Stat 206: Linear Models

Lecture 12

Nov. 6, 2019

Body Fat: Compare Models

Variables in Model	$\hat{\beta}_1$	$\hat{\beta}_2$	$s\{\hat{\beta}_1\}$	$s\{\hat{\beta}_2\}$	MSE
Model 1: X_1	0.8572	-	0.1288	-	7.95
Model 2: X_2	-	0.8565	-	0.1100	6.3
Model 3: X_1, X_2	0.2224	0.6594	0.3034	0.2912	6.47
Model 4: X_1, X_2, X_3	4.334	-2.857	3.016	2.582	6.15

- The regression coefficient for X_1 (X_2) depending on which other X variables are included in the model.
- The standard errors of the fitted regression coefficients are becoming when more X variables are included into the model.
- MSE tends to as additional X variables are added into the model.

R outputs

Body Fat: Compare Models

Variables in Model	$\hat{\beta}_1$	$\hat{\beta}_2$	$s\{\hat{\beta}_1\}$	$s\{\hat{\beta}_2\}$	MSE
Model 1: X_1	0.8572	-	0.1288	-	7.95
Model 2: X_2	-	0.8565	-	0.1100	6.3
Model 3: X_1, X_2	0.2224	0.6594	0.3034	0.2912	6.47
Model 4: X_1, X_2, X_3	4.334	-2.857	3.016	2.582	6.15

- The regression coefficient for X_1 (X_2) varies drastically depending on which other X variables are included in the model.
- The standard errors of the fitted regression coefficients are becoming inflated when more X variables are included into the model.
- MSE tends to decrease as additional X variables are added into the model.

R outputs

- $SSR(X_1) = 352.27$, $SSR(X_1|X_2) = 3.47$.
- The reason why $SSR(X_1|X_2)$ is so small compared to $SSR(X_1)$ is that X_1 and X_2 are with each other and with the response variable Y.
  - When X_2 is already in the model, the marginal contribution from X_1 in explaining Y is since X_2 contains much of the information as X_1 in terms of explaining Y.

What would happen if X_1 and X_2 were not correlated with Y, but were highly correlated among themselves?

- $SSR(X_1) = 352.27$, $SSR(X_1|X_2) = 3.47$.
- The reason why $SSR(X_1|X_2)$ is so small compared to $SSR(X_1)$ is that X_1 and X_2 are highly correlated with each other **and with the response variable Y.**
    - When X_2 is already in the model, the marginal contribution from X_1 in explaining Y is small since X_2 contains much of the same information as X_1 in terms of explaining Y.

What would happen if X_1 and X_2 were not correlated with Y, but were highly correlated among themselves?

In Model 4, none of the three X variables is statistically significant by the T-tests. However, the F-test for regression relation is highly significant. Is there a paradox?

- From the general linear test perspective, each T-test is a test, testing whether the of an X variable is significant given X variables being included in the model.
- The three tests of the marginal effects of X_1, X_2, X_3 together are to testing whether there is a regression relation between Y and (X_1, X_2, X_3).
- The reduced model for each individual test contains X variables and thus may lead to non-significant results due to .
- On the other hand, the reduced model for testing regression relation contains X variable.

Model 4

In Model 4, none of the three X variables is statistically significant by the T-tests. However, the F-test for regression relation is highly significant. Is there a paradox?

- From the general linear test perspective, each T-test is a marginal test, testing whether the marginal effect of an X variable is significant given **all other** X variables being included in the model.
- The three tests of the marginal effects of X_1, X_2, X_3 together are not equivalent to testing whether there is a regression relation between Y and (X_1, X_2, X_3).
- The reduced model for each individual test contains **all other** X variables and thus may lead to non-significant results due to multicollinearity.
- On the other hand, the reduced model for testing regression relation contains no X variable.

Model 4

Effects of Multicollinearity: Summary

- With multicollinearity, the estimated regression coefficients tend to have sampling variability (i.e., standard errors). This leads to:
  - confidence intervals.
  - It's possible that of the regression coefficients is statistically significant, but at the same time there is a regression relation between the response variable and the entire set of X variables.
- Multicollinearity does not prevent us from getting a of the data.

prediction

Effects of Multicollinearity: Summary

- With multicollinearity, the estimated regression coefficients tend to have large sampling variability (i.e., large standard errors). This leads to:
  - Wide confidence intervals.
  - It's possible that none of the regression coefficients is statistically significant, but at the same time there is a significant regression relation between the response variable and the entire set of X variables.
- Multicollinearity does not prevent us from getting a good fit of the data.

prediction

Interpretation of Regression Coefficients and ESS

In the presence of multicollinearity:

- The regression coefficient of an X variable which other X variables are also in the model.
- Therefore, a regression coefficient reflect any inherent effect of the corresponding X variable on the response variable, but only a given whatever other X variables are also in the model.
- Similarly, there is sum of squares that can be ascribed to any one X variable.
  - The reduction in the total variation in Y ascribed to an X variable must be interpreted as a given other X variables also included in the model.

Interpretation of Regression Coefficients and ESS

In the presence of multicollinearity:

- The regression coefficient of an X variable depends on which other X variables are also in the model.
- Therefore, a regression coefficient does **not** reflect any inherent effect of the corresponding X variable on the response variable, but only a marginal effect given whatever other X variables are also in the model.
- Similarly, there is **no** unique sum of squares that can be ascribed to any one X variable.
  - The reduction in the total variation in Y ascribed to an X variable must be interpreted as a margin reduction given other X variables also included in the model.

Quantify Multicollinearity: Variance Inflation Factor

Under the standardized model:

$$\sigma^2(\hat{\beta}^*) = \qquad\qquad\qquad\qquad .$$

- The kth diagonal element of the inverse correlation matrix $\mathbf{r}_{XX}^{-1}$ is called the **variance inflation factor (VIF)** for $\hat{\beta}_k^*$, denoted by VIF_k.
- The variance of the estimated regression coefficient $\hat{\beta}_k^*$:

$$\sigma^2(\hat{\beta}_k^*) = \qquad\qquad\qquad\qquad .$$

- The variance of the estimated regression coefficient $\hat{\beta}_k$ in the original model:

$$\sigma^2(\hat{\beta}_k) = \qquad\qquad\qquad\qquad .$$

Quantify Multicollinearity: Variance Inflation Factor

Under the standardized model:

$$\sigma^2(\hat{\boldsymbol{\beta}}^*) = \sigma^2 \begin{bmatrix} \frac{1}{n} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{r}_{XX}^{-1} \end{bmatrix}$$

- The kth diagonal element of the inverse correlation matrix $\mathbf{r}_{XX}^{-1}$ is called the **variance inflation factor (VIF)** for $\hat{\beta}_k^*$, denoted by VIF_k.
- The variance of the estimated regression coefficient $\hat{\beta}_k^*$:

$$\sigma^2(\hat{\beta}_k^*) = VIF_k \sigma^2, \quad k = 1, \cdots, p-1.$$

- The variance of the estimated regression coefficient $\hat{\beta}_k$ in the original model:

$$\sigma^2(\hat{\beta}_k) = VIF_k \times \frac{\sigma^2}{\sum_{i=1}^n (X_{ik} - \bar{X}_k)^2}, \quad k = 1, \cdots, p-1.$$

It can be shown that

$$VIF_k = \frac{1}{1 - R_k^2} (\geq 1), \qquad k = 1, \cdots, p-1,$$

where R_k^2 is the coefficient of multiple determination when X_k is regressed on the rest of X variables $\{X_j : 1 \leq j \neq k \leq p-1\}$.

- If X_k is uncorrelated with the rest of the X variables, then $R_k^2 =$ and $VIF_k =$.
- If $R_k^2 > 0$, then VIF_k , indicating an variance for $\hat{\beta}_k^*$ (eqv. $\hat{\beta}_k$) due to the between X_k and the other X variables.
- If X_k has a perfect linear association with the rest of the X variables, then $R_k^2 =$, $VIF_k =$ and so the variance of $\hat{\beta}_k^*$ (eqv. $\hat{\beta}_k$) is .
- In practice, $\max_k VIF_k > 10$ is often taken as an indication that multicollinearity is high.

It can be shown that

$$VIF_k = \frac{1}{1 - R_k^2} (\geq 1), \qquad k = 1, \cdots, p-1,$$

where R_k^2 is the coefficient of multiple determination when X_k is regressed on the rest of X variables $\{X_j : 1 \leq j \neq k \leq p-1\}$.

- If X_k is uncorrelated with the rest of the X variables, then $R_k^2 = 0$ and $VIF_k = 1$ (no inflation).
- If $R_k^2 > 0$, then $VIF_k > 1$, indicating an inflated variance for $\hat{\beta}_k^*$ (eqv. $\hat{\beta}_k$) due to the intercorrelation between X_k and the other X variables.
- If X_k has a perfect linear association with the rest of the X variables, then $R_k^2 = 1$, $VIF_k = \infty$ and so the variance of $\hat{\beta}_k^*$ (eqv. $\hat{\beta}_k$) is infinity (ill-defined).
- In practice, $\max_k VIF_k > 10$ is often taken as an indication that multicollinearity is high.

Body Fat

Correlation matrices.

$$\mathbf{r}_{XX} = \begin{bmatrix} 1.00 & 0.92 & 0.46 \\ 0.92 & 1.00 & 0.08 \\ 0.46 & 0.08 & 1.00 \end{bmatrix}, \quad \mathbf{r}_{XY} = \begin{bmatrix} 0.84 \\ 0.88 \\ 0.14 \end{bmatrix}.$$

X_1 and X_2 are highly correlated, X_1 and X_3 are moderately correlated, X_2 and X_3 are not much correlated. Moreover,

$$\mathbf{r}_{XX}^{-1} = \begin{bmatrix} 708.84 & -631.92 & -270.99 \\ -631.92 & 564.34 & 241.49 \\ -270.99 & 241.49 & 104.61 \end{bmatrix}$$

So,

Each predictor is with the rest of the predictors.

Body Fat

Correlation matrices.

$$\mathbf{r}_{XX} = \begin{bmatrix} 1.00 & 0.92 & 0.46 \\ 0.92 & 1.00 & 0.08 \\ 0.46 & 0.08 & 1.00 \end{bmatrix}, \quad \mathbf{r}_{XY} = \begin{bmatrix} 0.84 \\ 0.88 \\ 0.14 \end{bmatrix}.$$

X_1 and X_2 are highly correlated, X_1 and X_3 are moderately correlated, X_2 and X_3 are not much correlated. Moreover,

$$\mathbf{r}_{XX}^{-1} = \begin{bmatrix} 708.84 & -631.92 & -270.99 \\ -631.92 & 564.34 & 241.49 \\ -270.99 & 241.49 & 104.61 \end{bmatrix}$$

So,

$$R_1^2 = 0.9986, \quad R_2^2 = 0.9982, \quad R_3^2 = 0.9904.$$

Each predictor is highly intercorrelated with the rest of the predictors.

Coefficient of Partial Determination

It measures the marginal contribution in proportional reduction in SSE by adding one X variable into a model.

- Definition.

$$
\begin{aligned}
& R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1} \\
:= \quad & \frac{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1}) - SSE(X_1,\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})} \\
= \quad & \frac{SSR(X_j|X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}
\end{aligned}
$$

- Coefficients of partial determination are in between
  .
- For example, $R^2_{Y,1|2} =$ is

Coefficient of Partial Determination

It measures the marginal contribution in proportional reduction in SSE by adding one X variable into a model.

- Definition.

$$
\begin{aligned}
& R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1} \\
:= \quad & \frac{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1}) - SSE(X_1,\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})} \\
= \quad & \frac{SSR(X_j|X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}
\end{aligned}
$$

- Coefficients of partial determination are in between 0 and 1.
- For example, $R^2_{Y,1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$ is the proportional reduction in SSE by including X_1 into the model with X_2.

Body Fat

- From R outputs, we can obtain a number of coefficients of partial determination. E.g.:

$$R^2_{Y,2|1} =$$

$$R^2_{Y,1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{113.42 - 109.95}{113.43} = 3.1\%.$$

$$R^2_{Y,3|12} =$$

- When X_2 is added to the model containing X_1, SSE is reduced by ; When X_1 is added to the model containing X_2, SSE is reduced by ; When X_3 is added to the model containing X_1, X_2, SSE is reduced by .

R outputs

Body Fat

- From R outputs, we can obtain a number of coefficients of partial determination. E.g.:

$$R^2_{Y,2|1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)} = \frac{143.12 - 109.95}{143.12} = 23.2\%.$$

$$R^2_{Y,1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{113.42 - 109.95}{113.43} = 3.1\%.$$

$$R^2_{Y,3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.55}{109.95} = 10.5\%.$$

- When X_2 is added to the model containing X_1, SSE is reduced by 23.2%; When X_1 is added to the model containing X_2, SSE is reduced by 3.1%; When X_3 is added to the model containing X_1, X_2, SSE is reduced by 10.5%.

◂ R outputs

Interpretation of Coefficient of Partial Determination

- $SSR(X_j|X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{p-1})$ is the SSR when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \leq l \neq j \leq p\}$.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ is the

  between the two sets of residuals obtained by regressing Y and X_j to the rest of variables $X_{-(j)}$, respectively.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ measures the linear association between Y and X_j after

  have been adjusted for.

Interpretation of Coefficient of Partial Determination

- $SSR(X_j|X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{p-1})$ is the SSR when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \leq l \neq j \leq p\}$. (Discussed in the Lab session)
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ is the coefficient of simple determination (i.e., the squared correlation coefficient) between the two sets of residuals obtained by regressing Y and X_j to the rest of variables $X_{-(j)}$, respectively.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ measures the linear association between Y and X_j after the linear effects of $X_{-(j)}$ have been adjusted for.

Example. $R^2_{Y,1|2}$.

- Regress Y on X_2: $e_i(Y|X_2) = Y_i - \widehat{Y}_i(X_2), \quad i = 1, \cdots n$.
- Regress X_1 on X_2: $e_i(X_1|X_2) = X_{i1} - \widehat{X}_{i1}(X_2), \quad i = 1, \cdots n$.
- $R^2_{Y1|2}$ equals to the coefficient of simple determination between $e_i(Y|X_2)$ and $e_i(X_1|X_2)$.
- It measures the linear association between Y and X_1 after the linear effects of X_2 have been adjusted for.

Partial Correlations

The **signed** square-root of a coefficient of partial determination is called a partial correlation.

- The sign is the same as the sign of the corresponding fitted regression coefficient.
- Partial correlation is the
  between the .
- Partial correlations can be used to find the "best" X variable to be added next for inclusion in the regression model.

Partial Correlations

The **signed** square-root of a coefficient of partial determination is called a partial correlation.

- The sign is the same as the sign of the corresponding fitted regression coefficient.
- Partial correlation is the correlation coefficient between the two respective sets of residuals.
- Partial correlations can be used to find the “best” X variable to be added next for inclusion in the regression model.

Body Fat

- $r_{Y2|1} =$.
- $r_{Y1|2} =$.
- $r_{Y3|12} =$.

Body Fat

- $r_{Y2|1} = \sqrt{0.232} = 0.482$, since in Model 3, $\hat{\beta}_2 > 0$.
- $r_{Y1|2} = \sqrt{0.031} = 0.176$, since in Model 3, $\hat{\beta}_1 > 0$.
- $r_{Y3|12} = -\sqrt{0.105} = -0.324$, since in Model 4, $\hat{\beta}_3 < 0$.

LS Fitted Regression Coefficients as Partial Coefficients

The LS fitted regression coefficients $\hat{\beta}$ are indeed partial coefficients.

- Consider $p - 1$ X variables in the model. Let $\hat{\beta}_j$ be the LS fitted regression coefficient for X_j.
- Then $\hat{\beta}_j$ equals to the LS fitted regression coefficient when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \leq l \neq j \leq p\}$.

Confirm this numerically with some of homework data sets.