ជំពូកទី៦ គំរូព្យាករណ៍ច្រើនអថេរ

Multiple Linear Regression

រៀបរៀងនិងបង្រៀនដោយបណ្ឌិត ង៉ាន់ ស៊ុនដេត
សាស្ត្រាចារ្យនៃសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រនិងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- មេរៀនមុនយើងបានសិក្សាគំរូព្យាករណ៍ដែលមានមួយអថេរ។ នៅក្នុងសេដ្ឋកិច្ចកម្រមានបញ្ហាតែមួយដែលជាប់ពាក់ព័ន្ធគ្នាគឺវាមានកត្តាច្រើនជាងមួយដែលមានការពាក់ព័ន្ធគ្នាជាបណ្តាញសេដ្ឋកិច្ចមួយ។
- តទៅនេះ យើងនឹងសិក្សាគំរូវិសមីការវិភាគសេដ្ឋកិច្ចច្រើនអថេរ៖

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i, i = 1,2,\ldots,n, where\ u_i: error\ term$$

Y_i: Dependent variable, X_{ki}: Independent Variable, $k = 1,2,\ldots k$

β_k: Parameter of Model, $k = 0,1,2,\ldots k$; β_0: Intercept or Constant

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- ឧទាហរណ៍៖ គំរូប្រាក់ឈ្នួល(Wage)។ នៅក្នុងសេដ្ឋកិច្ចការងារ ការផ្តល់ប្រាក់ឈ្នួលសមរម្យមានការជាប់ពាក់ព័ន្ធជាមួយកត្តាច្រើនយ៉ាង ដូចជា ចំនួនឆ្នាំនៃការសិក្សាដែលបានរៀនចប់(edu) បទពិសោធន៍ការងារ(exp) ភេទ(sex) អាយុ(age) ធ្វើការថែមម៉ោង(ot)ជាដើម។
- ដូច្នេះគំរូវិភាគអាចកំណត់បានជាសមីការលីនេអ៊ែរ

$$wage = f(edu, exp, age, ot)$$

$$log(wage)_i = \beta_0 + \beta_1(edu)_i + \beta_2(exp)_i + \beta_3(sex)_i + \beta_4(ot)_i + u_i, i = 1,2, \dots, n$$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- តទៅ យើងនឹងសរសេរសមីការវិភាគនេះជារាងម៉ាទ្រីស៖

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_k X_{k1} + u_1$$

$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_k X_{k2} + u_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \cdots + \beta_k X_{kn} + u_n$$

- ដូច្នេះ យើងបានម៉ាទ្រីសអាចសរសេរ $Y_{(n,1)} = X_{(n,k+1)}\beta_{(k+1,1)} + U_{(n,1)}$.

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- តាងកន្សោមម៉ាទ្រីសនីមួយៗដោយ៖

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix}_{(n,1)} ,X = \begin{pmatrix} 1 & X_{11} \cdots & X_{k1} \\ 1 & X_{12} \cdots & X_{k2} \\ \vdots & \ddots & \vdots \\ 1 & X_{1n} \cdots & X_{kn} \end{pmatrix}_{(n,k+1)} ,\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_k \end{pmatrix}_{(k+1,1)} ,e = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix}_{(n,1)}$$

- យើងបានសមីការម៉ាទ្រីស $Y = X\beta + e$ (1) គឺជាសមីការម៉ាទ្រីសដែលមានចំនួន k អថេរពន្យល់ និង n ចំនួនអង្កេត។

គំរូព្យាករណ៍ច្រើនអេថេរ
Multiple Linear Regression

- ដើម្បីគណនាមេគុណប៉ារ៉ាម៉ែត្រ $\beta_0, \beta_1, \dots, \beta_k$ យើងអនុវត្តវិធីអប្បបរមាផលបូកការេតម្លៃតូច(OLS[Ordinary Least Square]) $min \sum e_i^2 = min(e')(e) = \min(Y - X\beta)'(Y - X\beta) = minS$ ដែល e'ជាម៉ាទ្រីសត្រង់ស្ប៉ូនៃម៉ាទ្រសផលសង $e = Y - X\beta$។
- ពន្លាតកន្សោម $S = (Y - X\beta)^2 = Y'Y - 2X'Y\beta + X'X\beta^2$ លក្ខខណ្ឌលំដាប់១នៃកម្រិតអប្បបរមាធៀប β ៖ $\frac{\partial S}{\partial \beta} = -2X'Y + 2X'X\beta = 0 \Rightarrow X'X\beta = X'Y$ ដែល X' ជាម៉ាទ្រីសត្រង់ស្ប៉ូនៃម៉ាទ្រីស X សំណុំទិន្នន័យអង្កេត។
- បើ $\det(X'X) \neq 0 \Rightarrow inverse(X'X) = (X'X)^{-1}$ យើងបានមេគុណ $\boldsymbol{\hat{\beta} = (X'X)^{-1}(X'Y)}$ ។

គំរូពយាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- កន្សោម $S=(Y-X\beta)^2=(Y'-\beta'X')(Y-X\beta)=Y'Y-Y'X\beta-\beta'X'\mathrm{Y}+\beta'X'X\beta=Y'Y-2\beta'X'\mathrm{Y}+\beta'X'X\beta$ គឺជាផលគុណនៃពីរកន្សោមម៉ាទ្រីសដែលមានលំដាប់មិនស្មើគ្នា។
- ម៉ាទ្រីសការេ $X'X$ ជាម៉ាទ្រីសកំណត់ពាក់កណ្ដាលវិជ្ជមានឬអាចជាម៉ាទ្រីសុីមេទ្រី។
- យើងបានសមីការជារាងម៉ាទ្រីសជាថ្មី៖ $Y=\hat{\beta}_0+\hat{\beta}_1X_{1i}+\hat{\beta}_2X_{2i}+\cdots+\hat{\beta}_kX_{ki}+e_i$.

  Intercept: $\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}_1-\hat{\beta}_2\bar{X}_2-\cdots-\hat{\beta}_k\bar{X}_k$
- បើយកអថេរថ្មី $x_i=X_i-\bar{X}\Rightarrow\sum(X_i-\bar{X})=\sum x_i=0\Rightarrow\bar{x}=0$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

i. ការសន្មតលើភាពចៃដន្យនៃទិន្នន័យ(stochastic hypothesizes)។

H1: តម្លៃនៃអថេរ X_{ki} ដែលបានមកពីការអង្កេតជាចំនួនថេរគ្មានលំអៀង។

H2: $E(e_i) = 0$ តម្លៃសង្ឃឹមគណិតនៃតួលំអៀងស្មើ 0។

H3: $E\left(e_i^2\right) = \sigma^2$ វ៉ារ្យង់នៃតម្លៃតួលំអៀងជាចំនួនថេរ$(\forall i)$(homoscedasticity)។

H4: $E(e_i e_{i'}) = 0, if\ i \neq i'$ គ្រប់តួលំអៀងមិនទាក់ទងគ្នា(ឯករាជ្យពីគ្នា)។

H5: $Cov(X_{ki}, e_i) = 0$ តម្លៃតួលំអៀងគឺឯករាជ្យពីអថេរពន្យល់។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ii. វចនាសម្ព័ន្ធនៃសម្មតិកម្ម

H6: អវត្តមានកូលីនេអ៊ែររវាងអថេរពន្យល់នានានាំឲ្យម៉ាទ្រីសផលគុណ $(X'X)$ មានតម្លៃជាទៀងទាត់ដែលនាំឲ្យមានម៉ាទ្រីសច្រាស $(X'X)^{-1}$។

H7: កន្សោមផលធៀប $\frac{(X'X)}{n}$ ខិតទៅរកតម្លៃកំណត់មិនទៀងទាត់។

H8: $n > k + 1$ បរិមាណចំនួនអង្កេតមានចំនួនច្រើនជាងអថេរពន្យល់។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

iii. លក្ខណៈរបស់មេគុណព្យាករណ៍

សន្មតថាមេគុណព្យាករណ៍បានរកឃើញ $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$ ។ សមីការម៉ាទ្រីសក្រោមទម្រង់ជាលីនេអ៊ែរមានច្រើនរបៀប៖

$$\left.\begin{array}{l} Y = X\beta + u_i \\ Y = X\hat{\beta} + e \\ \hat{Y} = X\hat{\beta} \end{array}\right\} \Rightarrow e = Y - \hat{Y}, where\ e = residual,$$ យើងបាន៖

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{X}'(X\beta + e)] = (\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{X}'(X\beta) + \boldsymbol{X}'\mathrm{e}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'X)\beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$$

$$= \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$$

ដូច្នេះ $E(\hat{\beta}) = \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}(\boldsymbol{e}) = \beta$ ពីព្រោះ $\boldsymbol{E}(\boldsymbol{e}) = 0$ ។ ដូច្នេះមេគុណ $E(\hat{\beta}) = \beta$ ។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- គណនាវ៉ារ្យង់របស់មេគុណព្យាករណ៍ $\beta_0, \beta_1, \ldots, \beta_k$

  យើងមាន $\widehat{\boldsymbol{\beta}} = \beta + (\boldsymbol{X'X})^{-1}\boldsymbol{X'e} \Rightarrow \widehat{\boldsymbol{\beta}} - \beta = (\boldsymbol{X'X})^{-1}\boldsymbol{X'e}$

- លក្ខណៈ H3 វ៉ារ្យង់របស់មេគុណព្យាករណ៍កំណត់ដោយ $E\left((\hat{\beta} - \beta)^2\right) = \Omega^2_{\hat{\beta}}$

យើងមាន $\widehat{\boldsymbol{\beta}} - \beta = (\boldsymbol{X'X})^{-1}\boldsymbol{X'e}$ នាំឲ្យ $(\hat{\beta} - \beta)' = ((\boldsymbol{X'X})^{-1}\boldsymbol{X'e})' = e'X(\boldsymbol{X'X})^{-1}$

យើងបាន $(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = (\boldsymbol{X'X})^{-1}\boldsymbol{X'e}\, e'X(\boldsymbol{X'X})^{-1}$ យើងទាញបានកន្សោមវ៉ារ្យង់មេគុណៈ

$\Omega^2_{\hat{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'E}(\boldsymbol{e}\, e')X(\boldsymbol{X'X})^{-1}$ ដោយសារ $\boldsymbol{E}(\boldsymbol{e}\, e') = \sigma^2 I$ ជាម៉ាទ្រីសវ៉ារ្យង់កូវ៉ារ្យង់តួលំអៀង ដោយយោងតាមលក្ខណៈ H3 និង H4 ។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

$$\Omega_u = E(ee') = \begin{bmatrix} E(e_1e_1) & E(e_1e_2) & \cdots & E(e_1e_n) \\ E(e_2e_1) & E(e_2e_2) & \cdots & E(e_2e_n) \\ . & . & \cdots & . \\ E(e_ne_1) & E(e_ne_2) & \cdots & E(e_ne_n) \end{bmatrix} = \begin{bmatrix} \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & 0 \\ . & . & \cdots & . \\ 0 & 0 & \cdots & \sigma_e^2 \end{bmatrix} = \sigma_e^2 I$$

នាំឲ្យ $\Omega_{\hat{\beta}}^2 = (\boldsymbol{X'X})^{-\mathbf{1}}\boldsymbol{X'}\sigma^2 IX(\boldsymbol{X'X})^{-\mathbf{1}} = \sigma^2(\boldsymbol{X'X})^{-\mathbf{1}}(\boldsymbol{X'}X)(\boldsymbol{X'X})^{-\mathbf{1}} = \sigma^2(\boldsymbol{X'X})^{-\mathbf{1}}$

ដូច្នេះយើងបាន $\boldsymbol{\Omega_{\hat{\beta}}^2 = \sigma^2(X'X)^{-1}}$ ។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- គណនាវ៉ារ្យង់នៃមេគុណព្យាករណ៍ $sd_{\widehat{\beta}_k}, k = 0,1, \dots, k$
  a. មុនដំបូងគណនាវ៉ារ្យង់នៃតួសំណល់ ឬតួលំអៀង

     យើងមានរូបមន្តស្រាប់ $\hat{\sigma}^2 = \frac{\sum e_i^2}{n-k-1} = \frac{e_i' e_i}{n-k-1}$
  b. បន្ទាប់មកគណនាវ៉ារ្យង់នៃមេគុណព្យាករណ៍

     រូបមន្តវ៉ារ្យង់មេគុណព្យាករណ៍គឺជាផលគុណនៃវ៉ារ្យង់តម្លៃលំអៀងនិងម៉ាទ្រីសច្រាស $(X'X)^{-1}$។ យើងបាន $\widehat{\boldsymbol{\Omega}}_{\widehat{\boldsymbol{\beta}}_k} = \boldsymbol{\hat{\sigma}^2}(\boldsymbol{X'X})^{-\mathbf{1}}$ ។

     តម្លៃធាតុម៉ាទ្រីសនៅលើអង្កត់ទ្រូងគោលតាមលំដាប់ជាតម្លៃវ៉ារ្យង់របស់មេគុណ $\beta_0, \beta_1, \dots, \beta_k$ តាមលំដាប់។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- តារាងវិភាគវ៉ារ្យង់

ដូចគំរូព្យាករណ៍មួយអថេរពន្យល់ដែរ យើងមាន៖

1) $\sum Y_i = \sum \hat{Y}_i \Rightarrow \bar{Y} = \bar{\hat{Y}}$

2) $\sum e_i = 0 \Rightarrow \bar{e} = 0$

តាមទំនាក់ទំនងទាំងពីរខាងលើនេះ យើងទាញបានសមីការវ៉ារ្យង់ដូចគំរូព្យាករណ៍មួយអថេរ៖

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

$$TSS \quad = \quad ESS \quad + \quad RSS$$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ចែកសមីការរាំរៀងនឹង TSS យើងបាន៖

$$1 = \frac{ESS+RSS}{TSS} = R^2 + \frac{RSS}{TSS} \Rightarrow R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$ ដែល $R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$

Sources of Variables	Sum of Square(SS)	Degree of Freedom(df)	Mean Square(MS)
$X_i, i = 1, 2, \ldots k$	ESS = $\sum(\hat{Y}_i - \bar{Y})^2$	K	$\frac{ESS}{k}$
Residual	RSS = $\sum e_i^2$	$n - k - 1$	$\frac{RSS}{n-k-1}$
TOTAL	TSS = $\sum(Y_i - \bar{Y})^2$	$n - 1$	

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- ពិន្ទុស្ថិតិ Fisher : $F^* = \frac{\sum(\hat{Y}_i - \bar{\hat{Y}})^2/K}{\sum(Y_i - \bar{Y})^2/(n-k-1)} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$ ។
- សម្មតិកម្មន៍រមលរបស់សំណល់លំអៀង នាំឲ្យយើងបានសម្មតិកម្ម H0: F^* ជាច្បាប់នៃ Fisher ។ ដូច្នេះយើងប្រៀបធៀប F^* នេះតាមការគណនាទៅនឹង F នៃទ្រឹស្តីចំពោះកម្រិតសេរី K និង$(n-k-1)$៖ បើ $F^* > F$ នោះយើងបដិសេធន៍សម្មតិកម្ម H0 គំរូព្យាករណ៍ពន្យល់បានទាំងស្រុង។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- តម្លៃ R^2 កែតម្រូវកំណត់សរសេរដោយ $\bar{R}^2$(Adjust R-Square)

យើងមានរូបមន្ត $R^2 = \frac{\sum(\hat{Y}_i - \bar{\hat{Y}})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$ សមភាពនៃការកែតម្រូវនេះនិងវាយតម្លៃថាគេអាចធ្វើការកែតម្រូវ R^2 តម្រូវទៅតាមកម្រិតសេរីនៃការប៉ាន់ប្រមាណ។ ដូច្នេះកាលណាកម្រិតសេរីទាប(ចំនួនអថេរពន្យល់និងចំនួនអង្កេតស្មើគ្នា កម្រិតសេរីស្មើ0 តម្លៃ $R^2 = 1 \Rightarrow$នៅពេលនោះកម្រិតនៃការពន្យល់ក្នុងគំរូវិភាគឥតមាន) នៅទីបំផុតវាត្រូវកែតម្រូវតម្លៃ R^2 ដោយគិតពីចំនួននៃការអង្កេតឡើងវិញដោយធៀបជាមួយអថេរពន្យល់។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

យើងបាន R^2កែត្រូវរបស់សមីការមានមួយអថេរពន្យល់កំណត់ដោយ $\bar{R}^2 = 1 - \frac{\frac{RSS}{n-2}}{\frac{TSS}{n-1}} \rightarrow$

$\bar{R}^2 = 1 - \frac{RSS}{TSS} \times \frac{n-1}{n-2} = 1 - \frac{TSS-ESS}{TSS} \times \frac{n-1}{n-2} \Rightarrow \boldsymbol{\bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-2}}$

ការគណនាតម្លៃមេគុណកែសម្រួល R^2 ($\boldsymbol{\bar{R}^2}$)មាន k អថេរពន្យល់ តាមនិយមន័យខាងលើ

កំណត់ដោយ $\boldsymbol{\bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1}}$។ ជាទូទៅយើងបាន $\bar{R}^2 < R^2$, ហើយកាលណា n

កាន់តែធំ នោះតម្លៃនៃ $\bar{R}^2 \approx R^2$ ។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

- ឧទាហរណ៍ គេឲ្យសមីការមានពអរថេរពន្យល់ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i, i = 1,2,\dots,14$ ។ សំណួរ៖ ១. សរសេរសមីការព្យាការណ៍ជារាងម៉ាទ្រីស, ២. គណនាមេគុណព្យាករណ៍, ៣. គណនាគម្លាតស្តង់ដារបស់មេគុណព្យាករណ៍នីមួយៗ, ៤. គណនាមេគុណកំណត់ទំនាក់ទំនង R^2 និង $\bar{R}^2$។ ខាងក្រោមជាទិន្នន័យដែលប្រើក្នុងការសិក្សាមេរៀននេះ។

ob	Y	X1	X2	X3	ob	Y	X1	X2	X3
1	12	2	45	121	**8**	19	5	33	147
2	14	1	43	132	**9**	21	5	41	128
3	10	3	43	154	**10**	16	8	38	163
4	16	6	47	145	**11**	19	4	32	161
5	14	7	42	129	**12**	21	9	31	172
6	19	8	41	156	**13**	25	12	35	174
7	21	8	32	132	**14**	21	7	29	180

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ក. សមីការម៉ាទ្រីសនៃគំរូវិភាគគឺ $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{u}$

ចំនួនអង្កេតមាន ១៤ ចំនួនអអេថរមាន ៤ គំរូវិភាគអាចសរសេរ៖

$$Y = \begin{pmatrix} 12 \\ 14 \\ . \\ . \\ . \\ 21 \end{pmatrix}, X = \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & 7 & 29 & 180 \end{pmatrix} \quad , \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_{14} \end{pmatrix}$$

លំដាប់ម៉ាទ្រីសនីមួយៗនៃសមីការម៉ាទ្រីសគឺ $\boldsymbol{Y}_{(\mathbf{14,1})} = \boldsymbol{X}_{(\mathbf{14,4})}\,\boldsymbol{\beta}_{(\mathbf{4,1})} + \boldsymbol{u}_{(\mathbf{14,1})}$

ខ. គណនាមេគុណប៉ារ៉ាម៉ែត្រនៃសមីការព្យាករណ៍

យើងមានរូបមន្តកំណត់មេគុណ $\boldsymbol{\hat{\beta}} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'Y})$ ។ បន្ទាប់មកគណនាកន្សោម $(\boldsymbol{X'X})$ និង $(\boldsymbol{X'X})^{-1}$ ៖

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ផលគុណ $X'X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 1 & \cdots & 7 \\ 45 & 43 & \cdots & . \\ 121 & 132 & \cdots & 180 \end{pmatrix} \begin{pmatrix} 1 & 2 & 45 & 121 \\ 1 & 1 & 43 & 132 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & 7 & 29 & 180 \end{pmatrix} = \begin{pmatrix} 14 & 85 & 532 & 2094 \\ 85 & 631 & 3126 & 13132 \\ 532 & 3126 & 20666 & 78683 \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}$

ម៉ាទ្រីសច្រាស $(\boldsymbol{X'X})^{-1} = \begin{pmatrix} 20.16865 & 0.015066 & -0.23145 & -0.07617 \\ 0.015066 & 0.013205 & 0.001194 & -0.00094 \\ -0.23145 & 0.001194 & 0.003635 & 0.000575 \\ -0.07617 & -0.00094 & 0.000575 & 0.000401 \end{pmatrix}$

ផលគុណ $X'Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2 & 1 & \cdots & 7 \\ 45 & 43 & \cdots & . \\ 121 & 132 & \cdots & 180 \end{pmatrix} \begin{pmatrix} 12 \\ 14 \\ . \\ . \\ . \\ 21 \end{pmatrix} = \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

គណនាមេគុណ $\boldsymbol{\beta}$

យើងបាន $\beta = (\boldsymbol{X'X})^{-1}X'Y = \begin{pmatrix} 20.16865 & 0.015066 & -0.23145 & -0.07617 \\ 0.015066 & 0.013205 & 0.001194 & -0.00094 \\ -0.23145 & 0.001194 & 0.003635 & 0.000575 \\ -0.07617 & -0.00094 & 0.000575 & 0.000401 \end{pmatrix} \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix} = \begin{pmatrix} 32.89132 \\ 0.801901 \\ -0.38136 \\ -0.03713 \end{pmatrix}$

ដូច្នេះយើងបានតម្លៃមេគុណប៉ារ៉ាម៉ែត្រ $\hat{\boldsymbol{\beta}}_0 = \mathbf{32.89}$, $\hat{\boldsymbol{\beta}}_1 = \mathbf{0.80}$, $\hat{\boldsymbol{\beta}}_2 = \mathbf{-0.38}$, $\hat{\boldsymbol{\beta}}_3 = \mathbf{-0.037}$

គ. គណនា វ៉ារ្យង់នៃតួលំអៀង $\boldsymbol{\sigma^2}$ និង គម្លាតស្តង់ដានៃមេគុណព្យាករណ៍ $\boldsymbol{sd_{\hat{\beta}}}$

ដំបូងគណនាវ៉ារ្យង់នៃតួលំអៀង $\sigma^2 = \sum e_i^2/(n-k-1) = (e_1^2 + \cdots + e_{14}^2)/10,$

$where \quad e_i = Y_i - (32.89 + 0.80X_{1i} - 0.38X_{2i} - 0.037X_{3i})$, i=1,2, … , 14

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

រាតាងគណនាតម្លៃលំអៀង៖ $e_i = Y_i - (32.89 + 0.80X_{1i} - 0.38X_{2i} - 0.037X_{3i})$, i=1,2, … , 14

Ob	Y	Predicted Y	Residuals	Sauqre ei
1	12	12.840795	-0.8407945	0.70694
2	14	12.393162	1.60683822	2.58193
3	10	13.18005	-3.1800496	10.1127
4	16	14.394494	1.60550591	2.57765
5	14	17.697326	-3.6973256	13.6702
6	19	17.878013	1.12198714	1.25886
7	21	22.201453	-1.2014526	1.44349
8	19	18.857402	0.14259837	0.02033
9	21	16.512019	4.487981	20.142
10	16	18.762173	-2.7621729	7.6296
11	19	17.917009	1.08299079	1.17287
12	21	21.899418	-0.8994182	0.80895
13	25	22.705406	2.29459405	5.26516
14	21	20.761282	0.23871792	0.05699
Sum=			0.00	**67.4477**

$$\hat{\sigma}_u^2 = \frac{ee'}{n-k-1} = \frac{\sum_{i=1}^{14} e_i^2}{14-3-} = \frac{67.4477}{10} =$$

$= 6.745$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ម៉ាទ្រីសវ៉ារ្យង់ កូវ៉ារ្យង់កំណត់ដោយកន្សោមផលគុណ៖

$$\widehat{\Omega}_{\hat{\beta}} = \hat{\sigma}_u^2 \begin{pmatrix} \mathbf{20.16865} & 0.015066 & -0.23145 & -0.07617 \\ 0.015066 & \mathbf{0.013205} & 0.001194 & -0.00094 \\ -0.23145 & 0.001194 & \mathbf{0.003635} & 0.000575 \\ -0.07617 & -0.00094 & 0.000575 & \mathbf{0.000401} \end{pmatrix}$$

យើងទាញបានតម្លៃវ៉ារ្យង់របស់មេគុណប៉ាន់ស្មាននីមួយៗតទៅនេះ៖

$$\hat{\sigma}^2_{\hat{\beta}_0} = 6.745 \times 20.168 = 136.04 \Rightarrow sd_{\hat{\beta}_0} = \sqrt{136.04} = 11.66$$

$$\hat{\sigma}^2_{\hat{\beta}_1} = 6.745 \times 0.013 = 0.087 \Rightarrow sd_{\hat{\beta}_1} = \sqrt{0.087} = 0.29$$

$$\hat{\sigma}^2_{\hat{\beta}_2} = 6.745 \times 0.0036 = 0.024 \Rightarrow sd_{\hat{\beta}_2} = \sqrt{0.024} = 0.15$$

$$\hat{\sigma}^2_{\hat{\beta}_3} = 6.745 \times 0.0004 = 0.0026 \Rightarrow sd_{\hat{\beta}_3} = \sqrt{0.0026} = 0.05$$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

គ. គណនាR^2 និង $\bar{R}^2$

យើងមាន$e'e = 67.45 \; and \; calculate \; \sum(Y_i - \bar{Y})^2 = 227.75$

យើងបាន $R^2 = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{67.45}{227.85} = 0.702$

មេគុណកំណត់ទំនាក់ទំនងកែតម្រូវគឺ

$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2) = 1 - \frac{14-1}{14-3-}(1 - 0.702) = 0.613$

យើងសង្កេតឃើញថា នៅពេលយើងកែប្រែកម្រិតសេរី វាមានការថយចុះនៃមេគុណកំណត់ទំនាក់ទំនង R^2 ។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ឧទាហរណ៍ទី២ មានទិន្នន័យនៃព្រឹត្តិការពាក់ព័ន្ធគ្នា Y_i, X_{1i}, X_{2i} ។

ក. សរសេរសមីការជារាងម៉ាទ្រីស និងគណនាមេគុណប៉ារ៉ាម៉ែត្រព្យាការណ៍។

ខ. គណនាតម្លៃលំអៀង បន្ទាប់មកសន្និដ្ឋានតម្លៃវ៉ារ្យង់លំអៀង។

គ. គណនាម៉ាទ្រីសវ៉ារ្យង់ កូវ៉ារ្យង់របស់មេគុណប៉ារ៉ាម៉ែត្រ។

ឃ. ធ្វើតេស្តមេគុណប៉ារ៉ម៉ែត្រព្យាករណ៍ទាំងពីរត្រង់ហានិភ័យ៥%។

ចម្លើយ

ក. គំរូវិភាគសរសេរជារាងម៉ាទ្រីស $Y = X\beta + e$. ដែលមានចំនួនអង្កេត n = 10, អថេរពន្យល់ k = 2.

យើងបានម៉ាទ្រីសនីមួយៗដូចខាងក្រោមនេះ៖

- ទិន្នន័យព្រឹត្តិការណ៍

No	Yi	X1i	X2i
1	12	7	48
2	21	9	40
3	24	11	18
4	24	12	28
5	13	7	40
6	17	9	32
7	21	12	31
8	26	14	24
9	31	19	22
10	30	21	25

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

$$Y_{(10;1)}=\begin{pmatrix}12\\21\\.\\.\\.\\30\end{pmatrix}, X_{(10;3)}=\begin{pmatrix}1&7&48\\1&9&40\\.&.&.\\.&.&.\\.&.&.\\1&21&25\end{pmatrix}, \beta_{(3;1)}=\begin{pmatrix}\beta_0\\\beta_1\\\beta_2\end{pmatrix}, e_{(10,1)}=\begin{pmatrix}e_1\\e_2\\.\\.\\.\\e_{10}\end{pmatrix},$$

ការកំណត់មេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍ត្រូវប្រើវិធី OLS ដែលបានផ្តល់តាមរូបមន្តៈ $\hat{\beta}=(X'X)^{-1}(X'Y)$ ដែល X' ជាម៉ាទ្រីសត្រង់ស្ប៉ូរបស់ម៉ាទ្រីស X ។ មេគុណរបស់វិធីការវ តម្លៃតូចគឺគ្មានលំអៀងនិងរួម គឺវាមានតម្លៃBLUE(The Best Linear Unbiaised Estimator).

$$X'_{(3;10)}=\begin{pmatrix}1&1\cdots&1\\7&9\cdots&21\\48&21\cdots&25\end{pmatrix}\Rightarrow(X'X)_{(3;3)}=\begin{pmatrix}10&121&308\\121&1667&3449\\308&3449&10282\end{pmatrix}\text{ និងផលគុណ}(X'Y)_{(3;1)}=\begin{pmatrix}219\\2904\\6291\end{pmatrix}$$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ម៉ាទ្រីសច្រាស $(X'X)^{-1} = \begin{pmatrix} 6.224592 & -0.21581 & -0.11407 \\ -0.21581 & 0.009443 & 0.003297 \\ -0.11407 & 0.003297 & 0.002408 \end{pmatrix}$

ដូច្នេះយើងបានមេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍គឺ $(X'X)^{-1}(X'Y) = \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 18.872 \\ 0.902 \\ -0.256 \end{pmatrix}$។

សមីការព្យាករណ៍កំណត់បាន៖ $Y_i = 18.872 + 0.902X_{1i} - 0.256X_{2i} + e_i \, , i = 1,2, \dots , 10.$

ខ. គណនាតម្លៃលំអៀង៖

យើងមានសមីការព្យាករណ៍ $Y_i = 18.872 + 0.902X_{1i} - 0.256X_{2i} + e_i \, , i = 1,2, \dots , 10.$

តម្លៃលំអៀង $e_i = Y_i - 18.872 - 0.902X_{1i} + 0.256X_{2i} \, , i = 1,2, \dots , 10.$

ដើម្បីគណនាតម្លៃលំអៀង e_i នីមួយៗគឺត្រូវជំនួសគូតម្លៃរបស់(Y_i, X_{1i}, X_{2i})ក្នុងសមីការព្យាករណ៍ម្តងមួយៗ បន្ទាប់មកគណនាវា នោះយើងនឹងបានតម្លៃលំអៀងចាប់ពីចំនុចទី១ដល់ទី១០។

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

ផលបូកការេតម្លៃតូច *RSS តាមវិធី OLS* គឺជាផលបូករបស់ការេតម្លៃលំអៀងនីមួយៗកំណត់ដោយ $RSS = \sum_{i=1}^{10} e_i^2$ = 31.469. ។ វ៉ារ្យង់របស់តម្លៃលំអៀងគឺ $\hat{\sigma}_e^2 = \frac{\sum_{i=1}^{10} e_i^2}{10-2-1} = \frac{31.469}{7} = 4.495$ ។

$\mathrm{Y} = X\beta + u$ សមីការម៉ាទ្រីសតាមទ្រឹស្តី

$\mathrm{Y} = X\hat{\beta} + e$ សមីការព្យាករណ៍មានលំអៀង

$\hat{Y} = X\hat{\beta}$ គំរូកែសម្រួលសម្រាប់ព្យាករណ៍

$e = \mathrm{Y} - X\hat{\beta} = Y - \hat{Y}$ =សមីការប៉ាន់ស្មានតម្លៃលំអៀង (estimated the residuals).

- excei

ob	Yi	Predicted Y	e_i	e_i^2
1	12	12.89753176	-0.897532	0.8055633
2	21	16.74915257	4.2508474	18.069704
3	24	24.18461944	-0.184619	0.0340843
4	24	22.52658377	1.4734162	2.1709554
5	13	14.9454438	-1.945444	3.7847516
6	17	18.7970646	-1.797065	3.2294412
7	21	21.75861676	-0.758617	0.5754994
8	26	25.35424856	0.6457514	0.4169949
9	31	30.37549849	0.6245015	0.3900021
10	30	31.41124024	-1.41124	1.991599
			Sum=	**31.469**

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

គ. ម៉ាទ្រីសវ៉ារ្យង់កូវ៉ារ្យង់របស់មេគុណព្យាករណ៍ $\widehat{\Omega}_{\widehat{\beta}} = \hat{\sigma}_e^2(X'X)^{-1}$

យើងបានគណនារួចហើយ $\hat{\sigma}_e^2$ និង $(X'X)^{-1}$ ក្នុងសំណួរក. ជាបន្តយើងបាន

$$\widehat{\Omega}_{\widehat{\beta}} = \hat{\sigma}_e^2(X'X)^{-1} = 4.495 \times \begin{pmatrix} 6.224592 & -0.21581 & -0.11407 \\ -0.21581 & 0.009443 & 0.003297 \\ -0.11407 & 0.003297 & 0.002408 \end{pmatrix} = \begin{pmatrix} 27.97954 & -0.97007 & -0.51273 \\ -0.97007 & 0.042446 & 0.014821 \\ -0.51273 & 0.014821 & 0.010825 \end{pmatrix}$$

វ៉ារ្យង់របស់មេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍នីមួយៗជាធាតុរបស់អង្កត់ទ្រូងគោលទី១ និងកូវ៉ារ្យង់ជាធាតុនៅខាងក្រៅអង្កត់ទ្រូងគោលទី១។ យើងបានវ៉ារ្យង់របស់មេគុណនីមួយៗគឺ៖

$\hat{\sigma}_{\widehat{\beta}_0}^2 = 27.979 \Rightarrow \hat{\sigma}_{\widehat{\beta}_0} = \sqrt{27.979} = 5.289,\ \hat{\sigma}_{\widehat{\beta}_1}^2 = 0.042 \Rightarrow \hat{\sigma}_{\widehat{\beta}_1} = \sqrt{0.042} = 0.206$

$\hat{\sigma}_{\widehat{\beta}_2}^2 = 0.0108 \Rightarrow \hat{\sigma}_{\widehat{\beta}_2} = \sqrt{0.0108} = 0.104.$

ឃ. ធ្វើតេស្តសម្មតិកម្មរបស់មេគុណអថេរពន្យល់ X_{1i} និង X_{2i}

សម្មតិកម្ម H_0: $\hat{\beta}_1 = 0\ and\ \hat{\beta}_2 = 0$ និងសម្មតិកម្ម H_1: $\hat{\beta}_1 \neq 0\ and\ \hat{\beta}_2 \neq 0$

គំរូពហុករណ៍ច្រើនអថេរ
Multiple Linear Regression

ratio student: $t^*_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.902}{0.206} = 4.377,\ and\ \ t^*_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{|-0.256|}{0.104} = 2.460$

ហើយ ratio student ដែលអានក្នុងតារាងត្រង់កម្រិតសេរី $n-k-1 = 10-2-1 = 7$:

$t_{(10;2)}^{5\%/2} = 2.365$ គឺមានតម្លៃតូចជាង ratio student របស់មេគុណ X_{1i} និង X_{2i} រៀងគ្នា។ ដូច្នេះមេគុណអថេរពន្យល់ទាំងពីរមានអត្ថន័យស្ថិតិគ្រប់គ្រាន់ខុសពី 0 ត្រង់ហានិភ័យ៥%ដែលគេឲ្យ។ អង្កត់ជឿជាក់របស់មេគុណនីមួយកំណត់តាមរូបមន្តៈ

$$\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\hat{\beta}_1}} = \mp t_{(10;2)}^{5\%/2} \Rightarrow \hat{\beta}_1 - \hat{\sigma}_{\hat{\beta}_1} t_{(10;2)}^{5\%/2} < \beta_1 < \hat{\beta}_1 + \hat{\sigma}_{\hat{\beta}_1} t_{(10;2)}^{5\%/2} \rightarrow \mathbf{(0.415 < \boldsymbol{\beta_1} < 1.389)}$$

$$\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\hat{\beta}_2}} = \mp t_{(10;2)}^{5\%/2} \Rightarrow \hat{\beta}_2 - \hat{\sigma}_{\hat{\beta}_2} t_{(10;2)}^{5\%/2} < \beta_2 < \hat{\beta}_2 + \hat{\sigma}_{\hat{\beta}_2} t_{(10;2)}^{5\%/2} \rightarrow \mathbf{(-0.502 < \boldsymbol{\beta_2} < -0.010)}$$

គំរូព្យាករណ៍ច្រើនអថេរ
Multiple Linear Regression

លំហាត់អនុវត្ត គេឲ្យទិន្នន័យអង្កេតស្តីពីការបរិភោគសាច់ជ្រូកប្រចាំឆ្នាំនៅសហរដ្ឋអាមេរិករយៈពេល២០ឆ្នាំ។

ក. ចូរកំណត់លក្ខណៈទូទៅរបស់អនុគមន៍តម្រូវការ $Q_i = D(P_i, Y_i)$និងអនុគមន៍ផ្គត់ផ្គង់ $Q_i = S(P_i, Z_i)$។ បកស្រាយអត្ថន័យមេគុណនីមួយៗក្នុងសមីការនីមួយៗ។

$P_i =$ ថ្លៃសាច់ជ្រូកគិតជាសេនក្នុង១ផោន(Cents per pound)

$Q_i =$ បរិមាណបរិភោគសាច់ជ្រូកគិតជាផោន(Pound per capita)

$Y_i =$ ចំណូលបំរុងទុករបស់មនុស្សម្នាក់ៗ

$Z_i =$ បរិមាណសាច់ជ្រូកដែលបានគ្រោងទុកក្នុងផលិតកម្មសាច់ជ្រូក

ខ. ធ្វើតេស្តមេគុណប៉ារ៉ាម៉ែត្ររបស់សមីការនីមួយៗដែលទទួលបាន។

គ. គណនាមេគុណកំណត់ទំនាក់ទំនងកែតម្រូវ $\bar{R}^2$។

ធ្វើសេចក្តីសន្និដ្ឋានលទ្ធផលរបស់ការអង្កេត។

- ទិន្នន័យ

ob	Pi	Yi	Qi	Pi	Zi
1	26.8	541	65.7	26.8	74.0
2	25.3	616	74.2	25.3	84.7
3	25.3	610	74.0	25.3	80.2
4	31.1	636	66.8	31.1	69.9
5	33.3	651	64.1	33.3	66.8
6	31.2	645	67.7	31.2	71.6
7	29.5	653	70.9	29.5	73.6
8	30.3	682	69.6	30.3	71.2
9	29.1	604	67.0	29.1	69.6
10	23.7	515	68.4	23.7	68.0
11	15.6	390	70.7	15.6	74.8
12	13.9	364	69.6	13.9	73.6
13	18.8	411	63.1	18.8	70.2
14	27.4	459	48.4	27.4	46.5
15	26.9	517	55.1	26.9	57.6
16	27.7	551	55.8	27.7	58.7
17	24.5	506	58.2	24.5	58.0
18	22.2	538	64.7	22.2	67.2
19	19.3	576	73.5	19.3	73.7
20	24.7	697	68.4	24.7	66.5

អ្នកដែលបានមកអង្គុយរៀនគឺ ទទួលបានលាភធំណាស់ពីមាតាបិតា។

បើមិនបញ្ចេញពលកម្ម
និងសមត្ថភាពទេ
តើបានអ្វីជាក់ក្រពះ?