

STA302/1001 - Methods of Data Analysis I

(Week 09 lecture notes)

Wei (Becky) Lin

Nov. 07, 2016



Last Week

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

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

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

That is

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

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

Last Week (contd..)

- If $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, then we have

- $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I})$
- $\mathbf{b} \sim N(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$
- $\hat{\mathbf{Y}} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{H})$
- $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 (\mathbf{I} - \mathbf{H}))$
- $\hat{\mathbf{Y}}_h \sim N(\mathbf{X}'_h \beta, \sigma^2 \mathbf{X}'_h (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h)$

}

*

- ANOVA in matrix form

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

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

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

- Mean response and prediction of new observation

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

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

Week 09- 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.

Chapter 6: Multiple Linear Regression

Multiple Regression Models

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 f(X_3) + \beta_4 D_i + \dots + \beta_p X_p$$

- 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)
 - E.g. Salary (Y) versus educational level (X_1) and years of experience (X_2).
 - X_1, X_2 measure different quantities that can affect Y.

Linear in X's

First-order Models with Two Predictor Variables

- Two predictor variables X_1, X_2 :

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

- 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 $E(\epsilon_i) = 0$:

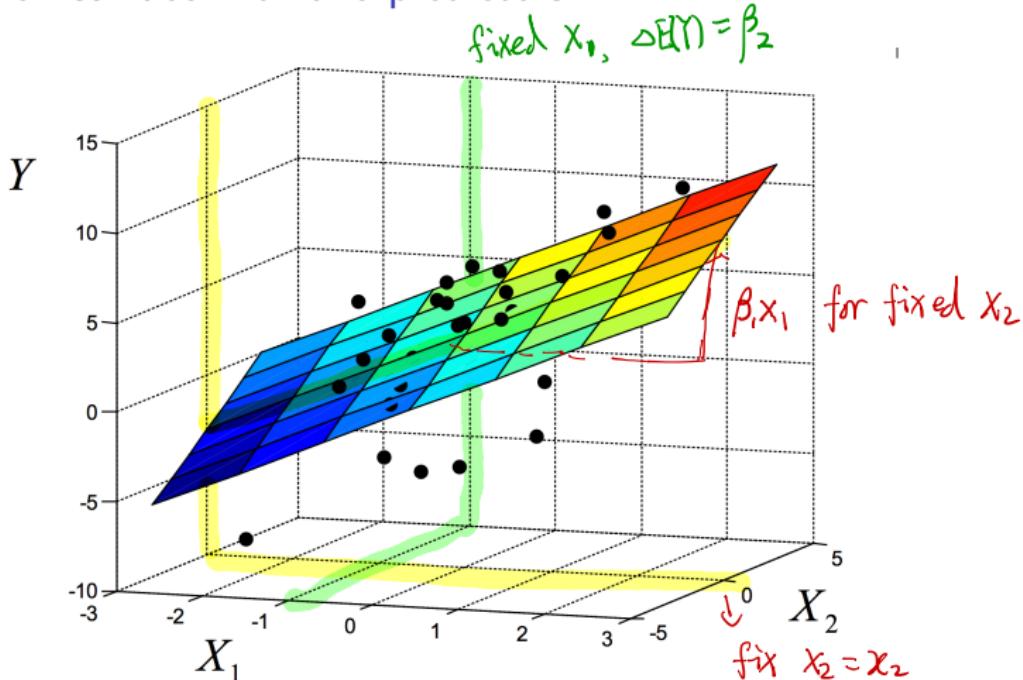
$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- 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 $\Delta X_1 = 1$ and $X_2 = \text{constant}$; $\partial E(Y)/\partial X_1 = \beta_1$
- β_2 : the change in the mean response with $X_1 = \text{constant}$ and $\Delta X_2 = 1$; $\partial E(Y)/\partial X_2 = \beta_2$
- β_1, β_2 : partial regression coefficients: they reflect the partial effect.

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

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

Regression surface with two predictors



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

Example: Regression surface with two predictors

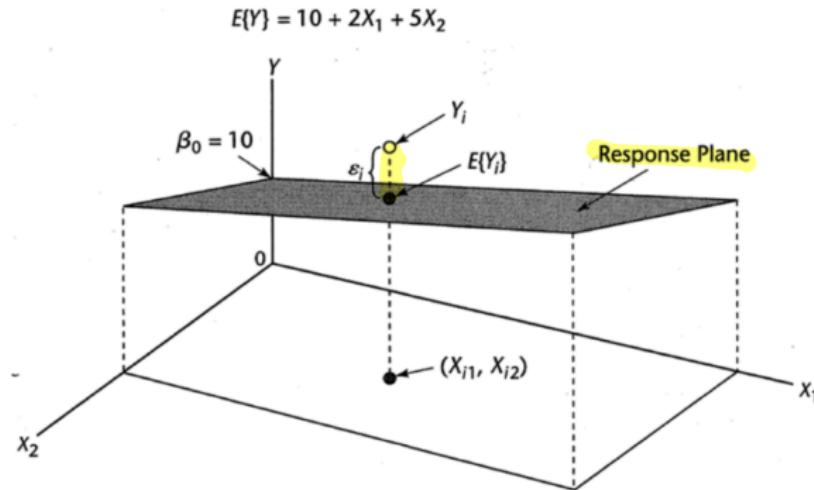


Figure : Response Function is a plane-Sales Promotion Example.

Multiple linear regression (MLR) model

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

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i(p-1)} + \epsilon_i \quad (1)$$

$$= \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \epsilon_i \quad (2)$$

$$= \sum_{k=0}^{p-1} \beta_k X_{ik} + \epsilon_i, \quad \text{where } X_{i0} = 1 \quad (3)$$

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

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

Multiple linear regression (MLR) model

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

- A **hyperplane**: in more than two dimensions; no longer possible to have a picture.
- **Additive** and do not interact.
- **Interpretation of regression coefficients**
 - β_0 : intercept in the regression hyper-plane:
 $X_1 = 0, X_2 = 0, \dots, X_{p-1} = 0$.
 - β_k : the change in the mean response with $\Delta X_k = 1$ 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: $b_0 = \hat{\beta}_0 = \bar{Y}(D=0)$, $b_1 = \hat{\beta}_1 = \bar{Y}(D=1)$

- MLR model:

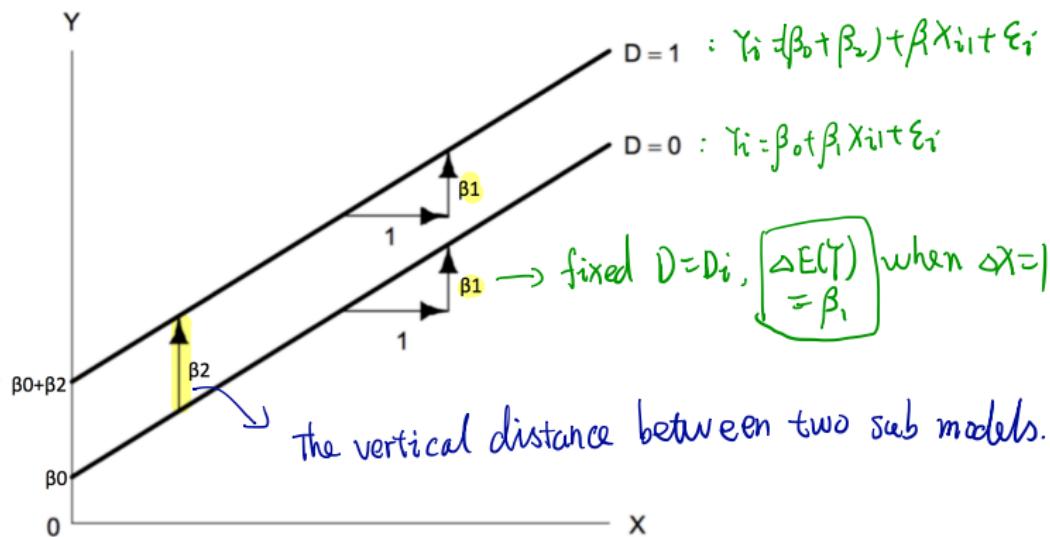
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 D_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

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

MLR model with dummy variable of binary levels

two sub models

$$\left\{ \begin{array}{l} D=0 : Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \\ D=1 : Y_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \epsilon_i \end{array} \right.$$
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 D_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$



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) professional; (2) white-collar; (3) blue-collar.

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

- SLR model with only X_3 :

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

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

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

Sub models

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

- 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: $Y_i = (\alpha + \gamma_3) + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$

- Interpretation 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.)

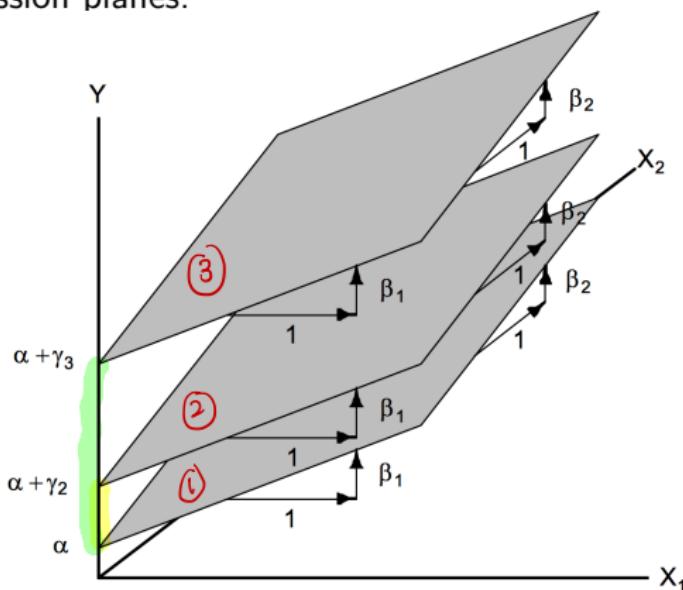
$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

- Interpretation 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)

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

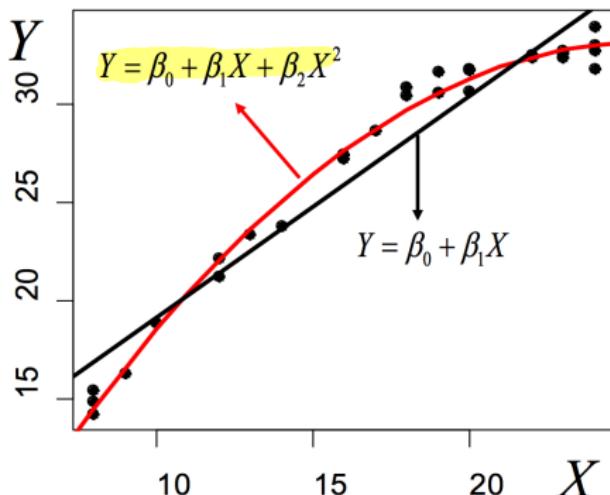
- Higher order

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

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

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i \quad (4)$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, X_{i3} = X_{i1} X_{i2} \quad (5)$$

- 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

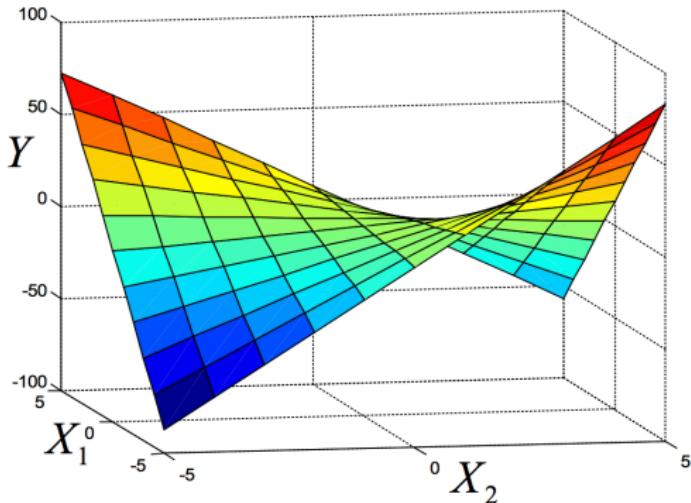
$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_3 X_1 X_2 X_3$$

Visualization: two-way interaction effect

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

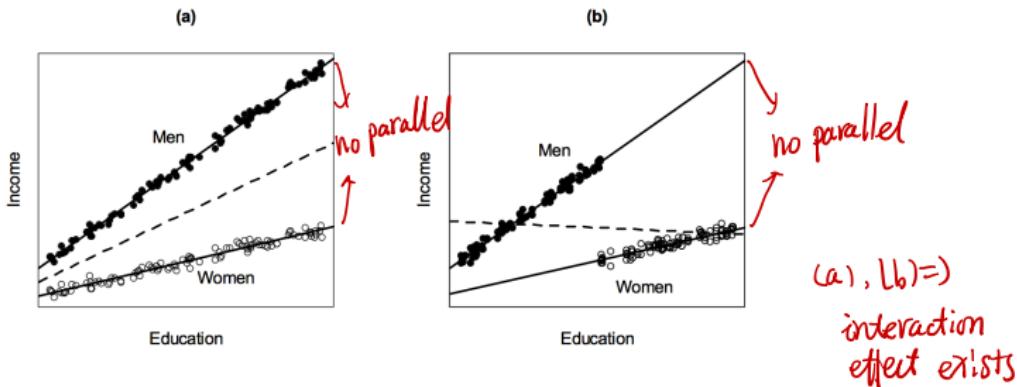
$$\frac{\partial E\{Y\}}{\partial X_1} = 2 - 3X_2$$

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



Example: two-way interaction effect

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

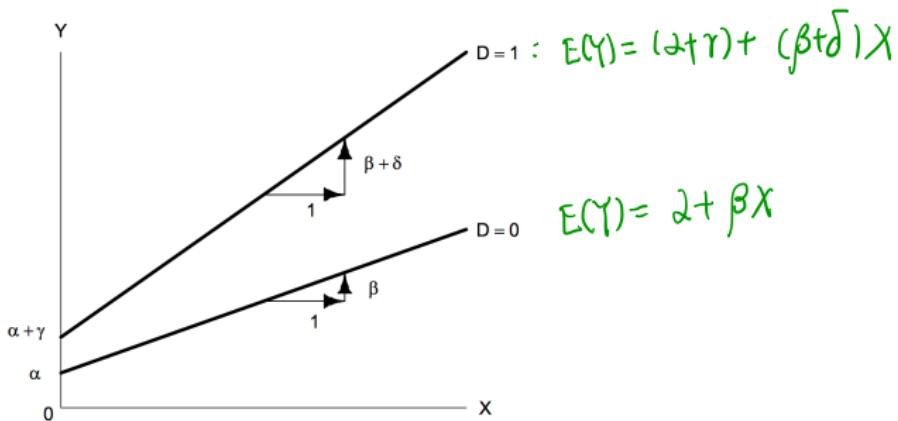


- In (a), gender and education are independent. In (b), women on average have more education than men.
- In both cases, gender and education interact in determining income. The within-gender regressions of income on education are not parallel: the slope for men is larger than the slope for women.
 -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



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

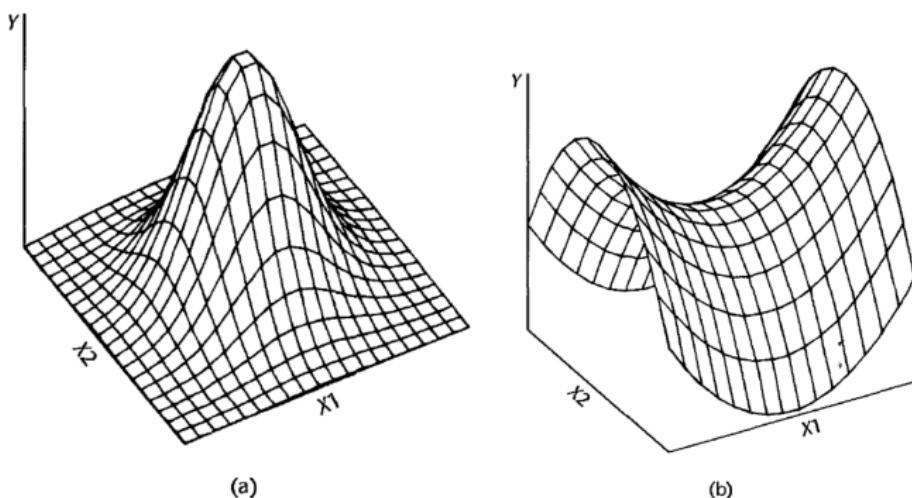
$$\begin{aligned}Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i && (\text{b}) \text{ next slide} \\&= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i2} + \epsilon_i\end{aligned}$$

- By cross-product term:

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

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

Additional Example of Mean Response Functions



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:

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

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

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

Matrix approach to MLR

MLR in Matrix Form

- The MLR model

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

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

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

MLR in Matrix Form (contd.)

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

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

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

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

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

- 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

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

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

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

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

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

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

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

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

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

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

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

Estimation of Regression Coefficients in MLR model (contd.)

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

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

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

$$A' = A^T$$

Fitted Values and Residuals

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & & & \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Rightarrow \hat{y}_i = (h_{i1} \ h_{i2} \ \cdots \ h_{in}) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{j=1}^n h_{ij} y_j$$

- The vector of \hat{Y}_i .

$$\hat{\mathbf{Y}}_{n \times 1} = \mathbf{X}\mathbf{b} = \mathbf{H}\mathbf{Y}, \quad \mathbf{H}_{n \times n} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

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

- The vector of residuals

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

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

Analysis of Variance Results (ANOVA)

The sums of squares for ANOVA

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

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

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

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

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

For SLR: $p=2 \Rightarrow$

$$\boxed{\begin{aligned} MSR &= \frac{SSR}{1} \\ MSE &= \frac{SSE}{n-2} \end{aligned}}$$

Analysis of Variance Results (ANOVA) (contd.)

Table: ANOVA table for MLR model

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

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

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

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

Practice problems after Week 09 lectures

- For MLR model, if $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, show that
 - $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I})$
 - $\mathbf{b} \sim N(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$
 - $\hat{\mathbf{Y}} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{H})$
 - $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 (\mathbf{I}-\mathbf{H}))$
 - $\hat{\mathbf{Y}}_h \sim N(\mathbf{X}'_h \beta, \sigma^2 \mathbf{X}'_h (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h)$
- 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