Regression using Matricies

We can now present linear regression in matrix terms. Begin with the linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad \text{where } i = 1, \ldots, n$$

This implies

$$\begin{aligned} Y_1 &= 1\beta_0 + \beta_1 X_1 + \epsilon_1 \\ Y_2 &= 1\beta_0 + \beta_1 X_2 + \epsilon_2 \\ &\vdots \\ Y_n &= 1\beta_0 + \beta_1 X_n + \epsilon_n \end{aligned}$$

We can now arrange these into appropriate vectors and matrices.

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \qquad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

We can write the model using matrix terms very compactly as

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}}$$

Dimension
of product

Since

$$E\{Y_i\} = \beta_0 + \beta_1 X_i$$

The expected values of **Y** are expressed in the **Xβ** product matrix

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \underset{n\times 1}{\mathbf{X}\boldsymbol{\beta}}$$

Next we turn to the assumptions of the errors. Since $\mathbf{E}\{\varepsilon_i\} = 0$,

$$\underset{n\times 1}{\mathbf{E}\{\boldsymbol{\epsilon}\}} = \underset{n\times 1}{\mathbf{0}}$$

We also make the assumption that the errors have constant variance σ^2 and that they are independent of one another $\sigma(\varepsilon_i, \varepsilon_j)=0$ for $i \neq j$. The **variance-covariance matrix of the error terms** can therefore be expressed as

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}} = \begin{bmatrix} \sigma_\epsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\epsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\epsilon^2 \end{bmatrix}$$

This scalar matrix can also be expressed as

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\}} = \sigma^2 \underset{n\times n}{\mathbf{I}}$$

Thus, the **regression model** can be completely expressed in matrix terms as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\varepsilon}$ is a vector of independent random variables with $\mathbf{E}\{\boldsymbol{\varepsilon}\}=0$ and $\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}= \sigma^2\mathbf{I}$

Solving for β

We can re-express the **error vector** as

$$\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

The *method of least squares* minimizes the **sum of squared errors** which is expressed as

$$\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \underset{1\times n}{\mathbf{Y}'}\ \underset{n\times 1}{\mathbf{Y}} - \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 1}{\mathbf{Y}} - \underset{1\times n}{\mathbf{Y}'}\ \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}} + \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}}$$

Note that each term is a 1x1 matrix, which means that each term is equal to its transpose. We will re-write the third term **Y'X β** as its transpose **β'X'Y**

$$= \underset{1\times n}{\mathbf{Y}'}\ \underset{n\times 1}{\mathbf{Y}} - \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 1}{\mathbf{Y}} - \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 1}{\mathbf{Y}} + \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}}$$

Combining the two middle terms we get

$$= \underset{1\times n}{\mathbf{Y}'}\ \underset{n\times 1}{\mathbf{Y}} - 2\underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 1}{\mathbf{Y}} + \underset{1\times 2}{\boldsymbol{\beta}'}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}}$$

To find the values for the elements in **β** that minimize the equation, we differentiate with respect to β_0 and β_1 (**β**)

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\underset{1\times n}{\mathbf{Y}'}\underset{n\times 1}{\mathbf{Y}} - 2\underset{1\times 2}{\boldsymbol{\beta}'}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + \underset{1\times 2}{\boldsymbol{\beta}'}\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}})$$

Differentiating, we get...

$$= -2\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + 2\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}$$

Solving this equation for zero...

$$\begin{aligned}\underset{2\times 1}{\mathbf{0}} &= -2\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + 2\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}} \\ &= -\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} + \underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}\end{aligned}$$

Adding **X'Y** to both sides

$$\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} = \underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}$$

Now we pre-multiply both sides of the equation by $(\mathbf{X'X})^{-1}$

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}}$$

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}} = \underset{2\times 2}{\mathbf{I}}\underset{2\times 1}{\boldsymbol{\beta}}$$

This means

$$\underset{2\times 1}{\boldsymbol{\beta}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\underset{2\times n}{\mathbf{X}'}\underset{n\times 1}{\mathbf{Y}}$$

The vector of regression coefficients can be obtained directly through manipulation of the design matrix and the vector of outcomes.

$$\underset{2\times 1}{\boldsymbol{\beta}} = \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{2\times n}{\mathbf{X}'} \underset{n\times 1}{\mathbf{Y}}$$

```
> lm.1 = lm(wage ~ 1 + age, data = myData)

> Y = myData$wage
> X = matrix(c(rep(1, 5), myData$age), ncol = 2)

> b = solve(t(X) %*% X) %*% t(X) %*% Y
> b

                  [,1]
(Intercept) -3.8982759
age          0.4993678

> summary(lm.1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.8983    19.3658  -0.201    0.853
age           0.4994     0.6185   0.807    0.479
```

$$\mathbf{b} = \begin{bmatrix} -3.8983 \\ 0.4994 \end{bmatrix}$$

When we estimate the $\boldsymbol{\beta}$ matrix from sample data we refer to it as $\mathbf{b}$

Fitted Values

Let the vector of **fitted values** be

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

In matrix notation

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}} \; \underset{2\times 1}{\mathbf{b}}$$

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}}$$

```
> X %*% b

       [,1]
1 12.081494
2 12.580862
3 12.081494
4 12.580862
5  9.085287

> fitted(lm.1)

        1         2         3         4         5 
12.081494 12.580862 12.081494 12.580862  9.085287 
```

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}}$$

Substituting the expression for $\mathbf{b}$

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\ \underset{2\times n}{\mathbf{X}'}\ \underset{n\times 1}{\mathbf{Y}}$$

This is often expressed as

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times n}{\mathbf{H}}\ \underset{n\times 1}{\mathbf{Y}}$$

where

$$\underset{n\times n}{\mathbf{H}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}}\ \underset{2\times n}{\mathbf{X}'}$$

H is a square *n* x *n* matrix that can be created completely from the design matrix and its transpose.

```
> h = X %*% solve(t(X) %*% X) %*% t(X)

> h

           1           2          3           4           5
1 0.21839080  0.24137931 0.21839080  0.24137931  0.08045977
2 0.24137931  0.29310345 0.24137931  0.29310345 -0.06896552
3 0.21839080  0.24137931 0.21839080  0.24137931  0.08045977
4 0.24137931  0.29310345 0.24137931  0.29310345 -0.06896552
5 0.08045977 -0.06896552 0.08045977 -0.06896552  0.97701149
```

The fitted values can be expressed as linear combinations of the response vector **Y** using coefficients found in **H**.

```
> h %*% Y

       [,1]
1 12.081494
2 12.580862
3 12.081494
4 12.580862
5  9.085287
```

Because of this, **H** is often referred to as the *hat matrix*.

H is a symmetric matrix and also has a special property called *idempotency*:

$$\underset{n\times n}{\mathbf{H}}\;\underset{n\times n}{\mathbf{H}} = \underset{n\times n}{\mathbf{H}}$$

```
> h %*% h

           1           2          3           4           5
1 0.21839080  0.24137931 0.21839080  0.24137931  0.08045977
2 0.24137931  0.29310345 0.24137931  0.29310345 -0.06896552
3 0.21839080  0.24137931 0.21839080  0.24137931  0.08045977
4 0.24137931  0.29310345 0.24137931  0.29310345 -0.06896552
5 0.08045977 -0.06896552 0.08045977 -0.06896552  0.97701149
```

Residuals

Let the vector of residuals be

$$\underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

In matrix notation

$$\begin{aligned} \underset{n\times 1}{\mathbf{e}} &= \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 2}{\mathbf{X}} \underset{2\times 1}{\mathbf{b}} \\ &= \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} \end{aligned}$$

Since the fitted values can be expressed as **HY**, the residual vector can also be expressed as

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times n}{\mathbf{H}} \; \underset{n\times 1}{\mathbf{Y}}$$

The residuals can also be expressed as linear combinations of the response vector **Y**.

Using the distributive property as it relates to matrices, we re-write this as

$$\underset{n\times 1}{\mathbf{e}} = \left(\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}} \right) \underset{n\times 1}{\mathbf{Y}}$$

The matrix (**1** - **H**), like **H**, is *symmetric* and *idempotent*.

$$\underset{n\times 1}{\mathbf{e}} = (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})\underset{n\times 1}{\mathbf{Y}}$$

```
> i = diag(5)

> (i - h) %*% Y

        [,1]
1 -0.08149425
2 -4.58086207
3  4.17850575
4  1.06913793
5 -0.58528736

> resid(lm.1)

          1           2           3           4           5
-0.08149425 -4.58086207  4.17850575  1.06913793 -0.58528736
```

Variance–Covariance Matrix of the Residuals

Since $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, the variance–covariance matrix of $\mathbf{e}$ is can be expressed as

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = \boldsymbol{\sigma}^2\{(\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})\underset{n\times 1}{\mathbf{Y}}\}$$

Using rules of variances, the right-hand side is re-expressed as

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})\,\boldsymbol{\sigma}^2\{\underset{n\times 1}{\mathbf{Y}}\}\,(\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})'$$

Now, for the model with normally distributed errors,

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\sigma}^2\{\mathbf{e}\} = \sigma^2 \mathbf{I}$$

Also, $(\mathbf{I} - \mathbf{H}) = (\mathbf{I} - \mathbf{H})'$ because of the symmetry of this matrix. Hence,

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})\, \sigma^2\{\underset{n\times n}{\mathbf{I}}\}\, (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})$$

Which reduces to

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = \sigma^2\, (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})\, (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})$$

Since $(\mathbf{1} - \mathbf{H})$ is idempotent, $(\mathbf{1} - \mathbf{H})\,(\mathbf{1} - \mathbf{H}) = (\mathbf{1} - \mathbf{H})$ and thus

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = \sigma^2\, (\underset{n\times n}{\mathbf{I}} - \underset{n\times n}{\mathbf{H}})$$

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = \sigma^2 \left(\underset{n \times n}{\mathbf{I}} - \underset{n \times n}{\mathbf{H}} \right)$$

```
# Obtain the estimate of the MSE
> MSE = 13.312

> MSE * (i - h)

          1         2         3         4         5
1 10.404782 -3.213241 -2.907218 -3.213241 -1.071080
2 -3.213241  9.410207 -3.213241 -3.901793  0.918069
3 -2.907218 -3.213241 10.404782 -3.213241 -1.071080
4 -3.213241 -3.901793 -3.213241  9.410207  0.918069
5 -1.071080  0.918069 -1.071080  0.918069  0.306023
```

The variance of residual e_i is

$$\sigma^2\{e_i\} = \sigma^2(1 - h_{ii})$$

where h_{ii} is the i^{th} element on the main diagonal of the hat matrix

The covariance between residuals e_i and ej $(i \neq j)$ is

$$\sigma\{e_i, e_j\} = \sigma^2(0 - h_{ij})$$

$$= -h_{ij}\sigma^2$$

where h_{ij} is the element in the i^{th} row and j^{th} column of the hat matrix

When the MSE is used as an estimator of σ²

$$s^2\{e_i\} = \mathrm{MSE}(1 - h_{ii})$$

$$s\{e_i, e_j\} = -h_{ij}\mathrm{MSE}$$

For our data

$$s^2\{e_1\} = 15.33642(1 - 0.33064516) = 10.26551$$

$$s\{e_1, e_2\} = -0.33064516(15.33642) = -5.070913$$

The main diagonal element h_{ii} of the hat matrix can be computed directly as

$$h_{ii} = \underset{1\times p}{\mathbf{X}_i'} \, (\underset{p\times n}{\mathbf{X}'} \, \underset{n\times p}{\mathbf{X}})^{-1} \, \underset{p\times 1}{\mathbf{X}_i}$$

where $\mathbf{X}_i$ is the vector of predictor values for the i^{th} case defined as

$$\underset{p\times 1}{\mathbf{X}_i} = \begin{bmatrix} 1 \\ X_{i,1} \\ \vdots \\ X_{i,p-1} \end{bmatrix}$$

$$h_{ii} = \underset{1\times p}{\mathbf{X}_i'} \, (\underset{p\times n}{\mathbf{X}'} \, \underset{n\times p}{\mathbf{X}})^{-1} \, \underset{p\times 1}{\mathbf{X}_i}$$

```r
# Obtain the predictor vector for the 1st observation
> X1 = X[1, ]

> h1 = t(X1) %*% solve(t(X) %*% X) %*% X1

> h1

          [,1]
[1,] 0.2183908
```

$$s^2\{e_i\} = \text{MSE}(1 - h_{ii})$$

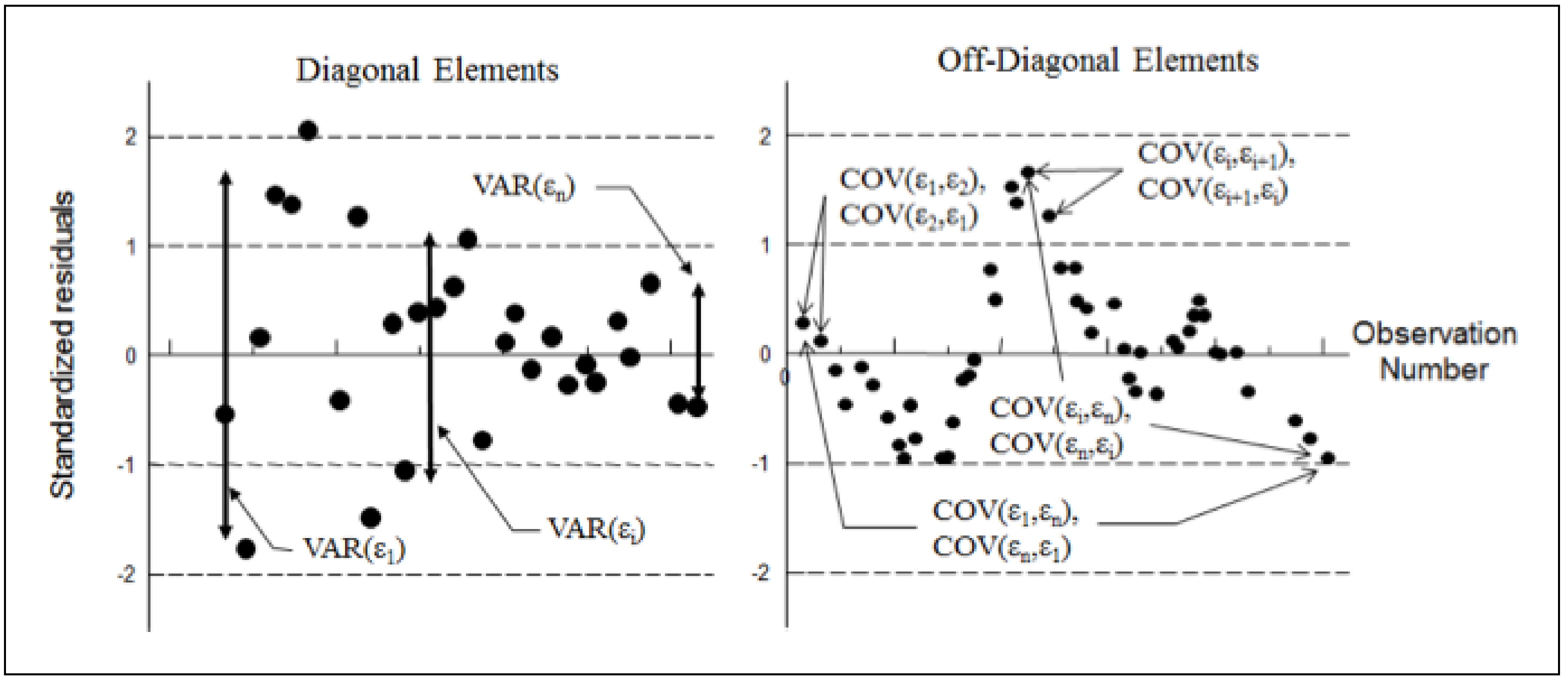
```r
# Compute the residual variance estimate for the 1st observation
> MSE * (1 - h1)

         [,1]
[1,] 10.40478
```

Recall the estimated variance–covariance matrix

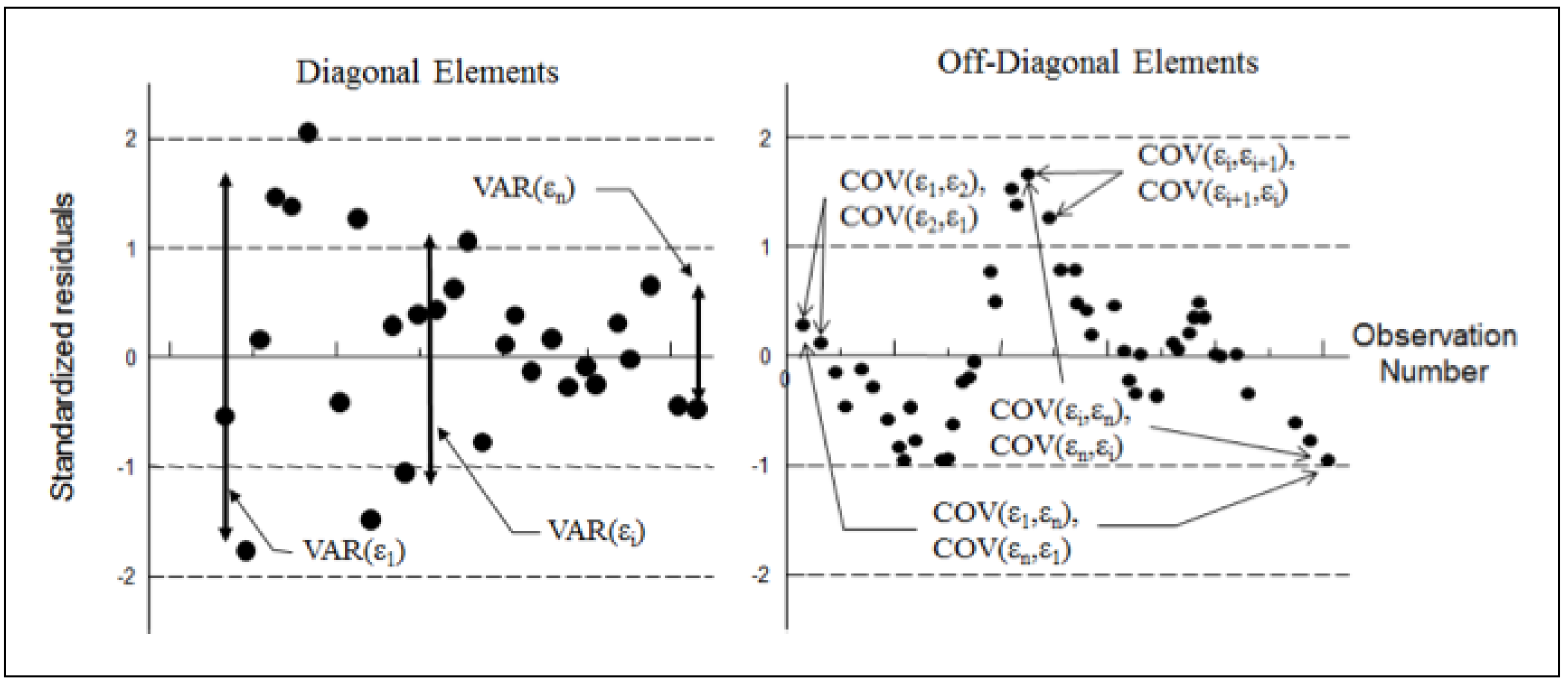
$$\hat{\sigma}^2\{\mathbf{e}\} = \begin{bmatrix} \boxed{10.404782} & -3.213241 & -2.907218 & -3.213241 & -1.071080 \\ -3.213241 & \boxed{9.410207} & -3.213241 & -3.901793 & 0.918069 \\ -2.907218 & -3.213241 & \boxed{10.404782} & -3.213241 & -1.071080 \\ -3.213241 & -3.901793 & -3.213241 & \boxed{9.410207} & 0.918069 \\ -1.071080 & 0.918069 & -1.071080 & 0.918069 & \boxed{0.306023} \end{bmatrix}$$

The **variances in residuals** (shown in the figure on the left) correspond to the diagonal elements of the variance-covariance matrix



$$\hat{\sigma}^2\{\mathbf{e}\} = \begin{bmatrix} 10.404782 & -3.213241 & -2.907218 & -3.213241 & -1.071080 \\ -3.213241 & 9.410207 & -3.213241 & -3.901793 & 0.918069 \\ -2.907218 & -3.213241 & 10.404782 & -3.213241 & -1.071080 \\ -3.213241 & -3.901793 & -3.213241 & 9.410207 & 0.918069 \\ -1.071080 & 0.918069 & -1.071080 & 0.918069 & 0.306023 \end{bmatrix}$$

The **covariances in residuals** (shown in the figure on the right) correspond to the off-diagonal elements of the variance-covariance matrix



The estimated variances are oftentimes substantially different so the magnitude of each residual is considered relative to its standard error

$$r_i = \frac{e_i}{s\{e_i\}} \qquad \text{where} \qquad s\{e_i\} = \sqrt{\text{MSE}(1 - h_{ii})}$$

These are referred to as **studentized residuals**.

For example, $$r_1 = \frac{-0.08149425}{\sqrt{10.404782}} = -0.02526449$$

```
# Compute the residual standard errors for all observations
> se = sqrt(diag(MSE * (1 - h)))

        1         2         3         4         5 
3.2256444 3.0676061 3.2256444 3.0676061 0.5531934 

# Compute studentized residuals
> resid(lm.1) / se

          1           2           3           4           5 
-0.02526449 -1.49330194  1.29540187  0.34852517 -1.05801571 

> rstandard(lm.1)
          1           2           3           4           5 
-0.02526434 -1.49329323  1.29539431  0.34852314 -1.05800954 
```

The hat matrix was defined earlier as

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

The fitted values, the residuals, and the variance–covariance matrix of the residuals can be expressed via the hat matrix

$$\underset{n \times 1}{\hat{\mathbf{Y}}} = \mathbf{H}\mathbf{Y}$$

$$\underset{n \times 1}{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\underset{n \times n}{\boldsymbol{\sigma}^2\{\mathbf{e}\}} = \sigma^2 \ (\mathbf{I} - \mathbf{H})$$

The hat matrix is also useful in identifying outlying X-values. In particular the elements on the main diagonal

$$h = \begin{matrix} \boxed{0.21839080} & 0.24137931 & 0.21839080 & 0.24137931 & 0.08045977 \\ 0.24137931 & \boxed{0.29310345} & 0.24137931 & 0.29310345 & -0.06896552 \\ 0.21839080 & 0.24137931 & \boxed{0.21839080} & 0.24137931 & 0.08045977 \\ 0.24137931 & 0.29310345 & 0.24137931 & \boxed{0.29310345} & -0.06896552 \\ 0.08045977 & -0.06896552 & 0.08045977 & -0.06896552 & \boxed{0.97701149} \end{matrix}$$

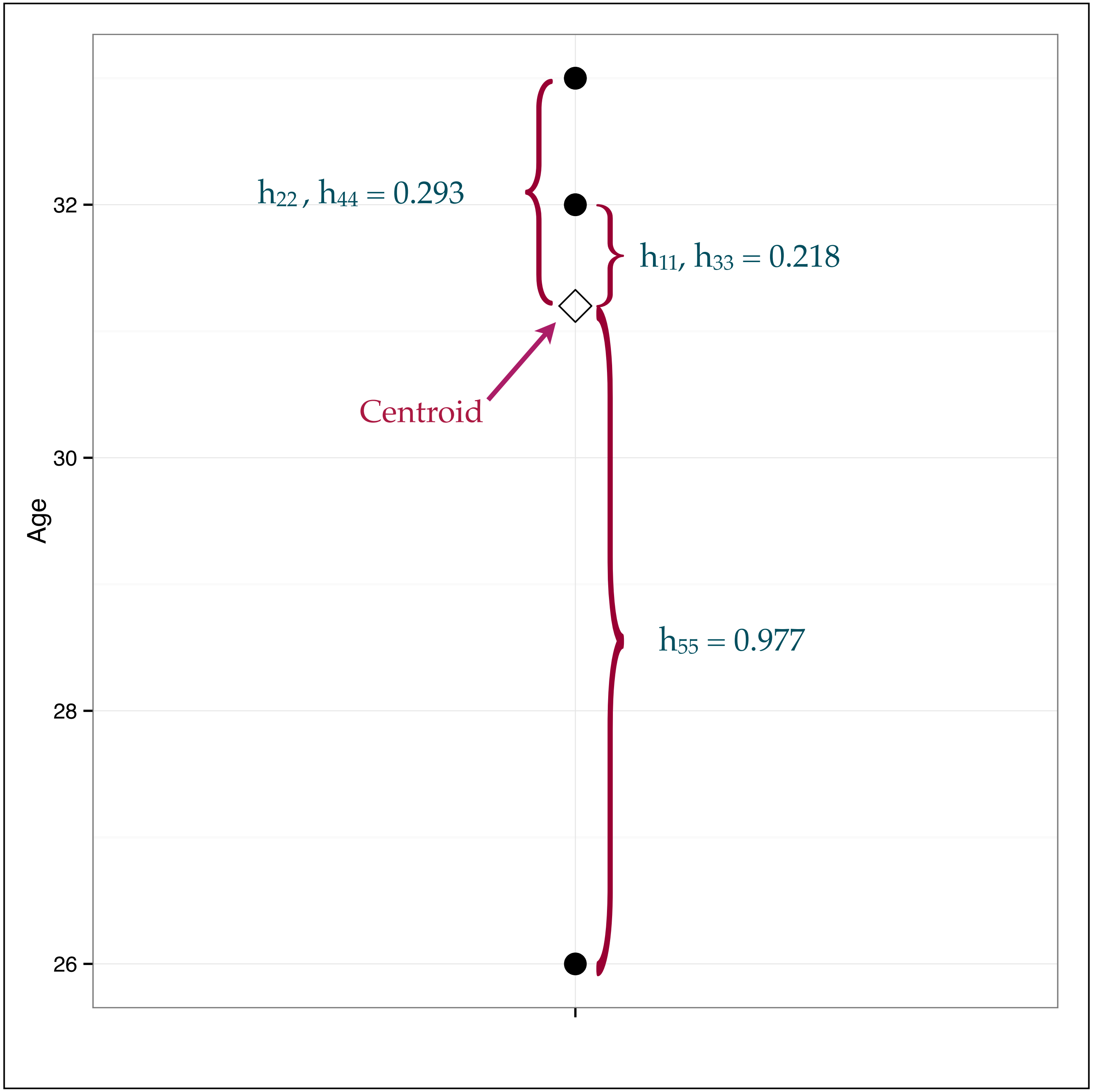
The h_{ii} elements measure the **multivariate distance** between the X-values for the i^{th} case and the means of the X-values for all n cases (i.e., the centroid)

The h_{ii} elements have the following properties

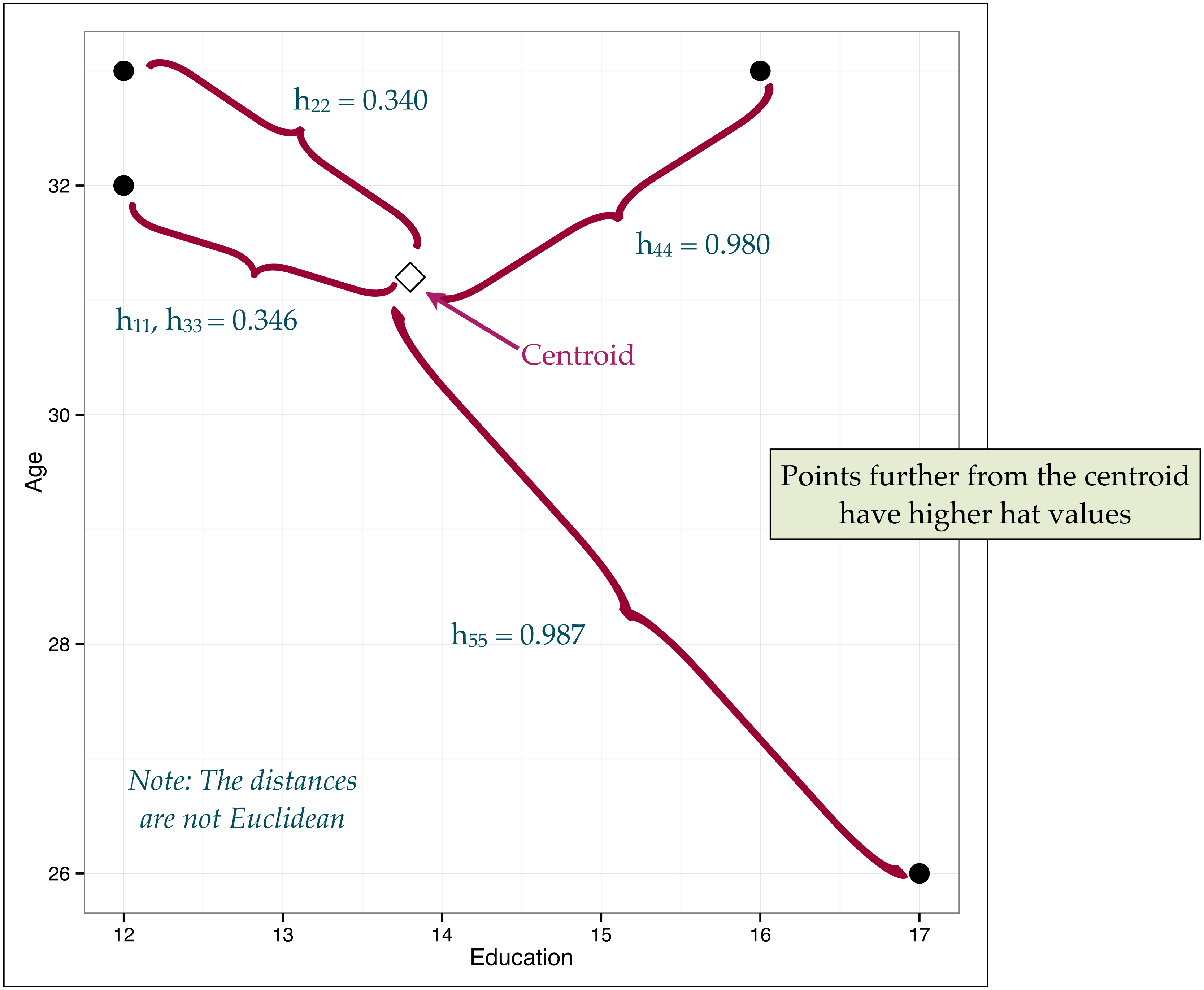
$$0 \leq h_{ii} \leq 1$$

$$\sum_{i=1}^{n} h_{ii} = p$$

The larger the value of h_{ii}, the smaller the variance of the residual e_i, hence the closer the fitted value will tend to be to Y_i



h_{22} , h_{44} = 0.293
h_{11}, h_{33} = 0.218
Centroid
h_{55} = 0.977
32
30
28
26
Age



$h_{22} = 0.340$
$h_{44} = 0.980$
$h_{11}, h_{33} = 0.346$
Centroid
$h_{55} = 0.987$
Points further from the centroid have higher hat values
Note: The distances are not Euclidean
Age
Education
32
30
28
26
12
13
14
15
16
17

Identifying High-Leverage Cases

The mean leverage value is

$$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n} = \frac{p}{n}$$

Leverage values over *twice as large* as the mean leverage are considered **outlying cases** with regard to their X-values

In the data we have, the **mean leverage value is 0.4**. Any observation with a *hat value* ≥ 0.8 would be considered an outlying case (high leverage) based on its X-values (in our case observations 4 and 5)

```
# Obtain the hat-values (leverage) values
> hatvalues(lm.1)

        1         2         3         4         5 
0.2183908 0.2931034 0.2183908 0.2931034 0.9770115 

# Mean hat value
> mhat = mean(hatvalues(lm.1))
> mhat
[1] 0.4

# Mean hat-value
> mean(hatvalues(lm.1))
[1] 0.4

# Is hat-value > mean hat-value
> hatvalues(lm.1) > mhat

    1     2     3     4     5 
FALSE FALSE FALSE FALSE  TRUE 
```

We can obtain regression results from just a few summary results, namely the correlation matrix, and standard deviations.

The **standardized regression coefficients** can be computed using

$$\hat{\boldsymbol{\beta}} = (\mathbf{r}_{\mathbf{XX}})^{-1}\mathbf{r}_{\mathbf{XY}}$$

To obtain the **unstandardized regression coefficients**

$$\mathbf{b} = \hat{\boldsymbol{\beta}}\frac{\hat{\sigma}_Y}{\hat{\sigma}_i}$$

We can obtain the multiple R^2 using

$$R^2 = \sum \hat{\boldsymbol{\beta}}\,\mathbf{r}_{\mathbf{XY}}$$

```
# Obtain the correlation matrix
> myCor = cor(myData[ , c("wage", "age", "educ")])
> myCor

           wage        age       educ
wage  1.0000000  0.4225006 -0.2314938
age   0.4225006  1.0000000 -0.6399446
educ -0.2314938 -0.6399446  1.0000000

# Matrix of r(XY)
> rXY = myCor[2:3, 1]

# Matrix of r(XX)
> rXX = myCor[2:3, 2:3]

# Compute regression coefficients
> b = solve(rXX) %*% rXY
> b

           [,1]
age  0.46464173
educ 0.06585118
```

These are the regression coefficients for the model

`z.wage ~ z.age + z.educ`

```
# Obtain R^2
> sum(rXY * b)
[1] 0.1810673

# Obtain the SDs for the outcome and predictors
> s = apply(myData[c("wage", "age", "educ")], MARGIN = 2, FUN = sd)

# Separate the SDs
> sY = s[1]
> sX = s[2:3]

# Compute unstandardized regression coefficients
> (sY / sX) * b

          [,1]
age  0.5491758
educ 0.0921978
```

These are the unstandardized regression coefficients from the model `wage ~ age + educ`

Multicollinearity

Several problems when predictors are highly correlated

- Adding/deleting a predictor changes the regression coefficients
- Additional SS associated with predictor variable varies depending on which variables are in model
- Estimated SEs become large
- Tests for the estimated coefficients may not be statistically significant even though the test for the model is

Indicators of multicollinearity include

- Large changes in estimated coefficient when a predictor is added or removed
- Nonsignificant results for important predictors (based on theory or prior experience)
- Estimated coefficients with an unexpected sign (based on theory or prior experience)
- Wide CIs for the regression coefficients that represent important predictors (based on theory or prior experience)

Variance Inflation Factor (VIF)

Variance inflation factors measure how much the variances of the estimated regression coefficients are inflated as compared to when the predictors are not linearly related (i.e., independent)

The precision of the OLS estimated regression coefficients as measured by their variances is

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

For purposes of measuring the impact of multicollinearity, it is **useful to work with the standardized regression model**. In this model, the **X'X** matrix is just the correlation matrix of the *X*-variables

The variances for the standardized coefficients are found by taking the k^{th} diagonal element of the inverse of the correlation matrix

```
# Fit a model with multiple predictors
> lm.2 = lm(wage ~ age + educ + status, data = myData)

# Get the design matrix
> X = model.matrix(lm.2)

# Inverse of the correlation matrix for the predictors
# Omit the column of ones
> solve(cor(X[ , -1]))

                 age      educ statusSingle
age         3.281682 1.6343002    1.8150271
educ        1.634300 1.8843962    0.6291698
statusSingle 1.815027 0.6291698    2.0743555

> vif(lm.2)

     age     educ   status
3.281682 1.884396 2.074355
```

The diagonal element $(VIF)_k$ is called the variance inflation factor for b'_k and is equal to

$$(\mathrm{VIF})_k = (1 - R_k^2)^{-1} \quad \text{where } k = 1, 2, \ldots, (p-1)$$

where R^2_k is the coefficient of multiple determination when x_k is regressed on the $p - 2$ other X variables in the model.

The VIF = 1 when $R^2_k = 0$
(x_k is not related to the other predictors)

The VIF > 1 when $R^2_k \neq 0$
which indicates an inflated variance for b'_k
(i.e., t-values too small; CIs too large)

The largest VIF is often used as a diagnostic indicator of the severity of multicollinearity. A maximum **VIF > 10** is usually considered problematic and is unduly influencing the OLS estimates

The mean VIF also provides information about multicollinearity, where the mean VIF is

$$\overline{\mathrm{VIF}} = \frac{\sum_{k=1}^{p-1} \mathrm{VIF}_k}{p-1}$$

Mean VIFs > 1 also indicate model problems due to multicollinearity