STAT 3008 Applied Regression Analysis
Tutorial 7.

XU Yongze

OCT 21&22&24

Contents

1 Review Questions **2**

1.1 Confidence ellipse (Ex5 Q3) 2

1.2 Find SSE for different models (Ex6 Q2) 4

2 Understanding parameter estimates **6**

3 Aliased and Multicollinearity **7**

4 Overfitted Model **8**

5 Underfitted Model (having lurking variable) **9**

5.1 Property of estimates 9

5.2 Relationship b/w mean functions of true and mis-specified model 9

6 Drawing conclusions **10**

7 More on R square **10**

1 Review Questions

1.1 Confidence ellipse (Ex5 Q3)

Confidence interval - confidence region for a single parameter;
Confidence ellipse - joint confidence region for two parameters.

♣ Question:

Let $n = 100$, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)' = (7, 1)'$, $\hat{\sigma} = 2$,

$$X'X = \begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}.$$

Find an inequality representing the 95% confidence ellipse for β. Stretch a graph for the ellipse. (Hints: Try to find some points that lie on the ellipse.)

♣ Answer:
$(1 - \alpha)$ confidence ellipse:

$$\frac{(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \le F(\alpha, p+1, n-p-1)$$

Here $p + 1 = 2, \alpha = 0.05$, so the 95% confidence region of β is:

$$\frac{(\beta-\widehat{\beta})^T(X^TX)(\beta-\widehat{\beta})}{2\widehat{\sigma^2}} \le F(0.05, 2, 100-2)$$

$$\frac{\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}-\begin{pmatrix}7\\ 1\end{pmatrix}\right)^T\begin{pmatrix}3 & 2\\ 2 & 5\end{pmatrix}\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}-\begin{pmatrix}7\\ 1\end{pmatrix}\right)}{2\times 2^2} \le 3.0892$$

$$(\beta_1-7 \quad \beta_2-1)\begin{pmatrix}3 & 2\\ 2 & 5\end{pmatrix}\begin{pmatrix}\beta_1-7\\ \beta_2-1\end{pmatrix} \le 3.0892\times 8$$

$$(3\beta_1+2\beta_2-23 \quad 2\beta_1+5\beta_2-19)\begin{pmatrix}\beta_1-7\\ \beta_2-1\end{pmatrix} \le 3.0892\times 8$$

$$3\beta_1^2+5\beta_2^2+4\beta_1\beta_2-46\beta_1-38\beta_2+180 \le 3.0892\times 8$$

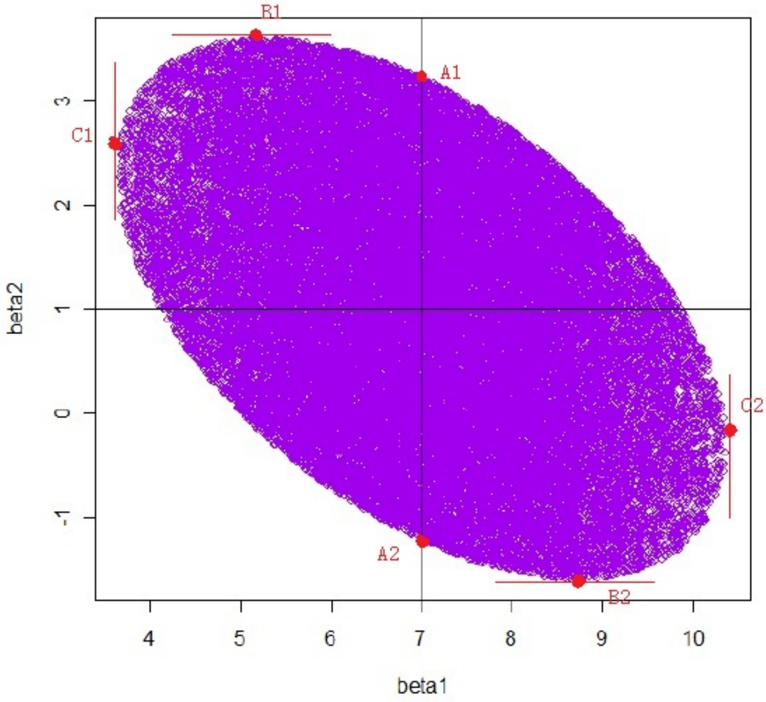
$$3\beta_1^2+5\beta_2^2+4\beta_1\beta_2-46\beta_1-38\beta_2+155.2864 \le 0$$

♣ Trick for calculation:

$$(\beta_1-7 \quad \beta_2-1)\begin{pmatrix}3 & 2\\ 2 & 5\end{pmatrix}\begin{pmatrix}\beta_1-7\\ \beta_2-1\end{pmatrix}$$

$\Rightarrow 3(\beta_1 - 7)^2 + 4(\beta_1 - 7)(\beta_2 - 1) + 5(\beta_2 - 1)^2$
$\Rightarrow 3\beta_1^2 + 5\beta_2^2 + 4\beta_1\beta_2 - (14 \times 3 + 4)\beta_1 - (4 \times 7 + 5 \times 2)\beta_2 + 3 \times 49 + 4 \times 7 + 5$
$\Rightarrow 3\beta_1^2 + 5\beta_2^2 + 4\beta_1\beta_2 - 46\beta_1 - 38\beta_2 + 180$

♣ How to find some special points and stretch a graph for the ellipse:



- Find points A1 and A2: solve

$$3(\beta_1 - 7)^2 + 4(\beta_1 - 7)(\beta_2 - 1) + 5(\beta_2 - 1)^2 - 3.0892 \times 8 = 0$$

  when $\beta_1 = 7$. Equivalently, to solve $5(\beta_2 - 1)^2 - 24.7136 = 0$. So $\beta_2 = \pm\sqrt{\frac{25.7136}{5}} + 1$. Hence $A1 = (7, 3.223223)$, $A2 = (7, -1.223223)$.

- Find points B1 and B2: to find the β_2's so that the equation

$$3\beta_1^2 + 5\beta_2^2 + 4\beta_1\beta_2 - 46\beta_1 - 38\beta_2 + 155.2864 = 0$$

  for β_1 has only one solution. The previous equation is equivalent to

$$3\beta_1^2 + (4\beta_2 - 46)\beta_1 + (5\beta_2^2 - 38\beta_2 + 155.2864) = 0.$$

  When $\triangle = (4\beta_2^2 - 46)^2 - 12(5\beta_2^2 - 38\beta 2 + 155.2864) = -44\beta_2^2 + 88\beta_2 + 252.5632 = 0$, i.e.,$\beta_2 = \frac{-88 \pm \sqrt{88^2 + 4 \times 44 \times 252.5632}}{-88} = 3.596165$ or -1.596165, the quadratic equation has only one solution $\beta_1 = -\frac{4\beta_2 - 46}{6} = 5.269223$ or 8.730777, respectively. Hence $B1 = (5.269223, 3.596165)$, $B2 = (8.730777, -1.5961665)$.

- Similarly, find pints C1 and C2.

1.2 Find SSE for different models (Ex6 Q2)

Let $Y = (29, 34, 19, 41, 36, 36, 24, 10)'$, $X1 = (1, 1, 3, 3, 5, 5, 7, 7)'$, $X2 = (10, 7, 6, 3, 2, -1, -2, -5)'$, $X3 = (2, 6, 0, 9, 7, 9, 4, 2)'$.

i Compare the regression model $Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + e$ and $Y = \beta_0 + \beta_2 X2 + e$ using a F-test. Which model will you use?

ii Test whether $\beta_1 = \beta_2$, i.e. both variables produce the same effect. (Hint: rewrite the model as

$$Y = \beta_0 + \beta_1(X1 + X2) + \beta_2 X2 + \beta_3 X3 + e \quad (1)$$

and do a t-test.)

ii Do a F-test to compare between model (1) and

$$Y = \beta_0 + \beta_1(X1 + X2) + \beta_3 X3 + e \quad (2)$$

Show numerically the F-statistic is equal to the square of the t-statistic.

Ans:

(i)

$H_0 : E(Y \mid X) = \beta_0 + \beta_2 X_2$

$H_1 : E(Y \mid X) = \beta_0 + \beta_2 X_2 E(Y \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Under H_1, $X = (1 \quad X_1 \quad X_2 \quad X_3) = \begin{pmatrix} 1 & 1 & 10 & 2 \\ 1 & 1 & 7 & 6 \\ 1 & 3 & 6 & 0 \\ 1 & 3 & 3 & 9 \\ 1 & 5 & 2 & 7 \\ 1 & 5 & -1 & 9 \\ 1 & 7 & -2 & 4 \\ 1 & 7 & -5 & 2 \end{pmatrix}$

$SSE_{H_1} = Y^T(I - H)Y = Y^T(I - X(X^T X)^{-1} X^T)Y = 2.2742$

$df_{H_1} = n - (p+1) = 8 - 4 = 4$

Under H_0, $X = \begin{pmatrix} 1 & X_2 \end{pmatrix} = \begin{pmatrix} 1 & 10 \\ 1 & 7 \\ 1 & 6 \\ 1 & 3 \\ 1 & 2 \\ 1 & -1 \\ 1 & -2 \\ 1 & -5 \end{pmatrix}$

$$SSE_{H_0} = Y^T(I-H)Y = Y^T(I - X(X^TX)^{-1}X^T)Y = 656.1994$$

$$df_{H_0} = n-(p+1) = 8-2 = 6$$

$$F_{(0.05,2,4)} = 6.9443$$

$$Obs.T.S. = \frac{\dfrac{SSE_{H_0} - SSE_{H_1}}{df_{H_0} - df_{H_1}}}{\dfrac{SSE_{H_1}}{df_{H_1}}} = 575.0883 > 6.9443$$

Therefore, H_0 is rejected at 0.05 significance level.

(ii) According to equation (1), $H_0 : \beta_1 = \beta_2 \Leftrightarrow H_0 : \beta_2 = 0$. To do t-test, first look at the 'full model'(alternative model) to find out $\hat{\beta}_i, \hat{\sigma}$ and $se(\hat{\beta}_i)$.

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

$$X = \begin{pmatrix} 1 & X_1 + X_2 & X_2 & X_3 \end{pmatrix} = \begin{pmatrix} 1 & 11 & 10 & 2 \\ 1 & 18 & 7 & 6 \\ 1 & 9 & 6 & 0 \\ 1 & 6 & 3 & 9 \\ 1 & 7 & 2 & 7 \\ 1 & 4 & -1 & 9 \\ 1 & 5 & -2 & 4 \\ 1 & 2 & -5 & 2 \end{pmatrix}$$

$$SSE = Y^T(I-H)Y = Y^T(I - X(X^TX)^{-1}X^T)Y = 2.2742$$

$$\widehat{\sigma^2} = \frac{SSE}{df} = \frac{2.2742}{8-4} = 0.5685$$

$$\widehat{Var}(\widehat{\beta} \mid X) = \widehat{\sigma^2}(X^T X)^{-1} = \begin{pmatrix} 6.2637 & -1.0494 & 0.5512 & -0.1537 \\ -1.0494 & 0.1842 & -0.0990 & 0.0204 \\ 0.5512 & -0.0990 & 0.0565 & -0.0101 \\ -0.1537 & 0.0204 & -0.0101 & 0.0095 \end{pmatrix}$$

$$se(\widehat{\beta}_2) = \sqrt{\widehat{Var}(\widehat{\beta}_2 \mid X)} = \sqrt{0.0565} = 0.2376$$

$$\widehat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} -3.1673 \\ 2.6774 \\ -0.3620 \\ 3.1373 \end{pmatrix}$$

$$t_{(0.025,4)} = 2.7764$$

$$obs.T.S. = \frac{\widehat{\beta}_2}{se(\widehat{\beta}_2)} = \frac{-0.3620}{0.2376} = -1.5233$$

$$|obs.T.S.| < 2.7764$$

Therefore, H_0 is not rejected at 5% significance level.

(iii)Refer to 'Ex6 Solutions'.

2 Understanding parameter estimates

$$\mathrm{E}(Y|X) = \beta_0 X_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$\mathrm{E}(Fuel|X) = 154.19 - 4.23\, Tax + 0.47\, Dlic - 6.14\, Income + 18.54 \log(Miles)$$

- Unit of βs:
  - unit of y/unit of x. (e.g. gallon/$1000 for -6.14)
- Unit of σ^2:
  - (unit of y)^2
- Meaning of β_i:
  - Rate of change of y on x_i, after adjusting for other variables
  - e.g.
    - Fuel decreased by 6.14 gallon when Income increased by $1000
    - Fuel increased by 18.54 gallon when Miles is doubled ($\log_2 2x = 1 + \log(x)$)
- Meaning of σ^2:
  - Variability that can't be explained by the regression line.

3 Aliased and Multicollinearity

3.1 Aliased:

Some predictors are linear combinations of other predictors
$\Rightarrow det(X'X) = 0 \Rightarrow X'X$ noninvertible

3.2 Multicollinearity:

Some predictors are highly correlated.
$\Rightarrow det(X'X) \approx 0 \Rightarrow (X'X)^{-1} = \frac{adj(X'X)}{det(X'X)}$ unstable

Perfect multicollinearity (correlation = 1) is equivalent to Aliased

♣ Consequence:
>$\hat{\beta}$ is unstable
>Var$(\hat{\beta})$ can be very large

♣ Explain intuitively:

1. Soma=1.59-0.116 WT2+0.056 WT9+0.048 WT18
2. Soma=1.59-(0.116+b) WT2+(0.056+b) WT9+0.048 WT18 - b DW9

- Any b in model 2 is essentially model 1!

So the estimates of coefficients are unstable and their variance can be very large.

4 Overfitted Model

- What happens when a bigger model is fit to the data from a smaller model?
  - Data:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+e$
  - Model:
    - $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+e$

$\boxed{Answer: \ \hat{\beta}_3 \approx 0, \hat{\beta}_4 \approx 0}$

Why this answer? In general,

(Over-fitted model)

True Model: $\quad \mathbf{Y} = \mathbf{X}_1\mathbf{b}_1 + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$

Mis-specified Model: $\quad \mathbf{Y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon} \quad \text{where} \quad \mathbf{X} = (\mathbf{X}_1 \ \ \mathbf{X}_2), \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}$

(a) What is $E(\widehat{\mathbf{b}})$?

(a) $E(\hat{\beta}) = E\left((X'X)^{-1}X'y\right) = (X'X)^{-1}X'E(y)$

$= (X'X)^{-1}X'(X_1\beta_1) = (X'X)^{-1}X'X\begin{pmatrix}\beta_1 \\ 0\end{pmatrix}$

$= \begin{pmatrix}\beta_1 \\ 0\end{pmatrix} \quad \Rightarrow \quad E(\hat{\beta}_1) = \beta_1$ and $E(\hat{\beta}_2) = 0$.

5 Underfitted Model (having lurking variable)

(Underfitted Model)

True Model: $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\gamma + \varepsilon, \quad E(\varepsilon) = \mathbf{0}, \quad |\mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$

Mis-specified Model: $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon, \quad E(\varepsilon) = \mathbf{0}, \quad \mathrm{Var}(\varepsilon) = \sigma^2\mathbf{I}$

Let the least squares estimate of $\mathbf{b}$ be $\widehat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

5.1 Property of estimates

(a) Prove that $\widehat{\mathbf{b}}$ is a biased estimator of $\mathbf{b}$ and find the bias.

(a) $$E(\hat{\beta}) = E[(X'X)^{-1}X'Y] = (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'(X\beta + Z\gamma) = \beta + C\gamma$$

when $C = (X'X)^{-1}X'Z$.

Thus, $\hat{\beta}$ is a biased estimate of β with bias $C\gamma$.

5.2 Relationship b/w mean functions of true and mis-specified model

♣ True model:

$$E(Y|X = x, L = l) = \beta_0 + \beta_1 x + \delta l$$

♣ Wrong model:

$$\begin{aligned} E(Y|X = x) &= E[E(Y|X = x, L = l)|X = x] \\ &= E(\beta_0 + \beta_1 x + \delta l|X = x) \\ &= \beta_0 + \beta_1 x + \delta E(L|X = x) \end{aligned}$$

If $E(L|X = x) = \gamma_0 + \gamma_1 x, E(Y|X = x) = (\beta_0 + \delta\gamma_0) + (\beta_1 + \delta\gamma_0)x$.

♣ Example (Ex7 Q3)

Suppose we fit a regression with the true mean function

$$E(Y|X_1 = x_1, X_2 = x_2) = 3 + 4x_1 + 2x_2$$

Provide conditions under which the mean function for $E(Y|X_1 = x_1)$ is linear but has a negative coefficient for x_1.

Ans:

$E(Y \mid X_1 = x_1, X_2 = x_2) = 3 + 4x_1 + 2x_2$

$E(Y \mid X_1 = x_1) = 3 + 4x_1 + 2E(X_2 \mid X_1 = x_1)$

Therefore, the above expectation is linear if $E(X_2 \mid X_1 = x_1) = a + bx_1$.

$E(Y \mid X_1 = x_1) = 3 + 4x_1 + 2(a + bx_1)$

$= 3 + 2a + (4 + 2b)x_1$

And the coefficient of x_1 is negative if $4 + 2b < 0 \rightarrow b < -2$.

Conclusion: estimated parameter may not tell the true effect of a variable.

6 Drawing conclusions

♣ Observational $\leftrightarrow$ experimental

$\Leftrightarrow$ cannot control the value of predictors $\leftrightarrow$ can control

♣ Drawing conclusions:

Observational studies - association

Experiments - causal relationship.

7 More on R square

- R^2 tends to be large if the X are dispersed, so try best to get a sample with larger range of X
- R^2 is useful to measure goodness of fit $\Leftrightarrow$ scatter plot looks like a sample from a bivariate normal distribution (elliptical shapled), i.e. without leverage point, curve, lurking variable etc.