សាកលវិទ្យាល័យជាតិគ្រប់គ្រង
National University of Management

មូលដ្ឋានគ្រឹះស្ថិតិអាជីវកម្ម

បណ្ឌិត សៅ ឡាយ   បណ្ឌិត ឆាយ ផាង

២០២១

សាកលវិទ្យាល័យជាតិគ្រប់គ្រង
National University of Management

មូលដ្ឋានគ្រឹះស្ថិតិអាជីវកម្ម

បណ្ឌិត សៅ ឡាយ   បណ្ឌិត ឆាយ ផាង

២០២១

បុព្វកថា

ដំណើរអភិវឌ្ឍនៃព្រះរាជាណាចក្រកម្ពុជានៅក្នុងយុគសម័យទំនើបនេះ ជាមេរៀនដ៏ជោគជ័យបំផុតមួយដែលចាប់ឫសគល់ចេញពីការបញ្ចប់របបប្រល័យពូជសាសន៍ ការបញ្ចប់សង្គ្រាម ការផ្សះផ្សារជាតិ ការកសាងមូលដ្ឋានរឹងមាំនៃសន្តិភាពនិងស្ថេរភាព និងការអភិវឌ្ឍសេដ្ឋកិច្ច។ នៅក្រោយពេលដែលសន្តិភាពត្រូវបានកើតឡើងដោយបរិបូណ៌នៅឆ្នាំ១៩៩៨ កម្ពុជាទទួលកំណើនសេដ្ឋកិច្ចខ្ពស់ គឺប្រមាណ ៨% ក្នុងមួយឆ្នាំ។ លើសពីនេះទៀត អត្រានៃភាពក្រីក្រត្រូវបានកាត់បន្ថយពីប្រមាណ ៥៣% នៅឆ្នាំ ២០០៤ មកនៅទាបជាង ១០% នៅឆ្នាំ២០១៩។ ដំណើរនៃការអភិវឌ្ឍជាតិជាសកម្មភាពដែលបន្តទៅមុខជាប់ជានិច្ច ហើយគោលនយោបាយថ្មីៗដែលមានលក្ខណៈអន្តរវិស័យគ្រប់ដណ្តប់ ក៏កំពុងលេចរូបរាងឡើង ដើម្បីតម្រង់ទិសកម្ពុជាឆ្ពោះទៅកាន់ប្រទេសមានប្រាក់ចំណូលមធ្យមកម្រិតខ្ពស់នៅឆ្នាំ២០៣០ និងឈានឡើងជាប្រទេសមានប្រាក់ចំណូលខ្ពស់ នៅឆ្នាំ២០៥០។ ការប្រែប្រួលឆាប់រហ័សនៃនិម្មាបនកម្មពិភពលោកនិងតំបន់ រួមទាំងទំនាក់ទំនងភូមិសាស្ត្រនយោបាយ បានផ្តល់កាលានុវត្តភាពសម្រាប់ការអភិវឌ្ឍឧស្សាហកម្មនៅកម្ពុជា ដែលត្រូវបានរាជរដ្ឋាភិបាលចាត់ទុកជាមូលដ្ឋានគ្រឹះនៃកំណើនសេដ្ឋកិច្ចកម្ពុជា។ រាជរដ្ឋាភិបាលកម្ពុជា បាននិងកំពុងបន្តពង្រឹង និងអភិវឌ្ឍវិស័យអប់រំឆ្ពោះទៅរកការស្រាវជ្រាវ និងនវានុវត្តន៍ ដើម្បីពង្រឹងសមត្ថភាព និងជំនាញរបស់ធនធានមនុស្សនៅកម្ពុជា ឱ្យស្របទៅនឹងបរិបទថ្មីនៃការអភិវឌ្ឍ ជាពិសេសការពង្រឹងសហគ្រិនភាពក្នុងការរៀបចំម៉ូដែលធុរកិច្ចថ្មីៗ។ ដើម្បីចាប់យកកាលានុវត្តភាពពីបដិវត្តន៍ឧស្សាហកម្មទី៤ និងសេដ្ឋកិច្ចឌីជីថលដែលកំពុងផុសផុលឡើង ប្រព័ន្ធអេកូឡូហ្ស៊ីដែលបង្កលក្ខណៈអំណោយផលដល់ការបង្កើតថ្មី នវានុវត្តន៍ ការស្រាវជ្រាវ និងអភិវឌ្ឍន៍ ត្រូវតែមានការកែលម្អ។

បណ្តាប្រទេសនៅទ្វីបអាស៊ីកំពុងនាំមុខក្នុងការវិនិយោគលើការស្រាវជ្រាវនិងអភិវឌ្ឍ ដោយមានភាគហ៊ុនប្រមាណ ៤៤% នៃការវិនិយោគទាំងមូលរបស់ពិភពលោក។ ប្រទេសចិនកំពុងបន្តកសាងហេដ្ឋារចនាសម្ព័ន្ធនៃការវិនិយោគលើការស្រាវជ្រាវនិងអភិវឌ្ឍ ក៏ដូចជាសមត្ថភាពមនុស្ស។ ផ្ទុយទៅវិញ ប្រទេសនៅទ្វីបអាមេរិកខាងត្បូង និងអាហ្វ្រិក កំពុងស្ថិតនៅឆ្ងាយពីការវិនិយោគនេះ ហើយជាលទ្ធផលប្រទេសទាំងនោះក៏ពុំមានកំណើនសេដ្ឋកិច្ចគួរឱ្យកត់សម្គាល់ដែរ។ ទុនវិនិយោគសរុបលើការស្រាវជ្រាវនិងអភិវឌ្ឍរបស់ប្រទេសនៅទ្វីបអាមេរិកខាងត្បូវ និងអាហ្វ្រិក មានប្រមាណ ៥% នៃការវិនិយោគ ទាំងមូលរបស់ពិភពលោក ក្នុងពេលដែលតំបន់ទាំង ២ នេះមានប្រជាជនប្រមាណ ២០% នៃប្រជាជន ពិភពលោក។ ប្រទេសចំនួន ៦ ដែលមានលំដាប់ខ្ពស់ជាងគេនៅក្នុងការវិនិយោគលើការស្រាវជ្រាវ និងអភិវឌ្ឍ រួមមានសហរដ្ឋអាមេរិក ចិន ជប៉ុន អាល្លឺម៉ង់ ឥណ្ឌា និងកូរ៉េខាងត្បូង ដែលស្មើនឹងប្រមាណ ៧០% នៃទុនវិនិយោគសរុបរបស់ពិភពលោក។

តើចំណេះដឹង ផលិតផល និងសេវាកម្មថ្មីទាំងនេះកើតឡើងពីអ្វី? ហើយកើតឡើងដោយរបៀបណា? ព្រះរាជាណាចក្រកម្ពុជាកំពុងតែកសាងមូលដ្ឋានសម្រាប់ការត្រៀមខ្លួនទទួល និងប្រកួតប្រជែងក្នុងយុគសម័យបដិវត្តឧស្សាហកម្មទី៤ នៅក្នុងសេដ្ឋកិច្ចដែលផ្អែកលើពុទ្ធិ ហើយដែលប្រការនេះចាំបាច់តម្រូវ

ឱ្យពលរដ្ឋកម្ពុជា ត្រូវក្លាយខ្លួនជាពលរដ្ឋឌីជីថល ពលរដ្ឋសកល និងពលរដ្ឋដែលប្រកបដោយការទទួលខុសត្រូវ ដែលមានសមត្ថភាពក្នុងការផលិត ចែកចាយ និងប្រើប្រាស់ពុទ្ធិដើម្បីទទួលមនុញ្ញផល និងរួមចំណែកក្នុងកំណើន។ ធនាគារពិភពលោកបានធ្វើការកត់សម្គាល់តាំងពីឆ្នាំ២០០២ នូវបម្លាស់ប្តូរនៃមូលដ្ឋានសេដ្ឋកិច្ច ពីសេដ្ឋកិច្ចដែលពឹងផ្អែកលើកម្លាំងពលកម្ម និងធនធានអតិកម្ម (Labour and Resource Based Economy) ទៅកាន់សេដ្ឋកិច្ចដែលពឹងផ្អែកលើពុទ្ធិ (Knowledge Based-Economy) ដែលក្នុងន័យនេះ ពុទ្ធិគឺជាគន្លឹះនៃការអភិវឌ្ឍ។ អាស្រ័យហេតុនេះ នៅលើគន្លងដែលកម្ពុជាកំពុងធ្វើដំណើរឆ្ពោះទៅកាន់សេដ្ឋកិច្ចឌីជីថល សង្គមកម្ពុជាត្រូវតែមានសមត្ថភាពក្នុងការផលិត ជ្រើសរើស បន្សាំ បង្កើតមុខរបរ និងប្រើប្រាស់ពុទ្ធិ ដើម្បីរក្សានិរន្តរភាពនៃកំណើន និងកែលម្អជីវភាពរស់នៅ។ សមត្ថភាពទាំងនេះ អាចកើតឡើងនៅពេលពលរដ្ឋកម្ពុជាមានឱកាសក្នុងការទទួលបានបទពិសោធន៍ពីការស្រាវជ្រាវ ការបណ្តុះគំនិតច្នៃប្រឌិត និងការស្វែងរកនវានុវត្តន៍។

កំណែទម្រង់វិស័យអប់រំ គឺជាការត្រួសត្រាយមាគ៌ាសម្រាប់ដំណើរឆ្ពោះទៅកាន់សង្គមប្រកបដោយពុទ្ធិ និងប្រជាពលរដ្ឋប្រកបដោយភាពរស់រវើក។ តាមរយៈមូលដ្ឋានអប់រំ សង្គមប្រកបដោយពុទ្ធិនឹងប្រមូលផ្តុំ បង្កើត និងចែករំលែក ទៅកាន់សមាជិកក្នុងសង្គមនូវសម្បទាអប់រំ ពិសេសគឺពុទ្ធិសម្បទា ក្នុងបុព្វហេតុនៃមនុស្សជាតិនិងឧត្តមប្រយោជន៍នៃប្រទេស។ សង្គមប្រកបដោយពុទ្ធិ គឺពុំគ្រាន់តែជាសង្គមដែលសម្បូរព័ត៌មានប៉ុណ្ណោះទេ តែជាសង្គមដែលប្រជាពលរដ្ឋអាចធ្វើបរិវត្តកម្មពីព័ត៌មានទៅជា មូលធនប្រកបដោយប្រសិទ្ធភាព។ ការរីកចម្រើនទៅមុខជាលំដាប់នៃបច្ចេកវិទ្យានិងតំណភ្ជាប់ បានពង្រីកព្រំដែននៃការចូលទៅកាន់ និងការទទួលបានព័ត៌មានជាសកល ហើយដែលក្នុងន័យនេះ ការអប់រំនឹងបន្តវិវត្តទៅមុខនិងមានការផ្លាស់ប្តូរ។ សង្គមមួយដែលមានអំណាន និងរបាប់ជាបុរេលក្ខខណ្ឌនៃជីវភាពប្រចាំថ្ងៃនៃប្រជាពលរដ្ឋ ពេលនោះបំណិននៃអំណាន និពន្ធ និងការគណនាលេខនព្វន្ត គឺជាចលករនៃការរៀនរបស់សិស្ស។ ជាតុដ៏ចម្បងមួយដែលស្ថិតនៅក្នុងការកសាងសង្គមដែលប្រកបដោយ ពុទ្ធិគឺសៀវភៅសិក្សា ហើយការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សាជាប្រចាំ គឺជានវានុវត្តន៍នៃវិស័យអប់រំដែលនាំទៅរកការសិក្សាពេញមួយជីវិត ការអភិវឌ្ឍសម្បទាអប់រំ និងការចែករំលែកចំណេះដឹង។ មូលដ្ឋានអប់រំ ជាពិសេសគឺគ្រឹះស្ថានឧត្តមសិក្សាត្រូវមានតួនាទីដែលប្រកបដោយការឆ្លើយតប ចំពោះតម្រូវការខាងលើនេះ។ សាស្ត្រាចារ្យ អ្នកស្រាវជ្រាវ និងបុគ្គលិកអប់រំត្រូវបន្តសិក្សាជាប់ជានិច្ច តាមរយៈការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ហើយដែលសៀវភៅសិក្សាទាំងនេះនឹងក្លាយជា ស្ពាននៃទំនាក់ទំនងរវាងនវានុវត្តន៍នៃបច្ចេកវិទ្យា និងការរៀននិងបង្រៀននៅក្នុងថ្នាក់រៀន។

សង្គមដែលប្រកបពុទ្ធិ ក៏ជាសង្គមដែលបណ្តុះឱ្យមានរចនាសម្ព័ន្ធទន់នៃសេដ្ឋកិច្ចដែលពឹងផ្អែកលើពុទ្ធិដែរ។ ឧទាហរណ៍ជាក់ស្តែងនៃបែបផែននេះរួមមាន Silicon Valley នៃសហរដ្ឋអាមេរិក សួនឧស្សាហកម្មវិស្វកម្មអាកាសយានយន្តនិងយានយន្តនៅទីក្រុង Munich ប្រទេសអាល្លឺម៉ង់ តំបន់ជីវបច្ចេកវិទ្យានៅក្រុង Hyderabad ប្រទេសឥណ្ឌា តំបន់ផលិតគ្រឿងអេឡិចត្រូនិកនិងសារគមនាគមន៍ ឌីជីថលនៅទីក្រុង Seoul ប្រទេសកូរ៉េខាងត្បូង ក៏ដូចជាសួនឧស្សាហកម្មថាមពល និងឥន្ទនគីមីសាស្ត្រនៃប្រទេស

ប្រេស៊ីល ហើយក៏នៅមានទីក្រុងនៃប្រទេសជាច្រើនទៀតនៅលើពិភពលោក។ លក្ខណៈសម្បត្តិនៃទីក្រុងទាំងនេះគឺការប្រើប្រាស់និន្នាការនៃការអភិវឌ្ឍដែលជំរុញ និងតម្រង់ទិសដោយចំណេះ ដឹង ហើយដែលចំណេះដឹងទាំងនោះកើតចេញជាដំបូង ពីការវិនិយោគទៅលើគ្រឹះស្ថានឧត្តមសិក្សាស្ថាប័នស្រាវជ្រាវមជ្ឈមណ្ឌលឧត្តមភាពនៃជំនាញជាន់ខ្ពស់ ការប្រកួតប្រជែងដោយគុណាធិបតេយ្យ និង ជាពិសេស គឺការបណ្ដុះវប្បធម៌អំណាននិងនិពន្ធសៀវភៅ។ ល្បឿននៃការរីកចម្រើនផ្នែកពុទ្ធិ និងបច្ចេកវិទ្យាកំពុងមានសន្ទុះលឿនជាងអ្វីដែលសិស្ស និងនិស្សិតអាចទទួលបានពីគ្រូនៅគ្រឹះស្ថានសិក្សា ដែលធ្វើឱ្យគោលដៅនៃការអប់រំនៅពេលបច្ចុប្បន្ននេះ មានការប្រឈមខ្លាំងជាងពេលណាទាំងអស់។ ឧទាហរណ៍ ក្នុងមួយឆ្នាំ មានសៀវភៅជាង ២,២លានចំណងជើង ត្រូវបានសរសេរនិងបោះពុម្ព ដែលក្នុងនោះប្រទេសចិនមាន ៤៤០ ពាន់ ចំណែកឯសហរដ្ឋអាមេរិកមាន ៣០៥ពាន់ និងប្រទេសរុស្ស៊ីមាន ១២០ពាន់ចំណងជើង។

ខណៈពេលដែលបច្ចេកវិទ្យាកំពុងរីកចម្រើនជារៀងរាល់ថ្ងៃ មធ្យោបាយសម្រាប់អំណានក៏មានច្រើន ជម្រើសសម្រាប់សិស្ស-និស្សិត និងសាធារណៈជន រួមមានការអានសៀវភៅ ការអានលើឧបករណ៍អេឡិចត្រូនិក ការអានដោយប្រើទូរសព្ទវៃឆ្លាត និងការអានលើកុំព្យូទ័រ ដែលសុទ្ធសឹងជាមធ្យោបាយសំខាន់ៗដែលនាំអ្នកអានទាំងឡាយឱ្យសម្រេចគោលបំណងអានរបស់ខ្លួន។ ម្យ៉ាងវិញទៀត អំណានដោយប្រើមធ្យោបាយបច្ចេកវិទ្យាទំនើប ចំណាយពេលតិច ងាយស្រួលអាន និងជួយដល់បរិស្ថានមួយកម្រិតទៀត។ នាពេលបច្ចុប្បន្ន សិស្ស-និស្សិត និងសាធារណៈជនកម្ពុជាដែលស្រឡាញ់អំណានកំពុងតែប្រើប្រាស់មធ្យោបាយអំណានទាំងនេះ។ បើយើងក្រឡេកមើលទៅប្រទេសជឿនលឿន ទោះបីជាបច្ចេកវិទ្យារីកចម្រើនខ្លាំងយ៉ាងណា អំណានតាមរយៈសៀវភៅនៅតែមានសន្ទុះដដែល។ ម្យ៉ាងវិញទៀត បច្ចេកវិទ្យាអានបែបទំនើបតាមរយៈឧបករណ៍ទំនើប អាស្រ័យលើលទ្ធភាពនៃធនធានអប់រំឌីជីថល និងមាតិកាឌីជីថលគ្រប់គ្រាន់ដែលបានផលិត និងបង្ហោះចែកចាយសម្រាប់អំណាន។

ក្នុងបរិបទកម្ពុជា ជាពិសេសក្នុងបរិការណ៍នៃការផ្ទុះរីករាលដាលនៃជំងឺកូវីដ-១៩ ក្រសួងអប់រំ យុវជន និងកីឡា បានជំរុញឱ្យមានបរិវត្តកម្មឌីជីថលនៅក្នុងអេកូស៊ីស្ដែមនៃការអប់រំ ជាពិសេសការអប់រំ តាមប្រព័ន្ធអេឡិចត្រូនិក និងការអប់រំពីចម្ងាយដើម្បីលើកកម្ពស់អំណាន តាមរយៈការផលិតមាតិកាឌីជីថលដែលមានភាពចំរុះ ការកសាងសមត្ថភាពផ្នែកតំណភ្ជាប់និងវេទិកាឌីជីថល ការពង្រីកវិសាលភាពនៃមជ្ឈមណ្ឌលទិន្នន័យ និងការលើកកម្ពស់គុណភាពនៃការផលិតធនធានអប់រំឌីជីថល គួបផ្សំ ជាមួយការចែកសន្លឹកកិច្ចការឱ្យសិស្សយកទៅរៀននៅផ្ទះ និងការចុះទៅជួបជាមួយសិស្សជាបណ្តុំនៅតាមសហគមន៍។ ក្នុងន័យលើកកម្ពស់អំណាន និងភាពសម្បូរបែបនៃធនធានសៀវភៅសិក្សា ឱ្យកាន់តែមានប្រសិទ្ធភាព និងភាពសក្ដិសិទ្ធិ និងផ្ដល់ឱកាសអំណានកាន់តែច្រើនថែមទៀតដល់សិស្សានុសិស្ស និស្សិត និងសាធារណៈជន ក្រសួងអប់រំ យុវជន និងកីឡាលើកទឹកចិត្តនូវចំណុចមួយចំនួនដូចខាងក្រោម៖

1. សាស្ត្រាចារ្យ អ្នកស្រាវជ្រាវ និងបុគ្គលិកអប់រំ សូមបន្តនិងបង្កើនការបោះពុម្ពស្នាដៃបន្ថែមទៀត ដើម្បីធ្វើឱ្យធនធានសម្រាប់អំណានកាន់តែសម្បូរបែប ជាពិសេសធនធានអំណានជាខេមរភាសា

2. គ្រឹះស្ថានឧត្តមសិក្សា សូមផ្តល់លទ្ធភាពគ្រប់បែបយ៉ាង ដើម្បីឱ្យបុគ្គលិកអប់រំគ្រប់លំដាប់ថ្នាក់ និងនិស្សិតគ្រប់កម្រិតសិក្សាអាចចូលរួមអាន និងសិក្សាស្រាវជ្រាវតាមគ្រប់លទ្ធភាពជាមួយធនធានអំណាន ជាពិសេសការរៀបចំឱ្យមានពេលវេលាសម្រាប់សហសិក្សា និងអំណានក្នុងបណ្ណាល័យ
3. សាស្ត្រាចារ្យតាមមុខវិជ្ជា និងអ្នកស្រាវជ្រាវតាមជំនាញឬវិស័យ ត្រូវរៀបចំដំណើរការរៀន បង្រៀន និងស្រាវជ្រាវដែលមានដាក់បញ្ចូលកិច្ចការស្វ័យសិក្សា សហសិក្សា ឬការស្រាវជ្រាវបណ្ណាល័យ ដែលតម្រូវឱ្យនិស្សិត ត្រូវអាននិងស្រាវជ្រាវជាមួយធនធានអំណាន
4. គ្រឹះស្ថានឧត្តមសិក្សា និងមជ្ឈមណ្ឌលស្រាវជ្រាវ ត្រូវខិតខំឱ្យអស់លទ្ធភាពក្នុងការបង្កើតបណ្ណាល័យ មជ្ឈមណ្ឌលរក្សាឯកសារ ឬមជ្ឈមណ្ឌលអប់រំឌីជីថល ជាដើម ដើម្បីឱ្យបុគ្គលិកអប់រំគ្រប់លំដាប់ថ្នាក់និងនិស្សិតគ្រប់កម្រិតសិក្សា អាចទទួលបាន និងស្វែងរកប្រភពសម្រាប់អំណានកាន់តែសម្បូរបែប និងមានភាពបត់បែន ឆ្លើយតបតាមតម្រូវការអ្នកអាន
5. និស្សិតគ្រប់កម្រិតសិក្សា ត្រូវខិតខំនិងចំណាយពេលអាន និងចាត់ទុកវប្បធម៌ និងអកប្បកិរិយា អំណានជាផ្នែកមួយ នៃពេលវេលានិងភាពស៊ីវិល័យនៃជីវិតប្រចាំថ្ងៃ
6. បងប្អូនជនរួមជាតិ ដែលជាមាតាបិតា ឬអ្នកអាណាព្យាបាល សូមជួយជំរុញនិងបង្កលក្ខណៈកាន់តែច្រើនថែមទៀត ជាពិសេសការលៃលកចំណាយនៅក្នុងគ្រួសារសម្រាប់ការទិញសម្ភារៈសិក្សា សៀវភៅអាន និងឧបករណ៍ សម្រាប់អំណានដល់កូនៗ ដែលចាត់ទុកជាការ វិនិយោគមួយដ៏សំខាន់ សម្រាប់បង្កើនចំណេះដឹង និងអនាគតរបស់ពួកគេ។

ដោយមានការគាំទ្រពីក្រសួងសេដ្ឋកិច្ច និងហិរញ្ញវត្ថុ នៅឆ្នាំ២០២០ ក្រសួងអប់រំ យុវជន និងកីឡា បានបង្កើតមូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍ ដែលហៅកាត់ថា "មូលនិធិ ស.គ.ន" និងហៅជាភាសាអង់គ្លេសថា The Research Creativity and Innovation Fund ដែលហៅកាត់ជាភាសាអង់គ្លេសថា "RCI Fund"។ គោលដៅចម្បងនៃមូលនិធិនេះ គឺរួមចំណែកលើកកម្ពស់វប្បធម៌នៃការស្រាវជ្រាវ បំផុសគំនិតច្នៃប្រឌិត និងជំរុញការធ្វើនវានុវត្ត ដើម្បីជាប្រយោជន៍ដល់វិស័យអប់រំ យុវជន និងកីឡា ដែលឆ្លើយតបទៅនឹងទីផ្សារពលកម្ម និងសាកលភាវូបនីយកម្ម។ មូលនិធិ ស.គ.ន បានសម្រេចកំណត់ប្រធានបទ ជាអាទិភាពសម្រាប់ការគាំទ្រដោយមូលនិធិចំនួន៣ រួមមាន ឌីជីថលនីយកម្មសម្រាប់បដិវត្តឧស្សាហកម្ម៤.0 (Digitalization for IR.4.0) ការស្រាវជ្រាវអនុវត្តលើវិស័យកសិកម្ម (Applied Agricultural Research) និងការស្រាវជ្រាវគរុកោសល្យសតវត្សទី២១ (21st Century Pedagogy Research)។

ដោយមានការធ្វើអាទិភាពរូបនីយកម្មទៅលើទិសដៅនៃការប្រើប្រាស់ថវិកាមូលនិធិ សម្រាប់ឆ្នាំ ២០២០ ក្រសួងសេដ្ឋកិច្ច និងហិរញ្ញវត្ថុ និងក្រសួងអប់រំ យុវជន និងកីឡា បានផ្តល់ការគាំទ្រដល់ការ

រៀបរៀង និពន្ធ និងកែលម្អ សៀវភៅសិក្សា(Text book) ដែលនឹងត្រូវប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា។ គោលបំណងនៃការរៀបរៀង និពន្ធ និងកែលម្អ សៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា គឺដើម្បីបង្កើន បរិមាណ លើកកម្ពស់គុណភាព និងពង្រីកសមធម៌នៃធនធានសិក្សាជាខេមរភាសា ជូនដល់និស្សិតដែល កំពុងបន្តការសិក្សា និងត្រៀមខ្លួនធ្វើការស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។ លើសពីនេះទៀត ការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា មានគោលដៅដូចខាងក្រោម ៖

- ឆ្លើយតបជាបន្ទាន់ចំពោះការខ្វះខាតធនធានសិក្សា ដែលជាតម្រូវការសិក្សារបស់និស្សិតនៅកម្រិតឧត្តមសិក្សា
- លើកកម្ពស់ទំនើបភាវូបនីយកម្ម និងឧត្តមានុវត្តន៍នៃការរៀននិងបង្រៀន និងការស្រាវជ្រាវនៅលើមុខវិជ្ជា កម្មវិធីសិក្សា ឬមុខជំនាញជាក់លាក់
- បង្កើនភាពស៊ីជម្រៅក្នុងការកសាងវិជ្ជាជីវៈនិងបទពិសោធន៍សម្រាប់បាឋនៈសាស្ត្រាចារ្យ និងអ្នកស្រាវជ្រាវ
- រួមចំណែកដល់ការកសាងភាពជាសហគមន៍វិជ្ជាជីវៈ ការចែករំលែកបទពិសោធន៍ និងវប្បធម៌នៃការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា។

ក្រសួងអប់រំ យុវជន និងកីឡា បានវាយតម្លៃខ្ពស់ចំពោះការបោះជំហាន ប្រកបដោយមនសិការវិជ្ជាជីវៈនៃគ្រឹះស្ថានឧត្តមសិក្សា និងបុគ្គលិកអប់រំទាំងអស់ ក្នុងការរៀបចំ រៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ដើម្បីបង្កើនបរិមាណ លើកកម្ពស់គុណភាព និងពង្រឹងសមធម៌នៃធនធានសិក្សាជាខេមរភាសា ជូននិស្សិតដែលកំពុងបន្តការសិក្សា និងត្រៀមខ្លួនធ្វើការស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។ សៀវភៅសិក្សាជាផ្នែកមួយនៃការទទួលស្គាល់គុណភាពអប់រំនៃគ្រឹះស្ថានឧត្តមសិក្សា និងជាធនធានសិក្សាដែលជាមូលដ្ឋានមួយដ៏សំខាន់ ក្នុងការគាំទ្រដល់ការបង្រៀន និងរៀន ហើយត្រូវមានបរិមាណគ្រប់គ្រាន់ ឆ្លើយតបទៅនឹងកម្មវិធីអប់រំ និងតម្រូវការសិក្សាស្រាវជ្រាវ។ ជាគោលការណ៍ គ្រឹះស្ថានឧត្តមសិក្សាទាំងអស់ ត្រូវមានសៀវភៅសិក្សាដែលប្រើជាគោលសម្រាប់មុខវិជ្ជានីមួយៗ។ ចំនួនសៀវភៅសិក្សាដែលគ្រប់គ្រាន់សម្រាប់ការស្រាវជ្រាវ និងការសិក្សារបស់និស្សិត ត្រូវមានយ៉ាងតិចមួយចំណងជើងក្នុងមួយមុខវិជ្ជា ហើយត្រូវតម្កល់យ៉ាងតិច ២ច្បាប់ នៅក្នុងបណ្ណាល័យ ឬអាចរកបានតាមប្រព័ន្ធអេឡិចត្រូនិក។ ក្រសួងអប់រំ យុវជន និងកីឡា លើកទឹកចិត្តបន្ថែមទៀតជូនដល់គ្រឹះស្ថានឧត្តមសិក្សារដ្ឋ និងឯកជន ដែលបានស្នើសុំថវិកាមូលនិធិរួច សូមចូលរួមបន្ថែមទៀតដើម្បីបង្កើនចំនួនចំណងជើងសៀវភៅ។ ចំណែកគ្រឹះស្ថានឧត្តមសិក្សារដ្ឋ និងឯកជនដែលពុំទាន់បានដាក់ពាក្យស្នើសុំ សូមចូលរួម ដើម្បីជាគុណប្រយោជន៍ដល់តម្រូវការដ៏ទទូច និងថ្លៃថ្លានៃនិស្សិតកម្ពុជាក្នុងការសិក្សា និងស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។

សេចក្តីបញ្ជាក់
នៃមូលនិធិការស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍

សៀវភៅសិក្សានេះជាលទ្ធផលនៃការស្នើសុំអនុវត្តថវិកាមូលនិធិការស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍ ក្នុងគម្រោងរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ដែលនឹងត្រូវប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា។ សៀវភៅសិក្សានេះ ត្រូវបានរៀបរៀង និពន្ធ ឬកែលម្អដោយមានការធានាអះអាងថាជាស្នាដៃរបស់អ្នក និពន្ធផ្ទាល់ និងបានឆ្លងកាត់ត្រួតពិនិត្យ ផ្តល់យោបល់ និងវាយតម្លៃដោយក្រុមប្រឹក្សាអប់រំ ក្រុមប្រឹក្សាស្រាវជ្រាវ ឬក្រុមប្រឹក្សាដែលមានតម្លៃស្មើនៃគ្រឹះស្ថានឧត្តមសិក្សា និងតាមរយៈកិច្ចសន្យាដែលបានធ្វើឡើង និងដែលបានតម្កល់ទុកនៅមូលនិធិការស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍។ រាល់ខ្លឹមសារ ការបកស្រាយ និងរូបភាព គឺជាជំហរនិងទស្សនៈផ្ទាល់របស់អ្នកនិពន្ធ ហើយ ពុំឆ្លុះបញ្ចាំង ឬជាតំណាងដល់មូលនិធិការស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍ នៃក្រសួងអប់រំ យុវជន និងកីឡា ឡើយ។

អារម្ភកថា

សៀវភៅមូលដ្ឋានគ្រឹះស្ថិតិនៃអាជីវកម្មនេះ ត្រូវបានតាក់តែងឡើងសម្រាប់និស្សិតថ្នាក់បរិញ្ញបត្រ ឬថ្នាក់ប្រហាក់ប្រហែល។ វាគ្របដណ្តប់ទៅលើផ្នែកសំខាន់ជាច្រើននៃស្ថិតិដែលអនុវត្តលើផ្នែកអាជីវកម្មនិងសេដ្ឋកិច្ច។ វាត្រូវបានរៀបរាងឡើងដោយមានលំដាប់លំដោយច្បាស់លាស់ ដោយចាប់ផ្តើមពីការបែងចែកប្រភេទអថេរ ឬទិន្នន័យ វិធីវាស់វែង គណនា បង្ហាញ ការធ្វើតេស្តលើសម្មតិកម្ម និង ការសន្និដ្ឋាន ដើម្បីធ្វើសេចក្តីសម្រេចចិត្តក្នុងអាជីវកម្ម ហើយលើសពីនេះទៅទៀត សៀវភៅនេះក៏បានបញ្ចូលផ្នែកព្យាករណ៍ខ្លះផងដែរ។ ផ្នែកនិមួយៗមានបញ្ចូលឧទាហរណ៍ច្បាស់លាស់ ដោយចាប់ផ្តើមពី បង្ហាញរូបមន្ត បន្ទាប់មកការអនុវត្តរូបមន្តតាមរយៈការគណនាដោយដៃ ព្រមទាំងការអនុវត្តជាមួយនឹងម៉ាស៊ីនគិតលេខ និងម៉ាស៊ីនកុំព្យូទ័រ តាមរយៈកម្មវិធី Excel និង SPSS ដើម្បីធ្វើការប្រៀបធៀប ឬផ្ទៀងផ្ទាត់លទ្ធផលជាមួយគ្នា ហើយនៅចុងជំពូកនិមួយៗគឺមានសំនួរ និងលំហាត់អនុវត្តដើម្បីឲ្យនិស្សិតរៀនគ្រាប់ធ្វើតាម។

ក្រោយពីបានអានសៀវភៅមូលដ្ឋានគ្រឹះស្ថិតិនៃអាជីវកម្មនេះរួច អ្នកសិក្សាមានចំណេះចេះដឹងអាចធ្វើកាអនុវត្តដោយផ្ទាល់ តាមរយៈការប្រមូលទិន្នន័យរួចធ្វើតេស្តលើគោលនយោបាយ សរសេរសារណា ឬសរសេរអត្ថបទវិទ្យាសាស្ត្រផ្សេងៗ និងអាចចេះព្យាករណ៍នៅពេលអនាគតថែមទៀតដែរ។

ទោះបីជាមូលដ្ឋានគ្រឹះស្ថិតិនៃអាជីវកម្មនេះ រៀបចំឡើងដោយសំរិតសំរាំងយ៉ាងណាក៏ដោយ រមែងតែមានចំណុចខ្វះខាតខ្លះពិតជាមិនខាន ហើយអ្នកចងក្រងសង្ឃឹមយ៉ាងមុតមាំថា វាអាចបំពេញនូវសេចក្តីត្រូវការរបស់និស្សិត និងអ្នកសិក្សាស្រាវជ្រាវមួយចំនួនធំ ក៏ដូចជាអ្នកអានជាច្រើនទៀតផងដែរ។ អាស្រ័យហេតុនេះយើងខ្ញុំដែលជាអ្នកចងក្រងស្នាដៃសៀវភៅនេះឡើង រង់ចាំទទួលនូវមតិរិះគន់ស្ថាបនាពីសំណាក់អ្នកអាន ក៏ដូចជាគណៈកម្មការត្រួតពិនិត្យ គ្រប់ពេលវេលានិងដោយក្តីរីករាយបំផុត ដើម្បីកែលំអរសៀវភៅនេះឲ្យកាន់តែមានភាពល្អប្រសើរឡើងនាពេលអនាគត។

សៀភៅនេះត្រូវបានតាក់តែងឡើងដោយគំនិតផ្តួចផ្តើម ជំនួយស្មារតីនិងហិរញ្ញវត្ថុពីប្រភពផ្សេងៗ ដូច្នេះយើងខ្ញុំគោរពនិងថ្លែងអំណាគុណរចំពោះ ឯកឧត្តម បណ្ឌិត ហោ ប៉េង សកលវិទ្យាធិការនៃសកលវិទ្យល័យជាតិគ្រប់គ្រង និង ឯកឧត្តម បណ្ឌិតសភាចារ្យ ហង់ ជួនណារ៉ុន រដ្ឋមន្ត្រី ក្រសួងអប់រំយុវជន និងកីឡា ដែលជាស្ថាបនិកនៃកម្មវិធីរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា សម្រាប់ប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា ព្រមទាំងម្ចាស់ជំនួយហិរញ្ញវត្ថុដ៏ទៃទៀត។ ហើយលើសពីនេះទៅទៀត ក៏សូមថ្លែងចំពោះភរិយា ស៊ឹម វិច្ឆិកា និង ចន្ថា សូដានី ដែលបានគាំទ្រនិងលើកទឹកចិត្តក្នុងរៀបរៀងនិងនិពន្ធសៀភៅនេះ។

ថ្ងៃចន្ទ ១៤រោច ខែពិសាខ ឆ្នាំឆ្លូវ ត្រីស័ក ព.ស ២៥៦៥
រាជធានីភ្នំពេញ ថ្ងៃទី១០ ខែឧសភា ឆ្នាំ២០២១

អំពីអ្នកនិពន្ធ

បណ្ឌិត **ឆាយ ផាង** បច្ចុប្បន្នជាព្រឹទ្ធបុរសមហាវិទ្យាល័យព័ត៌មានវិទ្យា នៃសាកលវិទ្យា ល័យជាតិគ្រប់គ្រង។ លោកបាននិងកំពុងបង្រៀនថ្នាក់បរិញ្ញាបត្រ ថ្នាក់បរិញ្ញាបត្រជាន់ខ្ពស់ និងថ្នាក់បណ្ឌិត នៅសកលវិទ្យាល័យជាតិគ្រប់គ្រង នូវមុខវិជ្ជាមួយចំនួនដូចជា Applied Statistics for Business and Economics, Management Information Technology, Quantitative Analysis, Inventory Management, Introduction to Statistics, និង Mathematical Analysis។ លោកបានទទួលបរិញ្ញាបត្រវិស្វករវិទ្យាសាស្ត្រកុំព្យូទ័រ ពីសាកលវិទ្យាល័យបច្ចេកវិទ្យាក្រុងហាណូយ វៀតណាម ឆ្នាំ ១៩៩៧ បរិញ្ញាបត្រជាន់ខ្ពស់ ជំនាញគ្រប់គ្រងអាជីវកម្ម និងព័ត៌មានវិទ្យា ពីសាកលវិទ្យាល័យ Universiti Utara Malaysia (UUM) សហការជាមួយសកលវិទ្យាល័យជាតិគ្រប់គ្រង ឆ្នាំ ២០០១ និងសញ្ញបត្របណ្ឌិត ជំនាញគ្រប់គ្រងអាជីវកម្ម ពីសកលវិទ្យា ល័យជាតិគ្រប់គ្រង ឆ្នាំ ២០១៧។ លើសពីនេះ លោកក៏បានសរសេរអត្ថបទបោះពុម្ពផ្សាយនៅ NUM Research Series បានចំនួន ៧ អត្ថបទ ។ ក្រៅពីការងារបង្រៀនជាង ២៣ឆ្នាំ និងការងារសរសេរបោះពុម្ពផ្សាយ លោកក៏បានចូលរួមដឹកនាំនិស្សិតក្នុងការសរសេរសារណាសម្រាប់ ថ្នាក់បរិញ្ញាបត្រ បរិញ្ញាបត្រជាន់ខ្ពស់ និងបណ្ឌិតថែមទៀតផង។

បណ្ឌិត **សៅ ឡាយ** បច្ចុប្បន្នជាប្រធានដេប៉ាតឺម៉ង់ច្បាប់សាធារណៈ នៃសាកលវិទ្យាល័យជាតិគ្រប់គ្រង។ លោកកំពុងបង្រៀនថ្នាក់បរិញ្ញាបត្រ ថ្នាក់បរិញ្ញាបត្រជាន់ខ្ពស់ និងថ្នាក់បណ្ឌិត នៅសាកលវិទ្យាល័យជាតិគ្រប់គ្រង នូវមុខវិជ្ជាមួយចំនួនដូចជា Econometrics for Finance, Econometrics, Applied Statistics for Business and Economics, Investment Analysis and Portfolio Management, និង Financial Risk Management។ លោកបានទទួលបរិញ្ញាបត្រនីតិសាស្ត្រ ពីសាកលវិទ្យាល័យភូមិន្ទនីតិសាស្ត្រ និងវិទ្យាសាស្ត្រសេដ្ឋកិច្ច ឆ្នាំ ១៩៩៧ បរិញ្ញាបត្រជាន់ខ្ពស់ ជំនាញគ្រប់គ្រងអាជីវកម្ម ពីសាកលវិទ្យាល័យ Universiti Utara Malaysia (UUM) សហការជាមួយសាកលវិទ្យាល័យជាតិគ្រប់គ្រង ឆ្នាំ ២០០១ និងសញ្ញបត្រ បណ្ឌិត ជំនាញគ្រប់គ្រងអាជីវកម្ម ពីសាកលវិទ្យាល័យជាតិគ្រប់គ្រង ឆ្នាំ ២០១២។ លោកបានសរសេរសៀវភៅ “លំនាំដើមនៃរដ្ឋបាលសធារណៈ” (បោះពុម្ព ៧លើក ចេញនៅឆ្នាំ២០០៦ និងចុងក្រោយក្រោយគឺឆ្នាំ ២០២១) សម្រាប់បង្រៀនថ្នាក់បរិញ្ញាបត្រ។ ក្រៅពីនេះលោកក៏បានសរសេរអត្ថបទបោះពុម្ពផ្សាយនៅ NUM Research Series បានចំនួន ៨ អត្ថបទ និងមួយអត្ថបទទៀត គឺបានបោះពុម្ពជាអន្តរជាតិ ជាទម្រង់ “Working Paper” នៅឆ្នាំ ១៩១៨។ ក្រៅពីការងារបង្រៀនជាង ២៣ឆ្នាំ និងការងារសរសេរបោះពុម្ពផ្សាយ លោកក៏បានចូលរួមដឹកនាំនិស្សិតក្នុងការសរសេរសារណាថ្នាក់បណ្ឌិតបានចំនួន ៣នាក់ និងថ្នាក់បរិញ្ញាបត្រជាន់ខ្ពស់ជាច្រើនទៀត។

តារាងមាតិកា

ជំពូកទី 1 លំនាំដើមនៃស្ថិតិ 1

គោលបំណង 1

1.1 ស្ថិតិគឺជាអ្វី? 1

1.2 ហេតុអ្វីចាំបាច់ត្រូវការរៀនស្ថិតិ? 1

1.3 វាក្យស័ព្ទមូលដ្ឋាននៃស្ថិតិ 2

1.4 ប្រភេទនៃអថេរ 3

1.5 រង្វាស់អថេរ 4

1.6 កម្មវិធីសម្រាប់គណនាស្ថិតិ 6

1.7 របៀបប្រើប្រាស់សៀវភៅនេះ 6

ជំពូកទី 2 ការរៀបចំទិន្នន័យ 8

គោលបំណង 8

2.1 ទិន្នន័យនិងប្រភពទិន្នន័យ 8

2.2 ការរៀចំទិន្នន័យអថេរប្រភេទ ឬដាច់ 8

2.3 អថេរប្រភេទដែលមានមួយអថេរ 9

2.2.2 អថេរប្រភេទដែលមានពីរឬច្រើនអថេរ 11

2.3 ការរៀបចំទិន្នន័យជាលេខ 14

2.3.1 លំដាប់លំដោយ និងបំណែងចែកហ្វ្រេកង់ចំពោះទិន្នន័យជាលេខ 14

2.4 គំនូសតាងអថេរជាលេខ 17

a. គំនូសតាងដើមនិងសន្លឹក 17

b. គំនូសតាងហ៊ីស្តូក្រាម 18

c. គំនូសតាងប៉ូលីហ្គូន 18

d. គំនូសតាងហាយ-ឡូ-ខ្លូស 20

e. គំនូសតាង ស្កាតតឺផ្លូត 22

f. គំនូសតាងជាស៊េរី 23
ជំពូកទី 3 ស្ថិតិបរិយាយ 29
គោលបំណង 29
3.1. ស្ថិតិបរិយាយ 29
3.2 រង្វាស់កណ្ដាល 29
3.2.1 មធ្យមពិជគណិត 30
3.2.2 មធ្យមទម្ងន់ 30
3.2.3 មធ្យមធរណីមាត្រ 31
3.2.4 មេដ្យាន 31
3.2.5 ម៉ូដ 32
3.3 រង្វាស់ភាពរាយប៉ាយ 33
3.3.2 វ៉ារីយ៉ង់និងគម្លាតស្តង់ដារ 33
3.3.3 មេគុណនៃភាពរាយប៉ាយ 37
3.3.4 សតភាគ ចតុភាគ ចន្លោះចតុភាគ និត គំនូសប្រអប់ 37
3.4 បំណែងចែកទម្រង់ទិន្នន័យ 41
3.5 កូវ៉ារីយ៉ង់និង មេគុណទំនាក់ទំនង 44
ជំពូកទី 4 បំណែងចែកប្រូបាប៊ីលីតេនៃអថេរដាច់ 49
4.1 បំណែងចែកប្រូបាប៊ីលីតេនៃអរថេរដាច់ 49
4.1.1 តម្លៃរំពឹងទុកនៃបំណែងចែកប្រូបាប៊ីលីតេរបស់អរថេរដាច់ 50
4.1.2 វ៉ារីយង់និងគម្លាតស្តង់ដារនៃអរថេរដាច់ 50
4.2 កូវ៉ារីយ៉ង និងការអនុវត្តន៍ក្នុងផ្នែកហិរញ្ញវត្ថុ 51
4.2.1 កូវ៉ារីយ៉ង់ 51
4.2.2 អត្រារំពឹងទុកនិងហានិភ័យនៃសម្លៀតវិនិយោគ 53
4.3 បំនែងចែកទ្វេធា 54

4.3.1 ប្រូបាប៊ីលីតេនៃបំនែងចែកទ្វេធានិងវិបាក 54
4.3.2 តម្លៃមធ្យម វ៉ារីយ៉ង់ និងគម្លាតស្តង់ដារនៃបំណែងចែកទ្វេធា 56
4.4 បំណែងចែកព័រសុង 56
4.5 បំណែងចែកអ៊ីពែកហ្សែនមីទ្រីស 57
4.6 បំណែងធម្មតា 59
4.6.1 បំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ 61
4.6.2 គណនាប្រូបាប៊ីលីតេនៃបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ 62
4.6.3 ការវាយតម្លៃ ន័រម៉ាល់លីតធី 64
4.7 បំណែងចែកឯកសណ្ឋាន 68
4.8 បំណែងចែកអេចស្ប៉ូណង់ស្យែល 69
ជំពូកទី 5 ការប្រមូលទិន្នន័យនិងបំណែងចែកប៉ាន់គំរូ 74
5.1 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូ 74
5.1.1 វិធីសាស្ត្រចៃដន្យ 75
5.1.2 វិធីសាស្ត្រប្រមូលប៉ានគំរូជាប្រព័ន្ធ 75
5.1.3 វិធីសាស្ត្រប្រមូលប៉ានគំរូតាមកូតា 76
5.1.4 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមក្រុម 77
5.2 វិធីសាស្ត្រមិនមែនប្រូបាប៊ីលីតេ 78
5.2.1 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមបែបវិនិច្ឆ័យ 78
5.2.2 វិធីសាស្ត្រងាយស្រួល 79
5.3 ការវាយតម្លៃលើការប្រមូលប៉ាន់គំរូក្នុងករណីស្ទង់មតិ 79
5.3.1 កំហុសក្នុងការស្ទង់មតិ 79
5.3.2 បញ្ហាក្រមសីលធម៌ 80
5.4 បំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូ 80
5.5 ភាពមិនលំអៀងនៃបំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូ 83

5.6 បំណែងចែកប៉ាន់គំរូនៃសមមាត្រ..85
ជំពូកទី 6 ការប៉ាន់ស្មានលើចន្លោះជឿជាក់..90
6.1 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលគម្លាតស្តង់ដារដឹង 90
6.2 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលគម្លាតស្តង់ដារមិនដឹង 92
6.2.1 វិបាកនៃបំណែងចែកអក្សរជី..92
6.2.2 ចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលដែលគម្លាតស្តង់ដារមិនដឹង 93
6.3 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់សមាមាត្រ..94
6.4 ការកំណត់ចំនួនប៉ាន់គំរូ 95
6.4.1 ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់តម្លៃមធ្យម 95
6.4.2 ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់សមាមាត្រ..96
6.5 ការអនុវត្តន៍ការប៉ាន់ស្មានចន្លោះជឿជាក់ទៅលើសាវនកម្ម និងការត្រួតពិនិត្យគុណភាព.......97
6.5.1 ការប៉ាន់ស្មានលើបរិមាណសរុបនៃចំនួនសកល 97
6.5.2 ការប៉ាន់ស្មានលើភាពខុសគ្នា..99
6.5.3 ការប៉ាន់ស្មានចន្លោះជឿជាក់កន្ទុយម្ខាង 100
6.6 ការប៉ាន់ស្មាចន្លោះជឿជាក់និងបញ្ហាសីលធម៌ 101
ជំពូកទី 7 ការធ្វើតេស្តលើសម្មតិកម្មសម្រាប់ប៉ាន់គំរូតែមួយ...104
7.1 នីតិវិធីសំខាន់នៃការធ្វើតេស្តសម្មតិកម្ម..104
7.2 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសងខាង 107
7.2.1 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសងខាងពេលគម្លាតស្តង់ដារដឹង 107
7.2.2 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសងខាងពេលគម្លាតស្តង់ដារមិនដឹង 111
7.3 ការធ្វើតេស្តកន្ទុយម្ខាង..113
ជំពូកទី 8 ការធ្វើតេស្តលើសម្មតិកម្មលើប៉ាន់គំរូពីរ..122
8.1 ការប្រៀបធៀបតម្លៃមធ្យមនៃចំនួនសកលពីរដែលឯករាជ្យ 122
8.1.1 តេស្តរកភាពខុសគ្នានៃតម្លៃសកលពីរដែលមានវ៉ារីយ៉ង់ស្មើគ្នា 122

8.1.2 ធីតេស្តរកភាពខុសគ្នារវាងតម្លៃមធ្យមនៃចំនួនសកលពីរដែលមានវ៉ារីយ៉ង់មិនស្មើ 125
8.2 ការប្រៀបធៀបតម្លៃមធ្យមរបស់ចំនួនសកលពីរដែលទាក់ទងគ្នា 127
8.2.1 ធីតេស្តទិន្នន័យជាគូរ 127
8.2.2 ប្រើតម្លៃភីវែលយូ 129
8.2.3 ប្រើចន្លោះជឿជាក់ 129
8.3 ការប្រៀបធៀបតម្លៃខុសគ្នារបស់សមាមាត្រ 130
8.3.1 ការប្រើតេស្តស្ថិតិ 131
8.3.2 ការប្រើតម្លៃភីវែលយូ 132
8.3.3 ការប្រើចន្លោះជឿជាក់ 132
8.4 ការធ្វើតេស្តលើវ៉ារីយ៉ង់ 133
8.4.1 ការប្រើតេស្តស្ថិតិ 135
8.4.2 ការប្រើតម្លៃភីវែលយូ 135
ជំពូកទី 9: ឆាយស្វែតេស្តនិងតេស្តដទៃទៀតដែលមិនមែនជាប៉ារ៉ាម៉ែត្រតេស្ត 141
9.1 ឆាយស្វែតេស្តចំពោះភាពអនុលោមតាម 141
9.2 ឆាយស្វែតេស្តសម្រាប់តារាងខ្វែង 144
9.2.1 ឆាយស្វែតេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរ 144
9.2.2 ឆាយស្វែតេស្តរកភាពខុសគ្នារវាងសមាមាត្រលើសពីពីរ 148
9.3 ឆាយស្វែតេស្តរកភាពឯករាជ្យ 152
9.4 ម៉ាកណេម៉ាតេស្ត 155
9.5 វីកូសុនតេស្តលើចំនួនសកលពីរឯករាជ្យ 160
9.5.1 វីកូសុនតេស្តលើចំនួនសកលពីរឯករាជ្យៈ ប៉ាន់គំរូតូច 161
9.5.2 វីកូសុនតេស្តលើប៉ាន់គំរូពីរឯករាជ្យៈ ប៉ាន់គំរូធំ 163
9.5.3 វីកូសុនតេស្តលើប៉ាន់គំរូពីរមិនឯករាជ្យៈ ប៉ាន់គំរូធំ 166
9.6 គ្រូស្កាល់វ៉ាលីសតេស្ត 169

ជំពូកទី 10 សមីការលីនេអ៊ែ 177
10.1 ការវិភាគសមីការលីនេអ៊ែ 177
10.2 សមីការលីននេអ៊ែរ 178
10.3 ការកំណត់ទំនាក់ទំនងលីននេអ៊ែរ 179
10.3.1 វិធីសាស្ត្រអឌីនណារីលិសស្វែរ 180
10.3.2 ការវាស់វែងបម្រែបម្រួល 186
10.3.3 ការគណនាបម្រែបម្រួល 187
10.3.4 គណនាអ៊ែការ៉េ 188
10.3.5 អ៊ែរើស្តង់ដារ 189
10.4 ការសន្និដ្ឋានអំពីមេគុណប្រាប់ទិស 190
10.5 អ៊ែហ្វតេស្តសម្រាប់ម៉ូដែលទាំងមូល 192
10.5 ល័ក្ខខ័ណ្ឌនៃសមីការលីនេអ៊ែ 192
10.5.1 ការវាយតម្លៃទំនាក់ទំនងជាបន្ទាត់ 193
10.5.2 ការវាយតម្លៃលើភាពឯករាជ្យ 196
10.5.3 ការវាយតម្លៃភាពថេរនៃវ៉ារីយ៉ង់និងដំណោះស្រាយ 202
10.6 ដំណោះស្រាយពេលវ៉ារីយ៉ង់មិនថេរ 205
10.7 ការព្យាករណ៍ 211
ឯកសារយោង 216
ឧបសម្ព័ន្ធ 218

ជំពូកទី 1 លំនាំដើមនៃស្ថិតិ

(Introduction to Statistics)

គោលបំណង

- កំណត់និយមន័យស្ថិតិ
- ពន្យល់មូលហេតុដែលនាំអោយរៀនពីស្ថិតិ
- ពន្យល់អំពីសារៈសំខាន់នៃស្ថិតិនៅក្នុងជំនួញ
- ពន្យល់អំពីប្រភេទនៃស្ថិតិ
- ពន្យល់អំពីកំហុសនៃការប្រើប្រាស់ស្ថិតិ

1.1 **ស្ថិតិគឺជាអ្វី**?

ស្ថិតិគឺផ្នែកមួយនៃគណិតវិទ្យាដែលបានប្រើ(អនុវត្ត)ក្នុងការប្រមូល រៀបចំ បង្ហាញ ឬពិពណ៌នា វិភាគនិង បកស្រាយទិន្នន័យចេញពីតំណាងប៉ាន់គំរូ ដែលមានបរិមាណតិច ដើម្បីធ្វើការសន្និដ្ឋានទៅលើសកល ដែលមានបរិមាណច្រើន[1]។

ផ្អែកតាមនិយមន័យខាងលើ ស្ថិតិត្រូវបានចែកជាពីរផ្នែក៖

- ស្ថិតិបែបពិពណ៌នា ឬ រិយាយ (descriptive statistics)៖ សិក្សាអំពីការប្រមូល រៀបចំ បង្ហាញ វិភាគទៅលើសំណុំទិន្នន័យណាមួយ។
- ស្ថិតិបែបសន្និដ្ឋាន (inferential statistics)៖ ការប្រើប្រាស់ទិន្នន័យដែលបានប្រមូលពីក្រុមតូច ឬប៉ាន់គំរូ ដើម្បីធ្វើកាសន្និដ្ឋានអំពីក្រុមធំ ឬហៅថាចំនួនសាកល (population)។

ឧទាហរណ៍1.1 និស្សិតមហាវិទ្យាល័យពត៌មានវិទ្យានៅសាកលវិទ្យាល័យជាតិគ្រប់គ្រង នៅឆ្នាំសិក្សា 2020-2021 មានចំនួនទាំងអស់ 500នាក់។ អ្នកស្ថិតិម្នាក់ចង់ដឹងថានិស្សិតទាំង500នាក់ចំណាយជាមធ្យមក្នុងមួយឆ្នាំអស់ប៉ុន្មានដុល្លា? គាត់សំរេចចិត្តសាកសួរនិស្សិតចំនួន 50នាក់អំពីចំណាយរបស់ពួកគេ ដោយសាកសួរម្តងម្នាក់ៗ រួចធ្វើការកត់ត្រា ប្រមូល រៀបចំ សង្ខេបទិន្នន័យទាំងនេះយកមកធ្វើបទបង្ហាញ ហៅថាស្ថិតិបែបពិពណ៌នា និងធ្វើការសនិដ្ឋានអំពីការចំណាយជាមធ្យមរបស់និស្សិតទាំងអស់(500 នាក់) ដោយផ្អែកលើប៉ាន់គំរូចំនួន 50នាក់ ហៅថាស្ថិតិបែបសន្និដ្ឋាន។

1.2 **ហេតុអ្វីចាំបាច់ត្រូវការរៀនស្ថិតិ**?

ជារៀងរាល់ថ្ងៃ មនុស្សយើងត្រូវការធ្វើសេចក្តីសម្រេចចិត្តដោះស្រាយការងារដូច្នេះស្ថិតិគឺមានសារៈសំខាន់ណាស់សម្រាប់មនុស្សទូទៅ ចាប់រាប់ពីអ្នកសិក្សា រហូតដល់ថ្នាក់ដឹកនាំរដ្ឋាភិបាល ។ ជាទូទៅមនុស្សភាគច្រើនចង់បានការធ្វើសេចក្តីសម្រេចចិត្តមួយដែលត្រឹមត្រូវ ហើយដើម្បីឲ្យការធ្វើសេចក្តីសម្រេចចិត្តនោះកើត វាតម្រូវអោយធ្វើតាមក្បួនខ្នាតបែបវិទ្យាសាស្ត្រនិង ប្រាកដនិយម មិនមែនតាមបែបរវើយវាយនោះទេ។

---

[1] និយមន័យដកស្រង់ចេញពី (1/5/2021) https://www.investopedia.com/terms/s/statistics.asp

ការធ្វើសេចក្ដីសំរេចចិត្តតាមបែបប្រាកដនិយមគឺត្រូវពឹងផ្អែកលើទិន្នន័យ ហើយទិន្នន័យនោះទៀតសោត មានគ្រប់ទីកន្លែង បើមិនយល់ដឹងអំពីស្ថិតិ របៀបរបបក្នុងការប្រមូល រៀបចំ បកស្រាយនោះទេ យើងមិនអាចបំលែងទិន្នន័យទាំងនេះទៅជាពត៌មាន ដើម្បីជំនួយដល់ការធ្វើសេចក្ដីសំរេចចិត្តឲ្យមានប្រសទ្ធិភាពបានឡើយ។ លើសពីនេះទៅទៀត ស្ថិតិអាចធ្វើឲ្យយើងធ្វើការការប្រមូលទិន្នន័យដែលកើតមាននៅអតីតកាល ដើម្បីធ្វើការសេន្និដ្ឋាន កែលំអរពីស្ថានភាពបច្ចុប្បន្ន និងអាចព្យាករណ៍អំពីអនាគតកាលទៀតផង។

1.3 **វាក្យស័ព្ទមូលដ្ឋាននៃស្ថិតិ**

- អថេរ (variable) គឺជាចរិកលក្ខណៈនៃវត្ថុ ឬបុគ្គលដែលអ្នកចង់វិភាគ ឧទាហរណ៍ អថេរ ចំណាយទៅលើការផ្សាយពាណិជ្ជកម្ម អាយុ ឬពិន្ទុរបស់និស្សិត។ល។
- ទិន្នន័យ (data) គឺជាតម្លៃដែលទាក់ទង ឬនៅជាប់ជាមួយអថេរ។ ឧទាហរណ៍ ការចំណាយទៅលើការផ្សាយពាណិជ្ជកម្ម (ជាអថេរ ហើយមានតែជាឈ្មោះតែប៉ុណ្ណោះ) គិតជាដុល្លាប្រចាំខែ (ទិន្នន័យជាខែ) អាយុគិតជាឆ្នាំ ទម្ងន់គិតជាគីឡូក្រាម។ល។
- និយមន័យប្រតិបត្តិការរបស់អថេរ (operational variable definition) សំដៅទៅលើវិធីដែលអ្នកកំណត់អត្ថន័យ និងវាស់វែងអថេរជាក់លាក់ណាមួយដែលអ្នកយកមកប្រើក្នុងការសិក្សារបស់អ្នក។ បើពុំមានកំណត់និយមន័យឲ្យបានច្បាស់លាស់នោះទេ វាអាចធ្វើអោយមានការយល់ច្រឡំអំពីអត្ថន័យរបស់អថេរមួយ ទៅអថេរមួយទៀត ពីព្រោះអថេរមួយអាចមានន័យលើសពីមួយ ឬពាក្យនោះមានន័យមិនច្បាស់លាស់ ។ ឧទាហរណ៍ កូនក្មេងចង់សំដៅលើមនុស្សដែលមានអាយុចន្លោះពី 6-12 ឆ្នាំ ហើយអ្នកមានការងារធ្វើគឺជាអ្នកដែលមានការងារធ្វើជាប់លាប់ 8ម៉ោងក្នុងមួយថ្ងៃ ឬ 48ម៉ោងក្នុងមួយសប្ដាហ៍យ៉ាងតិច។ល។
- ចំនួនសកល (population) សំដៅទៅលើក្រុមមនុស្សទាំងមូល វត្ថុទាំងអស់ ឬព្រឹត្តិការណ៍ទាំងមូល។ វាគឺជាក្រុមធំដែលអ្នកសិក្សាចង់សន្និដ្ឋាន ឬប៉ាន់ស្មាន។ ឧទាហរណ៍ និស្សិតMBAថ្នាក់សៅរ៍ អាទិត្យទាំងអស់នៅ NUM មានចំនួន 800នាក់នៅឆ្នាំសិក្សា 2020-2021។ ពាណិជ្ជករទាំងអស់នៅភ្នំពេញនៅឆ្នាំ 2021 មាន1លាននាក់។ល។
- ប៉ាន់គំរូ ឬ សំណាក(sample) គឺជាផ្នែកមួយរបស់ចំនួនសកលដែលជ្រើសរើសមកធ្វើការវិភាគ ។ វាតំណាងអោយក្រុមតូច ឧទាហរណ៍ និស្សិតMBAថ្នាក់សៅរ៍ អាទិត្យទាំងអស់នៅ NUMមានចំនួន 800នាក់នៅឆ្នាំសិក្សា 2020-2021 យើងសាកសួរចំនួន100នាក់ធ្វើជាប៉ាន់គំរូ។
- ប៉ារ៉ាម៉ែត្រ (parameter) គឺជារង្វាស់ជាលេខដែលបរិយាយអំពីចំនួនសកល
- ស្ថិតិ (statistics) គឺរវាស់ជាលេខដែលបរិយាយអំពីប៉ាន់គំរូ។

ឧទាហរណ៍ 1.2 យើងចង់សិក្សា ចំណូល (income) របស់ CEO ធនាគារទាំងអស់នៅប្រទេសកម្ពុជា នៅឆ្នាំ 2020 ដែលមានចំនួន 26 នាក់ មានឈ្មោះ A, B, C ដល់ Z (ជាចំនួនសកល) ហើយយើងសម្ភាសតែ 10នាក់(ជាប៉ាន់គំរូ) ដោយចៃដន្យ។ ចំណូលជាមធ្យមរបស់CEOទាំង26នាក់វាស់ដោយ (μ) ហើយចំណូលរបស់ប៉ាន់គំរូ 10នាក់វាស់ដោយ $\bar{x}$ ។ តារាង 1.1 បង្ហាញអំពីឧទាហរណ៍ រង្វាស់ផ្សេងៗនៃប៉ារាម៉ែត្រនិងស្ថិតិ។

តារាង 1.1 រង្វាស់ផ្សេងៗនៃចំនួនសកលនិងស្ថិតិ

Population	Sample
A B C D E F G H I J K L M N O P Q R S T V W X Y Z	A D G H K O R T W Y
រង្វាស់ប៉ារ៉ាម៉ែត្រ	**រង្វាស់ស្ថិតិ**
មធ្យម (μ)	មធ្យម ($\bar{x}$)
វ៉ារីយ៉ង់ (σ^2)	វ៉ារីយ៉ង់ (S^2)
គម្លាតស្តង់ដារ (σ)	គម្លាតស្តង់ដា (S)

1.4 ប្រភេទនៃអថេរ

ជាទូទៅអថេរត្រូវបានចែកចេញជាពីរប្រភេទ៖ អថេរជាគុណវិស័យ និងអថេរបរិមាណវិស័យ ។

- អថេរជាប្រភេទ ឬគុណវិស័យ (categorical or qualitative variable) គឺជាអថេរដែលមិនមានតម្លៃជាលេខ ឧទាហរណ៍ អថេរភេទ (gender variable) គឺមាន ស្រី/ប្រុស អថេរចំលើយ (answer variable) មាន៖ yes/no។
- អថេរបរិមាណវិស័យ (numerical or quantitative variable) គឺជាអថេរដែលមានតម្លៃជាលេខ ហើយអថេរនេះត្រូវបានបែងចែកជាពីរគឺ៖ លេខចំនួនគត់ឬអថេរដាច់ និងលេខដែលមានក្បៀសឬអថេរជាប់។
    - អថេរដាច់ ឬជាលេខចំនួនគត់ (discrete variable) ឧទាហរណ៍ ចំនួនកូននៅក្នុងបន្ទុក ឬចំនួនទូស័ព្ទ ចំនួនបន្ទប់កេងរបស់សណ្ឋាគារ។
    - អថេរជាប់ ឬជាលេខដែលអាចមានក្បៀស (continuous variable) ឧទាហរណ៍ ចំណូលរបស់គ្រូបង្រៀនគឺ $6.8/ម៉ោង ត្រីមានទម្ងន់ 2.5គីឡូ ឬចំងាយ 3.0 គីឡូម៉ែត្រពីផ្ទះទៅសាលារៀន។

1.5 រង្វាស់អថេរ

ជាទូទៅរង្វាស់អថេរគឺអាស្រ័យទៅនឹងប្រភេទនៃអថេរ៖

- បើសិនជាអថេរនោះជាអថេរដាច់ឬប្រភេទ (categorical variable) យើងអាចប្រើរង្វាស់ ណូមីណាល់ស្កេល (nominal scale) ឬប្រើរង្វាស់អរឌីណលស្កេល (ordinal scale)
- បើសិនជាអថេរជាប់ឬជាលេខ (continuous variable) យើងអាចប្រើរង្វាស់ ចន្លោះ (interval) ឬ អនុបាត (ratio) (មើលឧទាហរណ៍ក្នុងតារាង 1.4)

1. ណូមីណាល់ស្កេល (nominal scale) គឺត្រូវបានប្រើប្រាស់ដើម្បីវាស់វែងទៅលើអថេរប្រភេទ (categorical variable) ហើយវាបែងចែកទិន្នន័យជាប្រភេទ មិនមានលំដាប់លំដោយ ឧទាហរណ៍ ប្រុស/ ស្រី ឬ ស្រី ប្រុស ក៏បាន មិនប្រកាន់លំដាប់លំដោយនោះទេ។ ណូមីណាល់ស្កេល ត្រូវបាន

ចាត់ទុកថាជារង្វាស់ខ្សោយជាងគេក្នុងចំណោមរង្វាស់ទាំងបួន (មើលឧទាហរណ៍បន្ថែមនៅ តារាង 1.2)។

តារាង 1.2 ឧទាហរណ៍ រង្វាស់ណូមីណាល់ស្កេល

អថេរដាច់ឬប្រភេទ	រង្វាស់ ណូមីណាល់ស្កេល
ភេទ	ប្រុស / ស្រី
បោះឆ្នោត	គាំទ្រ/ ប្រឆាំង
ចម្រៀង	សម័យ/បុរាណ
សាសនា	ពុទ្ធ/ គ្រឹស/អុីស្លាម
សាំង	ធម្មតា / ស៊ុបពែ

2 អរឌីណាលស្កេល(ordinal scale) បានបែងចែកទិន្នន័យទៅតាមប្រភេទដូច ណូមីណាល់ស្កេលដែរ តែមានឬប្រកាន់លំដាប់លំដោយ។ រង្វាស់នេះខ្លាំងជាងរង្វាស់ ណូមីណាល់ស្កេល ក៏ប៉ុន្តែវានៅតែជាទម្រង់ខ្សោយដដែល ពីព្រោះថាតម្លៃខុសគ្នារវាងលំដាប់នីមួយៗគឺគ្មានន័យ ឧទាហរណ៍ ការឲ្យពិន្ទុលើផលិតផលមានបីកំរិតៈ អន់/ ល្មម/ ល្អ តម្លៃខុសគ្នារវាង អន់ និងល្មម គឺមិនអាចបកស្រាយបាន ឬគ្មានន័យ (មើលឧទាហរណ៍បន្ថែមនៅក្នុង តារាង 1.3)។

តារាង 1.3 ឧទាហរណ៍ រង្វាស់តាមលំដាប់លំដោយ

អថេរប្រភេទ	រង្វាស់តាមលំដាប់លំដោយ
និទ្ទេស	A / B/ C/ D/ E/ F
សារៈសំខាន់	មិនសំខាន់/ សំខាន់/ សំខាន់ណាស់
និស្សិតឆ្នាំទី	1/ 2/ 3/ 4
ពេញចិត្តចំពោះផលិតផល	មិនពេញចិត្ត/ គ្មានចំលើយ/ពេញចិត្ត
ការអោយពិន្ទុលើផលិតផល	អន់/ ល្មម/ ល្អ

3 រង្វស់ជាចន្លោះ (interval scale) គឺជារង្វាស់ជាលេខដែលមានលំដាប់លំដោយ និងមានចន្លោះថេរឬស្មើគ្នា ហើយតម្លៃខុសគ្នារវាងលំដាប់នីមួយៗគឺមានន័យពេញលេញ ប៉ុន្តែមិនមានតម្លៃសូន្យនោះទេ ឬសូន្យមានន័យមិនពិត។ ឧទាហរណ៍ សីតុណ្ហភាពខុសគ្នារវាង 30អង្សានិង 40អង្សា គឺមិនដូចគ្នារវាងសីតុណ្ហភាព 50អង្សា និង60អង្សា ពីព្រោះថាចន្លោះ 10អង្សាខុសគ្នានោះ គឺមានន័យខុសគ្នាក្នុងករណីទាំងពីរ។ ថ្វីបើចន្លោះ 10 អង្សាដូចគ្នាតែ10 អង្សានៅចន្លោះពី50ទៅ60អង្សាគឺក្តៅជាង ។ រង្វាស់ជាចន្លោះ មិនមែនថាល្អឥតខ្ចោះនោះទេ ពីព្រោះលេខសូន្យគឺនៅតែគ្មានន័យ ឧទាហរណ៍ 0អង្សាមិនមែនមានន័យថា មិនមានសីតុណ្ហភាព ឬទទេរនោះទេ នេះជាចំណុចសខ្សោយរបស់រង្វស់ជាចន្លោះ។ ជាភ័ព្វសំណាងល្អ ដែលនៅមានរង្វស់មួយទៀតមកបំពេញ ដូចជារង្វាស់ជាលេខឬអនុបាត (ratio) ដែលក្នុងនោះលេខសូន្យមាន

ន័យពេញលេញ សូន្យមានន័យថា គ្មាន ឬទទេរ ឧទាហរណ៍ ចំណូលស្មើសូន្យ មានន័យ គ្មានលុយ ដែលវាខុសពីសីតុណ្ហភាពសូន្យ ដែលបានរៀបរាប់ខាងលើ។

4 រង្វាស់ជាលេខអនុបាត (ratio) គឺរង្វាស់មានលំដាប់លំដោយ ខ្លាំងជាងរង្វាស់ផ្សេងៗទាំងអស់ហើយរង្វាស់ខុសគ្នាគឺមានលេខសូន្យ ឬលេខសូន្យមានន័យពិតប្រាកដ ឧទាហរណ៍ ចំណាយប្រចាំឆ្នាំរបស់និស្សិតគិតជាមធ្យមគឺ 300ដុល្លា បើចំណាយស្មើសូន្យគឺមានន័យថា គ្មានចំណាយ (មើលឧទាហរណ៍បន្ថែមនៅក្នុងតារា 1.4)។

តារាង 1.4 ឧទាហរណ៍ រង្វាស់ចន្លោះ (interval) និង លេខ (ratio)

អថេរជាប់ ឬជាលេខ (numerical variable)	រង្វាស់ (measurement)
សីតុណ្ហភាព	ចន្លោះ (interval)
ពិន្ទុ (និទ្ទេស) ឧ. 95-100=A	ចន្លោះ (interval)
ប្រាក់ខែ	ដុល្លា (ratio)
អាយុ	ឆ្នាំ (ratio)
ចំងាយ	ម៉ែត្រ ឬគីឡូម៉ែត្រ (ratio)
ទម្ងន់	គីឡូក្រាម (ratio)

1.6 **កម្មវិធីសម្រាប់គណនាស្ថិតិ**

រយៈពេល 5ទសវត្សរ៍ចុងក្រោយមានកម្មវិធីកុំព្យូទ័រ(software) ជាច្រើនត្រូវបានបង្កើតឡើងដើម្បីជំនួយដល់ការអនុវត្តន៍ស្ថិតិ។ ក្នុងចំណោមកញ្ចប់កម្មវិធីកុំព្យូទ័រជាច្រើនដូចជា Minitab, SAS, Eviews, R, Stata, SPSS និង កម្មវិធីកុំព្យូទ័រដ៏ទៃទៀត គេអង្កេតឃើញថា Minitab ត្រូវបានគេប្រើប្រាស់ជាងគេនៅតាមមហាវិទ្យាល័យ។ សម្រាប់សៀវភៅនេះ កម្មវិធី Excel, Eviews,និង SPSS ត្រូវបានយកមកជាឧបករណ៍ជំនួយដល់លើការអនុវត្តន៍ស្ថិតិ។

1.7 **របៀបប្រើប្រាសសៀវភៅនេះ**

សៀវភៅនេះត្រូវបានរៀបចំឡើងជាបួនផ្នែក។ ផ្នែកទីមួយមានជំពូកទី2 និង ទី3 បង្ហាញអំពីការបរិយាយ និងសង្ខេបទិន្នន័យ ផ្នែកទី2 មានជំពូកទី4 និង ទី5 បង្ហាញអំពីការប្រមូលទិន្នន័យ ផ្នែកទី3 ចាប់ពីជំពូកទី6 ដល់ទី9 បង្ហាញអំពីការធ្វើគ្គេសលើសម្មតិកម្ម និងផ្នែកទី4 ឬជំពូកទី 10 បង្ហាញអំពីការព្យាករណ៍។

ដោយហេតុថា សៀវភៅនេះជាសៀវភៅបែបអនុវត្តន៍ នៅបញ្ចាប់នៃជំពូកនីមួយៗគឺមានលំហាត់អនុវត្តន៍ និងមានចំឡើយខ្លីៗដែលបង្កលក្ខណៈងាយស្រួលដល់និស្សិតក្នុងការសិក្សា។

ការសិក្សាស្ថិតិត្រូវចាប់ផ្តើមពីការសិក្សាពាក្យបច្ចេកទេសជាមុន។ ពាក្យទាំងនោះអាចឲ្យអ្នកសិក្សាងាយយល់អំពីស្ថិតិនិងអាចជួយឲ្យអ្នកធ្វើសេចក្តីសំរេចចិត្តអោយកាន់តែប្រសើរឡើង។ អាជីវកម្មប្រើប្រាស់ស្ថិតិដើម្បីសង្ខេប ធ្វើសន្និដ្ឋានចេញពីទិន្នន័យ ដើម្បីកែលំអរលើនីតិវិធីជំនួញ និងដើម្បីធ្វើការព្យាករណ៍ដែលអាចទុកចិត្តបាន។ នៅពីជំពូកបន្ទាប់នេះ អ្នកនិងសិក្សាអំពីដំណើរការរៀបចំ និងស្ថិតិបែបបរិយាយ៕

សំណួរនិងលំហាត់

I. សំណួរ

1. អ្វីទៅជាភាពខុសគ្នារវាងប៉ាន់គំរូនិងចំនួនសកល?

2. អ្វីទៅជាភាពខុសគ្នារវាងស្ថិតិនិងប៉ារ៉ាម៉ែត្រ?

3. អ្វីទៅជាភាពខុសគ្នារវាងស្ថិតិបែបបរិយាយនិងបែបសន្និដ្ឋាន?

4. អ្វីទៅជាភាពខុសគ្នារវាងអថេរប្រភេទនិងអថេរជាលេខ?

5. អ្វីទៅជាភាពខុសគ្នារវាងអថេរដាច់និងអថេរជាប់?

6. អ្វីទៅជានិយមន័យប្រតិបត្តិការ? ហេតុអ្វីបានជានិយមន័យប្រតិបត្តិការមានសារៈសំខាន់

7. រង្វាស់ទាំងបួនមានអ្វីខ្លះ?

II. **សំណួរជ្រើសរើស** (multiple choice)

8. ចូលកំណត់អថេរនីមួយខាងក្រោមតើជាអថេរប្រភេទឬជាលេខ។ ប្រសិនជាលេខចូលកំណត់ថាលេខគត់ឬមានក្បៀស។ លើសពីនេះទៅទៀតចូលចំណត់ថាជារង្វាស់ជាស្កេល (scale)។

a. ទូរស័ព្ទក្នុងមួយគ្រួសារ

b. រយៈពេលហៅទូស័ព្ទ (គិតជានាទី) ក្នុងមួយខែ។

c. ចំនួនផ្ទះមាន វ៉ាយហ្វាយ

d. ល្បឿនអ៊ីនធឺណែត

9. ចំពោះអថេរខាងក្រោម ចូលកំណត់ថាតើជាអថេរប្រភេទ ឬជាលេខ។ បើជាលេខចូលកំណត់ថាជាលេខគត់ឬលេខមានក្បៀស ហើយលើសពីនេះទៅទៀត ចូលកំណត់ថាជារង្វាស់ស្កេល (scale) ចំពោះអថេរនិមួយៗ។

a. ឈ្មោះក្រុមហ៊ុនផ្គត់ផ្គង់អ៊ីនធឺណែត

b. រយៈពេល(ម៉ោង) ចំណាយក្នុងការប្រើអ៊ីធឺណែតក្នុងមួយសប្តារណ៍

c. ចំនួនសារអេលេចត្រូនិចទទួលបានក្នុងមួយសប្តារណ៍

d. ចំនួនដងទិញទំនិញលើបណ្តាញក្នុងមួយខែ

10. សន្មត់ថាពត៌មានខាងក្រោមគឺត្រូវបានប្រមូលពីពូសុខក្នុងការដាក់ពាក្យសុំខ្ចីលុយទិញផ្ទះជាមួយធនាគារ អេស៊ីលីដា។

a. សងឬទូទាត់ប្រចាំខែគឺ 500 ដុល្លា

b. ចំនួនការងារដែលបានធ្វើរយៈពេល10ឆ្នាំចុងក្រោយៈ 1

c. ចំណូលប្រចាំឆ្នាំរបស់គ្រួសារៈ 16,000 ដុល្លា

d. ស្ថានភាពគ្រួសារៈ រៀបការ

ជំពូកទី 2 ការរៀបចំទិន្នន័យ

(Organizing Data)

គោលបំណង

- ស្វែងរកប្រភពទិន្នន័យដែលប្រើប្រាស់ក្នុងជំនួញនិងសេដ្ឋកិច្ច
- បង្កើតគំនូសតាង និងតារាងចំពោះអថេរជាប់ឬជាលេខ
- បង្កើតគំនូសតាង និងតារាងចំពោះអថេរប្រភេទ ឬអថេរដាច់
- ពន្យល់អំពីគំនូសតាងនីមួយៗ
- ពន្យល់អំពីគោលការណ៍ក្នុងការប្រើគំនូសតាង

2.1 ទិន្នន័យនិងប្រភពទិន្នន័យ

ទិន្នន័យគឺជាពត៌មានដែលទាក់ទងជាមួយអង្គហេតុដែលអាចចេញមកពីប្រភពពីរ ៖ប្រភពទីមួយ (primary data) និងប្រភពទីពីរ (secondary data)។

- ប្រភពទីមួយ អាចប្រមូលតាមរយៈធ្វើការសម្ភាស (interview) ការស្ទង់មតិ (survey) ការធ្វើពិសោធន៍ ឬការធ្វើការសាកល្បង (experiment) ឬអង្កេត (observation) ជាដើម។
- ប្រភពទីពីរ ជាប្រភេទទិន្នន័យដែលគេបានចងក្រង ឬ បានផ្សព្វផ្សាយតាមរយៈទស្សនាវដ្តី ការសែត ឬ អ៊ីនតឺណែតជាដើម ឧទាហរណ៍ ទិន្នន័យពីការធ្វើជំរឿន តម្លៃភាគហ៊ុន ឬ GDP ជាដើម។

ចំណុចដែលខុសគ្នារវាងប្រភពទីមួយនិងប្រភពទីពីរគឺថា៖ ប្រភពទីមួយគឺជាពត៌មានដែលទាក់ទងជាមួយបច្ចុប្បន្នភាពរបស់អង្គហេតុ និងដោយផ្ទាល់ ហើយប្រភពទីពីរគឺជាពត៌មានដែលទាក់ទងជាមួយអតីតកាលហើយមានលក្ខណៈប្រយោលទៅនឹងប្រភពដើមនៃទិន្នន័យ។

2.2 ការរៀចំទិន្នន័យអថេរប្រភេទ ឬដាច់

ជាទូទៅអថេរបានចែកចេញជាពីរប្រភេទ (1) អរថេរគុណភាព (qualitative variable) មិនមែនជាលេខ ឬជាអថេរប្រភេទឬដាច់ (categorical variable) និង (2) អថេរបរិមាណ(quantitative variable) ឬ អថេរជាលេខ ហើយអថេរនេះត្រូវបានបែងចែកជាពីរបន្តទៀត គឺ អថេរលេខគត់ (discrete variable) និងអថេរជាប់ (continuous variable) ឬអថេរបន្ត។ ជាគោលការណ៍ក្នុងការរៀបចំទិន្នន័យប្រភេទត្រូវបានចែកចេញជាពីរ អាស្រ័យទៅនឹងចំនួនអថេរ។ បើសិន ទិន្នន័យមានតែមួយអថេរប្រភេទ (one categorical variable) យើងអាចរៀបចំទិន្នន័យជាតារាងសង្ខេប (summary table) ហើយបើអថេរមានចំនួនពីរឬច្រើនយើងត្រូវរៀបចំជាតារាងខ្វែង (contingency or cross table)។

2.3 អថេរប្រភេទដែលមានមួយអថេរ (one categorical variable)

អថេរប្រភេទដែលមានមួយអថេរ ត្រូវបានរៀបចំជាតារាងសង្ខេប ដែលបង្ហាញអំពីបរិមាណចំនួនហ្រ្វេកង់ និងភាគរយនៃប្រភេទនីមួយៗ ហើយធ្វើអោយយើងអាចញែកអោយដាច់ពីគ្នាថាប្រភេទនីមួយៗមានចំនួនហ្វ្រេកង់ប៉ុន្មាន និងភាគរយប៉ុន្មាន ។

ឧទាហរណ៍ 2.1 អ្នកស្រាវជ្រាវម្នាក់បានធ្វើការស្ទង់មតិទៅលើអតិថិជន100នាក់ ដោយចង់ដឹងអំពីចំណាង់ចំណូលចិត្តចំពោះការប្រើប្រាស់សេវាធនាគារអេស៊ីលីដា។ ហើយបានលទ្ធផលដូចខាងក្រោះ

តារាង 2.1 ទិន្នន័យចំណង់ចំណូលចិត្តក្នុងប្រើប្រាស់សេវាធនាគារ

អថេរប្រភេទ	ចំនួនអតិថិជន
អេធីអឹម	30
ទូរស័ព្ទ	50
ទៅធនាគារដោយផ្ទាល់	15
ផ្សេងៗ	5
សរុប	100

នេះជាអថេរប្រភេទ យើងមិនអាចបូកប្រភេទនីមួយចូលគ្នារួចគិតជាមធ្យមនោះទេ ពីព្រោះផលបូកអថេរប្រភទពិបាកបកស្រាយ យើងអាចវិភាគតាមរយៈហ្វ្រេកង់នៃអថេរនីមួយ និងភាគរយតែប៉ុណ្ណោះ ។

ចំនួនអតិថិជនប្រើប្រាស់ ATM មាន30នាក់ ឬស្មើនឹង 30% ($= \frac{30}{100} = 30\%$) និងចំនួនអតិថិជនប្រើប្រាស់ ទូរស័ព្ទ មាន50នាក់ ឬស្មើនឹង 50% ($= \frac{50}{100} = 50\%$) ហើយអ្នកទៅធនាគារដោយផ្ទាល់មាន 15% និងអ្នកប្រើប្រាស់សេវាផ្សេងៗមាន5%។ តារាង2.2 បង្ហាញអំពីលទ្ធផលសង្ខេបរបស់SPSS នៃការគណនានេះ។

តារាង2.2ហ្វ្រេកង់ចំណង់ចំណូលចិត្តចំពោះសេវាធនាគារ

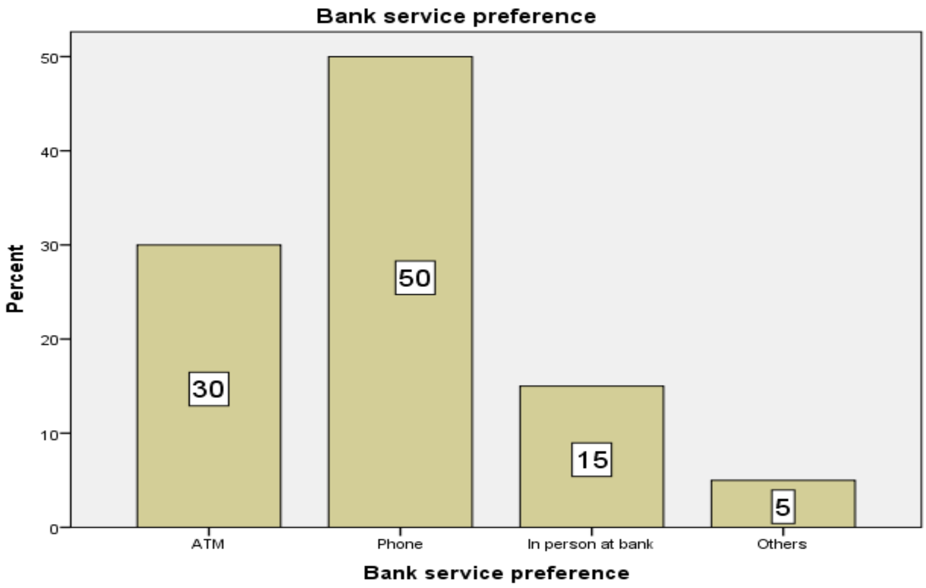
	Frequency	Percent	Valid Percent	Cumulative Percent
ATM	30	30.0	30.0	30.0
Phone	50	50.0	50.0	80.0
In person at bank	15	15.0	15.0	95.0
Others	5	5.0	5.0	100.0
Total	100	100.0	100.0	

គួរគត់សំគាល់ថា អតិថិជនចូលចិត្តសេវាធនាគារតាមរយៈទូរស័ព្ទ ច្រើនជាងគេរហូតដល់ 50% ហើយបន្ទាប់មកគឺសេវា ATM ដែលតំណាង30%។ សេវាទាំងពីរនេះតំណាងអោយ 80% នៃសេវាធនាគារទាំងអស់ (សូមមើលកូឡោនហ្វ្រេកង់គ្រួត "Column cumulative frequency" នៃតារាង2.2)។

តារាងសង្ខេប2.2អាចស្មុគស្មាញ ពិបាកមើលបន្តិច ជំរើសមួយទៀតយើងអាចយកលទ្ធផលនេះដាក់បង្ហាញលើគំនូសតាងក្រាហ្វិកម្តង។ ហើយគំនូសតាងក្រាហ្វិកដែលប្រើប្រាស់

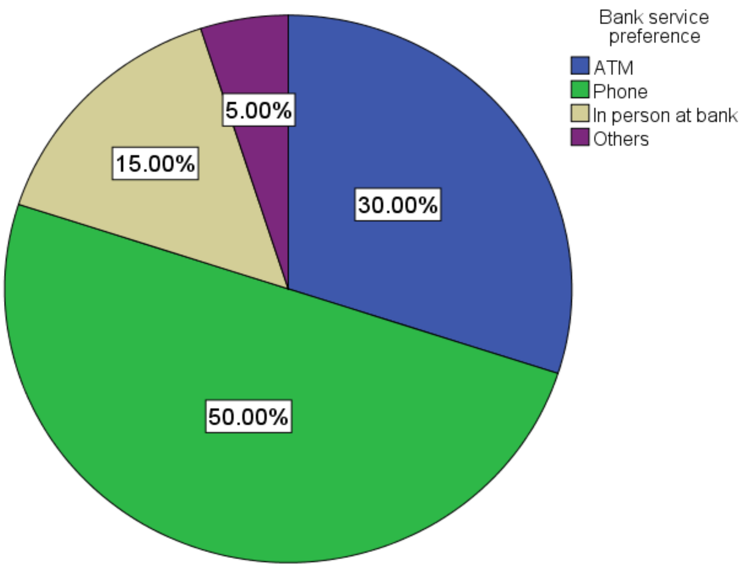
សម្រាប់អថេរប្រភេទដែលមានមួយអថេរនេះ (one categorical variable) គឺមានដូចជា បាឆាត (Bar Chart) ផាយឆាត (Pie Chart) និង ប៉ារាតូឆាត (Pareto Chart) ជាដើម។

រូបភាព 2.1 បាឆាត (Bar Chart) តំណាងចំណង់ចំណូលចិត្តចំពោះសេវាធនាគារ

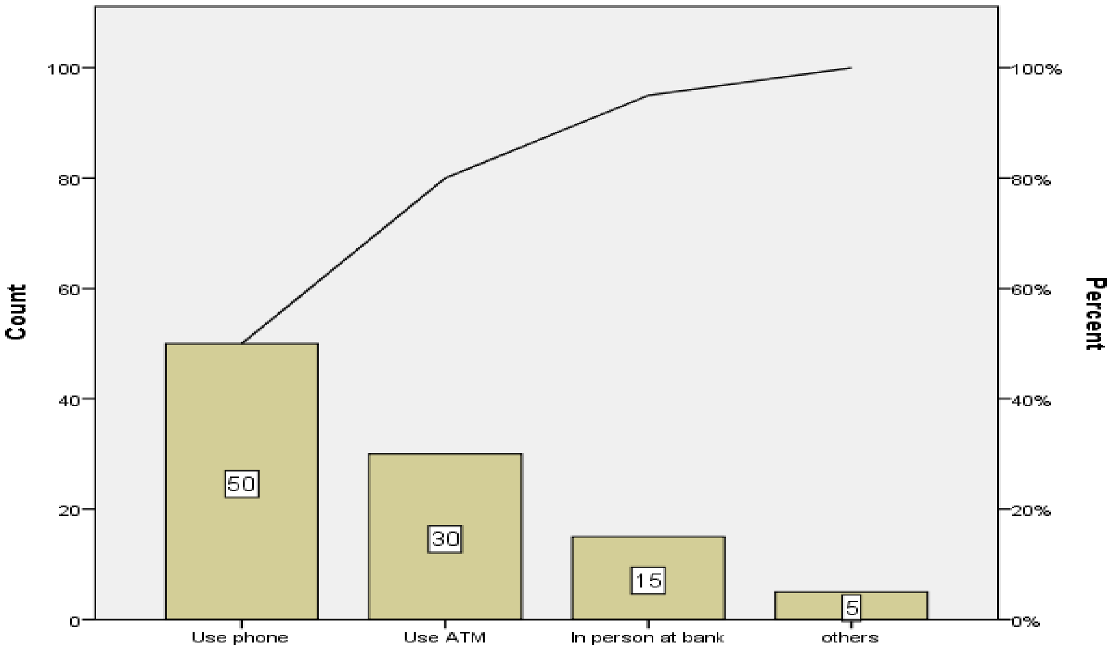


ចរិកលក្ខណៈនៃគំនូសតាង បាឆាត (រូប 2.1) គឺ អ័ក្ស Y តំណាងអោយចំនួនប្រេកង់ និងអ័ក្ស X តំណាងអោយប្រភេទសេវា ហើយបន្ទះនីមួយៗមិនជាប់គ្នានោះទេ។ បាឆាតធ្វើអោយអ្នកអានងាយយល់ បើប្រៀបធៀបទៅនឹងតារាងសង្ខេប 2.2 ខាងលើ។ ដោយសារតែវាជាអថេរប្រភេទដែលមានមួយអថេរ យើងអាចប្រើ ផាយឆាត (Pie Chart) មកប្រើជំនួស បាឆាតក៏បានដែរ (មើលរូប 2.2) ចំណែកការបកស្រាយគំនូសទាំងពីរគឺដូចជាមួយការបកស្រាយតាមតារាងសង្ខេប2.2ដែរ។

រូបភាព 2.3 Pie Chart តំណាងចំណង់ចំណូលចិត្តចំពោះសេវាធនាគារ



រូបភាព 2.4 ប៉ារ៉ាតូ (Pareto Chart) តំណាងចំណង់ចំណូលចិត្តទៅលើសេវាធនាគារ



បើយើងធ្វើការប្រៀបធៀបគំនូសតាងទាំងបី ប៉ារ៉ាតូឆាត មានលក្ខណៈពេញលេញជាងបាឆាត និង ផាយឆាត ។ ប៉ារ៉ាតូឆាត មានលក្ខណៈដូចទៅនិង បាឆាត ច្រើន តែខុសបន្តិច ត្រង់ថា ប៉ារ៉ាតូឆាត មានបញ្ជាក់អំពី ហ្រេ្វកង់ត្រួត។

ក្នុងការអនុវត្តន៍ជាក់ស្តែង អ្នកអាចជ្រើសរើសយកគំនូសតាងណាមួយក៏បាន ប៉ុន្តែត្រូវគួបផ្សំនឹងតារាងសង្ខេបពីព្រោះតារាងសង្ខេបមានបង្ហាញអំពីពត៌មានលំអិតជាងគំនូសតាងក្រាហ្វិកទាំងបីខាងលើ។

2.2.2 **អថេរប្រភេទដែលមានពីរឬច្រើនអថេរ**

ការសិក្សាមួយចំនួនអាចទាក់ទងជាមួយអថេរប្រភេទពីរ ឬ ច្រើន។ ក្នុងស្ថានភាពបែបនេះ យើងត្រូវប្រើប្រាស់តារាងខ្វែង ឬតារាង2អរថេរ(cross or contingency table)។ តារាងនេះតំណាងឲ្យចំឡើយរួមគ្នារវាងអថេរប្រភេទទីមួយ។ អថេរមួយស្ថិតនៅជួរដេក (row) និងអថេរមួយទៀតស្ថិតនៅជួរឈរ (column) ហើយកន្លែងប្រសព្វរវាងជួរទាំងពីរហៅថាសែល (cell) គឺតំណាងឲ្យចំឡើយរួម។

ឧទាហរណ៍ 2.2 កសិករម្នាក់បានជ្រើសរើសប៉ាន់គំរូកូនស្វាយចំនួន400ដើម ដោយមានមួយចំនួននៅរស់ ហើយមួយចំនួនក៏ងាប់ ឬបាត់ បន្ទាប់ពីដាំរួចបានមួយឆ្នាំ ហើយគាត់ចង់ដឹងភាគរយនៃកូនស្វាយដែលនៅរស់ និងបាត់បង់។ ឧបមាថា គាត់ធ្វើការបែងចែកកូនស្វាយ ជាបីប្រភេទទៅតាមអាយុ៖

- តិចជាងកន្លះឆ្នាំ មានសរុបចំនួន190 ដើម នៅរស់170ដើម ហើយងាប់ឬបាត់បង់ 20ដើម
- ធំជាងនកន្លះឆ្នាំ តែតិចជាងមួយឆ្នាំ មានសរុបចំនួន140 ដើម នៅរស់100ដើម ហើយងាប់ឬបាត់បង់ 40ដើម

- ធំជាងមួយឆ្នាំ តែតិចជាងមួយឆ្នាំកន្លះ មានសរុបចំនួន65 ដើម នៅរស់60ដើម ហើយងាប់ឬបាត់បង់5 ដើម។

ទិន្នន័យទាំងនេះបានបង្ហាញនៅក្នុងតារាង 2.3។ ជួរដេក (row) តំណាងអោយអាយុកូនស្វាយ រីឯជួរឈរ (column) តំណាងអោយចំនួនដើមស្វាយដែលនៅរស់និងងាប់។

តារាង 2.3 ចំនួនកូនស្វាយដែលនៅរស់ឬបាត់ និងអាយុកូនស្វាយ

កូនស្វាយ	នៅរស់	ងាប់ ឬបាត់	**សរុប**
អាយុ < 0.5 ឆ្នាំ	170	20	190
0.5 <=អាយុ<1.0 ឆ្នាំ	100	40	140
1.0<= អាយុ<1.5 ឆ្នាំ	65	5	70
សរុប	335	65	400

តារាង 2.3 ហៅថាតារាងខ្វែង (cross or contingency table) វាតំណាងឲ្យអថេរប្រភេទដែលមានពីរអថេរ។ លទ្ធផលនៃការគណនាភាគរយក្នុងឧទាហរណ៍នេះត្រូវបានបង្ហាញនៅក្នុងតារាង2.4 និងរូបភាព2.4 ខ្លាស់ស្ទើបាឆាត (Clustered Bar Chart)។

តារាង 2.4 តារាងខ្វែងបង្ហាញភាគរយនៃដើមស្វាយដែលរស់និងបាត់

Age of Tree * Survival or loss of Tree Cross Tabulation					
			Survival or loss		Total
			Lost	Survival	
Age	Age<0.50 year	Count	20	170	190
		% of Total	5.0%	42.5%	47.5%
	0.50<=Age<1.0 year	Count	40	100	140
		% of Total	10.0%	25.0%	35.0%
	1.0<= Age <1.5 year	Count	5	65	70
		% of Total	1.2%	16.2%	17.5%
Total		Count	65	335	400
		% of Total	16.2%	83.8%	100.0%

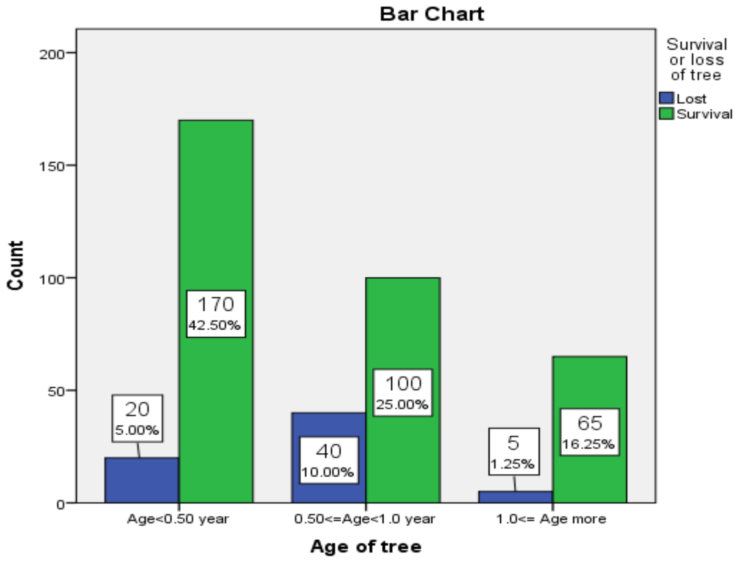
តារាង 2.4 បង្ហាញលទ្ធផលដូចតទៅ៖

- កូនស្វាយដែលមានអាយុតិចជាង 0.50 ឆ្នាំដែលនៅរស់មានចំនួន 170 ដើម ក្នុងចំណោមដើមស្វាយសរុបចំនួន 400ដើម ឬ ស្មើ 170/400=42.5%
- កូនស្វាយដែលមានអាយុតិចជា 0.50ឆ្នាំដែលបាត់ ស្មើ 20/400 =20%

ចំណែកឯភាគរយនៃកូនស្វាយដ៏ទៃទៀតត្រូវបានគណនាដោយវិធីសាស្ត្រដូចគ្នាដោយយកចំនួនកូនស្វាយនៅក្នុងប្រអប់ខ្វែងរវាងជួរដេកនិងជួរឈរចែកនិងចំនួនសរុប 400ដើម។

គួរកត់សំគាល់ថា បរិមាណដើមស្វាយដែលនៅរស់សរុបមាន 83.8% ហើយដើមស្វាយដែលបាត់ឬងាប់មាន16.2%។ កូនស្រាយដែលងាប់ច្រើនជាងគេគឺមានអាយុចន្លោះចាប់ពី 0.50ទៅដល់ 1.0 ឆ្នាំ ហើយមានរហូតដល់ 10% ឬ40ដើម ហើយកូនស្វាយដែលជាំរស់ច្រើនជាងគេគឺមានអាយុតិចជាង 0.50ឆ្នាំ ហើយមានរហូតដល់ 42.5% ឬ 170ដើម។ ដើម្បីឲ្យងាយយល់ជាងនេះ យើងអាចយកទិន្នន័យទាំងអស់ ដាក់បង្ហាញលើគំនូសតាង ខ្លាស់ស្នើបាធាត ឬសាយបាយសាយបាធាត (Clustered or Side-by Side Bar Chart)ដូចនៅក្នុងរូប 2.4។

រូប 2.4 ខ្លាស់ស្នើបាធាត បង្ហាញភាគរយនៃដើមស្វាយដែលរស់និងបាត់



2.3 ការរៀបចំទិន្នន័យជាលេខ

ដើម្បីចាត់ចែងទិន្នន័យជាលេខ យើងអាចរៀបចំទិន្នន័យនោះតាមលំដាប់លំដោយ (ordered array) ហើយរួចរៀបចំជាបំណែងចែកហ្វ្រេកង់ (frequency distribution) និងបំណែងចែកហ្វ្រេកង់គ្រួត (cumulative frequency distribution)។

2.3.1 លំដាប់លំដោយ និងបំណែងចែកហ្វ្រេកង់ចំពោះទិន្នន័យជាលេខ

សូមបញ្ជាក់ថា ទិន្នន័យជាលេខ ក៏អាចធ្វើការបែងចែកជាហ្វ្រេកង់បានដែរ តែយើងត្រូវរៀបចំទិន្នន័យនោះទៅតាមលំដាប់លំដោយជាមុនសិន។

ឧទាហរណ៍ 2.3 អ្នកស្រាវជ្រាវរបស់ផ្សារអ៊ីអន2 ចង់ធ្វើការសិក្សាលើចំណង់ចំណូលចិត្តរបស់អតិថិជនចំពោះផលិតផលមួយប្រភេទ ហើយអ្នកស្រាវជ្រាវជឿជាក់ថា ចំណង់ចំណូលចិត្តរបស់មនុស្សគឺមានទំនាក់ទំនងទៅនឹងអាយុ។ ក្មេងអាចមានចំណង់ចំណូលចិត្តខុសពីមនុស្សចាស់។ គាត់បានប្រមូលទិន្នន័យអតិថិជនចំនួន20នាក់ដែលមានអាយុដូចខាងក្រោម៖

24, 35, 17, 21, 24, 37, 26, 46, 57, 31, 33, 14, 10, 38, 41, 43, 44, 27, 53, 23

ដើម្បីគណនាប្រេកង់យើងត្រូវធ្វើតាមជំហានដូចខាងក្រោម៖

- ជំហានទី1៖ រៀងចំទិន្នន័យទៅតាមលំដាប់ពីតូចទៅធំ (ascending order) ឬ តាមលំដាប់ពីធំទៅតូច(descending order)

 10, 14, 17, 21, 24, 24, 26, 27, 31, 33, 34, 35, 37, 38, 41, 43, 44, 46, 53, 57
- ជំហានទី2៖ គណនារ៉ង់ (range) គឺស្មើចំនួនជាតុអតិបរមាដកជាតុអប្បរមា 57-10=47
- ជំហានទី3៖ កំណត់ចំនួនថ្នាក់ ឬក្រុម តាមទំនាក់ទំនង $2^C \geq n$ ដែល c ជាចំនួនថ្នាក់(ចំនួនគត់) n ជាចំនួនអង្កេត (number of observations) ដូចនេះយើងទទួលបានចំនួនថ្នាក់ស្មើ 5 ថ្នាក់៖

 $$C = \frac{\ln(n)}{\ln(2)} = \frac{\ln(20)}{\ln(2)} = 4.32 \approx 5$$
- ជំហានទី4៖ គណនាទំហំចន្លោះនៃថ្នាក់នីមួយៗ = 47/5 = 9.4 (បង្គត់ទៅជា10 ដើម្បីងាយស្រួល)
- ជំហានទី5៖ កំណត់អំពីដែនកំណត់របស់ថ្នាក់នីមួយៗ ដូចតារាង 2.5 ខាងក្រោម៖

តារាង 2.5 ដែននៃការកំណត់ថ្នាក់

ចំនួនថ្នាក់	ចន្លោះទិន្នន័យ
ថ្នាក់ទី 1	10 និងក្រោម 20 ឆ្នាំ
ថ្នាក់ទី 2	20 និងក្រោម 30 ឆ្នាំ
ថ្នាក់ទី 3	30 និងក្រោម 40 ឆ្នាំ
ថ្នាក់ទី 4	40 និងក្រោម 50 ឆ្នាំ
ថ្នាក់ទី 5	50 និងក្រោម 60 ឆ្នាំ

- ជំហានទី6៖ គណនាចំណុចកណ្តាល (midpoint) នៃថ្នាក់នីមួយៗ៖ 15, 25, 35, 45, 55
- ជំហានទី7៖ រាប់ចំនួនលេខខាងលើរួចដាក់បញ្ចូលតាមថ្នាក់នីមួយៗ
- ជំហ៊ានទី8៖ គណនាភាគរយនិងភាគរយនៃប្រេកង់គ្រួត

ទិន្នន័យសង្ខេបជំហ៊ានទាំង 7 បានដាក់បង្ហាញក្នុងតារាង 2.6

តារាង 2.6 ចំណុចកណ្តាលនិងប្រេកង់នៃអាយុអតិថិជន

ចំនួនថ្នាក់	ចន្លោះទិន្នន័យ	ចំណុចកណ្តាល (Midpoint)	ប្រេកង់ (frequency)
ថ្នាក់ទី 1	10 និងក្រោម 20 ឆ្នាំ	15	3
ថ្នាក់ទី 2	20 និងក្រោម 30 ឆ្នាំ	25	5
ថ្នាក់ទី 3	30 និងក្រោម 40 ឆ្នាំ	35	6
ថ្នាក់ទី 4	40 និងក្រោម 50 ឆ្នាំ	45	4
ថ្នាក់ទី 5	50 និងក្រោម 60 ឆ្នាំ	55	2

តារាង2.7 បង្ហាញលទ្ធផលប្រេកង់និងប្រេកង់ត្រួត។ អតិថិជនដែលមានអាយុចន្លោះពី30 និងក្រោម 40ឆ្នាំ គឺច្រើនជាងគេ មានហូតដល់ 30% ហើយបន្ទាប់មកគឺអាយុពី 20 និងក្រោម 30ឆ្នាំមាន25% ហើយទាំងពីរក្រុមគឺតំណាងឲ្យ 55%នៃអតិថិជនសរុប។ ចំណែកឯអតិថិជនដែលមានអាយុលើសពី50 ឆ្នាំគឺតិចជាងគេ មានតែ10% តែប៉ុណ្ណោះ។

តារាង 2.7 ប្រេកង់ ភាគរយនិងផលបូកតគ្នានៃភាគរយ

class interval					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10 but less than 20	3	15.0	15.0	15.0
	20 but less than 30	5	25.0	25.0	40.0
	30 but less than 40	6	30.0	30.0	70.0
	40 but less than 50	4	20.0	20.0	90.0
	50 or more	2	10.0	10.0	100.0
	Total	20	100.0	100.0	

2.4 **គំនូសតាងអថេរជាលេខ**

a. គំនូសតាងដើមនិងសន្លឹក

គំនូសតាងដើមនិងសន្លឹក (Stem-leaf)នេះអាចឲ្យយើងអង្កេត ឬកត់សំគាល់បំណែងចែកទិន្នន័យថាតើទិន្នន័យប្រមូលផ្តុំនៅត្រង់ណា?។

រូប 2.5 ដើមនិងសន្លឹក (Stem-leaf)តំណាងឲ្យអាយុអតិថិជនទាំងប្រាំក្រុម

```
frequency Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     3.00        1 .  047
     5.00        2 .  14467
     6.00        3 .  134578
     4.00        4 .  1346
     2.00        5 .  37

 Stem width:        10
 Each leaf:       1 case(s)
```

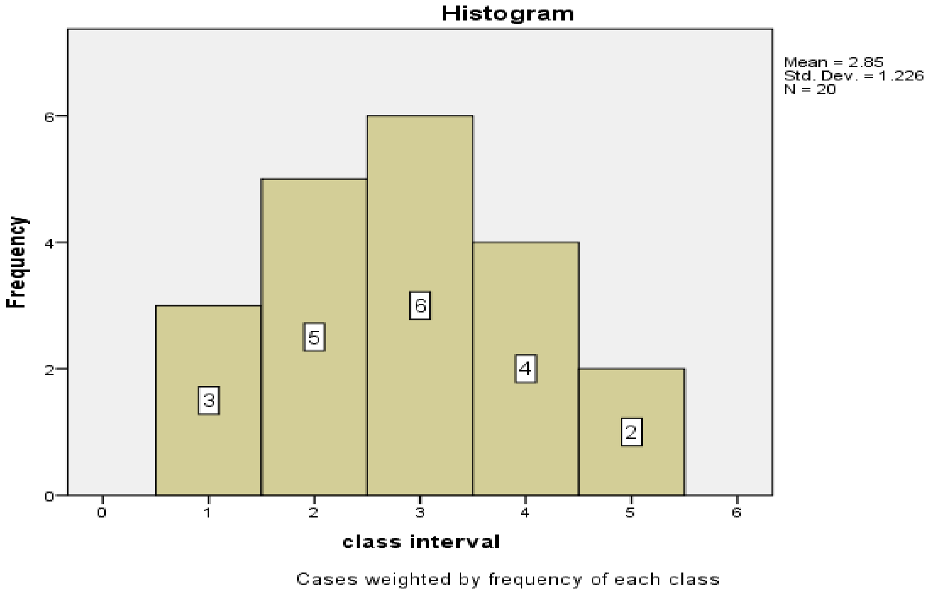
ដំហានដំបូង ទិន្នន័យត្រូវបានចាត់ជាក្រុម ហៅថា ដើម (stem) ហើយក្នុងក្រុមមានជាតុទិន្នន័យ ហៅថា សន្លឹក (leaf) ដែលស្ថិតនៅខាងស្តាំដៃ។ ឥឡូវយើងត្រឡប់ទៅមើលឧទាហរណ៍ 2.3 ដែលទាក់ទងជាមួយអាយុអតិថិជន 20នាក់ យើងបានគំនូសតាងដើមនិងសន្លឹក ដូចក្នុងរូបភាព 2.5 ។ រូប 2.5 ជាលទ្ធផល SPSS បង្ហាញឲ្យឃើញថា អតិថិជនដែលមានអាយុចន្លោះពី10 និងក្រោម 20ឆ្នាំ មាន 3នាក់ គឺអាយុ 10 14 និង អាយុ 17ឆ្នាំ ហើយអាយុចន្លោះពី30 និងក្រោម 40ឆ្នាំ គឺច្រើនជាងគេ មានរហូតដល់ 6នាក់ គឺអាយុ 31 33 34 35 37 និងអាយុ 38។ ចំណែកលេខដ៏ទៃទៀតគឺមានការបកស្រាយប្រហាក់ប្រហែលគ្នា។

b. គំនូសតាងហ៊ីស្តូក្រាម

ហ៊ីស្តូក្រាមឆាត (Histogram Chart) គឺជាប្រភេទ បាឆាត ដែលអ័ក្សដេកតំណាងឲ្យក្រុម (bin) ហើយកំពស់តំណាងឲ្យហ្វ្រេកង់។ សម្រាប់ឧទាហរណ៍របស់យើង អាយុ (មាន 5 ក្រុម) ហើយអ័ក្សឈរកំពស់តំណាងឲ្យចំនួនអតិថិជនក្នុងក្រុមនីមួយ ឬហ្វ្រេកង់។ សូមរំលឹកឡើងវិញថា ហ៊ីស្តូក្រាមឆាត ស្រដៀងគ្នាទៅនឹង បាឆាត តែខុសគ្នាត្រង់ ហ៊ីស្តូក្រាមឆាត មិនមានចន្លោះដូច បាឆាត នោះទេ។ ហើយ ហ៊ីស្តូក្រាមឆាត គេប្រើសម្រាប់អថេរជាលេខ ចំណែក បាឆាត ប្រើសម្រាប់អថេរប្រភេទ។

ឥឡូវយើងត្រឡប់ទៅមើលឧទាហរណ៍ 2.3 ខាងលើ យើងបានគំនូសតាង ហ៊ីស្តូក្រាមឆាត ដូចក្នុងរូបភាព 2.6 ខាងក្រោម។

រូប 2.5 ហ៊ីស្តូក្រាមឆាត អាយុអតិថិជនទាំងប្រាំក្រុម



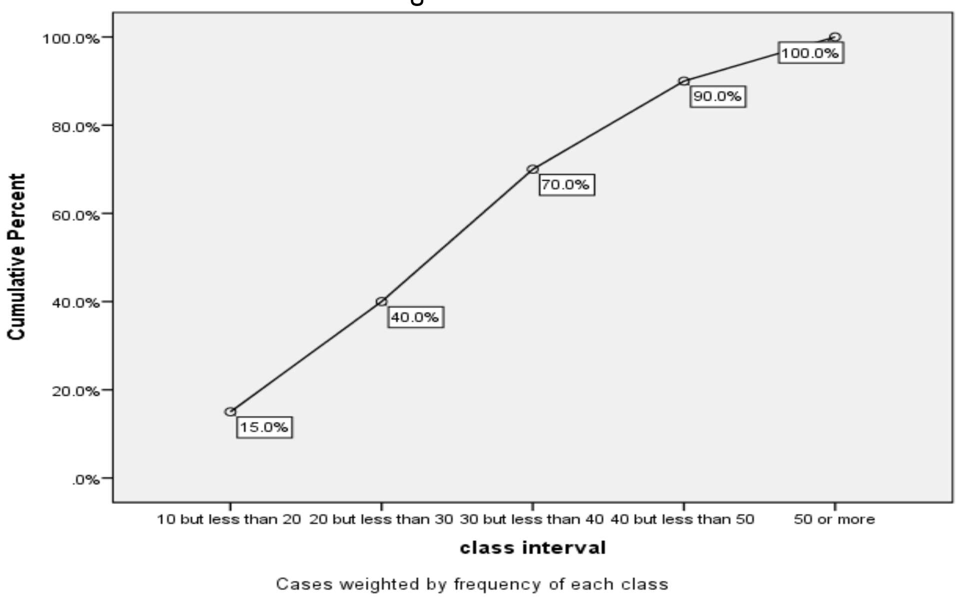
តាមរូប 2.5 ហ៊ីស្តូក្រាមឆាតខាងលើយើងអាចបកស្រាយតាមហ្វ្រេកង់។ ហើយការបកស្រាយគឺដូចទៅនឹងការបកស្រាយហ្វ្រេកង់មុនៗដែរ។

c. គំនូសតាងប៉ូលីហ្គុនជាភាគរយ

1) គំនូសតាងប៉ូលីហ្គុនជាភាគរយឬជាហ្រេ្វកង់ (percentage or frequency Polygon) ដែលមានមួយអថេរ (one numerical variable) ដូចទៅនឹងកំនូសតាងហ៊ីស្តូក្រាមជាភាគរយឬហ្រេ្វកង់ដែរ (percentage or frequency Histogram)។ គំនូសតាងប៉ូលីហ្គុនគូសភ្ជាប់ពីចំណុច midpoint នៃហ្រេ្វកង់នីមួយទៅនឹងចំនុចកណ្តាលនៃហ្រេ្វកង់មួយទៀត។ ចំណុចល្អរបស់ គំនូសតាងប៉ូលីហ្គុនលើស ហ៊ីស្តូក្រាម ត្រង់វាអាចបង្ហាញហ្រេ្វកង់គ្រួត (cumulative frequency polygon) ឬផលបូកភាគរយនៃហ្រេ្វកង់គ្រួត (cumulative percentage Polygon) ដែលហ៊ីស្តូក្រាម មិនបានបង្ហាញ។

ឧទាហរណ៍ អាយុអតិថិជន 20នាក់ខាងលើ យើងអាចបង្ហាញជាផលបូលជាភាគរយរបស់ហ្រេ្វកង់តាមរយៈ គំនូសតាងប៉ូលីហ្គុនជាភាគរយគ្រួត (cumulative percentage Polygon) ដូចដែលបានបង្ហាញក្នុងរូបទី2.6។

រូបទី2.6 ប៉ូលីហ្គុនជាភាគរយហ្រេ្វកង់គ្រួត នៃអាយុរបស់អតិថិជន



តាមរយៈរូបទី2.6 បង្ហាញឲឃើញថា របស់អតិថិជនដែលមានអាយុក្រោម30ឆ្នាំមាន40% ហើយ អាយុក្រោម40ឆ្នាំមាន 70% និង អាយុក្រោម50 ឆ្នាំមាន 90% ។

2) គំនូសតាងប៉ូលីហ្គុនដែលមានពីរ ឬច្រើនអថេរលេខ (two or more numerical variables)

ក្នុងករណីដែលអ្នកចង់ធ្វើការប្រៀបធៀបពីរក្រុមឡើងទៅដោយប្រើ ហ៊ីស្តូក្រាម នោះអាចជាន់គ្នា ដូច្នេះអ្នកអាចប្រើគំនូសតាង ប៉ូលីហ្គុនជាភាគរយជំនួសវិញ។ ប៉ូលីហ្គុនជាភាគរយ ឬហៅថា អូហ្គៀវ(Ogive) ប្រើចំណុចកណ្តាល (midpoint) របស់ចន្លោះថ្នាក់នីមួយៗតំណាងឲ្យក្រុមឬថ្នាក់នីមួយៗ។

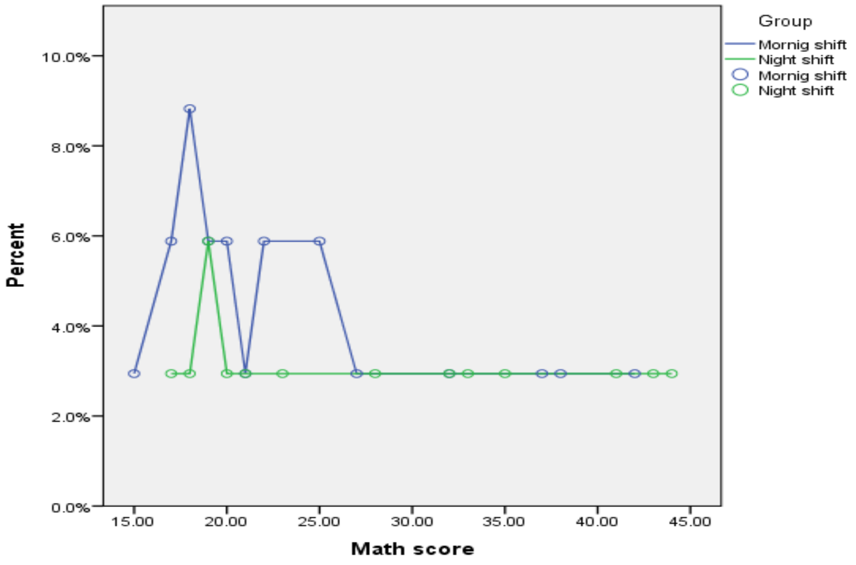
ឧទាហរណ៍ 2.4 អ្នកសិក្សាម្នាក់ចង់ប្រៀបធៀបពិន្ទុរបស់និស្សិតពីរក្រុម គឺវេនព្រឹកនិងវេនយប់។ គាត់ធ្វើការជ្រើសរើសពិន្ទុគណិតវិទ្យាថ្នាក់ព្រឹកចំនួន20នាក់និងថ្នាក់យប់ចំនួន14នាក់។

ពិន្ទុទាំងពីរវេនបានបង្ហាញនៅក្នុងតារាង 2.8 និងគំនូស ប៉ូលីហ្គូន គិតជាភាគរយគឺត្រូវបានដាក់បង្ហាញក្នុងរូប 2.6។

តារាង 2.8 ពិន្ទុគណិតវិទ្យារបស់និស្សិតវេនព្រឹកនិងវេនយប់

វេនយប់	15	17	17	18	18	18	19	19	20	20
	21	22	22	25	25	27	32	37	38	42
វេនព្រឹក	17	18	19	19	20	21	23	28	32	33
	35	41	43	44						

រូប 2.6 គំនូសតាងប៉ូលីហ្គូនជាភាគរយនៃពិន្ទុរបស់នីស្សិតទាំងពីរវេន



ដើម្បីធ្វើការវិភាគថា ក្រុមណាមួយសំបូរសិស្សពូកែមុខវិជ្ជាគណិតវិទ្យាច្រើនជាង ឥឡូវយើងសន្មត់ថា ពិន្ទុអតិបរមាគឺ 50 ហើយជាមធ្យមគឺ 25ពិន្ទុ។ តាមរយៈរូប 2.6 បង្ហាញឲ្យឃើញថានៅត្រង់ 20 ពីន្ទុក្រុបវេនព្រឹកមានរហូតដល់ 6% ចំណែកវេនយប់មានតែ 3%តែប៉ុណ្ណោះ ហើយចាប់ពីពិន្ទុ 25 ឡើងទៅ ក្រុមវេនយប់មានចំនួនច្រើជាងវេនព្រឹក ដូច្នេះសិស្សវេនព្រឹកសំបូរសិស្សខ្សោយមុខវិជ្ជាគណិតវិទ្យា ច្រើនជាងវេនយប់។

d. គំនូសតាងខ្ពស់-ទាប-បិទទ្វារ

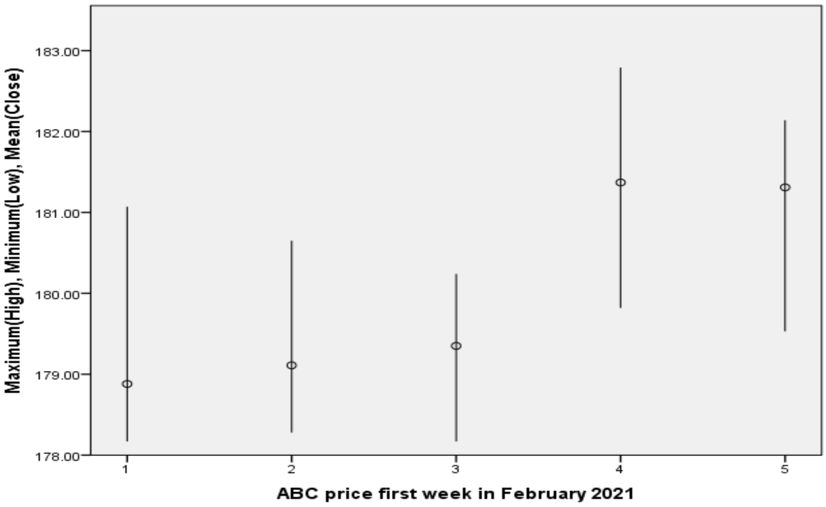
គំនូសតាង ខ្ពស់-ទាប-បិទទ្វារ (High-low-close chart) ជាប្រភេទគំនូសតាងដែលត្រូវបានគេប្រើដើម្បីបង្ហាញអំពីដំណើរប្រែប្រួលនៃតម្លៃមូលប័ត្រហិរញ្ញវត្ថុធៀបទៅនិងពេលវេលា។ វាគឺជាប្រភេទ បាធាតដែលមានអ័ក្សឈរឲ្យតម្លៃខ្ពស់ ទាប និងបិទទ្វារ ក្នុងគ្រាណាមួយ (ភាគច្រើនក្នុងមួយថ្ងៃ)។ តម្លៃបិទទ្វារ (closing price) គឺមានសារៈសំខាន់ដើម្បីអោយអ្នកវិភាគសិក្សាតម្លៃនៅពេលអនាគត។ ឧទាហរណ៍ 2.5

អ្នកវិភាគហិរញ្ញវត្ថុម្នាក់បានធ្វើការកត់ត្រាតម្លៃមូលប័ត្រ ABC នៅសប្តាហ៍ទីមួយនៃខែកម្ភៈ 2021 ដូចតារាង 2.9។

តារាង 2.9 តម្លៃភាគហ៊ុន (ដុល្លា) ABC នៅសប្តាហ៍ទីមួយនៃខែកុម្ភៈ 2021

ថ្ងៃ	ខ្ពស់	ទាប	បិទទ្វារ
Feb-1	181.07	178.17	178.88
Feb-2	180.65	178.28	179.11
Feb-3	180.24	178.17	179.35
Feb-4	182.79	179.82	181.37
Feb-5	182.14	179.53	181.31

រូប 2.7 គំនូសតាង ខ្ពស់-ទាប-បិទទ្វារ តម្លៃភាគហ៊ុន ($) ABC



រូប 2.7 បង្ហាញថា កំពូលនៃអង្កត់នីមួយៗតំណាងតម្លៃខ្ពស់ជាងគេនៅថ្ងៃនេះ (high) ហើយចុងអង្កត់តំណាងឲ្យតម្លៃទាបជាងគេ (low) និងចំណុចកណ្តាលនៃអង្កត់ គឺតំណាងឲ្យតម្លៃបិទទ្វារ (closing price) នៅថ្ងៃនេះ។ គំនូសតាង ខ្ពស់-ទាប-បិទទ្វារ មានសារៈសំខាន់ដើម្បីអោយព្យាករណ៍តម្លៃភាគហ៊ុននៅថ្ងៃស្អែក តែមិនបានវិភាគនៅក្នុងសៀវភៅនេះទេ។ បើសិនជានិស្សិតមានការចាប់អារម្មណ៍នូវការប្រើប្រាស់គំនូសនេះសូមអានសៀវភៅ Financial Behavioral Analysis។

e. គំនូសតាងស្កាតតឺផ្លត

គំនូសតាងស្កាតតឺផ្លត (the scatter plot) ជាគំនូសតាងដែលបង្ហាញអំពីទំនាក់ទំនងអថេរជាលេខដែលមានពីរអថេរ។ អថេរមួយស្ថិតនៅលើអ័ក្សY និងអថេរមួយទៀតស្ថិតនៅលើអ័ក្ស X។ ជាទូទៅ គំនូសតាងនេះត្រូវបានប្រើប្រាស់ជាជំហានទីមួយនៃការវិភាគសមីការលីនេអ៊ែ។

ឧទាហរណ៍ 2.6 អ្នកគ្រប់គ្រងទីផ្សារម្នាក់ចង់ដឹងអំពីទំនាក់ទំនង ឬប្រសិទ្ធភាពនៃការផ្សាយពាណិជ្ជកម្មរបស់ក្រុមហ៊ុនលក់កង់ ABC គាត់ធ្វើការកត់ត្រាលើចំណាយនៃការផ្សាយពាណិជ្ជកម្ម និងការលក់កង់ចំនួន 9 ត្រីមាស ដូចដែលបានបង្ហាញនៅក្នុងតារា 2.9។

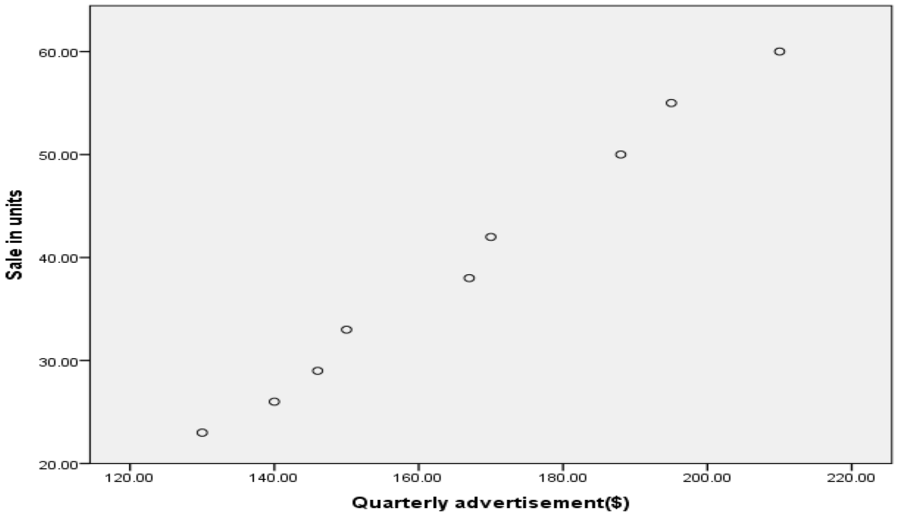
តារាង 2.9 ចំណាយនិងការលក់ចំនួន9ត្រីមាសរបស់ក្រុនហ៊ុនABC

ត្រីមាស	1	2	3	4	5	6	7	8	9
ការលក់ (ឯកតា)	23	26	29	33	38	42	50	55	60
ការផ្សាយពាណិជ្ជកម្ម($)	130	140	146	150	167	170	188	195	210

ដើម្បីពិនិត្យថាតើការលក់និងការផ្សាយពាណិជ្ជកម្មមានទំនាក់ទំងគ្នាយ៉ាងណា គំនូសតាងស្កាតតឺផ្លត នឹងអាចត្រូវបានគេប្រើ។ លទ្ធផលនៃ គំនូសតាងស្កាតតឺផ្លត ដូចបានបង្ហាញក្នុងរូប 2.8។

រូប 2.8 បង្ហាញអោយឃើញថា ការលក់ហាក់ដូចជាមានទំនាក់ទំនងវិជ្ជមានជាមួយការផ្សាយពាណិជ្ជកម្ម បើចំណាយនៃការផ្សាយពាណិជ្ជកម្មកើនឡើង ការលក់ក៏កើនឡើងដែរ ដូច្នេះយើងអាចសន្និដ្ឋានបានថា ការលក់ប្រែលជាមានទំនាក់ទំនងវិជ្ជមានជាមួយការផ្សាយពាណិជ្ជកម្ម។ បើក្នុងករណីដែលយើងចង់ដឹងថាកំរិតនៃការផ្សាយពាណិជ្ជកម្មទៅលើការលក់ លុះត្រាតែយើងធ្វើការវិភាកសមីការលីនេអ៊ែរ ។ ករណីនេះយើងនិងសិក្សានៅជំពូកទី10។

រូប 2.8 ស្កាតធើផ្លតតាងទំនាក់ទំនងរវាងការផ្សាយពាណិជ្ជកម្មនិងការលក់

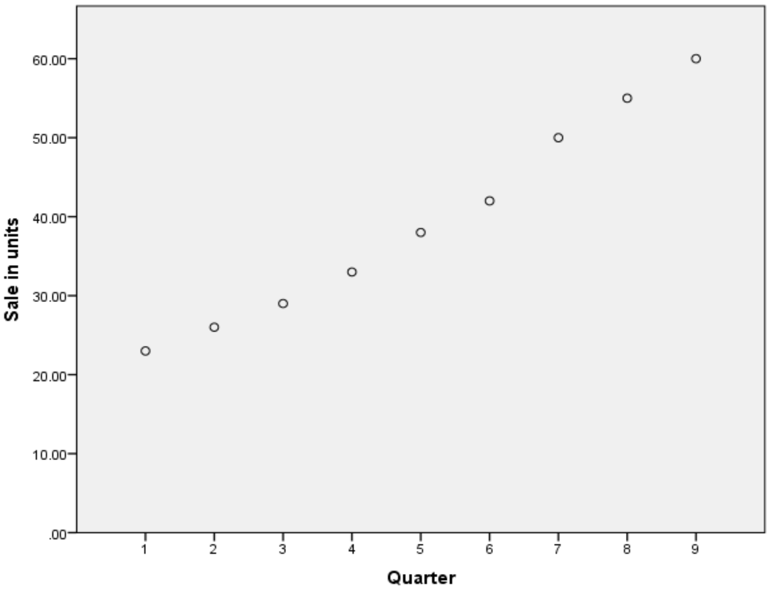


f. គំនូសតាងជាស៊េរី

គំនូសតាងពេលវេលាជាស៊េរីស្រដៀងទៅនឹងស្កាតតឺផ្លតដែរ វាបង្ហាញអំពីទំនាក់ទំនងអថេរមួយទៅនឹងពេលវាលា។ អថេរមួយជាលេខស្ថិតនៅលើអ័ក្សY និងអថេរមួយទៀតជាពេលពេលវេលាស្ថិតនៅលើអ័ក្សX។ ជាទូទៅ គំនូសតាងប្រភេទនេះត្រូវបានប្រើប្រាស់ជាជំហានទីមួយនៃការធ្វើវិភាគលីនេអ៊ែរ ដោយយកពេលវេលា (time or trend) ជំនួសអោយ X។ ឧបមាថា យើងចង់ដឹងថាតើការលក់ក្នុងត្រីមាសនីមួយកើនឡើងប៉ុន្មាន យើងអាចយកទិន្ន័យទាំងនេះដាក់លើ ស្កាតតឺផ្លត តាមពេលវេលា (the time-scatter plot)

ដោយយកបរិមាណលក់ដាក់នៅលើអ័ក្ស Y ហើយពេលវេលា (ត្រីមាស) ដាក់លើអ័ក្ស X ។ បើយើងយកទិន្នន័យក្នុងតារាង2.9 មកដាក់លើគំនូសតាងជាស៊េរី យើងបានរូបទី2.9។

រូប2.9 គំនូសតាងទំនាក់ទំនងការលក់នឹងពេលវេលា



រូប2.9 បញ្ជាក់ឲ្យឃើញថា ការលក់ពីត្រីមាសមួយទៅត្រីមាសមួយទៀត ហាក់ដូចជាមានការកើនឡើងពិតប្រាកដមែន។ ហើយដើម្បីចង់ដឹងថា មួយត្រីមាសកើនជាមធ្យមប៉ុន្មាន លុះត្រាតែយើងធ្វើ ការវិភាគតាមលីនេអ៊ែរ ដោយយក បរិមាណលក់តំណាងឲ្យ Y និងពេលវេលាត្រីមាស(quarter)តំណាងឲ្យ X ការវិភាគនេះនឹងលំអិតនៅជំពូកទី10។

សេចក្តីសង្ខេប

ការរៀបចំនិងបង្ហាញទិន្នន័យទាក់ទងជាមួយការរៀបចំជាតារាងសង្ខេប និងគំនួសតាងផ្សេងៗ ដើម្បីងាយស្រួលដល់ការធ្វើបទបង្ហាញនិងធ្វើសេចក្តីសន្និដ្ឋាន។ ការជ្រើសរើសប្រភេទតារាងសង្ខេប និងគំនួសតាង គឺអាស្រ័យទៅនឹងប្រភេទទិន្នន័យ។ នៅជំពូកទី3 អ្នកនិងធ្វើការសិក្សាទៅលើស្ថិតិបែបបរិយាយដែលជាផ្នែកមួយដ៏សំខាន់ក្នុងការវិភាគទិន្នន័យនិងសិក្សាអំពីការបកស្រាយទិន្នន័យ៕

សំណួរនិងលំហាត់

I. សំណួរ

1. តើ ហ៊ីស្តូក្រាម និង ប៉ូលីហ្គូន ខុសគ្នាយ៉ាងដូចម្តេចក្នុងការសង់និងប្រើប្រាស់?
2. តើគុណសម្បត្តិនិងគុណវិបត្តិមានអ្វីខ្លះក្នុងការប្រើគំនូស បាឆាត ផាយឆាត និង ប៉ារ៉ាតូឆាត?
2. ចូរប្រៀបធៀបចំណុចខុសគ្នារវាង បាឆាត និង ហ៊ីស្តូក្រាម ។
3. តើចំណុចខុសគ្នារវាងគំនូសស៊េរី និង ស្កាតឺផ្លត មានអ្វីខ្លះ?
4. តើតារាងខ្វែងប្រើប្រាស់នៅពេលណា?

II. លំហាត់

5. ចំណូលរបស់ក្រុមហ៊ុនវិនិយោគមួយ ត្រូវបានរាយការណ៍កាលពីឆ្នាំ 2020 មានប្រភពដូចខាងក្រោម៖

ប្រភព	ចំណូល (គិតជា $1000)
ស្រាបៀរ	5,000
សម្ភារៈសំណង់	3,000
កសិកម្ម	1,500
ផ្សេងៗ	500

a. ចូលគណនាភាគរយនៃប្រភេទនីមួយៗ?

b. តើអ្នកសន្និដ្ឋានយ៉ាងដូចម្តេចចំពោះប្រភពនៃចំណូលនៅឆ្នាំ 2020?

6. ការស្ទង់មតិមួយទៅលើអ្នកទិញទំនិញតាមបណ្តាញអ៊ីនធឺណែតចំនួន 400 នាក់ដែលទាក់ទងជាមួយគតិយាបទអតិថិជន។ ក្នុងចំណោមសំណួរទាំងនោះ មានសំណួរមួយត្រូវបានសួរថា៖ តើអ្នកចូលចិត្តទិញទំនិញតាមអ៊ីនធឺណែតដែរឬទេ? ចំលើយៈ បាទ ឬ ទេ។ លទ្ធផលត្រូវបានសង្ខេបក្នុងតារាងខាងក្រោម៖

ចូលចិត្ត	ភេទ		Total
	បុរស	ស្ត្រី	
បាទ	70	150	220
ទេ	80	100	180
Total	150	250	400

a. ចូលបង្កើតតារាងខ្វែងដោយផ្អែកលើភាគរយសរុប ភាគរយនៃជួរដេក និងភាគរយនៃជួរឈរ

b. តើអ្នកធ្វើការសន្និដ្ឋានយ៉ាងដូចម្តេចចំពោះគតិយាបទនៃការទិញនេះ?

6. សិប្បកម្មផលិតដែកសសៃមួយ បានផលិតសរសៃដែកដើម្បីផ្គត់ផ្គង់ទីផ្សារសំណង់។ ផលិតករបានកំណត់ថា ស្តង់ដារនៃសរសៃដែកគឺត្រូវស្ថិតនៅចន្លោះ 8.30 មីលីម៉ែត្រ ទៅ 8.60 មីលីម៉ែត្រ ។ ដើម្បីចង់ដឹងការផលិតរបស់គាត់នៅគោរពតាមស្តង់ដារដែលបានកំណត់ដែរឬយ៉ាងណា គាត់បានប្រមូលប៉ាន់គំរូចំនួន 40 សរសៃ ហើយទិន្នន័យទាំងនោះត្រូវបានកត់ត្រាដូចខាងក្រោម៖

8.317	8.484	8.489	8.429	8.405
8.383	8.403	8.414	8.46	8.439
8.348	8.414	8.481	8.412	8.411
8.41	8.419	8.415	8.42	8.427
8.351	8.385	8.479	8.41	8.42
8.373	8.465	8.429	8.405	8.498
8.481	8.498	8.458	8.323	8.409
8.422	8.447	8.462	8.42	8.511

a. ចូលគណនាបំណែងចែកប្រេកង់និងភាគរយនៃបំណែងចែកសរសៃដែក

b. ចូលគណនាផលបូកភាគរយនៃបំណែងចែក

c. តើអ្នកសន្និដ្ឋានយ៉ាងដូចម្តេចអំពីការផលិតសរសៃដែករបស់គាត់ដែលទាក់ទងជាមួយស្តង់ដារផលិតកម្មដែលតម្រូវឲ្យការផលិតសរសៃដែកគឺនៅចន្លោះពី 8.30 និង 8.60មីលីម៉ែត្រ?

7. ការស្ទង់មតិមួយទៅលើមនុស្សចំនួន 300 នាក់ទាក់ទងដល់កកត្តាដែលជះឥតទ្ធិពលទៅលើការធ្វើសេចក្តីសំរេចចិត្តក្នុងការទិញទំនិញរបស់ពួកគេ។ ទិន្នន័យទាំងនោះត្រូវបានកត់ត្រាក្នុងតារាងខាងក្រោម។

ឥតទ្ធិពល	ចំនួនអ្នកទិញ
ការផ្សាយបន្តផ្ទាល់	60
តាមរយៈមិត្តភ័ក្រ	30
បណ្តាញពត៌មាន	40
វេបសាយរបស់អ្នកលក់	120
បុគ្គលិករបស់អ្នកលក់	20
ផ្សេងៗ	30

a. ចូលសង់គំនូស បាធាត ផាយធាតនិង ហ៊ីស្តូក្រាម

b. តើគំនូសតាងណាមួយដែលអ្នកគិតថាសក្តិសមទៅនឹងទិន្នន័យប្រភេទនេះ?

c. ចូលធ្វើការសន្និដ្ឋានចំពោះកត្តាដែលមានឥតទ្ធិពលទៅលើការធ្វើសេចក្តីសំរេចចិត្តទាំងនោះ?

8. ខាងក្រោមនេះគឺជាចំណាយថ្លៃអគ្គីសនីប្រចាំខែ(គិតជា 1000 រៀល)ដែលត្រូវបានប្រមូលដោយចៃដន្យកាលពីឆ្នាំ 2000 របស់គ្រួសារទាំង 80 នៅក្នុងសង្កាត់មួយនៃរាជធានីភ្នំពេញ។

100	178	150	96	108	178	111	109
168	185	149	163	119	185	148	120
200	90	206	150	183	90	213	183
178	116	178	154	151	116	130	151
147	172	123	130	114	172	165	114
102	111	128	143	135	102	143	102
153	148	144	187	191	153	187	153
197	213	168	166	140	197	166	197
127	130	109	139	130	127	139	127
82	165	167	149	160	85	149	87

a. ចូលសង់គំនូសតាង ហីស្តូក្រាម និង ប៉ូលីហ្គូនគិតជាភាគរយ

b. ចូលសង់គំនូសតាង ប៉ូលីហ្គូន គិតជាភាគរយត្រួត

c. ចូលសង់គំនូសតាងដើមនិងសន្លឹក

c. តើចំណាយភាគច្រើនប្រមូលផ្តុំត្រង់តម្លៃប៉ុន្មាន ?

9. ខាងក្រោមនេះគឺជាពិន្ទុមុខវិជ្ជាស្ថិតិរបស់និស្សិតទាំងបីវេន។ ចូលប្រើគំនូស ប៉ូលីហ្គូន គិតជាភាគរយដើម្បីប្រៀបធៀបពិន្ទុនិងផ្តល់ការសន្និដ្ឋានទៅលើពិន្ទុរបស់និស្សិតទាំងបីវេន។

វេនព្រឹក	16	23	20	22	26	30	27	39	20	40
វេនថ្ងៃ	21	22	22	25	25	27	32	37	38	16
វេនយប់	17	20	39	25	26	21	23	28	32	33

ជំពូកទី 3 ស្ថិតិបរិយាយ

(Numerical Descriptive Statistics)

គោលបំណង

- រៀបរាប់អំពីវិបាកនៃទំនោរកណ្តាល បម្រែបម្រួល និងទម្រង់នៃទិន្នន័យជាលេខ
- បង្កើតនិងបកស្រាយអំពីរូបភាពនៃប្រអប់
- គណនាចតុភាគ និង សតភាគ
- គណនារង្វាស់សង្ខេបតាមបែបបរិយាយនៃចំនួនសកល
- គណនា ខូវ៉ារីយ៉ង់ និង មេគុណទំនាក់ទំនង
- លើកឧទាហរណ៍ផ្សេងៗនៃការប្រើប្រាស់ស្ថិតិ

3.1. ស្ថិតិបរិយាយ

ស្ថិតិបរិយាយ (descriptive statistics) គឺសេចក្តីសង្ខេបដែលរៀបរាប់ខ្លីៗអំពីសំណុំទិន្នន័យមួយ។ វាអាចជាសេចក្តីសង្ខេបដែលដំណាងឲ្យចំនួនសាកល (population) ឬតំណាងឲ្យប៉ាន់គំរូណាមួយ (a sample)។ ស្ថិតិបរិយាយត្រូវបានបំបែកជា រង្វាស់ទំនោរកណ្តាល (measure of center or tendency) និងរង្វាស់ភាពរាយប៉ាយ (variability or spread)។ ហើយរង្វាស់ទំនោរកណ្តាលរួមមាន៖ តម្លៃមធ្យម (mean) មេដ្យាន (median) និង ម៉ូដ (mode) ដោយឡែករង្វាស់ភាពរាយប៉ាយរួមមាន៖ គម្លាតស្តង់ដារ (standard deviation) វ៉ារីយ៉ង់ (variance) ចំនួនអប្បរម៉ា (minimum) ចំនួនអតិបរមា (maximum) កឺតូស៊ីស (kurtosis) និង ស្គូនេសskewness[2]។

3.2 រង្វាស់កណ្តាល

តម្លៃមធ្យម (mean or average) គឺជារង្វាស់ចំនុចកណ្តាលនៃទិន្នន័យដែលត្រូវបានគេប្រើញឹកញាប់ជាងគេ។ តម្លៃមធ្យមគឺត្រូវបានចែកចេញជាបីប្រភេទ៖ មធ្យមពិជគណិត (arithmetic mean) មធ្យមទម្ងន់ ឬសមាមាត្រ (weighted mean) និងមធ្យមធរណីមាត្រ (geometric mean)[3]។

3.2.1 មធ្យមពិជគណិត

មធ្យមពិជគណិត(arithmetic mean) អាចត្រូវបានប្រើដើម្បីគណនាទៅលើចំនួនគត់ (discrete data) ឬចំនួនដែលមានក្បៀស (continuous data) ដោយបូកតម្លៃនៃសំណុំ X ពីមួយដល់ n រួចចែកនឹង n។ រូបមន្តមធ្យមពិជគណិតត្រូវបានកំណត់ដូចតទៅ៖

---

[2]ទាញយកនៅថ្ងៃទី 5/5/2021 ចេញពី https://www.investopedia.com/terms/d/descriptive_statistics.asp#:~:text=Descriptive%20statistics%20are%20brief%20descriptive,measures%20of%20variability%20(spread).

[3] ទាញយកនៅថ្ងៃទី 5/5/2021 ចេញពី https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad (3.1a)$$

ដោយសារសមាមាត្រនៃមធ្យមពិជគណិតគឺស្មើគ្នា យើងអាចទាញ1/n ជាកត្តាបាន ដូច្នេះរូបមន្ត (4.1a) អាចសរសេរទៅជា៖

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (3.1b)$$

ឧទាហរណ៍ 3.1 ចំណូល5ថ្ងៃរបស់អ្នកលក់ម្នាក់គឺដូចខាងក្រោម៖

ថ្ងៃ	1	2	3	4	5
ចំណូល($)	12	13	14	15	101

ដូច្នេះ ចំណូលជាមធ្យមក្នុងមួយថ្ងៃរបស់អ្នកលក់គឺ35ដុល្លា៖

$$\bar{X} = \frac{12 + 13 + 14 + 15 + 16}{5} = \frac{1}{5}(12 + 13 + 14 + 15 + 16) = \$14$$

មធ្យមពិជគណិតមានសារៈសំខាន់ចំពោះទិន្នន័យដែលកើតរួច ហើយធាតុនីមួយៗនៃសំណុំ X មានតម្លៃមិនសូវខុសគ្នាខ្លាំង ក៏ប៉ុន្តែវាមានគុណវិបត្តិ ដោយសារតែវាងាយរងឥទ្ធិពលនៃធាតុមិនប្រក្រតី ឬធាតុខុសគេ (unusual or outlier)។ ក្នុងករណីដែលមានធាតុមិនប្រក្រតីណាមួយក្នុងសំណុំទិន្នន័យ វាអាចធ្វើឲ្យតម្លៃមធ្យមរបស់វាមានភាពប្រែប្រួលខ្លាំងពីការពិត។ ឧបមាថា ការចំណូល 5 ថ្ងៃខាងលើ បើថ្ងៃទី5 លក់បាន 101 ដុល្លាវិញ(មិនមែន 16ដុល្លាទេ) ពេលនេះចំនួនមធ្យមនៃការលក់ក្នុងមួយថ្ងៃគឺ 31 ដុល្លាឯណោះ គឺពិតជាមិនសមស្របទាល់តែសោះព្រោះថា មធ្យម 31 ដុល្លាក្នុងមួយថ្ងៃ គឺខុសពីការលក់ពីថ្ងៃទី1 ដល់ ថ្ងៃទី4ខ្លាំងណាស់។ ក្នុងរណីដែលមាន ធាតុមិនប្រក្រតី ឬធាតុខុសគេ អ្នកអាចប្រើមេដ្យានមកជំនួសវិញ។

3.2.2 មធ្យមទម្ងន់

មធ្យមពិជគណិតគឺជាករណីពិសេសរបស់មធ្យមទម្ងន់ ឬសមាមាត្រ(weighted average)។ មធ្យមពិជគណិតបានកំណត់ថា សមាមាត្រឬមេគុណនៃធាតុ X នីមួយៗគឺស្មើគ្នា (1/n)។ ផ្ទុយពីមធ្យមពិជគណិត មេគុណរបស់មធ្យមទម្ងន់នៃធាតុ X នីមួយៗ គឺមិនស្មើគ្នានោះទេ ហើយភាគច្រើន ផលបូកនៃមេគុណត្រូវតែស្មើ1 ពេលខ្លះក៏អាចខុសពី1ផងដែរ។ បើក្នុងករណីផលបូកមេគុណរបស់វាស្មើ 1 យើងអាចហៅថាមធ្យមសមាមាត្រ ប៉ុន្តែក្នុងករណីដែលផលបូកវាខុសពី 1 យើងអាចហៅថាមធ្យមទម្ងន់ ឧទាហរណ៍ ពិន្ទុគណិតវិទ្យាគឺមានមេគុណ 200% ហើយពិន្ទុមុខវិជ្ជាខ្មែរមានមេគុណ 150%។ ចំណុចខុសគ្នាមួយទៀតរវាងមធ្យមពិជគណិត និងមធ្យមទម្ងន់គឺថា មធ្យមពិជគណិតភាគច្រើនត្រូវបានអនុវត្តទៅលើទិន្នន័យ ឬព្រឹត្តិការណ៍ដែលកើតរួច(past data) ប៉ុន្តែបើក្នុងករណីដែលទិន្នន័យឬព្រឹត្តិការណ៍នោះមិនទាន់កើតមាន (future event) ឧទាហរណ៍ ដូចជាអាត្រាប្រាក់ចំណេញលើការវិនិយោគដែលរំពឹងទុកក្នុងពេលអនាគត គឺគេប្រើមធ្យមទម្ងន់ ឬសមាមាត្រជំនួសវិញ។ ក្នុងជីវភាពជាក់ស្តែង មធ្យមទម្ងន់ត្រូវបានអនុវត្តភាគច្រើននៅក្នុងពិភពហិរញ្ញវត្ថុ ។ មធ្យមទម្ងន់មានរូបមន្តដូចខាងក្រោម៖

$$\bar{X} = \frac{W_1X_1 + W_2X_2 + \cdots + W_nX_n}{W_1 + W_2 + \cdots + W_n} \qquad (3.2)$$

ហើយ $W_1 + W_2 + \cdots + W_n = 1$

ឧទាហរណ៍ 3.2 លោក A មានលុយសរុបចំនួន 500ដុល្លា គាត់យកលុយ 200ដុល្លា ទៅចងការឲ្យលោក B ដោយយកការប្រាក់ 5%ក្នុងមួយឆ្នាំ និង300ដុល្លាទៀតចងការឲ្យលោក C ដោយយកការប្រាក់តែ4%តែប៉ុណ្ណោះ។ តើ លោក A រំពឹងថានឹងទទួលបានការប្រាក់ប៉ុន្មាន?

$$\bar{X} = \frac{(200\text{x } 0.05) + (300\text{x}0.04)}{(200 + 300)} = \frac{200\text{x } 0.05}{500} + \frac{300\text{x}0.04}{500}$$

$$= \frac{200}{500} * 0.05 + \frac{300}{500} * 0.04 = 0.40(0.05) + 0.60(0.04) =$$

$0.044\ or\ 4.4\%$

ដូច្នេះការប្រាក់ដែលរំពឹងទុករបស់លោក A គិតជាមធ្យមគឺ 4.4%ឬ 22ដុល្លា។ ផ្ទុយមកវិញបើ លោក A ឲ្យលោក B និង លោក C ខ្ចីម្នាក់ៗ 250 ដុល្លាស្មើគ្នា។ ការប្រាក់រំពឹងទុករបស់លោកA គឺ 4.5% ពេលនេះមធ្យមទម្ងន់ក្លាយជាមធ្យមពិជគណិត ពីព្រោះសមាមាត្រស្មើគ្នា។

$$\bar{X} = \frac{250}{500} * 0.05 + \frac{250}{500} * 0.04 = \frac{1}{2} * 0.05 + \frac{1}{2} * 0.04 = \frac{1}{2} * (0.05 + 0.04) = 4.5\%$$

3.2.3 មធ្យមធរណីមាត្រ

មធ្យមធរណីមាត្រ (geometric mean) ជួនកាលហៅថា មធ្យមនៃអាត្រាផ្គួបប្រចាំឆ្នាំ (compound annual growth rate)។ វាគឺជាតម្លៃប្រាក់ដើមនិងការប្រាក់ក្នុងមួយដុល្លា រួចបំពាក់ឬសទី n ហើយដក1។ មធ្យមធរណីមាត្រ មានរូបមន្តដូចខាងក្រោម៖

$$\bar{X}_G = \sqrt[n]{(1 + r_1)(1 + r_2) \ldots (1 + r_n)} - 1 \qquad (3.3)$$

ឧទាហរណ៍ 3.3 Sophia ទិញភាគហ៊ុនរដ្ឋករទឹករយៈពេល 3ឆ្នាំ។ ឆ្នាំទី1 ចំណេញបាន 10% ឆ្នាំទី2 ចំនេញ 5% និងឆ្នាំទី3 ខាត 8%។ តើជាមធ្យមក្នុងមួយឆ្នាំ Sophia ចំណេញបានប៉ុន្មាន?

$$\bar{X}_G = \sqrt[3]{(1 + 0.10)(1 + 0.05)(1 - 0.08)} - 1 = 2.04\%$$

ដូច្នេះអាត្រាចំណេញរបស់ Sophia ទៅលើភាគហ៊ុនរបស់រដ្ឋករទឹក គិតជាមធ្យមគឺ 2.04%ក្នុងមួយឆ្នាំ។

3.2.4 មេដ្យាន

មេដ្យាន (median)គឺជាចំនុចកណ្តាលនៃទិន្នន័យ បន្ទាប់ពីទិន្នន័យនោះត្រូវបានរៀបចំតាមលំដាប់លំដោយរួច។ រូបមន្តមេដ្យានត្រូវបានចែកចេញជាពីរ អាស្រ័យទៅនឹងចំនួនធាតុនៃសំណុំ X ជាលេខសេស ឬ លេខគូ។

- ក្នុងករណីទិន្នន័យជាចំនួនសេស រូបមន្តមេដ្យានX៖

$$Median\ X(if\ n\ is\ even\ number) = X\frac{(n+1)}{2} \qquad (3.4a)$$

ហើយ X= សំណុំដែលមានធាតុ n

ឥឡូវយើងត្រឡប់ទៅមើលទិន្នន័យនៃការលក់ទាំង 5 (n=5) នៅក្នុងឧទាហរណ៍ 3.1ខាងលើ។ បន្ទាប់ពីយើងរៀបចំទិន្នន័យនោះតាមលំដាប់លំដោយរួច យើងបានរកឃើញមេដ្យានគឺ 14ដុល្លា (ធាតុទី3)។ គួរកត់សំគាល់ថា បើយើងប្រើមេដ្យានជារង្វាស់កណ្តាល វាហាក់ដូចជាមិនឃ្លាតឆ្ងាយពីអ្នកលក់ទី1 ទី2 និងទី4 ឡើយ បើធៀបជាមួយចំនួនមធ្យមពិជគណិត 31 ដុល្លាក្នុងករណីដែលថ្ងៃទី 5 លក់បាន 101ដុល្លាខាងលើ។

- ក្នុងករណីដែល n ជាចំនួនគត់ រូបមន្តមេដ្យានត្រូវបានកែសំរួលដូចខាងក្រោម៖

$$Median\ X(if\ n\ is\ odd\ number) = \frac{X_{n/2}+X_{(n/2)+1}}{2} \qquad (3.4b)$$

ឧទាហរណ៍ 3.4 ចំណូលរបស់អ្នកលក់ម្នាក់ក្នុងរយៈពេល6ថ្ងៃ (n=6) គឺដូចខាងក្រោម៖

ថ្ងៃ	1	2	3	4	5	6
ចំណូល($)	10	15	20	25	105	106

ដោយសារ n =6 គឺជាចំនួនគត់ មេដ្យាន ឬធាតុនៅកណ្តាលគឺមានដល់ទៅពីរ គឺធាតុទី 3 និងទី4ដូច្នេះយើងត្រូវប្រើរូបមន្ត (3.4b) នឹងត្រូវបានអនុវត្ត។

$$Median\ X(if\ n\ is\ odd\ number) = \frac{X_{6/2}+X_{(6/2)+1}}{2} = \frac{X_{(3+4)}}{2} = X_{3.5}$$

ដូច្នេះមេដ្យានគឺជាធាតុទី3.5 ឬ $=\frac{(20+25)}{2} = \$22.5$

3.2.5 ម៉ូដ

ម៉ូដ(Mode)គឺជាតម្លៃដែលកើតញឹកញាប់ជាងគេ។ ម៉ូដមានសារៈសំខាន់ចំពោះអថេរដាច់ ឬអថេរប្រភេទ (discrete or categorical variable) តែបើអថេរជាប់ ឬមានក្បៀសវិញ គឺវាមិនសូវត្រូវបានប្រើប្រាស់ឡើយ។ ដើម្បីរកម៉ូដ យើងត្រូវរៀបចំទិន្នន័យនោះតាមលំដាប់លំដោយជាមុនសិន រួចយើងធ្វើការអង្កេតទៅលើធាតុមួយណាដែលកើតញឹកញាប់ជាគេ។

ឧទាហរណ៍ 3.5 ការលក់ទឹកត្រចៀកកាំចំនួន 7 ថ្ងៃរបស់ Dany ត្រូវបានរៀបចំតាមលំដាប់លំដោយដូចខាងក្រោម៖

ថ្ងៃ	1	2	3	4	5	6	7
ចំណូល($)	10	12	12	12	15	15	16

ឧទាហរណ៍ 3.5 បញ្ជាក់ថា ក្នុងចំណោមការលក់ចំនួន 7ថ្ងៃ មាន 3ថ្ងៃលក់បាន 12ដុល្លា និត 2ថ្ងៃលក់បាន 15ដុល្លា ដូច្នេះម៉ូដគឺ 12ដុល្លា។ សូមបញ្ជាក់ថា ទិន្នន័យខ្លះ អត់មាន ម៉ូដក៏មានដែរ។

3.3 **រង្វាស់ភាពរាយប៉ាយ**

ជាទូទៅ ភាពរាយប៉ាយ(measure of variation)មានតម្លៃតូច បង្ហាញតម្លៃមធ្យម (mean or average) អាចទុកចិត្តបាន ហើយផ្ទុយមកវិញបើភាពរាយប៉ាយមានតម្លៃធំ គឺបញ្ជាក់ថាតម្លៃមធ្យមមិនគួរទុកចិត្តនោះទេ ។ រង្វាស់ភាពរាយប៉ាយរួមមាន៖ រ៉ង់ (range) វ៉ារីយ៉ង់(variance) គម្លាតស្តង់ដារ (standard deviation) និងមេគុណបម្រែបម្រួល (coefficient of variation)

3.3.1 រ៉ង់

រ៉ង់ (range) គឺជាគម្លាតពីធំទៅតូចដែលវាស់ភាពរាយប៉ាយដ៏សាមញ្ញមួយ។ រ៉ង់ត្រូវបានគណនាដោយយកតម្លៃធំជាងគេ (max) ដកនិងតម្លៃតូចជាងគេ (min)។ រូបមន្តរ៉ង់ដូចខាងក្រោម៖

Range = $X_{largest} - X_{smallest}$ (3.5)

ឧទាហរណ៍ 3.6 ចំណូលរបស់និស្សិត 11 រូប (គិតជាដុល្លា)ត្រូវបានកត់ត្រាដូចខាងក្រោម៖

0	1	2	3	4	5	6	7	8	9	10

Range = $Max - Min = 10 - 0 = 10$

ដូច្នេះគម្លាតនៃចំណូលពីធំទៅតូចរបស់និស្សិតទាំង11រូបគឺ 10ដុល្លា។

3.3.2 វ៉ារីយ៉ង់និងគម្លាតស្តង់ដារ

- វ៉ារីយ៉ង់នៃប៉ាន់គំរូ (sample variance) គឺជាគម្លាតចេញពីតម្លៃមធ្យមគិតជាការេ។ បើតម្លៃវ៉ារីយ៉ង់នៃប៉ាន់គំរូកាន់តែធំគឺបង្ហាញថា ទិន្នន័យមានភាពរាយប៉ាយកាន់តែច្រើន។ រូបមន្តវ៉ារីយ៉ង់នៃប៉ាន់គំរូត្រូវបានកំណត់ដូចតទៅៈ

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \qquad (3.6)$$

ហើយ $X_i =$ តម្លៃ X ដែលបានអង្កេត

$\bar{X} =$ តម្លៃមធ្យមរបស់ $X = \frac{\sum X_i}{n}$

$n =$ ចំនួនដែលបានអង្កេត

- គម្លាតស្តង់ដារ (standard deviation) គឺជាឫសការេនៃវ៉ារីយ៉ង់

$$S=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}} \qquad (3.7)$$

ឧទាហរណ៍ ចំណូលរបស់និស្សិតទាំង 11រូបខាងលើ យើងអាចគណនាវ៉ារីយ៉ង់ដោយប្រើរូបមន្ត 3.6 ហើយបានលទ្ធផលដូចក្នុងតារាង 3.1។

តារាង 3.1 នីតិវិធីគណនាវ៉ារីយ៉ង់ចំណូលរបស់និស្សិត 11នាក់

n	X_i	$\bar{X}$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	0	5	-5	25
2	1	5	-4	16
3	2	5	-3	9
4	3	5	-2	4
5	4	5	-1	1
6	5	5	0	0
7	6	5	1	1
8	7	5	2	4
9	8	5	3	9
10	9	5	4	16
11	10	5	5	25
Total	55			110

ដើម្បីគណនាវ៉ារីយ៉ង់ ជំហ៊ានដំបុងគេត្រូវគណនាតម្លៃមធ្យមជាមុនសិន។

➢ គណនាតម្លៃមធ្យមៈ $\bar{X} = \frac{(0+1+\cdots+10)}{11} = \frac{55}{11} = \$5$ (មើល column ទី3)

- គណនាវ៉ារីយ៉ង់ $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} = \frac{110}{11-1} = \$11$

មានន័យថាគម្លាតចំណូលគិតជាមធ្យមរបស់និស្សិតតាំង11នាក់គឺ11ដុល្លាការេ។ ជាទូទៅវ៉ារីយ៉ង់ពិបាកក្នុងការបកស្រាយ ដូច្នេះដើម្បីបកស្រាយឲ្យងាយយល់ គេបង្វែរវ៉ារីយ៉ង់ទៅជាគម្លាតស្តង់ដាវិញ។

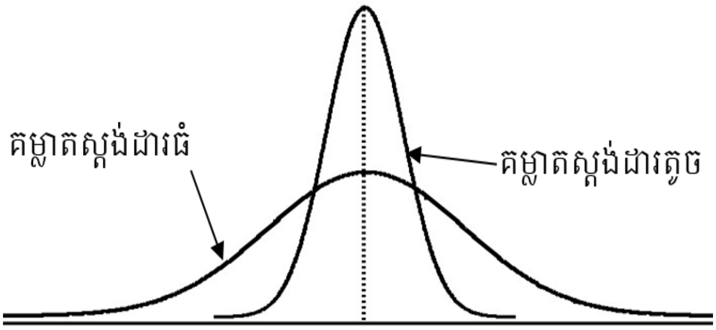
- ដូច្នេះគម្លាតស្តង់ដារនៃចំណូលរបស់និស្សិតទាំង11រូបគឺ៖

S=$\sqrt{\frac{110}{11-1}} = \$3.32$

សរុបមក ចំណូលរបស់និស្សិតទាំង11រូប គិតជាមធ្យមគឺ 5ដុល្លា ហើយគម្លាតស្តង់ដារ (គម្លាតចេញពីមធ្យម) គឺ3.32ដុល្លា។

ជាទូទៅគម្លាតស្តង់ដារកាន់តែធំបញ្ជាក់ថា ភាពរាយប៉ាយចេញពីតម្លៃមធ្យមកាន់តែឆ្ងាយ (សូមមើលរូប 3.1)។

រូប 3.1 ការប្រៀបធៀបគម្លាតស្តង់ដារ និង ខឺតូស៊ីស



តើគម្លាតស្តង់ដារប៉ុន្មានទើបចាត់ទុកថាធំ? ដើម្បីឆ្លើយនឹងសំណួរនេះ យើងអាចយកតម្លៃគម្លាតស្តង់ដារធៀបជាមួយតម្លៃមធ្យម។ សម្រាប់ឧទាហរណ៍ 3.6 តម្លៃមធ្យមគឺ 5 ដុល្លា រីឯគម្លាតស្តង់ដារវិញគឺ 3.32 ដុល្លា ហាក់ដូចជាធំណាស់ វាប្រហែលជា 66%នៃតម្លៃមធ្យម ដូច្នេះគម្លាតស្តងដារបែបនេះគឺត្រូវបានចាត់ទុកថាធំមែនទែន។

វិធានអក្សរសាស្ត្រ

ទិន្នន័យភាគច្រើនបានប្រមូលផ្តុំជុំវិញមេដ្យាន។ ក្នុងករណីដែលទិន្នន័យធៀងទៅខាងឆ្វេង ទិន្នន័យបានភាគច្រើនប្រមូលផ្តុំនៅខាងឆ្វេងហើយតម្លៃទាំងនោះគឺតូចជាងតម្លៃមធ្យម។ ក្នុងករណីដែលទិន្នន័យធៀងទៅខាងស្តាំ ទិន្នន័យបានប្រមូលភាគច្រើនប្រមូលផ្តុំនៅខាងស្តាំហើយតម្លៃទាំងនោះគឺធំជាងតម្លៃមធ្យម។ ក្នុងករណីស៊ីមេទ្រីគឺមេដ្យានស្មើទៅនឹងតម្លៃមធ្យម ហើយទិន្នន័យភាគច្រើនប្រមូលផ្តុំនៅត្រង់មេដ្យាននិងតម្លៃមធ្យមដែលបង្កើតបានបំណែងចែកមួយរាងរូបដូចជួង (មើលរូប3.4ខាងក្រោម)។ ករណីដែលទិន្នន័យមានបំណែងចែកមួយរាងរូបដូចជួងនោះ អ្នកអាចប្រើវិធានអក្សរសាស្ត្របាន (Empirical Rule) ហើយវិធាននេះបានកំណត់ថា៖

- ប្រហែលជា 68% នៃតម្លៃរបស់ទិន្នន័យស្ថិតនៅក្នុងចន្លោះ ± 1 ស្តង់ដារចេញពីតម្លៃមធ្យម

- ហែលជា 95% នៃតម្លៃរបស់ទិន្នន័យស្ថិតនៅក្នុងចន្លោះ ± 2 ស្តង់ដារចេញពីតម្លៃមធ្យម
- ហែលជា 99.7% នៃតម្លៃរបស់ទិន្នន័ស្ថិយតនៅក្នុងចន្លោះ ± 3 ស្តង់ដារចេញពីតម្លៃមធ្យម

ក្រៅពីប្រើសម្រាប់ធ្វើការបកស្រាយគម្លាតចេញពីតម្លៃមធ្យម វិធានអក្សរសាស្ត្រអាចត្រូវបានគេប្រើផងដែរដើម្បីកំណត់ជាតុមិនប្រក្រតី (outlier)។ ជាទូទៅបើជាតុណាមួយនៃទិន្នន័យដែលមិនស្ថិតនៅចន្លោះ $\mu \pm 3\sigma$ ត្រូវបានចាតទុកថា ជាតុមិនប្រក្រតីឬ outlier។

ឧទាហរណ៍ ចំនួនសកលនៃធុរេនកំពតប្រភេទលេខមួយមានទម្ងន់ μ= 4 គីឡូក្រាមក្នុងមួយផ្លែ ហើយមានគម្លាតស្តង់ដាររបស់ចំនួនសកល σ= 0.2គីឡូក្រាម។ ចូលគណនាចន្លោះ ± 1, ± 2, និង ± 3 ស្តង់ដារចេញពីតម្លៃមធ្យម ហើយបើសិនជាមានធុរេនមួយផ្លែទម្ងន់ 4.8 គីឡូក្រាម បើសិនគម្លាត ± 3 ស្តង់ដារត្រូវបានគេប្រើ តើធុរេននោះទម្ងន់មិនប្រក្រតីមែនទេ?

- $\mu \pm 1\sigma = 4 \pm 1(0.2) =$ (3.8, 4.2)
- $\mu \pm 2\sigma = 4 \pm 2(0.2) =$ (3.6, 4.4)
- $\mu \pm 3\sigma = 4 \pm 3(0.2) =$ (3.4, 4.6)

បកស្រាយវិធានអក្សរសាស្ត្រ:

- ប្រហែល 68% ទម្ងន់ធុរេនគឺស្ថិតនៅចន្លោះ 3.8 ទៅ 4.2 គីឡូក្រាម
- ប្រហែល 95% ទម្ងន់ធុរេនគឺស្ថិតនៅចន្លោះ 3.6 ទៅ 4.4 គីឡូក្រាម
- ប្រហែល 99.7% ទម្ងន់ធុរេនគឺស្ថិតនៅចន្លោះ 3.4 ទៅ 4.6 គីឡូក្រាម

ដោយសារតែគម្លាត 3ស្តង់ដារគឺនៅចន្លោះ 3.4 ទៅ 4.6 គីឡូក្រាម ហើយចន្លោះនេះមិនមានតម្លៃ 4.8 គីឡូក្រាម ដូច្នេះផ្លែធុរេនដែលមានទម្ងន់ 4.8 គីឡូក្រាម គឺជាផ្លែធុរេនពិសេសដែលត្រូវបានចាត់ទុកថាមានទម្ងន់មិនប្រក្រតី។

ទ្រឹស្តី The Chebyshev

ក្នុងករណីដែលទិន្ន័យមានលក្ខណៈធៀងខ្លាំង គួរប្រើទ្រឹស្តីរបស់ Chebyshev។ វិធាននេះបានចែងថា គ្រប់ទម្រង់ទិន្នន័យទាំងអស់ មិនទាក់ទងជាមួយបំណែងចែករបស់វានោះទេ តម្លៃគិតជាភាគរយនៃទិន្នន័យគឺស្ថិតនៅចម្ងាយ k ស្តង់ដារយ៉ាងតិចចេញពីតម្លៃមធ្យម។

$(1 - \frac{1}{K^2})$x 100%

អ្នកអាចទ្រឹស្តីនេះគ្រប់តម្លៃ K >1 (មិនបាច់ជាចំនួនគត់នោះទេ)។ ទ្រឹស្តីនេះបានចែងថា គ្រប់សំណុំទិន្នន័យជាលេខ[4]៖

- K=2: យ៉ាងហោចណាស់ $(1 - \frac{1}{2^2})$x 100%= 75% នៃទិន្នន័យគឺស្ថិតនៅក្នុងចន្លោះ 2 ស្តង់ដារនៃតម្លៃមធ្យម។ ហើយចន្លោះជឿជាក់សម្រាប់ប៉ាន់គំរូគឺ$\bar{X}$±2s ហើយសម្រាប់ចំនួនសកលគឺ ($\mu \pm 2\sigma$)

[4] ទ្រឹស្តី The Chebyshev ទាញយកនៅថ្ងៃទី 1/4/2021 ចេញពី https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_Introductory_Statistics_(Shafer_and_Zhang)/02%3A_Descriptive_Statistics/2.05%3A_The_Empirical_Rule_and_Chebyshev's_Theorem#:~:text=Chebyshev's%20Theorem%20is%20a%20fact,standard%20deviations%20of%20the%20mean.

- K=3: យ៉ាងហោចណាស់ $(1-\frac{1}{3^2})$x 100%= 88.9% នៃទិន្នន័យគឺស្ថិតនៅក្នុងចន្លោះ 3 ស្តង់ដារនៃតម្លៃមធ្យម។ ហើយចន្លោះជឿជាក់សម្រាប់ប៉ាន់គំរូគឺ$\bar{X}$±3s ហើយសម្រាប់ចំនួនសកលគឺ ($\mu \pm 3\sigma$)

ឧទាហរណ៍ផ្លែធុរេនខាងលើ μ=4 គីឡូក្រាម និងគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ s=0.2 គីឡូក្រាម។

- 75% នៃទម្ងន់ផ្លែធុរេនគឺនៅចន្លោះ 4 ± 2(0.2) ឬ ចាប់ពី 3.6 ដល់ 4.4 គីឡូក្រាម
- 88.9%នៃទម្ងន់ផ្លែធុរេនគឺនៅចន្លោះពី 4 ± 3(0.2) ឬ ចាប់ពី 3.4 ដល់ 4.6 គីឡូក្រាម

3.3.3 មេគុណនៃភាពរាយប៉ាយ

មេគុណនៃភាពរាយប៉ាយ (coefficient of variation, CV)គឺជារង្វាស់នៃភាពរាយប៉ាយដែលចេញពីតម្លៃមធ្យម។ ជាទូទៅមេគុណនេះត្រូវបានប្រើប្រាស់ដើម្បីប្រៀបធៀបម្រែបម្រួលនៃសំណុំទិន្នន័យពីរឬច្រើនដែលមានគម្លាតស្តង់ដារដូចគ្នា ឬរង្វាស់ខុសគ្នា និងត្រូវបានគណនាជាភាគរយ។ រូបមន្តមេគុណនៃភាពរាយប៉ាយគឺដូចខាងក្រោម៖

$$CV_i = \frac{S_i}{\bar{X}_i} * 100\% \qquad (3.8)$$

CV_i= មេគុណនៃភាពរាយប៉ាយរបស់វត្ថុណាមួយ (coefficient of variation of i item)

S_i= គម្លាតស្តង់ដាររបស់វត្ថុណាមួយ (standard deviation of i item)

$\bar{X}_i$= តម្លៃមធ្យមរបស់វត្ថុណាមួយ (average value of i item)

CV_i គឺជាប្រភេទអនុបាតអាក្រក់ (bad ratio) មានន័យថា តម្លៃ CV_i កាន់តែធំកាន់តែអន់ ឬអាក្រក់។ ឧបមាថា តម្លៃភាគហ៊ុន A ជាមធ្យម 50ដុល្លានិងតម្លៃភាគហ៊ុន B ជាមធ្យម 100ដុល្លា ហើយតម្លៃភាគហ៊ុន A និង B មានគម្លាតស្តង់ដាដូចគ្នាគឺ 5ដុល្លា។ តើភាគហ៊ុនមួយណាមានហានិភ័យខ្ពស់ជាង? សូមបញ្ជាក់ថានៅក្នុងទីផ្សារហិរញ្ញវត្ថុ គម្លាតស្តង់ដារគឺជារង្វាស់ហានិភ័យសរុប (total risk)។ ដើម្បីឆ្លើយនឹងសំណួរនេះ យើងត្រូវគណនា មេគុណនៃភាពរាយប៉ាយរបស់ភាគហ៊ុនទាំងពីរជាមុនសិន។

$$CV_A = \frac{5}{50} * 100\% = 10\%$$

$$CV_B = \frac{5}{100} * 100\% = 5\%$$

បើយើងពិនិត្យលើគម្លាតស្តង់ដារតែមួយមុខ យើងអាចសន្និដ្ឋានថា ភាគហ៊ុនទាំងពីរមានហានិភ័យស្មើគ្នា (គម្លាតស្តង់ស្មើគ្នា) តែបើយើងពិនិត្យឲ្យបានឱតដល់ ភាគហ៊ុនA មានហានិភ័យខ្ពស់ជាងភាគហ៊ុន B ពីព្រោះ $CV_A > CV_B$។

3.3.4 សតភាគ ចតុភាគ ចន្លោះចតុភាគ និង គំនូសប្រអប់

a. **សតភាគ** (percentile) គឺជាស្តង់ដាមួយដែលត្រូវបានប្រើប្រាស់ក្នុងគោលបំណងដើម្បីធ្វើការប្រៀបធៀបគម្លាតមួយ ឧទាហរណ៍ ការប្រៀបធៀបប្រាក់បៀវត្សបុគ្គលិក ពិន្ទុរបស់និស្សិត សុខភាពសាធារណៈ ធនាគារជាដើម។ ជាទូទៅ ទិន្នន័យត្រូវបានគេរៀបតាមលំដាប់លំដោយពីតូចទៅធំ ហើយអង្កន់ជា 100ភាគស្មើគ្នា ។ សូមកុំច្រឡំសតភាគ គឺជាភាគរយ[5]។ សតភាគចំពោះតម្លៃ X ត្រូវបានគណនាតាមរូបមន្តដូចតទៅ

[5] Allan G.B (2009) Statistics: A Step-by-Step Approach. Eighth Edition. McGraw-Hill.

$$\text{សតភាគ} = \frac{(\text{ចំនួនតម្លៃក្រោម } x) + 0.50}{\text{ចំនួនតម្លៃសរុប}} * 100 \qquad (3.9)$$

ឧទាហរណ៍ 3.7 គ្រូម្នាក់បានអោយពិន្ទុសិស្សចំនួន 10 នាក់ (ពិន្ទុ20/20) ដូចខាងក្រោម៖

17	15	11	5	8	2	4	5	20	12

ចូលគណនា សតភាគលំដាប់ (percentile rank) នៃពិន្ទុស្មើ 15?

- **ជំហានទី1:** រៀបចំទិន្នន័យតាមលំដាប់ពីតូចទៅធំ

2	4	5	5	8	11	12	15	17	20

- **ជំហានទី2:** គណនាសតភាគដោយប្រើរូបមន្ត (3.9)

ដោយសារ ចំនួនលេខនៅក្រោម X=15 ពិន្ទុគឺមាន 7លេខ ដោយរាប់ពីឆ្វេងទៅស្តាំ ដូច្នេះយើងបាន៖

$$\text{សតភាគ} = \frac{(7)+0.50}{10} * 100 = 75(\text{ទី}75\text{សតភាគ})$$

- **ជំហាន**ទី3: បំណកស្រាយ

ដូច្នេះចំពោះនិស្សិតដែលបានពិន្ទុ 15 គឺប្រសើរជាង 75 ភាគរយនៃនិស្សិតក្នុងថ្នាក់ដែលមានគ្នា10 នាក់។

b. **ចតុភាគ** (quartile) ស្រដៀងទៅនឹងសតភាគដែរ ប៉ុន្តែទិន្នន័យត្រូវបានគេបែងចែកបួនភាគស្មើៗគ្នា៖ 25% 50% និង75% ដូចរូប 3.2 ខាងក្រោម។

រូប 3.2 ចតុភាគ

Lower	Q_1	Second	Q_2	Third	Q_3	Upper
⇐ 25%⇒	I	⇐ 25% ⇒	II	⇐ 25%⇒	III	⇐ 25% ⇒

ចតុភាគ ជាញឹកញាប់ត្រូវបានគេប្រើប្រាស់ដើម្បីវាយតម្លៃទៅលើដំណើរការហិរញ្ញវត្ថុរបស់ធនាគារ និងភាគហ៊ុនជាដើម។ ដើម្បីគណនាចតុភាគ គេត្រូវរៀបចំទិន្នន័យតាមលំដាប់លំដោយពីតូចទៅធំ ហើយបន្ទាប់មកគេធ្វើការកំណត់នូវតម្លៃ៖

- Q_1តំណាងទិន្នន័យតូចជាង25%
- Q_2 តំណាងទិន្នន័យតូចជាង 50% (តំណាងមេដ្យាន)
- Q_3 តំណាងទិន្នន័យតូចជាង75%

ឧទាហរណ៍ 3.8 អ្នកវិភាគខាងហិរញ្ញវត្ថុម្នាក់ បានធ្វើការវិភាគទៅលើសំពៀតវិនិយោគមួយ (portfolio) ដែលមានប្រាក់ចំណេញរបស់ក្រុមហ៊ុនចំនួន12 ។ ប្រាក់ចំណេញ(ដុល្លា)ក្នុងមួយហ៊ុនរបស់ក្រុមហ៊ុននីមួយត្រូវបានរៀបតាមលំដាប់លំដោយពីតូចទៅធំដូចខាងក្រោម៖

Lower		Q_1	Second		Q_2	Third		Q_3	Upper		
⇐ 25%⇒			⇐ 25% ⇒			⇐ 25%⇒			⇐ 25% ⇒		
8	12	21	29	31	34	36	38	40	42	45	48

- ជំហានទី1៖ រៀបចំទិន្នន័យតាមលំដាប់លំដោយពីតូចទៅធំ (ដូចទិន្នន័យខាងលើ)
- ជំហានទី2៖ គណនាតម្លៃ Q_1, Q_2, Q_3
    - តម្លៃ Q_1គឺនៅចន្លោះ X_3 និង X_4, Q_1=(21+29)/2 =$25
    - តម្លៃ Q_2គឺនៅចន្លោះ X_6 និង X_7, Q_2=(34+36)/2 =$35
    - តម្លៃ Q_3គឺនៅចន្លោះ X_9 និង X_{10}, Q_3=(40+42)/2 =$41
- ជំហានទី3៖ បំណកស្រាយ
    - $Q_1 = \$25$ បញ្ជាក់ថា 25% ប្រាក់ចំណេញក្នុងមួយហ៊ុន ចំណេញមិនដល់ 25 ដុល្លា
    - $Q_2 = \$35$ បញ្ជាក់ថា 50% ប្រាក់ចំណេញក្នុងមួយហ៊ុន ចំណេញមិនដល់ 35 ដុល្លា
    - $Q_3 = \$41$ បញ្ជាក់ថា 75% ប្រាក់ចំណេញក្នុងមួយហ៊ុន ចំណេញមិនដល់ 41 ដុល្លា

ដើម្បីឲ្យដឹងថាសំពៀតវិនិយោគមួយណាប្រើសើរជាង អ្នកអាចយកតម្លៃ Q ទាំងនេះប្រៀបធៀបជា មួយនិង តម្លៃ Q នៃសំពៀតវិនិយោគដ៏ទៃទៀត។ ឧបមាថា កញ្ចប់សំពៀតមួយទៀតដែលមានតម្លៃ $Q_3 = 50$ ដុល្លា ដូច្នេះយើងអាចសន្និដ្ឋានបានថា សំពៀតទី2គឺល្អជាសំពៀតទីមួយពីព្រោះ 75%នៃសំពៀតទី2គឺចំណេញបានរហូតដល់ 50 ដុល្លា ហើយ75%នៃសំពៀតទី1គឺចំណេញបានតែ 41ដុល្លាតែប៉ុណ្ណោះ។

របៀបរកទីតាំង Q_1និង Q_3 ចំពោះទិន្នន័យដែលបានរៀបចំតាមលំដាប់លំដោយពីតូចទៅធំរួចរាល់ យើងអាចប្រើរូបមន្តដូចខាងក្រោម៖

$$Q_1 = \frac{(n+1)}{4} \text{ តម្លៃតាមលំដាប់លំដោយ} \quad (3.10)$$

$$Q_3 = \frac{3(n+1)}{4} \text{ តម្លៃតាមលំដាប់លំដោយ} \quad (3.11)$$

ហើយ n=លេខទីតាំង

- វិធានទី1: ឧបមាថា n=11 លេខគត់ ដូច្នេះ $Q_1 = \frac{11+1}{4} = 3$ ជាលេខដែលមានលេខរៀងទី3
- វិធានទី2: បើតម្លៃរបស់Q_1 ជាចំនួនមានក្បៀស ត្រូវបង្គ្រប់លេខនោះទៅជាលេខគូដែលនៅជិតបំផុត ឧបមាថា n=12 ដូច្នេះ Q= 3.25 បង្គត់ 3
- វិធានទី3: បើតម្លៃរបស់ Q_1 ជាចំនួនមានក្បៀស ឧទាហរណ៍ 2.5, ឬ 4.5, ឬ 7.5។ ឧបមាថា n=9 ដូច្នេះ $Q_1 = 2.5$ ករណីនេះយើងមិនអាចបង្គត់បានទេ គឺតម្លៃ Q_1ត្រូវស្ថិតនៅលំដាប់លេខរៀងទី 2 ទី3។

ឥឡូវយើងត្រឡប់មកឧទាហរណ៍ 3.8 ប្រាក់ចំណេញរបស់ក្រុមហ៊ុនចំនួន12 ខាងលើ ដែលទិន្នន័យត្រូវបានរៀបចំតាមលំដាប់លំដោយពីតូចទៅធំរួចរាល់។ ចូលគណនា Q_1និងQ_3?

$$Q_1 = \frac{(12+1)}{4} = 3.25$$

ដោយសា$Q_1 = 3.25$ យើងបង្គត់3។ ដូច្នេះក្រុមហ៊ុនដែលមានលេរៀងទី3 គឺមានប្រាក់ចំណេញ 21 ដុល្លាក្នុងមួយហ៊ុន (មើលរូប 3.2 ត្រង់ X មានលេខរៀងទី3) ហើយយើងដឹងហើយថា Q_1 គឺតំណាង25% ដូច្នេះយើងអាចបកស្រាយបន្តទៀតថា 25%នៃក្រុមហ៊ុនទាំងនោះមានប្រាក់ចំណេញតិចជាង 21ដុល្លាក្នុងមួយហ៊ុន ឬ75% នៃក្រុមហ៊ុនទាំងនោះមានប្រាក់ចំណេញច្រើនជាង 21ដុល្លាក្នុងមួយហ៊ុន។

$$Q_3 = \frac{3(12+1)}{4} = 9.75$$

ដោយសា$Q_3 = 9.75$ យើងបង្គត់10។ ដូច្នេះក្រុមហ៊ុនដែលមានលេរៀងទី10 គឺមានប្រាក់ចំណេញ 42 ដុល្លាក្នុងមួយហ៊ុន (មើលរូប 3.2 ត្រង់ X មានលេខរៀងទី10) ហើយយើងអាចបកស្រាយថា 75% នៃក្រុមហ៊ុនទាំងនោះមានប្រាក់ចំណេញតិចជាង 42ដុល្លាក្នុងមួយហ៊ុន ឬ25%នៃក្រុមហ៊ុនទាំងនោះមានប្រាក់ចំណេញច្រើនជាង 42ដុល្លាក្នុងមួយហ៊ុន។

c. **គម្លាតចតុភាគ** (interquartile range ឬ mid-spread) គឺជាតម្លៃខុសគ្នារវាង Q_3 និងQ_1 នៃសំណុំទិន្នន័យ។ វាតំណាងអោយ 50%នៃសំណុំទិន្នន័យដែលស្ថិតនៅកណ្តាល ហើយវាមិនរងឥទ្ធិពលដោយជាតុមិនប្រកត្តី ឬ outlier នោះទេ។

គម្លាតចតុភាគ ($Interquartile\ Range, IQR$)

$$IQR = Q_3 - Q_1 \qquad (3.12)$$

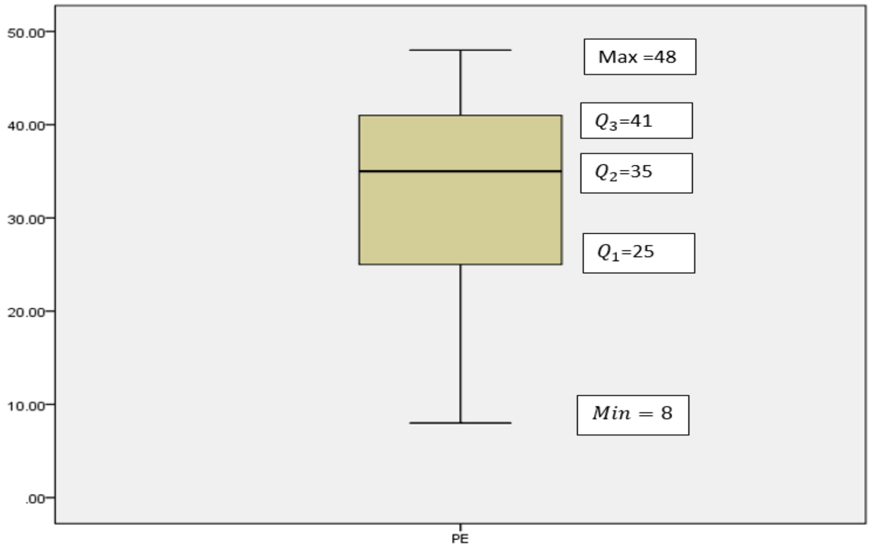
ឥឡូវយើងត្រឡប់មកឧទាហរណ៍ 3.8 របស់យើង។ ប្រាក់ចំណេញរបស់ក្រុមហ៊ុនចំនួន12 គឺ $Q_3 = \$42$ និង$Q_1 = \$21$ ដូច្នេះចន្លោះចតុភាគ IQR= 42-21 =$ 21 ឬ យើងអាចនិយាយម៉្យាងទៀតថា 50% នៃប្រាក់ចំណេញរបស់ក្រុហ៊ុនគឺស្ថិតនៅចន្លោះ 21ទៅ42ដុល្លា ឬមានគម្លាតចតុភាគ 21ដុល្លា។ ជាទូទៅបើគម្លាតចតុភាគកាន់តែធំ មានន័យថាទិន្នន័យរាយប៉ាយកាន់តែខ្លាំងចេញពីមេដ្យាន ឬគម្លាតស្តង់ដារកាន់តែខ្ពស់ ជាហេតុធ្វើអោយតម្លៃមធ្យម (mean) មិនអាចទុកចិត្តបាន។

d. **គំនូសប្រអប់**

គំនូសប្រអប់ (Box Plox) គឺជាវិធីសាស្រ្តបែងចែកទិន្នន័យជាលេខទៅតាមចតុភាគ (quartile) ដោយដាក់បង្ហាញវានៅលើក្រាហ្វិក។ វាជាឧបករណ៍ដ៏មានសារៈសំខាន់ក្នុងការរុករកទិន្នន័យ តួយ៉ាងដូចការរុកជាតុមិនប្រក្រត្តី ឬ outlier។ ដើម្បីសង់គំនូសប្រអប់ គេត្រូវពឹងផ្អែកលើទិន្នន័យចំនួន 5៖ តម្លៃអប្បរមា, Q_1, Q_2, Q_3 និង អតិបរមា។ បើយើងយកទិន្នន័យប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12 រួចដាក់ក្នុងគំនូសប្រអប់ យើងនិងបាន Box Plox ដូចរូប 3.3។

តាមរយៈរូប3.3 បញ្ជាក់អោយឃើញថា មេដ្យាន($Q_3 = \$35$) និងតម្លៃមធ្យម(mean=$32)មិនត្រួតស៊ីគ្នានោះទេ ដូច្នេះយើងអាចសន្និដ្ឋានបានថាទិន្នន្ន័យនេះអស៊ីមេទ្រី (តម្លៃមធ្យមនិងមេដ្យានមិនស្មើគ្នា) ហើយក៏ប៉ុន្តែតម្លៃមធ្យមខុសគ្នាជាមួយតម្លៃមេដ្យានតែ2ដុល្លាតែប៉ុណ្ណោះ។ ហើយតាមរយៈរូប3.3 ដែលនេះក៏បានបង្ហាញឲ្យដឹងថា សំណុំទិន្នន័យហាក់ដូចជាឆៀងទៅខាងឆ្វេងច្រើនជាង (negative skewed)។

រូបទី 3.3គំនូសប្រអប់នៃប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12



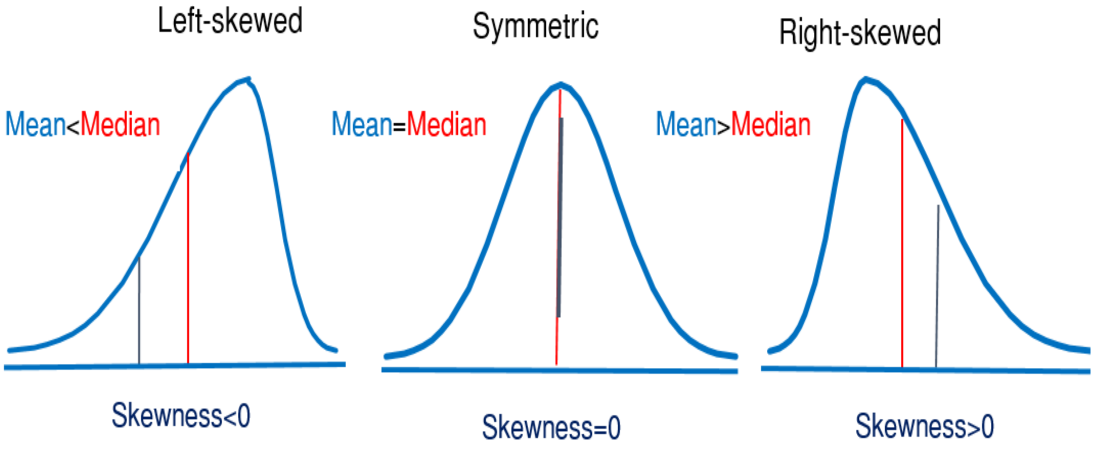
3.4 **បំណែងចែកទម្រង់ទិន្នន័យ**

បំណែងចែកទម្រង់ទិន្នន័យ (shape of distribution) គឺជារូបភាពទាំងមូលនៃបំណែងចែកតម្លៃទិន្នន័យ។ ពេលខ្លះបំណែងចែកតម្លៃទិន្នន័យអាចស៊ីមេទ្រី (symmetry) អស៊ីមេទ្រី(asymmetry) ឬធៀង (skewness)។ ស៊ីមេទ្រីគឺមានន័យថា តម្លៃនៅក្រោមតម្លៃមធ្យម និងលើមធ្យមគឺមានបរិមាណស្មើគ្នា ឬសមតុល្យ ឬ តម្លៃធៀងស្មើសូន្យ (skewness=0)។ ក្នុងករណីទិន្នន័យស៊ីមេទ្រី តម្លៃមធ្យម ស្មើនិងមេដ្យាន (mean= median) ហើយការប្រើប្រាស់តម្លៃមធ្យមគឺអាចទុកចិត្តបាន។

ភាពធៀង(skewness) និង ខឺតូស៊ីស (kurtosis) គឺជាទម្រង់ពីរទាក់ទងជាមួយស្ថិតិ។ ខឺតូស៊ីសវាស់អំពីភាពប្រមូលផ្តុំនៅកណ្តាលនៃសំណុំទិន្នន័យ រីឯភាពធៀងវាស់អំពីភាពស៊ីមេទ្រី។ ភាពធៀងអាចមានតម្លៃវិជ្ជមាន ពេលន្ទុយផ្នែកផ្នែកខាងស្តាំដៃវែងជាងខាងឆ្វេង (right skewed) និងអវិជ្ជមាន ពេលកន្ទុយផ្នែកឆ្វេងវែងជាងខាងស្តាំ (left skewed) (មើលរូបភាព 3.2)។ ដោយឡែក ខឺតូស៊ីស (kurtosis) គឺវាស់វែងអំពីកំពស់នៃបំណែងចែក (មើលរូបភាព 3.1)។ បើតម្លៃខឺតូស៊ីសកាន់តែធំ កំពស់កាន់តែខ្ពស់ តម្លៃរបស់ទិន្ន័យភាគច្រើនប្រមូលផ្តុំនៅត្រង់កណ្តាលឬតម្លៃមធ្យម ហើយផ្ទុយមកវិញ តម្លៃខឺតូស៊ីសកាន់តែតូច កំពស់ទាប តម្លៃរបស់ទិន្ន័យភាគច្រើនឃ្លាតឆ្ងាយពីតម្លៃមធ្យម ។ ដើម្បីអោយបំណែងចែកស៊ីមេទ្រី តម្លៃស្ថិតិខឺតូស៊ីសស្មើសូន្យ[6]។ ដោយឡែកសម្រាប់ ចាបេរ៉ាតេស្ត (Jarque-Bera test) ជាប្រភេទតេស្តរកបំណែងចែក ធម្មតា (normal distribution) គឺតម្រូវអោយតម្លៃខឺតូស៊ីសស្មើ3.0 ហើយវិជ្ជមានជានិច្ច។ JB-test មាននៅក្នុងកម្មវិធី Eviews ប៉ុន្តែនៅក្នុង SPSS ឬ Excel ដើម្បីអោយបំណែងចែកធម្មតានោះ គឺតម្រូវអោយតម្លៃ kurtosis ស្មើសូន្យ ហើយអាចអវិជ្ជមាន ដូច្នេះតម្លៃស្ថិតិខឺតូស៊ីស (kurtosis statistic) អាស្រ័យនឹងប្រភេទកម្មវិធី (software)។

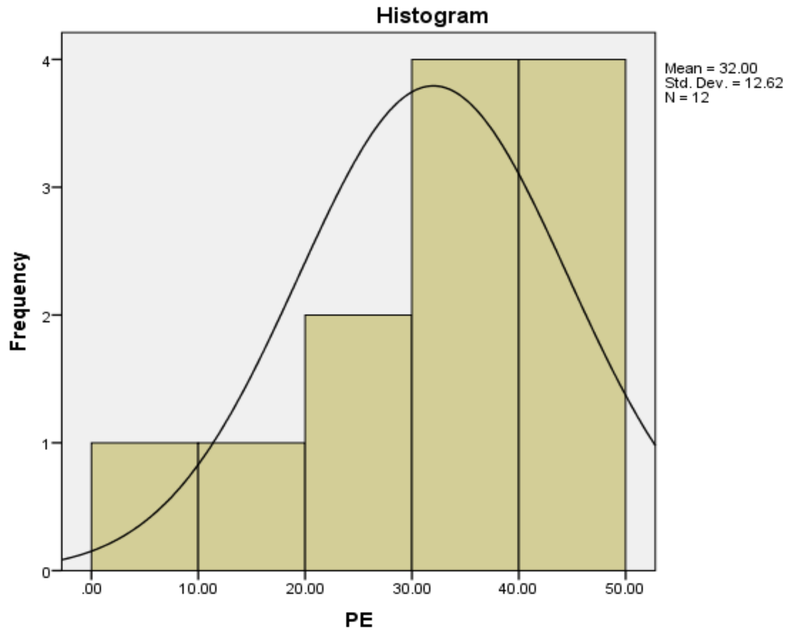
រូបភាព 3.4 ភាពធៀងនិងស៊ីមេទ្រី

[6] Allan G.B (2009) Elementary Statistics: A Step by Step Approach. Eight Edition McGraw-Hill



បើយើងយកទិន្នន័យប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12 មកពិនិត្យមើលបំណែងចែករបស់វា ដោយប្រើ SPSS យើងនិងទទួលបានបំណែងចែកទិន្នន័យដូចរូបភាព 3.5។

រូភាព 3.5 ប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12



រូប 3.5 បញ្ជាក់អោយឃើញថា ប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12 មានសភាពផៀងទៅខាងឆ្វេង តម្លៃស្ថិតិផៀងស្មើនឹង -0.819 ហើយ មេដ្យានស្មើនឹង 35ដុល្លា តម្លៃមធ្យមស្មើនឹង 32ដុល្លា និងខឹតូស៊ីស ស្មើនឹង -0.205 គឺអវិជ្ជមាន (មើលតារាងស្ថិតិសង្ខេប 3.2 ខាងក្រោម)។ ស្ថិតិទាំងនេះបញ្ជាក់ថា ទិន្នន័យនេះគឺអស៊ីមេទ្រី (asymmetric) ឬមិនមានបំណែងចែកធម្មតានោះទេ។ ដូច្នេះតម្លៃមធ្យម មិនសូវទុកចិត្តបានឡើយ ទោះជាយ៉ាងណាក្ដី តម្លៃមធ្យមនេអាចប្រើការបានពីព្រោះវាខុសជាមួយមេដ្យានតែ3ដុល្លាតែប៉ុណ្ណោះ។

តារាង 3.2 ស្ថិតិសង្ខេបនៃប្រាក់ចំណេញរបស់ក្រុមហ៊ុនទាំង12

Descriptives

			Statistic	Std. Error
PE	Mean		32.0000	3.64318
	95% Confidence Interval for Mean	Lower Bound	23.9814	
		Upper Bound	40.0186	
	5% Trimmed Mean		32.4444	
	Median		35.0000	
	Variance		159.273	
	Std. Deviation		12.62033	
	Minimum		8.00	
	Maximum		48.00	
	Range		40.00	
	Interquartile Range		18.50	
	Skewness		-.819	.637
	Kurtosis		-.205	1.232

3.5 កូវ៉ារីយ៉ង់និង មេគុណទំនាក់ទំនង

នៅក្នុងផ្នែកនេះយើងនឹងសិក្សាលើរង្វាស់ពីរដែលទាក់ទងនឹងអថេរជាលេខគឺៈ កូវ៉ារីយ៉ង់ និង មេគុណទំនាក់ទំនង (correlation)។

i) កូវ៉ារីយ៉ង់

កូវ៉ារីយ៉ង់ (covariance) វាស់វែងកំលាំងទំនាក់ទំនងជាបន្ទាត់រវាងអថេរជាលេខចំនួនពីរ (X និង Y) ហើយមានរូបមន្តដូចក្នុង សមីការ 3.13 ។

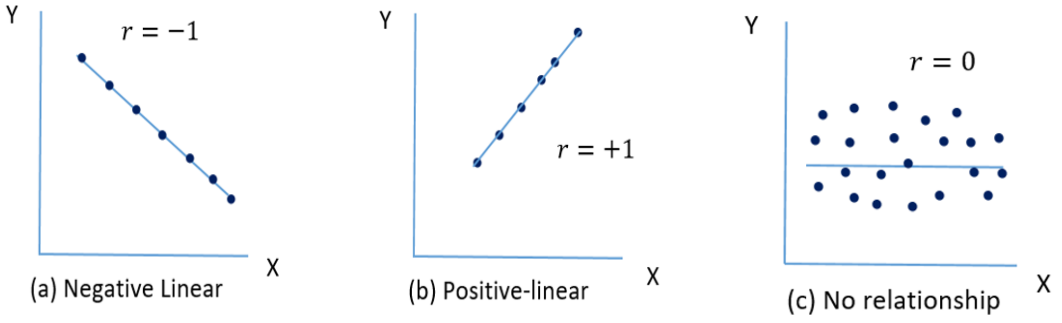
កូវ៉ារីយ៉ង់របស់ប៉ាន់គំរូ $Cov(X, Y)=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$ (3.13)

ii) មេគុណទំនាក់ទំនង

មេគុណទំនាក់ទំនង (the coefficient of correlation, r) វាស់វែងកំលាំងទំនាក់ទំនងជាបន្ទាត់គិតជាភាគរយរវាងអថេរជាលេខចំនួនពីរ (X និង Y)។ វាគឺជាទម្រង់ស្តង់ដាររបស់កូវ៉ារីយ៉ង់ (standardized form of covariance) ហើយមានតម្លៃនៅចន្លោះពី -1 ទៅ +1។

- បើ r=-1 ទំនាក់ទំនងរវាង X និង Y អវិជ្ជមាន 100% (រូប 3.6a)
- បើ r= 0 គ្មានទំនាក់ទំនងរវាង X និង Y (រូប 3.6c)
- បើ r=+1 ទំនាក់ទំនងរវាង X និង Y វិជ្ជមាន 100% (រូប 3.6c)

រូប 3.6 ទម្រង់ផ្សេងៗនៃមេគុណទំនាក់ទំនង



រូបមន្ត 3.14 កំណត់មេគុណទំនាក់ទំនង

មេគុណទំនាក់ទំនងរបស់ប៉ាន់គំរូ: $r = \frac{Cov(X,Y)}{S_X S_Y}$ ((3.14)

- Cov(X, Y) រូបមន្ត (3.13)
- $S_X = \sqrt{\frac{(X_i - \bar{X})^2}{n-1}}$
- $S_y = \sqrt{\frac{(Y_i - \bar{Y})^2}{n-1}}$
- n = ចំនួនប៉ាន់គំរូ

ឧទាហរណ៍ អាត្រាប្រាក់ចំណេញរយៈពេល5ថ្ងៃរបស់ភាគហ៊ុន X និង Y ត្រូវបានកត់ត្រាដូចក្នុងតារាង 3.3 ខាងក្រោម៖

តារាង 3.3 អត្រាប្រាក់ចំណេញរបស់ភាគហ៊ុន X និង Y រយៈពេល 5ថ្ងៃ

ថ្ងៃ	អត្រាចំណេញ របស់ X	អត្រាចំណេញ របស់ Y
1	1.1	3
2	1.7	4.2
3	2.1	4.9
4	1.4	4.1
5	0.2	2.5
មធ្យម	**1.30**	**3.74**

ចូលគណនានិងបកស្រាយ កូវ៉ារីយ៉ង់និងមេគុណទំនាក់ទំនង ?

ចំឡើយ៖

- កូវ៉ារីយ៉ង់: Cov(X, Y) $= \frac{(1.1-1.30)(3-3.74)+(1.7-1.30)(4.2-3.74)+\cdots+(0.20-1.3)(2.5-3.74)}{5-1} = 0.665$
- គម្លាតស្តង់ដារ X: $S_X = \sqrt{\frac{(1.1-1.30)^2+(1.7-1.30)^2+\cdots+(0.20-1.30)^2}{5-1}} = 0.72$
- គម្លាតស្តង់ដារ Y: $S_Y = \sqrt{\frac{(3-3.74)^2+(4.2-3.74)^2+\cdots+(2.5-3.74)^2}{5-1}} = 0.97$

ដូច្នេះមេគុណទំនាក់ទំនង $r = \frac{0.665}{0.72*0.97} = 0.95$

បំណកស្រាយ៖

អថេរ X និងY មានទំនាក់ទំនងវិជ្ជមានខ្លាំងរហូតដល់ 95%។ មានន័យថាបើភាគហ៊ុន X ចំណេញ នោះភាគហ៊ុន Yក៏ចំណេញដែរ ហើយបើខាតគឺខាតទាំងអស់គ្នា។

តារាង 3.4 ខាងក្រោមគឺមេគុណទំនាក់ទំនងដែលជាលទ្ធផលរបស់ Excel។ សូបធ្វើការផ្ទៀងផ្ទាត់លទ្ធផលនេះដោយខ្លួនឯង។

តារាង 3.4 មេគុណទំនាក់ទំនងរបស់ X និងY

	X	Y
X	1	
Y	0.95425	1

សេចក្តីសង្ខេប

ជំពូកនេះបានរៀបរាប់អំពីស្ថិតិបែបបរិយាយ ដែលទាក់ទងជាមួយ តម្លៃមធ្យម គម្លាតស្តង់ដារ មេដ្យាន សតភាគ ចតុភាគ និងចន្លោះចតុភាគ ដែលត្រូវបានគេប្រើប្រាស់ដើម្បីរៀបរាប់អំពីចរិកលក្ខណៈនៃទំនោរកណ្តាល បម្រែបម្រួលនិងទម្រង់។ ក្រៅពីនេះជំពូកនេះក៏បានរៀបរាប់ផងដែរអំពីខឺតូស៊ីស និងភាពធៀងដែលតំណាងឲ្យរបស់បំណែងចែកធម្មតារបស់សំណុំទិន្នន័យមួយ។

នៅជំពូកបន្ទាប់នេះអ្នកនឹងសិក្សាអំពីគោលការណ៍ជាមូលដ្ឋាននៃប្រូបាប៊ីលីតេ ដែលជាស្ពានតភ្ជាប់ពីស្ថិតិបែបបរិយាយទៅនិងស្ថិតតិបែបសន្និដ្ឋាន៕

សំណួរនិងលំហាត់

I. សំណួរ

1. ចូលរៀបរាប់អំពីវិបាករបស់ទ្រឹស្តីប្រមូលផ្តុំកណ្តាល?
2. ចូលរៀបរាប់អំពីលក្ខណៈខុសគ្នារវាងតម្លៃមធ្យម មេដ្យាន និងម៉ូដ និងគុណសម្បត្តិនិងគុណវិបត្តិរបស់វាទាំងបី។
3. ចូលធ្វើការបកស្រាយអំពីចតុភាគទីមួយ ទីពីរនិងទីបី?
4. តើតម្លៃ Z វាស់វែងអំពីអ្វី?
5. ចូលរៀបរាប់ចរិលក្ខណៈរបស់ វ៉េង វ៉ារីយ៉ង់ គម្លាតស្តង់ដារ និងមេគុណបម្រែបម្រួល។
6. តើវិធានអក្សរសាស្រ្តជួយបង្ហាញអំពីវិធីសាស្រ្តដែលទិន្នន័យជាលេខត្រូវបានគេបកស្រាយយ៉ាងដូចម្តេច?
7. តើវិធានអក្សរសាស្រ្តខុសគ្នាជាមួយវិធានរបស់លោក Chebyshev យ៉ាងដូចម្តេចខ្លះ?
8. អ្វីខ្លះជាវិបាកនៃទម្រង់របស់បំណែងចែក (the property of shape)?
9. តើកូវ៉ារីយ៉ង់និងមេគុណទំនាក់ទំនងខុសគ្នាយ៉ាងដូចម្តេច?

II. លំហាត់

10. ខាងក្រោមនេះគឺជាចំណាយក្នុងមួយថ្ងៃគិតជាដុល្លារបស់និស្សិត 10នាក់។

7	5	8	5	6	4	3	5	7	2

a. ចូលគណនាចំណាយជាមធ្យមនិងមេដ្យាន

b. ចូលគណនាវ៉ារីយ៉ង់ គម្លាតស្តង់ដារ វ៉ង់និងមេគុណបម្រែបម្រួល

c. តើទិន្នន័យនេះធៀងដែរឬទេ? បើធៀងតើទៅខាងណា?

d. ផ្អែកលើចំលើយ (a) ដល់ (c), តើអ្នកសន្និដ្ឋានយ៉ាងដូចម្តេច?

11. សន្មត់ថា ធនាគារមួយមានទីតាំងនៅភ្នំពេញមានការងារមមាញឹកខ្លាំង ហើយតែងតែធ្វើឲ្យអតិថិជនមានការរងចាំយូរនៅពេលទទួលសេវាម្តងៗ ។ ខាងក្រោមនេះគឺជារយៈពេលតម្រង់ជួរ(គិតជានាទី) ដែលបានប្រមូលពីអតិថិជន ចំនួន 16 នាក់។

28.06	15.10	18.02	15.40	18.33	14.25	18.01	8.35
10.49	16.40	15.50	14.08	16.17	19.41	15.47	17.20

a. ចូលគណនារយៈពេលមធ្យម និងមេដ្យាន

b. ចូលគណនាវ៉ារីយ៉ង់ គម្លាតស្តង់ដារ វ៉ង់ មេគុណបម្រែបម្រួល និងពិន្ទុ Z។

c. តើជាតុណាខ្លះដែលមិនធម្មតា? ចូលពន្យល់

d. តើទិន្នន័យនេះធៀងឬទេ? បើមាន តើធៀងយ៉ាងដូចម្តេច?

12. ខាងក្រោមនេះគឺជាចំណាយទៅលើអាហារពេលព្រឹករបស់និស្សិត 14នាក់ (គិតជាដុល្លា):

1.70	1.72	1.92	1.14	2.42	2.94	2.80
7.60	5.51	4.57	4.55	2.86	3.41	3.95

a. ចូលគណនាចតុភាគទី1 (Q1) និងចតុភាគទី 3 (Q3) និងចន្លោះចតុភាគ (the interquartile range)

b. ចូលបង្កើតគំនូសតាងប្រអប់ រួចបរិយាយទម្រង់របស់វា។

13. ធនាគារមួយមានទីតាំងនៅភ្នំពេញមានការងារមមាញឹកខ្លាំងតែងតែធ្វើឲ្យអតិថិជនមានការរងចាំយូរនៅពេលទទួលសេវាម្តងៗ ។ ខាងក្រោមនេះគឺជារយៈពេលតម្រង់ជួរ(គិតជានាទី) ដែលបានប្រមូលពីអតិថិជន ចំនួន 16 នាក់។

8.06	5.10	8.02	5.40	8.33	4.25	8.01	8.35
10.49	6.40	5.50	4.08	6.17	9.41	5.47	7.20

ធនាគារមួយទៀតដែលជាសាខារបស់ធនាគារខាងលើ មានទីតាំងស្ថិតនៅជាយក្រុង ហើយក៏មានការមមាញឹកដែរ។ ខាងក្រោមនេះគឺជារយៈពេលតម្រង់ជួរ(គិតជានាទី) ដែលបានប្រមូលពីអតិថិជន ចំនួន 16 នាក់:

7.06	5.10	7.05	5.40	8.30	4.25	7.00	8.30
8.49	6.40	5.50	5.00	6.17	9.41	5.47	7.20

a. ចូលបង្កើតគំនូសតាងប្រអប់និងបរិយាយអំពីទម្រង់នៃបំណែងចែករបស់ធនាគារទាំងពីរ

b. 75% នៃរយៈពេលរង់ចាំ តើធនាគារមានរយៈពេលយូរជាង?

c. ចូលប្រើគំនូសតាងប៉ូលីហ្គូនជាភាគរយ ដើម្បីធ្វើការប្រៀបធៀបរយៈពេលរង់ចាំរបស់ធនាគារទាំងពីរ

14. សន្ទស្សន៍ផ្សារហ៊ុនរបស់ S&P500 នៅឆ្នាំ 2020 មានអត្រាប្រាក់ចំណេញ μ=9% គម្លាតស្តង់ដារ σ=2%។

a. យោងតាមវិធានអក្សរសាស្ត្រ តើអត្រាំរំពឹកទុកដែលស្ថិតនៅចន្លោះ ± 1ស្តង់ដារ និង ± 2ស្តង់ដារគិតជាមធ្យមគឺប៉ុន្មានភាគរយ?

b. យោងតាមវិធាន Chebyshev តើអត្រាំរំពឹកទុកដែលស្ថិតនៅចន្លោះ ± 1, ± 2 និង ±3 ស្តង់ដារគិតជាមធ្យមគឺប៉ុន្មានភាគរយ?

15. ខាងក្រោមនេះគឺជាទិន្នន័យដែលបានកត់ត្រាទៅលើអថេរពីរគឺ X និង Y៖

X	8	6	9	4	9	14	15	20
Y	15	10	13	8	19	11	14	18

a. ចូលគណនាកូវ៉ារីយ៉ង់

b. ចូលគណនាមេគុណទំនាក់ទំនងរបស់អថេរទាំងពីរ

c. តើទំនាក់ទំនងរបស់អថេរទាំងពីរខ្លាំងកំរិតណា? ចូលពន្យល់

ជំពូកទី 4 បំណែងចែកប្រូបាប៊ីលីតេនៃអថេរដាច់

(Discrete Probability Distribution)

គោលបំណង

- សិក្សាអំពីវិបាកនៃបំណែងចែកប្រូបាប៊ីលីតេនៃអរថេរដាច់
- គណនាតម្លៃរំពឹងទុក និង វ៉ារីយ៉ង់បំណែងចែកប្រូបាប៊ីលីតេ
- គណនាកូវ៉ារីយ៉ង់និងការប្រើប្រាស់កូវ៉ារីយ៉ង់នៅក្នុងហិរញ្ញវត្ថុ
- បំណែងចែកទ្វេធា បំណែងចែកព័រស្សុង បំណែងចែកអ៊ីពែកហ្សែន
- បំណែងចែកធម្មតា
- ការប្រើប្រាស់បំណែងចែកទាំងនោះក្នុងបញ្ហាជំនួញ

4.1 បំណែងចែកប្រូបាប៊ីលីតេនៃអរថេរដាច់

នៅក្នុងមេរៀនមុនយើងបានដឹងហើយថា អថេរជាលេខ (numerical variable) ត្រូវបានបែងចែកជាពីរគឺអថេរលេខគត់ ឬដាច់ (discrete numerical variable) និងអថេរលេខបន្ត ឬមានក្បៀស (continuous numerical variable)។ អថេរលេខបន្ត បានមកពីការវាស់ ឧទាហរណ៍ កំពស់ ចំងាយ ទម្ងន់ ជាដើម។ ដោយឡែក ថេរលេខគត់ ឬដាច់ បានមកពីការរាប់ ឧទាហរណ៍ ចំនួនកូន ចំនួនបន្ទប់ ចំនួនផ្លែដូង។ល។ ឥឡូវនេះយើងសិក្សាអំពី បំណែងចែកប្រូបាប៊ីលីតេនៃអរថេរដាច់ (probability distribution for a discrete random variable) វាគឺជា បំណែចែកជាលេខដែលមានជំរើសតែមួយគត់ (numerical mutually exclusive) នៃលទ្ធផលទាំងអស់ដែលកើតមាន។

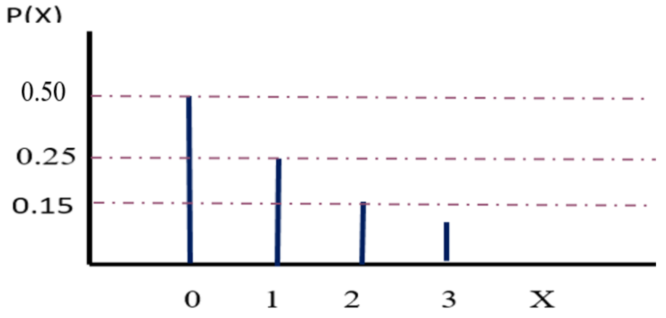
តារាង 4.1 ហ្វ្រេកង់នៃចរន្តអគ្គិសនីដាច់ក្នុងមួយខែ

ចំនួនដងដែលចរន្តអគ្គិសនីដាច់/មួយខែ	ករណីដែលដាច់	ប្រូបាប៊ីលីតេ
1	50	0.50
2	25	0.25
3	15	0.15
4	10	0.10
សរុប	100	1.00

ឧទាហរណ៍ 4.1 អ្នកសិក្សាម្នាក់ធ្វើការកត់ត្រានូវការដាច់ចរន្តអគ្គិសនីនៅរាជធានីភ្នំពេញ។ តារាង 4.1 បង្ហាញអំពីហ្វ្រេកង់ដែលបានកត់ត្រាប្រចាំខែនៃការដាច់ចរន្តអគ្គិសនីក្នុង100ខែ ឬករណី ចុងក្រោយ។

ជួរឈរទី1ក្នុងតារាង 4.1 បានបង្ហាញចំនួនដងដែលអគ្គិសនីបានដាច់ចរន្តក្នុងខែនីមួយៗ។ មួយខែដាច់ចរន្តអគ្គិសនី 1ដង មាន50ករណី ឬ50%(50/100) មួយខែដាច់ចរន្តអគ្គិសនី 2ដង មាន25ករណី ឬ

25%(25/100) មួយខែដាច់ចរន្តអគ្គិសនី 3ដង មាន15ករណីឬ15%(15/100) មួយខែដាច់ចរន្តអគ្គិសនី 4 ដង មាន10ករណី ឬ10% (10/100)។ បំណែងប្រូបាប៊ីលីតេទាំងនេះមានរាងដូច រូប 4.1។



រូប 4.1 បំណែចែកប្រូបាប៊ីលីតេនៃការដាច់ចរន្តអគ្គិសនីក្នុងមួយខែ

4.1.1 តម្លៃរំពឹងទុកនៃបំណែងចែកប្រូបាប៊ីលីតេរបស់អថេរដាច់

តម្លៃមធ្យម (μ)នៃបំណែងចែកប្រូបាប៊ីលីតេរបស់អថេរដាច់ គឺជាតម្លៃរំពឹងទុកនៃអថេរចៃដន្យ។ ដើម្បីគណនាតម្លៃរំពឹងទុក (E(X)) អ្នកត្រូវគុណលទ្ធផល (X) ជាមួយនិងប្រូបាប៊ីលីតេដែលត្រូវគ្នា P(X=x_i) រួចបូលផលគុណទាំងនោះចូលគ្នា ដូចរូបមន្តខាងក្រោម៖

តម្លៃមធ្យម (μ) នៃបំណែងចែកប្រូបាប៊ីលីតេនៃអថេរដាច់

$$\mu = E(X)=\sum_{i=1}^{N} x_i P(X = x_i) \qquad (4.1)$$

ហើយ x_i = លទ្ធផលទី (i) នៃអថេរចៃដន្យដាច់ X

P(X=x_i)=ប្រូបាប៊ីលីតេដែលកើតមានលទ្ធផល X

ឧទាហរណ៍ ការដាច់ចរន្តអគ្គិសនី (តារាង 4.1) តម្លៃរំពឹងទុក ឬចំនួនដងនៃការដាច់អគ្គិសនីគិតជាមធ្យមក្នុងមួយខែ៖

$$\mu = E(X)= (1)*(0.50) + (2)*(0.25) + (3*0.15) + (4)*(0.10) = 1.85$$

ដូច្នេះចំនួនដងនៃការដាច់ចរន្តអគ្គិសនីគិតជាមធ្យម ឬតម្លៃរំពឹងគឺ 1.85 ដងក្នុងមួយខែ។

4.1.2 វ៉ារីយ៉ង់និងគម្លាតស្តង់ដារនៃអថេរដាច់

អ្នកអាចគណនាវ៉ារីយ៉ង់នៃបំណែងចែកអថេរដាច់ ដោយគុណតម្លៃខុសគ្នាគិតជាការ៉េរវាងលទ្ធផលដែលអាចកើតមាននិងតម្លៃរំពឹងទុក ឬតម្លៃមធ្យម ជាមួយប្រូបាប៊ីលីតេដែលត្រូវគ្នា រួចបូកលទ្ធផលទាំងនោះបញ្ចូលគ្នា។

$$\sigma_i^2 = \sum_{i=1}^{N} p_i[x_i - E(X)]^2 \qquad (4.2)$$

ហើយ x_i = គឺជាលទ្ធផលនៃអថេរដាច់ចៃដន្យ X

p_i =P(X=x_i)=ប្រូបាប៊ីលីតេដែលកើតមានលទ្ធផល X

ក្នុងឧទាហរណ៍របស់យើង វ៉ារីយ៉ង់នៃការដាច់ចរន្តអគ្គិសនី៖

σ^2=0.50*$(1-1.85)^2$ +0.25*$(2-1.85)^2$ +0.15*$(3-1.85)^2$ +0.10*$(4-1.85)^2$ = 1.0275

និងគម្លាតស្តង់ដារ $\sigma = \sqrt{1.02755} = 1.0136$

នេះមានន័យថា ចំនួនដងនៃការដាច់ចរន្តអគ្គិសនីគិតជាមធ្យមគឺ 1.85 ដងក្នុងមួយខែ និងមានគម្លាតស្តង់ដារគឺ1.0136 ដង។

តារាង 4.2 គណនាវ៉ារីយ៉ង់និងគម្លាតស្តង់ដារនៃការដាច់ចរន្តអគ្គិសនី

Prob.(p_i)	x_i	E(X)	$[x_i - E(X)]^2$	$p_i[x_i - E(X)]^2$
0.50	1	1.85	$(1-1.85)^2$	$0.50*(1-1.85)^2$
0.25	2	1.85	$(2-1.85)^2$	$0.25*(2-1.85)^2$
0.15	3	1.85	$(3-1.85)^2$	$0.15*(3-1.85)^2$
0.10	4	1.85	$(4-1.85)^2$	$0.10*(4-1.85)^2$
				$\sigma^2 = 1.0275$

4.2 កូវ៉ារីយ៉ង និងការអនុវត្តន៍ក្នុងផ្នែកហិរញ្ញវត្ថុ

នៅផ្នែក4.1 យើងបានសិក្សាអំពី វ៉ារីយ៉ង់ និងគម្លាតស្តង់ដារ ព្រមទាំងតម្លៃរំពឹងទុក។ ដោយឡែកនៅផ្នែកនេះ យើងនឹងសិក្សាអំពី កូវ៉ារីយ៉ង់របស់អថេរពីរ ដែលត្រូវបានអនុវត្តជាញឹកញាប់នៅក្នុងហិរញ្ញវត្ថុ។

4.2.1 កូវ៉ារីយ៉ង់

កូវ៉ារីយ៉ង់ σ_{XY}(covariance) គឺកំលាំងទំនាក់ទំនងនៃអថេរជាលេខចំនួនពីរ X និង Y។ ជាទូទៅកូវ៉ារីយ៉ង់ពិបាកក្នុងការបកស្រាយណាស់ ដូច្នេះដើម្បីបកស្រាយ ភាគច្រើនគេបង្វែវាទៅជាកូរីលេសិន (correlation = ρ_{XY}) វិញ។ កូរីលេសិន ត្រូវបានគណនាដោយ យកកូវ៉ារីយ៉ង់ចែកនឹងផលគុណគម្លាតស្តង់ដារ៖ $\rho_{XY} = \sigma_{XY}/(\sigma_X * \sigma_X)$ ។ តម្លៃ កូរីលេសិន គឺនៅចន្លោះពី -1 ទៅ +1 ។ ពេលកូរីលេសិន វិជ្ជមាន មានន័យថាអថេរទាំងពីរមានទំនាក់ទំនងស្របគ្នា ហើយបើអវិជ្ជមានវិញ មានន័យថាថាអថេរទាំងពីរមានទំនាក់ទំនងច្រាស់ ហើយបើស្មើសូន្យ មានន័យថា អថេរទាំងពីរគឺមិនមានទំនាក់ទំនងជាមួយគ្នា។ សមីការ (4.3) បានកំណត់កូវ៉ារីយ៉ង់នៃបំណែងប្រូបាប៊ីលីតេរបស់អថេរដាច់។

$$\sigma_{XY} = \sum_{i=1}^{N} p_{XiYi}[x_i - E(X)] * [y_i - E(Y)] \quad (4.3)$$

ហើយ X= អថេរចៃដន្យដាច់ X

x_i = លទ្ធផលទី (i) នៃអថេរចៃដន្យដាច់ X

Y=អថេរចៃដន្យដាច់ Y

y_i = លទ្ធផលទី (i) នៃអថេរចៃដន្យដាច់ Y

p_{XiYi} = ប្រូបាប៊ីលីតេទី(i)ដែលកើតមានលទ្ធផល X និងទី(i)ដែលកើតមានលទ្ធផល Y

ឧទាហរណ៍ 4.2 ដូចបានដឹងមកហើយ សេដ្ឋកិច្ចបច្ចុប្បន្នគឺអាស្រ័យទៅនិងស្ថានភាព COVID-19 បើស្ថានភាព COVID-19 នៅតែរាលដាលខ្លាំង សេដ្ឋកិច្ចរំពឹងថានឹងធ្លាក់ខ្លាំងដែរ បើការរាលដាលតិច សេដ្ឋកិច្ចក៏ធ្លាក់តិចដែរ ហើយបើស្ថានភាព COVID-19 ឈប់រាលដាល សេដ្ឋកិច្ចក៏រំពឹងប្រសើរឡើងវិញដែរ។ ឥឡូវ

នេះយើងសន្មត់ថា តម្លៃភាគហ៊ុនរបស់រដ្ឋករ (PPWSA) និង អេស៊ីលីដា (ACLEDA) មានទំនាក់ទំនងជាមួយស្ថានភាព COVID-19 នៅឆ្នាំក្រោយ ហើយយើងសន្មត់បន្តទៀតថា ប្រូបាបីលីតេនៃ COVID-19ឆ្នាំក្រោយមានបីសណ្ឋាន ដូចក្នុងតារាង 4.3 ខាងក្រោម។

តារាង 4.3 អាត្រាប្រាក់ចំនេញនៃភាគហ៊ុន ACLEDA និង PPWSA ក្នុងស្ថានភាព COVID-19

p_{XiYi}	ស្ថានភាព COVID-19	អាត្រាចំនេញរបស់ ACLEDA	អាត្រាចំនេញរបស់PPWSA
0.5	ប្រសើរ	0.12	0.10
0.3	មានស្ថេរភាព	0.06	0.06
0.2	ធ្ងន់ធ្ងរ	-0.04	-0.03

❖ អនុវត្តរូបមន្ត (4.1) តម្លៃរំពឹងទុកនិងគម្លាតស្តង់ដាររបស់ភាគហ៊ុននីមួយៗដូចខាងក្រោម៖

តាង X=អាត្រាប្រាក់ចំណេញរបស់ ACLEDA និង Y= អាត្រាប្រាក់ចំណេញរបស់ PPWSA

❖ អនុវត្តរូបមន្ត (4.1) តម្លៃរំពឹងទុករបស់ភាគហ៊ុនទាំងពីរ៖

- E(X) = 0.5 (0.12) + 0.3(0.06) + 0.2(-0.04) = 0.07
- E(Y) = 0.5 (0.08) + 0.3(0.06) + 0.2(-0.03) = 0.062

❖ អនុវត្តរូបមន្ត (4.2) វ៉ារីយ៉ង់របស់ភាគហ៊ុនទាំងពីរគឺ៖

- VAR(X)=σ_X^2=$0.5(0.12-0.07)^2$+$0.3(0.08-0.07)^2+0.2(-0.04-0.07)^2$=0.0037
- គម្លាតស្តង់ដារ: SD(X)= σ_X=$\sqrt{0.0037}=0.06082$
- វ៉ារីយ៉ង់:VAR(Y)=σ_Y^2=$0.5(0.08-0.062)^2$+$0.3(0.05-0.062)^2+0.2(-0.01-0.062)^2$=0.002416
- គម្លាតស្តង់ដារ: SD(Y)= σ_Y=$\sqrt{0.002416}=0.04915$

❖ អនុវត្តរូបមន្ត (4.3)កូវ៉ារីយ៉ង់របស់ភាគហ៊ុនទាំងពីរ៖

- $\sigma_{XY}=0.5(0.12-0.07)*(0.10-0.062)+0.3(0.06-0.07)*(0.06-0.062)$
  $0.2(-0.04-0.07)*(-0.03-0.062)$
  =0.00298
- គណនាកូរីលេសិន

$\rho_{XY}=\sigma_{XY}/(\sigma_X*\sigma_X)$= 0.00298/(0.06082*0.04915)= 0.996

បំណកស្រាយ៖ ភាគហ៊ុន ACLEDA រំពឹងថាឆ្នាំក្រោយនឹងទទួលបានប្រាក់ចំណេញជាមធ្យម 7% និងគម្លាតស្តង់ដារ 0.06082 ហើយ PPWSA រំពឹងថាឆ្នាំក្រោយនិងទទួលបានប្រាក់ចំណេញជាមធ្យម 6.2% និងគម្លាតស្តង់ដារ 0.04915។ ភាគហ៊ុន ACLEDA គឺមានហានិភ័យខ្ពស់ជាងភាគហ៊ុន PPWSA ពីព្រោះថា ភាគហ៊ុន ACLEDA មានគម្លាតស្តង់ដារធំជាងភាគហ៊ុន PPWSA។ លើសពីនេះទៅទៀតភាគហ៊ុន ACLEDA អន់ជាភាគហ៊ុន PPWSA ពីព្រោះក្នុងមួយឯកតានៃហានិភ័យ ភាគហ៊ុន ACLEDA ទទួលបានប្រាក់ចំណេញតែ 1.15%តែប៉ុណ្ណោះ(0.07/0.06082) រីឯភាគហ៊ុន PPWSA ចំនេញដល់ 1.26%ឯណោះ (0.062/0.04915)។ ភាគហ៊ុនទាំងពីរមានទំនាក់ទំនងវិជ្ជមាន 99.6% ប្រហាក់ប្រហែល 1 មានន័យថា បើ

ចំណេញគឺចំណេញទាំងពីរ បើខាត គឺខាតទាំងពីរ។ តាមទ្រឹស្តីហិរញ្ញវត្ថុ បើសិនជាកូវ៉ារីយ៉ង់នៃអាត្រាប្រាក់ចំណេញរបស់ភាគហ៊ុនទាំងពីរស្មើ +1 ការទិញភាគហ៊ុនទាំងពីរបញ្ចូលគ្នា មិនជួយឲ្យកាត់បន្ថយហានិភ័យអ្វីទាំងអស់ បើកូវ៉ារីយ៉ង់រំកិលទៅ -1 ការទិញភាគហ៊ុនទាំងពីរបញ្ចូលគ្នា អាចជួយឲ្យកាត់បន្ថយហានិភ័យរហូតដល់សូន្យ។ បើកូវ៉ារីយ៉ង់កាន់តែតូចគឺកាន់តែប្រសើរ។

4.2.2 អាត្រារំពឹងទុកនិងហានិភ័យនៃសំពៀតវិនិយោគ

យើងបានសិក្សាអំពី កូវ៉ារីយ៉ង់របស់អថេរពីរ គម្លាតស្តង់ដារ និងតម្លៃរំពឹងរបស់អថេរពីរ ។

ឥឡូវនេះយើងសិក្សាតម្លៃរំពឹងនៃសំពៀតវិនិយោគវិញម្តង (portfolio)។ តម្លៃរំពឹងនៃសំពៀតតចំណោះការវិនិយោគទៅលើទ្រព្យសកម្មពីរ គឺស្មើនឹងសមាមាត្រដែលបានកំណត់ទៅលើ X គុណនិងតម្លៃរំពឹងនៃទ្រព្យសកម្ម X បូកជាមួយ សមាមាត្រដែលបានកំណត់ទៅលើ Y គុណនិងតម្លៃរំពឹងនៃទ្រព្យសកម្ម Y។

$$E(P) = wE(X) + (1-w)E(Y) \qquad (4.4)$$

ហើយ E(P) = តម្លៃរំពឹងនៃសម្លៀតវិនិយោគ P

w = សមាមាត្រវិនិយោគលើ X

(1-w) = សមាមាត្រវិនិយោគលើ Y

E(X) = តម្លៃរំពឹងនៃទ្រព្យសកម្ម X

E(Y) = តម្លៃរំពឹងនៃទ្រព្យសកម្ម Y

- វ៉ារីយ៉ង់នៃសំពៀតវិនិយោគ

$$\sigma_p^2 = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_{XY} \qquad (4.5)$$

- គម្លាតស្តង់ដារនៃសំពៀតវិនិយោគ $\sigma_p = \sqrt{\sigma_p^2}$

ឧបមាថា អ្នកមានប្រាក់ 1000 ដុល្លា ហើយអ្នកបានសំរេចចិត្តវិនិយោគលើភាគហ៊ុន ACLEDA ចំនួន 600 ដុល្លា ឬ 0.60 និង សល់ប៉ុន្មានវិនិយោគលើភាគហ៊ុន PPWSA។ ចូលគណនា តម្លៃរំពឹង និង គម្លាតស្តង់ដារនៃសំពៀតវិនិយោគ។

សម្មតិកម្មៈ W=600/1000=0.60, E(ACLEDA)=0.07, និង E(PPWSA)=0.062

- E(P) = 0.60(0.07) + (1-0.60)(0.062) = 0.0668

គណនាវ៉ារីយ៉ងនិងគម្លាតស្តង់ដារ របស់សំពៀតវិនិយោគ

សម្មតិកម្មៈ W=0.60, SD(ACLEDA)=0.06082, SD(PPWSA)=0.04915 និង $\sigma_{XY} = 0.00298$

ជំនួសទិន្នន័យទាំងនេះចូលក្នុងរូបមន្ត (4.5)

- $\sigma_p^2 = 0.60^2(0.06082)^2 + (1-0.60)^2(0.04915)^2 + 2(0.60)(1-0.60)(0.00298)$

$=0.003148$

- គម្លាតស្តង់ $\sigma_p = \sqrt{0.003148}=0.05611$

ដូច្នេះការសំរេចចិត្តវិនិយោគ 60% ទៅលើភាគហ៊ុន ACLEDA និង 40% ទៅលើភាគហ៊ុន PPWSA និងទទួលបានអាត្រាំពឹងទុកស្មើនឹង 6.68% ឬ 66.8 ដុល្លា និងមានគម្លាតស្តង់ដារ 5.611% ឬ 56.11ដុល្លា។

សូមបញ្ជាក់ថា សមីការ (4.5) គឺសម្រាប់វាស់វែងហានិភ័យ បើកូវ៉ារីយ៉ង់របស់ភាគហ៊ុនទាំងពីរកាន់តែតូច σ_{XY} ឬអវិជ្ជមាន តម្លៃនៃសមីការ (4.5)កាន់តែតូចដែរ មានន័យថាការវិនិយោគទៅលើទ្រព្យសកម្មពីរដែលមានកូវ៉ារីយ៉ង់អវិជ្ជជួយអោយយើងកាត់បន្ថយហានិភ័យកាន់តែតិច។

4.3 បំណែងចែកទ្វេធា

បំនែងចែកទ្វេធា (binomial distribution) គឺជាបំណែងចែកប្រូបាប៊ីលីតេនៃភាពជោគជ័យ ឬបរាជ័យ របស់លទ្ធផលនៃការធ្វើពិសោធន៍ ឬការស្ទង់មតិ ដែលត្រូវបានធ្វើឡើងជាច្រើនលើក។ ទ្វេធា គឺជាបំណែងដែលមានលទ្ធផលពីរ[7] ឧទាហរណ៍ ប្រលងជាប់ ឬធ្លាក់ អតិថេជនទិញ ឬមិនទិញ ជួលឬមិនជួល ឥណទានត្រូវបានអនុម័ត ឬ មិនអនុម័ត។ ល។

4.3.1 ប្រូបាប៊ីលីតេនៃបំណែងចែកទ្វេធានិងវិបាក

បំណែងចែកទ្វេធាមានវិបាកបួន៖

1. ចំនួនប៉ាន់គំរូគឺមានចំនួនថេរ (n)
2. ចំនួនប៉ាន់គំរូនីមួយៗមានលទ្ធផលពីរ តែជ្រើសរើសបានតែមួយគត់
3. ប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ដែលចាប់អារម្មណ៍ (π) គឺមានលក្ខណៈថេរពីការអង្កេតមួយទៅការអង្កេតមួយទៀត។ ដូច្នេះប្រូបាប៊ីលីតេនៃព្រឹត្តិការណ៍ដែលមិនចាប់អារម្មណ៍គឺ (1-π) ក៏ថេរដែរ
4. លទ្ធផលនៃការអង្កេតគឺឯករាជ្យពីមួយទៅមួយ

រូបមន្តបំណែងចែកទ្វេធាដែលទាក់ទងប្រូបាប៊ីលីតេជោគជ័យ X ដែលធ្វើពីសោធន៍ n ដង៖

$$P(X) = \frac{n!}{(n-X!)X!} * p^X * q^{n-X} \quad (4.6)$$

ហើយ P(X)=ប្រូបាប៊ីលីតេជោគជ័យ

p = តួរលេខនៃប្រូបាប៊ីលីតេជោគជ័យ

q = តួរលេខនៃប្រូបាប៊ីលីតេបរាជ័យ

X =ចំនួនជោគជ័យក្នុងចំណោមពិសោធន៍ n ដង

n = ចំនួនដងនៃការធ្វើពិសោធន៍

ឧទាហរណ៍4.3 អ្នកបោះកាក់បីក្នុងពេលតែមួយ (n=3) ចូលគណនាប្រូបាប៊ីលីតេដែលបានរូបពីរដង (X=2)។

[7] ទាញយកនៅថ្ងៃទី 10/5/2021 ចេញពី https://www.statisticshowto.com/probability-and-statistics/binomial-theorem/binomial-distribution-formula/

តាង H= មានរូប និង T=គ្មានរូប ហើយលទ្ធផលជា

តារាង 4.3 លទ្ធផលនៃប្រ៊ូបាប៊ីលីតេនៃការបោះកាក់3ក្នុងពេលតែមួយ

	លទ្ធផល		ប្រូបាប៊ីលីតេ
1	មានរូបទាំង3	HHH	1/8
2	កាក់ទី1និងទី2មានរូប ប៉ុន្តែទី3គ្មានរូប	**HHT**	1/8
3	កាក់ទី1មានរូប ទី2គ្មានរូប និងទី3មានរូប	**HTH**	1/8
4	កាក់ទី1គ្មានរូប និងទី2និងទី3មានរូប	**THH**	1/8
5	កាក់ទី1គ្មានរូប ទី2និងទី3មានរូប	TTH	1/8
6	កាក់ទី1គ្មានរូប ទី២និងទី3 មានរូប	THT	1/8
7	កាក់ទី1មានរូប ទី2និងទី3គ្មានរូប	HTT	1/8
8	គ្មានរូបទាំងបី	TTT	1/8
		ផលបូក	1.00

តាមលទ្ធផលក្នុងតារាង 4.4បង្ហាញថា លទ្ធផលដែលបោះកាក់3ក្នុងពេលតែមួយដែលផ្តល់ចំលើយមានរូប2គឺមានចំនួន3លើក៖ HHT, HTH និង THH

ដូច្នេះប្រ៊ូបាប៊ីលីតេនៃលទ្ធផលនេះគឺ= $\frac{3}{8}$=0.375

ឥឡូវយើងចង់ធ្វើការផ្ទៀងផ្ទាត់លទ្ធផលនេះ ដោយជំនួសធាតុចូលទាំងអស់ចូលក្នុងរូបមន្ត (4.6) យើងនឹងទទួលបានចំលើយដូចមុន។

$n = 3$, X= 2 (ចំនួនរូប), p=0.5 (ប្រ៊ូបាប៊ីលីតេ), q=0.5 (ប្រ៊ូបាប៊ីលីតេគ្មានរូប)

P(2 heads) =$\frac{n!}{(n-X!)X!} * p^X * q^{n-X}$=$\frac{3!}{(3-2!)2!} * 0.5^2 * 0.5^{(3-2)}$=0.375

ឧទាហរណ៍ 4.4 ការធ្វើជំរឿនមួយបង្ហាញថា មនុស្សម្នាក់ក្នុងចំណោម 5នាក់បានខ្ចីលុយធនាគារ។ បើសិនជាមនុស្ស10នាក់ត្រូវបានគេជ្រើសរើសដោយចៃដន្យ ចូលគណនាប្រ៊ូបាប៊ីលីតេរបស់មនុស្ស 3នាក់ដែលប្រាកដជាខ្ចីលុយធនាគារ។

ចំលើយ:

X=3, n=10, p = 1/5=0.20, q= (1-p)= (1-0.20) = 0.80 យើងបាន៖

P(3) = $\frac{10!}{(10-3!)3!} * 0.20^3 * 0.80^{10-3} = 0.201$

ដូច្នេះ ប្រ៊ូបាប៊ីលីតេរបស់មនុស្ស 3នាក់ក្នុងចំណោម 10នាក់ដែលប្រាកដជាខ្ចីលុយធនាគារ គឺ 20.1%។

4.3.2 តម្លៃមធ្យម វ៉ារីយ៉ង់ និងគម្លាតស្តង់ដារនៃបំណែងចែកទ្វេធា

តម្លៃមធ្យម : $\mu = n * p$ (4.7)

វ៉ារីយ៉ង់: $\sigma^2 = n * p * q$ (4.8)

គម្លាតស្តង់ដារ: $\sigma = \sqrt{n * p * q}$ (4.9)

ឥឡូវយើងត្រឡប់ទៅមើលឧទាហរណ៍ 4.4 ឡើងវិញ

- អ្នកខ្ចីលុយធនាគាគិតជាមធ្យម:$\mu = 10 * 0.20 = 2$ នាក់
- Variance: $\sigma^2 = 10 * 0.20 * 0.80 = 1.6$ នាក់
- គម្លាតស្តង់ដារ $\sigma = \sqrt{1.6}$=1.26នាក់

4.4 **បំណែងចែកព័រសុង**

បំណែងចែកព័រសុង (the poison distribution) គឺជាបំណែងចែកប្រូបាប៊ីលីតេដាច់ ហើយ n ធំនិង p តូច ហើយអថេរឯកកើតឡើងគ្រប់ពេល។ បំណែងចែកព័រសុងត្រូវបានប្រើប្រាស់ពេលដែលដង់ស៊ីតេនៃវត្ថុត្រូវបានរាយប៉ាយទៅលើតំបន់មួយ ឬ មួយចំនួន ដូចជាចំនួនកូនឈើដែលបានដាំក្នុងមួយហិកតា ឬ ចំនួនផលិតផលដែលខូចក្នុងរយៈពេលមួយនៃការផលិត។

ប្រូបាប៊ីលីតេនៃ X ដែលកើតមាននៅចន្លោះពេលវេលាមួយ បរិមាណ ឬ ផ្ទៃក្រឡា ចំពោះអថេរមួយ ហើយ λ គឺជាចំនួនមធ្យមដែលកើតឡើងក្នុងមួយឯកតា វេលាមួយ បរិមាណ ឬ ផ្ទៃក្រឡា ។ល។

$$P(X;\lambda) = \frac{e^{-\lambda}\lambda^X}{X!} \qquad (4.10)$$

ហើយ X= 0, 1, 2,...

e =ចំនួនថេរដែលមានតម្លៃប្រហែល 2.7183

ឧទាហរណ៍ 4.5 ការបោះពុម្ពសៀវភៅមួយមាន 500 ទំព័រ មានកំហុសចំនួន 200កន្លែង ចូលគណនាប្រូបាប៊ីលីតេនៃទំព័រមួយដែលមាន កំហុសចំនួន 3កន្លែង។

ចំលើយ៖

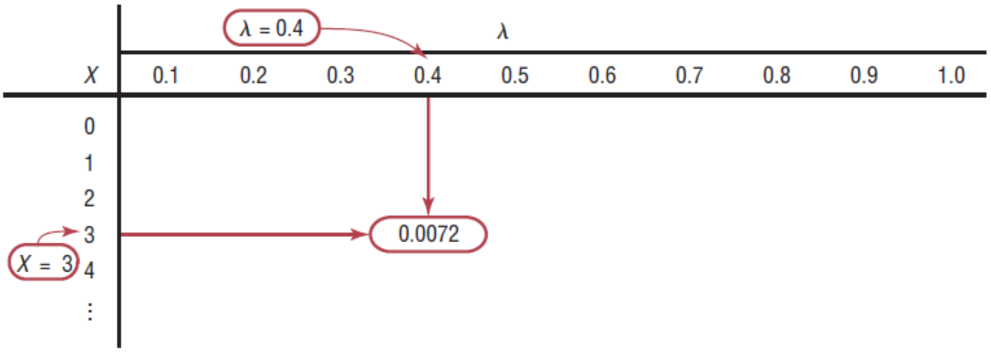
X= 3, កំហុសគិតជាមធ្យមក្នុងមួយទំព័រ $\lambda = \frac{200}{500} = 0.4$

$$P(X;\lambda) = \frac{2.7183^{-0.4}0.4^3}{3!} = 0.0072$$

ដូច្នេះ មួយទំព័រដែលមានកំហុសចំនួន 3កន្លែង មានប្រូបាប៊ីលីតេស្មើនឹង 0.72%

ជំរើសមួយទៀតអ្នកអាចរកចំឡើយខាងលើតាមរយៈការមើលតារាងព័រសុងដូចមាននៅក្នុងរូប4.1ខាងក្រម

រូប4.1 ផ្នែកមួយនៃតារាងព័រសុង



ប្រភព: Allan G. B (2009)

4.5 បំណែងចែកអ៊ីពែកហ្សែនមីទ្រីស

បំណែងចែកអ៊ីពែកហ្សែនមីទ្រីស (The Hypergeometric Distribution) គឺជាបំណែងចែកនៃអថេរមួយដែលមានលទ្ធផលពីរ វាត្រូវបានប្រើប្រាស់ពេលដែលការប្រមូលប៉ាន់គំរូមិនបានជំនួសមកវិញ មានន័យថាការប្រមូលប៉ាន់គំរូមិនឯករាជ្យ ឬចំនួនប៉ាន់គំរូតូច។ ក្នុងករណីចំនួនប៉ាន់គំរូតូច បំណែងចែកទ្វេជា ផ្តល់ប្រូបាប៊ីលីតេមិនត្រឹមត្រូវឡើយ។ ដូច្នេះគេអាចប្រើបំណែងចែកអ៊ីពែកហ្សែនមីទ្រីសមកជំនួសវិញ។

មានចំនួនសាកលមួយដែលមានតែវត្ថុពីរ ដូចជាប្រុសឬស្រី វត្ថុខូចឬប្រើបាន ជ័យជំនះ ឬបរាជ័យ ។ល។ ទាំងនេះហាល់ដូចជាមានវត្ថុ a និងវត្ថុមួយទៀត b ហើយ a+b គឺស្មើចំនួនសរុបនៃចំនួនសាកល ។ P(X) ប្រូបាប៊ីលីតេ នៃការជ្រើសរើសប៉ាន់គំរូ n (ដោយមិនមានការជំនួសវិញ) ដែលមានចំនួនវត្ថុ X ជារបស់ឬប្រភេទ a និង n-X មានចំនួនវត្ថុ ជារបស់ឬប្រភេទ b ប្រូបាប៊ីលីតេនៃX៖

$$P(X)=({}_aC_X * {}_bC_{n-X})/{}_{a+b}C_n \qquad (4.11)$$

ហើយ

${}_aC_X$ = វិធីសាស្ត្រជ្រើសរើសវត្ថុទី1

${}_bC_{n-X}$ = វិធីសាស្ត្រជ្រើសរើសវត្ថុទី2

${}_{a+b}C_n$ = វិធីសាស្ត្រជ្រើសរើសវត្ថុ n ចេញពីចំនួនសកល

ឧទាហរណ៍ 4.6 មានមនុស្ស 10 នាក់ដាក់ពាក្យធ្វើការជាអ្នកទីផ្សារ។ ប្រាំនាក់មានសញ្ញាបត្រមហាវិទ្យាល័យ(a=5) ហើយប្រាំនាក់ទៀតគឺគ្មាន (b=5)។ ប្រសិនជា ការជ្រើសរើសប៉ាន់គំរូបីនាក់ (n=3) ត្រូវបានជ្រើរើសដោយចៃដន្យ ចូលគណនាប្រូបាប៊ីលីតេដែលទាំងបីនាក់(X=3) មានសញ្ញាបត្រមហាវិទ្យល័យ?

ចំលើយ៖

a = 5 ប្រភទមានសញ្ញាបត្រ ហើយត្រូវជ្រើសរើសយកបីនាក់ (n=3) ហើយបីនាក់មានសញ្ញាបត្រ (X =3)

b = 5 ប្រភទមិនមានសញ្ញាបត្រ ហើយ (n- X)= 0

a+b =10

សំគាល់៖ យើងមានរូមន្តបន្សំ (combination): ${}_nC_r = \frac{n!}{(n-r)!r!}$

ជំនួសលេខទាំងនេះចូលក្នុងរូបមន្ត (4.11): $P(X)=({}_aC_X * {}_bC_{n-X})/{}_{a+b}C_n$

$$P(X)=({}_5C_3 * {}_5C_{3-3})/{}_{5+5}C_3$$

$$= [(5!/(5-3)!3!)*(5!/(5-0)!0!)]/(10!/(10-3)!3!)$$

$$= (10*1)/120$$

$$=0.0833$$

ដូច្នេះប្រូបាប៊ីលីតេដែលទាំងបីនាក់មានសញ្ញាបត្រមហាវិទ្យល័យគឺ 8.33%

ឥឡូវសន្មត់ថា យើងចង់គណនាប្រូបាប៊ីលីតេដែលទាំងពីរនាក់(X=2) មានសញ្ញាបត្រមហាវិទ្យល័យ?

$$P(X)=({}_5C_2 * {}_5C_{3-2})/{}_{5+5}C_3$$

= 5!/(5-2)!2!*(5!/(5-1)!1!)/(10!/(10-3)!3!)

= (10*5)/120

=0.42

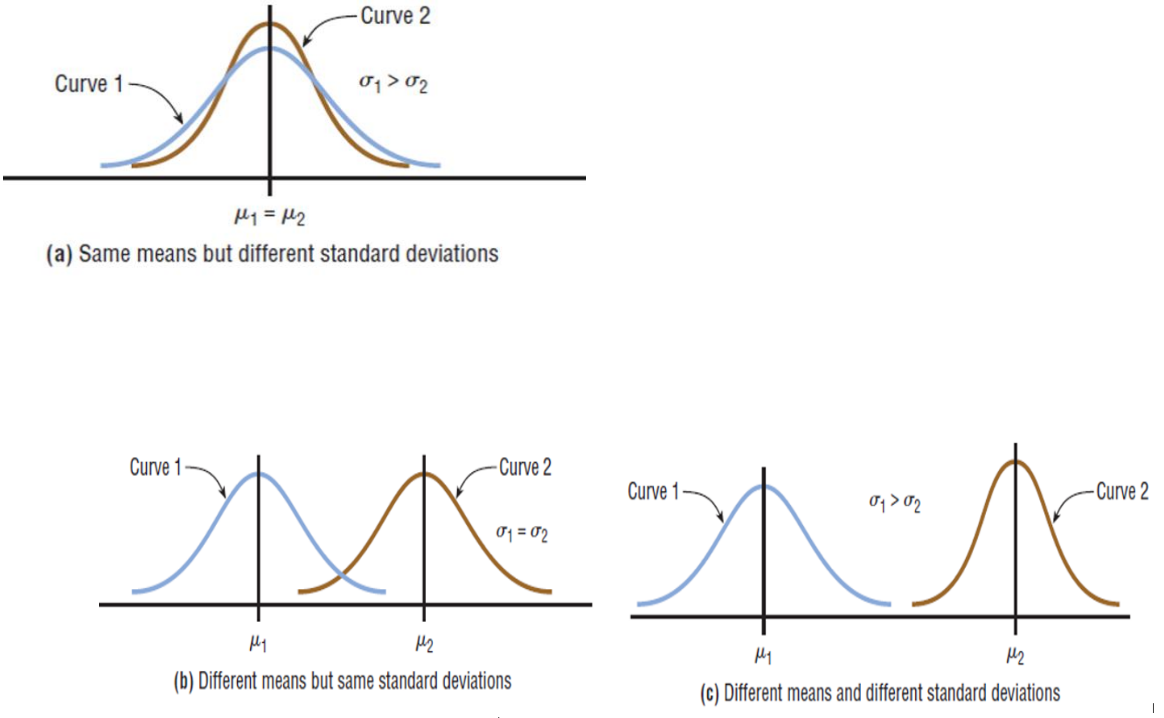
ដូច្នេះប្រូបាប៊ីលីតេដែលទាំងពីរនាក់មានសញ្ញាបត្រមហាវិទ្យាល័យគឺ 42%

4.6 **បំណែងធម្មតា**

បំណែងធម្មតា (normal distribution) ជាទូទៅត្រូវបានគេស្គាល់ថា the Gaussian distribution (1777–1855)គឺជាបំណែងចែកប្រូបាប៊ីលីតេដែលមានលក្ខណៈស៊ីមេទ្រី និងចំនួនប្រេកង់គឺស្ថិតនៅជុំវិញតម្លៃមធ្យម ហើយមានគំនូសតាងដូចរូបជួង (a bell curve)។ បំណែងធម្មតាអាចមានតម្លៃមធ្យមច្រើន (អាចវិជ្ជមាន ឬវិជ្ជមាន) និងគម្លាតស្តង់ដារគឺវិជ្ជមានជានិច្ច។ ទម្រង់និងទីតាំងនៃខ្សែកោងរបស់បំណែងចែកធម្មតាគឺពឹងផ្អែកលើទៅនឹងប៉ារ៉ាម៉ែត្រពីរគឺៈ តម្លៃមធ្យម (mean)និងគម្លាតស្តង់ដារ (standard deviation) ហើយប៉ារ៉ាម៉ែត្រនីមួយៗមានបំណែងចែកធម្មតាផ្ទាល់ខ្លួន អាស្រ័យលើតម្លៃមធ្យម និងគម្លាតស្តង់ដារ។ រូប 4.2a បង្ហាញពីបំណែងចែកធម្មតាពីរដែលមានតម្លៃមធ្យមស្មើគ្នា ប៉ុន្តែគម្លាតស្តង់ដារខុសគ្នា។ រូប 4.2b បង្ហាញពីបំណែងចែកធម្មតាពីរដែលមានគម្លាតស្តង់ដារស្មើគ្នា ប៉ុន្តែតម្លៃមធ្យមខុសគ្នា។ រូប 4.2c បង្ហាញពីបំណែងចែកធម្មតាពីរដែលមានតម្លៃមធ្យមខុសគ្នា និងគម្លាតស្តង់ដារខុសគ្នា។

សូមរំលឹកថា អថេរដាច់ (discrete variable) មិនមានតម្លៃនៅចន្លោះរវាងអថេរពីរនោះទេ ប៉ុន្តែសម្រាប់អថេរជាប់ (continuous variable) នៅចន្លោះអថេរនីមួយៗគឺមានតម្លៃ។

រូប 4.2 ទម្រង់ផ្សេងៗនៃបំណែងចែកធម្មតា



ប្រភព: Allan G. B (2009) ទំព័រ 303

សមីការបំណែងចែកធម្មតា (a normal distribution equation) គឺ៖

$$y = \frac{e^{-(X-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \qquad (4.12)$$

ហើយ

$e \approx 2.718$ ($\approx$ តម្លៃមធ្យមៈ $mean$)
$\pi = 3.14$
$\mu =$ តម្លៃមធ្យមនៃចំនួនសកល
$\sigma =$ គម្លាតស្តង់ដាររបស់ចំនួនសាកល

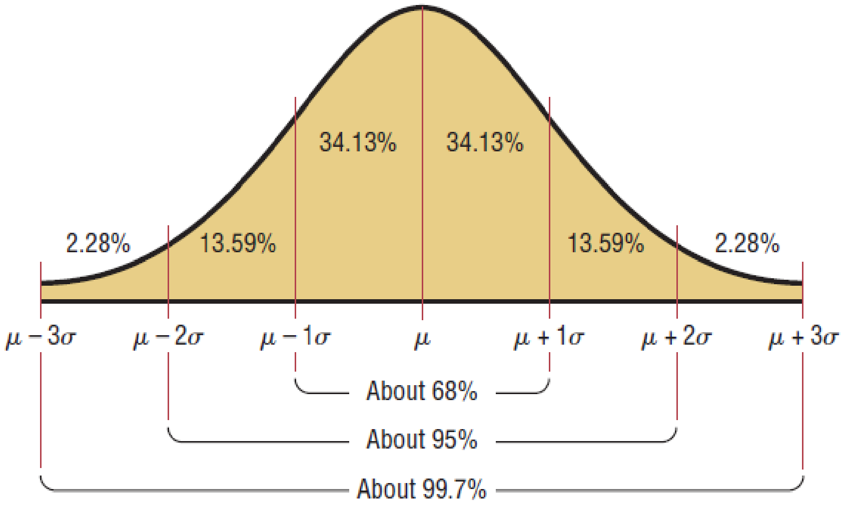
សមីការនេះមើលទៅស្មុគស្មាញណាស់ ប៉ុន្តែក្នុងការអនុវត្តន៍ យើងអាចប្រើតារាង normal ជំនួសសមីការនេះ។

ជាទូទៅក្នុងការអនុវត្តន៍ជាក់ស្តែង កម្រមានអថេរដែលមានបំណែងចែកដូចបំណែងចែកធម្មតាណាស់។ បំណែងចែកធម្មតាគ្រាន់តែជាទ្រឹស្តីតែប៉ុណ្ណោះ តែទោះបីជាយ៉ាងណាក៏ដោយ មានអថេរជាច្រើនមានបំណែងចែកខុសពីធម្មតាតិចទួចតែប៉ុណ្ណោះ។

សេក្តីសង្ខេបអំពីវិបាកនៃទ្រឹស្តីបំណែងចែកធម្មតា

1. បំណែងចែកធម្មតាមានរាងដូចជួង (រូប4.3)
2. តម្លៃមធ្យម ម៉ូដ និងមេដ្យានគឺស្មើគ្នា ស្ថិតនៅចំណុចកណ្តាលនៃរូប
3. បំណែងចែកធម្មតាមានម៉ូដតែមួយ
4. រូបមានលក្ខណៈស៊ីមេទ្រីត្រង់តម្លៃមធ្យម សងខាងនៃតម្លៃមធ្យមគឺស្មើគ្នា
5. ខ្សែកោងគឺមានលក្ខណៈជាប់គ្នា មិនដាច់ពីគ្នា ឬមិនមានប្រហោង
6. ខ្សែកោងគឺមិនដែលប៉ះគ្នាជាមួយអ័ក្ស X ឡើយ
7. ផ្ទៃនៃខ្សែកោងតំណាងឲ្យប្រូបាប៊ីលីតេ គឺស្មើ1 ឬ100%
8. ផ្ទៃដែលជាផ្នែកនៃខ្សែកោងស្ថិតនៅក្នុងរង្វង់1ស្តង់ដារនៃតម្លៃមធ្យមគឺប្រហែល 68% 2ស្តង់ដារគឺប្រហែល 95% និង 3 ស្តង់ដារគឺប្រហែល 99.7% (មើលរូប 4.3)

រូប 4.3 ផ្ទៃនៅក្នុងបំណែងចែកធម្មតា



ប្រភព: Allan G. B (2009) ទំព័រ 304

4.6.1 បំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ

ដូចរៀបរាប់ខាងដើមមកហើយ បំណែងចែកធម្មតាមានតម្លៃមធ្យមផ្ទាល់ខ្លួន (ច្រើន) ហើយគម្លាតស្តង់ដារក៏មានច្រើន ហើយទីតាំងនិងទម្រង់មានការប្រែប្រួល មានសភាពស្មុគស្មាញ។ ដើម្បីដោះស្រាយបញ្ហានេះ អ្នកស្ថិតិបានបង្កើតបំណែងចែកសាមញ្ញមួយដើម្បីប្រើប្រាស់ ដែលគេស្គាល់ថា បំណែងចែកធម្មតាដែលមានស្តង់ដារ (the Standard Normal Distribution)។ បំណែងចែកធម្មតាដែលមានស្តង់ដារ ត្រូវបានគេឲ្យនិយមន័យថា៖ គឺជាបំណែងចែកដែលមានតម្លៃមធ្យមស្មើ 0 និងគម្លាតស្តង់ដារស្មើ 1។

រូប 4.4 បង្ហាញអំពី បំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ។ ផ្ទៃដែលស្ថិតខាងលើនិងខាងក្រោមតម្លៃមធ្យម គឺផ្ទៃ 0.3413 ឬ 34.13%។

រូបមន្តបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ (the standard normal distribution equation)

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}} \qquad (4.13)$$

សូមបញ្ជាក់ថា គ្រប់អថេរដែលមានបំណែងចែកធម្មតា អាចបង្វែរទៅជាបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ ដោយប្រើរូបមន្តពិន្ទុស្តង់ដារ (standard score)។

$$z = \frac{X_i - \mu}{\sigma} \qquad (4.14)$$

តម្លៃ z គឺជាចំនួនស្តង់ដារដែលXណាមួយឃ្លាតចេញពីតម្លៃមធ្យម

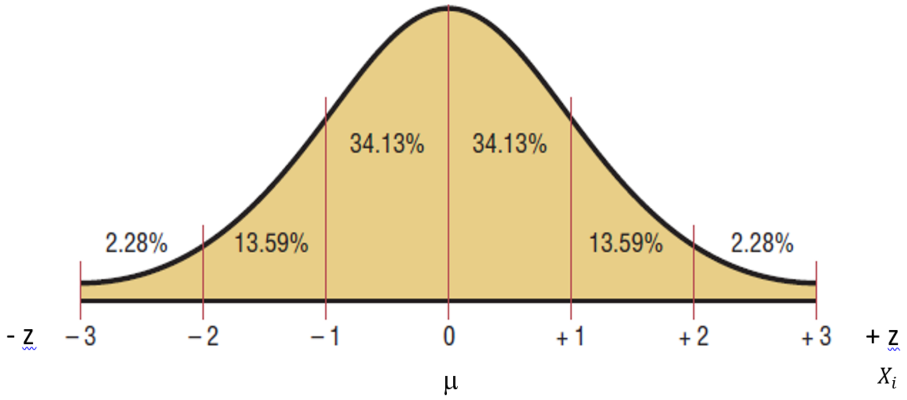
X_i =តម្លៃវត្ថុ (i)

μ =តម្លៃមធ្យម

σ =គម្លាតស្តង់ដារ

តារាង E ឧបសម្ព័ន្ធ C (Appendix C) បានបង្ហាញអំពីផ្ទៃដែលស្ថិតនៅក្រោមបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ ចំពោះតម្លៃ z ចាប់ពី-3.49 ដល់ 3.49។

រូប 4.4 បំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ



ប្រភព: Allan G. B (2009) ទំព័រ 304

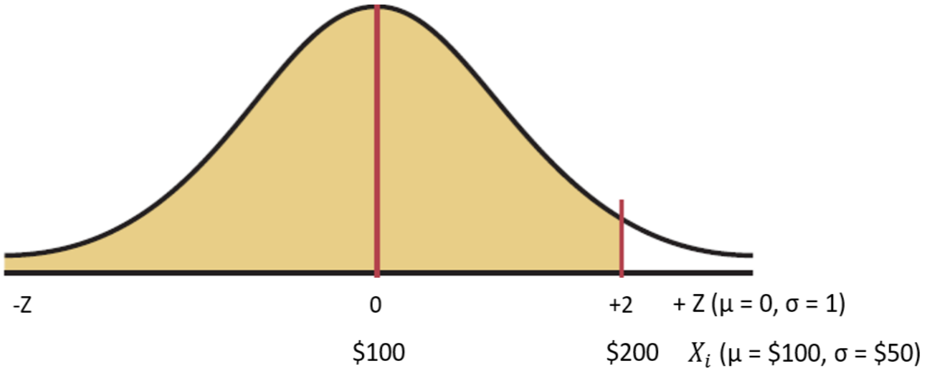
ឧទាហរណ៍ 4.7 ប្រសិនអថេរ X មានបំណែងចែកធម្មតា ហើយមានតម្លៃមធ្យម μ=100ដុល្លា និងគម្លាតស្តង់ដារ σ=50ដុល្លា។ ចូលគណនាតម្លៃ z ដែលមាន X=200ដុល្លា ?

អនុវត្តរូបមន្ត (4.14)

$$z = \frac{X_i-\mu}{\sigma} = \frac{200-100}{50} = 2.0$$

មានន័យថា X = 200 ដុល្លា គឺស្មើនឹង 2 ស្តង់ដារ (2 ដងនៃ 50 ដុល្លា) នៅពីលើតម្លៃមធ្យម 100 ដុល្លា (រូប 4.5)។

រូប 4.5 បំណែងចែកធម្មតា μ=100ដុល្លា σ=50ដុល្លានិង X=200ដុល្លា



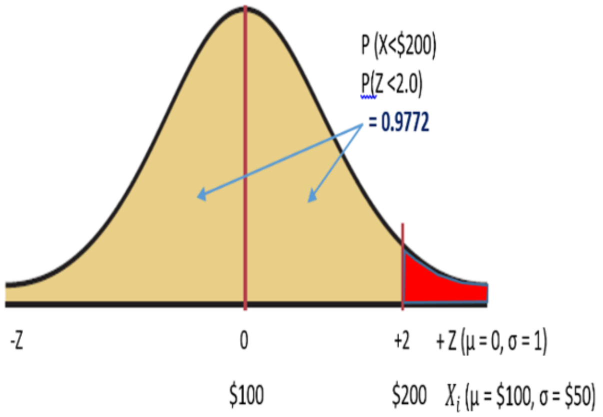
ប្រភព: ការបង្ហាញរបស់អ្នកនិពន្ធ

4.6.2 គណនាប្រូបាប៊ីលីតេនៃបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ

បន្ទាប់ពីគណនាតម្លៃ z បានហើយ យើងអាចគណនាប្រូបាប៊ីលីតេ ដែលជាផ្ទៃនៃខ្សែកោង។ ផ្ទៃនេះចែកចេញជា 3 តំបន់ 1) ផ្ទៃនៅខាងឆ្វេងនៃ z, 2) ផ្ទៃនៅខាងស្តាំនៃ z, និង 3) ផ្ទៃនៅចន្លោះនៃ z

1. ផ្ទៃនៅខាងឆ្វេងនៃ z

រូប 4.6 μ=$100ដុល្លា σ=$50ដុល្លា ហើយ p(X <$200)



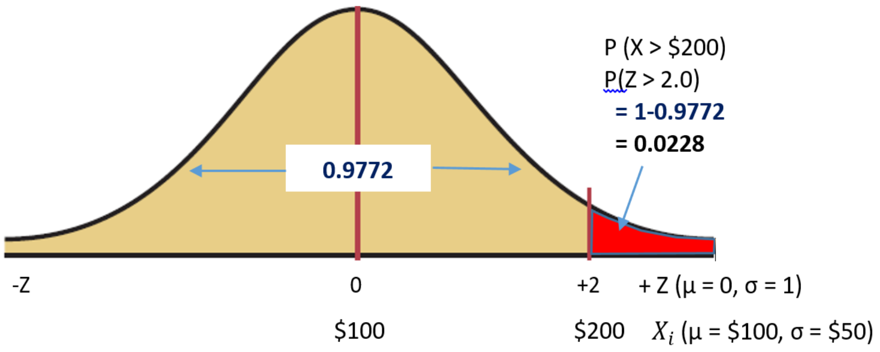
Z	0.00	0.01	0.02.......	0.09
0.00	....	....	....	....
0.01	....	....	....	....
0.02	....	....	....	....
......				
......				
2.00 →	0.9772	....	....	....
......	....	....	....	....

ប្រភព: ការបង្ហាញរបស់អ្នកនិពន្ធ

តារាងE ឧបសម្ព័ន្ធ C កំណត់អំពីផ្ទៃនៃបំណែងចែកធម្មតានៅខាងឆ្វេងនៃ z= 2.00 យើងនឹងបានផ្ទៃរបស់z គឺ 0.9772 ឬ 97.72% ដូចគំនូសតាងខាងលើ។ នេះមានន័យថា μ=100ដុល្លា σ=50ដុល្លា ប្រូបាប៊ីលីតេដែល X តូចជាង 200ដុល្លា (p(X <$200)) គឺស្មើនឹង 97.72% ។

2. ផ្ទៃនៅខាងស្តាំនៃ z

រូប 4.7 μ=\$100ដុល្លា σ=\$50ដុល្លា ហើយ p(X >\$200)



ប្រភព:ការបង្ហាញរបស់អ្នកនិពន្ធ

ឧបមាថា យើងចង់ដឹងអំពីកន្ទុយខាងស្តាំវិញម្តង គឺប្រូបាបីលីតេដែល X ធំជាជាង 200ដុល្លា (p(X> 200ដុល្លា)) គឺស្មើនឹង 1-0.9772 ឬ ស្មើនឹង 0.0228 នេះមានន័យថា μ=100ដុល្លា σ=50ដុល្លា ប្រូបាបីលីតេដែល X ធំជាង 200ដុល្លា (p(X > 200ដុល្លា)) គឺស្មើនឹង 2.28%

3. ផ្ទៃ z នៅចន្លោះ

ឧបមាថា μ=100ដុល្លា σ=50ដុល្លា ចូលគណនាប្រូបាបីលីតេដែល P(\$31.5 < X <\$200)

- ជំហ៊ានទី 1 គណនាតម្លៃ z ពេលដែល X = \$31.5

$$z = \frac{X_i - \mu}{\sigma} = \frac{31.5 - 100}{50} = -1.37$$

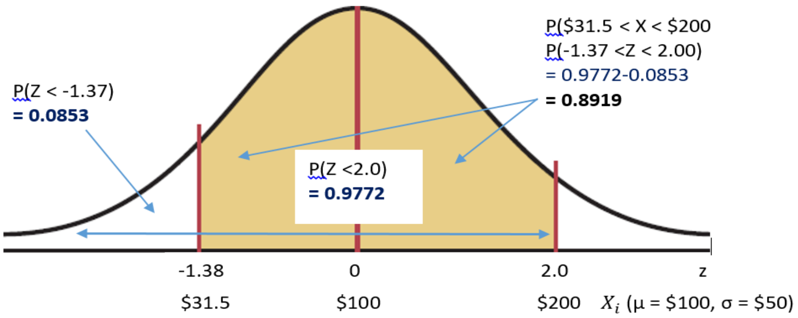
តាមតារាង E យើងបាន ប្រូបាបីលីតេ P(Z < -1.37) គឺស្មើនឹង 0.0853

- ជំហ៊ានទី 2 គណនាតម្លៃ z ពេលដែល X = \$200

តាមឧទាហរណ៍ខាងលើ យើងបានគណនាម្តងពេល X = \$200 យើងបាន z=2.0 ហើយ P (Z < 2.0) គឺស្មើនឹង 0.9772។

ដើម្បីរកប្រូបាបីលីតេ P(\$31.5 < X < \$200) គឺស្មើផ្ទៃដែលតូចជា \$200 ដកនិងផ្ទៃដែលតូចជាង \$31.5 ឬ P(-1.37 <Z < 2.00) = 0.9772-0.0853= 0.8919 ដូចរូប 4.8។ នេះមានន័យថាពេលដែល μ =\$100 σ=\$50 ប្រូបាបីលីតេ P(\$31.5 < X < \$200) គឺស្មើនឹង 89.19%។

រូប 4.8 μ=\$100ដុល្លា σ=\$50ដុល្លា ហើយ P(\$31.5 < X < \$200)



ប្រភព:ការបង្ហាញរបស់អ្នកនិពន្ធ

4.6.3 ការវាយតម្លៃន័រម៉ាល់លីតធី

មានអថេរជាប់ជាច្រើនត្រូវបានប្រើប្រាស់នៅក្នុងពាណិជ្ជកម្ម មានទម្រង់ដូចទៅនឹងបំណែងចែកធម្មតា។ ដើម្បីកំណត់ថា សំណុំទិន្នន័យមានទម្រង់ដូចទៅនឹងបំណែងចែកធម្មតាដែរឬយ៉ាងណា អ្នកអាចធ្វើការវាយតម្លៃន័រម៉ាល់លីតធី (normality) តាមរយៈការប្រៀបធៀបចរិកលក្ខណៈទិន្នន័យ ទៅនឹងវិបាកនៃបំណែងចែកធម្មតា។

1. វិបាកនៃបំណែងចែកធម្មតា៖
   - ស៊ីមេទ្រី គឺមានន័យថតម្លៃមធ្យមនិងមេដ្យានស្មើគ្នា
   - មានរាងដូចជួង
   - ចន្លោះនៃចតុភាគ $(Q_3 - Q_1)$ គឺស្មើនឹង1.33ដងនៃគម្លាតស្តង់ដារ
   - រ៉ង់ (range) គឺប្រហែលនឹង 6ដងនៃគម្លាតស្តង់ដារ

ក្រៅពីវិធីសាស្ត្រប្រៀបធៀបចរិកលក្ខណៈរបស់ទិន្នន័យទៅនឹងវិបាករបស់បំណែងចែកធម្មតា អ្នកអាចវាយតម្លៃ ន័រម៉ាល់លីតធី តាមវិធីសាស្ត្រមួយចំនួនទៀតដូចជា៖

- សង់ធាត (Charts) រួចធ្វើការអង្កេតទម្រង់របស់ធាត ទាំងនោះ តួយ៉ាងដូចជាធាតដើមនិងស្លឹក (stem-leave) គំនូសប្រអប់ (box-plot) ឬហ៊ីស្តូក្រាមធាត (histogram) ជាដើម។
- ការគណនាស្ថិតិបែបបរិយាយ (descriptive statistics) រួចធ្វើការប្រៀបធៀបស្ថិតិទាំងនោះទៅនឹងវិបាកនៃបំណែងចែកធម្មតា។ ប្រៀបធៀបតម្លៃមធ្យមនិងមេដ្យាន ផ្ទៀងផ្ទាត់ថា ចន្លោះចតុភាគ (interquartile range) គឺស្មើទៅនឹង 1.333ដង នៃគម្លាតស្តង់ដារដែរឬអត់? ហើយរ៉ង់(range) ស្មើទៅនឹង 6 ដងគម្លាតស្តង់ដារដែរឬអត់?

ឧទារណ៍ 4.8 ឥឡូវយើងត្រឡប់ទៅមើលបំណែងចែកអាយុក្នុងជំពូក 2 ទិន្នន័យនៅក្នុង Table 2-3។ រូបភាព 4.9 គឺជាលទ្ធផលនៃស្ថិតិបែបបរិយាយដែលបានមកពីការប្រើ SPSS។

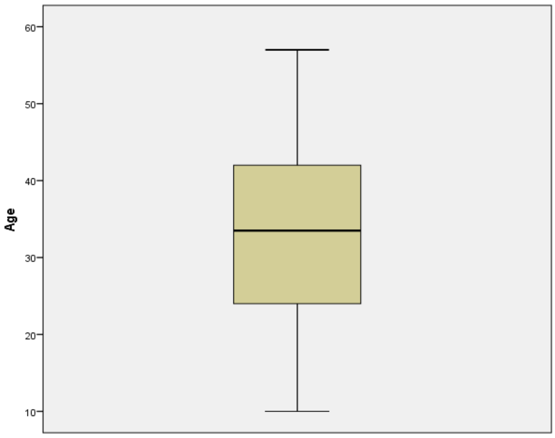
រូបភាព 4.9 ស្ថិតិបែបបរិយាយនៃបំណែងចែកអាយុអតិថិជន20នាក់

Statistics

Age

N	Valid	20
	Missing	0
Mean		32.75
Std. Error of Mean		2.817
Median		33.50
Std. Deviation		12.599
Skewness		.072
Std. Error of Skewness		.512
Kurtosis		-.487
Std. Error of Kurtosis		.992
Range		47
Percentiles	25	24.00
	50	33.50
	75	42.50

រូបភាព 4.10 គំនូសតាងអាយុអតិថិជន20នាក់



រូប 4.11 មែកនិងធាងនៃអាយុអតិថិជនទាំ20នាក់

```
Age

Age Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     3.00        1 .  047
     5.00        2 .  14467
     6.00        3 .  134578
     4.00        4 .  1346
     2.00        5 .  37

 Stem width:        10
 Each leaf:       1 case(s)
```

រូប 4.12 គំនូស q-q នៃអាយុអតិថិជនទាំ20នាក់

តាមរយៈរូប 4.9 ដល់ 4.12 យើងអាចធ្វើការអេះអាងដូចតទៅ៖

- តម្លៃមធ្យម 32.75 ខុសពីមេដ្យាន 33.50

- ចន្លោះនៃចតុភាគ ($Q_3 - Q_1$) = 50 គឺប្រហែល 50/12.6=3.97 (បំណែងចែកធម្មតាគឺតម្រូវឲ្យស្មើនឹង 1.33 ដងនៃគម្លាតស្តង់ដារ)
- គំនូសតាង Boxplot គឺមិនសូវផៀង (skewed =0.072) និងកឺតូស៊ីសស្មើ -0.487 (បំណែងចែកធម្មតាគឺតម្រូវឲ្យ skewed =0, kurtosis =0)
- រ៉េង (range) ស្មើនឹង 47 គឺស្មើនឹង 3.73 ដងនៃគម្លាតស្តង់ដា (បំណែងចែកធម្មតាគឺតម្រូវឲ្យស្មើនឹង 6ដងនៃគម្លាតស្តង់ដារ)
- រូប 4.11 គឺជា ដើមនិងសន្លឹក បញ្ជាក់ថាចំនួនប្រេកង់នៅកណ្តាលគឺមានចំនួនច្រើនជាងគេ ហើយវាបញ្ជាក់ឲ្យឃើញថាវាហាក់ដូចបំណែងចែកធម្មតា
- រូប 4.12 គំនូស q-q បញ្ជាក់ថា មេដ្យានហាក់ដូចស្ថិតនៅកណ្តាលនៃចតុភាគទី2និងទី3

ផ្អែកលើអំណះអំណាងទាំងអស់ខាងលើ យើងអាចសន្និដ្ឋានបានថា ទិន្នន័យអាយុអតិថិជនប្រហែលជាមានលក្ខណៈដូចបំណែងចែកធម្មតា។

2. តេស្តស្ថិតិ

ការប្រៀបធៀបស្ថិតិបរិយាយទៅនឹងវិបាករបស់បំណែងចែកធម្មតា ព្រមទាំងគំនួសតាងផ្សេងៗនៅមានលក្ខណៈស្រពិចស្រពិល មិនអាចធ្វើកំណត់បានថាតើទិន្នន័យនោះមានលក្ខណៈ ន័រម៉ាល់លីតធីឬ យ៉ាងណា? ឥឡូវមានតែការប្រើតេស្តស្ថិតិមួយទេដែលអាចអារកាត់បញ្ហានេះ។ មានតេស្តិតិជាច្រើនដើម្បីតេស្ត ន័រម៉ាល់លីតធី ក្នុងចំណោមតេស្តទាំងនោះ មានតេស្តមួយដែលគេនិយមជាងគេក្នុងពិភពអ្នកសេដ្ឋកិច្ច គឺ ចាកបេរ៉ាតេស្ត (Jarque-Bera, JB) ហើយវាមានរូបមន្តដូចខាងក្រោម៖

$$JB = \frac{n}{6} * \left(S^2 + \frac{(K-3)^2}{4}\right) \qquad (4.15)$$

ហើយ

S=ភាពផៀង (skewness)

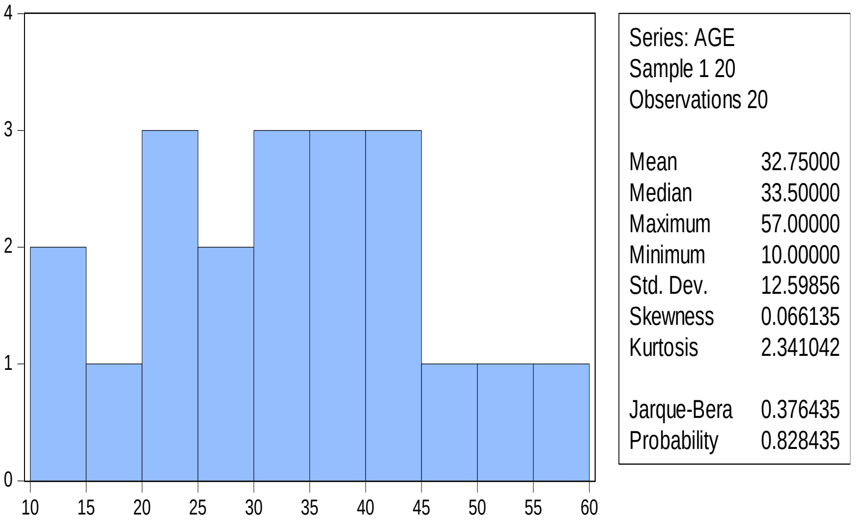
K =កឺតូស៊ីស kurtosis

$JB \sim \chi^2_2$

JB មានបំណែងចែកដូចបំណែងចែកឆាយស្ការដែលមានដឺគ្រីអ័វគ្រីដឹមស្មើពីរ (degree of freedom, df=2)។ ពេលតម្លៃភីវែលលូ (p-value) របស់ឆាយស្ការធំជាង5% ($p>0.05$) យើងអាចមិនបដិសេធនូវសម្មតិកម្មមោឃៈដែលថាអថេរនោះមានបំណែងចែកធម្មតា ផ្ទុយមកវិញ បើតម្លៃភីវែលលូរបស់ឆាយស្ការតូចជាង5% ($p <0.05$)គឺអថេរនោះមានបំណែងចែកមិនធម្មតា (not normal distribution)។

ឥឡូវយើងត្រឡប់មកឧទាហរណ៍អាយុអតិថិជនរបស់យើង។ យើងអាចគណនាតម្លៃ JB យើងត្រូវដឹងតម្លៃកឺតូស៊ីស (kurtosis) តម្លៃផៀង (skewness), និងចំនួនប៉ាន់គំរូ n។ យើងអាចរកតម្លៃជាតុចូលទាំងនេះតាមរយៈស្ថិតិបែបបរិយាយ(descriptive statistics) រួចជំនួសចូលក្នុងរូបមន្ត 4.15 ឬគណនាតាម Eviews។ ខាងក្រោមនេះគឺជាលទ្ធផល JB តេស្តដែលបានមកពី Eviews

រូប 4.12 លទ្ធផល JB តេស្ត ដោយ Eviews 9



JB =$\frac{20}{6}*(0.066^2+\frac{(-2.34-3)^2}{4}$)=0.376 (p=0.828 > 0.05)

លទ្ធផល JB តេស្តមានតម្លៃធាយឈ្មេស្មើនឹង 0.376 (p = 0.828> 0.05 ដូច្នេះ H_0: អថេរអាយុរបស់អតិថិជន គឺមានមានបំណែងចែកធម្មតាគឺត្រូវបានខកខានមិនត្រូវបានបដិសេធ។

សំគាល់ តម្លៃ S =0.066, និង K=2.34 នៅក្នុងតារាង 4.12 គឺខុសពីលទ្ធផលរបស់ Excel ឬ SPSS នៅក្នុងតារាង 4.9 ទោះជាទិន្នន័យតែមួយក៏ដោយព្រោះរូបមន្តគណនា S និង K ខុសគ្នាទៅតាមកម្មវិធីនីមួយៗ។

4.7 **បំណែងចែកឯកសណ្ឋាន**

បំណែងចែកឯកសណ្ឋាន (uniform distribution) គឺមានតម្លៃនៃប្រូបាប៊ីលីតេដូចគ្នាដែលនៅចន្លោះតម្លៃតូចជាងគេ a និងធំជាងគេ b ឬហៅថាបំណែងចែកចតុកោណកែង(មើលរូប 4.13)។ សមីការ (4.16) កំណត់អំពីដង់ស៊ីតេនៃប្រូបាពីលីតេឯកសណ្ឋាន។

$$f(X)=\frac{1}{b-a} \qquad (4.16)$$

បើ $a \le X \le b$ ហើយ 0 មិនស្ថិតនៅចន្លោះនេះ

ហើយ

$a =$ តម្លៃអប្បរមា

$b =$ តម្លៃអតិបរមា

- តម្លៃមធ្យមនៃបំណែងចែកឯកសណ្ឋាន

$$\mu=\frac{a+b}{2} \qquad (4.17)$$

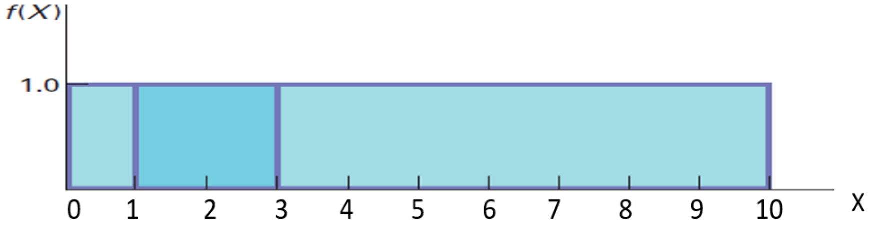
- វ៉ារីយ៉ង់និងគម្លាតស្តង់ដារនៃបំណែងចែកឯកសណ្ឋាន

$$\sigma^2=\frac{(b-a)^2}{12} \qquad (4.18)$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} \qquad (4.19)$$

ឧទាហរណ៍ 4.8 ចូលគណនាប្រូបាប៊ីលីតេនៃការជ្រើសរើសអថេរមួយដែលនៅចន្លោះ$1 និង $3 ចំពោះបំណែងចែកឯកសណ្ឋានដែលមាន a= $0និង b=$10

រូប 4.13 ប្រូបាប៊ីលីតេ P(1<X<3)ចំពោះបំណែងចែកឯកសណ្ឋាន a=0 និង b=10



រូប 4.13មានបណ្តោយ=3-1=2 និងទទឹង=1 ដូច្នេះP (1<X<3) គឺស្មើនឹងផ្ទៃ ចន្លោះ1ទៅ 3 ចែកនឹងផ្ទៃសរុប៖

P (1<X<3) = (2 * 1)/(1 * 10)=0.20 ឬ20%

តម្លៃមធ្យម: $\mu = \frac{0+10}{2} = 5$

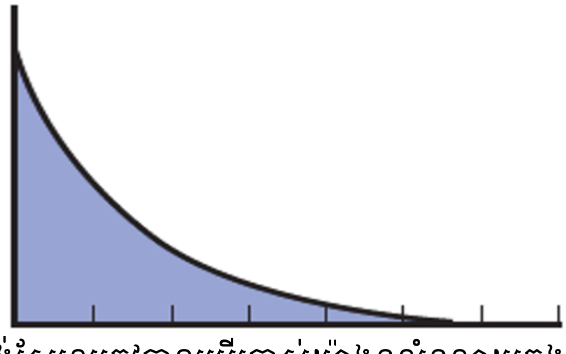
វ៉ារីយ៉ង់: $\sigma^2 = \frac{(10-0)^2}{12}$ =8.333

គម្លាតស្តង់ដារ:$\sigma = \sqrt{8.333} = 2.88$

4.8 បំណែងចែកអិចស្ប៉ូណង់ស្យែល

បំណែងចែកអិចស្ប៉ូណង់ស្យែល(the exponential distribution) គឺជាបំណែងចែកជាប់ដែលផៀងទៅខាងស្តាំចាប់ពី 0 រហូតដល់វិជ្ជមានមិនកំណត់ ដូចរូប4.14ខាងក្រោម។

រូប4.14 បំណែងចែកអិចស្ប៉ូណង់ស្យែល



បំណែងចែកអិចស្ប៉ូណង់ស្យែលត្រូវបានប្រើប្រាស់យ៉ាងទូលំទូលាយក្នុងទ្រឹស្តីរងចាំ តួរយ៉ាងដូចជា ការរងចាំនៅធនាគារ នៅមន្ទីរពេទ្យ ឬរងចាំសេវាផ្សេងៗ ជាដើម។

បំណែងចែកអិចស្បូណង់ស្យែលត្រូវបានកំណត់ដោយប៉ារ៉ាម៉ែត្រមួយ λ ចំនួនអ្នកមកដល់ក្នុងរយៈពេលមួយ។ វាមានទំនាក់ទំនងចេញពីនីតិវិធីព័រស្សុង។ ប្រូបាប៊ីលីតេចំពោះថេរវេលារវាងការមកដល់ត្រូវបានកំណត់ដោយសមីការ (4.20)៖

$$f(X) = \lambda e^{-\lambda x}\ for\ X > 0 \qquad (4.20)$$

ហើយ

$e = 2.7181$

$\lambda =$ ចំនួនមធ្យមនៃការមកដល់ក្នុងមួយឯកតានៃថេរវេលា

X=អថេរជាប់ដែលមានតម្លៃ $0 < X < \infty$

ហើយចំនួនមធ្យមនិងគម្លាតស្តង់ដារនៃការមកដល់ចន្លោះនៃពេលវេលាមួយ៖

$$\mu = \sigma = \frac{1}{\lambda} \qquad (4.21)$$

ពេល $\lambda = 1$ គេហៅថា បំណែងចែកអិចស្បូណង់ស្យែលដែលមានស្តង់ដារ។

$1/\lambda$=ចំនួនមធ្យមនៃចន្លោះពេលវេលាដែលមកដល់។

ឧបមាថា ចំនួនអ្នកមកដល់គិតជាមធ្យមក្នុងមួយនាទី $\lambda = 4$ ដូច្នេះរយៈពេលជាមធ្យមក្នុង**ចន្លោះ**នៃការមកដល់គឺ ¼=0.25 នាទីក្នុងមនុស្សម្នាក់។ សមីការ (4.22) កំណត់អំពីប្រូបាប៊ីលីតេត្រួតបន្តគ្នា (cumulative probability) នៃថេរវេលាមុននិងតិចជាងអ្នកបន្ទប់ Xមកដល់៖

Cumulative Exponential Probability

$$P(arrival\ time \leq X) = 1 - e^{-\lambda x} \qquad (4.22)$$

ឧទាហរណ៍ 4.9 អតិថិជនមកដល់ធនាគារ 20នាក់ក្នុងមួយម៉ោង បើសិនជាអតិថិជនមួយទើបមកដល់ គណនាប្រូបាប៊ីលីតេដែលអតិថិជនបន្ទប់នឹងមកដល់ក្នុងរយៈពេល6នាទីទៀត?

6នាទី X= 0.10ម៉ោង ហើយ $\lambda = 20$ (20 មកដល់ក្នុងមួយម៉ោង)

$$P(arrival\ time \leq X) = 1 - e^{-\lambda x} = 1 - 2.7182^{\ -20(0.10)} = 0.8647$$

ដូច្នេះប្រូបាប៊ីលីតេដែលអតិថិជនបន្ទាប់នឹងមកដល់ក្នុងរយៈពេល6នាទីទៀតគឺ 86.47%។

សេចក្តីសង្ខេប

ជំពូកនេះអ្នកបានសិក្សាបំណែងចែកប្រូបាប៊ីលីតេដាច់និងជាប់។ ប្រូបាប៊ីលីតេដាច់គឺជាបំណែងចែកនៃតម្លៃដែលអាចរាប់បាន មានចំនួនកំណត់និងមានតម្លៃវិជ្ជមាន។ រីឯបំណែងចែកប្រូបាប៊ីលីតេដាច់រួមមាន៖ បំណែងចែកប៊ីន័រមីល និងព័រស្សុង។ ក្នុងបំណែងចែកប្រូបាប៊ីលីតេដាច់ អ្នកបានសិក្សាតម្លៃរំពឹងទុក វ៉ារីយ៉ង់ និងគម្លាតស្តង់ដាររបស់វា។ ដោយឡែកចំពោះបំណែងចែកប្រូបាប៊ីលីតេជាប់វិញ អ្នកបានសិក្សាអំពី បំណែងចែកធម្មតា អ៊ិចស្ប៉ូណង់ស្យែល និតបំណែងចែកឯកសណ្ឋាន។ ចំពោះបំណែងចែកធម្មតា អ្នកក៏បានសិក្សាផងដែរអំពីរបៀបវាយតម្លៃ ដូចជាការប្រើគំនូសដើមនិងសន្លឹកនិង JB តេស្តថែមទៀត។

សំណួរនិងលំហាត់

I. **សំណួរ**

1. តើតម្លៃរំពឹងទុកនៃបំណែងចែកប្រូបាប៊ីលីតេមានន័យដូចម្តេច?

2. អ្វីទៅដែលហៅថាបំណែងចែកទ្វេធា ?

3. ពេលណាដែលត្រូវប្រើប្រាស់បំណែចែក hypergeometric ?

II. លំហាត់

4. ឧបមាថា អ្នកយកលុយចំនួន 1000 ដុល្លាទៅធ្វើការវិនិយោគទៅលើក្រុមហ៊ុនពីរ X និង Y ដែលស្ថិតនៅក្រោមស្ថានភាពសេដ្ឋកិច្ចនិងបំណែងចែកប្រូបាបីលីតេផ្សេងៗដូចតារាងខាងក្រោម៖

5.

		ប្រាក់ចំនេញ (%)	
ប្រូបាប៊ីលីតេ	ស្ថានភាពសេដ្ឋកិច្ច	Stock X	Stock Y
0.15	ធ្លាក់	-100	50
0.25	ធម្មតា	0	100
0.45	លូតលាស់មធ្យម	80	-20
0.15	រីកចំរើនខ្លាំង	150	-100

a. ចូលគណនាអាត្រាប្រាក់ចំណេញដែលរំពឹងទុកទៅលើការវិនិយោគភាគហ៊ុន X និង Y

b. គណនាគម្លាតស្តង់ដារនៃអាត្រាប្រាក់ចំណេញរបស់ភាគហ៊ុន X និង Y

c. គណនាកូវ៉ារីយ៉ង់អាត្រាប្រាក់ចំណេញរបស់ភាគហ៊ុន X និង Y និងធ្វើការពន្យល់

d. តើអ្នកនឹងវិនិយោគទៅលើភាគហ៊ុនមួយណា ?ចូលធ្វើការពន្យល់

5. បើ n=5 និង π=0.40 គណនាប្រូបាប៊ីលីតេដែល

a. X=4

b. $X \leq 3$

c. $X < 2$

d. X>1

5. អាជ្ញាធរអាកាសចរណ៍មួយបានរាយការណ៍ថា អាត្រានៃការមកដល់ទាន់ពេលវេលា (on- time arrival rate) នៃការហោះហើរក្នុងស្រុកមួយគឺ 0.825។ ចូលប្រើបំណែងចែកទ្វេធាដើម្បីគណនានូវប្រូបាប៊ីលីតេសម្រាប់ការមកដល់ 6 ជើងបន្ទាប់។

a. មាន 4 ជើងទាន់ពេលវេលា ?

b. ទាំង 6 ជើងទាន់ពេលវេលា ?

c. យ៉ាងហោចណាស់មាន 4 ជើងទាន់ពេលវេលា ?

d. ចូលគណនាចំនួនមកដល់ជាមធ្យម និងគម្លាតស្តង់ដារនៃការមកដល់ដែលទាន់ពេលវាលា។

6. R. H. Bruskin Associates Market Research បានរកឃើញថា 40% នៃពលរដ្ឋរបស់អាមេរិកមិនបានគិតថា ទាល់តែមានសញ្ញាបត្រមហាវិទ្យាល័យទើបធ្វើជំនួញបានជោគជ័យ។ ប្រសិនជាការជ្រើសរើសប៉ាន់គំរូមួយដោយចៃដន្យដែលមានចំនួនពលរដ្ឋអាមេរិក5 នាក់ ចូលគណនាប្រូបាប៊ីលីតេខាងក្រោម៖

a. មាន 2 នាក់យល់ស្របតាមការរកឃើញនោះ

b. យ៉ាងច្រើនបំផុត 3 នាក់នឹងយល់ព្រមតាមការឃើញនោះ

c. យ៉ាងហោចបំផុត 2 នាក់នឹងយល់ព្រមតាមការឃើញនោះ

d. តិចជាង 3 នាក់នឹងយល់ព្រមតាមការឃើញនោះ

ប្រភព: *American* by Daniel Evans Weiss.

ចំឡើយ a=0.346, b= 0.913, c=0.663, d= 0.683

7. សាវនករមួយរូបបានធ្វើការជ្រើសរើសប៉ាន់គំរូនៃប័ណ្ណបង់ពន្ធចំនួន 6 ដើម្បីធ្វើសាវនកម្ម។ ប្រើសិនជាមានប័ណ្ណបង់ពន្ធចំនួន 2 ឬច្រើន រកឃើញថាមិនត្រឹមត្រូវ នោះប័ណ្ណទាំងអស់ 100 ដែលជាចំនួនសកលនិងត្រូវបានត្រួតពិនិត្យ។ ចូលគណនាប្រូបាប៊ីលីតេដែលចំនួនសកលទាំងមូលនិងត្រូវបានត្រួតពិនិត្យ បើសិនជាចំនួនពិតប្រាកដដែលមិនត្រឹមត្រូវនៅក្នុងចំនួនសកលគឺ

a. 25 ?

b. 30 ?

c. 5 ?

d. 10 ?

8. អាត្រានៃការបោះបង់ការសិក្សានៅវិទ្យាល័យរបស់សិស្សអាមេរិក ប្រហែល 10.3% មុនបញ្ចប់ការសិក្សាទុតិយកភូមិ។ ប្រសិនជាសិស្ស 10 នាក់ដើរចូលសាលាដោយចៃដន្យ ចូលគណនាប្រូបាប៊ីលីតេះ

a. ដែលមានសិស្សបោះបង់ការសិក្សាមិនលើសពីពីរនាក់

b. យ៉ាថហោចណាស់មាន 6នាក់រៀនចប់

c. សិស្សទាំង 10 រៀនចប់

ប្រភព: www.infoplease.com

ចំឡើយ:

a. = 0.925

b. = 0.998

c. = 0.337

9. កាសិក្សាមួយទាក់ទងទៅនឹងចោរកម្មនៅតំបន់មួយចំនួនាពេលថ្មីៗនេះបានរកឃើញថា មានមួយករណីក្នុងចំណោមប្រជាជន 20,000 នាក់។ ប្រសិនជាទីក្រុងមួយមានប្រជាជន 80,000 នាក់ ចូលគណនាប្រូបាប៊ីលីតេដូចខាងក្រោមះ

a. ចោរកម្ម 0

b. ចោរកម្ម 1

c. ចោរកម្ម 2

d. ចោរកម្ម 3 ឬ ច្រើន

ប្រភពះ Allan G. B (2009)

ចំឡើយ:

a. = 0.0183

b. = 0.0733

c. = 0.1465

d. = 0.7619

ប្រភព: Allan G. B (2009)

10. ការត្រួតពិនិត្យការបញ្ចេញផ្សែងពុលរបស់រថយន្ត។ ប្រសិនជា 3% នៃរថយន្តទាំងអស់មិនបានគោរពតាមស្តង់ដារបញ្ចេញផ្សែងពល ចូលគណនាប្រូបាប៊ីលីតេដែលប៉ាន់គំរូរថយន្តចំនួន 90 ត្រូវបានជ្រើសរើស មានរថយន្ត 3គ្រឿងដែលមិនគោរពស្តង់ដារបញ្ចេញផ្សែងពុលដោយប្រើបំណែងចែក ព័រសុង។

ចំឡើយ:

0.2205

ប្រភព៖ Allan G. B (2009)

ជំពូកទី 5 ការប្រមូលទិន្នន័យនិងបំណែងចែកប៉ាន់គំរូ

(Sampling and Sample Distribution)

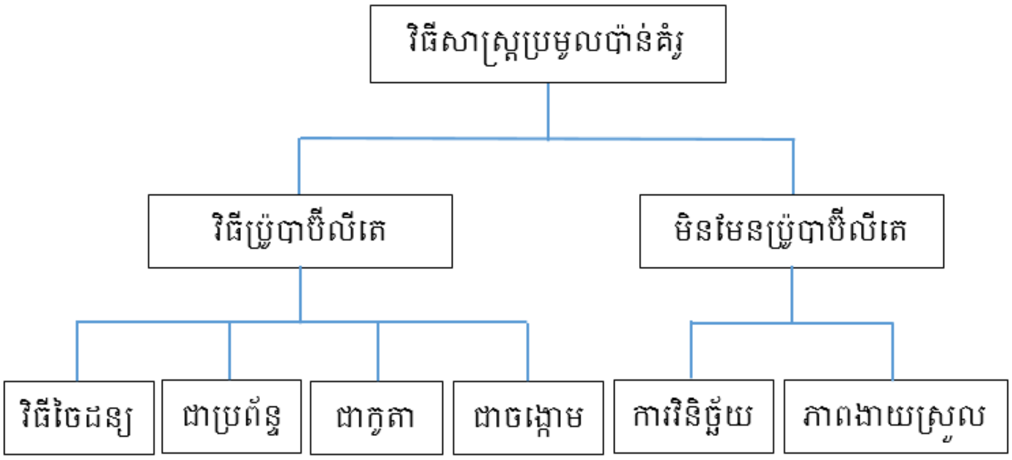
គោលបំណង

- ភាពខុសគ្នានៃវិធីសាស្ត្រប្រមូលប៉ាន់គំរូ
- ទ្រឹស្តីនៃប្រមូលប៉ាន់គំរូ
- គណនាប្រូបាប៊ីលីតេនៃប៉ាន់គំរូសម្រាប់តម្លៃមធ្យមនិងសមាមាត្រ
- សារៈសំខាន់នៃទ្រឹស្តីចំណុចកណ្តាល

5.1 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូ

ជាទូទៅវិធីសាស្ត្រប្រមូលប៉ាន់គំរូ (sampling method) មានពីរបែប៖ តាមបែបប្រូបាប៊ីលីតេ (probability sampling) និងមិនមែនប្រូបាប៊ីលីតេ (non-probability sampling)[8]មើលរូប 5.1។

រូប 5.1 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូ



ក្នុងវិធីសាសាស្ត្រប្រមូលប៉ាន់គំរូប្រូបាប៊ីលីតេ (probability sampling) គ្រប់វត្ថុឬបុគ្គលទាំងអស់ដែលជាសមាជិកនៃចំនួនសកល នឹងមានឱកាសស្មើគ្នាត្រូវបានធ្វើការជ្រើសរើស ហើយការជ្រើសរើសគឺផ្អែកលើលក្ខណៈចៃដន្យ (random selection)។ បើសិនជាអ្នកចង់បានលទ្ធផលដែលដំណាងឲ្យចំនួនសកលទាំងមូល។ វិធីសាសាស្ត្រប្រមូលប៉ាន់គំរូប្រូបាប៊ីលីតេ គឺជាជំរើសដែលមានសុពលភាពជាងគេ[9]។ ដោយឡែកវិធីសាស្ត្រប្រមូលប៉ាន់គំរូមិនមែនប្រូបាប៊ីលីតេ គឺជាវិធីដែលអ្នកស្រាវជ្រាវគិតថាងាយស្រួលក្នុងការ

---

[8] Barenson et all (2012) Basic Business Statistics: Concepts and Applications. Twelfth Edition. Prentice Hall

[9] សំរួលចេញពី៖ https://www.scribbr.com/methodology/sampling

ប្រមូល។ វិធីសាស្ត្រនេះចំណាយថវិការតិចក្នុងការប្រមូលទិន្នន័យ ប៉ុន្តែវាមានគុណវិបត្តិត្រង់ថា លទ្ធផលនៃការសិក្សាគឺមិនអាចធ្វើការអំណះអំណាងអំពីចំនួនសាកលបាន។

5.1.1 **វិធីសាស្ត្រចៃដន្យ**

នៅក្នុងវិធីសាស្ត្រចៃដន្យ គ្រប់វត្ថុឬបុគ្គលដែលជាគោលដៅក្នុងការជ្រើសរើសគឺមានឱកាសស្មើគ្នា។ ក្នុងវិធីសាស្ត្រប្រមូលប៉ានគំរូបែបនេះ អ្នកអាចធ្វើតាមការប្រមូលប៉ាន់គំរូដោយជំនួស (replacement) និងមិនជំនួស (without replacement) ក៏បាន ក៏ប៉ុន្តែអ្នកត្រូវការតារាងចៃដន្យ (random table)(មើលឧបសម្ព័ន្ធ E) ដើម្បីធ្វើការជ្រើសរើសប៉ាន់គំរូ។

ឧទាហរណ៍ 5.1 ចំនួននិស្សិតជំនាន់26សាកលវិទ្យាល័យជាតិគ្រប់គ្រងដែលត្រូវទទួលសញ្ញាបត្រនៅឆ្នាំ 2021 មានទាំងអស់ 3000នាក់ ហើយថ្នាក់ដឹកនាំសាកលវិទ្យាល័យចង់ធ្វើការជ្រើសរើសនិស្សិតចំនួន 50 នាក់ជាតំណាងដើម្បីឡើងទទួលសញ្ញាបត្រ។ សន្មត់ថាសាកលវិទ្យាល័យសំរេចជ្រើសរើសដោយប្រើវិធីសាស្ត្រចៃដន្យ។

- **ជំហានទី1៖** ត្រូវរៀបចំបញ្ជីឈ្មោះនិស្សិតទាំងអស់ក្នុងតារាងតែមួយដែលមានលេខរៀងពី 1 ដល់ 3000
- **ជំហានទី2៖** រកតារាងចៃដន្យ (random table) ឧទាហរណ៍ តារាង 5.1 ខាងក្រោម

តារាង 5.1 ផ្នែកមួយនៃតារាងចៃដន្យ(portion of a random table)

39087	71938	40355	54324	08401	26299	49420	59208
55700	24586	93247	32596	11865	63397	44251	43189
14756	23997	78643	75912	83832	32768	18928	57070
32166	53251	70654	92827	63491	04233	33825	69662

- **ជំហានទី3៖**យើងចាប់ផ្តើមធ្វើការជ្រើសរើស ឧបមាថា យើងចាប់ផ្តើមពីជួរដេកទី2និងជួរឈរទី2 គឺត្រូវនឹងលខ 24586។ ដោយសារតែយើងត្រូវការជ្រើសរើសប៉ាន់គំរូតែ 50 យើងប្រើតែពីរលេខតែប៉ុណ្ណោះ ឧបមាថាយើងយកពីរលេខខាងដើម (ពីរលេខខាងចុងក៏បាន) គឺលេខ 24 ដូច្នេះនិស្សិតដែលមានលេខរៀងទី24 គឺជាបេក្ខជនដែលនឹងត្រូវឡើងទទួលសញ្ញាបត្រ។ អ្នកបន្ទាប់ (ពីឆ្វេងទៅស្តាំ) គឺលេខរៀងទី 93 ហើយបន្ទាប់មកទៀតគឺលេខរៀងទី32។ យើងនឹងធ្វើការជ្រើសរើសបែបនេះរហូតដល់បាននិស្សិតចំនួន 50នាក់។

5.1.2 **វិធីសាស្ត្រប្រមូលប៉ាន់គំរូជាប្រព័ន្ធ**

វិធីសាស្ត្រប្រមូលប៉ាន់គំរូជាប្រព័ន្ធ (systematic sampling) ស្រដៀងទៅនឹងវិធីសាស្ត្រចៃដន្យដែរ តែមានលក្ខណៈងាយជាងបន្តិច សមាជិកគ្រប់រូបរបស់ចំនួនសាកលត្រូវរៀបចំឲ្យមានលេខរៀងត្រឹមត្រូវ បន្ទាប់មកគេនឹងធ្វើការជ្រើសរើសប៉ាន់គំរូតាមចន្លោះមួយយ៉ាងទៀងទាត់។ ឥឡូវនេះយើត្រឡប់ទៅមើលឧទាហរណ៍ 5.1 ឡើងវិញ។ ដំបូងយើងកំណត់ចំនួននិស្សិតដែលត្រូវឡើងទទួលសញ្ញាបត្រគឺ n=50 ចេញពីចំនួនសកល N=3000 នាក់ ហើយយើងតាង k=ចន្លោះពីបុគ្គលឬវត្ថុមួយ ទៅបុគ្គលឬវត្ថុមួយទៀតដែលត្រូវជ្រើសរើស (k ជាចំនួនគត់)។ ដូច្នេះ k =N/n = 3000/50= 60។ មានន័យថា ក្នុងបញ្ជីនិស្សិតដែលមានលេខ

រៀងពី1 ដល់ 3000 គឺ 60 យកមួយ។ ឧបមាថា យើងធ្វើការជ្រើសរើសអ្នកទីមួយចេញដំណើរពីអ្នកដែលមានលេខរៀងទី10 អ្នកទី2 ដែលត្រូវជ្រើសរើសគឺអ្នកមានលេខរៀងទី70 រីឯអ្នកទី3គឺមានលេខរៀងទី130 ធ្វើបែបនេះរហូតដល់បានប៉ាន់គំរូហូតដល់គ្រប់ចំនួន 50នាក់។ សូមបញ្ជាក់ថា ចំណុចសំខាន់ជាងគេគឺការជ្រើសរើសអ្នកទី1 តើត្រូវចាប់ផ្តើមពីលេខប៉ុន្មានទៅ? លេខនេះគឺត្រូវចាប់ផ្តើមដោយវិធីចៃដន្យ។

5.1.3 វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមកូតា

វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមកូតា (cluster or quota sampling) ទាក់ទងនិងការបែងចែកចំនួនសកល (N) ទៅជាក្រុមតូចៗ។ ហើយក្រុមតូចៗទាំងនោះគួរតែមានចរិកលក្ខណៈប្រហាក់ប្រហែលគ្នា ឧទាហរណ៍ ភេទ ឆ្នាំសិក្សា ឬជំនាញ ជាដើម។ បន្ទាប់មកយើងធ្វើការជ្រើសរើសប៉ាន់គំរូដោយចៃដន្យចេញពីក្រុមតូចទាំងនោះ រួចបូករួមបញ្ចូលគ្នាឲបានគ្រប់ចំនួនប៉ាន់គំរូដែលបានកំណត់។

ឧទាហណ៍ 5.2 អ្នកស្រាវជាវម្នាក់ចង់ធ្វើការសិក្សាអំពី គម្រោងអនាគតរបស់សនិស្សិត។ គាត់សំរេចចិត្តធ្វើការស្ទង់មតិទៅលើនិស្សិតនៅសាកលវិទ្យាល័យជាតិគ្រប់គ្រង ដែលមានចំនួននិស្សិតសរុប 10000នាក់ក្នុងឆ្នាំសិក្សា 2020-2021 ហើយចំនួននេះត្រូវបានបែងចែកជាបួនជំនាញ៖ គណនេយ្យមាន 5000 នាក់ ឬ 50% ហិរញ្ញវត្ថុនិងធនាគារមាន 2000 នាក់ឬ 20% សេដ្ឋកិច្ចមាន 1000នាក់ ឬ 10% និង គ្រប់គ្រង 2000 នាក់ឬ 20%។ ទិន្នន័យទាំងនេះបានបង្ហាញនៅក្នុង តារាង5.2។

តារាង5.2 លទ្ធផលនៃប្រមូលប៉ាន់គំរូតាមកូតា

	ចំនួននិស្សិត	សមាមាត្រ	ប៉ាន់គំរូ
គណនេយ្យ	5,000	0.50	500×0.50 =250
ហិរញ្ញវត្ថុនិងធនាគារ	2,000	0.20	500×0.20 =100
សេដ្ឋកិច្ច	1,000	0.10	500×0.10 = 50
គ្រប់គ្រង	2,000	0.20	500×0.20 =100
សរុប	10,000		500

វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមកូតាគឺមានពីរជំហាន៖ ជំហានដំបូងគឺត្រូវកំណត់ចំនួនប៉ាន់គំរូរបស់ជំនាញនីមួយៗសិន តាមរយៈការគណនាតម្លៃសមាមាត្រ ឧទាហរណ៍ ចំនួនប៉ាន់គំរូនិស្សិតគណនេយ្យគឺ 5000*0.50 = 250 នាក់ ហិរញ្ញវត្ថុនិងធនាគារគឺ 500*0.20 = 100 នាក់ (មើល column ចុងក្រោយ) ចំណែកឯចំនួនប៉ាន់គំរូជំនាញសេដ្ឋកិច្ចនិងគ្រប់គ្រងគឺត្រូវបានប្រើវិធីសាស្ត្រគណនាដូចគ្នា។ បន្ទាប់ពីបានកំណត់ចំនួនប៉ាន់គំរូរួច អ្នកស្រាជ្រាវអាចបន្តជំហានទីពីរ គឺត្រូវរៀបចំឈ្មោះនិស្សិតទាំង 10000 នាក់ក្នុងបញ្ជីមួយ ហើយធ្វើការជ្រើរើសតាមវិធីសាស្ត្រចៃដន្យ ឲ្យបានគ្រប់ចំនួនកូតាដែលបានកំណត់ក្នុងជំហានទីមួយ។

5.1.4 **វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមក្រុម**

វិធីសាស្ត្រប្រមូលប៉ានគំរូតាមក្រុម (cluster sampling) ជាវិធីសាស្ត្រប្រមូលតាមបែបប្រូបាប៊ីលីតេ ដែលនៅក្នុងនោះអ្នកសិក្សាបានធ្វើការបែងចែកចំនួនសកលជាក្រុមតូចៗហៅថា cluster រួចធ្វើការជ្រើសរើសចំនួនប៉ាន់គំរូចេញពីក្រុមទាំងនោះ។ វិធីសាស្ត្រនេះច្រើនតែអនុវត្តទៅលើចំនួនសកលដែលមានបរិមាណច្រើន ហើយស្ថិតនៅលើភូមិសាស្ត្រធំទូលាយ ដូចជាខេត្ត ក្រុង/ស្រុក ដែនដីចែកចាយផលិតផល ឬមណ្ឌលបោះឆ្នោតជាដើម។ លើសពីនេះទៅទៀត វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមក្រុមគឺត្រូវការចំនួនប៉ាន់គំរូច្រើនហើយអាចអនុវត្តតែមួយឬច្រើនជំហាន។

ឧទាហណ៍ 5.3 ការប្រមូលប៉ាន់គំរូតាមក្រុមច្រើនជំហាន(multistage cluster sampling)៖ ឧបមាថា អ្នកចង់សិក្សាអំពីចំណង់ចំណូលចិត្តក្នុងការជ្រើសរើសជំនាញរបស់សិស្សថ្នាក់ទី12។

- ជំហានទី1 កំណត់ចំនួនសកល (define N population)៖ ឧទាហរណ៍ សិស្សថ្នាក់ទី12ទាំងអស់មាន 150000នាក់នៅឆ្នាំសិក្សា 2020-2021
- ជំហ៊ានទី2៖ បែងចែកប៉ាន់គំរូទាំងអស់ទៅជាក្រុម (cluster)៖ វិទ្យាល័យគ្រប់ខេត្តក្រុងទាំងអស់ សន្មត់ថាមាន 50 វិទ្យាល័យទូទាំង 25ខេត្តក្រុង
- ជំហានទី3៖ ធ្វើការជ្រើសរើសដោយប្រើវិធីចៃដន្យទៅលើវិទ្យាល័យទាំង 50 នោះ ឧបមាថា ចំនួនប៉ាន់គំរូដែលត្រូវការគឺ10 វិទ្យាល័យ
- ជំហានទី4៖ ធ្វើការជ្រើរើសប៉ាន់គំរូ (និស្សិត) ចេញពីវិទ្យាល័យទាំង10 នោះ។ បន្ទាប់មកអ្នកអាចប្រើវិធីសាស្ត្រជ្រើសរើសតាមកូតាបាន។ ឧទាហរណ៍ ចំនួនប៉ាន់គំរូសរុបដែលកំណត់គឺ 1000នាក់ (ចំនួនប៉ាន់គំរូសិស្ស)។ បន្ទាប់មកទៀត អ្នកត្រូវកំណត់កូតាចេញពីវិទ្យាល័យនីមួយៗ ឧទាហរណ៍ វិទ្យាល័យទី1 យកកូតាប៉ុន្មាន? វិទ្យាល័យទី2 យកកូតាប៉ុន្មាន? ដូចវិធីសាស្ត្រកូតាដែលនារៀបរាប់ខាងលើ
- ជំហានទី5៖ អ្នកអាចចាប់ផ្តើមប្រវិធីសាស្ត្រចៃដន្យដើម្បីធ្វើការជ្រើសរើសសិស្សចេញពីវិទ្យាល័យនីមួយៗ ឧបមាថា វិទ្យាល័យទី1 មានកូតាដែលត្រូវជ្រើសរើស100នាក់ ហើយវិទ្យាល័យនេះមានសិស្ស សរុប500នាក់ ដូច្នេះអ្នកត្រូវរៀចំបញ្ជីសិស្សមួយដែលមានសិស្សទាំងអស់500នាក់ រួចជ្រើសរើសដោយចៃដន្យយកចំនួន100នាក់។ ចំណែកឯវិទ្យាល័យផ្សេងៗក៏ត្រូវប្រើវិធីសាស្ត្រដូចគ្នា រហូតដល់បានចំនួនប៉ាន់គំរូគ្រប់ចំនួន 1000 នាក់ ។

5.2 **វិធីសាស្ត្រមិនមែនប្រូបាប៊ីលីតេ**

វិធីសាស្ត្រមិនមែនប្រូបាប៊ីលីតេ (non-probability sampling) គឺត្រូវបានគេប្រើប្រាស់មានលក្ខណៈចង្អៀត ឧទាហរណ៍ដូចជាការសិក្សាស្រាវជ្រាវឬការធ្វើសារណារបស់និស្សិតជាដើម។ វិធីសាស្ត្រនេះមានពីរគឺ វិធីសាស្ត្រវិនិច្ឆ័យ និងវិធិងាយស្រួល។

5.2.1 **វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមបែបវិនិច្ឆ័យ**

វិធីសាស្ត្រប្រមូលប៉ាន់គំរូតាមបែបវិនិច្ឆ័យ(judgement sampling) ឬហៅថា ឯកទេស (expert sampling) គឺវិធីសាស្ត្រមិនមែនប្រូបាប៊ីលីតេ ដែលធ្វើឡើងផ្អែកលើមូលដ្ឋានចំណេះដឹងរបស់អ្នក

ស្រាវជ្រាវ ឬអ្នកឯកទេស។ ដោយសារតែការស្រាវជ្រាវនេះគាត់តែងឡើដោយអ្នកឯកទេស លទ្ធផលដែលបានកមកគឺខុសតិចតួចបំផុត[10]។ ជាទូទៅការប្រមូលប៉ាន់គំរូបែបនេះត្រូវបានប្រើប្រាស់ក្នុងករណីដែលចំនួនសកលមានកំណត់ តួយ៉ាងដូចជាអ្នកដឹកនាំ ឬនាយកប្រតិបត្តិដែលមានចំនួនតិច។

ឧទាហរណ៍ទី1អំពីចំនួនប៉ាន់គំរូ៖ នៅក្នុងតារាង 5.2បង្ហាញថាចំនួននិស្សិតសេដ្ឋកិច្ចដែលត្រូវសាកសួរគឺ50នាក់ អ្នកសិក្សាយល់ឃើញថា ចំនួននេះហាក់ដូចជាតិចពេក ហើយចំនួននិស្សិតគណនេយ្យនិងហិរញ្ញវត្ថុគឺ 250 ហាក់ដូចជាច្រើនពេក អ្នកស្រាវជ្រាវអាចប្រើការវិនិច្ឆ័យរបស់ខ្លួន រួចបង្កើនចំនួនប៉ាន់គំរូនិស្សិតសេដ្ឋកិច្ចឡើងដល់ 100នាក់ ហើយចំនួននិស្សិតគណនេយ្យនិងហិរញ្ញវត្ថុកាត់បន្ថយត្រឹម 200នាក់វិញ ដោយយល់ថា ចំនួន200 នាក់គឺគ្រប់គ្រាន់ ចៀសវាងលើសចំនួនឥតប្រយោជន៍។

ឧទាហរណ៍ទី2 អំពីគោលដៅប៉ាន់គំរូ៖ អ្នកសិក្សាម្នាក់ចង់សិក្សាវាយតម្លៃទៅលើ "គោលនយោបាយភូមិឃុំមានសុវត្ថិភាព" សំនួរមួយចោទសួរថា តើគាត់ទៅសួរអ្នកណាទើបត្រឹមត្រូវ? សួរប្រជាជនក្នុងភូមិ ឬសួរនគរបាល? ចំលើយ គឺគួរតែសួរប្រជាជនក្នុងភូមិដែលជាអ្នកទទួលសេវាទើបត្រឹមត្រូវ។ តើគាត់អាចសួរនគរបាលខ្លះ និងប្រជានខ្លះ ហើយយកចំលើយទាំងពីរបូកបញ្ចូលគ្នារួចធ្វើការវិភាគបានទេ? ចំលើយគឺអាចទៅរួច ក៏ប៉ុន្តែ មិនអាចបូកបញ្ចូលគ្នារួចវិភាគនោះទេ ត្រូវធ្វើការវិភាគដោយឡែកពីគ្នា ពីព្រោះម្ខាងជាចំលើយរបស់អ្នកទទួលសេវា ហើយម្ខាងទៀតជាចំឡើយរបស់អ្នកផ្តល់សេវា មិនអាចបូកចូលគ្នា រួចធ្វើការវិភាគបែបនោះទេ។

5.2.2 **វិធីសាស្រ្តងាយស្រួល**

វិធីសាស្រ្តងាយស្រួល (convenience sampling) ជួនកាលហៅថាការប្រមូលប៉ាន់គំរូដែលមានគោលបំណង (purposive sampling) ឬហៅថា (authoritative sampling) គឺវិធីសាស្រ្តមិនមែនប្រ៉ូបាប៊ីលីតេ ដែលផ្អែកលើមូលដ្ឋានភាពងាយស្រួលទាក់ទងឬមិនសូវចំណាយច្រើនក្នុងការប្រមូលទិន្នន័យ។ ឧទាហរណ៍ ថ្ងៃនេះហាងពីហ្សានៅផ្សារឡាក់គី ធ្វើការស្ទង់មតិចំពោះអ្នកដែលទៅហូបពីហ្សានៅទីនោះ ទាក់ទងជាមួយរស់ជាតិពីហ្សារបស់គេ។ ការធ្វើបែបនេះគឺចំណាយតិច ងាយស្រួល មានលក្ខណៈមិនមែនប្រ៉ូបាប៊ីលីតេ ពីព្រោះអតិថិជនទាំងអស់មិនមានឱកាសស្មើគ្នានោះទេ អ្នកដែលមានឱកាសគឺអ្នកដែលទៅហូបពីហ្សានៅថ្ងៃនេះ ចំណែកអ្នកដែលមិនបានទៅថ្ងៃនេះគឺមិនមានឱកាសត្រូវបានសាកសួរឡើយ។

5.3 **ការវាយតម្លៃលើការប្រមូលប៉ាន់គំរូក្នុងករណីស្ទង់មតិ**

ពេលដែលអ្នកស្រាវជ្រាវប្រើប្រាស់ការស្ទង់មតិដើម្បីប្រមូលទិន្នន័យ គាត់ត្រូវធ្វើការវាយតម្លៃលើការប្រមូលប៉ាន់គំរូដែលបានប្រមូលមកវិញ (survey sampling evaluation)។ ហើយជំហានបន្ទាប់ គាត់ត្រូវវាយតម្លៃថា ការស្ទង់មតិនោះគឺផ្អែកលើប្រ៉ូបាប៊ីលីតេ ឬមិនប្រ៉ូបាប៊ីលីតេ។ បើការស្ទង់មតិនោះគឺផ្អែកលើប្រ៉ូបាប៊ីលីតេ លទ្ធផលដែលបានមកអាចធ្វើការប៉ាន់ស្មានជាទូទៅបាន (generalize the result) តែបើមិនផ្អែកលើប្រ៉ូបាពីលីតេ គឺមិនអាចប៉ាន់ស្មានជាទូទៅលើចំនួនសកលបានឡើយ ពីព្រោះវាលំអៀង (bias) ក៏ប៉ុន្តែអ្នកអាចសន្និដ្ឋានក្នុងដែនកំណត់នៃប៉ាន់គំរូតែប៉ុណ្ណោះ។

---

[10] ដកស្រង់ចេញពី៖ https://www.questionpro.com/blog/judgmental-sampling/

5.3.1 កំហុសក្នុងការស្ទង់មតិ

ទោះបីជាការស្ទង់មតិប្រើប្រូបាប៊ីលីតេក៏អាចមានកំហុសដែរ(error)។ ជាទូទៅគេអង្កេតឃើញថា អ្នកសិក្សាអាចច្រើនតែមានកំហុសចំនួនបួន៖

1. ការខកខានមិនបានបញ្ចូលប៉ាន់គំរូគ្រប់គ្រាន់ (Coverage Error): កើតឡើងពេលដែលការប្រមូលប៉ានគំរូមិនត្រឹមត្រូវ ឬអ្នកដែលត្រូវឆ្លើយសំណួរគ្មានឱកាសចូលរួម ឬត្រូវបានទាត់ចោល។ វាអាចផ្តល់លទ្ធផលលំអៀងលើការជ្រើសរើស (selection bias)។ ឧទាហរណ៍ សួរតែស្រីស្អាត ឬអ្នកមាន មិនសួរអ្នកក្រ សួរតែមន្ត្រីរាជការ មិនសួរពលរដ្ឋជាដើម។
2. ការខកខានមិនបានឆ្លើយ (Non-response Error): មិនមែនគ្រប់អ្នកគ្រប់គ្នាឆ្លើយនឹងសំនួររបស់អ្នកឡើយ ឧទាហរណ៍ ដូចជាអ្នកមាន គេតែងតែរង្វល់ មិនមានពេលឆ្លើយនឹងសំណួររបស់អ្នក។
3. មិនបានផ្តល់ឱកាសស្មើគ្នា (Sample Error): សំដៅលើឱកាសចូលរួមឆ្លើយមិនស្មើគ្នា ឧទាហរណ៍ អ្នកប្រើវិធីសាស្ត្រងាយស្រួល (convenience sampling) អាចផ្តល់នូវលទ្ធផលប្រមូលប៉ាន់គំរូមានកំហុស។
4. ការវាស់វែងមិនត្រឹមត្រូវ (Measurement Error): ជួនកាលហៅថា observation error សំដៅលើគុណភាពនៃការតាក់តែងសំណួរ ដូចជាសំណួរមិនច្បាស់លាស់ ឬសំណួរឲ្យគំរ៉ុយមុន។ ឧទាហរណ៍ទី1 អ្នកស្រាវជ្រាវម្នាក់តាក់តែងសំណួរឡាយស្កឹតស្កេល (Likert scale) ដើម្បីវាយតម្លៃលើផលិតផលមួយ: 1=មិនពេញចិត្តទាល់តែសោះ, 2= មិនពេញចិត្ត, 3=ពេញចិត្ត, និង 4=ពេញចិត្តខ្លាំង។ សំណួរបែបនេះមិនទាន់គ្រប់ជ្រុងជ្រោយនោះទេ ពីព្រោះអ្នកគ្មានយោបល់មិនដឹងគូសលេខប៉ុន្មាន (គ្មានកន្លែងគូស)។ ឧទាហរណ៍ទី2 អ្នកឆ្លើយសំនួរគឺចេះតែឆ្លើយពីព្រោះគាត់ខ្លាចអ្នកសួរ ឬឆ្លើយដើម្បីឲអ្នកសម្ភាសសប្បាយចិត្ត ឧទាហរណ៍ទី3 ការឆ្លើយដោយសារនិន្នារឈ្នានីស។ល។

5.3.2 បញ្ហាក្រមសីលធម៌

បញ្ហាក្រមសីលធម៌គឺកើតចេញពីកំហុសបួនក្នុងពេលធ្វើការស្ទង់មតិ។

- ការខកខានមិនបានបញ្ចូលប៉ាន់គំរូគ្រប់គ្រាន់ (Coverage error) អាចកើតចេញនៅពេលពីការលំអៀងក្នុងការជ្រើសរើស (selection bias) ហើយក្លាយជាបញ្ហាសីលធម៌ បើសិនជាមានក្រុម ឬបុគ្គលណាមួយត្រូវបានទាត់ចោលចេញពីគម្រោងនៃការជ្រើសរើស ដោយសាតែចង់បានលទ្ធផលគាប់ចិត្តអ្នកឧបត្ថម្ភ។
- កំហុសមិនឆ្លើយតប (nonresponse error) នាំឲ្យមានការលំអៀងមិនឆ្លើយតប (nonresponse bias) ហើយក្លាយជាបញ្ហាសីលធម៌ បើសិនជាមានក្រុម ឬបុគ្គលណាមួយមានទំនងជាមានឱតិចជាងគេក្នុងការឆ្លើយសំណួរ។
- កំហុសនៃប៉ានគំរូក្លាយជាបញ្ហាសីលធម៌ បើសិនជាលទ្ធផលដែលត្រូវបង្ហាញមិនបានយោងទៅនឹងចំនួនប៉ាន់គំរូនិងភាគរយនៃលំអៀង ជាហេតុធ្វើឲអ្នកឧបត្ថម្ភអាចផ្សព្វផ្សាយពត៌មិនត្រឹមត្រូវដែលអាចធ្វើឲមហាជនមានការយល់ច្រឡំ។
- កំហុសនៃកាវាស់វែង (Measurement Error) ក្លាយជាបញ្ហាសីលធម៌តាមរយៈវិធីសាស្ត្របីយ៉ាង៖

1. អ្នកឧបត្ថម្ភជ្រើសរើសសំណួរដែលផ្តល់គំរ៉ូយឲ្យគូសចំឡើយណាមួយ
2. ឥរិយាបថរបស់អ្នកសម្ភាស និងសំលេងរបស់គាត់ មានបំណងបង្កើតឥទ្ធិពលឲ្យគូសចំលើយណាមួយ
3. អ្នកឆ្លើយមានចន្ទ្រះផ្តល់ពត៌មានខុស។

5.4 បំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូ

ជាទូទៅតម្លៃមធ្យមរបស់ប៉ាន់គំរូ (sample mean) ត្រូវបានប្រើប្រាស់យ៉ាងទូលំទូលាយ និងត្រូវបានយកទៅធ្វើការព្យាករណ៍តម្លៃមធ្យមរបស់ចំនួនសកលទៀតផង (mean of population)។ បំណែងចែកប៉ាន់គំរូគឺជាបំណែងចែកនៃតម្លៃទាំងអស់របស់ប៉ាន់គំរូស្ថិតិមួយដែលបានជ្រើសរើសចេញពីចំនួនសកល។

ឧទាហរណ៍ 5.3 ថ្នាក់រៀនកំរិតបណ្ឌិតមួយមាននិស្សិតទាំងអស់ចំនួន 4នាក់ (ចំនួនសកលមាន N=4) A B C និង D ហើយមានអាយុដូចក្នុងតារាង5.3៖

. តារាង5.3 អាយុរបស់និស្សិតទាំង 4នាក់

និស្សិត	X (អាយុគិតជាឆ្នាំ)
A	16
B	18
C	20
D	22

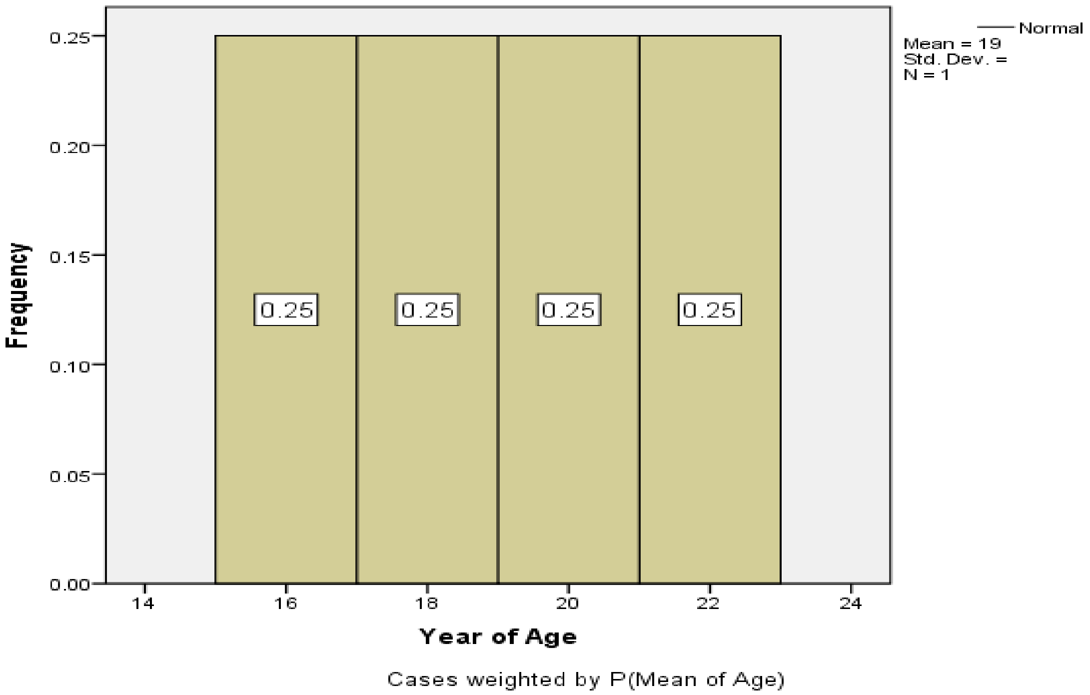
ឧបមាថាយើងធ្វើការសាកសួរចំនួនសកលទាំងមូល (4នាក់) យើបានតម្លៃមធ្យមរបស់ចំនួនសកលគឺស្មើនឹង19 ឆ្នាំនិងគម្លាតស្តង់ដាររបស់ចំនួនសកលគឺស្មើនឹង 2.236 ឆ្នាំ ហើយបំណែងចែកចំនួនសកលដូចបានបង្ហាញក្នុងរូប5.2៖

$$\mu = \frac{\sum X_i}{N} = \frac{(16+18+20+22)}{4} = 19 \text{ ឆ្នាំ}$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{(16-19)^2 + (18-19)^2 + (20-19)^2 + (22-19)^2}{4}} = 2.236 \text{ ឆ្នាំ}$$

រូប5.2 បំណែងចែកចំនួនសកល



តារាង 5.4 ចំនួន16 ប៉ាន់គំរូដែលអាចកើតមាន

1st Obs.	2nd Obs.			
	16	18	20	22
16	16, 16	16, 18	16, 20	16, 22
18	18, 16	18,18	18,20	18,22
20	20, 16	20,18	20,20	20,22
22	22,18	22,18	22,20	22,22

រូប5.2 បញ្ជាក់ថា បើយើងសួរទាំងបួននាក់ ឱកាស ឬប្រូបាប៊ីលីតេរបស់សិស្សម្នាក់ៗដែលត្រូវបានសួរគឺ ¼ ឬ 0.25 ស្មើគ្នា ដូច្នេះបំណែងចែកដើមរបស់ចំនួនសកលនេះមានលក្ខណៈឯកសណ្ឋាន (uniform distribution)។ ឥឡូវ ឧបមាថា យើងធ្វើការសាកសួរតែពីរនាក់ (sample=2) ដោយថៃដន្យ តើបំណែងចែកទៅជាយ៉ាងណា? តារាង5.4 បង្ហាញថា ករណីដែលសាកសួរឬជ្រើសរើសប៉ាន់គំរូពីរនាក់ដោយជំនួសវិញ (replacement) លទ្ធផលអាចមានរហូតដល់ 16 ករណី។

- ករណីទី1៖ យើងអាចសួរចំអ្នកទី1ដែលមានអាយុ16 ហើយអ្នកទី2 ក៏មានអាយុ16 ដូច្នេះយើងបានអាយុមធ្យមគឺ 16 ឆ្នាំ
- ករណីទី2៖ យើងអាចសួរចំអ្នកទី1 មានអាយុ16 អ្នកទី2 មានអាយុ18 ដូច្នេះយើងបានអាយុមធ្យមគឺ 17 ឆ្នាំ

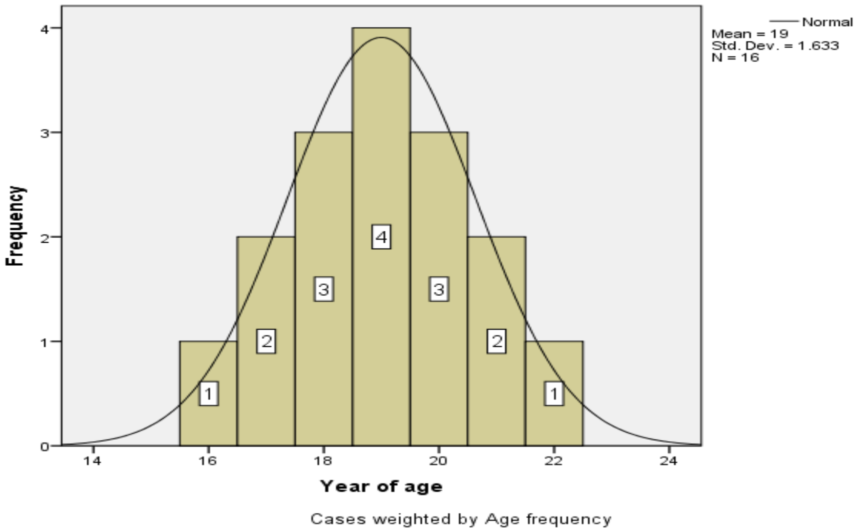
- ករណីទី3៖ យើងអាចសួរចំអ្នកទី1 មានអាយុ16 អ្នកទី2 មានអាយុ20 ដូច្នេះយើងបានអាយុមធ្យមគឺ 18 ឆ្នាំ។ ការសួរបែបនេះជាបន្តបន្ទាប់ រហូតដល់16ករណី យើងបានអាយុមធ្យមដូចក្នុងរាតា 5.4។

តារាង5.4 តម្លៃមធ្យមនៃប៉ាន់គំរូដែលអាចកើតមាន16ករណី

1st Obs.	2nd Obs.			
	16	18	20	22
16	16	17	18	19
18	17	18	19	20
20	18	19	20	21
22	19	20	21	22

តារាង5.4 បង្ហាញអំពីតម្លៃមធ្យមនៃប៉ាន់គំរូដែលអាចកើតមាន16ករណី ហើយបើយើងយកទិន្នន័យទាំងនេះមកបង្ហាញជាគំនូសតាងហ្ស្រេកង់ យើងនឹងបានរូប5.3។

រូប5.3 បំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូ



រូប5.3 បង្ហាញឲឃើញថា បំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូនៃអាយុរបស់និស្សិតពីរនាក់មានលក្ខណៈជាបំណែងចែកធម្មតា(normal distribution) ដូច្នេះយើងអាចសន្និដ្ឋានបានថា៖ ថ្វីបើបំណែងចែកសកល ឬបំណែងចែកដើម របស់វាមានលក្ខណៈឯកសណ្ឋាន (uniform distribution)ក៏ដោយ បំណែងចែករបស់ប៉ាន់គំរូដែលចេញមកពីចំនួនសកលនោះនឹងរំគិលទៅរកបំណែងចែកធម្មតា (approach to normal distribution)។ នេះជាផ្នែកមួយនៃទ្រឹស្តី The Central Limit Theorem។

5.5 ភាពមិនលំអៀងនៃបំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូ

មុននេះយើងបានគណនារួចមកហើយនូវតម្លៃមធ្យមនិងគម្លាតស្តង់ដាររបស់ចំនួនសកល ឥឡូវយើងនឹងធ្វើការគណនាតម្លៃមធ្យមនិងគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូវិញម្តង។

$$\mu_{\bar{X}} = \frac{\sum X_i}{N} = \frac{(16+17+17+18+18+18+\cdots+22)}{16} = 19 \text{ ឆ្នាំ}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum(X_i-\mu)^2}{N}}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{(16-19)^2+(17-19)^2+(7-19)^2+\cdots+(22-19)^2}{16}} = 1.58 \text{ ឆ្នាំ}$$

បន្ទាប់មកយើងធ្វើការប្រៀបធៀបតម្លៃមធ្យមព្រមទាំងគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ ជាមួយនឹងតម្លៃមធ្យមព្រមទាំងគម្លាតស្តង់ដាររបស់ចំនួនសកល ដើម្បីឲ្យកំណត់អំពីកំរិតលំអៀងរបស់ប៉ាន់គំរូ។ ការប្រៀបធៀបតម្លៃមធ្យមនិងគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូគឺត្រូវបានសង្ខេបនៅក្នុងតារាង 5.5។

តារាង 5.5 ប្រៀបធៀបតម្លៃនៃប៉ាន់គំរូនិងចំនួនសកល

ចំនួនសកល	ប៉ាន់គំរូ
μ= 19ឆ្នាំ	$\mu_{\bar{X}} = 19$ ឆ្នាំ
σ=2.236 ឆ្នាំ	$\sigma_{\bar{X}} = 1.58$

តារាង 5.5 ប្រៀបធៀបតម្លៃមធ្យមនិងគម្លាតស្តង់ដាររបស់ចំនួនសកល និងតម្លៃមធ្យមនិងគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ។ តាមរយៈតារាងនេះបញ្ជាក់ឲយើញថា តម្លៃមធ្យមរបស់ប៉ាន់គំរូគឺ19ឆ្នាំស្មើទៅនឹងតម្លៃមធ្យមរបស់ចំនួនសកលគឺ19ឆ្នាំដែរ ហើយគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូវិញគឺ 1.58 ឆ្នាំ តូចជាងគម្លាតស្តង់ដាររបស់ចំនួនសកលដែលមានតម្លៃស្មើនឹង 2.236 ឆ្នាំ គឺខុសគ្នាច្រើនគួរសម។

ក្រៅពីការការគណនាគម្លាតស្តង់របស់ប៉ាន់គំរូដែលបានរៀបរាប់ខាងលើ យើងអាចកំណាត់គម្លាតស្តង់របស់ប៉ាន់គំរូ តាមទំនាក់ទំនងគម្លាតស្តង់ដាររបស់សកល និងចំនួនប៉ាន់គំរូ ដូចក្នុងរូបមន្ត 5.1 ៖

រូបមន្តគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូក្នុងករណីដែលយើងស្គាល់គម្លាតស្តង់ដាររបស់សកល

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \qquad (5.1)$$

ហើយ

σ =គម្លាតស្តង់ដា (standard deviation)

n = ចំនួនប៉ាន់គំរូ (sample size)

រូបមន្ត 5.1 គឺជារូបមន្ត គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ ឬហៅថា លំអៀងស្តង់ដារនៃតម្លៃមធ្យម[11]។

[11] ទំព័រ 333 របស់ Allan G. B (2009) Elementary Statistics: A Step by Step Approach. Eight Edition. McGraw-Hill

ឥឡូវយើងត្រឡប់មកពិនិត្យឧទាហរណ៍ 5.3 របស់យើងឡើងវិញ។ ចំនួនប៉ាន់គំរូ n=2 និងគម្លាតស្តង់ដារចំនួនសកល $\sigma = 2.236$ ដូច្នេះមន្តស្តលំអៀងស្តង់ដារនៃតម្លៃមធ្យម៖

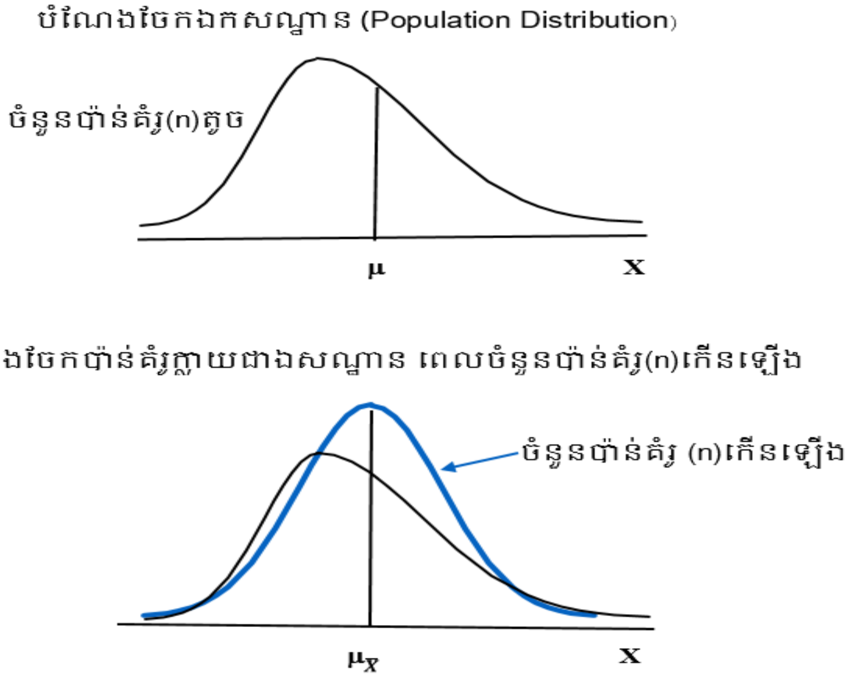
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.236}{\sqrt{2}} = 1.581$$

ដូច្នេះយើងបានតម្លៃលំអៀងស្តង់ដារស្មើនិង 1.581 ដូចតម្លៃគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូនៅក្នុងតារាង 5.5 ខាងលើ។

សំគាល់៖ ជាក់ពិតយើងអាចគណនាគម្លាតស្តង់របស់ប៉ាន់គំរូតារូបមន្តគម្លាតស្តង់ដារធម្មតាក៏បានដែរ ដោយមិនបាច់ប្រើរូបមន្ត លំអៀងស្តង់ដារនៃតម្លៃមធ្យមនោះទេ។ ការបង្កើតរូបមន្តគម្លាតស្តង់របស់ប៉ាន់គំរូ ជាអនុគមន៍នៃគម្លាតស្តង់ដាររបស់សកល និងចំនួនប៉ាន់គំរូគឺដើម្បីទាញលក្ខណៈទូទៅ ឬបង្កើតទ្រឹស្តីទូទៅមួយ ហៅថាទ្រឹស្តីកណ្តាលដែលមានដែនកំណត់ (The Central Limit Theorem,CLT)។ ទ្រឹស្តីនេះបានចែងថា៖ បើសិនជាយើងមានចំនួនសកលដែលមានតម្លៃមធ្យម (μ) និងគម្លាតស្តង់ដារ (σ) ហើយពេលដែលចំនួនប៉ាន់គំរូ (n) កើនឡើងដល់ចំនួនមួយគ្រប់គ្រាន់ ទម្រង់នៃបំណែងចែកតម្លៃមធ្យមរបស់ប៉ាន់គំរូដែលបានប្រមូលចេញពីចំនួនសលដោយមានការចំនួសវិញ នឹងរំកិលទៅរកបំណែចែកធម្មតា (ដូចរូប 5.2 និងរូប 5.3) ហើយបំណែងចែកនេះនឹងមានតម្លៃមធ្យម(μ) នឹងគម្លាតស្តង់ដារ $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ ។

វិបាករបស់ CLT បានបង្ហាញថា ទោះបីជាបំណែងចែកសកលមានលក្ខណៈឯកសណ្ឋានក៏ដោយ (ឬមានលក្ខណៈឆៀង) ក៏អាចផ្តល់នូវបំណែងចែកប៉ាន់គំរូមានលក្ខណៈធម្មតាដែរ បើចំនួនប៉ាន់គំរូ (n) ធំគ្រប់គ្រាន់។ ដូច្នេះ ពេលដែលបំណែងចែកសកលមានលក្ខណៈឯកសណ្ឋាន វិធីដែលធ្វើឲ្យបំណែងចែកប៉ាន់គំរូមានបំណែងចែកធម្មតាគឺត្រូវបង្កើនចំនួនប៉ាន់គំរូ ដូចបានបង្ហាញក្នុងរូប 5.5 ខាងក្រោម។

រូប 5.5 ទំនាក់ទំនងរវាងចំនួនប៉ាន់គំរូនិងបំណែងចែកឯកសណ្ឋាន



5.6 បំណែងចែកប៉ាន់គំរូនៃសមាមាត្រ

សូមរំលឹកថា បំណែងចែកប៉ាន់គំរូនៃសមាមាត្រ (the sampling distribution of proportion) គឺច្រើនមានទំនាក់ទំនងទៅនឹងអថេរប្រភេទចំនួនពីរ (two category variables)។ ឧទាហរណ៍ អតិថិជនចូលចិត្តផលិតផលរបស់អ្នក ឬចូលចិត្តផលិតផលគូរប្រគួតប្រជែង។ អថិតិជនបើកគណនី ឬមិនបើក ទិញឬមិនទិញផលិតផលរបស់អ្នក។

ជាទូទៅ សមាមាត្ររបស់ចំនួនសកលគឺតាងដោយ π សាមាមាត្ររបស់ប៉ាន់គំរូស្ថិតិគឺតាងដោយ p ហើយយើងប្រើ p ដើម្បីព្យាករណ៍ π។

រូបមន្តសមាមាត្ររបស់ប៉ាន់គំរូស្ថិតិកំណត់ដោយៈ

$$p = \frac{X}{n} = \frac{\text{ចំនួនវត្ថុដែលមានចំណាប់អារម្មណ៍}}{\text{ចំនួនប៉ាន់គំរូ}} \qquad (5.2)$$

$$0 \leq p \leq 1$$

និងលំអៀងស្តង់ដារ (standard deviation)របស់សមាមាត្រត្រូវកំណត់ដោយៈ

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \qquad (5.3)$$

សមាមាត្រ p គឺមានបំណែងចែកទ្វេជា (binomial distribution) ប៉ុន្តែ p មានបំណែងចែកប្រហែលបំណែងចែកធម្មតា (normal distribution)នៅពេលដែលចំនួនប៉ាន់គំរូធំ (យើងសន្មត់ថា ការប្រមូលទិន្នន័យមានការជំនួសវិញចំពោះចំនួនប៉ាន់គំរូមានការកំណត់ ឬមិនមានការជំនួសក្នុងករណីប៉ាន់គំរូមិនមានការកំណត់) ជាពិសេស $n\pi \geq 5$ និង$n(1-\pi) \geq 5$ ។ អ្នកអាចយករូបមន្ត z-score (4.14) នៅក្នុងជំពូក4 មកអនុវត្ត។

យើងជំនួស X ដោយ p, ជំនួស μ ដោយ π, និងជំនួស $\sigma/\sqrt{n}$ ដោយ $\sqrt{\pi(1-\pi)/n}$ យើងបានរូបមន្ត z សម្រាប់បំណែងចែកសមាមាត្រដូចខាងក្រោមៈ

$$z = \frac{X-\mu}{\sigma} = \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} \qquad (5.4)$$

ឧទាហរណ៍ 5. 4 អ្នកគ្រប់គ្រងសាខាអេស៊ីលីដាម្នាក់ដឹងហើយថា អ្នកដែលមានគណនីច្រើនលើសពីមួយ ទូទាំងសាខាទាំងមូលគឺមានសមាមាត្រ $\pi = 0.40$ ចូលគណនាប្រូបាប៊ីលីតេនៃការជ្រើសរើសយកចំនួនប៉ាន់គំរូ 200 នាក់ដែលមានគណនេយ្យលើសពីមួយតិចជាង 30%?

ល័ក្ខខ័ណ្ឌបំណែងចែកធម្មតា $n\pi \geq 5 = 200(0.40) = 80 \geq 5$ និង $n(1-\pi) \geq 5 = 0.40(1-0.40) = 120 \geq 5$ ដូច្នេះបំណែងចែកសមាមាត្រនេះគឺមានបំណែងចែកធម្មតា ហើយ តម្លៃ z គឺៈ

$$z = \frac{p-\pi}{\sqrt{\pi(1-\pi)/n}} = \frac{0.30-0.40}{\sqrt{0.40(1-0.40)/200}} = -2.89$$

យោងតាមតារាង E តារាង Normal Curve បានបង្ហាញថាផ្ទៃដែលតូចជាង -2.89 គឺ 0.0019 ឬ 0.19%។ ដូច្នេះ ប្រូបាប៊ីលីតេនៃការជ្រើសរើសយកចំនួនប៉ាន់គំរូ 200 នាក់ដែលមានគណនេយ្យលើសពីមួយតិចជាង 30% គឺមាន 0.19%៕

សេចក្តីសង្ខេប

អ្នកបានដឹងមកហើយថា ចំនួនសកលគឺមានទំហំធំធេងណាស់ ពេលខ្លះកំណត់បាន ពេលខ្លះកំណត់មិនបាន ហើយយើងមិនអាចប្រមូលគ្រប់តែបុគ្គលឬវត្ថុដែលស្ថិតនៅក្នុងចំនួនសកលនោះបានឡើយ។ ផ្ទុយមកវិញគឺយើងអាចប្រមូលតាមរយៈគំរូស្ថិតិដែលតំណាងអោយក្រុមតូច រួចប្រើលទ្ធផលនេះដើម្បីប៉ាន់ស្មានក្រុមធំ (ចំនួនសកល)។

ក្នុងជំពូកនេះអ្នកបានសិក្សាអំពីវិធីសាស្ត្រប្រមូលទិន្នន័យតាមវិធីសាស្ត្រប្រូបាប៊ីលីតេ ដូចជា៖ វិធីសាស្ត្រប្រូបាប៊ីលីតេសាមញ្ញ វិធីសាស្ត្រជាប្រព័ន្ធ វិធីសាស្ត្រកូតា និងវិធីសាស្ត្រជាក្រុម ហើយការជ្រើសរើសវិធីសាស្ត្រនីមួយមកប្រើគឺអាស្រ័យលើស្ថានភាព។ ក្រៅពីនេះអ្នកក៏បានសិក្សាផងដែរទៅលើទំនាក់ទំនងរវាងបំណែងចែកប៉ាន់គំរូនិងបំណែងចែកចំនួនសកល ពេលដែលចំនួនប៉ាន់គំរូកើនឡើងកាន់តែធំ ទំនាក់ទំនងនេះក៏បានលេចឡើងនូវទ្រឹស្តីកណ្តាលដែលមានដែនកំណត់ (CLT)។ តាមរយៈទ្រឹស្តី CLT យើងអាចដឹងថាតម្លៃមធ្យមរបស់ប៉ាន់គំរូគឺមិនលំអៀងទៅនឹងតម្លៃមធ្យមរបស់ចំនួនសកលនោះទេ ហើយតម្លៃមធ្យមរបស់សមាមាត្រក៏មិនលំអៀងទៅនឹងតម្លៃចំនួនសកលរបស់សមាមាត្រនោះដែរ។

.

សំនួរនិងលំហាត់

I. សំណួរ

1. ហេតុអ្វីបានជាតម្លៃមធ្យមរបស់ប៉ាន់គំរូមិនលំអៀងទៅនឹងតម្លៃមធ្យមរបស់ចំនួនសកល ?
2. ហេតុអ្វីបានជាស្តង់ដារអ៊ែរើរបស់តម្លៃមធ្យមថយចុះពេលដែលចំនួនប៉ាន់គំរូ n កើនឡើង ?
3. ហេតុអ្វីបានជាបំណែងចែកប៉ាន់គំរូនៃតម្លៃមធ្យមនឹងមានទម្រង់ជាបំណែងចែកធម្មតាពេលដែលចំនួនប៉ាន់គំរូកើនឡើងគ្រប់គ្រាន់ ទោះបីជាចំនួនសកលមិនមានបំណែងចែកធម្មតាក៏ដោយ ?
4. អ្វីខ្លះជាចំនុចខុសគ្នារវាងបំណែចែករបស់ចំនួនសកលនិងបំណែងចែករបស់ប៉ាន់គំរូ ?
5. តើក្រោមកាលៈទេសៈបែបណា ដែលបំណែងចែករបស់ប៉ាន់គំរូរបស់សមាមាត្រនឹងមានរាងដូចបំណែងចែកធម្មតា ?
6. អ្វីខ្លះជាចំនុចខុសគ្នារវាងការប្រមូលប៉ាន់គំរូតាមបែបប្រូបាប៊ីលីតេនិងការប្រមូលប៉ាន់គំរូបតាមបែបមិនមែនប្រូបាប៊ីលីតេ ?
7. អ្វីខ្លះជាចំនុចខុសគ្នារវាងការប្រមូលប៉ាន់គំរូដោយជំនួសនិងមិនជំនួសវិញ ?
8. អ្វីខ្លះជាចំនុចខុសគ្នារវាងការប្រមូលប៉ាន់គំរូតាមរយៈវិធីសាស្ត្រចៃដន្យធម្មតានិងការប្រមូលប៉ាន់គំរូជាលក្ខណៈប្រព័ន្ធ ?
9. អ្វីខ្លះជាចំនុចខុសគ្នារវាងការប្រមូលប៉ាន់គំរូតាមរយៈវិធីសាស្ត្រចៃដន្យធម្មតានិងការប្រមូលប៉ាន់គំរូជាក្រុម (a stratified sample) ?
10. អ្វីខ្លះជាចំនុចខុសគ្នារវាងការប្រមូលប៉ាន់គំរូតាមរយៈវិធីសាស្ត្រជាក្រុម និងការប្រមូលប៉ាន់គំរូតាមកូតា (a cluster sample) ?

II. លំហាត់

11. រយៈពេលដែលបុគ្គលិកធនាគារចំណាយជាមួយអតិថិជនម្នាក់គិតជាមធ្យមរបស់ចំនួនសកលគឺ μ=3.10 នាទី និងគម្លាតស្តង់ដារ σ = 0.40 នាទី។ ប្រសិនជាអ្នកជ្រើសរើសប៉ាន់គំរូអតិថិជន 16នាក់ដោយចៃដន្យ៖

a. ចូលគណនាប្រូបាប៊ីលីតេនៃរយៈចំណាយគិតជាមធ្យមសម្រាប់អតិថិជនម្នាក់យ៉ាងហោចណាស់ 3 នាទី?

b. ឱកាស 95% ដែលរយៈពេលមធ្យមរបស់ប៉ាន់គំរូតិចជាងប៉ុន្មាននាទី?

c. តើល័ក្ខខ័ណ្ឌ (assumption) អ្វីខ្លះដែលតម្រូវឲមានដើម្បីដោះស្រាយចំលើយ (a) និង (b)?

d. ប្រសិនអ្នកធ្វើការជ្រើសរើសដោយចៃដន្យនូវប៉ាន់គំរូអតិថិជន 64 នាក់ តើឱកាស 95% នៃរយៈពេលជាមធ្យមរបស់ប៉ាន់គំរូតិចជាងប៉ន្មាននាទី?

12.ការជ្រើសរើសប៉ាន់គំរូមួយដោយចៃដន្យទៅលើគ្រួសារចំនួន 50 ដែលទាក់ទងទៅនឹងសំនួរសំខាន់មួយគឺ "តើផ្ទះរបស់អ្នកមានប្រើអ៊ីនធឺណែតឬទេ?" មានផ្ទះចំនួន 20 ឆ្លើយថា "មាន" និង 30 ឆ្លើយថា "មិនមាន"។

a. ចូលគណនាសមាមាត្រនៃចំនួនផ្ទះដែលប្រើអ៊ីនធឺណែត?

b. បើសិនជាសមាមាត្ររបស់ចំនួនសកលស្មើនឹង 0.45 ចូលកំណត់អ៊ែរើស្តង់ដារនៃចំនួនសមាមាត្ររបស់សកល។

13. ការស្ទង់មតិមួយត្រូវបានគេធ្វើឡើង ដើម្បីឲ្យចង់ដឹងអំពីលទ្ធផលនៃការបោះឆ្នោតរបស់បេក្ខភាពពីររូប។ បើសិនជាមានបេក្ខជនមួយទទួលបានសំលេងយ៉ាងតិច 55% នៃសំលេងនៅក្នុងប៉ាន់គំរូ គឺបេក្ខជននោះត្រូវបានព្យាករណ៍ថាជាអ្នកឈ្នះទៅលើការបោះឆ្នោត។ ប្រសិនអ្នកជ្រើសរើសប៉ាន់គំរូដោយចៃដន្យចំនួនអ្នកទៅបោះឆ្នោត 100 នាក់ ចូលគណនាប្រូបាប៊ីលីតេដែលបេក្ខជនមួយនឹងត្រូវបានព្យាករណ៍ថាឈ្នះ៖

a. ពេលដែលភាគរយនៃចំនួនសកលរបស់គាត់ស្មើនឹង 50.1%?

b. ពេលដែលភាគរយនៃចំនួនសកលរបស់គាត់ស្មើនឹង 55%?

c. ពេលដែលភាគរយនៃចំនួនសកលរបស់គាត់ស្មើនឹង 49% (ហើយគាត់ប្រាកដជាចាញ់)?

d. ប្រសិនចំនួនប៉ាន់គំរូកើនឡើងដល់ចំនួន 500, ចូលគណនាសំនួរ (a) ដល់ (c) ឡើងវិញរួចធ្វើការពិភាក្សាទៅលើលទ្ធផលទាំងនោះ?

14. អ្នកផលិតនំប៉័ងម្នាក់បានផលិតនំរបស់ខ្លួនជារៀងរាល់ថ្ងៃ។ ទម្ងន់នំប៉័ងគាត់ដែលបានផលិតគឺមានបំណែងចែកប្រហែលទៅនឹងបំណែងចែកជម្មតា ហើយមានទម្ងន់មធ្យម 50 ក្រាម និងមានគម្លាតស្តង់ដារ 0.40 ក្រាម។ សន្មត់ថា អ្នកធ្វើការជ្រើសរើសចំនួនប៉ាន់គំរូ 30 នំ។

a. ចូលគណនាប្រូបាប៊ីលីតេដែលទម្ងន់មធ្យមរបស់ប៉ាន់គំរូនឹងតិចជាង 4.60 ក្រាម

b. បើប្រូបាប៊ីលីតេស្មើនឹង 70% ដែលទម្ងន់មធ្យមរបស់ប៉ាន់គំរូនឹងច្រើនជាងទម្ងន់មធ្យមរបស់សកលប៉ុន្មាន?

15. សញ្ញាប័ណ្ណប្រភេទ C មួយត្រូវបានគេរាយការណ៍ថា បានទទួលអាត្រាការប្រាក់ចំណេញគិតជាមធ្យមគឺ 57.5%កាលពីឆ្នាំកន្លង។ សន្មត់ថា អាត្រាប្រាក់ចំណេញមានបំណែងចែកជម្មតា ដែលមានអាត្រា

មធ្យមស្មើទៅនឹង 57.5% និងគម្លាតស្តង់ដារស្មើនឹង 20%។ ប្រសិនអ្នកធ្វើការជ្រើសរើសប៉ាន់គំរូដោយចៃដន្យចំនួន

16 ចេញពីចំនួនសកល តើតម្លៃប្រូបាប៊ីលីតេនៃប៉ាន់គំរូដែលមានអាត្រាប្រាក់ចំណេញជាមធ្យមៈ

a. តិចជាង 50%?

b. ចន្លោះពី 40% ទៅ 60%?

c. ច្រើនជាង 40%?

16. កាស្ទង់មតិមួយត្រូវបានធ្វើឡើងកាលពីពេលថ្មីៗនេះ ទៅលើឥតរិយាបទនៃការទិញរថយន្តថ្មីរបស់ពលរដ្ឋកម្ពុជា។ លទ្ធផលបង្ហាញថា 36% នៃប្រជាពលរដ្ឋខ្មែរនឹងទិញរថយន្តថ្មី (ចេញពីរោងចក្រ) ពេលដែលគេចង់ទិញរថយន្ត។ ឧបមាថាអ្នកធ្វើការជ្រើសរើសប៉ាន់គំរូអតិថិជនចំនួន 200 នាក់

a. ចូលគណនាប្រូបាប៊ីលីតេនៃប៉ាន់គំរូដែលនឹងនៅចន្លោះពី 30% និង 40% ជាអតិថិជនដែលចង់ទិញរងយន្តប្រភេទថ្មីចេញពីរោងចក្រ

b. ច្រើនជាង 30%?

ជំពូកទី 6 ការប៉ាន់ស្មានចន្លោះជឿជាក់

(Confidence Interval Estimation)

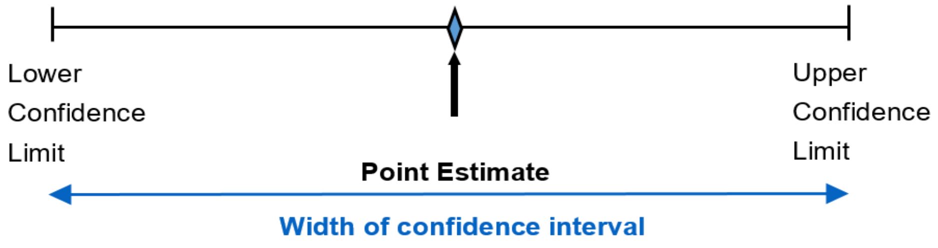
គោលបំណង

- ការប៉ាន់ស្មាននិងបកស្រាយចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមនិងសមាមាត្រ
- កំណត់អំពីចំនួនប៉ាន់គំរូចាំបាច់បង្កើតចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមនិងសាមមាត្រ
- ការប្រើប្រាស់ចន្លោះជឿជាក់ដើម្បីប៉ាន់ស្មានលើករធ្វើសាវនកម្ម
- ការប៉ាន់ស្មានចន្លោះជឿជាក់និងបញ្ហាសីលធម៌

6.1 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលគម្លាតស្តង់ដារដឹង

ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យម គឺត្រូវបានចែកជាពីរណើ៖ គម្លាតស្តង់ដារមិនដឹង និងដឹង។ ជាទូទៅការប៉ាន់ស្មានចន្លោះជឿជាក់របស់ប៉ារ៉ាម៉ែត្រនៃចំនួនសកល (ឧទាហរណ៍ដូចជា តម្លៃមធ្យម ឬសមាមាត្រ) គឺត្រូវផ្អែកលើតម្លៃមធ្យមរបស់ប៉ាន់គំរូ។ ជំហ៊ានទីមួយនេះគេត្រូវធ្វើការប៉ាន់ស្មានចំណុច (point estimation) និងបន្ទាប់មកគេចំណុចដែលបានប៉ាន់ស្មាននេះ ដើម្បីធ្វើការប៉ាន់ស្មានចន្លោះជឿជាក់ នៃចំនួនសកល (confidence interval estimation)។ រូប 6.1 និងតារាង 6.1 បង្ហាញអំពីភាពខុសគ្នារវាងចំនុចនិងចន្លោះជឿជាក់។

រូប6.1គំនូសតាងនៃការប៉ាន់ស្មានចំណុចនិងចន្លោះជឿជាក់



តារាង 6.1 ការប៉ាន់ស្មានចំណុចនិងចន្លោះជឿជាក់

ប៉ាន់គំរូស្ថិតិ (ការប៉ាន់ស្មានជាចំណុច)	ប៉ារ៉ាម៉ែត្រចំនួនសកល (ការប៉ាន់ស្មានចន្លោះជឿជាក់)
$\bar{X}$	$??? \le \mu \le ???$
p	$??? \le \pi \le ???$

រូបមន្តប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យម (គម្លាតស្តង់ដារដឹង)

នៅជំពូកទី 4 យើងមានរូបមន្ត (4.14) $z = \frac{X_i-\mu}{\sigma}$

ហើយ σ= គម្លាតស្តង់ដាររបស់ចំនួនសកល

យើងបានដឹងមកហើយថា $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ ឬ $\sigma = \sigma_{\bar{X}}\sqrt{n}$ បន្ទាប់មកយើងជំនួស តម្លៃ σ ចូលក្នុងរូបមន្ត (4.14) យើងបាន $z = \frac{X_i-\mu}{\sigma_{\bar{X}}\sqrt{n}}$ ⇒ $z(\sigma_{\bar{X}}\sqrt{n})$ = X-μ រួចទាញរកតម្លៃមធ្យមរបស់ចំនួនសកល μ = X - z ($\sigma_{\bar{X}}\sqrt{n}$)។ យើងចង់បានចន្លោះជឿជាក់ដែលមានដែលកំណត់អប្បរមានិងអតិបរមា (lower and upper limit) យើងត្រូវយកតម្លៃ z អវិជ្ជមានដាក់ខាងផ្នែកអប្បរមានិងវិជ្ជមានដាក់ខាងផ្នែកដែនកំណត់អតិបរមា ហើយបើដូច្នោះអាល់ហ្វាក៏ត្រូវចែកជាពីដែរ ហើយចន្លោះជឿជាក់នេះទៀតសោត គឺពឹងផ្អែកលើតម្លៃមធ្យម $\bar{X}$ ។ ជំនួសតម្លៃ z $\bar{X}$និងអាល់ហ្វាដូចការរៀបរាប់ខាងលើ យើងបានរូបមន្រ 6.1។

$$\mu = \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad (6.1)$$

រូបមន្ត 6.1 សម្រាប់ប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលគម្លាតស្តង់ដារដឹង។

ឧទាហរណ៍ 6.1 រោងម៉ាស៊ីនកិនស្រូវមួយបានច្រកអង្ករចំនួន 25បាវ ហើយមានទម្ងន់មធ្យម $\bar{X}$=500.3 ក្រាម និងសន្មត់ថាគម្លាតស្តង់ដារ σ=15 ក្រាម ចូលគណនាចន្លោះជឿជាក់ 95% សម្រាប់ចំនួនមធ្យមរបស់ចំនួនសកលនៃការច្រកអង្ករនេះ?

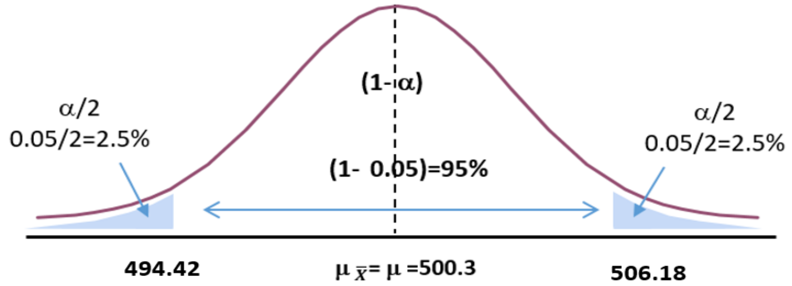
ជំនួសទិន្នន័យទាំងនេះចូលក្នុងរូមន្ត (6.1)

$$500.3 - 1.96\frac{15}{\sqrt{25}} \le \mu \le 500.3 + 1.96\frac{15}{\sqrt{25}}$$

$$494.42 \le \mu \le 506.8$$

បំណកស្រាយះ លទ្ធផលនេះបញ្ជាក់អោយដឹងថា 95% តម្លៃមធ្យមរបស់ចំនួនសកល (បរិមាណអង្ករជាមធ្យម) គឺនៅចន្លោះពី 494.42 ក្រាម ទៅ 506.18 ក្រាម។ នេះមានន័យថា បើយើងធ្វើការជ្រើសរើសប៉ាន់គំរូ ឬធ្វើការថ្លឹងអង្ករចំនួន 100 បាវដែលបានច្រករួច 95% ឬ 95 បាវ ជាមធ្យម គឺនៅចន្លោះពី 494.42 ក្រាម ទៅ 506.18 ក្រាម ហើយអាចមានខុស 5% ឬ 5 បាវ (ក្រោម 494.42 ក្រាម មាន 2.5%2.5បាវ និងលើ506.18 ក្រាម មាន2.5% ឬ2.5បាវ) សូមមើលរូប 6.2 ខាងក្រោម។

រូប 6.2 ចន្លោះជឿជាក់ 95%នៃទម្ងន់មធ្យមរបស់អង្ករ



6.2 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលគម្លាតស្តង់ដារមិនដឹង

ក្នុងជីវភាពជាក់ស្តែង គម្លាតស្តង់ដាររបស់ចំនួនសកលគឺកម្រដឹងណាស់ ដូច្នេះក្នុងករណីដែលគម្លាតស្តង់ដារមិនដឹង យើងនឹងអនុវត្តតាម student's t-distribution ឬហៅកាត់ថាបំណែងចែកអក្សរធី (t-

distribution)។ រូបមន្តនេះត្រូវបានបង្កើតឡើងដោយលោក William S. Gosset នៅសតវត្សទី 20 សម្រាប់គណនាចន្លោះជឿជាក់ពេលដែលគម្លាតស្តង់ដារមិនដឹង ហើយចំនួនប៉ាន់គំរូតូច[12]។

ប្រសិនជាអថេរ X មានបំណែងចែកធម្មតា បន្ទាប់មកតម្លៃ t-statistic:

$$t = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \qquad (6.2)$$

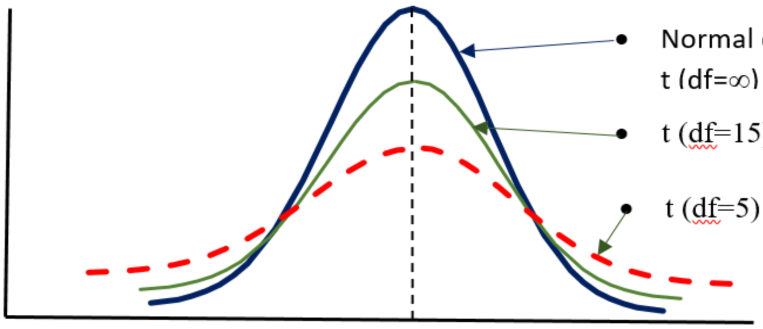
t មានចំនួនដឺគ្រីហ្វ្រីឌឹម (df) = n −1 (n=ចំនួនប៉ាន់គំរូ)

សមីការ (6.2) ស្រដៀងទៅនឹងរូបមន្ត z (សមីការ 6.1) លើកលែងតែតម្លៃគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ S ត្រូវបានគេប្រើដើម្បីប៉ាន់ស្មានគម្លាតស្តង់ដាររបស់ចំនួនសកល σ។

6.2.1 វិបាកនៃបំណែងចែកអក្សរធី

បំណែងចែកអក្សរធី (t) មានរូបរាងដូចបំណែងចែកធម្មតា (Normal Distribution) ។ បំណែកចែកទាំងពីរគឺស៊ីមេទ្រី ហើយមានរាងដូចរូបជួង ហើយមានតម្លៃមធ្យមស្មើនឹងមេដ្យាន។ ក៏ប៉ុន្តែបំណែងចែកអក្សរ (t) មានមានកំពស់ទាបជាង មិនសូវប្រមូលផ្តុំនៅនឹងកណ្តាលដូចបំណែងចែកធម្មតានោះទេ (មើលរូប 6.3) ។ នេះមានន័យថា បំណែងចែកអក្សរធី (t) មានគម្លាតស្តង់ដារធំជាងបំណែងចែកធម្មតា (S > σ) ហេតុដូច្នេះ គេប្រើ S ដើម្បីប៉ាន់ស្មាន σ ក្នុងករណី σ មិនដឹង។ ជាទូទៅបើចំនួនប៉ាន់គំរូចាប់ពី 120 ឡើងទៅ បំណែងចែកអក្សរ t និង z មិនសូវខុសគ្នានោះទេ(មើលរូប 6.3) ។

រូប 6.3 បំណែងចែកធម្មតានិងបំណែងចែកអក្សរ (t) ដែលមាន ចំនួនដឺគ្រីហ្វ្រីឌឹម (df) ផ្សេងៗគ្នា



តារាង 6.2 ប្រៀបធៀប តម្លៃ t និង Z ផ្អែកលើចន្លោះជឿជាក់និងចំនួនដឺគ្រីហ្វ្រីឌឹម (df) មួយចំនួន

Confidence level	$t\ (df=5)$	$t\ (df=15)$	$t\ (df=30)$	$Z\ (df=\infty)$
0.90	1.4398	1.3406	1.6973	1.645
0.95	1.9432	1.7531	2.0423	1.96
0.975	2.4469	2.1315	2.4573	2.24
0.99	3.1427	2.6025	2.750	2.33

តារាង 6.2 បញ្ជាក់ថា ពេលចំនួន df កើនឡើង តម្លៃ t-critical រំកិលទៅរកតម្លៃ Z។

[12] ទាញយកនៅថ្ងៃទី 2/5/2021 ចេញពី https://en.wikipedia.org/wiki/William_Sealy_Gosset

6.2.2 ចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលដែលគម្លាតស្តង់ដារមិនដឹង

សមីការ (6.1) ត្រូវបានប្រើប្រាស់ដើម្បីគណនាចន្លោះជឿជាក់ពេលដែលគម្លាតស្តង់ដារដឹង។ តាមរយៈសមីការ (6.1) យើងអាចទាញបានរូបមន្តចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមពេលដែលគម្លាតស្តង់ដារមិនដឹងដោយប្រើតម្លៃ t-critical ។

$$\mu = \overline{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (6.2)$$

ដែល t មានដឺគ្រីហ្វ្រីឌឹម df=n-1 ហើយចំណែកអថេរដ៏ទៃទៀតគឺមានអត្ថន័យដូចមុន។

ឥឡូវយើងត្រឡប់ទៅពិនិត្យឧទាហរណ៍ 6.1ឡើងវិញ រោងម៉ាស៊ីនកិនស្រូវមួយបានប្រកអង្ករចំនួន n= 25 បាវ ហើយមានទម្ងន់មធ្យម $\bar{X}$=500.3 ក្រាម និងសន្មត់ថាគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ S=15 ក្រាម ចូលគណនាចន្លោះជឿងជាក់ 95%សម្រាប់តម្លៃមធ្យមរបស់ចំនួនសកលនៃការប្រកអង្ករនេះ ?

ចន្លោះជឿជាក់ 95% តម្លៃធីដែលមានdf=25-1=24, $t_{\frac{0.05}{2};\,25-1}$ =2.0639 (មើលតារាង t-distribution) ជំនួសទិន្នន័យទាំងនេះចូលរូបមន្ត 6.2 យើងបាន៖

$$\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$500.3 - 2.0639\frac{15}{\sqrt{25}} \le \mu \le 500.3 + 2.0639\frac{15}{\sqrt{25}}$$

$$494.11 \le \mu \le 506.49$$

បំណកស្រាយ៖ លទ្ធផលនេះបញ្ជាក់ឲ្យដឹងថា 95% តម្លៃមធ្យមរបស់ចំនួនសកល (បរិមាណអង្ករគិតជាមធ្យម) គឺនៅចន្លោះពី 494.11 ក្រាម ទៅ 506.49 ក្រាម។ នេះមានន័យថា បើយើងធ្វើការជ្រើសរើសប៉ាន់គំរូ ឬធ្វើការថ្លឹងអង្ករចំនួន 100 បាវដែលបានប្រករួច 95% ជាមធ្យម គឺនៅចន្លោះពី 494.11 ក្រាម ទៅ 506.49 ក្រាម ហើយអាចមានខុស 5% (ក្រោម494.11 ក្រាម មាន 2.5%និង លើ 506.49 ក្រាម មាន2.5%) ។

6.3 ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់សមាមាត្រ

មុននេះយើងបានសិក្សាលើការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់សម្រាប់ទិន្នន័យជាលេខ (numerical variable) ឥឡូវយើងចូលដល់ ការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់សមាមាត្រ(confidence interval estimate for the proportion)។ នេះគឺជាការប៉ាន់ស្មានសម្រាប់ទិន្នន័យជាប្រភេទ (categorical data)។ ដំបូងយើងត្រូវគណនាការប៉ាន់ស្មានលើចំណុច (point estimation) គឺ p = X/n ហើយ p គឺជាសមាមាត្រដែលយើងចាប់អារម្មណ៍ ឧទាហរណ៍ និស្សិតរៀនជាប់ប៉ុន្មានភាគរយ (យើងចាប់អារម្មណ៍ឬចង់ដឹងលើអ្នកជោគជ័យ) ហើយ p ក៏អាចជាការបរាជ័យដែរ ឲតែយើងមានការចាប់អារម្មណ៍ឬចង់សិក្សា។ បន្ទាប់ពីយើងបានតម្លៃសមាមាត្រ p ហើយ យើងអាចប្រើទិន្នន័យនេះដើម្បីប៉ាន់ស្មានលើចន្លោះជឿជាក់នៃសមាមាត្ររបស់ចំនួនសកល(π)។ ខាងក្រោមនេះគឺជាចន្លោះជឿជាក់នៃចំនួនសកលសមាមាត្រ៖

$$\pi = p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \quad (6.3)$$

ហើយ $p =$ សមាមាត្ររបស់ប៉ាន់គំរូ $= \frac{X}{n} = \frac{\text{ចំនួនវត្ថុដែលចាប់អារម្មណ៍}}{\text{ចំនួនប៉ាន់គំរូ}}$

$\pi =$ សមាមាត្ររបស់ចំនួនសកល

$z_{\alpha/2} =$ តម្លៃ z នៅក្នុងតារាង $Normal$ ដែលយើងកំនត់យកមកប្រើ

ល័ក្ខខ័ណ្ឌតម្រូវ៖ X > 5 ហើយ n-X>5 ដើម្បីគោរពនូវល័ក្ខខ័ណ្ឌ normal distribution។

ឧទាហរណ៍ 6.2 ការសាកសួរនិស្សិត 100នាក់ដោយចៃដន្យបង្ហាញថា 20នាក់ចូលចិត្តរៀនមុខវិជ្ជាពត៌មានវិទ្យា ហើយ 80នាក់ទៀតមិនចូលចិត្ត។ ចូលធ្វើការប៉ាន់ស្មានចន្លោះជឿជាក់ 95% នៃសមាមាត្រពិតប្រាកដរបស់ចំនួនសកលចំពោះនិស្សិតដែលចូលចិត្តរៀនមុខវិជ្ជាពត៌មានវិទ្យា។

- គណនាសមាមាត្រនិស្សិតចូលចិត្តរៀនមុខវិជ្ជាពត៌មានវិទ្យា

$$p = \frac{X}{n} = \frac{20}{100} = 0.20$$

- ផ្ទៀងផ្ទាត់ល័ក្ខខ័ណ្ឌន័រម៉ាល់លីតធី (normality): X=20> 5, 100-20>5 (ផ្ទៀងផ្ទាត់)
- គណនាចន្លោះជឿជាក់ 95% (Z= 1.96)

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$0.20 - 1.96\sqrt{\frac{0.20(1-0.20)}{100}} \le \pi \le 0.20 + 1.96\sqrt{\frac{0.20(1-0.20)}{100}}$$

$$0.1216 \le \pi \le 0.2784$$

បំណកស្រាយ៖ 95% យើងជឿជាក់ថា និស្សិតដែលចូលចិត្តរៀនពត៌មានវិទ្យា គឺស្ថិតនៅចន្លោះ 12.16% ទៅ 27.84%។

6.4 **ការកំណត់ចំនួនប៉ាន់គំរូ**

ការកំណត់ចំនួនប៉ាន់គំរូ (determining the sample size) គឺមាសារៈសំខាន់ណាស់សម្រាប់គណនាចន្លោះជឿជាក់។ ជាទូទៅចំនួនប៉ាន់គំរូគឺរងឥទ្ធិពលដោយសារតែ ចន្លោះជឿជាក់ដែលអ្នកសិក្សាចង់បាន ឧទាហរណ៍អ្នកសិក្សាចង់សិក្សាចន្លោះជឿជាក់ 90%, 95% ឬ 99% និងកំរិតលំអៀងដែលអ្នកសិក្សាសុខចិត្តទទួលយក។

6.4.1 ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់តម្លៃមធ្យម

ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់ធ្វើការប៉ាន់ស្មានចន្លោះជឿជាក់នៃតម្លៃមធ្យម (sample size determination for confidence interval of mean) គឺយើងអាចទាញចេញពីរូបមន្ត (6.1)៖

$$\mu = \overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

តម្លៃដែលបូកបន្ថែម ឬដកចេញពីលើតម្លៃមធ្យម ($\bar{X}$) ត្រូវបានចាត់ទុកជាកំរិតលំអៀង (error) របស់ប៉ាន់គំរូចេញពីតម្លៃមធ្យម។ ដូច្នេះ កំរិតលំអៀង (e) ត្រូវបានកំណត់ដូចតទៅ៖

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

បន្ទាប់មកយើងទាញរក n ដែលតំណាងអោយប៉ាន់គំរូក្នុងការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យម (μ) គឺ៖

$$n = \frac{z_{\alpha/2}^2\sigma^2}{e^2} \qquad (6.4)$$

តាមរយៈរូបមន្ត (6.4) បង្ហាញឲ្យដឹងថា ដើម្បីគណនាចំនួនប៉ាន់គំរូក្នុងការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមគឺយើងត្រូវដឹងជាតុផ្សំបី៖

1. ចន្លោះជឿរជាក់ $(1-\alpha)$ កំណត់ដោយ $z_{\alpha/2}$ដែលបានមកពីតារាងបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ (the Standard Normal Distribution)
2. ការទទួលយកនូវកំរិតលំអៀងរបស់ប៉ាន់គំរូ (e)
3. គម្លាតស្តង់ដាររបស់ចំនួនសកល (σ)

ឧទាហរណ៍ 6.3 អង្គភាពកាំកុងត្រូលចង់ចុះទៅត្រួតពិនិត្យកុងទ័រស្ថានីយ៍សាំងមួយចំនួននៅរាជធានីភ្នំពេញដើម្បីឲដឹងថាស្ថនីយ៍ទាំងនោះចាក់សាំងឲអតិថិជនគ្រប់ចំណុះដែរឬអត់។ បើតាមទិន្នន័យចាក់សាំងកន្លងមកទូទាំងប្រទេសមានគម្លាតស្តង់ដារ 6 មីលីលីត្រ ក្នុងមួយលីត្រ។ ឥឡូវ កាំកុងត្រូលចង់កំណត់ចំនួនប៉ាន់គំរូដើម្បីប៉ាន់ស្មានចន្លោះជឿជាក់លើចំណុះសាំងជាមធ្យមក្នុងមួយលីត្រដែលចាក់ឲអតិថិជនដោយផ្អែកលើកំរិតជឿជាក់ 95% ហើយកាំកុងត្រូលសុខចិត្តទទួលកំរិតលំអៀង 5មីលីលីត្រ ។ តើប៉ាន់គំរូត្រូវប្រើប្រាស់គឺចំនួនប៉ុន្មាន?

$$n = \frac{z_{\alpha/2}^2\sigma^2}{e^2} = \frac{(1.96^2)(6^2)}{5^2} = 5.53 \approx 6$$

ដូច្នេះដើម្បីវាស់ចំណុះឬកំណត់ថាកុងទ័រចាក់សាំងនីមួយបានចាក់សាំងឲអតិថិជនគ្រប់លីត្រដែរឬអត់ កាំកុងត្រូលត្រូវយកប៉ាន់គំរូ (វាស់សាំងម្តងមួយលីត្រ) ចំនួន 6ដងចេញពីកុងទ័រនីមួយៗ

6.4.2 ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់សមាមាត្រ

ស្រដៀងទៅនឹងវិធីសាស្រ្តកំណត់ចំនួនប៉ាន់គំរូសម្រាប់តម្លៃមធ្យម ការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់សមាមាត្រ (sample size determination for proportion) គឺត្រូវដឹងអំពីចំនួនសកល (π)។ សូមរំលឹកថា ការកំណត់ប៉ាន់គំរូក្នុងការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យម (μ)គឺ៖

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើចន្លោះជឿជាក់របស់សមាមាត្រ យើងជំនួស σ ដោយ $\sqrt{\pi(1-\pi)}$ ដូច្នេះយើងបានកំរិតលំអៀងរបស់ប៉ាន់គំរូចេញពីសមាមាត្រដូចខាងក្រោម៖

$$e = z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}$$

បន្ទាប់មកយើងទាញរក n ដែលតំណាងអោយប៉ាន់គំរូក្នុងការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមរបស់សមាមាត្រ (π) គឺរូបមន្ត (6.5)៖

$$n = \frac{z_{\alpha/2}^2\pi(1-\pi)}{e^2} \qquad (6.5)$$

ដូច្នេះដើម្បីគណនាចំនួនប៉ាន់គំរូក្នុងការប៉ាន់ស្មានចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមរបស់សមាមាត្រ គឺយើងត្រូវដឹងកត្តាបី៖

1. ចន្លោះជឿរជាក់ (1-α) កំណត់ដោយ $z_{\alpha/2}$ដែលបានមកពីតារាងបំណែងចែកធម្មតាដែលមានលក្ខណះស្តង់ដារ (the Standard Normal Distribution)
2. ការទទួលយកនូវកំរិតលំអៀងរបស់ប៉ាន់គំរូ (e)
3. សមាមាត្ររបស់ចំនួនសកល (π)

នៅក្នុងការអនុវត្តជាក់ស្តែង តម្លៃ (π) អាចដឹងបានតាមរយៈនៃការពិសោធន៍ឬពត៌មានពីអតីតកាល ប៉ុន្តែបើគ្មានពត៌មានអំពី (π) នោះទេ យើងត្រូវកំណត់ចំនួនប៉ាន់គំរូអោយធំគ្រប់គ្រាន់។ ជាការកត់សំគាល់ពេល $\pi = 0.5$ ពេលនោះតម្លៃ $\pi(1-\pi)$ មានតម្លៃអតិបរមា។

ឥឡូវយើងត្រឡប់ទៅមើលឧទាហរណ៍ 6.2 ឡើងវិញដែលទាក់ទងទៅនឹងការសាកសួរនិស្សិត 100 នាក់ដោយចៃដន្យដែលលទ្ធផលបង្ហាញថា 20% ចូលចិត្តរៀនមុខវិជ្ជាពត៌មានវិទ្យា ហើយ 80%ទៀតមិនចូលចិត្ត។ បន្ទាប់មកយើងធ្វើការគណនាចំនួនប៉ាន់គំរូសម្រាប់សមាមាត្រនៃនិស្សិតដែលចូលចិត្តរៀនមុខវិជ្ជាពត៌មានវិទ្យាស្ថិតនៅចន្លោះជឿជាក់ 95% ដោយសន្មត់ថា យើងសុខចិត្តទួលយកភាពលំអៀងទៅលើការប៉ាន់ស្មាននេះ 10% ហើយយើងគ្មានពត៌មានទាក់ទងទៅនឹងចំនួនសកល (π) ឡើយ។

$$n = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{e^2} = \frac{(1.96^2) * 0.5(1-0.5)}{0.10^2} = 96.04$$

ដូច្នេះចំនួនប៉ាន់គំរូសម្រាប់ការសាកសួរនេះគឺ96នាក់។

សំគាល់៖ រូបមន្ត (6.3) និង (6.4) អាចប្រើសម្រាប់កំណត់ចំនួនប៉ាន់គំរូដើម្បីយកទៅប៉ាន់ស្មានចន្លោះជឿជាក់នៃតម្លៃមធ្យមនិងសមាមាត្រតែប៉ុណ្ណោះ។ រូបមន្តទាំងនេះមិនអាចយកទៅគណនាចំនួនប៉ាន់គំរូសម្រាប់សមីការលីនេអ៊ែរដែលមានច្រើនអថេរ (multiple regression) confirmatory factor analysis (CFA), structural equation modelling (SEM), exploratory factor analysis (EFA) និង principal components factor analysis (PCFA) នោះទេ។ បើអ្នកអានមានការចាប់អារម្មណ៍ទៅលើការកំណត់ចំនួនប៉ាន់គំរូសម្រាប់ម៉ូដែលទាំងនេះ សូមអានសៀងភៅរបស់អ្នកនិពន្ធ Tabachnick & Fidell (2007); Meyers, Gamst and Guarino (2013) និងអត្ថបទរបស់ Brooks G. P & Barcikowsk. R. S (2012)។

6.5 ការអនុវត្តន៍ការប៉ាន់ស្មានចន្លោះជឿជាក់ទៅលើសាវនកម្ម និងការត្រួតពិនិត្យគុណភាព

ជាទូទៅការងារសាវនកម្មគឺជាការងារប្រមូលភស្តុតាងដើម្បីធ្វើការវាយតម្លៃទៅលើការអនុវត្តន៍នាពេលកន្លងៗ ការងារនេះកម្រធ្វើការពិនិត្យលើពត៌មានទាំងអស់របស់ចំនួនសកលណាស់ គឺគេពឹងផ្អែកលើការប៉ាន់ស្មាននូវប្រូបាបីលីតេរបស់ប៉ាន់គំរូ ពីព្រោះ មានហេតុផលសំខាន់ដូចខាងក្រោម៖

- ចំនេញពេលវេលានិងថវិការ
- ផ្តល់លទ្ធផលអោយយើងអាចការពារខ្លួនបានចំពោះមុខច្បាប់
- អាចឲយើងដឹងនូវកំរិតលំអៀង
- អាចឲយើងសន្និដ្ឋានចំនួនសកលបានត្រឹមត្រូវ
- អាចឲសាវនករចងក្រងជាមួយឯកសាររបស់បុគ្គលផ្សេងទៀតដើម្បីធ្វើការវាយតម្លៃជាសមូហភាព

6.5.1 ការប៉ាន់ស្មានលើបរិមាណសរុបនៃចំនួនសកល

ការងារសវនកម្មគឺចាប់អារម្មណ៍លើការវាយតម្លៃទៅលើបរិមាណសរុបរបស់ចំនួនសកល (total amount of population) មិនមែនប៉ានស្មានលើចំនួនមធ្យមនោះទេ។ ហើយការងារប៉ាន់ស្មានលើចំនួនសរុបនៃចំនួនសកលគឺមានពីរជំហានៈ

ជំហានដំបូងយើងត្រូវធ្វើការប៉ាន់ស្មានលើចំនុចសរុបនៃចំនួនសកល (the point estimate for the population total) ជាមុនសិន

$$Total = N * \bar{X} \qquad (6.6)$$

ហើយ N = ចំនួនសរុបនៃ ចំនួនសកល (population size)

$\bar{X}$ =តម្លៃមធ្យម

ជំហានបន្ទាប់គឺធ្វើការប៉ាន់ស្មានចន្លោះជឿជាក់ចំពោះចំនួនសកលសរុប (confidence interval estimation for the total)

$$N * \bar{X} \pm N(t_{\frac{\alpha}{2}}) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad (6.7)$$

ហើយ $t_{\frac{\alpha}{2}}$ =តម្លៃ t នៅក្នុងតារាង t-distribution ដែលមានតម្លៃ $\alpha/2$ និងមានដឺក្រេហ្វ្រីឌឹម df=n-1

$\bar{X}$ = តម្លៃមធ្យមរបស់ប៉ាន់គំរូ

N=ចំនួនសកលសរុប

n = ចំនួនប៉ាន់គំរូ

S= គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ

ឧទាហរណ៍ 6.4 សវនករម្នាក់បានធ្វើសវនកម្មទៅលើក្រុមហ៊ុនមួយដែលមានប័ណ្ណចំណាយចំនួន 2000 ប័ណ្ណ ហើយគាត់ចង់ធ្វើការប៉ាន់ស្មានចំនួនសកលសរុបនៃប័ណ្ណចំណាយទាំងនោះតាមរយៈការជ្រើសរើសប័ណ្ណចំនួន 100 ហើយបានលទ្ធផលដូចតទៅៈ

តម្លៃមធ្យមរបស់ប័ណ្ណ $\bar{X} = \$1000$

គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ S=$250

ចូលបង្កើតចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើតម្លៃរបស់ចំនួនសកលសរុប? (សំគាល់៖ តារាង t-critical, 95%, $t_{\frac{0.05}{2},100-1}$=1.9842)

- ជំហានដំបូងគណនាចំនុចប៉ាន់ស្មាននៃចំនួនសកលសរុប (point estimate of the population total) ដោយប្រើរូមន្ត (6.6)

$$Total = N * \bar{X} = 2{,}000 * 1000 = \$2{,}000{,}000$$

- ជំហានបន្ទាប់គឺប្រើរូបមន្ត (6.7) ដើម្បីគណនាចន្លោះជឿជាក់នៃតម្លៃសរុបនៃចំនួនសកល

$N * \bar{X} \pm N(t_{\frac{\alpha}{2}}) \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

=2,000*1,000± 2,000(1.9842)$\frac{250}{\sqrt{100}} \sqrt{\frac{2{,}000-100}{2{,}000-1}}$

=2,000,000 ± 96,722.13

$1,903,277.87 ≤ Population total ≤ $2,096,722.13

ដូច្នេះនៅចន្លោះជឿជាក់ 95% សវនករបានប៉ាន់ស្មានថា តម្លៃសរុបរបស់ប័ណ្ណទាំង 2,000 គឺនៅចន្លោះពី $1,903,277.87 ដល់ $2,096,722.13។ បើសិនជារបាយការណ៍របស់ក្រុមហ៊ុនមិនបានបង្ហាញនូវតម្លៃសរុបទាំង 2,000 ប័ណ្ណនៅចន្លោះនេះទេ សវនករអាចមានភស្តុតាងដើម្បីធ្វើការចោទប្រកាន់បាន។

6.5.2 ការប៉ាន់ស្មានលើភាពខុសគ្នា

ការប៉ាន់ស្មានលើភាពខុសគ្នា (difference estimation) គឺត្រូវបានគេប្រើប្រាស់ពេលដែលសាវនករ ឬអ្នកត្រួតពិនិត្យគុណភាពយល់ឃើញថា មានភាពលំអៀងកើតមានចំពោះការប៉ាន់ស្មានធៀបទៅនិងទិន្នន័យជាក់ស្តែង ហើយគាត់ចង់ដឹងអំពីទំហំនៃភាពលំអៀងនោះដោយផ្អែកប៉ាន់គំរូ។

ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើភាពខុសគ្នា យើងត្រូវធ្វើតាមជំហ៊ានដូចខាងក្រោម[13]៖

1. កំណត់ចំនួនប៉ាន់គំរូដែលត្រូវការ
2. គណនាភាពលំអៀងរវាងតម្លៃដែលបានធ្វើសវនកម្មនិងតម្លៃក្នុងឯកសារ តម្លៃខុសគ្នាទី (i) យើងតាងដោយ D_i បើ $D_i = 0$ មានន័យថាតម្លៃដែលធ្វើសវនកម្មស្មើនឹងតម្លៃក្នុងឯសារដើម គឺត្រឹមត្រូវ ហើយវិជ្ជមានពេលតម្លៃដែលធ្វើសវនកម្មធំជាងតម្លៃឯសារដើម និងអវិជ្ជមានតម្លៃដែលធ្វើសវនកម្មតូចជាងតម្លៃឯសារដើម
3. គណនាតម្លៃមធ្យមនៃភាពខុសគ្នា ($\bar{D}$) ដោយចែកតម្លៃខុសគ្នាសរុបទៅនិងចំនួនប៉ាន់គំរូ ដូចសមីការ (6.8)៖

   តម្លៃខុសគ្នាជាមធ្យម: $\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n}$ (6.8)

   ហើយ D_i= តម្លៃលំអៀងនិមួយៗ
4. គណនាគម្លាតស្តង់ដារនៃតម្លៃខុសគ្នា S_D

   $$S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}} \quad (6.9)$$
5. ប្រើសមីការ (6.10) ដើម្បីធ្វើការប៉ាន់ស្មានលើចន្លោះជឿជាក់នៃតម្លៃខុសគ្នាសរុបរបស់ចំនួនសកល។

   $$N * \bar{D} \pm N(t_{\frac{\alpha}{2}}) \frac{S_D}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (6.10)$$

   ហើយ $t_{\frac{\alpha}{2}}$ =តម្លៃ t នៅក្នុងតារាង t-distribution ដែលមានតម្លៃ $\alpha/2$ និងមាន df=n-1

ឧទាហរណ៍ 6.5 ជារៀងរាល់សប្តាហ៍ស្ថានីយ៍លក់សាំង SOPHIA បានធ្វើការចាក់សាំង (លក់) 1,000 លើកទៅរថយន្ត (សម្រាប់តែកុងទ័រចាក់សាំងលេខ១តែប៉ុណ្ណោះ) ហើយអ្នកលក់មានការបារម្ភណ៍ចំពោះកុងទ័រចាក់សាំងមួយនេះ គាត់ក៏ធ្វើការប្រមូលប៉ាន់គំរូចំនួន 50 លើក ហើយសំគាល់ឃើញថាមានកំហុស 14លើក (អវិជ្ជមាន មានន័យថាខ្វះលីត្រ វិជ្ជមាន មានន័យថាលើសលីត្រ)។ ហើយទិន្នន័យត្រូវបានកត់ត្រាដូចខាងក្រោម៖

2	3	3.3	2.2	2.3	-1	-1.2	3	2.4	0.8	0.6	1.5	-0.8	0.9

[13] ដកស្រង់ចេញពី Berenson M. L, Levine D. M, & Krehbiel T. C. (2012). ទំព័រ 305-306

ចូលបង្កើតចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើបរិមាណលំអៀងសរុបចំនួន 1,000 លើក?

ចំឡើយ:

- គណនាតម្លៃមធ្យមខុសគ្នា $\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{(2+3+\cdots+0.9)}{50} = \frac{19}{50} = 0.38$
- គណនាគម្លាតស្តង់ដារ $S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}} = \sqrt{\frac{(2-0.38)^2+\cdots+(0.9-0.38)^2}{50-1}} = \frac{43.5016}{49} = 0.8878$
- គណនាចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើបរិមាណលំអៀងសរុបចំនួន 1,000 លើក (សំគាល់៖ តារាង t-critical, 95%, $t_{\frac{0.05}{2},50-1}$=2.0096)

$N * \bar{D} \pm N(t_{\frac{\alpha}{2}})\frac{S_D}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$=1,000 * 0.38 ± 1,000(2.0096) $\frac{0.8878}{\sqrt{50}}\sqrt{\frac{1,000-50}{1,000-1}}$

380 ± 246.047

133.953 ≤ total difference ≤ 626.047

ដូច្នេះនៅចន្លោះជឿជាក់ 95% កុងទ័រចាក់សាំងលេខ1នៅស្ថានីយ៍ SOPHIA ចាក់សរុប 1,000 លីត្រ (ចំនួនសកល) លំអៀងសរុបគឺនៅចន្លោះ 133.953 លីត្រ ដល់ 626.047 លីត្រ។ ចន្លោះនេះគឺវិជ្ជមានទាំងសង្ខាង មានន័យថាចាក់សាំងលើសលីត្រ។

6.5.3 ការប៉ាន់ស្មានចន្លោះជឿជាក់កន្ទុយម្ខាង

ការមិនគោរពតាមស្តង់ដារឬបទបញ្ញតិរបស់កាំកុងត្រូលគឺត្រូវប្រើចន្លោះជឿជាក់កន្ទុយម្ខាង(one-side confidence interval estimation)។ ជួនកាលប្រើកន្ទុយខាងឆ្វេង ជួនកាលខាងស្តាំ ឧទាហរណ៍ ម្ចាស់ស្ថានីយ៍បារម្មណ៍ខ្លាចលើលីត្រ រីឯកាំក្រុងត្រូលបារម្មណ៍ខ្លាចខ្វះលីត្រ។

សមីការ (6.11) កំណត់អំពីចន្លោះជឿជាក់កន្ទុយម្ខាង (ស្តាំ) សម្រាប់សមាមាត្ររបស់ម្ចាស់ស្ថានីយ៍ ដែលខ្លាចលើសលីត្រ៖

$$Upper\ bound = p + z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}} \qquad (6.8)$$

ឧទាហរណ៍ 6.6 ឥឡូវយើងពិនិត្យឡើងវិញនូវ ឧទាហរណ៍ 6.5 ឧបមារថា មានបទបញ្ជាផ្ទៃក្នុងរបស់ស្ថនីយ៍បានកំណត់ថា អត្រាអនុគ្រោះចំពោះការចាក់សាំងលើសលីត្រមិនឲហួស 10% (tolerable rate)។ ក្នុងករណីដែលលើស ម្ចាស់ស្ថានីយ៍មានចំណាត់ការជួសជុលកុងទ័រ។ បើសិនជាចន្លោះជឿជាក់ 95%ត្រូវបានយកមកប្រើប្រាស់ តើអ្នកសន្និដ្ឋានយ៉ាងដូចម្តេចចំពោះស្ថានីយ៍មួយនេះ?

ចំឡើយ:

ការចាក់សាំងនៅក្នុងឧទាហរណ៍ 6.5 មានកំហុស 14ដង ក្នុងប៉ាន់គំរូ n = 50 ក្នុងនោះ 12ដងលើសចំណុះ និង2ដងខ្វះលីត្រ។ ឥឡូវយើងចាប់អារម្មណ៍ទៅលើការចាក់សាំងលើសលីត្រ (កន្ទុយខាងស្តាំ) ដូច្នេះកំហុសគិតជាសមាមាត្រ p= 12/50 = 0.24 ឬ 24% ហើយបើ α=0.05 យើងបានតម្លៃគោល $z_\alpha = 1.645$

$$Upper\ bound = p + z_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}} = 0.24 + 1.645\sqrt{\frac{0.24(1-0.24)}{50}} * \sqrt{\frac{1,000-50}{1,000-1}}$$

$Upper\ bound$ =0.24 + 1.645*0.06* 0.9752 = 0.337

ដូច្នេះចន្លោះជឿជាក់ 95% យើងអាចសន្និដ្ឋានថា អត្រាចាក់សាំងលើសគឺ 33.7%ហើយកំរិតនេះបានលើសកំរិតដែលបទបញ្ជាផ្ទៃក្នុងដែលបានអនុញ្ញាត។ សូមផ្នែកបច្ចេកទេសពិនិត្យកុងទ័រចាក់សាំងលេខ1នេះឡើងវិញ។

6.6 ការប៉ាន់ស្មានចន្លោះជឿជាក់និងបញ្ហាសីលធម៌

បញ្ហាសីលធម៌គឺទាក់ទងដល់ការប៉ាន់ស្មានលើចំណុច (point estimation) ប៉ាន់គំរូ (sample size) និងចន្លោះជឿជាក់ (confidence interval)។ បើយើងខកខានមិនបានរាយការណ៍អំពីចន្លោះជឿជាក់ធ្វើឲ្យអ្នកអានយកចិត្តទុកដាក់តែការប៉ាន់ស្មានលើចំណុចឬចាប់អារម្មណ៍តែតម្លៃមធ្យមតែប៉ុណ្ណោះ ដែលជាហេតុធ្វើឲគេខ្វះការប្រុងប្រយ័ត្នទៅលើបំណែងចែកចំនួនសកល។

សេចក្ដីសង្ខេប

ជំពូកនេះបានពីភាគសារអំពីការប៉ាន់ស្មានលើចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យម និងសមាមាត្ររបស់ចំនួនសកល ព្រមទាំងចំនួនប៉ាន់គំរូដែលចាំបាច់សម្រាប់ការប៉ាន់ស្មាននោះ។ ករណីគម្លាតស្តង់ដារដឹងអ្នកប្រើ Z ហើយករណីមិនដឹងអ្នកប្រើបំណែងចែកអក្សរធី។ លើពីនេះអ្នកក៏បានសិក្សាផងដែរអំពីការប៉ាន់ស្មានទៅលើចន្លោះជឿជាក់ដែលអនុវត្តទៅលើបញ្ហាសវនកម្មនិងការត្រួតពិនិត្យគុណភាព។

សំនួរនិងលំហាត់

I. **សំនួរ**

1. ហេតុអ្វីបានអ្នកមិនដែលឮគេប្រើប្រាស់ចន្លោះជឿជាក់ 100% ដើម្បីប៉ាន់ស្មានចំនួនសកលដែលគេចាប់អារម្មណ៍។
2. តើពេលណាទើបអ្នកអាចប្រើបំណែងចែកអក្សរធី (t) ដើម្បីបង្កើតចន្លោះជឿជាក់ធ្វើការប៉ាន់ស្មានតម្លៃមធ្យមរបស់ចំនួនសកល?
3. ស្ថានភាពបែបណាទើបអ្នកប្រើចន្លោះជឿជាក់តែម្ខាងជំនួសចន្លោះជឿជាក់ទាំងសង្ខាង?
4. ពេលណាអ្នកចង់ធ្វើការប៉ាន់ស្មានទៅលើចំនួនសកលសរុប ជំនួសតម្លៃមធ្យមរបស់ចំនួនសកល?
5. តើការប៉ាន់ស្មានលើតម្លៃលំអៀងខុសពីការប៉ាន់ស្មានលើតម្លៃមធ្យមយ៉ាងដូចម្តេច?

II. **លំហាត់**

6. ខាងក្រោមនេះគឺជាចំណាយទៅលើអាហារពេលថ្ងៃត្រង់របស់បុគ្គលិកធនាគារចំនួន 10 នាក់ (គិតជាដុល្លា)

តម្លៃ	7.38	8.46	5.03	6.45	8.47	4.20	6.50	5.86	7.54	9.87

a. ចូលបង្កើតចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានចំនួនមធ្យមរបស់ចំនួនសកលនៃបរិមាណចំណាយសម្រាប់អាហារណ៍ថ្ងៃត្រង់ ដោយសន្មត់ថា បំណែងចែកមានលក្ខណៈធម្មតា

b. ចូលធ្វើការបកស្រាយចន្លោះជឿជាក់នៅក្នុងសំនួរ (a)

8.បើអ្នកចង់បានចន្លោះជឿជាក់ 95%ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើតម្លៃមធ្យមរបស់ចំនួនសកល ដែលស្ថិតនៅកំរិតអែររើ ± 5 និងគម្លាតស្តង់ដារស្មើ 15 ចូលគណនាចំនួនប៉ាន់គំរូដែលត្រូវការក្នុងការប៉ាន់ស្មាននេះ។

9. ការស្ទង់មតិមួយត្រូវបានធ្វើឡើងដើម្បីចង់ដឹងអំពីចំណាយទៅលើសេវាសុខភាពរបស់គ្រួសារនីមួយនៅក្រុងភ្នំពេញ។ អ្នកធ្វើការស្ទង់មតិចង់ដឹងនៅចន្លោះជឿជាក់ 95% ទៅលើការចំណាយនេះ។ ការសិក្សាពីមុនមកបញ្ជាក់ថា គម្លាតស្តង់ដារគឺប្រហែល 200 ដុល្លា។ បើអ្នកធ្វើការស្ទង់មតិចង់ឲ្យកំរិតអែររើ ±25 ដុល្លា តើចំនួនប៉ាន់គំរូដែលត្រូវការគឺប៉ុន្មានសម្រាប់ការសិក្សានេះ?

10. អ្នកស្រាវជ្រាវម្នាក់ចង់ប៉ាន់ស្មានលើចំណាយជាមធ្យមនៃថ្លៃអាគ្គិសនីប្រចាំខែសម្រាប់គ្រួសារនីមួយៗនៅខណ្ឌមួយរបស់រាជធានីភ្នំពេញ។ កន្លងមកធ្លាប់មានការសិក្សានៅខណ្ឌដ៏ទៃ ដែលមានគម្លាតស្រង់ដារ 25 ដុល្លា។ គាត់ចង់ធ្វើការប៉ាន់ស្មានក្នុងចន្លោះជឿជាក់ 99%ទៅលើចំណាយជាមធ្យម ហើយនៅចន្លោះអែររើ ±5 ដុល្លា។

a. តើចំនួនប៉ាន់គំរូដែលត្រូវការគឺប៉ុន្មានសម្រាប់ការសិក្សានេះ?

b. បើចន្លោះជឿជាក់ 95% ត្រូវបានសំរេចយក តើចំនួនប៉ាន់គំរូដែលត្រូវការគឺប៉ុន្មានសម្រាប់ការសិក្សានេះ?

11. ក្រុមហ៊ុនមួយមានបុគ្គលិកចំនួន 5000 នាក់ ចង់ធ្វើការប៉ាន់ស្មានទៅលើចំណាយប្រចាំឆ្នាំសម្រាប់សេវាសុខភាពរបស់គ្រួសារបុគ្គលិកនីមួយៗ ដើម្បីឲក្រុមហ៊ុនងាយស្រួលងាយធ្វើគម្រោងជានារាប់រងលើសេវាសុខភាពរបស់បុគ្គលិករបស់ក្រុមហ៊ុន។ បុគ្គលិកចំនួន 10 នាក់ត្រូវបានសាកសួរដោយចៃដន្យ បានបង្ហាញចំណាយលើសេវាសុខភាពសម្រាប់ឆ្នាំកន្លងមកដូចតទៅ៖

ចំណាយ	120	365	246	155	500	208	173	425	316	180

ចូលបង្កើតចន្លោះជឿជាក់ 90%ដើម្បីធ្វើការប៉ាន់ស្មានទៅលើចំណាយទៅលើសេវាសុខភាពសរុបរបស់បុគ្គលិកទាំងអស់កាលពីឆ្នាំមុន។

12. សវនករម្នាក់បានធ្វើការត្រួតពិនិត្យទៅលើវិក័យប័ត្ររបស់ក្រុមហ៊ុនមួយ ដោយយកប៉ាន់គំរូចំនួន 160 ចេញពីវិក័យប័ត្រសរុបចំនួន 5,000 ហើយបានរកឃើញថា មានការកត់ត្រាលំអៀង (លើស) ចំនួន 14 វិក័យប័ត្រ ហើយបរិមាណដែលលើសត្រូវបានកត់ត្រាជាដុល្លាដូចខាងក្រោម៖

Amount	28.60	20.32	97.36	200.63	104.18	84.92	98.76
	76.10	26.50	130.43	88.32	47.81	89.01	90.50

ចូលធ្វើការប៉ាន់ស្មានបរិមាណដុល្លាសរុបដែលបានកត់ត្រាលើសនៅចន្លោះជឿជាក់ 99% ។

13. មន្ត្រីសវកម្មមួយក្រុមបានធ្វើការត្រួតពិនិត្យទៅលើទៅលើវិក័យប័ត្រចំណាយរបស់ក្រុមហ៊ុនមួយ ដោយធ្វើការជ្រើសរើសប៉ាន់គំរូចំនួន 200 ចេញពីចំនួនសកល 10,000 វិក័យប័ត្រ។ តាមស្តង់ដារអនុគ្រោះរបស់សវនកម្មគឺ 0.05។ ក្រុមសវនករ បានរកឃើញថា មានវិក័យប័ត្រចំនួន 15 បានបំពានលើស្តង់ដារដែលបានកំណត់។

a. ចូលគណានចន្លោះជឿជាក់ 95% ផ្នែកខាងស្តាំ (the upper bound) ដើម្បីធ្វើការប៉ាន់ស្មាននូវអាត្រាមិនគោរពតាមស្តង់ដារអនុគ្រោះ។

b. ដោយផ្អែកលើចំលើយ (a) តើក្រុមសវនករគួរសន្និដ្ឋានយ៉ាងដូចម្តេច?

14. ការស្ទង់មតិមួយបានធ្វើទៅលើអ្នកប្រើប្រព័ន្ធផ្ទេរប្រាក់ចំនួន 1,000នាក់ ដោយរកឃើញថា 70% បានធ្វើការផ្ទេរតាមរយៈភ្នាក់ងារផ្ទេរប្រាក់ និង 50% បានផ្ទេរតាមរយៈទូរស័ព្ទ។

a. ចូលប្រើចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានលើសមាមាត្ររបស់ចំនួនសកលរបស់អ្នកផ្ទេរប្រាក់តាមភ្នាក់ងារ។

b. ចូលប្រើចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានលើសមាមាត្ររបស់ចំនួនសកលរបស់អ្នកផ្ទេរប្រាក់តាមទូរស័ព្ទ។.

c. ចូលសរសេរលទ្ធផលសង្ខេបនៃចំលើយ (a) និង (b)។

ជំពូកទី 7 ការធ្វើតេស្តលើសម្មតិកម្មសម្រាប់ប៉ាន់គំរូតែមួយ

(One-sample Hypothesis Testing)

គោលបំណង

- គោលការណ៍គ្រឹះនៃការធ្វើតេស្តសម្មតិកម្មនិងនីតិវិធី

- ការធ្វើតេស្តលើតម្លៃមធ្យម និងសមាមាត្រ
- ល័ក្ខខ័ណ្ឌចំពោះនីតិវិធីនៃកាតធ្វើតេស្តនីមួយៗ និងរបៀបវាយតម្លៃលើល័ក្ខខ័ណ្ឌ
- ការចៀសវាងកំហុសក្នុងការធ្វើតេស្តសម្មតិកម្ម

7.1 **នីតិវិធីសំខាន់នៃការធ្វើតេស្តសម្មតិកម្ម**

ការធ្វើតេស្តសម្មតិកម្មគឺជានីតិវិធីមួយដែលពឹងផ្អែកលើភស្តុតាងចេញពីប៉ាន់គំរូនិងទ្រឹស្តីនៃប្រូបាប៊ីលីតេ ដើម្បីកំណត់ថា ការអះអាងអំពីប៉ារ៉ាម៉ែត្រណាមួយរបស់ចំនួនសកលគួរតែបដិសេធ ឬមិនបដិសេធ។ ឧទាហរណ៍ ចៅសង្កាត់ម្នាក់បានអះអាងថា "គ្រួសារនីមួយៗក្នុងឃុំរបស់គាត់ មានម៉ូតូពីរគ្រឿងជាមធ្យមក្នុងមួយគ្រួសារ" នេះគឺជាសម្មតិកម្ម។

ការធ្វើតេស្តលើសម្មតិកម្មគឺអាចត្រូវបានបែងចែកប្រាំជំហានដូចតទៅៈ

1. កំណត់សម្មតិកម្មមោឃៈនិង សម្មតិកម្មជំនួស

ជំហានដំបូងនៃការធ្វើតេស្តគឺយើងត្រូវកំណត់អត្តសញ្ញាណសម្មតិកម្មដែលអះអាងថា ប៉ារ៉ាម៉ែត្រស្មើទៅនឹងតម្លៃណាមួយ សម្មតិកម្មនេះហៅថាសម្មតិកម្មមោឃៈ ឬសូន្យ (null hypothesis) ដែលសរសេរកាត់ H_0។ វាតំណាងឲស្ថានភាពដែលកំពុងកើតមាន (status quo) ហើយគេអាចសរសេរជាប្រយោគ ឬជាលេខក៏បាន។ ឧទាហរណ៍ (1) H_0: មនុស្សគ្រប់រូបសន្មត់ថាគ្មានទោស, (2) H_0: $\mu = 2$ (ម៉ូតូពីរគ្រឿង)។ បន្ទាប់ពីសម្មតិកម្មមោឃៈត្រូវបានបញ្ជាក់ គេត្រូវសរសេរសម្មតិកម្មជំរើស ឬផ្លាស់ (alternative hypothesis, H_1) ដែលជាសម្មតិកម្មផ្ទុយពីសម្មតិកម្មមោឃៈ ឧទាហរណ៍ $H_A \neq 2$ (ម៉ូតូពីរគ្រឿង)។

ជាទូទៅសម្មតិកម្មមោឃៈមានបីប្រភេទៈ

សម្មតិកម្មមោឃៈ	សម្មតិកម្មជំរើសឬផ្លាស់
H_0: $\mu = 2$ (ម៉ូតូពីរគ្រឿង)	$H_A \neq 2$ (ម៉ូតូពីរគ្រឿង)
H_0: $\mu \geq 2$ (ម៉ូតូពីរគ្រឿង)	H_0: $\mu < 2$ (ម៉ូតូពីរគ្រឿង)
H_0: $\mu \leq 2$ (ម៉ូតូពីរគ្រឿង)	H_0: $\mu > 2$ (ម៉ូតូពីរគ្រឿង)

លក្ខណៈទូទៅនៃសម្មតិកម្មមោឃៈគឺសំគាល់ដោយសញ្ញាស្មើ ដូចជា =, ≥ ឬ ≤

2. ការកំណត់តម្លៃអាល់ហ្វា ឬកំរិតសារៈសំខាន់

ជំហានទីពីរគឺយើងត្រូវកំណត់នូវកំរិតសារៈសំខាន់ (level of significance) ដែលត្រូវយកមកប្រើប្រាស់ក្នុងការធ្វើតេស្ត។ កំរិតសារៈសំខាន់ គឺជាប្រូបាប៊ីលីតេដែលធ្វើការបដិសេធ H_0ពេលដែលសម្មតិកម្មមោឃៈនោះពិត ហើយវាត្រូវបានតាងដោយ α ។ ជាទូទៅតម្លៃ α គឺនៅចន្លោះពី 0 ទៅ 1 ហើយតម្លៃ α ដែលត្រូវបានគេប្រើញឹកញាប់មាន បីកំរិត 10%, 5% និង 1%។ ជាទម្លាប់ α =1% គឺសម្រាប់គ្រប់គ្រងគុណភាព α =5% សម្រាប់ការសិក្សាលើផ្នែកទីផ្សារ និង α =10% គឺសម្រាប់ការស្ទង់មតិលើការបោះឆ្នោត។

3. ការកំណត់ប្រភេទតេស្តនិងតម្លៃស្ថិតិគោល

ដើម្បីធ្វើការតេស្តឲ្យបានត្រឹមត្រូវ អ្នកត្រូវកំណត់ថា ប្រភេទតេស្តមួយណាដែលត្រូវប្រើ តួយ៉ាងដូចជា t, ឬ Z, F ឬ χ^2 ជាដើម។ ការជ្រើសរើសប្រភទតេស្តយកមកប្រើគឺអាស្រ័យលើប្រភេទទិន្នន័យ

និងការសន្មត់ឬល័ក្ខខ័ណ្ឌ។ ឧទាហរណ៍ សម្រាប់ប៉ាន់គំរូដែលមានមួយអថេរ ហើយការធ្វើតេស្តទៅលើតម្លៃមធ្យម បើយើងសន្មត់ថាគម្លាតស្តង់ដារដឹងគឺប្រើ Z-test ហើយបើគម្លាតស្តង់ដារមិនដឹងគឺប្រើ t-test។

បន្ទាប់ពីដឹងថាប្រភេទតេស្តណាមួយនឹងត្រូវប្រើ ជំហានបន្ទប់គឺរកតម្លៃស្ថិតិគោល (critical value) ។ តម្លៃស្ថិតិគោល គឺជាចំណុចមួយនៃបំណែងចែកតំបន់បដិសេធ និងមិនបដិសេធ។ ហើយវាក៏ជាតម្លៃដែលតម្លៃតេស្តស្ថិតិត្រូវបានប្រៀបធៀបជាមួយ ដើម្បីសំរេចថាត្រូវបដិសេធសម្មតិកម្មមោឃៈ (H_0) ឬមិនបដិសេធ។ តម្លៃស្ថិតិគោលគឺអាស្រ័យទៅនឹងកំរិតសារៈសំខាន់ α និងការធ្វើតេស្តសងខាង ឬកន្ទុយតែម្ខាង។ ឧទាហរណ៍ បើសិនជា កំរិតសារៈសំខាន់ឬ $\alpha = 5\%$ ត្រូវបានជ្រើសរើស យើងនឹងបានតម្លៃស្ថិតិគោលពីតារាងបំណែងចែកធម្មតាសម្រាប់កន្ទុយម្ខាងគឺ $Z_c = 1.645$ (រូប 7.1) និងសម្រាប់កន្ទុយសង្ខាងគឺ $Z_c = 1.96$ (រូប 7.2)

4. គណនាតម្លៃតេស្តស្ថិតិ

តម្លៃតេស្តស្ថិតិ(test statistic)គឺជាតម្លៃមួយដែលត្រូវបានគណនាចេញពីប៉ាន់គំរូ ហើយវាត្រូវបានគេប្រើប្រាស់ដើម្បីកំណត់ថាតើបដិសេធសម្មតិកម្មមោឃៈដែរឬអត់ តាមរយៈការប្រៀបធៀបតម្លៃតេស្តនេះទៅនឹងតម្លៃស្ថិតិគោលនៅក្នុងជំហានទី4។ ចំណុចខុសគ្នារវាង តេស្តស្ថិតិ (test-statistic) ស្ថិតិគោល (critical value) គឺថា ដើម្បីបានតម្លៃតេស្តស្ថិតិ គេត្រូវប្រមូលទិន្នន័យរួចយកមកគណនា ដោយឡែក ដើម្បីបានតម្លៃស្ថិតិគោល គេត្រូវមើលតារាងស្ថិតិ។ ជាទូទៅការតេស្តសម្មតិកម្មសម្រាប់ប៉ាន់គំរូតែមួយគឺមានតេស្តពីរប្រភេទដែលអាចប្រើបានគឺ z-test និង t-test ។ z-test គឺត្រូវបានប្រើក្នុងករណីដែលគម្លាតស្តង់ដារដឹង (known σ) ហើយមានរូបមន្តដូច (7.1) ហើយ t-test ត្រូវបានប្រើក្នុងករណីដែលគម្លាតស្តង់ដារមិនដឹង (unknown σ)។

z-test មានរូបមន្តដូច (7.2) និង t-test មានរូបមន្តដូច (7.1) ។

$$Z - statistic = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \qquad (7.1)$$

$$t - statistic = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \qquad (7.2)$$

5. ការកំណត់អំពីវិធាននៃការធ្វើសេចក្តីសំរេចចិត្ត

បើសិនតម្លៃដាច់ខាតរបស់តេស្តស្ថិតិ (test-statistic) ធំជាងតម្លៃដាច់ខាតរបស់ស្ថិតិគោល (critical value) គឺ H_0 ត្រូវបានបដិសេធ និងផ្ទុយមកវិញ បើសិនតម្លៃដាច់ខាតេស្តស្ថិតិតូចជាងតម្លៃស្ថិតិគោល គឺ H_0មិនត្រូវបដិសេធ។

i) ការធ្វើតេស្តកន្ទុយម្ខាង (one-tailed test)

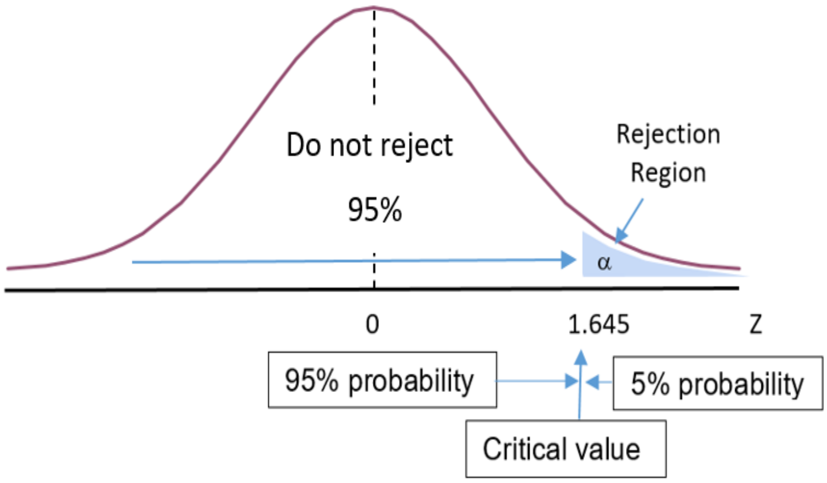
ការធ្វើតេស្តកន្ទុយម្ខាង ជួនកាលអាចជាកន្ទុយខាងស្តាំ ឬកន្ទុយខាងឆ្វេង គឺអាស្រ័យលើសម្មតិកម្មឧបមាថា ចៅសង្កាត់ម្នាក់បានអះអាងថា "គ្រួសារនីមួយៗក្នុងភូមិរបស់គាត់មានម៉ូតូពីរយ៉ាងតិច"($H_0 \geq 2$) ការធ្វើតេស្តគឺត្រូវធ្វើឡើងនៅកន្ទុយខាងឆ្វេង។ ម៉្យាងវិញទៀត បើគាត់អះអាងថា "គ្រួសារនីមួយៗក្នុងភូមិរបស់គាត់មានម៉ូតូពីរយ៉ាងច្រើន"($H_0 \leq 2$) ការធ្វើតេស្តគឺត្រូវធ្វើឡើងនៅកន្ទុយខាងស្តាំវិញម្តង។ ហេតុអ្វី? ពីព្រោះថា អ្នកសិក្សា ឬអ្នកធ្វើតេស្តជានិច្ចកាលតែងតែចង់បដិសេធនូវសម្មតិកម្ម H_0ដើម្បីបង្ហាញស្នាដៃរបស់គេ។ មានវិធីមួយទៀតដែលងាយចំណាំជាងនេះគឺការធ្វើតេស្តកន្ទុយខាងឆ្វេងត្រូវបានគេប្រើបើសិនតម្លៃ test-

statistic អវិជ្ជមាន ហើយផ្ទុយមកវិញការធ្វើតេស្តកន្ទុយខាងស្តាំត្រូវបានគេប្រើបើសិនតម្លៃ test-statistic វិជ្ជមាន វិធីសាស្ត្រនេះលុះត្រាតែអ្នកធ្វើការគណនា t-statistic រួចទើបដឹង។

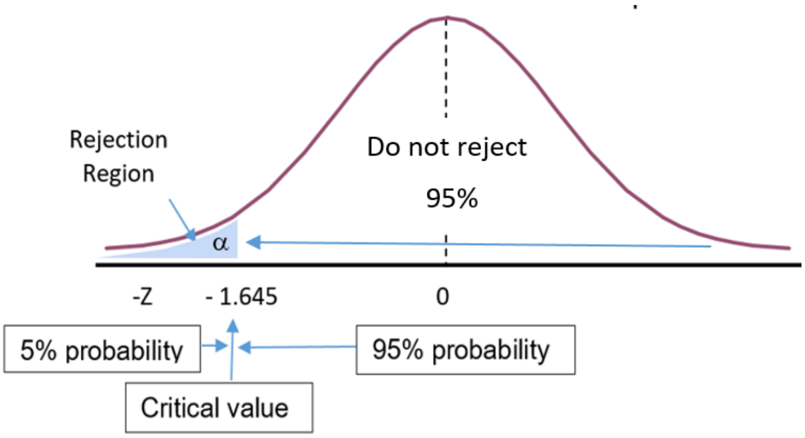
a. **ការធ្វើតេស្តកន្ទុយខាងស្តាំ** (right or upper tailed test): ឧបមាថា អាល់ហ្វាត្រូវបានគេជ្រើសរើសនូវកំរិត 5% តម្លៃស្ថិតិគោល (critical value) គឺ $Z_c = 1.645$ (មើរូប 7.1) ហើយបើតម្លៃ Z-statistic ធំជាង 1.645 (នៅខាងស្តាំដៃ1.645) យើងបដិសេធសម្មតិកម្មមោឃៈ (H_0) ហើយផ្ទុយមកវិញ បើតម្លៃ Z-statistic តូចជាង 1.645 (នៅខាងឆ្វេងដៃ1.645) គឺសំរេចចិត្តមិនបដិសេធ (H_0) នូវកំរិតអាល់ហ្វា 5%។

b. **ការធ្វើតេស្តកន្ទុយខាងឆ្វេង(left or lower-tailed test):** ឧបមាថា អាល់ហ្វាត្រូវបានគេជ្រើសរើសនូវកំរិត 5% ហើយការធ្វើតេស្តគឺត្រូវធ្វើឡើងនៅកន្ទុយខាងឆ្វេងវិញម្តង (រូប 7.2) គឺតម្លៃ Z-critical = -1.645។ បើ Z-statistic < -1.645 សម្មាតិកម្មមោឃៈត្រូវបានបដិសេធនៅកំរិតអាល់ហ្វា 5% ហើយផ្ទុយមកវិញ Z-statistic > -1.645 គឺសម្មាតិកម្មមោឃៈមិនត្រូវបានបដិសេធនោះឡើយ។ ។

រូប 7.1 បំណែងចែក របស់ Z នៅកំរិតសំខាន់ α=5% កន្ទុយស្តាំ



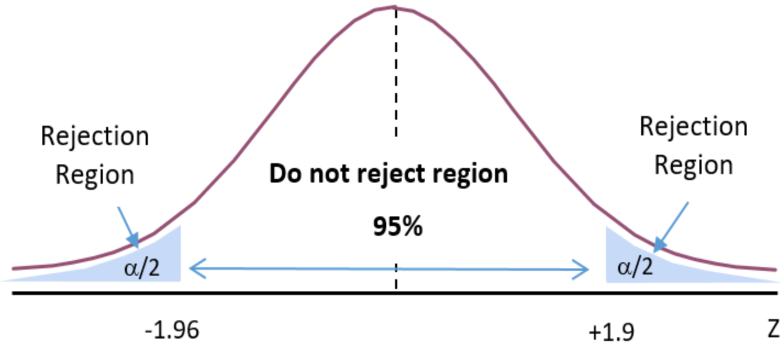
រូប 7.2 បំណែងចែក របស់ Z នៅកំរិតសំខាន់ α=5% កន្ទុយឆ្វេង



ii) ការធ្វើតេស្តកន្ទុយសងខាង (two-tailed test)

បើសិនអាល់ហ្វាត្រូវបានគេជ្រើសរើសនូវកំរិត 5% សម្រាប់ការធ្វើតេស្តកន្ទុយសងខាង តម្លៃស្ថិតិគោល (critical value) $Z_c = \pm 1.96$ (មើរូប 7.3)។ បើតម្លៃដាច់ខាត Z-statistic នៅក្នុងរូបមន្ត (7.1) ធំជាងតម្លៃដាច់ខាត 1.96 គឺយើងបដិសេធសម្មតិកម្មមោឃៈ (H_0) និងផ្ទុយមកវិញ បើតម្លៃដាច់ខាត Z-statistic តូចជាងដាច់ខាត 1.96 ឬ តម្លៃ Z-statistic នៅចន្លោះ= ±1.96 គឺសំរេចចិត្តមិនបដិសេធ (H_0) នូវកំរិតអាល់ហ្វា 5%។

រូប 7.3 បំណែងចែក របស់ Z នៅកំរិតសំខាន់ α=5% កន្ទុយសង្ខាង



7.2 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសងខាង

7.2.1 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសងខាងពេលគម្លាតស្តង់ដារដឹង

ក្នុងករណីដែលគម្លាតស្តង់ដារដឹង យើងប្រើ Z-test ហើយបើករណីដែលគម្លាតស្តង់ដារមិនដឹង យើងប្រើ t-test (t-test និងវិភាគនៅពេលបន្តិចទៀតនេះ)។

ឧទាហរណ៍ 7.1 សិប្បកម្មផលិតដបទឹកមួយប្រភេទដែលធ្វើពីដើមឫស្សីបានកំពស់តាមស្តង់ដារសកលគិតជាមធ្យមគឺ 30.15 សង់ទីម៉ែត្រ និងគម្លាតស្តង់ដារ σ = 0.8 សង់ទីម៉ែត្រ។ ប៉ាន់គំរូចំនួន 100 ត្រូវបានគេប្រមូល ហើយបានផ្តល់នូវកំពស់ជាមធ្យម 30 សង់ទីម៉ែត្រ ។ តើការផលិតនេះគោរពតាមស្តង់ដារសកលដែរ ឬទេ?

ចំឡើយ:

i) ការប្រើតេស្តស្ថិតិ

1. កំណត់សម្មតិកម្មមោឃៈនិង សម្មតិកម្មជំនួស

$H_0: \mu = 30.15$ សង់ទីម៉ែត្រ $H_1: \mu \neq 30.15$ សង់ទីម៉ែត្រ (តេស្តកន្ទុយសង្ខាង)

2. ការកំណត់តម្លៃអាល់ហ្វា ឬកំរិតសារៈសំខាន់

សន្មត់ថា α=0.05 ត្រូវបានគេជ្រើសរើស សម្រាប់ការធ្វើតេស្តនេះ និង n=100 (ទិន្នន័យនៅក្នុង Table 7.1)

3. ការកំណត់តម្លៃស្ថិតិគោល និងប្រភេទតេស្ត

ដោយសារ α=0.05 តាមតារាង the Standard Normal Distribution យើងបានតម្លៃស្ថិតិគោល Z-critical = ± 1.96

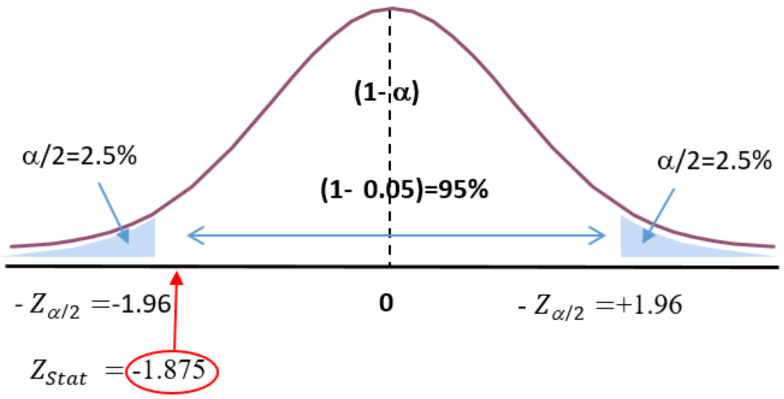
4. គណនាតម្លៃតេស្តស្ថិតិ

បើគម្លាតស្តង់ដារដឹងយើងប្រើ Z-test, n=100, σ=0.8, $\bar{X}$=30, μ=30.15 ជំនួសចូល រូមន្ត (7.1) យើងបាន $Z - statistic = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{30-30.15}{\frac{0.8}{\sqrt{100}}} = -1.875$

5. ការកំណត់អំពីវិធាននៃការធ្វើសេចក្តីសំរេចចិត្ត

បើសិនតម្លៃដាច់ខាតរបស់ Z-test តូចជាង 1.96 មិនបដិសេធ H_0ក៏ប៉ុន្តែបើតម្លៃដាច់ខាតរបស់ Z-test ធំជាង1.96 បដិសេធH_0

ក្នុងឧទាហរណ៍របស់យើង Z-statistic = -1.875 តូចជាង Z-critical = -1.96 ស្ថិតនៅតំបន់មិនបដិសេធ (រូប 7.4)



ដោយសារ $Z_{stat} = -1.875$ < -1.96 សម្មតិកម្មមោឃៈមិនត្រូវបានបដិសេធ ដូច្នេះយើងសន្និដ្ឋានថាមានភស្តុតាងមិនគ្រប់គ្រាន់ដែលបញ្ជាក់ថា កំពស់ដបជាមធ្យមគឺខុសពី 30.15 cm នោះទេ (insufficient evidence that true mean of the bamboo bottle's height is different from 30.15 cm)។

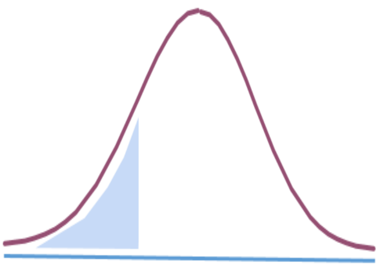
i) ការប្រើតម្លៃភីវែលយូ (p-value)

ក្រៅពីវិធីសាស្ត្រប្រៀបធៀបតម្លៃZ_{Stat} ទៅនឹងតម្លៃ Z-critical យើងមានវិធីសាស្ត្រជាច្រើនទៀតដើម្បីធ្វើសេចក្តីសម្រេចចិត្តថាតើត្រូវបដិសេធសម្មតិកម្មមោឃៈ ឬមិនត្រូវបដិសេធ តួរយ៉ាងដូចជា ការគណនាចន្លោះជឿជាក់ និងការមើលតម្លៃ p-value ជាដើម។ សម្រាប់ពេលនេះគឺយើងពិភាក្សាតម្លៃ p-value

ជាមុនសិន។ បើសិនជាតម្លៃ p តូចជាងអាល់ហ្វា ($p < \alpha$) យើងបដិសេធ សម្មតិកម្មមោឃៈ ហើយផ្ទុយមកវិញបើ $p > \alpha$ យើងមិនបដិសេធសម្មតិកម្មមោឃៈនោះឡើយ។

p-value ត្រូវបានឲ្យនិយមន័យថា ជាប្រូបាប៊ីលីតេដែលបដិសេធសម្មតិកម្មមោឃៈ (H_0)។ វាគឺជាផ្ទៃខាងចុងបំផុតនៃរូប បើនិយាយឲងាយយល់ (ឧទាហរណ៍ Z-test) p-value គឺបង្ហាញអំពីទីតាំងរបស់ Z_{Stat} ថាតើវានៅត្រង់ណានៃទីតាំងណាមួយនៃ Z-critical? បើ p-value តូចជាងអាល់ហ្វា ($p < \alpha$) Z_{Stat} ស្ថិតនៅតំបន់បដិសេធ ហើយផ្ទុយមកវិញ បើ p-value ធំជាងអាល់ហ្វា ($p > \alpha$) Z_{Stat} ស្ថិតនៅតំបន់មិនបដិសេធ។ ជាទូទៅតម្លៃ Z_{Stat}កាន់តែធំ p-value កាន់តែតូច កាន់តែបដិសេធសម្មតិកម្មមោឃៈ (H_0) (មើលរូប 7.4)។

ឥឡូវយើងពិនិត្យឧទាហរណ៍7.1 របស់យើងឡើងវិញ $Z_{Stat} = -1.87$ ហើយបើយើងមើលតារាង The Cumulative Standardized Normal Distribution (ដូចបានបង្ហាញក្នុងតារាង 7.1) យើងនឹងទទួលបានតម្លៃ p-value =0.0307 ដោយសារយើងតេស្តសងខាង តម្លៃ p-value ត្រូវគុណនឹង 2*0.0307= 6.14% > α=5% ដូច្នេះ យើងមិនមានភស្តុតាងគ្រប់គ្រាន់ដើម្បីធ្វើការបដិសេធសម្មតិកម្មមោឃៈនោះទេ[14]។

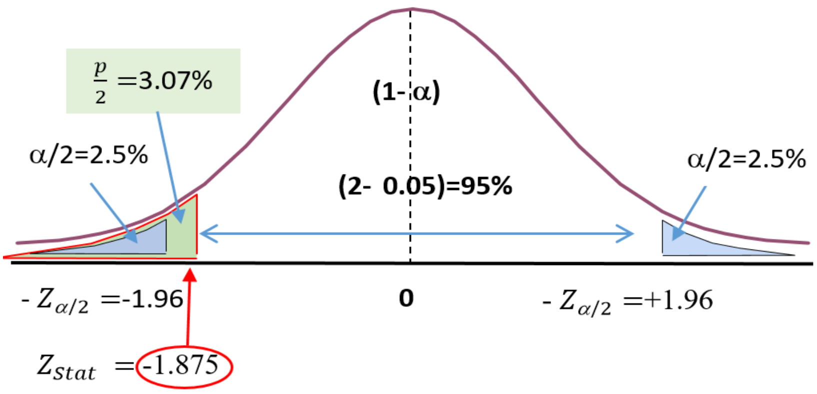


តារាង 7.1 The Cumulative Standardized Normal Distribution

Z	0.00	0.01	...	...	0.07	0.08	0.09
-6.0	...	...	...	...	...	...	...
-5.5	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
-1.9	...	...	...	...	...	...	...
-1.8	...	...	...	...	**0.0307**	0.0301	0.0204
...	...	...	...	...	...	...	...

រូប 7.4 តម្លៃ p-value ពេល Z-statistic= -1.875

14 តារាង The Cumulative Standardized Normal Distribution គឺជាតារាងដែលផ្តល់ប្រូបាប៊ីលីតេតែកន្ទុយម្ខាងទេ ហើយបើយើងធ្វើការតេស្តកន្ទុយសងខាង យើងត្រូវយកតម្លៃក្នុងតារាងនេះគុណនិងពីរ រួចធៀបវាទៅនិងតម្លៃអាល់ហ្វា។ ហើយក្នុងករណីធ្វើកាតេស្តកន្ទុយម្ខាង យើងអាចយកតម្លៃក្នុងតារាង The Cumulative Standardized Normal Distribution ធៀបជាមួយអាល់ផហ្វាតែម្តង មិនបាច់គុណនឹង2នោះទេ។



ii) ការប្រើចន្លោះជឿជាក់

ដូចដែលបញ្ជាក់ពីខាងលើខ្លះមកហើយ ក្រៅពីមើលតម្លៃ p-value យើងអាចមើលទៅលើចន្លោះជឿជាក់ ដើម្បីកំណត់ថាតើសម្មតិកម្មមោឃៈនិងត្រូវបានបដិសេធ ឬយ៉ាងណា។ សម្រាប់ឧទាហរណ៍ 7.1 របស់យើង $\bar{X}$ = 30, σ=0.8, n=100។ រូបមន្តចន្លោះជឿជាក់ជំពូក 6 សមីការ (6.1)៖

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$30 - 1.96\frac{0.8}{\sqrt{100}} \leq \mu \leq 30 + 1.96\frac{0.8}{\sqrt{100}}$$

$$29.8432 \leq \mu \leq 30.1568$$

ដោយហេតុថា ចន្លោះជឿជាក់នេះមានតម្លៃសម្មតិកម្មមោឃៈ ($\mu = 30.15$) ដូច្នេះយើងមិនបដិសេធសម្មតិកម្មមោឃៈនូវកំរិត α=5% នោះទេ។

សរុបមក ដើម្បីពិនិត្យទៅលើលទ្ធផលនៃការតេស្តសម្មតិកម្ម យើងមានជំរើសបី ៖ ទីមួយ គឺយើងយកតម្លៃ test-statistic ធៀបជាមួយតម្លៃ critical-value ទីពីរគឺយកតម្លៃ p-value ធៀបជាមួយតម្លៃអាល់ហ្វា និងទីបីគឺពិនិត្យទៅលើចន្លោះជឿជាក់។ ជំរើសទាំងបីនេះគឺផ្តល់នូវចំលើយតែមួយ ប៉ុន្តែ p-value មានលក្ខណៈទូលំទូលាយជាង។

7.2.2 ការធ្វើតេស្តលើសម្មតិកម្មកន្ទុយសង្ខាងពេលគម្លាតស្តង់ដារមិនដឹង

ក្នុងជីវភាពជាក់ស្តែង គម្លាតស្តង់ដារភាគច្រើនគឺមិនដឹង។ ក្នុងករណីដែលគម្លាតស្តង់ដារមិនដឹង យើងអាចប្រើ t-test ជំនួស Z-test វិញ។

ឧទាហរណ៍ 7.2 របាយការណ៍របស់ក្រសួងទេសចរណ៍មួយបានអះអាងថា តម្លៃសណ្ឋាគារថ្នាក់កណ្តាលនៅក្រុងសៀមរាបគិតជាមធ្យមគឺ $66 ក្នុងមួយយប់។ ដើម្បីតេស្តទៅលើអំណះអំណាងនេះ អ្នកស្រាវជ្រាវម្នាក់នៅសកលវិទ្យាល័យជាតិគ្រប់គ្រង បានប្រមូលទិន្នន័យទៅលើសណ្ឋាគារថ្នាក់កណ្តាលចំនួន 25 ហើយបានរកឃើញថា តម្លៃសណ្ឋាគារនៅក្រុងសៀមរាប ជាមធ្យមគឺ $\bar{X}$ = $72.50 ក្នុងមួយយប់ ហើយគម្លាតស្តង់ដារ S=$15.41 (ទិន្នន័យនៅក្នុង Table 9.2)។ បន្ទាប់មកគាត់ក៏ធ្វើការតេស្តទៅលើសម្មតិកម្មនេះ។

ដោយសារតែគម្លាតស្តង់ដាររបស់ចំនួនសកលមិនបានដឹង (σ unknown) ហើយយើងមានតែគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូតែប៉ុណ្ណោះ (S=$15.41)។ ដូច្នេះ t-test នឹងត្រូវបានប្រើជំនួស Z-test វិញ។ ចំណែកឯជំហ៊ានក្នុងការធ្វើតេស្តនេះគឺដូចគ្នាទៅនិង Z-test ដែរ៖

ចំឡើយ:

i) ការប្រើតេស្តស្ថិតិ

1. កំណត់សម្មតិកម្មមោឃៈនិង សម្មតិកម្មជំនួស

   H_0: $\mu = \$66$ H_1: $\mu \neq \$66$ (តេស្តកន្ទុយសងខាង)

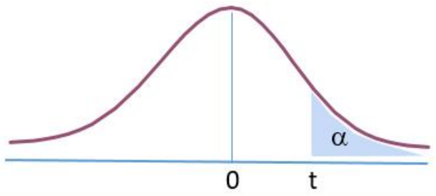
2. ការកំណត់តម្លៃអាល់ហ្វា ឬកំរិតសារៈសំខាន់

   សន្មត់ថា យើងសំរេចយក α=0.05 សម្រាប់ការធ្វើតេស្តនេះ

3. ការកំណត់ប្រភេទតេស្តនិងតម្លៃស្ថិតិគោល

   ដោយសារតែគម្លាតស្តង់ដារមិនបានដឹង ហើយយើងដឹងតែគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ S=$15.41 ដូច្នេះតេស្តដែលត្រូវប្រើគឺ t-test។ បើសិន α=0.05 ហើយ t មាន df= n-1 = 25-1 = 24 មើលតារាងបំណែងចែកអក្សរ t (តារាង 7.2 ខាងក្រោម) ត្រង់ចំណុចប្រសព្វគ្នារវាង df=24 និងប្រូបាប៊ីលីតេកន្ទុយខាងស្តាំ 0.025 យើងនិងបានតម្លៃស្ថិតិគោល t-critical = 2.0639

តារាង 7.2 Critical Values of t



Degrees of Freedom	Cumulative Probabilities					
	0.75	0.90	0.95	0.975	0.99	0.995
	Upper-Tail Areas					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
......						
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874

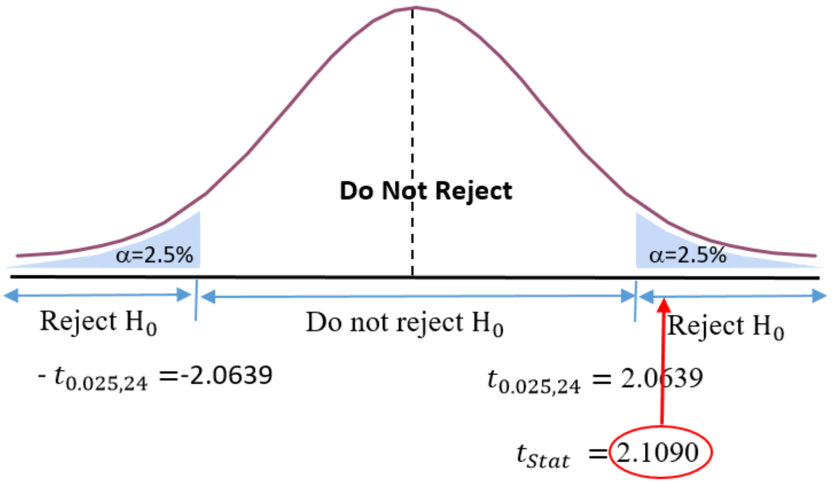
4. គណនាតម្លៃតេស្តស្ថិតិ

   ចំនួនប៉ាន់គំរូ n=25, S=$15.41, $\bar{X} = \$72.50$ និង $\mu = \$66$

   យើងប្រើរូមន្ត (7.2) $t - statistic = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{72.50-66}{\frac{15.41}{\sqrt{25}}} = 2.1090$

5. ការកំណត់អំពីវិធាននៃការធ្វើសេចក្តីសំរេចចិត្ត

   រូប 7.5 តម្លៃ $t_{0.025,24}$ = 2.0639 និង t-statistic =2.1090



បើតម្លៃដាច់ខាតរបស់ t-test តូចជាងតម្លៃដាច់ខាតរបស់ 2.0639 មិនបដិសេធ H_0ក៏ប៉ុន្តែបើតម្លៃដាច់ខាតរបស់ t-test ធំជាងតម្លៃដាច់ខាតរបស់ 2.0639 បដិសេធ H_0

តាមរយៈរូប 7.5 បញ្ជាក់ឲដឹងថា តម្លៃ t-statistic (2.1090) ធំជាងតម្លៃ t-critical (2.0639) យើងបដិសេធ H_0។ ដូច្នេះ យើងសន្និដ្ឋានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា តម្លៃបន្ទប់របស់សណ្ឋាគារថ្នាក់កណ្តាលគឺខុសពី $66 ដុល្លាក្នុងមួយយប់ ហើយរបាយការណ៍នោះមិនត្រឹមត្រូវនោះទេ។

ii) ការប្រើវិធីសាស្រ្តភីវែលយូ (p-value)

ដោយសារតម្លៃ t-statistic =2.1090 ហើយតម្លៃតេស្តនេះមាន df=24 តាមតារាង Excel យើងបានតម្លៃភីវែលយូ p =0.0456 < 0.05 ដូច្នេះយើងបដិសេធ H_0 ។

iii) ការប្រើចន្លោះជឿជាក់ (confidence interval)

សូមរំលឹកថា នៅជំពូកទី6 យើងធ្លាប់ដឹងរូបមន្តចន្លោះជឿជាក់រួចមកហើយ សម្រាប់ពេលនេះយើងយករូបមន្តនោះមកអនុវត្តទីនេះតែម្តង។

ឥឡូវនេះយើងពិនិត្យឧទាហរណ៍ 7.2 របស់យើងឡើងវិញ។ n=25, S=$15.41, $\bar{X}$ = $72.50 ។ ប្រសិន α=0.05 ហើយ t មាន df= n-1 = 25-1 = 24 តាមរយៈតារាងបំណែងចែកអក្សរ t (កន្ទុយសងខាង) យើងនិងបានតម្លៃស្ថិតិគោល t-critical = 2.0639

$$\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}} \qquad (6.2)$$

$$72.50 - 2.0639\frac{15.41}{\sqrt{25}} \leq \mu \leq 72.50 + 2.0639\frac{15.41}{\sqrt{25}}$$

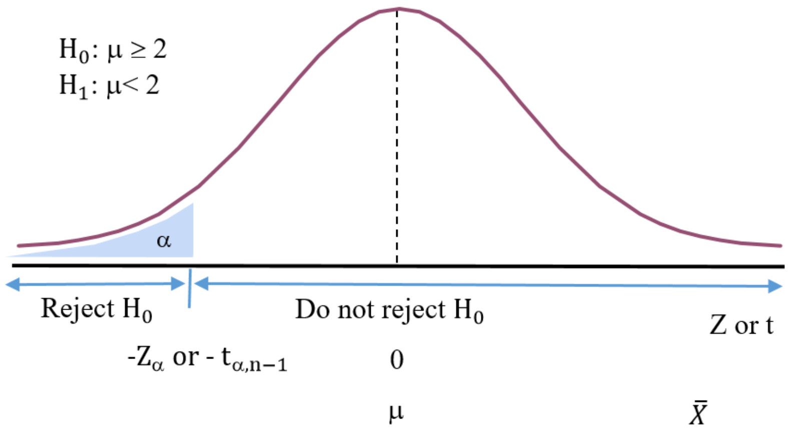
$$\$66.14 \leq \mu \leq \$78.86$$

ដោយហេតុថា 95% នៃតម្លៃបន្ទប់សណ្ឋាគារថ្នាក់កណ្តាលគឺស្ថិតនៅចន្លោះ $66.14 ដល់ $78.86 ហើយចន្លោះនេះពុំមានលេខ 66 នោះឡើយ ដូច្នេះយើងយើងបដិសេធ $H_0: \mu = \$66$ នៅកំរិត α=0.05 ហើយធ្វើការសន្និដ្ឋានដូចមុន។

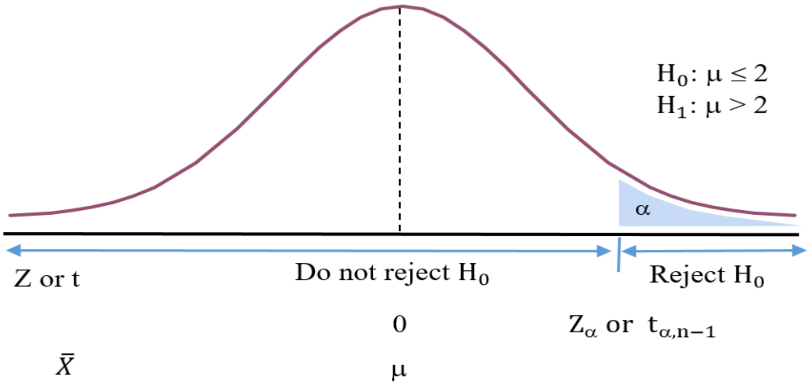
7.3 **ការធ្វើតេស្តកន្ទុយម្ខាង**

ការធ្វើតេស្តកន្ទុយម្ខាង (one-tail tests) ត្រូវបានគេធ្វើឡើងពេលដែលអ្នកសិក្សាចាប់អារម្មណ៍ទៅលើទិសដៅតែម្ខាង។ ការធ្វើតេស្តនេះអាចមានកន្ទុយខាងឆ្វេង ឬខាងស្តាំ វាអាស្រ័យលើសម្មតិកម្មដែលមាន។ ហើយការធ្វើតេស្តនេះទៀតសោតគឺអាចប្រើ t-test ឬ Z-test គឺ អាស្រ័យលើគម្លាតស្តង់ដាររបស់ចំនួនសកលដឹងឬមិនដឹង។ ឧបមាថា ចៅសង្កាត់មួយរូបបានអះអាងថា ប្រជាជននៅភូមិគាត់មានម៉ូតូ 2 គ្រឿងយ៉ាងតិចក្នុងមួយគ្រួសារ ($H_0: \mu \geq 2$) ដើម្បីតេស្តទៅលើសម្មតិកម្មនេះ យើងត្រូវតេស្តកន្ទុយខាងឆ្វេង ពីព្រោះថា អ្នកធ្វើតេស្តគឺជាអ្នកដែលមិនយល់ស្រប (បើយល់ស្របគឺមិនចាំបាច់ធ្វើតេស្តនោះទេ) ហើយមានបំណងបដិសេធសម្មតិកម្មនេះ។

រូប 7.6 ការតេស្តកន្ទុយខាងឆ្វេង (lower-tail tests)



រូប 7.7 ការតេស្តកន្ទុយខាងឆ្វេង (upper-tail tests)



ដើម្បីបដិសេធបានលុះត្រាតែគ្រួសារមួយមានម៉ូតូតិចជាង 2គ្រឿងក្នុងមួយគ្រួសារ ។ បើ $\mu < 2$ គឺកន្ទុយខាងឆ្វេងដូចបង្ហាញក្នុងរូប 7.6។ ហើយបើចៅសង្កាត់រូបនោះបានអះអាងថា ប្រជាជននៅភូមិគាត់មានម៉ូតូ 2គ្រឿងយ៉ាងច្រើនបំផុតក្នុងមួយគ្រួសារ ($H_0: \mu \leq 2$) ក្នុងករណីនេះ ការធ្វើតេស្តគឺត្រូវធ្វើឡើងនៅកន្ទុយខាងស្តាំង ដូចបង្ហាញក្នុងរូប 7.7 ដោយសំអាងលើហេតុផលស្រដៀងគ្នាទៅនិងការធ្វើតេស្តលើកន្ទុយខាងឆ្វេង។

ឧទាហរណ៍ 7.3 t-test កន្ទុយខាងស្តាំ (គម្លាតស្តង់ដារមិដឹង): ឥឡូវយើងធ្វើការពិនិត្យទៅលើឧទាហរណ៍ 7.2 របស់យើងឡើងវិញ ឧបមាថា របាយការណ៍របស់ក្រសួងទេសចរណ៍អះអាងថា តម្លៃសណ្ឋាគារ

ថ្នាក់កណ្ដាលនៅសៀមរាបយ៉ាងច្រើនបំផុតគឺ $66 ក្នុងមួយយប់ ($H_0$: $\mu \le \$66$)។ ដើម្បីបញ្ជាក់ថាអំណះអំណាងនេះពិតឬមិនពិត យើងធ្វើការសាកសួរសណ្ឋាគារចំនួន n=25, យើងបានតម្លៃមធ្យម $\bar{X} = \$72.50$ ក្នុងមួយយប់ ហើយគម្លាតស្តង់ដារ S=$15.41 (ទិន្នន័យនៅក្នុង Table 9.2) ហើយយើងធ្វើការតេស្តកន្ទុយខាងស្តាំ ហើយនីតិវិធីក្នុងការតេស្តគឺដូចមុន។

i) ការប្រើតេស្តស្ថិតិ

ចំឡើយ:

1. កំណត់សម្មតិកម្មមោឃៈនិង សម្មតិកម្មជំនួស

   H_0: $\mu \le \$66$ H_1: $\mu > \$66$ (តេស្តកន្ទុយខាងស្តាំ)

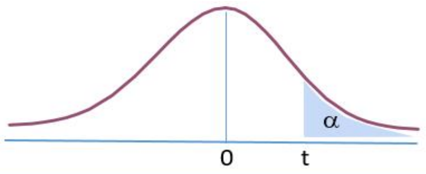
2. ការកំណត់តម្លៃអាល់ហ្វា ឬកំរិតសារៈសំខាន់

   សន្មត់ថា យើងសំរេចយក α=0.05 សម្រាប់ការធ្វើតេស្តនេះ

3. ការកំណត់ប្រភេទតេស្តនិងតម្លៃស្ថិតិគោល

   ដោយសារតែគម្លាតស្តង់ដារមិនបានដឹង ហើយយើងដឹងតែគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ S=$15.41 ដូច្នេះតេស្តដែលត្រូវប្រើគឺ t-test។ បើ α=0.05 ហើយ t មាន df= n-1 = 25-1 = 24 មើលតារាងបំណែងចែកអក្សរ t (តារាង 7.3 ខាងក្រោម) ត្រង់ចំណុចប្រសព្វគ្នារវាង df=24 និងប្រូបាប៊ីលីតេកន្ទុយខាងស្តាំ 0.05 យើងនិងបានតម្លៃស្ថិតិគោល t-critical = 1.7109

តារាង 7.3 Critical Values of t



Degrees of Freedom	Cumulative Probabilities					
	0.75	0.90	0.95	0.975	0.99	0.995
	Upper-Tail Areas					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
...	...	...	...	...	...	...
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874

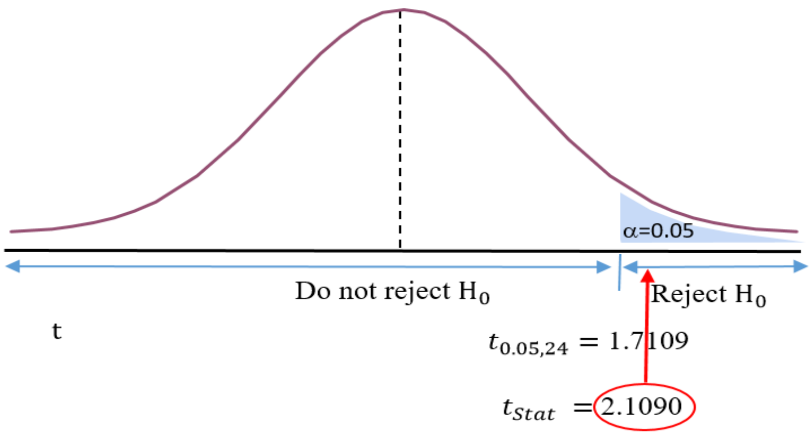
4. គណនាតម្លៃតេស្តស្ថិតិ

   ចំនួនប៉ាន់គំរូ n=25, S=$15.41, $\bar{X} = \$72.50$ និង $\mu = \$66$

   យើងប្រើរូបមន្ត (7.2) $t - statistic = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{72.50-66}{\frac{15.41}{\sqrt{25}}} = 2.1090$

5. ការកំណត់អំពីវិធាននៃការធ្វើសេចក្តីសំរេចចិត្ត

   រូប 7.8 តម្លៃ $t_{0.05,24}$ = 1.7109 និង t-statistic =2.1090



បើតម្លៃ t-test តូចជាងតម្លៃដាច់ខាតរបស់ 1.7109 មិនបដិសេធ H_0 ហើយបើតម្លៃ t-test ធំជាងតម្លៃ 1.7109 បដិសេធ H_0។

លទ្ធផល t-test = 2.1090 > 1.7109 ដូច្នេះយើងបដិសេធ H_0 ហើយសន្និដ្ឋានថា មិនមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា តម្លៃសណ្ឋាគារថ្នាក់កណ្តាលនៅក្រុងសៀមរាបមានតម្លៃយ៉ាងច្រើនបំផុតគឺ $66នោះទេ។

ii) ការប្រើភីវ៉ែលយូ (p-value)

ក្រៅពីមើលតម្លៃ t-test ធៀបទៅនឹង t-critical ដើម្បីបដិសេធ H_0 យើងមានវិធីមួយទៀតគឺមើលតម្លៃ p-value របស់ t-test រួចធៀបវាទៅនឹងតម្លៃអាល់ហ្វា។ សម្រាប់ឧទាហរណ៍របស់យើង ដោយសារ លទ្ធផល t-test = 2.1090 ហើយ t មាន df=24 តាម Excel យើងទទួលបាន p-value = 0.02278 < α= 0.05[15] ដូច្នេះយើងបដិសេធ H_0។

iii) ការប្រើចន្លោះជឿជាក់

ជាទូទៅការធ្វើតេស្តកន្ទុយម្ខាង ទោះជាខាងស្តាំ ឬឆ្វេងក៏ដោយ គឺមានតែវិធីពីរតែប៉ុណ្ណោះ គឺតាមរយៈ test-statistic និង p-value ហើយគេកម្រមើលគណនាចន្លោះជឿជាក់ (confidence interval) ណាស់ លើកលែងតែករណីចាំបាច់ ឧទាហរណ៍ដូចជា សវនកម្ម ឬការត្រួតពិនិត្យគុណភាពដូចបានសិក្សានៅជំពូកមុន។

7.4 តេស្តសម្មតិកម្មលើសមាមាត្រ

សូមរំលឹកថា សមាមាត្រគឺទាក់ទងទៅនឹងលទ្ធផលពីរ ឧទាហរណ៍ដូចជា អតិថិជនបើកគណនេយ្យ ឬមិនបើកគណនេយ្យ ទិញឬមិនទិញ យល់ព្រម ឬមិនយល់ព្រម។ល។ ការតេស្តលើសមាមាត្រក៏អាចធ្វើឡើងដោយប្រើកន្ទុយសងខាង ឬម្ខាង ដូចការធ្វើតេស្តសម្រាប់តម្លៃមធ្យមដែរ។

នៅជំពូកទី6 យើងធ្លាប់ដឹងមកហើយ សមាមាត្រមានរូបមន្តដូចខាងក្រោម៖

$$p = \text{សមាមាត្រនៃប៉ាន់គំរូ} = \frac{X}{n} = \frac{\text{ចំនួនវត្ថុដែលចាប់អារម្មណ៍}}{\text{ចំនួនប៉ាន់គំរូ}} \qquad (6.3)$$

[15] ករណីនេះមិនចាំបាច់គុណនឹង 2ទេពីព្រោះយើងតេស្តកន្ទុយតែម្ខាង

ពេល X > 5 ហើយ n-X>5 នោះ p មានលក្ខណះប្រហាក់ប្រហែល normal distribution ដែលមានតម្លៃមធ្យមនិងគម្លាតស្តង់ដារ[16]៖

$$\mu_p = \pi \qquad (7.1)$$

$$\sigma_{p}=\sqrt{\frac{\pi(1-\pi)}{n}} \qquad (7.2)$$

ដោយសារតែសមាមាត្រនៃប៉ាន់គំរូ p មានបំណែងចែក ប្រហាក់ប្រហែល normal distribution ដូច្នេះ Z-test និងត្រូវបានប្រើ៖

$$Z - statisitc = \frac{p-\pi}{\sigma_p} = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \qquad (7.3)$$

ឧទាហរណ៍ 7.4 ការធ្វើតេស្តលើសមាមាត្រដោយប្រើកន្ទុយសងខាង៖

របាយការណ៍សិក្សាមួយបានអះអាងថា មានសិស្សវិទ្យាល័យចំនួន 15% ចូលចិត្តមុខវិជ្ជាគណិតវិទ្យា។ ដើម្បីតេស្តលើអំណះអំណាងនេះ អ្នកស្រាវជ្រាវម្នាក់បានធ្វើការស្ទង់មតិទៅលើសិស្សវិទ្យាល័យចំនួន 500នាក់ដោយចៃដន្យ ហើយបានរកឃើញថា មានសិស្សចំនួន 30នាក់ឆ្លើយថាចូលចិត្ត។ សន្មត់ថា α=0.05 ត្រូវបានប្រើប្រាស់ តើរបាយការណ៍ខាងលើពិតឬមិនពិត?

ចំឡើយ:

i) ការប្រើតេស្តស្ថិតិ

1. កំណត់សម្មតិកម្មមោឃៈនិង សម្មតិកម្មជំនួស

   H_0: $\pi = 0.15$ $\qquad$ H_1: $\pi \neq 0.15$ $\qquad$ (តេស្តកន្ទុយសង្ខាង)

2. ការកំណត់តម្លៃអាល់ហ្វា ឬកំរិតសារៈសំខាន់

   សន្មត់ថា យើងសំរេចយក α=0.05 សម្រាប់ការធ្វើតេស្តនេះ

3. ការកំណត់ប្រភេទតេស្តនិងតម្លៃស្ថិតិគោល

   ជាទូទៅការធ្វើតេស្តទៅលើសមាមាត្រ គឺ Z-statistic នឹងត្រូវបានគេប្រើ។ ដោយហេតុថា x = 30 > 5 ហើយ n-X= 470 > 5 បំណែងចែកសមាមាត្រនេះមានលក្ខណះ normal ដូច្នេះបើ α=0.05 យើងបានតម្លៃស្ថិតិគោល Z-critical = ±1.96

4. គណនាតម្លៃតេស្តស្ថិតិ

   ចំនួនប៉ាន់គំរូ n=500, X=30, $p = \frac{X}{n} = \frac{30}{500} = 0.06$

   យើងប្រើរូមន្ត (7.3)

   $$Z - statisitc = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.06-0.15}{\sqrt{\frac{0.15(1-0.15)}{500}}} = -5.636$$

5. ការកំណត់អំពីវិធាននៃការធ្វើសេចក្ដីសំរេចចិត្ត

[16] រូបមន្ត (7.3) នឹងប្រើបានក្នុងករណីដែល X > 5 ហើយ n-X>5 ប៉ុន្តែក្នុងករណីដែលចំនួនប៉ាន់គំរូតូច ហើយមានលក្ខណៈមិន normal ជាពិសេស X < 5 ហើយ n-X< 5 រូបមន្តផ្សេងទៀតនឹងអនុវត្ត តែយើងមិនលើកមកសិក្សាក្នុងសៀវភៅនេះទេ ព្រោះរូបមន្តមានសភាពស្មុគស្មាញ ហើយភាគច្រើនគឺមាននៅក្នុង software ផ្សេងៗ។

បើតម្លៃដាច់ខាត Z-statistic < 1.96 មិនបដិសេធ H_0 ហើយបដិសេធ បើសិនជាតម្លៃដាច់ខាត Z-statistic > 1.96។

ដោយសារតម្លៃដាច់ខាតរបស់ Z-statistic = 5.636 > 1.96 H_0 ត្រូវបានបដិសេធ ដូច្នេះ យើងអាចសន្និដ្ឋានបានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា សមាមាត្រសិស្សវិទ្យាល័យចូលចិត្តគណិតវិទ្យាខុសពី 15% ក៏ប៉ុន្តែយើងមិនដឹងថា របាយករណ៍នេះរាយការណ៍លើសឬខ្វះនោះទេ។

ii) ការប្រើភីវែលយូ (p-value)

ដោយសារ តម្លៃ Z-statistic = -5.636 តាមរយៈ Excel Calculator បានផ្តល់ឲ p-value = 1.74E-08 < α=0.05 ដូច្នេះ H_0 ត្រូវបានបដិសេធ

iii) ការប្រើចន្លោះជឿជាក់

បើ α=0.05 ត្រូវបានគេអនុម័ត យើងទទួលបាន Z-critical = ± 1.96។ តាមរូបមន្ត (6.3) នៅជំពូកទី 6 យើងបានចន្លោះជឿជាក់ 95% សម្រាប់សមាមាត្ររបស់ចំនួនសកលៈ

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

$$0.06 - 1.96\sqrt{\frac{0.06(1-0.06)}{500}} \le \pi \le 0.06 + 1.96\sqrt{\frac{0.06(1-0.06)}{500}}$$

$$0.0392 \le \pi \le 0.0808$$

ដោយហេតុថា ចន្លោះជឿជាក់មិនមានតម្លៃ $H_0: \pi = 0.15$ ដូច្នេះយើងបដិសេធH_0 នូវកំរិត α=5%។

សរុបមកវិធីសាស្ត្រទាំងបី៖ test-statistic (ក្នុងករណីនេះគឺ Z-statistic) p-value និង វិធីសាស្ត្រ interval គឺបានបង្ហាញលទ្ធផលដូចគ្នា។

សេចក្តីសង្ខេប

ក្នុងជំពូកនេះ អ្នកបានសិក្សាទៅលើវិធីសាស្ត្រពីរសំខាន់ៗ ក្នុងការតេស្តសម្មតិកម្ម ភីវែលយូ និងចន្លោះជឿជាក់។ អ្នកបានដឹងអំពីនីតិវិធីក្នុងការធ្វើតេស្ត តួយ៉ាងដូចជា ការសរសេរសម្មតិកម្ម និងការជ្រើសរើសប្រភេទតេស្តដែលត្រូវប្រើ និងវិធានករណ៍ក្នុងការធ្វើសេចក្តីសំរេចចិត្ត។ លើពីនេះទៅទៀត អ្នកក៏បានដឹងផងដែរអំពីរបៀបប្រើប្រាស់តេស្តកន្ទុយម្ខាង និងសង្ខាង។

សំនួរនិងលំហាត់

I. សំនួរ

1. អ្វីទៅជាលក្ខណៈខុសគ្នារវាងសម្មតិកម្មមោឃៈនិងសម្មតិកម្មឆ្លាស់
2. អ្វីទៅជាលក្ខណៈខុសគ្នារវាងការតេស្តកន្ទុយម្ខាងនិងសង្ខាង
3. អ្វីទៅភីវែលយូ (p-value)?

II. លំហាត់

4. អ្នកត្រួតពិនិត្យគុណភាពអំពូលអគ្គិសនីម្នាក់ចង់ដឹងថា រយៈពេលមធ្យមនៃអំពូលដែលអាចប្រើប្រាស់បានស្មើទៅនឹង 400 ម៉ោងដែរឬយ៉ាងណា។ គាត់ប្រមូលប៉ាន់គំរូចំនួន 80 អំពូល គិតជាមធ្យមគឺប្រើប្រាស់បាន 385 ម៉ោង។ សន្មត់ថា គម្លាតស្តង់ដាររបស់ចំនួនសកលគឺ 90 ម៉ោង។

a. នៅកំរិតអាល់ហ្វាស្មើនឹង 5% តើមានភស្តុតាងបញ្ជាក់ថាជីវិតមធ្យមរបស់អំពូលខុសពី 400 ម៉ោងដែរឬអត់?

b. ចូលគណនានិងបកស្រាយតម្លៃ p-value

c. ចូលបង្កើតចន្លោះជឿជាក់ 95% ដើម្បីធ្វើការប៉ាន់ស្មានរយៈពេលប្រើប្រាស់ជាមធ្យមរបស់ចំនួនសកលរបស់អំពូល

d. ចូលធ្វើការប្រៀបធៀបចំលើយ (a) និង (c) រួចធ្វើការសន្និដ្ឋាន

5. អ្នកគ្រប់គ្រងភោជនីយដ្ឋានម្នាក់បានធ្វើការកត់ត្រារយៈពេលរងចាំរបស់អតិថិជន (ចាប់ពីពេលហៅម្ហូបរហូតដល់ទទួលបានម្ហូប) គឺមានរះពេល 4 នាទី។ គាត់ធ្វើការជ្រើសរើសប៉ាន់គំរូចំនួន 70 ដោយចៃដន្យ បានបង្ហាញថា រយៈពេលរងចាំជាមធ្យមគឺ 3.50 នាទី និងមានគម្លាតស្តង់ដារ 0.9 នាទី។

a. នៅកំរិតអាល់ហ្វាស្មើនឹង 5% តើមានភស្តុតាងបញ្ជាក់ថារយៈពេលរងចាំខុសពី 4 នាទីដែរឬអត់?

b. ចូលគណនានិងបកស្រាយតម្លៃភីវ៉ែលយូ (p-value)

c. ដោយសារតែចំនួនប៉ាន់គំរូមានតែ 70 តើអ្នកមានក្តីបារម្ភណ៍អំពីទម្រង់នៃបំណែងចែកសកលដែរឬទេពេលដែលអ្នកប្រើ t តេស្ត? ចូលពន្យល់

6. អ្នកគ្រប់គ្រងស្ថានីយ៍លក់សាំងមួយកន្លែងបានអះអាងថា ការលក់សាំងទៅឲរថយន្តនៅស្ថានីយ៍របស់គាត់គឺយ៉ាងតិចបំផុតគឺ 1,000 លីត្រក្នុងមួយថ្ងៃ។ អ្នកស្រាវជ្រាវម្នាក់បានធ្វើការប្រមូលប៉ាន់គំរូដោយចៃដន្យចំនួន 20 ថ្ងៃ ជាមធ្យមលក់បាន 1,020 លីត្រ និងគម្លាតស្តង់ដារ 15 លីត្រ។

a. នៅកំរិតអាល់ហ្វា 5% តើមានភស្តុតាងបញ្ជាក់ថាការអះអាងរបស់អ្នកគ្រប់គ្រងខាងលើពិត ឬមិនពិត?

b. ចូលកំណត់ភីវ៉ែលយូ (p-value) និងធ្វើការបកស្រាយ

7. ខាងក្រោមនេះគឺជាទិន្នន័យចំណាយ(គិតជាដុល្លា) ទៅលើអាហារថ្ងៃត្រង់របស់បុគ្គលិក 10 នាក់នៃក្រុមហ៊ុនមួយ៖

ចំណាយ	4.80	6.20	5.40	5.03	6.45	3.47	4.20	5.86	4.54	5.60

a. នៅកំរិតអាល់ហ្វាស្មើនឹង 5% តើមានភស្តុតាងបញ្ជាក់ថា ចំណាយលើអាហារថ្ងៃត្រង់របស់បុគ្គលិកគិតជាមធ្យមខុសពី 6.80 ដុល្លាឬទេ?

b. ចូលកំណត់ភីវ៉ែលយូ (p-value) ក្នុងចំលើយ (a) និងធ្វើការបកស្រាយ

c. តើល័ក្ខខ័ណ្ឌអ្វីខ្លះដែលត្រូវដាក់លើបំណែងចែករបស់ចំនួនសកល ដើម្បីឲការប្រើប្រាស់ t តេស្តនៅក្នុងចំលើយ (a) មានសុពលភាព?

8. អ្នកគ្រប់គ្រងសេវាលាងរថយន្តមួយកន្លែងចង់ធ្វើឲប្រសើរឡើងនូវសេវារបស់ខ្លួន។ សន្មត់ថា រយៈពេលលាងសំអាតរថយន្តនៅក្នុងឧស្សាហកម្មមួយនេះគឺមានរយៈពេល 20 នាទី។ ហើយថ្មីៗនេះគាត់បានក៏បានធ្វើការកត់ត្រាលើសេវាលាងរថយន្តរបស់គាត់ចំនួន 90 ករណី គិតជាមធ្យមគឺមានរយៈពេល 18 នាទី និងគម្លាតស្តង់ដារ 1.5 នាទី។

a. នៅកំរិតអាល់ហ្វាស្មើនឹង 5% តើមានភស្តុតាងបញ្ជាក់ថាសេវាកម្មលាងរថយន្តរបស់គាត់ប្រសើរជាងឧស្សាហកម្មដែរឬអត់? ចុះបើកំរិតអាល់ហ្វាស្មើនឹង 1% វិញ?

b. ចូលកំណត់ភីវែលយូ p-value ក្នុងចំលើយ (a) និងធ្វើការបកស្រាយ

9. អ្នកគ្រប់គ្រងធនាគារមួយបានប្រមូយចង់ធ្វើឲប្រសើរឡើយនូវសេវាកម្មរបស់ខ្លួនឲកាន់តែមានប្រសិទ្ធភាពជាងមុន។ គាត់បានធ្វើការប្រមូលរយៈពេលរងចាំ (គិតជានាទី) ដោយចៃដន្យទៅលើអតិថិជនចំនួន 16 នាក់នៅចន្លោះម៉ោង 11:30 ទៅ ម៉ោង 1:00។ ទិន្នន័យទាំងនោះបានបង្ហាញដូចខាងក្រោម៖

រយៈពេលរងចាំ	5.22	4.55	3.02	5.13	4.15	6.34	3.54	3.20
	3.50	5.10	3.38	5.12	6.46	6.19	3.79	6.20

a. នៅកំរិតអាល់ហ្វាស្មើនឹង 5% តើមានភស្តុតាងបញ្ជាក់ថារយៈពេលរងចាំតិចជាង 6នាទីដែរឬទេ?
b. ចុះបើនៅកំរិតអាល់ហ្វាស្មើនឹង 10% តើមានភស្តុតាងបញ្ជាក់ថារយៈពេលរងចាំតិចជាង 6នាទីដែរឬទេ?
c. តើល័ក្ខខ័ណ្ឌអ្វីខ្លះដែលត្រូវដាក់លើបំណែងចែករបស់ចំនួនសកល ដើម្បីឲការប្រើប្រាស់ t តេស្តនៅក្នុងចំលើយ (a) មានសុពលភាព?
d. ចូលបង្កើតគំនូសប្រអប់ដើម្បីវាយតម្លៃលើលក្ខខណ្ឌ normality ក្នុងចំលើយ (b)។

ជំពូកទី 8 ការធ្វើតេស្តលើសម្មតិកម្មលើប៉ាន់គំរូពីរ

(Two-sample Tests)

គោលបំណង

ក្នុងជំពូកនេះ អ្នកនិងធ្វើការសិក្សាអំពីរបៀបតេស្តសម្មតិកម្មទៅលើ៖

- តម្លៃមធ្យមនៃចំនួនសកលពីរឯករាជ្យ
- តម្លៃមធ្យមនៃចំនួនសកលពីរដែលទាក់ទងគ្នា
- សមាមាត្រនៃចំនួនសកលពីរឯករាជ្យ
- សមាមាត្រនៃចំនួនសកលពីរដែលទាក់ទងគ្នា
- វ៉ារីយ៉ង់ពីរ

8.1 ការប្រៀបធៀបតម្លៃមធ្យមនៃចំនួនសកលពីរដែលឯករាជ្យ

សឹងតែគ្រប់ករណីទាំងអស់ គឺយើងមិនដឹងអំពីគម្លាតស្តង់ដារបស់ចំនួនសកលនោះទេ ទោះបីជាចំនួនសកលនោះមានតែមួយ ឬពីរក៏ដោយ។ ករណីចំនួនសកលពីរ គឺយើងរិតតែមិនដឹងអំពីគម្លាតស្តង់ដារ ឬវ៉ារីយង់របស់វាទាំងពីរនោះស្មើគ្នាឬអត់ថែមទៀត។ ដើម្បីសិក្សាទៅលើករណីនេះ យើងត្រូវបែងចែកជាពីរករណី៖ ករណីទី1 គឺយើងអាចសន្មត់ថា វ៉ារីយ៉ង់របស់ចំនួនសកលទាំពីរស្មើគ្នា និងករណីទី2 គឺវ៉ារីយ៉ង់មិនស្មើគ្នា ហើយចុងក្រោយយើងធ្វើតេស្តបញ្ជាក់ថា វ៉ារីយ៉ង់ទាំងពីរស្មើគ្នាឬយ៉ាងណា។

8.1.1 តេស្តរកភាពខុសគ្នានៃតម្លៃសកលពីរដែលមានវ៉ារីយ៉ង់ស្មើគ្នា

ដើមសំរេចថា សម្មតិកម្មមោឃៈត្រូវបានបដិសេធដែរឬយ៉ាងណា យើងប្រើវិធីសាស្ត្របីដូចនៅក្នុងជំពូកទី6។ ករណីនេះយើងសន្មត់ថា ប៉ាន់គំរូគឺយកចេញពីចំនួនសកលពីរដោយចៃដន្យ ហើយចំនួនសកលទាំងពីរគឺឯករាជ្យពីគ្នា និងមានវ៉ារីយ៉ង់ស្មើគ្នា។ ក្នុងករណីបែបនេះ អ្នកអាចប្រើធីតេស្តដែលមានវ៉ារីយ៉ង់រួមបញ្ចូលគ្នា (pooled variance t-test) ដើម្បីកំណត់ថា តម្លៃមធ្យមរបស់ចំនួនសកលពីរនោះខុសគ្នាដែរឬយ៉ាងណា។

ឥឡូវយើងសន្មត់ថា តម្លៃមធ្យមនៃចំនួនសកលទី1 គឺ μ_1និងតម្លៃមធ្យមនៃចំនួនសកលទី2 គឺμ_2។ ដើម្បីតេស្តរកភាពខុសគ្នារវាងតម្លៃមធ្យមទាំងពីរ យើងអាចប្រើកន្ទុយសង្ខាង ឬ ម្ខាង (ខាងស្តាំ ឬឆ្វេង) ហើយសម្មតិកម្មទាំងនោះអាចសរសេរដូចខាងក្រោម៖

- តេស្តសម្មតិកម្មកន្ទុយសង្ខាង

  សម្មតិកម្មមោឃៈ $H_0: \mu_1 = \mu_2$ ឬ $H_0: \mu_1 - \mu_2 = 0$

  សម្មតិកម្មជំរើស $H_1: \mu_1 \neq \mu_2$ ឬ $H_1: \mu_1 - \mu_2 \neq 0$

- តេស្តសម្មតិកម្មកន្ទុយខាងឆ្វេង

  សម្មតិកម្មមោឃៈ $H_0: \mu_1 \geq \mu_2$ ឬ $H_0: \mu_1 - \mu_2 \geq 0$

  សម្មតិកម្មជំរើស $H_1: \mu_1 < \mu_2$ ឬ $H_1: \mu_1 - \mu_2 < 0$

- តេស្តសម្មតិកម្មកន្ទុយខាងស្តាំ

  សម្មតិកម្មមោឃៈ $H_0: \mu_1 \leq \mu_2$ ឬ $H_0: \mu_1 - \mu_2 \leq 0$

  សម្មតិកម្មជំរើស $H_1: \mu_1 > \mu_2$ ឬ $H_1: \mu_1 - \mu_2 > 0$

ដើម្បីតេស្តសម្មតិកម្មមោឃៈ យើងប្រើធីតេស្តដែលមានវ៉ារីយ៉ង់រួមបញ្ចូលគ្នា (pooled variance t-test) ដូចសមីការ (8.1)។

$$t - statisic = \frac{\overline{(X_1 - X_2)} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad (8.1)$$

ហើយ $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$

$S_p^2 =$ វ៉ារីយ៉ង់បញ្ចូលគ្នា

$\overline{X_1} =$ តម្លៃមធ្យមរបស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី1

$S_1^2 =$វ៉ារីយ៉ង់របស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី1

n_1=ប៉ាន់គំរូរបស់ចំនួនសកលទី1

$\overline{X_2} =$ តម្លៃមធ្យមរបស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី2

$S_2^2 =$វ៉ារីយ៉ង់របស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី2

n_2=ប៉ាន់គំរូរបស់ចំនួនសកលទី2

t-statistic មានបំណែងចែកអក្សរធី (t-distribution) ដែលមានដឺគ្រីហ្វ្រីឌឹម df= $n_1 + n_2 - 2$

សំគាល់៖ ក្នុងករណី $n_1 = n_2$ វ៉ារីយ៉ង់រួមបញ្ចូលគ្នា $S_p^2 = \frac{S_1^2 + S_2^2}{2}$

រីឯចន្លោះជឿជាក់សម្រាប់ $\mu_1 - \mu_2$គឺ៖

$$\overline{(X_1} - \overline{X_2}) \pm t_{\alpha/2}\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})} \qquad (8.2)$$

ឧទាហរណ៍ 8.1 អ្នកស្រាវជ្រាវម្នាក់ចង់ធ្វើការប្រៀបធៀបអត្រាប្រាក់ចំណេញរបស់សន្ទស្សន៍ផ្សារហ៊ុនពីរ គឺផ្សារ NYSE និងផ្សារ NASDAQ ដែលមានទិន្នន័យដូចក្នុងតារាង 8.1 ៖

តារាង 8.1 ទិន្នន័យផ្សារហ៊ុន NYSE និង NASDAQ

	ផ្សារ NYSE	ផ្សារ NASDAQ
ចំនួនប៉ាន់គំរូ	21	25
អត្រាប្រាក់ចំណេញ	3.27%	2.53%
គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ	1.30%	1.16%

ឥឡូវនេះយើងសន្មត់ថា ចំនួនសកលទាំងពីរគឺមានបំណែងចែកប្រហាក់ប្រហែលទៅនឹងបំណែងចែកធម្មតា ហើយមានវ៉ារីយ៉ង់របស់វាគឺស្មើគ្នា។ បើសិនជា α=5% ត្រូវបានគេប្រើប្រាស់ តើតម្លៃមធ្យមនៃអត្រាប្រាក់ចំណេញ របស់ចំនួនសកលទាំពីរស្មើគ្នាដែរឬទេ?

សំគាល់៖ រាល់ការធ្វើតេស្តលើសម្មតិកម្មគឺយើងអាចដើរតាមជំហ៊ានទាំង5ដូចដែលបានរៀបរាប់នៅក្នុងជំពូកទី7 ។ សម្រាប់ជំពូកនេះគឺអាចខុសខ្លះពីជំពូកមុនៗដោយសារតែភាពស្មុគ្គស្មាញនៃការគណនាតម្លៃតេស្ត។

ចំឡើយ:

i) ប្រើតេស្តស្ថិតិ

- ជំហានទី1: សរសេរសម្មតិកម្ម

  សម្មតិកម្មមោឃ: $H_0: \mu_1 = \mu_2$ ឬ $H_0: \mu_1 - \mu_2 = 0$

  សម្មតិកម្មជំរើស $H_1: \mu_1 \neq \mu_2$ ឬ $H_1: \mu_1 - \mu_2 \neq 0$

- ជំហានទី2: គណនាវ៉ារីយ៉ង់រួមបញ្ចូលគ្នារវាងចំនួនសកលទាំងពីរ

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{(21-1)1.3^2 + (25-1)1.16^2}{(21-1)+(25-1)} = 1.5021$$

- ជំហានទី3: បន្ទាប់មកជំនួសវ៉ារីយ៉ង់រួមបញ្ចូលគ្នារវាងចំនួនសកលទាំងពីរចូលក្នុងរូបមន្ត (8.1)

$$t - statisic = \frac{\overline{(X_1 - X_2)} - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(3.27-2.53)-0}{\sqrt{1.5021\left(\frac{1}{21}+\frac{1}{25}\right)}} = 2.040$$

- ជំហានទី4: រកតម្លៃស្ថិតិគោល

  យើងសន្មត់ថា α=0.05 និង ហើយ df= $n_1 + n_2 - 2$=21+25-2=44 នោះយើងបាន t-critical = 2.0154 (មើលតារាង t-distribution)

- ជំហាន**ទី5:** ធ្វើសេចក្តីសំរេចចិត្តៈ ដោយសារ t-statistic = 2.040 > t-critical = 2.0154 ដូច្នេះយើងបដិសេធ H_0នៅកំរិតα=0.05 ហើយសន្និដ្ឋានថា វាមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា អាត្រាប្រាក់ចំនេញនៅលើទីផ្សារទាំពីគឺមិនស្មើគ្នានោះទេ។

ii) ការប្រើតម្លៃភីវែលយូ p-value

ក្រៅពីប្រើ test-statistic ឥឡូវយើងប្រើ p-value ម្តង។ តម្លៃ t-statistic ដែលយើងគណនាខាងលើគឺ 2.040 និងមាន df= $n_1 + n_2 - 2$=21+25-2=44 តាម Excel Calculator យើងអាចរកតម្លៃ p-value = 0.047 ។ ដោយសារតម្លៃ p-value នេះតូចជាង *5%* ដូច្នេះយើងបដិសេធ H_0 ហើយរក្សាការសន្និដ្ឋាននៅដូចមុន។

iii) ការប្រើចន្លោះជឿជាក់ (confidence interval)

ក្រៅពីប្រើ test-statistic និង p-value ដើម្បីធ្វើសេចក្តីសំរេចចិត្ត ឥឡូវយើងប្រើចន្លោះជឿជាក់វិញម្តង តាមរយៈនៃការប្រើរូបមន្ត (8.2)។ ដើម្បីគណនា ចន្លោះជឿជាក់95%ទៅលើឧទាហរណ៍ 8.1 វិញម្តង។

$$\overline{(X_1 - X_2)} \pm t_{\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad (8.2)$$

$$(3.27 - 2.53) \pm 2.0154 * \sqrt{1.5021\left(\frac{1}{21}+\frac{1}{25}\right)}$$

$$0.0087 \leq \mu_1 - \mu_2 \leq 1.4710$$

ផលដកនៃចំនួនសកលទី1 និងទី2 គឺនៅចន្លោះ 0.0087 និង 1.4710 នៅចន្លោះ ពុំមានលេខសូន្យនោះទេ ដូចច្នេះយើងសន្និដ្ឋានថា នៅចន្លោះជឿជាក់ 95% អាត្រាប្រាក់ចំណេញនៅលើទីផ្សារទាំងពីរគឺមិនស្មើគ្នានោះទេ។ ដោយសារតែយើងយកតម្លៃមធ្យមនៃចំនួនសលទីមួយដកតម្លៃមធ្យមនៃចំនួនសលទីពីរ

ហើយយើងបានលទ្ធផលវិជ្ជមានទាំងសង្ខាង។ លទ្ធផលនេះអាចឲយើងដឹងបន្ថែមទៀតថា អាត្រាប្រាក់ចំនេញនៅលើទីមួយចំនេញច្រើនជាងផ្សារទីពីរ គឺ 95%នៅចន្លោះពី 0.0087% ដល់ 1.4710%។

8.1.2 ធីតេស្តរកភាពខុសគ្នារវាងតម្លៃមធ្យមនៃចំនួនសកលពីរដែលមានវ៉ារីយ៉ង់មិនស្មើគ្នា

ប្រសិនជាអ្នកមិនអាចធ្វើការសន្មត់ថា វ៉ារីយ៉ង់មិនស្មើគ្នានៃចំនួនសកលឯករាជ្យពីរ អ្នកមិនអាចប្រើវ៉ារីយ៉ង់រួមបញ្ចូលគ្នា (S_p^2) បាននោះទេ ប៉ុន្តែអ្នកត្រូវប្រើវ៉ារីយ៉ង់ដោយឡែករបស់ធីតេស្ត (separate-variance t-test) ដែលបង្កើតឡើងដោយលោក Battherwaite។ សមីការ (8.3) កំណត់តេស្តស្ថិតិសម្រាប់វ៉ារីយ៉ង់ដោយឡែករបស់ធីតេស្តដើម្បីរកភាពខុសគ្នារវាងតម្លៃមធ្យម។

$$t - statisic = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})}} \quad (8.3)$$

ហើយ $\overline{X_1}$ = តម្លៃមធ្យមរបស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី1

S_1^2 =វ៉ារីយ៉ង់របស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី1

n_1=ប៉ាន់គំរូរបស់ចំនួនសកលទី1

$\overline{X_2}$ = តម្លៃមធ្យមរបស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី2

S_2^2 =វ៉ារីយ៉ង់របស់ប៉ាន់គំរូដែលបានមកពីចំនួនសកលទី2

n_2=ប៉ាន់គំរូរបស់ចំនួនសកលទី2

ធីតេស្តសម្រាប់វ៉ារីយ៉ង់ដោយឡែកក្នុងសមីការ (8.3) ស្រដៀងធីតេស្តសម្រាប់វ៉ារីយ៉ង់រួមនៃសមីការ (8.1) ដែរ តែខុសគ្នាត្រង់ភាគបែងតែប៉ុណ្ណោះ ហើយសមីការ (8.3) មាន df ស្រដៀងទៅនឹងបំណែងចែកអក្សរធី (t-distribution) ដែលមាន df ជាចំនួនគត់តាមសមាមាត្រដូចខាងក្រោម៖

$$V = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{S_2^2}{n_2})^2}{n_2 - 1}}$$

ពេលនេះយើងពិនិត្យឧទាហរណ៍ 8.1 របស់យើងឡើងវិញ(ទិន្នន័យដូចក្នុងតារាង 8.1) ដោយសន្មត់ថា វ៉ារីយង់នៃអាត្រាប្រាក់ចំណេញនៅលើទីផ្សារទាំងពីរគឺមិនស្មើគ្នា។

តារាង 8.1 ទិន្នន័យផ្សារហ៊ុន NYSE និង NASDAQ

	ផ្សារ NYSE	ផ្សារ NASDAQ
ចំនួនប៉ាន់គំរូ	21	25

អត្រាប្រាក់ចំណេញ	3.27%	2.53%
គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ	1.30%	1.16%

បន្ទាប់មកយើងគណនា df ដែលប្រើជាមួយ ធីតេស្តសម្រាប់គណនាវ៉ារីយ៉ង់ដោយឡែក៖

$$V=\frac{(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1-1}+\frac{(\frac{S_2^2}{n_2})^2}{n_2-1}}=\frac{(\frac{1.30^2}{21}+\frac{1.16^2}{25})^2}{\frac{(\frac{1.30^2}{21})^2}{21-1}+\frac{(\frac{1.16^2}{25})^2}{25-1}}=\frac{(0.0805+0.0538)^2}{\frac{(0.0805)^2}{20}+\frac{(0.0538)^2}{24}}=\frac{0.0180}{0.00044}=40.48$$

បើ α= 0.05 ហើយ df=40។ តាមតារាង t-distribution យើងបានតម្លៃ t-critical = 2.0211

បន្ទាប់ទៀតគឺប្រើសមីការ (8.3) ដើម្បីគណនាធីតេស្ត៖

$$t-statistic=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)}}=\frac{(3.27-2.53)-0}{\sqrt{\left(\frac{1.30^2}{21}+\frac{1.16^2}{25}\right)}}=\frac{0.74}{0.1343}=2.0193$$

ដោយសារតម្លៃ t-statistic (2.0193) តូចជាងតម្លៃ t-critical (2.0211) ដូច្នេះ H_0 គឺខកខានមិនត្រូវបានបដិសេធ។ លទ្ធផលនេះគឺផ្ទុយទៅនឹងករណីវ៉ារីយ៉ង់រួមដែលបានគណនាខាងលើ។

ឥឡូវយើងធ្វើការផ្ទៀងផ្ទាត់លទ្ធផលរបស់យើងទៅនឹងលទ្ធផលរបស់ SPSS វិញម្តង។ តារាង 8.2 គឺជាលទ្ធផល t-test ដែលបានមកពីការប្រើកម្មវិធី SPSS។ កម្មវិធីនេះបានផ្តល់ឲយើងនូវលទ្ធផលទាំងពីរតែម្តង ទាំងវ៉ារីយ៉ង់ស្មើគ្នា (equal variance) និងវ៉ារីយ៉ង់មិនស្មើគ្នា (unequal variance)។

តារាង 8.2 Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Return (%)	Equal variances assumed	.849	.362	**2.040**	44	**.047**	.7399200	.3627905	**.0087638**	**1.4710762**
	Equal variances not assumed			**2.019**	40.570	**.050**	.7399200	.3664768	**-.0004331**	**1.4802731**

លទ្ធផលរបស់តាមតារាង (8.2) បញ្ជាក់ឲដឹងថា តេស្តទាំងពីរគឺខុសគ្នា បើវ៉ារីយ៉ង់ត្រូវបានសន្មត់ថាស្មើគ្នា គឺលទ្ធផលនៃតេស្តបង្ហាញថា អាត្រាប្រាក់ចំណេញរបស់ទីផ្សារទាំងពីរគឺខុសគ្នា ហើយបើវ៉ារីយ៉ង់ត្រូវបានសន្មត់ថាមិនស្មើគ្នា គឺលទ្ធផលនៃតេស្តបង្ហាញថា អាត្រាប្រាក់ចំណេញរបស់ទីផ្សារទាំងពីរគឺមិនខុសគ្នានោះទេ។ តើតេស្តមួយណាត្រឹមត្រូវ? ដើម្បីឆ្លើយនិងសំនួរនេះ យើងត្រូវធ្វើការតេស្តមួយទៀតគឺតេស្តដើម្បីឲ

ដឹងថា វ៉ារីយ៉ង់របស់ចំនួនសកលទាំងពីរស្មើគ្នាឬយ៉ាងណា ទើបយើងអាចរកចំឡើយមួយដ៏ត្រឹមត្រូវបាន។ ការតេស្តលើវ៉ារីយ៉ង់គឺត្រូវប្រើ F-test ដែលនឹងត្រូវបានបង្ហាញនៅខាងចុងជំពូកនេះ។

8.2 **ការប្រៀបធៀបតម្លៃមធ្យមរបស់ចំនួនសកលពីរដែលទាក់ទងគ្នា**

មានស្ថានភាពពីរដែលទាក់ទងជាមួយចំនួនសកលពីរដែលមិនឯករាជ្យពីគ្នា៖ ករណីទីមួយគឺវង្វាស់ដដែលៗ (repeated measures) គឺជាការប្រមូលប៉ាន់គំរូចេញពីសំណុំវត្ថុឬបុគ្គលដដែលៗ ដែលស្ថិតនៅលក្ខខណ្ឌ ឬពេលវេលាផ្សេងគ្នា ឧទាហរណ៍ គ្រូពេទ្យវាស់កំដៅអ្នកជំងឺប្រាំម៉ោងម្តងៗ និងករណីទីពីរ គឺទាក់ទងទៅនឹងទិន្នន័យជាគូ (paired sample) ឧទាហរណ៍ ការធ្វើតេស្តពិន្ទុគណិតវិទ្យា មុនចូលរៀននិងក្រោយចូលរៀន ឬវិធីផ្សាយពាណិជ្ជកម្មពីរផ្សេងគ្នាលើផលិតផលតែមួយ។ ក្នុងស្ថានភាពទាំងពីរនេះ អ្នកកំពង់ចាប់អារម្មណ៍លើការប្រៀបធៀបតម្លៃមធ្យមពីរដែលទាក់ទងគ្នា មិនមែនធ្វើការប្រៀបធៀបតម្លៃមធ្យមពីរដែលឯករាជ្យពីគ្នានោះទេ។

8.2.1 ធីតេស្តទិន្នន័យជាគូ

ធីតេស្តទិន្នន័យជាគូរ (paired sample t-test) ជូនកាលហៅថាធីតេស្តលើទិន្នន័យមិនឯករាជ្យ (dependent sample t-test) គឺជានីតិវិធីតេស្តស្ថិតិដើម្បីកំណត់ថាតម្លៃមធ្យមពីរខុសគ្នាឬយ៉ាងណា។ ធីតេស្តទិន្នន័យជាគូ តម្រូវឲ្យ ឬបុគ្គលដែលត្រូវបានសាកសួរនោះ ត្រូវធ្វើឡើងពីរដង ទើបចេញលទ្ធផលជាគូរ។

- សមីការ (8.4a) កំណត់តម្លៃខុសគ្នាជាគូទី ith (the paired difference) គឺ D_i៖

$$D_i = X_{1i} - X_{2i} \qquad (8.4a)$$

- តម្លៃខុសគ្នាគិតជាមធ្យមនៃចំនួនសកល μ_D គឺ៖

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} \qquad (8.4b)$$

- គម្លាតស្តង់ដាររបស់ប៉ាន់គំរូ៖

$$S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}} \qquad (8.4c)$$

ហើយ n=ចំនួនគូរនៃប៉ាន់គំរូ

- តេស្តស្ថិតិសម្រាប់តម្លៃ μ_D គឺ៖

$$t - statistic = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \qquad (8.4d)$$

ដែល t-statistic មាន df= n-1

- ចន្លោះជឿជាក់សម្រាប់តម្លៃ μ_D គឺ៖

$$\bar{D} \pm t_{\frac{\alpha}{2}} * \frac{S_D}{\sqrt{n}} \qquad (8.4e)$$

ឧទាហរណ៍ 8.2 ធីតេស្តទិន្នន័យជាគូ៖ ក្រុមហ៊ុនមួយបានបញ្ជូនបគ្គលិកផ្នែកលក់ប្រាំនាក់ឲ្យទៅចូលរួមកម្មវិធីវិកហ្វឹនមួយដើម្បីឲ្យលក់កាន់តែមានប្រសិទ្ធភាពជាងមុន ហើយក្រុមហ៊ុនបានធ្វើការកត់ត្រាចំនួនពាក្យបណ្តឹង មុននិងក្រោយចូលរួមកម្មវិធីហ្វឹកហ្វឺន ដូចក្នុងតារាង 8.3 ខាងក្រោម៖

តារាង 8.3 ចំនួនពាក្យបណ្តឹងរបស់បុគ្គលិកទាំងប្រាំ

	ចំនួនពាក្យបណ្តឹង	ខុសគ្នា=2-1

បុគ្គលិក	មុនហ្វឹកហ្វឺន (1)	ក្រោយហ្វឹកហ្វឺន (2)	(D_i)
A	6	4	-2
B	20	6	-14
C	3	2	-1
D	0	0	0
E	4	0	-4
សរុប			-21

ចំឡើយ:

i) ប្រើតេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម

$H_0: \mu_D = 0$
$H_1: \mu_D \neq 0$

- **ជំហានទី2:** គណនាតម្លៃខុសគ្នាគិតជាមធ្យមនៃចំនួនសកលៈ

$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{21}{5} = -4.2$$

- **ជំហានទី3:** គណនាគម្លាតស្តង់ដាររបស់ប៉ាន់គំរូៈ

$$S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(-2+4.2)^2 + \cdots + (-4+4.2)^2}{n-1}} = 5.67$$

- **ជំហានទី4:** គណនាតេស្តស្ថិតិសម្រាប់តម្លៃ μ_D គឺៈ

$$t - statistic = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} = \frac{-4.2 - 0}{\frac{5.67}{\sqrt{5}}} = -1.66$$

- **ជំហានទី5:** រកតម្លៃគោល

ឥឡូវសន្មត់ថា $\alpha = 5\%$ ហើយ df=n-1=4, តាមតារាង t-distribution យើងបាន t-critical= ± 2.7764 (two-tail)

- **ជំហានទី6:** សេចក្តីសំរេច

ដោយសារតម្លៃដាច់ខាតរបស់ t-test គឺស្មើនឹង 1.66 តូចជាងតម្លៃ t-critical (2.7764) ដូច្នេះនូវកំរិត$\alpha = 5\%$ $H_0: \mu_D = 0$ ខកខានមិនបានបដិសេធឡើយ ហើយយើងសន្និដ្ឋានថា ការហ្វឹកហ្វឺនមិនមានប្រសិទ្ធភាពទេ។

8.2.2 ប្រើតម្លៃ p-value

សូមរំលឹកថា ដើម្បីសំរេចថា H_0 ត្រូវបានបដិសេធ ឬយ៉ាងណា យើងមានវិធីសាស្ត្របីៈ ទីមួយគឺ ប្រើតេស្តស្ថិតិ ទីពីរគឺ p-value និងទីបីគឺ ប្រើចន្លោះជឿជាក់ ។ សម្រាប់ឧទាហរណ៍របស់យើង t-statistic=-1.66, df= តម្លៃ p-value =0.173(Sig. 2-tailed)>0.05 ដូច្នេះ H_0 គឺត្រូវបានខកខានមិនត្រូវបានបដិសេធនោះទេ ហើយការសន្និដ្ឋានរបស់យើងគឺនៅដដែល។

8.2.3 ប្រើចន្លោះជឿជាក់

ឥឡូវនេះយើងប្រើចន្លោះជឿជាក់វិញម្តង។ សម្រាប់ឧទាហរណ៍របស់យើង ចន្លោះជឿជាក់ 95% សម្រាប់តម្លៃ μ_D គឺ៖

$$\bar{D} \pm t_{\frac{\alpha}{2}} * \frac{S_D}{\sqrt{n}} = -4.2 \pm 2.7764\,(\frac{5.67}{\sqrt{5}})$$

$$-11.2401 \leq \mu_D \leq 2.8401$$

ដូច្នេះនៅចន្លោះជឿជាក់ 95% បានបញ្ជាក់ថា តម្លៃខុសគ្នានៃចំនួនពាក្យបណ្តឹងគឺស្ថិតនៅចន្លោះ -11.2401 ទៅ 2.8401។ ដោយសារចន្លោះជឿជាក់នេះមានលេខសូន្យ ដូច្នេះនូវកំរិត $\alpha = 5\%$ យើងខកខានមិនបានបដិសេធH_0 នោះឡើយ ហើយយើងសន្និដ្ឋានគឺនៅដដែល។

តារាង 8.4 ខាងក្រោមគឺជាលទ្ធផលតេស្តទិន្នន័យជាគូរបស់ SPSS។ សូមអ្នកធ្វើការផ្ទៀងផ្ទាត់តម្លៃ t-test, ចន្លោះជឿជាក់ និង p-value ទៅនឹងការគណនាខាងលើដោយខ្លួនឯង។

តារាង 8.4 Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 After - Before	-4.200	5.675	2.538	-11.246	2.846	-1.655	4	.173

8.3 ការប្រៀបធៀបតម្លៃខុសគ្នារបស់សមាមាត្រ

ជាញឹកញាប់ បើយើងចង់ធ្វើការប្រៀបធៀបតម្លៃខុសគ្នារវាងសមាមាត្រពីរ គឺយើងអាចប្រើតេស្តស្ថិតិពីរគឺ Z-test ឬ ឆាយស្វែតេស្ត ($\chi^2 - \text{test}$) ហើយលទ្ធផលរបស់តេស្តទាំងពីរគឺប្រហាក់ប្រហែលគ្នា។ សម្រាប់ពេលនេះ យើងនឹងសិក្សាតែ Z-test តែប៉ុណ្ណោះ។ Z-test សម្រាប់តម្លៃខុសគ្នារវាងសមាមាត្រពីរគឺត្រូវបានកំណត់នៅក្នុងសមីការ 8.7 ខាងក្រោម៖

$$\text{Z} - \text{statistic} = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad (8.7)$$

និង

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2};\ p_1 = \frac{X_1}{n_1}\ ;\ p_2 = \frac{X_2}{n_2}$$

ហើយ p_1 = សមាមាត្រនៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងប៉ាន់គំរូទីមួយ

X_1 = ចំនួននៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងប៉ាន់គំរូទីមួយ

n_1 =ចំនួនប៉ាន់គំរូទីមួយ

π_1 = សមាមាត្រនៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងចំនួនសកលទីមួយ

p_2 = សមាមាត្រនៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងប៉ាន់គំរូទីពីរ

X_2 = ចំនួននៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងប៉ាន់គំរូទីពីរ

n_2 =ចំនួនប៉ាន់គំរូទីពីរ

π_2 = សមាមាត្រនៃវត្ថុដែលចាប់អារម្មណ៍ក្នុងចំនួនសកលទីពីរ

$\bar{p}$=តម្លៃមធ្យមនៃសមាមាត្រទាំងពីររបស់ចំនួនសកលដែលចាប់អារម្មណ៍

Z-test មានបំណែងចែកធម្មតាដែលមានលក្ខណៈស្តង់ដារ (standard normal distribution) ។ ហើយការតេស្តនេះយើងអាចធ្វើឡើងដោយប្រើកន្ទុយសងខាង (two-tailed test) ឬកន្ទុយម្ខាង ជួនកាលកន្ទុយខាងស្តាំ ឬឆ្វេង (upper- or lower- tailed) ដូចបានរៀបរាប់ត្រង់ចំនុច 8.1 ហើយការជ្រើសរើសតេស្តខាងស្តាំ ឬឆ្វេង គឺអាស្រ័យលើសម្មតិកម្ម។

ឧទាហរណ៍ 8.3 ការស្ទង់មតិមួយត្រូវបានធ្វើឡើងដើម្បីចង់ដឹងថា សមាមាត្របុរសនិងស្ត្រីនឹងគាំទ្របេក្ខភាព A ខុសគ្នាឬយ៉ាងណា។ ការប្រមូលប៉ាន់គំរូដោយចៃដន្យបានបង្ហាញថា បុរស 36 នាក់ក្នុងចំណោម 72 នាក់ និងស្ត្រី 35នាក់ក្នុងចំណោម50នាក់ បានបោះឆ្នោតគាំទ្របេក្ខភាព A។ បើសិន α=5% តើសមាមាត្រទាំងពីរខុសគ្នាដែរឬទេ?

ចំឡើយ:

សូមរំលឹកថា ក្នុងការអនុវត្តជាក់ស្តែងដើម្បីសំរេចចិត្តបដិសេធ H_0 ឬមិនបដិសេធ ក្រៅពីធ្វើការប្រៀបធៀប Z-test ទៅនឹងZ-critical យើងយើងមានជំរើសពីរទៀតគឺតាមរយៈ តម្លៃ p-value និង ការគណនាចន្លោះជឿជាក់។

8.3.1 ការប្រើតេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម

  $H_0: \pi_1 = \pi_2$ or $H_0: \pi_1 - \pi_2 = 0$

  $H_1: \pi_1 \neq \pi_2$ or $H_1: \pi_1 - \pi_2 \neq 0$

- **ជំហានទី2:** គណនាសមាមាត្ររបស់ប៉ាន់គំរូ

$$p_1 = \frac{X_1}{n_1} = \frac{36}{72} = 0.50$$

$$p_2 = \frac{X_2}{n_2} = \frac{35}{50} = 0.70$$

- **ជំហានទី3:** គណនាតម្លៃមធ្យមនៃសមាមាត្រទាំងពីរ

  $$\bar{p} = \frac{X_1+X_2}{n_1+n_2} = \frac{(36+35)}{(72+50)} = 0.582$$

- **ជំហានទី4:** គណនា Z-statistic

  $$\text{Z-statistic}=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{(0.50-0.70)-(0)}{\sqrt{0.582(1-0.582)(\frac{1}{72}+\frac{1}{50})}} = -2.2026$$

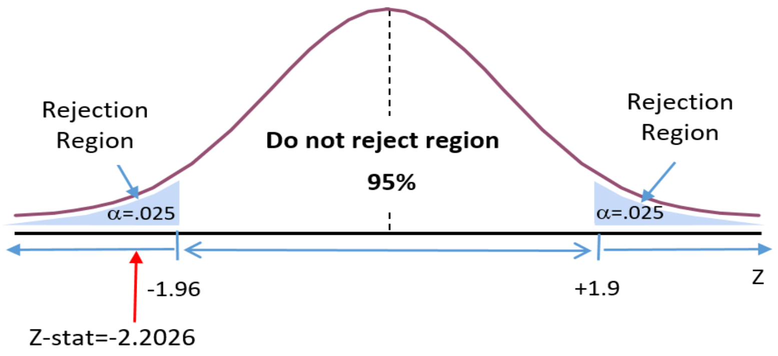
- **ជំហានទី5:** កំណត់តម្លៃ Z-critical

  នៅត្រង់ α=5% យើងបាន Z-critical=$\pm$ 1.96

- **ជំហានទី6:** ធ្វើសេចក្តីសំរេចចិត្ត

ដើម្បីធ្វើការសំរេចចិត្តនៅក្នុងវិធីសាស្ត្រនេះ យើងធ្វើការប្រៀបធៀបតម្លៃ Z-statistic ទៅនឹង Z-critical។ ក្នុងឧទាហរណ៍របស់យើង Z-statistic =-2.20 < Z-critical =-1.96 ដូច្នេះ H_0ត្រូវបានបដិសេធនៅកំរិត α=5% (មើលរូប 8.1) ហើយយើងសន្និដ្ឋានថាមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា មានភាពខុសគ្នារវាងសមាមាត្របុរសនិងស្ត្រីក្នុងការគាំទ្របេក្ខភាព A។

រូប 8.1 តម្លៃ Z-statistic នៃសមាមាត្ររបស់ស្ត្រីនិងបុរស



សូមបញ្ជាក់ថា SPSS មិនបានបង្កើត Z-test សម្រាប់តេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រនៃចំនួនសកលពីរនោះទេ តែ SPSS បានបង្កើត ឆាយឈ្វែតេស្តជំនួសវិញ។ ចំពោះតេស្តនេះយើងនឹងលើកមកសិក្សាម្តងទៀតនៅជំពូកក្រោយៗ។

8.3.2 ការប្រើ p-value

នៅខាងលើយើងបានគណនាតម្លៃតេស្តរួចមកហើយ គឺ Z-test=-2.2026 តាមរយៈតម្លៃ Z-test នេះយើងអាចរក p-value =0.0276។ ដោយសារ p=0.0276 (two-tailed) < 0.05 ដូច្នេះ H_0 ត្រូវបានបដិសេធនៅកំរិត α=5% ហើយការសន្និដ្ឋានគឺដូចមុន។

8.3.3 ការប្រើចន្លោះជឿជាក់

ចន្លោះជឿជាក់សម្រាប់សមាមាត្រខុសគ្នានៃចំនួនសកលពីរ $\pi_1 - \pi_2$ គឺត្រូវបានកំណត់ដោយរូបមន្ត 8.8៖

$$(p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \qquad (8.8)$$

ចន្លោះជឿជាក់ 95% សម្រាប់សមាមាត្រខុសគ្នានៃចំនួនសកលពីរក្នុងឧទាហរណ៍របស់យើងគឺ

$$(0.50 - 0.70) \pm 1.96\sqrt{\frac{0.50(1-0.50)}{72} + \frac{0.70(1-0.70)}{50}}$$

$$-0.20 \pm 1.96\ (0.08759)$$

$$-0.37167 \le \pi_1 - \pi_2 \le -0.02832$$

សម្រាប់ឧទាហរណ៍របស់យើង ចន្លោះជឿជាក់ 95% នៃសមាមាត្រខុសគ្នារបស់ចំនួនសកលគឺនៅចន្លោះ -0.37167 និង -0.02832 ក្នុងចន្លោះនេះមិនមានលេខសូន្យនោះទេ គឺអវិជ្ជមានទាំងសង្ខាង ដូច្នេះ H_0 ត្រូវបានបដិសេធនៅកំរិត α=5% ហើយយើងសន្និដ្ឋានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា សមាមាត្រនៃមនុស្សស្រីនិងមនុស្សប្រុសដែលបោះឆ្នោតគាំទ្របេក្ខភាព A គឺខុសគ្នា។

8.4 ការធ្វើតេស្តលើវ៉ារីយ៉ង់

ដូចដែលបានរៀបរាប់ខាងលើរួចមកហើយថា ការប្រៀបធៀបតម្លៃមធ្យមនៃចំនួនសកលពីរដែលឯករាជ្យពីគ្នា (two independent populations) គឺយើងប្រើធីតេស្ត (t-test) ហើយធីតេស្តនេះទៀតសោតគឺមានពីរប្រភេទ អាស្រ័យទៅនឹងការសន្មត់ថា វ៉ារីយ៉ង់របស់ចំនួនសកលទាំងពីរនោះស្មើគ្នាឬមិនស្មើគ្នា។ ដូច្នេះដើម្បីធ្វើការជ្រើសរើស ធីតេស្តណាមួយយកមកប្រើឲ្យបានត្រឹមត្រូវនោះ គឺយើងចាំបាច់ត្រូវធ្វើតេស្តលើវ៉ារីយ៉ង់របស់ចំនួនសកលទាំងនោះជាមុនសិន។

ជាញឹកញាប់ ផលធៀបនៃវ៉ារីយ៉ង់របស់ប៉ាន់គំរូ (S_1^2/S_2^2) ត្រូវបានគេធ្វើតេស្តដើម្បីកំណត់ភាពស្មើគ្នាឬមិនស្មើគ្នានៃវ៉ារីយ៉ង់។ S_1^2គឺជាវ៉ារីយ៉ង់ដែលបានមកពីប៉ាន់គំរូទីមួយហើយត្រូវបានសន្មត់ថាធំជាង S_2^2គឺជាវ៉ារីយ៉ង់ដែលបានមកពីប៉ាន់គំរូទីពីរ។ ផលធៀបនេះមានបំនែងចែកតាមបំណែងចែក អក្សរ F-distribution (មើលតារាង E.5)។ F-distribution មានចំនួន df ចំនួនពីរ គឺ df របស់ភាគយក (numerator) និង df របស់ភាគបែង (denominator)។ F-statistic សម្រាប់ផលធៀបនៃវ៉ារីយ៉ង់ពីរគឺត្រូវបានកំណត់នៅក្នុងសមីការ 8.9។

$$F_{stat} = \frac{S_1^2}{S_2^2} \qquad (8.9)$$

ហើយ S_1^2 = វ៉ារីយ៉ង់របស់ប៉ាន់គំរូទីមួយ

S_2^2 = វ៉ារីយ៉ង់របស់ប៉ាន់គំរូទីពីរ

n_1 =ចំនួនប៉ាន់គំរូទីមួយ

n_2 =ចំនួនប៉ាន់គំរូទីពីរ

n_{1-1} = ចំនួន df របស់ភាគយក

n_{2-1} = ចំនួន df របស់ភាគបែង

សម្មាតិកម្មរបស់ F-test គឺ

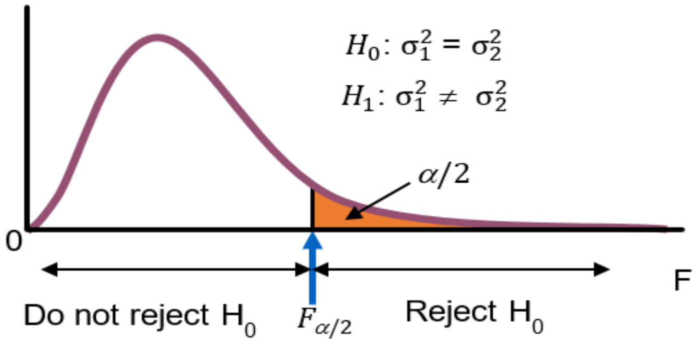
$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

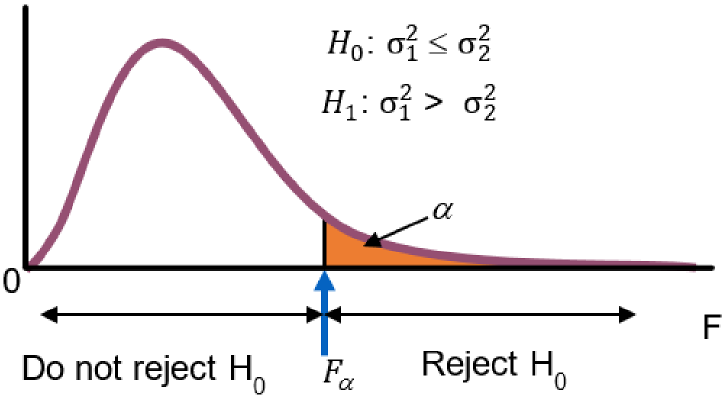
ដូចតេស្តផ្សេងៗទៀតដែរ យើងនឹងបដិសេធ H_0 ប្រសិន F-statistic > F-critical ហើយផ្ទុយមកវិញ បើ F-statistic < F-critical គឺយើងមិនបដិសេធ H_0 នោះឡើយ។

សូមបញ្ជាក់ថា ការធ្វើតេស្តនេះអាចតេស្តកន្ទុយសងខាង និងកន្ទុយម្ខាង គឺខាងស្តាំតែមួយគត់ (upper-tailed only) មិនអាចកន្ទុយខាងឆ្វេងនោះទេ ពីព្រោះ F-statistic=S_1^2/S_2^2 មិនអាចតូចជាង 1 ឬអវិជ្ជមានទេ (ពីព្រោះ $S_1^2 > S_2^2$ ជានិច្ច)។ គំនូសបង្ហាញក្នុងរូប 8.2a ជាការតេស្តដោយប្រើកន្ទុយសងខាង (ប្រើ α/2) ហើយ 8.2b ជាការតេស្តដោយប្រើកន្ទុយខាងស្តាំ (ប្រើ α)។

រូប 8.2a F-critical ដែលតេស្តកន្ទុយសងខាង



រូប 8.2b F-critical ដែលតេស្តកន្ទុយម្ខាង



ឥឡូវយើងធ្វើការពិនិត្យឧទាហរណ៍ 8.1ឡើងវិញ ដែលនៅក្នុងនោះយើងធ្វើការប្រៀបធៀបអាត្រាប្រាក់ចំណេញនៅលើទីផ្សារ NYSE និងផ្សារ NASDAQ ហើយយើងបានធ្វើតេស្តម្តងរួចមកហើយ ទាំងករណីសន្មត់ថាវ៉ារីយ៉ង់ស្មើគ្នា និងមិនស្មើគ្នា។ លទ្ធផលនៃការធ្វើតេស្តបញ្ជាក់ថា បើ α=5% ករណីវ៉ារីយ៉ង់ស្មើគ្នា H_0 ត្រូវបានបដិសេធ (t-statistic =2.040, p=0.047) ហើយបើ α=5% ករណីវ៉ារីយ៉ង់មិនស្មើគ្នា H_0 មិនត្រូវបានបដិសេធនោះទេ (t-statistic =2.019, p=0.050)។ សំនួរមួយអាចសួរថា តើតេស្តមួយណាត្រឹមត្រូវ? ចំលើយ គឺសូមរងចាំងការធ្វើតេស្តណាទៅលើវ៉ារីយ៉ង់សិន។

សូមរំលឹកថា នៅក្នុងឧទាហរណ៍ 8.1៖ ផ្សារ NYSE: n =21, $\bar{X}$ = 3.27, និង S_1=1.30%

ហើយផ្សារ NASDAQ: n =25, $\bar{X}$ = 2.53, និង S_2=1.16%។

8.4.1 ការប្រើតេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម:

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

- **ជំហានទី2:** គណនា F-statistic:

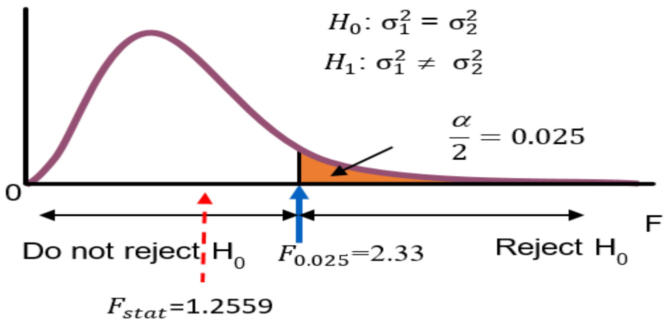
$$F_{stat} = \frac{S_1^2}{S_2^2} = \frac{1.30^2}{1.16^2} = \frac{1.69}{1.3456} = 1.2559$$

- **ជំហានទី3:** កំណត់តម្លៃសារៈសំខាន់និង តម្លៃស្ថិតិគោល

សន្មត់ថា α=5%, df របស់ភាគយក (numerator) = n_1-1 = 21-1=20, ហើយ df របស់ភាគបែង (denominator) = n_2-1 = 25-1=24 យោងតាមតារាង F-distribution (E.5) យើងបាន តម្លៃ F-critical, $F_{0.025,20,24} = 2.33$។

- **ជំហានទី4:** ការធ្វើសេចក្ដីសំរេចិត្ត

ដោយសារតែ F-statistic = 1.2559 < F-critical = 2.33 (រូប 8.3) គឺ H_0មិនត្រូវបាន បដិសេធ ហើយយើងសន្និដ្ឋានថា មិនមានភស្តុតាងគ្រាប់គ្រាន់បញ្ជាក់ថា វ៉ារីយ៉ង់ទាំងពីរខុសគ្នានោះទេ ហើយ t-statistic =2.040 (p=0.047) គឺជាតេស្តដែលត្រឹមត្រូវ (សូមមើលតារាង 8.2 និងការសន្និដ្ឋាននៅត្រង់ចំនុច 8.1.1)។



8.4.2 ការប្រើ p-value

បន្ទាប់ពីបានតម្លៃ F-statistic រួចហើយ យើងអាចរក តម្លៃ p-value របស់ F-statistic នោះ។ សូមរំលឹកថា p-value គឺជាប្រូបាប៊ីលីតេដែលអាចបដិសេធសម្មាតិកម្មមោឃៈ ហើយមានទីតាំងស្ថិតនៅខាងស្ដាំងដៃរបស់ F-statistic។ វិធីសាស្រ្តងាយស្រួលក្នុងការរកតម្លៃ p-value គឺយើងប្រើ Excel Calculator (អ្នកអាចរកបានតាមរយៈ Google)។ ក្នុងឧទាហរណ៍របស់យើង F-statistic=1.2559, degree freedom (df) របស់ភាគយកស្មើនឹង 20 និង df របស់ភាគបែងស្មើនឹង 24 ប្រើ Excel Calculator យើងទទួលបាន p-value =0.2944 > 0.05។ ដូច្នេះ H_0: ត្រូវបានបដិសេធ។

សេចក្ដីសង្ខេប

ជំពូកនេះបានណែនាំអោយអ្នកស្គាល់អំពីវិធីសាស្ត្រផ្សេងៗនៃការប្រៀបធៀបប៉ាន់គំរូពីរ។ ក្នុងករណីដែលប៉ាន់គំរូពីរឯករាជ្យពីគ្នា ហើយមានវ៉ារីយ៉ង់ស្មើគ្នា អ្នកអាចប្រើធីតេស្តដោយរួមបញ្ចូលវ៉ារីយ៉ង់ទាំងពីរចូលគ្នា (pooled variance) ប្រើចន្លោះជឿជាក់ និង p-value ជាដើម។ ប៉ុន្តែក្នុងករណីដែលវ៉ារីយ៉ង់មិនស្មើគ្នាវិញ នីតិវិធីនៃការប្រៀបធៀបនៅរក្សាដដែល ក៏ប៉ុន្តែរូបមន្តធីតេសនិង df មានការកែប្រែបន្តិចបន្តួច។ ក្រៅពីនេះអ្នកក៏បានសិក្សាអំពីការធ្វើតេស្តលើប៉ាន់គំរូពីរដែលជាប់ទាក់ទងគ្នា ឬការប្រៀបធៀបជាគូ និងសមាមាត្រផងដែរ។ ទីបំផុតអ្នកបានស្គាល់អំពីវិធីសាស្ត្រក្នុងការធ្វើតេស្តដើម្បីកំណត់ថា វ៉ារីយ៉ង់របស់ប៉ាន់គំរូពីរស្មើគ្នាឬអត់ តេស្តនេះធ្វើឡើងក្នុងគោលបំណងដើម្បីកំណត់ថា តេស្តមួយណាដែលត្រូវប្រើ៖ តេស្តដែលមានវ៉ារីយ៉ង់ស្មើគ្នា ឬមិនស្មើគ្នា។

សំនួរនិងលំហាត់

I. សំនួរ

1. ចូលសរសេរនិតិវិធីលើការធ្វើតេស្តលើសម្មតិកម្ម
2. តើនៅក្នុងល័ក្ខខ័ណ្ឌអ្វីខ្លះដែលវ៉ារីយ៉ង់រួម t តេស្តត្រូវបានប្រើប្រាស់ដើម្បីរកតម្លៃខុសគ្នានៃរមធ្យមពីររបស់ចំនួនសកលដែលឯករាជ្យពីគ្នា ?
3. តើនៅក្រោមលល័ក្ខខ័ណ្ឌអ្វីខ្លះដែលអ្នកគួរប្រើ F តេស្ត ដើម្បីរកភាពខុសគ្នានៃវ៉ារីយ៉ង់បស់ចំនួនសកលពីរឯករាជ្យ ?
4. អ្វីខ្លះជាលក្ខណះខុសគ្នារវាងចំនួនសកលពីរឯករាជ្យនិងចំនួនសកលពីរទាក់ទងគ្នា ?
5. អ្វីខ្លះជាលក្ខណះខុសគ្នារវាងរង្វាស់ច្រំដែលនិងរង្វាស់បន្ស៊ីឬមុននិងក្រោយ ?
6. ពេលដែលអ្នកមានចំនួនសកលពីរឯករាជ្យពីគ្នា ចូលពន្យល់លក្ខណះដូចគ្នានិងខុសគ្នារវាងការតេស្តសម្មតិកម្មនិងការតេស្តដោយប្រើចន្លោះជឿជាក់ដើម្បីរកតម្លៃមធ្យមរបស់ចំនួនសកលពីរខុសគ្នា ?
7. តើនៅក្រោមលក្ខខណ្ឌអ្វីខ្លះដែលអ្នកគួរប្រើ t តេស្តជាគូរដើម្បីរកតម្លៃមធ្យមពីរខុសគ្នារបស់ចំនួនសកលពីរដែលជាប់ទាក់ទងគ្នា ?

II. លំហាត់

8. ថ្នាក់ដឹកនាំធនាគារចង់ធ្វើការកែលំអរគុណភាពសេវារបស់ធនាគារ។ គាត់បានធ្វើការកត់ត្រារយះពេលរងចាំ (គិតជានាទី) របស់អតិថិជនចំនួន16 នាក់ នៅសាខាទីមួយ ដូចខាងក្រោម៖

4.21	5.55	3.02	5.13	4.30	2.34	3.54	3.20
4.50	6.10	0.38	5.12	6.46	6.19	3.39	5.30

ហើយគាត់ក៏ធ្វើការកត់ត្រារយះពេលរងចាំរបស់របស់អតិថិជន 16នាក់នៅសាខាទីពីរៈ

9.30	5.40	8.02	5.79	8.13	3.82	8.01	8.35
10.49	6.18	5.24	4.08	6.17	9.31	5.47	7.12

a. សន្មត់ថាវ៉ារីយ៉ង់នៃទិន្នន័យរបស់ធនាគារពីរស្មើគ្នា តើមានភស្តុតាងបញ្ជាក់ថារយៈពេលរង់ចាំរបស់សាខានៃធនាគារទាំងពីរខុសគ្នាដែរទេ?

b. ចូលគណនា p-value ក្នុងសំនួរ (a) និងធ្វើការបកស្រាយអត្ថន័យរបស់វា

c. ក្រៅពីសន្មត់ថាវ៉ារីយ៉ង់នៃទិន្នន័យរបស់ធនាគារពីរស្មើគ្នា តើមានល័ក្ខខ័ណ្ឌអ្វីបន្ថែមទៀតដែលចាំបាច់នៅក្នុងការធ្វើតេស្តសំនួរ (a)?

d. ចូលគណនាចន្លោះជឿជាក់ 95% និងធ្វើការបកស្រាយរយៈពេលរង់ចាំនៃសាខាទាំងពីររបស់ធនាគារ

9. អ្នកជំនាញខាងតុបតែងចំនួន 9 រូបត្រូវបានអញ្ជើញចូលរួមក្នុងការវាយតម្លៃទៅលើសោភ័ណភាពហាងកាហ្វេពីរកន្លែង ដោយឲ្យពិន្ទុចាប់ពី 1 ដល់ 10។ ទិន្នន័យទាំងនោះមានដូចតទៅ៖

អ្នកជំនាញ	1	2	3	4	5	6	7	8	9
A	7	7	8	6	6	8	9	6	7
B	6	7	9	6	7	9	10	7	8

a. នៅកំរិតអាល់ហ្វា 0.05 តើមានភស្តុតាងបញជាក់ថា ហាងកាហ្វេទាំងពីរមានពិន្ទុខុសគ្នាឬទេ?

b. តើត្រូវមានលក្ខខណ្ឌអ្វីខ្លះដែលតម្រូវឲមាន ក្នុងការធ្វើតេស្តប្រភេទនេះ?

c. ចូលគណនា p-value ក្នុងសំនួរ (a) និងធ្វើការបកស្រាយអត្ថន័យរបស់វា

d. ចូលគណនាចន្លោះជឿជាក់ 95% និងធ្វើការបកស្រាយពិន្ទុខុសគ្នារបស់ហាងកាហ្វេទាំងពីរ។

10. ការស្ទង់មតិមួយត្រូវបានធ្វើទៅលើអ្នកទិញទំនិញតាមបណ្តាញអ៊ីនតឺណែតចំនួន 400 នាក់ដែលទាក់ទងជាមួយឥរិយាបទអតិថិជន។ ក្នុងចំណោមនោះ មានសំនួរមួយត្រូវបានសួរ "តើអ្នកមានការពេញចិត្តក្នុងការទិញទំនិញតាមបណ្តាញអ៊ីនតឺណែតដែរឬទេ?" ក្នុងចំណោមបុរស 150 នាក់ មាន 100 នាក់ឆ្លើយថា "ពេញចិត្ត" ហើយក្នុងចំណោមស្ត្រី 250 នាក់ មាន 200 នាក់បានឆ្លើយថា "ពេញចិត្ត"។

a. នៅកំរិតអាល់ហ្វា 0.01 តើមានភស្តុតាងបញ្ជាក់ថា សមាមាត្រ"ពេញចិត្ត" របស់បុរសខុសគ្នាជាមួយនឹងស្ត្រីដែរឬទេ?

b. ចូលគណនា p-value ក្នុងសំនួរ (a) និងធ្វើការបកស្រាយអត្ថន័យរបស់វា

c. ចូលគណនាចន្លោះជឿជាក់ 99% និងធ្វើការបកស្រាយសមាមាត្រ"ពេញចិត្ត" ខុសគ្នារវាងបុរសនិងស្ត្រីក្នុងការកទិញទំនិញតាមបណ្តាញអ៊ីនតឺណែត។

11. សិប្បកម្មផលិតសណ្តែកដីកំប៉ុងមួយកន្លែងមានខ្សែសង្វាក់ផលិតកម្មពីរ A និង B។ អ្នកគ្រប់គ្រងសង្វាក់ផលិតកម្មបានកំណត់ថា ទម្ងន់សណ្តែកដីគិតជាមធ្យមនៅក្នុងកំប៉ុងនីមួយៗគឺ 200ក្រាម។ ថ្មីៗនេះគាត់មានការសង្ស័យថា បម្រែបម្រួលនៃការច្រកកំប៉ុងអាចខុសគ្នា ហើយគាត់បានប្រមូលប៉ាន់គំរូមួយចំនួនដូចខាងក្រោម៖

	ចង្វាក់ A	ចង្វាក់ B

ទម្ងន់មធ្យម	200.5 ក្រាម	198.8 ក្រាម
គម្លាតស្តង់ដារ	0.40	0.16
n	14	20

a. នៅកំរិតអាល់ហ្វា 0.05 តើមានភស្តុតាងបញ្ជាក់ថា វ៉ារីយ៉ង់របស់សង្វាក់ផលិតកម្ម A ធំជាងវ៉ារីយ៉ង់របស់សង្វាក់ផលិតកម្ម B?

b. ចូលគណនា p-value ក្នុងសំនួរ (a) និងធ្វើការបកស្រាយអត្ថន័យរបស់វា

c. តើលក្ខខណ្ឌចាំបាច់អ្វីខ្លះដែលតម្រូវឲមានក្នុងចំនួនសកលទាំងពីរ ដើម្បីធ្វើ F តេស្តនៅក្នុងសំនួរ (a) មានសុពលភាព?

12. ខាងក្រោមនេះគឺជាទិន្នន័យដែលទាក់ទងជាមួយរយៈពេល (គិតជាម៉ោង) ប្រើប្រាស់ថ្មពីរប្រភេទ A និង B ទៅលើកុំព្យូទ័រប្រភេទតែមួយ មុននឹងត្រូវបានសាកថ្មជាថ្មី (សន្មត់ថាលក្ខខណ្ឌនៃការប្រើគឺដូចគ្នា)។ ប៉ាន់គំរូថ្ម A និង B គឺ 18 ស្មើគ្នា។

A	3.35	3.10	2.50	2.45	2.50	2.00	3.30	2.40	2.40
	2.35	2.20	2.55	2.10	3.15	2.34	2.40	2.10	2.05
B	3.20	3.30	2.50	2.40	3.50	2.40	3.30	2.20	3.15
	2.30	2.50	2.55	3.20	3.25	2.00	2.150	3.00	2.45

a. នៅកំរិតអាល់ហ្វា 0.05 តើមានភស្តុតាងបញ្ជាក់ថា វ៉ារីយ៉ង់របស់របស់ថ្មទាំងពីរខុសគ្នាឬទេ?

b. ចូលគណនា p-value ក្នុងសំនួរ (a) និងធ្វើការបកស្រាយអត្ថន័យរបស់វា

c. តើលក្ខខណ្ឌចាំបាច់អ្វីខ្លះដែលតម្រូវឲមានក្នុងចំនួនសកលទាំងពីរ ដើម្បីធ្វើ F តេស្តនៅក្នុងសំនួរ (a) មានសុពលភាព?

d. ផ្អែកលើចំលើយ (a) តើ t ប្រភេទណាមួយដែលនិងយកមកប្រើក្នុងការប្រៀបធៀបរយៈពេលប្រើប្រាស់ជាមធ្យមរបស់ថ្មទាំងពីរ?

ជំពូកទី 9

ឆាយស្វែរតេស្ត និងតេស្តដទៃទៀតដែលមិនមែនជាប៉ារ៉ាម៉ែត្រតេស្ត

(Chi-square Test and Other Non-parametric Tests)

គោលបំណង

ជំពូកនេះអ្នកនឹងសិក្សាអំពី

- របៀបប្រើឆាយស្វែរតេស្តលើសមាមាត្រ
- ឆាយស្វែរតេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរ និងលើសពីពីរ
- ឆាយស្វែរតេស្តរកភាពឯករាជ្យ
- របៀបប្រើម៉ាកណេម៉ាតេស្ត
- វីកូសុនតេស្តលើប៉ាន់គំរូពីរឯករាជ្យ
- គ្រូស្កាល់វ៉ាលីសតេស្ត

9.1 ឆាយស្វែរតេស្តចំពោះភាពអនុលោមតាម

ឆាយស្វែរតេស្តចំពោះភាពអនុលោមតាម ឬភាពស័ក្តិសមទៅនឹងទម្រង់ណាមួយ (Chi-square test for goodness of fit) គឺត្រូវបានគេប្រើប្រាស់សម្រាប់តេស្តដើម្បីចង់ដឹងថា បំណែងចែកហ្វ្រេកង់មួយអនុលោមតាម ឬគោរពតាមទម្រង់ណាមួយជាក់លាក់ ឬយ៉ាងណា។ ឧទាហរណ៍ ផលិតផលបានឆ្លើយតមទៅនឹងតម្រូវការរបស់អតិថិជនដែរឬអត់ ការអនុវត្តន៍ការងារបានអនុលោមតាម ឬគោរពតាមគោលនយោបាយ ឬផែនការដែរឬអត់ ។ ឆាយស្វែរតេស្តសម្រាប់ភាពអនុលោមតាម ឬភាពសំស្របតាមត្រូវបានកំណត់នៅក្នុងសមីការ 9.1។

$$\chi^2_{Stat} = \sum \frac{(O-E)^2}{E} \qquad (9.1)$$

ហើយ $O =$ ហ្វ្រេកង់ដែលបានអង្កេត (observed frequency)

E= ហ្វ្រេកង់រំពឹងទុក (expected frequency): E = $n * p_i$

$n =$ ចំនួនប៉ាន់គំរូ

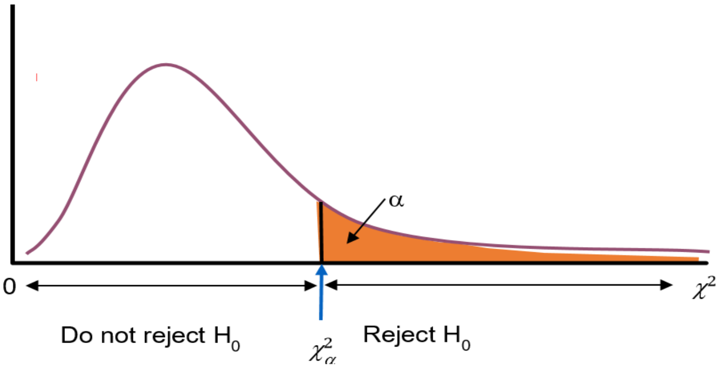
$p_i =$ ប្រូបាប៊ីលីតេនៃអថេរប្រភេទនីមួយៗដែលបានបញ្ជាក់ដោយ H_0

χ^2_{Stat}មានទម្រង់ជាបំណែងចែកប្រហាក់ប្រហែលទៅនឹង χ^2-distribution ដែលមាន degree of freedom, df= ចំនួនអថេរប្រភេទដកមួយ (number of category variable -1) ។ បើχ^2_{Stat} > χ^2-critical នោះ H_0 នឹងត្រូវបានបដិសេធ ហើយផ្ទុយមកវិញ បើតូចជាង គឺH_0នឹងមិនត្រូវបានបដិសេធនោះទេ (មើលរូប 9.1)។

កាសន្មត់នៃការប្រើឆាយឈ្វេតេស្តភាពអនុលោមតាម ឬភាពសំស្របតាមទម្រង់អ្វីមួយគឺ៖

- ការជ្រើសរើសប៉ាន់គំរូត្រូវធ្វើឡើងតាមលក្ខណៈចៃដន្យ
- ប្រេកង់រំពឹងទុកនៃអថេរប្រភេទនីមួយៗត្រូវធំជាង 5

រូប 9.1 តំបន់បដិសេធនិងមិនបដិសេធក្នុងការប្រើឆាយឈ្វេដើម្បីតេស្តសមាមាត្រ



ឧទាហរណ៍ 9.1 គោលនយោបាយរបស់ធនាគារជាតិកម្ពុជាដែលដាក់លើធនាគារពាណិជ្ជឲអនុវត្តតាមគឺ 60% នៃការផ្តល់កម្ចីទៅលើអជីវកម្ម 30% ទៅលើបុគ្គល និង 10% ទៅលើកសិកម្ម។ ដើម្បីវាយតម្លៃថា ធនាគារពាណិជ្ជមួយគោរពគោលនយោបាយនេះដែរឬអត់ សាវនករម្នាក់បានចុះទៅប្រមូលប៉ាន់គំរូពីធនាគារនោះដោយចៃដន្យទៅលើកម្ចីចុងក្រោយនេះចំនួន85 ហើយនៅក្នុងចំណោមកម្ចីទាំង 85នោះ គាត់បានរកឃើញថា កម្ចីចំនួន 62ត្រូវបានផ្តល់ឲអាជីវកម្ម កម្ចីចំនួន 13បានផ្តល់ឲបុគ្គល និងកម្ចីចំនួន 10បានផ្តល់ឲ្យកសិករ។ នៅកំរិត α=0.05 តើធនាគារពាណិជ្ជមួយនេះបានគោរព ឬដើមតាមគោលនយោបាយរបស់ធនាគារជាតិដែរឬទេ?

ចំឡើយ៖

i) តេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម

H_0: គោលនយោបាយរបស់ធនាគារជាតិត្រូវបានគោរព

H_1: គោលនយោបាយរបស់ធនាគារជាតិមិនត្រូវបានគោរព

- **ជំហានទី2:** រកតម្លៃ $\chi^2 - critical$, បើ α= 0.05 ហើយចំនួន df= ចំនួនអថេរប្រភេទ-1 ឬ 3-1=2 តាមតារាង χ^2-distribution យើងបាន $\chi^2_{0.05,2} = 5.991$
- **ជំហានទី3:** គណនាតម្លៃរំពឹងទុក E = $n * p_i$

តម្លៃរំពឹងទុករបស់កម្ចីអជីវកម្ម: E = 85 x 0.60 = 51

តម្លៃរំពឹងទុករបស់កម្ចីបុគ្គល: E = 85 x 0.30 = 25.5

តម្លៃរំពឹងទុករបស់កម្ចីកសិករ៖ E = 85 x 0.10 = 8.5

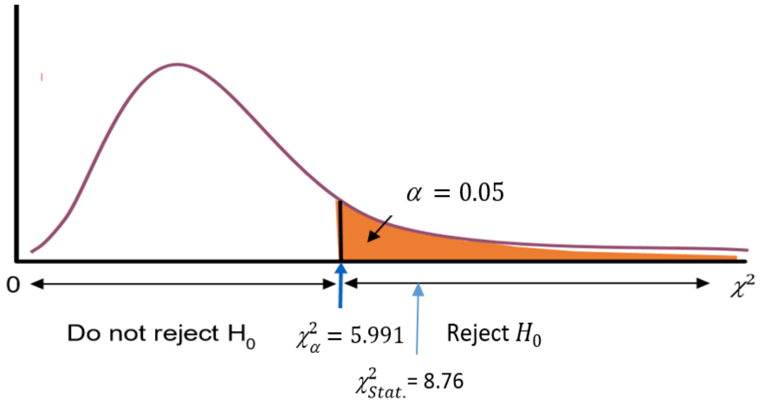
- **ជំហានទី4:** គណនា$\chi^2_{Stat.}$ ដោយប្រើរូបមន្ត (9.1): $\chi^2_{Stat} = \sum\frac{(O-E)^2}{E}$

$\chi^2_{Stat.} = \frac{(62-51)^2+}{51} + \frac{(13-25.5)^2}{25.5} + \frac{(10-8.5)^2}{8.5} = 8.76$ (p=0.012)

- **ជំហានទី5:** ការធ្វើសេចក្តីសំរេចចិត្ត

ដោយ $\chi^2_{Stat.}$=8.76 > $\chi^2_{0.05,2}$ = 5.991 ដូច្នេះ H_0 ត្រូវបានបដិសេធ (មើលរូប 9.2)

រូប 9.2 បំណែងចែកឆាយស្ែចំពោះការតេស្តលើគោលនយោបានរបស់ធនាគារពាណិជ្ជ



- **ជំហានទី6:** សន្និដ្ឋាន

នៅកំរិត α= 0.05 សវនករមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ការអនុវត្តន៍របស់ធនាគារពាណិជ្ជនេះ មិន គោរពតាមគោលនយោបានរបស់ធនាគាជាតិនោះឡើយ។

ii) ការប្រើភីវែលយូ (p-value)

បន្ទាប់ពីបានតម្លៃឆាយស្ែតេស្តរួច យើងអាចរកតម្លៃ ភីវែលយូ ។ $\chi^2_{Stat.}$=8.76 ហើយមាន df=2 តាមរយៈ Excel យើងបាន p-value =0.012 < 0.05 ដូច្នេះ H_0 ត្រូវបានបដិសេធ ហើយយើងអាចសន្និដ្ឋាន ដូចមុន។

iii) ការប្រើកម្មវិធីសូវែ(Software)

តារាង 9.1 គឺជាលទ្ធផលឆាយស្ែតេស្តរបស់ SPSS នៅក្នុងឧទាហរណ៍របស់យើង។ សូមអ្នកអានធ្វើការផ្ទៀងផ្ទាត់តម្លៃឆាយស្ែតេស្តនិង p-value នៅក្នុងតារាងនេះដោយខ្លួនឯង។

តារាង 9.1 លទ្ធផលនៃឆាយស្វែតេស្តរបស់ SPSS

Test Statistics

	Loan
Chi-Square	8.765[a]
df	2
Asymp. Sig.	.012

a. 0 cells (0.0%) have expected frequencies less than 5.

The minimum expected cell frequency is 8.5.

9.2 ឆាយស្កែតេស្តសម្រាប់តារាងខ្វែង

9.2.1 ឆាយឈ្វេតេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរ

នៅក្នុងជំពូកមុន អ្នកបានប្រើ Z-test ដើម្បីតេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរ តែសម្រាប់ពេលនេះគឺអ្នកនិងសិក្សាអំពីឆាយឈ្វេតេស្តរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរវិញម្តង។ លទ្ធផលនៃតេស្តទាំងពីគឺស្រដៀងគ្នា។

ឆាយឈ្វេតេស្តគឺផ្អែកលើតារាងខ្វែង (cross or contingency table) ដែលអ្នកបានសិក្សានៅជំពូកទីពីរ។ សូមរំលឹកថា តារាងខ្វែងគឺទាក់ទងទៅនឹងអថេរប្រភេទ (categorical variables) ហើយប្រសិនជាមានអថេរប្រភេទពីរនោះ យើងនឹងបានតារាងខ្វែង 2 x 2 (ជួរដេកពីរ និងជួរឈរពីរ) ដូចបានបង្ហាញនៅក្នុងតារាង 9.2។

តារាង 9.2 ទម្រង់នៃតារាងខ្វែង 2 x 2

	ក្រុម		
អថេរ	1	2	សរុប
វត្ថុដែលចាប់អារម្មណ៍	X_1	X_2	X
វត្ថុដែលមិនចាប់អារម្មណ៍	$n_1 - X_1$	$n_2 - X_2$	n-X
សរុប	n_1	n_2	n

ហើយ X_1 = ចំនួនវត្ថុដែលចាប់អារម្មណ៍នៅក្នុងក្រុមទីមួយ

X_2= ចំនួនវត្ថុដែលចាប់អារម្មណ៍នៅក្នុងក្រុមទីពីរ

$n_1 - X_1$= ចំនួនវត្ថុដែលមិនចាប់អារម្មណ៍នៅក្នុងក្រុមទីមួយ

$n_2 - X_2$= ចំនួនវត្ថុដែលមិនចាប់អារម្មណ៍នៅក្នុងក្រុមទីពីរ

X = $X_1 + X_2$ = ចំនួនវត្ថុដែលចាប់អារម្មណ៍សរុប

n_1 = ចំនួនប៉ាន់គំរូនៃក្រុមទីមួយ

n_2= ចំនួនប៉ាន់គំរូនៃក្រុមទីមួយ

$n = n_1 + n_2$ = ចំនួនប៉ាន់គំរូសរុប

ដើម្បីតេស្តលើសមាមាត្រពីរ យើងត្រឡប់ទៅឧទាហរណ៍ 8.3 នៅក្នុងជំពូកទី8 ដែលទាក់ទងទៅនិងសមាមាត្របុរសនិងស្ត្រីគាំទ្របេក្ខភាព A។ ហើយឥឡូវយើងរៀបចំទិន្នន័យទាំអនោះជាតារាងខ្វែង 2 x 2 ដូចខាងក្រោម៖

តារាង 9.3 សមាមាត្របុរសនិងស្រ្តីគាំទ្របេក្ខភាព A

	ក្រុម		
អរថេរ	បុរស	ស្រ្តី	សរុប
គាំទ្រ	36	35	71
មិនគាំទ្រ	36	15	51
សរុប	72	50	122

ដើម្បីកំណត់ថា សមាមាត្ររបស់បុរសដែលបានគាំទ្របេក្ខភាព A ខុសពីសមាមាត្រស្រ្តីដែលបានគាំទ្របេក្ខភាព A នោះ យើងប្រើឆាយឈ្វែតេស្តចំពោះភាពខុសគ្នារវាងសមាមាត្រពីរ (Chi-square test for difference between two proportion)។ សម្មតិកម្មមោឃៈរបស់តេស្តនេះបានកំណត់ថា H_0: $\pi_1 = \pi_2$ ហើយសមីការរបស់តេស្តនេះត្រូវបានកំណត់នៅក្នុងសមីការ 9.2។

$$\chi^2_{Stat} = \sum_{all\ cell} \frac{(f_0 - f_e)^2}{f_e} \qquad (9.2)$$

ហើយ f_0 = ចំនួនហ្វ្រេកង់អង្កេត (observed frequency) នៅក្នុងសែល (cell) នៃតារាងខ្វែង

f_e = ចំនួនហ្វ្រេកង់ដែលបានរំពឹងទុកក្នុងសែល (cell) នៃតារាងខ្វែង

χ^2_{Stat}មានទម្រង់ជាបំណែងចែកប្រហាក់ប្រហែល χ^2-distribution ដែលមាន df នៅក្នុងតារាងខ្វែងស្មើទៅនឹងចំនួនជួរដេកដកមួយ គុណចំនួនជួរឈរដកមួយ (number of row -1 times number of column -1)។

បើχ^2_{Stat} ធំជាង χ^2-critical នោះ H_0: $\pi_1 = \pi_2$ នឹងត្រូវបានបដិសេធ ហើយផ្ទុយមកវិញ បើតូចជាង គឺH_0នឹងមិនត្រូវបានបដិសេធនោះទេ (មើលរូប 9.1)។

ដើម្បីបង្ហាញថាឆាយឈ្វែត្រូវបានតេស្តរបៀបណា សូមយើងត្រឡប់ទៅឧទាហរណ៍ 8.3 របស់យើងឡើងវិញ។

i) តេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម

  H_0: $\pi_1 = \pi_2$ សមាមាត្របុរសនិងស្រ្តីគាំទ្របេក្ខភាព A គឺស្មើគ្នា

  H_0: $\pi_1 \neq \pi_2$ សមាមាត្របុរសនិងស្រ្តីគាំទ្របេក្ខភាព A គឺមិនស្មើគ្នា

- **ជំហានទី2:** បើ α= 0.05 ត្រូវបានគេជ្រើសរើសហើយ ឆាយឈ្វែដែលមាន df= (Row-1)*(Column-1)=(1*1)=1 ដូច្នេះយើងបាន χ^2-critical= 3.841
- **ជំហានទី3:** គណនាមធ្យមសមាមាត្រ (រូបមន្ត 8.7 ជំពូកទី 8)

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n} = \frac{37 + 35}{72 + 50} = \frac{71}{122} = 0.582$$

$\bar{p}$ ត្រូវបានប្រើប្រាស់ដើម្បីប៉ាន់ស្មាន π ដែលជាសមាមាត្រតំណាងអោយការគាំទ្របេក្ខភាព A។ ដោយផ្អែកលើចំនួនប៉ាន់គំរូសរុប 122 សមាមាត្រដែលគាំទ្របេក្ខភាព A គឺ 58.2% ហើយសមាមាត្រដែលមិនគាំទ្រគឺ 1- 0.582 ឬ 47.2%.

- **ជំហានទី4:** គណនាប្រេកង់ដែលបានរំពឹងទុក

តារាង 9.4 ប្រេកង់ដែលបានរំពឹងទុក

	ប្រេកង់ដែលបានរំពឹងទុក (f_e) (គាំទ្រ) = $\bar{p} * n_i$	ប្រេកង់ដែលមិនបានរំពឹងទុក (f_e) (មិនគាំទ្រ) = (1- $\bar{p}$) $* n_i$
បុរស	0.582*72 = 41.902	(1-0.582)*72 = 30.096
ស្ត្រី	0.582* 50 = 29.1	(1-0.582)* 50 = 20.9

- **ជំហានទី5:** បញ្ចូលតារាងរំពឹងទុកចូលក្នុងតារាងអង្កេតជាក់ស្តែង

តារាង 9.5 តារាងរំពឹងទុកជាមួយនិងតារាងអង្កេតជាក់ស្តែង

	ក្រុម		
អថេរ	បុរស	ស្ត្រី	សរុប
គាំទ្រ	f_0= 36 f_e=41.902	f_0 = 35 f_e= 29.1	71
មិនគាំទ្រ	f_0= 36 f_e= 30.096	f_0= 15 f_e= 20.9	51
សរុប	72	50	122

- **ជំហាន6:** គណនាឆាយឈ្វេតេស្ត

$$\chi^2_{Stat} = \sum_{all\ cell} \frac{(f_0-f_e)^2}{f_e} = \frac{(36-41.902)^2}{41.902} + \frac{(35-29.1)^2}{29.1} + \frac{(36-30.096)^2}{30.096} + \frac{(15-20.9)^2}{20.9}$$

= 0.8313 + 1.1962 + 1.1582 + 1.6655

= 4.8512

- **ជំហានទី7:** កាធ្វើសេចក្តីសំរេចចិត្ត

ដោយសារតែតម្លៃឆាយឈ្វេតេស្ត χ^2_{Stat} = 4.8512 > χ^2-critical= 3.841, ដូច្នេះ H_0: $\pi_1 = \pi_2$ ត្រូវបានបដិសេដ ហើយយើងសន្និដ្ឋានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា សមាមាត្របុរសនិងស្ត្រីដែលគាំទ្របេក្ខភាព A គឺមិនដូចគ្នាទេ។

ii) ការប្រើភីវែលយូ (p-value)

បន្ទាប់ពីបានតម្លៃ ឆាយឈ្វេតេស្តរួចមក យើងអាចគណនាតម្លៃ ភីវែលយូតាមរយៈ Excel ។ សម្រាប់ឧទាហរណ៍របស់យើង χ^2_{Stat} = 4.8512 ហើយមាន df=1 យើងបាន p =0.0276 <0.05 ដូច្នេះ H_0ត្រូវបានបដិសេធនៅកំរិតអាលហ្វា 5% ហើយយើងធ្វើការសន្និដ្ឋានដូចមុន។

iii) ការប្រើកម្មវិធីសូវែ (Software)

តារាង 9.6 គឺជាលទ្ធផលឆាយឈ្វេតេស្តរបស់ SPSS នៅក្នុងឧទាហរណ៍របស់យើង។ សូមអ្នកអានធ្វើការផ្ទៀងផ្ទាត់តម្លៃឆាយឈ្វេតេស្តនិង p-value នៅក្នុងតារាងនេះដោយខ្លួនឯង។

តារា 9.6 លទ្ធផលឆាយឈ្វេតេស្តតាមតារាងខ្វែងរបស់ SPSS

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.852[a]	1	.028		
Continuity Correction[b]	4.064	1	.044		
Likelihood Ratio	4.935	1	.026		
Fisher's Exact Test				.040	.021
Linear-by-Linear Association	4.812	1	.028		
N of Valid Cases	122				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 20.90.

b. Computed only for a 2x2 table

9.2.2 ឆាយឈ្វែតេស្តរកភាពខុសគ្នារវាងសមាមាត្រលើសពីពីរ

នៅក្នុងផ្នែកនេះយើងធ្វើការពង្រីកការប្រើប្រាស់ឆាយឈ្វែតេស្តទៅលើសមាមាត្រនៃនចំនួនសកលឯករាជ្យលើសពីពីរ។ វាស្រសដៀងទៅនឹងការធ្វើតេស្តទៅលើចំនួនសមាមាត្រពីរដែរ ក៏ប៉ុន្តែយើងនឹងធ្វើការពង្រីកចំនួនគូឡោន (c) នៅក្នុងតារាងខ្វែង (contingency table)។ ហើយសម្មតិកម្មមោឃៈគឺ មិនមានភាពខុសគ្នារវាងសមាមាត្រនៃនួនសកលពីរនោះទេ។

$H_0: \pi_1 = \pi_2 = \cdots \pi_C$

H_1:មិនមែន π_J ទាំងអស់ស្មើគ្នានោះទេ (j=1,2,...c)

ហើយរូបមន្តឆាយឈ្វែគឺរូបមន្ត (9.1) មុន

$$\chi^2_{Stat} = \sum_{all\ cell} \frac{(f_0 - f_e)^2}{f_e} \qquad (9.1)$$

ហើយ f_0 = ចំនួនប្រេក្វង់ដែលបានអង្កេត នៅក្នុងសែល(cell) នៃតារាងខ្វែង 2×c

f_e =ចំនួនប្រេក្វង់ដែលបានរំពឹងទុកក្នុងសែល(cell)នៃតារាងខ្វែង

χ^2_{Stat}សម្រាប់ 2 x c ដែលមាន df= (2-1)x (c-1)= c-1 ដោយសន្មត់ថានៅក្នុងសែល(cell) នៃតារាងខ្វែងនីមួយមានយ៉ាងតិចមួយ។

បើχ^2_{Stat} ធំជាង χ^2-critical នោះ H_0នឹងត្រូវបានបដិសេធ ហើយផ្ទុយមកវិញ បើតូចជាង គឺH_0នឹងមិនត្រូវបានបដិសេធនោះទេ។

ដើម្បីគណនាសមាមាត្រដែលមាន c ក្រុម យើងត្រូវពង្រីកសមីកាត 9.2 ទៀតទៅ៖

$$\bar{p} = \frac{X_1 + X_{2+..} + X_c}{n_1 + n_2 + \cdots + n_c} = \frac{X}{n} \qquad (9.3)$$

បន្ទាប់ពីបានតម្លៃ $\bar{p}$ យើងអាចយកមកគណនាតម្លៃសមាមាត្រដែលយើងចាប់អារម្មណ៍ និងសមាមាត្រដែលមិនចាប់អារម្មណ៍។

- តម្លៃសមាមាត្រដែលចាប់អារម្មណ៍ (expected frequency of interest proportion) គឺ៖

$$f_e = n_1 * \bar{p} \qquad (9.4)$$

- តម្លៃសមាមាត្រដែលមិនចាប់អារម្មណ៍ (expected frequency of not interest proportion) គឺ៖

$$f_{e1} = n_i * (1 - \bar{p}) \qquad (9.5)$$

ឧទាហរណ៍ 9.1 កាស្ទង់មតិមួយបានធ្វើទៅលើសិស្សចំនួនសរុប 500នាក់ ចេញពីវិទ្យាល័យចំនួនបី A B និង C ទាក់ទងដល់ចំណង់ចំណូលចិត្តរបស់គេចំពោះមុខវិទ្យាគណិតវិទ្យា។ ចូលចិត្ត (yes) ឬមិនពេញចិត្ត (no) ។ ទិន្នន័យទាំងអស់នេះបានបង្ហាញនៅក្នុងតារាង 9.4។

តារាង 9.7 តារាងអង្កេត 2*3 របស់ប៉ាន់គំរូ (f_o)

	វិទ្យាល័យ			
	A	B	C	សរុប
ចូលចិត្ត	190	160	110	460
មិនចូលចិត្ត	50	40	50	140
សរុប	240	200	160	600

បន្ទាប់មកយើងគណនា សមាមាត្រដែលចូលចិត្តគណិតវិទ្យារបស់វិទ្យាល័យទាំងបី ($\bar{p}$) ដើម្បីប៉ាន់ស្មានតម្លៃសមាមាត្ររបស់ចំនួនសកល (π) ដោយប្រើរូបមន្ត (9.3)។

$$\bar{p} = \frac{X_1 + X_{2+..} + X_c}{n_1 + n_2 + \cdots + n_c} = \frac{(190 + 160 + 110)}{(240 + 200 + 160)} = \frac{460}{600} = 0.7667$$

ដូច្នេះសិស្សដែលចូលចិត្តមុខវិជ្ជាគណិតវិទ្យាមាន 76.67% ហើយសិស្សដែលមិនចូលចិត្តមុខវិជ្ជានេះមាន (1-$\bar{p}$) ឬ 23.33%។

បន្ទាប់មកយើងគណនា តម្លៃសមាមាត្រដែលចាប់អារម្មណ៍ដោយប្រើរូបមន្ត 9.4

- សម្រាប់ឧទាហរហ៍របស់យើងតម្លៃសមាមាត្រដែលចាប់អារម្មណ៍គឺ ការចូលចិត្តមុខវិទ្យាគណិតរួមមាន៖
  - តម្លៃរំពឹងទុករបស់វិទ្យាល័យ A: $f_e = n_1 * \bar{p}$= 240*0.7667 = 184
  - តម្លៃរំពឹងទុករបស់វិទ្យាល័យ B: $f_e = n_2 * \bar{p}$= 200*0.7667 = 153.33
  - តម្លៃរំពឹងទុករបស់វិទ្យាល័យ C: $f_e = n_3 * \bar{p}$= 160*0.7667 = 122.67
- ហើយសមាមាត្ររំពឹងទុករបស់សិស្សដែលមិនចូលចិត្តមុខវិទ្យាគណិត គឺត្រូវគណនាតាមរូបមន្ត៖

  $$f_{e1} = n_i * (1 - \bar{p}) \qquad (9.5)$$

  - តម្លៃរំពឹងទុកមិនចូលចិត្តរបស់វិទ្យាល័យ A: f_{e1} = 240*(1-0.7667)= 56
  - តម្លៃរំពឹងទុកមិនចូលចិត្តរបស់វិទ្យាល័យ B: f_{e1} = 200*(1-0.7667)=46.67
  - តម្លៃរំពឹងទុកមិនចូលចិត្តរបស់វិទ្យាល័យ C: f_{e1} = 160*(1-0.7667)= 37.33

តារាង 9.8 តារាងរំពឹងទុក 2*3 (f_e)

	វិទ្យាល័យ			
	A	B	C	សរុប
ចូលចិត្ត	184	153.33	122.67	460
មិនចូលចិត្ត	56	46.67	37.33	140
សរុប	240	200	160	600

- ជំហ៊ានបន្ទាប់យើងសរសេរសម្មតិកម្មនៃតម្លៃសមាមាត្រ

  $H_0: \pi_1 = \pi_2 = \pi_3$

  H_1: មិនមែន π_j ស្មើគ្នា, (ហើយ $j = 1, 2, 3$)
- វិធិដែលងាយស្រួលក្នុងការគណនាឆាយឈ្វែតេស្ត χ^2_{Stat} គឺយើងបញ្ចូលតារាង អង្កេត និងរំពឹងទុកចូលក្នុងតារាងតែមួយ។

តារាង 9.10 សមាមាត្រអង្កេតនិងរពឹងទុក

f_o	f_e	(f_o-f_e)	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
190	184	6	36	0.195652
160	153.33	6.67	44.4889	0.290151
110	122.67	-12.67	160.5289	1.308624
50	56	-6	36	0.642857
40	46.67	-6.67	44.4889	0.953265
50	37.33	12.67	160.5289	4.300265
សរុប				7.69

$$\chi^2_{Stat} = \frac{(f_o - f_e)^2}{f_e} = 7.69$$

- បើ α = 0.05 និងឆាយឈ្វែដែលមាន df= c-1= 3-1 =2, តាមតារាង E 4 យើងបានតម្លៃ $\chi^2_{0.05,2} = 5.991$
- ការធ្វើសេចក្ដីសំរេចចិត្តៈ ដោយសារ $\chi^2_{Stat} = 7.69 > \chi^2_{0.05,2} = 5.991$ ដូច្នេះ $H_0: \pi_1 = \pi_2 = \pi_3$ ត្រូវបានបដិសេធនូវកំរិត α = 0.05 ហើយយើងសន្និដ្ឋានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថាចំណង់ចំណូលចិត្តរបស់សិស្សវិទ្យាល័យទាំងបីគឺមិនស្មើគ្នានោះទេ។

ii) ការប្រើភីវែលយូ (p-value)

បន្ទាប់ពីបានតម្លៃ ឆាយឈ្វែតេស្តរួចមក យើងអាចគណនាតម្លៃ p-value តាមរយៈ Excel Calculator។ សម្រាប់ឧទាហរណ៍របស់យើង $\chi^2_{Stat} = 7.69$ ហើយមាន df=2 យើងបាន p-value =0.021 <0.05 ដូច្នេះ H_0ត្រូវបានបដិសេធនៅកំរិតអាលហ្វា 5% ហើយយើងធ្វើការសន្និដ្ឋានដូចមុន។

iii) ការប្រើកម្មវិធីសូវែ (Software)

តារាង 9.7 គឺជាលទ្ធផលធាយឈ្វែតេស្តរបស់ SPSS នៅក្នុងឧទាហរណ៍របស់យើង។ សូមអ្នកអានធ្វើការផ្ទៀងផ្ទាត់តម្លៃធាយឈ្វែតេស្តនិង p-value នៅក្នុងតារាងនេះដោយខ្លួនឯង។

តារាង 9.11 លទ្ធផលធាយឈ្វែតេស្តរបស់ SPSS

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.686[a]	2	.021
Likelihood Ratio	7.383	2	.025
N of Valid Cases	600		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 37.33.

9.3 ធាយឈ្វែតេស្តរកភាពឯករាជ្យ

នៅផ្នែកមុនយើងបានសិក្សាអំពីការប្រើប្រាស់ធាយឈ្វែតេស្តដើម្បីរកតម្លៃខុសគ្នារវាងសមាមាត្រពីរ ឬច្រើនដោយផ្អែកលើតារាងខ្វែង។ ឥឡូវយើងបន្តប្រើតារាងខ្វែងដែលមានជួរដេក (r) និងជួរឈរ (c) ដើម្បីបង្កើតករណីទូទៅមួយសម្រាប់ការប្រើធាយឈ្វែដើម្បីធ្វើតេសរកភាពឯករាជ្យ (Chi-square test for independence) នៃចំនួនសកលដែលមានលក្ខណៈសមាមាត្រពីរ។

- សម្មតិកម្មនៃកាតធ្វើតេស្តឯករាជ្យ

  H_0: អថេរប្រភេទពីរឯករាជ្យពីគ្នា (មិនទាក់ទងគ្នា)

  H_1: អថេរប្រភេទពីរមិនឯករាជ្យពីគ្នា (ទាក់ទងគ្នា)

- ចំណែករូបមន្តគឺនៅដែលគឺរូបមន្ត 9.1៖

$$\chi^2_{Stat} = \sum_{all\ cell} \frac{(f_0 - f_e)^2}{f_e} \qquad (9.1)$$

ហើយ $f_0 =$ ចំនួនប្រេកង់ដែលបានអង្កេត (observed frequency) នៅក្នុងសែល (cell) នៃតារាងខ្វែង (ដូចមុន)

$f_e =$ ចំនួនប្រេកង់ដែលបានរំពឹងទុកក្នុងស៊ែលនៃតារាងខ្វែង (ដូចមុន)

- ការតេស្តគឺនៅដដែលដូចមុន បើχ^2_{Stat} ធំជាង χ^2-critical នោះ $H_0: \pi_1 = \pi_2$ នឹងត្រូវបានបដិសេធ ហើយផ្ទុយមកវិញ បើតូចជាង គឺH_0នឹងមិនត្រូវបានបដិសេធនោះទេ។
- ដោយឡែកប្រេកង់ដែលរំពឹងទុកត្រូវបានគណនាដូចរូបមន្ត 9.6៖

$$f_e = \frac{Row\ Total\ x\ Column\ Total}{n} \qquad (9.6)$$

ហើយ ចំនួនរ៉ូសរុប (row total)= ផលបូកនៃប្រេកង់ទាំងអស់នៅក្នុងជួរដេក

ចំនួនជួរឈរសរុប (column total)=ផលបូកនៃប្រេកង់ទាំងអស់នៅក្នុងជួរឈរ

$n =$ ចំនួនប៉ាន់គំរូ

ឧទាហរណ៍ 9.3 ការស្ទង់មតិមួយដើម្បីសិក្សាទៅលើការធ្លាក់ចុះនៃចំនួនភ្ញៀវក្នុងការស្នាកនៅសណ្ឋាគារក្នុងតំបន់មួយនៃក្រុងព្រះសីហនុ។ ដើម្បីឲ្យដឹងអំពីមូលហេតុពិតប្រាកដនៃការធ្លាក់ចុះនេះ អ្នកសិក្សាម្នាក់បានធ្វើការសាកសួរទៅលើប៉ាន់គំរូសណ្ឋាគារចំនួនបីប្រភេទ (A=ទំហំធំ B=កណ្តាល និង C=តូច)៖ A ចំនួន 100 នាក់ B ចំនួន 40 នាក់ និងC ចំនួន 80នាក់ ដែលទាក់ទងជាមួយមូលហេតុដែលធ្វើឲ្យភ្ញៀវមិនត្រឡប់មកស្នាក់នៅជាថ្មី៖

1)តម្លៃ

2)ទីតាំង

3)លក្ខណៈបន្ទប់

4)និងកត្តាផ្សេងៗ

ទិន្នន័យទាំងនេះមាននៅក្នុងតារាង 9.10។

តារាង 9.12 តារាងខ្វែង ហេតុផលមិនត្រឡប់មកស្នាក់នៅជាថ្មី

	ប្រភេទសណ្ឋាគារ			
ហេតុផលមិនត្រឡប់	A	B	C	សរុប
តម្លៃ	30	8	40	78
ទីតាំង	40	16	12	68
បន្ទប់	15	7	18	40
ផ្សេងៗ	15	9	10	34
សរុប	100	40	80	220

i) ការប្រើតេស្តស្ថិតិ

- **ជំហានទី1:** សរសេរសម្មតិកម្ម

H_0: ប្រភេទសណ្ឋាគារគឺឯករាជ្យទៅនឹងហេតុផលមិនត្រឡប់មកវិញ (មិនទាក់ទងគ្នា)

H_1: ប្រភេទសណ្ឋាគារគឺមិនឯករាជ្យទៅនឹងហេតុផលមិនត្រឡប់មកវិញ (ទាក់ទងគ្នា)

- **ជំហានទី2:** គណនាតម្លៃរំពឹងទុករបស់សែលនីមួយៗតាមរូបមន្ត 9.6
  - សែល (ថ្លៃរបស់ A)= (78x100)/220 = 35.45
  - សែល (ទីតាំងរបស់ A)= (68x100)/220 = 30.91
  - សែល (បន្ទប់របស់ A)= (40x100)/220 = 18.18
  - សែល (ផ្សេងៗរបស់ A)= (34x100)/220 = 15.45

ចំពោះតម្លៃរំពឹងទុកនៃសែល (cell) ដ៏ទៃទៀតរបស់សណ្ឋាគារ B និង C គឺត្រូវបានគណនាដោយប្រើវិធីសាស្ត្រដូចគ្នា ហើយលទ្ធផលទាំងនោះត្រូវបានសង្ខេបនៅក្នុងតារាង 9.11 ខាងក្រោម៖

តារាង 9.13 តម្លៃរពឹងទុករបស់សណ្ឋាគារនីមួយៗ

	សណ្ឋាគារ			
ហេតុផលមិនត្រឡប់	A	B	C	សរុប
តម្លៃ	35.45	14.18	28.36	78
ទីតាំង	30.91	12.36	24.73	68
បន្ទប់	18.18	7.27	14.54	40
ផ្សេងៗ	15.45	6.18	12.36	34
សរុប	100	40	80	220

- **ជំហានទី3**: គណនាឆាយឈ្វែតេស្ត

តារាង 9.14 គណនាឆាយឈ្វែតេស្តលើភាពឯករាជ្យ

សែល (cell)	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
តម្លៃរបស់ A	30	35.45	-5.45	29.7025	0.83787
ទីតាំងរបស់ A	40	30.91	9.09	82.6281	2.673183
បន្ទប់របស់ A	15	18.18	-3.18	10.1124	0.556238
ផ្សេងៗរបស់ A	15	15.45	-0.45	0.2025	0.013107
តម្លៃរបស់ B	8	14.18	-6.18	38.1924	2.693399
ទីតាំងរបស់ B	16	12.36	3.64	13.2496	1.071974
បន្ទប់របស់ B	7	7.27	-0.27	0.0729	0.010028
ផ្សេងៗរបស់ B	9	6.18	2.82	7.9524	1.286796
តម្លៃរបស់ C	40	28.36	11.64	135.4896	4.777489
ទីតាំងរបស់ C	12	24.73	-12.73	162.0529	6.552887
បន្ទប់របស់ C	18	14.54	3.46	11.9716	0.823356
ផ្សេងៗរបស់ C	10	12.36	-2.36	5.5696	0.450615
					21.74694

$$\chi^2_{Stat} = \sum_{all\ cell} \frac{(f_0 - f_e)^2}{f_e} = 21.74694 \quad (p= 0.00134)$$

- **ជំហានទី4**: ធ្វើសេចក្តីសំរេចចិត្ត

ឆាយឈ្វែដែលមាន df= (r-1)*(c-1) = (4-1)(3-1)= 6

បើសិន α=0.05, ដូច្នេះតាមតារាងបំណែងចែកឆាយឈ្វែយើងបាន $\chi^2_{0.05,6} = 12.592$។

ដោយ$\chi^2_{Stat} = 21.74694 > \chi^2_{0.05,6} = 12.592$។ ដូច្នេះយើងបដិសេធ H_0 ហើយសន្និដ្ឋានថា យើងមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ប្រភេទសណ្ឋាគារ (A=ទំហំធំ B=កណ្ដាល និង C=តូច) មានទំនាក់ទំនងទៅនិងហេតុផលមិនត្រឡប់មកជួលសណ្ឋាគារវិញ។

ii) ការប្រើភីវ៉េលយូ (p-value)

បន្ទាប់ពីបានតម្លៃ ឆាយស្ក្វែតេស្តរួចមក យើងអាចគណនាតម្លៃ p-value តាមរយៈ Excel Calculator។ សម្រាប់ឧទាហរណ៍របស់យើង$\chi^2_{Stat} = 21.74694$ ហើយមាន df=6 យើងបាន p-value =0.001 <0.05 ដូច្នេះ H_0ត្រូវបានបដិសេធនៅកំរិតអាលហ្វា 5% ហើយយើងធ្វើការសន្និដ្ឋានដូចមុន។

iii) ការប្រើកម្មវិធីសូវ៉ែ (Software)

តារាង 9.15 គឺជាលទ្ធផលឆាយស្ក្វែតេស្តរបស់ SPSS នៅក្នុងឧទាហរណ៍របស់យើង។ សូមអ្នកអានធ្វើការផ្ទៀងផ្ទាត់តម្លៃឆាយស្ក្វែតេស្តនិង p-value នៅក្នុងតារាងនេះដោយខ្លួនឯង។

តារាង 9.15 ឆាយស្ក្វែតេសរកភាពឯករាជ្យនៃសណ្ឋាគារផ្ដល់ដោយ SPSS

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21.739[a]	6	.001
Likelihood Ratio	22.830	6	.001
Linear-by-Linear Association	1.002	1	.317
N of Valid Cases	220		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.18.

9.4 ម៉ាកណេម៉ាតេស្ត

នៅពេលដែលអ្នកប្រមូលប៉ាន់គំរូជាគូរ ឬវាស់វែងច្រំដែល (repeated sample) ឧទាហរណ៍ ដូចជា មុនប្រឡង និងក្រោយប្រឡង គឺយើងមិនអាចប្រើឆាយស្ក្វែតេស្តធម្មតាបានទេ។ អ្នកត្រូវប្រើម៉ាកណេម៉ាតេស្ត (McNemar's Test)។ ទោះបីឆាយស្ក្វែតេស្តនិងម៉ាកណេម៉ាតេស្តត្រូវបានប្រើសម្រាប់តេស្តលើសមាមាត្រ តែតេស្តទាំងពីរគឺខុសគ្នា ឆាយស្ក្វែតេស្តគឺតម្រូវឲ្យទិន្នន័យឯករាជ្យពីគ្នា (independent sample) ចំណែកម៉ាកណេម៉ាតេស្តគឺតម្រូវឲ្យទិន្នន័យទាក់ទងគ្នា (related sample) ហើយការរៀបចំទិន្នន័យក៏មិនដូចឆាយស្ក្វែតេស្តធម្មតាដែរ។ លើសពីនេះទៅទៀត ម៉ាកណេម៉ាតេស្តមានទម្រង់ដូចបំណែងចែកធម្មតា (normal distribution)។

តារាង 9.16 កំណត់អំពីទម្រង់របស់ម៉ាកណេម៉ាតេស្ត។ តារាងនេះគឺជាតារាងខ្វែងដូចគណីរបស់ឆាយស្ក្វែដែរប៉ុន្តែរបៀបគណនាសមាមាត្រគឺខុសគ្នា។

តារាង 9.16 តារាងខ្វែង 2 x 2 សម្រាប់ម៉ាកណេម៉ាតេស្ត

	After Condition (group 2)	

Before Condition (group 1)	Yes	No	Total
Yes	A	B	A+B
No	C	D	C+D
Total	A+C	B+D	n

ហើយ A= ចំនួនអ្នកឆ្លើយបាទ ចំពោះល័ក្ខខ័ណ្ឌទីមួយ និងបាទចំពោះលក្ខខណ្ឌទីពីរ

B=ចំនួនអ្នកឆ្លើយបាទ ចំពោះល័ក្ខខ័ណ្ឌទីមួយ និងទេចំពោះល័ក្ខខ័ណ្ឌទីពីរ

C=ចំនួនអ្នកឆ្លើយទេ ចំពោះល័ក្ខខ័ណ្ឌទីមួយ និងបាទចំពោះល័ក្ខខ័ណ្ឌទីពីរ

D=ចំនួនអ្នកឆ្លើយទេ ចំពោះល័ក្ខខ័ណ្ឌទីមួយ និងទេ ចំពោះល័ក្ខខ័ណ្ឌទីពីរ

n=ចំនួនអ្នកឆ្លើយក្នុងប៉ាន់គំរូ

- សមាមាត្រនៃប៉ាន់គំរូដែលចាប់អារម្មណ៍គឺៈ

  $p_1 = \frac{A+B}{n}$=សមាមាត្ររបស់អ្នកឆ្លើយ "បាទ" ចំពោះល័ក្ខខ័ណ្ឌទីមួយ

  $p_2 = \frac{A+C}{n}$=សមាមាត្ររបស់អ្នកឆ្លើយ "បាទ" ចំពោះល័ក្ខខ័ណ្ឌទីពីរ
- សម្មតិកម្មត្រូវតេស្តលើសមាមាត្រអាចមានកន្ទុយសងខាងនិងកន្ទុយម្ខាងៈ
  - $H_0: \pi_1 = \pi_2$ និង $H_1: \pi_1 = \pi_2$ ឬ
  - $H_0: \pi_1 \geq \pi_2$ និង $H_1: \pi_1 < \pi_2$ ឬ
  - $H_0: \pi_1 \leq \pi_2$ និង $H_1: \pi_1 >= \pi_2$
- ម៉ាកណេម៉ាតេស្តស្ថិតិអាចតេស្តបានពីររបៀប ហើយករណីទាំងពីរគឺពឹងផ្អែកលើតារាងខ្វែង 2x2 តែមួយ

1. McNemar's Z-test

$$Z_{stat.} = \frac{B-C}{\sqrt{(B+C)}} \qquad (9.7)$$

ហើយ $Z_{stat.}$ មានបំណែងចែកដូច the normal distribution

B = មុនឆ្លើយ YES ក្រោយឆ្លើយ NO បើ B កាន់តែធំ ក្រោយដាក់ល័ក្ខខ័ណ្ឌ ឆ្លើយ NO កាន់តែច្រើន។

C = មុនឆ្លើយ NO ក្រោយឆ្លើយ YES បើ C កាន់តែធំ ក្រោយដាក់ល័ក្ខខ័ណ្ឌឆ្លើយ YES កាន់តែច្រើន ។

- បើ B > C ឬតម្លៃ $Z_{stat.} > 0$ ហើយបើ $Z_{stat.}$ មាន p-value < 0.05 ថែមទៀត នោះមានន័យថា បន្ទាប់ពីដាក់ល័ក្ខខ័ណ្ឌរួច NO > YES ឬល័ក្ខខ័ណ្ឌគ្មានប្រសិទ្ធភាព។
- បើ B < C ឬតម្លៃ $Z_{stat.} < 0$ ហើយបើ $Z_{stat.}$ មាន p-value > 0.05 ថែមទៀត នោះមានន័យថា បន្ទាប់ពីដាក់លក្ខខណ្ឌ NO < YES ឬល័ក្ខខ័ណ្ឌមានប្រសិទ្ធភាព។

- ការសំរេចចិត្របដិសេធដូចករណីធម្មតាដែរ៖
    - Z-stat. > Z_α បដិសេធ H_0
    - Z-stat. < Z_α មិនបដិសេធ H_0

2. McNemar's χ^2-test[17]

$$\chi^2 = \frac{(B-C)^2}{B+C} \qquad (9.10)$$

បើសិនជាសែល C និង B ធំគ្រប់គ្រាន់ តម្លៃ χ^2មានទម្រង់ដូចបំណែងចែកឆាយឈ្វែដែលមាន df=1។

ប្រហែលជាមួយឆ្នាំបន្ទាប់ពី Quinn McNemar បានចេញផ្សាយរូបមន្តតេស្តរបស់គាត់ លោក Edwards បានស្នើឲមានការកែរូបមន្តឆាយឈ្វែទៅជា(9.11) ដែលយើងប្រើរហូដល់បច្ចុប្បន្ន។

ឆាយឈ្វែដែលធ្វើកំណែត្រឹមត្រូវ (Chi-square Continuity Corrected Version)

$$\chi^2 = \frac{(|B-C|-1)^2}{B+C} \qquad (9.11)$$

ឧទាហរណ៍ 9.4 ឧបមាថា អ្នកធ្វើការសាកសួរទៅលើអតិថិជនចំនួន 400 នាក់ ទាក់ទងដល់ចំណាប់អរម្មណ៍នៃការទទួលយកឥទានសារជាថ្មីនៃការបង់រំលួសផ្ទះ មុននឹងប្តូរល័ក្ខខ័ណ្ឌ។ បន្ទប់ពីកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួស ហើយដើម្បីកំណត់ថា ការកែប្រែល័ក្ខខ័ណ្ឌក្នុងការបង់រំលួសមានប្រសិទ្ធភាព អ្នកធ្វើការសាកសួរទៅលើ400នាក់ដដែល ហើយពួកគេបានឆ្លើយដូចខាងក្រោម៖

- មុនកែប្រែលក្ខខណ្ឌ ឆ្លើយ yes និងក្រោយកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ yes មាន 140នាក់

តារាង 9.17 ភាពស្មោះត្រង់ជាមួយម៉ាកសញ្ញា

មុនកែប្រែល័ក្ខខ័ណ្ឌ	ក្រោយពីកែប្រែល័ក្ខខ័ណ្ឌ		
	yes	no	សរុប
yes	140	40	180
no	20	200	220
សរុប	160	240	400

- មុនកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ yes និងក្រោយកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ no មាន 40នាក់
- មុនកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ no និងក្រោយកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ yes មាន 20នាក់
- មុនកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ no និងក្រោយកែប្រែល័ក្ខខ័ណ្ឌ ឆ្លើយ no មាន 200នាក់។

ចំឡើយទាំងនេះត្រូវបានរៀបចំជាតារាងខ្វែងដូចក្នុងតារាង 9.17។

---

[17] McNemar's test for paired nominal data derived from http://rasbt.github.io/mlxtend/user_guide/evaluate/mcnemar/#:~:text=The%20McNemar%20test%20statistic%20(%22chi,with%20one%20degree%20of%20freedom.

តើការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសមិនមានប្រសិទ្ធភាពដែរឬអត់?

ចំឡើយ៖

i) ការប្រើតេស្តស្ថិតិ

- ជំហានទី1: សម្មតិកម្មមោឃៈត្រូវសន្មត់ថាការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសមិនធ្វើឲ្យសមាមាត្រនៃចំនួនអ្នកខ្ចីមានការប្រែប្រួលនោះទេ ឬគ្មានប្រសិទ្ធភាព មានន័យថា សមាមាត្រនៃចំនួនអ្នកខ្ចីមុននិងការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសគឺស្មើគ្នា

  $H_0: \pi_1 = \pi_2$ ការប្តូរល័ក្ខខ័ណ្ឌគឺមិនធ្វើឲ្យសមាមាត្រអ្នកខ្ចីមានការប្រែប្រួល

  $H_0: \pi_1 \neq \pi_2$ ការប្តូរល័ក្ខខ័ណ្ឌគឺមិនធ្វើឲ្យសមាមាត្រអ្នកខ្ចីមានការប្រែប្រួល

- ជំហានទី2: គណនាសមាមាត្រនៃប៉ាន់គំរូដែលចាប់អារម្មណ៍គឺ
  - សមាមាត្ររបស់អ្នកឆ្លើយថា ខ្ចី មុនប្តូរល័ក្ខខ័ណ្ឌ

    $p_1 = \frac{A+B}{n} = \frac{140+40}{400} = \frac{180}{400} = 0.45$
  - សមាមាត្ររបស់អ្នកឆ្លើយថា ខ្ចី ក្រោយប្តូរល័ក្ខខ័ណ្ឌ

    $p_2 = \frac{A+C}{n} = \frac{140+20}{400} = \frac{160}{400} = 0.40$

❖ សំគាល់: p_1បាន់ស្មានឬតំណាង π_1ហើយ p_2បាន់ស្មានឬតំណាង π_2 ហើយ

  បើ $\pi_1 \geq \pi_2$ ការក្រោយប្តូរល័ក្ខខ័ណ្ឌគឺគ្មានប្រសិទ្ធភាព

- ជំហានទី3: ជ្រើសរើស Z-critical បើ α=0.05 ត្រូវបានជ្រើសរើសនោះតាមតារាង normal distribution យើងបានតម្លៃ $Z_{0.05} = \pm 1.645$
- ជំហានទី4: គណនា Z-test

$$Z_{stat.} = \frac{B-C}{\sqrt{(B+C)}} = \frac{40-20}{\sqrt{40+20}} = 2.582$$

- ជំហានទី5: ធ្វើសេចក្តីសំរេចចិត្ត

  ដោយសារតែ $Z_{stat.} = 2.582 > Z_{0.05} = 1.645$ (B > C) ដូច្នេះ H_0 ត្រូវបានបដិសេធ ហើយយើងសន្និដ្ឋាជាបឋមថា មានស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសធ្វើឲ្យសមាមាត្រនៃចំនួនអ្នកខ្ចីមានការប្រែប្រួល។ មានន័យថា សមាមាត្រមុននិងក្រោយកែប្រែល័ក្ខខ័ណ្ឌគឺមិនស្មើគ្នានេះទេ ដោយសារតែH_0 ត្រូវបានបដិសេធ ក៏ប៉ុន្តែយើងមិនទាន់ដឹងថាការកែប្រែល័ក្ខខ័ណ្ឌនេះ ធ្វើឲ្យអ្នកខ្ចីកើនច្រើនជាងមុន ឬថយច្រើនជាងមុន។ បើកើនច្រើនជាងមុនគឺមានប្រសិទ្ធភាព ក៏ប៉ុន្តែបើថយ ឬមិនប្រែប្រួលគឺមានន័យថាគ្មានប្រសិទ្ធភាព។ ដូច្នេះជំហានបន្ទាប់ដែលត្រូវពិនិត្យលើតម្លៃនៃ B-C តាមរយៈនៃការប្រើ Z-test៖

  - B-C > 0 ឬ $Z_{stat.} > 0$, ការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសគ្មានប្រសិទ្ធភាព
  - B-C < 0 ឬ $Z_{stat.} > 0$, ការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសមានប្រសិទ្ធភាព

ចំពោះករណីរបស់យើង $Z_{stat.} = 2.582 > 0$ (p= 0.0098) ដូច្នេះយើងអាចសន្និដ្ឋាជាចុងក្រោយថា ការកែប្រែល័ក្ខខ័ណ្ឌបង់រំលួសមានប្រសិទ្ធភាពនោះទេ។

ឥឡូវយើងប្រើរូបមន្ត 9.11 ដើម្បីធ្វើការតេស្តលើសម្មាតិកម្មទៅលើឧទាហរណ៍របស់យើងវិញម្តង តើលទ្ធផលដូច Z-test ឬយ៉ាងណា?

$$\chi^2 = \frac{(|B-C|-1)^2}{B+C} = \frac{(|40-20|-1)^2}{40+20} = 6.0167 \qquad (p=0.01417)$$

បើ α=0.05 ត្រូវបានជ្រើសរើសនោះតាមតារាង Chi-square distribution យើងបានតម្លៃឆាយឈ្វេលគោលដែលមាន df=1 គឺ $\chi^2_{0.025,1}$= 5.024។ ដោយសារ χ^2-statistic = 6.0167 > $\chi^2_{0.025,1}$= 5.024 ដូច្នេះ H_0 ត្រូវបានបដិសេធ ហើយយើងសន្និដ្ឋាថា ការកែប្រែលក្ខខ័ណ្ឌបង់រំលួស ធ្វើឲ្យប៉ះពាល់ដល់ឥទាន (សមាមាត្រអ្នកខ្ចីមុននិងក្រោយកែប្រែលក្ខខ័ណ្ឌបង់រំលួសមិនស្មើគ្នា) ក៏ប៉ុន្តែមយើងមិនដឹងថាមានប្រសិទ្ធភាពឬយ៉ាងណានោះទេ ពីព្រោះតម្លៃឆាយឈ្វេមិនមានតម្លៃអវិជ្ជមាននោះទេ ហើយករណីនេះវាខុសពី ការប្រើ Z-test ខាងលើ។

ក្នុងឧទាហរណ៍ 9.4 តែមួយ យើងអាចប្រើ ប្រើឆាយឈ្វេតេស្ត ឬ Z-test ក៏បាន ហើយលទ្ធផលគឺបដិសេធH_0ដូចគ្នា។ Z-test មានលក្ខណៈងាយយល់និងស៊ីជំរៅជាងឆាយឈ្វេតេស្ត ក្នុងឧទាហរណ៍នេះ។

- ❖ SPSS មិនបានបង្កើត McNema's test (Z-test) នោះទេ តែ SPSS បានង្កើត ឆាយឈ្វេសម្រាប់ McNema's test ជំនួសវិញ។ ហើយបើយើងប្រើទិន្នន័យក្នុងតារាង 9.4 ក្នុង SPSS យើងនឹងបានលទ្ធផលដូចតារាង 9.18 ខាងក្រោម។

តារាង 9.18 លទ្ធផល McNema's Test លើការកែប្រែកលក្ខខ័ណ្ឌបង់រំលួស

Test Statistics[a]

	Before change & After cchange
N	400
Chi-Square[b]	6.017
Asymp. Sig.	.014

a. McNemar Test

b. Continuity Corrected

9.5 វិធីសុនតេស្តលើចំនួនសកលពីរឯករាជ្យ

នៅក្នុងជំពូកទី8 អ្នកបានសិក្សា t-test សម្រាប់តេស្តរកតម្លៃខុសគ្នារវាងចំនួនសកលពីរ ហើយក្នុងការតេស្តនោះបានសន្មត់ថា ទិន្នន័យមានបំណែងចែកធម្មតា។ ក្នុងករណីបើទិន្នន័យមិនមានលក្ខណៈធម្មតា នោះអ្នកមានជំរើសពីរ៖

- ប្រើវីកូសុនរ៉ែងសាំតេស្ត (Wilcoxon Rank Sum Test) ឬ
- ប្រើធីតេស្តដែលមានវ៉ារីយ៉ង់រួមបញ្ចូលគ្នា (pooled variance t test) (សៀវភៅនេះមិនបានរៀបរាប់ករណីនេះទេ)។

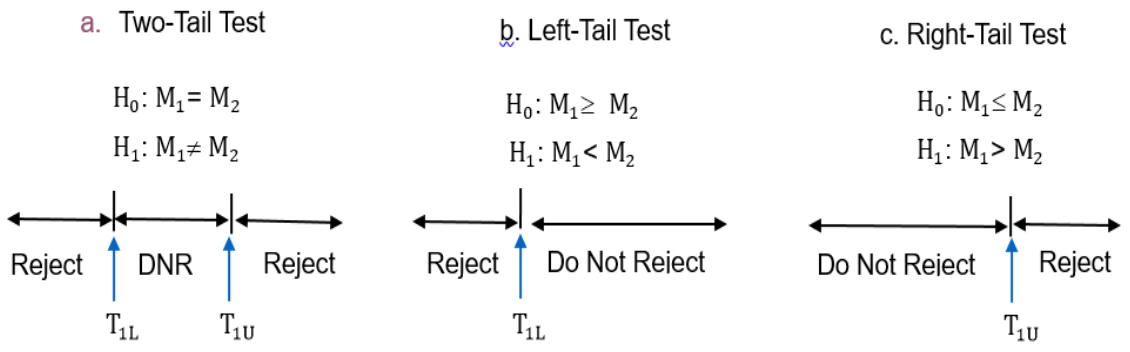
នៅផ្នែកនេះនឹងរៀបរាប់អំពី Wilcoxon Rank Sum Test ជាប្រភេទតេស្តមួយដែលតេស្តរកភាពខុសគ្នារវាងមេដ្យានពីរ របស់ចំនួនសកល (difference between two median of population)។

សូមបញ្ជាក់ថា Wilcoxon test គឺមានពីរ៖ Wilcoxon Rank Sum Test ឬ ហៅថា Mann-Whitney Test ក្នុង SPSS ប្រើប៉ាន់គំរូឯករាជ្យ (two independent samples)ហើយ Wilcoxon Signed-Rank Test គឺប្រើប៉ាន់គំរូមិនឯករាជ្យ ឬទិន្នន័យជាគូ (dependent/ related or paired data)។ ដំបូងនេះ យើងលើកយក Wilcoxon Rank Sum Test ជជែកជាមុន និង Wilcoxon Signed-Rank Test ជជែកនៅពេលបន្ទាប់។

វីកូសុនរ៉េងសាំតេស្ត មានអានុភាពដូច t-test ដែលបានរៀបរាប់នៅជំពូកទី 8 ដែរ ក៏ប៉ុន្តវាមិនត្រូវការទិន្នន័យមានបំណែងចែកធម្មតានោះទេ។ អ្នកអាចប្រើ Wilcoxon Rank Sum Test បាន ដ៏រាបណាទិន្នន័យរបស់អ្នកជាប្រភេទដែលមានលំដាប់លំដោយ (only ordinal data)។ នេះជាករណីដែលកើតឡើងជាញឹកញាប់នៅក្នុងការសិក្សាលើឥរិយាបថអតិថិជននិងការសិក្សាទីផ្សារ[18] (consumer behavior and marketing research)

ដើម្បីប្រើ Wilcoxon តេស្ត អ្នកត្រូវប្រមូលទិន្នន័យចេញពីប៉ាន់គំរូពីរ n_1 និង n_2 ដែលចំនួនប៉ាន់គំរូសរុប n = $n_1 + n_2$។ បន្ទាប់មកអ្នកត្រូវកំណត់រ៉េង (rank) នៃប៉ាន់គំរូពីតូចទៅធំ។ លេខតូចជាងគេគឺកំណត់លំដាប់ទី1 និងបន្តបន្ទាប់ រហូតដល់លេខធំជាងគេគឺ $n_1 + n_2$។ បើសិនមានលេខជាន់គ្នា អ្នកត្រូវគិតជាតម្លៃមធ្យមកំអោយលំដាប់ជាន់គ្នា។

រូប9.3 Wilcoxon តេស្តលើមេដ្យានរបស់ចន្លួនសកលពីរ



ក្នុងករណីដែលចំនួនប៉ាន់គំរូមិនស្មើគ្នា n_1គឺប៉ាន់គំរូដែលតូចជាង n_2។ Wilcoxon តេស្តគឺ៖

- T_1 = ផលបូកនៃលំដាប់លំដោយដែលបានកំណត់ក្នុងប៉ាន់គំរូទី1 n_1ដែលជាប៉ាន់គំរូដែលតូចជាងគេ (បើប៉ាន់គំរូស្មើគ្នា$n_1 = n_2$ នោះ T_1គឺយកតម្លៃមួយណាក៏បាន) ហើយ
- T_2 =ផលបូកនៃលំដាប់លំដោយដែលបានកំណត់ក្នុងប៉ាន់គំរូទី2 n_2។

អ្នកអាចធ្វើការផ្ទៀងផ្ទាត់នូវផលបូករ៉េង (rank) របស់ T ទាំងពីរគឺត្រូវទៅនឹង៖

$$T_1 + T_2 = \frac{n(n+1)}{2} \qquad (9.12)$$

Wilcoxon តេស្តអាចមានកន្ទុយសងខាង ឬម្ខាង (ស្តាំឬឆ្វេង) ដូចក្នុងរូបទី 9.3។

---

[18] Page 683 of Allan G. B (2009) Elementary Statistics: A Step by Step Approach. Eight Edition. McGraw-

9.5.1 វិក្លុសុនតេស្តលើចំនួនសកលពីរឯករាជ្យៈ ប៉ាន់គំរូតូច

ក្នុងករណីដែលប៉ាន់គំរូតូច n_1និង n_2 ≤ 10 បន្ទាប់ពីបានតម្លៃ T_1អ្នកត្រូវយលតម្លៃ T_1 ទៅប្រៀបធៀបជាមួយវីក្លុសុនគោល (Wilcoxon-critical) ដែលបានកំណត់ក្នុងតារាង E.6

- ករណីតេស្តកន្ទុយសងខាង (រូប 9.3a) បើតម្លៃ $T_1> T_{1U}$ ឬ $T_1< T_{1L}$ បដិសេធ H_0
- ករណីតេស្តកន្ទុយខាងឆ្វេង (រូប 9.3b) បើតម្លៃ $T_1< T_{1L}$ បដិសេធ H_0
- ករណីតេស្តកន្ទុយខាងស្តាំ (រូប 9.3c) បើតម្លៃ $T_1> T_{1U}$ បដិសេធ H_0

ដូចអ្នកដឹងមកហើយ Wilcoxon's Rank Sum Test គឺចេញមកពីធីតេស្តលើប៉ាន់គំរូពីរ (two sample t-test)។ ដូច្នេះ លក្ខខណ្ឌ (assumptions) របស់វាក៏ដូច t-test ដែរ[19]។ ខាងក្រោមនេះគឺជាការរំលឹកលើងវិញនៃលក្ខខណ្ឌរបស់ t-test៖

- ប៉ាន់គំរូត្រូវតែឯករាជ្យពីគ្នា
- ប៉ាន់គំរូទាំងពីរត្រូវមានវ៉ារីយ៉ង់ស្មើគ្នា
- ចំនួនសកលទាំងពីរត្រូវមានលក្ខណៈដូចបំណែងចែកធម្មតា

ក្នុងករណីដែលប៉ាន់គំរូតូច មិនអាចបំពេញបានលក្ខខណ្ឌទាំងបីនេះបាន គឺអ្នកមិនអាចប្រើ t-test បានទេ។ បើបំពេញបានលក្ខខណ្ឌពីរខាងដើមគេ Wilcoxon's Rank Sum Test នឹងត្រូវបានប្រើជំនួស t-test វិញ។

ឧទាហរណ៍9.5 Wilcoxon's Rank Sum Test ក្នុងករណីប៉ាន់គំរូតូច n_1និង $n_2 \leq 10$: ឧបមាថាអ្នកចង់ធ្វើការវាយតម្លៃទៅលើគុណភាពរបស់សកលវិទ្យាល័យពីរ ដោយប្រើអ័រឌីណាល់ស្កែល (ordinal scale) ពិន្ទុចាប់ពី 1 ដល់ 100។ ដោយសាកសួរទៅលើសកវិទ្យាធិការ 9 រូបផ្សេងគ្នា។ បួនរូបវាយតម្លៃលើសកលវិទ្យល័យ A ($n_1 = 4$) និងប្រាំរូប (n_2 =5) វាយតម្លៃលើសកលវិទ្យាល័យ B (ប៉ាន់គំរូពីរឯករាជ្យពីគ្នា)

- សកលវិទ្យាល័យ A មាន 5 រូបបានពិន្ទុៈ 69, 72, 80, 88, 90
- សកលវិទ្យាល័យ B មាន 4 រូបបានពិន្ទុៈ 70, 80, 86, 96

នៅកំរិត α=0.05 តើសាកលវិទ្យាល័យទាំងពីរមានចំណាត់ថ្នាក់ស្មើគ្នាឬយ៉ាងណា?

- **ជំហានទី1:** ដាក់រ៉ង់ឲទិន្នន័យ (rank) (មើលតារាង 9.19)។

  ដូច្នេះយើងបាន$T_1 = 21.5$
  - សកលវិទ្យល័យ B: $n_1 = 4$
  - សកលវិទ្យល័យ A: n_2 =5
- **ជំហានទី2:** កំណាត់តម្លៃវីក្លុសុនគោល (Wilcoxon's critical) បើ α=0.05 យោងតាមតារាង E6-តម្លៃទាប-ខ្ពស់ របស់ T_1 សម្រាប់ Wilcoxon Rank Sum Test យើងនឹងបាន $T_{1L} = 11$និង $T_{1U} = 29$ (មើលតារាង 9.20)។

[19] https://data.library.virginia.edu/the-wilcoxon-rank-sum-test/

តារាង 9.19 Wilcoxon Rank Sum Test ចំពោះប៉ាន់គំរូតូច

សាកលវិទ្យាល័យ		រ៉ង់ (rank)	
A	B	A	B
69		1	
	70		2
72		3	
80		4.5	
	80		4.5
	86		6
88		7	
90		8	
	96		9
ផលបូករ៉ង់ (rank sum)		23.5	21.5

តារាង 9.20 តម្លៃទាប-ខ្ពស់ របស់ T_1 សម្រាប់ Wilcoxon Rank Sum Test

n_2	α		n_1	
	One-Tailed	Two-Tailed	4	5
4				
5	.05	.10	12, 28	19, 36
	.025	.05	11, 29	17, 38
	.01	.02	10, 30	16, 39
	.005	.01	---, ---	15, 40
6				

- **ជំហានទី3:** សេចក្តីសំរេចចិត្ត

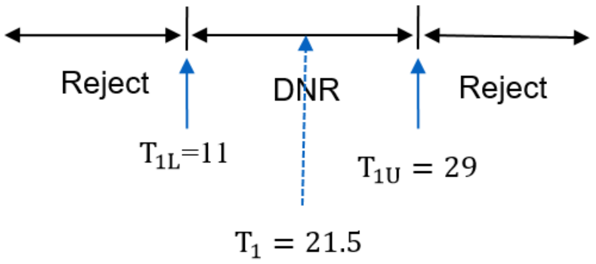
រូប 9.4 បញ្ជាក់ថា $T_1 = 21.5$ ស្ថិតនៅចន្លោះ $T_{1L} = 11$និង $T_{1U} = 29$ គឺតំបន់មិនបដិសេធ ដូច្នេះ $H_0: M_1 = M_2$ មិនត្រូវបានបដិសេធ ហើយយើងសន្និដ្ឋានថា មិនមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ចំណាត់ថ្នាក់របស់សកលវិទ្យាល័យទាំងពីរខុសគ្នានោះទេ។

រូប9.4 Wilcoxon Tail Test

α=0.05, $n_1 = 4$, $n_2 = 5$

H_0: $M_1 = M_2$

H_1: $M_1 \neq M_2$



DNR= Do Not Rejection តំបន់មិនបដិសេធ

9.5.2 វិក្វសុនតេស្តលើប៉ាន់គំរូពីរឯករាជ្យៈ ប៉ាន់គំរូធំ

ឥឡូវយើងធ្វើការពិភាក្សាលើវិក្វសុនតេសត្រសំរាប់ប៉ាន់គំរូពីរឯករាជ្យ Wilcoxon Rank Sum Test for two independent sample វិញម្តង។ តារាង E6- Wilcoxon Test អាចប្រើបានក្នុងករណីដែលប៉ាន់គំរូ n_1និង n_2 មានត្រឹម 10 តែប៉ុណ្ណោះ។ បើក្នុងករណីដែលប៉ាន់គំរូធំ n_1និង n_2 > 10 T_1 មានបំណែងចែកប្រហាក់ប្រហែលទៅនឹងបំណែងចែកធម្មតា ដែលមានតម្លៃធ្យម μ_{T_1} ស្មើទៅនឹង៖

$$\mu_{T1} = \frac{n_1(n+1)}{2} \qquad (9.12a)$$

និងគម្លាតស្តង់ដារ៖

$$\sigma_{T1} = \sqrt{\frac{n_1 n_2 (n+1)}{12}} \qquad (9.12b)$$

បន្ទាប់មក Wilcoxon′ Z-statistic សម្រាប់ចំនួនប៉ាន់គំរូធំ៖

$$Z_{Stat.} = \frac{T_1 - \mu_{T1}}{\sigma_{T1}} \qquad (9.12c)$$

ហើយ $Z_{Stat.}$ មានមានបំណែងចែកប្រហាក់ប្រហែលទៅនឹងបំណែងចែកធម្មតា

សមីការ 9.12 គឺត្រូវបានប្រើប្រាស់ដើម្បីតេស្តលើសម្មតិកម្មមោឃៈពេលដែលចំនួនប៉ាន់គំរូមិនស្ថិតនៅក្នុងតារាង E6 ឬ n_1≥10 និង n_2≥10 និងប៉ាន់គំរូមិនមានលក្ខណៈ normal។

ឧទាហរណ៍9.6 Wilcoxon′s Rank Sum Test ក្នុងករណីប៉ាន់គំរូធំៈ ឧបមាថាយើងធ្វើការពង្រីកចំនួនប៉ាន់គំរូក្នុងការវាយតម្លៃសកលវិទ្យាល័យ A និង B ដោយយក $n_1 = 11$ និង $n_2 = 11$ ហើយទិន្នន័យទាំងនេះបានរៀបចំជារ៉ែងនៅក្នុងតារាង 9.21។

ចំឡើយ៖

- ជំហានទី1: សម្មាតិកម្ម

  $H_0: M_1 = M_2$ និង $H_1: M_1 \neq M_2$

- ជំហានទី2: រកតម្លៃស្ថិតិគោល

  ដោយ $\alpha = 0.05$ តាមតារាង normal distribution យើងបាន $Z = \pm 1.96$

  តារាង 9.19 Wilcoxon Rank Sum Test ចំពោះប៉ាន់គំរូធំ

University		Rank	
A $n_1 = 11$	B $n_2 = 11$	A	B
	50		1
52		2	
	54		3
55		4.5	
	55		4.5
58		6	
69		7	
	70		8
72		9	
80		10.5	
	80		10.5
	86		12
88		13	
90		14	
	91		15
92		16	
	93		17
94		18	
	96		19
	98		20
	99		21
100		22	
Rank Sum		122	130

- ជំហានទី3: គណនាតម្លៃតេស្ត

a. គណនា T_1

បន្ទាប់ពីរៀងរង់រួច (មើលតារាង 9.19) យើងគណនា៖

T_1 =2+4.5+6+7+9+10.5+13+14+16+18+22=122

T_2 = 1+3+4.5+8+10.5+12+15+7+19+20+21=130

សូមរំលឹកថា T_1 = ផលបូកនៃលំដាប់លំដោយដែលបានកំណត់ក្នុងប៉ាន់គំរូទី1 n_1ដែលជាប៉ាន់គំរូដែលតូចជាងគេ (បើប៉ាន់គំរូស្មើគ្នា$n_1 = n_2$ នោះ T_1តំណាងឲ្យប៉ាន់គំរូមួយណាក៏បាន)។

b. តម្លៃមធ្យម μ_{T1}

ក្នុងករណីដែលចំនួនប៉ាន់គំរូធំ n_1និង $n_2 > 10$ យើមិនមានតារាង Wilcoxon ដើម្បីមើល Wilcoxon's critical នោះទេ ដូច្នេះយើងត្រូវប្រើ Z-test។

$$\mu_{T1} = \frac{n_1(n+1)}{2} = \frac{11(22+1)}{2} = 126.5$$

c. គណនាគម្លាតស្តង់ដារ σ_{T1}

$$\sigma_{T1} = \sqrt{\frac{n_1 n_2(n+1)}{12}} = \sqrt{\frac{11*11(22+1)}{12}} = 15.2288$$

d. គណនា Z-test ដោយយក T_1

$$Z_{Stat.}(T_1) = \frac{T_1 - \mu_{T1}}{\sigma_{T1}} = \frac{122 - 126.5}{15.2288} = -0.2955$$

- ជំហានទី 4: ធ្វើសេចក្តីសំរេចចិត្ត

$Z_{Stat.} = -0.2955 > -1.96$ មិនបដិសេធH_0។

- ជំហានទី 5: សន្និដ្ឋាន

មិនមានភាស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា សកលវិទ្យាល័យទាំងពីរនោះមានចំណាត់ថ្នាក់ខុសគ្នានោះទេ។

តារាង 9.20 លទ្ធផលនៃការប្រៀបធៀបចំណាត់ថ្នាក់សកលវិទ្យាល័យ

Test Statistics[a]

	Rating
Mann-Whitney U	56.000
Wilcoxon W	122.000
Z	-.296
Asymp. Sig. (2-tailed)	.767
Exact Sig. [2*(1-tailed Sig.)]	.797[b]

a. Grouping Variable: University

b. Not corrected for ties.

សូមបញ្ជាក់ថា Wilcoxon Rank Sum Test នៅក្នុង SPSS ហៅថា Mann-Whitney Test។ លទ្ធផលនៃ Mann-Whitney Test ក្នុងឧទាហរណ៍របស់យើងគឺបានបង្ហាញក្នុបតារាង 9.20។ សូមធ្វើការផ្ទៀងផ្ទាត់តម្លៃ Z-statistic ក្នុងតារាងនេះជាមួយនឹងឧទាហរណ៍របស់ខាងលើដោយខ្លួនឯង។

9.5.3 វិក្វសុនតេស្តលើប៉ាន់គំរូពីរមិនឯករាជ្យៈ ប៉ាន់គំរូធំ

ក្នុងករណីប៉ាន់គំរូពីរមិនឯករាជ្យឬជាគូ (two related/matched pair sample) យើងអាចប្រើ Wilcoxon Signed-Rank Test ជំនួស Wilcoxon Rank Sum Test វិញ។ តេស្តប្រភេទនេះបានចាត់ទុកជាតេស្ត មុននិងក្រោយព្រឹត្តិការណ៍ទៅលើកម្មវត្ថុតែមួយ ឬជាគូ(before and after test using the same subject or pair test)។ Wilcoxon Signed-Rank Test អាចប្រើជំនួស t-test សម្រាប់ប៉ានគំរូដែលទាក់ទងគ្នា (related sample)[20]។

ឧទាហរណ៍ 9.7 Wilcoxon's Signed-Rank Test

អ្នកគ្រប់គ្រងម្នាក់ចង់ធ្វើការវាយតម្លៃទៅលើកម្មវិធីបំប៉នទៅលើផ្នែកសេវាទំនាក់ទំនងរបស់ធនាគារមួយដើម្បីឲ្យដឹងថាកម្មវិធីនេះមានប្រសិទ្ធភាពឬយ៉ាងណា។ គាត់បានធ្វើការតត់ត្រាចំនួនពាក្យបណ្តឹងពីអតិថិជនទៅលើផ្នែកនេះចំនួន 8 សប្តាហ៍មុនចាប់ផ្តើមវគ្គបំប៉ននិង 8 សប្តាហ៍ ក្រោយពីបំប៉នរួច។ ទិន្នន័យទាំងនោះត្រូវបានកត់ត្រានៅក្នុងតារាង 9.21។

តារាង 9.21 ពិន្ទុវាយតម្លៃផ្នែទំនាក់ទំនងរបស់ធនាគារ

N	Before(B)	After(A)
1	85	75
2	70	50
3	40	50
4	65	40
5	80	20
6	75	65
7	55	40
8	20	25

នៅកំរិត $\alpha = 1\%$ តើមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា កម្មវិធីបំប៉ននេះមានប្រសិទ្ធភាពដែរឬទេ?

ចំឡើយៈ

- ជំហានទី 1: សរសេរសម្មតិកម្ម

[20] Page 688 of Allan G. B (2009) Elementary Statistics: A Step by Step Approach. Eight Edition. McGraw-Hill

H_0: ចំនួនពាក្យបណ្ដឹងមុន ≤ ចំនួនពាក្យបណ្ដឹងក្រោយវគ្គបំប៉ន (កម្មវិធីបំប៉នគ្មានប្រសិទ្ធភាព)

H_1: ចំនួនពាក្យបណ្ដឹងមុន > ចំនួនពាក្យបណ្ដឹងក្រោយវគ្គបំប៉ន (កម្មវិធីបំប៉នមានប្រសិទ្ធភាព)

- ជំហាន **ទី 2:** រកតម្លៃស្ថិតិគោលចេញពីតារាង the Wilcoxon Signed-Rank Test Critical= W_C
  - បើ α = 5% និងចំនួនប៉ាន់គំរូ n =8 តេស្តកន្ទុយខាងឆ្វេង (one-tailed test) យើងបានតម្លៃគោលរបស់ Wilcoxon, W_C= 6។
  - យើងនឹងបដិសេធ H_0 បើសិនតម្លៃតេស្តស្ថិតិ W_S < 6 (តាមការណែនាំរបស់តារាងនេះ)។
- ជំហាន **ទី 3:** គណនាតម្លៃ Wilcoxon Signed-Rank Test, W_S

a. គណនាពិន្ទុខុសគ្នារវាងមុនបំប៉ននិងក្រោយបំប៉ន (B-A) (ជួរឈរទី 4 តារាង 9.22)

b. ដាក់តម្លៃដាច់ខាតទៅលើពិន្ទុខុសគ្នា (B-A) (ជួរឈរទី 5)

c. តំរៀបតាមលំដាប់ពីទាបទៅខ្ពស់ (rank) នូវតម្លៃដាច់ខាត (B-A) (ជួរឈរទី 6)

d. កំនត់សញ្ញា (-/+) ឲ្យស្របតាមពិន្ទុខុសគ្នារវាង B-A (ជួរឈរទី 7)

e. បូកបញ្ចូលគ្នារវាងតម្លៃ (+) និង (+) ហើយ តម្លៃ (-) និង (-) នៅត្រង់កូឡោន (rank)

ផលបូករ៉ង់វិជ្ជមាន: W+= (3) + (6) + (7) + (8) + (3) + (5) = + 32

ផលបូករ៉ង់អវិជ្ជមាន: W- = (-3) + (-1) = - 4

តារាង 9.22 តម្លៃតេស្ត

N	Before (B)	After (A)	D=B-A	|D|	Rank	Signed Rank
1	85	75	10	10	3	+ 3
2	70	50	20	20	6	+ 6
3	40	50	-10	10	3	- 3
4	65	40	25	25	7	+ 7
5	80	20	60	60	8	+ 8
6	75	65	10	10	3	+ 3
7	55	40	15	15	5	+ 5
8	20	25	-5	5	1	-1

f. ជ្រើសរើសតម្លៃដាក់ខាតដែលតូចជាងគេក្នុងចំណោមផលបូកទាំងពីរធ្វើជាតំណាងឲ្យតម្លៃ Wilcoxon តេស្ត។ ដូច្នេះ Wilcoxon Signed Rank Test, W_S= | - 4| = 4

- ជំហាន **ទី 4:** ការធ្វើសេចក្ដីសំរេចចិត្ត

  ដោយសារតម្លៃ Wilcoxon test, W_S= 4 < W_C= 6 ដូច្នេះ H_0 ត្រូវបានបដិសេធ
- ជំហាន **ទី 5:** សន្និដ្ឋាន

  យើងមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា វគ្គបំប៉នគឺមានប្រសិទ្ធភាពនូវកំរិត α = 5%។

❖ **សំគាល់:** ក្នុងករណីដែលចំនួនប៉ាន់គំរូ n ≥ 30 (តារាង Wilcoxon Signed Rank Test មានត្រឹម n=30) នោះ តម្លៃតេស្ត W_S ត្រូវបានជំនួសដោយ Z-test វិញ ពីព្រោះវាត្រូវបានចាត់ទុកដូចបំណែចែកធម្មតា ហើយមានរូមន្តដូចបានកំណត់ក្នុងសមីការ 9.13។

តម្លៃមធ្យម: $\mu = \frac{n(n+1)}{4}$ (9.13a)

គម្លាតស្តង់ដារ: $\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$ (9.13b)

$$Z_{Stat.} = \frac{W_S - \mu_{T_1}}{\sigma} = \frac{W_S - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (9.13c)$$

ហើយ n = ចំនួនគូរដែលមានតម្លៃផលដកខុសពី 0

W_S = តម្លៃដាច់ខាងនៃរង់ដែលតូចជាងគេ

ឥឡូវយើសាក្យល្បងអនុវត្តទិន្នន័យក្នុងឧទាហរណ៍របស់យើង។ ដោយសារតែក្នុងឧទាហរណ៍របស់យើងមានរ៉ងស្មើគ្នា (tied ranks) ដូច្នេះយើងត្រូវកាត់បន្ថយ σ ដោយ $\frac{t^3-t}{48}$ សម្រាប់រ៉ង់នីមួយដែលស្មើគ្នា[21]។ សម្រាប់ឧទាហរណ៍របស់យើងគឺមានរ៉ង់ដែលស្មើគ្នា: 3,3,3 ដូច្នេះត្រូវកាត់បន្ថយ σ ដោយ $\frac{3^3-3}{48}$= 0.5

$$Z_{Stat.} = \frac{4 - \frac{8(8+1)}{4}}{\sqrt{\frac{8(8+1)(2*8+1)}{24} - 0.5}} = \frac{-14}{\sqrt{50.5}} = -1.97$$

នៅកំរិត α=0.05 $Z_{Stat.} = -1.97 < Z_{Critical} = -1.96$ ដូច្នេះ H_0 ត្រូវបានបដិសេធ ហើយយើងរក្សាការសន្និដ្ឋានដូចមុន។

តារាង 9.23 គឺជាលទ្ធផល Wilcoxon Signed Rank Test ដែលគណនាដោយ SPSS។ ចំលើយរបស់ $Z_{Stat.}$ =-1.97 គឺដូចលទ្ធផលរបស់យើងដែរ។

តារាង 9.24 Wilcoxon Signed Rank Test Result (SPSS)

Test Statistics[a]

	After training score - Before training score
Z	-1.970[b]
Asymp. Sig. (2-tailed)	.049

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

9.6 គ្រូស្កាល់វ៉ាលីសតេស្ត

ដើម្បីធ្វើកាប្រៀបធៀបតម្លៃមធ្យមនៃចំនួនសកលបីឬច្រើនយើងប្រើ One-way ANOVA F-test ដោយសន្មត់ថាចំនួនសកលទាំងនោះមានបំណែងចែកធម្មតា និងវ៉ារីយ៉ង់របស់ចំនួនសកលនោះស្មើគ្នា។ បើសិនជាការសន្មត់នេះមិនបានឆ្លើយតបទេនោះ គ្រូស្កាល់វ៉ាលីសហាស់តេស្ត (The Kruskal-Wallis H Test) អាច

[21] Derived from https://www.statisticshowto.com/wilcoxon-signed-rank-test/

ត្រូវបានគេប្រើដើម្បីជំនួស One-way ANOVA F-test ។ The Kruskal-Wallis H Test បានមកពីការពង្រីករូបមន្តមែនវ៉ាយនីយូតេស្ត (an extension of Mann-Whitney U Test) ដើម្បីធ្វើការប្រៀបធៀបតម្លៃមធ្យមនៃចំនួនសកលបីឬច្រើន។

នៅក្នុងការប្រើ គ្រូស្កាល់វ៉ាលីសហាស់តេស្ត ទិន្នន័យត្រូវជាលេខ ឬ លំដាប់លំដោយ ឬ ជាចន្លោះ (continuous, ordinal scale or interval) ហើយត្រូវបំពេញល័ក្ខខ័ណ្ឌមួយចំនួនដូចតទៅ (certain assumptions)៖

- ❖ ល័ក្ខខ័ណ្ឌទី1: ត្រូវមានអថេរមិនឯករាជ្យមួយ (one dependent variable) ជាលេខ ឬ លំដាប់លំដោយ ឬ ជាចន្លោះ
- ល័ក្ខខ័ណ្ឌទី2: ត្រូវមានអថេរឯករាជ្យជាប្រភេទពីរ ច្រើនក្រុម (two or more categorical groups or variables)
- ល័ក្ខខ័ណ្ឌទី3: ត្រូវមានភាពឯករាជ្យ ឬមិនមានទំនាក់ទំនងនៅក្នុងក្រុម ឬរវាងក្រុម (no relationship between the observations in each group or between the groups themselves)។ ឧទាហរណ៍ ម្នាក់ចូលរួមបានតែមួយក្រុមប៉ុណ្ណោះមិនអាចចូលពីរឬច្រើនក្រុមបានឡើយ។ បើសិនជាល័ក្ខខ័ណ្ឌទី3 មិនត្រូវបានផ្ទៀងផ្ទាត់ អ្នកមិនអាចប្រើ The Kruskal-Wallis H Test បានឡើយ ហើយអ្នកអាចប្រើតេស្តផ្សេង ឧទាហរណ៍ដូចជា a Friedman Test ជាដើម[22]។

រូបមន្ត The Kruskal-Wallis H Test ត្រូវបានកំណត់ដោយសមិការ 9.14៖

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right] - 3(n+1) \qquad (9.14)$$

ហើយ $n =$ តម្លៃសរុបនៃប៉ាន់គំរូដែលបានរួមបញ្ចូលគ្នា

$n_j =$ តម្លៃសរុបនៃប៉ាន់គំរូទី j ($j = 1,2 \dots c$)

$T_j =$ ផលបូកនៃរ៉ង់ដែលបានដាក់ឲ្យប៉ាន់គំរូទី j

$T_j^2 =$ ផលបូកនៃរ៉ង់ដែលបានដាក់ឲប៉ាន់គំរូទី j លើកជាការ៉េ

$c =$ចំនួនក្រុម

បើសិចំនួនប៉ាន់គំរូធំ (យ៉ាងតិចបំផុតគឺ 5) H-test មានបំណែងចែកប្រហាក់ប្រហែល ឆាយឈ្វែដែលមាន df=c-1។ ហើយយើងនឹងបដិសេធ H_0 ប្រសិន H-test > χ_α^2។

ឧទាហរណ៍ 9.8 The Kruskal-Wallis H Test: គណៈកម្មាការប្រកួតចង់ធ្វើការប្រៀបធៀបពិន្ទុគណិតវិទ្យារបស់និស្សិតទាំងបីក្រុម ហើយក្រុមនីមួយៗមានសមាជិកប្រាំនាក់។ ហើយគណៈកម្មាការបានធ្វើការកត់ត្រាពិន្ទុសមាជិកក្រុមនីមួយៗរួចដាក់ចំណាត់ថ្នាក់ឬ រ៉ង់ដូចនៅក្នុងតារាង 9.23 ៖

ចំឡើយ៖

- **ជំហានទី1:** កំណត់សម្មតិកម្ម

[22] Derived from https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php

H_0:ពិន្ទុគណិតវិទ្យារបស់ក្រុមទាំងបីស្មើគ្នា

H_1: ពិន្ទុគណិតវិទ្យារបស់ក្រុមទាំងបីស្មើគ្នា

- **ជំហានទី2:** រកតម្លៃស្ថិតិគោល

បើ α =5% ត្រូវបានជ្រើសរើស ហើយឆាយឈ្មៃដែលមាន df=c-1=3-1=2 នោះ យោងតាមតារាង Chi-square distribution យើងបាន $\chi^2_{0.05,2}$=5.991

- **ជំហានទី3:** បានកំណត់រ៉ង់សម្រាប់ពិន្ទុនីមួយៗ

ចំពោះពិន្ទុដែលទាបជាងគេគឺ 40 គឺត្រូវបានដាក់រ៉ង់លេខ 1 ហើយរ៉ង់លេខ 2 ត្រូវបានដាក់លើពិន្ទុ 41 ចំណែកពិន្ទុ 50 មានរហូតដល់ទៅ 2 ត្រូវបានកំនត់ទី (5+6)/2= 5.5 សម្រាប់ពិន្ទុ 50 ទាំងពីរ និងរ៉ង់ដ៏ទៃទៀត ត្រូវបានកំណត់តាមវិធីសាស្ត្រដូចគ្នាជាបន្តបន្ទាប់ (មើលតារាង 9.23 ខាងលើ)។

- **ជំហានទី4:** គណនាតម្លៃតេស្ត

a. រគណនាផលបូកនៃរ៉ង់សម្រាប់ក្រុមនីមួយៗ។

T_1= 7+11+1+9.5+15 = 43.5

T_2= 13+8+3+14+5.5= 43.5

T_3= 5.5+4+12+9.5+2 = 33

តារាង 9.23 ពិន្ទុគណិតរបស់និស្សិតទាំង 3 ក្រុម

ចំនួន	ពិន្ទុ	ក្រុម	រ៉ង់
1	51	1	7
2	70	1	11
3	40	1	1
4	65	1	9.5
5	80	1	15
6	75	2	13
7	55	2	8
8	46	2	3
9	76	2	14
10	50	2	5.5
11	50	3	5.5
12	48	3	4
13	72	3	12
14	65	3	9.5
15	41	3	2

បញ្ជាក់៖ ដើម្បីបញ្ជាក់ថាការគណនាផលបូករ៉ង់នីមួយត្រឹមត្រូវដែរឬយ៉ាងណា អ្នកអាចធ្វើការផ្ទៀងផ្ទាត់ដោយប្រើរូបមន្តៈ

$$T_1+T_2+T_3=\frac{n(n+1)}{2}$$

$$43.5+43.5+33=\frac{15(15+1)}{2}$$

120 = 120 (ការដាក់រ៉ង់គឺត្រឹមត្រូវ)

b. គណនា H-test

$$H=\left[\frac{12}{n(n+1)}\sum_{j=1}^{c}\frac{T_j^2}{n_j}\right]-3(n+1)$$

$$H=\left[\frac{12}{15(15+1)}\left(\frac{43.5^2}{5}+\frac{43.5^2}{5}+\frac{33^2}{5}\right)\right]-3(15+1)$$

H = 0.05 (974.7)-48 =0.735

- **ជំហានទី5:** ការធ្វើសេចក្ដីសំរេចចិត្ត

ដោយ H-test =-0.735 < $\chi^2_{0.05,2}$=5.991 ដូច្នេះ H_0 មិនត្រូវបានបដិសេធ។

- **ជំហានទី6:** សន្និដ្ឋាន

មិនមានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ពិន្ទុគណិតវិទ្យារបស់ក្រុមទាំងបីខុសគ្នានោះឡើយ។

តារាង 9.24 គឺជាលទ្ធផល The Kruskal-Wallis H Test របស់ SPSS។ សូមបញ្ជាក់ថា លទ្ធផលនេះគឺដូចជាមួយលទ្ធផលដែលចេញមកពីរូបមន្ត 9.14 ដែរ។

តារាង 9.24 គឺជាលទ្ធផល The Kruskal-Wallis ទៅលើពិន្ទុគណិតវិទ្យា

Test Statistics[a,b]

	Math score
Chi-Square	.738
df	2
Asymp. Sig.	.692

a. Kruskal Wallis Test

b. Grouping Variable: Group

សេចក្ដីសង្ខេប

នៅក្នុងជំពូកនេះអ្នកបានសិក្សាប្រភេទឆាយស្ក្វែតេស្តជាច្រើនដែលអនុវត្តលើស្ថានភាពផ្សេងគ្នាដូចជា៖ ឆាយស្ក្វែតេស្តចំពោះការអនុលោមតាម ឬភាពសំស្របតាម ដែលត្រូវបានគេប្រើប្រាស់សម្រាប់តេស្តដើម្បីចង់ដឹងថាបំណែងចែកហ្វ្រេកង់អនុលោមតាម ឬគោរពតាមទំរង់ណាមួយជាក់លាក់ដែរឬយ៉ាងណា។ លើសពីនេះទៅទៀតអ្នកបានសិក្សាអំពីការប្រើឆាយស្ក្វែតេស្តទៅលើតម្លៃខុសគ្នារវាងសមាមាត្រពីឬច្រើនដោយផ្អែកលើតារាងខ្វែង និងឆាយស្ក្វែតេស្តដែលចង់ដឹងអំពីភាពទាក់ទងគ្នារវាងសមាមាត្រពីរ។

ក្រៅពីនេះអ្នកក៏បានសិក្សាអំពីការប្រើម៉ាកណោម៉ាតេស្តដើម្បីសិក្សាលើស្ថានភាពប៉ាន់គំរូពីរមិនឯករាជ្យពីគ្នា ពេលដែលទិន្នន័យបំពានលើល័ក្ខខ័ណ្ឌបំណែងចែកធម្មតា។ លើសពីនេះទៅទៀតអ្នកបានសិក្សាលើផលបូករ៉ង់វីកូសុនតេស្ត និងវីកូសុនសញ្ញារ៉ង់តេស្ត ពេលដែលធីតេស្ត បំពានលើល័ក្ខខ័ណ្ឌបំណែងចែកធម្មតា។ ហើយចុងក្រោយអ្នកបានសិក្សាគ្រូស្កាល់វ៉ាលីសតេស្តពេលដែលល័ក្ខខ័ណ្ឌ One Way ANOVA F-test ត្រូវបានបំពាន។

សំណួរនិងលំហាត់

I. សំណួរ

1. តើនៅក្រោមល័ក្ខខ័ណ្ឌអ្វីសខ្លះដែលអ្នកគួរតែប្រើតេស្តដើម្បីកំណត់ថាសមាមាត្រនៃចំនួនសកលពីរដែលឯករាជ្យពីគ្នា មានតម្លៃស្មើគ្នាឬយ៉ាងណា?
2. តើនៅក្រោមល័ក្ខខ័ណ្ឌអ្វីសខ្លះដែលអ្នកគួរតែប្រើតេស្តដើម្បីកំណត់ថាសមាមាត្រនៃចំនួនសកលពីរឬច្រើនដែលឯករាជ្យពីគ្នា ខុសគ្នាឬយ៉ាងណា?
3. តើនៅក្រោមល័ក្ខខ័ណ្ឌអ្វីសខ្លះដែលអ្នកគួរតែប្រើតេស្តដើម្បីកំណត់ថាចំនួនសកលពីរ ឯករាជ្យពីគ្នា?
4. តើនៅក្រោមល័ក្ខខ័ណ្ឌអ្វីសខ្លះដែលអ្នកគួរតែប្រើ McNemar តេស្ត?
5. អ្វីខ្លះគឺជានីតិវិធីនៃការធ្វើតេស្តមិនមែនជាប៉ារ៉ាម៉ែត្រតេស្ត?
6. តើនៅក្រោមល័ក្ខខ័ណ្ឌអ្វីសខ្លះដែលអ្នកគួរតែប្រើ Wilcoxon rank sum តេស្តជំនួស t តេស្តដើម្បីកំណត់តម្លៃមធ្យមពីរខុសគ្នា?

II. លំហាត់

7. ក្នុងចំណោមអ្នកទិញទំនិញតាមបណ្តាយអ៊ីតឺណែត 400 នាក់ ត្រូវបានគេធ្វើការស្ទង់មតិដើម្បីចង់ដឹងអំពីឥរិយាបទរបស់ពួកគាត់។ មានសំណួរជាច្រើនត្រូវបានសួរ ក្នុងនោះមានសំណួរមួយទាក់ទងទៅនឹង "តើអ្នកពេញចិត្តទៅនឹងការទិញទំនិញតាមអ៊ីនតឺណែតដែរឬទេ?" លទ្ធផលនៃការស្ទង់មតិត្រូវបានសង្ខេបនៅក្នុងតារាងខ្វែងដូចខាងក្រោម៖

ពេញចិត្តទៅនឹងការទិញទំនិញ	ភេទ		សរុប
	បុរស	ស្ត្រី	
បាទ	60	200	260
ទេ	40	100	140
សរុប	100	300	400

a. នៅកំរិតអាល់ហ្វា 1% តើសមាមាត្រនៃការពេញចិត្តទិញទំនិញតាមអ៊ីនតឺណែតរបស់បុរសនិងស្ត្រីខុសគ្នាដែរឬទេ?

b. ចូលកំណត់ p-value នៅចំឡើយ (a) និងបកស្រាយអត្ថន័យរបស់វា។

c. ចូលគណនាចំឡើយរបស់អ្នកនៅចំនួរ (a) និង (b) បើបុរស 40 ចូលចិត្តនិង 60 នាក់មិនចូលចិត្តវិញ?

8. ធនាគារមួយចង់ដឹងថា តើមានទំនាក់ទំនងរវាងការងារថែមម៉ោងនិង កំរិតនៃ stress ដែរឬយ៉ាងណា។ ការសិក្សាមួយបានប្រមូលទិន្នន័យទៅលើបុគ្គលិកចំនួន 100 នាក់ បានបង្ហាញទិន្នន័យដូចតទៅៈ

រយៈពេលថែមម៉ោង	កំរិតស្ត្រេស			សរុប
	ខ្លាំង	មធ្យម	តិច	
តិចជាង 30 នាទី	10	8	10	28
30 < 60 នាទី	12	10	10	32
60 នាទីឡើងទៅ	18	12	10	40
សរុប	40	30	30	100

a. នៅកំរិតអាល់ហ្វា 1% តើមានភស្តុតាងបញ្ជាក់ថាមានទំនាក់ទំនងរវាងរយៈពេលថែមម៉ោងនិងកំរិតស្ត្រេសដែរឬទេ?

b. តើចំលើយ (a) ទៅជាយ៉ាងណាបើអាល់ហ្វាស្មើនឹង 0.05?

9. ធនាគារមួយចង់ដឹងអំពីប្រសិទ្ធភាពគោលនយោបាយនៃការផ្តល់កម្ចីថ្មីរបស់ខ្លួន។ មុនកែប្រែគោលនយោបាយចាស់ ធនាគារសួរអតិថិជន 100 នាក់ 60 នាក់ឆ្លើយថា "ខ្ចី" ហើយ 40 ឆ្លើយថា "មិនខ្ចី"។ ក្រោយពីកែប្រែល័ក្ខខណ្ឌឲ្យខ្ចី 70 នាក់ឆ្លើយថា "ខ្ចី" ហើយ 30 ឆ្លើយថា "មិនខ្ចី"។ ទិន្នន័យទាំងនេះបានបង្ហាញក្នុងតារាងខាងក្រោមៈ

មុនកែប្រែលក្ខខណ្ឌឲ្យខ្ចី	ក្រោយពីកែប្រែលក្ខខណ្ឌខ្ចី		សរុប
	ខ្ចី	មិនខ្ចី	
ខ្ចី	55	5	60
មិនខ្ចី	15	25	40
សរុប	70	30	100

a. នៅកំរិតអាល់ហ្វា 5% តើមានភស្តុតាងបញ្ជាក់ថាការកែប្រែលក្ខខណខ្ចីមានប្រសិទ្ធភាពដែរឬទេ? (សមាមាត្រយល់ព្រមខ្ចីមុនកែប្រែល័ក្ខខ័ណ្ឌ តិចជាងសមាមាត្រយល់ព្រមខ្ចីក្រោយការកែប្រែល័ក្ខខ័ណ្ឌ(

b. ចូលកំណត់ p-value នៅចំលើយ (a) និងបកស្រាយអត្ថន័យ

10. អ្នកស្រាវជ្រាវរថយន្តម្នាក់ចង់ដឹងអំពីទំនាក់ទំនងរវាងចំណាយនិងចំងាយបើបររបស់អ្នកប្រើក្នុ។ គាត់បានសួរទៅលើអតិថិជនចំនួន 30 នាក់ ហើយបានគណនាចំងាយប្រើប្រាស់ជាមធ្យមរបស់រថយន្តមួយឬអតិថិជនម្នាក់គឺ 10,000 គីឡូម៉ែត្រ និងគម្លាតស្តង់ដារ 2,400 គីឡូម៉ែត្រ។ តាមបទពិសោធកន្លងមក គម្លាតស្តង់ដារគឺតែ 2,000 គីឡូម៉ែត្រតែប៉ុណ្ណោះ។

a. នៅកំរិតអាល់ហ្វា 5% តើមានភស្តុតាងបញ្ជាក់ថាគម្លាតស្តង់ដាររបស់ចំនួនសកលខុសពី 2,000 គីឡូម៉ែត្រដែរឬទេ?

b. តើត្រូវមានល័ក្ខខ័ណ្ឌអ្វីខ្លះដែលតម្រូវឲ្យមានចំពោះការធ្វើតេស្តនេះ?

c.ចូលកំណត់ p-value នៅចំលើយ (a) និងបកស្រាយអត្ថន័យ។

11. ខាងក្រោមនេះគឺជាពត៌មានទាក់ទងទៅនឹងប៉ាន់គំរូពីរដែលឯករាជ្យពីគ្នា ហើយមានទម្រង់ប្រហែលគ្នា៖

n_1	2.2	2.3	2.9	3.6	12	
n_2	2.7	4.4	4.6	5.6	7	16

a. ចូលដាក់រ៉ង់ (rank ឲ្យប៉ាន់គំរូ)1=រ៉ង់តូចជាងគេ ហើយប៉ាន់គំរូ n= $n_1 + n_2 = 11$)

b. គណនាតម្លៃ តេស្តស្ថិតិ T_1

c. គណនាតម្លៃ តេស្តស្ថិតិ T_2 (ផលបូកប៉ាន់គំរូដែលធំជាងគេ)

d. ចូលផ្ទៀងផ្ទាត់ភាពត្រឹមត្រូវនៃចំលើយដាក់រ៉ង់របស់អ្នកដោយប្រើរូបមន្តដូចតទៅ៖

$$T_1 + T_2 = \frac{n(n+1)}{2}$$

12. អ្នកគ្រប់គ្រងស្ថានីយ៍ធៀកឡានរបស់ក្រសួងសាធារណៈចង់ធ្វើការសិក្សាទៅលើរយៈពេលរងចាំធៀកឡានមួយលើកៗ ក្នុងគោលបំណងកែលំអរយៈពេលធៀកឲមានប្រសិទ្ធភាពជាងមុន ។ គាត់បានធ្វើការកត់ត្រារយៈពេលរងចាំ (គិតជានាទី) រហូតដល់ធៀកចប់ របស់ស្ថានីយ៍បួនកន្លែង ដោយមួយកន្លែជ្រើសរើស ប៉ាន់គំរូចំនួន10 ឡាន។ ទិន្នន័យទាំងនេះត្រូវបានកត់ត្រាក្នុងតារាងខាងក្រោម៖

ស្ថានីយ៍ទី1	ស្ថានីយ៍ទី2	ស្ថានីយ៍ទី3	ស្ថានីយ៍ទី4
60.20	61.33	37.18	38.82
78.15	46.40	68.13	36.35
63.13	53.44	51.08	32.33
75.45	72.30	50.21	52.34
47.24	53.09	58.47	34.13
70.18	27.27	86.29	69.37
48.53	52.46	62.10	78.52
55.43	40.24	44.14	55.15
64.06	53.50	64.17	49.21
64.28	37.28	50.28	66.40

a. នៅកម្រិតអាល់ហ្វា 5% តើមានភស្តុតាងបញ្ជាក់ថារយៈពេលធៀកនៃស្ថានីយ៍នីមួយៗខុសគ្នាឬទេ?

ជំពូកទី 10 សមីការលីនេអ៊ែរ

(Simple Linear Regression)

គោលបំណង

- កំណត់អថេរឯករាជ្យ និងអថេរមិនឯករាជ្យ
- ការអនុវត្តន៍សមីការលីនេអ៊ែរដើម្បីប៉ាន់ស្មានទំនាក់ទំនងជាបន្ទាត់រវាងអថេរពីរ
- ការគណនាមេគុណប៉ាន់ស្មាន
- ការបកស្រាយអំពីមេគុណទំនាក់ទំនង
- ការបកស្រាយលទ្ធផលសមីការលីនេអ៊ែរ
- ការវាស់វែងអំពីបម្រែបម្រួល
- ការគណនាអ៊ែការ៉េ ធីតេស្ត និងអែហ្វតេស្ត
- ការវាយតម្លៃលើល័ក្ខខ័ណ្ឌលីនេអ៊ែរ
- វិធីសាស្ត្រដោះស្រាយពេលមានល័ក្ខខ័ណ្ឌត្រូវបានបំពាន
- ការព្យាករណ៍ជាតម្លៃមធ្យមនិងចន្លោះជឿជាក់របស់អថេរឯករាជ្យ
- កំហុសនៃការប៉ាន់ស្មានសមីការលីនេអ៊ែរ

10.1 ការវិភាគសមីការលីនេអ៊ែរ

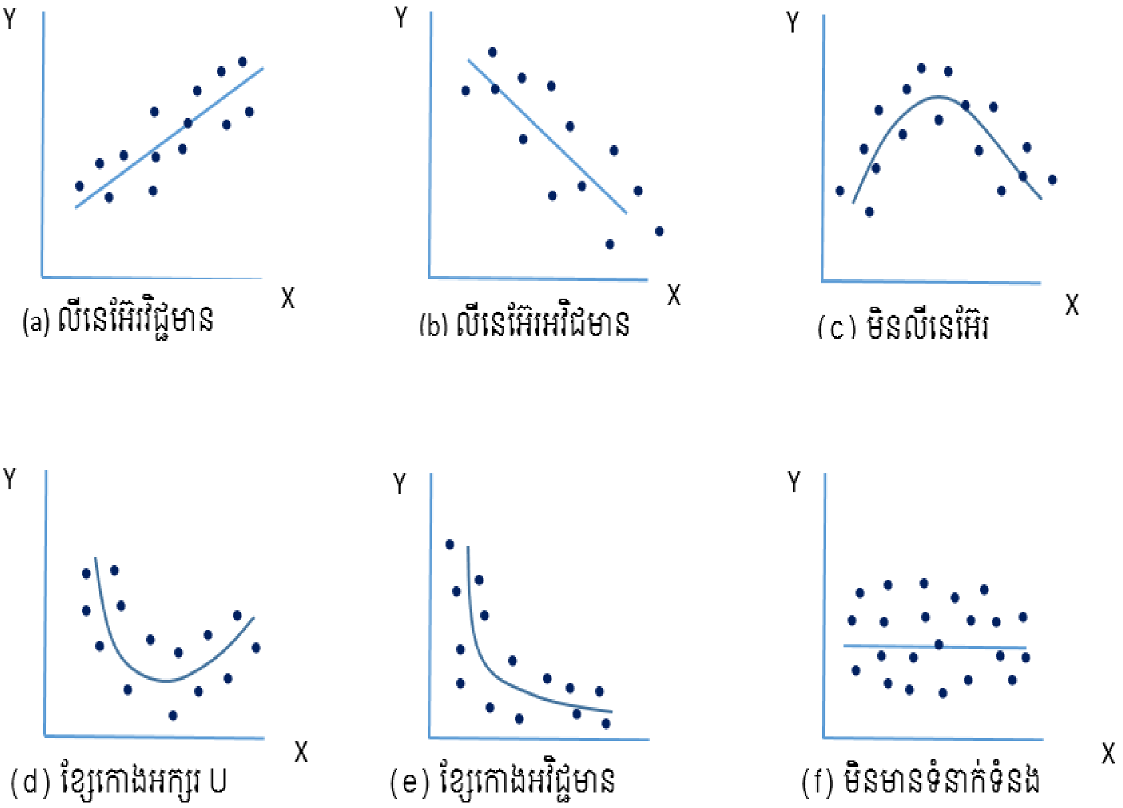
ការវិភាគសមីការលីនេអ៊ែរ(linear regression analysis) គឺជានីតិវិធីនៃការប៉ាន់ស្មានទំនាក់ទំនងរវាងអថេរឯករាជ្យ (dependent variable, regressand or outcome variable) និងអថេរមិនឯករាជ្យមួយឬច្រើន (independent variable, repressor, explanatory, covariate or predictor variable)។ ឧទាហរណ៍ដូចជាទំនាក់ទំនងរវាងការលក់និងការផ្សាយពាណិជ្ជកម្ម៖

ការលក់= អថេរឯករាជ្យ ហើយការផ្សាយពាណិជ្ជកម្ម=អថេរមិនឯករាជ្យ។

ដើម្បីប៉ាន់ស្មានទំនាកទំនងរវាងអថេរទាំងពីរ ជំហានដំបូងគេត្រូវយកទិន្នន័យរបស់ប៉ាន់គំរូមកដាក់បង្ហាញលើរូបភាពស្កាត់តឺផ្លត (មើលក្នុងជំពូក2) បន្ទាប់គេធ្វើការអង្កេតលើចំណុចទាំងឡាយដែលមាននៅលើរូបនោះដើម្បីងាយស្រួលកំណត់អំពីទំនាក់ទំនងរបស់អថេរទាំងពីរនោះជាបន្ទាត់ឬយ៉ាងណា។ បើសិនជាស្កាត់តឺផ្លតបង្ហាញថាមានទំនាក់ទំនងជាបន្ទាប់ហើយ បន្ទាប់ទៀតគឺត្រូវជ្រើសរើសម៉ូដែលគណិតឡាយណាដែលស័ក្តិសមទៅនឹងរូបនោះមកអនុវត្ត។

រូប (a) និង (b) បង្ហាញទំនាក់ទំងជាបន្ទាត់ (វិជ្ជមាន និងអវិជ្ជមាន) ហើយចំណែករូប (c), (d) និង (e) បង្ហាញទំនាក់ទំនងជាខ្សែកោង និងចុងក្រោយរូប (f) បង្ហាញថាមិនមានទំនាក់ទំនងរវាងអថេតទាំងពីរ។

រូប10.1 ស្កាត់តឺផ្លត(scatter plot) បង្ហាញអំពីទម្រង់ផ្សេងៗនៃទំនាក់ទំនងរវាងអថេរ



10.2 សមីការលីនេអ៊ែរ

តាមរយៈរូប Scatter plot ខាងលើបង្ហាញឲ្យឃើញថា ទំនាក់ទំនងរវាងអថេរឯករាជ្យនិងមិនឯករាជ្យ គឺមានទម្រង់ច្រើន ចាប់ពីទម្រង់សាមញ្ញរហូតដល់ស្មុគស្មាញ។ ទំនាក់ទំនងធម្មតាគឺជាទំនាក់ទំនងរាងជាបន្ទាត់ ឬ លីនេអ៊ែរ (simple linear relationship) ដូចបានបង្ហាញក្នុងរូប 10.2 ហើយសមីការលីនេអ៊ែត្រូវបានកំណត់នៅក្នុងសមីការ 10.1។

សមីការលីនេអ៊ែរធម្មតា (simple linear regression model)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (10.1)$$

ហើយ

Y_i= អថេរឯករាជ្យ (dependent variable)

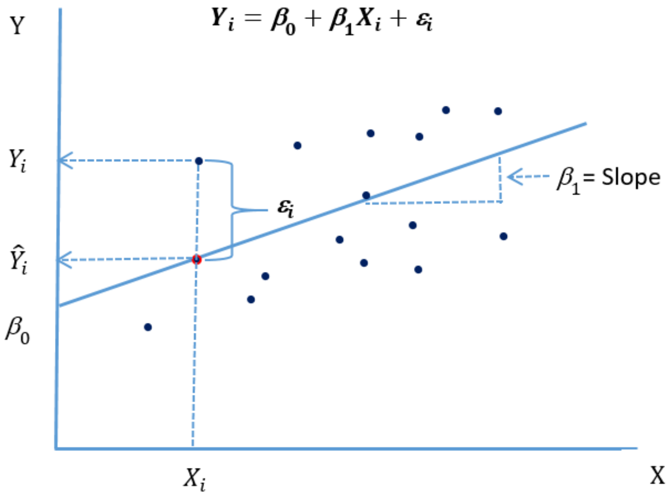
β_0= ចំនួនថេរ (intercept or constant)

β_1= មេគុណរបស់ចំនួនសកល (slope of population)

X_i= អថេរមិនឯករាជ្យ (dependent)

ε_i= អ៊ែរើ ឬលំអៀង (error term)

រូប 10.2 ទំនាក់ទំនងជាបន្ទាត់រវាងអថេរ X និង Y



10.3 ការកំណត់ទំនាក់ទំនងលីនេអ៊ែរ

មានវិធីសាស្ត្រពីរដែលមានដែលត្រូវបាននិយមជាងគេក្នុងការប៉ាន់ស្មានម៉ូដែលលីនេអ៊ែរ (linear regression model) គឺ (Ordinary Least Square, OLS) និង ម៉ាស់ស៊ីម៉ាយឡាយលីហ៊ូត (Maximize Likelihood Estimation, MLE)។ គោលការណ៍របស់ OLS គឺធ្វើឲ្យផលបូកគិតជាការ៉េនៃតម្លៃអ៊ែរើ (តម្លៃអង្កេតដកតម្លៃព្យាករណ៍) តូចបំផុតនៃសំណុំទិន្នន័យដែលបានមក (minimizing the sum of the squares of residuals in given data set) រីឯ MLE គឺជាវិធីសាស្ត្រប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រដោយផ្អែកលើកាសន្មត់លើបំណែងចែកប្រូបាប៊ីលីតេនៃអ៊ែរើ (an assumption about the probability distribution of the disturbance term, u_i) តាមរយៈការធ្វើអតិបរមាកម្មលើមុខងារប្រូបាប៊ីលីតេ (maximizing the probability function) ហើយ MLE គឺមានសភាពស្មុគស្មាញជាង OLS។ ទាំង OLS និង MLE ត្រូវការអ៊ែរើ (errors or residuals) មានបំណែងចែកធម្មតា (normally distributed)។ បើល័ក្ខខ័ណ្ឌ normal distribution ត្រូវបានបំពេញ លោក Hayashi (2000) អះអាងថា OLS មានប្រសិទ្ធភាពជាង[23] ហើយចំណុចមួយទៀតដែលធ្វើឲ្យ OLS ខុសពី MLE គឺថា OLS អាចប៉ាន់ស្មានបានល្អលុះត្រាតែទំនាក់ទំនងនៃប៉ារ៉ាម៉ែត្រទាំងនោះជាបន្ទាត់ ក៏ប៉ុន្តែវិធីសាស្ត្រ MLE អាចអនុវត្តបានដូច OLS ដែរ ហើយថែមទាំងទំនាក់ទំនងនៃប៉ារ៉ាម៉ែត្រនោះមិនមែនជាបន្ទាត់ក៏បាន (can also be applied to regression models that are nonlinear in the parameters[24]) ។ ដោយឡែកលោក Guajarati (2008) បានអះអាងថា ការប្រើ OLS និង MLE ដើម្បីធ្វើការប៉ាន់ស្មានលើមេគុណរបស់ ឬប៉ារ៉ាម៉ែត្រ X គឺមានលទ្ធផលដូចគ្នា ទាំងលីនេអ៊ែរដែលមាន X តែមួយ និង X ច្រើន (simple multiple regression) ប៉ុន្តែ ML estimator នៃ $\sigma^2 = \sum \hat{u}_i^2 / n$ គឺលំអៀង (biased) ដោយឡែក OLS estimator នៃ $\sigma^2 = \sum \hat{u}_i^2 / (n-2)$ គឺមិនលំអៀង (unbiased) នោះឡើយ។ MLE ជាទូទៅ ត្រូវបានគេហៅថា វិធីសាស្ត្រដែលប្រើប៉ាន់គំរូធំ (a large sample method) បើ

[23] Hayashi, Fumio (2000). *Econometrics*. Princeton University Press. p. 15.

[24] ដកស្រង់ពី https://openclassrooms.com/en/courses/5873596-design-effective-statistical-models-to-understand-your-data/

សិនជា ចំនួនប៉ាន់គំរូ n កាន់តែធំ វិធីសាស្ត្រទាំងពីគឺដូចគ្នា (ពីព្រោះ n ≈ n-2) ប៉ុន្តែការណី n តូច OLS មានប្រសិទ្ធភាពជាង[25]។ លើសពីនេះទៅទៀត OLS បានដាក់ល័ក្ខខ័ណ្ឌបន្ថែមមួយទៀតគឺ អ៊ែររើ (errors or residuals, u_i) ត្រូវមានបំណែងចែកធម្មតា (normal distribution) ដែលជាល័ក្ខខ័ណ្ឌសម្រាប់ធ្វើការប៉ាន់ស្មាន និងតេស្តសម្មតិកម្មនៃសមីការលីនេអ៊ែរ។

ដូច្នេះសម្រាប់សៀវភៅនេះយើងនិងធ្វើការវិភាគលើសមីការលីនេអ៊ែរដែលប្រើវិធីសាស្ត្រ OLS តែប៉ុណ្ណោះពីព្រោះ 1) OLS ងាយយល់ ងាយអនុវត្តជាង 2) ការប្រើ MLE និង OLS ដើម្បីធ្វើការប៉ាន់ស្មានលើ β_0និងβ_1 គឺបានលទ្ធផលដូចគ្នា (រួមទាំងប៉ារ៉ាម៉ែត្ររបស់សមីការលីនេអ៊ែរដែលមាន X ច្រើនផង) 3) ដោយសារយើងប្រើចំនួនប៉ាន់គំរូទំហំមធ្យម (moderately large sample) ការប្រើ OLS ឬ MLE ដើម្បីប៉ានស្មានលើវ៉ារីយ៉ង់ σ^2គឺមិនខុសគ្នានោះទេ។

10.3.1 វិធីសាស្ត្រអ័រឌីនែរីលីសស្វែរ (Ordinary Least Square)

សមីការ 10.1 គឺជាសមីការតំណាងឲចំនួនសកល ជាទូទៅតម្លៃប៉ារ៉ាម៉ែត្រ (β_0និងβ_1) របស់សមីការនោះគឺយើងមិនស្គាល់នោះទេ ដូច្នេះដើម្បីធ្វើការប៉ាន់ស្មានលើតម្លៃប៉ារ៉ាម៉ែត្រទាំងនោះ យើងត្រូវផ្អែកលើប៉ាន់គំរូ (sample)។ សមីការរបស់ប៉ាន់គំរូ ត្រូវបានហៅថាលសមីការព្យាករណ៍ (the prediction line equation) ដែលត្រូវបានកំណត់ក្នុងសមីការ 10.2។

សមីការព្យាករណ៍:

$\hat{Y}_i = b_0 + b_1X_i$ (10.2)

ហើយ

$\hat{Y}_i$ = តម្លៃព្យាករណ៍សម្រាប់ប៉ាន់គំរូទី(i)

b_0 =ចំនួនថេររបស់ Y

b_1= មេគុណនៃប៉ាន់គំរូ

X_i =តម្លៃ X សម្រាប់ប៉ាន់គំរូទី(i)

ដូចបានជជែកខាងលើមកហើយ យើងនឹងប្រើវិធី OLS ដើម្បីធ្វើការប៉ាន់ស្មាន b_0និងb_1។ សូមរំលឹកថា វិធីសាស្ត្រ OLS គឺធ្វើឲ្យអ៊ែររើគិតជាការ៉េអប្បរមា៖

Minimize $u_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - b_0 + b_1X_i)^2$ (10.3)

- ជំហ៊ានទី 1: បន្ទាប់មកយើងធ្វើដេរីវេតាមផ្អែកលើសមីការ (10.3) (take partial derivative) ដោយឲ្យដេរីវេនេះស្មើសូន្យ។ ដំបូងនេះយើងសន្មត់ថា b_0 Y_i និងX_i គឺជាចំនួនថេរ ហើយ b_1គឺជាអថេរ។

សូមរំឭក យើងមានរូបមន្តដេរីវេ Y = $U^n \Rightarrow$ Y′ = $nU'U^{n-1}$

$\Rightarrow \frac{\delta \sum u_i^2}{\delta b_1} = 2\sum(0 - 0 - X_i)(Y_i - b_0 - b_1X_i) = 0$

[25] Guajarati D. N, & Porter D. C (2008) Basic Econometrics. Fifth Edition. McGraw-Hill. p. 102.

$= 2\sum(-X_i Y_i + b_0X_i + b_1X_i^2) = 0$

តាមពិត b_0និងb_1 គឺជាចំនួនថេរ យើងអាចទាញចេញពីសញ្ញាផលបូកបាន ហើយ 2 សំរួលចោលរួចធ្វើដេរីវេទីមួយស្មើសូន្យ ដូច្នេះយើងបាន៖

$\sum(-X_i Y_i) + b_0\sum X_i + b_1\sum X_i^2 = 0$

ដោយ $\bar{X} = \frac{\sum X}{n} \Rightarrow \sum X = n\bar{X}$ ជំនួសតម្លៃផលបូកនេះចូលសមីការខាងលើ យើងបាន៖

$\sum(-X_i Y_i) + b_0 n\bar{X} + b_1\sum X_i^2 = 0$

$\Rightarrow b_0 n\bar{X} = \sum(X_i Y_i) - b_1\sum X_i^2$

$\Rightarrow b_0 = \frac{\sum(X_iY_i) - b_1\sum X_i^2}{n\bar{X}}$ ឬ $\frac{\sum(X_iY_i)}{n\bar{X}} - \frac{b_1\sum X_i^2}{n\bar{X}}$ (10.4)

- ជំហ៊ានទី2: បន្ទាប់យើងធ្វើដេរីវេដោយផ្អែកលើសមីកាត (10.3) ម្តងទៀត ស្រដៀងទៅនឹងករណីទីមួយ ដោយចាត់ទុក b_1 Y_i និងX_i គឺជាចំនួនថេរ ហើយ b_0គឺជាអថេរវិញម្តង។

Minimize $u_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - b_0 + b_1X_i)^2$ (10.3)

$\Rightarrow \frac{\delta\sum u_i^2}{\delta b_1} = 2\sum(0 - 1 - 0)(Y_i - b_0 - b_1X_i) = 0$

$= 2\sum(-Y_i + b_0 + b_1X_i) = 0$

ដោយសារ b_0និងb_1 គឺជាចំនួនថេរ យើងអាចទាញចេញពីសញ្ញាផលបូកបាន ហើយ 2 សំរួលចោលរួចធ្វើដេរីវេទីមួយស្មើសូន្យ ដូច្នេះយើងបាន៖

$-\sum Y_i + b_0\sum 1 + b_1\sum X_i) = 0$

ដោយយើងបានដឹងហើយថា $\sum 1 = n$, $\sum X = n\bar{X}$ និង $\sum Y = n\bar{Y}$ ជំនួសទិន្នន័យទាំងនេះចូលដេរីវេខាងលើ យើងបាន៖

$-n\bar{Y} + b_0 n + b_1 n\bar{X} = 0$

សំរួល n ចោល យើងបាន

$\bar{Y} = b_0 + b_1\bar{X}$ (10.5)

- ជំហ៊ានទី3: ជំនួសសមីកាត 10.4 $b_0 = \frac{\sum(X_iY_i)}{n\bar{X}} - \frac{b_1\sum X_i^2}{n\bar{X}}$ ចូលក្នុងសមីការ 10.5 យើងបាន៖

$$\bar{Y} = \frac{\sum(X_iY_i)}{n\bar{X}} - \frac{b_1\sum X_i^2}{n\bar{X}} + b_1\bar{X}$$

ទាញ b_1 ជាកត្តា យើងបាន

$$\bar{Y} = \frac{\sum(X_iY_i)}{n\bar{X}} + b_1(\bar{X} - \frac{\sum X_i^2}{n\bar{X}})$$

បន្ទាប់មកតម្រូវភាគបែងរួចលប់ភាគបែងចោល

$$n\bar{X}\bar{Y} = \sum X_i Y_i + b_1(n\bar{X}^2 - \sum X_i^2)$$

$\Rightarrow b_1 = \frac{n\bar{X}\bar{Y} - \sum X_iY_i}{(n\bar{X}^2 - \sum X_i^2)}$ (10.6)

ដោយ $\sum X = n\bar{X}$ និង $\sum Y = n\bar{Y}$ ដូច្នេះសមីការ (10.6) អាចសរសេរទៅជា៖

$b_1 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$ (10.7)

សមីការ 10.4 និង 10.7 ត្រូវបានគេស្គាល់ថា ជាឧបករណ៍ប៉ាន់ស្មានរបស់ OLS (OLS estimators)

សំគាល់៖

1. សមីការ (10.6) អាចសរសេរជា៖

$$\bar{Y}=b_{0+}b_1\bar{X} \qquad (10.8)$$

2. សមីការ (10.8) អាចសរសេរទៀតជា៖

$$b_0=\frac{\sum Y}{n}-b_1\frac{\sum X}{n} \qquad (10.9)$$

3. ផលគុណនៃអថេរ X ជាមួយនិងផលបូកនៃអ៊ែរើ ឬសំណល់គឺស្មើសូន្យ

$$\sum u_i X_i=0 \qquad (10.10)$$

4. ផលគុណនៃអថេរ Y ជាមួយនិងផលបូកនៃអ៊ែរើ ឬសំណល់គឺស្មើសូន្យ

$$\sum u_i Y_i=0 \qquad (10.11)$$

5. i =ប៉ាន់គំរូទី i ដែលស្ថិតនៅក្នុងទិន្នន័យប្រមូលក្នុងគ្រាតែមួយ (cross sectional data) តែបើទិន្នន័យនោះជាស៊េរីវិញ i ត្រូវបានជំនួសដោយ t វិញ។

ឧទាហរណ៍ 10.1 ការលក់ Online និងការផ្សាយពាណិជ្ជកម្ម

ភ្នាក់ងារស្រាវជ្រាវម្នាក់ចង់ធ្វើការពិនិត្យទំនាក់ទំនងរវាងការលក់និងការចំណាយលើការផ្សាយពាណិជ្ជកម្ម តាមរយៈ Facebook Boost page ហើយគាត់ធ្វើការកត់ត្រានូវចំណូលពីការលក់និងចំណាយលើការផ្សាយពាណិជ្ជកម្មរយៈពេល 16 ខែចុងរបស់ក្រុមហ៊ុនមួយ ហើយទិន្នន័យទាំងនោះមានដាក់ក្នុងតារាង 10.1។

តារាង 10.1 ចំណូលនិងចំណាយលើការផ្សាយពាណិជកម្ម

t	Y($1,000)	X ($)	t	Y($1,000)	X ($)
1	245	140	9	319	140
2	312	150	10	255	160
3	279	170	11	282	170
4	308	180	12	308	190
5	199	110	13	199	120
6	219	155	14	223	158
7	405	235	15	405	280
8	324	245	16	324	220

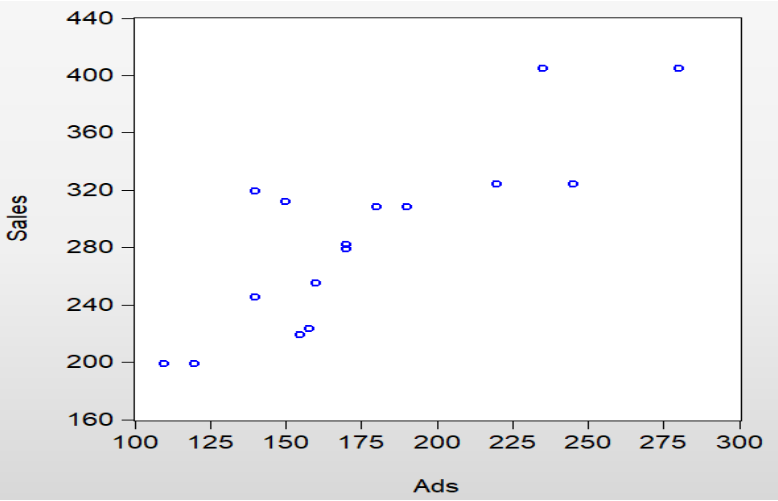
Y= ចំណូលពីការលក់គិតជា ($1000)

X= ចំណាយលើការ Facebook Boost page ឬផ្សាយពាណិជ្ជកម្មគិតជា ($)

t = គិតជាខែ

ជំហានទី1៖ សង់ Scatter Plot

រូប 10.3 Scatter Plot



តារាង 10.2 ជាទិន្នន័យសម្រាប់គណនា b_0 និងb_1

t	Y_i	X_i	X_i^2	Y_i^2	X_iY_i
1	245	140	19600	60025	34300
2	312	150	22500	97344	46800
3	279	170	28900	77841	47430
4	308	180	32400	94864	55440
5	199	110	12100	39601	21890
6	219	155	24025	47961	33945
7	405	235	55225	164025	95175
8	324	245	60025	104976	79380
9	319	140	19600	101761	44660
10	255	160	25600	65025	40800
11	282	170	28900	79524	47940
12	308	190	36100	94864	58520
13	199	120	14400	39601	23880
14	223	158	24964	49729	35234
15	405	280	78400	164025	113400
16	324	220	48400	104976	71280
Total	ΣY=4606	ΣX=2823	$\sum X^2$= 531139	$\sum Y^2$= 1386142	ΣXY= 850074

ជំហានទី2៖ គណនា b_0 និងb_1

ដើម្បីចៀសវាងភាពស្មុគស្មាញក្នុងគណនា b_0 និងb_1 យើងត្រូវអនុវត្តតាមរូបមន្ត 10.7 និង 10.9។ តារាង 10.2 គឺជាទិន្នន័យដែលនឹងយកទៅជំនួសក្នុងរូបមន្តទាំងពីរ។

- គណនា b_1

$$b_1 = \frac{n\sum X_t Y_t - \sum X_t \sum Y_t}{n\sum X_t^2 - (\sum X_t)^2} = \frac{16\,(\mathbf{850074}) - (\mathbf{2823})(\mathbf{4606})}{16(\mathbf{531139}) - (\mathbf{2823})^2} = \frac{598446}{528895} = 1.13$$

- គណនា b_0

$$b_0 = \frac{\sum Y}{n} - b_1\frac{\sum X}{n} = \frac{4606}{16} - 1.13\frac{2823}{16} = 88.23$$

បង្កើតសមីការប៉ាន់ស្មាន (predicted line)

$$\hat{Y}_t = 88.23 + 1.13\,X_t \qquad (10.12)$$

តារាង 10.3 លទ្ធផលសង្ខេបរបស់ Excel

SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.83853				
R Square	0.70313				
Adjusted R Square	0.68193				
Standard Error	35.7254				
Observations	16				
ANOVA					
	df	*SS*	*MS*	*F*	*Sig. F*
Regression	1	42321.45	42321.45	33.1593	4.9E-05
Residual	14	17868.3	1276.307		
Total	15	60189.75			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	88.2355	35.8011	2.4646	0.0272	11.449	165.021
Ads	1.13150	0.19649	5.7584	4.9E-05	0.7100	1.55294

តារាង 10.3 បានផ្តល់លទ្ធផល $b_0 = 88.23$ (coefficient of intercept) និង $b_1 = 1.13$ (coefficient of X) ដូចទៅនិងលទ្ធផលដែលយើងបានគណនាខាងលើ។

- ជំហានទី3៖ សរសេរសមីការប៉ានស្មានៈនិងធ្វើការបកស្រាយ។

  សមីការប៉ានស្មានៈ $\hat{Y}_t = 88.23 + 1.13\,X_t$

- ការបកស្រាយ b_1: មេគុណរបស់ X ត្រូបបានបកស្រាយថា តម្លៃប៉ាន់ស្មានរបស់ Y គិតជាមធ្យម ពេលដែល X កើនឡើងមួយឯកតា។ សម្រាប់ឧទាហរណ៍របស់យើង ពេលដែលការចំណាយទៅលើការការផ្សាយពាណិជ្ជកម្មកើនឡើង $1 ការលក់គិតជា មធ្យមកើនឡើង 1.13 ឯកតា ឬ $1,130 (ការលក់ត្រូវបានកត់ត្រាជា $1000)។ លទ្ធផលនេះគឺជាពត៌មានល្អមួយពីព្រោះវាស្របទៅនឹងទ្រឹស្តី ឬដូចអ្វីដែលយើងបានរំពឹងទុក បើការផ្សាយពាណិជ្ជកម្មកើនឡើង ការលក់រំពឹងថានឹងកើនឡើង។ បើសិន

ជាសញ្ញានៃមេគុណរបស់ X ឬ b_1 ផ្ទុយទៅនិងទ្រឹស្តី លទ្ធផលនៃការប៉ាន់ស្មានប្រាកដជាមានបញ្ហា អ្នកត្រូវរកដំណោះស្រាយជាបន្ទាន់តាមរយៈការពិនិត្យទិន្នន័យឡើងវិញ ថែមទិន្នន័យ ឬថែមអថេរ X ឬអ្នកត្រូវការរកឯកសារយោងដើម្បីមកការពារលទ្ធផលនេះ។

- ការបកស្រាយ b_0: ចំនួនថេរត្រូវបានបកស្រាយថា តម្លៃប៉ាន់ស្មានរបស់ Y គិតជាមធ្យមពេលដែល X=0។ សម្រាប់ឧទាហរណ៍របស់យើង ពេលដែលការចំណាយលើការផ្សាយពាណិជ្ជកម្មស្មើសូន្យ ការលក់គិតជាមធ្យមគឺ 88.23 ឯកតា ឬ $88,230 (ការលក់កត់ត្រា $1000) មានន័យថាការលក់ Online អាចលក់បាន $88,230 ក្នុងមួយខែដោយមិនបាច់ចំណាយទៅលើ Boost Page។ ពេលខ្លះការបកស្រាយb_0 អាចមានន័យសេដ្ឋកិច្ច ហើយពេលខ្លះក៏មិនមានម័យសេដ្ឋកិច្ចដែរ ឧទាហណ៍ យើងមានសមីការប៉ាន់ស្មានមួយៈ $\widehat{\text{តម្លៃផ្ទះ}} = 100 + 500 * \text{ផ្ទៃក្រឡា}\ (m^2)$។ ដើម្បីបកស្រាយចំនួនថេរ (100) គេត្រូវឲ្យផ្ទៃក្រឡាស្មើសូន្យ។ ក្នុងឧទាហណ៍នេះ បើផ្ទះមួយមានផ្ទៃក្រឡាស្មើ 0 ម៉ែត្រការ៉េ តម្លៃផ្ទះគិតជាមធ្យមគឺ $100។ តាមការពិតវាគ្មានន័យជាសេដ្ឋកិច្ចនោះទេ ពីព្រោះមិនមានផ្ទះណាដែលមានផ្ទៃក្រឡាស្មើសូន្យនោះឡើយ។

10.3.2 ការវាស់វែងបម្រែបម្រួល

ពេលប្រើវិធីសាស្រ្ត OLS ដើម្បីកំណត់មេគុណរបស់អថេរ អ្នកត្រូវធ្វើការវាស់វែងអំពីបម្រែបម្រួល (measurement of variation)។ បម្រែបម្រួលសរុប (total variation) ត្រូវបានចែកជាពីរផ្នែកៈ មួយផ្នែកពន្យល់បាន (explained variation) រីឯមួយផ្នែកទៀតគឺពន្យល់មិនបាន(unexplained variation) ឬអ៊ែរើ (errors or residuals)។ ជាទូទៅការវាស់បម្រែបម្រួលគឺត្រូវបានលើកជាការ៉េពីព្រោះមានបម្រែបម្រួល(អ៊ែរើ)ខ្លះវិជ្ជមាន ខ្លះទៀតអវិជ្ជមាន ដែលអាចធ្វើឲ្យផលបូកបម្រែបម្រួលស្មើសូន្យដោយសារតែបញ្ហាសញ្ញា ដូច្នេះអ្នកត្រូវលើកវាជាការ៉េដើម្បីដោះស្រាយបញ្ហានេះ។

- បម្រែបម្រួលសរុបគិតជាការ៉េ (total sum square, TSS) ស្មើនិងបម្រែបម្រួលការ៉េដែលពន្យល់បាន (regression sum square, SSR) បូកនិងបម្រែបម្រួលពន្យល់មិនបាន ឬអ៊ែរើ (residuals or error sum square, SSE)។ រូប 10.4 បង្ហាញអំពីទំនាក់ទំនងនេះ។

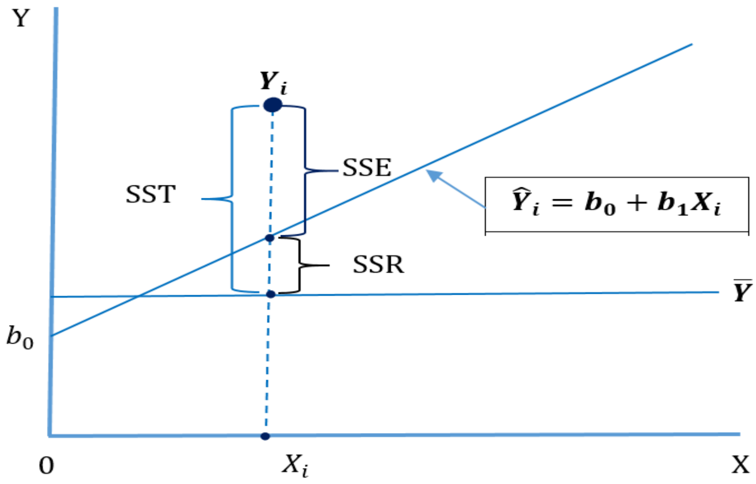
  TSS = SSR + SSE (10.13)
  - $TSS = \sum(Y_i - \bar{Y})^2$
  - $SSR = \sum(\hat{Y} - \bar{Y})^2$
  - $SSE = \sum(Y_i - \hat{Y})^2$

  ហើយ Y = តម្លៃអង្កេតរបស់ Y

  $\bar{Y}$ = តម្លៃមធ្យមរបស់ Y

  $\hat{Y}$ = តម្លៃព្យាករណ៍របស់ Y

រូប10.4 រង្វាស់បម្រែបម្រួល



10.3.3 ការគណនាបម្រែបម្រួល

ឥឡូវយើងត្រឡប់មកពិនិត្យឧទាហរណ៍របស់យើង ហើយធ្វើការគណនាបម្រែបម្រួល។ លទ្ធផលនៃការគណនាបម្រែបម្រួលទាំងនោះមានបង្ហាញក្នុងតារាង 10.4។

តារាង 10.4 គណនាបម្រែបម្រួល

n	Y_i	X_i	$\bar{Y}$	$\hat{Y}$	$(Y_i - \hat{Y})^2$	$(\hat{Y} - \bar{Y})^2$	$(Y_i - \bar{Y})^2$
1	245	140	287.875	246.65	2.709	1700	1838
2	312	150	287.875	257.96	2920	894.9	582
3	279	170	287.875	280.59	2.531	53.06	78.77
4	308	180	287.875	291.91	259	16.25	405
5	199	110	287.875	212.70	187.7	5651	7899
6	219	155	287.875	263.62	1991	588.4	4744
7	405	235	287.875	354.14	2587	4391	13718
8	324	245	287.875	365.45	1718	6018	1305
9	319	140	287.875	246.65	5235	1700	968.8
10	255	160	287.875	269.28	203.8	345.9	1081
11	282	170	287.875	280.59	1.985	53.06	34.52
12	308	190	287.875	303.22	22.84	235.5	405
13	199	120	287.875	224.02	625.8	4078	7899
14	223	158	287.875	267.01	1937	435.2	4209
15	405	280	287.875	405.06	0.003	13731	13718
16	324	220	287.875	337.17	173.3	2430	1305
	$\bar{Y}$ =287.87				SSE=17,868	SSR=42,321	SST=60,190

- $SSE = \sum(Y_i - \hat{Y})^2$= 17,868
- $SSR = \sum(\hat{Y} - \bar{Y})^2$= 42,321

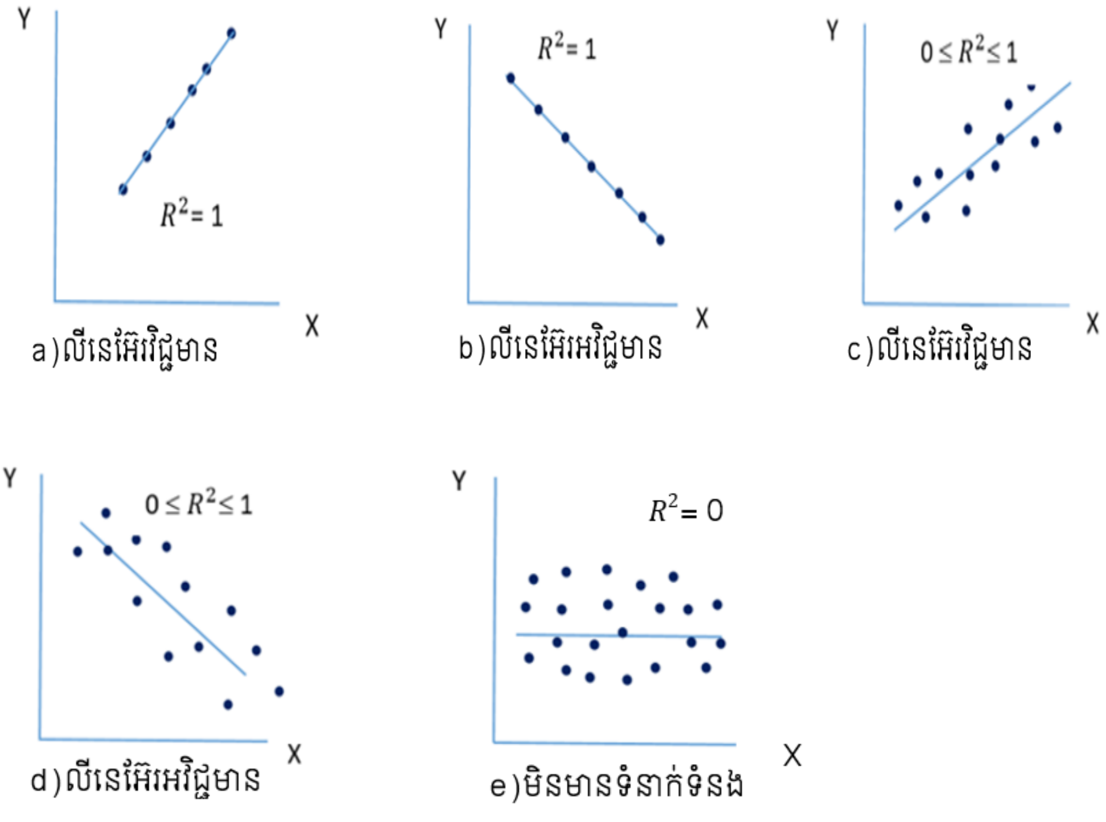
- $TSS = \sum(Y_i - \bar{Y})^2$= 17,868+ 42,321=60,190

10.3.4 គណនាអ៊ែការ៉េ

អ៊ែការ៉េ (R^2)គឺផ្នែកមួយនៃបម្រែបម្រួលសរុបរបស់អថេរឯករាជ្យ ដែលត្រូវបានពន្យល់ដោយបម្រែបម្រួលរបស់អថេរមិនឯករាជ្យ។ ផ្អែកតាមនិយមន័យនេះ អ៊ែការ៉េគឺត្រូវស្ថិតនៅចន្លោះ 0 និង 1 វាវាស់អំពីភាពល្អឬសុក្រិតរបស់ម៉ូដែល (model fit) បើតម្លៃអ៊ែការ៉េកិលខិតជិត 1 មានន័យថាភាពល្អរបស់ម៉ូដែលគឺមានកំរិតខ្ពស់ ហើយបើអ៊ែការ៉េកិលទៅ0 នោះបញ្ជាក់ថា ម៉ូដែលមិនសូវល្អ(រូប10.5 តំណាងឲ្យតម្លៃផ្សេងនៃអ៊ែការ៉េ)។

តើអ៊ែការ៉េត្រូវមានតម្លៃប៉ុន្មានទើបយកជាការបាន? មិនមានចំលើយច្បាស់លាស់នោះទេ។ តម្លៃអ៊ែការ៉េដែលត្រឹមត្រូវគឺអាស្រ័យលើការសិក្សា។ បើកាសិក្សាទៅលើវត្ថុអរូបិយ ឧទាហរណ៍ដូចឥរិយាបថរបស់មនុស្ស គឺអ៊ែការ៉េតិចជាង 50% ក៏ប៉ុន្តែការសិក្សាលើរូបិយវិញ គឺអ៊ែការ៉េតិចជាង 90%[26]។ ជាញឹកញាប់ការសិក្សាលើទិន្នន័យជាស៊េរី អ៊ែការ៉េតែងតែខ្ពស់ជានិច្ច រហូតដល់ជាង 90% ក៏ប៉ុន្តែការសិក្សាលើទិន្នន័យដែលប្រមូលតែមួយគ្រា អ៊ែការ៉េភាគច្រើនទាបជា 90%។

រូប 10.5 តម្លៃផ្សេងៗនៃអ៊ែការ៉េ



[26] ដកស្រង់ចេញពី: https://statisticsbyjim.com/regression/how-high-r-squared/

X

- អ៊ែកាវ៉េត្រូវបានកំណត់ដោយរូបមន្ត 10.13 ខាងក្រោម៖

$$R^2 = \frac{SSR}{SST} \qquad (10.14)$$

- សម្រាប់ឧទាហរណ៍របស់យើង អ៊ែកាវ៉េគឺ៖

$$R^2 = \frac{42,321}{60,190} = 0.70$$ (សូមមើលតម្លៃអ៊ែកាវ៉េក្នុងតារាង10.3)

- **បំណកស្រាយ៖** បម្រែបម្រួលនៃការលក់ត្រូវបានពន្យល់ដោយបម្រែបម្រួលនៃការផ្សាយពាណិជ្ជកម្ម ប្រហែល 70% ហើយនៅសល់ប្រហែល 30%ទៀតដែលការផ្សាយពាណិជ្ជកម្មមិនអាចពន្យល់បាន នេះបណ្តាលមកពីកត្តាផ្សេងទៀតដែលយើងមិនបានបញ្ចូលក្នុងម៉ូដែល។

10.3.5 អ៊ែរើស្តង់ដារ

អ៊ែរើស្តង់ដារ (standard error, SE) គឺជាការវាស់គម្លាតរបស់តម្លៃ X_i ចេញពីបន្ទាត់ព្យាករណ៍ ឬប៉ាន់ស្មាន ($\hat{Y}$)។ ចំពោះរីហ្គ្រេសិនណាដែលមានអ៊ែរើស្តង់ដារតូច ម៉ូដែលនោះព្យាករណ៍បានល្អ និងផ្ទុយមកវិញ បើមានអ៊ែរើស្តង់ដារធំ គឺម៉ូដែលមិនសូវល្អ។ សូមកុំយល់ច្រឡំអ៊ែរើស្តង់ដារទៅនិងគម្លាតស្តង់ដារ (standard deviation)។ គម្លាតស្តង់ដារគឺវាស់បម្រែបម្រួលចេញពីតម្លៃមធ្យម ហើយវាមានគោលដៅចង់ដឹងអំពីភាពល្អ ឬភាពរាយប៉ាយនៃសំណុំទិន្នន័យកំរិតណា។ ដោយឡែកអ៊ែរើស្តង់គឺ វាស់គម្លាតចេញពីបន្ទាត់ព្យាករណ៍ ហើយវាត្រូវបានគេប្រើប្រាស់ដើម្បីធ្វើការព្យាករណ៍ ដូចជាការព្យាករណ៍តម្លៃអថេរមិនឯករាជ្យ ដោយប្រើវិធីសាស្រ្តចន្លោះជឿជាក់ជាដើម (predict interval for Y)។ អ៊ែរើស្តង់ដារត្រូវបានកំណត់នៅក្នុងរូបមន្ត 10.15។

$$SE_{YX} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{SSE}{n-2}} \qquad (10.15)$$

ហើយ

SSE = ផលបូកអ៊ែរើគិតជាការ៉េ

n = ចំនួនប៉ាន់គំរូ

ចំពោះឧទាហរណ៍របស់យើង អ៊ែរើស្តង់ដារគឺ៖

$$SE_{YX} == \sqrt{\frac{17,868}{16-2}} = 35.725$$

អ៊ែរើស្តង់ដារ ក្នុងឧទាហរណ៍របស់យើងគឺ 35.725 ឬ $35,725 (សូមផ្ទៀងផ្ទាត់ទៅនិងតម្លៃ standard error ក្នុងតារាង 10.3 ខាងលើ)។ ដោយសារតែតម្លៃរបស់ Y គឺនៅចន្លោះពីរ $199,000 ទៅ $405,000 ហើយ អ៊ែរើស្តង់ដារស្មើនិង $35,725 គឺត្រូវបានចាត់ទុកថាតូចនៅឡើយ។

10.4 ការសន្និដ្ឋានអំពីមេគុណរបាប់ទិស

ដើម្បីកំណាត់ថាអថេរ X មានសារៈសំខាន់ក្នុងការព្យាករណ៍អថេរ Y ដែរឬអត់ អ្នកត្រូវតែធ្វើតេស្តទៅលើមេគុណរបស់ X ឬហៅថាស្លូប (slope of the regression) ដែលតាងដោយ β_1។ ជាទូទៅ ការកំណត់សារៈសំខាន់នេះត្រូវបានធ្វើឡើងបីបែប 1) t-test និង 2)p-value 3) ការប៉ានស្មានចន្លោះជឿជាក់នៃមេគុណ។

- ធីតេស្តសម្រាប់មេគុណ

ជាទូទៅស្លូបរបស់ចំនួនសកល (slope of the population regression) គឺត្រូវបានសន្មត់ថាស្មើសូន្យ ដូច្នេះសម្មតិកម្មមោឃៈនិងសម្មតិកម្មផ្ទុយគឺ៖

$H_0: \beta_1 = 0$ (មិនមានទំនាក់ទំនងជាបន្ទាត់រវាង X និង Y ឬ ស្លូបស្មើសូន្យ)

$H_0: \beta_1 \neq 0$ (មានទំនាក់ទំនងជាបន្ទាត់រវាង X និង Y ឬ ស្លូបខុសពីសូន្យ)

ដើម្បីតេស្តលើសម្មតិកម្មនេះអ្នកអាចប្រើ t-test ដូចដែលបានកំណត់ក្នុងសមីការ 10.16 ខាងក្រោម៖

$$t_{Statistic} = \frac{b_1 - \beta_1}{S_{b_1}} \qquad (10.16a)$$

t មាន $df. = n - 2$

ហើយ

b_1 = ស្លូបរបស់ regression

β_1 = សម្មាតិកម្មរបស់ស្លូប ($\beta_1 = 0$)

S_{b_1} = អ៊ែរើស្តង់ដាររបស់ស្លូប ដែលត្រូវបានកំណត់ដោយសមីការ 10.16b

$$S_{b_1} = \frac{S_{YX}}{\sqrt{\sum(X_i - \bar{X})^2}} \qquad (10.16b)$$

និង S_{YX} = អ៊ែរើស្តង់ដារនៃការប៉ាន់ស្មាន ដែលកំណត់ដោយសមីការ 10.15 ខាងលើ៖

$$SE_{YX} = \sqrt{\frac{SSE}{n-2}} \qquad (10.15)$$

ឥឡូវយើងពិនិត្យឧទាហរណ៍ 10.1 របស់យើងឡើងវិញ $S_{YX} = 35.726$, $b_1 = 1.1315$ (លទ្ធផលខាងលើ) $\sqrt{\sum(X_i - \bar{X})^2}$= 181.8129 (គណនាដោយខ្លួនឯង)។ ជំនួសទិន្នន័យទាំងនេះចូលក្នុងសមីការ 10.16b យើងបាន៖

$$S_{b_1} = \frac{S_{YX}}{\sqrt{\sum(X_i - \bar{X})^2}} = \frac{35.726}{181.8129} = 0.1965$$

បន្ទាប់មកទៀតយើងគណនា ធីតេស្ត

$$t_{Stat} = \frac{b_1 - \beta_1}{S_{b_1}} = \frac{1.1315 - 0}{0.1965} = 5.7583$$

បើ α= 0.05 ត្រូវបានគេជ្រើសរើស df=16-2=14 តាមតារាង t-distribution យើងបាន $t_{0.05/2,14}$ = 2.1448។ ដោយសារ $t_{Stat} = 5.7583 > t_{0.05/2,14} = 2.1448$ ដូច្នេះ $H_0: \beta_1 = 0$ ត្រូវបានបដិសេធនូវកំរិត α = 0.05 ហើយយើងសន្និដ្ឋានថា មានភស្តុតាងគ្រប់គ្រាន់បញ្ជាក់ថា ការផ្សាយពាណិជ្ជកម្មមានឥទ្ធិពលលើការលក់។

- វិធីសាស្ត្រភីវែលយូ (p-value)

ក្រៅពីការប្រៀបធៀបតម្លៃ t-statistic ទៅនិង t-critical យើអាចមើលតម្លៃ p រួចធ្វើការប្រៀបធៀបទៅនឹង α ។ បើ p-value < α បញ្ជាក់ថា H_0ស្ថិតនៅតំបន់បដិសេធ ហើយផ្ទុយមកវិញ បើ p-value > α គឺ បញ្ជាក់ថា H_0ស្ថិតនៅតំបន់មិនបដិសេធ។ សម្រាប់ឧទាហរណ៍របស់យើង p-value សម្រាប់ t-statistic= 4.9E-05 ឬ = 0.000049 (មើលតារាង 10.3) ដូច្នេះ H_0 ត្រូវបានបដិសេធគ្រប់កំរិតនៃអាល់ហ្វា ទោះជា α= 0.01, α= 0.05 ឬ α= 0.10 ក៏ដោយ។ សូមបញ្ជាក់ថា ការជ្រើសរើសយកវិធីសាស្ត្រ p-value មកធ្វើការវិនិច្ឆ័យ គឺមានលក្ខណៈទូលាយជាងការប្រើ t-statistic ធៀបទៅនិង t-critical។

- ចន្លោះជឿជាក់របស់ស្លូប

ក្រៅពីវិធីសាស្ត្រ t-statistics និង p-value យើងមានវិធីមួយផ្សេងទៀតក្នុងការកំណត់សារៈសំខាន់របស់អថេរ X គឺចន្លោះជឿជាក់ (confidence interval , CI for β_j) របស់ស្លូប ដែលត្រូវបានកំណត់នៅក្នុងរូបមន្ត 10.17។

CI for $\beta_1 = b_1 \pm t_{\alpha/2} S_{b_1}$ (10.17)

t មាន $df. = n - 2$

សម្រាប់ឧទាហរណ៍របស់យើង បើ$\alpha = 0.05$ ត្រូវបានគេជ្រើសរើស យើងបាន $t_{0.05/2,14} = 2.1448$ ហើយ $S_{b_1} = 0.1965$ នោះយើងបាន ចន្លោះជឿជាក់សម្រាប់ β_1គឺ៖

CI for β_1 = $1.1315 \pm 2.1448(0.1965)$

$0.7100 \le \beta_1 \le 1.5529$

ពីព្រោះថាការកត់ត្រាការលក់ជាខ្នង់$1000 ដូច្នេះយើងជឿជាក់ថា 95% ថា β_1 នៅចន្លោះពី 0.7100 (000) ទៅ 1.5529 (000) ឬមានន័យថា ក្នុងមួយដុល្លានៃការផ្សាយពាណិជ្ជកម្មមានឥទ្ធិពលលើការលក់នៅចន្លោះ $710 ទៅ $1,552.9។ ដោយសារនៅចន្លោះនេះមិនមានលេខសូន្យ ដូច្នេះយើងសន្និដ្ឋានថា ការផ្សាយពាណិជ្ជកម្មមានឥទ្ធិពលើការលក់។

10.5 **អែហ្វតេស្តសម្រាប់ម៉ូដែលទាំងមូល**

ជាគោលការណ៍ F-test គឺជាតេស្តជាក្រុម (group test) ហើយត្រូវបានកំណត់ក្នុងរូបមន្ត 10.18។ សម្រាប់សមីការលីនេអ៊ែរ F-test សម្រាប់ម៉ូដែលទាំងមូល (overall significance test) ហើយតេស្តលើសម្មាតិកម្មដូចខាងក្រោម៖

H_0: $\beta_1 = \beta_2 = \cdots \beta_j = 0$ (ម៉ូដែលទាំងមូលមិនមានសារៈសំខាន់)

H_0: យ៉ាងហោចណាស់មាន $\beta \neq 0$ (ម៉ូដែលទាំងមូលមានសារៈសំខាន់)

$$F - Statistic = \frac{\frac{SSR}{K}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE} \quad (10.18)$$

F-statistic មានដឺគ្រីហ្វ្រីឌឹម df. របស់ភាគយក (numerator) ស្មើ k និង ភាគបែង n-k-1

n = ចំនួនប៉ាន់គំរូ

k = អថេរឯករាជ្យ

សម្រាប់ឧទាហរណ៍របស់យើង តាមតារាង 10.3 យើងបាន MSR=42,321.45, MSE=1,276.307, n=16, k=1 នោះយងបាន៖

$$F - Statistic = \frac{42{,}321.45}{1{,}276.307} = 33.1593 \quad (p=4.9E-05)$$

ដោយ p-value =4.5E-05 ឬ =0.000049 < 0.05 ដូច្នេះ H_0 ត្រូវបានបដិសេធ ហើយយើងសន្និដ្ឋានថា ម៉ូដែលទាំងមូលគឺមានសារៈសំខាន់ (significance)។

10.5 **លក្ខខណ្ឌនៃសមីការលីនេអ៊ែរ** (Regression's Assumptions)

នៅក្នុងជំពូកមុនៗ អ្នកបានសិក្សាអំពីការតេស្តលើសម្មតិកម្ម ហើយដើម្បីឲ្យតេស្តទាំងនោះមានសុពលភាព វាត្រូវតែគោរពតាមលក្ខខណ្ឌមួចំនួនដូចជា៖ ទិន្នន័យត្រូវតែមានបំណែងចែកធម្មតា ឬវ៉ារីយ៉ង់ត្រូវ

តែស្មើគ្នាជាដើម។ ស្រដៀងគ្នានោះដែរ ដើម្បីឲ្យលីនេអ៊ែរអាចទុកចិត្តបាន វាត្រូវគោរពតាមល័ក្ខខ័ណ្ឌមួយចំនួនដូចខាងក្រោម៖

- ទំនាក់ទំនងជាបន្ទាត់ (linearity)
- អ៊ែរើត្រូវតែឯករាជ្យ (independence of errors)
- អ៊ែរើត្រូវតែមានលក្ខណៈន័រម៉ាល់ (normality of errors)
- អ៊ែរើត្រូវតែថេរឬដូចគ្នា (equality of variance or errors)

ដូចបានដឹងពីខាងដើមមកហើយ អ៊ែរើ ឬ រេស៊ីតយួល (errors or residuals, u_i) គឺជាតម្លៃខុសគ្នារវាងតម្លៃអង្កេត (Y_i) និងតម្លៃព្យាករណ៍ ($\hat{Y}_i$)។ ការដាក់ល័ក្ខខ័ណ្ឌគឺសំដៅលើតែអ៊ែរើមួយមុខតែប៉ុណ្ណោះ។

10.5.1 ការវាយតម្លៃទំនាក់ទំនងជាបន្ទាត់

ពាក្យថា ទំនាក់ទំនងជាបន្ទាត់ (linearity) សម្រាប់សមីការលីនេអ៊ែរគឺសំដៅទៅលើពីរចំនុច៖

- ទី1: ទំនាក់ទំនងរវាង X និង Y ត្រូវមានលក្ខណៈជាបន្ទត់ (X មិនមានស្វ័យគុណ) មិនមែនជាសមីការដឺក្រេទី2 ទី3 ឬជា exponential នោះឡើយ។ នោះហៅថា ទំនាក់ទំនងរវាងអថេរនិងអថេរ (ឧទាហរណ៍ X និង Y)ត្រូវមានលក្ខណៈជាបន្ទាត់ (linearity in variables)
- ទី2: ទំនាក់ទំនងជាបន្ទាត់ក៏ចង់សំដៅទៅលើប៉ារ៉ាម៉ែត្រ ដូចជា $\beta_1, \beta_2, .. \beta_j$ ផលដែរ។ ដូច្នេះប៉ារ៉ាម៉ែត្រមិនត្រូវមានស្វ័យគុណនោះទេ។ ក្នុងករណីដែលប៉ារ៉ាម៉ែត្រមានជាប់ស្វ័យគុណ ឧទាហរណ៍ សមីការ Cobb–Douglas (CD) production function: $Q_i = \beta_1 L_i^{\beta_2} K_i^{\beta_3}$ ដែល Q= ទិន្នផល (output), L= កំលាំងពលកម្ម (Labor) និង K=មូលធន (capital)។ សមីការបែបនេះមិនមានលក្ខណៈបន្ទាត់នោះទេ ដូច្នេះប៉ាន់ស្មានសមីការនេះ គេត្រូវបំលែងវាឲ្យទៅជាបន្ទាត់សិន៖ ln(Q_i)=ln(β_1)+ $\beta_2 \ln(L_i)$+ $\beta_3 \ln(K_i)$។ ចំណែកឯការប៉ាន់ស្មាននិងការបកស្រាយលទ្ធផលនៃសមីការនេះ សូមអានសៀវភៅ Econometrics By Examples របស់លោក Guajarati (2012)។

សម្រាប់លោក Guajarati គឺថា ម៉ូដែលដែលមានរាងជាបន្ទាត់ (Linear Regression Model) លុះត្រាតែម៉ូដែលនោះ៖

- រាងជាបន្ទាត់ទាំងប៉ារ៉ាម៉ែត្រនិងទាំងអថេរ
- រាងជាបន្ទាត់តែប៉ារ៉ាម៉ែត្រមិនចំបាច់អថេរក៏បាន

ហើយការកំណត់ម៉ូដែលដែលមានរាងជាបន្ទាត់បានរៀបរាប់ក្នុងតារាង 10.5។

តារាង 10.5ការកំណត់ម៉ូដែលដែលមានរាងជាបន្ទាត់

ម៉ូដែលដែលមានប៉ារ៉ាម៉ែត្រជាបន្ទាត់?	ម៉ូដែលដែលមានអថេរជាបន្ទាត់?	
	Yes	No

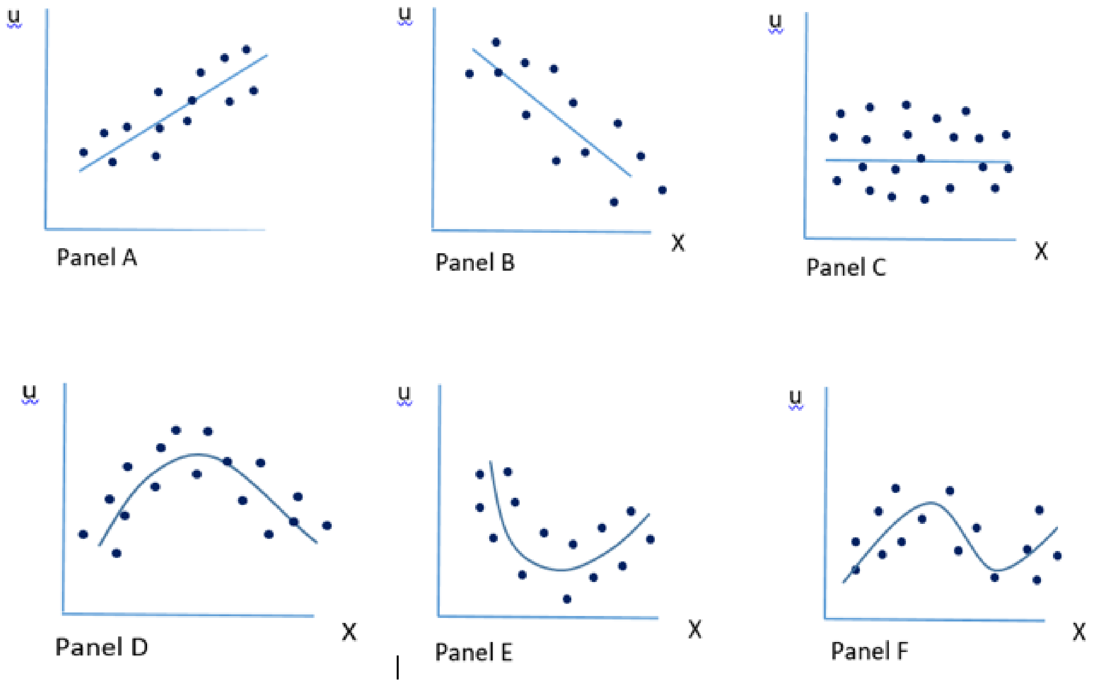
Yes	LRM	LRM
No	NLRM	NLRM

LRM= Linear Regression Model (លីនេអ៊ែរម៉ូដែល)

NLRM= Non-Linear Regression Model (មិនលីនេអ៊ែរម៉ូដែល)

ប្រភព: Guajarati (2008) ទំព័រ 39

រូប10.6 ទម្រង់ផ្សេងៗនៃ Scatter Plotរបស់ Residuals



ដើម្បីវាយតម្លៃទំនាក់ទំនងជាបន្ទាត់នៃសមីការលីនេអ៊ែរ គេមានវិធីសាស្ត្រវាយដ៏សាមញ្ញមួយគឺ បន្ទាប់ពីបាន់ស្មានសមីការព្យាករណ៍រួច អ្នកត្រូវយកអ៊ែររើ ឬ រេស៊ីតយួល (residuals, $u_i = Y_i - \hat{Y}_i$) មកតាងជា Scatter Plot ដោយដាក់រេស៊ីតយួលស្ថិតលើអ័ក្សឈរ ហើយ Xស្ថិតនៅលើអ័ក្សដេក (មើលរូប10.6) រួចធ្វើការអង្កេតលើរូបរាង ឬទម្រង់នៃ Scatter Plot នោះ។ បើ residuals មានរាងដូចរូប A B និង C យើងអាចសន្និដ្ឋានថា ទំនាក់ទំនងរវាង X និង Y មានលក្ខណៈជាបន្ទាត់ ហើយបើមានរាងដូចរូប D E និង F នោះ ទំនាក់ទំនងរវាង X និង Y មិនមានលក្ខណៈជាបន្ទាត់ អាចជាសមីការដឺក្រេទី2 ទី3 ឬ exponentialថែមទៀតផង។

សម្រាប់ឧទាហរណ៍របស់យើង residuals គឺបានគណនាក្នុងតោរាង10. 6 (ជួរឈរទី5) ដាក់ង្ហាញលើរូប 10.7។

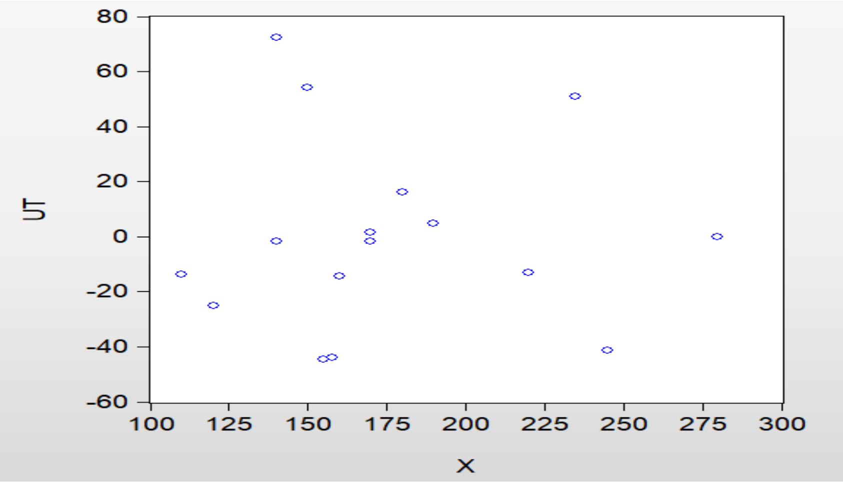
តាមរយៈ residuals ក្នុងរូប 10.7 យើងអាចអង្កេតឃើញថា ទំនាក់ទំនងរវាង X និង Y មាន រាងដូចរូប10.6 C ដូច្នេះយើងសន្និដ្ឋានថា ទំនាក់ទំនងរវាង X និង Y គឺប្រហែលជាបន្ទាត់។

តារាង10. 6 លទ្ធផលនៃ residuals

n	Y_i	X_i	$\hat{Y}$	$\hat{u}_t = Y_i - \hat{Y}$

1	245	140	246.65	-1.646
2	312	150	257.96	54.039
3	279	170	280.59	-1.591
4	308	180	291.91	16.094
5	199	110	212.7	-13.7
6	219	155	263.62	-44.62
7	405	235	354.14	50.861
8	324	245	365.45	-41.45
9	319	140	246.65	72.354
10	255	160	269.28	-14.28
11	282	170	280.59	1.409
12	308	190	303.22	4.779
13	199	120	224.02	-25.02
14	223	158	267.01	-44.01
15	405	280	405.06	-0.056
16	324	220	337.17	-13.17

រូប 10.7 គំនូសតាងScatter Plot នៃ Residuals



10.5.2 ការវាយតម្លៃលើភាពឯករាជ្យ

ភាពឯករាជ្យគឺចង់សំដៅលើរ៉េស៊ីតយួល (residuals) នៃគ្រានីមួយៗមិនទាក់ទងគ្នា ឬនិយាយមួយបែបទៀតគឺ កូវ៉ារីយ៉ង់របស់រ៉េស៊ីតយួល ស្មើសូន្យ $[Cov\left(u_i, u_j | X_i, X_j\right) = Cov\left(u_i, u_j\right) = 0]$។ បើ រ៉េស៊ីតយួល មិនឯករាជ្យឬទាក់ទងគ្នា គេហៅថាអូតូកូរីលេសិន ឬសីរាលកូរីលេសិន (auto-correlation or serial

correlation)។ ឧទាហរណ៍ កូដកម្មនៅត្រីមាសទីមួយមិនជះឥទ្ធិពលទៅលើផលិតផលនៅត្រីមាសទីពីរ។ ដូច្នេះយើងមិនជឿថា បើផលិតផលទាបនៅត្រីមាសទីមួយ វានឹងទាបនៅត្រីមាសទីពីរនោះទេ។ ឧទាហរណ៍ទាំងពីរនេះគឺបញ្ជាក់ថារ៉េស៊ីតយូលនៃគ្រានីមួយៗមិនទាក់ទងគ្នា។

នៅក្នុងទិន្នន័យដែលប្រមូលតែមួយគ្រា (cross sectional data) រ៉េស៊ីតយូលអាចមានទំនាក់ទំនងពីអង្គភាពមួយទៅអង្គភាពមួយទៀត។ ទំនាក់ទំនងនេះហៅថាស្បាស្យ៉ាលកូរីលេសិន (spatial correlation)។ មិនមែនទាក់ទងជាមួយពេលវេលានេះទេ តែវាទាក់ទងជាមួយនឹងចន្លោះ ពីអង្គភាពមួយទៅមួយទៀត (correlation in space rather than over time)។ ករណីនេះច្រើនតែកើតចេញពីការប្រមូលទិន្នន័យជាក្រុម (cluster sampling) ជាពិសេសទាក់ទងជាមួយរង្វាស់ដូចគ្នា វាស់វែងលើអថេរតែមួយ តួយ៉ាងដូចជា ការសាកសួរសមាជិកច្រើននាក់នៅក្នុងគ្រួសារតែមួយ ឬក្រុមហ៊ុនជាច្រើនធ្វើប្រត្តិបត្តិការនៅក្នុងទីផ្សារតែមួយ។ ឧទាហរណ៍ទី1 អ្នកធ្វើការសាកសួរដោយចៃដន្យទៅលើមនុស្សបីនាក់ទាក់ទងនឹងការចំណាយក្នុងគ្រួសារ ហើយមនុស្សទាំងបីនាក់នោះហូបបាយរួមគ្នាក្នុងគ្រួសារតែមួយ ដូច្នេះចំណាយក្នុងគ្រួសាររបស់អ្នកទាំងបីគឺតែមួយ (ទាក់ទងគ្នា) ។ ឧទាហរណ៍ទី2 ចំណូលរបស់ថៅកែនិងបុគ្គលិកគឺទាក់ទងគ្នា បើសិនជាថៅកែប្រើវិធីសាស្ត្រលើកទឹកចិត្តបុគ្គលិកតាមរយៈបរិមាណលក់ បើលក់បានច្រើន បុគ្គលិកបាន bonus ច្រើន។ អ្នកទាំងពីរហាក់ដូចជាធ្វើប្រតិបត្តិការក្នុងទីផ្សារតែមួយ។ ដូច្នេះដើម្បីឲ្យការវិភាគលីនេអ៊ែរឲ្យមានន័យ អ្នកត្រូវកំណត់ថា ស្បាស្យ៉ាលកូរីលេសិន មានវត្តមាននៅក្នុងទិន្នន័យរបស់ដែរអត់។ ការធ្វើតេស្តលើស្បាស្យ៉ាលកូរីលេសិន គឺមានភាពស្មុគស្មាញជាងការធ្វើតេស្តលើទិន្នន័យជាសេរី ហើយមិនបានបរិយាយក្នុងសៀវភៅនេះទេ បើអ្នកអានមានចំណាប់អារម្មណ៍ សូមមើលសៀវភៅរបស់លោក Anselin (1988)។

ដោយឡែកចំពោះទិន្នន័យជាសេរី (time series data) វិញ ការប្រមូលទិន្នន័យមានទំនាក់ទំនងតាមពេលវេលា ឬទាក់ទងពីគ្រាមួយ ទៅគ្រាមួយទៀត ជាពិសេសចន្លោះគ្រានីមួយៗខ្លី ដូចជា មួយថ្ងៃម្តង ឬមួយសប្តាហ៍ម្តង ជាដើម។ ទិន្នន័យទាំងនេះហាក់ដូចបង្ហាញថាមានទំនាក់ទំនងនិងគ្នា (inter-correlation) ។ ក្នុងស្ថានភាពបែបនេះ ល័ក្ខខ័ណ្ឌ no auto-correlation ហាក់ដូចត្រូវបានបំពាន។ ឧទាហរណ៍ផ្សេងទៀតដូចជា ការធ្វើជាកូដកម្មនៅត្រីមាសមុនជះ ឥទ្ធិពលទៅលើផលិតផលនៃត្រីមាសក្រោយ ឬ ចំណូលគ្រួសារមួយនៅគ្រាមុនទាក់ទងទៅនឹងចំណូលគ្រួសារនោះនៅពេលក្រោយ គេហៅថា serial correlation ឬ auto-correlation។

- មូលហេតុដែលបណ្តាលឲ្យមានសឺរាលកូរីលេសិន (serial correlation[27]):
    a. ប្រើទម្រង់មិនត្រឹមត្រូវ (functional form)
    b. ភ្លេចអថេរសំខាន់នៅក្នុងម៉ូដែល (specification error)

---

[27] ដកស្រង់ចេញពី Guajarati D. N & Porter D. C (2008) Basic Econometrics. Fifth Edition. McGraw-Hill (p.452)

c. បាតុភូតខូបវែប (Cobweb) ឧទាហរណ៍ការសំរេចចិត្តដាំដំណាំរបស់កសិករឆ្នាំនេះ គឺរងឥទ្ធិពលពីតម្លៃកាលពីឆ្នាំមុន
d. បំលែងទិន្នន័យ (data massaging or manipulation) ឧទាហរណ៍យើងអាចគណនាទិន្នន័យប្រចាំខែតាមរយៈទិន្នន័យត្រីមាសចែកនិងបី ឬ បំលែងទិន្នន័យ (data transformation) ដូចជាការបំលែងទិន្នន័យដែលកត់ត្រាជម្មតាឲ្យទៅជាទិន្នន័យបម្រែបម្រួល (taking first difference)។

- ផលវិបាកដែលបណ្តាលមកពីសឺរាលកូរីលេសិនៈ ទោះបីជាមានវត្តមានរបស់ serial ឬ auto-correlation នៅក្នុងទិន្នន័យ ក៏មិនធ្វើអោយមេគុណប៉ាន់ស្មានលំអៀងដែរ (unbiased estimator) ក៏ប៉ុន្តែមេគុណប៉ាន់ស្មានគ្មានប្រសិទ្ធភាព (no longer efficient) ទីបំផុតបណ្តាលឲ្យ t-test, F-test និង χ^2 អនុវត្តមិនស្របតាមល័ក្ខខ័ណ្ឌ (cannot legitimately applied) ដូច្នេះអ្នកត្រូវរកដំណោះស្រាយ។
- រកដំណោះស្រាយពេលមាន សឺរាលកូរីលេសិន គឺអាស្រ័យលើភាពទាក់ទងគ្នាជាមួយរ៉េស៊ីតយួលដោយសារតែ សឺរាលកូរីលេសិនគឺបណ្តាលមកពី រ៉េស៊ីតយួលជាប់ទាក់ទងគ្នាពីគ្រាមួយទៅគ្រាមួយទៀត ដូច្នេះដើម្បីដោះស្រាយ សឺរាលកូរីលេសិនក្នុងករណីដែលយើងដឹងមេគុណទំនាក់ទំនង (ρ) យើងអាចប្រើយន្តការលំដាប់ទីមួយនៃអូតូរីហ្គ្រេស៊ីវ (first order auto-regressive, AR(1)) ឬអូតូរីហ្គ្រេស៊ីវច្រើនដង AR(p) វាជាទម្រង់របស់នីតិវិធីបម្រែបម្រួល (differencing procedure) ជំរើសទីពីរគឺយើងអាចប្រើយន្តការមូវិញអាវើរេច (moving average, MA) ឬ MA លាយជាមួយ AR យន្តការនេះហៅថា ARMA។ បើករណីដែលយើងមិនដឹងមេគុណទំនាក់ទំនង (ρ) យើងអាចប៉ាន់ស្មានវាតាមរយៈ Durbin-Watson ឬវិធិសាស្តដ៏ទៃទៀត និង ជំរើសដោយឡែកមួយទៀតគឺយើងអាចប្រើយន្តការ Newey–West HAC ។ យន្តការចុងក្រោយនេះមិនត្រឹមតែកែ serial correlation ប៉ុណ្ណោះទេ តែអាចកែហិតេរ៉ូស្តេដាស្ទីកស៊ីតធី (Heteroscedasticity) ថែមទៀត។ សម្រាប់សៀវភៅនេះ មិនបាននិយាយដល់ដំណោះស្រាយទាំងនេះទេ បើអ្នកអានមានចំណាប់អារម្មណ៍ សូមអានសៀវភៅរបស់លោក Guajarati (2008) ជំពូកទី12។

ដើម្បីវាយតម្លៃលើល័ក្ខខ័ណ្ឌនេះយើងមានវិធីសាស្រ្តពីរៈ ការដាក់បង្ហាញ residuals លើរូបភាព (graphical method) និង ប្រើតេស្តស្ថិតិ (statistic test) ដូចជា the Run test (តេស្តនេះមិនត្រូវការល័ក្ខខ័ណ្ឌ normality of error term), the Durbin-Watson D test, និង the **Breusch–Godfrey test ឬ** ហៅថា **Lagrange multiplier (LM) test** ជាដើម។ តែសម្រាប់សៀវភៅនេះ និងលើកយកតែវិធីសាស្រ្តរូបភាព និង Durbin-Watson មកបរិយាយតែប៉ុណោះ ចំណែកឯតេស្តដ៏ទៃទៀត អ្នកអានអាចមើលក្នុងសសៀវភៅរបស់លោក Guajarati។

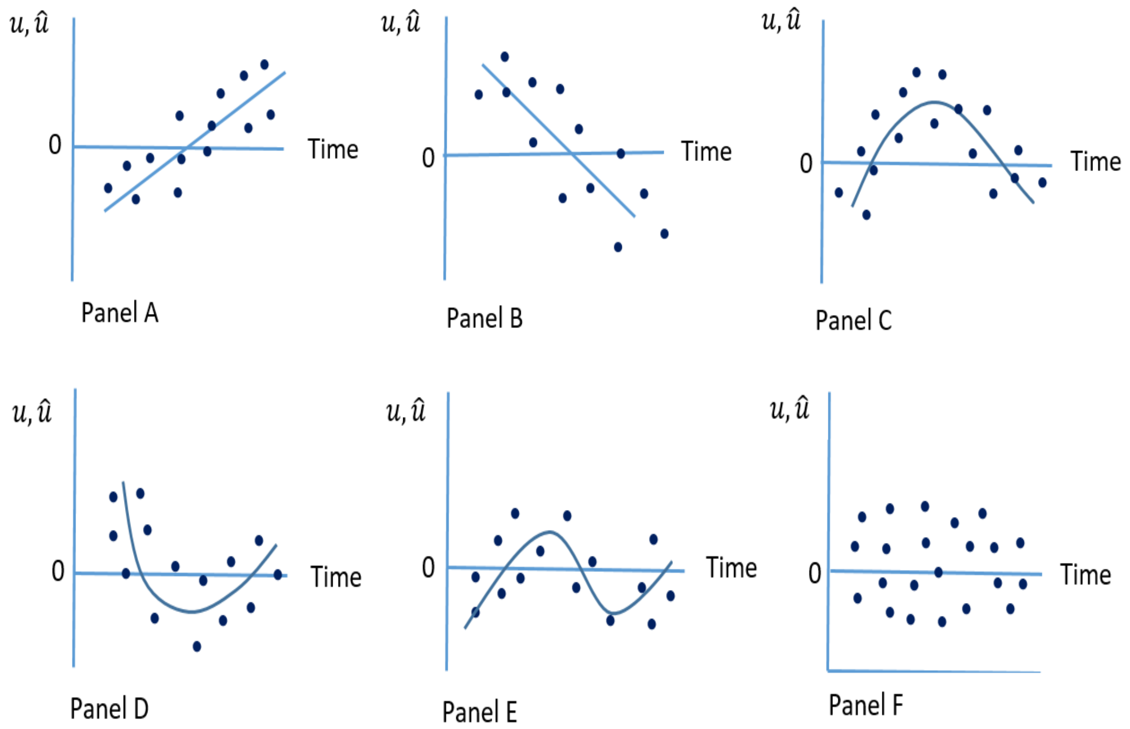
វិធីសាស្ត្ររូបភាព (graphical method)

បន្ទាប់ពីបានសមីការប៉ាន់ស្មានរួច អ្នកយករ៉េស៊ីតយួលរបស់សមីការនោះទៅដាក់លើរូបស្កាត់តឺផ្លត(Scatter Plot) ដោយយក រ៉េស៊ីតយួល (u, $\hat{u}$) ដាក់លើអ័ក្សឈរ និងពេលវេលា (time) ដាក់លើអ័ក្សដេក (មើលរូប 10.8) រួចធ្វើការអង្កេតទម្រង់រូបទាំងនោះ។

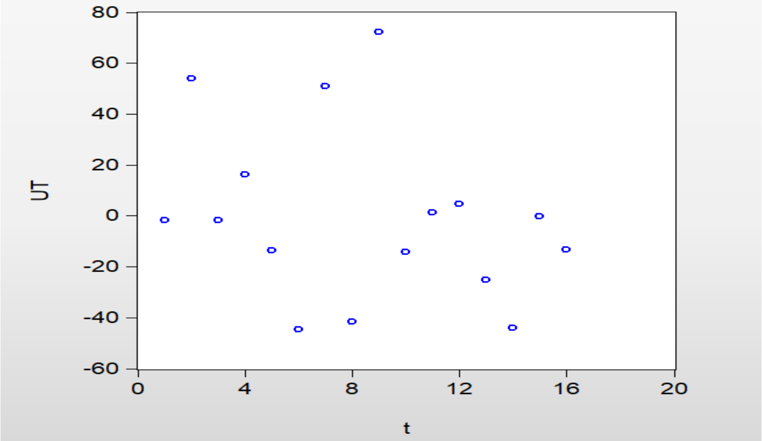
ក្នុងរូប 10.8 រូប A និង B បង្ហាញពីទម្រង់ឡើងចុះ ហើយ C និង D បង្ហាញទម្រង់ជាសមីការដឺក្រេទីពីរ និងរូប E បង្ហាញទម្រង់ឡើងចុះ ហើយមានតែរូប F មួយទេ ដែលមិនមានទំរង់អ្វីទាំងអស់។ រូប F បង្ហាញថា គ្មានបញ្ហាអូតូកូរីលេសិន (no autocorrelation)។

ចំពោះឧទាហរណ៍ 10.1 របស់យើង រ៉េស៊ីតយួល ត្រូវបានដាក់បង្ហាញលើរូប 10.9។ តាមរយៈរូបនេះបង្ហាញថា រ៉េស៊ីតយួលហាក់ដូចមានទម្រង់រូប F ដែលចំនុចនីមួយៗមានលក្ខណៈឆ្លាស់ចុះឡើងយ៉ាងទៀងទាត់ ឬ មានលក្ខណៈចៃដន្យ (random) ដូច្នេះយើងសន្និដ្ឋានថា មិនមានវត្តមាន អូតូកូរីលេសិនក្នុងរ៉េស៊ីតយួលនោះទេ។

រូប 10.8 ទម្រង់ auto-correlation និង non-autocorrelation



រូប 10.9 Residuals Plot



ដឺប៊ីនវ៉តសុនតេស្ត

តេស្តភាគច្រើនដែលរកទំនាក់ទំនងជាស៊េរី (serial correlation)គឺ ដឺប៊ីនវ៉តសុនតេស្ត (Durbin –Watson d statistic)។ តេស្តនេះត្រូវបានបង្កើតឡើងដោយអ្នកស្ថិតិហើយត្រូវបានកំណត់នៅក្នុងសមីការ 10.19។

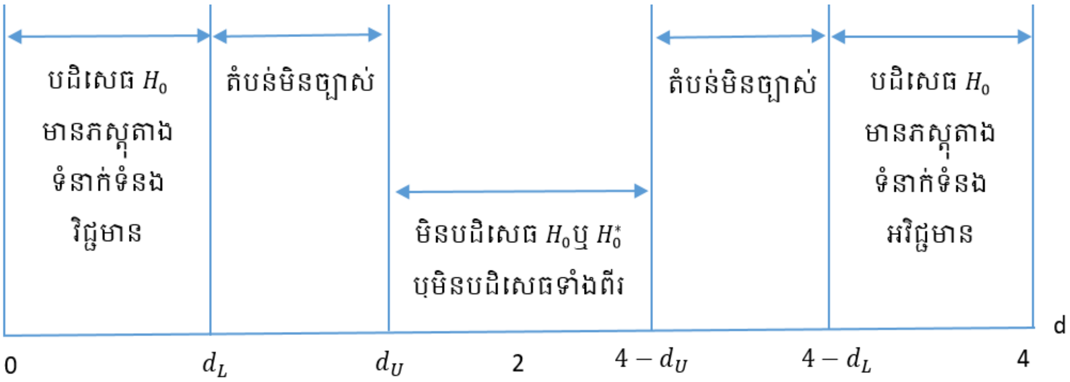
$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2} \qquad (10.19)$$

ហើយ $\hat{u}_t$ = Residuals នៅគ្រា t ($\hat{u}_t = Y_i - \hat{Y}$)

$\hat{u}_{t-1}$= Residuals នៅគ្រា t-1 ឬក្រោយមួយគ្រា

បន្ទាប់ពីបានគណនាតម្លៃ ដឺប៊ីនវ៉តសុន (d) រួចមក អ្នកយកវាទៅធៀបជាមួយដឺប៊ីនវ៉តសុនក្នុងតារាង 10.7។ បើសិនតម្លៃ d នៅចន្លោះជួរឈរទី1 និងទី5 គឺមិនបដិសេធH_0 ហើយសន្និដ្ឋានថា មានស្តុតាងបញ្ជាក់ថា ស៊េរីតយួលមានទំនាក់ទំនងវិជ្ជមាន (positive serial correlation) ហើយបើ d នៅចន្លោះជួរឈរទី2 និងទី4 មិនច្បាស់លាស់ ហើយ d នៅចន្លោះជួរឈរទី3 គឺមិនបដិសេធH_0 ឬ/និង H_0^* រួចសន្និដ្ឋានថា មិនមានវត្តមានទំនាក់ទំនងវិជ្ជមានឬអវិជ្ជមានក្នុងស៊េរីតយួលនោះទេ (no positive serial correlation or negative serial correlation)។

តារាង 10.7 ដឺប៊ីនវ៉តសុន(Durbin –Watson d statistic)



H_0 = មិនមានទំនាក់ទំនងវិជ្ជមានដោយស្វ័យប្រវត្តិ (no positive autocorrelation)

H_0^* = មិនមានទំនាក់ទំនងអវិជ្ជមានដោយស្វ័យប្រវត្តិ(no negative autocorrelation)

d_L = តម្លៃក្នុងតារាងដឺប៊ីនវតសុនកំរិតទាប (Durbin-Watson Statistic lower bound)

d_U = តម្លៃក្នុងតារាងដឺប៊ីនវតសុនកំរិតខ្ពស់ (Durbin-Watson Statistic upper bound)

ឥឡូវយើងវិលមកឧទាហរណ៍ 10.1 របស់យើងបន្តទៀត ដើម្បីចង់ដឹងថា រ៉េស៊ីតយួលរបស់ សមីការលីនេអ៊ែររបស់យើងឯករាជ្យឬយ៉ាងណា។ តារាង 10.8 បង្ហាញពីទិន្នន័យដើម្បីគណនា ដឺប៊ីនវតសុនតេស្ត។

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t-\hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2} = \frac{50057.4}{17868} = 2.80$$

បន្ទាប់មកយើងតម្លៃ Durbin d=2.80 នេះទៅផ្ទៀងផ្ទាត់នៅក្នុងតារាង ដឺប៊ីនវតសុន (តារាង 10.9) ត្រង់ K=1, $d_L = 1.10$ និង d_U =1.37, នោះយើងបានតារាង 10.8 រ៉េស៊ីតយួលនៃសមីការប៉ាន់ស្មាន (10.12)

n	$\hat{u}_t$	$\hat{u}_{t-1}$	${\hat{u}_t}^2$	$(\hat{u}_t - \hat{u}_{t-1})^2$
1	-1.646	NA	2.7089	NA
2	54.039	-1.646	2920.2	3100.82
3	-1.591	54.039	2.5311	3094.7
4	16.094	-1.591	259.02	312.76
5	-13.7	16.094	187.71	887.73
6	-44.62	-13.7	1990.8	955.89
7	50.861	-44.62	2586.9	9116.39
8	-41.45	50.861	1718.4	8522.06
9	72.35	-41.45	5235.1	12952.2
10	-14.28	72.354	203.8	7504.77
11	1.41	-14.28	1.98	246.02
12	4.78	1.409	22.84	11.35
13	-25.02	4.78	625.79	887.73
14	-44.01	-25.02	1937.1	360.89
15	-0.06	-44.01	0.003	1932.19
16	-13.17	-0.06	173.35	171.87
			$\sum \hat{u}_t^2 = 17868$	$\sum(\hat{u}_t - \hat{u}_{t-1})^2 = 50057.4$

- 4-d_U =4-1.37 =2.63, និង 4-d_L =4-1.10=2.90 ដូច្នេះ d=2.80 គឺនៅចន្លោះ
- 4-d_U=2.63 < d= 2.80 < 4-d_L=2.90 ដែលជាតំបន់មិនច្បាស់លាស់។ ដូច្នេះត្រូវប្រមូលទិន្នន័យបន្ថែមសម្រាប់ការធ្វើតេស្តនេះ។

តារាង 10.9 រកតម្លៃគោលរបស់ដឺប៊ីនវតសុន(Durbin-Watson Statistic)

$\alpha = 0.05$										
	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
→16	1.10	1.37	.98	1.54	.86	1.73	.74	1.93	.62	2.15
17	1.13	1.38	1.02	1.54	.90	1.71	.78	1.90	.67	2.10
18	1.16	1.39	1.05	1.53	.93	1.69	.82	1.87	.71	2.06

រូប10.10 លទ្ធផលសមីការលីនេអ៊ែរ

Dependent Variable: Y
Method: Least Squares
Date: 04/25/21 Time: 14:20
Sample: 2020M01 2021M04
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	88.23553	35.80115	2.464600	0.0273
X	1.131502	0.196496	5.758410	0.0000

R-squared	0.703134	Mean dependent var	287.8750
Adjusted R-squared	0.681929	S.D. dependent var	63.34548
S.E. of regression	35.72545	Akaike info criterion	10.10607
Sum squared resid	17868.30	Schwarz criterion	10.20265
Log likelihood	-78.84858	Hannan-Quinn criter.	10.11102
F-statistic	33.15929	Durbin-Watson stat	2.801463
Prob(F-statistic)	0.000050		

កម្មវិធី Eviews បានបញ្ចូលតម្លៃ ដឺប៊ីនវតសុនតេស្តនៅក្នុងលទ្ធផលសមីការលីនេអ៊ែរតែម្តង។ តារាង10.10 បង្ហាញអំពីលទ្ធផលសមីការលីនេអ៊ែរផលបូករ៉េស៊ីតយួល គិតជាការ៉េ និងតម្លៃ ដឺប៊ីនវតសុនតេស្តគឺដូចលទ្ធផលខាងលើដែលបានគណនារួច (សូមធ្វើការផ្ទៀងផ្ទាត់ដោយខ្លួនឯង)។

10.5.3 ការវាយតម្លៃភាពថេរនៃវ៉ារីយ៉ង់និងដំណោះស្រាយ

I. ការវាយតម្លៃភាពថេរនៃវ៉ារីយ៉ង់

ភាពដូចគ្នានៃវ៉ារីយ៉ង់ (homogeneity of variance) ឬវ៉ារីយ៉ង់ថេរ គឺជាល័ក្ខខ័ណ្ឌដ៏សំខាន់សម្រាប់ការប្រៀបធៀបវ៉ារីយ៉ង់ ឬហៅថា analysis of variance (ANOVA) បើសិនវ៉ារីយ៉ង់មិនថេរ ឬល័ក្ខខ័ណ្ឌនេះត្រូវបានបំពាន ប្រ៉ូបាប៊ីលីតេដែលបដិសេធសម្មតិកម្មមោឃៈ(H_0: វ៉ារីយ៉ង់ស្មើគ្នា)ទាំងមានកំហុសគឺមានច្រើន។ នៅក្នុងសមីការលីនេអ៊ែរវិញ ល័ក្ខខ័ណ្ឌប្រភេទនេះត្រូវបានដាក់កំហិតទៅលើអ៊ែរើ ឬ រ៉េស៊ីតយួល (u_i) ដែលតម្រូវឲ្យ មានវ៉ារីយ៉ង់ដូចគ្នា (homogeneity) ហើយបើល័ក្ខខ័ណ្ឌនេះត្រូវបានបំពាន គេចាត់ទុកថាមានបញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី (heteroscedasticity) នៅក្នុងរ៉េស៊ីតយួល។ ប្រសិនមាន ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធីកើតឡើង OLS នៅតែមិនលំអៀង (unbiased) និងនៅតែស៊ីសង្វាក់ឬលូនដដែល

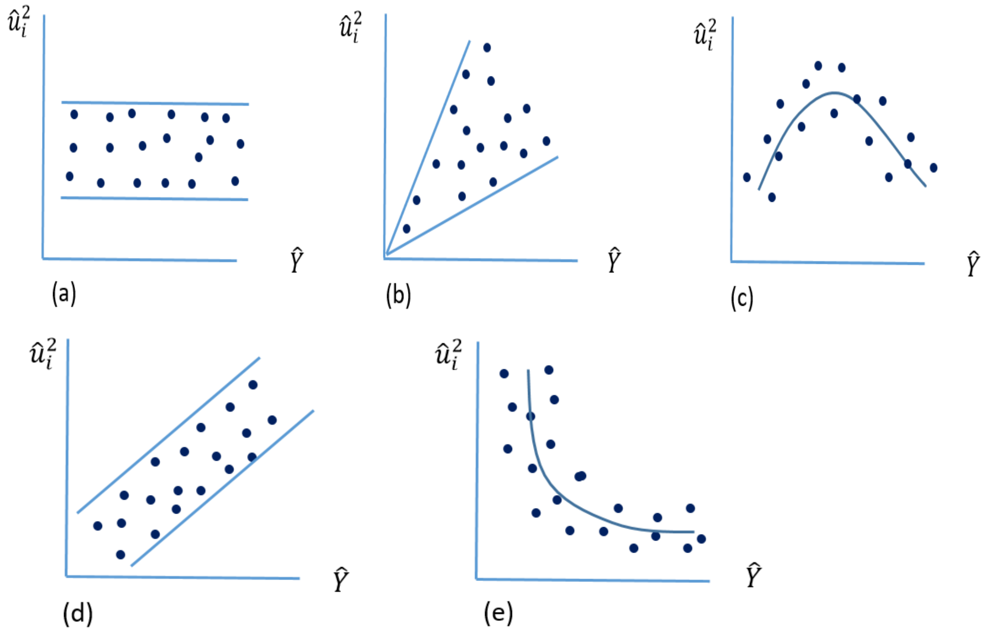
(consistency) ប៉ុន្តែវាមិនបានធ្វើឲ្យវ៉ារីយ៉ង់តូចបំផុតនោះឡើយ (estimators are no longer minimum variance) ដូច្នេះ OLS លែងជាបន្ទាត់ប្រសើរបំផុតទៀតហើយ(not BLUE: not best linear unbiased estimator) ទៀតហើយ។ ដោយសារតែ ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី គឺទាក់ទងជាមួយវ៉ារីយ៉ង់មិនថេរ ដូច្នេះផលប៉ះពាល់មួយផ្សេងទៀតគឺ F-test និង t-test មិនអាចទុកចិត្តបានដែលជាហេតុនាំឲ្យយើងមានការយល់ច្រឡំ ឬសន្និដ្ឋានខុស[28] ។

អ្នកអាចវាយតម្លៃភាពភាពដូចគ្នានៃវ៉ារីយ៉ង់តាមរយៈវិធីសាស្ត្ររូបភាពនិងតេស្តស្ថិតិ។

1) វិធីសាស្ត្ររូបភាព (Graphical Method)

បន្ទាប់ពីធ្វើការប៉ាន់ស្មានសមីការបន្ទាត់រួចមក អ្នកទទួលបានរ៉េស៊ីតយួល បន្ទាប់មកលើកវាជាការ៉េ ($\hat{u}_i^2$) ហើយយកវាទៅគូសតាងជាមួយតម្លៃព្យាករណ៍ ($\hat{Y}_i$) ដោយដាក់ $\hat{u}_i^2$ នៅលើអ័ក្សឈរ ហើយដាក់ $\hat{Y}_i$ នៅលើអ័ក្សដេក(ដូចរូប 10.12)។

រូប 10.12 ទម្រង់ផ្សេងៗនៃរ៉េស៊ីតយួលដែលប៉ាន់ស្មាន



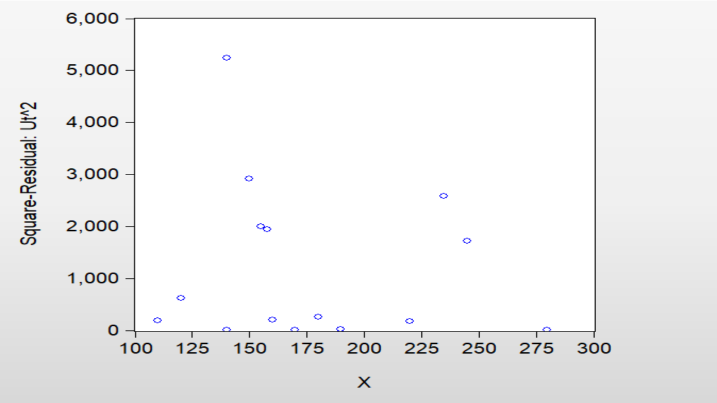
បន្ទាប់មកទៀត ធ្វើការអង្កេតលើទំនាក់ទំនងរបស់វា។បើសិនទំនាក់ទំនង $\hat{u}_i^2$ និង $\hat{Y}_i$ មានរៀងដូចរូប (a) យើងអាចសន្និដ្ឋានថាប្រហែលជាវ៉ារីយ៉ង់ដូចគ្នា។ ហើយបើទំនាក់ទំនង$\hat{u}_i^2$ និង $\hat{Y}_i$ មានរៀងដូចរូប b ដល់ e ប្រហែលជាវ៉ារីយ៉ង់មិនដូចគ្នាទេ ឬមានបញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី។

[28] បញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី បង្កឲ្យមានផលវិបាកស្រដៀងទៅនឹងបញ្ហាទំនាក់ទំនងជាស៊េរី ដែរ ក៏ប៉ុន្តែទំនាក់ទំនងជាស៊េរីភាគច្រើនកើតមានតែក្នុងទិន្នន័យជាស៊េរីតែប៉ុណ្ណោះ រីឯបញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី វិញអាចកើតមានលើទិន្នន័យទាំងពីរប្រភេទ គឺទិន្នន័យជាស៊េរីផងនិងទិន្នន័យប្រមូលក្នុងគ្រាតែមួយផង។ ហើយបញ្ហាទាំងពីរសុទ្ធតែទាក់ទងជាមួយរ៉េស៊ីតយួលទាំងអស់។ ទំនាក់ទំនងជាស៊េរីទាក់ទងជាមួយបញ្ហា រ៉េស៊ីតយួលជាប់ទាក់ទងគ្នាពីគ្រាមួយទៅគ្រាមួយទៀង រីឯ ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ទាក់ទងជាមួយរ៉េស៊ីតយួលមិនថេរ។

វិធីសាស្ត្ររូបភាពមួយផ្សេងទៀតស្រដៀងទៅនឹងវិធីនេះដែរ ដោយអ្នកគ្រាន់តែយក $\hat{u}_i^2$ ទៅតាងជាមួយអថេរឯករាជ្យ (X) ណាមួយដែលអ្នកសង្ស័យថាអថេរនោះមានទំនាក់ទំនងជាមួយ រ៉េស៊ីតយួល រួចធ្វើការអង្កេតទំនាក់ទំនងរបស់វា ហើយការសន្និដ្ឋានស្រដៀងទៅនឹងទំនាក់ទំនង $\hat{u}_i^2$ និង $\hat{Y}_i$ ដែរ។

ចំពោះឧទាហរណ៍របស់យើង ការវាយតម្លៃនៃភាពដូចគ្នា ឬថេរនៃវ៉ារីយ៉ង់គឺត្រូវបានយក$\hat{u}_i^2$មកដាក់បង្ហាញជាមួយ (X) ដូចក្នុងរូប 10.13។ តាមរយៈរូបនេះ យើងអាចអង្កេតឃើញថា ចំណុចទាំងអស់ក្នុងរូបមិនមានរាងទម្រង់ណាមួយដូចក្នុងរូប 10.2 ចាប់ពី b ដល់ c ឡើយ។ ប៉ុន្តែវាមានប្រហាក់ប្រហែលទៅនិងរូប a ទៅវិញ។ ដូច្នេះយើងអាចសន្និដ្ឋានថា រ៉េស៊ីតយួល ក្នុងសមីការប៉ាន់ស្មាន 10. 12 ប្រហែលជាថេរ ឬដូចគ្នា។

រូប 10.13 ទំនាក់ទំនងរ៉េស៊ីតយួលលើកជាការ៉េជាមួយ X



2) ការប្រើតេស្តស្ថិតិជាផ្លូវការ

មានតេស្តជាច្រើនត្រូវបានបង្កើតឡើងដើម្បីតេស្តលើហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ហើយក្នុងចំណោមតេស្ត ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធីខាងក្រោម មានតែតេស្តរបស់លោក Park មួយទេដែលប្រើទំនាក់ទំនងរវាង (u_i) ជាមួយ Y ចំណែកតេស្តដ៏ទៃទៀតគឺប្រើ ទំនាក់ទំនង រ៉េស៊ីតយួល (u_i) ជាមួយ X ។

i. Park Test

បន្ទាប់ពីប៉ាន់ស្មានសមីការ 10.12 រួចមក អ្នកត្រូវប៉ាន់ស្មានសមីការលើនៃអ៊ែរជំនួយ (auxiliary regression) មួយដោយយក $\hat{u}_i^2$តំណាងឲ្យ Y ហើយ អថេរឯករាជ្យតំណាងដោយដោយតម្លៃប៉ាន់ដ ($\hat{Y}_i$)។ ដូច្នេះយើងបានសមីការ Part Test: $\hat{u}_i^2 = \alpha_0 + \alpha_1 \hat{Y}_i$ បើα_1 មានសារៈសំខាន់ (significance) មានន័យថា $\hat{u}_i^2$ មានទំនាក់ទំនងជាមួយ $\hat{Y}_i$ ហើយយើងសន្និដ្ឋានថា វ៉ារីយ៉ង់របស់សមីការប៉ាន់ស្មានគឺមិនដូចគ្នានោះទេ ឬសមីការ 10.12 មានបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី។

ii. Harvey Test

Harvey Test គឺយកលោកនេពៃនៃ រ៉េស៊ីតយួលលើកជាការ៉េជំនួស Y ហើយអថេរឯករាជ្យគឺនៅដដែល: ln($\hat{u}_i^2$)= $\alpha_0 + \alpha_1 X_i$ បើ α_1 មានសារៈសំខាន់ (significance) នោះមានន័យថា $\hat{u}_i^2$ មាន

ទំនាក់ទំនងជាមួយ X ហើយយើងសន្និដ្ឋានថា វ៉ារីយ៉ង់របស់សមីការប៉ាន់ស្មានគឺមិនដូចគ្នានោះទេ ឬមានន័យថាសមីការប៉ាន់ស្មានបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី។

iii. Glejser Test

ស្រដៀងទៅនិងនីតិបិធីរបស់ Harvey ដែរ គឺ Y ជំនួសដោយតម្លៃដាច់ខាតនៃ រ៉េស៊ីតយួល ហើយហើយអថេរឯករាជ្យគឺនៅដដែល: $|\hat{u}_i|$ $\alpha_0+\alpha_1 X_i$ បើ α_1 ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី នោះមានន័យថា រ៉េស៊ីតយួលមានទំនាក់ទំនងជាមួយ X ហើយយើងសន្និដ្ឋានថា វ៉ារីយ៉ង់របស់សមីការប៉ាន់ស្មានគឺបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី។

iv. Breusch–Pagan–Godfrey តេស្តរកបញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ដោយយក $\hat{u}_i^2$ តំណាងឲ Y ដូចលោក Park ដែរ ក៏ប៉ុន្តែអថេរឯករាជ្យ មិនមែនតំណាងដោយ $\hat{Y}_i$ នោះទេ តែរក្សាអថេរ X ដូចដើម។ រីឯ White's heteroscedasticity ស្រដៀងទៅនឹង Breusch–Pagan–Godfrey ដែរ ដោយយក $\hat{u}_i^2$ តំណាងឲ Y ហើយរក្សាអថេរ X នៅដដែល ដោយថែម X^2 មួយទៀត ក្នុងករណីសមីការលីនេអ៊ែរដែលមានXតែមួយ (simple regression) ហើយបើករណីសមីការលីនេអ៊ែរដែលមានXច្រើន (multiple regression) គឺបន្ថែមអន្តរកម្ម (interaction term) ដូចជា $X_1 * X_2$ ថែមទៀត។ តេស្តទាំងពីរនេះចង់កំណត់ទំនងរវាងរ៉េស៊ីតយួលជាមួយ X ប៉ុន្តែវាមានលក្ខណៈស្មុគស្មាញជាងតេស្តបីខាងដើម ហើយប្រសិនអ្នកអានមានចំណាប់អារម្មណ៍សូមអានសៀវភៅ Basic Econometrics (2008) របស់លោក Guajarati។ សូមបញ្ជាក់ថាតេស្តទាំងនេះភាគច្រើនត្រូវបានបង្កើតឡើងនៅក្នុងកម្មវិធី Eviews។

10.6 ដំណោះស្រាយពេលវ៉ារីយ៉ង់មិនថេរ

ជាទូទៅដំណោះស្រាយចំពោះវ៉ារីយ៉ងមិនថេរត្រូវបានចែកជាពីរៈ ករណីវ៉ារីយ៉ង់បានដឹង និង មិនបានដឹង។

1. ករណីវ៉ារីយ៉ង់បានដឹង (when σ_i^2 is known)

ករណីនេះយើងអាចប្រើវិធីសាស្ត្រសមាមាត្រគិតជាការ៉េដែលធ្វើឲ្យអ៊ែរើតូចបំផុត (The Method of Weighted Least Square, WLS)។ ឧបមាថា យើងមានសមីការដូចខាងក្រោម៖

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (10.16)$$

ឥឡូវសន្មត់ថា វ៉ារីយ៉ង់របស់ ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ត្រូវបានដឹង σ_i^2 បន្ទាប់មកយើងចែកសមីការ 10.16 ជាមួយ σ_i (អាចធ្វើនៅក្នុង Excel បាន)

$$\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i} \qquad (10.17)$$

សូមបញ្ជាក់ថា សមីការ (10.17) គឺជាសមីការកាត់តាមគល់ ០។ សមីការនេះត្រូវបានស្គាល់ថា WLS estimator ឬ the Generalized Least Square (GLS)។ ក្នុងបរិបហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ពាក្យទាំងពីរអាចប្រើជំនួសគ្នាបាន (Guajarati & Porter, 2008)។ ពេលប៉ាន់ស្មានសមីការ (10.17) រួចរាល់ យើងគុណសមីការនេះនឹងσ_i ពេលនោះយើងនិងទទួលបានមេគុណ β_0និង β_1ជាស្រេច។

ឧទាហរណ៍ 10.2 អ្នកសិក្សាម្នាក់ចង់ដឹងអំពីទំនាក់ទំនងរវាងចំណាយលើការស្រាវជ្រាវនិងប្រាក់ចំណេញរបស់ឧស្សាហកម្ម។ ទិន្នន័យទាំងអស់មានដាក់បង្ហាញក្នុងតារាង 10.11។

តារាង 10.11 ប្រាក់ចំណេញនិងចំណាយលើការស្រាវជ្រាវ

n	Y	X	σ_i	Y_i/σ_i	X_i/σ
1	190	6400	1200	0.16	5.33
2	1570	11700	230	6.83	50.86
3	280	14700	1750	0.16	8.40
4	2830	22000	200	14.15	110
5	230	26500	2760	0.08	9.60
6	3750	32400	280	13.39	115.71
7	2890	35000	800	3.61	43.75
8	4650	40500	530	8.77	76.41
9	5040	70800	1500	3.36	47.20
10	13900	80600	6500	2.138462	12.40

ហើយ Y= ប្រាក់ចំណេញជាមធ្យមគិតជា $1000 របស់ឧស្សាហកម្មទាំង10

X=ចំណាយលើការស្រាវជ្រាវគិតជាមធ្យមរបស់ឧស្សាហកម្មទាំង10

σ_i= គម្លាតស្តង់ដារនៃប្រាក់ចំណេញរបស់ផឧស្សាហកម្មទាំង10

- ជំហានទី1: បង្កើតសមីការប៉ាន់ស្មាន

ដំបូងនេះយើងធ្វើការប៉ាន់ស្មានសមីការដែលតំណាងឲ្យទំនាក់ទំនងរវាងប្រាក់ចំណេញនិងការចំណាយលើការអភិវឌ្ឍន៍ជាមុនសិន បន្ទាប់មកទៀតយើងធ្វើតេស្តលើសមីការនោះដើម្បីកំណត់ថា តើវាមានបញ្ហាហេតេរ៉ូស្កេដាស្ទីកស៊ីធី ដែរឬអត់។

លទ្ធផលនៃការប៉ាន់ស្មានទំនាក់ទំនងប្រាក់ចំណេញនិងចំណាយលើការស្រាវជ្រាវបានបង្ហាញដោយសមីការ 10.17។

$\hat{Y}_i$= -1,363.94 +0.1438X_i (10.17)

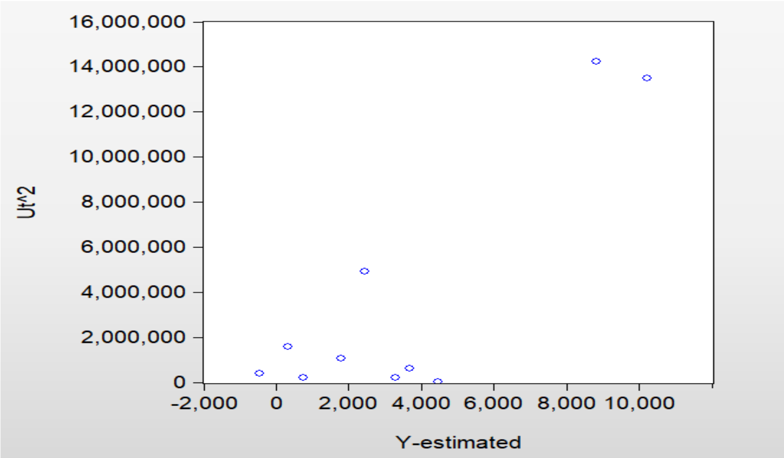
se= (1,202.78) (0.0292)

t= (-1.1340) (2.9293)

- ជំហានទី 2: បង្កើតគំនូសតាងរ៉េស៊ីតយួល (plot residuals)

បន្ទាប់មកយើងលើ រ៉េស៊ីតយួល ជាការ៉េ ($\hat{u}_i^2$) រួច ដាក់បង្ហាញនៅលើក្រាហ្វិកជាមួយតម្លៃប៉ាន់ស្មាន $\hat{Y}_i$ (Y-estimated) យើងបាន រូប10.14។

រូប 10.4 ទំនាក់ទំនង $\hat{u}_i^2$ និង តម្លៃប៉ាន់ស្មាន$\hat{Y}_i$



តាមរយៈរូបនេះយើងសង្ស័យថាមានបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ក្នុង OLS របស់យើង។

- ជំហានទី3: សន្មត់ថាយើងជ្រើសរើស Glejser test ដើម្បីរកបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី តាមរយៈនៃការប៉ាន់ស្មានសមីការ: $|\hat{u}_i|= \alpha_0+\alpha_1 X_i$ នោះយើងបានលទ្ធផលដូចខាងក្រោម៖

$|\hat{u}_i|=$ 15.768 + 0.042 X_i (10.18)

$se =$ (496.37) (0.0120)

$t =$ (0.0317) (3.4929)

តាមរយៈលទ្ធផលនេះបង្ហាញឲ្យឃើញថា មេគុណរបស់ X គឺ មានសារៈសំខាន់ (significance) ខ្លាំងតែម្តង ដូច្នេះយើងសន្និដ្ឋានថា មានបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី នៅក្នុង OLS របស់យើង

- ជំហានទី4: ប្រើ WLS ដើម្បីដោះស្រាយ ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី តាមរយៈការប៉ាន់ស្មានសមីការ 10.17: $\frac{Y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{X_i}{\sigma_i}$ ។ នោះយើងបាន៖

$\frac{Y_i}{\sigma_i} = 566.20(\frac{1}{\sigma_i}) + 0.0953(\frac{X_i}{\sigma_i})$ (10.19)

$se =$ (323.94) (0.01316)

$t =$ (1.7479) (7.2421)

រួចយើងគុណសមីការ 10.19 និង σ_i រួចសំរួល σ_i ចោល នោះយើងទទួលបានសមីការ 10.20

$Y_i=$ 566.20 + 0.0953 X_i (10.20)

ជំហានទី 5: យើងប្រើ Glejser test ដើម្បីរកបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ឡើងវិញ នោះយើងបាន លទ្ធផលសមីការ 10.20៖

$$|\hat{u}_i| = 0.6702 + 5.44\text{E-06}\ X_i \qquad (10.21)$$

$$se = (0.2064)\quad (5.01\text{E-06})$$

$$t = (3.2465)\quad (1.0864)$$

នៅកំរិត α =0.05 df=10-2=8 តេស្តកន្ទុយសងខាង(two-tailed test) យើងបាបតម្លៃធីគោល t-critical =2.3060។ ដោយ t-statistic = 1.0864 < t-critical =2.3060 ដូច្នេះមេគុណរបស់ X គឺមិនសំខាន់ (insignificance)។ យើងសន្និដ្ឋានថា សមីការប៉ាន់ស្មាន 10.20 លែងមានបញ្ហាហេតេរ៉ូស្កេតដាស្ទីកស៊ីធីទៀតហើយ។ ដូច្នេះតេសកតទាំងឡាយដូចជា F-test និង t-test នៅក្នុងសមីការនេះអាចទុកចិត្តបាន។

2. ករណីវ៉ារីយ៉ង់មិនដឹង (unknown σ_i^2)

មានករណីជាច្រើន ការសិក្សាមិនទាក់ទងជាមួយតម្លៃមធ្យមរបស់ Y ឬ X នោះទេ មានន័យយើងមិនអាច ដឹង σ_i^2 ហើយបើសមីការប៉ាន់ស្មានរបស់អ្នកជួបបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធី ថែមទៀត អ្នកត្រូវប្រើនីតិវិធីស្រដៀងទៅនិង WLS ឬ GLS ប៉ុន្តែមិនមែនចែកសមីការនឹង σ_i នោះទេ តែត្រូវចែក $\sqrt{X_i}, X_i$, ឬ X_i^2 ពេលចែករួចរាល់ហើយ ទើបយើងធ្វើការប៉ាន់ស្មានសមីការលីនេអ៊ែរ។ បើចែក$\sqrt{X_i}$ ពេលពីប៉ាន់ស្មានរួចអ្នកត្រូវគុណនឹង $\sqrt{X_i}$ មកវិញ បើចែកជាមួយអថេរណា ពេលប៉ាន់ស្មានរួចត្រូវគុណជាមួយអថេរនោះ ដើម្បីឲ្យបានសមីការដើមវិញ ។ ពេលខ្លះអ្នកក៏ត្រូវការចែកនឹង $\hat{Y}_i$ ហើយគុណនិង $\hat{Y}_i$ វិញក៏មាន (សូមអានសៀវភៅ Basic Econometrics របស់លោក Guajarati, 2008)។ ជំរើសមួយទៀតពេលដែលអ្នកជួបបញ្ហា ហេតេរ៉ូស្កេតដាស្ទីកស៊ីធីក្នុងការប៉ាន់ស្មាន OLS អ្នកអាចប្រើ សមីការរបស់លោក White ដែលជាទូទៅគេហៅថា "White heteroskedasticity-consistent standard errors & covariance"។ លទ្ធផលនៃការប្រើសមីការនេះ គឺអ្នកនឹងទទួលបានមេគុណដូច OLS ធម្មតា តែខុសគ្នាត្រង់ អ៊ែររើស្តង់ដារ (standard error) និង t-test ។ សមីការនេះមាននៅក្នុងកម្មវិធី Eviews ដោយអ្នកគ្រាន់តែជ្រើសរើសមុខងារនេះមកប្រើជាការស្រេច។

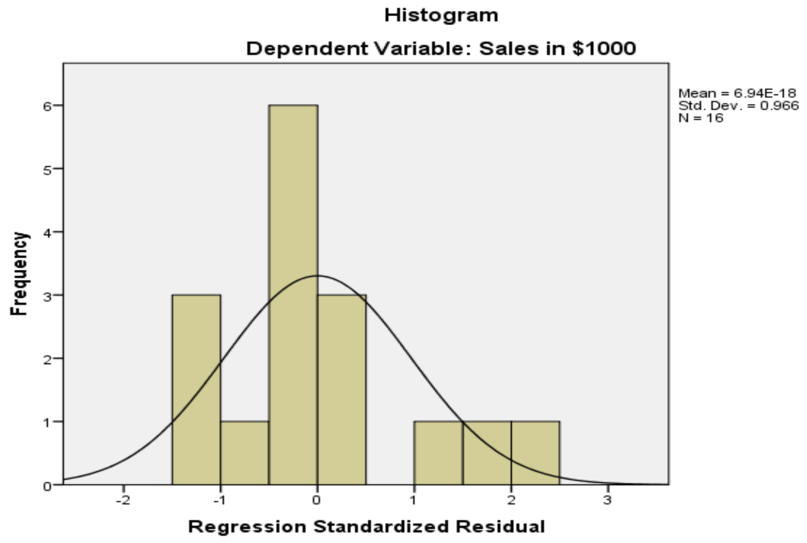
10.5.4 ការវាយតម្លៃន័រម៉ាល់លីតធី

ន័រម៉ាល់លីតធី (Normality) តេស្តគឺមានគោលដៅកំណត់ថាតើ រ៉េស៊ីតយួល ដែលបានមកពីសមីការប៉ាន់ស្មាននោះមានលក្ខណៈន័រម៉ាល់លីតធី ឬអត់។ មានតេស្តមួយចំនួនដែលតម្រូវឲ្យមានលក្ខណៈ ន័រម៉ាល់លីតធី ដូចជា t-test និង F-test ជាដើម។ ដោយសារតែនៅក្នុងសមីការលីនេអ៊ែរគឺមានប្រើប្រាស់ t-test និង F-test (យើងនឹងធ្វើការរៀបរាប់នៅពេលបន្តិចទៀតនេះ) ហេតុដូច្នេះហើយបានជាការវាយតម្លៃលើ ន័រម៉ាល់លីតធី ទៅលើរ៉េស៊ីតយួល គឺជាល័ក្ខខ័ណ្ឌចំបាច់។

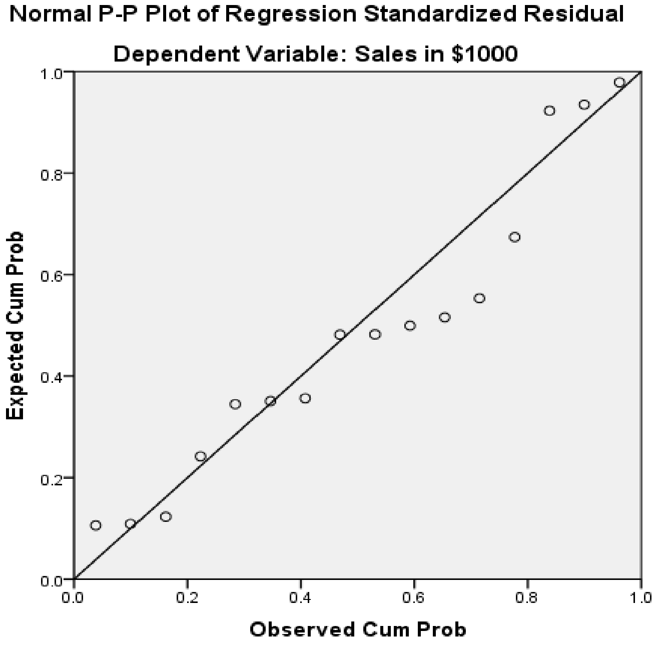
ដើម្បីវាយតម្លៃទៅលើល័ក្ខខ័ណ្ឌ ន័រម៉ាល់លីតធី អ្នកមានជំរើសច្រើន ដូចជាការយករ៉េស៊ីតយួលមកដាក់បង្ហាញជាគំនូសហ៊ីស្តូក្រាមនិងគំនូសដើមនិងស្លឹក (Histogram, Stem and Leave) ឬ ដាក់ជាគំនូសន័រម៉ាល់លីតធី (normality plot) ជាដើម។

សម្រាប់ឧទាហរណ៍ 10.1 របស់យើងរ៉េស៊ីតយួលត្រូវបានដាក់បង្ហាញនៅលើ គំនូសហ៊ីស្តូក្រាម និង គំនូសន័រម៉ាល់លីតធី ក្នុងរូប 10.8 និង 10.9។

រូប 10.8 គំនូសហ៊ីស្តូក្រាម (Histogram)



រូប10.9 គំនូសន័រម៉ាល់លីតធី



រូប10.8 គំនូសហ៊ីស្តូក្រាម មានរាងមិនសូវច្បាស់ថាតើដូចរូបជួង ឬមិនដូច ហើយមានចំណុចមួយចំនួននៃ រ៉េស៊ីតយូលក្នុងរូប 10.9 ស្ថិតនៅជិតបន្ទាត់ ហើយខ្លះក៏នៅឆ្ងាយពីបន្ទាត់ ដែលធ្វើឲ្យពិបាកកំណត់ថា រ៉េស៊ីតយូល របស់សមីការប៉ាន់ស្មាន 10.12 មានលក្ខណៈ ន័រម៉ាល់លីតធី ដែរឬអត់។ ក្នុងករណីបែបនេះមានតែតេស្តស្ថិតិទេដែលអាចអារកាត់បាន។ ចាកបេរ៉ាតេស្ត (Jarque-Bera, JB) គឺត្រូវបានបរិយាយនៅជំពូកមុនម្តងហើយ វាគឺជាតេស្តមួយដែលមានប្រជាប្រិយភាពជាងគេក្នុងកំណត់អំពី ន័រម៉ាល់លីតធី របស់រ៉េស៊ីតយូល។ JB តេស្តគឺងាយយល់និងងាយអនុវត្ត ប៉ុន្តែតេស្តនេះអាចប្រើប្រាស់បានសមស្របក្នុង

ករណីដែលចំនួនប៉ាន់គំរូមានចំនួនច្រើនតែប៉ុណ្ណោះ ហើយក្នុងករណីដែលប៉ាន់គំរូមានចំនួនតិច តេស្តនេះមិនស័ក្តសមឡើយ[29]។

JB តេស្ត គឺតេស្តលើសម្មតិកម្ម H_0: រ៉េស៊ីតយួល គឺមានលក្ខណៈ ន័រម៉ាល់លីតធី ហើយ JB តេស្តត្រូវបានកំណត់ដូចក្នុងសមីការ 10.5។

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \sim \chi_2^2 \qquad (10.5)$$

ហើយ n= ចំនួនប៉ាន់គំរូ

S= ភាពធៀង (Skewness)

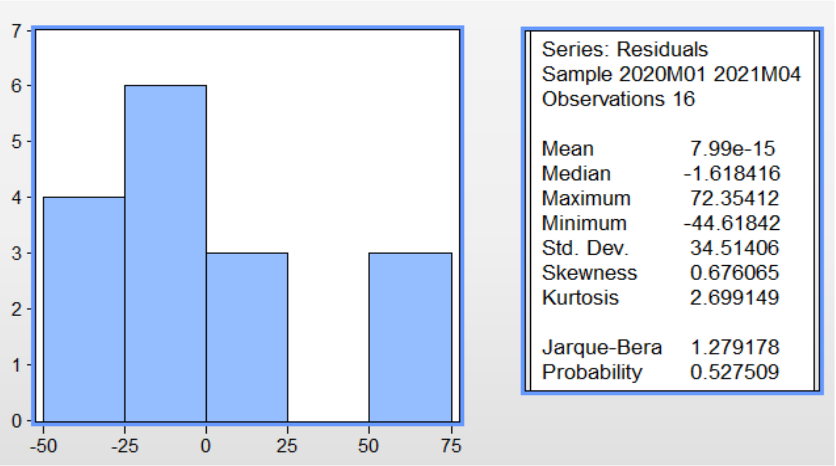
K= កំពស់របស់ហ៊ីស្តូក្រាម (Kurtosis)

χ_2^2 = JB មានបំណែងចែកដូចឆាយស្ក្វែដែលមាន df=2

ពេលដែលភីវ៉ែលយូ (p-value) របស់ឆាយស្ក្វែធំជាង 0.05 រ៉េស៊ីតយួល គឺមានលក្ខណៈន័រម៉ាល់លីតធី ហើយផ្ទុយមកវិញពេល p<0.05 រ៉េស៊ីតយួលគឺមានលក្ខណៈមិន ន័រម៉ាល់លីតធី ។

ឥឡូវយើងសន្មត់ថានៅក្នុងឧទាហរណ៍របស់យើងមានចំនួនប៉ាន់គំរូធំគ្រប់គ្រាន់ ដូច្នេះយើងអាចប្រើ JB តេស្តបាន។

រូប10.11 លទ្ធផល JB តេស្ត



តាមរយៈស្ថិតិបែបពិយាយរបស់ រ៉េស៊ីតយួលយើងបាន S=0.3381, K=1.7855, និង n=10

$$JB = 16\left[\frac{0.676^2}{6} + \frac{(2.699-3)^2}{24}\right] = 1.279 \qquad (p=0.5275)$$

---

[29] ដកស្រង់ចេញពី Gujarati. D (2012) Econometrics by Example, First Edition, Palgrave Macmillan, UK (p.128)

យើងបាន JB=1.279 ឬ χ_2^2=1.279 ហើយមាន p-value =0.5275 ដូច្នេះ H_0: រ៉េស៊ីតយួលមានលក្ខណៈ ន័រម៉ាល់លីតធី គឺត្រូវបានខកខានមិនបានបដិសេធ ហើយយើងអាចសន្និដ្ឋានថារ៉េស៊ីតយួល គឺមានលក្ខណៈន័រម៉ាល់លីតធី។ លទ្ធផលនៃការតេស្ត ន័រម៉ាល់លីតធី ដែលផ្តល់ដោយ Eviews មានបង្ហាញក្នុងរូប10.11 ។

10.7 **ការព្យាករណ៍**

បន្ទាប់ពីបានសមីការប៉ាន់ស្មានរួច អ្នកអាចយកសមីការនេះទៅធ្វើការព្យាករណ៍បានលុះត្រាណាតែមិនមានល័ក្ខខ័ណ្ឌលីនេអ៊ែរណាមួយត្រូវបានបំពាន។ ជាទូទៅការព្យាករណ៍មានពីរបៀបៈ ការព្យាករណ៍ជាចំណុច (point estimate) និងការព្យាករណ៍ជាចន្លោះជឿជាក់របស់ Y (interval estimate for Y)។ ហើយការព្យាករណ៍ជាចន្លោះជឿជាក់សម្រាប់ Y ត្រូវបានបែងចែកជាពីរទៅទៀតគឺ ការព្យាករណ៍ចន្លោះជឿជាក់តម្លៃមធ្យម Y (confidence interval estimate for mean of Y) និងការព្យាករណ៍ចន្លោះជឿជាក់តម្លៃបុគ្គលរបស់ Y (confidence interval estimated for individual value of Y)។ សម្រាប់ពេលនេះ យើងធ្វើការពិភាក្សាតែការព្យាករណ៍ជាចំណុច ហើយចំពោះការព្យាករណ៍ជាចន្លោះជឿជាក់នៃតម្លៃមធ្យមរបស់ Y តែប៉ុណ្ណោះ ពីព្រោះថា ការព្យាករណ៍ជាចន្លោះជឿជាក់តម្លៃមធ្យម Y មានប្រជាប្រិយភាពជាងការព្យាករណ៍ចន្លោះជឿជាក់តម្លៃបុគ្គលរបស់ Y៕

1. ការព្យាករណ៍ជាតម្លៃមធ្យម (point estimate)

ឥឡូវយើងត្រឡប់មកមើលឧទាហរណ៍របស់យើងឡើងវិញ ឧបមាថា អ្នកលក់ចង់ដឹងថា បើសិនជាគាត់ចំណាយលើការផ្សាយពាណិជកម្មចំនួន $200 តើការលក់ជាមធ្យមគឺប៉ុន្មាន?

ចំឡើយៈ

យើងមានសមីការប៉ាន់ស្មាន $\hat{Y}_i = 88.23 + 1.13\,X_i$ បន្ទាប់មកយើងជំនួស X=$200 ចូលក្នុងសមីការនេះយើងបានៈ $\hat{Y}_i = 88.23 + 1.13(200) = \$314.23$ (000) ឬ $314,230 គិតជាមធ្យមកក្នុងមួយខែ។

ជាទូទៅការប្រើសមីការប៉ាន់ស្មានដើម្បីធ្វើការព្យាករណ៍ត្រូវគោរពតាមគោលការណ៍មួយគឺ អ្នកអាចព្យាករណ៍ក្នុងវង្វង់ឬដែនកំណត់របស់ប៉ាន់គំរូតែប៉ុណ្ណោះ (only interpolation) ហើយអ្នកមិនអាចធ្វើការព្យាករណ៍ក្រៅវង្វង់ឬដែនកំណត់នោះបានឡើយ (extrapolation)។ ចំពោះឧទាហរណ៍របស់យើង តម្លៃប៉ាន់គំរូ X គឺនៅចន្លោះពី $110 ទៅ $280 តែប៉ុណ្ណោះ (មើលតារាង 10.1) ដូច្នេះអ្នកមិនអាចយកតម្លៃ X ក្រៅពីចន្លោះនេះ (តូចជាង $110 ឬ ធំជាង $280) មកព្យាករណ៍ការលក់បានឡើយ ពីព្រោះតម្លៃទាំងនោះជាករណីពិសេសដែលអ្នកមិនបានសិក្សា។

2. ការព្យាករណ៍ចន្លោះជឿជាក់

ការព្យាករណ៍ចន្លោះជឿជាក់ចំពោះតម្លៃមធ្យម (μ) របស់ Y ត្រូវបានកំណត់ដូចរូបមន្ត 10.19។

Confidence Interval for $\mu_{Y|X=X_i}$: $\hat{Y} \pm t_{\alpha/2} S_{YS}\sqrt{h_i}$ (10.19)

ហើយ $h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{SSX}$

$SSX = \sum_{i=1}^{n}(X_i - \bar{X})^2$

$\hat{Y}_i$ = តម្លៃព្យាករណ៍

S_{YX}=អែរើស្តង់ដាររបស់ Y

n =ប៉ាន់គំរូ

X_i= តម្លៃ X ដែលចង់ព្យាករណ៍ Y

$\mu_{Y|X=X_i}$= តម្លៃមធ្យមរបស់ Y ពេល X=X_i

ឥឡូវយើងបន្តការព្យាករណ៍ក្នុងឧទាហរណ៍របស់យើងពេល X_i=\$200 នោះ $\hat{Y}$ = \$314.23 និង $\bar{X}$ = 176.4375។ បើ α=0.05, df=16-2=14 នោះយើងបានតម្លៃជីគោល $t_{\frac{0.05}{2},14}$=2.1448 និង S_{YX}=35.72 (មើលតារាង 10.3)

$SSX = \sum_{i=1}^{n}(X_i - \bar{X})^2$=33,055.94 (គណនាខ្លួនឯង) បន្ទាប់មកយើងគណនា h៖

$h_i = \frac{1}{n} + \frac{(X_i-\bar{X})^2}{SSX} = \frac{1}{16} + \frac{(200-176.4375)^2}{33,055.94}$=0.0793

គណនាចន្លោះជឿជាក់សម្រាប់តម្លៃមធ្យមរបស់ Y:

$\hat{Y} \pm t_{\alpha/2} S_{YS}\sqrt{h_i}$ = \$314.23± 2.1448 (35.72) $\sqrt{0.0793}$

= \$314.23 ±21.57

292.66 ≤ $\mu_{Y|X=X_i}$ ≤ 335.80

ដូច្នេះ 95% យើងជឿជាក់ថា ពេល X=\$200 Y គិតជាមធ្យមគឺនៅចន្លោះពី 292.66 ទៅ 335.8 ឬ \$292,600 ដល់ \$335,800

10.7 **កំហុសក្នុងការវិភាគលីនេអ៊ែរ**

- ការប្រើសមីការលីនេអ៊ែរសិនមិនដឹងល័ក្ខខ័ណ្ឌឬ បើសិនល័ក្ខខ័ណ្ឌណាមួយនៃសមីការលីនេអ៊ែរត្រូវបានបំពាន ហើយអ្នកសិក្សាមិនដឹងវិធីដោះស្រាយ ឬខកខានមិនបានដំណោះស្រាយ នោះការប៉ាន់ស្មាន និងការព្យាករណ៍អាចមិនត្រឹមត្រូវនោះឡើយ។
- ការប្រើសមីការលីនេអ៊ែរមិនមានទ្រឹស្តី ឬមានចំណោះដឹងដែលទាក់ទងជាមួយកម្មវត្ថុនៃសមីការលីនេអ៊ែរ ហាក់ដូចជាវង្វេលនៅក្នុងបាទសមុទ្រអញ្ចឹង
- ការព្យាករណ៍ត្រូវធ្វើឡើងតែនៅក្នុងវង្វង់របស់ប៉ាន់គំរូ (interpolation) តែប៉ុណ្ណោះ ហើយមិនគួរធ្វើការព្យារណ៍នៅខាងក្រៅវង្វង់របស់ប៉ាន់គំរូនៃ X (extrapolation) នោះទេ ពីព្រោះវាជាករណីពិសេសដែលអ្នកមិនបានសិក្សា។

សេចក្តីសង្ខេប

នៅក្នុងជំពូកនេះអ្នកបានស្គាល់សមីការលីនេអ៊ែ និងមិនលីនេអ៊ែ ការទាញយករូបមន្តគណនាមេគុណរបស់អថេរឯករាជ្យនិងគណនាចំនួនថេររបស់សមីការបន្ទាត់។ ហើយបន្ទាប់មកអ្នកបានស្គាល់អំពីការគណនានិងការបកស្រាយ អ៊ែការ៉េ F-test, និង t-test ថែមទៀត។ ក្នុងជំពូកនេះអ្នកបានដឹងផងដែរអំពីរបៀបវាយតម្លៃលើល័ក្ខខ័ណ្ឌទាំងប្រាំងរបស់សមីការលីនេអ៊ែរនិងរបៀបដោះស្រាយ បើសិនជាមានល័ក្ខខ័ណ្ឌណាមួយត្រូវបានបំពាន ជាពីសេសវិធីសាស្រ្តដោះស្រាយក្នុងករណីដែលមានបញ្ហាទំនាក់ទំនងជាស៊េរី និង បញ្ហាហេតេ

រ៉ូស្តេតជាស្ទិចស៊ីតធីតាម WLS ឬ GLS ដើម។ ហើយចុងក្រោយអ្នកបានសិក្សាអំពីរបៀបព្យាករណ៍ជាចំណុច និងចន្លោះជឿជាក់ ព្រមទាំងល័ក្ខខ័ណ្ឌនៃការព្យាករណ៍ថែមទៀត។

សំណួរនិងលំហាត់

I. សំណួរ

1. ចូលធ្វើការបកស្រាយចំនួនថេររបស់ Y និង មេគុណ X នៃសមីការលីនេអ៊ែរដែលមានមួយអថេរ?
2. តើអ៊ែរការេ (R-square) មានន័យដូចម្តេច?
3. ពេលណាដែលបម្រែបម្រួលមិនអាចពន្យល់បានស្មើសូន្យ?
4. ពេលណាដែលបម្រែបម្រួលដែលអាចពន្យល់បានស្មើសូន្យ?
5. ហេតុអ្វីបានជាយើងគួរវិភាគទៅលើរ៉េស៊ីឌ្យូល (residuals) របស់សមីការលីនេអ៊ែរ?
6. តើល័ក្ខខ័ណ្ឌ (assumption) របស់សមីការលីនេអ៊ែរមានអ្វីខ្លះ?
7. តើពេលណាដែលអ្នកប្រើស្ថិតិ Durbin-Watson និងប្រើយ៉ាងដូចម្តេច?

II. លំហាត់

8. ភ្នាក់ងារអចលនវត្ថុមួយចង់ធ្វើការព្យាករណ៍តម្លៃជួលអចលនវត្ថុប្រចាំខែ (គិតជាដុល្លា) ទៅលើអាប៉ាតឺម៉ង់ដោយផ្អែកលើផ្ទៃក្រឡាអាប៉ាតឺម៉ង់ (ម៉ែត្រការ៉េ)។ គាត់បានប្រមូលប៉ាន់គំរូចំនួន 20 អាប៉ាតឺម៉ង់ ដែលបានកត់ត្រាដូចខាងក្រោម៖

ប៉ាន់គំរូ	ថ្លៃជួល	ម៉ែត្រការ៉េ	ប៉ាន់គំរូ	ថ្លៃជួល	ម៉ែត្រការ៉េ
1	960	840	11	1700	1258
2	1500	1450	12	1400	1175
3	1200	1000	13	1440	1235
4	1400	1230	14	1200	1255
5	970	720	15	1800	1262
6	1650	1486	16	1300	1250
7	1550	1400	17	1250	996
8	939	724	18	1500	1261
9	775	710	19	1650	1040
10	1250	956	20	1200	965

a. ចូលសង់គំនូស scatter plot.

b. ចូលប្រើវិធីសាស្ត្រលិសស្ទ័រ (least-squares) ដើម្បីកំណត់ចំនួនថេរ (b0) និងមេគុណ X(b1)

c. ចូលបកស្រាយ (b0) និងមេគុណ X (b1)

d. ព្យាករណ៍ថ្លៃជួលប្រចាំខែចំពោះអាប៉ាតឺម៉ង់មួយដែលមានផ្ទៃក្រឡា 1,000 ម៉ែត្រការ៉េ

e. ចូលគណនា អ៊ែការ៉េ និងបកស្រាយអត្ថន័យរបស់វា។

f. ចូលគណនាអ៊ែរើស្តង់ដារ

g. តើអ្នកយល់យ៉ាងដូចម្ដេចចំពោះសមីការលីនេអ៊ែរនេះពេលដែលប្រើវាដើម្បីធ្វើការព្យាករណ៍ ?

f. ចូលធ្វើការវាយតម្លៃល័ក្ខខ័ណ្ឌរបស់សមីការលីនេអ៊ែរតើមានការបំពានដែរឬទេ ?

9. អ្នកលក់តាមអ៊ីតឺណែកម្នាក់ចង់កំណត់ទំនាក់ទនងរវាងការចំណាយលើការផ្សាយពាណិជ្ជកម្ម (boost page) និងចំណូលដែលបានមកពីការលក់ប្រចាំខែ។ គាត់ធ្វើការកត់ត្រាចំណូលនិងចំណាយ (គិតជាដុល្លា) នេះរយៈពេល 24 ខែដូចខាងក្រោម៖

ខែ	ចំណាយ	ចំណូល	ខែ	ចំណាយ	ចំណូល
1	50	4020	13	63	3980
2	70	3810	14	72	4430
3	90	5310	15	59	3965
4	60	4260	16	80	4580
5	70	4300	17	95	5590
6	80	4100	18	60	3450
7	50	3215	19	91	5100
8	75	4810	20	94	5750
9	80	5240	21	69	4300
10	79	4730	22	54	3720
11	71	4415	23	90	5390
12	52	2920	24	67	4100

a. សន្មត់ថាមានទំនាក់ទំនងជាបន្ទាត់ ចូលប្រើវិធីសាស្ត្រលិសស្វ៊ែរ (LS) ដើម្បីកំណត់ចំនួនថេរ (b0) និងមេគុណ X (b1)

b. ចូលធ្វើការព្យាករណ៍ការលក់ពេលដែលចំណាយទៅលើការផ្សាយពាណិជ្ជកម្មស្មើនឹង 80 ដុល្លា។

c. ចូលបង្កើតគំនូសតាង (plot) residuals ជាមួយនឹងពេលវេលា

d. ចូលគណនាស្ថិតិ Durbin-Watson។ នៅកំរិតអាល់ហ្វា 0.05 តើមានភស្តុតាងបញ្ជាក់ថា មានបញ្ហាទំនាក់ទំនងជាស៊េរីនៅក្នុងរ៉េស៊ីតយួលដែរឬទេ ?

e. ផ្អែកតាមចំលើយ (c) និង (d) តើអ្នកសន្និដ្ឋានសុពលភាពរបស់ម៉ូដែលយ៉ាងដូចម្ដេច ?

10. ចូលប្រើទិន្នន័យក្នុងលំហាត់មុនដើម្បីឆ្លើយទៅនឹងសំនួរខាងក្រោម៖

a. ចូលគណនា តម្លៃ F-ស្ថិតិ

b. ផ្អែកលើចំលើយ (a) និង (b), តើអ្នកសំរេចចិត្តយ៉ាងដូចម្ដេច ?

c. ចូលគណនា អ៊ែការ៉េ (R-square)

d. ចូលបង្កើតចន្លោះជឿជាក់ 95% ដើម្បីប៉ាន់ស្មានតម្លៃមធ្យមរបស់ចំនួនសកលពេលដែល ចំណាយ= 70 ដុល្លា។

ឯកសារយោង

I. Book

- Allan G. B (2009) Elementary Statistics: A Step by Step Approach. Eight Edition. McGraw-Hill
- Berenson M. L, Levine D. M, & Krehbiel T. C. (2012). Basic Business Statistics: Concepts and Applications. Twelfth Edition. Prentice Hall
- Brooks C. (2014) Introductory Econometrics for Finance. Third Edition, Cambridge University Press, UK
- Brooks G. P & Barcikowsk. R. S (2012) Multiple Linear Regression Viewpoints, 2012, Vol. 38(2)
- Gujarati. D (2012) Econometrics By Example, First Edition, Palgrave Macmillan, UK
- Guajarati D. N, & Porter D. C (2008) Basic Econometrics. Fifth Edition. McGraw-Hill
- Hayashi, Fumio (2000). *Econometrics*. Princeton University Press.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Sage Publications, IncHayashi, Fumio (2001). *Econometrics*. Princeton University Press
- Pollant J. (2011) SPSS SURVIVAL MANUAL: A step by step guide to data analysis using SPSS. 4th edition. Allen & Unwin, Australia
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

II. Web Base

- Binomial distribution derived from https://www.statisticshowto.com/probability-and-statistics/binomial-theorem/binomial-distribution-formula/
- Descriptive statistics derived from https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php
- Nonlinear in the parametersderived from
- Kruskal-Wallis-test derived from https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php
- McNemar's test for paired nominal data derived from http://rasbt.github.io/mlxtend/user_guide/evaluate/mcnemar/#:~:text=The%20McNemar%20test%20statistic%20(%22chi,with%20one%20degree%20of%20freedom.

- Nonlinear in the parameters in regression model derived from https://openclassrooms.com/en/courses/5873596-design-effective-statistical-models-to-understand-your-data/
- Non-probability Sampling derived from https://www.questionpro.com/blog/judgmental-sampling/
- Probability and Non-probability Sampling derived from https://www.scribbr.com/methodology/sampling
- R-square value derived from https://statisticsbyjim.com/regression/how-high-r-squared/
- Variability or spread definition derived from https://www.investopedia.com/terms/d/descriptive_statistics.asp#:~:text=Descriptive%20statistics%20are%20brief%20descriptive,measures%20of%20variability%20(spread).
- What is statistic？The definition is adapted from https://www.investopedia.com/terms/s/statistics.asp
- Wilcoxon-Signed Test derived from https://www.statisticshowto.com/wilcoxon-signed-rank-test/
- William S. Gosset´s t-distribution derived from https://en.wikipedia.org/wiki/William_Sealy_Gosset

ឧបសម្ព័ន្ធ

Table A: A Binomial Probability Distribution

n	r	0.01	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
	0	0.82	0.36	0.12	0.04	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0.17	0.38	0.27	0.14	0.06	0.02	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0.02	0.19	0.29	0.23	0.14	0.07	0.03	0.01	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0.06	0.19	0.24	0.21	0.13	0.07	0.03	0.01	0	0	0	0	0	0	0	0	0	0	0
	4	0	0.01	0.09	0.18	0.22	0.19	0.13	0.07	0.04	0.01	0.01	0	0	0	0	0	0	0	0	0
	5	0	0	0.03	0.1	0.18	0.2	0.18	0.13	0.08	0.04	0.02	0.01	0	0	0	0	0	0	0	0
	6	0	0	0.01	0.05	0.11	0.17	0.19	0.17	0.12	0.08	0.04	0.02	0.01	0	0	0	0	0	0	0
	7	0	0	0	0.02	0.06	0.11	0.16	0.18	0.17	0.12	0.07	0.04	0.02	0.01	0	0	0	0	0	0
	8	0	0	0	0.01	0.02	0.06	0.11	0.16	0.18	0.16	0.12	0.07	0.04	0.01	0	0	0	0	0	0
	9	0	0	0	0	0.01	0.03	0.07	0.12	0.16	0.18	0.16	0.12	0.07	0.03	0.01	0	0	0	0	0
20	10		0	0	0	0	0.01	0.03	0.07	0.12	0.16	0.18	0.16	0.12	0.07	0.03	0.01	0	0	0	0
	11		0	0	0	0	0	0.01	0.03	0.07	0.12	0.16	0.18	0.16	0.12	0.07	0.03	0.01	0	0	0
	12		0	0	0	0	0	0	0.01	0.04	0.07	0.12	0.16	0.18	0.16	0.11	0.06	0.02	0.01	0	0
	13		0	0	0	0	0	0	0.01	0.02	0.04	0.07	0.12	0.17	0.18	0.16	0.11	0.06	0.02	0	0
	14		0	0	0	0	0	0	0	0.01	0.02	0.04	0.08	0.12	0.17	0.19	0.17	0.11	0.05	0.01	0
	15		0	0	0	0	0	0	0	0	0.01	0.02	0.04	0.08	0.13	0.18	0.2	0.18	0.1	0.03	0
	16		0	0	0	0	0	0	0	0	0	0.01	0.01	0.04	0.07	0.13	0.19	0.22	0.18	0.09	0.01
	17		0	0	0	0	0	0	0	0	0	0	0	0.01	0.03	0.07	0.13	0.21	0.24	0.19	0.06
	18		0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.03	0.07	0.14	0.23	0.29	0.19
	19		0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.02	0.06	0.14	0.27	0.38
	20		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.04	0.12	0.36

https://www.sjsu.edu/people/saul.cohn/courses/stats/s0/BinomialProbabTable.pdf

Table B: The Poisson Cumulative Distribution

		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.2	1.4	1.6	1.8
x=	0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679	0.3012	0.2466	0.2019	0.1653
	1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358	0.6626	0.5918	0.5249	0.4628
	2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197	0.8795	0.8335	0.7834	0.7306
	3	1	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.981	0.9662	0.9463	0.9212	0.8913
	4	1	1	1	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963	0.9923	0.9857	0.9763	0.9636
	5	1	1	1	1	1	1	0.9999	0.9998	0.9997	0.9994	0.9985	0.9968	0.994	0.9896
		2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.5	5	5.5
x=	0	0.1353	0.1108	0.0907	0.0743	0.0608	0.0498	0.0408	0.0334	0.0273	0.0224	0.0183	0.0111	0.0067	0.0041
	1	0.406	0.3546	0.3084	0.2674	0.2311	0.1991	0.1712	0.1468	0.1257	0.1074	0.0916	0.0611	0.0404	0.0266
	2	0.6767	0.6227	0.5697	0.5184	0.4695	0.4232	0.3799	0.3397	0.3027	0.2689	0.2381	0.1736	0.1247	0.0884
	3	0.8571	0.8194	0.7787	0.736	0.6919	0.6472	0.6025	0.5584	0.5152	0.4735	0.4335	0.3423	0.265	0.2017
	4	0.9473	0.9275	0.9041	0.8774	0.8477	0.8153	0.7806	0.7442	0.7064	0.6678	0.6288	0.5321	0.4405	0.3575
	5	0.9834	0.9751	0.9643	0.951	0.9349	0.9161	0.8946	0.8705	0.8441	0.8156	0.7851	0.7029	0.616	0.5289
		6	6.5	7	7.5	8	8.5	9	9.5	10	11	10	12	14	15
x=	0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0	0	0	0	0	0
	1	0.0174	0.0113	0.0073	0.0047	0.003	0.0019	0.0012	0.0008	0.0005	0.0002	0.0005	0.0001	0	0
	2	0.062	0.043	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0012	0.0028	0.0005	0.0001	0
	3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0049	0.0103	0.0023	0.0005	0.0002
	4	0.2851	0.2237	0.173	0.1321	0.0996	0.0744	0.055	0.0403	0.0293	0.0151	0.0293	0.0076	0.0018	0.0009
	5	0.4457	0.369	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0375	0.0671	0.0203	0.0055	0.0028

https://www.math.uh.edu/~dlabate/poisson_cdf.pdf

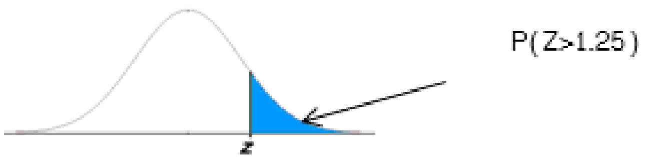


Table C: The Normal Distribution

0.Z5	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

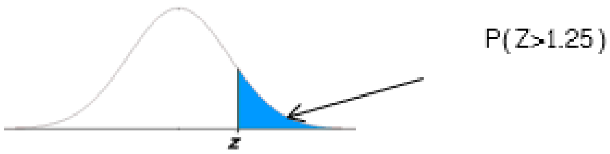


Table C: The Normal Distribution

0.Z5	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table B: The Poisson Cumulative Distribution

		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.2	1.4	1.6	1.8
x=	0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679	0.3012	0.2466	0.2019	0.1653
	1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358	0.6626	0.5918	0.5249	0.4628
	2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197	0.8795	0.8335	0.7834	0.7306
	3	1	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.981	0.9662	0.9463	0.9212	0.8913
	4	1	1	1	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963	0.9923	0.9857	0.9763	0.9636
	5	1	1	1	1	1	1	0.9999	0.9998	0.9997	0.9994	0.9985	0.9968	0.994	0.9896
		2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.5	5	5.5
x=	0	0.1353	0.1108	0.0907	0.0743	0.0608	0.0498	0.0408	0.0334	0.0273	0.0224	0.0183	0.0111	0.0067	0.0041
	1	0.406	0.3546	0.3084	0.2674	0.2311	0.1991	0.1712	0.1468	0.1257	0.1074	0.0916	0.0611	0.0404	0.0266
	2	0.6767	0.6227	0.5697	0.5184	0.4695	0.4232	0.3799	0.3397	0.3027	0.2689	0.2381	0.1736	0.1247	0.0884
	3	0.8571	0.8194	0.7787	0.736	0.6919	0.6472	0.6025	0.5584	0.5152	0.4735	0.4335	0.3423	0.265	0.2017
	4	0.9473	0.9275	0.9041	0.8774	0.8477	0.8153	0.7806	0.7442	0.7064	0.6678	0.6288	0.5321	0.4405	0.3575
	5	0.9834	0.9751	0.9643	0.951	0.9349	0.9161	0.8946	0.8705	0.8441	0.8156	0.7851	0.7029	0.616	0.5289
		6	6.5	7	7.5	8	8.5	9	9.5	10	11	10	12	14	15
x=	0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0	0	0	0	0	0
	1	0.0174	0.0113	0.0073	0.0047	0.003	0.0019	0.0012	0.0008	0.0005	0.0002	0.0005	0.0001	0	0
	2	0.062	0.043	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0012	0.0028	0.0005	0.0001	0
	3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0049	0.0103	0.0023	0.0005	0.0002
	4	0.2851	0.2237	0.173	0.1321	0.0996	0.0744	0.055	0.0403	0.0293	0.0151	0.0293	0.0076	0.0018	0.0009
	5	0.4457	0.369	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0375	0.0671	0.0203	0.0055	0.0028

https://www.math.uh.edu/~dlabate/poisson_cdf.pdf

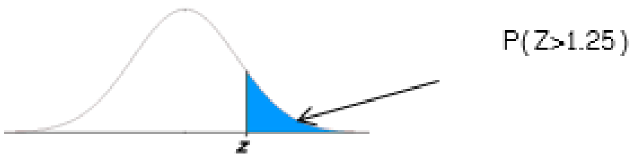
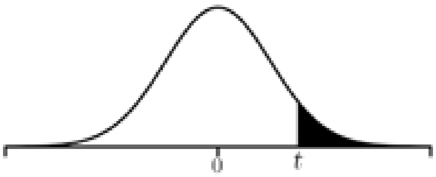


Table C: The Normal Distribution

0.Z5	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table D: The t-Distribution 

df \ t	0.01	0.02	0.05	0.1	0.2	0.3	0.5	0.7	0.9	Two tails test
	0.005	0.01	0.025	0.05	0.1	0.15	0.25	0.35	0.45	One tail test
1	63.657	31.821	12.706	6.314	3.078	1.963	1.000	0.510	0.158	
2	9.925	6.965	4.303	2.920	1.886	1.386	0.816	0.445	0.142	
3	5.841	4.541	3.182	2.353	1.638	1.250	0.765	0.424	0.137	
4	4.604	3.747	2.776	2.132	1.533	1.190	0.741	0.414	0.134	
5	4.032	3.365	2.571	2.015	1.476	1.156	0.727	0.408	0.132	
6	3.707	3.143	2.447	1.943	1.440	1.134	0.718	0.404	0.131	
7	3.499	2.998	2.365	1.895	1.415	1.119	0.711	0.402	0.130	
8	3.355	2.896	2.306	1.860	1.397	1.108	0.706	0.399	0.130	
9	3.250	2.821	2.262	1.833	1.383	1.100	0.703	0.398	0.129	
10	3.169	2.764	2.228	1.812	1.372	1.093	0.700	0.397	0.129	
11	3.106	2.718	2.201	1.796	1.363	1.088	0.697	0.396	0.129	
12	3.055	2.681	2.179	1.782	1.356	1.083	0.695	0.395	0.128	
13	3.012	2.650	2.160	1.771	1.350	1.079	0.694	0.394	0.128	
14	2.977	2.624	2.145	1.761	1.345	1.076	0.692	0.393	0.128	
15	2.947	2.602	2.131	1.753	1.341	1.074	0.691	0.393	0.128	
16	2.921	2.583	2.120	1.746	1.337	1.071	0.690	0.392	0.128	
17	2.898	2.567	2.110	1.740	1.333	1.069	0.689	0.392	0.128	
18	2.878	2.552	2.101	1.734	1.330	1.067	0.688	0.392	0.127	
19	2.861	2.539	2.093	1.729	1.328	1.066	0.688	0.391	0.127	
20	2.845	2.528	2.086	1.725	1.325	1.064	0.687	0.391	0.127	
21	2.831	2.518	2.080	1.721	1.323	1.063	0.686	0.391	0.127	
22	2.819	2.508	2.074	1.717	1.321	1.061	0.686	0.390	0.127	
23	2.807	2.500	2.069	1.714	1.319	1.060	0.685	0.390	0.127	
24	2.797	2.492	2.064	1.711	1.318	1.059	0.685	0.390	0.127	
25	2.787	2.485	2.060	1.708	1.316	1.058	0.684	0.390	0.127	
26	2.779	2.479	2.056	1.706	1.315	1.058	0.684	0.390	0.127	
27	2.771	2.473	2.052	1.703	1.314	1.057	0.684	0.389	0.127	
28	2.763	2.467	2.048	1.701	1.313	1.056	0.683	0.389	0.127	
29	2.756	2.462	2.045	1.699	1.311	1.055	0.683	0.389	0.127	
30	2.750	2.457	2.042	1.697	1.310	1.055	0.683	0.389	0.127	
40	2.704	2.423	2.021	1.684	1.303	1.050	0.681	0.388	0.126	
60	2.660	2.390	2.000	1.671	1.296	1.045	0.679	0.387	0.126	
120	2.617	2.358	1.980	1.658	1.289	1.041	0.677	0.386	0.126	
∞	2.577	2.327	1.960	1.645	1.282	1.037	0.675	0.385	0.126	

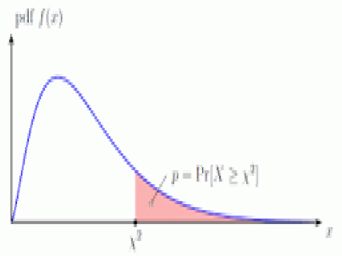


Table E: The Chi-Square Distribution

α / df	0.001	0.005	0.01	0.02	0.025	0.05	0.1	0.2	0.3	0.7	0.9	0.95	0.975	0.99	0.995	0.999
1	10.828	7.879	6.635	5.412	5.024	3.841	2.706	1.642	1.074	0.148	0.016	0.004	0.001	0.000	0.000	0.000
2	13.816	10.597	9.210	7.824	7.378	5.991	4.605	3.219	2.408	0.713	0.211	0.103	0.051	0.020	0.010	0.002
3	16.266	12.838	11.345	9.837	9.348	7.815	6.251	4.642	3.665	1.424	0.584	0.352	0.216	0.115	0.072	0.024
4	18.467	14.860	13.277	11.668	11.143	9.488	7.779	5.989	4.878	2.195	1.064	0.711	0.484	0.297	0.207	0.091
5	20.515	16.750	15.086	13.388	12.833	11.070	9.236	7.289	6.064	3.000	1.610	1.145	0.831	0.554	0.412	0.210
6	22.458	18.548	16.812	15.033	14.449	12.592	10.645	8.558	7.231	3.828	2.204	1.635	1.237	0.872	0.676	0.381
7	24.322	20.278	18.475	16.622	16.013	14.067	12.017	9.803	8.383	4.671	2.833	2.167	1.690	1.239	0.989	0.598
8	26.124	21.955	20.090	18.168	17.535	15.507	13.362	11.030	9.524	5.527	3.490	2.733	2.180	1.646	1.344	0.857
9	27.877	23.589	21.666	19.679	19.023	16.919	14.684	12.242	10.656	6.393	4.168	3.325	2.700	2.088	1.735	1.152
10	29.588	25.188	23.209	21.161	20.483	18.307	15.987	13.442	11.781	7.267	4.865	3.940	3.247	2.558	2.156	1.479
11	31.264	26.757	24.725	22.618	21.920	19.675	17.275	14.631	12.899	8.148	5.578	4.575	3.816	3.053	2.603	1.834
12	32.909	28.300	26.217	24.054	23.337	21.026	18.549	15.812	14.011	9.034	6.304	5.226	4.404	3.571	3.074	2.214
13	34.528	29.819	27.688	25.472	24.736	22.362	19.812	16.985	15.119	9.926	7.042	5.892	5.009	4.107	3.565	2.617
14	36.123	31.319	29.141	26.873	26.119	23.685	21.064	18.151	16.222	10.821	7.790	6.571	5.629	4.660	4.075	3.041
15	37.697	32.801	30.578	28.259	27.488	24.996	22.307	19.311	17.322	11.721	8.547	7.261	6.262	5.229	4.601	3.483
16	39.252	34.267	32.000	29.633	28.845	26.296	23.542	20.465	18.418	12.624	9.312	7.962	6.908	5.812	5.142	3.942
17	40.790	35.718	33.409	30.995	30.191	27.587	24.769	21.615	19.511	13.531	10.085	8.672	7.564	6.408	5.697	4.416
18	42.312	37.156	34.805	32.346	31.526	28.869	25.989	22.760	20.601	14.440	10.865	9.390	8.231	7.015	6.265	4.905
19	43.820	38.582	36.191	33.687	32.852	30.144	27.204	23.900	21.689	15.352	11.651	10.117	8.907	7.633	6.844	5.407
20	45.315	39.997	37.566	35.020	34.170	31.410	28.412	25.038	22.775	16.266	12.443	10.851	9.591	8.260	7.434	5.921
21	46.797	41.401	38.932	36.343	35.479	32.671	29.615	26.171	23.858	17.182	13.240	11.591	10.283	8.897	8.034	6.447
22	48.268	42.796	40.289	37.659	36.781	33.924	30.813	27.301	24.939	18.101	14.041	12.338	10.982	9.542	8.643	6.983
23	49.728	44.181	41.638	38.968	38.076	35.172	32.007	28.429	26.018	19.021	14.848	13.091	11.689	10.196	9.260	7.529
24	51.179	45.559	42.980	40.270	39.364	36.415	33.196	29.553	27.096	19.943	15.659	13.848	12.401	10.856	9.886	8.085
25	52.620	46.928	44.314	41.566	40.646	37.652	34.382	30.675	28.172	20.867	16.473	14.611	13.120	11.524	10.520	8.649
26	54.052	48.290	45.642	42.856	41.923	38.885	35.563	31.795	29.246	21.792	17.292	15.379	13.844	12.198	11.160	9.222
27	55.476	49.645	46.963	44.140	43.195	40.113	36.741	32.912	30.319	22.719	18.114	16.151	14.573	12.879	11.808	9.803
28	56.892	50.993	48.278	45.419	44.461	41.337	37.916	34.027	31.391	23.647	18.939	16.928	15.308	13.565	12.461	10.391
29	58.301	52.336	49.588	46.693	45.722	42.557	39.087	35.139	32.461	24.577	19.768	17.708	16.047	14.256	13.121	10.986
30	59.703	53.672	50.892	47.962	46.979	43.773	40.256	36.250	33.530	25.508	20.599	18.493	16.791	14.953	13.787	11.588
40	73.402	66.766	63.691	60.436	59.342	55.758	51.805	47.269	44.165	34.872	29.051	26.509	24.433	22.164	20.707	17.916
50	86.661	79.490	76.154	72.613	71.420	67.505	63.167	58.164	54.723	44.313	37.689	34.764	32.357	29.707	27.991	24.674
60	99.607	91.952	88.379	84.580	83.298	79.082	74.397	68.972	65.227	53.809	46.459	43.188	40.482	37.485	35.534	31.738
70	112.317	104.215	100.425	96.388	95.023	90.531	85.527	79.715	75.689	63.346	55.329	51.739	48.758	45.442	43.275	39.036
80	124.839	116.321	112.329	108.069	106.629	101.879	96.578	90.405	86.120	72.915	64.278	60.391	57.153	53.540	51.172	46.520
90	137.208	128.299	124.116	119.648	118.136	113.145	107.565	101.054	96.524	82.511	73.291	69.126	65.647	61.754	59.196	54.155
100	149.449	140.169	135.807	131.142	129.561	124.342	118.498	111.667	106.906	92.129	82.358	77.929	74.222	70.065	67.328	61.918

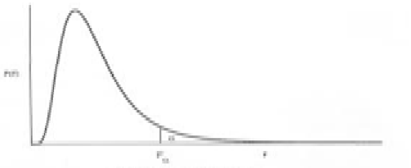


Table F: The F-Distribution

	α=0.05									
Denominator	The numerator Degrees of Freedom									
df	10	12	15	20	24	30	40	60	120	∞
1	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.29
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.41
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.66
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.26
5000	1.83	1.75	1.67	1.57	1.52	1.46	1.40	1.32	1.22	1.05

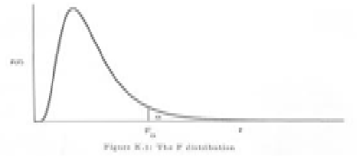
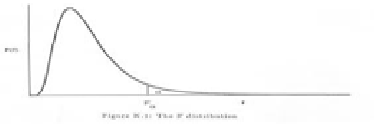


Table F: The F-Distribution

	α=0.025									
Denominator	The numerator Degrees of Freedom									
df	1	2	3	4	5	6	7	8	9	∞
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	1
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	2
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	3
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	4
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	5
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	6
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	7
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	8
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	9
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	10
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	11
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	12
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	13
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	14
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	15
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	16
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	17
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	18
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	19
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	20
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	21
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	22
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	23
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	24
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	25
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	26
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	27
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	28
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	29
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	30
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	40
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	60
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	120
5000	5.03	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.12	5000

Table F: The F-Distribution



	α=0.01								
Denominator	The numerator Degrees of Freedom								
df	1	2	3	4	5	6	7	8	9
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
5000	6.64	4.61	3.79	3.32	3.02	2.81	2.64	2.51	2.41

Table G: Critical Values for the Wilcoxon Signed-Rank Test

n	Two-Tailed Test		One-Tailed Test	
	α = .05	α = .01	α = .05	α = .01
5	--	--	0	--
6	0	--	2	--
7	2	--	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

http://users.stat.ufl.edu/~winner/tables/wilcox_signrank.pdf

គាំទ្រលើថវិការៀបរៀង និពន្ធ និងកែលម្អដោយ
ក្រសួងអប់រំ យុវជន និងកីឡា

មូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍

CAMBODIA