Multiple Linear Regression: Hypothesis Testing

Key terms

- Extra sums-of-squares

Key ideas/results

1. $R(\mathcal{B}|\mathcal{A})$ represents the regression sums-of-squares accounted for by variables $\mathcal{B}$ after variables $\mathcal{A}$ have already been fitted.

   It decomposes as follows:

$$\begin{aligned} SS_{\text{Reg}} = R(\beta_1, \beta_2, \ldots, \beta_k|\beta_0) &= R(\beta_1|\beta_0) + R(\beta_2|\beta_1, \beta_0) + \\ &\quad R(\beta_3|\beta_2, \beta_1, \beta_0) + \cdots + \\ &\quad R(\beta_k|\beta_{k-1}, \beta_{k-2}, \ldots, \beta_1, \beta_0) \end{aligned}$$

2. BLUE/Gauss-Markov

   For the general linear model, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, if $\mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\mathsf{Var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$ then the least squares estimators $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ are Best Linear Unbiased Estimators (BLUE). That is, among the class of linear unbiased estimators, $\hat{\boldsymbol{\beta}}$ has the smallest variance. (This result is known as the Gauss-Markov theorem).

3. UMVUE For the general linear model, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, if $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ then the least squares estimators $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ are the Uniform Minimum Variance Unbiased Estimators (UMVUE). That is, among the class of all unbiased estimators (not just linear ones), $\hat{\boldsymbol{\beta}}$ has the smallest variance.

Added Variable Plot Redux

As mentioned last time, added variable plots are the only way to precisely visualize how parameters are estimated in multiple linear regression. The notation yesterday (and in Weisberg) is confusing, so here is another attempt at motivating added variable plots.

We have a 2-predictor multiple linear regression model; we are trying to predict FUEL consumption with TAX rate and percentage of licensed drivers (DLIC). In vector notation we write this

$$\text{FUEL} = \beta_0 \mathbf{1} + \beta_1 \text{TAX} + \beta_2 \text{DLIC} + \boldsymbol{\epsilon}.$$

Now, *orthogonalize* with respect to DLIC. That is, for each component of the expression above, take the residuals from fitting a regression on DLIC:

$$\text{FUEL}_{\perp\text{DLIC}} = \beta_0 \mathbf{1}_{\perp\text{DLIC}} + \beta_1 \text{TAX}_{\perp\text{DLIC}} + \beta_2 \text{DLIC}_{\perp\text{DLIC}} + \boldsymbol{\epsilon}_{\perp\text{DLIC}}.$$

Where

$$
\begin{aligned}
(\text{FUEL}_{\perp\text{DLIC}})_i &= \text{FUEL}_i - \hat{\alpha}_0 - \hat{\alpha}_1 \text{DLIC}_i \\
&\quad \text{The residuals from fitting } \mathsf{E}(\text{FUEL}_i) = \alpha_0 + \alpha_1 \text{DLIC}_i \\
(\mathbf{1}_{\perp\text{DLIC}})_i &= \mathbf{1}_i - \hat{\gamma}_0 - \hat{\gamma}_1 \text{DLIC}_i \\
&\quad \text{The residuals from fitting } \mathsf{E}(\mathbf{1}_i) = \gamma_0 + \gamma_1 \text{DLIC}_i \\
&\quad \text{but...}\underline{\hspace{6cm}} \\
(\text{TAX}_{\perp\text{DLIC}})_i &= \text{TAX}_i - \hat{\delta}_0 - \hat{\delta}_1 \text{DLIC}_i \\
&\quad \text{The residuals from fitting } \mathsf{E}(\text{TAX}_i) = \delta_0 + \delta_1 \text{DLIC}_i \\
(\text{DLIC}_{\perp\text{DLIC}})_i &= \text{DLIC}_i - \hat{\eta}_0 - \hat{\eta}_1 \text{DLIC}_i \\
&\quad \text{The residuals from fitting } \mathsf{E}(\text{DLIC}_i) = \eta_0 + \eta_1 \text{DLIC}_i \\
&\quad \text{but...}\underline{\hspace{6cm}} \\
(\boldsymbol{\epsilon}_{\perp\text{DLIC}})_i &= \epsilon_i - \hat{\theta}_0 - \hat{\theta}_1 \text{DLIC}_i \\
&\quad \text{The residuals from fitting } \mathsf{E}(\epsilon_i) = \theta_0 + \theta_1 \text{DLIC}_i
\end{aligned}
$$

Hence we're left with

$$\text{FUEL}_{\perp\text{DLIC}} = \beta_1 \text{TAX}_{\perp\text{DLIC}} + \boldsymbol{\epsilon}_{\perp\text{DLIC}}$$

and so we can estimate β_1 just by fitting a simple linear regression of $\text{FUEL}_{\perp\text{DLIC}}$ on $\text{TAX}_{\perp\text{DLIC}}$, which is exactly what an added variable plot is!

$R(\cdot|\cdot)$ notation

The expression for partitioning of variability holds in multiple linear regression, just as it did in in simple linear regression: $SS_{\text{Tot}} = SS_{\text{Reg}} + SS_{\text{Err}}$. However, since we have many predictor variables running around it will be useful to consider how each predictor contributes to SS_{Reg}. The $R(\cdot|\cdot)$ does this. $R(\mathcal{B}|\mathcal{A})$ represents the regression sums-of-squares accounted for by variables $\mathcal{B}$ after variables $\mathcal{A}$ have already been fitted.

The contribution from each predictor can be decomposed as follows:

$$
\begin{aligned}
SS_{\text{Reg}} = R(\beta_1, \beta_2, \ldots, \beta_k | \beta_0) &= R(\beta_1|\beta_0) + R(\beta_2|\beta_1, \beta_0) + \\
&\quad R(\beta_3|\beta_2, \beta_1, \beta_0) + \cdots + \\
&\quad R(\beta_k|\beta_{k-1}, \beta_{k-2}, \ldots, \beta_1, \beta_0)
\end{aligned}
$$

This can be interpreted as "SS_{Reg} is sum of: The regression sum-of-squares from first fitting X_1, plus the regression sum-of-squares from adding X_2 to the model with just X_1, plus the regression sum-of-squares from adding X_3 to the model with X_1 and X_2...