Purdue-NCKU program

Lecture 7
Simple Linear Regression

Dr. Qifan Song

Goals of Regression Analysis

Regression: use data (Y_i, X_i) to find out a relationship

$$E(Y) = f_\beta(X),$$

or median, mode of Y if possible.

- Serve three purposes
  - Describes an association between X and Y
    * In some applications, the choice of which variable is X and which is Y can be arbitrary
    * Association generally does not imply causality
  - In experimental settings, helps select X to control Y at the desired level
  - Predict a future value of Y at a specific value of X

Straight Line Mean Equation

- Formula for a straight line

$$E(Y_i) = \beta_0 + \beta_1 X_i, \text{ or } E(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

  - β_0 is the intercept
  - β_1 is the slope

- Need to **estimate** β_0 and β_1
  i.e. determine their plausible values from the data

- Will use method of **least squares** (OLS estimator).

Simple Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- β_0 is the intercept
- β_1 is the slope
- ε_i is the i^{th} random error term
  - Mean 0, i.e. $E(\varepsilon_i) = 0$
  - Constant Variance σ^2, i.e. $Var(\varepsilon_i) = \sigma^2$
  - Uncorrelated, i.e. $Cov(\varepsilon_i, \varepsilon_j) = 0$
  - Independent to X_i if X_i is random

Estimation of Regression Function

- Consider the deviation of observed data Y_i from a straight line with slope a and intercept b,

$$Y_i - (aX_i + b)$$

  it measures how good the line $ax + b$ fits the data (X_i, Y_i) in terms of vertical distance

- Method of least squares (smallest sum of squared derivation)
  - Find the value of a and b which minimize

$$Q = \sum_{i=1}^{n} [Y_i - (aX_i + b)]^2$$

  - Motivated by $E(Y) = \arg\min_b E(Y - b)^2 \approx \arg\min_b \sum (Y_i - b)^2/n$.

Estimating β's

- β_1 is the true unknown slope
  - Defines change in $E(Y)$ for change in X, i.e.,

$$\beta_1 = \frac{\Delta E(Y)}{\Delta X}$$

- b_1 is the least squares estimate of β_1

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- β_0 is the true unknown intercept
  - β_0 is the expected value of Y under $X = 0$

$$E(Y) = \beta_1 X + \beta_0 = \beta_1 \times X + \beta_0 = \beta_0$$

- b_0 is the least squares estimate of β_0

$$b_0 = \overline{Y} - b_1 \overline{X}$$

  that is, the fitted line goes through $(\overline{X}, \overline{Y})$.

Properties of Estimates

- b_1 is a linear estimator, i.e., a linear combination of Y_i's.

$$\begin{aligned} b_1 &= \sum_{i=1}^{n} \frac{(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ &= \sum_{i=1}^{n} \frac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} Y_i \\ &= \sum_{i=1}^{n} k_i Y_i \end{aligned}$$

where $k_i = (X_i - \overline{X})/\sum_{i=1}^{n}(X_i - \overline{X})^2$

- Note that $\sum k_i = 0$, $\sum k_i X_i = 1$, thus

$$\begin{aligned} E(b_1) &= \sum k_i E(Y_i) = \sum k_i(\beta_0 + \beta_1 X_i) \\ &= \beta_0 \sum k_i + \beta_1 \sum k_i X_i \\ &= 0 + \beta_1, \end{aligned}$$

- b_0 is also a linear combination of Y_i's,

$$\begin{aligned} b_0 &= \overline{Y} - b_1\overline{X} = \sum_{i=1}^{n} \frac{1}{n} Y_i - \overline{X} \sum_{i=1}^{n} k_i Y_i \\ &= \sum_{i=1}^{n} \left(\frac{1}{n} - \frac{\overline{X}(X_i - \overline{X})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right) Y_i \\ &= \sum_{i=1}^{n} \tilde{k}_i Y_i \end{aligned}$$

where

$$\tilde{k}_i = \frac{1}{n} - \frac{\overline{X}(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$$

- Note $\sum \tilde{k}_i = 1$, $\sum \tilde{k}_i X_i = 0$.

$$\begin{aligned} E(b_0) &= \sum \tilde{k}_i E(Y_i) = \sum \tilde{k}_i (\beta_0 + \beta_1 X_i) \\ &= \beta_0 \sum \tilde{k}_i + \beta_1 \sum \tilde{k}_i X_i \\ &= \beta_0 + 0, \end{aligned}$$

Estimated Regression Line

- Using the estimated parameters, the fitted regression line is

$$\widehat{Y}_i = b_0 + b_1 X_i$$

where $\widehat{Y}_i$ is the estimated value at X_i (Fitted value).

- Fitted value $\widehat{Y}_i$ is also an estimate of the mean response $E(Y_i)$
- $\widehat{Y}_i = \sum_{j=1}^{n}(\tilde{k}_j + X_i k_j)Y_j = \sum_{j=1}^{n} \check{k}_{ij} Y_j$ is also a linear estimator
- $E(\widehat{Y}_i) = E(b_0 + b_1 X_i) = E(b_0) + E(b_1)X_i = \beta_0 + \beta_1 X_i = E(Y_i)$

- *Gauss-Markov* theorem: b_0, b_1 and $\widehat{Y}_i$ have minimum variance among all unbiased linear estimators.

Residuals

- The *residual* is the difference between the observed and fitted values

$$e_i = Y_i - \widehat{Y}_i$$

- This is not the error term $\varepsilon_i = Y_i - E(Y_i)$
- The e_i is observable while ε_i is not
  - $\sum e_i$=0
  - $\sum Y_i = \sum \widehat{Y}_i$
  - $\sum X_i e_i = 0$
  - $\sum \widehat{Y}_i e_i = 0$

Estimation of Error Variance

- In regression model

$$s^2 = \frac{\sum(Y_i - \widehat{Y}_i)^2}{n-2}$$

  - Also known as the *mean square error* (MSE)
  - Two df lost by using (b_0, b_1) in place of (β_0, β_1)
  - unbiased estimation

$$\begin{aligned}
\sum(Y_i - \widehat{Y}_i)^2 &= \sum e_i(Y_i - \widehat{Y}_i) = \sum e_i Y_i \\
&= \sum e_i(\beta_0 + \beta_1 X_i + \varepsilon_i) = \sum e_i \varepsilon_i \\
&= \sum Y_i \varepsilon_i - \sum \widehat{Y}_i \varepsilon_i, \\
E(Y_i \varepsilon_i) &= E(\beta_0 + \beta_1 X_i + \varepsilon_i)\varepsilon_i = E(\varepsilon^2) = \sigma^2, \\
E(\widehat{Y}_i \varepsilon_i) &= E(\sum_{j=1}^{n} \breve{k}_{ij} Y_j)\varepsilon_i = E\breve{k}_{ii} Y_i \varepsilon_i = \breve{k}_{ii}\sigma^2
\end{aligned}$$

Note that $\breve{k}_{ii} = \tilde{k}_i + X_i k_i$ and the properties of k_i and $\tilde{k}_i$, we can show that $E(s^2) = \sigma^2$

Normal Error Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \varepsilon_i \sim^{iid} N(0, \sigma^2)$$

- the random error term is assumed to be **independent normally** distributed
- Defines distribution of random variable Y_i

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

- The normality assumption will greatly simplifies the theory of analysis beyond estimations, allows us to construct confidence intervals / perform hypothesis tests
- Most inferences are only sensitive to large departures from normality

Sampling Distribution of b_i's

- Under the normality assumption, b_1 also follows a normal distribution since it is a linear combination of normal r.v.s.

- It is sufficient to figure the first two moments of b_1:

$$
\begin{aligned}
E(b_1) &= \beta_1, \\
Var(b_1) &= \sum k_i^2 Var(Y_i) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}.
\end{aligned}
$$

- Therefore

$$b_1 \sim N(\beta_1, \sigma^2 / \sum_{i=1}^n (X_i - \overline{X})^2)$$

- b_0 is also a linear combination of Y_i's, $b_0 = \sum \tilde{k}_i Y_i$, thus by normal assumption, $b_0 \sim N(E(b_0), Var(b_0))$, where

$$
\begin{aligned}
E(b_0) &= \beta_0, \\
Var(b_0) &= \sum \tilde{k}_i^2 Var(Y_i) = \sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right].
\end{aligned}
$$

t-test for $H_0 : \beta_1 = \beta_1^0$

- Consider statistics $\frac{b_1-\beta_1^0}{se(b_1)}$, where standard error of b_1 means an estimation for the $\sqrt{Var(b_1)}$.
- $se(b_1) = \sqrt{s^2/\sum_{i=1}^n (X_i - \overline{X})^2}$, that is to replace the unknown σ^2 by its unbiased estimator $s^2 = MSE = \sum(Y_i - \widehat{Y}_i)^2/(n-2)$,
- The test statistics is

$$\frac{b_1 - \beta_1^0}{se(b_1)} = \frac{b_1 - \beta_1^0}{\sqrt{s^2/\sum_{i=1}^n (X_i - \overline{X})^2}}$$

- $(n-2)s^2/\sigma^2 \sim \chi^2_{n-2}$ and s^2 is independent to b_1 (will be proved later), then $\frac{b_1-\beta_1^0}{se(b_1)} \sim t_{n-2}$

t-test for $H_0 : \beta_1 = \beta_1^0$

t-test statistics $t^* = \frac{b_1 - \beta_1^0}{se(b_1)}$ under level α

- Reject if $|t^*| \geq t(1 - \alpha/2, n-2)$ or p-value $P(|t_{n-2}| \geq |t^*|) \leq \alpha$

C.I for β_1

$b_1 \pm t(1 - \alpha/2, n-2) se(b_1)$

Similar inference for β_0.

Comments

- When errors not normal, procedures are generally reasonable approximations
- Procedures can be modified for one-sided test / confidence intervals
- To obtain an accurate interval estimation, at design stage, choose X_i such that
  - $\sum(X_i - \overline{X})^2$ is large $\rightarrow$ smaller margin of error for β_1
  - $\sum(X_i - \overline{X})^2$ is large and $|\overline{X}|$ is small $\rightarrow$ smaller margin of error for β_0

Interval Estimation of $E(Y_h)$

- Often interested in estimating the mean response for particular X_h, i.e., the parameter of interests is $E(Y_h) = \beta_0 + \beta_1 X_h$.
- Unbiased estimation is $\widehat{Y}_h = b_0 + b_1 X_h$.
- Derive the sampling distribution of $b_0 + b_1 X_h$ in order to make test and CI.
  - $\widehat{Y}_h = \sum \breve{k}_i Y_i$ where $k_i = \frac{1}{n} + \frac{(X_h - \overline{X})(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$
  - $E(\widehat{Y}_h) = \beta_0 + \beta_1 X_h$
  - $Var(\widehat{Y}_h) = \sigma^2 \left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]$, $se^2(\widehat{Y}_h) = s^2 \left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right]$.
  - Test: $(\widehat{Y}_h -$null value$)/se(\widehat{Y}_h)$; CI: $\widehat{Y}_h \pm t(1 - \alpha/2, n-2) se(\widehat{Y}_h)$

Interval Prediction of $Y_{h(new)}$

- Predicting future observation $Y_{h(new)} = E[Y_h] + \varepsilon_{h(new)}$
- The prediction interval for a unknown r.v., i.e., $P(L < Y_{h(new)} < U) = 1 - \alpha$
- Comparing with CI of $E[Y_h]$, one need to take account of future error $\varepsilon_{h(new)}$.
  - $E(b_0 + b_1 X_h) = E(Y_{h(new)})$
  - $Var(Y_{h(new)} - b_0 + b_1 X_h) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right]$
  - $se^2(Y_{h(new)} - b_0 + b_1 X_h) = s^2 \left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right]$
  - PI: $b_0 + b_1 X_h \pm t(1 - \alpha/2, n-2)\sqrt{s^2 \left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right]}$

Confidence Band for Response Means

- Consider the *entire regression line*
- Want to define a likely region within which this unknown real line lies
- Rigorously, $P(L(x) < \beta_0 + \beta_1 x < U(x) \text{ for all } x) \geq 1 - \alpha$
- One can show

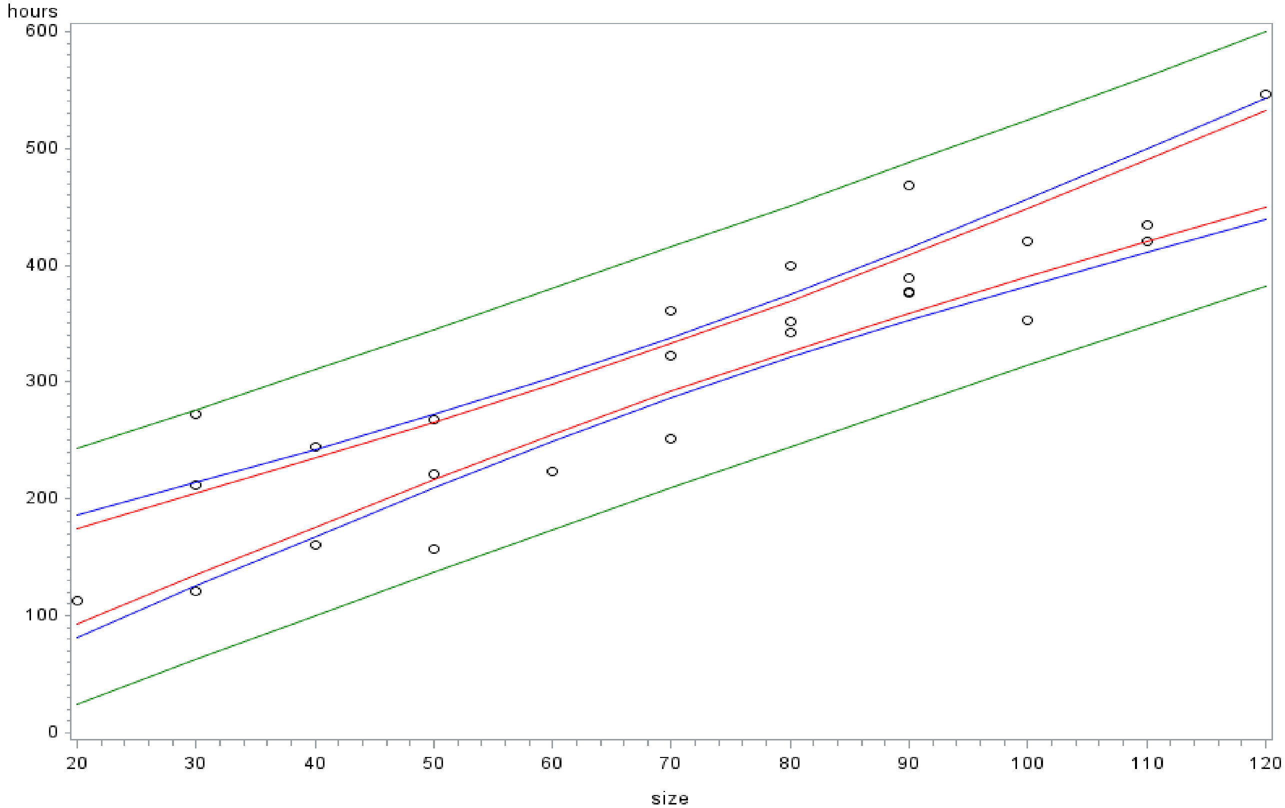
$$\max_x \left[\frac{b_0 + b_1 x - (\beta_0 + \beta_1 x)}{se(\widehat{Y}_h)(x)} \right]^2 \sim 2F_{2,n-2}$$

- Replace $t(1-\alpha/2, n-2)$ with Working-Hotelling value in each confidence interval

$$W = \sqrt{2F(1-\alpha; 2, n-2)} \Rightarrow \widehat{Y}_h \pm W \times se(\widehat{Y}_h)$$

- Boundary values define a hyperbola

Confidence Band vs C.I. vs P.I.



- Blue – 95% confidence band; widest when $X_h - \overline{X}$ is large
- Red – 95% confidence interval for the mean; always narrowest
- Green – 95% confidence interval for the individual prediction; widest when $X_h - \overline{X}$ is small

ANOVA approach

- Organizes results arithmetically
- The total sum of squares in Y is defined

$$\text{SSTO} = \sum (Y_i - \overline{Y})^2$$

- Can partition the total sum of squares into
  - Model (explained by regression)
  - Error (unexplained / residual)

$$\begin{aligned} \sum (Y_i - \overline{Y})^2 &= \sum (Y_i - \hat{Y}_i + \hat{Y}_i - \overline{Y})^2 \\ &= \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \overline{Y})^2 \\ \text{SSTO} &= \quad \text{SSE} \quad + \quad \text{SSR} \end{aligned}$$

Sum of Squares

- Can also express

$$\begin{aligned} \text{SSR} &= \sum(\hat{Y}_i - \overline{Y})^2 \\ &= \sum(b_0 + b_1 X_i - b_0 - b_1 \overline{X})^2 \\ &= b_1^2 \sum(X_i - \overline{X})^2 \end{aligned}$$

- Degrees of freedom is 1 due to normality of b_1 for some δ.
- SSR large when $\hat{Y}_i$'s are different from $\overline{Y}$
- Error sum of squares is equal to the sum of squared residuals

$$\text{SSE} = \sum(Y_i - \hat{Y}_i)^2 = \sum e_i^2$$

- Degrees of freedom is $n-2$ due to using (b_0, b_1) in place of (β_0, β_1), and SSE $\sim \sigma^2\chi^2_{n-2}$
- The MSE = SSE/$(n-2)$ and represents an unbiased estimate of σ^2 when taking X into account

F Test

- Can use this structure to test $H_0 : \beta_1 = 0$
- Consider

$$F^* = \frac{\text{MSR}}{\text{MSE}}$$

- If $\beta_1 = 0$, then F^* should be near one, since both denominator and numerator are of mean
$sigma^2$.
- Need sampling distribution of F^* under H_0 to obtain p-value.
-

$$F^* sim \;\; F_{1,n-2}$$

- When H_0 is false, F^* tends to be large
- p-value $= Pr(F(1, n-2) > F^*)$
- Reject when $F^* > F_{1,n-2,1-\alpha}$, or p-value $< \alpha$

General Linear Test

- A third way to test for linear association
- Consider **two** models
  - Full model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - Reduced model: $Y_i = \beta_0 + \varepsilon_i$
- Will compare models using SSE's
  - Error sum of squares of the full model will be labeled SSE(F)
  - Error sum of squares of the reduced model will be labeled SSE(R)
- Note: SSTO is the same under each model

- Reduced model corresponds to $H_0 : \beta_1 = 0$
- Can be shown that SSE(F) $\leq$ SSE(R)
- Idea: more parameters provide better fit
- If SSE(F) is not much smaller than SSE(R), full model doesn't better explain Y.

$$
\begin{aligned}
F^* &= \frac{(\text{SSE(R)} - \text{SSE(F)})/(df_R - df_F)}{\text{SSE(F)}/df_F} \\
&= \frac{(\text{SSTO} - \text{SSE})/1}{\text{SSE(F)}/(n-2)}
\end{aligned}
$$

- Same test as before, but will have a more general use in multiple regression

Diagnose

- use a residual plots (e_i vs $\widehat{Y}_i$) to check
  - linearity
  - constant variance
  - never plot e_i vs Y_i
  - normality
- Boxcox transformation

Matrix Approach

- Consider example with $n = 4$
- Consider expressing observations:

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\
Y_3 &= \beta_0 + \beta_1 X_3 + \varepsilon_3 \\
Y_4 &= \beta_0 + \beta_1 X_4 + \varepsilon_4
\end{aligned}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \beta_0 + \beta_1 X_3 \\ \beta_0 + \beta_1 X_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\mathbf{X}$ is called the design matrix

Regression Matrices

- Can express observations

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- Both $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ are random vectors

$$\begin{aligned} E(\mathbf{Y}) &= \mathbf{X}\boldsymbol{\beta} + E(\boldsymbol{\varepsilon}) \\ &= \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\sigma}^2(\mathbf{Y}) &= 0 + \boldsymbol{\sigma}^2(\boldsymbol{\varepsilon}) \\ &= \sigma^2\mathbf{I} \end{aligned}$$

Least Squares

- Express quantity Q

$$\begin{aligned} Q &= (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}'\mathbf{Y}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}-\mathbf{Y}'\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{Y}'\mathbf{Y}-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

  - $(\mathbf{X}\boldsymbol{\beta})' = \boldsymbol{\beta}'\mathbf{X}'$

- Taking derivative $\longrightarrow -2\mathbf{X}'\mathbf{Y}+2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=0$
  - $\frac{\partial}{\partial\beta}\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y}=\mathbf{X}'\mathbf{Y}$
  - $\frac{\partial}{\partial\beta}\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$

- This means $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Fitted Values

- The fitted values $\hat{\mathbf{Y}} = \mathbf{X}b = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- Matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is called the *hat matrix*
  - $\mathbf{H}$ is symmetric, i.e., $\mathbf{H}' = \mathbf{H}$
  - $\mathbf{H}$ is idempotent, i.e., $\mathbf{HH} = \mathbf{H}$
- Equivalently write $\hat{\mathbf{Y}} = \mathbf{HY}$
- Matrix $\mathbf{H}$ used in diagnostics (Chapter 9)

Residuals

- Residual matrix

$$\begin{aligned} \mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} \\ &= \mathbf{Y} - \mathbf{HY} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{Y} \end{aligned}$$

- $\mathbf{e}$ is a random vector

$$\begin{aligned} E(\mathbf{e}) &= (\mathbf{I} - \mathbf{H}) \times E(\mathbf{Y}) \\ &= (\mathbf{I} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \boldsymbol{\sigma}^2(\mathbf{e}) &= (\mathbf{I} - \mathbf{H}) \times \boldsymbol{\sigma}^2(\mathbf{Y}) \times (\mathbf{I} - \mathbf{H})' \\ &= (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})' \\ &= (\mathbf{I} - \mathbf{H})\sigma^2 \end{aligned}$$

Inference

- Vector $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{A}\mathbf{Y}$
- The mean and variance are

$$\begin{aligned} E(\mathbf{b}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

$$\begin{aligned} \boldsymbol{\sigma}^2(\mathbf{b}) &= \mathbf{A} \times \boldsymbol{\sigma}^2(\mathbf{Y}) \times \mathbf{A}' \\ &= \mathbf{A} \times \sigma^2\mathbf{I} \times \mathbf{A}' \\ &= \sigma^2\mathbf{A}\mathbf{A}' \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- Thus, $\mathbf{b}$ is *multivariate* Normal($\boldsymbol{\beta}$, $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$)
- $SSE = (\mathbf{e})'(\mathbf{e}) = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \sim \sigma^2\chi^2_{n-2}$
- Since $\mathbf{b}$ and $(\mathbf{e})$ are indep, SSE and $\mathbf{b}$ are independent

Chapter Review

- Simple linear regression
- OLS estimation and prediction
- Inference and confidence band