Multiple Regression Models

Let's now consider two input variables, X_1, X_2.

model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

(Y_i = ith response, $X_i = (X_{i1}, X_{i2})$ = ith input)

example: (photography studio, new locations)

X_1 = number of persons aged 16 or older (adults)

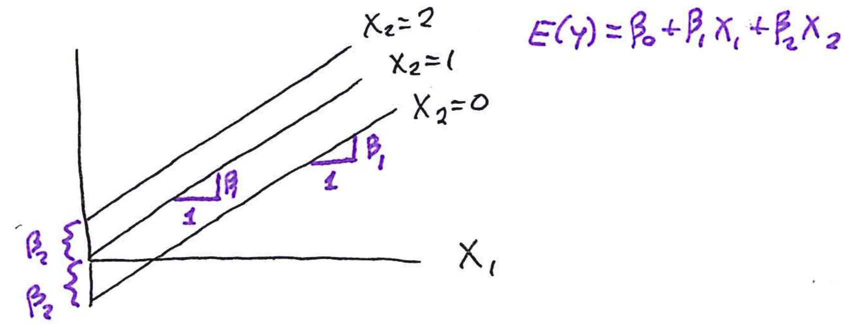
X_2 = per capita disposable income (income)

Y = sales

interaction plot:

(graphing the data)

(understanding the model)



Note that $\dfrac{\partial E(Y)}{\partial X_i} = \beta_i$ (difference in mean response from a 1 unit increase in X_i, with all other input levels held fixed)

(**partial effects**)

In general, we observe $\{(X_{i1},\dots,X_{ir}, Y_i), i=1,\dots,n\}$ modeled as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_r X_{ir} + \varepsilon_i.$$

Let $p = r+1$ be the number of regression parameters.

③

Matrix Approach to Linear Regression

Write
$$\underset{(n\times 1)}{\underline{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$ as the <u>response vector</u>,

$$\underset{(n\times p)}{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1r} \\ \vdots & & \vdots & \\ 1 & X_{n1} & \cdots & X_{nr} \end{bmatrix} = \begin{bmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_n' \end{bmatrix}$$ as the <u>input matrix</u> ([redacted] design matrix)

and
$$\underset{(p\times 1)}{\underline{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}$$ as the <u>parameter vector</u>.

<u>model</u>: $\underset{n\times 1}{\underline{Y}} = \underset{n\times p}{X}\, \underset{p\times 1}{\underline{\beta}} + \underset{n\times 1}{\underline{\varepsilon}}$, where $\varepsilon_1, \dots, \varepsilon_n \sim \text{iid } N(0, \sigma^2)$

(recall matrix multiplication and addition)

<u>example</u>: <u>simple linear regression</u>

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

③

Random vector

$$\underset{(n\times1)}{\underline{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

mean vector

$$E(\underline{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \underset{(n\times1)}{\underline{\mu}}$$

covariance matrix

$$Cov(\underline{Y}) = \begin{bmatrix} V(Y_1) & C(Y_1,Y_2) & \cdots & C(Y_1,Y_n) \\ C(Y_2,Y_1) & V(Y_2) & & \\ \vdots & & \ddots & \\ C(Y_n,Y_1) & & & V(Y_n) \end{bmatrix} = \underset{(n\times n)}{\Sigma}$$

(Σ is symmetric. That is, $\Sigma' = \Sigma$)

regression example: $E(\underline{\varepsilon}) = \underline{0}$, $Cov(\underline{\varepsilon}) = \sigma^2 I$ ← identity matrix

---

Fact: Let $\underset{m\times1}{\underline{W}} = \underset{m\times n}{A}\ \underset{n\times1}{\underline{Y}} + \underset{m\times1}{\underline{b}}$.

Then $\underline{\mu}_W = A\underline{\mu}_Y + \underline{b}$ and $\Sigma_W = A\Sigma_Y A'$

example: $\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$

Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, so $\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} Y_1 - Y_2 \\ Y_1 + Y_2 \end{bmatrix}$

Then

$$\underline{\mu}_W = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}, \quad \Sigma_W = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 + 2\sigma_{12} \end{bmatrix}$$

(compare to the usual equations for means and variances)

Regression model in matrix notation :

$$Y = X\beta + \varepsilon \;, \qquad \varepsilon \sim N_n(0, \sigma^2 I)$$

least squares criterion:

Recall that $|a-b|^2 = (a-b)'(a-b) = \sum (a_i - b_i)^2$

$$Q(\beta) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_r X_{ir} \right)^2 = \left| Y - X\beta \right|^2$$

$Q(\beta)$ is minimized at $b = (X'X)^{-1} X' Y$

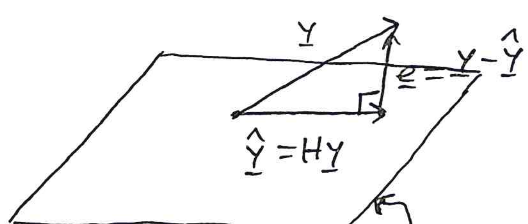
fitted values : $\hat{Y} = Xb = X(X'X)^{-1}X'Y = HY$

($H = X(X'X)^{-1}X'$ is called the hat matrix.)

$$SSE = \left| Y - Xb \right|^2 , \qquad MSE = \frac{SSE}{n-p}$$

($e = Y - Xb$ is the residual vector)

geometric interpretation : $|Y|^2 = |\hat{Y}|^2 + |e|^2$



$c(X) = \{ \mu : \mu = X\beta \}$

p-dimensional subspace of $\mathbb{R}^n$

We can use the rules for covariance matrices to show

$$\text{Cov}(\underline{b}) = \sigma^2 (X'X)^{-1}. \quad \text{Thus, } \quad \underline{\widehat{\text{Cov}}(\underline{b}) = MSE(X'X)^{-1}}$$

example: studios, Y = sales, X_1 = adults, X_2 = income

$$\underline{b} = \begin{bmatrix} -68.86 \\ 1.45 \\ 9.37 \end{bmatrix}, \quad \widehat{\text{Cov}}(\underline{b}) = \begin{bmatrix} \widehat{Var}(b_0) & \hat{C}(b_0,b_1) & \hat{C}(b_0,b_2) \\ & \hat{V}(b_1) & \hat{C}(b_1,b_2) \\ & & \hat{V}(b_2) \end{bmatrix}$$

$$SE(\underline{b}) = \begin{bmatrix} \_\_\_ \\ 0.212 \\ 4.064 \end{bmatrix} \qquad \left(SE(b_k) = \sqrt{\widehat{\text{Cov}}(\underline{b})_{k+1,k+1}} \right)$$

$$\widehat{\text{Corr}}(b_1, b_2) = -0.78 \qquad \left(\text{Corr}(b_1,b_2) = \frac{\text{Cov}(b_1,b_2)}{\sqrt{V(b_1)V(b_2)}} \right)$$

Each of the input variables (pop size, income) has a positive effect on total sales.

Estimating a mean response:

$$\mu_h = E(Y|\underline{X}_h) = \beta_0 + \beta_1 X_{h1} + \dots + \beta_r X_{hr}$$

$$= \underline{X}_h' \underline{\beta}, \quad \text{where } \underline{X}_h' = [1 \ X_{h1} \cdots X_{hr}].$$

$$\underline{\hat{Y}_h = \underline{X}_h' \underline{b} = b_0 + b_1 X_{h1} + \dots + b_r X_{hr}}$$

$$Var(\hat{Y}_h) = \sigma^2 \underline{X}_h' (X'X)^{-1} \underline{X}_h, \quad \underline{SE(\hat{Y}_h) = \sqrt{MSE \cdot \underline{X}_h' (X'X)^{-1} \underline{X}_h}}$$

example: $X_{h1} = 65.4$, $X_{h2} = 17.6$ CI for $\mu_h = [185.29, 196.92]$ estimate of mean sales for all stores

6

Simple linear regression in matrix form

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$$

$$\underline{b} = (X'X)^{-1}X'\underline{Y}, \quad X'X = \begin{bmatrix} n & \sum_1^n X_i \\ \sum_1^n X_i & \sum_1^n X_i^2 \end{bmatrix}, \quad X'\underline{Y} = \begin{bmatrix} \sum_1^n Y_i \\ \sum_1^n X_i Y_i \end{bmatrix}$$

$$Cov(\underline{b}) = \sigma^2 (X'X)^{-1}$$

$$= \begin{bmatrix} Var(b_0) & Cov(b_0, b_1) \\ Cov(b_0, b_1) & Var(b_1) \end{bmatrix}$$

$$\hat{Y}_h = b_0 + b_1 X_h = \begin{bmatrix} 1 & X_h \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \underline{X}_h' \underline{b}$$

$$Var(\hat{Y}_h) = \underline{X}_h' Cov(\underline{b}) \underline{X}_h$$

$$= \begin{bmatrix} 1 & X_h \end{bmatrix} \begin{bmatrix} V(b_0) & C(b_0, b_1) \\ C(b_0, b_1) & V(b_1) \end{bmatrix} \begin{bmatrix} 1 \\ X_h \end{bmatrix}$$