ជំពូកទី៣ ជំរឿនខាយ

Simple Linear Regression

រៀបរៀងនិងបង្រៀនដោយបណ្ឌិត ង៉ាន់ ស៊ុនដេត
សាស្ត្រាចារ្យនៃសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រនិងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច

ជំរឿនងាយ

Simple Linear regression

- ទិន្នផលកសិកម្ម(yield)សណ្តែកមានជាប់ពាក់ព័ន្ធនឹងបរិមាណជី(fertilizer)ដែលមានក្នុងដីចំការរបស់កសិករកំណត់ដោយសមីការ៖ $yield_i = \beta_0 + \beta_1 fertilizer_i + u_i, i = 1,2, \dots, n$ ដែល u_i ជាតួលំអៀងកំណត់មិនបានមួយដែលមានឥទ្ធិពលដល់ទិន្នផលសណ្តែកដែរ។
- ប្រាក់ឈ្នួល ($wage$) របស់បុគ្គលម្នាក់ៗជាប់ពាក់ព័ន្ធនឹងចំនួនឆ្នាំសិក្សា (edu) ដែលបានរៀនចប់ថ្នាក់នៃឆ្នាំសិក្សានីមួយៗកំណត់ដោយ៖ $wage_i = \beta_0 + \beta_1 edu_i + u_i, i = 1,2, \dots, n$ ដែល u_iជាតួលំអៀងកំណត់មិនបានមួយដែលមានឥទ្ធិពលដល់ប្រាក់ឈ្នួលដែរ។
- ពិន្ទុប្រឡងបញ្ចប់ការសិក្សា(score)មានជាប់ពាក់ព័ន្ធនឹងចំនួនវត្តមាន(attend)ដែលបានចូលរៀននិងស្តាប់គ្រូពន្យល់គ្រប់មុខវិជ្ជាកំណត់ដោយ៖ $score_i = \beta_0 + \beta_1 attend_i + u_i, i = 1,2, \dots, n$ ដែល u_iជាតួលំអៀងកំណត់មិនបានមួយដែលមានឥទ្ធិពលដល់ចំនួនពិន្ទុសរុបដែរ។
- គំរូទាំងបីខាងលើនេះគឺជាគំរូជម្រឿនងាយ (Simple Linear Regression)។

ជម្រើនងាយ

Simple Linear regression

- អថេរនៅអង្គខាងឆ្វេងនៃសមីការតាងដោយ Y_i ហៅថា *អថេរត្រូវពន្យល់* (Dependent Variable)។ អថេរនៅអង្គខាងស្តាំនៃសមីការតាងដោយ X_i ហៅថា *អថេរពន្យល់* (Independent Variable)។
- នៅក្នុងការអង្កេតរបស់ព្រឹត្តិការណ៍អ្វីមួយ យើងអាចកំណត់យកគូ $(X_i, Y_i), i = 1,2, \ldots, n$ ។ នៅពេលនេះយើងកំណត់បានសមីការជម្រៀនងាយមួយគឺ $Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1,2, \ldots, n$ ។
- ប៉ារ៉ាម៉ែត្រព្យាករណ៍ β_0 និង β_1 ជាមេគុណព្យាករណ៍ដែលត្រូវកំណត់
- u_i ជាតួលំអៀងមួយដែលមានឥទ្ធិពលនៅក្នុសមីការព្យាករណ៍។ ក្នុងការអនុវត្តន៍តួលំអៀង u_i ត្រូវប្តូរទៅជាតម្លៃលំអៀងវិញកំណត់ដោយ e_i ។
- វិធីកំណត់ប៉ារ៉ាម៉ែត្រព្យាករណ៍ β_0 និង β_1 គឺប្រើលក្ខខណ្ឌ *អប្បបរមាផលបូកការេតម្លៃតូច* (OLS) (Ordinary Least Square)។

ជម្រើនងាយ

Simple Linear regression

- មានសមីការលីនេអ៊ែរទូទៅ៖ $\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + u_i, i = 1, 2, \dots, n}$ ។
- មានកំណត់គោលការណ៍មួយចំនួនដែលស្មើរតែក្លាយជាសម្មតិកម្មទៅហើយ ដែលសម្រាប់ធ្វើជាមូលដ្ឋានសំខាន់ក្នុងការកំណត់ប៉ារ៉ាម៉ែត្រព្យាករណ៍ β_0 និង β_1 ៖
  - H1: Data generating process is $\boldsymbol{Y_i = \beta_0 + \beta_1 X_i + e_i}$.
  - H2: The *n* observations of $\boldsymbol{X_i}$ are fixed numbers.
  - H3: The *n* error terms $\boldsymbol{e_i}$ are random, with E($\boldsymbol{e_i}$)=0.
  - H4: The variance of *n* errors is fixed, $E\left(e_i^2\right) = \sigma^2$.
  - H5: The errors are uncorrelated, $E\left(e_i e_j\right) = 0, for\ all\ i \neq j$.
  - H6: β_0 និង β_1 are unknown, but fixed for all *n* observation.
  - H7: $e_1, e_2, \dots, e_n$ are jointly normally distribution, with H3, H4, H5; $e_i \sim N(0, \sigma^2)$.

ជម្រឿនងាយ

Simple Linear regression

- ឧបមាថាយើងមានសមីការទូទៅកំណត់ដោយ $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i, i = 1,2,\dots,n.$ (1) និង សមីការព្យាករណ៍កំណត់ដោយ $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1,2,\dots,n.$ (2)
- ធ្វើប្រមាណវិធីដកសមីការ (1) និង (2) យើងបាន៖ $Y_i - \hat{Y}_i = e_i \Rightarrow e_i = Y_i - \beta_0 - \beta_1 X_i.$
- សម្មតិកម្ម H3: $E(e_i) = 0 \Rightarrow \sum_{i=1}^{n} e_i = 0 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} e_i = \bar{e} = 0 \to \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0$

  $\Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}\beta_0 - \beta_1\frac{1}{n}\sum_{i=1}^{n} X_i = 0 \to \bar{Y} - \beta_0 - \beta_1\bar{X} = 0 \Rightarrow \boldsymbol{\beta_0 = \bar{Y} - \beta_1\bar{X}}.$
- ម្យ៉ាងទៀត $e_i = Y_i - \beta_0 - \beta_1 X_i = Y_i - (\bar{Y} - \boldsymbol{\beta_1}\bar{X}) - \beta_1 X_i = (Y_i - \bar{Y}) - \boldsymbol{\beta_1}(X_i - \bar{X})$
- សម្មតិកម្ម H4: $E\left(e_i^2\right) = \sigma^2 =$ ថេរ $= \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}[(Y_i - \bar{Y}) - \boldsymbol{\beta_1}(X_i - \bar{X})]^2 \Rightarrow$

  $\Rightarrow \sigma^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 + \sum_{i=1}^{n}\boldsymbol{\beta_1}^2(X_i - \bar{X}))^2 - 2\sum_{i=1}^{n}\boldsymbol{\beta_1}(X_i - \bar{X})(Y_i - \bar{Y})$ (3)

ជម្រើនងាយ

Simple Linear regression

- គណនាដេរីវេលំដាប់១របស់កន្សោមសមីការ (3) ធៀបនឹង β_1 យើងបាន៖

$$0 = 0 + 2\boldsymbol{\beta_1} \sum_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})^2 - 2\sum_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})(Y_i - \overline{\boldsymbol{Y}}) \Rightarrow$$

$$\Rightarrow 2\boldsymbol{\beta_1} \sum_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})^2 = 2\sum_{i=1}^{n}(X_i - \overline{\boldsymbol{X}})(Y_i - \overline{\boldsymbol{Y}}) \Rightarrow \widehat{\boldsymbol{\beta}}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

- ដូច្នេះ យើងបានស្រាយបញ្ជាក់វិធីកំណត់មេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍របស់គំរូវិភាគទូទៅ៖

$$\widehat{\boldsymbol{\beta}}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad , \quad \widehat{\boldsymbol{\beta}}_0 = \overline{\boldsymbol{Y}} - \widehat{\boldsymbol{\beta}}_1 \overline{\boldsymbol{X}}$$

ជម្រើនងាយ

Simple Linear regression

- លក្ខណៈរបស់មេគុណព្យាករណ៍

យើងមានសមីការទូទៅ $Y_i = \beta_0 + \beta_1 X_i + e_i \Rightarrow$ តម្លៃមធ្យម $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{e}$

ដកសមីការទូទៅនឹងតម្លៃមធ្យម $\bar{Y}$, => $Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (e_i - \bar{e})$ បន្ទាប់មកជំនួសក្នុង

កន្សោមមេគុណ β_1៖ $\boldsymbol{\hat{\beta}_1} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})[\beta_1(X_i-\bar{X})+(e_i-\bar{e})]}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{\beta_1 \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(X_i-\bar{X})(e_i-\bar{e})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \rightarrow$

$= \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i-\bar{X})(e_i-\bar{e})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(X_i-\bar{X})e_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2} - \frac{\sum_{i=1}^{n}(X_i-\bar{X})\bar{e}}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$

ដោយសារ $\sum_{i=1}^{n}(X_i - \bar{X})\bar{e} = \bar{e}(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X}) = \bar{e}n\bar{X} - \bar{e}n\bar{X} = 0$

ដូច្នេះយើងនៅសល់តែកន្សោម $\Rightarrow \hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i-\bar{X})e_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$

ជម្រើនងាយ

Simple Linear regression

- បើពិនិត្យតាមសម្មតិកម្ម H3: $E(e_i) = 0 \Rightarrow E(\hat{\beta}_1) = E(\beta_1) + \frac{\sum_{i=1}^{n}(X_i - \bar{X})E(e_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = E(\beta_1) = \beta_1$

  ដូច្នេះយើងបាន $E(\hat{\beta}_1) = \beta_1$ ។

- ស្រាយបញ្ជាក់ដូចគ្នានេះ $E(\hat{\beta}_0) = \beta_0$

  យើងមាន $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$ និង $\bar{Y} = \beta_0 + \beta_1\bar{X} + \bar{e} \Rightarrow \hat{\beta}_0 + \hat{\beta}_1\bar{X} = \beta_0 + \beta_1\bar{X} + \bar{e} \Rightarrow$

  $\Rightarrow \hat{\beta}_0 = \beta_0 + \bar{e} - \bar{X}(\hat{\beta}_1 - \beta_1) \Rightarrow E(\hat{\beta}_0) = E(\beta_0) + E(\bar{e}) + E[(\hat{\beta}_1 - \beta_1)\bar{X}]$

  ដោយសារ $E(e_i) = \frac{1}{n}\sum_{i=1}^{n} e_i = \bar{e} = 0 \Rightarrow E(\bar{e}) = 0$ និង $E(\hat{\beta}_1) = \beta_1 \Rightarrow E(\hat{\beta}_1 - \beta_1) = 0, E(\beta_0) = \beta_0$

  ដូច្នេះយើងបានស្រាយបញ្ជាក់លក្ខណៈ: $E(\hat{\beta}_0) = \beta_0$ ។

ជម្រើនងាយ

Simple Linear regression

- ទ្រឹស្តីបទ៖ វ៉ារ្យង់គំរូសំណាករបស់មេគុណព្យាករណ៍ $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ និង

$$Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)។$$

ស្រាយបញ្ជាក់

យើងមាន $Var(\hat{\beta}_1) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_1 - \bar{\beta}_1)^2 = E\left(\hat{\beta}_1 - E\left(\hat{\beta}_1\right)^2\right) = E(\hat{\beta}_1 - \beta_1)^2 = E\{\frac{\sum_{i=1}^{n}(X_i - \bar{X})e_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\}^2 =$

$$= E\left\{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 e_i^2}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2}\right\} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 E(e_i^2)}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} = \frac{Var(e_i)}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

ផម្រើនងាយ

Simple Linear regression

- ផលវិបាកនៃសម្មតិម្មរបស់លំអៀង Normal ៖

យើងមាន $e_i = Y_i - \hat{\beta}_1 X_i - \hat{\beta}_0 \Rightarrow e_i = Y_i - \hat{\beta}_1 X_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 \bar{X}$ ដោយធ្វើវិធីថែមថយតួ

យើងអាចសរសេរម្តងទៀត $e_i = (Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X})$, ដោយយក $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$

ម្យ៉ាងទៀត $Y_i - \bar{Y} = \beta_0 + \beta_1 X_i + e_i - \beta_0 - \beta_1 \bar{X} - \bar{e} = \beta_1 (X_i - \bar{X}) + (e_i - \bar{e})$

យើងបាន $e_i = \beta_1 (X_i - \bar{X}) + (e_i - \bar{e}) - \hat{\beta}_1 (X_i - \bar{X}) = \left(\beta_1 - \hat{\beta}_1\right)(X_i - \bar{X}) + (e_i - \bar{e})$

លើកអង្គទាំងពីរជាការេក្នុងចំនួនអង្កេត n យើងបាន៖

$$\sum_{i=1}^{n} e_i^2 = (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} (e_i - \bar{e})^2 + 2(\beta_1 - \hat{\beta}_1) \sum_{i=1}^{n} (X_i - \bar{X})(e_i - \bar{e})$$

ដោយសារមាន $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} (X_i - \bar{X})(e_i - \bar{e})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \Rightarrow \sum_{i=1}^{n} (X_i - \bar{X})(e_i - \bar{e}) = -(\beta_1 - \hat{\beta}_1) \sum_{i=1}^{n} (X_i - \bar{X})^2$

ជម្រើនងាយ

Simple Linear regression

យើងបាន $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(e_i - \bar{e})^2 - (\beta_1 - \hat{\beta}_1)^2 \sum_{i=1}^{n}(X_i - \bar{X})^2$

ដាក់កន្សោមចុងក្រោយនេះជាសង្ឃឹមគណិត យើងបាន៖

$$\mathrm{E}\left[\sum_{i=1}^{n} e_i^2\right] = E\left[\sum_{i=1}^{n}(e_i - \bar{e})^2\right] - E\left[(\beta_1 - \hat{\beta}_1)^2\right] \sum_{i=1}^{n}(X_i - \bar{X})^2$$

យើងសិក្សាកន្សោមនេះជាពីករណី៖

i. ករណី $E\left[\sum_{i=1}^{n}(e_i - \bar{e}\,)^2\right] = E\left[\sum_{i=1}^{n}(e_i^2 - 2\bar{e}e_i + \bar{e}^2)\right] = E\left[\sum_{i=1}^{n} e_i^2 - 2\bar{e}\sum_{i=1}^{n} e_i + \sum_{i=1}^{n} \bar{e}^2\right]$

$$= E\left[\sum_{i=1}^{n} e_i^2 - 2\bar{e}n\bar{e} + n\bar{e}^2\right] = E\left[\sum_{i=1}^{n} e_i^2 - 2n\bar{e}^2 + n\bar{e}^2\right] = E\left[\sum_{i=1}^{n} e_i^2 - n\bar{e}^2\right] =$$

$$= E\left[\sum_{i=1}^{n} e_i^2 - \bar{e}\sum_{i=1}^{n} e_i\right] = E\left[\sum_{i=1}^{n} e_i^2 - \frac{\left[\sum_{i=1}^{n} e_i\right]^2}{n}\right] = \left[\sum_{i=1}^{n} E\left(e_i^2\right) - \frac{1}{n}E\left(\sum_{i=1}^{n} e_i\right)^2\right]$$

ជម្រើនងាយ

Simple Linear regression

យើងដឹងថា $E\left(e_i^2\right) = \sigma^2 \Rightarrow \sum_{i=1}^{n} E\left(e_i^2\right) = n\sigma^2$ និង $\frac{1}{n}E(\sum_{i=1}^{n} e_i)^2 = \frac{1}{n}E(e_1 + e_2 + \cdots + e_i)^2 = \frac{1}{n}n\sigma^2$

$\Rightarrow E[\sum_{i=1}^{n}(e_i - \bar{e})^2] = n\sigma^2 - \frac{1}{n}n\sigma^2 = (n-1)\sigma^2.$

ii. យើងមានកន្សោម $Var\left(\hat{\beta}_1\right) = E(\hat{\beta}_1 - \beta_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$ (ស្រាយបញ្ជាក់ខាងលើ)

$\Rightarrow E(\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma^2$

ដូច្នេះយើងបាន $\mathrm{E}\left[\sum_{i=1}^{n} e_i^2\right] = (n-1)\sigma^2 - \sigma^2 = (n-2)\sigma^2 \Rightarrow \sigma_e^2 = \frac{E\left[\sum_{i=1}^{n} e_i^2\right]}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$.

នៅចុងបញ្ចប់យើងបានវ៉ារ្យង់របស់ស្វ័យលំអៀង$\boldsymbol{u_i}$គឺ $\boldsymbol{\hat{\sigma}^2 = \hat{\sigma}_e^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2}$ ។

ជម្រើនងាយ

Simple Linear regression

- នៅក្នុងការអនុវត្តន៍យើងជំនួសវ៉ារ្យង់របស់មេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍នីមួយៗទៅជា៖

$$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\hat{\boldsymbol{\beta}}_{\mathbf{1}}} = \frac{\hat{\sigma}_e^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{1}{n-2} \times \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \Rightarrow sd\colon \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_{\mathbf{1}}} = \sqrt{\frac{1}{n-2} \times \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

$$\hat{\boldsymbol{\sigma}}^{\mathbf{2}}_{\hat{\boldsymbol{\beta}}_{\mathbf{0}}} = \hat{\sigma}_e^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \frac{\mathbf{1}}{\boldsymbol{n-2}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{e_i^2}\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) \rightarrow$$

$$\Rightarrow sd\colon \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_{\mathbf{0}}} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n-2}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{e_i^2}\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)}$$

ជម្រើនងាយ

Simple Linear regression

- សម្មតិកម្មភាពធម្មតានៃលំអៀងបង្ហាញថា៖ $\frac{\hat{\beta}_1-\beta_1}{\sigma_{\hat{\beta}_1}}$ និង $\frac{\hat{\beta}_0-\beta_0}{\sigma_{\hat{\beta}_0}}$ គោរពតាម *ច្បាប់ណ័រម៉ាល* (Normal Law) បង្រួមកណ្ដាលមធ្យមស្មើសូន្យ $N(0,1)$។

- យើងមាន $\hat{\sigma}_e^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 \Rightarrow \sum_{i=1}^{n} e_i^2 = (n-2)\hat{\sigma}_e^2 \Rightarrow \boldsymbol{\frac{\sum_{i=1}^{n} e_i^2}{\sigma_e^2} = (n-2)\frac{\hat{\sigma}_e^2}{\sigma_e^2}}$

កន្សោមផលធៀបនេះគឺជាច្បាប់ χ^2 (Chi-Square) មានកម្រិតសេរី(degree of freedom) $(n-2)$(ផលបូកការ៉េនៃ $(n-2)$ អថេរចៃដន្យឯករាជ្យ *ច្បាប់ណ័រម៉ាលបង្រួមកណ្ដាលមធ្យម*) ដោយសារគំរូវិភាគមានប៉ារ៉ាម៉ែត្រព្យាករណ៍តែពីរ β_0និង β_1។

ជម្រើនងាយ

Simple Linear regression

- យើងកត់សម្គាល់ឃើញថា $(n-2)\frac{\hat{\sigma}_e^2}{\sigma_e^2} = (n-2)\frac{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2}{\hat{\sigma}_{\hat{\beta}}^2 \sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{\sum_{i=1}^{n} e_i^2}{\hat{\sigma}_{\hat{\beta}}^2 \sum_{i=1}^{n}(X_i-\bar{X})^2}$

  ដូច្នេះកន្សោមនេះជាច្បាប់ χ^2 (Chi-Square) មានកម្រិតសេរី $(n-2)$ ដែរ។ ជាលទ្ធផលនៅក្នុង

  ការអនុវត្តន៍កន្សោម ៖ $\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\hat{\beta}_1}}$ និង $\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\hat{\beta}_0}}$ គោរពតាមច្បាប់*Student* មាន $(n-2)$ កម្រិតសេរី។
- ឥឡូវនេះយើងអាចរៀបចំការធ្វើតេស្តស្ថិតិបានដើម្បីផ្តល់ចម្លើយចំពោះបញ្ហាដូចជា៖
  - ✓ ការប្រៀបធៀបនៃមេគុណតំរែតំរង់ទៅនឹងមេគុណប៉ារ៉ាម៉ែត្រថេរមួយដែលគេឲ្យ។
  - ✓ ការប្រៀបធៀបមេគុណតំរែតំរង់របស់គំរូពីរផ្សេងគ្នា។
  - ✓ កំណត់ចន្លោះជឿជាក់សម្រាប់មេគុណ។

ដំណើរការ

Simple Linear regression

- **ការធ្វើតេស្តទ្វេភាគ**(Two-tailed tests)

ត្រូវសាកល្បងធ្វើតេស្តចាប់ផ្តើមជាមុនពីហានិភ័យ ៥%, ហើយសម្មតិកម្មត្រូវវាយតម្លៃរវាង៖

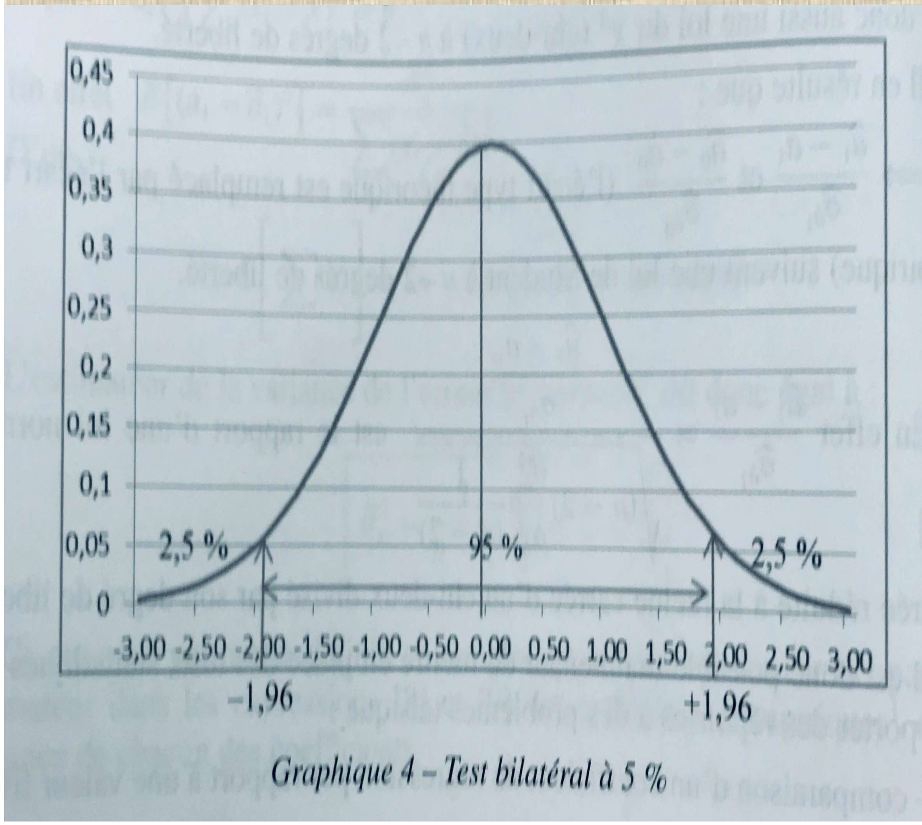
$$Hypothesis\ \boldsymbol{H_0}: \beta_1 = 0$$

$$Hypothesis\ \boldsymbol{H_1}: \beta_1 \neq 0$$

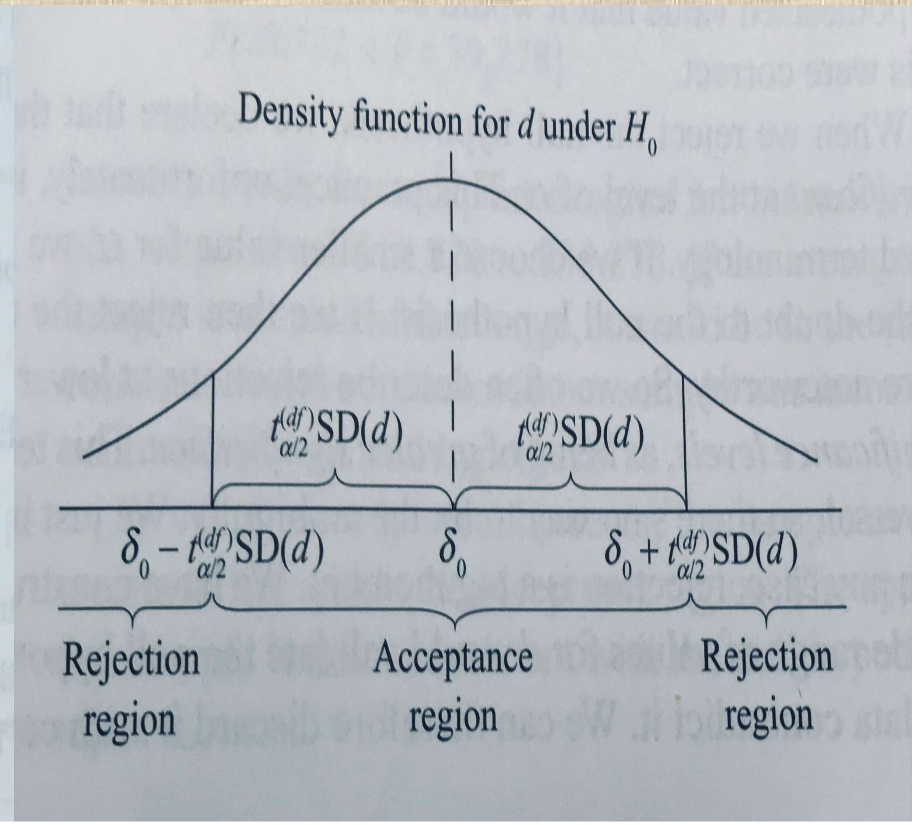
យើងដឹងថា $\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}}$ ជាច្បាប់*Student* មាន $(n-2)$ កម្រិតសេរី។ ករណីបើ $\boldsymbol{H_0}: \beta_1 = 0$ ផលធៀប $\frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}}$ ហៅថា *ratio student* ដែលជាច្បាប់*Student* មាន $(n-2)$ កម្រិតសេរី។ ដូច្នេះការធ្វើតេស្តទ្វេភាគត្រូវរៀបចំតាម *ratio student* ក្នុងតម្លៃពិន្ទុ $t - stat = t^* = \frac{|\hat{\beta}_1|}{\hat{\sigma}_{\hat{\beta}_1}}$ ទៅនឹងពិន្ទុ $t -$ *ratio student* អានក្នុងតារាងមាន $(n-2)$ កម្រិតសេរី ហើយចាប់ផ្តើមពីកម្រិតហានិភ័យ *៥%*។ ប្រសិនបើ $n - 2 > 30$, $t_{+\infty}^{5\%/2} = 1.96$ ។ ប្រសិនបើ $t^* > t_{+\infty}^{5\%/2} = 1.96$ នាំឲ្យយើងបដិសេធសម្មតិកម្ម $\boldsymbol{H_0}$ ហើយមេគុណដែលស្គាល់ $\hat{\beta}_1$ មានអត្ថន័យស្ថិតិគ្រប់គ្រាន់ខុសពី0។

ដម្រឿនងាយ

Simple Linear regression



Graphique 4 – Test bilatéral à 5 %



ជម្រើនងាយ

Simple Linear regression

- សំណង់អង្កត់ជឿជាក់(The construction of confidence intervals=CI)

យើងមាន $\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\hat{\beta}_1}}$ ជាច្បាប់*Student* មាន $(n-2)$ កម្រិតសេរី។ ប្រសិនបើ $\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\hat{\beta}_1}} = t_{(n-2)}^{\propto\%/2}$ អង្កត់ជឿជាក់អាចកំណត់បានតាមនក្រោម៖

$$\frac{\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1}{\hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1}} = \pm \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} \Rightarrow \left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} < \boldsymbol{\beta}_1 < \hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2}\right)$$ និង

$$\frac{\hat{\boldsymbol{\beta}}_0-\boldsymbol{\beta}_0}{\hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_0}} = \pm \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} \Rightarrow \left(\hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_0} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} < \boldsymbol{\beta}_0 < \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_0} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2}\right)$$

ជម្រើនងាយ

Simple Linear regression

ឧទាហរណ៍១ ការអង្កេតមួយទាក់ទងចំនួនឆ្នាំសិក្សានិងចំណូលដែលទទួលបានគិតUSDក្នុងមួយឆ្នាំ។ ការអង្កេតនេះផ្តោតចំពោះបុគ្គលដែលអាចទាក់ទងបាន២០នាក់។ រាល់ទិន្នន័យដែលបានផ្តល់ឲ្យសន្មត់អាចទុកចិត្តបាន។

សំណួរ

ក. កំណត់មេគុណជាប់ពាក់ព័ន្ធរវាងការសិក្សានិងចំណូល

ខ. សរសេរសមីការព្យាករណ៍និងបកស្រាយមេគុណប៉ារ៉ាម៉ែត្រដែលទទួលបាន។

គ. ធ្វើតេស្តមេគុណប៉ារ៉ាម៉ែត្រនិងកំណត់អង្កត់ជឿជាក់។

ឃ. តើអ្នកគិតថាសមីការនេះអាចទទួលយកបានឬទេ?

ទិន្នន័យអង្កេតទាក់ទងចំនួនឆ្នាំសិក្សានិងចំណូលដែលទទួលបានគិតUSDក្នុងមួយឆ្នាំ

Ob	Edu=Xi	Income=Yi
1	0	0
2	0	0
3	8	10,500
4	10	0
5	11	0
6	11	29,000
7	11	0
8	12	50,000
9	12	3,800
10	12	0
11	12	12,500
12	13	27,500
13	13	64,000
14	13	0
15	14	57,000
16	16	30,000
17	16	92,000
18	16	80,000
19	16	50,000
20	18	62,000

មុនដំបូងយើងត្រូវរៀបចំតារាងគណនាទិន្នន័យឲ្យបានរួចរាល់សិន បន្ទាប់មកទើបគណនាបន្តបន្ទាប់

Ob	Edu=Xi	$X_i - \bar{X}$	Income=Yi	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	0	-11.7	0	-28415	332455.5	136.89	807412225
2	0	-11.7	0	-28415	332455.5	136.89	807412225
3	8	-3.7	10,500	-17915	66285.5	13.69	320947225
4	10	-1.7	0	-28415	48305.5	2.89	807412225
5	11	-0.7	0	-28415	19890.5	0.49	807412225
6	11	-0.7	29,000	585	-409.5	0.49	342225
7	11	-0.7	0	-28415	19890.5	0.49	807412225
8	12	0.3	50,000	21585	6475.5	0.09	465912225
9	12	0.3	3,800	-24615	-7384.5	0.09	605898225
10	12	0.3	0	-28415	-8524.5	0.09	807412225
11	12	0.3	12,500	-15915	-4774.5	0.09	253287225
12	13	1.3	27,500	-915	-1189.5	1.69	837225
13	13	1.3	64,000	35585	46260.5	1.69	1266292225
14	13	1.3	0	-28415	-36939.5	1.69	807412225
15	14	2.3	57,000	28585	65745.5	5.29	817102225
16	16	4.3	30,000	1585	6815.5	18.49	2512225
17	16	4.3	92,000	63585	273415.5	18.49	4043052225
18	16	4.3	80,000	51585	221815.5	18.49	2661012225
19	16	4.3	50,000	21585	92815.5	18.49	465912225
20	18	6.3	62,000	33585	211585.5	39.69	1127952225
Sum=	234	0	568,300	0	1,684,990.00	416.20	17,682,945,500.00
Average=	11.70	0	28,415	0	88,684		
Variance=	21.90526316		930681342.1				
	4.680305883		30507.07036			20.40	132,977.24
	B*D=	142782.4209					2712865.997
	R=F24/(B*d)=	0.6211107				R=	0.621110664

ដំណោះស្រាយ

Simple Linear regression

ក. កំណត់មេគុណទំនាក់ទំនងរវាងការសិក្សានិងចំណូល

$$R = \frac{\sum_{i=1}^{20}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{20}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{20}(Y_i - \bar{Y})^2}} = \frac{1,684,990.00}{\sqrt{416.20}\sqrt{17,682,945,500.00}} = \frac{1,684,990.00}{20.40 \times 132,977.24} = 0.621$$

យើបានមេគុណទំនាក់ទំនងរវាងការសិក្សានិងចំណូល $R = 0.621$ ជាចំនួនវិជ្ជាមានមានន័យថាចំណូលបុគ្គលម្នាក់ៗជាអនុគន៍កើននៃការសិក្សា។

ខ. សមីការមេគុណព្យាករណ៍កំណត់ដោយៈ $\boldsymbol{incom_i = \beta_0 + \beta_1 edu_i + e_i}, i = 1,2,\ldots,20$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{20}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{20}(X_i - \bar{X})^2} = \frac{1,684,990.00}{416.20} = 4,048.51$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 28,415 - 4,048.51 \times 11.70 = -18,952.57$$

ជម្រើនងាយ

Simple Linear regression

បកស្រាយមេគុណប៉ារ៉ាម៉ែត្រព្យាករណ៍៖ បើយើងធ្វើដេរីវេអនុគមន៍ប្រាក់ចំណូលធៀបនឹងអថេរការសិក្សាគឺ $\widehat{incom} = -18952.57 + 4048.51(edu) \Rightarrow (incom)' = \hat{\beta}_1 = 4048.51 > 0$ មានន័យ ថាផលិតភាពចំណូលធៀបនឹងការសិក្សាកើនបាន 4048.51បើការសិក្សាកើនបន្ថែម១ឆ្នាំ។ ករ ណីចំនួនថេរ $\hat{\beta}_0 < 0$ អាចនិយាយបានថាត្រូវបង់ប្រាក់ 18952.57 ជាមុនសិនបន្ទាប់មកទើបអាចទទួលបានប្រាក់ឈ្នួលជាក្រោយ ប៉ុន្តែការជាក់ស្តែងពុំអាចអនុវត្តទៅកើតទេ។

គ. ធ្វើតេស្តមេគុណសំខាន់ដែលជាប់ពាក់ព័ន្ធនឹងការសិក្សា $\hat{\beta}_1$៖

ដំបូងគណនាវ៉ារ្យង់របស់មេគុណ $\hat{\beta}_1$,

$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2_e}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$, ហើយវ៉ារ្យង់លំអៀង $\hat{\sigma}^2_e = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{10861246076.406}{18} = 603402559.80$

ជម្រើនងាយ

Simple Linear regression

$\Rightarrow \hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2_e}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{603402559.80}{416.20} = 1{,}449{,}789.91 \Rightarrow \sigma_{\hat{\beta}_1} = 1{,}204.0722$

យើងបាន ratio-student: $t^* = \frac{|\hat{\beta}_1|}{\sigma_{\hat{\beta}_1}} = \frac{4048.51}{1{,}204.0722} = 3.362 > t_{18}^{5\%/2} = 2.101$ ជាលទ្ធផលគឺបដិសេធសម្មតិកម្ម Ho និងមេគុណ $\hat{\beta}_1$ មានអត្ថន័យស្ថិតិគ្រប់គ្រាន់ខុសពី0 ត្រង់ហានិភ័យ៥%។

កំណត់អង្កត់ជឿជាក់

យើងមាន $\frac{\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1}{\hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1}} = \pm \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} \Rightarrow \left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2} < \boldsymbol{\beta}_1 < \hat{\boldsymbol{\beta}}_1 + \hat{\boldsymbol{\sigma}}_{\hat{\boldsymbol{\beta}}_1} \times \boldsymbol{t}_{(\boldsymbol{n}-2)}^{\propto\%/2}\right)$

ជំនួសតម្លៃលេខ $(4048.51 - 1{,}204.0722 \times 2.101 < \beta_1 < 4048.51 + 1{,}204.0722 \times 2.101)$

យើងបានអង្កត់ជឿជាក់របស់មេគុណ edu គឺ $(\mathbf{1518.84} < \boldsymbol{\beta}_1 < \mathbf{6578.17})$ ។

គ. ការទទួលយកគំរូវិភាគនេះត្រូវមានការប្រុងប្រយ័ត្នពីព្រោះមេគុណប៉ារ៉ាម៉ែត្រថេរជាចំនួនអវិជ្ជមាន។ កត្តានេះនាំឲ្យមានការបកស្រាយមិនបានច្បាស់លាស់ជាពិសេសការកំណត់អថេរពន្យល់និងសមីការគំរូវិភាគ ឬការកំណត់រង្វាស់ទិន្នន័យមានភាពលំអៀងជាដើម។

ជម្រើនងាយ

Simple Linear regression

- ឧទាហរណ៍២. ផលិតផលអាហារថ្មីមួយប្រភេទដែលទើបនឹងចូលប្រកួតក្នុងទីផ្សារដំបូងមានតម្រូវការទាបនៅឡើយ។ អ្នកទីផ្សារចាប់ផ្តើមប្រើយុទ្ធសាស្ត្រថ្លៃនិងធ្វើការអង្កេតលទ្ធផលនៃការលក់ចំនួន១៥សប្តាហ៍ ទិន្នន័យបានបង្ហាញមកយ៉ាងនេះ។

សំណួរ

ក. កំណត់មេគុណជាប់ពាក់ព័ន្ធរវាងតម្រូវការនិងថ្លៃ

ខ. សរសេរសមីការតម្រូវការផលិតផលជាអនុគមន៍ថ្លៃ និងបកស្រាយមេគុណប៉ារ៉ាម៉ែត្រថ្លៃដែលទទួលបាន រួចគណនាភាពយឺតនៃតម្រូវការធៀបនឹងថ្លៃ។

គ. ធ្វើតេស្តមេគុណប៉ារ៉ាម៉ែត្រថ្លៃ និងកំណត់អង្កត់ជឿជាក់។

ឃ. តើអ្នកគិតថាអគមន៍តម្រូវការនេះអាចទទួលយកបានឬទេ?

- ទិន្នន័យ

Obser	Quantity	Price
1	3	18
2	3	16
3	7	17
4	6	12
5	10	15
6	15	15
7	16	4
8	13	13
9	9	11
10	15	6
11	9	8
12	15	10
13	12	7
14	18	7
15	21	7

ជម្រើនងាយ

Simple Linear regression

- អនុគមន៍ប្រើប្រាស់(Consumption function)

  ❖ អនុគមន៍មានប្រភពចេញមកពីទ្រឹស្តី Keynes។ អនុគមន៍ប្រើប្រាស់របស់ទ្រឹស្តី Keynes អាចសង្ខេបយកតែចំនុចសំខាន់ចំបាច់ខ្លះដែលទាក់ទងនឹងការវិភាគជាមួយសមីការព្យាករណ៍តាមវិធីសេដ្ឋកិច្ចមាត្រសាស្រ្ត៖

  1. ការប្រើប្រាស់ជាក់ស្តែងគឺជាអនុគមន៍មានស្ថេរភាពសមធម៌(fairly stable)នៃចំណូលពិតបច្ចុប្បន្ន i.e $C = f(Y)$, ដែល C និង Y គឺជាការប្រើប្រាស់ជាក់ស្តែងនិងចំណូលជាក់ស្តែងរបស់អ្នកប្រើប្រាស់ម្នាក់ៗ,
  2. ទំនោរលំអៀងនៃការប្រើប្រាស់(MPC)នៅចន្លោះ ០និង១ i.e $0 = \frac{\partial C}{\partial Y} < 1.$
  3. មេគុណ MPC តូចជាងទំនោរមធ្យមនៃការប្រើប្រាស់(APC), i.e $\frac{\partial C}{\partial Y} < \frac{C}{Y}$.
  4. នៅពេលប្រាក់ចំណូលពិតកើនឡើង MPC ធ្លាក់ចុះ។

ដម្រឿនងាយ

Simple Linear regression

ឧទាហរណ៍ ជាមួយទិន្នន័យនេះ ចូរឆ្លើយសំណួរ៖

1. សរសេរអនុគមន៍នៃការប្រើប្រាស់ពាន់ព័ន្ធនឹងចំណូលបុគ្គលនិងបកស្រាយអត្ថន័យ។
2. គណនាមេគុណព្យាករណ៍របស់សមីការ
3. ធ្វើតេស្តមេគុណត្រង់ហានិភ័យ ៥%។
4. កំណត់អង្កត់ជឿជាក់

- គេឲ្យ

Year	C	INC
1992	931979.3	1161063.2
1993	940104.7	1149570.9
1994	950702.3	1173503.5
1995	963526.4	1201128.0
1996	980879.4	1214274.2
1997	989912.0	1243432.7
1998	1015774.4	1287787.5
1999	1046365.2	1330917.8
2000	1079458.9	1385073.3
2001	1105375.6	1413487.4
2002	1132562.2	1430752.1
2003	1150003.4	1442220.8
2004	1175804.2	1475653.4

ក្រាភិកបកស្រាយ Keynesian' consumption function