Multiple Linear Regression

Simple linear regression: One predictor — X

Multiple linear regression: Several predictors — $X_1, \ldots, X_k$

Linear (regression) model:
$E(Y|X_1 = x_1, \ldots, X_k = x_k) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$ — models **mean response** as a function of predictors

Examples: linear model?

- $E(Y|x) = \beta_0 + \beta_1 x$ — yes
- $E(Y|\mathbf{x}) = \beta_0 + \beta_1 x + \beta_2 x^2$ - yes. $(x_1 = x,\ x_2 = x^2)$ — Note: A linear model is capturing a non-linear relationship b/w y and x.
- $E(Y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 * x_2)$ — yes.
- $E(\log(Y)|x) = \beta_0 + \beta_1 \log(x)$ — yes.
- $E(Y|x) = \beta_0 + (\beta_1 x)^{-1} = \beta_0 + \frac{1}{\beta_1} \cdot \frac{1}{x}$ — NO, but can be made a linear model by taking $\gamma_1 = \frac{1}{\beta_1}$
- $E[Y|X] = \frac{1}{\beta_0 + \beta_1 x}$ — No

Note: "Linear" refers to **linear in regression coefficients**

Linear model: $E(Y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

Y - response, $X_1, \ldots, X_k$ — predictors

Interpretation of $k+1$ regression coefficients:

- $\beta_0 = E(Y|\mathbf{x} = \mathbf{0})$ — intercept
- $\beta_j = E(Y|x_1, \ldots, \underline{x_j + 1}, \ldots, x_k) - E(Y|x_1, \ldots, \underline{x_j}, \ldots, x_k)$ — slope of x_j, i.e., change in mean response when jth predictor increases by 1, while keeping other predictors fixed, $j = 1, \ldots, k$. $\longrightarrow$ β_j represents the effect of the jth predictor after adjusting for the effects of the other predictors.

Data: n independent subjects, ith subject gives $(Y_i, X_{1i}, X_{2i}, \ldots, X_{ki})$, $i = 1, \ldots, n$.

Linear model for data: For $i = 1, \ldots, n$,

$$E(\underline{Y_i}|x_{i1}, \ldots, x_{ik}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

Subject #	Y	X_1	$\ldots$	X_k
1	Y_1	X_{11}	$\ldots$	X_{1k}
2	Y_2	X_{21}	$\ldots$	X_{2k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
i	Y_i	X_{i1}	$\ldots$	X_{ik}
$\vdots$	$\vdots$			
n	Y_n	X_{n1}	$\ldots$	X_{nk}

— $E[Y_i|x_i]$

Alternative form: $Y_i = \boxed{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}} + \epsilon_i$

Assumptions:

- $E(\epsilon_i) = 0$, $\text{var}(\epsilon_i) = \sigma^2$, and ϵ_i are independent. — as before.
- $k + 1 < n$ — i.e., have more observations than the number of regression coefficients
- The predictors are considered fixed and are measured without error

These imply:

- $E(Y_i|x_{i1}, \ldots, x_{ik}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$
- $\text{var}(Y_i) = \sigma^2$
- $Y_1, \ldots, Y_n$ are independent.

— as before.

Linear Model in Matrix Notation

Define:

response vector; intercept; design matrix or regression matrix.; coeff. vector

$$\mathbf{Y} = \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{n\times 1}, \mathbf{X} = \begin{array}{cccc} \uparrow & x_1 & \cdots & x_k \\ \end{array} \underbrace{\begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}}_{n\times(k+1)}, \boldsymbol{\beta} = \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{(k+1)\times 1}, \boldsymbol{\epsilon} = \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{n\times 1}$$

error vector

- $Y_i = (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) + \epsilon_i = (\text{i-th row of } X) * \boldsymbol{\beta} + \epsilon_i$
- $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $E(\mathbf{Y}|\mathbf{X}) = X\beta + E[\epsilon] = X\beta$, $\quad E[\epsilon] = \begin{bmatrix} E[\epsilon_1] \\ \vdots \\ E[\epsilon_n] \end{bmatrix}$.
- rank of X is full, i.e., $(\mathbf{X}'\mathbf{X})^{-1}$ exists.
- $\hat{\boldsymbol{\beta}}$ = estimator of $\boldsymbol{\beta}$
- $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ = fitted (or predicted) response

$\widehat{E[Y|X]}$

$$X_0 = \underbrace{\begin{bmatrix} x_{01} \\ \vdots \\ x_{0k} \end{bmatrix}}_{k\times 1}$$

Predicted response when $\mathbf{x} = \mathbf{x}_0$: $\hat{Y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$

Estimated mean response from the model at the given predictor values.

Least Squares Estimation of β

$\beta_0, \beta_1, \ldots, \beta_k$

As before: Minimize $\sum_{i=1}^{n} \epsilon_i^2$ with respect to $\beta_0, \beta_1, \beta_k$ to get $\hat{\beta}$

- Least squares estimator: $\boxed{\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}}$ — solution of the system of equations $\frac{\partial \sum_{i=1}^{n} \epsilon_i^2}{\partial \beta_0} = 0$, $\frac{\partial \sum_{i=1}^{n} \epsilon_i^2}{\partial \beta_j} = 0$, $j = 1, 2, \ldots, k$
- Minimum value of $\sum_{i=1}^{n} \epsilon_i^2$ is $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}) = SS_{\text{ERR}}$ — **error (or residual) sum of squares**

Properties of $\hat{\beta}$:

- Linear in $\mathbf{Y}$
- Unbiased, i.e., $E[\hat{\beta}] = E[(X'X)^{-1}X'Y] = (X'X)^{-1}X'\underbrace{E(Y)} = \underbrace{(X'X)^{-1}}\underbrace{X'(X\beta)} = \beta$ for all β.
- $\text{var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ $(k+1)\times(k+1)$
- $\text{var}(\hat{\beta}_0) = \sigma^2 \times$ first diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$
- $\text{var}(\hat{\beta}_j) = \sigma^2 \times (j+1)$th diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$
- $\hat{\sigma}^2 = SS_{\text{ERR}}/(n-k-1) = MS_{\text{ERR}}$ is unbiased for σ^2.

unknown parameter = $(k+1)$ reg coeff $+ \underset{\uparrow \sigma^2}{1} = (k+2)$.

ANOVA table

As before:

- $SS_{\text{TOT}} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = (\mathbf{Y} - \overline{\mathbf{Y}})'(\mathbf{Y} - \overline{\mathbf{Y}})$, where

$$\overline{\mathbf{Y}} = \overline{Y}\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}$$

- $SS_{\text{REG}} = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = (\hat{\mathbf{Y}} - \overline{\mathbf{Y}})'(\hat{\mathbf{Y}} - \overline{\mathbf{Y}})$
- $SS_{\text{ERR}} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}})$

ANOVA identity: $SS_{REG} + SS_{ERR} = SS_{TOT}$.

\# slopes.

Source	SS	d.f.	MS	F
Model	SS_{REG}	k	$MS_{\text{REG}} = \frac{SS_{\text{REG}}}{k}$	$\frac{MS_{\text{REG}}}{MS_{\text{ERR}}}$
Error	SS_{ERR}	$n-k-1$	$MS_{\text{ERR}} = \frac{SS_{\text{ERR}}}{n-k-1}$	
Total	SS_{TOT}	$n-1$		

$n -$ # reg. coeff.