STA302/1001 - Methods of Data Analysis I

(Week 05 - Lecture A)

Wei (Becky) Lin

June 12-16, 2017

UNIVERSITY OF TORONTO

Last Lecture

- Review on matrices
- Simple linear regression model in matrix form.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, \ldots, n$$

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

That is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- $\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- Fitted: $\hat{\boldsymbol{Y}}$=$\mathbf{Xb}$=$\mathbf{HY}$, where $\boldsymbol{H = X(X'X)^{-1}X', (HH = H, H' = H)}$.
- Residuals: $\mathbf{e}$=$\mathbf{Y}$-$\hat{\boldsymbol{Y}}$=$\mathbf{(I\text{-}H)Y}$

Last Lecture (contd..)

- If $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, then we have
    - $\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\beta, \boldsymbol{\sigma}^2\mathbf{I})$
    - $\mathbf{b} \sim \mathrm{N}(\beta, \boldsymbol{\sigma}^2 (\boldsymbol{X'X})^{-1})$
    - $\hat{\boldsymbol{Y}} \sim \mathrm{N}(\mathbf{X}\beta, \boldsymbol{\sigma}^2\mathbf{H})$
    - $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\sigma}^2(\mathbf{I}\text{-}\mathbf{H}))$
    - $\hat{\boldsymbol{Y}}_{\boldsymbol{h}} \sim \mathrm{N}(\boldsymbol{X}_{\boldsymbol{h}}'\beta, \boldsymbol{\sigma}^2 \boldsymbol{X}_{\boldsymbol{h}}'(\boldsymbol{X'X})^{-1}\boldsymbol{X}_{\boldsymbol{h}})$
- ANOVA in matrix form

$$SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - (\sum Y_i)/n = \mathbf{Y}'(\mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J})\mathbf{Y}$$

$$SSE = e'e = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}'[\mathbf{H} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J}]\mathbf{Y}$$

- Mean response and prediction of new observation

$$V(\hat{Y}_h) = \sigma^2[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2}],$$

$$\Rightarrow s^2(\hat{Y}_h) = MSE(X_h'(X'X)^{-1}X_h), \; s_{pred}^2 = MSE(1 + X_h'(X'X)^{-1}X_h).$$

Week 05 Lecture A - Learning objectives & Outcomes

- First order model with two predictor variables.
- Multiple Linear regression (MLR).
    - MLR with dummy variable.
    - MLR with regressor f(X).
        - Polynomial regression.
        - With interaction effect.
        - With transformed predictor.
- Matrix approach to MLR

Chapter 6:
Multiple Linear Regression

Multiple Regression Models

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 f(X_3) + \beta_4 D_i + \cdots \beta_{p-1} X_p$$

- Different options for regressor variables (X's) lead to different model interpretations.
  - Distinct X's.
  - Functions of the X's.
  - Indicator variables.
- Distinct variables (first order model) → Linear in X's
  - E.g. Salary (Y) versus educational level (X_1) and years of experience (X_2).
  - X_1, X_2 measure different quantitles that can affect Y.

First-order Models with Two Predictors Variables

- Two predictor variables X_1, X_2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, i = 1, \ldots, n$$

  - first-order model with two predictor variables: linear in predictor variables.
  - X_1, X_2: additive effect or not to interact (no $X_{i1}X_{i2}$ term in model).

- Assume $E(\epsilon_i)$=0:

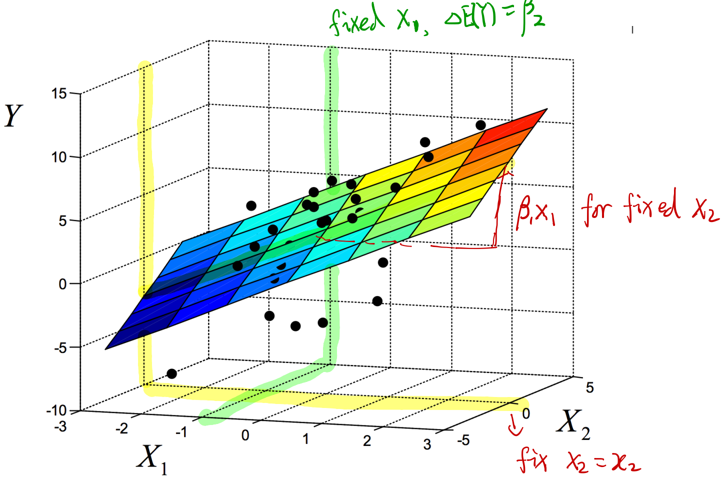
$$E(\mathbf{Y}) = \beta_0 + \beta_1 \boldsymbol{X_1} + \beta_2 \boldsymbol{X_2}$$

  - Mean response is on regression surface or a response surface
  - β_0: intercept in the regression plane: $X_1 = 0, X_2 = 0$.
  - β_1: the change in the mean response with $\triangle X_1 = 1$ and $X_2 =$ constant; $\partial E(Y)/\partial X_1 = \beta_1$
  - β_2: the change in the mean response with $X_1 =$ constant and $\triangle X_2 = 1$; $\partial E(Y)/\partial X_2 = \beta_2$
  - β_1, β_2: partial regression coefficients: they reflect the partial effect.

First-order Models with Two Predictor Variables (contd.)

Model: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Regression surface with two predictors



First-order Models with Two Predictor Variables (contd.)

Example: Regression surface with two predictors

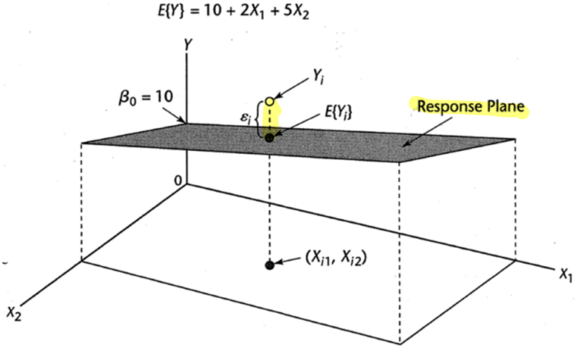


Figure : Response Function is a plane-Sales Promotion Example.

Multiple linear regression (MLR) model

- MLR with $p-1$ predictor variables $(X_1, \ldots, X_{p-1})$

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i(p-1)} + \epsilon_i && (1) \\
&= \beta_0 + \sum_{k=1}^{p-1} \beta_i X_{ik} + \epsilon_i && (2) \\
&= \sum_{k=0}^{p-1} \beta_i X_{ik} + \epsilon_i, \quad \text{where } X_{i0} = 1 && (3)
\end{aligned}
$$

  - Regression coefficient parameters: $\beta_0, \beta_1, \ldots \beta_{p-1}$. $\dim(\beta) = p$
  - Known constants: $X_{i0}, X_{i1}, X_{i2}, \ldots X_{i(p-1)}$
  - $\epsilon_i \sim N(0, \sigma^2)$ and being independent.

- Since $E(\epsilon_i) = 0$, the response mean function, $E(\mathbf{Y})$ is a linear combination of (p-1) predictors which defines a hyper-plane.

$$E(\mathbf{Y}) = \beta_0 + \beta_1 \mathbf{X_1} + \beta_2 \mathbf{X_2} + \ldots + \beta_{p-1} \mathbf{X_{p-1}}$$

Multiple linear regression (MLR) model

$$E(\mathbf{Y}) = \beta_0 + \beta_1 \boldsymbol{X_1} + \beta_2 \boldsymbol{X_2} + \ldots + \beta_{p-1} \boldsymbol{X_{p-1}}$$

- A hyperplane: in more than two dimensions; no long possible to have a picture.
- Additive and do not interact.
- Interpretation of regression coefficients
  - β_0: intercept in the regression hyper-plane: $X_1 = 0, X_2 = 0, \ldots, X_{p-1} = 0$.
  - β_k: the change in the mean response with $\triangle X_k = 1$ and all other predictor variables are held constant.

MLR Model with More than One Predictor Variables

- We have more than one predictor variable in the MLR model.
  - MLR with factor variable (dummy variable)
    - factor variable is a qualitative explanatory variable (or dummy variable), with categories (also called levels).
  - MLR with regressor f(X): a function of X's.
    - Polynomial regression

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2$$

    - Interaction effect

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2}$$

    - Transformed X's

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \exp\{X_{i1}\}$$

MLR model with dummy variable of binary levels.

- Response: Y be the income.
- Predictors: X_1 be the total education years, X_2 be the gender information

$$D_i = \begin{cases} 1, & X_2 = \text{male} \\ 0, & X_2 = \text{female} \end{cases}$$

- SLR model:

$$Y_i = \beta_0 + \beta_2 D_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

  - $\beta_0 = E(Y|D = 0)$, $\beta_1 = E(Y|D = 1) - E(Y|D = 0)$
  - Reg. Coef. estimators: $b_0 = \hat{\beta}_0 = \bar{Y}(D = 0), b_1 = \hat{\beta}_1 = \bar{Y}(D = 1) - \bar{Y}(D = 0)$

- MLR model:

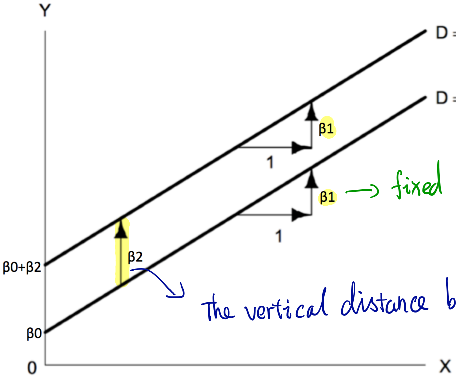
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 D_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

  - $\beta_0 = E(Y|X_1 = 0, D = 0)$,
  - $\beta_1 = E(Y|X_1 = x + 1, D = d_0) - E(Y|X_1 = x, D = d_0), d_0 = 0/1$
  - $\beta_2 = E(Y|X_1 = x_0, D = 1) - E(Y|X_1 = x_0, D = 0)$

MLR model with dummy variable of binary levels

two sub models
$$\begin{cases} D=0: & Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \\ D=1: & Y_i = (\beta_0+\beta_2) + \beta_1 X_{i1} + \varepsilon_i \end{cases}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 D_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$



$: Y_i = (\beta_0+\beta_2) + \beta_1 X_{i1} + \varepsilon_i$

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

fixed $D = D_i$, $\boxed{\Delta E(Y) = \beta_1}$ when $\Delta X = 1$

The vertical distance between two sub models.

The parameters in the additive dummy-regression model.

MLR model with factor variable of multiple levels.

- Response: Y be the income.
- Predictors: X_1 be the total education years. X_2 be the total years of experience, X_3 be the occupations which can be classified into 3 levels: (1) professonal; (2) white-collar; (3) blue-collar.

$$D_{i2} = \begin{cases} 1, & X_3 = \text{white-collar} \\ 0, & X_3 = \text{otherwise} \end{cases} \qquad D_{i3} = \begin{cases} 1, & X_3 = \text{professional} \\ 0, & X_3 = \text{otherwise} \end{cases}$$

- SLR model with only X_3:

$$Y_i = \alpha + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

  - $\alpha = E(Y_{blue-collar})$,
  - $\gamma_2 = E(Y_{white-collar}) - E(Y_{blue-collar})$
  - $\gamma_3 = E(Y_{professional}) - E(Y_{blue-collar})$
  - Regression Coefficient estimators:
    - $b_0 = \hat{\alpha} = \bar{Y}$(blue-collar),
    - $b_1 = \hat{\gamma}_2 = \bar{Y}$(white-collar)-$\bar{Y}$(blue-collar),
    - $b_2 = \hat{\gamma}_3 = \bar{Y}$(professional)-$\bar{Y}$(blue-collar)

MLR model with factor variable of multiple levels (contd.)

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

sub models

- This model describes three parallel regression planes, which can differ in their intercepts
  - Blue collar: $Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$
  - White collar: $Y_i = (\alpha + \gamma_2) + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$
  - Professional collar: $Y_i = (\alpha + \gamma_3) + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$
- Interpreation of regression coefficients
  - $\alpha = E(Y|X_1 = 0, X_2 = 0, X_3 = \text{blue-collar})$: gives the intercept for blue-collar model.
  - $\gamma_2 = E(Y|D_2 = 1, \text{other fixed}) - E(Y|D_2 = 0, \text{other fixed})$: represents the constant vertical distance between the parallel regression planes for white-collar and blue-collar occupations (fixing the values of education and income).
  - $\gamma_3 = E(Y|D_3 = 1, \text{other fixed}) - E(Y|D_3 = 0, \text{other fixed})$: represents the constant vertical distance between the parallel regression planes for professional and blue-collar occupations (fixing the values of education and income).

MLR model with factor variable of multiple levels (contd.)

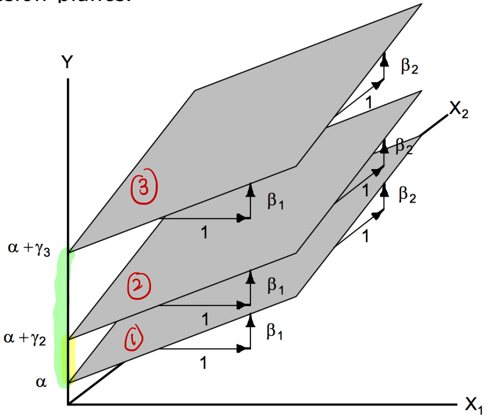
$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

- Interpreation of regression coefficients (contd.)
  - $\beta_1 = E(Y|X_1 = x_1 + 1, X_2 = x_2, X_3 = x_3) - E(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3)$: holding all other predictors fixed, the change in the mean income for one more year of education.
  - $\beta_2 = E(Y|X_1 = x_1, X_2 = x_2 + 1, X_3 = x_3) - E(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3)$: holding all other predictors fixed, the change in the mean income for one more year of experience

MLR model with factor variable of multiple levels (contd.)

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

- The additive dummy-regression model showing three parallel regression planes.



① Blue-collar model

$E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2$

② white-collar model

$E(Y) = (\alpha + \gamma_2) + \beta_1 X_1 + \beta_2 X_2$

③ Professional model

$E(Y) = (\alpha + \gamma_3) + \beta_1 X_1 + \beta_2 X_2$

MLR model: Polynomial Regression (Ch8)

- Special cases of the general linear regression model or MLR model.
- Squared and higher-order terms of the predictors $\Rightarrow$ curvilinear.
- Polynomial regression is useful for describing curvilinear relationships.
  - Quadratic (2nd order)

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2$$

  - Cubic (3rd order)

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

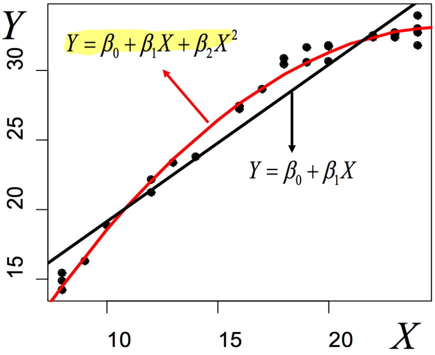
  - Higher order

$$E(Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots$$

Example: Polynomial Regression

- Data
  - Y: female steroid level.
  - X: age.
- Fitted regression model

$$\hat{Y} = -6.21 + 3.08\text{age} - 0.06\text{age}^2$$



MLR with interaction term

- Two-way interaction

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i & (4) \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, X_{i3} = X_{i1} X_{i2} & (5) \end{aligned}$$

- Two predictor variables interact in determining a response variable when the partial effect of one depends on the value of the other.
  - $\partial E(Y)/\partial X_1 = \beta_1 + \beta_3 X_2$. The partial effect of X_1 depends on the value of X_2. Or, the effect of X_1 varies by X_2.
  - Interaction is a symmetric concept — the effect of X_1 varies by X_2, and the effect of X_2 by X_1($\partial E(Y)/\partial X_2 = \beta_2 + \beta_3 X_1$)
- If the regressions in different categories of a dummy variable are not parallel, then the dummy variable interacts with one or more of the quantitative predictor variables. Move to slides 23 -24
- The dummy-regression model can be modified to reflect interactions.
- Three-way interaction

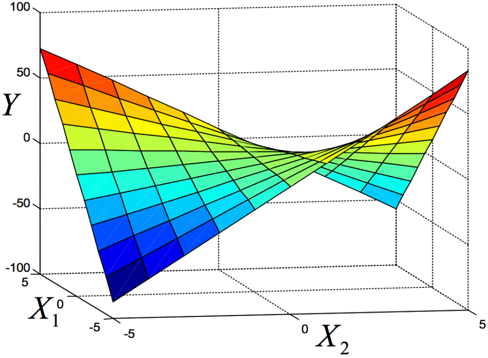
$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_3 X_1 X_2 X_3$$

Visualization: two-way interaction effect

$$E(Y) = 3 + 2X_1 + 2X_2 - 3X_1X_2$$

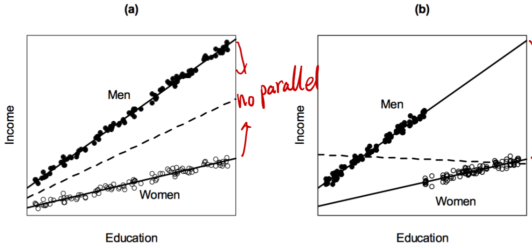
$$\frac{\partial E\{Y\}}{\partial X_1} = 2 - 3X_2$$

$$\frac{\partial E\{Y\}}{\partial X_2} = 2 - 3X_1$$



Example: two-way interaction effect

- Response (Y): income.
- Predictors: X_1 denotes the eduation years; X_2 denotes the gender (D=1 for male,D=0 for female).



no parallel

no parallel

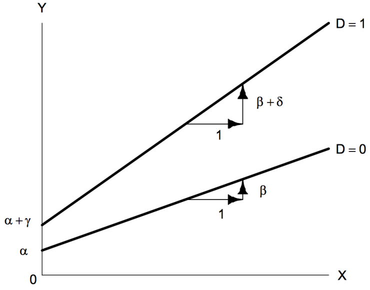
(a), (b) => interaction effect exists

- In (a), gender and education are independent. In (b), women on average have more education than men.
- In both cases, gender and education interact in determing income. The within-gender regressions of income on education are not parallel: the slope for men is larger than the slope for women.
  - the effect of education varies by gender, or the gender effect varies by education.

Example: two-way interaction effect (contd.)

$$Y_i = \alpha + \beta X_i + \gamma D_i + \delta (X_i D_i) + \epsilon_i$$

- Model for female: $Y_i = \alpha + \beta X_i + \epsilon_i$
- Model for male: $Y_i = (\alpha + \gamma) + (\beta + \delta) X_i + \epsilon_i$
    - γ gives the difference in intercepts between the male and female groups.
    - δ gives the difference in slopes between the male and female groups.



D = 1 : $E(Y) = (\alpha+\gamma) + (\beta+\delta)X$

$E(Y) = \alpha + \beta X$

The parameters in the dummy-regression model with interaction.

MLR with Transformed Variables

- Include regressors that are functions of one or several X's
  - E.g. $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 \log(X_2/X_1)$
- Apply some transformation to Y.
  - Curvilinear response mean functions
    - $\log(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$, $Y' = \log(Y)$
    - $Y = 1/(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon)$, $Y' = 1/Y$.

MLR with Interaction Effects and Combination of Cases

- An example of non-additive regression model

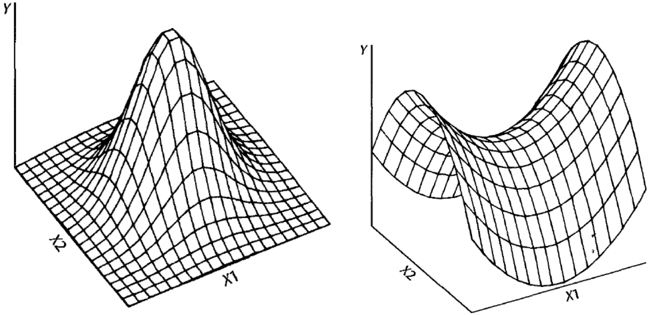
$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i \\
&= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i
\end{aligned}
$$

- By cross-product term:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \epsilon_i \\
&= \beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4 Z_{i4} + \beta_5 Z_{i5} + \epsilon_i
\end{aligned}
$$

MLR with Interaction Effects and Combination of Cases (contd.)

Additional Example of Mean Response Functions



(a) (b)

Meaning of Linear in MLR model

- MLR refers to Multiple Linear Regression.
- MLR is also called General Linear Regression Model (GLRM).
- A regression is linear in parameters:

$$Y_i = \beta_0 C_{i0} + \beta_1 C_{i1} + \beta_2 C_{i2} + \ldots + \beta_{p-1} C_{i(p-1)} + \epsilon_i$$

  - where C_{ik}, k$=0, 1, \ldots p_1$ are coefficients involving the predictor variables, X's.

- Ilustration: nonlinear

$$Y_i = \beta_0 \exp\left(\beta_1 X_i\right) + \epsilon_i$$

Matrix approach to MLR

MLR in Matrix Form

- The MLR model

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i(p-1)} + \epsilon_i, i = 1, \ldots, n$$

- In matrix form $\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}}$

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1(p-1)} \\ 1 & X_{21} & X_{22} & \ldots & X_{2(p-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \ldots & X_{n(p-1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

MLR in Matrix Form (contd.)

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\beta} + \underset{n\times 1}{\boldsymbol{\epsilon}}$$

- $\mathbf{Y}$: vector of responses.
- β: vector of parameters.
- $\mathbf{X}$: matrix of constants (design matrix).
- $\boldsymbol{\epsilon}$: vector of independent normal random variables.

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_{n\times n})$$

- Expectation and variance-covariance matrix of $\mathbf{Y}$:

$$\underset{n\times 1}{\mathbf{E(Y)}} = \mathbf{X}\beta; \quad Var(\mathbf{Y}) = \sigma^2 I_{n\times n}$$

  - That is, $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 I_{n\times n})$.

Estimation of Regression Coefficients in MLR model

- The MLR model

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i(p-1)} + \epsilon_i, i = 1, \ldots, n$$

- The Least squares method: the values of $\beta_0, \ldots, \beta_{p-1}$ minimizes Q

$$Q = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{b})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{b})^{\cancel{T}}$$

$$\Rightarrow \frac{\partial Q}{\partial \mathbf{b}} = -2Y'X + 2b^T X^T X$$

- Setting this equation to zero: $-2Y'X + 2b^T X^T X = 0$
- Transpose both sides of the equation, we have

$$\underset{p\times p}{X'X}\ \underset{p\times 1}{b} = \underset{p\times 1}{X'Y}$$

- Solving it for $\hat{\beta}$. Assume $X'X$ is invertible

$$\Rightarrow LSE: \underset{p\times 1}{b} = \underset{p\times p}{(X'X)^{-1}} \underset{p\times 1}{X'Y}$$

Estimation of Regression Coefficients in MLR model (contd.)

- The method of MLE leads to the same estimators of normal error regression model.
- The likelihood function

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big\{ - \frac{1}{2\sigma^2} \sum_1^n (Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i(p-1)})^2 \Big\}$$

Maximizing the likelihood function w.r.t $\beta_0, \beta_1, \ldots, \beta_{p-1}$ leads to the the same LS estmators $\mathbf{b}$.

Estimation of Regression Coefficients in MLR model (contd.)

$* \begin{cases} ① \ X'X: \text{ is symmetric matrix} \\ ② \ ((X'X)^{-1})^T = (X'X)^{-1} \end{cases}$

$$\begin{aligned} Var(b) &= Var(\underline{(X'X)^{-1}X'} Y) \\ &= \underline{(X^TX)^{-1}X^T}\ Var(Y)\ [\underline{(X^TX)^{-1}X^T}]^T \\ &= \sigma^2 (X^TX)^{-1} \underline{X^T X} (X^TX)^{-1} \\ &= \sigma^2 (X^TX)^{-1} \end{aligned}$$

$$\boxed{\underset{p\times 1}{b} = \underset{p\times p}{(X'X)^{-1}}\underset{p\times 1}{X'Y}} \ *$$

$\boxed{A' = A^T}$

- E(**b**)=E($\underset{p\times p}{(\boldsymbol{X'X})}^{-\mathbf{1}}\underset{p\times 1}{X'Y}$)=$(\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'}E(\mathbf{Y}) = \beta$
- Var(**b**)=Var($(\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'}\mathbf{Y}$)=$\sigma^2(\boldsymbol{X'X})^{-\mathbf{1}}$,
- Var(**b**) is estimated by s^2(**b**)=MSE$(\boldsymbol{X'X})^{-\mathbf{1}}$
- **b**$\sim N(\beta, \sigma^2(\boldsymbol{X'X})^{-\mathbf{1}})$

Fitted Values and Residuals

$$\hat{Y} = \begin{pmatrix}\hat{Y}_1\\ \vdots \\ \hat{Y}_n\end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n}\\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \\ h_{n1} & h_{n2} & \cdots & h_{nn}\end{pmatrix}\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix}$$

$$\Rightarrow \hat{Y}_i = (h_{i1}\ \ h_{i2}\ \cdots\ h_{in})\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix} = \sum_{j=1}^{n} h_{ij}Y_j$$

- The vector of $\hat{Y}_i$.

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \mathbf{X}\boldsymbol{b} = \mathbf{HY},\ \underset{n\times \cancel{1}n}{\boldsymbol{H}} = \boldsymbol{X(X'X)^{-1}X'}$$

  - $\text{E}(\hat{\boldsymbol{Y}})=\text{E}(\mathbf{HY})=\mathbf{HX}\beta=\boldsymbol{X}\beta$
  - $\text{Var}(\hat{\boldsymbol{Y}})=\text{Var}(\mathbf{HY})=\sigma^2\ \mathbf{H}$.
  - $\hat{\mathbf{Y}} \sim N(\boldsymbol{X}\beta, \sigma^2\mathbf{H})$

- The vector of residuals

$$\underset{n\times 1}{\mathbf{e}} = \mathbf{Y} - \hat{\boldsymbol{Y}} = \mathbf{(I\text{-}H)Y}$$

  - $\text{E}(\mathbf{e})=\text{E}(\mathbf{(I\text{-}H)Y})=\mathbf{X}\overset{\text{beta}}{\cancel{\boldsymbol{b}}} - \boldsymbol{X}\beta=\mathbf{0}$
  - $\text{Var}(\mathbf{e})=\text{Var}(\mathbf{(I\text{-}H)Y})=\sigma^2\mathbf{(I\text{-}H)}$,estimated by $s^2(e)=\text{MSE}\mathbf{(I\text{-}H)}$.
  - $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{(I\text{-}H)})$

Analysis of Variance Results (ANOVA)

The sums of squares for ANOVA

$$SSTO = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - (\sum Y_i)/n = \mathbf{Y}'(\mathbf{I} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J})\mathbf{Y}$$

$$SSE = e'e = (\mathbf{Y} - \mathbf{Xb})'(\mathbf{Y} - \mathbf{Xb}) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SSR = \sum(\hat{Y}_i - \bar{Y})^2 = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{Y}'\mathbf{J}\mathbf{Y} = \mathbf{Y}'[\mathbf{H} - \frac{\mathbf{1}}{\mathbf{n}}\mathbf{J}]\mathbf{Y}$$

$$MSR = \frac{SSR}{p-1}$$

$$MSE = \frac{SSE}{n-p}$$

For SLR: $p=2 \Rightarrow$ $\boxed{MSR = \frac{SSR}{1} \quad MSE = \frac{SSE}{n-2}}$

Analysis of Variance Results (ANOVA) (contd.)

Table: ANOVA table for MLR model

Source of Variation	SS	d.f.	MS
Regression	SSR=$\mathbf{b'X'Y}$-$\frac{1}{n}\mathbf{Y'JY}$	p-1	MSR=$\frac{SSR}{p-1}$
Error	SSE=$\mathbf{Y'Y}$-$\mathbf{b'X'Y}$	n-p	MSE=$\frac{SSE}{n-p}$
Total	SSTO= $\mathbf{Y'Y}$-$\frac{1}{n}\mathbf{Y'JY}$	n-1	

- E{MSE}=σ^2
- p-1=2 $\Rightarrow$ $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

$$E\{MSR\} = \sigma^2 + \frac{1}{2}[\beta_1^2 \sum (X_{i1} - \bar{X}_1)^2 + \beta_2^2 \sum (X_{i2} - \bar{X}_2)^2 + 2\beta_1\beta_2 \sum (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)]$$

If $\beta_1 = \beta_2 = 0 \Rightarrow$ E{MSR}=σ^2, otherwise, E{MSR}$\geq \sigma^2$,

Practice problems after lectures

- For MLR model, if $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, show that
  - $\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\beta, \boldsymbol{\sigma}^2\mathbf{I})$
  - $\mathbf{b} \sim \mathrm{N}(\beta, \boldsymbol{\sigma}^2 \boldsymbol{(X'X)^{-1}})$
  - $\boldsymbol{\hat{Y}} \sim \mathrm{N}(\mathbf{X}\beta, \boldsymbol{\sigma}^2\mathbf{H})$
  - $\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\sigma}^2(\mathbf{I}\text{-}\mathbf{H}))$
  - $\boldsymbol{\hat{Y}_h} \sim \mathrm{N}(\boldsymbol{X_h'}\beta, \boldsymbol{\sigma}^2 \boldsymbol{X_h'(X'X)^{-1}X_h})$
- Practice problem in Chapter 6: 6.2, 6.3, 6.5 (a-d), 6.6, 6.7 (a), 6.8, 6.18 (b-e), 6.19, 6.20, 6.21, 6.22, 6.25, 6.27.

Upcoming topics

- Chapter 6:
  - F test for regression coefficients.
  - Coefficient of Multiple Determination.
  - Inferences about Regression Parameters.
  - Interval Estimation of β_k, $E(Y_h)$.
  - F test of lack of fit
- Chapter 7:
  - The extra sum of squares