

ជំពូក១ : ប្រូបាប៊ីលីតេ

សង្ខេបនៃប្រូបាប៊ីលីតេ

ប្រូបាប៊ីលីតេគឺជាផ្នែកមួយយ៉ាងសំខាន់ដែលផ្សារភ្ជាប់ស្ថិតិអធិប្បាយ (descriptive statistics) ទៅនឹងស្ថិតិវិភាគ (statistical analysis) ។

ការសិក្សាទៅលើ descriptive statistics ផ្តល់នូវការរួមចំណែកយ៉ាងធំធេងក្នុងការបង្កើត mathematics models ដើម្បីសិក្សាទៅលើបាតុភូតចៃដន្យ ដែលតែងតែកើតឡើងនៅក្នុងជីវភាពរស់នៅរបស់យើង ។ លក្ខណៈចំបងនៃបាតុភូតចៃដន្យទាំងនេះត្រូវបានឆ្លើយតបទៅនឹងការគណនានៃ ប្រូបាប៊ីលីតេដែលបានរីកចំរើនយ៉ាងលឿននៅក្នុងគណិតវិទ្យា ។

ការសិក្សាទៅលើមូលដ្ឋានចំបងនៃទ្រឹស្តី ប្រូបាប៊ីលីតេបានកើតឡើងនៅក្នុងសតវត្សទី XVI-XVII ដោយលោក Kardana, Ferma, Pascal និងអ្នកប្រាជ្ញមួយចំនួនទៀត ។

១- តើប្រូបាប៊ីលីតេសិក្សាអំពីអ្វី?

ប្រូបាប៊ីលីតេមានតួនាទីយ៉ាងសំខាន់ក្នុងការសិក្សាទៅលើបាតុភូតចៃដន្យក៏ដូចជាការសិក្សាតាមដានទៅលើព្រឹត្តិការណ៍ ។

យើងចែកព្រឹត្តិការណ៍ ជា 3 ផ្នែកធំ ៗ ៖

- ព្រឹត្តិការណ៍ពិតប្រាកដ៖

ជាព្រឹត្តិការណ៍មួយដែលចាំបាច់ត្រូវតែកើតឡើងក្រោមលក្ខខណ្ឌមួយច្បាស់លាស់ ។

ឧទាហរណ៍ ៖

ទឹកនៅក្នុងសម្ពាធបរិយាកាសធម្មតា កកនៅ 0°C ។ សម្ពាធបរិយាកាសធម្មតា ជាលក្ខខណ្ឌច្បាស់លាស់ចំណែកឯ កកនៅ 0°C ជាព្រឹត្តិការណ៍ពិតប្រាកដ ។

- ព្រឹត្តិការណ៍មិនអាចមានឬមិនពិត៖

ក្រោមលក្ខខណ្ឌមួយច្បាស់លាស់ព្រឹត្តិការណ៍នេះមិនអាចកើតមានឡើងទេ ។

ឧទាហរណ៍:

នៅក្នុងសម្ពាធបរិយាកាសធម្មតា ទឹក កក នៅ 20°C ។ លក្ខខណ្ឌរបស់យើងនៅពេលនេះគឺ សម្ពាធបរិយាកាសធម្មតា ចំណែក ទឹក កកនៅ 20°C ជាព្រឹត្តិការណ៍មិនពិត ។

- ព្រឹត្តិការណ៍ចែងន្យ:

ជាព្រឹត្តិការណ៍មួយដែលកើតឡើងហើយដែលយើងពុំអាចកំណត់ទុកនូវលទ្ធផលរបស់វាជាមុនបាន ។

ឧទាហរណ៍ :

នៅក្នុងការបោះគ្រាប់ឡូតឡាក់ យើងពុំអាចកំណត់ជាមុនបាននូវលទ្ធផលរបស់វា ។ ព្រឹត្តិការណ៍ដែលនឹងកើតឡើងជាព្រឹត្តិការណ៍ចែងន្យ ។

ដូច្នេះយើងអាចនិយាយថា ប្រូបាប៊ីលីតេសិក្សាទៅលើច្បាប់បំបែកនៃព្រឹត្តិការណ៍ដែលនឹងកើតមានឡើងនៅក្នុងការពិសោធន៍ ។

វិញ្ញាសា (ការពិសោធន៍) និងព្រឹត្តិការណ៍ចែងន្យ:

និយមន័យ:

វិញ្ញាសាជាទង្វើមួយដែលគេ ពុំអាចកំណត់ច្បាស់លាស់ជាមុននូវលទ្ធផលរបស់វា ។

ឧទាហរណ៍ :

នៅក្នុងថង់មួយមានកូនឃ្លីពណ៌ជាច្រើន ។ គេដកកូនឃ្លីមួយពីក្នុងថង់នោះ ។ ការដកកូនឃ្លីចេញពីក្នុងថង់ជា **វិញ្ញាសា** ចំណែកការកំណត់ពណ៌សំបុររបស់កូនឃ្លីដែលដកចេញមកនេះជា **ព្រឹត្តិការណ៍ចែងន្យ** ។

ព្រឹត្តិការណ៍មិនចុះសំរុងគ្នា:

គេនិយាយថាព្រឹត្តិការណ៍ពីរជាព្រឹត្តិការណ៍មិនចុះសំរុងគ្នាកាលណាព្រឹត្តិការណ៍ទាំងពីរនេះពុំអាចកើតឡើងព្រមពេលជាមួយគ្នាបានឡើយ ។

ឧទាហរណ៍:

នៅក្នុងការបោះកាក់មួយដងយើងអាចទទួលបានឬមួយផ្នែក រូប ឬមួយផ្នែកខាងលេខ ។ បានសេចក្តីថា រូប និងលេខពុំអាចកើតឡើងព្រមពេលជាមួយគ្នាទេ ។ ព្រឹត្តិការណ៍ទាំងពីរនេះជាព្រឹត្តិការណ៍មិនចុះសំរុងគ្នា ។

ចំណាំ:

គេតែងតាងព្រឹត្តិការណ៍ដោយអក្សរ A,B,C,D,... ។

តាមន័យរបស់គណិតវិទ្យា បើ A និង B ជាព្រឹត្តិការណ៍ពីរមិនចុះសំរុងគ្នា នោះ:

$$A \cap B = \varnothing$$

សាកលៈ

សាកល Ω គឺជាសំនុំលទ្ធផលដែលអាចកើតមានឡើងនៅក្នុងការពិសោធន៍ (នៃវិញ្ញាសា) ។ លទ្ធផលនីមួយៗដែលអាចកើតមានឡើងនៃវិញ្ញាសាអោយឈ្មោះថា ព្រឹត្តិការណ៍ឯកធាតុ (elementary events) ។ ព្រឹត្តិការណ៍ឯកធាតុច្រើនតាងដោយ $\omega_1, \omega_2, \omega_3, \dots, \omega_n$ ។ គ្រប់ព្រឹត្តិការណ៍ឯកធាតុដែលយើងចង់អោយកើតមានឡើងនៃវិញ្ញាសាអោយឈ្មោះថា ព្រឹត្តិការណ៍នៃ ករណីស្រប ចំណែកព្រឹត្តិការណ៍ឯកធាតុដែលអាចកើតមានឡើងនៃវិញ្ញាសាអោយឈ្មោះថា ព្រឹត្តិការណ៍នៃករណីអាច។

ឧទាហរណ៍ៈ

គេមានសំនួរនៃការអង្កេតដូចខាងក្រោមៈ

តើអ្នកចូលចិត្តទៅកំសាន្តនៅតំបន់ទេសចរណ៍ណាជាងគេនៅក្នុងប្រទេសកម្ពុជា?

១ តំបន់ឆ្នេរ ១ តំបន់ភ្នំ ១ អង្គរវត្ត ១ ផ្សេងពីនេះ

ក្នុងករណីនេះសាកលព័ត៌មានដែលអាចទទួលបានពីការអង្កេតមាន តំបន់ឆ្នេរ តំបន់ភ្នំ អង្គរវត្ត ផ្សេងពីនេះ បង្កើតបានជាសាកល Ω ។

និយមន័យៈ

ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ A គឺជាផលធៀបរវាងចំនួនព្រឹត្តិការណ៍នៃករណីស្របនឹងចំនួនព្រឹត្តិការណ៍នៃករណីអាច ។

$$P(A) = \frac{m}{n}$$

ឬ $P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$

m -ចំនួនព្រឹត្តិការណ៍នៃករណីស្រប ; n -ចំនួនព្រឹត្តិការណ៍នៃករណីអាច ។ ក្នុងលក្ខខណ្ឌនេះយើងសន្មតថា ព្រឹត្តិការណ៍ឯកធាតុទាំងអស់មិនចុះសំរុងគ្នា មានសមភាពនឹងគ្នាហើយបង្កើត បានជាសាកល Ω ។

ចំណាំៈ

តាមការបង្ហាញខាងលើប្រូបាប៊ីលីតេគឺជាតំលៃលេខ ដែលកំណត់លទ្ធភាពនៃការកើតឡើងនៃព្រឹត្តិការណ៍ ។

២- លក្ខណៈសំគាល់នៃប្រូបាប៊ីលីតេ:

- ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ពិតប្រាកដស្មើមួយ
- ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍នៃព្រឹត្តិការណ៍មិនអាចមានស្មើសូន្យ
- ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ចែងនូវជំនួសវិជ្ជមានដែលកំណត់នៅក្នុងចន្លោះសូន្យនិងមួយ ។

$$0 \leq P(A) \leq 1$$

៣- តំរៀប(arrangement) , ចំណាស់(Permutation), និងបន្សំ(Combination):

- តំរៀប(មិនសារឡើងវិញ)

គេអោយសំនុំរាប់អស់ E ដែលមាន n ធាតុ ។ ដែលហៅថាតំរៀបនៃ m ធាតុនៃសំនុំ E គឺជាអនុវត្តន៍ប្រកាន់និមួយៗពី សំនុំ [1,m] ទៅសំនុំ E ។

$$A_n^m = n(n-1)(n-2)\dots(n-m+1) = \frac{n!}{(n-m)!}$$

Splus : choose(n,m,order.matters=T)

ឧទាហរណ៍:

តើគេអាចសរសេរលេខ 2 ខ្ទង់ដែលបង្កើតឡើងដោយលេខ 1,2,3 បានប៉ុន្មានរបៀប?

$$A_3^2 = 3 * 2 = 6$$

$$\{12,13,21,23,31,32\}$$

ចំណាំ: ក្នុងករណីនៃតំរៀបសារឡើងវិញ $A_n^m = n^m$

- ចំណាស់ (Permutation)

បើ $n=m$ តំរៀប A_n^m ហៅថាចំណាស់នៃ n ធាតុដែលតាងដោយ P_n ដែលកំណត់ :

$$P_n = A_n^n = n! = 1.2.3\dots n$$

ឧទាហរណ៍:

តើគេអាចសរសេរលេខ 3 ខ្ទង់ដែលបង្កើតឡើងដោយលេខ 1,2,3 បានប៉ុន្មានរបៀប ?

$$A_3^3 = P_3 = 1 * 2 * 3 = 6$$

$$\{123,132,213,231,312,321\}$$

- បន្សំ (Combination)

គេអោយសំនុំរាប់អស់ E ដែលមាន n ធាតុ ។ ដែលហៅថាបន្សំនៃ m ធាតុយកក្នុងសំនុំ E គឺជាផ្នែកមួយនៃសំនុំ E ដែលមាន m ធាតុ ហើយ តាងដោយ C_n^m or $\binom{n}{m}$ ។

$$C_n^m = \frac{n!}{m!(n-m)!}$$

S-plus: choose(n,m,order.matters=F)

ឧទាហរណ៍៖

តើគេអាចទាញបៀវត្ស 2 សន្លឹកពីក្នុងបៀវត្សដែលមាន 32 សន្លឹកបានប៉ុន្មានរបៀប?

$$C_{32}^2 = \frac{32!}{2!30!}$$

ចំណាំ :

ក្នុងការដោះស្រាយបញ្ហាដែលទាក់ទងទៅនឹងបន្សំ គេច្រើនអនុវត្តក្បួនដូចខាងក្រោម៖

- បើសិនណាវត្ថុ A ត្រូវបានជ្រើសរើសដោយ n របៀប ហើយវត្ថុ B ត្រូវបានជ្រើសរើសដោយ m របៀប ក្នុងចំណោមវត្ថុមួយចំនួននោះ ការជ្រើសរើសវត្ថុ A ឬ B អាចមាន n+m របៀប ។
- បើសិនណាវត្ថុ A ត្រូវបានជ្រើសរើសដោយ n របៀបហើយបន្ទាប់ពីការជ្រើសរើសវត្ថុ A វត្ថុ B ត្រូវបានជ្រើសដោយ m របៀប នោះគូ (A,B) ត្រូវបានជ្រើសរើសដោយ n+m របៀប ។

ឧទាហរណ៍មួយចំនួនក្នុងការគណនាប្រូបាប៊ីលីតេ :

ឧទាហរណ៍១៖

បុរសម្នាក់មានបំណងទូរស័ព្ទ ទៅមិត្តរបស់គាត់ ប៉ុន្តែគាត់ភ្លេចលេខដំបូងមួយខ្ទង់ ។ រកប្រូបាប៊ីលីតេដើម្បីអោយបុរសរូបនេះចុចលេខដែលគាត់ត្រូវការ ។

តាង A ជាព្រឹត្តិការណ៍ដែលបុរសរូបនេះចុចលេខដែលគាត់ត្រូវការ ។ ចំនួនព្រឹត្តិការណ៍នៃករណីអាចស្មើ 10 ដោយទូរស័ព្ទមាន 10 លេខ ដូច្នេះ៖

$$P(A) = \frac{1}{10}$$

ឧទាហរណ៍ ២ :

រៀបរៀងដោយ ម៉ាក កាមេរ៉ាន (Statistician, Mathematician).

បុរសម្នាក់បានភ្លេចលេខទូរស័ព្ទ 2 ខ្ទង់ចុងក្រោយនៃលេខទូរស័ព្ទរបស់មិត្តគាត់ ហើយគាត់ដឹងឡើងថា លេខទាំងពីរខ្ទង់ចុងក្រោយនេះខុសគ្នា ។ រកប្រូបាប៊ីលីតេដើម្បីអោយបុរសនេះ ចុចលេខដែលគាត់ត្រូវការនៅលើកទីមួយ ?

តាង A ជាព្រឹត្តិការណ៍ដែលបុរសម្នាក់នេះចុចលេខដែលគាត់ត្រូវការ ។

$$P(A) = \frac{1}{A_{10}^2}$$

ឧទាហរណ៍ 3 :

គេបោះគ្រាប់ឡកឡាក់ 2 គ្រាប់ ។ រកប្រូបាប៊ីលីតេដើម្បីអោយគ្រាប់ឡកឡាក់ទាំងពីរនេះស្មើ 4 ។

តាង A ជាព្រឹត្តិការណ៍ដើម្បីអោយគ្រាប់ឡកឡាក់ទាំងពីរមានផលបូកស្មើ 4 ។ យើងឃើញថាគ្រាប់ឡកឡាក់នីមួយៗអាចចេញបាន 6 របៀប ។ ដូច្នេះគ្រាប់ឡកឡាក់ទាំងពីរនេះអាចចេញបាន $6*6$ របៀប ។ ចំនួនព្រឹត្តិការណ៍នៃករណីអាចស្មើ 36 របៀប ឯចំនួនព្រឹត្តិការណ៍នៃករណីស្របស្មើ 3 គឺ:

$$\{(1,3),(3,1),(2,2)\}$$

$$P(A) = \frac{3}{36}$$

ឧទាហរណ៍អនុវត្តន៍:

១-នៅក្នុងសហគ្រាសមួយមានលោហធាតុ 10 ប្រភេទហើយដែលមាន 7 ប្រភេទជាប្រភេទលេខមួយ ។ រកប្រូបាប៊ីលីតេដើម្បីអោយក្នុងចំណោមលោហធាតុ 6 ប្រភេទដែលគេយកចេញពីសហគ្រាស នោះមានលោហធាតុ 4 ប្រភេទជាប្រភេទលេខមួយ ។

២-គេទាញប៊៊ូរ 5 សន្លឹកពីក្នុងប៊៊ូរ 52 សន្លឹក ។ តើលទ្ធផលនៃការទាញប៊៊ូរនេះអាចមានប៉ុន្មានរបៀបខុសៗគ្នា ?

៣-សំនុំមួយមាន 5 ធាតុ តើគេអាចបង្កើតសំនុំរងដែលមាន 3 ធាតុបានប៉ុន្មានរបៀប?

៤- S-plus: *Factorial, Combinations, Permutation*

ទំរង់ code ក្នុងការប្រើប្រាស់

```
factorial(n)
choose(n, k, order.matters=F)
choose.multinomial(n, m)
```

ARGUMENTS ដែលចាំបាច់:

n integer vector. Must be length 1 for choose.multinomial. Need not be integer for

factorial.
 k integer vector.
 m integer vector, which sums to n.

តំលៃ:

a numeric vector
 factorial(n) is $n!$
 choose(n, k) is $n! / (k! (n-k)!)$, i.e. the binomial coefficients.
 choose(n, k, order.matters=T) is $n! / (n-k)!$, i.e. the number of ordered subsets of length k from a set with n distinct elements.
 choose.multinomial(n, m) is $n! / \text{prod}(m!)$, the multinomial coefficients.

ឧទាហរណ៍

factorial(5)
 choose(5, 2)
 choose(5, -1:6)
 choose.multinomial(6, c(3,1,2))
 choose(5, 2, order=T)



ប្រមាណវិធីមួយចំនួននៃប្រូបាប៊ីលីតេ

១- ប្រមាណវិធីបូក

- បើ A និង B ជាព្រឹត្តិការណ៍ពីរមិនចុះសំរុងគ្នា (disjoint) នោះ :

$$P(A + B) = P(A) + P(B)$$

ជាទូទៅបើ A_1, A_2, \dots, A_n ជាព្រឹត្តិការណ៍មិនចុះសំរុងគ្នានោះ :

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

- បើ A_1, A_2, \dots, A_n ជាព្រឹត្តិការណ៍ឯកធាតុដែលបង្កើតបានជាសាកល Ω នោះ :
- ព្រឹត្តិការណ៍ពីរជាព្រឹត្តិការណ៍ផ្ទុយគ្នាកាលណា ព្រឹត្តិការណ៍ទាំងពីរនេះបង្កើតបានជាសាកល Ω ។ គេតែងតាងព្រឹត្តិការណ៍ផ្ទុយគ្នាដោយ \bar{A} ។ ដូច្នេះបើ \bar{A} និង A ជាព្រឹត្តិការណ៍ពីរផ្ទុយគ្នានោះ :

$$P(A) + P(\bar{A}) = 1$$

២- ស្វ័យសក្យនៃប្រូបាប៊ីលីតេ :

ក- $\forall A \subset \Omega \Rightarrow 0 \leq P(A) \leq 1$

ខ- $P(\Omega) = 1$

គ- $\forall A, B \subset \Omega, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ ។

៣- ប្រមាណវិធីគុណនៃព្រឹត្តិការណ៍ :

ផលគុណព្រឹត្តិការណ៍ពីរ A និង B គឺជាព្រឹត្តិការណ៍ដែលកើតឡើងរួមគ្នារវាងព្រឹត្តិការណ៍ទាំងពីរ ។

ឧទាហរណ៍:

បើ A ជាលោហៈធាតុប្រភេទលេខមួយ ហើយ B ជាលោហៈធាតុដែលលាបពណ៌ នោះលោហៈធាតុ

AB ជាលោហៈធាតុប្រភេទលេខមួយហើយលាបពណ៌ ។

៤- ប្រូបាប៊ីលីតេមានលក្ខខណ្ឌ (conditional probability):

ប្រូបាប៊ីលីតេមានលក្ខខណ្ឌ $P(A|B)$ គឺជា ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ A ដែលបានកើតឡើងក្នុងស្ថានភាពមួយនៅពេលដែលព្រឹត្តិការណ៍ B បានកើតរួច មកហើយ ។

ឧទាហរណ៍ :

នៅក្នុងថង់មួយមានប៊ូលពណ៌ស 3 និងប៊ូលពណ៌ខ្មៅ 3 ។ គេយកប៊ូលចេញពីក្នុងថង់មួយៗចំនួនពីរ លើកដោយមិនដាក់ចូលទៅវិញ ។ រកប្រូបាប៊ីលីតេដើម្បីអោយការដកចេញប៊ូលពីក្នុងថង់លើកទីពីរជាប៊ូលពណ៌ស បើសិនណាការដកចេញលើកទីមួយជាប៊ូលពណ៌ខ្មៅ ។

បើ A ជាព្រឹត្តិការណ៍នៃប៊ូលពណ៌ខ្មៅ

B ជាព្រឹត្តិការណ៍នៃប៊ូលពណ៌ ស

$$P(A) = \frac{3}{6} = \frac{1}{2}; P(A * B) = P(A \cap B) = \frac{3 * 3}{A_6^2} = \frac{9}{30} = \frac{3}{10} \text{ ។}$$

$$P(B \text{ ដោយស្ម័គ្រចំពោះ } A) = \frac{\frac{3}{10}}{\frac{1}{2}} = \frac{3}{5}$$

ប្រូបាប៊ីលីតេនេះហៅថាប្រូបាប៊ីលីតេ B មានលក្ខខណ្ឌ A ហើយគេច្រើនតែតាងដោយ $P(B|A)$ ឬ

$$P_A(B) \text{ ។}$$

និយមន័យ :

ប្រូបាប៊ីលីតេមានលក្ខខណ្ឌនៃព្រឹត្តិការណ៍ B ក្រោមលក្ខខណ្ឌដែលព្រឹត្តិការណ៍ A បានកើតឡើងរួចមកហើយ កំណត់ដោយ:

$$P(B|A) = \frac{P(AB)}{P(A)}, 0 < P(A) \leq 1$$

$$P(AB) = P(B|A) * P(A)$$

ចំណាំ:

$$P(AB) = P(BA) \Leftrightarrow P(A) * P(B|A) = P(B)P(A|B)$$

វិបាក:

$$P(A_1 \cdot A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1})$$

ឧទាហរណ៍ :

១-អ្នកលេងកូនឃ្លីម្នាក់មានឃ្លីស 3 គ្រាប់ និងឃ្លីក្រហម 7 គ្រាប់ ។ គាត់បានដកឃ្លីចេញពីហោប៉ៅ 1 គ្រាប់ (ដោយមិនដាក់ទៅវិញ) បន្ទាប់មកដកមួយគ្រាប់ទៀត ។ រកប្រូបាប៊ីលីតេដើម្បីអោយកូនឃ្លីដែលគាត់ដកចេញលើកទីមួយជាឃ្លីពណ៌ស និងលើកទីពីរជាឃ្លីពណ៌ក្រហម ។

ចម្លើយ: 7/30

២- គេបោះគ្រាប់ឡកឡាក់ពីរគ្រាប់ គ្រាប់ឡកឡាក់ទីមួយធ្វើអំពីភ្នំកនិងគ្រាប់ឡកឡាក់ទីពីរធ្វើអំពីស្នែង ។

ក- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយផលបូកគ្រាប់ឡកឡាក់ទាំងពីរស្មើ 8 ។

ខ- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយផលបូកគ្រាប់ឡកឡាក់តូចជាង 10 ។

~~~~~\*~~~~~

## ទ្រឹស្តីបទមួយចំនួននៃប្រូបាប៊ីលីតេ

### ១-ទ្រឹស្តីបទ ១

បើ A និង B ជាព្រឹត្តិការណ៍ចៃដន្យពីរ នោះ:

$$P(A + B) = P(A) + P(B) - P(AB)$$

### សំរាយបញ្ជី:

$A+B$  កើតមានឡើងនៅពេលណាដែលមានការកើតឡើងនៃព្រឹត្តិការណ៍ណាមួយក្នុងចំណោមព្រឹត្តិការណ៍ទាំងពីរ:

$$\overline{AB}, \overline{A}B, AB$$

ដូច្នេះ :  $P(A + B) = P(\overline{AB}) + P(\overline{A}B) + P(AB)$  (1)

$$P(A) = P(\overline{AB}) + P(AB) \Rightarrow P(\overline{AB}) = P(A) - P(AB), (2)$$

$$P(B) = P(\overline{A}B) + P(AB) \Rightarrow P(\overline{A}B) = P(B) - P(AB), (3)$$

ជំនួស (2) និង (3) ក្នុង(1) យើងបាន:

$$P(A + B) = P(A) + P(B) - P(AB)$$

### ២-លក្ខណៈសំគាល់មួយចំនួន:

- បើ A និង B ជាព្រឹត្តិការណ៍មិនទាក់ទងគ្នា (independent events)

$$P(A + B) = P(A) + P(B) - P(A) * P(B)$$

- បើ A និង B ជាព្រឹត្តិការណ៍ទាក់ទងគ្នា (dependent events)

$$P(A + B) = P(A) + P(B) - P(A)P(B \setminus A)$$

- បើ A និង B ជាព្រឹត្តិការណ៍មិនចុះសំរុងគ្នា (disjoint)

$$P(A + B) = P(A) + P(B) \quad \text{ព្រោះ } P(AB) = 0$$

### ៣-ទ្រឹស្តីបទ ២:

ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ A ដែលកើតឡើងក្រោមលក្ខខណ្ឌនៃការកើតឡើងព្រឹត្តិការណ៍ណាមួយក្នុងចំណោមព្រឹត្តិការណ៍មិនចុះសំរុងគ្នា  $B_1, B_2, \dots, B_n$  ដែលបង្កើតជាសាកលកំណត់ដោយ:

$$\begin{aligned} P(A) &= P(B_1)P(A \setminus B_1) + P(B_2)P(A \setminus B_2) + \dots + P(B_n)P(A \setminus B_n) \\ &= \sum_{i=1}^n P(B_i)P(A \setminus B_i) \end{aligned}$$

### ឧទាហរណ៍១:

នៅក្នុងកេសទីមួយមានចង្កៀង 20 គ្រឿងដែលក្នុងនោះមាន 18 គ្រឿងអាចប្រើប្រាស់បាន។ នៅក្នុងកេសទីពីរមានចង្កៀង 10 គ្រឿងដែលក្នុងនោះមាន 9 គ្រឿងអាចប្រើប្រាស់បាន។ គេយកចង្កៀងមួយពីកេសទីពីរដាក់ទៅក្នុងកេសទីមួយ។ រកប្រូបាប៊ីលីតេដើម្បីអោយចង្កៀងមួយដែលយកចេញពីកេសទីមួយជាចង្កៀងអាចប្រើប្រាស់បាន។

### ចម្លើយ:

តាង A ជាព្រឹត្តិការណ៍ “ចង្កៀងដែលយកចេញពីកេសទីមួយជាចង្កៀងដែលអាចប្រើប្រាស់បាន” ហើយ  $B_1$  ជាព្រឹត្តិការណ៍ “ចង្កៀងដែលយកចេញពីកេសទីពីរដាក់ចូលកេសទីមួយអាចប្រើប្រាស់បាន” និង  $B_2$  ជាព្រឹត្តិការណ៍ “ចង្កៀងដែលយកចេញពីកេសទីពីរដាក់ចូលកេសទីមួយមិនអាចប្រើប្រាស់បាន”។

$$P(B_1) = \frac{9}{10}; P(B_2) = \frac{1}{10}$$

ដូច្នេះប្រូបាប៊ីលីតេដើម្បីអោយចង្កៀងដែលយកចេញពីកេសទីមួយអាចប្រើប្រាស់បានជាប្រូបាប៊ីលីតេមានលក្ខខណ្ឌដែលកំណត់ដោយ  $P(A \mid B_1) = \frac{19}{21}$  (ក្រោមលក្ខខណ្ឌ  $B_1$ )។ ដូចគ្នានេះដែរ ប្រូបាប៊ីលីតេដើម្បីអោយចង្កៀងដែលយកចេញពីកេសទីមួយអាចប្រើប្រាស់បានក្រោមលក្ខខណ្ឌ  $B_2$  កំណត់ដោយ:

$$P(A \mid B_2) = \frac{18}{21}$$

$$\begin{aligned} \text{ដូច្នេះ } P(A) &= P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) \\ &= \frac{9}{10} * \frac{19}{21} + \frac{1}{10} * \frac{18}{21} = 0.9 \end{aligned}$$

### ឧទាហរណ៍២:

មានមនុស្សចំនួន 6 នាក់បានចូលលេងល្បែងដ៏សាហាវមួយ ឈ្មោះថា “Roulette Russe”។ អ្នកទាំង 6 នោះមានឈ្មោះ A,B,C,D,E,F ដែលត្រូវលេងតាមលំដាប់នេះ។ នៅក្នុងល្បែងនេះគេយកកាំភ្លើង “revolver” ដែលអាចដាក់បាន 6 គ្រាប់។ អ្នកទីមួយត្រូវដាក់គ្រាប់កាំភ្លើងមួយគ្រាប់ហើយ បង្វិលដុំកាំភ្លើងដោយចៃដន្យរួចបាញ់ខ្លួនឯង។ បើអ្នកទីមួយមិនស្លាប់ អ្នកទីពីរត្រូវដាក់បន្ថែមមួយគ្រាប់ទៀតរួចបាញ់ខ្លួនឯង។ បើអ្នកទីពីរមិនស្លាប់ទៀត គេនឹងបន្ថែមមួយគ្រាប់ទៀត ហើយបើគ្មានអ្នកស្លាប់ទេ គេបន្តធ្វើរបៀបនេះរហូតដល់អ្នកទី 6។ តើអ្នកមានលេខរៀងទីប៉ុន្មានដែលមានសំណាងជាងគេនៅក្នុងទង្វើ នេះ?

**ចំណើយ :**

តាង A ជាព្រឹត្តិការណ៍ " បុគ្គល A ត្រូវស្លាប់ "

$\bar{A}$  ជាព្រឹត្តិការណ៍ " បុគ្គល A មិនស្លាប់ "

**ទីតាំងទី ១:**

$$P(A) = \frac{1}{6}$$

**ទីតាំងទី ២:**

$$P(\bar{A} \cap B) = P(\bar{A})P(B \setminus \bar{A}) = \frac{5}{6} * \frac{2}{6} = \frac{5}{18}$$

**ទីតាំងទី៣:**

$$P(\bar{A} \cap \bar{B} \cap C) = P(\bar{A})P(\bar{B} \setminus \bar{A})P(C \setminus \bar{A} \cap \bar{B}) = \frac{5}{6} * \frac{4}{6} * \frac{3}{6} = \frac{5}{18}$$

**ទីតាំងទី៤:**

$$P(\bar{A} \cap \bar{B} \cap \bar{C} \cap D) = P(\bar{A})P(\bar{B} \setminus \bar{A})P(\bar{C} \setminus \bar{A} \cap \bar{B})P(D \setminus \bar{A} \cap \bar{B} \cap \bar{C})$$

$$P(\bar{A} \cap \bar{B} \cap \bar{C} \cap D) = \frac{5}{6} * \frac{4}{6} * \frac{3}{6} * \frac{4}{6} = \frac{5}{27}$$

**ទីតាំងទី៥:**

$$\begin{aligned} P(\bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D} \cap E) &= P(\bar{A})P(\bar{B} \setminus \bar{A})P(\bar{C} \setminus \bar{A} \cap \bar{B})P(\bar{D} \setminus \bar{A} \cap \bar{B} \cap \bar{C})P(E \setminus \bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D}) \\ &= \frac{5}{6} * \frac{4}{6} * \frac{3}{6} * \frac{2}{6} * \frac{5}{6} = \frac{5^2}{18^2} \end{aligned}$$

**ទីតាំងទី៦:**

$$\begin{aligned} P(\overline{ABCDEF}) &= P(\bar{A})P(\bar{B} \setminus \bar{A})P(\bar{C} \setminus \bar{A} \cap \bar{B})P(\bar{D} \setminus \bar{A} \cap \bar{B} \cap \bar{C})P(\bar{E} \setminus \bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D})P(\bar{F} \setminus \bar{A} \cap \bar{B} \cap \bar{C} \cap \bar{D} \cap \bar{E}) \\ &= \frac{5}{6} * \frac{2}{3} * \frac{1}{2} * \frac{1}{3} * \frac{1}{6} * \frac{6}{6} = \frac{5}{18^2} \end{aligned}$$

យើងសង្កេតឃើញថាទីតាំងដែលល្អបំផុតចំពោះអ្នកលេងទាំងនេះគឺជាទីតាំងចុងក្រោយបំផុត ។ យើងក៏ អាចផ្ទៀងផ្ទាត់ផងដែរថាក្នុងចំណោមអ្នកលេងទាំង 6 នាក់នេះត្រូវមានម្នាក់ជាអ្នកស្លាប់ បានសេចក្តីថា:

$$P(A \cup B \cup C \cup D \cup E \cup F) = P(A) + P(B) + P(C) + P(D) + P(E) + P(F) = 1$$

**S-plus:**

```
function()
{ b<-1
  z<-1/6
  VeryLuck<-z
  for(i in 2:6)
    { a<-i/6
      b<-b*((6-i+1)/6)
      z[i]<-a*b
      if( VeryLuck>z[i])
      {
        VeryLuck <-z[i]
        Num<-i
      }
    }
  Person<-c(1,2,3,4,5,6)
  z<-cbind(Person,z)
  VeryLuck<-cbind(Num,VeryLuck)
  return(z,VeryLuck)
}
```

**Result:**

|      | Person | z          |
|------|--------|------------|
| [1,] | 1      | 0.16666667 |
| [2,] | 2      | 0.27777778 |
| [3,] | 3      | 0.27777778 |
| [4,] | 4      | 0.18518519 |
| [5,] | 5      | 0.07716049 |
| [6,] | 6      | 0.01543210 |

**\$VeryLuck:**

|      | Num | VeryLuck  |
|------|-----|-----------|
| [1,] | 6   | 0.0154321 |

## រូបមន្ត Bayes

### ១- ទ្រឹស្តី Bayes

នៅក្នុងក្រុមហ៊ុន SIMCO មានបុគ្គលិក 20% មានសញ្ញាបត្រ Management ។ ក្នុងចំណោមបុគ្គលិកទាំង 20% នេះ មាន 70% មានប័ណ្ណបញ្ជីអចិន្ត្រៃយ៍ ។ យើងដឹងទៀតថាក្នុងចំណោមបុគ្គលិកដែល គ្មានសញ្ញាបត្រ Management 15% មានប័ណ្ណបញ្ជីអចិន្ត្រៃយ៍ ។ បើសិនជាបុគ្គលិកម្នាក់ដែលមានប័ណ្ណបញ្ជីអចិន្ត្រៃយ៍ត្រូវបានជ្រើសរើសដោយចៃដន្យពីក្រុមហ៊ុននេះ តើប្រូបាប៊ីលីតេដើម្បីអោយគាត់មាន សញ្ញាបត្រខាង Management ស្មើប៉ុន្មាន ?

#### ចំណេះដឹង :

បុគ្គលិកទាំងអស់ត្រូវចែកជាពីរផ្នែកមិនចុះសំរុងគ្នា (disjoints) គឺ:

$E_1$  : បុគ្គលិកដែលមានសញ្ញាបត្រ Management

$E_2$  : បុគ្គលិកដែលគ្មានសញ្ញាបត្រ Management

$$P(E_1) = 0.20, P(E_2) = 0.80$$

$A$  : បុគ្គលិកម្នាក់ដែលបានជ្រើសរើសដោយចៃដន្យមានប័ណ្ណបញ្ជីអចិន្ត្រៃយ៍

$$P(A \setminus E_1) = 0.70, P(A \setminus E_2) = 0.15$$

អ្វីដែលយើងចង់រកនៅពេលនេះគឺ  $P(E_1 \setminus A)$  :

$$P(E_1 \setminus A) = \frac{P(E_1 \cap A)}{P(A)} \Rightarrow P(E_1 \cap A) = P(A)P(E_1 \setminus A)$$

ម្យ៉ាងវិញទៀត  $A$  បង្កើនពីព្រឹត្តិការណ៍ពីរដែលមិនចុះសំរុងគ្នា:

$$A = (E_1 \cap A) \cup (E_2 \cap A)$$

ដូច្នេះ:

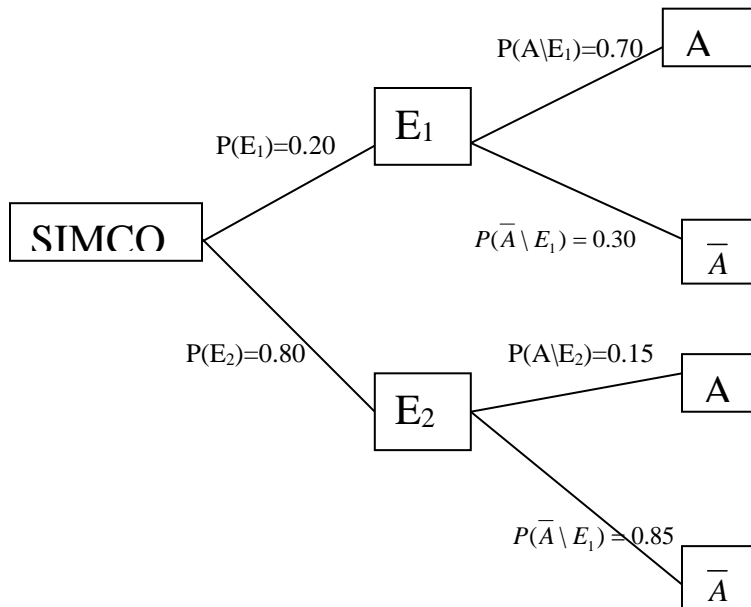
$$P(A) = P(E_1 \cap A) + P(E_2 \cap A)$$

ដោយ  $P(E_1 \cap A) = P(E_1)P(A \setminus E_1)$  និង  $P(E_2 \cap A) = P(E_2)P(A \setminus E_2)$

$$\Rightarrow P(E_1 \setminus A) = \frac{P(E_1)P(A \setminus E_1)}{P(E_1)P(A \setminus E_1) + P(E_2)P(A \setminus E_2)}$$

រូបមន្តនេះហៅថា រូបមន្ត Bayes ។

ដូច្នេះ : 
$$P(E_1 \setminus A) = \frac{0.20 * 0.70}{0.20 * 0.70 + 0.80 * 0.15} = 0.5384$$



ជាទូទៅ :

$$P(E_i \setminus A) = \frac{P(E_i)P(A \setminus E_i)}{\sum_{i=1}^n P(E_i)P(A \setminus E_i)}$$

រូបមន្តនេះអោយឈ្មោះថា រូបមន្ត Bayes ។

## ២-លំហាត់អនុវត្តន៍:

- ១- ប្រូបាប៊ីលីតេដើម្បីអោយយន្តហោះចេញតាមពេលកំណត់ដែលបានព្រៀងនៅលើតារាងពេលវេលា (scheduled flight departs)  $P(D)=0.83$  ។ ដូចគ្នានេះដែរ ប្រូបាប៊ីលីតេដើម្បីអោយយន្តហោះនេះមកដល់តាមពេលកំណត់  $P(A)=0.82$  ហើយប្រូបាប៊ីលីតេដើម្បីអោយយន្តហោះនេះចេញតាមពេលកំណត់ និងមកដល់តាមពេលកំណត់  $P(DA)=0.78$  ។ កំណត់ប្រូបាប៊ីលីតេ:
- a)  $P(A \setminus D)$  ; b)  $P(D \setminus A)$

- ២- នៅក្នុងភូមិដ៏តូចមួយមាននិស្សិតមួយចំនួនដែលបានបញ្ចប់ការសិក្សាថ្នាក់មធ្យមផ្នែកបច្ចេកទេសកសិកម្ម ហើយដែលមាននៅក្នុងតារាងស្ថិតិដូចខាងក្រោម:



|      | មានការងារធ្វើ | គ្មានការងារធ្វើ | សរុប |
|------|---------------|-----------------|------|
| បុរស | 460           | 40              | 500  |
| ស្រី | 140           | 260             | 400  |
| សរុប | 600           | 300             | 900  |

និស្សិតម្នាក់ត្រូវបានជ្រើសរើសពីក្នុងភូមិនេះដោយចៃដន្យដើម្បីទៅសិក្សាបន្តនៅបរទេសសិក្សាស្វែងយល់ពីបច្ចេកវិជ្ជាមួយចំនួនក្នុងការបង្កើតរោងចក្រមួយក្នុងភូមិនេះ ។ កំណត់ប្រូបាប៊ីលីតេ៖

ក- និស្សិតម្នាក់នេះជាបុរសក្រោមលក្ខខ័ណ្ឌដែលគាត់មានការងារធ្វើ?

ខ- និស្សិតម្នាក់នេះជាស្រីក្រោមលក្ខខ័ណ្ឌដែលគាត់មានការងារធ្វើ?

**ចម្លើយ៖**

តាង M ជាបុរសដែលបានជ្រើសរើស

F ជាស្រីដែលបានជ្រើសរើស

E និស្សិតម្នាក់នេះមានការងារធ្វើ

$$P(M \setminus E) = \frac{P(E \cap M)}{P(E)} = \frac{\frac{460}{900}}{\frac{600}{900}} = \frac{23}{30}; P(F \setminus E) = \frac{P(E \cap F)}{P(E)}$$

៣- អ្នកទទួលខុសត្រូវនៃការិយាល័យធុនធានមនុស្សមួយកន្លែងបានតំកល់ទុកប្រវត្តិរូបសង្ខេបរបស់បញ្ជីន្ត

ចំនួន 16,000 នាក់ ។ គេបានធ្វើការសិក្សាទៅលើអាយុនិងភេទរបស់បញ្ជីន្តទាំងនេះដូចខាងក្រោម៖

| អាយុ \ ភេទ                | ប្រុស(M) | ស្រី(F) | សរុប  |
|---------------------------|----------|---------|-------|
| ក្រោម 30 ឆ្នាំ (A)        | 1200     | 1700    | 2900  |
| ចន្លោះ 30 ទៅ 40 ឆ្នាំ (B) | 2600     | 4200    | 6800  |
| ច្រើនជាង 40 ឆ្នាំ (C)     | 4000     | 2300    | 6300  |
| សរុប                      | 7800     | 8200    | 16000 |

- ក-បើប្រវត្តិរូបសង្ខេបរបស់បញ្ជីន្ត ម្នាក់ត្រូវបានជ្រើសរើសដោយចៃដន្យ តើប្រូបាប៊ីលីតេដើម្បីអោយ  
មនុស្សម្នាក់នេះមានអាយុតិចជាង 30 ឆ្នាំស្មើប៉ុន្មាន ?
- ខ- ប្រូបាប៊ីលីតេដើម្បីអោយមនុស្សស្រីម្នាក់ដែលត្រូវបានជ្រើសរើសដោយចៃដន្យមានអាយុតិចជាង 30  
ឆ្នាំ ស្មើប៉ុន្មាន ?
- គ- ប្រូបាប៊ីលីតេដើម្បីអោយមនុស្សម្នាក់ដែលត្រូវបានជ្រើសរើសដោយចៃដន្យជាមនុស្សប្រុសនិងមានអាយុ  
ច្រើនជាង 40 ឆ្នាំ ស្មើប៉ុន្មាន ?
- ឃ- ប្រូបាប៊ីលីតេដើម្បីអោយមនុស្សម្នាក់ដែលត្រូវបានជ្រើសរើសដោយចៃដន្យជាមនុស្សស្រីនិងមានអាយុ  
តិចជាង 40 ឆ្នាំ ស្មើប៉ុន្មាន?

**ចម្លើយ :**

$$\text{ក- } P(A) = \frac{2900}{16000} = 0.1813$$

$$\text{ខ- } P(A \setminus F) = \frac{P(A \cap F)}{P(F)} = \frac{\frac{1700}{16000}}{\frac{8200}{16000}} = 0.2073$$

$$\text{គ- } P(M \cap C) = \frac{4000}{16000} = 0.25 \quad \text{ឬ} \quad P(M \cap C) = P(M)P(C \setminus M) = \frac{7800}{16000} * \frac{4000}{7800} = 0.25$$

$$\text{ឃ- } P(F \cap (A \cup B)) = P(F \cap A) + P(F \cap B) = 0.36875$$

៣- ឧទាហរណ៍ Splus:

ដោយអនុវត្តទៅលើតារាង ទិន្នន័យ sales នៅក្នុង SPSS ចូរសរសេរ Program Splus បង្កើតតារាង វាយតម្លៃ Probability , conditional probability ដូចខាងក្រោម:

Table1

|       | Government | Commercial | Academic | Total |
|-------|------------|------------|----------|-------|
| North | $P(G N)$   | $P(C N)$   | $P(A N)$ | 100%  |
| South | $P(G S)$   | $P(C S)$   | $P(A S)$ | 100%  |
| East  | $P(G E)$   | $P(C E)$   | $P(A E)$ | 100%  |
| West  | $P(G W)$   | $P(C W)$   | $P(A W)$ | 100%  |
| Total | $P(G)$     | $P(C)$     | $P(A)$   | 100%  |

Table2

|       | Government | Commercial | Academic | Total  |
|-------|------------|------------|----------|--------|
| North | $P(N G)$   | $P(N C)$   | $P(N A)$ | $P(N)$ |
| South | $P(S G)$   | $P(S C)$   | $P(S A)$ | $P(S)$ |
| East  | $P(E G)$   | $P(E C)$   | $P(E A)$ | $P(E)$ |
| West  | $P(W G)$   | $P(W C)$   | $P(W A)$ | $P(W)$ |
| Total | 100%       | 100%       | 100%     | 100%   |

Table3

|       | Government | Commercial | Academic | Total  |
|-------|------------|------------|----------|--------|
| North | $P(G*N)$   | $P(C*N)$   | $P(A*N)$ | $P(N)$ |
| South | $P(G*S)$   | $P(C*S)$   | $P(A*S)$ | $P(S)$ |
| East  | $P(G*E)$   | $P(C*E)$   | $P(A*E)$ | $P(E)$ |
| West  | $P(G*W)$   | $P(C*W)$   | $P(A*W)$ | $P(W)$ |
| Total | $P(G)$     | $P(C)$     | $P(A)$   | 100%   |



## ជំពូក២: រោងប្រូបាប៊ីលីតេ

### អថេរចៃដន្យ និងរោងប្រូបាប៊ីលីតេ

(Random variables and probability distribution)

#### ១- អថេរចៃដន្យ:

នៅក្នុងការបោះគ្រាប់ឡកឡាក់យើងពុំអាចកំណត់បានជាមុនថា លេខមួយណាកើតមុនទេ ព្រោះវាទាក់ទងទៅនឹងលក្ខខណ្ឌជាច្រើន។ នៅក្នុងន័យនេះលេខ 1,2,3,4,5,6 ជាតំលៃលេខដែលអាចកើតឡើងដោយចៃដន្យ។

#### និយមន័យទី១:

អថេរចៃដន្យ (random variable) គឺជាអថេរដែលនៅក្នុងលទ្ធផលនៃវិញ្ញាសានីមួយៗ អាចទទួលបានតំលៃតែមួយគត់ដែលជាតំលៃមួយពុំអាចកំណត់បានជាមុន ហើយដែលទាក់ទងទៅនឹងលក្ខខណ្ឌចៃដន្យជាច្រើន។

#### ឧទាហរណ៍ :

ក្នុងការបោះគ្រាប់ឡកឡាក់បើសិនណាយើងទទួលបានលេខ 1 ក្នុងការបោះលើកទីមួយ យើងនិយាយថា លេខ 1 ជាអថេរចៃដន្យក្នុងការពិសោធលើកទីមួយ។ ការបោះគ្រាប់ឡកឡាក់បន្តទៀតអាចទទួលបានតំលៃលេខជាហូរហែ តំលៃទាំងនេះអោយឈ្មោះថា **អថេរចៃដន្យ**។ ដូច្នេះក្នុងការបោះគ្រាប់ឡកឡាក់ អថេរចៃដន្យអាចមាន 6 កំណើ:  $\{1,2,3,4,5,6\}$  ។

#### និយមន័យទី២:

អថេរចៃដន្យដាច់ (discrete random variable) គឺជាអថេរចៃដន្យដែលអាចទទួលបានតំលៃដាច់ ដោយឡែកពីគ្នា ហើយមានប្រូបាប៊ីលីតេមួយកំណត់។ ចំនួនតំលៃលេខដែលអាចនៃអថេរចៃដន្យដាច់ជា ចំនួនអាចរាប់អស់ ឬជាចំនួនរាប់មិនអស់។

#### ឧទាហរណ៍ :

ចំនួនក្មេងប្រុសដែលទើបនឹងកើតក្នុងចំណោមក្មេងដែលទើបនឹងកើត 200 នាក់ ជាអថេរចៃដន្យដែលតាងដោយ X ហើយអថេរចៃដន្យអាចទទួលបានតំលៃ 0,1,2,3,...,200 ។

### និយមន័យទី៣:

អថេរចៃដន្យជាប់ (continuous random variables) គឺជាអថេរចៃដន្យដែលអាចទទួលបាននូវគ្រប់តំលៃនៅក្នុងចន្លោះមួយកំណត់ ឬមួយក្នុងចន្លោះមួយមិនកំណត់ ។ ចំនួនតំលៃនៃអថេរចៃដន្យជាប់ ជាចំនួនរាប់មិនអស់ ។

### ឧទាហរណ៍:

ចំងាយចរនៃការបាញ់គ្រាប់កាំភ្លើងមួយគ្រាប់គឺជាអថេរចៃដន្យជាប់ ។ ជាការពិតណាស់ចំងាយចរនៃគ្រាប់កាំភ្លើងនេះវាទាក់ទងទៅនឹងលក្ខខណ្ឌជាច្រើន៖ កំលាំងធាក់នៃកាំភ្លើង ទិសដៅនៃខ្យល់ និងសម្ពាធបរិយាកាសផ្សេងៗទៀត ដែលយើងពុំអាចគិតទុកជាមុន ។ តំលៃដែលអាចនៃអថេរចៃដន្យនេះស្ថិតនៅក្នុងចន្លោះ (a,b) ។ យើងឃើញថាគ្រាប់កាំភ្លើងអាចធ្លាក់ដោយចៃដន្យនូវគ្រាប់ចំនុចនៃចន្លោះ (a,b) ។

### ២- ឧទាហរណ៍ នៃរបាយប្រូបាប៊ីលីតេ (probability distribution) នៃចំនួនអាំងឡែប៊ុយម៉ា (switch) ដែលខូច:

សហគ្រាស Microtek បានផលិត switch ដែលមានពន្លឺ ។ សំរុងស្ថិតិមួយបានបង្ហាញមាន 5% នៃ switch ដែលផលិតដោយ Microtek ត្រូវខូចប្រើប្រាស់ពុំបាន ។ សន្មតថាគេជ្រើសរើសដោយចៃដន្យនូវ switch ចំនួន 2 ។ សន្មតថាអថេរចៃដន្យតាងដោយ X ចំនួន switch ដែលខូចនៅក្នុងគំរូស្ថិតិដែលគេធ្វើការដកស្រង់នេះ ។ ដើម្បីបង្កើតរបាយប្រូបាប៊ីលីតេ នៃអថេរចៃដន្យនេះយើងត្រូវស្គាល់:

ក- តំលៃអថេរចៃដន្យទាំងនេះ

ខ- តំលៃប្រូបាប៊ីលីតេនៃអថេរនីមួយៗ

ជាដំបូងយើងត្រូវកំណត់នូវគ្រប់ព្រឹត្តិការណ៍ឯកធាតុនៃវិញ្ញាណនេះ ដោយតាង D ជា " switch ដែលខូច " និង B ជា " switch ដែលល្អ " ដូច្នេះព្រឹត្តិការណ៍ទាំងនេះគឺ:

$$\Omega = \{BB, BD, DB, DD\}$$

| ព្រឹត្តិការណ៍ឯកធាតុ | តំលៃ X ដែលទាក់ទងទៅនឹងព្រឹត្តិការណ៍ទាំងនេះ |
|---------------------|-------------------------------------------|
| BB                  | 0                                         |
| BD                  | 1                                         |
| DB                  | 1                                         |
| DD                  | 2                                         |

យើងដឹងថាមាន switch ចំនួន 5% ខូច (95% អាចប្រើប្រាស់បាន) ដូច្នេះប្រូបាប៊ីលីតេនៃ switch ដែលខូចស្មើ 0.05 និងប្រូបាប៊ីលីតេនៃ switch ដែលអាចប្រើប្រាស់បានស្មើ 0.95 ។ ដោយសន្មតព្រឹត្តិការណ៍ខាងលើមិនទាក់ទងគ្នា :

$$P(BB)=P(B)*P(B)=0.95*0.95=0.9025$$

$$P(BD)=P(B)*P(D)=0.95*0.05=0.0475$$

$$P(DB)=P(D)*P(B)=0.05*0.95=0.0475$$

$$P(DD)=P(D)*P(D)=0.05*0.05=0.0025$$

ដូច្នេះយើងអាចសង្ខេបតារាងរបាយប្រូបាប៊ីលីតេដូចខាងក្រោម :

| ព្រឹត្តិការណ៍ឯកធាតុ | តំលៃ $x_i$ នៃអថេរចៃដន្យ X | ប្រូបាប៊ីលីតេ $P(X=x_i)$ |
|---------------------|---------------------------|--------------------------|
| BB                  | $x_1=0$                   | $P(X=0)=0.9025$          |
| BD                  | $x_2=1$                   | $P(X=1)=0.0950$          |
| DB                  |                           |                          |
| DD                  | $x_3=2$                   | $P(X=2)=0.0025$          |
|                     |                           | ផលបូក = 1                |

ដូច្នេះអថេរចៃដន្យទាំងនេះជាអថេរចៃដន្យដាច់ដែល :

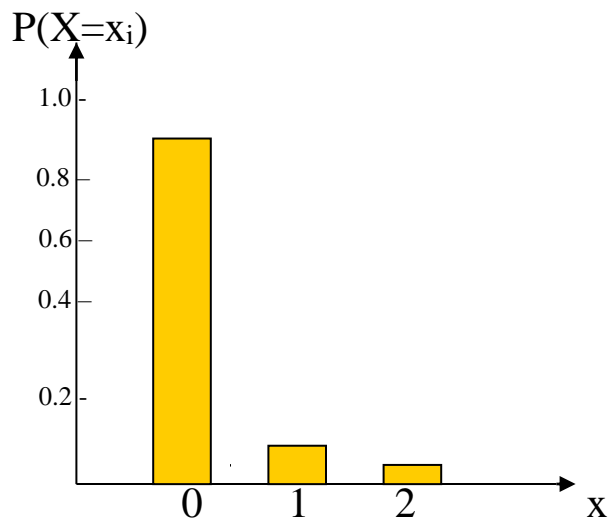
X : 0 1 2

### ៣- ក្រាហ្វិកនិងច្បាប់របាយនៃប្រូបាប៊ីលីតេ (Graphics and probability distribution)

#### ដ្យាក្រាមដំបង (bar chart) :

គេអាចតារាងរបាយប្រូបាប៊ីលីតេនៃអថេរចៃដន្យដាច់ដោយដ្យាក្រាមដំបង ។ ដោយអនុវត្តលើឧទាហរណ៍ខាងលើយើងអាចបានដ្យាក្រាមដំបងដូចខាងក្រោម ដោយប្រើកម្មវិធី S-plus:

```
function()
{ x<-c(0,1,2)
  probability<-c(0.9025,0.0950,0.0025)
  barplot(probability,names=levels(factor(x)))
  return()
}
```



#### ៤-អនុគមន៍រំបាយ (Distribution function)

អនុគមន៍រំបាយកំណត់ដោយ:  $F(x_i) = P(X \leq x_i)$

នៅក្នុងករណីនៃអថេរចៃដន្យដាច់:

$$F(x_i) = P(X \leq x_i) = P(X = x_1) + P(X = x_2) + \dots + P(X = x_i)$$

ក្រាហ្វិចនៃអនុគមន៍រំបាយជាអនុគមន៍កាំជណ្តើរ (step function) ។

#### ឧទាហរណ៍:

នៅក្នុងក្រុមហ៊ុន sangamex គេបានសិក្សាទៅលើកំហុសនៃការតំឡើងឧបករណ៍ electronic ដែលកំណត់ក្នុងតារាងខាងក្រោម :

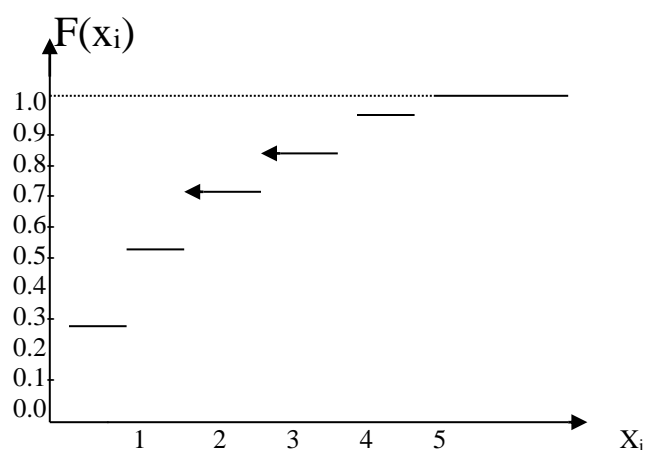
| ចំនួននៃកំហុស $x_i$ | $P(X=x_i)$ |
|--------------------|------------|
| 0                  | 0.30       |
| 1                  | 0.25       |
| 2                  | 0.18       |
| 3                  | 0.14       |
| 4                  | 0.10       |
| 5                  | 0.03       |

ក- កំណត់អនុគមន៍រំបាយ និងសង់ក្រាហ្វិច :

ក្នុងន័យនេះយើងត្រូវកំណត់  $P(X \leq x_i) = F(x_i)$  នៃអថេរចៃដន្យ ” ចំនួនកំហុសទៅតាមឧបករណ៍

electronic នីមួយៗ ” ។

| ចំនួនកំហុស $x_i$ | $F(x_i) = P(X \leq x_i)$ |
|------------------|--------------------------|
| 0                | 0.30                     |
| 1                | 0.55                     |
| 2                | 0.73                     |
| 3                | 0.87                     |
| 4                | 0.97                     |
| 5                | 1.00                     |



ខ- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយចំនួនកំហុសនៃការតំឡើងឧបករណ៍ electronic នេះតូចជាង ឬ ស្មើ 2?

$$F(2) = P(X \leq 2) = 0.73$$

គ- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយចំនួនកំហុសនៃការតំឡើងឧបករណ៍ electronic នេះធំជាង 1?

$$P(X > 1) = 1 - P(X \leq 1)$$

$$= 1 - F(1)$$

$$= 1 - 0.55$$

$$= 0.45$$

Graph នៃអនុគមន៍របាយខាងលើអាចទទួលបានដោយប្រើកម្មវិធី S-plus:

```
function()
{ x<-c(0,1,2,3,4,5)
  dist.fun<-c(0.30,0.55,0.73,0.87,0.97,1)
  plot(stepfun(x,dist.fun),type="l")
  return()
}
```



**លំហាត់:**

នៅក្នុងការអង្កេតមួយដែលធ្វើឡើងនៅក្នុងហាងលក់ទំនិញដ៏ធំមួយបានអោយដឹងថា ក្នុងចំណោម អតិថិជន 200 នាក់ មាន 120 នាក់បានបញ្ជាក់ថាគាត់មកហាងលក់ទំនិញនេះដោយសារគាត់ឃើញការផ្សាយពាណិជ្ជកម្ម ។ មួយចំនួនទៀតមកទីនេះដោយពុំបានឃើញការផ្សាយពាណិជ្ជកម្មទេ ។ ក្នុងចំណោមអតិថិជនទាំង 200 នាក់នេះដែលត្រូវបានចោទសួរមាន 60 នាក់បានទិញសម្ភារៈមួយចំនួន ហើយក្នុងចំណោមអ្នកទាំង 60 នាក់នេះ មាន 20 នាក់បានឃើញការផ្សាយពាណិជ្ជកម្ម ។

ក- ចូរបង្កើតតារាង two-way table

ខ- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយអតិថិជន ដែលពុំបានឃើញការផ្សាយពាណិជ្ជកម្ម បានទិញសម្ភារៈមួយចំនួនក្នុងហាងនេះ ។

គ- កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយអតិថិជន ដែលបានឃើញការផ្សាយពាណិជ្ជកម្ម បានទិញសម្ភារៈមួយចំនួនក្នុងហាងនេះ ។

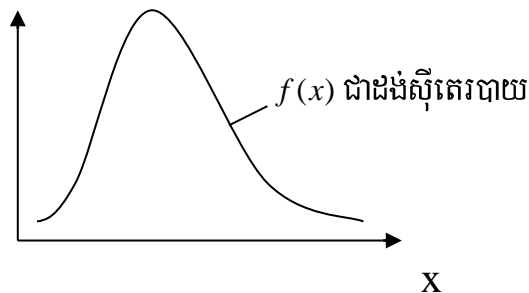
# ប្រូបាប៊ីលីតេនៃអថេរចៃដន្យជាប់

## (Probability of continuous random variables)

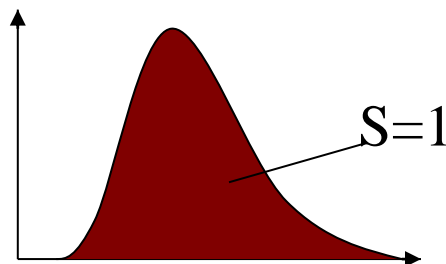
បើ  $X$  ជាអថេរចៃដន្យជាប់ (continuous random variables) នោះវាមានរបាយដង់ស៊ីតេ (density distribution)  $f(x), x \in \mathcal{R}$  ដែល  $F(x) = \int_{-\infty}^{+\infty} f(t)dt$  ហើយ  $F(x)$  ជា អនុគមន៍របាយ (distribution function) ។

### ១-លក្ខណៈសំគាល់នៃដង់ស៊ីតេរបាយ :

ក-  $f(x) \geq 0 \forall x \in \mathcal{R}$  បានន័យថាខ្សែកោងរបាយដង់ស៊ីតេជានិច្ចកាលស្ថិតនៅខាងលើខ្សែអក្សរអាប៉ូស៊ីស ។



ខ-  $\int_{-\infty}^{+\infty} f(t)dt = 1$  បានន័យថាក្រលាផ្ទៃដែលខ័ណ្ឌដោយខ្សែកោងរបាយដង់ស៊ីតេ នឹងអ័ក្សអាប៉ូស៊ីស ស្មើមួយ ។



គ-  $P(a \leq X \leq b)$  ជាក្រលាផ្ទៃដែលបង្កើតឡើងដោយខ្សែកោងរបាយដង់ស៊ីតេនឹងអ័ក្សអាប៉ូស៊ីស

ខ័ណ្ឌដោយបន្ទាត់ឈរកាត់តាមចំនុចអាប់ស៊ីស  $a$  និង  $b$  ។

**ចំណាំ :** នៅក្នុងករណីដែល  $X$  ជាអថេរចៃដន្យជាប់ នោះ :

$$\text{ក- } P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = F(b) - F(a)$$

$$\text{ខ- } P(X = a) = P(a \leq X \leq a) = F(a) - F(a) = 0$$

## **២-សង្ខ័យគណិត វ៉ារ្យង់ និង គំលាតស្តង់ដា នៃអថេរចៃដន្យជាប់ :**

**ក- សង្ខ័យគណិត (Expectation):**

$X$  ជាអថេរចៃដន្យជាប់ដែលអាចទទួលបានតំលៃ  $X_1, X_2, \dots, X_n$  ហើយដែលមានប្រូបាប៊ីលីតេរៀង

$P(X=x_i)=p_i$  ,  $i=1,2,3,\dots,n$  ។ សង្ខ័យគណិតនៃអថេរចៃដន្យ  $X$  កំណត់ដូចខាងក្រោម :

$$E(X) = \sum_{i=1}^n x_i p_i$$

តំលៃសង្ខ័យគណិតជាចំនួនពិត ១ ។ នៅក្នុងស្ថិតិតេច្រើនប្រើនិមិត្តសញ្ញា  $\mu$  ជំនួសតំលៃ  $E(X)$  ។

**ខ-វ៉ារ្យង់ (Variance) និងគំលាតស្តង់ដា (Standard deviation) :**

បើ  $X$  ជាអថេរចៃដន្យជាប់ដែលអាចទទួលបានតំលៃ  $X_1, X_2, \dots, X_n$  នោះវ៉ារ្យង់នៃ  $X$  កំណត់ដោយ:

$$V(x) = \sum_{i=1}^n (x_i - E(X))^2 p_i$$

ដែល  $p_i = P(X = x_i), i = 1, 2, \dots, n$  ។

រឹសការេនៃវ៉ារ្យង់អោយឈ្មោះថា standard deviation ដែលតាងដោយ  $\sigma(X)$  :

$$\sigma(X) = \sqrt{V(X)}$$

ជាទូទៅវ៉ារ្យង់នៃអថេរចៃដន្យ  $X$  កំណត់ដោយ:

$$V(X) = E(X - E(X))^2 = E(X^2) - E^2(X)$$

## **៣-សង្ខ័យគណិត និងវ៉ារ្យង់នៃអថេរចៃដន្យជាប់ (continuous random variables)**

នៅក្នុងករណីដែល  $X$  ជាអថេរចៃដន្យជាប់ នោះសង្ខ័យគណិតនិងវ៉ារ្យង់នៃ  $X$  កំណត់ដោយ :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx$$

ដែល  $f(x)$  ជាដង់ស៊ីតេបាយនៃ  $X$  ។

### ចំណាំ:

ក្នុងករណីដែលយើងធ្វើការស្រង់ទិន្នន័យក្នុងគំរូតាង (sample) សំរាប់ការវាយតម្លៃ នោះតម្លៃ

Expectation Variance និង Standard deviation ត្រូវបានគណនាតាមរូបមន្តបំប្លែងដូចខាងក្រោម:

$$E(X) \approx \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{sample mean}$$

$$\sigma(X) \approx \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2} = S \quad \text{sample standard deviation}$$

### ឧទាហរណ៍ :

យើងមានទិន្នន័យដូចខាងក្រោម

| profit (\$) | probability |
|-------------|-------------|
| 25000       | 0.25        |
| 30000       | 0.30        |
| 35000       | 0.20        |
| 40000       | 0.15        |
| 45000       | 0.10        |

$E(\text{profit}) = ?$

$$\sum_{i=1}^5 x_i p_i = 25000 * 0.25 + 30000 * 0.30 + 35000 * 0.20 + 40000 * 0.15 + 45000 * 0.10$$

= 32750 ដោយគណនាតាមតារាងរបាយប្រូបាប៊ីលីតេ ។

ក្នុងករណីដែលតារាងរបាយប្រូបាប៊ីលីតេពុំស្គាល់  $E(X)$  កំណត់ដោយ:

$$E(\text{profit}) = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{25000 + 30000 + 35000 + 40000 + 45000}{5} = 35000$$

ក្នុងករណីដែល  $n$  ធំ  $\sum_{i=1}^n x_i p_i \cong \frac{1}{n} \sum_{i=1}^n x_i$  ។

Standard deviation(profit) = ?

$$\text{Std.deviation} = \sqrt{\sum_{i=1}^5 (x_i - E(X))^2 * p_i} = 6417.749$$

ក្នុងករណីដែលពុំស្គាល់តារាងរបាយប្រូបាប៊ីលីតេ std.deviation កំណត់ដោយ:

$$\text{std.deviation} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 = 7905.694$$

$$\text{ក្នុងករណីដែល } n \text{ ធំ } \sqrt{\sum_{i=1}^n (x_i - E(X))^2 * p_i} \cong \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

#### ៤-មេគុណបំរែបំរួល (coefficient of variation):

ជាទូទៅមេគុណបំរែបំរួលកំណត់ដោយ :

$$\text{c.v\%} = \frac{\sigma}{E(X)} * 100$$

នៅក្នុងការអនុវត្តផ្នែកសេដ្ឋកិច្ច coefficient of variation គឺជា Risk ។ កាលណា តំលៃ coefficient of variation ធំបានសេចក្តីថាគំរោងការងារនេះមានហានិភ័យខ្ពស់។ កាលណា coefficient of variation មានតំលៃតូចជាង 15% បានសេចក្តីថា Risk តូច របាយការណ៍ នៃការអង្កេតមានលក្ខណៈស្មើសាច់ (មានភាពប្រែប្រួលតិចពីអថេរមួយទៅអថេរមួយ) យើងក៏អាចនិយាយផងដែរថា standard deviation តូច ។

#### ឧទាហរណ៍:

សហគមន៍ “**optigestion**” មានគំរោងទិញភាគហ៊ុននៅក្នុងសហគ្រាសមួយកន្លែងដែលផលិត micro-informatics ។ គេបានពិនិត្យមើលលើតារាងលទ្ធផលក្នុងរយៈពេលមួយឆ្នាំ ដែលមាន ស្រាប់នៅក្នុងសហគ្រាសនេះ ។

| Output % | Probability p <sub>i</sub> |
|----------|----------------------------|
| 28.0     | 0.05                       |
| 23.0     | 0.21                       |
| 18.6     | 0.34                       |
| 15.0     | 0.22                       |
| 12.0     | 0.10                       |
| 8.0      | 0.08                       |

ក-តើសហគមន៍ **optigestion** អាចនឹងសង្ឃឹមថាទទួលបាន output ប៉ុន្មានដែរ ?

ខ-គណនាគំរោងគំរូ (standard deviation) និងមេគុណបំរែបំរួល (cv) នៃទិន្នផល (output) ។

គ- សហគមន៍នេះបានសិក្សាផងដែរទៅលើការទិញកម្ចីប័ណ្ណរដ្ឋាភិបាល (Government obligation) ដែលមានទិន្នផល 12% ថេរ ។ គណនាមេគុណបំរែបំរួល (C.V) ។

ឃ-តើការទិញប័ណ្ណភាគហ៊ុនប្រភេទណាដែលមាន Risk ធំជាងគេ?

**ចំណើយ:**

ដោយប្រើ program S-plus ដូចខាងក្រោម :

```
function()
{
  output <- c(28, 23.4, 18.6, 15, 12, 8)
  prob <- c(0.05, 0.21, 0.34, 0.22, 0.1, 0.08)
  output.square <- output^2
  prod <- (output^2) * prob
  Table <- cbind(output, output.square, prob, prod)
  expectation <- sum(output * prob)
  std.dev <- sqrt(sum((output - expectation)^2 * prob))
  C.V <- (std.dev/expectation) * 100
  return(Table, expectation, std.dev, C.V)
}
```

យើងទទួលបានចំណើយដូចខាងក្រោម :

```
$Table:
      output output.square prob  prod
[1,]  28.0      784.0    0.05 39.2000
[2,]  23.4      547.56    0.21 114.9876
[3,]  18.6      345.96    0.34 117.6264
[4,]  15.0      225.00    0.22  49.5000
[5,]  12.0      144.00    0.10  14.4000
[6,]   8.0       64.00    0.08   5.1200
```

\$expectation:

[1] 17.778

\$std.dev:

[1] 4.977622

\$C.V:

[1] 27.99877

យោងទៅលើចំណើយខាងលើយើងបាន :

$$\text{ក-} E(X) = \sum_{i=1}^6 x_i p_i = 17.778$$

$$\text{ខ-} V(X) = E(X - E(X))^2 = E(X^2) - E^2(X) \Rightarrow \sigma(X) = 4.977622$$

$$\text{គ-} C.V \% = \frac{\sigma(X)}{E(X)} * 100 = 27.998$$

ឃ-ដោយ Government obligation មានទិន្នផលជាក់លាក់ 12% ថេរ នោះ

standard deviation ស្មើ 0 នាំអោយ  $C.V=0\%$  ។

| Investment                                     | Risk           |
|------------------------------------------------|----------------|
| Investment នៅក្នុងសហគ្រាស<br>Micro-informatics | $C.V=28.998\%$ |
| Government obligation                          | $C.V=0\%$      |

#### ៥-លក្ខណៈសំគាល់នៃសង្ខ័យគណិត និង វ៉ារ្យង់ :

##### ក- សង្ខ័យគណិត:

$E(X+c)=E(X)+c$  ដែល  $c$  ជាតំលៃថេរ

$E(aX)=aE(X)$  ដែល  $a$  ជាតំលៃថេរ

$E(aX+c)=aE(X)+c$

##### ខ- វ៉ារ្យង់:

$V(X+c)=V(X)$  ដែល  $c$  ជាចំនួនថេរ

$V(aX)=a^2V(X)$

$V(aX+c)=a^2V(X)$

##### គ- គំលាតស្តង់ដា (standard deviation)

$\sigma(X+c)=\sigma(X)$

$\sigma(aX)=a\sigma(X)$

$\sigma(aX+c)=a\sigma(X)$

##### ឆ- វិសមភាព Tchebychev:

សន្មតថា  $X$  ជាអថេរចៃដន្យ ដែលមានច្បាប់ប្រូបាប៊ីលីតេមួយកំណត់ មានសង្ខ័យគណិត  $E(X)$  និង

វ៉ារ្យង់  $V(X)=\sigma^2(X)$  ។

វិសមភាព Tchebychev កំណត់ដូចខាងក្រោម :

$$P[E(X)-k\sigma(X) \leq X \leq E(X)+k\sigma(X)] \geq 1 - \frac{1}{k^2}$$

ដែល  $k$  ជាចំនួនថេរ ។

វិសមភាពនេះផ្ទៀងផ្ទាត់នូវគ្រប់ច្បាប់ប្រូបាប៊ីលីតេនៃអថេរចៃដន្យ (ដាច់ ឬ ជាប់) ក្នុងលក្ខខណ្ឌដែល

យើងអាចកំណត់បាននូវតំលៃ  $E(X)$  និង តំលៃ  $\sigma(X)$  ។

## Splus និងការវិភាគទិន្នន័យ(data analysis)

Splus ជា software មួយប្រភេទដូច SPSS ដែរ ប៉ុន្តែវាមានលក្ខណៈវិភាគស៊ីជម្រៅជាង SPSS និង ជា programming language ។ Splus ជាប្រភេទ software ដែលផ្តល់លក្ខណៈងាយស្រួលក្នុងការបង្កើត graphics ។

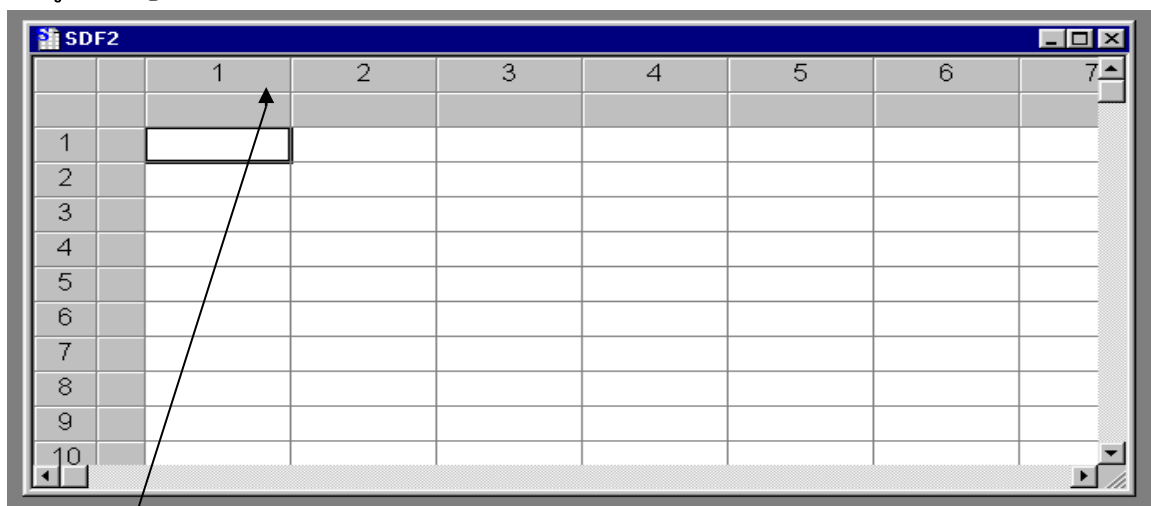
### ១- ការបញ្ចូលទិន្នន័យ (data input) :

ដើម្បីបញ្ចូល ទិន្នន័យ យើងត្រូវហៅ data editor window :



new data frame

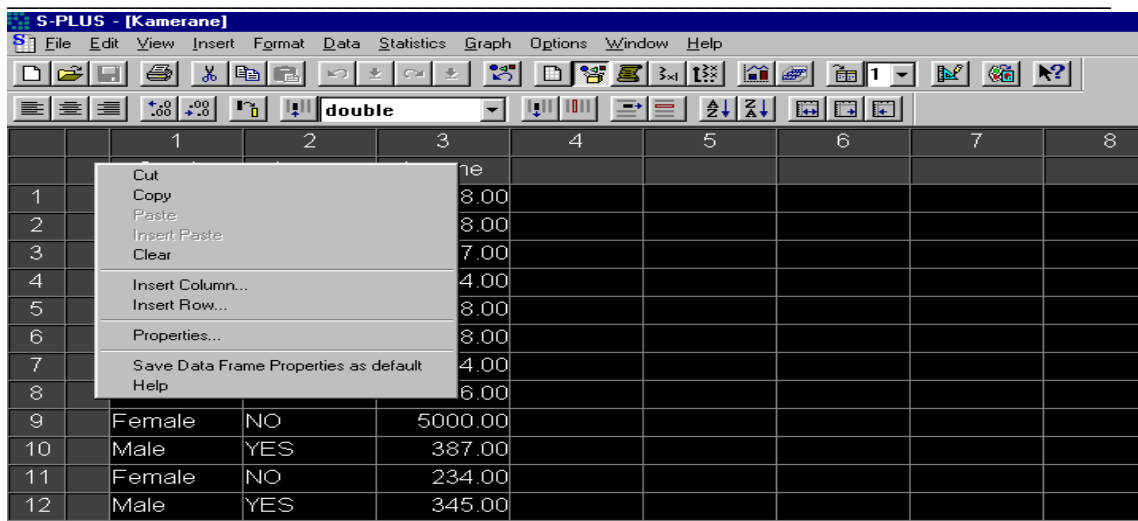
ចុចទៅលើ Tool ដែលមានឈ្មោះថា new data frame យើងទទួលបាន data editor window ដែលមានរាងដូចខាងក្រោម:



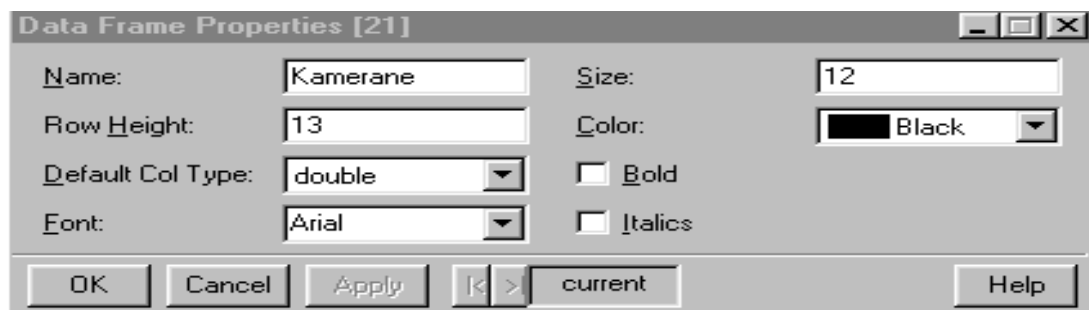
ចុចពីរដងទៅលើ ថតខាងលើដើម្បីបញ្ចូល variable name ។

ដូចគ្នានេះដែរចុចពីរដងទៅលើថតនៃជួរដេកនេះដើម្បីបញ្ចូល name of observation ។





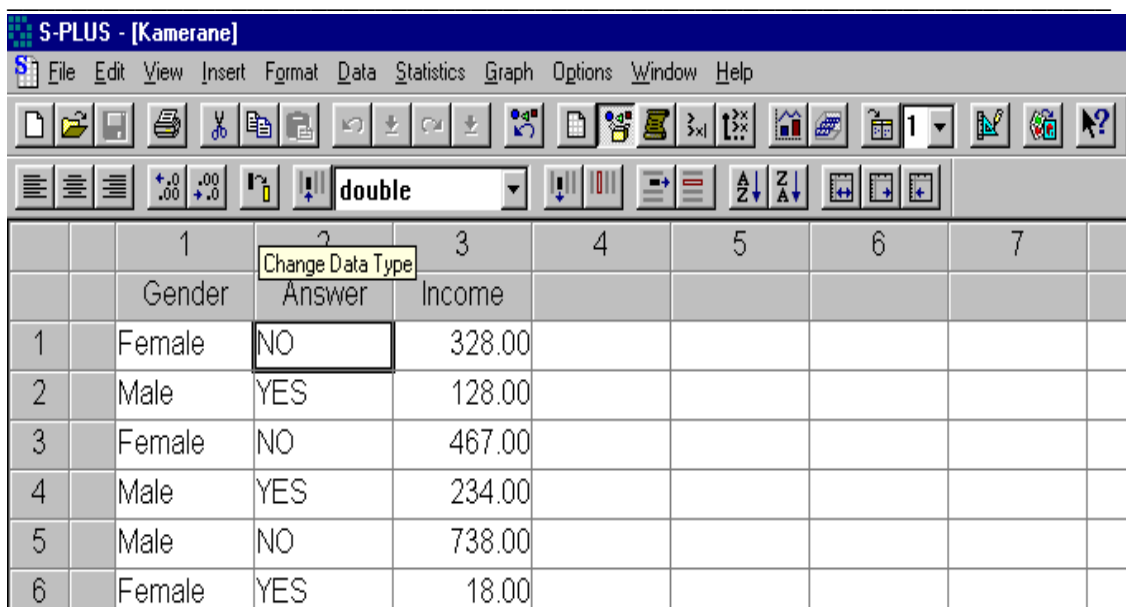
នៅពេលបញ្ចូលទិន្នន័យចប់ ចូរសង្កេតមើលឈ្មោះទិន្នន័យនៅផ្នែកខាងលើនៃ data editor window ដែលអោយដោយ default របស់ computer ។ យើងអាចដូរឈ្មោះទិន្នន័យនេះដោយ គ្រាន់តែ select data editor window ហើយចុច mouse ខាងស្តាំ select យក properties ហើយដូរឈ្មោះដែលយើងចង់បាន:



ពេលនេះយើងអាចប្តូរឈ្មោះ data editor window នៅក្នុងថត Name តាមដែលយើងចង់បាន ។

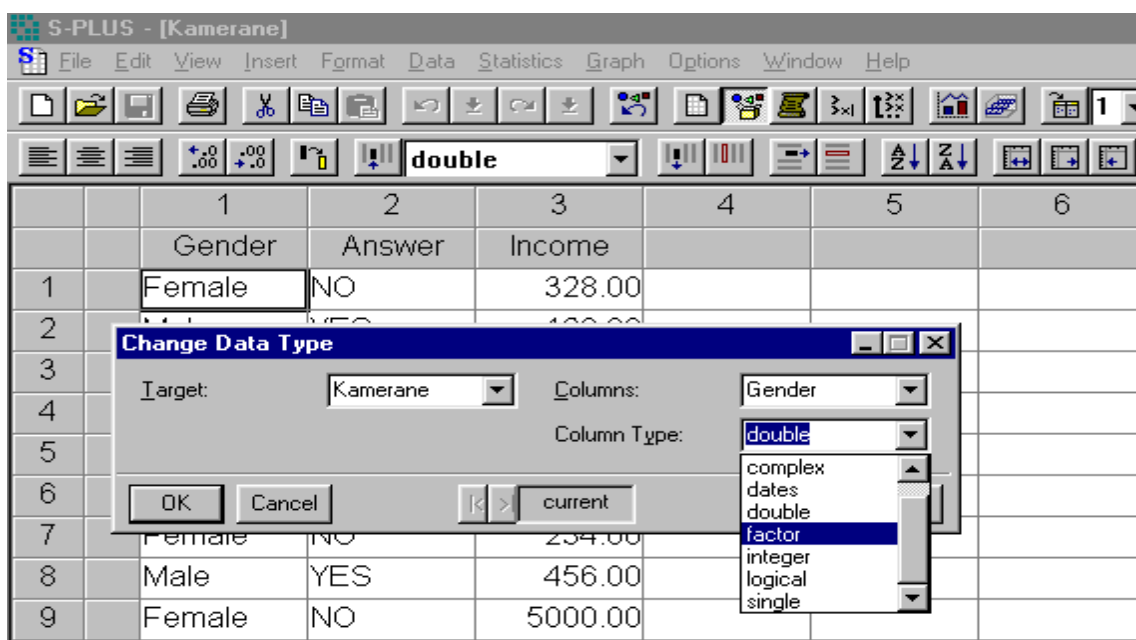
## ២-លក្ខណៈនៃទិន្នន័យ (Data type) :

ទិន្នន័យដែលយើងប្រើប្រាស់ក្នុងកញ្ចប់មានលក្ខណៈជា numeric, string, dates,...និងមានមួយចំនួនទៀត ដែលយើងអាចប្រើបាននៅក្នុង change data type tool ដែលមានរូបរាងដូចខាងក្រោម:



|   | 1      | 2      | 3      | 4 | 5 | 6 | 7 |
|---|--------|--------|--------|---|---|---|---|
|   | Gender | Answer | Income |   |   |   |   |
| 1 | Female | NO     | 328.00 |   |   |   |   |
| 2 | Male   | YES    | 128.00 |   |   |   |   |
| 3 | Female | NO     | 467.00 |   |   |   |   |
| 4 | Male   | YES    | 234.00 |   |   |   |   |
| 5 | Male   | NO     | 738.00 |   |   |   |   |
| 6 | Female | YES    | 18.00  |   |   |   |   |

មុននឹងបញ្ចូលទិន្នន័យយើងត្រូវសង្កេតទៅលើ data type សិន តើ វាជា numeric , string ឬប្រភេទទិន្នន័យបែបណា ។ ឧទាហរណ៍នៅក្នុងតារាងខាងលើនៅ column ទីមួយទិន្នន័យមានលក្ខណៈ ជា string ដូច្នេះនៅ ក្នុងការបញ្ចូលទិន្នន័យនេះ យើងត្រូវបញ្ចូល variable name រួច change data type:



បន្ទាប់មក select យក factor ព្រោះថាទិន្នន័យរបស់យើងមាន factor ពីរ គឺ Female និង Male ។  
ដូចគ្នានេះដែរចំពោះ column ទីពីរ ប៉ុន្តែ column ទី 3 ទិន្នន័យមានលក្ខណៈជា numerical data ដូច្នេះ  
ការបញ្ចូលពុំត្រូវបានឆ្លងកាត់ការ change data type ទេ ។

### ៣- ស្ថិតិពិពណ៌នា (Descriptive analysis) :

ការវិភាគទិន្នន័យមានលក្ខណៈស្មុគស្មាញទៅតាមប្រធានបទផ្សេងៗគ្នា ។ នៅពេលនេះយើងលើកយកតែ  
ការវិភាគដែលមានលក្ខណៈជា Descriptive analysis ។ ឧទាហរណ៍យើងមានទិន្នន័យដូចខាងក្រោម:

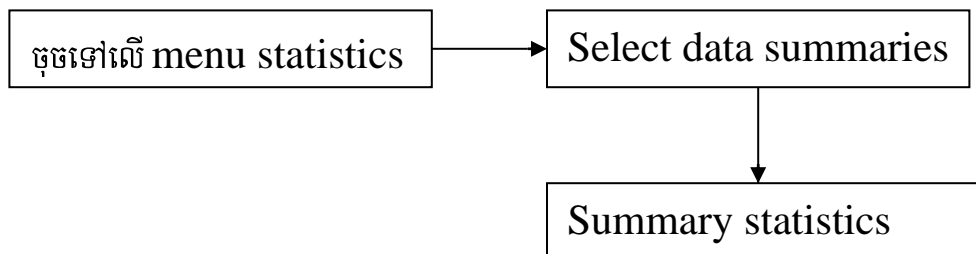
| CUSTOMER              | REVENUE    | REGION | INDUSTRY   |
|-----------------------|------------|--------|------------|
| 1 Regular customer    | 3786.8670  | West   | Government |
| 2 Regular customer    | 1733.6005  | South  | Commercial |
| 3 Preferred customer  | 2125.7356  | West   | Government |
| 4 Preferred customer  | 2259.1379  | West   | Commercial |
| 5 Preferred customer  | 1586.6104  | North  | Government |
| 6 Regular customer    | NA         | West   | Academic   |
| 7 Preferred customer  | 1838.4591  | South  | Government |
| 8 Regular customer    | 1846.7495  | West   | Government |
| 9 Regular customer    | 1714.1120  | East   | Academic   |
| 10 Preferred customer | 1717.5371  | South  | Academic   |
| 11 Preferred customer | 4388.0708  | West   | Government |
| 12 Preferred customer | 12346.0000 | North  | Commercial |
| 13 Preferred customer | 3833.5526  | West   | Government |
| 14 Preferred customer | 2290.6734  | North  | Commercial |
| 15 Preferred customer | 4139.7650  | South  | Commercial |
| 16 Preferred customer | 2193.9056  | South  | Academic   |
| 17 Preferred customer | 2938.3083  | South  | Government |
| 18 Regular customer   | 3313.3158  | East   | Commercial |
| 19 Preferred customer | 3327.0176  | East   | Academic   |
| 20 Regular customer   | 1449.3379  | East   | Government |
| 21 Preferred customer | 2695.6158  | North  | Academic   |
| 22 Preferred customer | 3272.2086  | South  | Commercial |
| 23 Regular customer   | 4096.3530  | North  | Academic   |
| 24 Regular customer   | 4339.9974  | South  | Academic   |
| 25 Regular customer   | 2462.3752  | West   | Government |
| 26 Preferred customer | 3629.5420  | East   | Commercial |
| 27 Preferred customer | 465.1717   | West   | Academic   |
| 28 Preferred customer | 1870.9232  | North  | Commercial |
| 29 Preferred customer | 1376.5828  | West   | Academic   |
| 30 Regular customer   | 3796.7922  | South  | Government |
| 31 Preferred customer | 4253.4813  | North  | Academic   |
| 32 Preferred customer | 3057.0979  | South  | Government |
| 33 Preferred customer | 3394.7124  | West   | Government |
| 34 Regular customer   | 1387.3440  | West   | Government |
| 35 Preferred customer | 2148.4856  | North  | Commercial |
| 36 Preferred customer | 2204.7234  | West   | Commercial |
| 37 Preferred customer | 2021.8557  | South  | Academic   |
| 38 Regular customer   | 123.0000   | East   | Government |
| 39 Preferred customer | 3301.1004  | West   | Academic   |

|    |                    |            |       |            |
|----|--------------------|------------|-------|------------|
| 40 | Regular customer   | 1855.5423  | South | Government |
| 41 | Preferred customer | 2885.1827  | East  | Government |
| 42 | Preferred customer | 3353.5838  | East  | Academic   |
| 43 | Preferred customer | 2263.7547  | East  | Commercial |
| 44 | Regular customer   | 1620.7525  | West  | Government |
| 45 | Regular customer   | 3398.7307  | East  | Academic   |
| 46 | Preferred customer | 3987.9912  | West  | Government |
| 47 | Preferred customer | 3808.7805  | East  | Academic   |
| 48 | Preferred customer | 3605.4659  | West  | Academic   |
| 49 | Regular customer   | 3226.4753  | West  | Commercial |
| 50 | Regular customer   | 4356.6054  | South | Commercial |
| 51 | Regular customer   | 2120.6041  | East  | Academic   |
| 52 | Preferred customer | 3943.9258  | North | Government |
| 53 | Regular customer   | 3689.8859  | West  | Commercial |
| 54 | Regular customer   | 430.1497   | West  | Commercial |
| 55 | Preferred customer | 2267.0971  | North | Government |
| 56 | Preferred customer | 1123.1291  | South | Academic   |
| 57 | Preferred customer | 2384.9484  | West  | Commercial |
| 58 | Regular customer   | 5011.5981  | West  | Academic   |
| 59 | Preferred customer | 34566.0000 | West  | Commercial |
| 60 | Preferred customer | 2742.6422  | West  | Commercial |
| 61 | Preferred customer | 2608.7393  | North | Commercial |
| 62 | Preferred customer | 2261.5493  | West  | Commercial |
| 63 | Regular customer   | 2744.7494  | West  | Academic   |
| 64 | Regular customer   | 4024.3463  | North | Commercial |
| 65 | Preferred customer | 4728.5242  | South | Commercial |
| 66 | Preferred customer | 3185.4868  | North | Academic   |
| 67 | Regular customer   | 2406.9697  | East  | Commercial |
| 68 | Preferred customer | 5417.3652  | East  | Commercial |
| 69 | Regular customer   | 3875.5855  | East  | Government |
| 70 | Preferred customer | 1902.2245  | East  | Commercial |
| 71 | Regular customer   | 836.1970   | East  | Government |
| 72 | Regular customer   | 3750.3950  | South | Academic   |
| 73 | Preferred customer | 3418.8871  | East  | Government |
| 74 | Regular customer   | 2068.3873  | North | Commercial |
| 75 | Regular customer   | 3810.8523  | West  | Commercial |
| 76 | Preferred customer | 432.0000   | South | Academic   |
| 77 | Preferred customer | 430.4231   | South | Academic   |
| 78 | Preferred customer | 718.5732   | North | Commercial |
| 79 | Preferred customer | 2570.7565  | North | Commercial |
| 80 | Preferred customer | 2671.3214  | West  | Commercial |
| 81 | Preferred customer | 2241.6868  | South | Academic   |
| 82 | Regular customer   | 2409.5362  | South | Government |
| 83 | Regular customer   | 54322.0000 | West  | Government |
| 84 | Regular customer   | 1547.8898  | East  | Academic   |
| 85 | Preferred customer | 2233.1350  | West  | Government |
| 86 | Regular customer   | 1283.8048  | South | Government |
| 87 | Preferred customer | 2658.7289  | North | Government |
| 88 | Regular customer   | 2823.8458  | East  | Government |
| 89 | Preferred customer | 336.6098   | East  | Commercial |
| 90 | Preferred customer | 2120.1780  | East  | Academic   |
| 91 | Preferred customer | 2917.7317  | East  | Government |
| 92 | Preferred customer | 2370.9864  | North | Commercial |
| 93 | Regular customer   | 2696.9629  | East  | Academic   |
| 94 | Regular customer   | 1975.5370  | East  | Academic   |

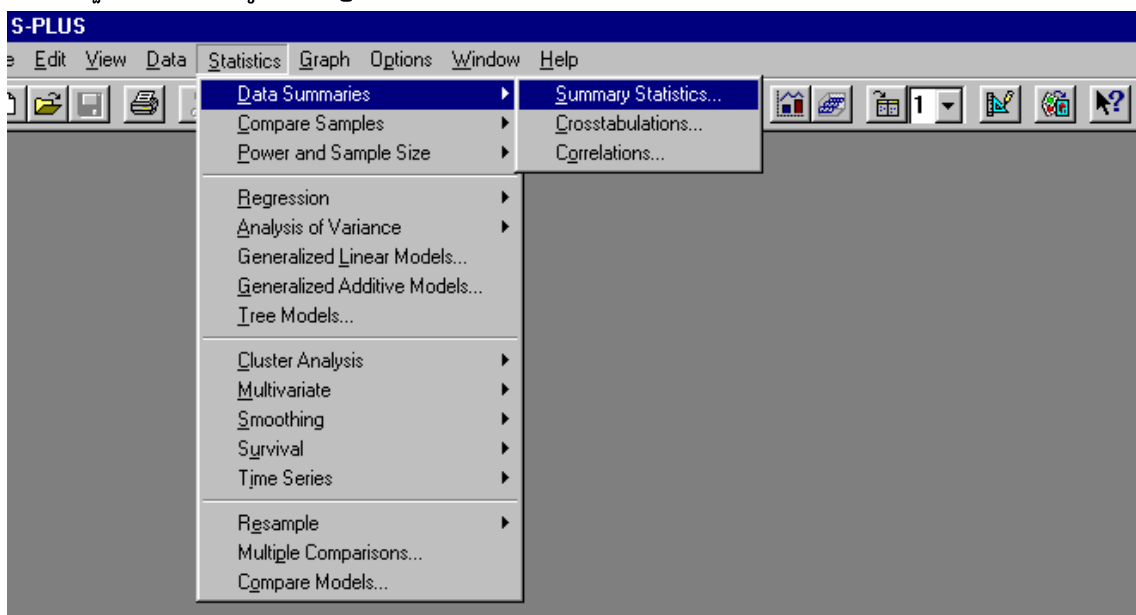
|     |                    |           |       |            |
|-----|--------------------|-----------|-------|------------|
| 95  | Regular customer   | 2704.2350 | East  | Government |
| 96  | Regular customer   | 3658.2319 | North | Government |
| 97  | Regular customer   | 3501.6697 | East  | Government |
| 98  | Preferred customer | 2895.8984 | North | Government |
| 99  | Regular customer   | 1429.5769 | East  | Commercial |
| 100 | Regular customer   | 3913.8769 | South | Academic   |

ដោយ column ទី 1,3,4 ទិន្នន័យមានលក្ខណៈជា string (factor) ដូច្នេះយើងធ្វើ descriptive analysis

ទៅលើ columns ទាំងនេះដូចខាងក្រោម:



យើងទទួលបានតារាងដូចខាងក្រោម :



ជ្រើសរើសយក variable names ដែលយើងចង់ធ្វើការវិភាគ ហើយរក្សាទុកនូវផ្នែកខាងក្រោមគឺ Result នូវ summarize Categorical Variables និង Print Results ព្រោះទិន្នន័យមានលក្ខណៈជា string ដូច្នេះការវិភាគត្រូវធ្វើឡើងតាម frequency analysis ។ នៅក្នុងតារាងខាងលើយើង select Customer បន្ទាប់មក select Region and Industry រួចចុច OK យើងទទួលបានចំណេញដូចខាងក្រោម៖

**\*\*\* Summary Statistics for data in: sales \*\*\***

**CUSTOMER**

**Preferred customer:58**

**Regular customer:42**

**\*\*\* Summary Statistics for data in: sales \*\*\***

**REGION**

**East:28**

**North:20**

**South:22**

**West:30**

**INDUSTRY**

**Academic:31**

**Commercial:34**

**Government:35**

ដូច្នេះយើងឃើញថា summary statistics បានរាប់ចំនួន factor និងមួយៗ នៅក្នុង variable names ដែលយើងធ្វើការវិភាគ ។

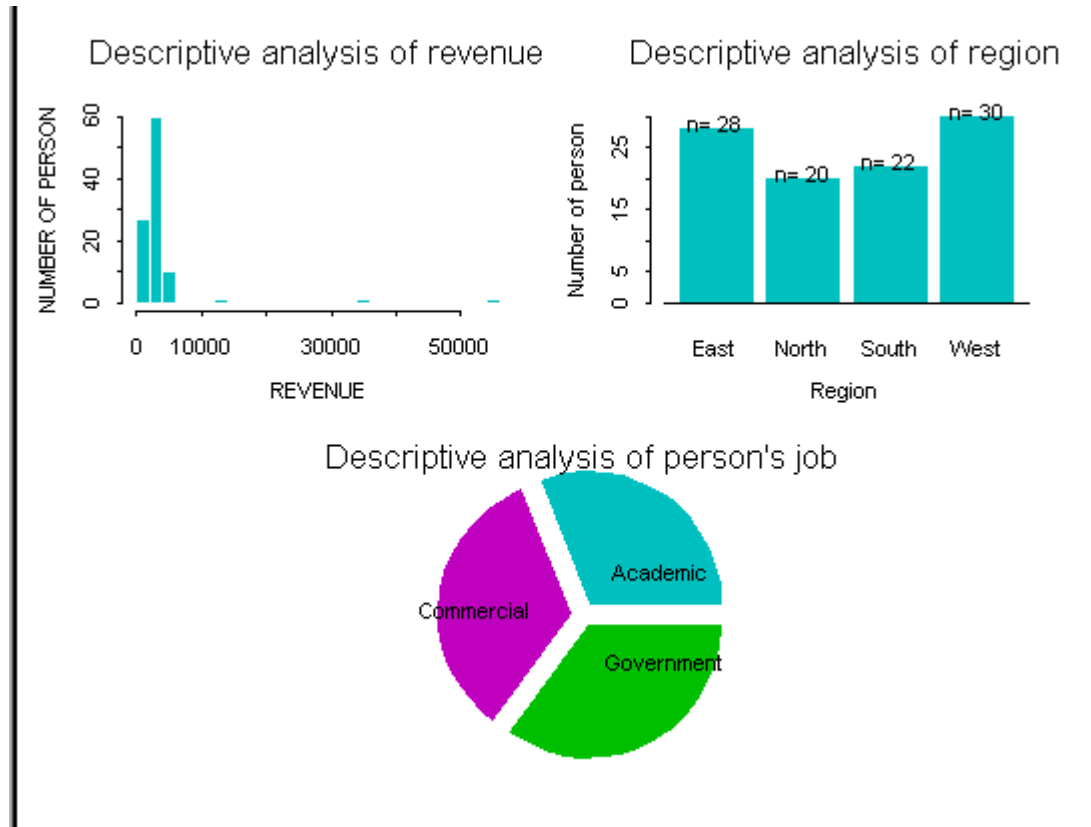
#### **៤- វិភាគក្រាហ្វិច (Graphical analysis):**

ដោយអនុវត្តទៅលើទិន្នន័យខាងលើហើយដោយប្រើកម្មវិធី S-plus (programming) យើងអាចធ្វើ data summary និង graphical analysis ដូចខាងក្រោម:

```
function(sales)
{
  i <- 1:3
  analysis <- summary(sales[, i])
  par(mfg = c(1, 1, 2, 2))
  graph1 <- hist(sales[, 1], nclass = 30, xlab =
    "REVENUE", ylab = "NUMBER OF PERSON",
    main = "Descriptive analysis of revenue",
  )
  par(mfg = c(1, 2, 2, 2))
  graph2 <- barplot(summary(sales[, 2]), names =
    levels(factor(sales[, 2])), xlab =
    "Region", ylab = "Number of person",
    main = "Descriptive analysis of region"
  )
  text(graph2, summary(sales[, 2]) + 1, paste(
    "n=", summary(sales[, 2]))
  )
  par(mfg = c(2, 1, 2, 1))
  graph3 <- pie(summary(sales[, 3]), names =
    levels(factor(sales[, 3])), col = 2:4,
    explode = 1:3, size = 1.3, rotate = F,
    main = "Descriptive analysis of person's job"
  )
  return(analysis)
}
```

#### **output of data analysis :**

| REVENUE       | REGION   | INDUSTRY      |
|---------------|----------|---------------|
| Min.: 123     | East:28  | Academic:31   |
| 1st Qu.: 1890 | North:20 | Commercial:34 |
| Median: 2670  | South:22 | Government:35 |
| Mean: 3570    | West:30  |               |
| 3rd Qu.: 3640 |          |               |
| Max.:54300    |          |               |





## ច្បាប់ណរម៉ាល់និងការអនុវត្ត

អថេរចៃដន្យជាប់  $X$  គោរពតាមច្បាប់ណរម៉ាល់កាលណាវាមានរបាយដង់ស៊ីតេដូចខាងក្រោម៖

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < +\infty$$

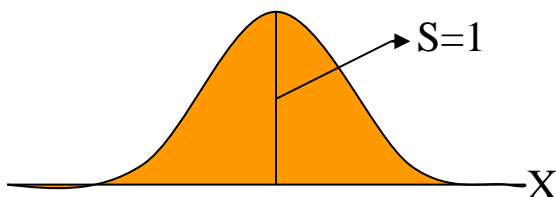
ច្បាប់ណរម៉ាល់អាស្រ័យទៅនឹងតំលៃ  $\mu$  និង  $\sigma$  ។

**ចំណាំ៖**

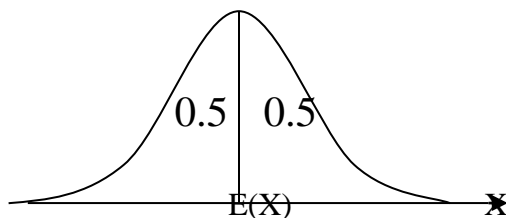
$X$  គោរពតាមច្បាប់ណរម៉ាល់ដែលមានសង្ខេបគណិត  $E(X) = \mu$  និង វ៉ារ្យង់  $\sigma(X) = \sigma$  សរសេរ  $N(\mu, \sigma^2)$  ។

### ១- លក្ខណៈសំគាល់អនុគមន៍របាយដង់ស៊ីតេ Normal:

ក-ដោយច្បាប់ណរម៉ាល់ជាច្បាប់ប្រូបាប៊ីលីតេនោះក្រលាផ្ទៃដែលខ័ណ្ឌដោយខ្សែកោងណរម៉ាល់និងអ័ក្សអាប់ស៊ីសស្មើ 1 ។



ខ- ខ្សែកោងណរម៉ាល់មានលក្ខណៈឆ្លុះធៀបនឹងតំលៃសង្ខេបគណិត  $E(X)$  ។



គ- ដោយសារខ្សែកោងណ័រម៉ាល់មានលក្ខណៈឆ្លុះនោះ ៖

$$\text{mean}(X)=\text{Median}(X)=\text{Mode}(X)$$

### ចំណាំ៖

#### Median :

Median(X) គឺជាតំលៃដែលចែក population ជាពីរផ្នែកស្មើគ្នា (X ត្រូវបានរៀបតាមលំដាប់កើន ឬ ចុះ) ។

#### ឧទាហរណ៍៖

គេមានរបាយទិន្នន័យដូចខាងក្រោម៖

X: 270 240 150 600 450 500 670

រៀបទិន្នន័យតាមលំដាប់កើន៖

X: 150 240 270 **450** 500 600 670

**Median(X)= 450**

នៅក្នុង S-plus គេសរសេរ ៖

>X<-c(270 ,240, 150, 600, 450, 500, 670)

> median(X)

[1] **450**

ក្នុងករណីដែល X មានរបាយទិន្នន័យដែលមានបរិមាណជាចំនួនគូ

X: 150 240 270 **450 500** 600 670 700

**Median(X)=(450+500)/2**

នៅក្នុង S-plus សរសេរ ៖

>median(X)

[1] **475**

#### Mode :

Mode(X) គឺជាតំលៃអថេរចែងនូវ X ដែលមានប្រេកង់ (frequency) ខ្ពស់ជាងគេ ។

#### ឧទាហរណ៍៖

គេមានរបាយទិន្នន័យ X ដូចខាងក្រោម៖

X: **20** 30 25 **20** 30 **20 20** 15 **20** 35 **20** 54 28

**Mode(X)=20**

**ឃ-** 
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

## ឧទាហរណ៍៖

លទ្ធផលនៃការអង្កេតនៅក្នុងទីក្រុងភ្នំពេញបានអោយដឹងថា ការចំណាយប្រចាំខែជាមធ្យមរបស់គ្រួសារនីមួយៗ មានរបាយទិន្នន័យ Normal  $N(\$250,(\$100)^2)$  ។ តើមានគ្រួសារប៉ុន្មានភាគរយដែលរស់នៅក្នុងទីក្រុងភ្នំពេញដែលមានការចំណាយប្រចាំខែជាមធ្យមប្រែប្រួលនៅក្នុងចន្លោះ[\$180,\$320]?

បើ  $X$  ជាអថេរនៃការចំណាយជាមធ្យមប្រចាំខែរបស់គ្រួសារនីមួយៗ នោះ៖

$$P(\$180 \leq X \leq \$320) = \int_{180}^{320} \frac{1}{100\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-250}{100}\right)^2} dx$$

ចំណេញនៃអាំងតេក្រាលនេះតំណាងអោយភាគរយនៃគ្រួសារដែលមានការចំណាយជាមធ្យមប្រចាំខែនៅក្នុងចន្លោះពី \$180 ទៅ \$320 ។

**S-plus:**

**pnorm(x,mean,std.dev)** អាចប្រើជំនួយសំរាប់ការគណនាអាំងតេក្រាលនៃរបាយទិន្នន័យ Normal ដែលមានទំរង់ដូចខាងក្រោម៖

$$P(X \leq x) = \int_{-\infty}^x f(x)dx, \text{ ដែល } f(x) \text{ ជាអនុគមន៍ដង់ស៊ីតេ Normal}$$

ដោយអនុវត្តទៅលើឧទាហរណ៍ខាងលើ យើងអាចសរសេរ៖

**>pnorm(320,250,100)-pnorm(180,250,100)**  
**[1] 0.5160727**

$$P(\$180 \leq X \leq 320) = 0.5160 = 51.60\%$$

**ង-** បើ  $X$  ជាអថេរនៃរបាយទិន្នន័យជាប់ នោះ ៖

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

មានន័យថាសញ្ញា “=” គ្មានន័យអ្វីទាំងអស់។

## **២-ច្បាប់ណរម៉ាល់បង្រួមកណ្តាល** (Standard normal distribution)

គ្រប់អថេរចៃដន្យ  $X$  នៃច្បាប់ណរម៉ាល់ដែលមានតំលៃមធ្យម  $\mu$  និង វ៉ារ្យង់  $\sigma^2$  អាចបំប្លែងទៅជាអថេរណរម៉ាល់បង្រួមកណ្តាល (Standard random variable of normal distribution) ដែលកំណត់ដោយប្រមាណវិធីដូចខាងក្រោម ៖

$$Z = \frac{X - \mu}{\sigma}$$

ដែល  $E(Z)=0$  និង  $V(Z)=1$  ។

ដូច្នេះរបាយដង់ស៊ីតេនៃច្បាប់ណរម៉ាល់ប្រែមកណ្តាលកំណត់ដោយ៖

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < +\infty$$

អថេរចែងណរម៉ាល់

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$Z = \frac{X - \mu}{\sigma}$$

អថេរចែងណរម៉ាល់ប្រែមកណ្តាល

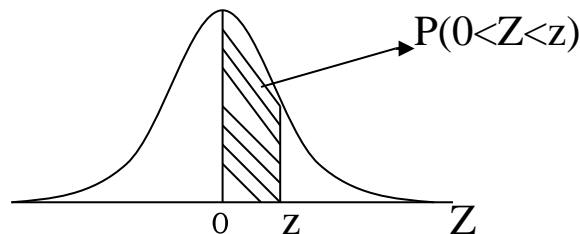
$$Z \sim N(0,1)$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

### ឧទាហរណ៍

សន្មតថា  $Z \sim N(0,1)$  ចូរកំណត់  $P(0 < Z < 0.5)$ ?

ដើម្បីគណនាតំលៃប្រូបាប៊ីលីតេ នេះយើងអាចប្រើតារាងណរម៉ាល់ប្រែមកណ្តាលដែលមានស្រាប់ ឬក៏ប្រើ S-plus ដើម្បីសំរួលក្នុងការគណនា។ តារាង ណរម៉ាល់ប្រែមកណ្តាលអាចអនុញ្ញាតអោយយើងគណនាក្រលាផ្ទៃ ដែល  $Z$  ប្រែប្រួលពីតំលៃ 0 ទៅតំលៃ  $Z$  ណាមួយដែលមានរូបភាពដូចខាងក្រោម៖



### S-plus:

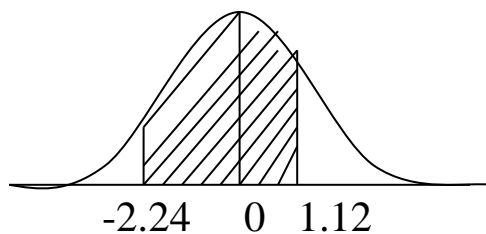
នៅក្នុង S-plus យើងអាចប្រើ command **pnorm(z)** ដែលអនុញ្ញាតអោយគណនាក្រឡាផ្ទៃដែល  $Z$  ប្រែប្រួលពី  $-\infty$  មកតំលៃ  $Z$  ណាមួយ។

```
>pnorm(0.5, 0, 1)- 0.5
```

```
[1] 0.1914625
```

**$P(0 < Z < 0.5) = 0.1915$**

ដោយប្រើតារាង Normal standard ចូរគណនា  $P(-2.24 < Z < 1.12)$ ?



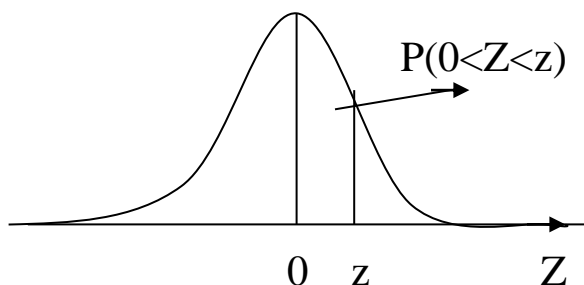
$$\begin{aligned} P(-2.24 < Z < 1.12) &= P(-2.24 < Z < 0) + P(0 < Z < 1.12) \\ &= P(0 < Z < 2.24) + P(0 < Z < 1.12) \\ &= 0.4875 + 0.3686 = 0.8561 \end{aligned}$$

| z   | 0.00 | 0.01 | 0.02   | 0.03   | 0.04 | 0.05 | ..... | 0.09 |
|-----|------|------|--------|--------|------|------|-------|------|
| 0.0 |      |      |        |        |      |      |       |      |
| 0.1 |      |      |        |        |      |      |       |      |
| 0.2 |      |      |        |        |      |      |       |      |
| 1.0 |      |      |        |        |      |      |       |      |
| 1.1 |      |      | 0.3686 |        |      |      |       |      |
| 1.2 |      |      |        |        |      |      |       |      |
| 2.0 |      |      |        |        |      |      |       |      |
| 2.1 |      |      |        |        |      |      |       |      |
| 2.2 |      |      |        | 0.4875 |      |      |       |      |

**Splus:**

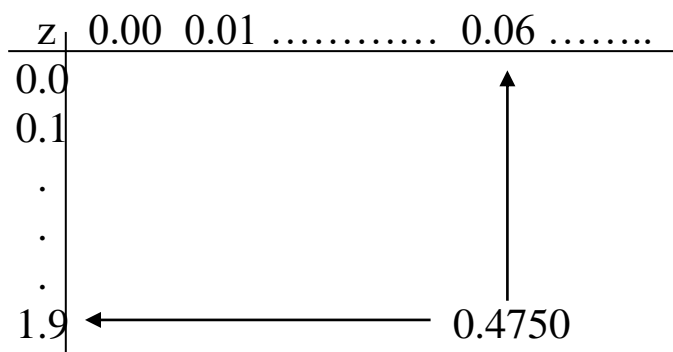
```
> pnorm(1.12)-pnorm(-2.24)
> [1] 0.8560977
```

ក្នុងករណីដែលយើងស្គាល់ក្រលាផ្ទៃដែលខ័ណ្ឌដោយខ្សែកោងណរម៉ាល់បង្រួមកណ្តាលនិងអ័ក្សអាប់ស៊ីស យើង អាចកំណត់តំលៃ  $z$  នៅលើអ័ក្ស  $Z$  ដោយប្រើតារាងឬ ដោយអនុវត្តតាម S-plus programming ដូចខាងក្រោម:



**ឧទាហរណ៍ :** ចូរគណនា  $z$  ដែល  $P(0 < Z < z) = 0.4750$  ។

ដោយប្រើតារាងណរម៉ាល់បង្រួមកណ្តាលនាំអោយ  $z = 1.96$



**S-plus:**

qnorm= quantile of normal distribution

qnorm(prob,mean,std.dev)

>qnorm(0.475,0,1)

[1] 1.959964

ដោយប្រើ S-plus programming យើងបង្កើតតារាងណរម៉ាល់បង្រួមកណ្តាលដែលអនុញ្ញាតអោយយើង

អាចគណនា  $P(0 < Z < z)$  ដែល  $Z$  ជាអថេរណរម៉ាល់បង្រួមកណ្តាល:

```
function()
{
  z1 <- seq(0, 0.09, 0.01)
  z2 <- seq(0, 3.9, 0.1)
  z <- seq(0, 3.99, 0.01)
  prob <- pnorm(z) - 0.5
  Table <- t(array(prob, dim = c(length(z1),
    length(z2))))
  Table <- cbind(c(z2), Table)
  Table <- rbind(c(NA, z1), Table)
  return(Table)
}
```

Standard of Normal Distribution  $P(0 < Z < z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$

| Z   | 0.00   | 0.01    | 0.02    | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|-----|--------|---------|---------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.0000 | 0.00399 | 0.00798 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.04380 | 0.04776 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.08317 | 0.08706 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.12172 | 0.12552 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.15910 | 0.16276 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.19497 | 0.19847 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.22907 | 0.23237 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.26115 | 0.26424 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.29103 | 0.29389 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.31859 | 0.32121 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.34375 | 0.34614 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.36650 | 0.36864 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.38686 | 0.38877 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.40490 | 0.40658 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.42073 | 0.42220 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.43448 | 0.43574 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.44630 | 0.44738 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.45637 | 0.45728 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.46485 | 0.46562 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.47193 | 0.47257 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.47778 | 0.47831 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.48257 | 0.48300 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.48645 | 0.48679 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.48956 | 0.48983 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.49202 | 0.49224 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.49396 | 0.49413 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.49547 | 0.49560 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.49664 | 0.49674 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.49752 | 0.49760 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4979 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.49819 | 0.49825 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.49869 | 0.49874 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |
| 3.1 | 0.4990 | 0.49906 | 0.49910 | 0.4991 | 0.4992 | 0.4992 | 0.4992 | 0.4992 | 0.4993 | 0.4993 |
| 3.2 | 0.4993 | 0.49934 | 0.49936 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4995 | 0.4995 | 0.4995 |
| 3.3 | 0.4995 | 0.49953 | 0.49955 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4997 |
| 3.4 | 0.4997 | 0.49968 | 0.49969 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4998 |
| 3.5 | 0.4998 | 0.49978 | 0.49978 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 |
| 3.6 | 0.4998 | 0.49985 | 0.49985 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.7 | 0.4999 | 0.49990 | 0.49990 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.8 | 0.4999 | 0.49993 | 0.49993 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.9 | 0.5000 | 0.49995 | 0.49996 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |

### ៣-លំហាត់អនុវត្តន៍:

១- លទ្ធផលនៃការសិក្សាវិភាគទៅលើពិន្ទុនិស្សិត ដែលសិក្សាមុខវិជ្ជា “Introduction to computer “

ដែលមកពី department ផ្សេងៗ បានអោយដឹងថាពិន្ទុនេះគោរពតាមច្បាប់ណរម៉ាល់ ដែលមាន សង្ឃឹមគណិតស្មើ 70 និងវ៉ារ្យង់ ស្មើ 100 ។

ក- កំណត់មេគុណបំរែបំរួល (coefficient of variation) នៃពិន្ទុនិស្សិតទាំងនេះ ។

ខ- បើ X ជាពិន្ទុនិស្សិតទាំងនេះ តើមាននិស្សិតប៉ុន្មានភាគរយដែលមានពិន្ទុខ្ពស់ជាង 82?

គ- តើគេត្រូវជ្រើសរើសពិន្ទុតូចជាងប៉ុន្មានដើម្បីអោយបាននិស្សិត 25% ?

២- ជណ្តើរយន្តនៃអាគារមួយកន្លែងអាចផ្ទុកទំងន់បានតែ 800kg ។ សន្មតថាទំងន់នៃអ្នកប្រើប្រាស់ជណ្តើរ

នេះគោរពតាមច្បាប់ន័រម៉ាល់ដែលមាន មធ្យមស្មើ 80kg និង វ៉ារ្យង់ 100kg ។ តើគេត្រូវអនុញ្ញាត

អោយមនុស្សយ៉ាងច្រើនប៉ុន្មាននាក់ជិះជណ្តើរនេះ បើសិនណាគេចង់អោយប្រូបាប៊ីលីតេនៃការលើស

បន្ទុកក្នុងជណ្តើរនេះមិនលើសពី  $10^{-4}$ ?

#### ចំណាំ:

បើ  $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \dots, X_n \sim N(\mu_n, \sigma_n^2)$

ហើយ  $X_1, X_2, \dots, X_n$  មិនអាស្រ័យគ្នា នោះ  $Y = X_1 + X_2 + \dots + X_n$

$Y \sim N(\mu_1 + \mu_2 + \dots + \mu_n, \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2)$  ។

៣- ក្រុមហ៊ុនមួយនៅទីក្រុងភ្នំពេញមានបំណងចង់នាំចូលនូវប្រភេទម៉ូតូ YAMAHA មួយចំនួន ។ ដើម្បី

អោយទីផ្សារនេះមានលក្ខណៈល្អប្រសើរសំរាប់ក្រុមហ៊ុននេះ លុះត្រាតែ 18% នៃអ្នករស់នៅក្នុងទីក្រុង

ភ្នំពេញមានប្រាក់ចំណូលប្រចាំខែ យ៉ាងតិច \$300 ។ សន្មត ថានៅក្នុងទីក្រុងភ្នំពេញ ចំណូលប្រចាំខែរបស់

គ្រួសារនីមួយៗ គោរពតាមច្បាប់ណរម៉ាល់ដែលមានតំលៃមធ្យម \$250 និងគំលាតស្តង់ដា (standard

deviation) \$100 ។ តើ ភ្នំពេញអាចក្លាយទៅជាទីផ្សារល្អប្រសើររបស់ ក្រុម ហ៊ុននេះ ដែរ ឬទេ?

៤- ក្រុមហ៊ុនមួយមានបំណងចង់បញ្ចេញផលិតផលរបស់ខ្លួននៅក្នុងទីផ្សារមួយកន្លែង ។ ដើម្បីអោយទីផ្សារ

នេះជាតំបន់មួយដែលមានលក្ខណៈល្អប្រសើរសំរាប់ក្រុមហ៊ុននេះលុះត្រាតែ 55% នៃគ្រួសារដែលរស់



នៅក្នុងតំបន់នេះមានប្រាក់ចំណូលប្រចាំឆ្នាំ \$12500 យ៉ាងតិច ។ សន្មតថានៅក្នុងតំបន់នេះប្រាក់ចំណូលប្រចាំ ឆ្នាំជាមធ្យមរបស់គ្រួសារនីមួយៗ ស្មើ \$12000 ហើយដែលមានគំនាតស្តង់ដា (standard deviation) ស្មើ \$2500 ។

តើតំបន់នេះជាទីផ្សារល្អប្រសើរសំរាប់ក្រុមហ៊ុននេះដែរឬទេ ? បើសិនណាប្រាក់ចំណូលប្រចាំឆ្នាំរបស់គ្រួសារនីមួយៗគោរពតាមច្បាប់ ណរម៉ាល់ (normal) ។

៤- ក្រុមហ៊ុនសាងសង់មួយបានបញ្ជាក់ថាដើម្បីសាងសង់អាគារមួយដែលមាន 10 ល្វែង កំពស់ 4 ជាន់ ព្រងទាំងមាន ក្បាច់រចនាល្អប្រណិតត្រូវចំណាយពេលជាមធ្យម 50 សប្តាហ៍ និងគំនាតគំរូ (standard deviation) 5 សប្តាហ៍ ។ ឥឡូវនេះក្រុមហ៊ុននេះមានបំណងចង់ដេញថ្លៃលើគំរោងសាងសង់មួយដែលមានលក្ខណៈដូចខាងលើ ដោយបញ្ជាក់ពីពេលវេលាច្បាស់លាស់ក្នុងការសាងសង់នៅក្នុងកិច្ចសន្យាហើយដែលមានប្រូបាប៊ីលីតេនៃការគោរពកិច្ចសន្យាស្មើ 90% ។

តើក្រុមហ៊ុននេះត្រូវបញ្ជាក់ក្នុងកិច្ចសន្យាចំនួនប៉ុន្មានសប្តាហ៍ដើម្បីសំរេចការសាងសង់នេះបើសិនណាគេសន្មតថាពេលវេលាចាំបាច់ក្នុងការសាងសង់គោរពតាមច្បាប់ណរម៉ាល់(normal) ។



# ប្រព័ន្ធគណិតសម្រាប់វិទ្យាសាស្ត្រ

## ១-ច្បាប់ទ្វេភាព (The Binomial distribution)

ច្បាប់នេះអនុវត្តតែចំពោះបាតុភូតទាំងឡាយណាដែលមានលទ្ធផលតែពីរ៖ ជោគជ័យ (success) ឬ បរាជ័យ (failure) ; បាន ឬមិនបាន ; អាច ឬមិនអាច ; វត្តមាន ឬអវត្តមាន ។ ល។

### អថេរចៃដន្យ Bernoulli :

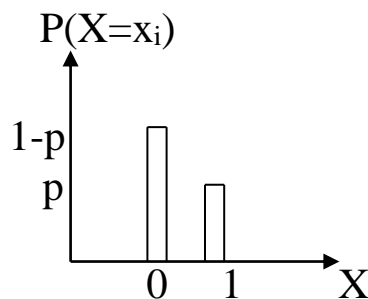
**និយមន័យ :** អថេរចៃដន្យដាច់ (Discrete random variables) ទាំងឡាយណាដែលទទួល តំលៃ 1 និង 0 ហើយដែលមានប្រូបាប៊ីលីតេរៀង  $p$  និង  $1-p$  ហៅថាអថេរចៃដន្យ **Bernoulli** ។

ក-  $X$  ជាអថេរចៃដន្យ Bernoulli  $X$  អាចទទួលតំលៃដូចខាងក្រោម៖

$$X = \begin{cases} 1, (success) \\ 0, (failure) \end{cases}$$

ខ-តារាងរបាយប្រូបាប៊ីលីតេនៃអថេរចៃដន្យ Bernoulli :

| $X$        | 1   | 0     |
|------------|-----|-------|
| $P(X=x_i)$ | $p$ | $1-p$ |



គ- សង្ខេបគណិតនិងរ៉ាំរ៉ៃនៃអថេរ **Bernoulli**:

$$E(X) = \sum_{i=1}^n x_i p_i \quad \text{ដែល } p_i = P(X = x_i)$$

$$E(X) = 1 * p + 0 * (1 - p) = p$$

$$V(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i = (1 - p)^2 * p + (0 - p)^2 * (1 - p)$$

$$= (1 - p)[(1 - p)p + p^2] = p(1 - p)$$

ដូច្នេះ:

$$\begin{aligned} E(X) &= p \\ V(X) &= p(1 - p) \end{aligned}$$

## **២-វិញ្ញាសា Bernoulli និងច្បាប់ទ្វេធា :**

សន្មតថាគេធ្វើការអង្កេត  $n$  ដងនៃវិញ្ញាសា **Bernoulli** ។ កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយមានការកើតឡើងនៃ  $X$  ព្រឹត្តិការណ៍ក្នុងគំរូស្ថិតិ (sample)  $n$  បើសិនណាប្រូបាប៊ីលីតេនៃការកើតឡើងនៃព្រឹត្តិការណ៍ក្នុងការសាកល្បងម្តងៗ ស្មើ  $p$  ។ ការគណនាប្រូបាប៊ីលីតេបែបនេះអោយឈ្មោះថា "ច្បាប់ទ្វេធា (Binomial) ។

## **៣-រូបមន្តសំរាប់ការគណនាប្រូបាប៊ីលីតេនៃច្បាប់ទ្វេធា:**

សន្មតថាយើងមាន  $n$  វិញ្ញាសា **Bernoulli** (មិនអាស្រ័យនឹងគ្នា) ដែលនៅលទ្ធផលនៃវិញ្ញាសានីមួយៗ ការកើតឡើងនៃព្រឹត្តិការណ៍មានប្រូបាប៊ីលីតេ  $p$  និងផ្ទុយពីនេះមានប្រូបាប៊ីលីតេ  $1-p$  ដូច្នេះប្រូបាប៊ីលីតេនៃ  $X$  ព្រឹត្តិការណ៍ដែលកើតឡើងក្នុង  $n$  វិញ្ញាសា **Bernoulli** កំណត់ដោយ :

$$P(X = x) = C_n^x p^x (1-p)^{n-x}$$

$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n; 0 \leq p \leq 1$$

ច្បាប់ **Binomial** អាស្រ័យទៅតាមតំលៃ  $n$  និង  $p$  ។

### **ឧទាហរណ៍ទី១:**

និស្សិតម្នាក់ត្រូវឆ្លងកាត់ការប្រឡងមួយដែលមាន 120 សំណួរ ដែលនៅក្នុងសំណួរនីមួយៗមាន 4 ចំណើយ ក្នុងនោះចំណើយដែលត្រឹមត្រូវមានតែមួយប៉ុណ្ណោះ ។ ដោយសារនិស្សិតរូបនេះពុំមានពេលវេលាគ្រប់គ្រាន់សំរាប់ការសិក្សា គាត់សំរេចចិត្តទៅប្រឡងដោយជ្រើសរើសយកចំណើយនៅក្នុងសំណួរនីមួយៗដោយចៃដន្យ ។

និស្សិតរូបនេះអាចប្រឡងជាប់ ដរាបណាគាត់អាចឆ្លើយត្រូវយ៉ាងតិច 50% នៃសំណួរទាំងអស់ ។

ក- កំណត់រកប្រូបាប៊ីលីតេដើម្បីអោយនិស្សិតរូបនេះប្រឡងជាប់ ។

ខ- តើនិស្សិតរូបនេះអាចឆ្លើយត្រូវជាមធ្យមប៉ុន្មានសំណួរ?

គ- តើគំលាតស្តង់ដារនៃសំណួរដែលអាចឆ្លើយត្រូវស្មើប៉ុន្មាន?

### **ចំណើយ:**

ក- សន្មតថា  $X$  ជាអថេរនៃចំនួនសំណួរដែលអាចឆ្លើយត្រូវ

$$P(X \geq 60) = \sum_{x=60}^{120} \binom{120}{x} (0.25)^x (1-0.25)^{120-x}$$

**S-plus:**

**dbinom:** density of Binomial Distribution

**dbinom(x,n,p)**

x ជាចំនួនព្រឹត្តិការណ៍ដែលកើតឡើងក្នុងគំរូស្ថិតិ n

n ជាគំរូស្ថិតិ (sample size)

p ជាប្រូបាប៊ីលីតេនៃការកើតឡើងនៃព្រឹត្តិការណ៍ក្នុងវិញ្ញាសា Bernoulli ម្តងៗ

ដោយអនុវត្តទៅលើឧទាហរណ៍ខាងលើយើងអាចគណនាបាន៖

```
function()
{ x<- 60:120
  n<-120
  p<-1/4
  prob<-sum(dbinom(x,n,p))
  return(prob)
}
```

**ចំណេញ:** [1] 3.423767e-009 = 0

ដូច្នេះយើងអាចសន្និដ្ឋានបានថានិស្សិតរូបនេះពុំអាចមានលទ្ធភាពប្រលងជាប់នោះទេ ។

ទោះបីជា និស្សិតនេះពុំអាចប្រលងជាប់យ៉ាងណាក្តី ក៏គាត់អាចឆ្លើយត្រូវខ្លះដែរ ចំនួនសំនួរជាមធ្យម

ដែលអាចឆ្លើយត្រូវ នឹងត្រូវគណនានៅក្នុង Expectation ខាងក្រោម ។

**៤-សង្ខេបគណិត និង វារ្យង់នៃអថេរទេធា :**

បើ  $X \sim B(x,n,p)$  នោះ :

$$E(X) = np$$

$$V(X) = np(1 - p), \sigma(X) = \sqrt{np(1 - p)}$$

ដោយអនុវត្តទៅលើឧទាហរណ៍ខាងលើ ចំនួនសំនួរជាមធ្យម ដែលអាចឆ្លើយត្រូវគឺ :

$$E(X) = 120 * 1/4 = 30 \text{ សំនួរ}$$

ចំណែកឯគំរូស្ថិតិសង្ខេបនៃសំនួរដែលការឆ្លើយត្រូវ គឺ :

$$V(X) = \sqrt{n * p * (1 - p)} = 120 * 1/4 * 3/4 = 4.743416 \cong 5 \text{ សំនួរ}$$

**ឧទាហរណ៍ ទី ២ :**

ក្រោយពីការវិភាគទៅលើចារឹកលក្ខណៈនៃអតិថិជនតាមរយៈនៃការធ្វើ questionnaire បានអោយដឹងថា ក្នុងចំណោមអ្នកទិញ 5 នាក់មាន 3 នាក់ ( $p=0.60$ ) បានរងឥទ្ធិពលនៃម៉ាកពាណិជ្ជកម្ម ។ នាយក Marketing នៃ ហាងទំនិញដ៏ធំមួយ បានធ្វើប្រជាមតិទៅលើអ្នកទិញសម្ភារៈចំនួន 20 នាក់ដោយ ចៃដន្យ ។

កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយមានអ្នកទិញតិចជាង 10 នាក់រងឥទ្ធិពលនៃម៉ាកពាណិជ្ជកម្ម ។

### ចំណើយ:

សន្មតថា  $X$  ជាចំនួនអ្នកទិញដែលរងឥទ្ធិពលនៃម៉ាកពាណិជ្ជកម្មក្នុងការទិញសម្ភារៈប្រើប្រាស់ ។ ដូច្នេះ  $X$  អាចទទួលតំលៃ ពី 0 ទៅ 20 ។ ដោយអនុវត្តច្បាប់ទ្វេធា យើងបាន:

$$P(X = x) = \binom{20}{x} (0.60)^x (0.4)^{20-x}$$

$$\binom{20}{x} = C_{20}^x$$

ដូច្នេះប្រូបាប៊ីលីតេដើម្បីអោយអ្នកទិញតិចជាង 10 នាក់រងឥទ្ធិពលនៃម៉ាកពាណិជ្ជកម្មកំណត់ដោយ:

$$P(X \leq 9) = \sum_{x=0}^9 \binom{20}{x} (0.60)^x (1 - 0.40)^{20-x}$$

$$= C_{20}^0 (0.60)^0 (0.40)^{20} + C_{20}^1 (0.60)^1 (0.40)^{19} + \dots + C_{20}^9 (0.60)^9 (0.40)^{11}$$

$$= 0.1273$$

```
function()
{
  x <- 0:9
  proba <- sum(dbinom(x, 20, 0.6))
  return(proba)
}
solution : proba=0.1273
```

### ៥-លំហាត់អនុវត្តន៍ :

- ១- ក្នុង College មួយកន្លែង កំរិតនៃការប្រលងធ្លាក់មុខវិជ្ជា “ Introduction to computer” ស្មើ 40% ។ ក្នុងចំណោមនិស្សិត 15 នាក់ដែលចុះឈ្មោះរៀនមុខវិជ្ជានេះ កំណត់:
- ក- ប្រូបាប៊ីលីតេដើម្បីអោយមាននិស្សិតម្នាក់គត់ប្រលងធ្លាក់មុខវិជ្ជានេះ ។
- ខ- ប្រូបាប៊ីលីតេដើម្បីអោយក្នុងចំណោមនិស្សិតទាំងនេះមានលើសពាក់កណ្តាលប្រលងធ្លាក់មុខវិជ្ជា នេះ ។
- គ- កំណត់ចំនួននិស្សិតក្នុងចំណោមនិស្សិតទាំង 15 នាក់នេះដើម្បីអោយប្រូបាប៊ីលីតេដើម្បីធ្លាក់មាន តំលៃ អតិបរមា ។

ចំណើយ: ក-  $p=0.0047$ , ខ-  $p=0.2131$ , គ-  $n=6$

- ២- អ្នកទទួលខុសត្រូវនៃក្រុមហ៊ុនយន្តហោះ “ swiss air” បានបញ្ជាក់ថា 10% នៃភ្ញៀវរបស់គាត់  
កក់ សំបុត first class ។ ក្នុងចំណោមភ្ញៀវ 5 នាក់ ដែលនឹងមកកក់ សំបុត្រយន្តហោះ ចូរកំណត់៖
- ក-ប្រូបាប៊ីលីតេដើម្បីអោយគ្មាននរណាម្នាក់កក់ first class ។
- ខ-ប្រូបាប៊ីលីតេដើម្បីអោយមានពីរនាក់គត់កក់ first class ។
- គ-ក្នុងចំណោមភ្ញៀវ 5 នាក់ដែលនឹងមកកក់ សំបុត្រ កំណត់ចំនួនអតិថិជនដែលកក់ first class ហើយ  
ដែលប្រូបាប៊ីលីតេមានតំលៃអតិបរមា។
- ចំណើយ: ក- 0.5905, ខ- 0.0729, គ-  $n=0$





## (Hypergeometric distribution)

### ១- របាយទិន្នន័យ អ៊ីពែធរណីមាត្រ

សន្មតថានៅក្នុង Population មួយដែលមាន  $N$  ធាតុមាន  $a$  ធាតុដែលត្រូវកំណត់ថាជាជោគជ័យ (success) និង  $b$  ធាតុបរាជ័យ (failure) ហើយដែល  $N=a+b$  ។ គេធ្វើការជ្រើសរើសនូវគំរូស្ថិតិ (sample) ដែលមាន  $n$  ធាតុ ( $n$  តូចជាងឬស្មើ  $N$ ) ពីក្នុង population នេះ ។ កំណត់តំលៃប្រូបាប៊ីលីតេដើម្បីអោយមាន  $x$  ជោគជ័យនៅក្នុងគំរូស្ថិតិនេះ ។ តំលៃប្រូបាប៊ីលីតេនេះត្រូវបានកំណត់ដោយច្បាប់អ៊ីពែធរណីមាត្រ ដែលកំណត់ដូចខាងក្រោម៖

$$P(X = x) = \frac{C_a^x * C_b^{n-x}}{C_N^n}$$

$$\text{ដែល } x = \begin{cases} 0, 1, 2, \dots, n; n \leq a \\ 0, 1, 2, \dots, a; n > a \end{cases}$$

$$A=N*p \text{ និង } b=N(1-p)$$

ដែល  $p$  ជាផលធៀបនៃជោគជ័យក្នុង population នេះ ។

### ២- សង្ខេបគណិត និង វារ្យង់

បើ  $X$  ជាអថេរចៃដន្យអ៊ីពែធរណីមាត្រ (hypergeometric random variable) នោះ៖

$$E(X) = np$$

$$V(X) = np(1-p) * \frac{N-n}{N-1}$$

### ឧទាហរណ៍១៖

អ្នកទទួលខុសត្រូវនៃមន្ទីរពិសោធន៍មួយកន្លែងមានបុគ្គលិក 18 នាក់ដែលក្នុងនោះមានមនុស្សស្រី 10 នាក់ និងមនុស្សប្រុស 8 នាក់ ។ គាត់ត្រូវជ្រើសរើសបុគ្គលិករបស់គាត់ចំនួនពីរនាក់ដើម្បីចូលរួមក្នុងគណៈកម្មការការពារសន្តិសុខក្នុងមន្ទីរពិសោធន៍នេះ ។ ការជ្រើសរើសត្រូវធ្វើឡើងដោយចៃដន្យ ។  
ក-តើគេអាចជ្រើសរើសមនុស្សពីរនាក់នេះបានប៉ុន្មានរបៀប ព្រមទាំងកំណត់ប្រូបាប៊ីលីតេនៃការជ្រើសរើសនេះ?

ខ-សន្មតថា  $X$  ជាចំនួនមនុស្សស្រីដែលគេជ្រើសរើសក្នុងកុំរូស្តីតិនេះ តើ  $X$  អាចទទួលតំលៃណាខ្លះ ហើយកំណត់របាយប្រូបាប៊ីលីតេនៃ  $X$ ?

គ-កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយមនុស្សពីរនាក់ដែលត្រូវបានជ្រើសរើសជាមនុស្សស្រី ។

**ចំណាំ:** ១-  $n=153$  របៀប ; ២-  $x : 0,1,2$  និង  $P(X = x) = \frac{C_{10}^x * C_8^{2-x}}{C_{18}^2}$

៣-  $p=0.2941$

### **៣-S-plus function:**

dhypcr : density of hypergeometric distribution

**dhypcr(x,a,b,n)**

អនុវត្តលើឧទាហរណ៍ខាងលើ:

```
function()
{
  #Question 1
  n <- choose(18, 2)
  proba1 <- 1/n
  # Question 2
  x <- 0:2
  proba2 <- dhypcr(x, 10, 8, 2)
  Table <- rbind(x, proba2)
  #Question 3
  proba3 <- dhypcr(2, 10, 8, 2)
  return(n, proba1, Table, proba3)
}
```

**solution:**

**\$n:**

**[1] 153**

**\$proba1:**

**[1] 0.006535948**

**\$Table:**

|               |                  |                  |                  |
|---------------|------------------|------------------|------------------|
| <b>x</b>      | <b>0.0000000</b> | <b>1.0000000</b> | <b>2.0000000</b> |
| <b>proba2</b> | <b>0.1830065</b> | <b>0.5228758</b> | <b>0.2941176</b> |

**\$proba3:**

**[1] 0.2941176**

### **ឧទាហរណ៍២:**



អ្នកលក់ទំនិញនៃហាងមួយកន្លែងបានបញ្ជាក់ថា 10% នៃ printer ដែលគាត់ទទួលយកមកលក់ត្រូវបានជួសជុលមុននឹងលក់អោយអតិថិជនរបស់គាត់ ។ បើសិនណាគេជ្រើសរើសដោយចៃដន្យនូវ printer ចំនួន 5 ពីក្នុងចំណោម printer ចំនួន 20 គ្រឿង កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយបាន printer ដែលខូចទាល់តែសោះ ?



### (POISSON DISTRIBUTION)

ច្បាប់ព័រសុងត្រូវបានគេអនុវត្តយ៉ាងញឹកញាប់នៅក្នុងបណ្តាក្រុមហ៊ុនសហគ្រាសនានាដូចជា៖ នៅក្នុងការគ្រប់គ្រងក្នុងរោងចក្រឧស្សាហកម្ម (ចំនួនអ្នករងគ្រោះថ្នាក់នៅក្នុងពេលបំពេញការងារ, ការផ្ទៀងផ្ទាត់នូវបញ្ជីគណនេយ្យ ។ល ។), operation research (ការសិក្សាអំពីការបញ្ជូនព័ត៌មានដែលមានលក្ខណៈបន្តបន្ទាប់គ្នា), ការធ្វើចរាចរនៅលើផ្លូវសាធារណៈ (ចំនួនយានយន្តដែលត្រូវកកស្ទះនៅចំណុចណាមួយ), ប្រជាសាស្ត្រ (ការផ្តល់កំណើតដែលធ្វើអោយមានការកើនឡើងនៃប្រជាជន, ការស្លាប់នៅក្នុង population មួយដែលគេសិក្សា), ការស្រាវជ្រាវអំពីវេជ្ជសាស្ត្រ (ការសិក្សាទៅលើកំណកំណើតនៃបាក់តេរី), ... ។

នៅក្រោមលក្ខខណ្ឌខ្លះច្បាប់ព័រសុងមានកំហិតទៅរកច្បាប់ទ្វេធា ។

#### ១-ច្បាប់ព័រសុង៖

អថេរចៃដន្យ  $X$  ដែលអាចទទួលតំលៃ  $0, 1, 2, 3, \dots, n, \dots$  ហើយដែលមានប្រូបាប៊ីលីតេ

$$P(X = x) = \frac{e^{-\lambda} * \lambda^x}{x!}, \lambda > 0 \text{ គោរពទៅតាមច្បាប់ព័រសុង } P(X = x | \lambda) \text{ ដែលមានប៉ារ៉ាម៉ែត្រ}$$

$\lambda$  ។

#### ២-សង្ឃឹមគណិត វ៉ារ្យង់ និង គំណាតស្តង់ដា (Expectation, variance and Standard deviation)

បើ  $X$  ជាអថេរចៃដន្យព័រសុង នោះ:

$$E(X) = \lambda$$

$$V(X) = \lambda$$

$$\sigma(X) = \sqrt{\lambda}$$

### ចំណាំ:

$X$  គោរពតាមច្បាប់ព័រសុងអាចសរសេរ:

$$X \sim P(X = x \mid \lambda)$$

$$X \sim P(x; \lambda)$$

$$X \sim P(\lambda)$$

$$X \sim P(x, n, p)$$

### លំហាត់:

អ្នកទទួលខុសត្រូវនៃគណៈកម្មការសន្តិសុខសហគ្រាស NICOM បានសិក្សាទៅលើចំនួនគ្រោះថ្នាក់ការងារដែលបានកើតឡើងក្នុងរយៈពេល 2 ឆ្នាំកន្លងទៅ ។ តាមរបាយការណ៍នៃការសិក្សានេះបានអោយ ដឹងថាចំនួនមធ្យមនៃគ្រោះថ្នាក់ការងារមាន 1.6 ក្នុងមួយថ្ងៃ ។

ក- ដោយសន្មតថាចំនួនគ្រោះថ្នាក់នៅក្នុងមួយថ្ងៃគោរពតាមច្បាប់ព័រសុង កំណត់រូបមន្តមួយដែលអាចអោយយើងគណនានូវប្រូបាប៊ីលីតេនៃ  $X$  គ្រោះថ្នាក់ការងារក្នុងមួយថ្ងៃ ។

$$E(X) = \lambda = 1.6 \Rightarrow P(X = x \mid \lambda = 1.6) = \frac{e^{-1.6} * (1.6)^x}{x!}, x = 0, 1, 2, \dots$$

ខ- គំណាតស្តង់ដារនៃ  $X$ :

$$\sigma(X) = \sqrt{1.6} = 1.2625 \text{ គ្រោះថ្នាក់ការងារ/1 ថ្ងៃ}$$

គ- ប្រូបាប៊ីលីតេនៃការមានគ្រោះថ្នាក់ការងារលើសពី 2 ដងក្នុងមួយថ្ងៃ

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[ \frac{e^{-1.6} (1.6)^0}{0!} + \frac{e^{-1.6} (1.6)^1}{1!} + \frac{e^{-1.6} (1.6)^2}{2!} \right] \\ &= 0.2167 = 21.67\% \end{aligned}$$

ឃ- គណនាប្រូបាប៊ីលីតេនៃចំនួនគ្រោះថ្នាក់ការងារនៅក្នុងចន្លោះ  $[E(X) - \sigma(X), E(X) + \sigma(X)]$

$$E(X) - \sigma(X) = 1.6 - 1.265 = 0.335$$

$$E(X) + \sigma(X) = 1.6 + 1.265 = 2.865$$

$$P(E(X) - \sigma(X) \leq X \leq E(X) + \sigma(X)) = P(0.335 \leq X \leq 2.865)$$

ដោយអថេរចៃដន្យពីរស្តង់ដារអថេរចៃដន្យដាច់ដែលអាចទទួលបានតំលៃ  $0, 1, 2, 3, \dots, n, \dots$  ដូច្នេះ:

$$\begin{aligned} P(0.335 \leq X \leq 2.865) &= P(0 \leq X \leq 2) \\ &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.7833 \end{aligned}$$

### S-plus:

**dpois : density of poisson distribution**

**dpois(x, λ)**

ដោយអនុវត្ត S-plus លើឧទាហរណ៍ខាងលើ យើងអាចសរសេរ program ដូចខាងក្រោម :

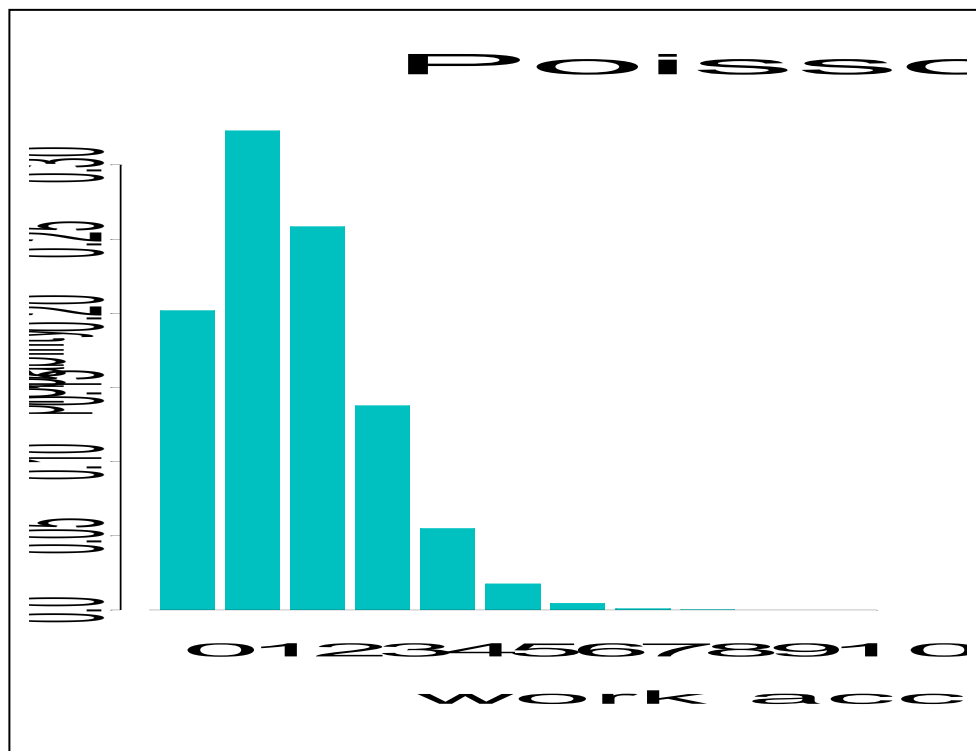
```
function()
{
#Question 1
  lambda <- 1.6
  print(" Poisson(x,lambda)=(exp(-lambda)*lambda^x)/factorial(x) ; lambda=1.6 ;
        x=0,1,2,.....")
#Question 2
  sigma <- sqrt(lambda)
#Question 3
  x <- 0:2
  proba3 <- 1 - sum((exp( - lambda) * lambda^x)/factorial(x))
#Question 4
  born.inf <- as.integer(lambda - sigma)
  born.sup <- as.integer(lambda + sigma)
  x <- born.inf:born.sup
  proba4 <- sum((exp( - lambda) * lambda^x)/factorial(x))
#Question 5
  x<-0:10
  proba5<-(exp(-lambda)*lambda^x)/factorial(x)
  barplot(proba5,names=factor(levels(x)),xlab="work accidence per day",ylab=
    "probability",main="Poisson distribution")
  return(sigma, proba3, proba4)
}
```

### output

```
[1] " Poisson(x,lambda)=(exp(-lambda)*lambda^x)/factorial(x) ; lambda=1.6 ;
x=0,1,2,....."
$sigma:
[1] 1.264911
```

```
$proba3:  
[1] 0.2166415
```

```
$proba4:  
[1] 0.7833585
```



**ឧទាហរណ៍ :** ការអនុវត្តច្បាប់ **Poisson** ទៅលើការលក់ប្រចាំថ្ងៃនៃក្រុមហ៊ុន **Computer** ។ ការលក់ប្រចាំថ្ងៃ  $X_1$  នៃក្រុមហ៊ុន computer មួយកន្លែងគោរពតាមច្បាប់ Poisson ដែលមានតំលៃ មធ្យម  $\lambda_1 = 4.2$  គ្រឿង ។

ក- កំណត់ វ៉ារ្យង់នៃអថេរ  $X_1$  ?

ដោយការលក់ប្រចាំថ្ងៃ  $X_1$  គោរពតាមច្បាប់ Poisson ដូច្នេះ :

$$E(X_1) = \text{Var}(X_1) = 4.2 \text{ (គ្រឿង)}^2$$

ខ- កំណត់សមមាត្រធ្យូប (proportion) នៃថ្ងៃដែលលក់បាន :

-មួយគ្រឿង :

$$P(X = 1, \lambda = 4.2) = \frac{4.2^1 * e^{-4.2}}{1!} = 0.0630$$

-ពី 4 ទៅ 5 គ្រឿង :

$$P(4 \leq X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6) = 0.1944 + 0.1633 + 0.1143 = 0.472$$

គ- នៅក្នុងរយៈពេល 250 ថ្ងៃដែលបើកលក់តើមានប្រហែលប៉ុន្មានថ្ងៃដែលការលក់ប្រចាំថ្ងៃដាច់បាន 3 គ្រឿង ?

$$\lambda \approx n * p = 250 * P(X = 3) = 250 * 0.1852 \approx 46 \text{ ថ្ងៃ}$$

### ចំណាំ:

-បើ  $X_1$  និង  $X_2$  ជាអថេរចៃដន្យពីរស្តង់ដែលមិនអាស្រ័យនឹងគ្នា ហើយដែលមានប្រូបាប៊ីលីតេរៀង

$P(X_1, \lambda_1)$  និង  $P(X_2, \lambda_2)$  នោះ :

$$Y = X_1 + X_2 \sim P(Y, \lambda_1 + \lambda_2)$$

- ជាទូទៅបើ  $X_1, X_2, \dots, X_n$  ជាអថេរចៃដន្យមិនអាស្រ័យនឹងគ្នាហើយដែលមានប្រូបាប៊ីលីតេរៀង

$P(X_1, \lambda_1), P(X_2, \lambda_2), \dots, P(X_n, \lambda_n)$  នោះ :

$$Y = X_1 + X_2 + \dots + X_n \sim P(Y, \lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n)$$

### ៣-កំហិតនៃច្បាប់ ទ្វេធាទៅរកច្បាប់ពីរស្តង់ :

សន្មតថាលក្ខខណ្ឌនៃការអនុវត្តទៅលើច្បាប់ទ្វេធាត្រូវបានបំពេញ ប៉ុន្តែនៅ ក្នុងករណីនេះទំហំ

sample នៃការ សាកល្បង  $n$  មានទំហំធំហើយប្រូបាប៊ីលីតេ  $p$  មានតំលៃតូចដែលធ្វើអោយ  $n * p$

មានតំលៃមួយ តូចធៀបទៅនឹងតំលៃ  $n$  ដូច្នេះយើងអាចធ្វើកំហិតពីច្បាប់ ទ្វេធាទៅច្បាប់ពីរស្តង់ដូចខាង

ក្រោម :

$$b(x; n; p) \approx p(x; \lambda), \lambda = np$$

នៅក្នុងការអនុវត្តកំហិតពីច្បាប់ទ្វេធា ទៅរកច្បាប់ពីរស្តង់អាចអនុវត្តបានលុះត្រា :

$$n > 20, p \leq 0.10 \text{ និង } np \leq 5$$

$$\lim_{\substack{n \rightarrow +\infty \\ p \rightarrow 0}} \binom{n}{x} p^x (1-p)^{n-x} = \frac{(np)^x e^{-np}}{x!} = \frac{\lambda^x e^{-\lambda}}{x!}$$

យើងនឹងលើកយកឧទាហរណ៍មួយដើម្បីបញ្ជាក់អំពីករណីខាងលើ ដោយប្រើ S-plus programming

language ដូចខាងក្រោម :

function(n, p)

{

    kam2 <- function(n, p)

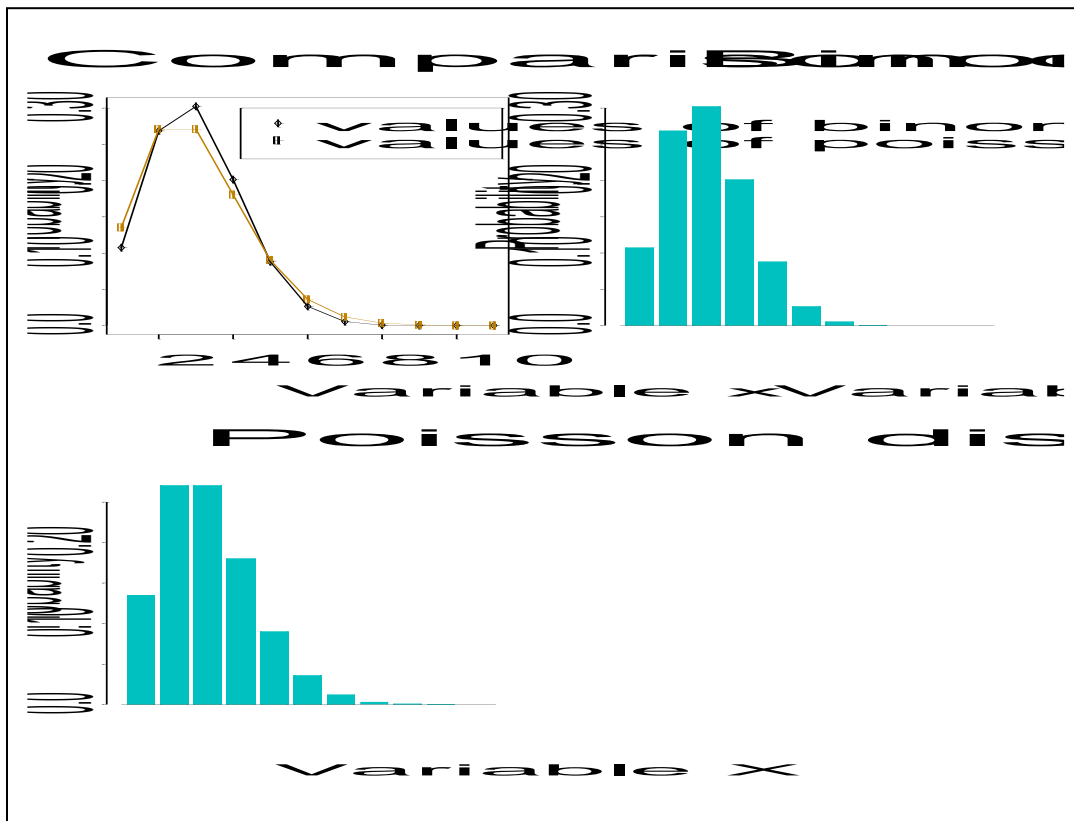
រៀបរៀងដោយ ម៉ាក កាម៉េរ៉ាន (Statistician, Mathematician).

```
{
  x <- 0:10
  binomial <- dbinom(x, n, p)
  lambda <- n * p
  poisson <- dpois(x, lambda)
  table <- cbind(x, binomial, poisson)
  par(mfrow = c(2, 2))
  y1 <- plot(binomial, pch = 9, xlab =
    "Variable x", ylab =
    "Probability", main =
    "Comparison of Binomial with Poisson distribution"
  )
  y2 <- points(poisson, col = 5, pch = 12
  )
  lines(binomial)
  lines(poisson, col = 5)
  legend(4.2, 0.3, c("values of binomial",
    "values of poisson"), mark = c(
    9, 12))
  barplot(binomial, xlab = "Variable X",
    ylab = "Probability", main =
    "Binomial distribution")
  barplot(poisson, xlab = "Variable X",
    ylab = "Probability", main =
    "Poisson distribution")
  return(table, y1, y2)
}
result.1 <- kam2(10, 0.2)
result.2 <- kam2(20, 0.1)
result.3 <- kam2(30, 0.05)
return(result.1, result.2, result.3)
}
```

**ក- ករណីដែល  $n=10, p=0.20, b(x,10,0.20)$  :**

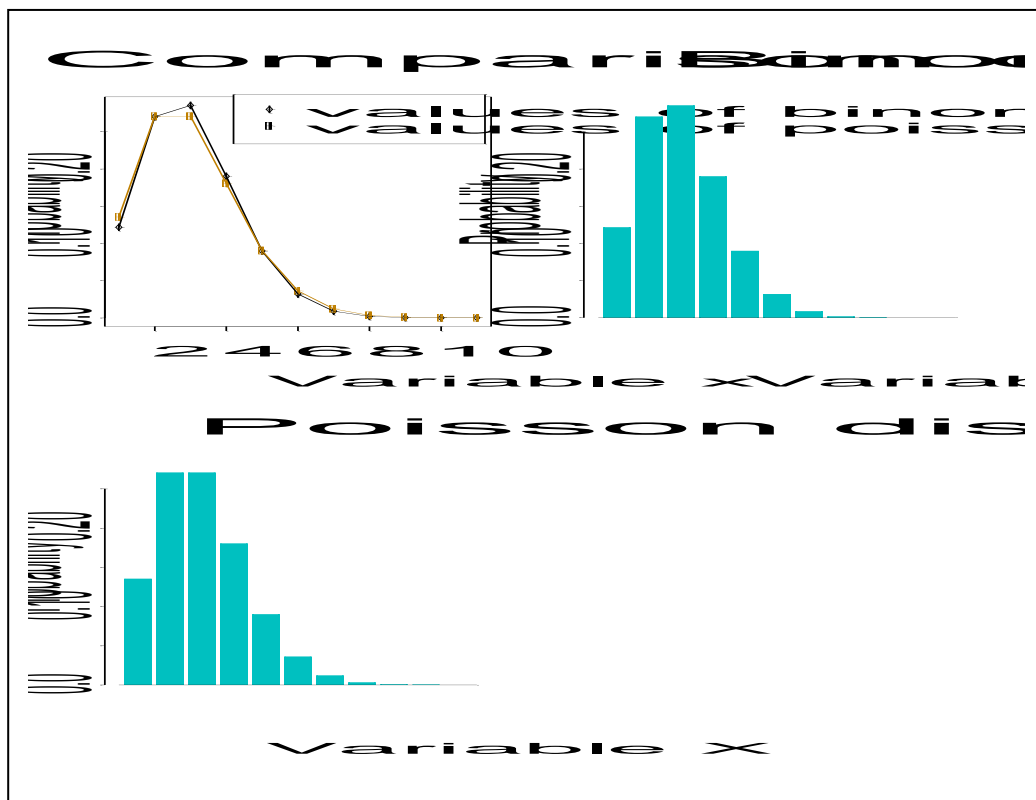
រៀបរៀងដោយ ម៉ាក

| x | binomial     | poisson       |
|---|--------------|---------------|
| 0 | 0.1073741824 | 0.13533528324 |
| 1 | 0.2684354560 | 0.27067056647 |
| 2 | 0.3019898880 | 0.27067056647 |
| 3 | 0.2013265920 | 0.18044704432 |
| 4 | 0.0880803840 | 0.09022352216 |
| 5 | 0.0264241152 | 0.03608940886 |
| 6 | 0.0055050240 | 0.01202980295 |
| 7 | 0.0007864320 | 0.00343708656 |
| 8 | 0.0000737280 | 0.00085927164 |
| 9 | 0.0000040960 | 0.00019094925 |



ខ- ករណីដែល  $n=20, p=0.10, b(x,20,0.10)$

| x | binomial      | poisson       |
|---|---------------|---------------|
| 0 | 1.215767e-001 | 0.13533528324 |
| 1 | 2.701703e-001 | 0.27067056647 |
| 2 | 2.851798e-001 | 0.27067056647 |
| 3 | 1.901199e-001 | 0.18044704432 |
| 4 | 8.977883e-002 | 0.09022352216 |
| 5 | 3.192136e-002 | 0.03608940886 |
| 6 | 8.867045e-003 | 0.01202980295 |
| 7 | 1.970454e-003 | 0.00343708656 |
| 8 | 3.557765e-004 | 0.00085927164 |

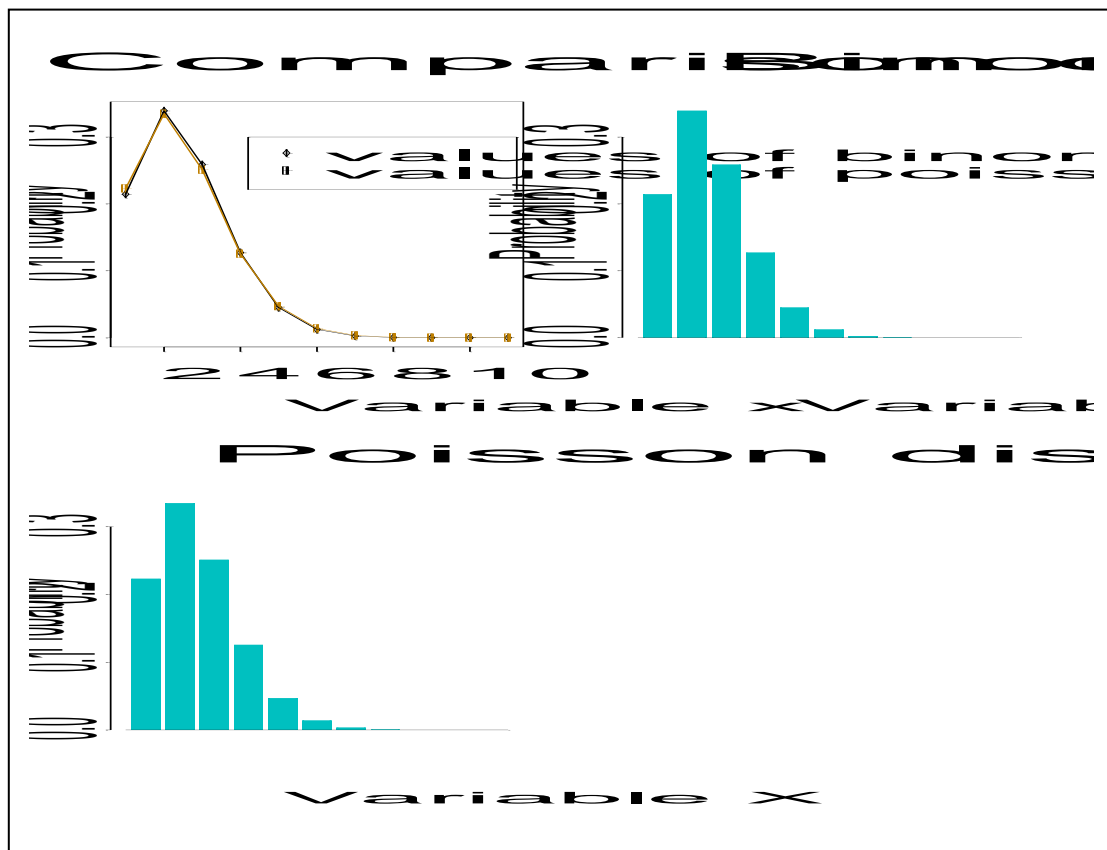


ត-ករណី ដែល  $n=30, p=0.05, b(x, 30, 0.05)$  :

រៀបរៀងដោយ

| x | binomial      | poisson       |
|---|---------------|---------------|
| 0 | 2.146388e-001 | 2.231302e-001 |
| 1 | 3.389033e-001 | 3.346952e-001 |
| 2 | 2.586367e-001 | 2.510214e-001 |
| 3 | 1.270496e-001 | 1.255107e-001 |
| 4 | 4.513605e-002 | 4.706652e-002 |
| 5 | 1.235302e-002 | 1.411996e-002 |
| 6 | 2.708997e-003 | 3.529989e-003 |
| 7 | 4.888415e-004 | 7.564262e-004 |





#### ៥-លំហាត់អនុវត្តន៍:

- ១- ដោយប្រើកម្មវិធី S-plus ចូរបង្ហាញនិងពន្យល់ពីតួនាទីនីមួយៗ នៃ command ខាងក្រោម:  
 $\text{dbinom}(x, 30, 0.10)$  ,  $x \leftarrow 0:12$
- ខ-  $\text{Plot}(\text{data})$
- គ-  $\text{par}(\text{mfrow} = c(3, 2))$

ឃ- តើ points ខុស ពី plot យ៉ាងដូចម្តេច?

ង- return()

ច- par(mfg=c(1,2,1,2))

២- ដោយប្រើកម្មវិធី S-plus ចូរសរសេរ program ដើម្បីដោះស្រាយបញ្ហាខាងក្រោម:

និស្សិតម្នាក់ត្រូវឆ្លងកាត់ការប្រលងមួយដែលមាន 15 សំណួរ ។ វិធីសាស្ត្រនៃការឆ្លើយសំណួរ គឺ ពិតឬមិនពិត ។ ដោយសារនិស្សិតនេះពុំបានរៀបចំមេរៀនបានល្អ គាត់សំរេចចិត្តបោះកាក់ក្នុងការឆ្លើយ បើកាក់ចេញផ្ទៃកខាងរូប គាត់ឆ្លើយ "ពិត" បើកាក់ចេញផ្ទៃកខាងលេខ គាត់ឆ្លើយ "មិនពិត" ។ និស្សិតនេះប្រលងជាប់ លុះត្រាតែគាត់ ទទួលបានពិន្ទុ 60% ។ កំណត់ប្រូបាប៊ីលីតេដើម្បីអោយនិស្សិតរូបនេះប្រលងជាប់ ។

៣- ក្រុមហ៊ុនសាងសង់មួយបានបញ្ជាក់ថាដើម្បីសាងសង់អាគារមួយដែលមាន 10 ល្វែង កំពស់ 4 ជាន់ព្រម

ទាំងមានក្បាច់រចនាល្អប្រណីតត្រូវចំណាយពេលជាមធ្យម 50 សប្តាហ៍ និងគំនាតស្តង់ដា (standard deviation) 5 សប្តាហ៍ ។

ឥឡូវនេះក្រុមហ៊ុននេះមានបំណងចង់ដេញថ្លៃលើគំរោងសាងសង់មួយដែលមានលក្ខណៈដូចខាងលើដោយ បញ្ជាក់ពីពេលវេលាច្បាស់លាស់ក្នុងការសាងសង់នៅក្នុងកិច្ចសន្យាហើយដែលមានប្រូបាប៊ីលីតេនៃការគោរព កិច្ចសន្យាស្មើ 90% ។ តើក្រុមហ៊ុននេះត្រូវបញ្ជាក់ក្នុងកិច្ចសន្យាចំនួនប៉ុន្មាន សប្តាហ៍ដើម្បីសំរេចការសាងសង់នេះ? បើសិនណាគេសន្មតថាពេលវេលាចាំបាច់ក្នុងការសាងសង់គោរពតាមច្បាប់ណរម៉ាល់ (normal) ។

៤- គេមានតារាងទិន្នន័យដូចខាងក្រោម :

| Output(\$) | Probability |
|------------|-------------|
| 27000      | 0.11        |
| 23000      | 0.14        |
| 20000      | 0.24        |
| 22000      | 0.21        |
| 16000      | 0.30        |

ដោយប្រើកម្មវិធី S-plus ចូរបង្កើត table នៃទិន្នន័យខាងលើនេះ និងគណនាមធ្យមនៃ output ដែល នឹងអាចទទួលបានព្រមទាំងគំលាតគំរូ (standard deviation) និង risk ។

៥- ដោយប្រើកម្មវិធី S-plus ចូរសង់ program ដើម្បីដោះស្រាយបញ្ហាខាងក្រោម៖

តាមការអង្កេតមួយដែលបានធ្វើឡើងដោយទស្សនាវដ្តី “Gestion” បានអោយដឹងថា 28% នៃបុគ្គលិក របស់ក្រុមហ៊ុន “ AT&T ” មានសញ្ញាបត្រ “Master in management” ។ សន្មតថាបុគ្គលិក 15នាក់ត្រូវ បានជ្រើសរើសដោយចៃដន្យ ។

ក- តើជាមធ្យមមានបុគ្គលិកប៉ុន្មាននាក់ដែលមានសញ្ញាបត្រ “Master in management” ក្នុងចំណោម គំរូស្ថិតិខាងលើ?

ខ- ប្រូបាប៊ីលីតេដើម្បីអោយ បុគ្គលិក 5 នាក់មានសញ្ញាបត្រ “Master in management” ក្នុងចំណោម គំរូស្ថិតិខាងលើ ។

៦-បុរសម្នាក់មានបំណងចង់វិនិយោគថវិកាមួយចំនួនដោយការទិញប័ណ្ណភាគហ៊ុន ។ គាត់សង្កេតឃើញថា ក្រុមហ៊ុន A អាចផ្តល់ទិន្នផល 16% បើគ្មានការធ្លាក់ចុះនៃសេដ្ឋកិច្ច ផ្ទុយទៅវិញគាត់អាចទទួលបាន 10% បើ មានការធ្លាក់ចុះនៃសេដ្ឋកិច្ច ។ គាត់សង្កេតឃើញផងដែរក្រុមហ៊ុន B ផ្តល់ទិន្នផលថេរ 12% ។

តើការធ្លាក់ចុះនៃសេដ្ឋកិច្ចត្រូវតែចាត់ទុកជាប៉ាន់ស្មានភាគរយ ដើម្បីអោយការទិញប័ណ្ណភាគហ៊ុននៃក្រុមហ៊ុន A ទទួលបានទិន្នផលខ្ពស់ជាងក្រុមហ៊ុន B?

៧-Security Alarm System អាចដំណើរការបានឬមិនបានអាស្រ័យនឹងការវាយបញ្ចូល 3-digits numerical code ត្រឹមត្រូវនៅក្នុង digital panel នៃប្រព័ន្ធនេះ ។

ក- តើចំនួនលេខ code ទាំងអស់ដែលអាចមានចំនួនប៉ុន្មាន បើក្នុងចំណោមលេខ code នីមួយៗ គ្មាន numerical digit ណាមួយដែលត្រូវបានប្រើពីរដងនោះទេ?

ខ-តើចំនួនលេខ code ទាំងអស់ដែលអាច មានចំនួនប៉ុន្មាន បើក្នុងចំណោមលេខ code នីមួយៗ numerical digit នីមួយៗអាចប្រើលើសពី 1 ដង?

(លេខ code ខាងលើនេះត្រូវបានសរសេរនៅក្នុងប្រព័ន្ធរបាប់គោល 10)

៨- គេមានតារាងទិន្នន័យ DATA ដែលស្រង់បានពីការអង្កេតទៅលើ Salary/year របស់បុគ្គលិកមួយចំនួន នៅក្នុងក្រុមហ៊ុនមួយកន្លែងទៅតាម ភេទ តួនាទី និងកម្រិតការសិក្សា ហើយដែលមាននៅក្នុង object explorer នៃ S-plus ដែល៖

- ជួរឈរទី១ តាងអោយ Gender ( Male, Female)
- ជួរឈរទី២ តាងអោយ Jobcategories (Manager, Custodial, Clerical)
- ជួរឈរទី៣ តាងអោយ Level of education (<HS, HS, BA, MBA)
- ជួរឈរទី៤ តាងអោយ Salary/year (Numerical Data)

ដោយផ្ដាច់យកតែ Records នៃបុគ្គលិកដែលមានកម្រិតការសិក្សាមិនលើសពី HS និងមានតួនាទីជាClerical ចូរសរសេរ function S-plus បង្កើតតារាង Crosstabs ដែលវាយតម្លៃជាលក្ខណៈខ្វែងគ្នារវាង Variable “Jobcategories” និង “Gender” ដោយយក “Jobcategories” ជាជួរដេកនៃ Crosstabs ។

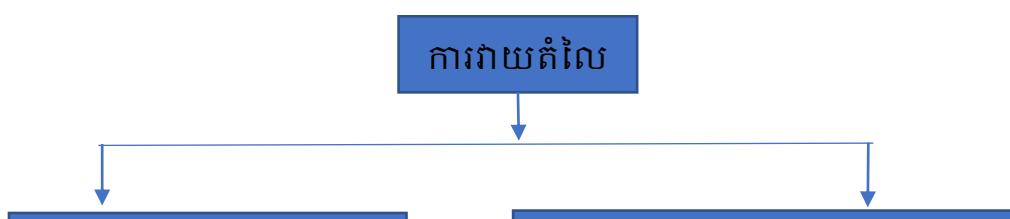
៩-ដោយអនុវត្តទៅលើតារាងទិន្នន័យនៃ **លំហាត់ទី ៨** ហើយសន្មតថា variable Salary/year មានរបាយ ទិន្នន័យ Normal, ចូរដក 10% នៃទិន្នន័យដែលតូចជាងគេចេញពី Salary/year ហើយសរសេរ function S-plus វាយតម្លៃជាមធ្យមទៅលើ Salary/year របស់បុគ្គលិកនីមួយៗដែលមានទំរង់៖

| N | Mean | Variance | Std deviation |
|---|------|----------|---------------|
|   |      |          |               |



**ជំពូក៣៖**

**ការវាយតម្លៃ**  
(Estimation)



### ១. របៀបវារៈនៃទិន្នន័យ(Measurement of Data):

ជាទូទៅនៅក្នុងការស្រាវជ្រាវ ទិន្នន័យចែកចេញជាពីរផ្នែកធំៗ៖

Quantitative data និង Qualitative data

#### ក. Quantitative data:

Quantitative data គឺជាប្រភេទទិន្នន័យដែលសំគាល់បរិមាណ ហើយដែលអាចធ្វើ

ប្រមាណវិធីពិជគណិត(+ , - , \* , / ) បាន ។

#### ឧទាហរណ៍៖

ការចំណាយជាមធ្យមប្រចាំសប្តាហ៍របស់និស្សិតនីមួយៗ

រយៈពេលជាមធ្យមនៃការធ្វើកិច្ចការរបស់និស្សិតនីមួយៗ

#### ខ. Qualitative data:

ជាប្រភេទទិន្នន័យដែលសំគាល់សភាពលក្ខណៈនៃអ្វីមួយហើយដែលមិនអាចធ្វើ

ប្រមាណវិធីពិជគណិតបាន។

### ឧទាហរណ៍៖

ភេទរបស់និស្សិតនីមួយៗ ៖ Male or Female

Size of T-shirt : Small Medium Large Extra-Large

យ៉ាងណាម៉េញ Qualitative data ចែកចេញជាពីរផ្នែក៖ Nominal & Ordinal ។

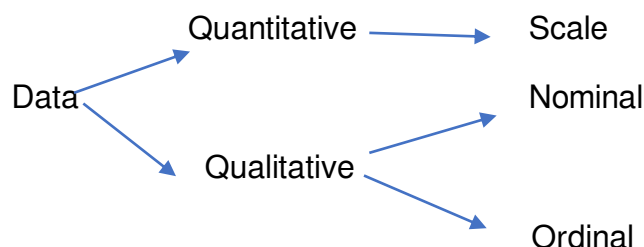
**Nominal data:** ជាប្រភេទ Qualitative data ដែលគ្មានលំដាប់នៃការប្រៀបធៀប។

ឧទាហរណ៍៖ Male or Female; Yes or No

**Ordinal data:** ជាប្រភេទទិន្នន័យ Qualitativeដែលមានលំដាប់នៃការប្រៀបធៀប។

ឧទាហរណ៍៖ កម្រិតការសិក្សា <HS , HS, Bachelor, Master, Ph.D ។

### សង្ខេប៖



សំគាល់៖ នៅក្នុង software SPSS ឬក៏ Software ផ្សេងទៀត រង្វាស់នៃទិន្នន័យចែកចេញជា ៣ ៖ scale, Nominal & Ordinal ។

## ២. ពាក្យបច្ចេកទេសស្ថិតិ(Statistics Terminology):

**Population:**

ជាសាកលនៃព័ត៌មានទាំងអស់ដែលអាចកើតមាននៅក្នុងការសិក្សាស្រាវជ្រាវ។

**Sample**

(គំរូតាង ឬ សំណាក) ៖ ជាផ្នែកមួយនៃ population ដែលត្រូវដកស្រង់ ដោយចៃដន្យ សំរាប់ ការសិក្សា ។

**Sample size:**

ជាទំហំនៃការអង្កេត។

**Parameter of population:**

ជាលក្ខណៈសំគាល់អ្វីមួយនៃ population ដែលត្រូវសិក្សាវាយតម្លៃ។

**Estimator**

ជារង្វាស់ទទួលបាននៅក្នុងសំណាក ដែលប្រើសំរាប់វាស់ parameter ។

**Unbiased estimator**

ជារង្វាស់ទទួលបាននៅក្នុងសំណាក ដែលវាស់ parameter បានល្អប្រសើរ បំផុត។

**Std.error (លំអៀងស្តង់ដារ) ៖**

ជាលំអៀងជាមធ្យមដែលកើតឡើងនៅក្នុងការវាយតម្លៃ។

**Margin of error:**

ជាលំអៀងធំបំផុតដែលកើតឡើងក្នុងការវាយតម្លៃ។

**Level of confidence:**

កំរិតជឿជាក់។

**Level of significance:**

កំរិតភាគរយនៃ population ដែលខុសនឹងការវាយតម្លៃ ហើយដែល

$$\text{Level of confidence} + \text{Level of significance} = 1$$

**Outliers:**

តំលៃមិនប្រក្រតីដែលមាននៅក្នុងទិន្នន័យ។

### ៣. ការវាយតម្លៃដោយចំនុច(Point estimates):

គឺជាសិក្សាវាយតម្លៃទៅលើប៉ារ៉ាម៉ែត្រ នៃ population  $\theta$  តាមរយៈតម្លៃ estimator  $\hat{\theta}$  ដែល ទទួលបានពីការសំយោគពីទិន្នន័យស្រង់បាននៅក្នុង សំណាកតែមួយ។

$\theta$  ជា parameter of population

$\hat{\theta}$  ជា estimator

គេនិយាយថា  $\hat{\theta}$  ជា unbiased estimator (រង្វាស់ដែលវាស់បានប្រសើរបំផុត) នៃ  $\theta$ , កាលណា  $E(\hat{\theta}) - \theta = 0$  ។

ទោះបី  $\hat{\theta}$  វាយតម្លៃ  $\theta$  បានល្អប្រសើរយ៉ាងណាក៏ដោយ ក៏នៅលទ្ធផលនៃការវាយតម្លៃ នៅតែមានលំអៀងដដែល ហើយ លំអៀងនោះ អោយឈ្មោះថា លំអៀងស្តង់ដាតាងដោយ  $S(\hat{\theta})$ ។

$$\hat{\theta} \xrightarrow[\text{S}(\hat{\theta})]{\text{វាយតម្លៃ}} \theta, \quad \text{S}(\hat{\theta}) \text{ ជាលំអៀងស្តង់ដា}$$

### ក. ការវាយតម្លៃជាមធ្យម ( $\mu$ ):

នៅក្នុងករណីនៃការវាយតម្លៃជាមធ្យម, ទិន្នន័យដែលស្រង់នៅក្នុងសំណាកដាច់ខាតត្រូវតែជាប្រភេទ Quantitative ។

ប៉ារ៉ាម៉ែត្រនៃ population តំណាងដោយតម្លៃមធ្យម  $\mu$  ហើយ unbiased estimator តំណាងដោយតម្លៃ  $\bar{x}$  ដែលជាតម្លៃមធ្យមដែលទទួលបានពីសំណាក (sample mean)។

ក្នុងករណីនេះ លំអៀងស្តង់ដានៃការវាយតម្លៃកំណត់ ដោយ  $S(\bar{x}) = \frac{S(X)}{\sqrt{n}}$  ។

**ចំណាំ៖**  $\bar{x}$  ជា unbiased estimator នៃ  $\mu$ , ព្រោះ  $E(\bar{X}) - \mu = 0$  (យោងតាម

Central limit theorem)។

បើ  $X \sim N(\mu, \sigma^2)$ , នោះ  $\bar{X} \sim N(\mu, \sigma^2/n)$  ។ ក្នុងករណីនេះ  $E(\bar{X}) = \mu$   
 $\Rightarrow E(\bar{X}) - \mu = 0$ , គេអាចនិយាយថា  $\bar{X}$  ជា unbiased estimator នៃ  $\mu$  ។



## ២.១.១- ឧទាហរណ៍នៃការវាយតម្លៃជាមធ្យមនៅក្នុង SPSS

នៅក្នុង SPSS ដើម្បីធ្វើការវាយតម្លៃជាមធ្យម គេអនុវត្តទៅលើ Tool “Descriptive analysis” ។ Descriptive analysis គឺជាការសិក្សាវិភាគដែលអនុវត្តទៅលើប្រភេទទិន្នន័យដែលជា Quantitative ហើយដែលក្នុងនោះអាចផ្តល់នូវតម្លៃមធ្យម (Sample Mean) គំលាតស្តង់ដា (Standard deviation) កំហុសនៃការវាយតម្លៃទៅលើតម្លៃមធ្យម (Standard error mean) និងតម្លៃមួយចំនួន ទៀត ។

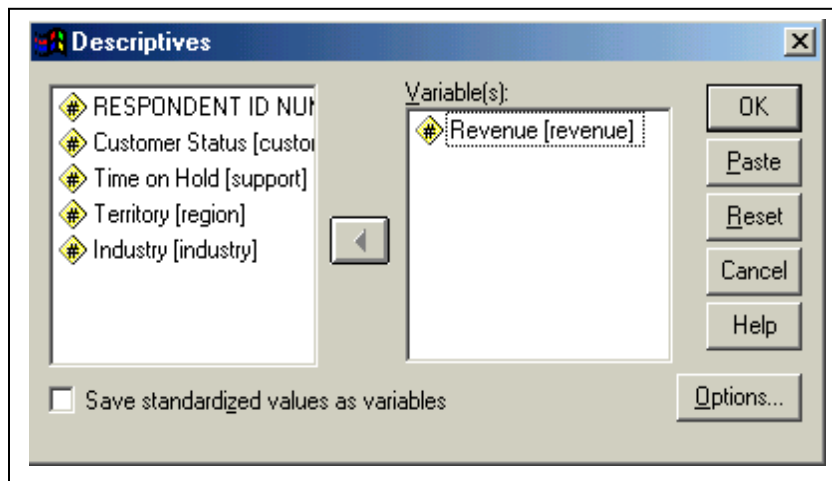
### ឧទាហរណ៍៖

សន្មតថាមានបំណងធ្វើការវាយតម្លៃទៅលើប្រាក់ចំណូលជាមធ្យមរបស់គ្រួសារដែលរស់នៅក្នុងតំបន់មួយ កន្លែង ។ ក្នុងករណីនេះគេធ្វើការអង្កេតទៅលើគ្រួសារមួយចំនួនដែលរស់នៅក្នុងតំបន់នេះដោយចៃដន្យមាន នយ័តម្ភច្រើនសិបនូវគំរូតាងស្ថិតិមួយសំរាប់ការវាយតម្លៃ ។

ទិន្នន័យដែលទទួលបានពីការអង្កេតនេះមានឈ្មោះថា “sales” ដែលកំណត់ទុកនៅក្នុង data file នៃ SPSS ។

ដើម្បីធ្វើការវាយតម្លៃទៅលើលទ្ធផលនៃការអង្កេតនេះ គេអនុវត្តដូចខាងក្រោម៖

Analyze → Descriptive statistics → Descriptives



លទ្ធផលដែលទទួលបានពីការវិភាគនេះមានដូចខាងក្រោម៖

| Descriptive Statistics |           |           |           |            |            |           |
|------------------------|-----------|-----------|-----------|------------|------------|-----------|
|                        | N         | Minimum   | Maximum   | Mean       |            | Std.      |
|                        | Statistic | Statistic | Statistic | Statistic  | Std. Error | Statistic |
| Revenue                | 1461      | \$500     | \$6,213   | \$2,557.53 | \$25.02    | \$956.334 |
| Valid N (listwise)     | 1461      |           |           |            |            |           |

រៀបរៀងដោយ ម៉ាក កាមេរ៉ាន (Statistician, Mathematician).

### ការបកស្រាយ:

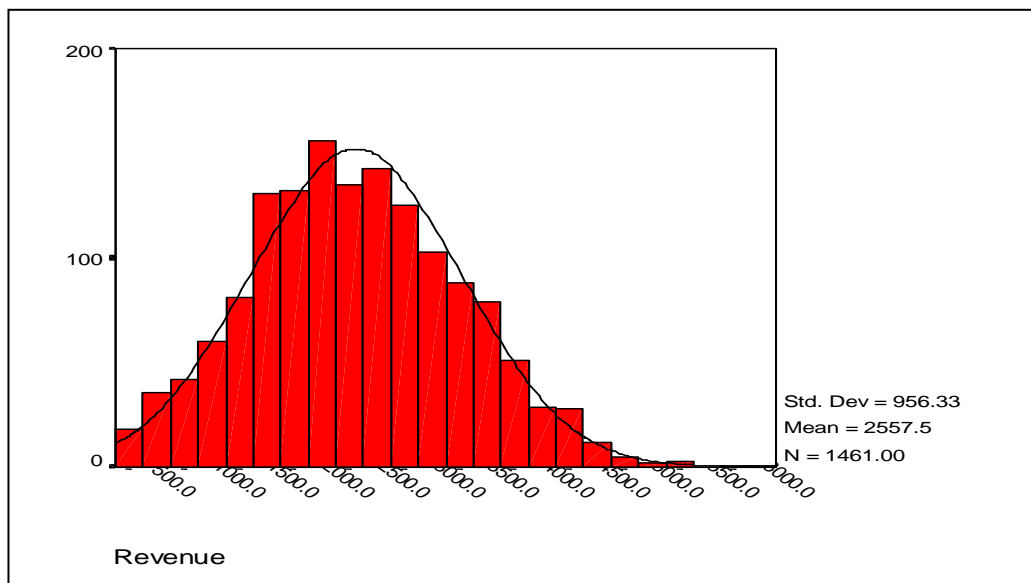
ការវិភាគនេះបានបង្ហាញអោយឃើញថាក្នុងចំណោម 1461 គ្រួសារដែលយើងបាន ធ្វើការអង្កេត នេះ គ្រួសារដែលមានប្រាក់ចំណូលទាបជាងគេគឺទទួលបាន \$500និងគ្រួសារដែលទទួលបានប្រាក់ចំណូលខ្ពស់ ជាងគេគឺ \$6213 ។

តាមរយៈនៃលទ្ធផលនេះផងដែរយើងអាចវាយតម្លៃថានៅក្នុងតំបន់នេះគ្រួសារនីមួយៗមានប្រាក់ ចំណូលជា មធ្យមប្រចាំឆ្នាំ \$2557.53 ហើយដែល មានគ្រួសារខ្លះទទួលបានលើសនេះ ឬ ទាបជាង នេះដោយគំណាតមធ្យម \$956.334 ។ យើងសង្កេតឃើញថា កំហុសជាមធ្យមនៃការវាយតម្លៃទៅ លើ មធ្យមនៃប្រាក់ចំណូលរបស់គ្រួសារនីមួយៗគឺ \$25.02

### **Graphics:**

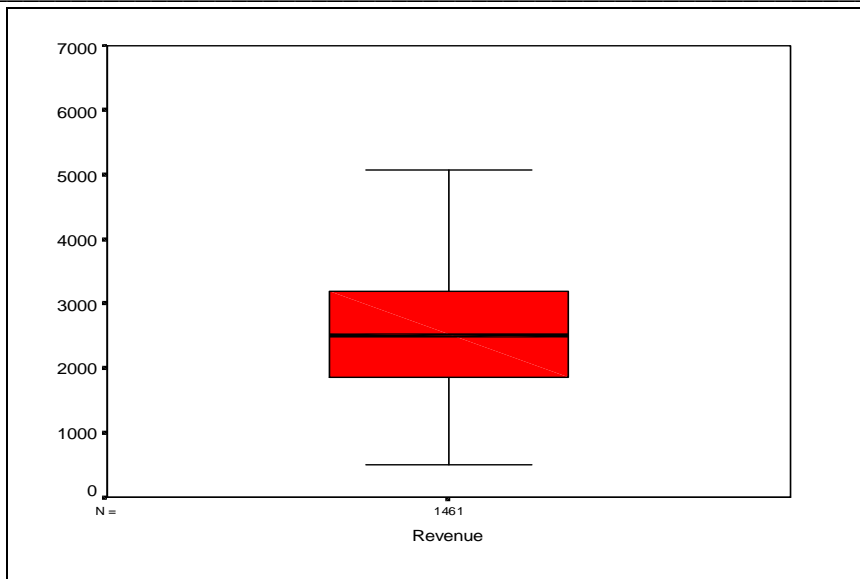
មាន Graphics ពីរប្រភេទដែលគេប្រើញឹកញាប់ក្នុងការបកស្រាយទៅលើលទ្ធផលនៃ ការវិភាគតាម រយៈ:Descriptive analysis គឺ Histogram ឬ Boxplot ។

### **Histogram:**



Graphicនេះបានបង្ហាញអោយឃើញថាភាគច្រើននៃអ្នករស់នៅក្នុងតំបន់នេះមានប្រាក់ចំណូលប្រចាំ ឆ្នាំប្រែប្រួលពី \$1500 ទៅ \$3600 ។ ចំណែកអ្នកមានប្រាក់ចំណូលប្រចាំឆ្នាំក្រោម \$1500 ឬ ខ្ពស់ ជាង \$3600 មានមិនជាច្រើនប៉ុន្មានទេ ។

### **Boxplot:**



### ចំណាំ:

គេអាចធ្វើការវិភាគវាយតម្លៃជាមធ្យមនៅក្នុង SPSS តាមរយៈ Syntax:

#### DESCRIPTIVES

/ Variable=Revenue

/ Statistics= Min Max Mean Stddev SEmean.

### ២.១.២- ឧទាហរណ៍នៃការវាយតម្លៃដោយប្រៀបធៀបជាមធ្យម:

ក្នុងករណីដែលគេមានបំណងវាយតម្លៃដោយប្រៀបធៀប តម្លៃមធ្យមទៅតាមក្រុម ( ក្នុង ឧទាហរណ៍ខាងលើ គឺប្រៀបធៀបទៅតាមតំបន់ ) នោះគេអាចវិភាគតាមរយៈ Syntax ខាងក្រោម:

#### MEANS

/ Table= Revenue BY Region

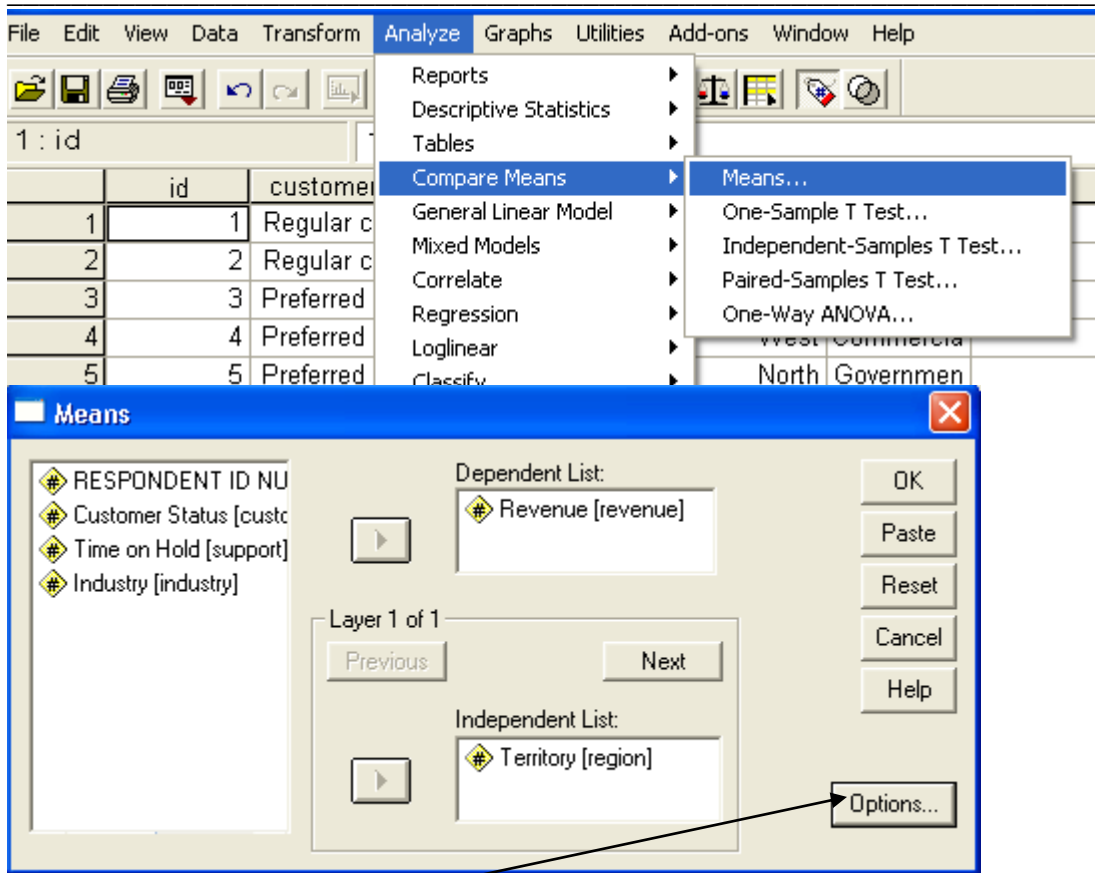
/Cells= Min Max Mean Stddev SEmean.

#### Report

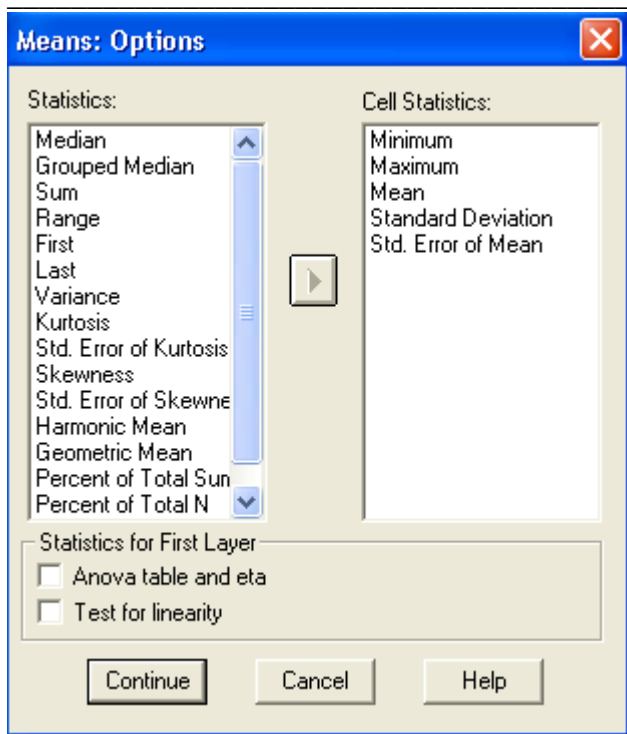
| Revenue   |         |         |            |                |                    |
|-----------|---------|---------|------------|----------------|--------------------|
| Territory | Minimum | Maximum | Mean       | Std. Deviation | Std. Error of Mean |
| North     | \$509   | \$6,024 | \$2,476.90 | \$963.195      | \$49.806           |
| South     | \$545   | \$6,213 | \$2,575.66 | \$981.426      | \$53.701           |
| East      | \$500   | \$5,417 | \$2,560.01 | \$932.849      | \$48.694           |
| West      | \$500   | \$5,012 | \$2,617.59 | \$948.074      | \$48.256           |
| Total     | \$500   | \$6,213 | \$2,557.53 | \$956.334      | \$25.020           |

យើងក៏អាចប្រើការវិភាគតាមរយៈ Tool ដូចខាងក្រោម:

រៀបរៀងដោយ ម៉ាក កាមេរ៉ាន (Statistician, Mathematician).



ដោយ Click ទៅលើ button **options** ដើម្បីជ្រើសរើសយកតំលៃដែលយើងត្រូវការ:



### ២.១.៣- អនុវត្តនៅក្នុង S-plus:

នៅក្នុង S-plus យើងមាន Command ជាច្រើនសំរាប់ជំនួយក្នុងការវាយតម្លៃជាមធ្យម:

**mean(X,na.rm=T)** សំរាប់កំណត់រកតម្លៃមធ្យម

**var(X,na.method="available")** សំរាប់កំណត់រកតម្លៃវ៉ារ្យង់

**stdev(X,na.rm=T)** ប្រើសំរាប់កំណត់រកតម្លៃ Standard deviation

**function(sales)**

```
{
  Revenue<- sales[,3][!is.na(sales[,3])]
  Min<- min(Revenue)
  Max<- max(Revenue)
  Mean<- mean(Revenue, na.rm=T)
  Variance<- var(Revenue,na.method="available")
  Stddev<- stdev(Revenue,na.rm=T)
  SE.mean<-Stddev/sqrt(length(Revenue))
  Result<-c(Min,Max,Mean,Variance, Stddev, SE.mean)
  names(Result)<-c("Min","Max","Mean","Variance","Stddev",
                  "SE.mean")
  return(Result)
}
```

ក្នុងករណីដែលយើងមានបំណងក្នុងការប្រៀបធៀបតំលៃមធ្យមនៃ **Revenue** ទៅតាមតំបន់ (Region) នីមួយៗ យើងសរសេរ៖

```
function(sales)
{
  PROG<-function(sales,Region)
  {
    Revenue<- sales[,3][!is.na(sales[,3]) & sale[,5]==Region]
    Min<- min(Revenue)
    Max<- max(Revenue)
    Mean<- mean(Revenue, na.rm=T)
    Variance<- var(Revenue,na.method="available")
    Stddev<- stdev(Revenue,na.rm=T)
    SE.mean<-Stddev/sqrt(length(Revenue))
    Result<-c(Min,Max,Mean,Variance, Stddev, SE.mean)
    return(Result)
  }
  Res1<- PROG(sales,"North")
  Res2<- PROG(sales,"South")
  Res3<- PROG(sales,"East")
  Res4<- PROG(sales,"West")
  Table<-rbind(Res1,Res2,Res3,Res4)
  Rowname<-c("North","South","East","West")
  Colname<- c("Min","Max","Mean","Variance","Stddev",
              "SE.mean")
  dimnames(Table)<-list(Rowname, Colname)
  return(Table)
}
```

## ២.២- ការវាយតំលៃជាភាគរយ (Point estimates for proportion)

ក្នុងករណីនៃការវាយតំលៃជាភាគរយ ទិន្នន័យដែលស្រង់នៅក្នុងគំរូតាងស្ថិតិត្រូវតែជាប្រភេទ ទិន្នន័យបែបសភាពលក្ខណៈ (Qualitative data) ។

បើ  $P$  ជាប៉ារ៉ាម៉ែត្រនៃប្រូបាប៊ីលីតេដែលត្រូវវាយតំលៃ នោះអេស្ទីម៉ាទ័រប្រសើរបំផុតក្នុង ការវាយតំលៃនេះ (Unbiased estimator) គឺជាភាគរយដែលទទួលបាននៅក្នុងគំរូតាងស្ថិតិដែលតាងដោយ  $\hat{p}$  ដែលបំពេញលក្ខខណ្ឌ៖

$$E(\hat{p}) - P = 0$$

### បំណកស្រាយ៖

យោងតាមទ្រឹស្តី Binomial គេមាន ៖ បើ  $X$  ជាអថេរ Binomial នោះ  $E(X)=n \cdot p \Rightarrow$

$$\frac{E(X)}{n} = p \Rightarrow E(X/n)=p; \text{ ដោយ } p = \frac{X}{n}, \text{ នោះ } E(p) = p \Rightarrow E(p) - p = 0 \Rightarrow p \text{ ជា}$$

unbiased estimator នៃ  $p$  ។

នៅក្នុងការវាយតម្លៃនេះ លំអៀងស្តង់ដារ (Std. Error) ត្រូវបានកំណត់ដោយរូបមន្តដូចខាង

$$\text{ក្រោម៖ } S(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{។}$$

### **២.២.១- ឧទាហរណ៍នៃការវាយតម្លៃជាភាគរយនៅក្នុង SPSS:**

ការវាយតម្លៃជាភាគរយនៅក្នុង SPSS ធ្វើឡើងតាមរយៈ **Frequency analysis** ។

**FREQUENCY ANALYSIS:**

**Frequency analysis:** គឺជាការសិក្សាវិភាគវាយតម្លៃជាភាគរយឬការរាប់ចំនួនដែលអនុវត្ត

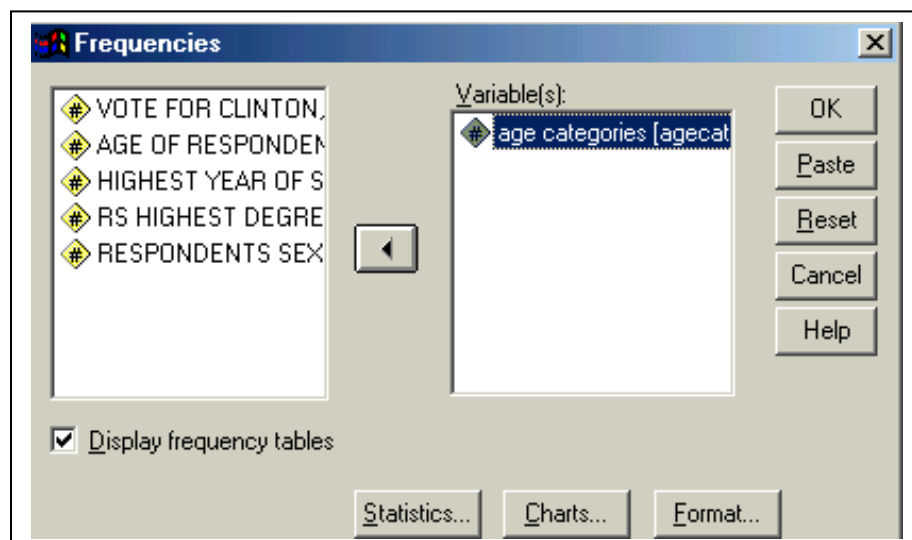
ទៅលើ qualitative data ។

### ឧទាហរណ៍៖

សន្មតថាគេមានបំណងវាយតម្លៃទៅលើភាគរយនៃអាយុអ្នកដែលនឹងចូលរួមទៅក្នុងការបោះឆ្នោតនៅក្នុងតំបន់ដ៏ធំមួយ ដោយធ្វើការបែងចែកទៅតាមថ្នាក់នៃអាយុ៖ “Lt 35”, “35-44”, “45-64” & “65+” ក្នុងករណីនេះគេបានធ្វើការអង្កេតទៅលើមនុស្សមួយចំនួនដែលមានសិទ្ធិបោះឆ្នោតដោយចៃដន្យនៅក្នុងតំបន់នេះ ហើយទិន្នន័យដែលស្រង់បានកំណត់ទុកនៅក្នុង data file នៃ SPSS ដែលមានឈ្មោះថា **voter** ។

ដោយធ្វើ “frequencies analysis” យើងបាន៖

Analyze → Descriptive statistics → Frequencies



រៀបរៀងដោយ ម៉ាក កាមេរ៉ាន (Statistician, Mathematician).

គេទទួលបានលទ្ធផលដូចខាងក្រោម៖

| age categories |         |           |         |               |                    |
|----------------|---------|-----------|---------|---------------|--------------------|
|                |         | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid          | lt 35   | 438       | 23.7    | 23.7          | 23.7               |
|                | 35 - 44 | 444       | 24.0    | 24.0          | 47.8               |
|                | 45 - 64 | 617       | 33.4    | 33.4          | 81.2               |
|                | 65 +    | 348       | 18.8    | 18.8          | 100.0              |
|                | Total   | 1847      | 100.0   | 100.0         |                    |

### ការបកស្រាយ៖

Frequency គឺជាបរិមាណនៃការអង្កេតដែលកើតឡើងនៅក្នុងគំរូស្ថិតិដែលយើងសិក្សា ហើយដែលបង្ហាញអំពីបរិមាណនៃអ្នកដែលនឹងចូលរួមនៅក្នុងការបោះឆ្នោតទៅតាមថ្នាក់នៃអាយុនីមួយៗ ។

$$Percent = \frac{Frequency}{Total}$$

$$Valid\ percent = \frac{Frequency}{Total - Missing}$$

ក្នុងករណីខាងលើនេះ Missing ស្មើ 0 , ដូច្នេះ Percent និង Valid percent មានតំលៃដូចគ្នា ។

ក្នុងករណីនៃការវាយតំលៃទៅលើការអង្កេត ជាទូទៅគេប្រើ Valid percent ។

តាមរយៈនៃការអង្កេតខាងលើគេអាចវាយតំលៃបានថាក្នុងចំណោមអ្នកដែលនឹងចូលរួមបោះឆ្នោតនាពេល ខាងមុខក្នុងតំបន់នេះនឹងមានអ្នកដែលមានអាយុ៖

“ ក្រោម 35 ឆ្នាំ “ ចំនួន 23.7%

“ 35-44 ឆ្នាំ “ ចំនួន 24%

“ 45-64 ឆ្នាំ “ ចំនួន 33.4%

“ 65 ឆ្នាំឡើងទៅ “ ចំនួន 18.8%

យ៉ាងណាមិញ តំលៃ Cumulative percent ផ្តល់លទ្ធភាពអោយយើងអាចវាយតំលៃជាបណ្តុំពីលើចុះក្រោម ។



**ឧទាហរណ៍ :**

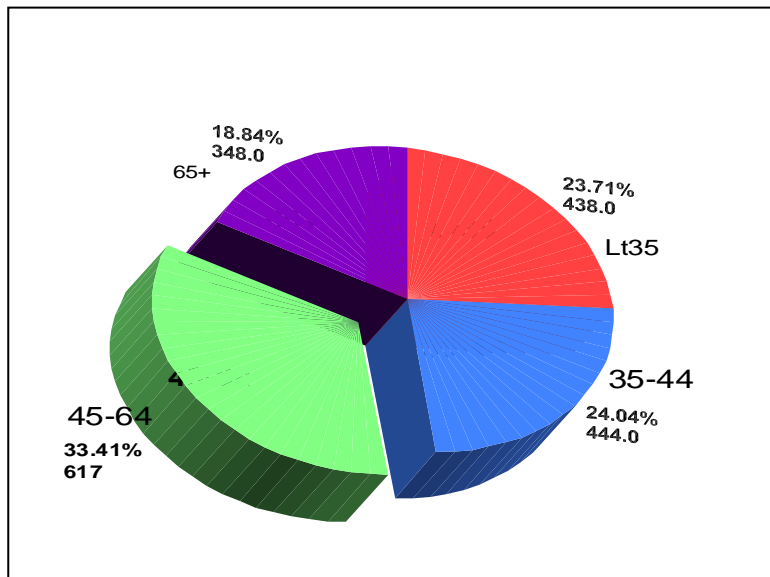
Cumulative percent=47.8% មានន័យថាជាភាគរយអ្នកដែលមានអាយុ ៤៤ ឆ្នាំចុះ

Graphics ដែលប្រើ ប្រាស់សំរាប់សំរួលការពន្យល់ទៅលើលទ្ធផលនៃ Frequencies analysis មាន

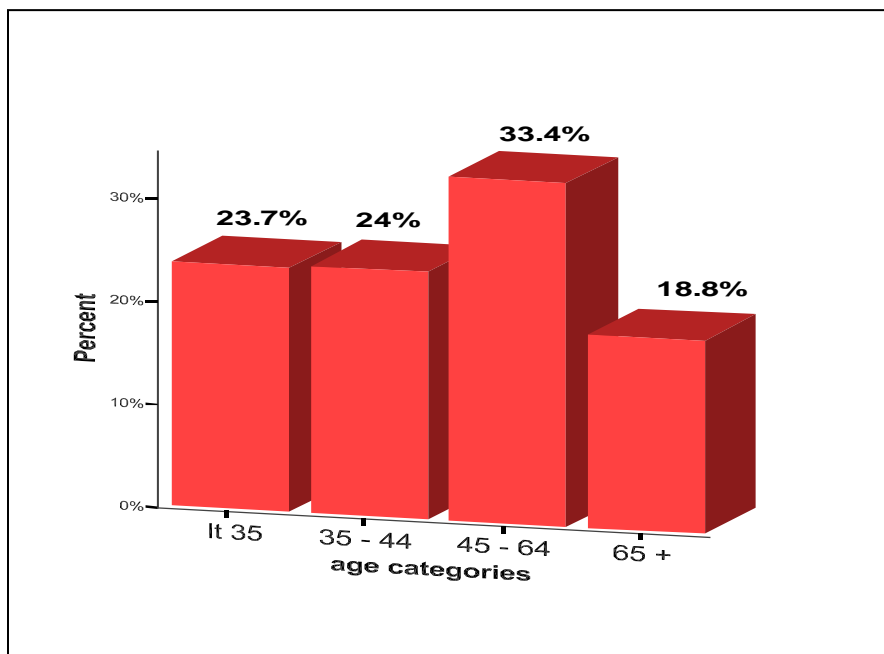
Piechart ឬ Barchart ។

**Graphics:**

PIECHART



BARCHART



### ២.២.២- ការវិភាគតាមរយៈ S-plus:

នៅក្នុង S-plus ការវាយតម្លៃជាភាគរយធ្វើឡើងតាមរយៈនៃការបង្កើត **function** ដោយខ្លួនឯងដោយគោរពទៅតាមទ្រឹស្តីនៃការកំណត់រក **Frequency, Valid percent, Cumulative percent** ។

```
function(voter)
{
  Age<-voter[,3][!is.na(voter[,3])]
  Frequency<-summary(Age)
  Percent<-Frequency/length(Age)
  Table<-cbind(Frequency,Percent)
  return(Table)
}
```

#### Output

| Age     | Frequency | Percent   |
|---------|-----------|-----------|
| Lt 35   | 438       | 0.2371413 |
| 35 - 44 | 444       | 0.2403898 |
| 45 - 64 | 617       | 0.3340552 |
| 65 +    | 348       | 0.1884136 |

### ២.២.៣- ការវាយតម្លៃជាភាគរយតាមរយៈ Crosstabs

Crosstabs ជាបច្ចេកទេសដែលអាចវាយតម្លៃជាភាគរយជាលក្ខណៈខ្វែងគ្នារវាងអថេរពីរ ដែលប្រភេទទិន្នន័យជា Qualitative ។ នៅលទ្ធផលនៃការវិភាគ យើងអាចទទួលបានតម្លៃភាគរយដែលគណនាតាមគោលការណ៍នៃ Probability និងតាមរយៈទ្រឹស្តីនៃប្រូបាប៊ីលីតេមានលក្ខខណ្ឌ (Conditional probability) ។

#### ឧទាហរណ៍:

សន្មតថាមានបំណងវាយតម្លៃទៅលើអ្នកដែលនឹងចូលរួមទៅក្នុងការបោះឆ្នោតនៅក្នុងតំបន់ដ៏ធំមួយតាមរយៈសំណួរខាងក្រោម:

ក- ភេទ

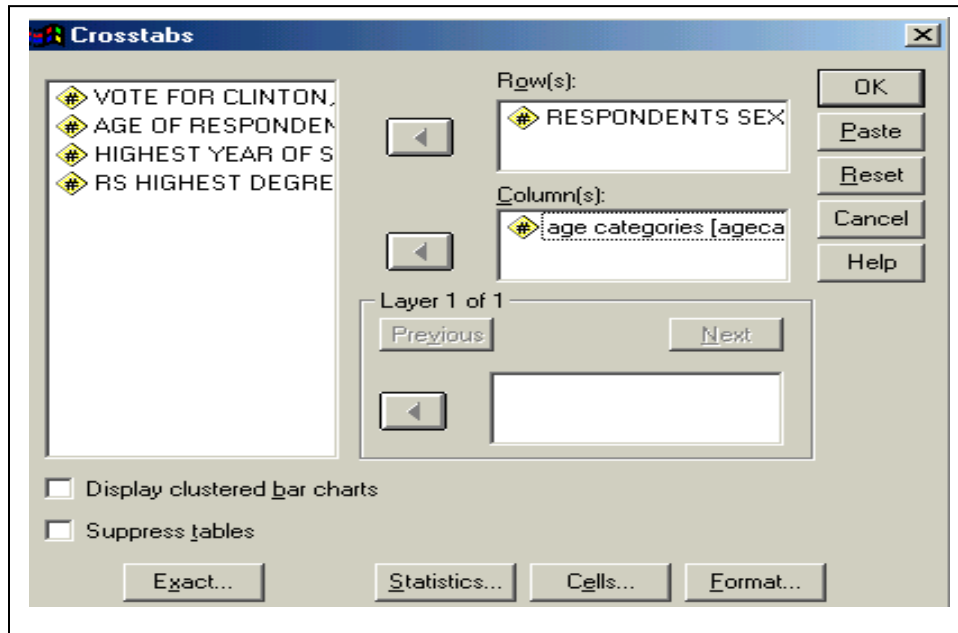
☐Male ☐Female

ខ- អាយុ:

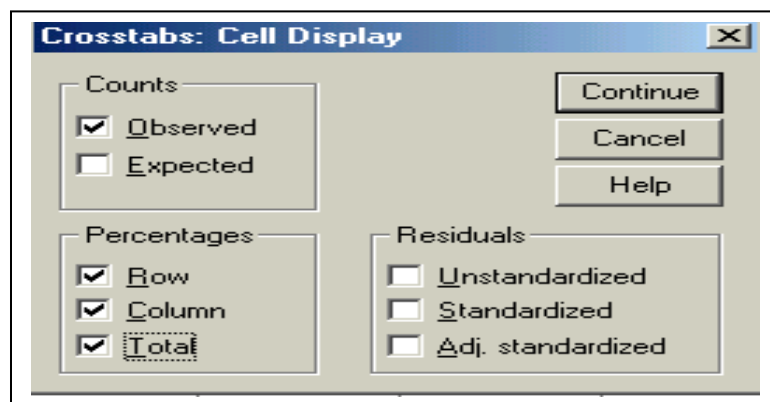
☐Lt35 ☐35-44 ☐45-64 ☐65+

ដោយធ្វើការអង្កេតទៅលើមនុស្សមួយចំនួនដែលមានសិទ្ធិបោះឆ្នោតនៅក្នុងតំបន់នេះដោយចៃដន្យ ហើយទិន្នន័យដែលស្រង់បានកំណត់ទុកនៅក្នុង data file នៃ SPSS ឈ្មោះ “voter” ។ ដោយប្រើ SPSS ក្នុងការវិភាគ៖

Analyze→Descriptive statistics→crosstabs



ដោយចុចទៅលើ button “cell” នៃប្រអប់ Crosstabs យើងបាន៖



តាមរយៈនៃការវិភាគនេះយើងអាចទទួលបានតារាងរបាយការណ៍ 4 ខុសៗគ្នា៖

- **តារាងរបាយការណ៍ទី១:** គឺជាតារាងសង្ខេបនូវចំនួន Frequency ជាលក្ខណៈខ្លាំងគួររវាងអថេរទាំងពីរ ។ ក្នុងករណីនេះយើងជ្រើសរើសយកតែ “observed” តែប៉ុណ្ណោះ យើងទទួលបាន:

| RESPONDENTS SEX * age categories Crosstabulation |        |                |         |         |      |       |
|--------------------------------------------------|--------|----------------|---------|---------|------|-------|
| Count                                            |        |                |         |         |      |       |
|                                                  |        | age categories |         |         |      | Total |
|                                                  |        | lt 35          | 35 - 44 | 45 - 64 | 65 + |       |
| RESPONDENTS                                      | male   | 181            | 199     | 287     | 137  | 804   |
| SEX                                              | female | 257            | 245     | 330     | 211  | 1043  |
| Total                                            |        | 438            | 444     | 617     | 348  | 1847  |

- **តារាងរបាយការណ៍ទី២:**  
គឺជាតារាងប្រើប្រាស់សំរាប់ការវាយតម្លៃជាភាគរយធៀបទៅនឹងជួរដេកនៃតារាងរបាយការណ៍ ។ ដើម្បីទទួលបានតារាងរបាយការណ៍នេះយើងជ្រើសរើសនៅក្នុងប្រអប់ “cell” នូវពាក្យ “ROW” យើងបាន:

| RESPONDENTS SEX * age categories Crosstabulation |        |                |         |         |       |        |
|--------------------------------------------------|--------|----------------|---------|---------|-------|--------|
| % within RESPONDENTS SEX                         |        |                |         |         |       |        |
|                                                  |        | age categories |         |         |       | Total  |
|                                                  |        | lt 35          | 35 - 44 | 45 - 64 | 65 +  |        |
| RESPONDENTS                                      | male   | 22.5%          | 24.8%   | 35.7%   | 17.0% | 100.0% |
| SEX                                              | female | 24.6%          | 23.5%   | 31.6%   | 20.2% | 100.0% |
| Total                                            |        | 23.7%          | 24.0%   | 33.4%   | 18.8% | 100.0% |

ក្នុងតារាងនេះយើងអាចប្រៀបធៀបបានតែភាគរយធៀបនឹងជួរដេកតែប៉ុណ្ណោះ មិនអាចប្រៀបធៀបភាគរយធៀបនឹងជួរឈរទេ ។

### ការបកស្រាយ:

22.5% មានន័យថាក្នុងចំណោមមនុស្សប្រុសទាំងអស់ដែលនឹងចូលរួមក្នុងការបោះឆ្នោត មាន

22.5% មានអាយុក្រោម 35 ឆ្នាំ ។

ដូចគ្នានេះដែរយើងអាចនិយាយថាក្នុងចំណោមមនុស្សប្រុសទាំងអស់ដែលនឹងចូលរួមក្នុងការ

បោះឆ្នោតមាន 22.5% មានអាយុក្រោម 35ឆ្នាំ 24.8% មានអាយុនៅចន្លោះពី 35 ទៅ 44 ឆ្នាំ 35.7% មានអាយុនៅចន្លោះពី 45 ទៅ 64ឆ្នាំ និង 17% មានអាយុចាប់ពី 65ឆ្នាំ ឡើងទៅ ។ ដូចនេះយើងអាចសន្និដ្ឋានបានថាក្នុងចំណោមមនុស្សប្រុសដែលនឹងចូលរួមក្នុងការបោះឆ្នោត មានមនុស្សប្រុសដែលមាន អាយុ “45-64” ច្រើនជាងគេរហូតដល់ 37.5% ។ ចំណែកជួរដេក “Female” ក៏យើងធ្វើការសន្និដ្ឋានដូចខាងលើដែរ ។

ចំណែកឯភាគរយដែលស្ថិតនៅជួរដេក “Total” ជាភាគរយដែលពន្យល់ទៅលើ ភាគរយអ្នកដែលនឹងចូលរួមនៅក្នុងការបោះឆ្នោត តាមថ្នាក់នៃ អាយុនីមួយៗ ។

● **តារាងរបាយការណ៍ទី៣:**

គឺជាតារាងប្រើប្រាស់សំរាប់ការវាយតម្លៃជាភាគរយធៀបទៅនឹងជួរឈរនៃ តារាងរបាយការណ៍ ។ ដើម្បីទទួលបានតារាងរបាយការណ៍នេះយើង ជ្រើសរើសនៅក្នុងប្រអប់ “cell” នូវពាក្យ “Column” យើងបាន:

| RESPONDENTS SEX * age categories Crosstabulation |        |                |         |         |        |
|--------------------------------------------------|--------|----------------|---------|---------|--------|
| % within age categories                          |        | age categories |         |         |        |
|                                                  |        | lt 35          | 35 - 44 | 45 - 64 | 65 +   |
| RESPONDENTS SEX                                  | male   | 41.3%          | 44.8%   | 46.5%   | 39.4%  |
|                                                  | female | 58.7%          | 55.2%   | 53.5%   | 60.6%  |
| Total                                            |        | 100.0%         | 100.0%  | 100.0%  | 100.0% |

ក្នុងតារាងនេះយើងអាចប្រៀបធៀបបានតែភាគរយធៀបនឹងជួរឈរតែប៉ុណ្ណោះ មិនអាចប្រៀបធៀបភាគរយធៀបនឹងជួរដេកទេ ។

**ការបកស្រាយ:**

យើងអាចនិយាយថាក្នុងចំណោមអ្នកដែលមានអាយុក្រោម 35ឆ្នាំទាំងអស់ដែលនឹងចូលរួមក្នុងការបោះឆ្នោតមាន 41.3% ជាមនុស្សប្រុស និង 58.7% ជាមនុស្សស្រី ចំណែកជួរឈរផ្សេងទៀតក៏យើងធ្វើការសន្និដ្ឋានដូចគ្នានេះដែរ ។ ចំណែកឯភាគរយដែលស្ថិតនៅជួរឈរ “Total” ជាភាគរយដែលអាចពន្យល់ថានឹងមានមនុស្សប្រុស 43.5% និងមនុស្សស្រី 56.5% ដែលនឹងចូលរួមនៅក្នុងការបោះឆ្នោតនាពេលខាងមុខ ។

• **តារាងរបាយការណ៍ទី៤:**

គឺជាតារាងប្រើប្រាស់សំរាប់ការវាយតម្លៃជាភាគរយធៀបទៅចំនួនសរុបនៃការអង្កេត ។ ដើម្បីទទួលបានតារាងរបាយការណ៍នេះយើងជ្រើសរើសនៅក្នុងប្រអប់ “cell” នូវពាក្យ “Total” យើងបាន:

| RESPONDENTS SEX * age categories Crosstabulation |        |                |         |         |       |        |
|--------------------------------------------------|--------|----------------|---------|---------|-------|--------|
| % of Total                                       |        | age categories |         |         |       | Total  |
|                                                  |        | lt 35          | 35 - 44 | 45 - 64 | 65 +  |        |
| RESPONDENTS                                      | male   | 9.8%           | 10.8%   | 15.5%   | 7.4%  | 43.5%  |
| SEX                                              | female | 13.9%          | 13.3%   | 17.9%   | 11.4% | 56.5%  |
|                                                  | Total  | 23.7%          | 24.0%   | 33.4%   | 18.8% | 100.0% |

ក្នុងករណីនេះភាគរយទាំងអស់ដែលជាប្រសព្វរវាងជួរដេកនិងជួរឈរគឺជាភាគរយដែលវាយតម្លៃធៀបនឹងបរិមាណនៃការអង្កេតទាំងអស់ ។

**ការបកស្រាយ:**

ក្នុងចំណោមអ្នកដែលនឹងចូលរួមនៅក្នុងការបោះឆ្នោតខាងមុខមាន:

- 9.8% ជាមនុស្សប្រុសដែលមានអាយុក្រោម 35 ឆ្នាំ
- 10.8% ជាមនុស្សប្រុសដែលមានអាយុ 35-44 ឆ្នាំ
- 15.5% ជាមនុស្សប្រុសដែលមានអាយុ 45-64 ឆ្នាំ
- 7.4% ជាមនុស្សប្រុសដែលមានអាយុ 65 ឆ្នាំឡើងទៅ
- 13.9% ជាមនុស្សស្រីដែលមានអាយុ ក្រោម 35 ឆ្នាំ
- 13.3% ជាមនុស្សស្រីដែលមានអាយុ 35-44 ឆ្នាំ
- 17.9% ជាមនុស្សស្រីដែលមានអាយុ 45-64 ឆ្នាំ
- 11.4% ជាមនុស្សស្រីដែលមានអាយុ 65 ឆ្នាំឡើងទៅ

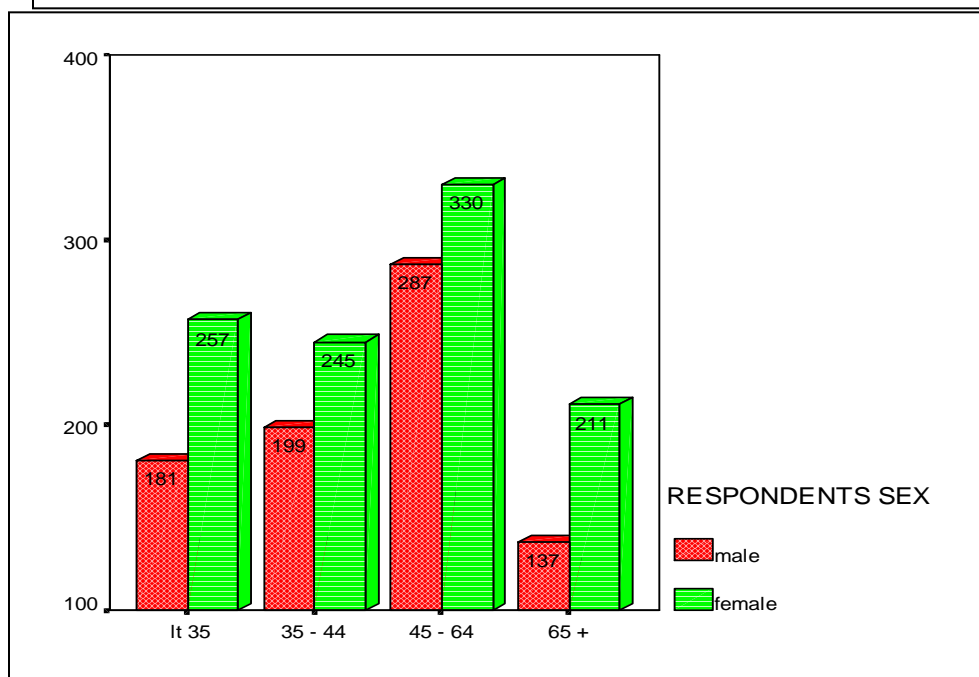
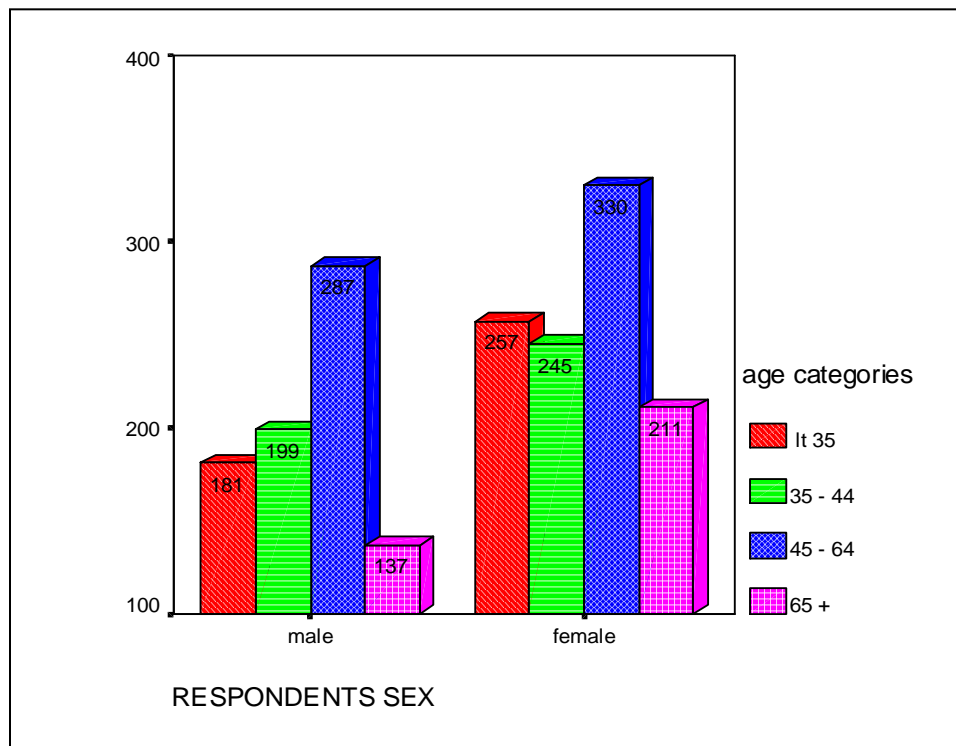
ចំណែកអ្នកដែលនឹងចូលរួមក្នុងការបោះឆ្នោតនាពេលខាងមុខមានមនុស្សប្រុសចំនួន 43.5% និងមនុស្សស្រីចំនួន 56.5% ។ ដូចគ្នានេះគេអាចធ្វើការបកស្រាយទៅលើភាគរយសរុបនៃថ្នាក់អាយុរបស់អ្នកដែលចូលរួមនៅក្នុងការបោះឆ្នោត:

- 23.7% មានអាយុក្រោម 35 ឆ្នាំ
- 24% មានអាយុ 35-44 ឆ្នាំ

- 33.4% មានអាយុ 45-64 ឆ្នាំ
- 18.8% មានអាយុចាប់ពី 65 ឆ្នាំឡើងទៅ

### GRAPHICS:

Graphic ដែលបកស្រាយល្បឿននិងភេទនៃអ្នកចូលរួមក្នុងការបោះឆ្នោត:



## ២.២.៤- Function S-plus សំរាប់ការវាយតម្លៃជាភាគរយតាមរយៈ Crosstabs

ដើម្បីបង្កើតការវិភាគវាយតម្លៃជាភាគរយជាលក្ខណៈខ្លែងគ្នាតាមរយៈនៃ Crosstabs យើងមាន

Command បញ្ជាដូចខាងក្រោម៖

**crosstabs(~Variable1+ Variable2, na.action=na.exclude)**

ដែល :

Variable1 ជា qualitative variable ដែលមានតួនាទីជា factor row

Variable2 ជា qualitative variable ដែលមានតួនាទីជា factor column

na.action=na.exclude ប្រើសំរាប់លុបតម្លៃ NA ពី តួនៃអថេរ

```
function(voter)
{
  Age<- voter[,3]
  Gender<- voter[,6]
  Table<-crosstabs(~Age+Gender,na.action=na.exclude)
  return(Table)
}
```

នៅក្នុងថត (Cell) នីមួយៗនៃតារាង Crosstabs ដែលជាលទ្ធផលនៃ function ខាងលើមាន 4

តម្លៃក្នុងនោះមាន៖

- ចំនួន (Observed frequency)
- ប្រូបាប៊ីលីតេមានលក្ខខណ្ឌធៀបនឹងជួរដេក
- ប្រូបាប៊ីលីតេមានលក្ខខណ្ឌធៀបនឹងជួរឈរ
- ប្រូបាប៊ីលីតេនៃផលគុណ

|            |  |
|------------|--|
|            |  |
| N          |  |
| N/RowTotal |  |
| N/ColTotal |  |
| N/Total    |  |
|            |  |



**output:**

Call:

crosstabs( ~ Age + Gender, na.action = na.exclude)  
1847 cases in table

**age categories \* RESPONDENTS SEX Crosstabulation**

|                |         | RESPONDENTS SEX |        | Rowtotal |
|----------------|---------|-----------------|--------|----------|
|                |         | male            | female |          |
| age categories | lt 35   | 181             | 257    | 438      |
|                |         | 41.3%           | 58.7%  | 100.0%   |
|                |         | 22.5%           | 24.6%  | 23.7%    |
|                |         | 9.8%            | 13.9%  | 23.7%    |
|                | 35 - 44 | 199             | 245    | 444      |
|                |         | 44.8%           | 55.2%  | 100.0%   |
|                |         | 24.8%           | 23.5%  | 24.0%    |
|                |         | 10.8%           | 13.3%  | 24.0%    |
|                | 45 - 64 | 287             | 330    | 617      |
|                |         | 46.5%           | 53.5%  | 100.0%   |
|                |         | 35.7%           | 31.6%  | 33.4%    |
|                |         | 15.5%           | 17.9%  | 33.4%    |
|                | 65 +    | 137             | 211    | 348      |
|                |         | 39.4%           | 60.6%  | 100.0%   |
|                |         | 17.0%           | 20.2%  | 18.8%    |
|                |         | 7.4%            | 11.4%  | 18.8%    |
| ColTotal       |         | 804             | 1043   | 1847     |
|                |         | 43.5%           | 56.5%  | 100.0%   |
|                |         | 100.0%          | 100.0% | 100.0%   |
|                |         | 43.5%           | 56.5%  | 100.0%   |

## ២.២.៥- ការ Recode ទិន្នន័យសំរាប់ការវិភាគ

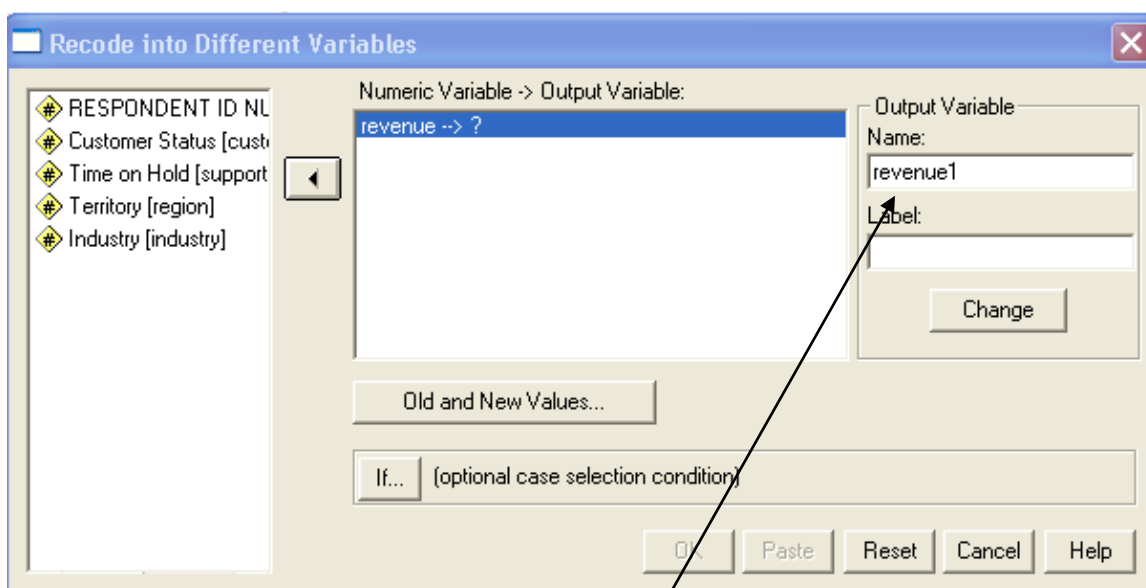
នៅពេលខ្លះទិន្នន័យពុំមានលក្ខណៈសមប្រកបសំរាប់ការវាយតម្លៃជាភាគរយ ដែលតម្រូវអោយយើងធ្វើ Recode នៃទិន្នន័យនេះជាចាំបាច់ដើម្បីតម្រូវការវិភាគដែលយើងចង់បាន ។

### ឧទាហរណ៍:

ដោយអនុវត្តទៅលើតារាងទិន្នន័យ sales នៃ SPSS ដែលទទួលបានពីការអង្កេតទៅលើបុគ្គលមួយចំនួន ចូរធ្វើការវាយតម្លៃជាភាគរយទៅលើ Revenue/year របស់បុគ្គលនីមួយៗដែលរស់នៅក្នុងតំបន់ផ្សេងៗគ្នា ទៅតាមទំរង់តារាងដូចខាងក្រោម:

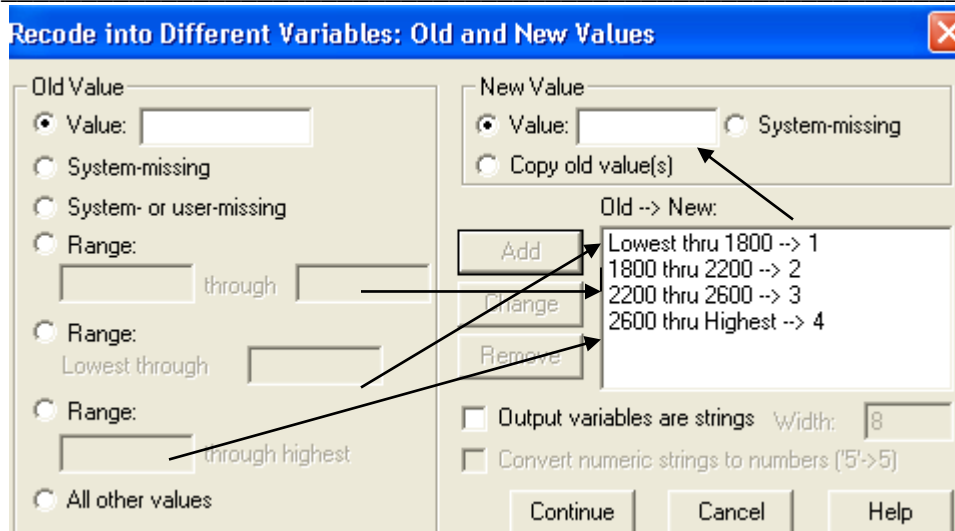
|       | Revenue/year |               |               |         |       |
|-------|--------------|---------------|---------------|---------|-------|
|       | <\$1800      | \$1800-\$2200 | \$2200-\$2600 | >\$2600 | Total |
| North |              |               |               |         |       |
| South |              |               |               |         |       |
| East  |              |               |               |         |       |
| West  |              |               |               |         |       |
| Total |              |               |               |         |       |

ដោយសារទិន្នន័យនៅក្នុងជួរឈរ Revenue ជាប្រភេទ Quantitative ដែលតម្រូវអោយយើងធ្វើការបំប្លែងទៅជា Qualitative តាមរយៈនៃការធ្វើ Recode នៅក្នុង SPSS:  
Transform→Recode→Into different variable



កំណត់ឈ្មោះ: អថេរថ្មី **Revenue1**

ដើម្បីផ្តល់ code អោយទិន្នន័យថ្មីនៅក្នុងជួរឈរ Revenue1 ចុចទៅលើ Button “Old & New value” យើងទទួលបាន:



ដោយចុចទៅលើ Continue យើងទទួលបាន Code នៃទិន្នន័យថ្មីនៅក្នុង variable “Revenue1” ។

ដោយផ្តល់ Value label អោយ code នៃទិន្នន័យក្នុង Variable Revenue1 យើងបាន៖

| Name     | Type    | Width | Decimals | Label         | Values         |
|----------|---------|-------|----------|---------------|----------------|
| id       | Numeric | 4     | 0        | RESPONDEN     | None           |
| customer | Numeric | 1     | 0        | Customer Stat | {1, Regular cu |
| revenue  | Dollar  | 8     | 0        | Revenue       | None           |
| support  |         |       |          |               | Minute)        |
| region   |         |       |          |               | th}...         |
| industry |         |       |          |               | vernmen        |
| revenue1 |         |       |          |               |                |

| Value | Value Label     |
|-------|-----------------|
| 1.00  | "<\$1800"       |
| 2.00  | "\$1800-\$2200" |
| 3.00  | "\$2201-\$2600" |

ក្រោយពីទទួលបានទិន្នន័យដែលយើងត្រូវការសំរាប់ការវិភាគសមស្រប យើងអាចទទួលបានតារាងវាយតម្លៃជាភាគរយដែលមានទំរង់ជាតារាង Crosstabs ដែលមាន 4 ប្រភេទ ។ តាមរយៈ Syntax ខាងក្រោម យើងអាចទទួលបានតារាង 4 ខុសៗគ្នា៖

**Syntax:**

**CROSSTABS**  
**/Tables= Region BY Revenue1**  
**/Format= avalue Tables**  
**/Cells=Count Row Column Total.**

**Count :** Frequency table

**Row :** Percent within row

**Column :** Percent within column

**Total :** Percent of total

### ៣. ការវាយតម្លៃដោយចន្លោះជឿជាក់ (Confidence interval estimates)

ទោះបី  $\hat{\theta}$  ជា unbiased estimator ក្នុងការវាយតម្លៃទៅលើប៉ារ៉ាម៉ែត្រនៃប្រព័ន្ធប្រជាប្រិយស្វ័យ  $\theta$  យ៉ាងណាក៏ដោយក៏នៅតែមានលំអៀង (Error) រវាងតម្លៃអង្កេតនិងតម្លៃពិតដែលគេអោយឈ្មោះថា “Margin of Error” ដែលតាងដោយ  $E$ ,  $E = |\hat{\theta} - \theta|$  ជាលំអៀងធំបំផុតនៃការវាយតម្លៃ ។ ដោយសារ  $\hat{\theta}$  វាយតម្លៃអោយ  $\theta$  មានលំអៀង ធំបំផុត  $E$  នោះវាអាចមានលំអៀង ខាងឆ្វេងនៃតម្លៃ  $\theta$  ឬខាងស្តាំតម្លៃ  $\theta$  ដូច្នេះ :

$$\hat{\theta} - E \leq \theta \leq \hat{\theta} + E$$

ចន្លោះនេះអាចជឿជាក់បាន លុះត្រាតែ ៖

$$P(\hat{\theta} - E \leq \theta \leq \hat{\theta} + E) = (1 - \alpha) \quad \text{ជាតម្លៃដែលខិតទៅរក 1 ។}$$

ក្នុងករណីនេះ  $(1 - \alpha)$  ជា កម្រិតជឿជាក់ (level of Confidence )

$\hat{\theta} - E$  អោយឈ្មោះថា Lower Bound (LB)

$\hat{\theta} + E$  អោយឈ្មោះថា Upper Bound (UB)

ដូច្នេះដើម្បីបង្កើត ចន្លោះជឿជាក់សំរាប់ការវាយតម្លៃជាមធ្យម ឬ ភាគរយ យើងត្រូវចាំបាច់កំណត់រក Margin of Error  $E$  :

$$E = ?$$

### ៣.១- ការវាយតម្លៃជាមធ្យមដោយចន្លោះជឿជាក់ (Confidence interval estimates for mean)

ក្នុងករណីនេះ ទិន្នន័យដែលទទួលបានពីការអង្កេតជាប្រភេទ Quantitative data ដូច្នេះ មធ្យមនៃគំរូតាង (sample)  $\bar{X}$  ជា unbiased estimator សំរាប់ការវាយតម្លៃទៅលើប៉ារ៉ាម៉ែត្រនៃប្លូតូយឡាស្យុង  $\mu$  ។

$$P(\bar{X} - E \leq \mu \leq \bar{X} + E) = (1 - \alpha)$$

ចូរកំណត់រក E:

បំប្លែងពីរបាយទិន្នន័យ Normal មករបាយទិន្នន័យ Normal Standard យើងបាន:

យោងតាមទ្រឹស្តី *Central Limit Theorem* គេមាន៖

$$\bar{X} \sim N(\mu, \sigma^2 / n) \rightarrow Z = \frac{\bar{X} - E(\bar{X})}{\sigma(\bar{X})} \rightarrow Z \sim N(0,1)$$

$$P(\bar{X} - E \leq \mu \leq \bar{X} + E) = (1 - \alpha), (1) \quad P(-Z_{\alpha/2} \leq \frac{\bar{X} - E(\bar{X})}{\sigma(\bar{X})} \leq Z_{\alpha/2}) = (1 - \alpha)$$

ដោយ៖

$$\bar{X} \sim N(\mu, \sigma^2 / n)$$

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

$$\Rightarrow \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx S(\bar{X}) = \frac{S}{\sqrt{n}}$$

$$P(-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{S / \sqrt{n}} \leq Z_{\alpha/2}) = (1 - \alpha)$$

$$P(-Z_{\alpha/2} S / \sqrt{n} \leq \bar{X} - \mu \leq Z_{\alpha/2} S / \sqrt{n}) = (1 - \alpha)$$

$$P(-\bar{X} - Z_{\alpha/2} S / \sqrt{n} \leq -\mu \leq -\bar{X} + Z_{\alpha/2} S / \sqrt{n}) = (1 - \alpha)$$

$$P(\bar{X} + Z_{\alpha/2} S / \sqrt{n} \geq \mu \geq \bar{X} - Z_{\alpha/2} S / \sqrt{n}) = (1 - \alpha)$$

$$\text{ដូច្នេះ ៖ } P(\bar{X} - Z_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{S}{\sqrt{n}}) = (1 - \alpha), (2)$$

ដោយប្រៀបធៀប (1) & (2) គេបាន៖

$$E = Z_{\alpha/2} \frac{S}{\sqrt{n}},$$

$Z_{\alpha/2}$  អាចកំណត់តម្លៃបានកាលណា យើងស្គាល់តម្លៃ  $(1 - \alpha)$

ក្នុងករណីនេះ Confidence interval កំណត់

$$\text{Lower bound} = \bar{X} - Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$$\text{Upper bound} = \bar{X} + Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

**ចំណាំ៖**

តាមការស្រាយបញ្ជាក់ខាងលើ រូបមន្តទូទៅនៃតំលៃ Margin of Error ត្រូវបានកំណត់ដោយ៖

$$\text{Margin of Error (E)} = Z_{\alpha/2} * \text{Std.error}$$

### ៣.១.១- ការកំណត់រក Confidence interval for mean នៅក្នុង SPSS

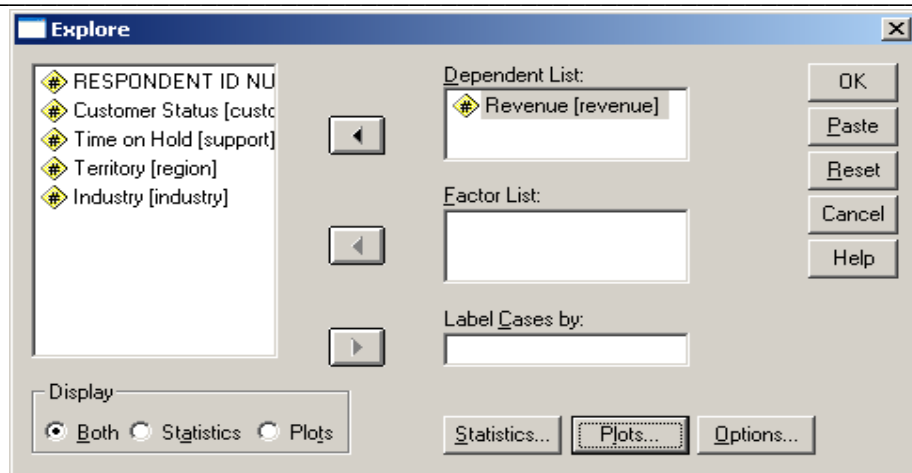
ជាទូទៅការវាយតំលៃជាមធ្យម ឬដោយចន្លោះជឿជាក់គឺធ្វើឡើងតាមរយៈ “Explore analysis” tool ដែលមានលក្ខណៈវិភាគលំអិតជាង Descriptives analysis tool ។

Explore analysis គឺជាការសិក្សាវិភាគវាយតំលៃដូច Descriptive analysis ដែរតែ Explore មានលក្ខណៈលំអិតជាងដោយអាចបញ្ចេញនូវតំលៃ Descriptive ជាច្រើនដូចជា៖ Mean, Std error, confidence interval, range, inter-quartile range,... ម្យ៉ាងវិញទៀត Explore អាចសិក្សាវិភាគវាយតំលៃទៅលើ Quantitative variable ដោយបំបែកការសិក្សានេះតាម Levels (Categories) ដែលមាន នៅក្នុង qualitative variable មួយផ្សេងទៀត ។

**ឧទាហរណ៍៖**

ដោយអនុវត្តទៅលើតារាងទិន្នន័យ “sales” នៃ SPSS, ចូរធ្វើ Explore analysis ទៅលើ Variable “Revenue” ដែលទទួលបានពីការអង្កេតទៅលើគ្រួសារមួយចំនួនដែលរស់នៅក្នុងតំបន់មួយកន្លែង ។

Analyze → Descriptive Statistic → Explore



### Descriptives

|         |                                  |             | Statistic  | Std. Error |
|---------|----------------------------------|-------------|------------|------------|
| Revenue | Mean                             |             | \$2,557.53 | \$25.020   |
|         | 95% Confidence Interval for Mean | Lower Bound | \$2,508.45 |            |
|         |                                  | Upper Bound | \$2,606.61 |            |
|         | 5% Trimmed Mean                  |             | \$2,542.49 |            |
|         | Median                           |             | \$2,508.07 |            |
|         | Variance                         |             | 914575.3   |            |
|         | Std. Deviation                   |             | \$956.334  |            |
|         | Minimum                          |             | \$500      |            |
|         | Maximum                          |             | \$6,213    |            |
|         | Range                            |             | \$5,713    |            |
|         | Interquartile Range              |             | \$1,327.62 |            |
|         |                                  |             |            |            |
|         |                                  |             |            |            |
|         |                                  |             |            |            |

### Interpretation:

\* **Mean**=\$2557.53 មានន័យថាគ្រួសារនីមួយៗដែលរស់នៅក្នុងតំបន់នេះមាន Revenue

ជាមធ្យម \$2557.53

$$Std.error = \frac{Std.deviation}{\sqrt{Sample}} = \$25.020$$

មានន័យថា ERROR ដែលកើតឡើងក្នុងការវាយតម្លៃទៅលើ Revenue ជាមធ្យមរបស់គ្រួសារនីមួយៗមានជាមធ្យម \$25.020 ។

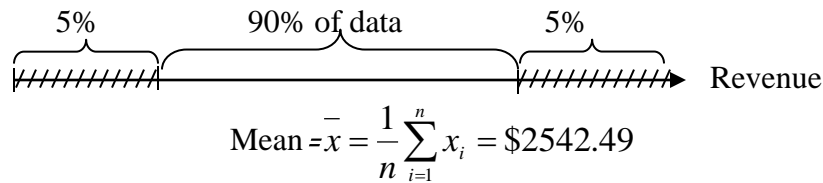
\* **95% Confidence interval**

**Lower bound** : Mean-1.96\*Std.deviation/ $\sqrt{sample\ size}$  =\$2508.45

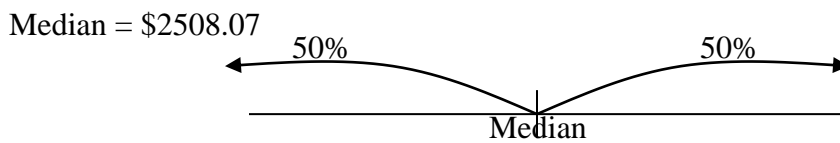
**Upper Bound** : Mean+1.96\*Std.deviation/ $\sqrt{sample\ size}$  =\$2606.61

មានន័យថា 95% នៃគ្រួសារដែលរស់នៅក្នុងតំបន់នេះមាន Revenue ជាមធ្យមប្រចាំឆ្នាំប្រែប្រួល ពី \$2508.45 ទៅ \$2606.61.15 ។

\* 5% trimmed mean

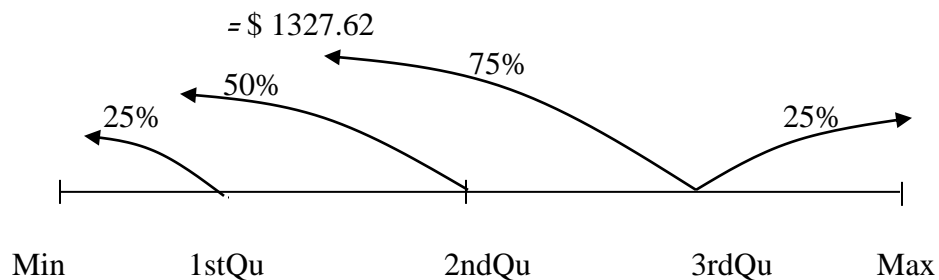


គេអាចប្រើតម្លៃ 5% trimmed mean ជំនួសតម្លៃ Mean ក្នុងការវាយតម្លៃជាមធ្យមទៅលើ Revenue របស់គ្រួសារនីមួយៗ ក្នុងករណីដែលគេសង្កេតឃើញថា នៅក្នុងទិន្នន័យ សំបូរទៅដោយ តម្លៃ “Outliers” ។



មាន 50% នៃគ្រួសារដែលយើងធ្វើការអង្កេត មាន Revenue ទាបជាង \$2508.07 និង 50% ទៀត ខ្ពស់ជាង ។

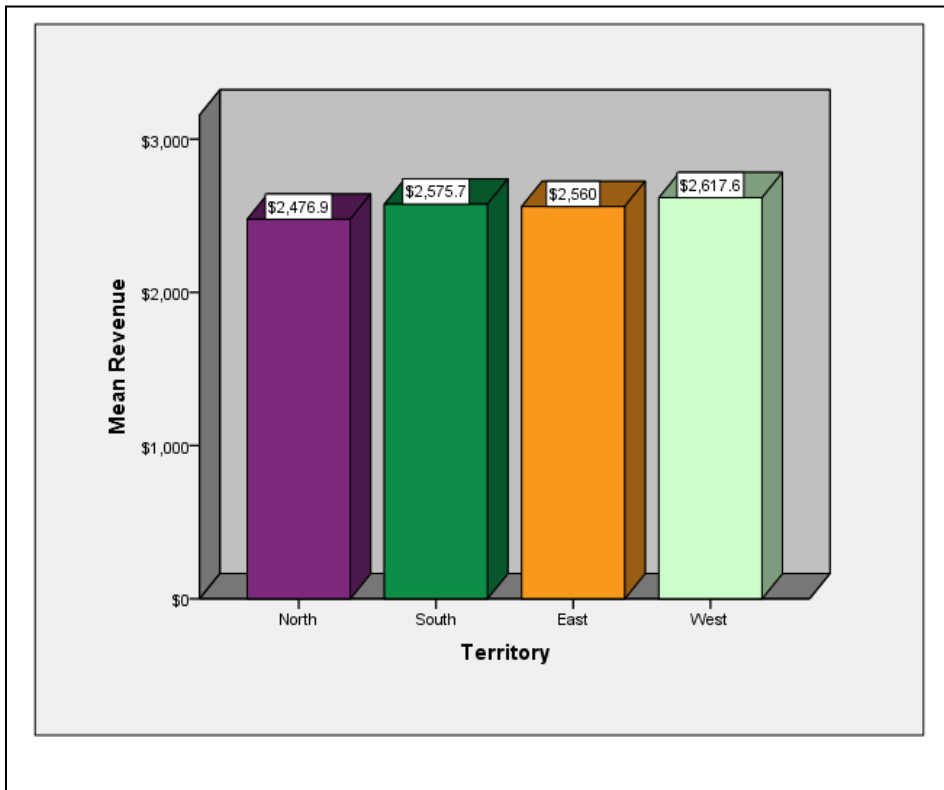
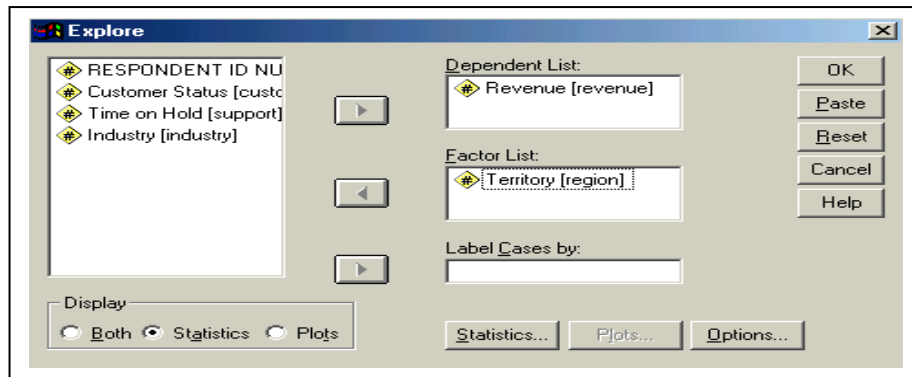
- Range = Maximum – Minimum = \$5713  
តំលាតរវាងគ្រួសារដែល Revenue ខ្ពស់ជាងគេនិងទាបជាងគេនៅក្នុងការអង្កេតនេះគឺ \$5713
- Inter-quartile range= 3rd Qu - 1st Qu , **Qu.=Quartile=1/4**



### ឧទាហរណ៍:

ដោយអនុវត្តទៅលើឧទាហរណ៍ខាងលើ ចូរធ្វើ Explore Analysis ទៅលើ Variable Revenue ដដែលដោយបំបែកការសិក្សានេះ ទៅតាមតំបន់នីមួយៗ (North, South, East & West ) នៃ Variable “Region”?



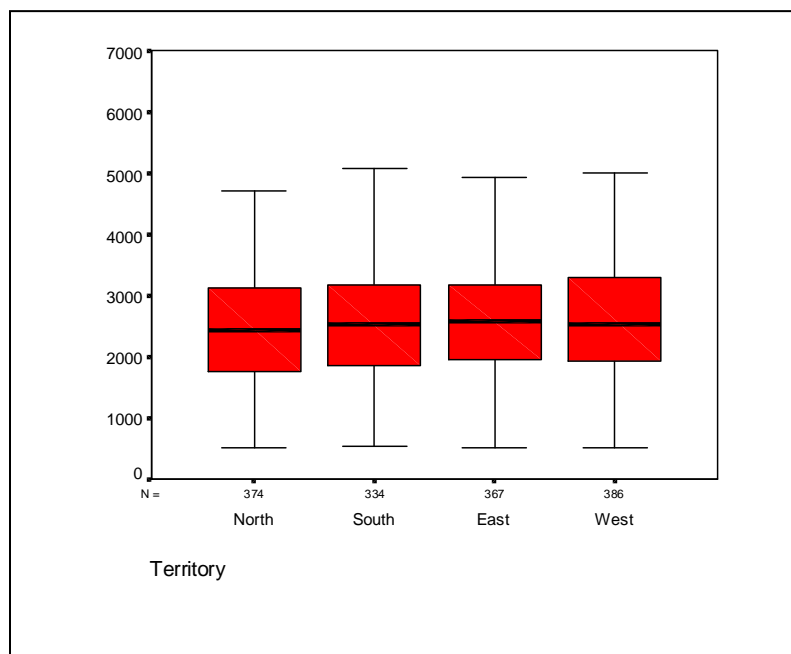


រូបភាព Barchart ដែលពន្យល់ការប្រៀបធៀបតំលៃមធ្យមទៅតាមតំបន់នីមួយៗ។

| Descriptives |       |                                  |             |            |            |
|--------------|-------|----------------------------------|-------------|------------|------------|
| Territory    |       |                                  |             | Statistic  | Std. Error |
| Revenue      | North | Mean                             |             | \$2,476.90 | \$49.806   |
|              |       | 95% Confidence Interval for Mean | Lower Bound | \$2,378.97 |            |
|              |       |                                  | Upper Bound | \$2,574.84 |            |
|              |       | 5% Trimmed Mean                  |             | \$2,460.48 |            |
|              |       | Median                           |             | \$2,416.78 |            |
|              |       | Variance                         |             | 927743.8   |            |
|              |       | Std. Deviation                   |             | \$963.195  |            |
|              |       | Minimum                          |             | \$509      |            |
|              |       | Maximum                          |             | \$6,024    |            |
|              |       | Range                            |             | \$5,515    |            |
|              |       | Interquartile Range              |             | \$1,358.61 |            |
|              | South | Mean                             |             | \$2,575.66 | \$53.701   |
|              |       | 95% Confidence Interval for Mean | Lower Bound | \$2,470.02 |            |
|              |       |                                  | Upper Bound | \$2,681.30 |            |
|              |       | 5% Trimmed Mean                  |             | \$2,548.34 |            |
|              |       | Median                           |             | \$2,536.03 |            |
|              |       | Variance                         |             | 963197.0   |            |
|              |       | Std. Deviation                   |             | \$981.426  |            |
|              |       | Minimum                          |             | \$545      |            |
|              |       | Maximum                          |             | \$6,213    |            |
|              |       | Range                            |             | \$5,669    |            |
|              |       | Interquartile Range              |             | \$1,317.22 |            |
|              | East  | Mean                             |             | \$2,560.01 | \$48.694   |
|              |       | 95% Confidence Interval for Mean | Lower Bound | \$2,464.26 |            |
|              |       |                                  | Upper Bound | \$2,655.77 |            |
|              |       | 5% Trimmed Mean                  |             | \$2,550.21 |            |
|              |       | Median                           |             | \$2,565.10 |            |
|              |       | Variance                         |             | 870207.4   |            |
|              |       | Std. Deviation                   |             | \$932.849  |            |
|              |       | Minimum                          |             | \$500      |            |
|              |       | Maximum                          |             | \$5,417    |            |
|              |       | Range                            |             | \$4,917    |            |
|              |       | Interquartile Range              |             | \$1,247.68 |            |
|              | West  | Mean                             |             | \$2,617.59 | \$48.256   |
|              |       | 95% Confidence Interval for Mean | Lower Bound | \$2,522.71 |            |
|              |       |                                  | Upper Bound | \$2,712.47 |            |
|              |       | 5% Trimmed Mean                  |             | \$2,610.33 |            |
|              |       | Median                           |             | \$2,532.37 |            |
|              |       | Variance                         |             | 898844.3   |            |
|              |       | Std. Deviation                   |             | \$948.074  |            |
|              |       | Minimum                          |             | \$500      |            |
|              |       | Maximum                          |             | \$5,012    |            |
|              |       | Range                            |             | \$4,511    |            |
|              |       | Interquartile Range              |             | \$1,384.27 |            |

លទ្ធផលខាងលើនេះត្រូវបកស្រាយពន្យល់ដូចឧទាហរណ៍ខាងលើដែរក៏ប៉ុន្តែវិភាគទៅតាមតំបន់ នីមួយៗ ។

Graphic ដែលងាយពន្យល់ទៅលើការប្រៀបធៀបលទ្ធផលនេះគឺប្រភេទ boxplot:



### ៣.១.២- ការកំណត់ Confidence interval for mean នៅក្នុង S-plus

នៅក្នុង S-plus ការកំណត់ Confidence interval សំរាប់ការវាយតម្លៃជាមធ្យមធ្វើឡើងតាមរយៈ

Command ខាងក្រោម៖

**t.test(variable, conf.level=(1- $\alpha$ ))\$conf.int**

ដែល៖

**variable** ជា Quantitative

**(1- $\alpha$ )** ជា confidence level ដែលត្រូវផ្តល់តម្លៃ

**\$conf.int** ប្រើសំរាប់កំណត់ Lower Bound និង Upper Bound នៃ Confidence interval

ដោយអនុវត្តទៅលើឧទាហរណ៍ខាងលើ យើងបង្កើត function S-plus សំរាប់ការគណនាតម្លៃ Quartiles និង Confidence interval សំរាប់ការវាយតម្លៃទៅលើ variable “Revenue” នៃតារាងទិន្នន័យ “sales” ។

---

```

function(sales)
{
  PROG<-function(sales,Region)
  {
    Revenue<- sales[,3][!is.na(sales[,3])& sales[,5]==Region]
    Report<-summary(Revenue)
    Confint<-t.test(Revenue,conf.level=0.95)$conf.int
    return(Report,Confint)
  }
  North<-PROG(sales,"North")
  South<-PROG(sales,"South")
  East<-PROG(sales,"East")
  West<-PROG(sales,"West")
  return(North, South, East, West)
}

```

**Output:****\$North:****\$North\$Report:**

| Min.  | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|------|---------|------|
| 273.6 | 1752    | 2409   | 2449 | 3101    | 6024 |

\$North\$Confint:

[1] 2349.392 2548.718

attr(\$North\$Confint, "conf.level"):

[1] 0.95

**\$South:****\$South\$Report:**

| Min.  | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|------|---------|------|
| 34.05 | 1789    | 2497   | 2521 | 3159    | 6213 |

\$South\$Confint:

[1] 2410.814 2630.615

attr(\$South\$Confint, "conf.level"):

[1] 0.95

**\$East:****\$East\$Report:**

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|------|---------|------|
| 12.6 | 1905    | 2542   | 2529 | 3148    | 5417 |

\$East\$Confint:

[1] 2430.479 2627.169

attr(\$East\$Confint, "conf.level"):

[1] 0.95

\$West:

\$West\$Report:

Min. 1st Qu. Median Mean 3rd Qu. Max.  
127.6 1886 2505 2566 3276 5012

\$West\$Confint:

[1] 2467.734 2664.776

attr(\$West\$Confint, "conf.level"):

[1] 0.95

## ៣.២ -ការវាយតម្លៃជាភាគរយដោយ Confidence interval

យើងដឹងថា  $\hat{P}$  ជា Unbiased estimator of  $P$  មានន័យថា  $\hat{P}$  វាយតម្លៃអោយ  $P$  មាន

លក្ខណៈប្រសើរ ។ ក្នុងករណីនៃការវាយតម្លៃ ដោយ Confidence interval ទៅលើ  $P$  យើងត្រូវកំណត់រក

Lower bound & Upper bound ឆ្លើយឆ្លងណាអោយ :

$$P(\text{Lower bound} \leq P \leq \text{Upper bound}) = (1 - \alpha)$$

$$P(\hat{p} - E \leq P \leq \hat{p} + E) = (1 - \alpha)$$

ដែល  $(1 - \alpha)$  ជា Confidence Level ។

យើងដឹងថា ៖ Margin of Error,  $E = Z_{\alpha/2} * \text{Std.Error}$

ដោយ  $\hat{P}$  វាយតម្លៃអោយ  $P$  មានការកើតឡើងនូវ Std.error  $S(\hat{P})$  ដូច្នេះ

$$\text{Lower bound} = \hat{p} - z_{\alpha/2} * s(\hat{p})$$

$$\text{Upper bound} = \hat{p} + z_{\alpha/2} * s(\hat{p})$$

កំណត់រក  $S(\hat{P})$  ?

ដោយការសិក្សាទៅលើការវាយតម្លៃជាភាគរយមានទំនាក់ទំនងនឹង របាយទិន្នន័យ Binomial ដូច្នេះ

យើងអាចប្រើគោលការណ៍នៃ Binomial ក្នុងការគណនា  $S(\hat{P})$  ។

បើ  $X$  ជាអថេរ Binomial, នោះ

$$V(X) = n * p * (1 - p)$$

$$\frac{V(X)}{n} = p * (1 - p)$$

$$V\left(\frac{X}{n}\right) = \frac{p * (1 - p)}{n}$$

ដូច្នេះ 
$$\hat{P} = \frac{X}{n}$$

$$V(\hat{P}) = \frac{\hat{p}^*(1-\hat{p})}{n}$$

$$\sigma(\hat{P}) = \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}} \approx S(\hat{P}) = \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}}$$

$\hat{P}$  ដោយ វាយតម្លៃ  $P$  មានលំអៀងជាមធ្យម (Std. Error) កំណត់៖

$$S(\hat{P}) = \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}} \Rightarrow E = Z_{\alpha/2} \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}}$$

ដូច្នេះ Confidence interval សំរាប់ការវាយតម្លៃទៅលើ  $P$  កំណត់ ៖

$$\text{Lower\_bound} = \hat{P} - Z_{\alpha/2} \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}}$$

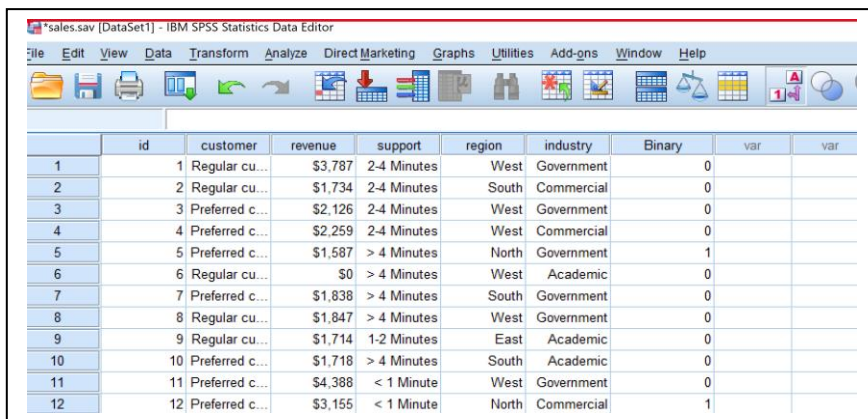
$$\text{Upper\_bound} = \hat{P} + Z_{\alpha/2} \sqrt{\frac{\hat{p}^*(1-\hat{p})}{n}}$$

### ឧទាហរណ៍៖

ការសិក្សាមួយបានធ្វើឡើងនៅក្នុងទីក្រុងមួយ ក្នុងគោលបំណងវាយតម្លៃទៅលើភាគរយនៃបុគ្គលដែលកំពុងបំពេញការងារនៅក្នុងទីក្រុងនេះដែលមកពីភាគខាងជើងនៃប្រទេសតាងដោយប៉ារ៉ាម៉ែត្រ  $P_{\text{North}}$  ។ ក្នុងករណីនេះ គេបានធ្វើការអង្កេតទៅលើបុគ្គលដែលកំពុងបំពេញការងារនៅក្នុងទីក្រុងនេះចំនួន  $n=1500$  នាក់ដោយប្រព័ន្ធចែជន្យហើយទិន្នន័យដែលទទួលបានកំណត់នៅក្នុង SPSS មានឈ្មោះថា “sales”។ ចូរធ្វើការវាយតម្លៃទៅលើប៉ារ៉ាម៉ែត្រ  $P_{\text{North}}$  ព្រមទាំងគណនា Confidence interval នៃការវាយតម្លៃនេះ?

- ជំហានទី១៖

Recode variable “Region” ទៅជា Binary code 0 ឬ 1 ដោយរក្សាលេខមួយសំរាប់តំបន់ North និង លេខ 0 សំរាប់តំបន់ផ្សេងៗ ។

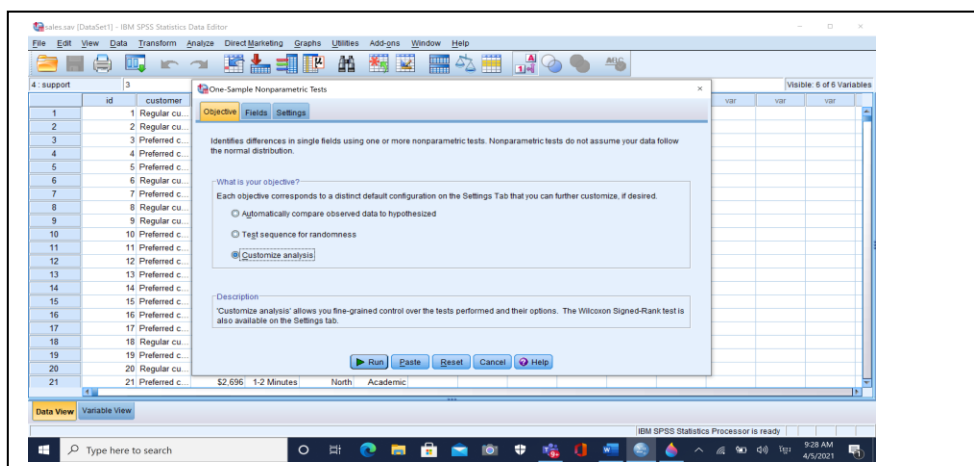


|    | id | customer       | revenue | support     | region | industry   | Binary | var | var |
|----|----|----------------|---------|-------------|--------|------------|--------|-----|-----|
| 1  | 1  | Regular cu...  | \$3,787 | 2-4 Minutes | West   | Government | 0      |     |     |
| 2  | 2  | Regular cu...  | \$1,734 | 2-4 Minutes | South  | Commercial | 0      |     |     |
| 3  | 3  | Preferred c... | \$2,126 | 2-4 Minutes | West   | Government | 0      |     |     |
| 4  | 4  | Preferred c... | \$2,259 | 2-4 Minutes | West   | Commercial | 0      |     |     |
| 5  | 5  | Preferred c... | \$1,587 | > 4 Minutes | North  | Government | 1      |     |     |
| 6  | 6  | Regular cu...  | \$0     | > 4 Minutes | West   | Academic   | 0      |     |     |
| 7  | 7  | Preferred c... | \$1,838 | > 4 Minutes | South  | Government | 0      |     |     |
| 8  | 8  | Regular cu...  | \$1,847 | > 4 Minutes | West   | Government | 0      |     |     |
| 9  | 9  | Regular cu...  | \$1,714 | 1-2 Minutes | East   | Academic   | 0      |     |     |
| 10 | 10 | Preferred c... | \$1,718 | > 4 Minutes | South  | Academic   | 0      |     |     |
| 11 | 11 | Preferred c... | \$4,388 | < 1 Minute  | West   | Government | 0      |     |     |
| 12 | 12 | Preferred c... | \$3,155 | < 1 Minute  | North  | Commercial | 1      |     |     |

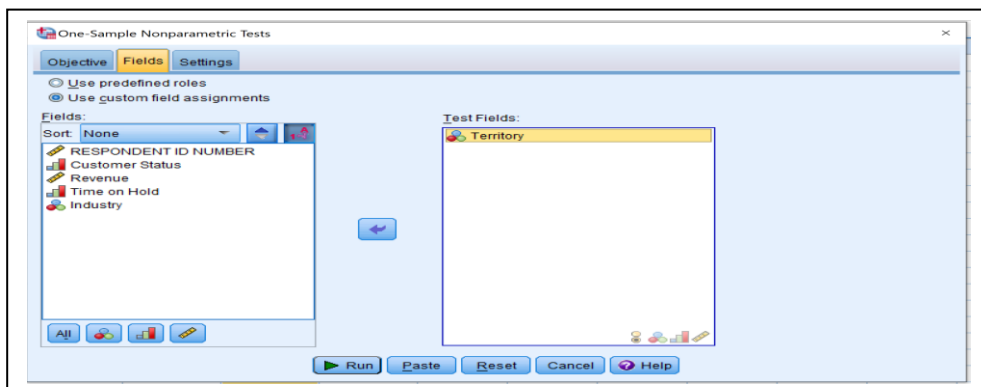
- ជំហានទី២ ៖

ប្រើ SPSS Analysis tool ទៅលើ អថេរ Binary

Analyze -----> Non-parametric test -----> One sample

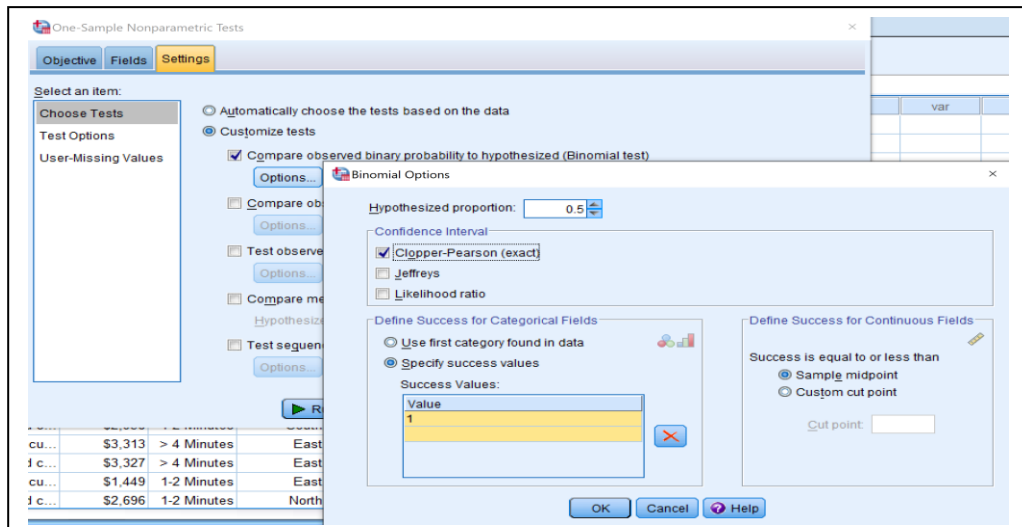


ជ្រើសរើសយក **Button Fields** ហើយរក្សាអថេរ **Region** នៅក្នុង **Test fields**



រៀបរៀងដោយ មាក កាម៉ាន (Statistician, Mathematician).

ជ្រើសរើសយក **Button Settings** ហើយអនុវត្តដូចរូបភាពខាងក្រោម៖



Click ទៅលើ **button OK** ហើយ **click Run** យើងទទួលបាន output ដូចខាងក្រោម៖

| Confidence Interval Summary                        |                                 |          |                         |       |
|----------------------------------------------------|---------------------------------|----------|-------------------------|-------|
| Confidence Interval Type                           | Parameter                       | Estimate | 95% Confidence Interval |       |
|                                                    |                                 |          | Lower                   | Upper |
| One-Sample Binomial Success Rate (Clopper-Pearson) | Probability (Territory=North ). | .255     | .233                    | .278  |

### ការបកស្រាយ៖

ឆ្លងកាត់ការសិក្សានេះ យើងអាចធ្វើសេចក្តីសន្និដ្ឋានថា នៅក្នុងទីក្រុងនេះមានបុគ្គលដែលមកពីភាគខាងជើងនៃប្រទេសចំនួន **25.50%** ដែលកំពុងបំពេញការងារនៅក្នុងទីក្រុងនេះ ហើយភាគរយនេះអាចនឹងមានការប្រែប្រួលក្នុងគំលាតពី **23.30%** ទៅ **27.80%** នៅក្នុងកំរិតជឿជាក់ **95%** ។

**S-plus:**

`prop.test(x,n, conf.level=(1- $\alpha$ ))$conf.int`

ដែល៖

x ជាចំនួនព្រឹត្តិការណ៍ជោគជ័យក្នុងការអង្កេត n ដង

n ជាចំនួននៃការអង្កេត



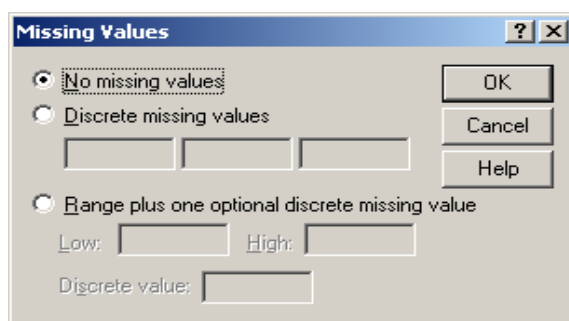
### ៣.៣ - ការប្រើប្រព័ន្ធ Missing នៅក្នុងការវាយតម្លៃ

នៅពេលខ្លះការវាយតម្លៃបង្កឡើងដោយភាពស្មុគស្មាញដែលទាមទារអោយយើងយល់ដឹងពីការប្រើប្រាស់ប្រព័ន្ធ Missing ដើម្បីគ្រប់គ្រងទិន្នន័យអោយចូលរួម ឬមិនអោយចូលរួមនៅក្នុងការវិភាគ ។

នៅក្នុង SPSS Variable នីមួយៗមានប្រព័ន្ធ Missing រៀងៗខ្លួន ។

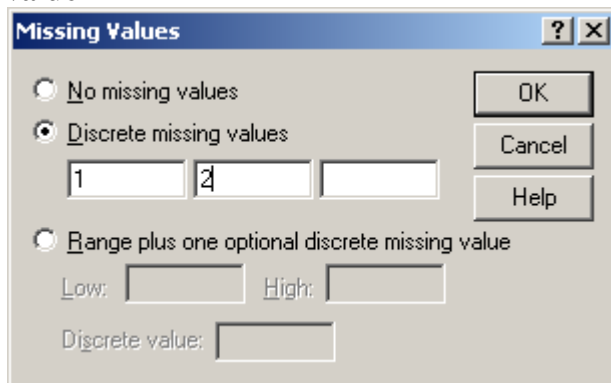
Missing System {  
 No missing value  
 Discrete missing value  
 Rang plus one optional discrete missing value

#### a. No missing value:



កាលណាយើងជ្រើសយក “No missing value” នៅក្នុងប្រអប់ Missing នៃ Variable ណា មួយ នោះមានន័យថាវាទទួលបានទិន្នន័យទាំងអស់ដែលមាននៅក្នុង Variable នេះចូលរួមទាំងស្រុងទៅក្នុងការ វិភាគ ឬ មានន័យថាប្រព័ន្ធ Missing មិនដំណើរការ (Default system) ។

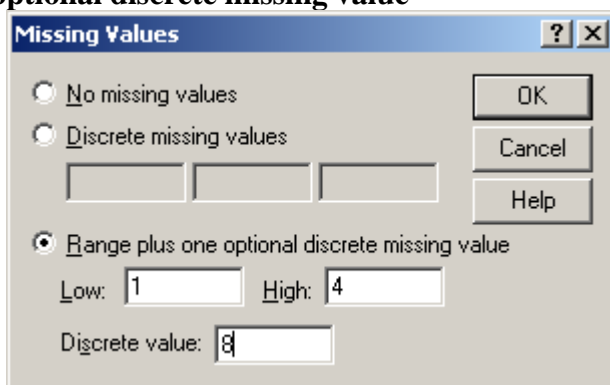
#### b- Discrete missing value



កាលណាយើងជ្រើសរើសយកទីតាំងទី២ នៃប្រព័ន្ធ Missing មានន័យថាយើងអាចបិទទិន្នន័យ (code) បាន 3 តំលៃដាច់ (Discrete) យ៉ាងច្រើនមិនអោយចូលរួមក្នុងការវិភាគ ។

ប្រអប់ខាងលើបានបង្ហាញអោយឃើញថា code លេខ 1 និងលេខ 2 មិនចូលរួមទៅក្នុងការវិភាគ នៃទិន្នន័យទេ ។

#### c-Ranges plus one optional discrete missing value



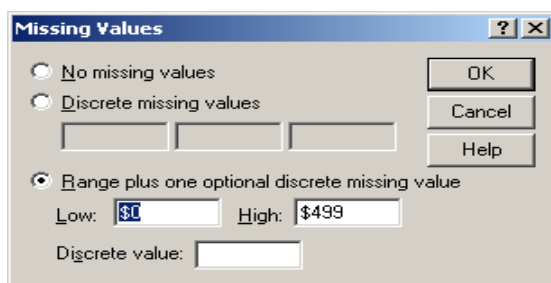
លេខ 1, 2, 3, 4 និង 8 មិនចូលរួមទៅក្នុងការវិភាគនៃទិន្នន័យ ។

នៅត្រង់ទីតាំងទី 3 នៃប្រអប់ Missing ផ្តល់លទ្ធភាពអោយយើងអាចបិទទិន្នន័យ (Code) មួយចន្លោះរៀងគ្នា មិនអោយចូលរួមទៅក្នុងការវិភាគ ហើយព្រមទាំងអាចបិទបាន 1 តំលៃដាច់ (Code) មួយផ្សេងទៀត ។

#### ឧទាហរណ៍:

ដោយអនុវត្តទៅលើតារាងទិន្នន័យនៃ SPSS ចូរធ្វើ Descriptive analysis ទៅលើ Variable “Revenue” ដោយចាត់ទុកថា រាល់តំលៃទាំងឡាយណាដែលតូចជាង \$500 តំលៃជាតំលៃ “Outliers” ដែលត្រូវដកចេញពីការវិភាគនេះ ។

#### ជំហានទី១ : ប្រើ Missing system នៃ variable “Revenue”



## ជំហានទី២: Descriptive analysis

### Syntax

**DESCRIPTIVES**

**/Variable=Revenue**

**/Statistics=Min Max Mean SEmean.**

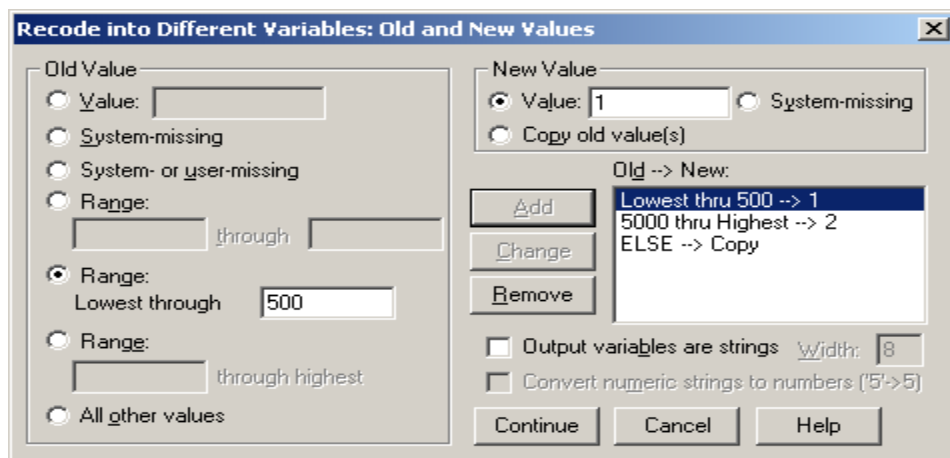
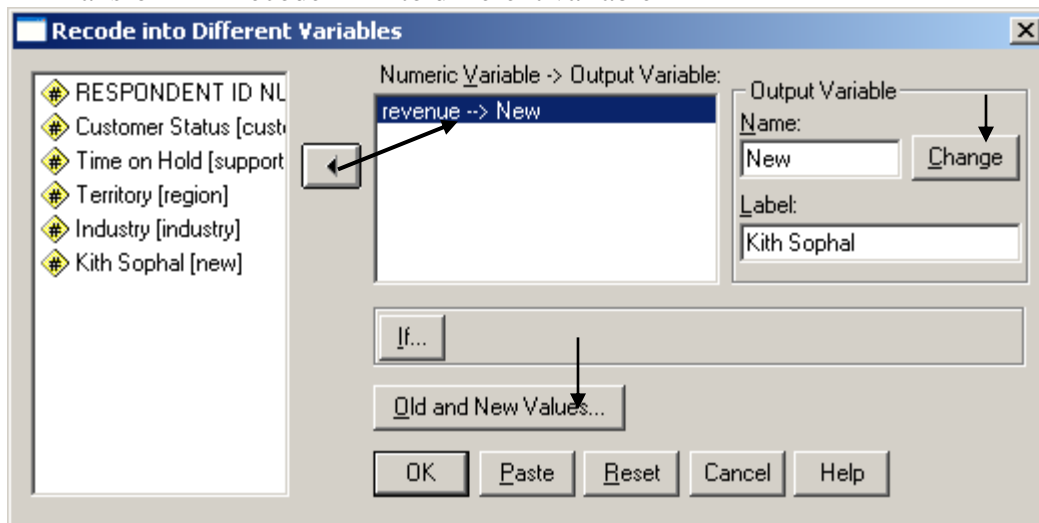
## ឧទាហរណ៍៖

ដោយអនុវត្តទៅលើតារាងទិន្នន័យនៃ SPSS ចូរធ្វើ Description analysis ទៅលើ Variable “Revenue” ដោយចាត់ទុកថារាល់តំលៃទាំងឡាយណាដែលតូចជាង \$500 ធំជាង \$5000ជាប្រភេទ

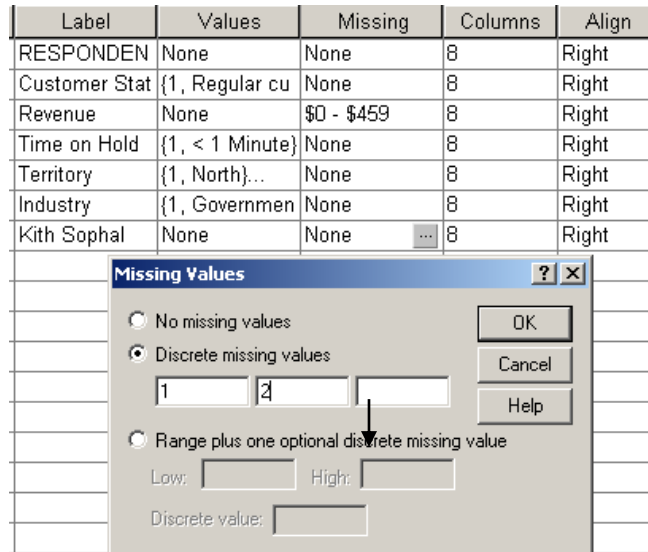
“Outliers” ដែលត្រូវដកចេញពីការវិភាគនេះ?

## ជំហានទី១: ដោយប្រើ Recode ទៅលើ Variable Revenue:

Transform → Recode → Into different variable



**ជំហានទី៣:** ប្រើប្រព័ន្ធ Missing នៃ Variable “New”



**ជំហានទី ៤:** Descriptive analysis

DESCRIPTIVES  
/Variable=New  
/Statistics=Min Max Mean Stddev SEmean.

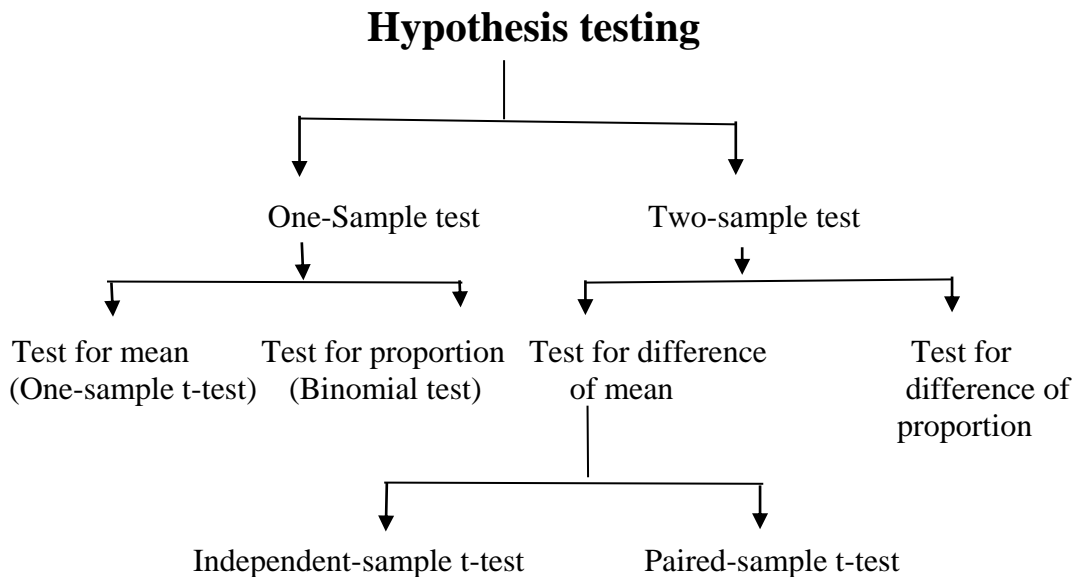
**Descriptive Statistics**

|                    | N         | Minimum   | Maximum   | Mean      |            |
|--------------------|-----------|-----------|-----------|-----------|------------|
|                    | Statistic | Statistic | Statistic | Statistic | Std. Error |
| Kith Sophal        | 1452      | 500.11    | 4946.12   | 2539.3164 | 24.4138    |
| Valid N (listwise) | 1452      |           |           |           |            |

នេះជាលទ្ធផលចុងក្រោយដែលយើងចង់បាន ។

\*\*\*\*\*

## ជំពូក៤: ការផ្ទៀងផ្ទាត់សម្មតិកម្ម (Hypothesis testing)



### I. One-sample test:

#### ១-ការបង្កើតសម្មតិកម្មនៅក្នុងសំណាកតែមួយ:

នៅក្នុងការផ្ទៀងផ្ទាត់សម្មតិកម្មក្នុងការវិភាគស្ថិតិ គេបង្កើតសម្មតិកម្មពីរ  $H_0$  &  $H_a$  ដែល:

$H_0$  អោយឈ្មោះថាសម្មតិកម្មសូន្យ (Null hypothesis)

$H_a$  អោយឈ្មោះថាសម្មតិកម្មប្រឆាំងឬផ្ទុយ (Alternative hypothesis)

សម្មតិកម្ម  $H_0$  &  $H_a$  ជាសម្មតិកម្មពីរដែលប្រឆាំងគ្នា ឬផ្ទុយគ្នា ។

នៅលទ្ធផលនៃការផ្ទៀងផ្ទាត់សម្មតិកម្ម ជាទូទៅ បើយើងទទួលយកសម្មតិកម្មសូន្យ  $H_0$  នោះយើងត្រូវបដិសេធជាដាច់ខាតនូវសម្មតិកម្មប្រឆាំង  $H_a$  ហើយបើយើងទទួលយកសម្មតិកម្មប្រឆាំង  $H_a$  នោះយើងបដិសេធជាដាច់ខាតនូវសម្មតិកម្មសូន្យ  $H_0$  ។

**ចំណាំ:** សម្មតិកម្មទាំងឡាយណាដែលលើកឡើងសំរាប់ការផ្ទៀងផ្ទាត់មានផ្ទុកសញ្ញា  $=, \geq$  or  $\leq$ , នោះសម្មតិកម្ម ទាំងនេះមានតួនាទីជាសម្មតិកម្មសូន្យ  $H_0$  ។ សម្មតិកម្មទាំងឡាយណាដែលមានផ្ទុកសញ្ញា  $\neq, >$  or  $<$ , នោះវាដើរតួនាទីជាសម្មតិកម្មប្រឆាំង  $H_a$  ។

សម្មតិកម្មដែលលើកឡើងសំរាប់ការផ្ទៀងផ្ទាត់អាចដើរតួនាទីជាសម្មតិកម្មសូន្យ ឬ សម្មតិកម្មប្រឆាំង។

## ២-ទំរង់ទូទៅសម្មតិកម្ម

ជាទូទៅសម្មតិកម្មសំរាប់ការផ្ទៀងផ្ទាត់ក្នុងការវិភាគស្ថិតិ ចែកចេញជាពីរទំរង់:

**ក- Two-tailed:** (ការផ្ទៀងផ្ទាត់តាមទិសដៅពីរ)

គឺជាការផ្ទៀងផ្ទាត់សម្មតិកម្មដោយគ្រាន់តែចង់ដឹងថាប៉ារ៉ាម៉ែត្រ  $\theta$  នៃលក្ខណៈសំគាល់ណាមួយ នៅក្នុង population មានតំលៃស្មើ ឬខុសពីតំលៃដែលយើងលើកឡើងក្នុងសម្មតិកម្ម  $\theta_0$  ( test proportion or test value) ។

**ឧទាហរណ៍:**

$$H_0 : \theta = \theta_0$$

VS

$$H_a : \theta \neq \theta_0$$

**ខ- One-tailed:** (ការផ្ទៀងផ្ទាត់តាមទិសដៅតែមួយ)

គឺជាការផ្ទៀងផ្ទាត់សម្មតិកម្មដោយគ្រាន់តែចង់ដឹងថា ប៉ារ៉ាម៉ែត្រ  $\theta$  នៃលក្ខណៈសំគាល់ណាមួយ នៃ population មានតំលៃធំជាង ឬតូចជាង តំលៃដែលយើងលើកឡើងក្នុងសម្មតិកម្ម  $\theta_0$  ( test proportion or test value) ដែលក្នុងនោះមាន **Left-tailed (less)** ឬ **Right-tailed(greater)**

**Left-tailed**

**Right-tailed**

$$H_0 : \theta = \theta_0 (>)$$

$$H_0 : \theta = \theta_0 (<)$$

VS

VS

$$H_a : \theta < \theta_0$$

$$H_a : \theta > \theta_0$$

## ៣. តំលៃ t-statistics ៖

ដើម្បីធ្វើការសម្រេចចិត្តនៅក្នុងការផ្ទៀងផ្ទាត់សម្មតិកម្ម ថា  $H_0$  ឬ  $H_a$  ពិត ឬ មិនពិត, គេត្រូវការ កំណត់ជាចាំបាច់នូវតំលៃ t-statistics ។ តំលៃ t-statistics ជាមេគុណណ័រម៉ាល់ស្តង់ដាដែលបង្ហាញអំពីការប្រៀបធៀបតំលៃដែលរកឃើញក្នុងសំណាក ទៅនឹងតំលៃដែលលើកឡើង ក្នុងសម្មតិកម្មធៀបនឹងលំអៀងស្តង់ដាក្រោមលក្ខខណ្ឌ  $H_0$ , កំណត់៖  $t = \frac{\hat{\theta} - \theta_0}{S_{H_0}(\hat{\theta})}$ ,

ដែល  $\theta_0$  ជា test value ឬ test proportion

$S_{H_0}(\hat{\theta})$  ជា លំអៀងស្តង់ដាក្រោមលក្ខខណ្ឌ  $H_0$ ;  $\hat{\theta}$  ជា unbiased estimator;

#### ៤. តារាងនៃការសម្រេចចិត្ត៖

| Hypothesis testing & Decision                                                              |                                     |                                     |                                                                                                          |
|--------------------------------------------------------------------------------------------|-------------------------------------|-------------------------------------|----------------------------------------------------------------------------------------------------------|
| Formulation of Hypotheses                                                                  | Rule                                |                                     | Decision                                                                                                 |
| <b>2-tailed test</b><br>$H_0: \theta = \theta_0$<br>vs<br>$H_a: \theta \neq \theta_0$      | Rejection Region<br>$\alpha / 2$    | Acceptance Region<br>$(1 - \alpha)$ | If $-Z_c \leq t \leq Z_c$ then we do not reject $H_0$<br>if $t < -Z_c$ or $t > Z_c$ then we reject $H_0$ |
| <b>Left-tailed test</b><br>$H_0: \theta = \theta_0 (>)$<br>vs<br>$H_a: \theta < \theta_0$  | Rejection Region<br>$\alpha$        | Acceptance Region<br>$(1 - \alpha)$ | If $t \geq -Z_c$ then we do not reject $H_0$<br>If $t < -Z_c$ then we reject $H_0$                       |
| <b>Right-tailed test</b><br>$H_0: \theta = \theta_0 (<)$<br>vs<br>$H_a: \theta > \theta_0$ | Acceptance Region<br>$(1 - \alpha)$ | Rejection Region<br>$\alpha$        | If $t \leq Z_c$ then we do not reject $H_0$<br>If $t > Z_c$ then we reject $H_0$                         |

#### ចំណាំ៖

- Rejection region មានន័យស្មើ Critical region
- $Z_c$  ជា critical value ដែលខណ្ឌចែក Rejection region ពី Acceptance region
- នៅក្នុង 2-tailed test  $Z_c$  មានតំលៃស្មើ  $Z_{\alpha/2}$  ប្រែប្រួលតំលៃទៅតាមកំរិតជឿជាក់
- នៅក្នុង 1-tailed test  $Z_c$  មានតំលៃស្មើ  $Z_{\alpha}$  ប្រែប្រួលតំលៃទៅតាមកំរិតជឿជាក់

#### វិធីសាស្ត្រសំរាប់កំណត់តំលៃ $Z_c$ ៖

បើតំលៃ level of confidence  $(1 - \alpha) = 95\%$  ឬ កំរិតលំអៀងនៃ population (level of significance)  $\alpha = 5\%$ , នោះ  $Z_c$  ត្រូវបានគណនាតាមរយៈ SPSS Syntax ដូចខាងក្រោម៖

- Two-tailed test:

$$Z_c = \text{IDF.NORMAL}(0.975, 0, 1) = 1.96 \quad (Z_{\alpha/2})$$

- One-tailed:

$$Z_c = \text{IDF.NORMAL}(0.95, 0, 1) = 1.64 \quad (Z_{\alpha})$$

## ៥ - One sample t-test (Test for mean):

នៅក្នុងករណីនេះទិន្នន័យដែលស្រង់នៅក្នុងសំណាក ឬ គំរូតាងស្ថិតិត្រូវតែជាប្រភេទ Quantitative ហើយប៉ារ៉ាម៉ែត្រនៃ population តំណាងអោយតំលៃមធ្យមដែលសម្មតិកម្មអាច មាន 3 ទំរង់ដូចខាងក្រោម៖

| ទំរង់នៃសម្មតិកម្ម                                 |                                                    |                                                    |
|---------------------------------------------------|----------------------------------------------------|----------------------------------------------------|
| 2-tailed test                                     | Left-tailed test                                   | Right-tailed test                                  |
| $H_0: \mu = \mu_0$<br>vs<br>$H_a: \mu \neq \mu_0$ | $H_0: \mu = \mu_0 (>)$<br>vs<br>$H_a: \mu < \mu_0$ | $H_0: \mu = \mu_0 (<)$<br>vs<br>$H_a: \mu > \mu_0$ |

$\mu_0$  ជា test value

តំលៃ t-statistics:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, S_{H_0}(\bar{X}) = \frac{S}{\sqrt{n}}$$

ដែល:  $\bar{X}$  ជា sample mean

$\mu_0$  ជា test value

$S$  ជា std. deviation

$n$  ជាទំហំនៃការអង្កេត

### ៥.១ - ឧទាហរណ៍អនុវត្តនៅក្នុង SPSS:

ដោយអនុវត្តទៅលើតារាងទិន្នន័យ “sales” នៃ SPSS ដែលទទួលបានពីការអង្កេតទៅលើ Revenue ប្រចាំឆ្នាំរបស់គ្រួសារមួយចំនួននៅក្នុងតំបន់ដីធ្លីមួយ ។ ចូរធ្វើការផ្ទៀងផ្ទាត់សម្មតិកម្មដែលលើកឡើងថា គ្រួសារនីមួយៗ ដែលរស់ នៅក្នុងតំបន់មាន Revenue ជាមធ្យមប្រចាំឆ្នាំយ៉ាងតិច \$2000 ។ កំណត់យក Significance level  $\alpha = 5\%$  ។

ទំរង់នៃ test of hypothesis:

$H_0: \mu = 2000$ (>) , Claim$   
 vs

$H_a: \mu < 2000$$

$\mu$  ជាប្រាក់ចំណូលជាមធ្យមប្រចាំឆ្នាំរបស់គ្រួសារនីមួយៗ



### វិភាគ:

#### Syntax:

t-test

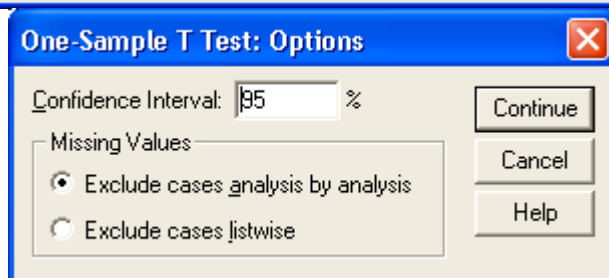
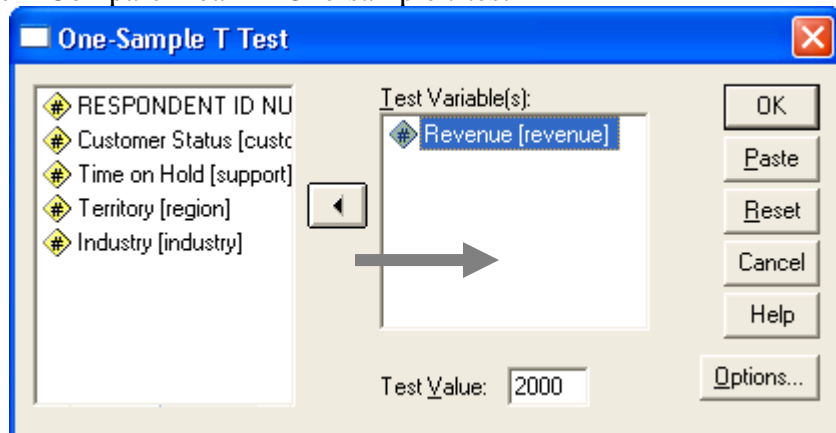
/Testval=2000

/Variable=Revenue

/Criteria=Cin(0.95).

#### Application tool:

Analyze → Compare mean → One-sample t-test

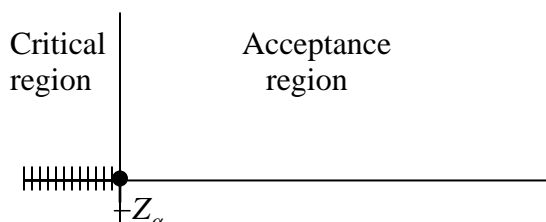


One-Sample Test

| Test Value = 2000 |        |      |                 |                 |                                           |          |
|-------------------|--------|------|-----------------|-----------------|-------------------------------------------|----------|
|                   | t      | df   | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference |          |
|                   |        |      |                 |                 | Lower                                     | Upper    |
| Revenue           | 20.035 | 1487 | .000            | \$516.58        | \$466.00                                  | \$567.15 |

#### **Interpretation:**

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = 20.035$$



រៀបរៀងដោយ ម៉ាក កាមេរ៉ាន (Statistician, Mathematician).

$$\alpha = 5\% \Rightarrow Z_\alpha = 1.64$$

ដើម្បីកំណត់រកតម្លៃ critical value យើងអាចប្រើ Syntax ខាងក្រោម៖

COMPUTE Z=IDF.NORMAL(0.95,0,1).  
EXECUTE.

$$t = 20.035 > -Z_\alpha = -1.64 \Rightarrow \text{Accept } H_0$$

មានន័យថា 95% នៃគ្រួសារដែលរស់នៅក្នុងតំបន់នេះមាន Revenue ជាមធ្យមប្រចាំឆ្នាំយ៉ាងតិច \$2000

ដូច្នេះសម្មតិកម្មដែលលើកឡើងជាសម្មតិកម្មដែលត្រឹមត្រូវ ។

\*df=degree of freedom

=sample size-1

=1487

\* sig(2-tailed) =significance probability of 2-tailed test  
= 0.000

### ចំណាំ៖

**sig.(2-tailed)** តម្លៃនេះអាចប្រើជំនួសតម្លៃ t-statistic ក្នុងការសំរេចចិត្តនៃ 2-tailed test ។ ម្យ៉ាងវិញ

ទៀត តម្លៃ sig(2-tailed) ប្រើសំរាប់ដោះស្រាយបញ្ហាក្នុងករណីដែលយើង Accept “ $\geq$ ” ឬ “ $\leq$ ” ។

### Rule:

**បើ**  $\text{sig}(2\text{-tailed}) \geq \text{significance level } \alpha \Rightarrow \text{Accept “=”}$

**បើ**  $\text{sig}(2\text{-tailed}) < \text{significance level } \alpha \Rightarrow \text{Reject “=”}$

ដោយ  $\text{sig}(2\text{-tailed}) = 0.000 < \alpha = 0.05 \Rightarrow \text{Reject “=”}$

ដូច្នេះគេអាចនិយាយយ៉ាងច្បាស់ថា 95% នៃគ្រួសារដែលរស់នៅក្នុងតំបន់នេះមាន Revenue ជាមធ្យម

ខ្ពស់ជាង \$2000/Year ។

\*Mean Difference =Mean-test value  
= \$516.18

គេអាចធ្វើការសន្និដ្ឋានថា គ្រួសារនីមួយៗដែលរស់នៅក្នុងតំបន់នេះមាន Revenue/Year ជាមធ្យម

ខ្ពស់ ជាង \$2000 ក្នុងគំណត់ជាមធ្យម \$516.18 ។

\*95% Confidence Interval of the Difference  $(\bar{X} - \mu_0)$

$$\text{Lower bound: } (\bar{X} - \mu_0) - Z_{\alpha/2} \frac{S}{\sqrt{n}} = \$466.00$$

$$\text{Upper bound: } (\bar{X} - \mu_0) + Z_{\alpha/2} \frac{S}{\sqrt{n}} = \$567.15$$

95% នៃគ្រួសារដែលរស់នៅក្នុងតំបន់សិក្សានេះមាន Revenue/Year ជាមធ្យមខ្ពស់ជាង \$2000 ក្នុងគំណត់ជាមធ្យមប្រែប្រួលពី \$466 ទៅ \$567.15 ។

### ការបកស្រាយ:

តាមរយៈនៃការជ្រាវស្រាវទៅលើប្រាក់ចំណូលជាមធ្យមប្រចាំឆ្នាំរបស់គ្រួសារនីមួយៗ ដែលរស់នៅក្នុងតំបន់នេះបានបង្ហាញអោយឃើញថាគ្រួសារនីមួយៗមានប្រាក់ចំណូលជាមធ្យមប្រចាំឆ្នាំខ្ពស់ជាង \$2000 ក្នុងគំណត់ជាមធ្យម \$516.58 ។ យើងអាចនិយាយបានទៀតថា 95% នៃគ្រួសារដែលរស់ក្នុងតំបន់នេះមានប្រាក់ចំណូលជាមធ្យមប្រចាំឆ្នាំខ្ពស់ ជាង \$2000 ក្នុងគំណត់ជាមធ្យមប្រែប្រួលពី \$466 ទៅ \$567.15 ។

### ៥.២ - ការអនុវត្ត t-test ក្នុង S-plus:

សំរាប់ការអនុវត្ត t-test ជាមួយ S-plus យើងមាន Command ដូចខាងក្រោម:

`t.test(data , alternative= " two.sided ", conf.level=(1-  $\alpha$  ),mu= $\mu_0$ )`

ឬ `less`

ឬ `greater`

ដោយអនុវត្តជាមួយឧទាហរណ៍ខាងលើយើងបាន:

```
function(sales)
{
  Revenue<-sales[,3][!is.na(sales[,3])]
  Report<-t.test(Revenue,alternative="less",conf.level=0.95,mu=2000)
  return(Report)
}
```

**Output:**

#### One-sample t-Test

*data: Revenue*

*t = 20.0353, df = 1487, p-value = 1*

*alternative hypothesis: true mean is less than 2000*

*95 percent confidence interval:*

*NA 2559.015*

*sample estimates:*

*mean of x*

*2516.57*

### ឧទាហរណ៍៖

ការសិក្សាមួយធ្វើឡើងក្នុងគោលបំណងផ្ទៀងផ្ទាត់ទៅលើប្រសិទ្ធភាពនៃម៉ាស៊ីនស្វ័យប្រវត្តិ ដែលបានដាក់អោយដំណើរការនៅក្នុងថ្នាក់ផលិតកម្មនៃការច្រកនំស្រួយ (Cereal) ក្នុងប្រអប់ ដោយកំណត់កម្មវិធីអោយច្រកចូលប្រអប់ក្នុងទំងន់ជាមធ្យម 1000 g ។

ដើម្បីផ្ទៀងផ្ទាត់ប្រសិទ្ធភាពនៃម៉ាស៊ីននេះ គេបាននាំស្រីសយក Cereal ប្រអប់ដោយ ចៃដន្យចំនួន 30 ប្រអប់ពីក្នុងចំណោមផលិតផលដែលផលិតដោយម៉ាស៊ីននេះ រួចថ្លឹងទម្ងន់នំ នៅក្នុងប្រអប់នីមួយៗ។ ទិន្នន័យដែលទទួលបានមានដូចខាងក្រោម៖

990 980 1000 950 970 1050 1000 990 1100 970 985 1000 1010 1020 975  
985 970 1010 995 1000 990 995 980 1020 970 960 1010 985 960 1020

យោងលើទិន្នន័យនេះតើម៉ាស៊ីនស្វ័យប្រវត្តិនេះមានប្រសិទ្ធភាពនៅក្នុងដំណើរការផលិតកម្ម នេះដែរឬទេ ក្នុងកំរិត level of significance 5%?

### ចម្លើយ៖

ទំរង់សម្មតិកម្មដែលសមស្រប៖

$$H_0: \mu = 1000g, \text{ Claim}$$

vs

$$H_a: \mu \neq 1000g$$

$\mu$  ជាទម្ងន់ជាមធ្យមនៃ Cereal ប្រអប់

វិភាគ៖ ដោយវិភាគតាមរយៈ SPSS គេទទួលបានលទ្ធផលដូចខាងក្រោម៖

| One-Sample Test      |                   |    |                 |                 |                                           |        |
|----------------------|-------------------|----|-----------------|-----------------|-------------------------------------------|--------|
|                      | Test Value = 1000 |    |                 |                 |                                           |        |
|                      | t                 | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference |        |
|                      |                   |    |                 |                 | Lower                                     | Upper  |
| Weight of Cereal box | -.997             | 29 | .327            | -5.33333        | -16.2742                                  | 5.6076 |

ដោយការផ្ទៀងផ្ទាត់សម្មតិកម្មរបស់យើងជាប្រភេទ 2-tailed test នោះយើងអាចប្រើតំលៃ Sig.(2-tailed) ជំនួសតំលៃ t-statistics សំរាប់ការសំរេចចិត្ត។

ដោយ Sig.(2-tailed)=0.327 >  $\alpha=0.05$  នោះយើងទទួលយកសម្មតិកម្មសូន្យ មានន័យថា ម៉ាស៊ីនស្វ័យប្រវត្តិនេះមានដំណើរការប្រក្រតីក្នុងដំណើរការនៃចង្វាក់ផលិតកម្មនេះ ។

**ចំណាំ:** ក្នុងករណីដែលយើងទទួលយកសញ្ញា “=” នោះតំលៃផ្សេងៗនៅក្នុងតារាង  
លទ្ធផលគ្មានន័យក្នុងការបកស្រាយទេ។

**S-plus:**

ដោយសន្មតថា នៅក្នុង DataFrame នៃ Object explorer នៃ S-plus មានតារាងទិន្នន័យ  
“Weight” ដែលជាទំងន់ Cereal ទាំង 30 ប្រអប់ ដូច្នេះគេអាចសរសេរ S-plus syntax  
ដូចខាងក្រោម៖

```
> X<-Weight[,1]
> t.test(X, alternative="two.sided", conf.level=0.95, mu=1000)
```

| One-sample t-Test                                   |                  |
|-----------------------------------------------------|------------------|
| data: X                                             |                  |
| <b>t = -0.997</b> , df = 29, p-value = <b>0.327</b> |                  |
| alternative hypothesis: mean is not equal to 1000   |                  |
| 95 percent confidence interval:                     |                  |
| <b>983.7258</b>                                     | <b>1005.6076</b> |
| sample estimates:                                   |                  |
| mean of x                                           |                  |
| <b>994.6667</b>                                     |                  |

## ៦. One-sample test for proportion (Binomial test):

គឺជាការសិក្សាផ្ទៀងផ្ទាត់ទៅលើសម្មតិកម្ម ដែលប៉ារ៉ាម៉ែត្រ  $p$  នៃ population តំណាង  
ដោយភាគរយ តាមរយៈតំលៃ unbiased estimator  $\hat{p}$  ទទួលបានពីទិន្នន័យបែប Qualitative  
ស្រង់ក្នុងសំណាក ឬ គំរូតាងស្ថិតិតែមួយ។

### ៦.១ ទំរង់សម្មតិកម្ម:

| ទំរង់នៃសម្មតិកម្ម                         |                                            |                                            |
|-------------------------------------------|--------------------------------------------|--------------------------------------------|
| 2-tailed test                             | Left-tailed test                           | Right-tailed test                          |
| $H_0: p = p_0$<br>vs<br>$H_a: p \neq p_0$ | $H_0: p = p_0 (>)$<br>vs<br>$H_a: p < p_0$ | $H_0: p = p_0 (<)$<br>vs<br>$H_a: p > p_0$ |

$p_0$  ជាតំលៃ test proportion

## ៦.២ តំលៃ t-statistics :

$$t = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \text{ ដែល } \hat{p} \text{ ជា unbiased estimator}$$

$p_0$  ជាតំលៃ test value

$n$  ជាទំហំនៃការអង្កេត

## ៦.៣ ឧទាហរណ៍៖

អ្នកនយោបាយម្នាក់មានការចាប់អារម្មណ៍ទៅលើអាយុនៃអ្នកដែលត្រូវចូលរួម នាពេលបោះឆ្នោតខាងមុខ ហើយបានលើកជាអំណះអំណាងមួយថា នឹងមានអ្នកបោះឆ្នោតដែលមានអាយុមិនលើស 35 ឆ្នាំដែលនឹងចូលរួមការបោះឆ្នោតនាពេលខាងមុខមានកំរិតទាបជាង 30%។ ដើម្បីផ្ទៀងផ្ទាត់ទៅនឹងអំណះអំណាងនេះ គេបានធ្វើការអង្កេតទៅលើអ្នកមានសិទ្ធិបោះឆ្នោតមួយចំនួននៅក្នុងតំបន់ហើយទិន្នន័យដែលស្រង់បានកំណត់នៅក្នុង SPSS (V.13.0) មានឈ្មោះថា "voter" ។ ចូរធ្វើការផ្ទៀងផ្ទាត់អំណះអំណាងនេះក្នុងកំរិត level of significance  $\alpha = 5\%$ ?

## ចម្លើយ៖

- លើកសម្មតិកម្មដែលសមស្រប៖

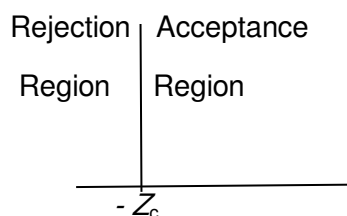
$$H_0: p \geq 30\%$$

V.S

$$H_a: p < 30\%$$

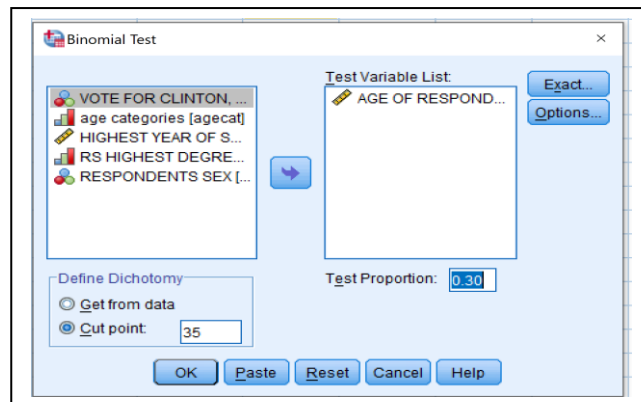
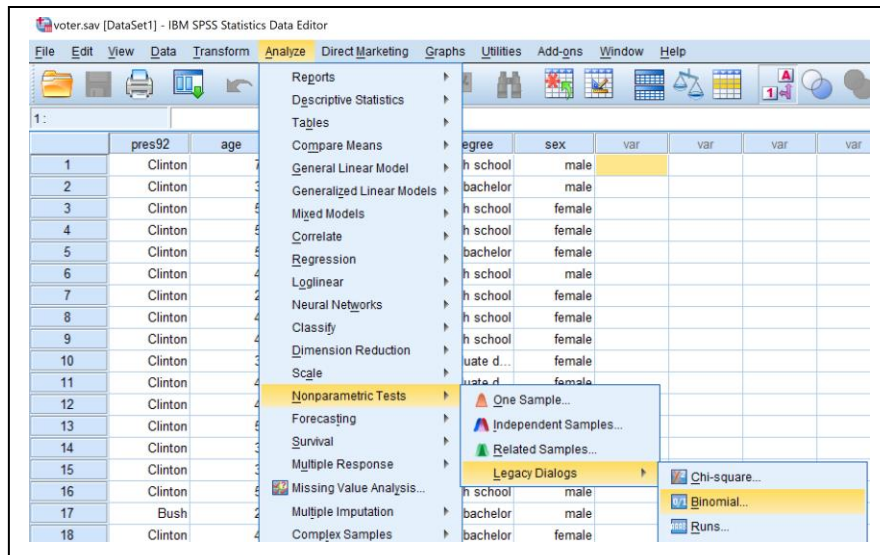
ដែល  $p$  ជាភាគរយអ្នកដែលមានអាយុមិនលើស 35 ឆ្នាំ

- ក្បួននៃការសម្រេចចិត្ត៖



- ការវិភាគតាមរយៈ SPSS :

Analyze -----> Non-parametric test -----> Legacy dialogs -----> Binomial



កំណត់យក "Cut point " ស្មើនឹង 35 ក្នុងគោលបំណងបែងចែកអាយុជាពីរក្រុម៖ ក្រុមអាយុ 35 ចុះ និង អាយុច្រើនជាង 35 ឆ្នាំ។ ចូរចំណាំថា Binomial test ជានិច្ចជាកាលធ្វើការផ្ទៀងផ្ទាត់តែ ចំពោះក្រុម ទី១ តែប៉ុណ្ណោះ។

ក្នុងករណីនេះ Output ដែលទទួលបានពីការវិភាគមានដូចខាងក្រោម៖

| Binomial Test     |               |      |                |            |                       |
|-------------------|---------------|------|----------------|------------|-----------------------|
|                   | Category      | N    | Observed Prop. | Test Prop. | Exact Sig. (1-tailed) |
| AGE OF RESPONDENT | Group 1 <= 35 | 478  | .3             | .3         | .000 <sup>a</sup>     |
|                   | Group 2 > 35  | 1369 | .7             |            |                       |
|                   | Total         | 1847 | 1.0            |            |                       |

a. Alternative hypothesis states that the proportion of cases in the first group < .3.

### ចំណាំ៖

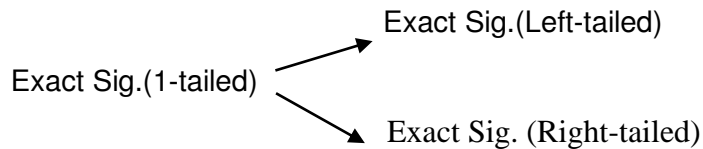
នៅក្នុងលទ្ធផលនៃ Binomial test គេប្រើតំលៃ **Exact Sig.** ដើម្បីធ្វើការផ្ទៀងផ្ទាត់សម្មតិកម្ម ។

Exact Sig. = Exact Significance Probability Value

**Rule:** If Exact Sig.  $\geq$  Level of significance  $\alpha$  then " Do not reject  $H_0$  "

If Exact Sig.  $<$  Level of significance  $\alpha$  then " Reject  $H_0$  "

Exact Sig. មាន 2 ករណី ខុសគ្នា៖ Exact Sig.(2-tailed) , Exact Sig.(1-tailed) , ដែល ៖



Exact Sig. នីមួយៗប្រើសំរាប់ការផ្ទៀងផ្ទាត់សម្មតិកម្មតាមទំរង់រៀងៗខ្លួនតាមគោលការណ៍ខាងលើ។

ចំពោះ Exact Sig.(Left-tailed) ឬ Exact Sig.(Right-tailed), គេអាចដឹងដោយចែកដាច់ពីគ្នាតាមរយៈ Message នៅក្រោមតារាង Output :

- បើមាន **Message** "a. Alternative hypothesis states that the proportion of the cases in the first group < test proportion " (ក្នុងករណីរបស់យើង test proportion=0.3) នោះវាជា Exact Sig.(Left-tailed) ។
- ក្នុងករណីផ្ទុយពីនេះ ជាប្រភេទ Exact Sig.(Right-tailed) ហើយដែល  $\text{Exact Sig. (Left-tailed)} + \text{Exact Sig. (Right-tailed)} = 1$  ។

ដោយតំលៃ Exact Sig.(1-tailed) ជាប្រភេទ Exact Sig.(Left-tailed) ហើយសម្មតិកម្មដែលយើងលើកឡើងជាប្រភេទ Left-tailed test នោះយើងអាចប្រើ "Rule of Exact Sig." បាន ។

### សន្និដ្ឋាន៖

ក្នុងករណីនៃឧទាហរណ៍ខាងលើ តំលៃ Exact Sig.(Left-tailed)=0.000 និងតំលៃ  $\alpha = 0.05$  នោះយើង Reject  $H_0$  មានន័យថា៖

ភាគរយអ្នកបោះឆ្នោតដែលមានអាយុ 35 ចុះដែលនឹងចូលរួមក្នុងការបោះឆ្នោតនាពេលខាងមុខមានកំរិតទាបជាង 30% ដូច្នេះអំណះអំណាងរបស់អ្នកនយោបាយនេះពិតជាត្រឹមត្រូវ ។



**S-plus:** ជាទូទៅនៅក្នុង S-plus គេមាន Syntax ដូចខាងក្រោម៖

`prop.test(x,n,alternative="....." , conf.level=(1- $\alpha$ ) , p=p0)`

ដែល x ជាចំនួនករណីជោគជ័យក្នុងការអង្កេត

n ជាទំហំនៃការអង្កេត

Alternative អាចទទួលយកតំលៃក្នុង៖ "two.sided" or "less" or "greater"

(1- $\alpha$ ) ជា confidence level ដែលត្រូវផ្តល់តំលៃ

p<sub>0</sub> ជាតំលៃដែលលើកឡើងក្នុងសម្មតិកម្ម

ក្នុងករណីនេះ យើងអាចសរសេរ ៖

`prop.test(478,1847,alternative= "less", conf.level=0.95 , p= 0.30)`

#### 1-sample proportions test with continuity correction

**data:** 478 out of 1847, null probability 0.3

*X-square* = 14.7352, *df* = 1, **p-value** = 0.0001

**alternative hypothesis:** *P(success) in Group 1 is less than 0.3*

**95 percent confidence interval:**

0.000000 0.276181

**sample estimates:**

*prop'n in Group 1*

0.2587981

## II- Two-sample test

### 9. Independent sample t-test

គេមាន population ពីរមិនអាស្រ័យគ្នា។ គេមានបំណងប្រៀបធៀបប៉ារ៉ាម៉ែត្រនៃ

population ទាំងពីរ ។

$\mu_1$  ជាតំលៃមធ្យមដែលតំណាងអោយប៉ារ៉ាម៉ែត្រនៃ population ទី១

$\mu_2$  ជាតំលៃមធ្យមដែលតំណាងអោយប៉ារ៉ាម៉ែត្រនៃ population ទី២

គេមានបំណងប្រៀបធៀប  $\mu_1$  និង  $\mu_2$  តាមរយៈនៃការប្រៀបធៀប  $\bar{X}_1$  &  $\bar{X}_2$

ដែលជាមធ្យមនៃសំណាក (sample mean) ដកស្រង់ចេញពី population ទាំងពីររៀងគ្នា។

ដូច្នេះ គេអាចសរសេរបានថា៖

$$(\bar{X}_1 - \bar{X}_2) \text{ វាយតម្លៃ } (\mu_1 - \mu_2) \text{ មានលំអៀងស្តង់ដា } S_{H_0}(\bar{X}_1 - \bar{X}_2)$$

ដូច្នេះតម្លៃ t-statistics សរសេរ៖

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{H_0}(\bar{X}_1 - \bar{X}_2)}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

ដែល៖

$\bar{X}_1$  ជា Sample mean នៅក្នុង Sample1

$\bar{X}_2$  ជា Sample mean នៅក្នុង Sample2

$S_1$  ជា Std.deviation នៃ data នៅក្នុង Sample1

$S_2$  ជា Std.deviation នៃ data នៅក្នុង Sample2

$n_1$  ជា Sample Size នៃ Sample1

$n_2$  ជា Sample Size នៃ Sample2

**ទម្រង់នៃ test of hypothesis:**

| ទម្រង់នៃសម្មតិកម្ម                                    |                                                        |                                                        |
|-------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------|
| 2-tailed test                                         | Left-tailed test                                       | Right-tailed test                                      |
| $H_0: \mu_1 = \mu_2$<br>vs<br>$H_a: \mu_1 \neq \mu_2$ | $H_0: \mu_1 = \mu_2 (>)$<br>vs<br>$H_a: \mu_1 < \mu_2$ | $H_0: \mu_1 = \mu_2 (<)$<br>vs<br>$H_a: \mu_1 > \mu_2$ |

Confidence interval for  $(\mu_1 - \mu_2)$  ?

$$\text{ដោយ } S(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

នោះ confidence interval for  $(\mu_1 - \mu_2)$  កំណត់៖

$$\text{Lower bound : } (\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$\text{Upper bound : } (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

### ១.១.-ឧទាហរណ៍អនុវត្តនៅក្នុង SPSS:

ការសិក្សាធ្វើឡើងដើម្បីផ្ទៀងផ្ទាត់ សម្មតិកម្ម ដែលលើកថានិស្សិតដែលបញ្ចប់ការសិក្សាផ្នែក Engineering ទទួលបាន Salary ជាមធ្យមប្រចាំឆ្នាំខ្ពស់ជាងនិស្សិតដែលបានបញ្ចប់ការសិក្សាផ្នែក Business Administration ។ ទិន្នន័យដែលស្រង់បានពីការអង្កេតលើបញ្ហានេះកំណត់ទុកនៅក្នុងតារាង ទិន្នន័យ “University of Florida graduate Salaries” នៃ SPSS (V.13.0) ។

ចូរធ្វើការផ្ទៀងផ្ទាត់ក្នុងកំរិត Significance level  $\alpha = 5\%$  ។

test of hypothesis

$$H_0 : \mu_E = \mu_{BA} (<)$$

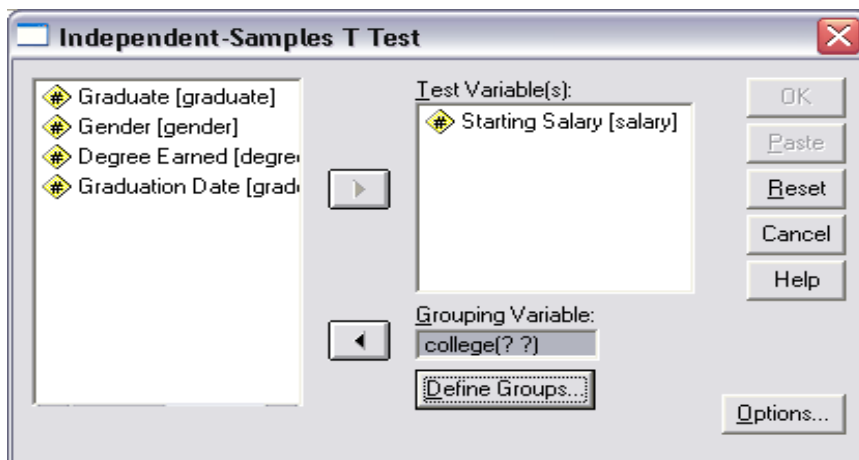
vs

$$H_a : \mu_E > \mu_{BA}$$

$\mu_E$  : Salary ជាមធ្យមប្រចាំឆ្នាំរបស់និស្សិតដែលបញ្ចប់ការសិក្សាផ្នែក Engineering

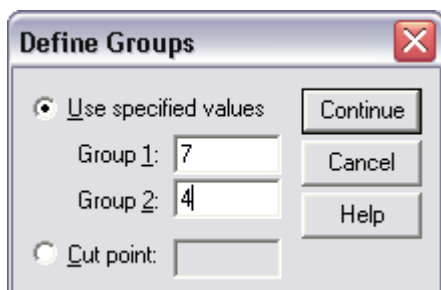
$\mu_{BA}$  : Salary ជាមធ្យមប្រចាំឆ្នាំរបស់និស្សិតដែលបញ្ចប់ការសិក្សាផ្នែក “Business Administration”

Analyze → Compare mean → Independent-Sample t-test



Click ទៅលើ button “ Define group” ដើម្បីកំណត់លេខ codeនៃ Category ដែលត្រូវប្រៀបធៀប៖

រៀបរៀងដោយ មាក កាមេរ៉ាន (Statistician, Mathematician).



The image shows the 'Define Groups' dialog box in SPSS. The 'Use specified values' radio button is selected. Below it, 'Group 1:' is set to 7 and 'Group 2:' is set to 4. The 'Continue', 'Cancel', and 'Help' buttons are visible on the right.

លេខ code 7 តាងអោយ Engineering

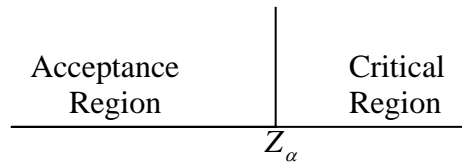
លេខ code 4 តាងអោយ Business Administration

| Independent Samples Test     |                                           |                         |
|------------------------------|-------------------------------------------|-------------------------|
|                              |                                           | Starting Salary         |
|                              |                                           | Equal variances assumed |
| t-test for Equality of Means | t                                         | 13.787                  |
|                              | df                                        | 601                     |
|                              | Sig. (2-tailed)                           | .000                    |
|                              | Mean Difference                           | 6062.82                 |
|                              | Std. Error Difference                     | 439.750                 |
|                              | 95% Confidence Interval of the Difference |                         |
|                              |                                           | Lower                   |
|                              |                                           | Upper                   |
|                              |                                           | 5199.181                |
|                              |                                           | 6926.450                |

| Group Statistics |                         |     |          |                |                 |
|------------------|-------------------------|-----|----------|----------------|-----------------|
| College          |                         | N   | Mean     | Std. Deviation | Std. Error Mean |
| Starting Salary  | Engineering             | 281 | 30876.87 | 5189.219       | 309.563         |
|                  | Business Administration | 322 | 24814.05 | 5553.360       | 309.477         |

Interpretation:

$$t = \frac{\bar{X}_E - \bar{X}_{BA}}{\sqrt{\frac{S_E^2}{n_E} + \frac{S_{BA}^2}{n_{BA}}}} = 13.787$$



$$\alpha = 5\% \Rightarrow Z_\alpha = 1.64$$

$$t = 13.787 > Z_\alpha = 1.64 \Rightarrow \text{Reject } H_0$$

មានន័យថា និស្សិតដែលបានបញ្ចប់ការសិក្សាផ្នែក Engineering ហើយដែលកំពុងបំពេញការងារនៅក្នុងសង្គមទទួលបាននូវ Salary ជាមធ្យមប្រចាំឆ្នាំខ្ពស់ជាងនិស្សិតផ្នែក Business Administration ។

ដូច្នេះ អំណះអំណាងដែលលើកឡើងជាអំណះអំណាងដែលត្រឹមត្រូវ ។

\*Mean difference= 6062.82 មានន័យថានិស្សិតផ្នែក Engineering ទទួលបាន salary

ខ្ពស់ជាងនិស្សិតផ្នែក Business Administration ក្នុងគំណាតជាមធ្យម 6062.82 ។

\*Std error difference=439.750 មានន័យថា error ដែលកើតឡើងនៅក្នុងការវាយ តម្លៃទៅលើ Mean difference មាន 439.750 ។

\* 95% confidence interval of the difference

$$\text{Lower bound : } (\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 5199.181\$$$

$$\text{Upper bound: } (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 6926.45\$$$

យើងអាចនិយាយបានថា 95% នៃនិស្សិតដែលបានបញ្ចប់ការសិក្សាផ្នែក Engineering ទទួលបាន salary ជាមធ្យមប្រចាំឆ្នាំខ្ពស់ជាងនិស្សិតដែលបានបញ្ចប់ការសិក្សាផ្នែក Business Administration ក្នុងគំណាតជាមធ្យមប្រែប្រួលពី 5199.181 ទៅ 6926.45 ។

## ៤.១.២- S-plus application

ដើម្បីធ្វើការប្រៀបធៀបតម្លៃមធ្យមតាមរយៈ Independent-sample t-test នៅក្នុង S-plus គេមាន

Command ដូចខាងក្រោម:

`t.test(data1,data2, alternative=".....",conf.level=(1- $\alpha$ ),var.equal=T)`

ដែល data1 ជា quantitative variable ស្រង់នៅក្នុង sample1

data2 ជា quantitative variable ស្រង់នៅក្នុង sample2

var.equal=T មានន័យថា "Assumed variance are equal"

var.equal=F មានន័យថា "Assumed variance are not equal"

សន្មតថា Data តំណាងអោយតារាងទិន្នន័យ University of Florida graduate salaries នៅក្នុង

S-plus) ។

```
function(Data)
{
  Sal.Eng<-Data[,4][!is.na(Data[,4]) & Data[,3]=="Engineering"]
  Sal.Bus<- Data[,4][!is.na(Data[,4]) & Data[,3]=="Business Administration"]
  Report<-t.test(Sal.Eng,Sal.Bus,alternative="greater",conf.level=0.95,var.equal=T)
  return(Report)
}
```

**Output:**

#### Pooled variance Two-Sample t-Test

data: Sal.Eng and Sal.Bus

t = 13.7869, df = 601, p-value = 0

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

5338.374 NA

sample estimates:

|           |           |
|-----------|-----------|
| mean of x | mean of y |
| 30876.87  | 24814.05  |

#### ឧទាហរណ៍៖

អំណះអំណាងមួយដែលបានលើកឡើងនៅក្នុងសង្គមការងារបានអះអាងថាអ្នកដែលបំពេញការងារនៅក្នុងស្ថាប័ន "Commercial" ទទួលបានប្រាក់ចំណូលប្រចាំខែជាមធ្យមខ្ពស់ជាងអ្នកដែលបំពេញការងារនៅក្នុងស្ថាប័ន "Academic" ។ ផ្ទៀងផ្ទាត់ទៅនឹងអំណះអំណាងនេះ គេបានសិក្សាទៅលើបុគ្គលមួយចំនួននៅក្នុងតំបន់នេះដោយចែងនូវហើយទិន្នន័យដែលស្រង់បាន កំណត់នៅក្នុង SPSS (V.13.0) មានឈ្មោះថា "sales" ។ ចូរផ្ទៀងផ្ទាត់អំណះអំណាងនេះក្នុងកំរិត level of significance  $\alpha = 5\%$  ?

ទំរង់សម្មតិកម្មសមស្រប៖

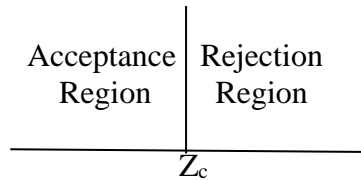
$$H_0: \mu_C = \mu_A (<)$$

vs

$$H_a: \mu_C > \mu_A$$

$\mu_C$  ជាប្រាក់ចំណូលជាមធ្យមនៃបុគ្គលដែលបំពេញការងារក្នុងស្ថាប័ន Commercial

$\mu_A$  ជាប្រាក់ចំណូលជាមធ្យមនៃបុគ្គលដែលបំពេញការងារក្នុងស្ថាប័ន Academic



**Output:**

| Independent Samples Test     |                                           |                         |
|------------------------------|-------------------------------------------|-------------------------|
|                              |                                           | Revenue                 |
|                              |                                           | Equal variances assumed |
| t-test for Equality of Means | t                                         | -1.010                  |
|                              | df                                        | 990                     |
|                              | Sig. (2-tailed)                           | .313                    |
|                              | Mean Difference                           | -\$64.429               |
|                              | Std. Error Difference                     | \$63.822                |
|                              | 95% Confidence Interval of the Difference | Lower -\$189.670        |
|                              |                                           | Upper \$60.812          |

$\alpha = 5\% \Rightarrow Z_c = 1.64$  , ដូច្នេះ ៖

$t = -1.010 < Z_c = 1.64$  នោះយើងទទួលយកសម្មតិកម្មសូន្យ ។ គេអាចសន្និដ្ឋានបានថា៖ អ្នកដែលបំពេញការងារនៅក្នុងស្ថាប័ន Commercial មិនទទួលបានប្រាក់ចំណូលជាមធ្យមប្រចាំខែខ្ពស់ជាងអ្នកដែលបំពេញការងារក្នុងស្ថាប័ន Academic នោះទេ។ ដូច្នេះអំណះអំណាង ដែលលើកឡើងពិតជាមិនត្រឹមត្រូវទេ។

ដោយតម្លៃ  $\text{sig. (2-tailed)} = 0.313 > \alpha = 0.05$  នោះ យើងទទួលយកសញ្ញា " = " មានន័យថា អ្នកដែលបំពេញការងារនៅក្នុងស្ថាប័នទាំងពីរមានកំរិតប្រាក់ចំណូលជាមធ្យមប្រចាំខែដូចគ្នា ។

S-plus:

```
function(sales)
{ Rev.C<- sales[,3][!is.na(sales[,3]) & sales[,6]==" Commercial" ]
  Rev.A<- sales[,3][!is.na(sales[,3]) & sales[,6]=="Academic" ]
  Report<- t.test(Rev.C, Rev.A, alternative="greater" ,
                  conf.level=0.95, var.equal=T)
  return(Report)
}
```

Output:

```
      Pooled-Variance Two-Sample t-Test
data: Rev.C and Rev.A
t = -1.0095, df = 990, p-value = 0.8435
alternative hypothesis: difference in means is greater than 0
95 percent confidence interval:
-169.5043      NA
sample estimates:
mean of x    mean of y
 2481.21    2545.639
```



## ៤.២- Dependent sample t-test(paired-sample t-test)

ក្នុងករណីនៃ Paired sample t-test ទិន្នន័យដែលជ្រើសរើសសិរាបការផ្ទៀងផ្ទាត់ស្ថិតក្នុង Sample ដែលអាស្រ័យគ្នា។ ខាងក្រោមនេះជាទិន្នន័យដែលស្រង់បានពីការអង្កេត:

| Obs | X <sub>1</sub>  | X <sub>2</sub>  | D=X <sub>2</sub> -X <sub>1</sub> | X <sub>1</sub> , X <sub>2</sub> : quantitative variable<br>D ជាគំលាតរវាង X <sub>2</sub> & X <sub>1</sub> |
|-----|-----------------|-----------------|----------------------------------|----------------------------------------------------------------------------------------------------------|
| 1   | X <sub>11</sub> | X <sub>21</sub> | d <sub>1</sub>                   |                                                                                                          |
| 2   | X <sub>12</sub> | X <sub>22</sub> | d <sub>2</sub>                   |                                                                                                          |
| 3   | X <sub>13</sub> | X <sub>23</sub> | d <sub>3</sub>                   |                                                                                                          |
| .   | .               | .               | .                                |                                                                                                          |
| .   | .               | .               | .                                |                                                                                                          |
| .   | .               | .               | .                                |                                                                                                          |
| n   | X <sub>1n</sub> | X <sub>2n</sub> | d <sub>n</sub>                   |                                                                                                          |

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n d_i \text{ mean of the difference}$$

កំណត់រកតម្លៃ t-statistic:

$$t = \frac{\bar{D}}{S(\bar{D})}, \quad \text{ដែល } S(\bar{D}) \text{ ជា std.error}$$

$$S(\bar{D}) = \frac{S(D)}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \cdot \sum_{i=1}^n (d_i - \bar{D})^2}$$

ដូច្នេះតម្លៃ t-statistic កំណត់ :

$$t = \frac{\bar{D}}{\sqrt{\frac{1}{n(n-1)} \cdot \sum_{i=1}^n (d_i - \bar{D})^2}}$$

### ទំរង់សម្មតិកម្ម៖

| ទំរង់សម្មតិកម្ម      |                      |                      |
|----------------------|----------------------|----------------------|
| 2-tailed test        | Left-tailed test     | Right-tailed test    |
| $H_0 : \mu_D = 0$    | $H_0 : \mu_D \geq 0$ | $H_0 : \mu_D \leq 0$ |
| vs                   | vs                   | vs                   |
| $H_1 : \mu_D \neq 0$ | $H_1 : \mu_D < 0$    | $H_1 : \mu_D > 0$    |

$\mu_D$  ជាមធ្យមនៃគំលាតនៃប្លុយឡាស្យុង ( Mean of the difference of the population)

### Confidence interval for $\mu_D$ (Paired difference)

ដោយ  $\bar{D}$  វ៉ាយតំលៃអោយ  $\mu_D$  មាន std error  $S(\bar{D})$  នោះ confidence interval for  $\mu_D$

កំណត់:

$$\text{Lower bound : } \bar{D} - Z_{\alpha/2} \cdot \sqrt{\frac{1}{n(n-1)} \cdot \sum_{i=1}^n (d_i - \bar{D})^2}$$

$$\text{Upper bound : } \bar{D} + Z_{\alpha/2} \cdot \sqrt{\frac{1}{n(n-1)} \cdot \sum_{i=1}^n (d_i - \bar{D})^2}$$

### ៤.២.១-លំហាត់អនុវត្តនៅក្នុង SPSS:

ខាងក្រោមនេះជាតារាងទិន្នន័យដែលជាពិន្ទុរបស់និស្សិតមួយក្រុមដែលស្ថិតនៅក្នុងកម្មវិធីសិក្សាតែមួយដូចគ្នា ហើយដែលត្រូវបានធ្វើ Test មុននិងក្រោយវគ្គ “training of developing knowledge” ទៅលើ C++ programming Language ។

|        |    |    |    |    |    |    |    |    |    |    |    |    |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|
| Before | 60 | 50 | 45 | 35 | 80 | 85 | 75 | 65 | 55 | 90 | 95 | 55 |
| After  | 65 | 90 | 50 | 50 | 80 | 90 | 85 | 80 | 85 | 90 | 95 | 75 |

តើវគ្គ “Training” នេះមានប្រសិទ្ធភាពដែរឬទេ ក្នុងកំរិត Significance Level  $\alpha = 5\%$  ?

Test of hypothesis:

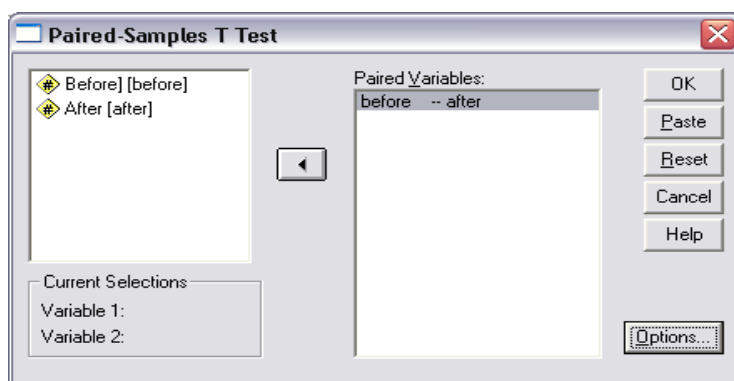
$$H_0 : \mu_D = 0 (>)$$

vs

$$H_1 : \mu_D < 0$$

$\mu_D$  គឺជាកំណាតជាមធ្យមនៃពិន្ទុរបស់និស្សិតមុននិងក្រោយវគ្គ “training”

Analyze → Compare mean → Paired-Sample t-test



| Paired Samples Test |                                           | Pair 1                                   |
|---------------------|-------------------------------------------|------------------------------------------|
|                     |                                           | Before the Training - After the training |
| Paired Differences  | Mean                                      | -12.083                                  |
|                     | Std. Deviation                            | 12.695                                   |
|                     | Std. Error Mean                           | 3.665                                    |
|                     | 95% Confidence Interval of the Difference |                                          |
|                     | Lower                                     | -20.150                                  |
|                     | Upper                                     | -4.017                                   |
| t                   |                                           | -3.297                                   |
| df                  |                                           | 11                                       |
| Sig. (2-tailed)     |                                           | .007                                     |

Interpretation:

$$t = \frac{\bar{D}}{\sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (d_i - \bar{D})^2}} = -3.29$$

Critical  
Region

Acceptance  
Region

$$-Z_{\alpha} = -1.64$$

$t = -3.29 < -1.64 \Rightarrow \text{Reject } H_0$  ក្នុងកំរិត 95% ពិតយើងអាចសន្និដ្ឋានថា វគ្គ Training មានប្រសិទ្ធភាពក្នុងការបង្កើន ចំណេះដឹងដល់និស្សិតទាំងនោះ ។

$$\text{Mean} = \frac{1}{n} \cdot \sum_{i=1}^n d_i = -12.08 \text{ ក្រោយពីឆ្លងកាត់វគ្គ "Training of Developing$$

Knowledge" in C++ Programming Language" និស្សិតនីមួយៗ មានកំនើនពិន្ទុជាមធ្យម 12.08 ។

$$\text{std.deviation} = \sqrt{\frac{1}{n-1} \cdot \sum_{u=1}^n (d_i - \bar{D})^2} = 12.695$$

$$\text{std.error.mean} = \sqrt{\frac{1}{n(n-1)} \cdot \sum_{u=1}^n (d_i - \bar{D})^2} = 3.664$$

ការវាយតម្លៃទៅលើ កំនើនពិន្ទុរបស់និស្សិតនីមួយៗមាន ERROR ជាមធ្យម 3.664

95% Confidence interval of the difference:

$$\text{Lower Bound : } \bar{D} - Z_{\alpha/2} \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (d_i - \bar{D})^2} = -20.149$$

$$\text{Upper Bound : } \bar{D} + Z_{\alpha/2} \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (d_i - \bar{D})^2} = -4.017$$

ក្រោយពីការឆ្លងកាត់វគ្គ “training of developing Knowledge in C++ Programming Language”, យើងសង្កេតឃើញថា 95% នៃនិស្សិតបានទទួលនូវកំនើនចំណេះដឹងគិតជាពិន្ទុជាមធ្យម ប្រែប្រួលពី 4.017 ទៅ 20.149 ។

### ៤.២.២-លំហាត់អនុវត្តនៅក្នុង S-plus:

យើងអាចធ្វើការវិភាគ Paired-sample t-test នៅក្នុង S-plus ដោយប្រើ Command ដូចខាង ក្រោម:

**t.test(data1,data2,alternative=".....",conf.level=(1- $\alpha$ ),paired=T)**

ដែល data1 តាងអោយទិន្នន័យដែលស្រង់មុនសកម្មភាពអ្វីមួយ

data2 តាងអោយទិន្នន័យដែលស្រង់ក្រោយសកម្មភាពអ្វីមួយ

paired=T បញ្ជាក់ថា data1 & data2 អាស្រ័យនឹងគ្នា

```
function( )
{
  data1<-c(60 , 50, 45, 35, 80, 85, 75, 65, 55, 90, 95, 55 )
  data2<-c(65 , 90, 50, 50, 80, 90, 85, 80, 85, 90, 95, 75)
  report<-t.test(data1,data2,alternative="less",conf.level=0.95,paired=T)
  return(report)
}
```

### Output

#### Paired t-Test

data: data1 and data2

t = -3.2971, df = 11, p-value = 0.0036

alternative hypothesis: true mean of differences is less than 0

95 percent confidence interval:

NA -5.50167

sample estimates:

mean of x - y

-12.08333

**ចំណាំ**

ក្នុងករណីដែលយើងមានបំណងកំណត់រក Confidence interval of the difference នោះយើងគ្រាន់

តែកំណត់យក alternative="two.sided" យើងនឹងទទួលបាននូវចំណេះដឹង ។

ឧទាហរណ៍៖

```
function( )
{
  data1<-c(60 , 50, 45, 35, 80, 85, 75, 65, 55, 90, 95, 55 )
  data2<-c(65 , 90, 50, 50, 80, 90, 85, 80, 85, 90, 95, 75)
  report<- t.test(data1,data2,alternative="two.sided",conf.level=0.95,paired=T)
  return(report)
}
```

**Output:**

Paired t-Test

data: data1 and data2

t = -3.2971, df = 11, p-value = 0.0071

alternative hypothesis: mean of differences is not equal to 0

95 percent confidence interval:

**-20.149632      -4.017034**

sample estimates:

mean of x - y

-12.08333

## ជំពូកទី៖ Chi-square

### ១- Chi-square as goodness-of-fit test

Chi-square (Goodness-of-Fit) ប្រើសំរាប់សិក្សាផ្ទៀងផ្ទាត់របាយនៃទិន្នន័យដែលទទួលបានពីការអង្កេតទៅនឹងរបាយទិន្នន័យណាមួយ ដោយធ្វើការប្រៀបធៀប observed frequency ទៅនឹង expected frequency នៃរបាយទិន្នន័យដែលយើងចង់បាន ។

#### ទំរង់នៃសម្មតិកម្ម (Hypothesis testing)

$H_0$  : របាយទិន្នន័យដែលទទួលបានពីការអង្កេតគោរពទៅតាមច្បាប់នៃរបាយទិន្នន័យណាមួយ  
Vs

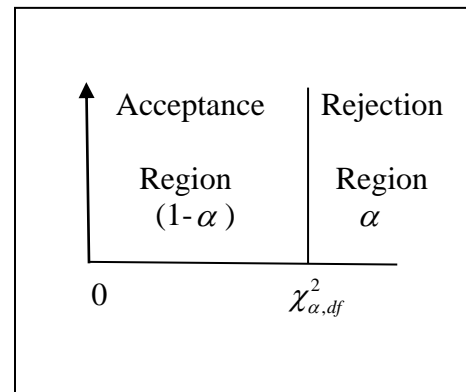
$H_1$  : របាយទិន្នន័យដែលទទួលបានពីការអង្កេត មិនគោរពទៅតាមច្បាប់នៃរបាយទិន្នន័យណាមួយ

#### តំលៃ Chi-square

$$\chi^2 = \sum_{i=1}^n \frac{(fo_i - fe_i)^2}{fe_i}$$

ដែល  $fo_i$  ជា observed frequency

$fe_i$  ជា expected frequency



#### ការសម្រេចចិត្ត

បើ  $\chi^2 \leq \chi^2_{\alpha, df} \Rightarrow$  Do not reject  $H_0$  (Accept  $H_0$ )

បើ  $\chi^2 > \chi^2_{\alpha, df} \Rightarrow$  Reject  $H_0$

#### ចំណាំ៖

តំលៃ  $\chi^2_{\alpha, df}$  អោយឈ្មោះថា critical value of Chi-square ដែល :

$\alpha$  ជា significance level

df ជា degree of freedom ដែល  $df = (\text{Number of group}) - 1$

តំលៃ  $\chi^2_{\alpha;df}$  អាចគណនាតាមរយៈ Syntax ដូចខាងក្រោម៖

**Compute Chi=IDF.CHISQ(1- $\alpha$  , df).**

**Execute.**

### ឧទាហរណ៍៖

គេមានបំណងសិក្សាទៅលើគ្រាប់ឡកឡាក់មួយដើម្បីចង់ដឹងថាទ្រង់ទ្រាយនៃគ្រាប់ឡកឡាក់នេះមានលក្ខណៈស្មើសាច់ល្អដែរឬទេ នៅលើផ្ទៃមុខខាង មានន័យថាលទ្ធផលនៃគ្រាប់ឡកឡាក់នេះចេញដោយគោរពទៅតាមច្បាប់នៃភាពចៃដន្យ (Random) ដែរឬទេក្នុងកំរិត  $\alpha = 5\%$  ?

ខាងក្រោមនេះជាលទ្ធផលនៃការសាកល្បងបោះគ្រាប់ឡកឡាក់ចំនួន 120 ដង៖

| លេខ | ចំនួន<br>(Observed<br>Frequency) |
|-----|----------------------------------|
| 1   | 18                               |
| 2   | 22                               |
| 3   | 15                               |
| 4   | 25                               |
| 5   | 20                               |
| 6   | 20                               |

### ទំរង់សម្មតិកម្ម៖

$H_0$  : លទ្ធផលនៃការបោះគ្រាប់ឡកឡាក់គោរពទៅតាមច្បាប់នៃភាពចៃដន្យ

$V_s$

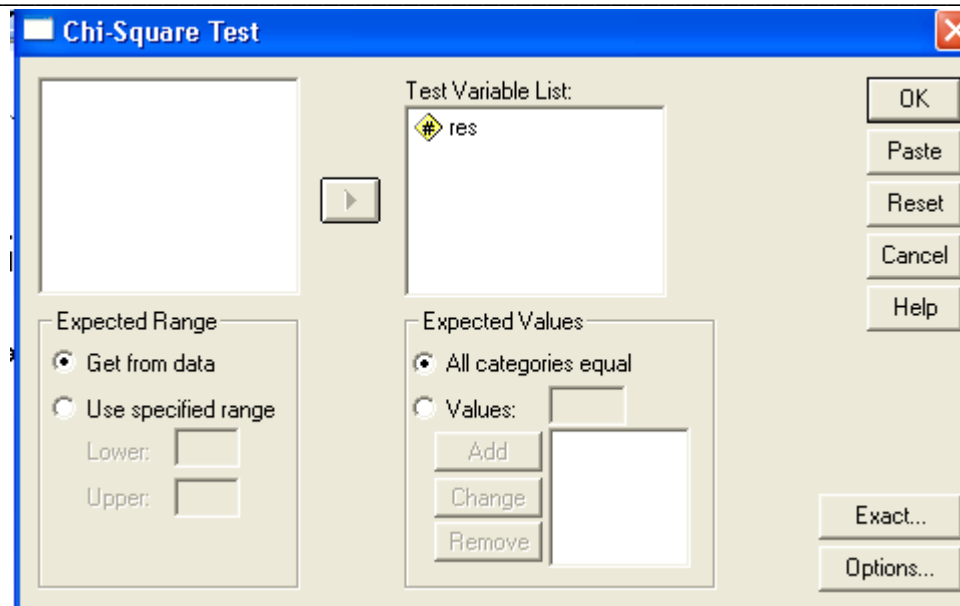
$H_1$ : លទ្ធផលនៃការបោះគ្រាប់ឡកឡាក់មិនគោរពទៅតាមច្បាប់នៃភាពចៃដន្យ

### Syntax:

**NPAR TEST**  
**/CHISQUARE=res**  
**/EXPECTED=EQUAL**  
**/MISSING ANALYSIS.**

### Application tool:

Analyze→Non-parametric→ legacy dialogs → Chi-square



Output:

| res   |            |            |          |
|-------|------------|------------|----------|
|       | Observed N | Expected N | Residual |
| 1     | 18         | 20.0       | -2.0     |
| 2     | 22         | 20.0       | 2.0      |
| 3     | 15         | 20.0       | -5.0     |
| 4     | 25         | 20.0       | 5.0      |
| 5     | 20         | 20.0       | .0       |
| 6     | 20         | 20.0       | .0       |
| Total | 120        |            |          |

#### Test Statistics

|                         | res   |
|-------------------------|-------|
| Chi-Square <sup>a</sup> | 2.900 |
| df                      | 5     |
| Asymp. Sig.             | .715  |

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 20.0.

កំណត់រកតម្លៃ Critical value ក្នុងកម្រិត significance level 5% :

$$\chi^2_{0.95;5} = 11.07$$

ដោយតម្លៃ  $\chi^2 = 2.90 < \chi^2_{\alpha;df} = 11.07$  នាំអោយយើងទទួលយកសម្មតិកម្មសូន្យ មានន័យថា លទ្ធផលនៃការបោះត្រាប់ឡកឡាក់នេះចេញដោយគោរពទៅតាមច្បាប់នៃភាពចៃដន្យ ។



ដូចគ្នានេះដែរ គេអាចប្រើតំលៃ Asymp.sig ជំនួសយំលៃ  $\chi^2$  ក្នុងការសំរេចចិត្តក្នុងការផ្ទៀងផ្ទាត់នេះ ដោយគោរពទៅតាមគោលការណ៍នៃតំលៃ Sig. ។

ដោយ Asymp.sig = 0.715 >  $\alpha = 0.05$  នោះ គេទទួលយកសម្មតិកម្មសូន្យ ។

ចំណាំ៖ Asymp.sig មានន័យថា Asymptotic Significance Probability Value ។

## ២- Chi-square as independent test

ក្នុងករណីនេះ Chi-square សិក្សាអំពីទំនាក់ទំនងដែលកើតឡើងរវាងអថេរពីរ ដែលជាប្រភេទ Qualitative data ។ តាមរយៈនៃការសិក្សានេះយើងអាចដឹងថាអថេរមួយអាចមានឥទ្ធិពលទៅលើការប្រែប្រួលនៃអថេរមួយទៀតដែរឬទេ ហើយដែលការសិក្សានេះធ្វើឡើងតាមរយៈនៃ Crosstabs analysis ។

### ទំរង់នៃសម្មតិកម្ម៖

$H_0$  : A និង B មិនអាស្រ័យគ្នា

Vs

$H_1$  : A និង B អាស្រ័យគ្នា

តំលៃ Chi-square ត្រូវបានគណនាតាម រូបមន្ត ៖

$$\chi^2 = \sum_{i,j} \frac{(fo_{ij} - fe_{ij})^2}{fe_{ij}}$$

ដែល  $fo_{ij}$  ជា observed frequency នៅក្នុងតារាង Crosstab

$fe_{ij}$  ជា expected frequency ដែលបានគណនាក្នុងតារាង Crosstabs តាមរូបមន្តខាងក្រោម៖

ក្រោម៖

$$fe_{ij} = (\text{Total in row } i) * (\text{Total in column } j) / (\text{Total of observation})$$

ក្នុងករណីនេះ degree of freedom (df) ត្រូវបានគណនាតាមរូបមន្ត៖

$$df = (\text{Number of rows} - 1)(\text{Number of Column} - 1)$$

ការសំរេចនៃ Test of hypothesis ធ្វើឡើងដូចនៅក្នុងចំណុចខាងលើដែរ ។

## ឧទាហរណ៍

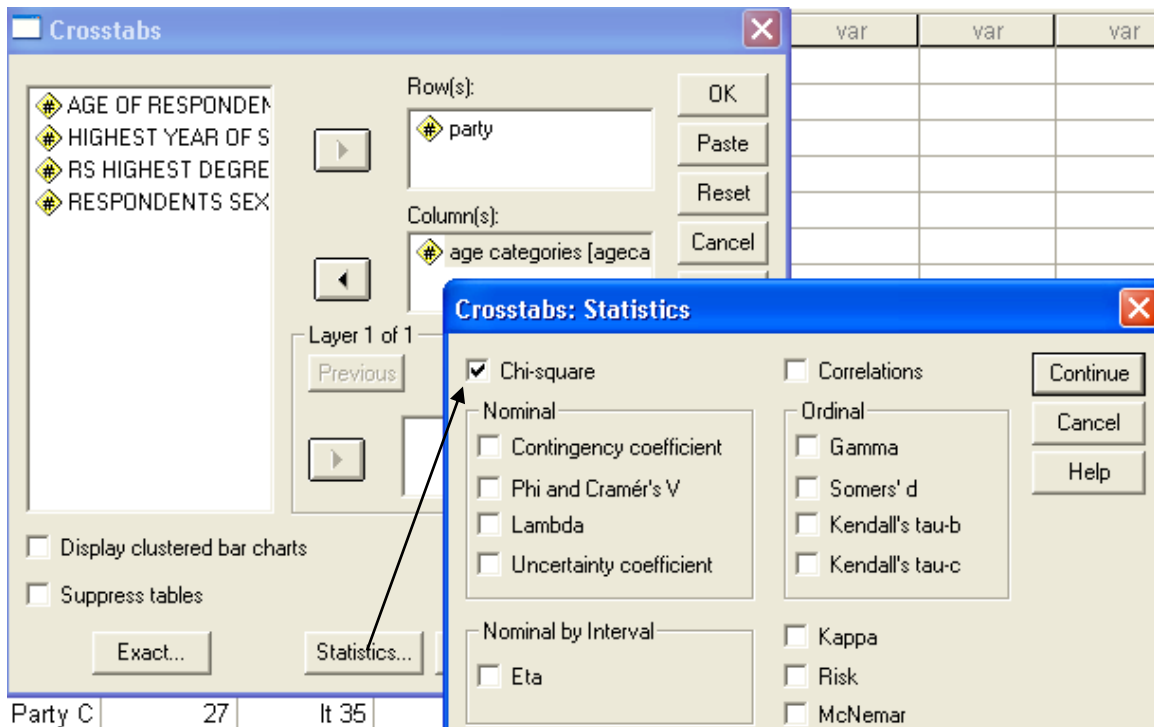
គេមានបំណងចង់ដឹងថាតើកត្តាអាយុ (agecat) នៃអ្នកដែលនឹងចូលរួមក្នុងការបោះឆ្នោតមានឥទ្ធិពលទៅលើជំរើសបេក្ខភាពគណបក្សនយោបាយ(Party) ដែរឬទេ? ក្នុងករណីនេះការអង្កេតមួយបានធ្វើនៅក្នុងតំបន់ហើយទិន្នន័យដែលទទួលបានឈ្មោះ **voter** នៅក្នុង SPSS ។

### ទំរង់សម្មតិកម្មសមស្រប៖

$H_0$ : ជំរើសបេក្ខភាពគណបក្សនយោបាយ មិនអាស្រ័យនឹងវ័យនៃអ្នកចូលរួមក្នុងការបោះឆ្នោត  
v.s

$H_a$ : ជំរើសបេក្ខភាពគណបក្សនយោបាយ អាស្រ័យនឹងវ័យនៃអ្នកចូលរួមក្នុងការបោះឆ្នោត

Analyze→ Descriptives statistics→ Crosstabs



**Output:**

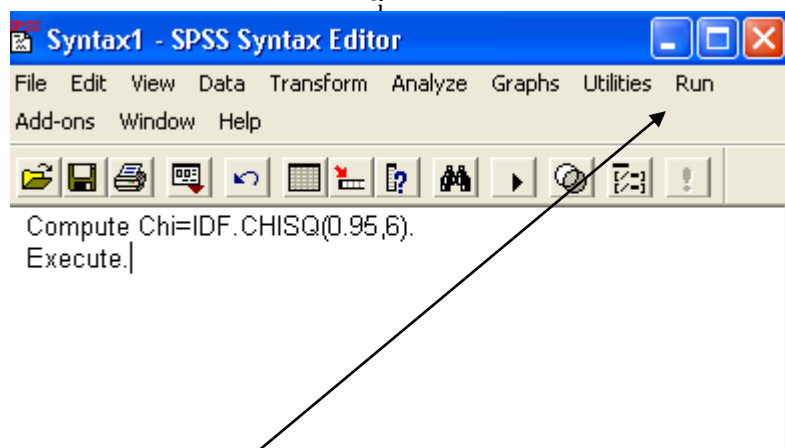
**Chi-Square Tests**

|                              | Value               | df | Asymp. Sig. (2-sided) |
|------------------------------|---------------------|----|-----------------------|
| Pearson Chi-Square           | 56.531 <sup>a</sup> | 6  | .000                  |
| Likelihood Ratio             | 63.654              | 6  | .000                  |
| Linear-by-Linear Association | 3.689               | 1  | .055                  |
| N of Valid Cases             | 1847                |    |                       |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 52.38.

តំលៃ Chi-square

កំណត់រកតំលៃ Critical value of chisquare ក្នុងករណីដែល  $df=6$  និង  $\alpha=5\%$



ប្រើ run all

$$\chi^2_{0.95;6} = 12.59$$

| Acceptance Region | Rejection Region |
|-------------------|------------------|
|-------------------|------------------|

$$\chi^2_{0.95;6} = 12.59$$

ដោយតំលៃ  $\chi^2 = 56.531 > \chi^2_{0.95;6} = 12.59$  នោះបានសេចក្តីថាជំរើសបេក្ខភាពគណបក្សនយោបាយអាស្រ័យនឹងអាយុនៃអ្នកចូលរួមនៅក្នុងការបោះឆ្នោត ។

ដូចគ្នានេះដែរ យើងអាចប្រើតំលៃ Asymp.sig(2-sided) សំរាប់ការសំរេចចិត្តក្នុងការផ្ទៀងផ្ទាត់សម្មតិកម្ម ។ ដោយ  $\text{Asymp.sig}(2\text{-sided}) = 0.000 < \alpha = 0.05$  នោះគេបដិសេធសម្មតិកម្មសូន្យ ។

ក្នុងករណីនេះយើងអាចពន្យល់ភាពអាស្រ័យនឹងគ្នា តាមរយៈការវាយតម្លៃជាភាគរយនៅក្នុង តារាង

Crosstabs ខាងក្រោម៖

**Party \* age categories Crosstabulation**

% within age categories

|         | age categories |         |         |        | Total  |
|---------|----------------|---------|---------|--------|--------|
|         | lt 35          | 35 - 44 | 45 - 64 | 65 +   |        |
| Party A | 34.9%          | 35.1%   | 35.5%   | 38.2%  | 35.8%  |
| Party B | 22.6%          | 18.2%   | 13.3%   | 4.6%   | 15.1%  |
| Party C | 42.5%          | 46.6%   | 51.2%   | 57.2%  | 49.2%  |
| Total   | 100.0%         | 100.0%  | 100.0%  | 100.0% | 100.0% |

ក្នុងតារាងនេះបានបង្ហាញអោយឃើញថា ការគាំទ្រទៅលើបេក្ខភាព Party C មានការកើនឡើង កាលណាវ័យនៃអ្នកចូលរួមក្នុងការបោះឆ្នោតកាន់តែមានវ័យចំណាស់។ ចំណែកឯបេក្ខភាព Party B កំរិតភាគរយនៃការគាំទ្រកាន់តែថយចុះកាលណាវ័យ របស់អ្នកបោះឆ្នោតកាន់តែមានវ័យចំណាស់។ ចំពោះ Party A ភាគរយនៃការគាំទ្រមិនអាស្រ័យទៅនឹងវ័យនៃអ្នកបោះឆ្នោតទេ។

## ជំពូក៦: Correlation and simple linear regression

### 1). Correlation:

Correlation គឺជាការសិក្សាកំណត់នូវកំរិតទំនាក់ទំនង (កំរិតសហព័ន្ធ) និងទិសដៅបំប្លែង ដែលកើតឡើងរវាងអថេរពីរដែល ជា quantitative(Scale) data ។

#### ឧទាហរណ៍:

- គេនិយាយថាមាន Correlation ដែលកើតឡើងរវាង “Income Sales” & Advertising expenditure ។
- មាន Correlation ដែលកើតឡើងរវាងចំនួនអត្តមានរបស់និស្សិតទៅលទ្ធផលរបស់គាត់ ក្នុងការសិក្សាទៅលើមុខវិជ្ជានីមួយៗ ។

គេអាចសិក្សាអំពី Correlation តាមរយៈ Graphic ដែលមានឈ្មោះថា Scatter diagram

សន្និដ្ឋានថាមានគូទិន្នន័យ X & Y ដែលជាប្រភេទ Quantitative :

| Obs. | X  | Y  |
|------|----|----|
| 1    | X1 | Y1 |
| 2    | X2 | Y2 |
| 3    | X3 | Y3 |
| .    | .  | .  |
| .    | .  | .  |
| .    | .  | .  |
| n    | Xn | Yn |

(a) **Positives Correlation** មានន័យថា X & Y ប្រែប្រួល តាមទិសដៅ ស្របគ្នា ដែល តម្លៃមួយទៅតាម ទិសនៃបន្ទាត់វិជ្ជមាន ៖

បើ X កើន  $\Rightarrow$  Y កើន

បើ X ថុន  $\Rightarrow$  Y ថុន

(b) **Negatives Correlation** មានន័យថា X & Y ប្រែប្រួលតាមទិសដៅ ផ្ទុយគ្នាហើយត្រឹមស្របតាមទិសនៃបន្ទាត់អវិជ្ជមាន ៖

បើ X កើន  $\Rightarrow$  Y ថុន

បើ X ថុន  $\Rightarrow$  Y កើន

(c) **No Correlation** រវាង X & Y មានន័យថា (X,Y) នៅរាយប៉ាយពេញលេញដោយមិនបានបង្ហាញអំពីទិសដៅនៃការប្រែប្រួលណាមួយនោះទេ,  $\text{Correlation}=0$  ។

(d) មានការកើតឡើងនូវ Perfect Positive Correlation មានន័យថាទិន្នន័យទាំងអស់ត្រូវរួមនៅលើបន្ទាត់ដែលមានទិសវិជ្ជមាន ។

100% នៃ X កើន  $\Rightarrow$  100% នៃ Y កើន

100% នៃ X ថុះ  $\Rightarrow$  100% នៃ Y ថុះ

**Correlation=+1**

(e) មានការកើតឡើងនូវ Perfect Negatives Correlation មានន័យថាទិន្នន័យទាំងអស់ត្រូវរួមនៅលើបន្ទាត់ដែលមានទិសអវិជ្ជមាន ។

100% នៃ X កើន  $\Rightarrow$  100% នៃ Y ថុះ

100% នៃ X ថុះ  $\Rightarrow$  100% នៃ Y កើន

**Correlation = -1**

ដូច្នេះ តំលៃ Correlation រវាង X & Y ដែលគាងដោយ  $R_{xy}$  ឬ  $\text{Corr}(X,Y)$  កំណត់:

$$-1 \leq R_{xy} \leq +1$$

នៅក្នុង SPSS ឬ S-plus តំលៃ Correlation ត្រូវបានគណនាតាមរូបមន្ត:

$$R_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

ដែល:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ Mean of X}$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \text{ Mean of Y}$$

## ឧទាហរណ៍៖

ខាងក្រោមនេះជាទិន្នន័យដែលស្រង់ក្នុងរយៈពេល ១០ឆ្នាំ បន្តបន្ទាប់គ្នាដែលបង្ហាញអំពីការប្រែប្រួលនៃ

Annual profit & Spent on Research and development ទៅតាមឆ្នាំនីមួយៗរបស់ក្រុមហ៊ុនដ៏ធំ

មួយ ។

| Year                                 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
|--------------------------------------|------|------|------|------|------|------|------|------|------|------|
| Annual Profit<br>(Millions dollars)  | 5    | 4    | 7    | 8    | 10   | 9    | 4    | 11   | 12   | 13   |
| Spent on R & D<br>(Millions dollars) | 15   | 12   | 20   | 19   | 22   | 21   | 13   | 25   | 30   | 32   |

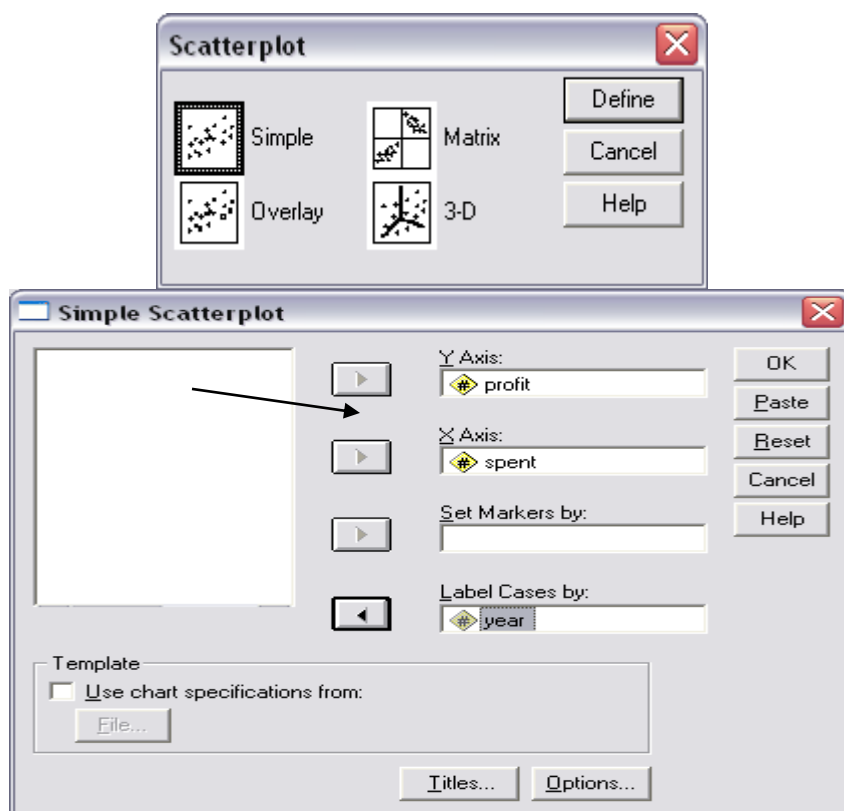
ក-សង់ Sactter diagram & គណនាតម្លៃ Correlation ព្រមទាំងធ្វើការបកស្រាយ ។

ខ-បង្កើត Model regression(Model of prediction) ទៅលើ Annual Profit ។

ចំណើយ

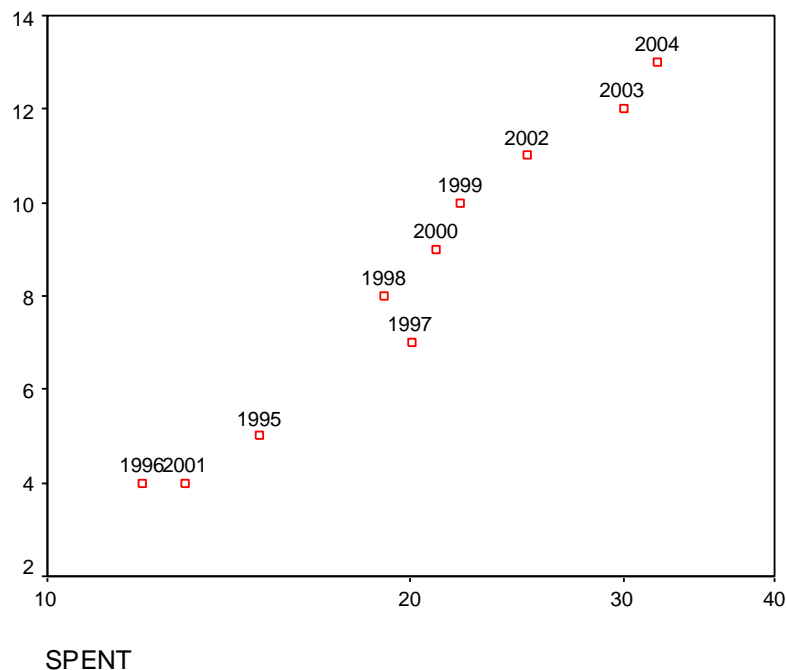
ក- សង់ Scatter diagram

Graphs → Scatter → Simple → Define



ចុចពីរដងទៅ diagram

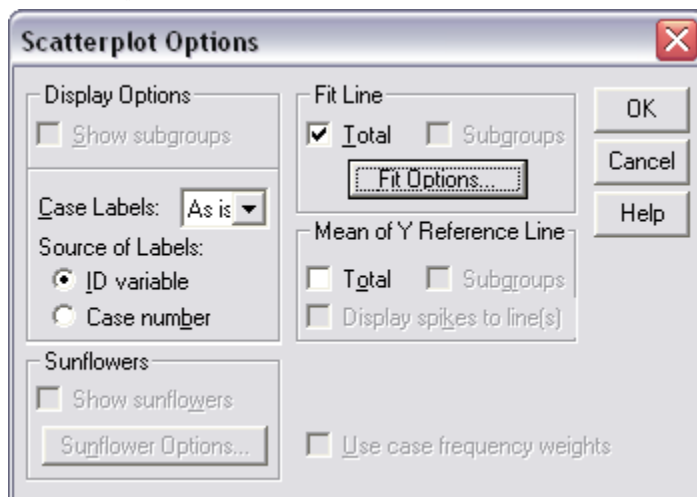
ចុចទៅលើ Point id tool ហើយចុចទៅលើចំណុចនីមួយៗ :



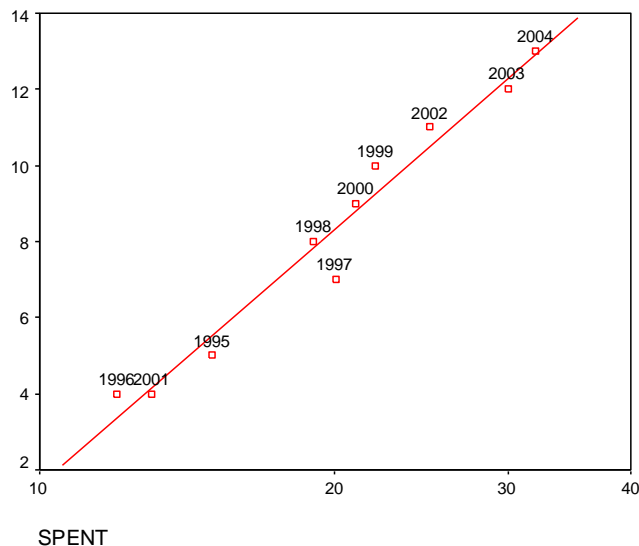
ចុចពីរដងទៅលើ diagram

ជ្រើសរើសយក Chart → Option

ជ្រើសរើសយក ☒ Total





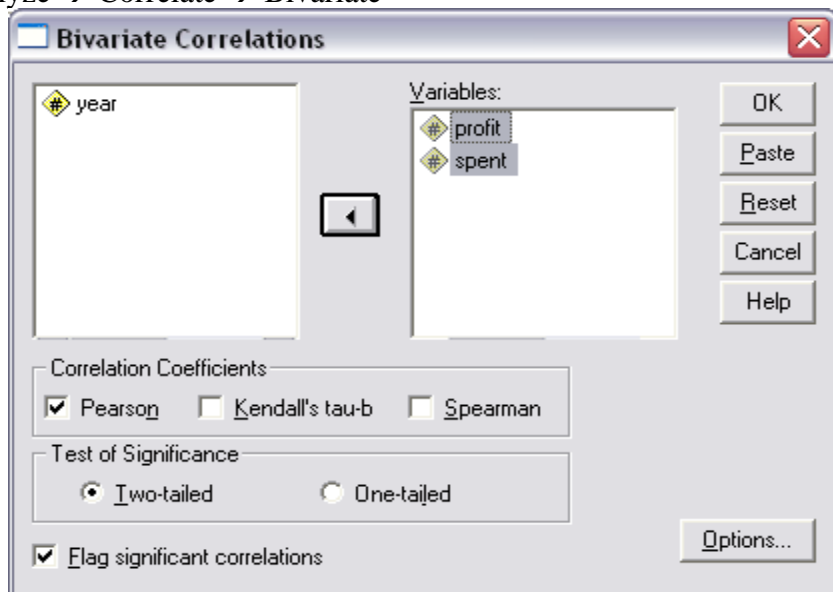


### ការបកស្រាយ:

តាមរយៈនៃ Scatter diagram យើងសង្កេតឃើញថា “Annual profit” & “Spent on R & D” ប្រែប្រួលតាមទិសដៅស្របគ្នា មានន័យថាកាលណា “Spent on R & D” មានកំរិតកាន់តែខ្ពស់នោះ Annual profit ក៏កាន់តែកើនដែរ ។ ដូច្នេះយើងអាចនិយាយថា Annual profit ដែលក្រុមហ៊ុនទទួលបាន អាស្រ័យទៅនឹង “Spent on R & D” ។

កំនត់រកតម្លៃ Correlation:

Analyze → Correlate → Bivariate



## Correlations

| Correlations                      |                     |                                  |                                   |
|-----------------------------------|---------------------|----------------------------------|-----------------------------------|
|                                   |                     | Annual profit in million dollars | Spent on R & D in million dollars |
| Annual profit in million dollars  | Pearson Correlation | 1                                | .975**                            |
|                                   | Sig. (2-tailed)     |                                  | .000                              |
|                                   | N                   | 10                               | 10                                |
| Spent on R & D in million dollars | Pearson Correlation | .975**                           | 1                                 |
|                                   | Sig. (2-tailed)     | .000                             |                                   |
|                                   | N                   | 10                               | 10                                |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

តំលៃ Correlation:

$$R_{xy} = 0.975 = 97.50\%$$

ដោយតំលៃ Correlation មានកំរិតរហូតដល់ 97.5% មានន័យថា “Annual profit” មានទំនាក់ទំនងយ៉ាងខ្លាំងទៅនឹង “Spent on R & D”។ គេអាចនិយាយថាបើគេបន្ថយការចំណាយទៅលើ R & D នោះ Annual profit ក៏មានការថយចុះដែរ ផ្ទុយទៅវិញបើគេបង្កើន R & D នោះគេនឹងសង្ឃឹមថា Annual profit ក៏កាន់តែកើនឡើងដែរ។

**Test of Signification:**

$$H_0 : R_{xy}=0$$

VS

$$H_1 : R_{xy} \neq 0$$

$$\text{Sig (s - tailed)} = 0.000 < \alpha = 0.01 \Rightarrow \text{Reject } H_0$$

មានន័យថា **Correlation is significant at the 0.01 Level** ។

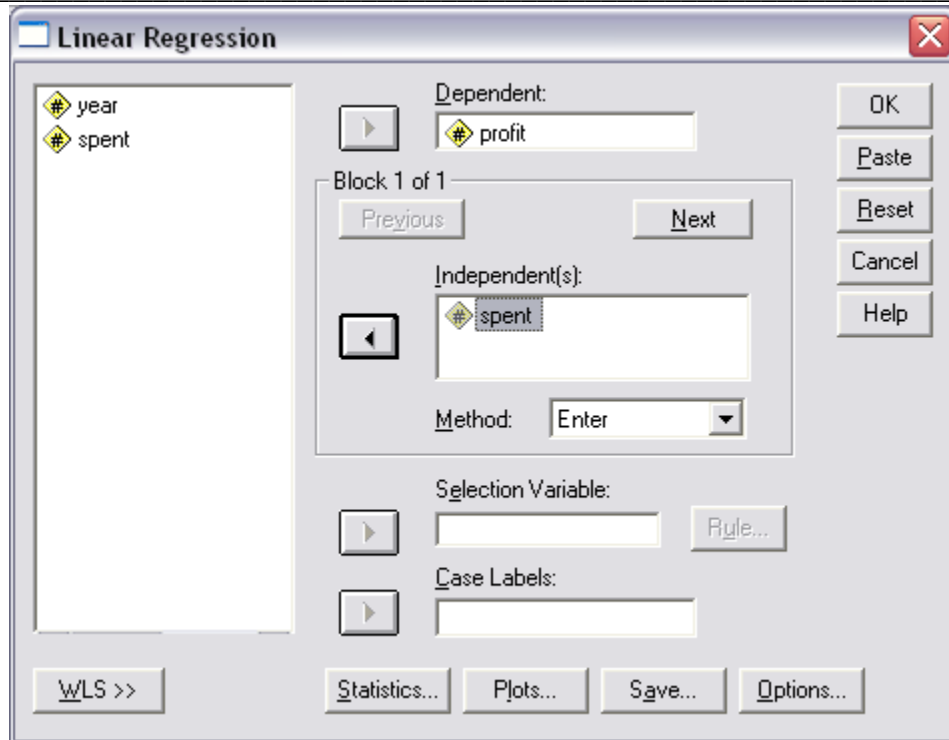
កំនត់រក Regression model (Model of prediction)

**Model:**

$$\text{Annual profit} = b_0 + b_1 * (\text{Spend on R\&D})$$

**Analysis:**

Analyze → Regression → Linear



Coefficients(a)

| Model |            | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|--------|------|
|       |            | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant) | -1.626                      | .838       |                           | -1.939 | .088 |
|       | SPENT      | .475                        | .038       | .975                      | 12.376 | .000 |

a. Dependent Variable: PROFIT

$$b_0 = -1.626$$

$$b_1 = 0.475$$

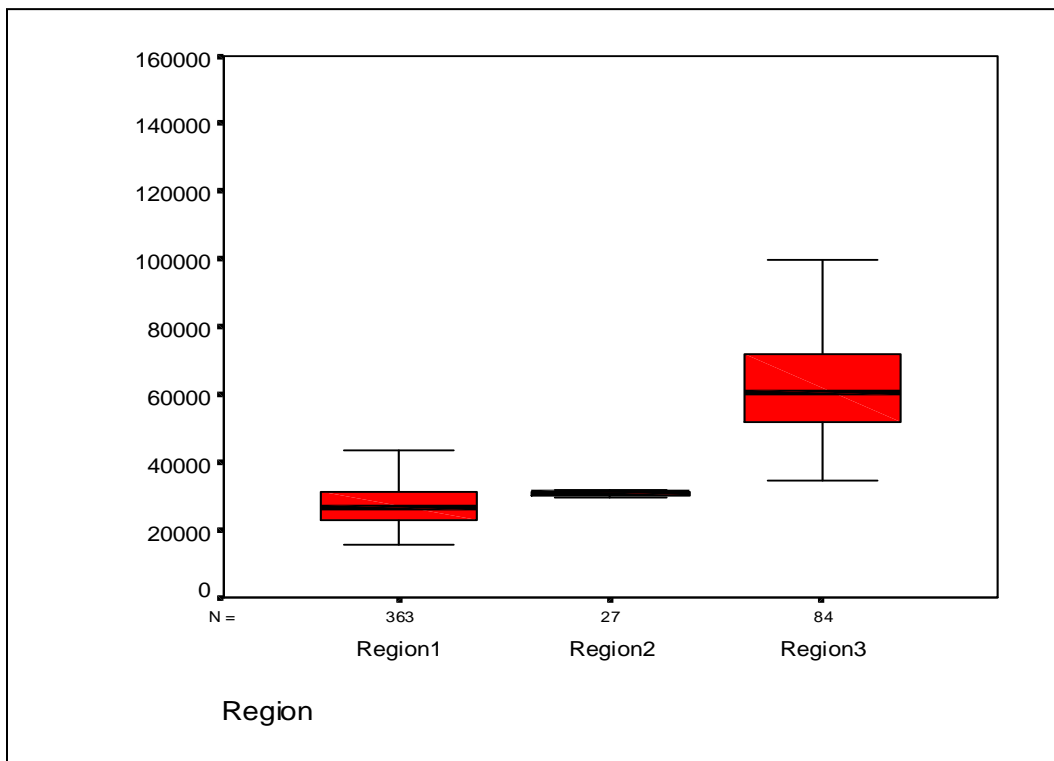
$$\text{Annual profit (Millions dollars)} = -1.626 + 0.475 * (\text{Spent on R \& D})$$

តាមរយៈនៃ Model ខាងលើយើងអាចនិយាយបានថាក្នុងការប្រែប្រួលមួយឯកតានៃការចំណាយ ទៅលើការស្រាវជ្រាវ និងការអភិវឌ្ឍ (Spent on R&D) របស់ក្រុមហ៊ុន នោះប្រាក់ចំណេញប្រចាំឆ្នាំ (Annual profit) នឹងមានការកើនឡើង 0.475 millions ។

## ជំពូក៦: លំហាត់អនុវត្តន៍

### វិញ្ញាសាទី១:

គេបានធ្វើការអង្កេតទៅលើកំរិតជីវភាពរស់នៅប្រចាំឆ្នាំរបស់គ្រួសារមួយចំនួនដែលរស់នៅក្នុងតំបន់ ៣ កន្លែងខុសៗគ្នា ហើយតាមរយៈនៃការវិភាគ គេទទួលបាននូវក្រាហ្វិច Boxplot ដូចរូបភាពខាងក្រោម ចូរធ្វើការពន្យល់និងបកស្រាយអោយបានក្បោះក្បាយ ។



### វិញ្ញាសាទី២

គេបានធ្វើការអង្កេតទៅលើបុគ្គលមួយចំនួនដែលមានសិទ្ធិបោះឆ្នោតនៅក្នុងតំបន់មួយកន្លែងដើម្បីធ្វើការវាយតម្លៃទៅលើភាគរយនៃអ្នកគាំទ្រគណៈបក្ស A,B,C ទៅតាមថ្នាក់នៃអាយុនៃបុគ្គលទាំងនោះ ។ លទ្ធផលនៃការវិភាគនេះមានដូចក្នុងតារាងខាងក្រោម:

| PARTY * age categories Crosstabulation |         |                |         |         |              |        |
|----------------------------------------|---------|----------------|---------|---------|--------------|--------|
| % within PARTY                         |         | age categories |         |         |              | Total  |
|                                        |         | Less than 35   | 35 - 44 | 45 - 64 | More than 64 |        |
| PARTY                                  | Party A | 23.1%          | 23.6%   | 33.1%   | 20.1%        | 100.0% |
|                                        | Party B | 35.6%          | 29.1%   | 29.5%   | 5.8%         | 100.0% |
|                                        | Party C | 20.5%          | 22.8%   | 34.8%   | 21.9%        | 100.0% |
|                                        | Total   | 23.7%          | 24.0%   | 33.4%   | 18.8%        | 100.0% |

ចូរធ្វើការពន្យល់ទៅលើលទ្ធផលវិភាគក្នុងតារាងខាងលើ ។

### វិញ្ញាសាទី៣

គេបានធ្វើការអង្កេតទៅលើគ្រួសារមួយចំនួនដែលរស់នៅតំបន់ដាច់ស្រយាលមួយកន្លែងដោយធ្វើការស្រង់ព័ត៌មានទៅលើបរិមាណស្រូវដែលគាត់ប្រមូលផលបានក្នុងរយៈពេលពេញមួយឆ្នាំ ។ ដោយសង្កេតឃើញថាប្រជាជនមានកង្វះខាតស្បៀងដោយសារភាពរាំងស្ងួត រាជរដ្ឋាភិបាលបានធ្វើការអភិវឌ្ឍវិស័យធារាសាស្ត្រនៅតំបន់នេះ ។ រយៈពេលមួយឆ្នាំក្រោយមកគេបានធ្វើការអង្កេតទៅលើគ្រួសារមួយចំនួននេះម្តងទៀតហើយលទ្ធផលនៃការវិភាគទិន្នន័យមានដូចខាងក្រោម៖

| Paired Samples Test |                                           |                              |        |
|---------------------|-------------------------------------------|------------------------------|--------|
|                     |                                           | Pair 1                       |        |
|                     |                                           | After(Tonne) - Before(Tonne) |        |
| Paired Differences  | Mean                                      |                              | 1.5333 |
|                     | Std. Deviation                            |                              | .78571 |
|                     | Std. Error Mean                           |                              | .17146 |
|                     | 95% Confidence Interval of the Difference | Lower                        | 1.1757 |
|                     |                                           | Upper                        | 1.8910 |
| t                   |                                           |                              | 8.943  |
| df                  |                                           |                              | 20     |
| Sig. (2-tailed)     |                                           |                              | .000   |

តើក្រោមការជួយដោះស្រាយរបស់រាជរដ្ឋាភិបាលក្រសួងដែលរស់នៅតំបន់នេះទទួលបាននូវទិន្នផលស្រូវប្រសើរជាងមុនដែរឬទេ ដោយធ្វើការសន្និដ្ឋានក្នុងកំរិតកំហុស (significance level) 5% ?

ចូរធ្វើការបកស្រាយអោយ បានក្បោះក្បាយ ។

### វិធានសាទី៤

គេបានធ្វើការសិក្សាទៅលើប្រជាជនមួយចំនួនដែលមានសិទ្ធិចូលរួមការបោះឆ្នោតនាពេលខាងមុខដើម្បីចង់ដឹងថាការជ្រើសរើសយកគណបក្សនយោបាយណាមួយអាស្រ័យទៅនឹងថ្នាក់ នៃអាយុអ្នកចូលរួមក្នុងការបោះឆ្នោតដែរឬទេ? លទ្ធផលនៃការវិភាគទៅលើទិន្នន័យដែលទទួលបានពីការអង្កេតមានដូចខាងក្រោម៖

| age categories * Election Crosstabulation |         |          |         |         |        |
|-------------------------------------------|---------|----------|---------|---------|--------|
| % within age categories                   |         |          |         |         |        |
|                                           |         | Election |         |         | Total  |
|                                           |         | PARTY A  | PARTY B | PARTY C |        |
| age                                       | lt 35   | 34.9%    | 22.6%   | 42.5%   | 100.0% |
| categories                                | 35 - 44 | 35.1%    | 18.2%   | 46.6%   | 100.0% |
|                                           | 45 - 64 | 35.5%    | 13.3%   | 51.2%   | 100.0% |
|                                           | 65 +    | 38.2%    | 4.6%    | 57.2%   | 100.0% |
| Total                                     |         | 35.8%    | 15.1%   | 49.2%   | 100.0% |

| Chi-Square Tests             |                     |    |                       |
|------------------------------|---------------------|----|-----------------------|
|                              | Value               | df | Asymp. Sig. (2-sided) |
| Pearson Chi-Square           | 56.531 <sup>a</sup> | 6  | .000                  |
| Likelihood Ratio             | 63.654              | 6  | .000                  |
| Linear-by-Linear Association | 3.689               | 1  | .055                  |
| N of Valid Cases             | 1847                |    |                       |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 52.38.

ចូរធ្វើការបកស្រាយពន្យល់អោយបានក្បោះក្បាយ ។ (significance level=0.05)

## វិញ្ញាសាទី៥

យើងសង្កេតឃើញថានៅខេត្តសៀមរាបជីវភាពប្រជាជនមួយភាគធំអាស្រ័យទៅនឹងវិស័យទេសចរណ៍ ។ ដោយនៅចុងឆ្នាំ២០០៥នេះគេសង្កេតឃើញថាបរិមាណនៃភ្ញៀវទេសចរណ៍មានការកើនឡើង ដូច្នេះគេបាន ធ្វើការសិក្សាចង់ដឹងអំពីកំណើនជីវភាពប្រជាជននៅក្នុងតំបន់នោះ ធៀបនឹងឆ្នាំ ២០០៤ ដោយធ្វើការ ប្រៀបធៀបប្រាក់ចំណូលរបស់គ្រួសារនីមួយៗដែលបានអង្កេត ដោយចែងនៅក្នុងឆ្នាំ ២០០៤ និង ឆ្នាំ ២០០៥ ។ លទ្ធផលវិភាគមានដូចខាងក្រោម៖

| Group Statistics |           |     |             |                |                 |
|------------------|-----------|-----|-------------|----------------|-----------------|
| College          |           | N   | Mean        | Std. Deviation | Std. Error Mean |
| Revenue/year     | Year 2005 | 281 | \$30,876.87 | \$5,189.219    | \$309.563       |
|                  | Year 2004 | 415 | \$23,780.00 | \$7,678.715    | \$376.933       |

| Independent Samples Test     |                                           |             | Revenue/year            |
|------------------------------|-------------------------------------------|-------------|-------------------------|
|                              |                                           |             | Equal variances assumed |
| t-test for Equality of Means | t                                         |             | 13.539                  |
|                              | df                                        |             | 694                     |
|                              | Sig. (2-tailed)                           |             | .000                    |
|                              | Mean Difference                           |             | \$7,096.87              |
|                              | Std. Error Difference                     |             | \$524.186               |
|                              | 95% Confidence Interval of the Difference | Lower Upper | \$6,067.687 \$8,126.050 |

ចូរធ្វើការពន្យល់និងបកស្រាយអោយបានក្លៀនក្លាយ ។ (significance level=5%)

---

**ឯកសារយោង**

- 1- **Mendenhall, William** (1988), Statistics for the engineering and computer science.
- 2- **Richard I. Levin & David S. Rubin** (1998), Statistics for Management.
- 3- **Andy Field**(2000), Discovering Statistics using SPSS for window.