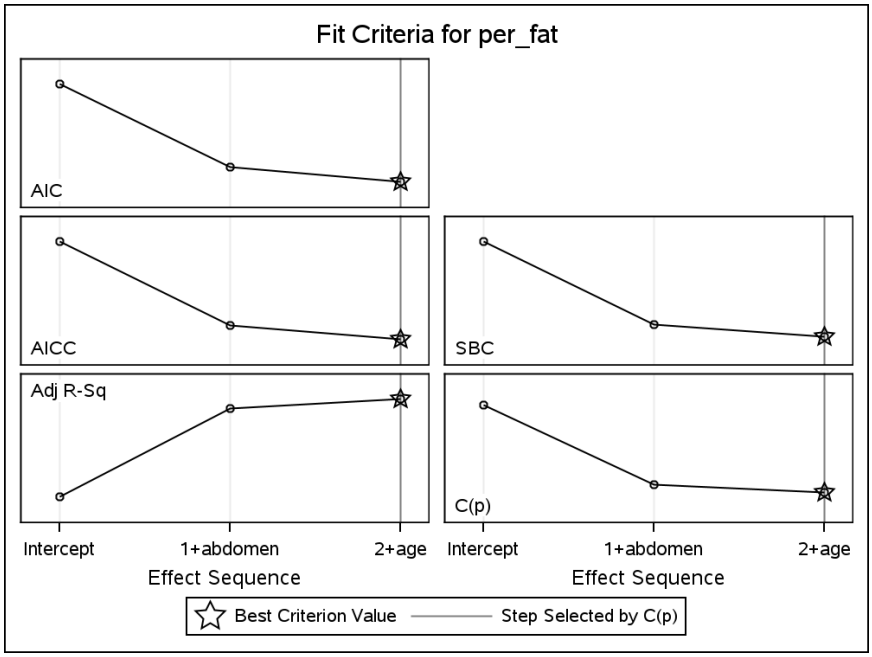
The GLMSELECT Procedure

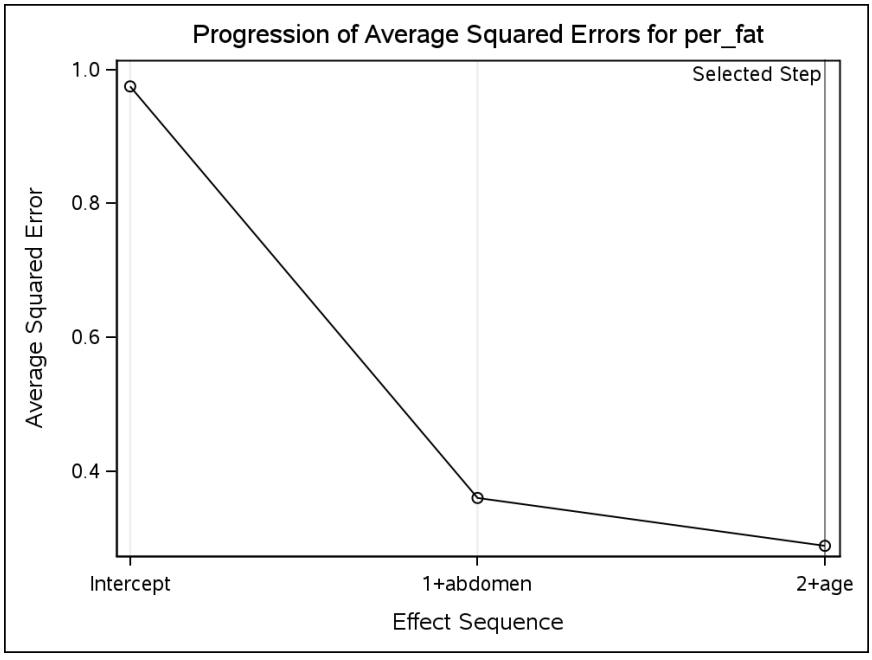
LASSO Selection Summary					
Step	*Effect Entered*	*Effect Removed*	*Number Effects In*	*CP*	*SBC*
0	*Intercept*		1	94.0674	2.6762
1	*abdomen*		2	12.7016	−33.5390
2	*age*		3	5.0045*	−38.7315*
* Optimal Value of Criterion					

Selection stopped at a local minimum of the SBC criterion.

Stop Details				
Candidate For	*Effect*	*Candidate SBC*		*Compare SBC*
Entry	ankle	−35.0892	>	−38.7315




Fit Criteria for per_fat
AIC
AICC
SBC
Adj R-Sq
C(p)
Intercept
1+abdomen
2+age
Effect Sequence
Best Criterion Value
Step Selected by C(p)


Progression of Average Squared Errors for per_fat
Selected Step
Average Squared Error
1.0
0.8
0.6
0.4
Intercept
1+abdomen
2+age
Effect Sequence

Body Fat Data

Lasso Regression

Standardize the data

The GLMSELECT Procedure

Selected Model

Effects:	Intercept abdomen age

Analysis of Variance				
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*
Model	2	27.48183	13.74092	44.14
Error	37	11.51817	0.31130	
Corrected Total	39	39.00000		

Root MSE	0.55794
Dependent Mean	2.55351E−16
R-Square	0.7047
Adj R-Sq	0.6887
AIC	−1.79816
AICC	−0.65531
BIC	−41.64633
C(p)	5.00447
SBC	−38.73153

Parameter Estimates		
Parameter	*DF*	*Estimate*
Intercept	1	−1.15369E−15
abdomen	1	0.710335
age	1	0.118158

Body Fat Data

Correlation and PCR

The CORR Procedure

9 Variables:	per_fat thigh knee ankle abdomen biceps forearm age wrist

Simple Statistics						
Variable	*N*	*Mean*	*Std Dev*	*Median*	*Minimum*	*Maximum*
per_fat	40	20.28750	8.46414	21.25000	3.70000	40.10000
thigh	40	59.80500	4.38371	59.30000	51.70000	68.10000
knee	40	38.95000	2.22665	38.60000	34.70000	44.20000
ankle	40	23.36000	2.10016	22.95000	20.40000	33.70000
abdomen	40	93.68000	10.81974	93.40000	74.60000	113.90000
biceps	40	32.71500	2.69791	32.50000	27.90000	38.20000
forearm	40	28.95500	1.65327	29.05000	25.90000	33.10000
age	40	46.22500	12.51765	45.00000	26.00000	74.00000
wrist	40	18.35500	0.91370	18.25000	16.50000	20.90000

Pearson Correlation Coefficients, N = 40											
	per_fat	*thigh*	*knee*	*ankle*	*abdomen*	*biceps*	*forearm*	*age*	*wrist*	*Pper_fat*	*Pthig*
per_fat	1.00000	0.47372	0.43853	0.25554	0.82907	0.48619	0.37513	0.46200	0.33300		0.002
thigh	0.47372	1.00000	0.80987	0.25976	0.70771	0.69621	0.66683	−0.10114	0.43710	0.0020	
knee	0.43853	0.80987	1.00000	0.37828	0.69172	0.66624	0.64610	0.15929	0.67465	0.0046	<.000
ankle	0.25554	0.25976	0.37828	1.00000	0.21690	0.29313	0.26096	0.23404	0.36971	0.1115	0.105
abdomen	0.82907	0.70771	0.69172	0.21690	1.00000	0.62800	0.55114	0.38013	0.48249	<.0001	<.000
biceps	0.48619	0.69621	0.66624	0.29313	0.62800	1.00000	0.81422	−0.07299	0.52900	0.0015	<.000
forearm	0.37513	0.66683	0.64610	0.26096	0.55114	0.81422	1.00000	−0.14458	0.59679	0.0171	<.000
age	0.46200	−0.10114	0.15929	0.23404	0.38013	−0.07299	−0.14458	1.00000	0.40243	0.0027	0.534
wrist	0.33300	0.43710	0.67465	0.36971	0.48249	0.52900	0.59679	0.40243	1.00000	0.0358	0.004

Spearman Correlation Coefficients, N = 40											
	per_fat	*thigh*	*knee*	*ankle*	*abdomen*	*biceps*	*forearm*	*age*	*wrist*	*Pper_fat*	*Pthig*
per_fat	1.00000	0.47933	0.45617	0.30345	0.80955	0.51088	0.38282	0.46023	0.33012		0.001
thigh	0.47933	1.00000	0.82082	0.58577	0.75522	0.65418	0.67606	−0.04593	0.50559	0.0018	
knee	0.45617	0.82082	1.00000	0.68705	0.72661	0.64756	0.66582	0.12865	0.69781	0.0031	<.000
ankle	0.30345	0.58577	0.68705	1.00000	0.40165	0.47763	0.54526	0.09175	0.63911	0.0570	<.000
abdomen	0.80955	0.75522	0.72661	0.40165	1.00000	0.66814	0.59345	0.34460	0.49035	<.0001	<.000
biceps	0.51088	0.65418	0.64756	0.47763	0.66814	1.00000	0.79333	−0.07152	0.54926	0.0008	<.000
forearm	0.38282	0.67606	0.66582	0.54526	0.59345	0.79333	1.00000	−0.09070	0.61477	0.0148	<.000
age	0.46023	−0.04593	0.12865	0.09175	0.34460	−0.07152	−0.09070	1.00000	0.31752	0.0028	0.778
wrist	0.33012	0.50559	0.69781	0.63911	0.49035	0.54926	0.61477	0.31752	1.00000	0.0375	0.000

Kendall Tau b Correlation Coefficients, N = 40											
	per_fat	*thigh*	*knee*	*ankle*	*abdomen*	*biceps*	*forearm*	*age*	*wrist*	*Pper_fat*	*Pthig*
per_fat	1.00000	0.32990	0.30859	0.22251	0.61104	0.33205	0.23149	0.33335	0.23179		0.002
thigh	0.32990	1.00000	0.63172	0.45137	0.56114	0.48774	0.48503	−0.03641	0.33553	0.0028	
knee	0.30859	0.63172	1.00000	0.50552	0.57641	0.48223	0.48989	0.07427	0.52716	0.0053	<.000
ankle	0.22251	0.45137	0.50552	1.00000	0.31784	0.34975	0.40707	0.06789	0.45217	0.0449	<.000
abdomen	0.61104	0.56114	0.57641	0.31784	1.00000	0.47558	0.41302	0.22540	0.36675	<.0001	<.000
biceps	0.33205	0.48774	0.48223	0.34975	0.47558	1.00000	0.62764	−0.05455	0.39116	0.0026	<.000
forearm	0.23149	0.48503	0.48989	0.40707	0.41302	0.62764	1.00000	−0.07218	0.45850	0.0377	<.000
age	0.33335	−0.03641	0.07427	0.06789	0.22540	−0.05455	−0.07218	1.00000	0.22076	0.0027	0.743
wrist	0.23179	0.33553	0.52716	0.45217	0.36675	0.39116	0.45850	0.22076	1.00000	0.0377	0.002

Body Fat Data

Correlation and PCR

The PRINCOMP Procedure

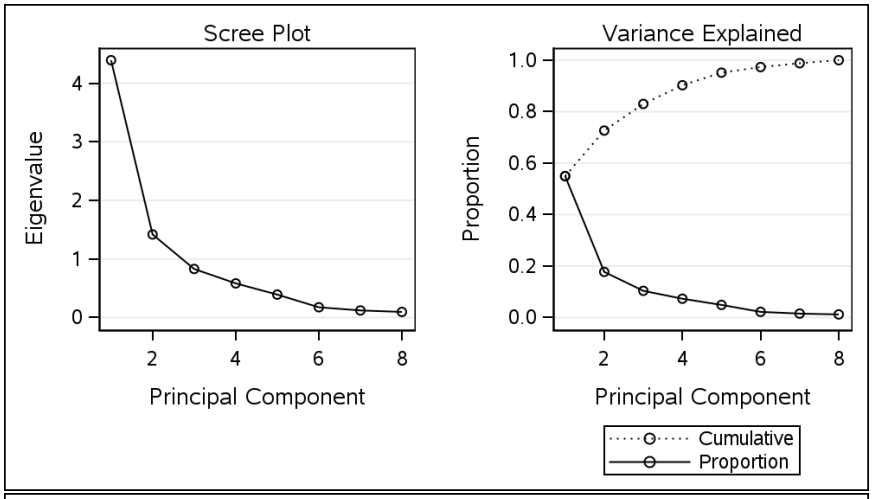
Observations	40
Variables	8

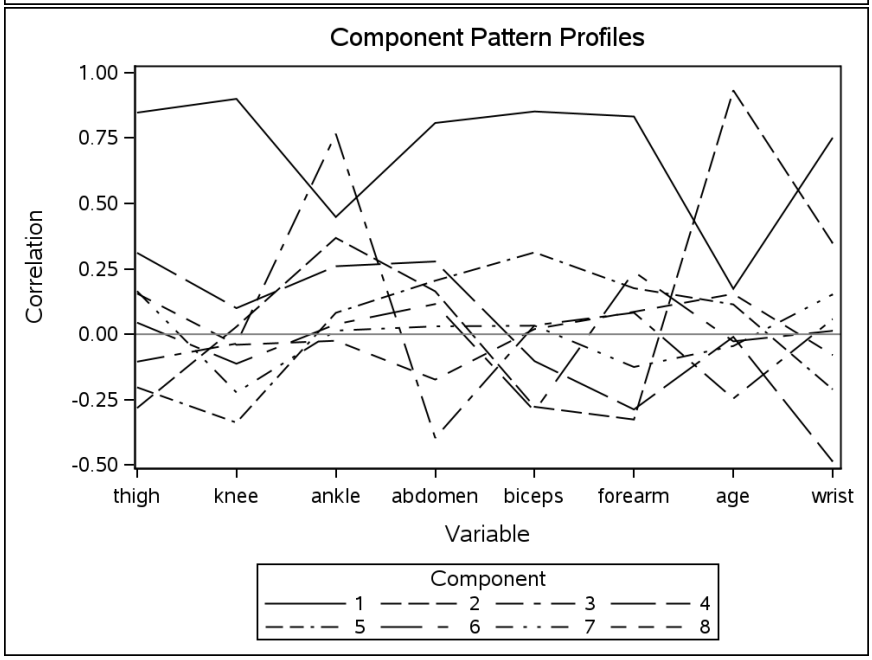
Simple Statistics								
	thigh	*knee*	*ankle*	*abdomen*	*biceps*	*forearm*	*age*	*wrist*
Mean	59.80500000	38.95000000	23.36000000	93.68000000	32.71500000	28.95500000	46.22500000	18.35500000
StD	4.38370818	2.22664517	2.10015872	10.81973979	2.69791467	1.65327181	12.51765420	0.91369916

Correlation Matrix								
	thigh	*knee*	*ankle*	*abdomen*	*biceps*	*forearm*	*age*	*wrist*
thigh	1.0000	0.8099	0.2598	0.7077	0.6962	0.6668	−.1011	0.4371
knee	0.8099	1.0000	0.3783	0.6917	0.6662	0.6461	0.1593	0.6746
ankle	0.2598	0.3783	1.0000	0.2169	0.2931	0.2610	0.2340	0.3697
abdomen	0.7077	0.6917	0.2169	1.0000	0.6280	0.5511	0.3801	0.4825
biceps	0.6962	0.6662	0.2931	0.6280	1.0000	0.8142	−.0730	0.5290
forearm	0.6668	0.6461	0.2610	0.5511	0.8142	1.0000	−.1446	0.5968
age	−.1011	0.1593	0.2340	0.3801	−.0730	−.1446	1.0000	0.4024
wrist	0.4371	0.6746	0.3697	0.4825	0.5290	0.5968	0.4024	1.0000

Eigenvalues of the Correlation Matrix				
	Eigenvalue	*Difference*	*Proportion*	*Cumulative*
1	4.39402697	2.97672436	0.5493	0.5493
2	1.41730261	0.58913360	0.1772	0.7264
3	0.82816901	0.24613648	0.1035	0.8299
4	0.58203253	0.19161336	0.0728	0.9027
5	0.39041917	0.21644990	0.0488	0.9515
6	0.17396927	0.05420012	0.0217	0.9732
7	0.11976915	0.02545787	0.0150	0.9882
8	0.09431129		0.0118	1.0000

Eigenvectors								
	Prin1	*Prin2*	*Prin3*	*Prin4*	*Prin5*	*Prin6*	*Prin7*	*Prin8*
thigh	0.404208	−.237052	−.115280	0.407865	−.325031	0.106336	0.476021	0.507179
knee	0.429302	0.024187	−.039453	0.130803	−.540938	−.270181	−.643088	−.131433
ankle	0.213791	0.309322	0.841908	0.340738	0.131026	0.091302	0.040033	−.081211
abdomen	0.385338	0.137843	−.435578	0.364530	0.327610	0.276500	0.086907	−.564420
biceps	0.406302	−.233101	0.037327	−.134983	0.500812	−.705042	0.094521	0.064330
forearm	0.397118	−.273967	0.090584	−.377010	0.281431	0.573251	−.362082	0.279239
age	0.082848	0.783370	−.270240	−.013216	0.182398	−.065103	−.129982	0.501856
wrist	0.358435	0.292366	0.063154	−.637981	−.337579	0.031099	0.440174	−.257599


Scree Plot
Eigenvalue
Principal Component
Variance Explained
Proportion
Principal Component
Cumulative
Proportion


Component Pattern Profiles
Correlation
thigh
knee
ankle
abdomen
biceps
forearm
age
wrist
Variable
Component
1
2
3
4
5
6
7
8

Body Fat Data

Correlation and PCR

The REG Procedure

Model: MODEL1

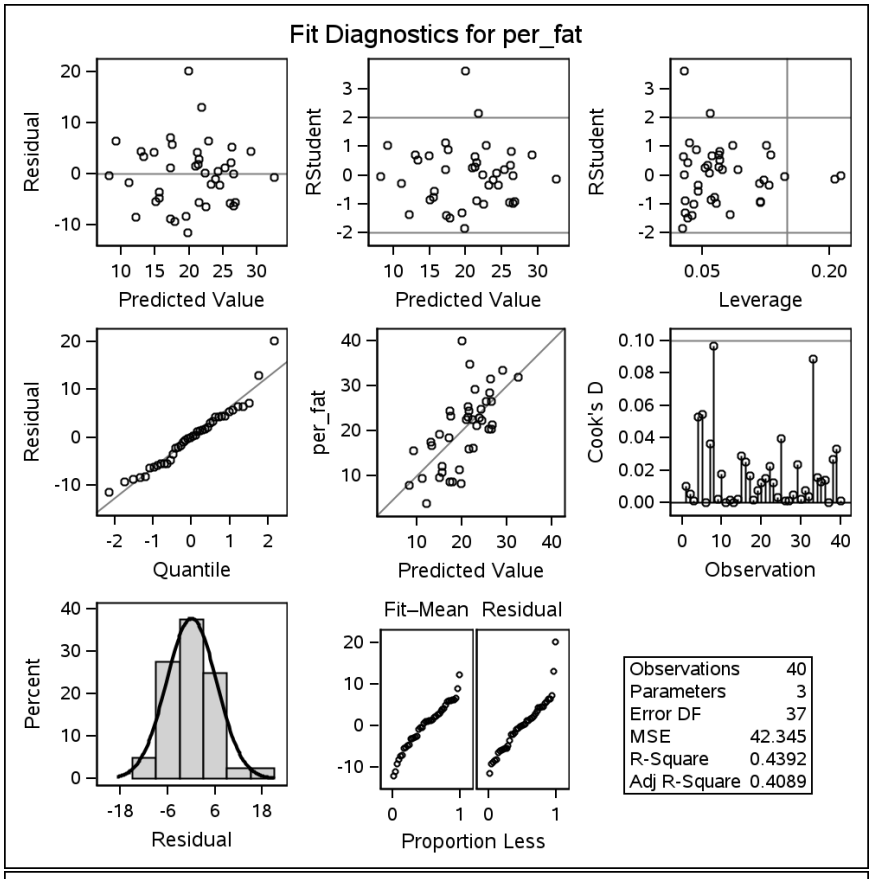
Dependent Variable: per_fat

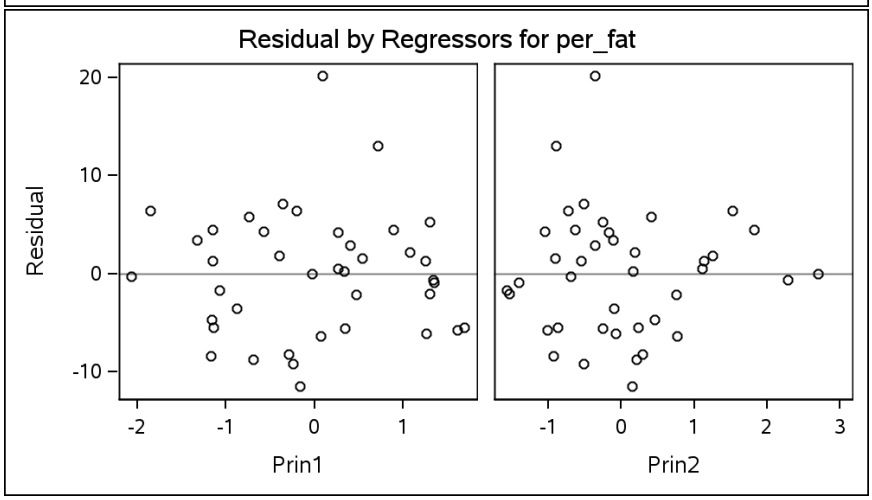
Number of Observations Read	40
Number of Observations Used	40

Analysis of Variance					
Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	2	1227.27023	613.63511	14.49	<.0001
Error	37	1566.75352	42.34469		
Corrected Total	39	2794.02375			

Root MSE	6.50728	*R-Square*	0.4392
Dependent Mean	20.28750	*Adj R-Sq*	0.4089
Coeff Var	32.07532		

Parameter Estimates						
Variable	*DF*	*Parameter Estimate*	*Standard Error*	*t Value*	*Pr > |t|*	*Type I SS*
Intercept	1	20.28750	1.02889	19.72	<.0001	16463
Prin1	1	5.07960	1.04200	4.87	<.0001	1006.29283
Prin2	1	2.38035	1.04200	2.28	0.0282	220.97740


Fit Diagnostics for per_fat
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
per_fat
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean Residual
Proportion Less
Observations 40
Parameters 3
Error DF 37
MSE 42.345
R-Square 0.4392
Adj R-Square 0.4089


Residual by Regressors for per_fat
Residual
Prin1
Prin2

C(p) Selection Method

Number of Observations Read	40
Number of Observations Used	40

Number in Model	*C(p)*	*R-Square*	*Variables in Model*
5	6.0597	0.7471	Prin1 Prin2 Prin3 Prin4 Prin5
6	7.0000	0.7550	Prin1 Prin2 Prin3 Prin4 Prin5 Prin6
4	14.6825	0.6682	Prin1 Prin2 Prin3 Prin5
4	14.7117	0.6680	Prin1 Prin3 Prin4 Prin5
5	15.6228	0.6761	Prin1 Prin2 Prin3 Prin5 Prin6
5	15.6520	0.6759	Prin1 Prin3 Prin4 Prin5 Prin6
4	15.8785	0.6594	Prin1 Prin2 Prin4 Prin5
5	16.8188	0.6672	Prin1 Prin2 Prin4 Prin5 Prin6
4	23.0823	0.6059	Prin1 Prin2 Prin3 Prin4
3	23.3345	0.5892	Prin1 Prin3 Prin5