**University of California, Los Angeles**
**Department of Statistics**

**Statistics 100C** **Instructor: Nicolas Christou**

## Centering and scaling
## Multicollinearity

The usual multiple regression model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. If we partition $\mathbf{X} = (\mathbf{1}, \mathbf{X}_{(\mathbf{0})})$ and $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta}_{(\mathbf{0})} \end{pmatrix}$ we can write the model as $\mathbf{y} = \beta_0 \mathbf{1} + \mathbf{X}_{(\mathbf{0})}\boldsymbol{\beta}_{(\mathbf{0})} + \boldsymbol{\epsilon}$. Suppose now we add and subtract $\frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X}_{(\mathbf{0})}\boldsymbol{\beta}_{(\mathbf{0})}$. The model now is

$$\begin{aligned}
\mathbf{y} &= \mathbf{1}\left(\beta_0 + \frac{1}{n}\mathbf{1}'\mathbf{X}_{(\mathbf{0})}\boldsymbol{\beta}_{(\mathbf{0})}\right) + (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{X}_{(\mathbf{0})}\boldsymbol{\beta}_{(\mathbf{0})} + \boldsymbol{\epsilon} \\
\mathbf{y} &= \mathbf{1}\left(\beta_0 + \bar{\mathbf{x}}'\boldsymbol{\beta}_{(\mathbf{0})}\right) + (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{X}_{(\mathbf{0})}\boldsymbol{\beta}_{(\mathbf{0})} + \boldsymbol{\epsilon} \\
\mathbf{y} &= \gamma_0 \mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{(\mathbf{0})} + \boldsymbol{\epsilon}
\end{aligned}$$

This is called the centered model, where,

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}, \ \boldsymbol{\beta}_{(\mathbf{0})} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \\ \beta_k \end{pmatrix}, \ \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{n-1} \\ \epsilon_n \end{pmatrix}, \ \mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & z_{13} & \cdots & z_{1k} \\ z_{21} & z_{22} & z_{23} & \cdots & z_{2k} \\ z_{31} & z_{32} & z_{33} & \cdots & z_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & z_{n3} & \cdots & z_{nk} \end{pmatrix}.$$

Another way to express the model above is to look at the regression model equation for each $y_i$.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i.$$

After we add and subtract $\beta_j \bar{x}_j$ it becomes the "centered model":

$$\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} \pm \beta_1 \bar{x}_1 + \beta_2 x_{i2} \pm \beta_2 \bar{x}_2 + \ldots + \beta_k x_{ik} \pm \beta_k \bar{x}_k + \epsilon_i \\
y_i &= (\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \ldots + \beta_k \bar{x}_k) + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \ldots + \beta_k(x_{ik} - \bar{x}_k) + \epsilon_i \\
y_i &= \gamma_0 + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \ldots + \beta_k(x_{ik} - \bar{x}_k) + \epsilon_i
\end{aligned}$$

We can expressed $y_i$ as

$$\begin{aligned}
y_i &= \gamma_0 + \beta_1(x_{i1} - \bar{x}_1) + \beta_2(x_{i2} - \bar{x}_2) + \ldots + \beta_k(x_{ik} - \bar{x}_k) + \epsilon_i \qquad (1) \\
y_i &= \gamma_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_k z_{ik} + \epsilon_i
\end{aligned}$$

Estimation of the centered model:

Centering and scaling:
We can now scale the predictors as follows. We multiply and divide each centered predictor in equation (1) by the quantity $\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ for $j = 1 \ldots k$ to get:

$$\begin{aligned}
y_i &= \gamma_0 + \beta_1 \frac{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}}(x_{i1}-\bar{x}_1) + \ldots + \beta_k \frac{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}}(x_{ik}-\bar{x}_k) + \epsilon_i \text{ or} \\
y_i &= \gamma_0 + \delta_1 \frac{z_{i1}}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} + \ldots + \delta_k \frac{z_{ik}}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} + \epsilon_i \text{ or} \\
y_i &= \gamma_0 + \delta_1 Z_{si1} + \ldots + \delta_k Z_{sik} + \epsilon_i,
\end{aligned}$$

This is the centered and scaled model, where, $\delta_j = \beta_j \sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$ and $Z_{sij} = \frac{x_{ij}-\bar{x}_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$.

We can also obtain the centered and scaled model in matrix form as follows.

$$\begin{aligned}
\mathbf{y} &= \gamma_0 \mathbf{1} + \mathbf{Z}\boldsymbol{\beta}_{\mathbf{(0)}} + \boldsymbol{\epsilon} \\
\mathbf{y} &= \gamma_0 \mathbf{1} + \mathbf{Z}\mathbf{D}^{\mathbf{-1}}\mathbf{D}\boldsymbol{\beta}_{\mathbf{(0)}} + \boldsymbol{\epsilon} \\
\mathbf{y} &= \gamma_0 \mathbf{1} + \mathbf{Z_s}\boldsymbol{\delta}_{\mathbf{(0)}} + \boldsymbol{\epsilon}
\end{aligned}$$

where, $\mathbf{Zs} = \mathbf{ZD}^{-1}$ and $\boldsymbol{\delta}_{\mathbf{(0)}} = \mathbf{D}\boldsymbol{\beta}_{\mathbf{(0)}}$. The matrix $\mathbf{D}$ is defined as

$$\mathbf{D} = \begin{pmatrix}
\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}
\end{pmatrix}$$

Estimation of the centered and scaled model: If we regress $y$ on $Z_{s1}, Z_{s2}, \ldots, Z_{sk}$ we will obtain estimates for $\gamma_0$ and $\boldsymbol{\delta_{(0)}}$. Therefore,

$$\begin{pmatrix} \hat{\gamma}_0 \\ \hat{\boldsymbol{\delta}}_{\mathbf{(0)}} \end{pmatrix} = \left[ \begin{pmatrix} \mathbf{1}' \\ \mathbf{Zs}' \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{Zs} \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{1}' \\ \mathbf{Zs}' \end{pmatrix} \mathbf{y} = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{Zs} \\ \mathbf{Zs}'\mathbf{1} & \mathbf{Zs}'\mathbf{Zs} \end{pmatrix} \begin{pmatrix} \mathbf{1}' \\ \mathbf{Zs}' \end{pmatrix} \mathbf{y}$$

But, $\mathbf{1}'\mathbf{Zs} = \mathbf{0}$ and $\mathbf{Zs}'\mathbf{1} = \mathbf{0}$. Therefore,

$$\begin{pmatrix} \hat{\gamma}_0 \\ \hat{\boldsymbol{\delta}}_{\mathbf{(0)}} \end{pmatrix} = \begin{pmatrix} n & \mathbf{0} \\ \mathbf{0} & \mathbf{Zs}'\mathbf{Zs} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}' \\ \mathbf{Zs}' \end{pmatrix} \mathbf{y} = \begin{pmatrix} \frac{1}{n} & \mathbf{0} \\ \mathbf{0} & (\mathbf{Zs}'\mathbf{Zs})^{-1} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \mathbf{Zs}'\mathbf{y} \end{pmatrix}$$

It follows that, $\hat{\gamma}_0 = \bar{y}$ and $\hat{\boldsymbol{\delta}}_{\mathbf{(0)}} = (\mathbf{Zs}'\mathbf{Zs})^{-1}\mathbf{Zs}'\mathbf{y}$. But, $\mathbf{Zs}'\mathbf{Zs} = \mathbf{R}$ (correlation matrix of the $k$ predictors - see page 3). Finally, $\hat{\boldsymbol{\delta}}_{\mathbf{(0)}} = \mathbf{R}^{-1}\mathbf{Zs}'\mathbf{y}$ and $var(\hat{\boldsymbol{\delta}}_{\mathbf{(0)}}) = \sigma^2\mathbf{R}^{-1}$.

Summary:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{non-centered model} \\ \mathbf{y} &= \beta_0\mathbf{1} + \mathbf{X_{(0)}}\boldsymbol{\beta_{(0)}} + \boldsymbol{\epsilon} \\ \\ \mathbf{y} &= \gamma_0\mathbf{1} + \mathbf{Z}\boldsymbol{\beta_{(0)}} + \boldsymbol{\epsilon} \quad \text{centered model} \\ \\ \mathbf{y} &= \gamma_0\mathbf{1} + \mathbf{Z_s}\boldsymbol{\delta_{(0)}} + \boldsymbol{\epsilon} \quad \text{centered and scaled model} \end{aligned}$$

These three models have the same
fitted values
residuals
$SSR$
$SSE$
$R^2$
$F$ statistic for testing the overall significance of the model
$t$ statistics for testing individual $\beta_i$ coefficients.

Notes:
$\mathbf{1}'\mathbf{Z} = 0'$ and $\mathbf{Z}'\mathbf{1} = \mathbf{0}$
$\mathbf{1}'\mathbf{Z_s} = 0'$ and $\mathbf{Z_s}'\mathbf{1} = \mathbf{0}$

We can verify that $\mathbf{Zs}'\mathbf{Zs} = \mathbf{R}$ from the following:

$$\mathbf{Zs}'\mathbf{Zs} = \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \frac{x_{21}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \cdots & \cdots & \frac{x_{n1}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} \\ \frac{x_{12}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \frac{x_{22}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \cdots & \cdots & \frac{x_{n2}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} \\ \frac{x_{13}-\bar{x}_3}{\sqrt{\sum_{i=1}^{n}(x_{i3}-\bar{x}_3)^2}} & \frac{x_{23}-\bar{x}_3}{\sqrt{\sum_{i=1}^{n}(x_{i3}-\bar{x}_3)^2}} & \cdots & \cdots & \frac{x_{n3}-\bar{x}_3}{\sqrt{\sum_{i=1}^{n}(x_{i3}-\bar{x}_3)^2}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{x_{1k}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} & \frac{x_{2k}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} & \cdots & \cdots & \frac{x_{nk}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} \end{pmatrix} \begin{pmatrix} \frac{x_{11}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \frac{x_{12}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \cdots & \cdots & \frac{x_{1k}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} \\ \frac{x_{21}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \frac{x_{22}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \cdots & \cdots & \frac{x_{2k}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} \\ \frac{x_{31}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \frac{x_{32}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \cdots & \cdots & \frac{x_{3k}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{x_{n1}-\bar{x}_1}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}} & \frac{x_{n2}-\bar{x}_2}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} & \cdots & \cdots & \frac{x_{nk}-\bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}} \end{pmatrix}$$

Therefore,

$$\mathbf{Zs}'\mathbf{Zs} = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} & \cdots & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & r_{24} & \cdots & \cdots & r_{2k} \\ r_{31} & r_{32} & 1 & r_{34} & \cdots & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & r_{k4} & \cdots & \cdots & 1 \end{pmatrix} = \mathbf{R}.$$

**Multicollinearity - theory**

Using the centered and scaled model we showed that the variance covariance matrix of $\hat{\boldsymbol{\delta}}_{\mathbf{(0)}}$ is equal to $var(\hat{\boldsymbol{\delta}}_{\mathbf{(0)}}) = \sigma^2 \mathbf{R}^{-1}$. We want to find an expression for $var(\hat{\delta}_1)$. This is equal to $\sigma^2 \times$(position (1,1) of $\mathbf{R^{-1}}$). First we will partition $\mathbf{R}$ as follows:

$$\mathbf{R} = \left( \begin{array}{ccccccc} 1 & r_{12} & r_{13} & r_{14} & \dots & \dots & r_{1k} \\ r_{21} & 1 & r_{23} & r_{24} & \dots & \dots & r_{2k} \\ r_{31} & r_{32} & 1 & r_{34} & \dots & \dots & r_{3k} \\ \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \dots & \dots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & r_{k4} & \dots & \dots & 1 \end{array} \right) = \left( \begin{array}{cc} 1 & \mathbf{r}' \\ \mathbf{r} & \mathbf{R_{22}} \end{array} \right).$$

To find the inverse of the partitioned matrix we will use the following result from linear algebra:

$$\left( \begin{array}{cc} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{array} \right)^{-1} = \left( \begin{array}{cc} \mathbf{C_{11}}^{-1} & -\mathbf{C_{11}}^{-1}\mathbf{C_{12}} \\ -\mathbf{C_{21}}\mathbf{C_{11}}^{-1} & \mathbf{A_{22}}^{-1} + \mathbf{C_{21}}\mathbf{C_{11}}^{-1}\mathbf{C_{12}} \end{array} \right).$$

where,

$$\begin{aligned} \mathbf{C_{11}} &= \mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}}^{-1}\mathbf{A_{21}} \\ \mathbf{C_{12}} &= \mathbf{A_{12}}\mathbf{A_{22}}^{-1} \\ \mathbf{C_{21}} &= \mathbf{A_{22}}^{-1}\mathbf{A_{21}} \end{aligned}$$

Using this result we can find the inverse of the partitioned $\mathbf{R}$ matrix. In particular, we are interested in finding the element at position (1,1) of $\mathbf{R}^{-1}$. It will correspond to $\mathbf{C_{11}}^{-1}$ and it is equal to $(1-\mathbf{r}'\mathbf{R_{22}}^{-1}\mathbf{r})^{-1}$. Therefore, $var(\hat{\delta}_1) = \frac{\sigma^2}{1-\mathbf{r}'\mathbf{R_{22}}^{-1}\mathbf{r}}$. We will show now that $var(\hat{\delta}_1) = \frac{\sigma^2}{1-R_1^2}$, where $R_1^2$ is the $R^2$ of the regression of $x_1$ on $x_2, x_3, \dots, x_k$.

Find $R_1^2$ using the centered and scaled model:

$$Zs_{i1} = \alpha_0 + \alpha_2 Zs_{i2} + \alpha_3 Zs_{i3} + \dots + \alpha_k Zs_{ik} + \epsilon_i$$

As always, $R_1^2 = \frac{SSR}{SST}$. But here, $SST = \sum_{i=1}^n (Zs_{i1} - \bar{Z}s_1)^2 = \sum_{i=1}^n Zs_1^2 = \mathbf{Zs_1}'\mathbf{Zs_1} = 1$. This is true because

$$\mathbf{Zs_1} = \left( \begin{array}{c} \frac{x_{11}-\bar{x}_1}{\sqrt{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}} \\ \frac{x_{21}-\bar{x}_1}{\sqrt{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}} \\ \vdots \\ \vdots \\ \frac{x_{n1}-\bar{x}_1}{\sqrt{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}} \end{array} \right).$$

So far we showed that $R_1^2 = SSR$.

Now let's find $SSR$. We know that $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2$. In the model we are using here, the response variable is $Zs_1$, and because $\bar{Z}s_1 = 0$ it follows that $SSR = \hat{\mathbf{Zs_1}}'\hat{\mathbf{Zs_1}} = \mathbf{Zs_1}'\mathbf{H}\mathbf{Zs_1}$, where $\mathbf{H}$ is the hat matrix constructed using the centered and scaled variables $Zs_2, Zs_3, \dots, Zs_k$. Therefore, $SSR = \mathbf{Zs1}'\mathbf{Zs}^*(\mathbf{Zs}^{*\prime}\mathbf{Zs}^*)^{-1}\mathbf{Zs}^{*\prime}\mathbf{Zs_1} = \mathbf{r}'\mathbf{R_{22}}^{-1}\mathbf{r}$, where $\mathbf{Zs}^*$ is $\mathbf{Zs}$ without $\mathbf{Zs_1}$. So $R_1^2 = \mathbf{r}'\mathbf{R_{22}}^{-1}\mathbf{r}$.

We just showed that $var(\hat{\delta}_1) = \frac{\sigma^2}{1-\mathbf{r}'\mathbf{R_{22}}^{-1}\mathbf{r}} = \frac{\sigma^2}{1-R_1^2}$. Since $\delta_j = \beta_j\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$, it follows that $var(\hat{\beta}_1) = \frac{\sigma^2}{(1-R_1^2)\sum_{i=1}^n (x_{i1}-x_1)^2}$. We see that if the $R^2$ of the regression of predictor $j$ on the other $k-1$ predictors is large (close to 1) the variance of the predictor of $\hat{\beta}_j$ will be inflated, and therefore the corresponding $t$ statistic will be small.

**Variance inflation factor (VIF)**

The variance inflation factor is given by $VIF_j = \frac{1}{1-R_j^2}$, and because $var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_1^2)\sum_{i=1}^n (x_{ij}-x_j)^2}$ it can be expressed as $VIF_j = \frac{\sum_{i=1}^n (x_{ij}-\bar{x}_j)^2}{\sigma^2} var(\hat{\beta}_j) = \frac{\sum_{i=1}^n (x_{ij}-\bar{x}_j)^2}{\sigma^2}\sigma^2 V_{jj} = (n-1)S_{xj}^2 V_{jj}$, where, $V_{jj}$ is the $(j,j)$th element of $(\mathbf{X}'\mathbf{X})^{-1}$.

**Geometric interpretation**

Since $R_1^2 = SSR = SST - SSE = 1 - SSE$ it follows that $SSE = 1 - R_1^2$. But $SSE = \mathbf{e}'\mathbf{e}$ which represent the squared length of the residual vector of the model $Zs_{i1} = \alpha_0 + \alpha_2 Zs_{i2} + \alpha_3 Zs_{i3} + \ldots + \alpha_k Zs_{ik} + \epsilon_i$. Therefore, if $R_1^2 \approx 1$ it follows that the squared length of the residual vector is close to zero, which means the fitted vector (in this case $\hat{\mathbf{Zs_1}}$) will be "close" to the subspace spanned by the columns of $\mathbf{Zs_2}, \mathbf{Zs_3}, \ldots, \mathbf{Zs}_k$.