STA302/1001 - Methods of Data Analysis I

(Week 05- Lecture B)

Wei (Becky) Lin

June 12-16, 2016

UNIVERSITY OF TORONTO

Last Lecture

- 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

Week 5 Lecture B - Learning objectives & Outcomes

- Geometric perspective of least squares regression
- F test for regression coefficients.
- Coefficient of Multiple Determination.
- Inferences about Regression Parameters.
- Interval Estimation of β_k, $E(Y_h)$.
- Extra sum of squares

Geometric perspective of LS regression

Review of Orthogonal Projection

- Two vectors U,V$\in R^n$ are said to be orthogonal (denoted U$\perp$ V) if $< U, V >$=U$\cdot$ V=0.
- A vector U$\in R^n$ is said to be orthogonal to a nonempty set $X \subset R^n$ (denoted U$\perp$ X) if $< U, x >= U \cdot x = 0$ for all $x \in X$.
- Nonempty sets $U, V \in R^n$ are said to be orthogonal (denoted U$\perp$ V) if $u \cdot v$ =0 for any $u \in U, v \in V$.
- Let S be a subset of R^n, the orthogonal complement of S, denoted of $S^{\perp}$, is the set of all vectors $x \in R^n$ that are orthogonal to S.
- Theorem: Let V be a subspace of R^n. Then
    - $V^{\perp}$ is also a subspace of R^n.
    - $V \cap V^{\perp} = \{0\}$.
    - dim V+ dim $V^{\perp}$ = n.
    - $R^n = V \oplus V^{\perp}$ (direct sum), which means that any vector $x \in R^n$ is uniquely represented as $x = p + o$ where $p \in V$ and $o \in V^{\perp}$.

Geometric perspective of LS regression

Reiew of Orthogonal Projection (contd.)

- Projetion matrix P: $\boldsymbol{P = P^T = P^2 = PP}$.
  - Project vector y to x:

$$P = \frac{xx^T}{x^Tx}, P_x(y) = Py = \frac{xx^T}{x^Tx}y = \frac{x^Ty}{x^Tx}x = \frac{\langle X, Y \rangle}{\langle X, X \rangle} X$$

  - Project vector y to the column space of X: $P = X(X^TX)^{-1}X^T$

$$X = \begin{bmatrix} \vdots & \vdots \\ x_1 & x_2 \\ \vdots & \vdots \end{bmatrix}, P_X(Y) = PY = \hat{Y} = b_1 \begin{bmatrix} \vdots \\ x_1 \\ \vdots \end{bmatrix} + b_2 \begin{bmatrix} \vdots \\ x_2 \\ \vdots \end{bmatrix} = Xb$$

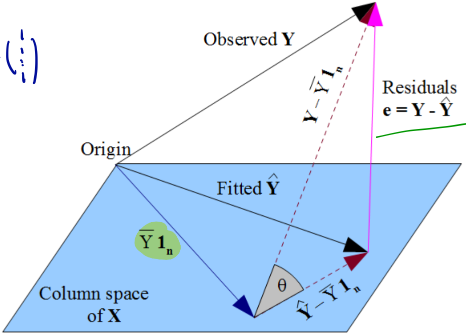
Geometric perspective of LS regression

- LS regression: orthogonal projection of Y to the column space of X

$$\hat{Y} = HY = X(X^TX)^{-1}X^TY = Xb, H = X(X^TX)^{-1}X^T$$

project Y to $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$$= \frac{\langle Y, 1 \rangle}{\langle 1, 1 \rangle} \mathbf{1} = \frac{\sum Y_i}{n} \mathbf{1} = \bar{Y} \mathbf{1}$$



$e \perp col(X)$

$\Downarrow$

$e \perp \forall v \in col(X)$

eg ① $e \perp 1' \Leftrightarrow \langle e, 1 \rangle = 0$

② $e \perp x \Leftrightarrow e \cdot x = 0$

http://stats.stackexchange.com/questions/194523/why-does-the-sum-of-residuals-equal-0-from-a-graphical-perspective

- $\sum_i e_i = 0 \Leftrightarrow < e, 1 >= e \cdot 1 = 0$, the vector of residuals **e** is orthogonal to the vector of ones, **1**. $\because 1 \in col(X)$
- $\sum_i x_i e_i = 0 \Leftrightarrow < e, x >= e \cdot x = 0$, the vector of residuals **e** is orthogonal (perpendicular) to the vector of **x**. $\because x \in col(X)$

Geometric perspective of LS regression

- LS regression: orthogonal projection of Y to the column space of **X**

$$\hat{\boldsymbol{Y}} = \boldsymbol{HY} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{Xb}, \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

- Since $\boldsymbol{H}$ is an orthogonal projection onto the column space of **X**, can show
  - $\boldsymbol{HXb} = \boldsymbol{Xb} = \hat{Y}$ $\Rightarrow$ $Xb \in$ column space of X. Project Xb onto $Col(X) = Xb$

    $HXb = X(X^TX)^{-1}X^TXb = XIb = Xb = \hat{Y}$

  - $\boldsymbol{HX} = \boldsymbol{X}$ $\Rightarrow$ $X \in col(X) \Rightarrow HX = X$

    $X(X^TX)^{-1}X^TX = X$

  - $\boldsymbol{H}(\underbrace{\boldsymbol{Y} - \boldsymbol{Xb}}_{=e}) = \boldsymbol{0}$ $\Rightarrow$ $e \perp col(X) \Rightarrow$ Project e onto $Col(X) = 0$

    ① $HY - HXb = \hat{Y} - Xb = \hat{Y} - \hat{Y} = 0$

    ② $HY - H\hat{Y} = \hat{Y} - \hat{Y} = 0$ since $\cdot\ HY = \hat{Y}$ and $H\hat{Y} = \hat{Y}$

Multiple Linear regression in matrix form

- We write the general linear regression model in the matrix as following

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

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

$$\underset{p\times 1}{\boldsymbol{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = \boldsymbol{(\underset{p\times p}{X^T X})^{-1}(\underset{p\times 1}{X^T Y})}$$

Multiple Linear regression in matrix form

- The sum of squares for the analysis of variance in matrix form are

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

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

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

$$MSR = \frac{SSR}{p-1}$$ p-1 = # of X in MLR

$$MSE = \frac{SSE}{n-p}$$ n-p = n − # of β's

where

$$\underset{n\times n}{\boldsymbol{J}} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} = \mathbb{1}\mathbb{1}^T$$

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

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

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

True or not by examining $\frac{MSR}{MSE}$

F test for regression relation in MLR

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{X_1} + \ldots + \beta_{p-1} \boldsymbol{X_{p-1}} + \boldsymbol{\epsilon}$$

- We set the following hypotheses to test whether there is a regression relation between the response variable Y and the set of X variables $X_1, X_2, \ldots, X_{p-1}$:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

versus

$$H_a : \text{not all } \beta_k (k = 1, 2, \ldots, p-1) \text{ equal zero}$$

- We use the following test statistic

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

$F_{p-1,n-p} = \frac{1}{F_{n-p,p-1}}$

- Reject H_0 at α significance level if

$$F^* > F_{1-\alpha; p-1, n-p}$$

Inference about Regression Parameters

Test of β_k

- The least squares and maximum likelihood estimators in **b** are unbiased

$$E(\boldsymbol{b}) = \beta$$

- The variance-covariance matrix **b**

$$\underset{p\times 1}{\boldsymbol{b}} = \sigma^2\boldsymbol{(X^TX)^{-1}} = \begin{bmatrix} V(b_0) & Cov(b_0, b_1) & \dots & cov(b_0, b_{p-1}) \\ Cov(b_0, b_1) & V(b_1) & \dots & cov(b_1, b_{p-1}) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(b_{p-1}, b_0) & Cov(b_{p-1}, b_2) & \dots & V(b_{p-1}) \end{bmatrix}$$

- The estimated variance-covariance matrix of **b**

$$\underset{p\times p}{s^2\{\boldsymbol{b}\}} = \text{MSE}(\boldsymbol{X^TX})^{-1}$$

Inference about Regression Parameters

Test of β_k

- The null and alternative hypotheses are

$$H_0 : \beta_k = 0 \quad versus \quad H_a : \beta_k \neq 0$$

- The test statistic is

$$t^* = \frac{b_k}{s\{b_k\}}$$

$s^2(b_k) = k^{th}$ element on diagonal of $MSE(X^TX)^{-1}$

- Decision Rule: Reject H_0 at α significance level if

$$|t^*| > t_{1-\alpha/2;n-p}$$

Inference about Regression Parameters

Interval of β_k

- For the normal error MLR model

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t_{n-p}, k = 0, 1, \ldots, p-1$$

- Hence, $100(1-\alpha)\%$ confidence interval is

$$b_k \pm t_{1-\alpha/2;n-p} s\{b_k\}$$

- If g parameters are to be estimated jointly or simultaneously(where $g \leq p$), the confidence limits with family confidence coefficient $1-\alpha$ are

$$b_k \pm Bs\{b_k\},$$

where

$$B = t_{1-\alpha/(2g);n-p}$$

Interval Estimation of $E\{Y_h\}$

- To estimate the mean response at $X_{h1}, \ldots, X_{h(p-1)}$, let us define the vector as

$$\underset{p\times 1}{\boldsymbol{X_h}} = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h(p-1)} \end{bmatrix}$$

- The mean response to be estimated is

$$E\{Y_h\} = \boldsymbol{X_h^T}\beta = \beta_0 + \beta_1 X_{h1} + \beta_2 X_{h2} + \cdots + \beta_{p-1} X_{h(p-1)}$$

- The estimated mean response corresponding to $\boldsymbol{X_h}$

$$\hat{Y}_h = \boldsymbol{X_h^T b}$$

- It is unbiased and its variance is

$$E\{\hat{Y}_h\} = \boldsymbol{X_h^T}\boldsymbol{\beta} = E\{Y_h\}$$

$$Var\{\hat{Y}_h\} = \boldsymbol{X_h^T}\mathbf{Var\{\boldsymbol{b}\}}\boldsymbol{X_h} = \sigma^2\boldsymbol{X_h^T(X^TX)^{-1}X_h}$$

Interval Estimation of $E\{Y_h\}$ (contd.)

- The estimated Variance of $\boldsymbol{\hat{Y}_h}$ in matrix notation is

$$S^2\{\boldsymbol{\hat{Y}_h}\} = \boldsymbol{X_h^T} s^2\{b\} X_h = \text{MSE}\Big(\boldsymbol{X_h^T (X^T X)^{-1} X_h}\Big)$$

- The $1-\alpha$ confidence limits for $E\{Y_h\}$ are

$$\hat{Y}_h \pm t_{1-\alpha/2;n-p} s\{\hat{Y}_h\}$$

Prediction for New observation $Y_{h(new)}$

- The $1-\alpha$ confidence limits for $Y_{h(new)}$ corresponding to $\boldsymbol{X_h}$, the specified values of $\boldsymbol{X}$ variables, are:

$$\hat{Y}_h \pm t_{1-\alpha/2;n-p} s_{pred}$$

where

$$s_{pred} = \text{MSE}(1 + \boldsymbol{X_h^T (X^T X)^{-1} X_h})$$

- Simultaneous or joint Prediction limits
  - Sheffe method:for g new observations at g different levels $\boldsymbol{X_h}$ with family confidence coefficient $1-\alpha$

$$\hat{Y}_h \pm S s_{pred}$$

where

$$S^2 = gF_{1-\alpha;g,n-p}$$

  - Bonferroni simultaneous predication limits for g new observations at g different levels $\boldsymbol{X_h}$ with family confidence coefficient $1-\alpha$ are given by

$$\hat{Y}_h \pm B s_{pred}$$

where

$$B = t_{1-\alpha/(2g);n-p}$$

Coefficient of Multiple Determination

- The coefficient of multiple determinant R^2

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}}$$

- Measures the percentage reduction of total variation in Y associated with $X_1, \ldots, X_{p-1}$
- $0 \leq R^2 \leq 1$.
- Adding more X variable $\Rightarrow R^2$ goes up.
    - SSE can never become larger with more X variables.
    - SSTO is always the same of a given set of responses.

← more X, higher chance that partial of SSTO can be explained
$\Rightarrow$ left over = SSE is ↓.

The adjusted Coefficient of Multiple Determination

- The adjusted coefficient of multiple determination: R_a^2

$$R_a^2 = 1 - \frac{\text{SSE}/(n-p)}{\text{SSTO}/(n-1)} = 1 - \frac{n-1}{n-p}\frac{\text{SSE}}{\text{SSTO}}$$

  - $R_a^2 \leq R^2$ and R_a^2 doen't increase as fast as the R^2 does when more X's are introduced in the model.

- Coefficient of Multiple Correlation: R

$$R = \sqrt{R^2}$$

Data example: Housing price

- Data set contain 546 observations on sales prices of houses sold during July, August and September, 1987 in the city of Windsor, Canada.
- Variables
  - price: sale price of a house
  - lotsize: the lot size of a property in square feet
  - bedrooms: number of bedrooms
  - bathrms: number of bathrooms
  - driveway: dummy, 1 if the house has a driveway

Variable	Obs	Mean	Std. Dev.	Min	Max
price	546	68121.6	26702.67	25000	190000
lotsize	546	5150.266	2168.159	1650	16200
bedrooms	546	2.965201	0.7373879	1	6
bathrms	546	1.285714	0.5021579	1	4
stories	546	1.807692	0.8682025	1	4
driveway	546	0.8589744	0.3483672	0	1

Example: Housing price (contd.)

```r
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
house$driveway = as.factor(house$driveway)
str(house)
```

```
## 'data.frame':    546 obs. of  6 variables:
##  $ price   : num  42000 38500 49500 60500 61000 66000 66000 69000 83800 8850
##  $ lotsize : int  5850 4000 3060 6650 6360 4160 3880 4160 4800 5500 ...
##  $ bedrooms: int  3 2 3 3 2 3 3 3 3 3 ...
##  $ bathrms : int  1 1 1 1 1 1 2 1 1 2 ...
##  $ stories : int  2 1 1 2 1 1 2 3 1 4 ...
##  $ driveway: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

Example: Housing price (contd.)

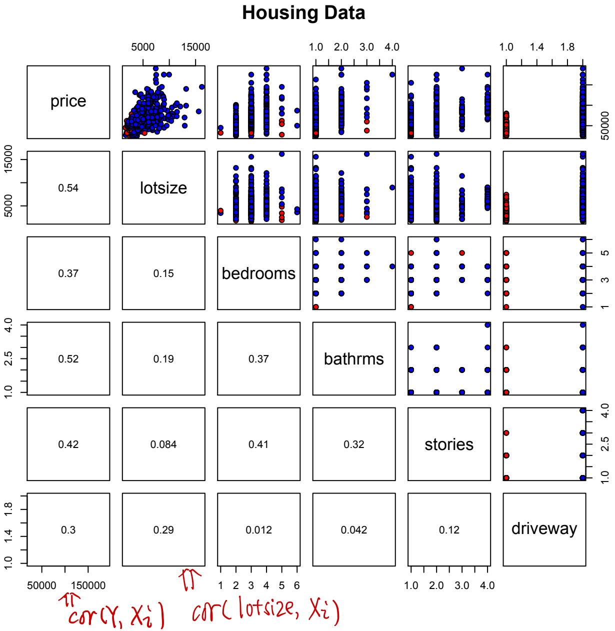
```r
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
house$driveway = as.factor(house$driveway)

pairs(house[1:6], main = "Housing Data ", pch = 21,
      bg = c("red","blue")[unclass(house$driveway)] )

# change the upper pannel with Person correlation coefficient
panel.pearson <- function(x, y, ...) {
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;
  text(horizontal, vertical, format(abs(cor(x,y)), digits=2))
  }

pairs(house[1:6], main = "Housing Data ", pch = 21,
      bg = c("red","blue")[unclass(house$driveway)],
      lower.panel=panel.pearson)
```

Example: Housing price (contd.)



Example: Housing price (contd.)

```
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
house$driveway = as.factor(house$driveway)
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)
model.matrix(fin)[1:20,]
```

Find $X_{n\times p}$ = design matrix

```
##    (Intercept) log(lotsize) bedrooms bathrms stories factor(driveway)1
## 1            1     8.674197        3       1       2                 1
## 2            1     8.294050        2       1       1                 1
## 3            1     8.026170        3       1       1                 1
## 4            1     8.802372        3       1       2                 1
## 5            1     8.757784        2       1       1                 1
## 6            1     8.333270        3       1       1                 1
## 7            1     8.263590        3       2       2                 1
## 8            1     8.333270        3       1       3                 1
## 9            1     8.476371        3       1       1                 1
## 10           1     8.612503        3       2       4                 1
## 11           1     8.881836        3       2       1                 1
## 12           1     8.006368        2       1       1                 0
## 13           1     7.438384        3       1       2                 1
## 14           1     7.965546        3       1       1                 0
## 15           1     8.188689        2       1       1                 1
## 16           1     8.066208        2       1       1                 1
## 17           1     8.101678        3       1       2                 0
## 18           1     8.556414        4       1       3                 1
## 19           1     8.146130        1       1       1                 1
## 20           1     8.290544        2       2       1                 0
```

Example: Housing price (contd.)

```
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)
anova(fin)
```

$F^* = \frac{SSR(X_i)/1}{MSE} \sim F_{1,n-p}$ for: $H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

```
## Analysis of Variance Table
##
## Response: log(price)
##                    Df Sum Sq Mean Sq F value    Pr(>F)
## log(lotsize)        1 25.368 25.3677 419.860 < 2.2e-16 ***
## bedrooms            1  6.163  6.1629 102.002 < 2.2e-16 ***
## bathrms             1  6.459  6.4589 106.900 < 2.2e-16 ***
## stories             1  3.317  3.3168  54.895 4.929e-13 ***
## factor(driveway)    1  1.480  1.4804  24.502 9.935e-07 ***
## Residuals         540 32.627  0.0604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SSE

```
dim(anova(fin))
```

→ 6x5 matrix

```
## [1] 6 5
```

```
sum(anova(fin)[-6,2]) # find SSR
```

take all from 2nd column but leave $a_{6,2}$ out.

```
## [1] 42.78664
```

$= 25.368 + 6.163 + 6.459 + 3.317 + 1.480$, $df(SSR) = 1+1+1+1+1 = 5$

```
sum(anova(fin)[,2])   # find SSTO
```

sum 2nd column

```
## [1] 75.41317
```

$= SSR + SSE = 42.78664 + 32.627$

Example: Housing price (contd.)

```
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)
summary(fin)
```

```
##
## Call:
## lm(formula = log(price) ~ log(lotsize) + bedrooms + bathrms +
##     stories + factor(driveway), data = house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8303 -0.1479  0.0102  0.1465  0.7325
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.88824    0.23138  29.771  < 2e-16 ***
## log(lotsize)       0.40512    0.02877  14.080  < 2e-16 ***
## bedrooms           0.05818    0.01642   3.544 0.000428 ***
## bathrms            0.20616    0.02337   8.822  < 2e-16 ***
## stories            0.09150    0.01369   6.684 5.81e-11 ***
## factor(driveway)1  0.16036    0.03240   4.950 9.94e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2458 on 540 degrees of freedom
## Multiple R-squared:  0.5674, Adjusted R-squared:  0.5634
## F-statistic: 141.6 on 5 and 540 DF,  p-value: < 2.2e-16
```

Handwritten annotations: b_k (over Estimate), $s(b_k)$ (over Std. Error), $t^* = \frac{b_k}{s(b_k)} \sim_{H_0} t_{n-p}$, $H_0: \beta_k = 0 \quad H_a: \beta_k \neq 0$, $= \sqrt{MSE}$ (0.2458), $= n-p$ (540), $R^2 =$ (0.5674), $= R_a^2$ (0.5634)

$$\begin{cases} H_0: \beta_1 = \cdots = \beta_5 = 0 \\ H_1: \text{some } \beta_k \neq 0 \end{cases}$$

$$F^* = \frac{MSR}{MSE} \sim F_{p-1, n-p}$$

```
2*(1-pt(3.544,540))
```

← two tails

```
## [1] 0.0004282715
```

Example: Housing price (contd.)

```r
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)
options(scipen=6) # disable scientific output
vcov(fin)  # estimated var-cov matrix of b= mse*(X'X)^{-1}
```

```
##                    (Intercept)   log(lotsize)       bedrooms
## (Intercept)        0.05353462397 -0.00648606510 -0.00017887714
## log(lotsize)      -0.00648606510  0.00082787396 -0.00004753673
## bedrooms          -0.00017887714 -0.00004753673  0.00026946280   = s^2(b_3)
## bathrms            0.00056378230 -0.00010326530 -0.00009984940
## stories           -0.00006901554  0.00001006112 -0.00007493807
## factor(driveway)1  0.00165482470 -0.00031165004  0.00005376683
##                        bathrms        stories factor(driveway)1
## (Intercept)        0.00056378230 -0.00006901554     0.00165482470
## log(lotsize)      -0.00010326530  0.00001006112    -0.00031165004
## bedrooms          -0.00009984940 -0.00007493807     0.00005376683
## bathrms            0.00054606398 -0.00006376076     0.00002302511
## stories           -0.00006376076  0.00018739631    -0.00005906873
## factor(driveway)1  0.00002302511 -0.00005906873     0.00104956084   = s^2(b_5)
```

```r
sqrt(diag(vcov(fin)))   => s(b_k)
```

```
##       (Intercept)       log(lotsize)          bedrooms            bathrms
##        0.23137550 = s(b_0)   0.02877280 = s(b_1)   0.01641532 = s(b_2)   0.02336801 = s(b_3)
##           stories factor(driveway)1
##        0.01368928 = s(b_4)   0.03239693 = s(b_5)
```

Example: Housing price (contd.)

```r
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
house$driveway = as.factor(house$driveway)
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)
confint(fin)
```

```
##                          2.5 %     97.5 %
## (Intercept)         6.43372903 7.34274174   → β0
## log(lotsize)        0.34859876 0.46163942   → β1
## bedrooms            0.02593479 0.09042622   → β2
## bathrms             0.16026031 0.25206700   → β3
## stories             0.06460618 0.11838770   → β4
## factor(driveway)1   0.09672481 0.22400370   → β5
```

Example: Housing price (contd.)

```r
house = read.table("/Users/Wei/Downloads/housing.txt",sep="",header=T)
house$driveway = as.factor(house$driveway)
fin =lm(log(price)~log(lotsize)+bedrooms+bathrms+stories+factor(driveway),data=house)

newX = list(lotsize=5150,bedrooms=3,stories=2,bathrms=2,driveway=1)
predict(fin,newdata=newX,interval="confidence") # same result using predict.lm(...)
```

```
##        fit     lwr      upr
## 1 11.28091 11.2425 11.31933
```

```r
predict(fin,newdata=newX,interval="predict")
```

```
##        fit      lwr      upr
## 1 11.28091 10.79654 11.76529
```

```r
class(newX)
```

```
## [1] "list"
```

```r
class(house)
```

```
## [1] "data.frame"
```

Chapter 7: Multiple Regression II

Motivation for the extra sum of squares

- F test for all partial effect is zero

$$H_0 : \beta_1 = \ldots = \beta_{p-1} = 0, F^* = MSR/MSE$$

- t test for one partial effect is zero

$$H_0 : \beta_k = 0, t^* = b_k/s\{b_k\}$$

- Question: whether certain X variables can be dropped from model

$$H_0 : \text{some } \beta_k = 0$$

eg: $H_0: \beta_1 = \beta_3 = \beta_6 = 0$ H_a: at least one of $(\beta_1, \beta_3, \beta_6)$ is $\neq 0$

Data Example

A study of the relation of amount of body fat

- A sample of 20 healthy females: 25-35 years old
- Y: body fat
- X_1: triceps skinfold thickness
- X_2: thigh circumference
- X_3: midarm circumference

Want to study how some of all these predictor variables can be used to provide reliable estimate of body fat.

Data Example (contd.)

```
body=read.table("/Users/Wei/TA/Teaching/0-STA302-2016F/Week10-Nov14/bodyfat.txt",header=T)
m0 = lm(Y~1,data=body); anova(m0)
```

no X in model → SSTO = SSE

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 19 495.39  26.073
```

SSTO = SSE = 495.39

```
SSTO=anova(m0)[1,2] # SSTO=SSE=495.39, SSR=0 since no X in model

m1 =lm(Y~X1,data=body); anova(m1)
```

SSTO = 495.39

SSR(X_1) = 352.27

SSE(X_1) = 143.12

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value      Pr(>F)
## X1         1 352.27  352.27  44.305 0.000003024 ***
## Residuals 18 143.12    7.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
SSR1=anova(m1)[1,2] # SSR(X1)=352.27,SSE(X1)=143.12
```

Data Example (contd.)

```r
body=read.table("/Users/Wei/TA/Teaching/0-STA302-2016F/Week10-Nov14/bodyfat.txt",header=T)
m2 = lm(Y~X1+X2,data=body); anova(m2)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value      Pr(>F)
## X1         1 352.27  352.27 54.4661 0.000001075 ***
## X2         1  33.17   33.17  5.1284      0.0369 *
## Residuals 17 109.95    6.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$SSR(X_1)$, $SSR(X_2|X_1)$

$SSTO = 495.39$

$SSR(X_1) = 352.27$

$SSE(X_1)$

$SSR(X_2|X_1) = 33.17$

$SSE(X_1,X_2) = 109.95$

$\Rightarrow SSTO = SSR(X_1) + \underbrace{SSR(X_2|X_1) + SSE(X_1,X_2)}_{SSE(X_1)}$

$SSE(X_1) = SSR(X_2|X_1) + SSE(X_1,X_2)$

```r
SSR2=sum(anova(m2)[-3,2]) # SSR(X1,X2)=385.44, SSE(X1,X2)=109.95
m3 =lm(Y~X1+X2+X3,data=body); anova(m3)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value      Pr(>F)
## X1         1 352.27  352.27 57.2768 0.000001131 ***
## X2         1  33.17   33.17  5.3931     0.03373 *
## X3         1  11.55   11.55  1.8773     0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$SSR(X_3|X_1,X_2)$

$SSR(X_1)$

$SSE(X_1) \Rightarrow SSR(X_2|X_1)$, $SSE(X_1,X_2) \Rightarrow SSR(X_3|X_1,X_2)$, $SSE(X_1,X_2,X_3)$

```r
SSR3 =sum(anova(m3)[-4,2]) # SSR(X1,X2,X3)=396.98, SSE(X1,X2,X3)=98.40
```

Extra sum of squares

Notation:

- Assume X_1 is in the model
  - SSR(X_1): the regression sum of squares.
  - SSE(X_1): the error sum of squares.
- SSR($X_2|X_1$): the extra sum of squares. This measures the marginal effect of adding another variable to the regression model when X_1 is already in the model.
- Data of body example:

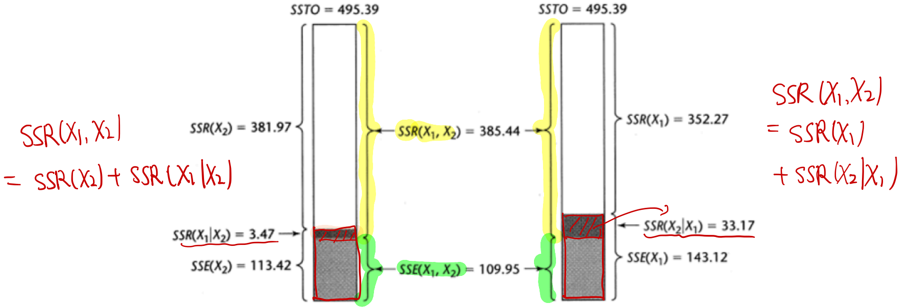
$$\begin{aligned} SSR(X_2|X_1) &= SSE(X_1) - SSE(X_1, X_2) = SSE(m1) - SSE(m2) \\ &= 143.12 - 109.95 = 33.17 \\ &= SSR(X_1, X_2) - SSR(X_1) = SSR(m2) - SSR(m1) \\ &= (352.27 + 33.17) - (353.27) \end{aligned}$$

Extra sum of squares

- SSR($X_2|X_1$)
  - Reflects the additional or extra reduction in the error sum of squares (SSE) associated with X_2, given that X_1 is already in the model.
  - the marginal increase in the regression sum of squares (SSR)
- We observed that
  - the marginal reduction in the SSE =the marginal increase in SSR.
  - SSTO=SSR+SSE
    - SSTO: measure the variability of Y which does not depend on the regression model fitted.
    - Any reduction in SSE implies an identical increase in SSR

Extra sum of squares

- Body fat example



$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17$$
$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1) = 385.44 - 352.27 = 33.17$$

Extra sum of squares

- Body fat example

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17$$
$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1) = 385.44 - 352.27 = 33.17$$

An extra sum of squares: adding X_3

$$SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = 109.95 - 98.41 = 11.54$$
$$SSR(X_3|X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2) = 396.98 - 385.44 = 11.54$$

An extra sum of squares: adding X_2, X_3

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) = 143.21 - 98.41 = 44.71$$
$$SSR(X_2, X_3|X_1) = SSR(X_1, X_2, X_3) - SSR(X_1) = 396.98 - 352.27 = 44.71$$

Basic idea: extra sums of squares

$$SSTO = SSR + SSE$$

- An extra sum of squares measures the marginal decrease in the error sum of squares when one or several predictor variables are added to the regression model, given that other variables are already in the model.
- Equivalently, one can view the extra sum of squares as measuring the marginal increase in the regression sum of squares.
- Extra: SSE goes down, SSR goes up.

Definition: Extra Sums of Squares of two/three variables

Two variables

- If X_1 is the extra variable

$$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2) = SSR(X_1, X_2) - SSR(X_2)$$

- If X_2 is the extra variable

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = SSR(X_1, X_2) - SSR(X_1)$$

Three or more variables

- If X_3 is the extra variable

$$SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$$

- If X_2, X_3 are the extra variables

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_1, X_2, X_3) - SSR(X_1)$$

Decomposition of SSR into Extra Sums of Squares

Consider two X variables

$$SSTO = SSR(X_1) + SSE(X_1) = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2)$$

$$SSTO = SSR(X_1, X_2) + SSE(X_1, X_2)$$

$$\Rightarrow SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$$

Decomposition of SSR into Extra Sums of Squares

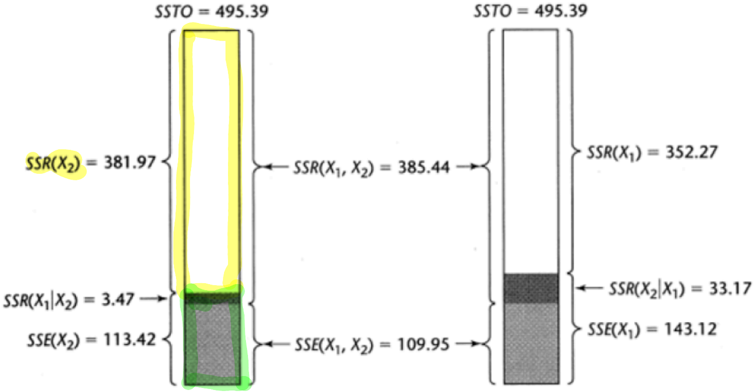
Decomposition: $SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$

- $SSR(X_1)$: measures the contribution by including X_1 alone in the model.
- $SSR(X_2|X_1)$: measures the additional contribution when X_2 is included, given that X_1 is already in the model.
    - The order of the X variables is arbitrary

$$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)$$
$$= SSR(X_1) + SSR(X_2|X_1)$$

Decomposition of SSR into Extra Sums of Squares

- Revisit: Body fat example



Decomposition of SSR into Extra Sums of Squares

- When the regression model contains three X variables, X_1, X_2, X_3.

$$
\begin{aligned}
SSR(X_1, X_2, X_3) &= SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) : 1 + 2 + 3 \\
&= SSR(X_2) + SSR(X_3|X_2) + SSR(X_1|X_2, X_3) : 2 + 3 + 1 \\
&= SSR(X_3) + SSR(X_1|X_3) + SSR(X_2|X_1, X_3) : 3 + 1 + 2 \\
&= SSR(X_1) + SSR(X_2, X_3|X_1) : 1 + (2 + 3)
\end{aligned}
$$

- The number of possible decompositions become vast as the number of X variables in the regression model increases.

Table: Decomposition of SSR into Extra Sums of Squares

ANOVA Table

Source of Variance	SS	df	MS
Regression	$SSR(X_1, X_2, X_3)$	3	$MSR(X_1, X_2, X_3)$
X_1	$SSR(X_1)$	1	$MSR(X_1)$
$X_2 \mid X_1$	$SSR(X_2 \mid X_1)$	1	$MSR(X_2 \mid X_1)$
$X_3 \mid X_1, X_2$	$SSR(X_3 \mid X_1, X_2)$	1	$MSR(X_3 \mid X_1, X_2)$
Error	$SSE(X_1, X_2, X_3)$	n-4	$MSR(X_1, X_2, X_3)$
Total	SSTO	n-1	

- Each extra sum of squares involving
  - a single extra X variable has associated with it one d.f.
  - two extra X variables have two d.f.
- Mean squares

$$MSR(X_2|X_1) = \frac{SSR(X_2|X_1)}{1}$$

$$MSR(X_2, X_3|X_1) = \frac{SSR(X_2, X_3|X_1)}{2}$$

Example of body fat data: Decomposition of SSR

```r
body=read.table("/Users/Wei/TA/Teaching/0-STA302-2016F/Week10-Nov14/bodyfat.txt",header=T)
m3 =lm(Y~X1+X2+X3,data=body)
anova(m3)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value      Pr(>F)
## X1 SSR(X1)  1 352.27  352.27 57.2768 0.000001131 ***
## X2 SSR(X2|X1)  1  33.17   33.17  5.3931     0.03373 *
## X3 SSR(X3|X1,X2)  1  11.55   11.55  1.8773     0.18956
## Residuals 16  98.40    6.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Upcoming topics

- Ch7.2: Use of Extra Sum of Squares in Tests for Regression coefficients
- Ch7.3: Summary of Tests Concerning Regression Coefficients
- Ch7.5: Standardized Multiple Regression Model
- Ch7.6: Multicollinearity and Its Effects

Practice problems after lectures

- Keep trying problems in Ch6:
  - 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.
- Try all problems that we have covered today in Ch7:
  - 7.2, 7.3, 7.7, 7.8, 7.12, 7.15, 7.20, 7.22, 7.23, 7.27, 7.31.