ជំពូកទី៥ សមីការវ៉ារ្យង់

ANOVA

រៀបរៀងនិងបង្រៀនដោយបណ្ឌិត ង៉ាន់ ស៊ុនដេត

សាស្ត្រាចារ្យនៃសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រនិងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច

សមីការវ៉ារ្យង់
Equation of Variance

- ស្រាយកន្សោមដូចជាមុនៈ **1.** $\boldsymbol{\sum e_i = 0}$, **2.** $\boldsymbol{\bar{Y} = \bar{\hat{Y}}}$, **3.** $\boldsymbol{\sum X_i e_i = 0}$, **4.** $\boldsymbol{\sum \hat{Y}_i e_i = 0}$

1. $\boldsymbol{\sum e_i = 0}$

   មានសមីការព្យាករណ៍ $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \Rightarrow \sum Y_i = \sum \hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \sum e_i \rightarrow$

$$\sum Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum X_i = \sum e_i \quad (1)$$

   ហើយជំនួសតម្លៃ $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \Rightarrow \hat{\beta}_0 = \frac{1}{n}\sum Y_i - \hat{\beta}_1 \frac{1}{n}\sum X_i$ ក្នុង (1)

   $\Rightarrow \sum Y_i - \sum Y_i + \hat{\beta}_1 \sum X_i - \hat{\beta}_1 \sum X_i = \boldsymbol{\sum e_i = 0 \Rightarrow \bar{e} = 0}$ ។

2. $\boldsymbol{\bar{Y} = \bar{\hat{Y}}}$

   យើងមាន $Y_i = \hat{Y}_i + e_i \Rightarrow Y_i - \hat{Y}_i = e_i \Rightarrow \sum Y_i - \sum \hat{Y}_i = \sum e_i = 0 \Rightarrow \boldsymbol{\bar{Y} = \bar{\hat{Y}}}$ ។

សមីការវ៉ារ្យង់
Equation of Variance

3. $\sum \boldsymbol{X_i e_i = 0}$ ជំនួសតម្លៃ $e_i = Y_i - \beta_0 - \beta_1 X_i$ យើងបាន

$$\sum X_i e_i = \sum X_i (Y_i - \beta_0 - \beta_1 X_i) = \sum X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum X_i^2$$

យើងមានលក្ខខណ្ឌលំដាប់មួយនៃផលបូកការេតម្លៃតូច $\frac{\partial RSS}{\partial \beta_1} = 0$

$\frac{\partial R}{\partial \beta_1} = -2\sum X_i Y_i + 2\beta_0 \sum X_i + 2\beta_1 \sum X_i^2 = 0$ ចែកអង្គទាំងពីរនឹង - 2 យើងបាន

$\sum X_i e_i = \sum X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum X_i^2 = 0$ ។

4. $\sum \boldsymbol{\widehat{Y}_i e_i = 0}$ យើងមាន $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \Rightarrow$

$\sum \hat{Y}_i e_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 X_i)\, e_i = \sum \hat{\beta}_0 e_i + \sum \hat{\beta}_1 X_i e_i = \hat{\beta}_0 \sum e_i + \hat{\beta}_1 \sum X_i\, e_i = 0$ ។

សមីការវ៉ារ្យង់
Equation of Variance

- សមីការវ៉ារ្យង់

យើងមាន $Y_i = \hat{Y}_i + e_i \leftrightarrow Y_i - \bar{Y} = \hat{Y}_i - \bar{\hat{Y}} + e_i \Rightarrow (Y_i - \bar{Y})^2 = \left((\hat{Y}_i - \bar{\hat{Y}}) + e_i\right)^2$

ពន្លាតកន្សោម និងធ្វើផលបូកយើងបានលទ្ធផលជាបន្តបន្ទាប

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

$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{\hat{Y}})^2 + \sum e_i^2 + 2\sum \hat{Y}_i e_i - 2\sum \bar{\hat{Y}}\, e_i$ ពីព្រោះ $\sum \hat{Y}_i e_i = 0, and\ \bar{\hat{Y}} \sum e_i = 0$

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

តាង TSS= $\sum(Y_i - \bar{Y})^2$, ESS= $\sum(\hat{Y}_i - \bar{\hat{Y}})^2$, RSS= $\sum e_i^2$

នៅទីបញ្ចប់ យើងបានសមីការវ៉ារ្យង់គឺ *TSS = ESS + RSS*

សមីការវ៉ារ្យង់
Equation of Variance

- ការវិភាគវ៉ារ្យង់(ANOVA)

យើងមានសមីការវ៉ារ្យង់ *TSS = ESS + RSS* បើចែកអង្គទាំងពីរនឹង TSS ផល

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

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

R^2 ហៅថាមេគុណទំនាក់ទំនង ហើយ *R* ជាមេគុណការជាប់ពាន់ព័ន្ធរវាងគូរ (X_i, Y_i)។

សមីការវ៉ារ្យង់
Equation of Variance

❑ តារាងវិភាគវ៉ារ្យង់

Source of variable ប្រភពអថេរ	**Sum of Square** ផលបូកការេ	**Degree of freedom (df)** កម្រិតសេរី	**Mean Square (MS)** ការេមធ្យម
X	**ESS** $= \sum(\hat{Y}_i - \bar{\bar{Y}})^2$	1	$\frac{\text{ESS}}{\mathbf{1}}$
Residual	**RSS** $= \sum e_i{}^2$	n – 2	$\frac{\text{RSS}}{\boldsymbol{n-2}}$
Total	**TSS** $= \sum(Y_i - \bar{Y})^2$		

សមីការវ៉ារ្យង់
Equation of Variance

- ការធ្វើតេស្ត $H_0 : \hat{\beta}_1 = 0$ មានន័យថាជាការធ្វើតេស្ត $H_0 : ESS = 0$ អថេរពន្យល់ X_i មិនមាននៅក្នុងនៅក្នុងគំរូវិភាគទេ។
- ផ្ទុយមកវិញ បើបដិសេធន៍ $H_0: ESS = 0$ មានន័យថា សម្មតិកម្ម $H_1: ESS \neq 0$ ស្ថិតិនៃការធ្វើតេស្តគំរូវិភាផ្តល់ឲ្យតាមតម្លៃពិន្ទុ Fisher(Fisher Score)៖

$$F^* = \frac{\frac{ESS}{df_{ESS}}}{\frac{RSS}{df_{RSS}}} = \frac{\frac{\sum(\hat{Y}_i - \bar{\hat{Y}})^2}{1}}{\frac{\sum e_i^2}{n-2}} \Rightarrow \boldsymbol{F^*} = \frac{\frac{ESS}{df_{ESS}}}{\frac{RSS}{df_{RSS}}} = \frac{\frac{ESS}{1}}{\frac{TSS-ESS}{n-2}} = \frac{\frac{(\frac{ESS}{TSS})}{1}}{\frac{(TSS-ESS)/TSS}{n-2}} = \frac{\frac{\boldsymbol{R^2}}{\boldsymbol{1}}}{\frac{\boldsymbol{1-R^2}}{\boldsymbol{n-2}}}$$

- ស្ថិតិ $\boldsymbol{F^*}$ គឺជាផលធៀបនៃមធ្យមការេរបស់អថេរពន្យល់និងមធ្យមការេរបស់សំណល់។

សមីការវ៉ារ្យង់
Equation of Variance

- អថេរ X_i គឺមានសារៈសំខាន់ជាអថេរពន្យល់ជាក់លាក់។
- ***F^**** គឺជាស្ថិតិរបស់ Fisher មានកម្រិតសេរីពី 1 ដល់ n – 2 ។ បើ $F^* > F_{1;n-2}^{\propto\%}$ យើងបដិសេធសម្មតិកម្ម H_0 ត្រង់ហានិភ័យ $\propto$ % ហើយអថេរ X_i មានអត្ថន័យស្ថិតិ។ ករណីផ្ទុយមកវិញបើមិនអាចបដិសេធសម្មតិកម្ម H_0 អថេរ X_i មិនមានការពន្យល់ចំពោះអថេរ Y_i ។
- យើងអាចធ្វើសេចក្តីសន្និដ្ឋានថា $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} \Rightarrow \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$ នាំឲ្យកន្សោម

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i - \hat{\beta}_0 - \hat{\beta}_1\bar{X} = \hat{\beta}_1(X_i - \bar{X}) \Rightarrow \sum\left(\hat{Y}_i - \bar{Y}\right)^2 = \hat{\beta}_1^2 \sum(X_i - \bar{X})^2$$

សមីការវ៉ារ្យង់
Equation of Variance

- ដូច្នេះ $F^* = \frac{\frac{\sum(\hat{Y}_i-\bar{\hat{Y}})^2}{1}}{\frac{RSS}{n-2}} = \frac{\frac{\hat{\beta}_1^2 \sum\left(X_i-\bar{X}\right)^2}{1}}{\frac{\sum e_i^2}{n-2}}$ ។ ម្យ៉ាងទៀតយើងមាន $t^* = \frac{R}{\sqrt{\frac{1-R^2}{n-2}}}$ លើកជាការេយើងបាន $\boldsymbol{(t^*)^2} = \frac{R^2}{\frac{1-R^2}{n-2}} = \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}} = \boldsymbol{F^*} = \frac{\frac{\hat{\beta}_1^2 \sum(X_i-\bar{X})^2}{1}}{\frac{\sum e_i^2}{n-2}} \Rightarrow \boldsymbol{F^* = (t^*)^2}$ ។
- គួរកត់សម្គាល់ដែរថា តម្លៃស្ថិតនៃ $\boldsymbol{F^*}$ កាន់តែធំគំរូព្យាករណ៍កាន់តែមានភាពប្រសើរ។

សមីការវ៉ារ្យង់
Equation of Variance

- ម្យ៉ាងទៀតបើយើងយកកន្សោម $\frac{RSS}{TSS} = \frac{\frac{RSS}{n-2}}{\frac{TSS}{n-1}} \Rightarrow \frac{RSS}{TSS} = \frac{Var(e_i)}{Var(Y_i)}$
- យើងមានកន្សោម $1 = \frac{ESS+R}{TSS} = \frac{ESS}{TSS} + \frac{RSS}{TSS} \Rightarrow \frac{ESS}{TSS} \cong 1 - \frac{Var(e_i)}{Var(Y_i)} \Rightarrow \bar{R}^2 = 1 - \frac{\frac{RSSS}{n-2}}{\frac{TSS}{n-1}} \rightarrow$

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

- ដូច្នេះយើងបាន $\bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-2}$. ។ តម្លៃមេគុណ $\boldsymbol{\bar{R}^2}$ ហៅថា R^2 កែតម្រូវ។
- ជាទូទៅ $\boldsymbol{\bar{R}^2} < R^2$ ជានិច្ច, ប៉ុន្តែវាអាច $\boldsymbol{\bar{R}^2} = R^2$ កាលណាចំនួនអង្កេត $\boldsymbol{n}$ ធំគ្រប់គ្រាន់។

សិក្សាឧទាហរណ៍

- ឧទាហរណ៍២៖ ក្សេតវិទូបានសិក្សាដី៨៥ក្បាលលើទិន្នផលពោត y_i មានការពាក់ព័ន្ធជាមួយអត្រានៃជាតិរ៉ែ bauxite x_i ដែលមានក្នុងដី។ លទ្ធផលបានបង្ហាញតាមសមីការវិភាគខាងក្រោម៖

$$y_i = 132.80 - 1.1x_i + e_i,\ \ i = 1,2, \dots 85$$

(4.3) (10.2)

(.): ratio of Student, $\sum e^2 = 6\ 234.32$

សំណួរ ក. បង្ហាញថា តេស្តសម្មតិកម្មH_0: $\beta_1 = 0$ បានមកពីតេស្តសម្មតិកម្ម $R = 0$ ដែល R ជាមេគុណនៃការជាប់ទាក់ទងលីនេអែររវាង x_iនិង y_i ដែលត្រូវគណនា។ ខ. បង្កើតតារាងវិភាគវ៉ារ្យង់និងផ្ទៀងផ្ទាត់លទ្ធផលដែលទទួលបានក្នុងសំណួរ ក. ចេញពីតេស្ត Fisher។ គ. គណនាមេគុណកែតម្រូវ $\bar{R}^2$ ។

ចម្លើយ

ក. បង្ហាញតេស្តសម្មតិកម្ម H_0: $R = 0$, ដោយហេតុថា $t^* = \frac{|\hat{\beta}_1|}{sd_{\hat{\beta}_1}} \Rightarrow sd_{\hat{\beta}_1} = \frac{|\hat{\beta}_1|}{t^*} = \frac{|-1.1|}{10.2} = 0.107$

យើងមាន $R = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}} \Rightarrow R^2 = \frac{\hat{\beta}_1 \sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} = \frac{ESS}{TSS}$

សិក្សាឧទាហរណ៍

ហើយមេគុណ $\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \Rightarrow$

$$\hat{\beta}_1 \sum(X_i - \bar{X})(Y_i - \bar{Y}) = \hat{\beta}_1 \times \hat{\beta}_1 \sum (X_i - \bar{X})^2 = \hat{\beta}_1{}^2 \sum (X_i - \bar{X})^2 = \sum(\hat{Y}_i - \bar{Y})^2 = ESS$$

សម្រាប់គំរូជម្រៀនងាយ យើងបានស្រាយបញ្ជាក់រួចមកហើយតម្លៃស្ថិតិ Fisher,

$$F^* = \frac{R^2}{(1-R^2)/(n-2)} = (t^*)^2 \Rightarrow t^* = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \Rightarrow 10.2 = \frac{R\sqrt{85-2}}{\sqrt{1-R^2}},$$

ក្រោយគណនា យើងបាន $R^2 = 0.556 \Rightarrow |R| = 0.745$

នេះនាំឲ្យយើងធ្វើតេស្តបានថា បើទំនាក់ទំនងរវាង x_i និង y_i ជាវិជ្ជមាន ឬបើអវិជ្ជមាន មេគុណនៃការជាប់ពាក់ព័ន្ធមានអត្ថន័យស្ថិតិខុសពី0 ត្រង់ហានិភ័យ ៥%។

ដូច្នេះ មេគុណ $R = 0.745 \neq 0$ ។

សិក្សាឧទាហរណ៍

ខ. តារាងវ៉ារ្យង់

ដើម្បីសង់តារាងវិភាគវ៉ារ្យង់យើងគម្បីស្គាល់តម្លៃនៃកម្រោមវិភាគនីមួយៗ

$ESS = \sum(X_i - \bar{X})^2$, $RSS = \sum e_i^2$, $TSS = \sum(Y_i - \bar{Y})^2$ and $R^2 = 1 - \frac{RSS}{TSS}$, យើងបានដឹងខ្លះ

$R^2 = 0.556, RSS = \sum e_i^2$=6234.32, $TSS = \frac{RSS}{1-R^2} = \frac{6234.32}{1-0.556} = 14041.26,$
$ESS = TSS - RSS = 7806.94$

តារាងវ៉ារ្យង់បង្កើតបានខាងក្រោម៖

Source of variable	Sum Square	Degree of Freedom	Mean Square
X_i	ESS = 7806.94	1	7806.94
Residual	RSS = 6234.32	85 - 2	75.11
Total	TSS = 14041.26	85 - 1	

សិក្សាឧទាហរណ៍

- យើងកត់សម្គាល់ឃើញថា

$$F^* = \frac{ESS/1}{RSS/(n-2)} = \frac{7806.94/1}{6234.32/83} = 103.94 > F^{5\%}_{1;83} = 3.96$$

ហើយយើងក៏សម្គាល់ដែរថា $F^* = (t^*)^2 = 10.2^2 \approx 104$

ដូច្នេះ នៅក្នុងគំរូព្យាករណ៍ងាយ ឬជម្រឿនងាយ វាមានការធ្វើតេស្តសម្មតិកម្មទូទៅមួយគឺ

Ho	H1
$\beta_1 = 0$	$\beta_1 \neq 0$
$R_{x,y} = 0$	$R_{x,y} \neq 0$
$ESS = 0$	$ESS \neq 0$

តេស្តទី១ គឺទៅលើមេគុណបន្ទាត់នៃជម្រឿន តេស្តទី២ គឺតេស្តលើមេគុណទំនាក់ទំនងរវាង x_i និង y_i និងចុងបញ្ចប់វាយតម្លៃអត្ថន័យស្ថិតិរបស់កិច្ចការទាំងស្រុង។

សិក្សាឧទាហរណ៍

គ. មេគុណកំណត់ទំនាក់ទំនងកែតម្រូវ $\bar{R}^2$

យើងមាន $\bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-2} = 1 - (1 - 0.556)\frac{85-1}{85-2} = 0.550$

ដូច្នេះ តម្លៃមេគុណ $\bar{R}^2 = 0.550 < R^2 = 0.556$