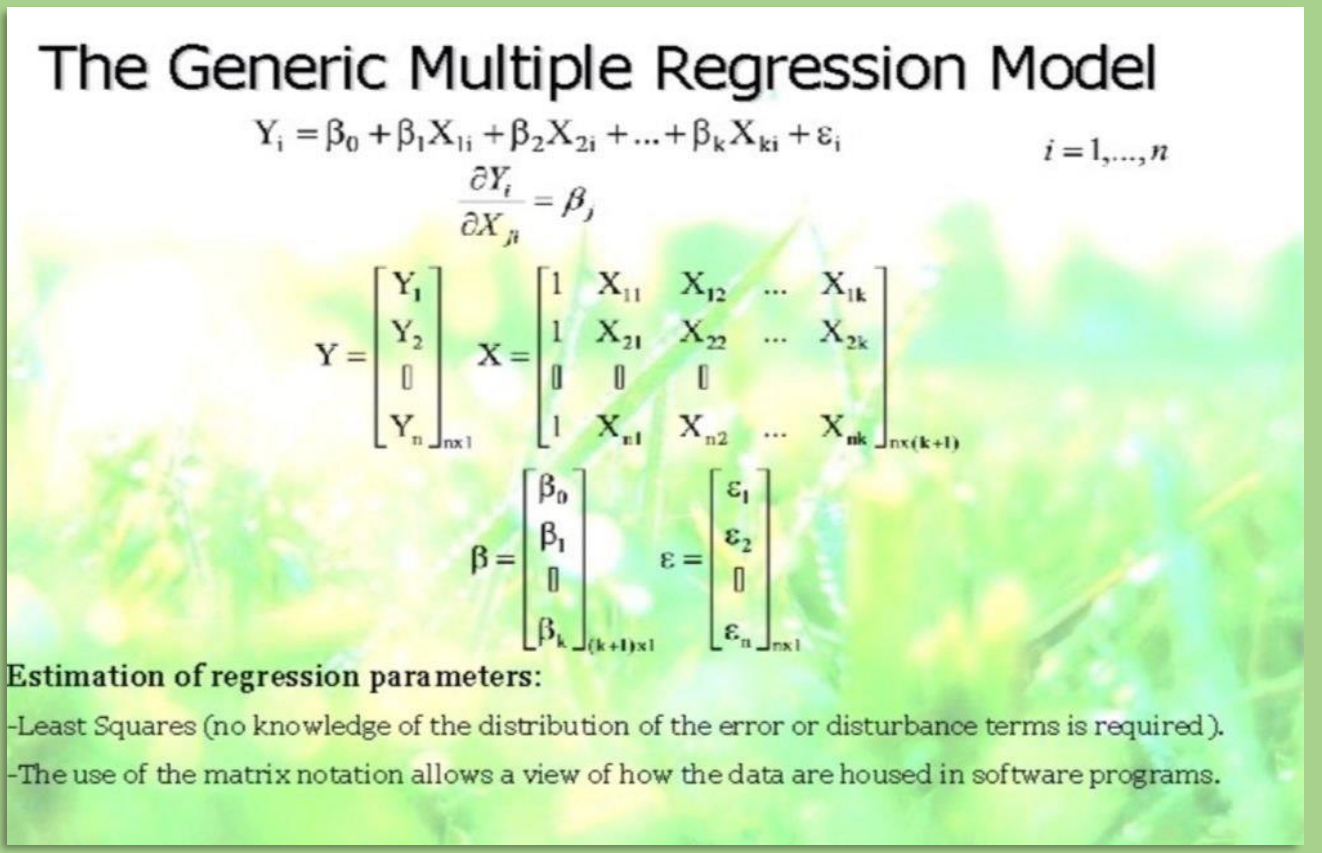
សាកលវិទ្យាល័យជាតិគ្រប់គ្រង

មូលដ្ឋានគ្រឹះនៃម៉ូដែលរឺគ្រេស្យុង



បណ្ឌិត ឈុន ង៉ាឌី

បណ្ឌិត តាយ ផាង

២០២១

មូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេសស្យុង

FUNDAMENTALS OF REGRESSION MODEL

បណ្ឌិត ឈុន រ៉ាឌី

បណ្ឌិត ឆាយ ផាន

ឆ្នាំ ២០២១

រក្សាសិទ្ធិ

រៀបរៀងដោយ បណ្ឌិត **ឈុន ង៉ាឌី**

បណ្ឌិត **ឆាយ ផាង**

គណៈកម្មការត្រួតពិនិត្យ ឯកឧត្តមបណ្ឌិត **ហោ ម៉ៃង** ប្រធាន

លោកបណ្ឌិត **លី សុខហេង** អនុប្រធាន

លោកបណ្ឌិត **តឹម រុនស៊ីណារិទ្ធិ** សមាជិក

លោកស្រី **ប៉ុក ភារុន** សមាជិក

បោះពុម្ពផ្សាយលើកទី១

ភ្នំពេញ ឆ្នាំ២០២១

រក្សាសិទ្ធិគ្រប់យ៉ាង

បុព្វកថា

ដំណើរអភិវឌ្ឍនៃព្រះរាជាណាចក្រកម្ពុជានៅក្នុងយុគសម័យទំនើបនេះ ជាមេរៀនដ៏ជោគជ័យបំផុតមួយ ដែលចាប់ឫសគល់ចេញពីការបញ្ចប់របបប្រល័យពូជសាសន៍ ការបញ្ចប់សង្គ្រាម ការផ្សះផ្សារជាតិ ការកសាងមូលដ្ឋានរឹងមាំនៃសន្តិភាពនិងស្ថេរភាព និងការអភិវឌ្ឍសេដ្ឋកិច្ច។ នៅក្រោយពេលដែលសន្តិភាពត្រូវបានកើតឡើងដោយបរិបូរណ៍នៅឆ្នាំ១៩៩៨ កម្ពុជាទទួលកំណើនសេដ្ឋកិច្ចខ្ពស់ គឺប្រមាណ៨% ក្នុងមួយឆ្នាំ។ លើសពីនេះទៀត អត្រានៃភាពក្រីក្រត្រូវបានកាត់បន្ថយពីប្រមាណ៥៣% នៅឆ្នាំ២០០៤ មកនៅទាបជាង១០% នៅឆ្នាំ២០១៩។ ដំណើរនៃការអភិវឌ្ឍជាតិជាសកម្មភាពដែលបន្តទៅមុខជាប់ជានិច្ច ហើយគោលនយោបាយថ្មីៗដែលមានលក្ខណៈអន្តរវិស័យគ្របដណ្តប់ក៏កំពុងលេចរូបរាងឡើង ដើម្បីតម្រង់ទិសកម្ពុជាឆ្ពោះទៅកាន់ប្រទេសមានប្រាក់ចំណូលមធ្យមកម្រិតខ្ពស់នៅឆ្នាំ២០៣០ និងឈានឡើងជាប្រទេសមានប្រាក់ចំណូលខ្ពស់ នៅឆ្នាំ២០៥០។ ការប្រែប្រួលឆាប់រហ័សនៃនិម្មាបនកម្មពិភពលោកនិងតំបន់ រួមទាំងទំនាក់ទំនងភូមិសាស្ត្រនយោបាយ បានផ្តល់កាលានុវត្តភាពសម្រាប់ការអភិវឌ្ឍឧស្សាហកម្មនៅកម្ពុជា ដែលត្រូវបានរាជរដ្ឋាភិបាលចាត់ទុកជាមូលដ្ឋានគ្រឹះនៃកំណើនសេដ្ឋកិច្ចកម្ពុជា។ រាជរដ្ឋាភិបាលកម្ពុជាបាន និងកំពុងបន្តពង្រឹងនិងអភិវឌ្ឍវិស័យអប់រំឆ្ពោះទៅរកការស្រាវជ្រាវនិងនវានុវត្តន៍ ដើម្បីពង្រឹងសមត្ថភាពនិងជំនាញរបស់ធនធានមនុស្សនៅកម្ពុជា ឱ្យស្របទៅនឹងបរិបទថ្មីនៃការអភិវឌ្ឍ ជាពិសេសការពង្រឹងសហគ្រិនភាពក្នុងការរៀបចំម៉ូដែលធុរកិច្ចថ្មីៗ។ ដើម្បីចាប់យកកាលានុវត្តភាពពីបដិវត្តន៍ឧស្សាហកម្មទី៤ និងសេដ្ឋកិច្ចឌីជីថលដែលកំពុងផុសផុលឡើង ប្រព័ន្ធអេកូឡូហ្ស៊ីដែលបង្កលក្ខណៈអំណោយផលដល់ការបង្កើតថ្មី នវានុវត្តន៍ ការស្រាវជ្រាវ និងអភិវឌ្ឍន៍ត្រូវតែមានការកែលម្អ។

បណ្តាប្រទេសនៅទ្វីបអាស៊ីកំពុងនាំមុខក្នុងការវិនិយោគលើការស្រាវជ្រាវនិងអភិវឌ្ឍ ដោយមានភាគហ៊ុនប្រមាណ៤៤% នៃការវិនិយោគទាំងមូលរបស់ពិភពលោក។ ប្រទេសចិនកំពុងបន្តកសាងហេដ្ឋារចនាសម្ព័ន្ធនៃការវិនិយោគលើការស្រាវជ្រាវនិងអភិវឌ្ឍ ក៏ដូចជាសមត្ថភាពមនុស្ស។ ផ្ទុយទៅវិញ ប្រទេសនៅទ្វីបអាមេរិកខាងត្បូងនិងអាហ្វ្រិក កំពុងស្ថិតនៅឆ្ងាយពីការវិនិយោគនេះ ហើយជាលទ្ធផលប្រទេសទាំងនោះក៏ពុំមានកំណើនសេដ្ឋកិច្ចគួរឱ្យកត់សម្គាល់ដែរ។ ទុនវិនិយោគសរុបលើការស្រាវជ្រាវនិងអភិវឌ្ឍរបស់ប្រទេសនៅទ្វីបអាមេរិកខាងត្បូងនិងអាហ្វ្រិក មានប្រមាណ៥%នៃការវិនិយោគទាំងមូលរបស់ពិភពលោក ក្នុងពេលដែលតំបន់ទាំង២នេះមានប្រជាជនប្រមាណ២០%នៃប្រជាជនពិភពលោក។ ប្រទេសចំនួន៦ដែលមានលំដាប់ខ្ពស់ជាងគេនៅក្នុងការវិនិយោគលើការស្រាវជ្រាវនិងអភិវឌ្ឍ រួមមានសហរដ្ឋអាមេរិក ចិន ជប៉ុន អាល្លឺម៉ង់ ឥណ្ឌា និងកូរ៉េខាងត្បូង ដែលស្មើនឹងប្រមាណ៧០%នៃទុនវិនិយោគសរុបរបស់ពិភពលោក។

តើចំណេះដឹង ផលិតផល និងសេវាកម្មថ្មីទាំងនេះកើតឡើងពីអ្វី? ហើយកើតឡើងដោយរបៀបណា? ព្រះរាជាណាចក្រកម្ពុជាកំពុងតែកសាងមូលដ្ឋានសម្រាប់ការត្រៀមខ្លួនទទួល និងប្រកួតប្រជែងក្នុងយុគសម័យបដិវត្តឧស្សាហកម្មទី៤ នៅក្នុងសេដ្ឋកិច្ចដែលផ្អែកលើពុទ្ធិ ហើយដែលប្រការនេះ

ចាំបាច់តម្រូវឱ្យពលរដ្ឋកម្ពុជា ត្រូវក្លាយខ្លួនជាពលរដ្ឋឌីជីថល ពលរដ្ឋសកល និងពលរដ្ឋដែលប្រកបដោយការទទួលខុសត្រូវ ដែលមានសមត្ថភាពក្នុងការផលិត ចែកចាយ និងប្រើប្រាស់ពុទ្ធិដើម្បីទទួលមនុញ្ញផល និងរួមចំណែកក្នុងកំណើន។ ធនាគារពិភពលោកបានធ្វើការកត់សម្គាល់តាំងពីឆ្នាំ ២០០២នូវបម្លាស់ប្តូរនៃមូលដ្ឋានសេដ្ឋកិច្ច 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 ពិសេសគឺពុទ្ធិសម្បទាក្នុងបុព្វហេតុនៃមនុស្សជាតិនិងឧត្តមប្រយោជន៍នៃប្រទេស។ សង្គមប្រកបដោយពុទ្ធិ គឺពុំគ្រាន់តែជាសង្គមដែលសម្បូរព័ត៌មានប៉ុណ្ណោះទេ តែជាសង្គមដែលប្រជាពលរដ្ឋអាចធ្វើបរិវត្តកម្មពីព័ត៌មានទៅជាមូលធនប្រកបដោយប្រសិទ្ធភាព។ ការរីកចម្រើនទៅមុខដ៏លំដាប់នៃបច្ចេកវិទ្យានិងតំណភ្ជាប់ បានពង្រីកព្រំដែននៃការចូលទៅកាន់ និងការទទួលបានព័ត៌មានជាសកល ហើយដែលក្នុងន័យនេះ ការអប់រំនឹងបន្តវិវត្តទៅមុខនិងមានការផ្លាស់ប្តូរ។ សង្គមមួយដែលមានអំណាន និងរបាប់ជាបុរេលក្ខខណ្ឌនៃជីវភាពប្រចាំថ្ងៃនៃប្រជាពលរដ្ឋ ពេលនោះបំណិននៃអំណាន និពន្ធ និងការគណនាលេខនព្វន្ត គឺជាចលករនៃការរៀនរបស់សិស្ស។ ធាតុដ៏ចម្បងមួយដែលស្ថិតនៅក្នុងការកសាងសង្គមដែលប្រកបដោយពុទ្ធិគឺសៀវភៅសិក្សា ហើយការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សាជាប្រចាំ គឺជានវានុវត្តន៍នៃវិស័យអប់រំដែលនាំទៅរកការសិក្សាពេញមួយជីវិត ការអភិវឌ្ឍសម្បទាអប់រំ និងការចែករំលែកចំណេះដឹង។ មូលដ្ឋានអប់រំ ជាពិសេសគឺគ្រឹះស្ថានឧត្តមសិក្សាត្រូវមានតួនាទីដែលប្រកបដោយការឆ្លើយតប ចំពោះតម្រូវការខាងលើនេះ។ សាស្ត្រាចារ្យ អ្នកស្រាវជ្រាវ និងបុគ្គលិកអប់រំត្រូវបន្តសិក្សាជាប់ជានិច្ច តាមរយៈការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ហើយដែលសៀវភៅសិក្សាទាំងនេះនឹងក្លាយជាស្ពាននៃទំនាក់ទំនងរវាងនវានុវត្តន៍នៃបច្ចេកវិទ្យា និងការរៀននិងបង្រៀននៅក្នុងថ្នាក់រៀន។

សង្គមដែលប្រកបពុទ្ធិ ក៏ជាសង្គមដែលបណ្តុះឱ្យមានរចនាសម្ព័ន្ធទន់នៃសេដ្ឋកិច្ចដែលពឹងផ្អែកលើពុទ្ធិដែរ។ ឧទាហរណ៍ជាក់ស្តែងនៃបែបផែននេះរួមមាន Silicon Valley នៃសហរដ្ឋអាមេរិក សួនឧស្សាហកម្មវិស្វកម្មអាកាសយានយន្តនិងយានយន្តនៅទីក្រុង Munich ប្រទេសអាល្លឺម៉ង់ តំបន់ជីវបច្ចេកវិទ្យានៅក្រុង Hyderabad ប្រទេសឥណ្ឌា តំបន់ផលិតគ្រឿងអេឡិចត្រូនិកនិងសារគមនាគមន៍ឌីជីថលនៅទីក្រុង Seoul ប្រទេសកូរេខាងត្បូង ក៏ដូចជាសួនឧស្សាហកម្មថាមពល និងឥន្ធនគីមីសាស្ត្រនៃប្រទេសប្រេស៊ីល ហើយក៏នៅមានទីក្រុងនៃប្រទេសជាច្រើនទៀតនៅលើពិភពលោក។ លក្ខណៈសម្បត្តិ

នៃទីក្រុងទាំងនេះគឺការប្រើប្រាស់និន្នាការនៃការអភិវឌ្ឍដែលជំរុញ និងតម្រង់ទិសដោយចំណេះដឹង ហើយដែលចំណេះដឹងទាំងនោះកើតចេញជាដំបូងពីការវិនិយោគទៅលើគ្រឹះស្ថានឧត្តមសិក្សា ស្ថាប័នស្រាវជ្រាវ មជ្ឈមណ្ឌលឧត្តមភាពនៃជំនាញជាន់ខ្ពស់ ការប្រកួតប្រជែងដោយគុណាធិបតេយ្យ និងជាពិសេសគឺការបណ្តុះវប្បធម៌អំណាននិងនិពន្ធសៀវភៅ។ ល្បឿននៃការរីកចម្រើនផ្នែកពុទ្ធិ និងបច្ចេកវិទ្យាកំពុងមានសន្ទុះលឿនជាងអ្វីដែលសិស្ស និងនិស្សិតអាចទទួលបានពីគ្រូនៅគ្រឹះស្ថានសិក្សា ដែលធ្វើឱ្យគោលដៅនៃការអប់រំនៅពេលបច្ចុប្បន្ននេះ មានការប្រឈមខ្លាំងជាងពេលណាទាំងអស់។ ឧទាហរណ៍ ក្នុងមួយឆ្នាំ មានសៀវភៅជាង២,២លានចំណងជើង ត្រូវបានសរសេរនិងបោះពុម្ព ដែលក្នុងនោះ ប្រទេសចិនមាន៤៤០ពាន់ ចំណែកឯសហរដ្ឋអាមេរិកមាន៣០៥ពាន់ និងប្រទេសរុស្ស៊ីមាន១២០ពាន់ចំណងជើង។

ខណៈពេលដែលបច្ចេកវិទ្យាកំពុងរីកចម្រើនជារៀងរាល់ថ្ងៃ មធ្យោបាយសម្រាប់អំណានក៏មានច្រើនជម្រើសសម្រាប់សិស្ស-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 បានជំរុញឱ្យមានបរិវត្តកម្មឌីជីថលនៅក្នុងអេកូស៊ីស្តែមនៃការអប់រំ ជាពិសេសការអប់រំតាមប្រព័ន្ធអេឡិចត្រូនិកនិងការអប់រំពីចម្ងាយ ដើម្បីលើកកម្ពស់អំណាន តាមរយៈការផលិតមាតិកាឌីជីថលដែលមានភាពចំរុះ ការកសាងសមត្ថភាពផ្នែកតំណភ្ជាប់និងវេទិកាឌីជីថល ការពង្រីកវិសាលភាពនៃមជ្ឈមណ្ឌលទិន្នន័យ និងការលើកកម្ពស់គុណភាពនៃការផលិតធនធានអប់រំឌីជីថល គួបផ្សំជាមួយការចែកសន្លឹកកិច្ចការឱ្យសិស្សយកទៅរៀននៅផ្ទះ និងការចុះទៅជួបជាមួយសិស្សជាបណ្តុំនៅតាមសហគមន៍។ ក្នុងន័យលើកកម្ពស់អំណាន និងភាពសម្បូរបែបនៃធនធានសៀវភៅសិក្សា ឱ្យកាន់តែមានប្រសិទ្ធភាពនិងភាពសក្តិសិទ្ធិ និងផ្តល់ឱកាសអំណានកាន់តែច្រើនថែមទៀតដល់សិស្សានុសិស្ស និស្សិត និងសាធារណៈជន ក្រសួងអប់រំ យុវជន និងកីឡាលើកទឹកចិត្តនូវចំណុចមួយចំនួនដូចខាងក្រោម៖

១. សាស្ត្រាចារ្យ អ្នកស្រាវជ្រាវ និងបុគ្គលិកអប់រំ សូមបន្តនិងបង្កើនការបោះពុម្ពស្នាដៃបន្ថែមទៀត ដើម្បីធ្វើឱ្យធនធានសម្រាប់អំណានកាន់តែសម្បូរបែប ជាពិសេសធនធានអំណានជាខេមរភាសា

២. គ្រឹះស្ថានឧត្តមសិក្សា សូមផ្តល់លទ្ធភាពគ្រប់បែបយ៉ាង ដើម្បីឱ្យបុគ្គលិកអប់រំគ្រប់លំដាប់ថ្នាក់ និងនិស្សិតគ្រប់កម្រិតសិក្សាអាចចូលរួមអាន និងសិក្សាស្រាវជ្រាវតាមគ្រប់លទ្ធភាពជាមួយធនធានអំណាន ជាពិសេសការរៀបចំឱ្យមានពេលវេលាសម្រាប់សហសិក្សា និងអំណានក្នុងបណ្ណាល័យ

៣. សាស្ត្រាចារ្យតាមមុខវិជ្ជា និងអ្នកស្រាវជ្រាវតាមជំនាញឬវិស័យ ត្រូវរៀបចំដំណើរការរៀនបង្រៀន និងស្រាវជ្រាវដែលមានដាក់បញ្ចូលកិច្ចការស្វ័យសិក្សា សហសិក្សា ឬការស្រាវជ្រាវបណ្ណាល័យដែលតម្រូវឱ្យនិស្សិត ត្រូវអាននិងស្រាវជ្រាវជាមួយធនធានអំណាន

៤. គ្រឹះស្ថានឧត្តមសិក្សា និងមជ្ឈមណ្ឌលស្រាវជ្រាវ ត្រូវខិតខំឱ្យអស់លទ្ធភាពក្នុងការបង្កើតបណ្ណាល័យ មជ្ឈមណ្ឌលរក្សាឯកសារ ឬមជ្ឈមណ្ឌលអប់រំឌីជីថលជាដើម ដើម្បីឱ្យបុគ្គលិកអប់រំគ្រប់លំដាប់ថ្នាក់និងនិស្សិតគ្រប់កម្រិតសិក្សាអាចទទួលបាន និងស្វែងរកប្រភពសម្រាប់អំណានកាន់តែសម្បូរបែប និងមានភាពបត់បែន ឆ្លើយតបតាមតម្រូវការអ្នកអាន

៥. និស្សិតគ្រប់កម្រិតសិក្សាត្រូវខិតខំនិងចំណាយពេលវេលាដើម្បីអាន និងចាត់ទុកវប្បធម៌និងអកប្បកិរិយាអំណានជាផ្នែកមួយ នៃពេលវេលានិងភាពស៊ីវិល័យនៃជីវិតប្រចាំថ្ងៃ

៦. បងប្អូនជនរួមជាតិ ដែលជាមាតាបិតា ឬអ្នកអាណាព្យាបាល សូមជួយជំរុញនិងបង្កលក្ខណៈកាន់តែច្រើនថែមទៀត ជាពិសេសការលៃលកចំណាយនៅក្នុងគ្រួសារសម្រាប់ការទិញសម្ភារៈសិក្សា សៀវភៅអាន និងឧបករណ៍សម្រាប់អំណានដល់កូនៗ ដែលចាត់ទុកជាការវិនិយោគមួយដ៏សំខាន់ សម្រាប់ បង្កើនចំណេះដឹង 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 រួមមានឌីជីថលនីយកម្មសម្រាប់បដិវត្តឧស្សាហកម្ម៤.0 (Digitalization for IR.4.0) ការស្រាវជ្រាវអនុវត្តលើវិស័យកសិកម្ម (Applied Agricultural Research) និងការស្រាវជ្រាវគរុកោសល្យសតវត្សទី២១ (21st Century Pedagogy Research)។

ដោយមានការធ្វើអាទិភាពវូបនីយកម្មទៅលើទិសដៅ នៃការប្រើប្រាស់ថវិកាមូលនិធិសម្រាប់ឆ្នាំ២០២០ ក្រសួងសេដ្ឋកិច្ច និងហិរញ្ញវត្ថុ និងក្រសួងអប់រំ យុវជន និងកីឡា បានផ្តល់ការគាំទ្រដល់ការ**រៀបរៀង និពន្ធ និងកែលម្អ សៀវភៅសិក្សា (Text book) ដែលនឹងត្រូវប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា**។ គោលបំណងនៃការរៀបរៀង និពន្ធ និងកែលម្អ សៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា គឺដើម្បីបង្កើនបរិមាណ លើកកម្ពស់គុណភាព និងពង្រីកសមធម៌នៃធនធានសិក្សាជាខេមរភាសា ជូនដល់និស្សិត

ដែលកំពុងបន្តការសិក្សា និងត្រៀមខ្លួនធ្វើការស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។ លើសពីនេះទៀត ការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា មានគោលដៅដូចខាងក្រោម ៖

១. ឆ្លើយតបជាបន្ទាន់ចំពោះការខ្វះខាតធនធានសិក្សា ដែលជាតម្រូវការសិក្សារបស់និស្សិតនៅកម្រិតឧត្តមសិក្សា

២. លើកកម្ពស់ទំនើបភាវូបនីយកម្ម និងឧត្តមានុវត្តន៍នៃការរៀននិងបង្រៀន និងការស្រាវជ្រាវនៅលើមុខវិជ្ជា កម្មវិធីសិក្សា ឬមុខជំនាញជាក់លាក់

៣. បង្កើនភាពស៊ីជម្រៅក្នុងការកសាងវិជ្ជាជីវៈនិងបទពិសោធន៍សម្រាប់ហានៈសាស្ត្រាចារ្យ និងអ្នកស្រាវជ្រាវ

៤. រួមចំណែកដល់ការកសាងភាពជាសហគមន៍វិជ្ជាជីវៈ ការចែករំលែកបទពិសោធន៍ និងវប្បធម៌នៃការរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា។

ក្រសួងអប់រំ យុវជន និងកីឡា បានវាយតម្លៃខ្ពស់ចំពោះការបោះជំហានប្រកបដោយមនសិការវិជ្ជាជីវៈនៃគ្រឹះស្ថានឧត្តមសិក្សា និងបុគ្គលិកអប់រំទាំងអស់ ក្នុងការរៀបចំ រៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ដើម្បីបង្កើនបរិមាណ លើកកម្ពស់គុណភាព និងពង្រឹងសមធម៌នៃធនធានសិក្សាជាខេមរភាសា ជូននិស្សិតដែលកំពុងបន្តការសិក្សា និងត្រៀមខ្លួនធ្វើការស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។ សៀវភៅសិក្សាជាផ្នែកមួយនៃការទទួលស្គាល់គុណភាពអប់រំនៃគ្រឹះស្ថានឧត្តមសិក្សា និងជាធនធានសិក្សាដែលជាមូលដ្ឋានមួយដ៏សំខាន់ ក្នុងការគាំទ្រដល់ការបង្រៀន និងរៀន ហើយត្រូវមានបរិមាណគ្រប់គ្រាន់ ឆ្លើយតបទៅនឹងកម្មវិធីអប់រំ និងតម្រូវការសិក្សាស្រាវជ្រាវ។ ជាគោលការណ៍ គ្រឹះស្ថានឧត្តមសិក្សាទាំងអស់ ត្រូវមានសៀវភៅសិក្សាដែលប្រើជាគោលសម្រាប់មុខវិជ្ជានីមួយៗ។ ចំនួនសៀវភៅសិក្សាដែលគ្រប់គ្រាន់សម្រាប់ការស្រាវជ្រាវ និងការសិក្សារបស់និស្សិត ត្រូវមានយ៉ាងតិចមួយចំណងជើងក្នុងមួយមុខវិជ្ជា ហើយត្រូវតម្កល់យ៉ាងតិច២ច្បាប់នៅក្នុងបណ្ណាល័យ ឬអាចរកបានតាមប្រព័ន្ធអេឡិចត្រូនិក។ ក្រសួងអប់រំ យុវជន និងកីឡា លើកទឹកចិត្តបន្ថែមទៀតជូនដល់គ្រឹះស្ថានឧត្តមសិក្សារដ្ឋ និងឯកជនដែលបានស្នើសុំថវិកាមូលនិធិ ស.គ.ន រួច សូមចូលរួមបន្ថែមទៀតដើម្បីបង្កើនចំនួនចំណងជើងសៀវភៅ។ ចំណែកគ្រឹះស្ថានឧត្តមសិក្សារដ្ឋ និងឯកជនដែលពុំទាន់បានដាក់ពាក្យស្នើសុំថវិកាមូលនិធិ ដើម្បីរៀបរៀង និពន្ធ និងកែលម្អ សៀវភៅសិក្សានៅកម្រិតឧត្តមសិក្សា សូមរួសរាន់ចូលរួមដើម្បីជាគុណប្រយោជន៍ដល់តម្រូវការដ៏ទទូចនិងថ្លៃថ្លានៃនិស្សិតកម្ពុជាក្នុងការសិក្សា និងស្រាវជ្រាវនៅកម្រិតឧត្តមសិក្សា។

សេចក្តីបញ្ជាក់
នៃមូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍

សៀវភៅសិក្សានេះជាលទ្ធផលនៃការស្នើសុំអនុវត្តថវិកាមូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍ ក្នុងគម្រោង**រៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា ដែលនឹងត្រូវប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា។** សៀវភៅសិក្សានេះ ត្រូវបានរៀបរៀង និពន្ធ ឬកែលម្អដោយមានការធានាអះអាងថាជាស្នាដៃរបស់អ្នកនិពន្ធផ្ទាល់ និងបានឆ្លងកាត់ត្រួតពិនិត្យ ផ្តល់យោបល់ និងវាយតម្លៃដោយក្រុមប្រឹក្សាអប់រំ ក្រុមប្រឹក្សាស្រាវជ្រាវ ឬក្រុមប្រឹក្សាដែលមានតម្លៃស្មើនៃគ្រឹះស្ថានឧត្តមសិក្សា និងតាមរយៈកិច្ចសន្យាដែលបានធ្វើឡើង និងដែលបានតម្កល់ទុកនៅមូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍។ ពាល់ខ្លឹមសារ ការបកស្រាយ ឬរូបភាព ដែលមាននៅក្នុងសៀវភៅនេះ គឺជាជំហរនិងទស្សនៈផ្ទាល់របស់អ្នកនិពន្ធ ហើយពុំឆ្លុះបញ្ចាំង ឬជាតំណាងដល់មូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និងនវានុវត្តន៍ នៃក្រសួងអប់រំ យុវជន និងកីឡាឡើយ។

អារម្ភកថា

ស្នាដៃចងក្រង និង រៀបរៀងមូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងមួយនេះ គឺ យើងបានបកស្រាយលំអិតអំពីរ៉ឺហ្គ្រេស្សុងជាមួយទិន្នន័យបែបcross-sectional ដែលគ្របដណ្តប់លើវិសាលភាព និង តម្រូវការតាមលំដាប់លំដោយនៃការវិភាគទិន្នន័យដោយបង្ហាញវិធីសាស្ត្រ និង បកស្រាយលទ្ធផល។ សៀវភៅនេះ បានណែនាំឲ្យស្គាល់ អំពីមូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងសម្រាប់ប្រើប្រាស់ក្នុងជីវភាពជាក់ស្តែង ជាពិសេសក្នុងផ្នែកធុរកិច្ច និង សេដ្ឋកិច្ច ដោយលើកយកការវាស់វែង ការគណនា ការប៉ាន់ប្រមាណ ការព្យាករណ៍ ការធ្វើតេស្តលើសម្មតិកម្ម ការបង្ហាញ និង ឈានទៅដល់ការសន្និដ្ឋាន ដើម្បីធ្វើការសម្រេចចិត្ត។

ក្រោយពីបានអានសៀវភៅមូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងនេះរួច អ្នកនឹងអាចធ្វើការសាកល្បងក្នុងការវិភាគទិន្នន័យដើម្បីរកឲ្យឃើញនូវប្រសិទ្ធភាពនៃការអនុវត្តន៍គោលនយោបាយនៅក្នុងអង្គភាពរបស់អ្នក និង ក៏អាចព្យាករណ៍ទៅលើលទ្ធផលនៅពេលអនាគតថែមទៀតដែរ។

មូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងនេះពិតជានៅមានការខ្វះខាតខ្លះពិតមែន ប៉ុន្តែគឺអាចបំពេញនូវសេចក្តីត្រូវការរបស់អាជីកម្ម របស់អ្នកសិក្សាស្រាវជ្រាវ ក៏ដូចជាអ្នកអានជាច្រើនទៀតផងដែរ។ អាស្រ័យហេតុនេះ យើងខ្ញុំដែលជាអ្នកចងក្រងស្នាដៃសៀវភៅនេះឡើង រង់ចាំទទួលនូវមតិរិះគន់ដើម្បីស្តាបនា ពីសំណាក់អ្នកអានគ្រប់គ្នា ដែលបានចំណាយពេលវេលាដ៏មានតម្លៃ និង យកចិត្តទុកដាក់ក្នុងការអានសៀវភៅមូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងមួយនេះដែរ។

ជាទីបញ្ចប់យើងខ្ញុំ សូមថ្លែងអំណរគុណជាអតិបរមាចំពោះគណៈកម្មការត្រួតពិនិត្យ ដែលបានចំណាយពេលវេលាដើម្បីតាមដាន និង ពិនិត្យសៀវភៅមូលដ្ឋានគ្រឹះនៃម៉ូដែលរ៉ឺហ្គ្រេស្សុងនេះ។

រាល់ការផ្តល់ជាមតិយោបល់ល្អៗសម្រាប់កែលម្អ យើងខ្ញុំនឹងទទួលដោយក្តីរីករាយបំផុត ដើម្បីធ្វើឲ្យសៀវភៅមួយនេះមានសុក្រឹតភាពក្នុងការបោះពុម្ពលើកក្រោយបន្តទៀត។

សៀវភៅនេះត្រូវបានតាក់តែងឡើងដោយមានជំនួយស្មារតី និង ហិរញ្ញវត្ថុពីប្រភពផ្សេងៗ ដូច្នេះយើងខ្ញុំគោរព និងថ្លែងអំណគុណរចំពោះ ឯកឧត្តម បណ្ឌិត ហោ ប៉េង សាកលវិទ្យាធិការនៃសាកលវិទ្យាល័យជាតិគ្រប់គ្រង និង ឯកឧត្តម បណ្ឌិតសភាចារ្យ ហង់ ជួន ណារ៉ុន រដ្ឋមន្ត្រី ក្រសួងអប់រំ យុវជន និង កីឡា ដែលជាស្ថាបនិកនៃកម្មវិធីរៀបរៀង និពន្ធ និងកែលម្អសៀវភៅសិក្សា សម្រាប់ប្រើប្រាស់នៅកម្រិតឧត្តមសិក្សា ព្រមទាំងម្ចាស់ជំនួយហិរញ្ញវត្ថុដទៃទៀត។ ហើយលើសពីនេះទៅទៀតក៏សូមថ្លែងចំពោះក្រុមគ្រួសាររបស់យើងខ្ញុំ ដែលបានគាំទ្រ និង លើកទឹកចិត្តក្នុងការចងក្រង និង រៀបរៀង សៀវភៅនេះ។

ថ្ងៃចន្ទ ៥ រោច ខែ ជេស្ឋ ឆ្នាំឆ្លូវ ត្រីស័ក ព.ស ២៥៦៥

រាជធានីភ្នំពេញ ថ្ងៃទី ៣១ ខែឧសភា ឆ្នាំ២០២១

មាតិកា

មេរៀនទី១
ចរិតនៃរ៉ឺគ្រ្រេស្យុង

១.១ អ្វីទៅជារ៉ឺគ្រ្រេស្យុង? ០១
១.២ ការវិភាគដោយអំណះអំណាង ០២
១.៣ ម៉ូដែលរ៉ឺគ្រ្រេស្យុង ០២
១.៤ គោលបំណងនៃម៉ូដែលរ៉ឺគ្រ្រេស្យុង ០៩
សំណួរ ១០
ករណីសិក្សា..... ១១

មេរៀនទី២
រ៉ឺគ្រ្រេស្យុងសាមញ្ញ

២.១ សេចក្តីផ្តើម ១២
២.២ មេគុណប៉ាន់ស្មាន ១៤
២.៣ វាយតម្លៃកំរិតភាពជឿជាក់នៃការព្យាករណ៍ ១៩
២.៤ ការគណនាតម្លៃR^2 ..... ២០
២.៥ ឯកតាវាស់វែង និងទម្រង់(functional form)នៃម៉ូដែលរ៉ឺគ្រ្រេស្យុង ២២
    ២.៥.១ ការជះឥទ្ធិពលនៃបំរែបំរួលឯកតាវាស់វែងទៅលើOLS ២២
    ២.៥.២ ការបញ្ចូលNonlinearitiesក្នុងម៉ូដែលរ៉ឺគ្រ្រេស្យុងសាមញ្ញ ២៣
២.៦ លក្ខខណ្ឌនៃម៉ូដែលរ៉ឺគ្រ្រេស្យុង ២៥
លំហាត់ ២៦

មេរៀនទី៣
រ៉ឺគ្រ្រេស្យុងពហុ: ការប៉ាន់ស្មាន

៣.១ រ៉ឺគ្រ្រេស្យុងពហុ (multiple regression) ៣០
    ៣.១.១ ម៉ូដែលមានអញ្ញត្តិឯករាជ្យពីរ ៣១
    ៣.១.២ ម៉ូដែលជាមួយ"k"អញ្ញត្តិឯករាជ្យ (k independent variables) ៣៣
៣.២ យន្តការ និង ការបកស្រាយ Ordinary Least Squares ៣៤
៣.៣ តម្លៃព្យាករណ៍ និង សំណល់ (OLS Fixed Values and Residuals) ៣៦
៣.៤ ការគណនាR^2 (Goodness-of-Fit) ៣៧
៣.៥ លក្ខខណ្ឌនៃរ៉ឺគ្រ្រេស្យុងពហុ (Multiple regression model) ៣៨
៣.៦ ការប៉ាន់ស្មានវ៉ារីយ៉ង់ (σ^2)នៃម៉ូដែលOLS ៣៩

សំណួរ និង លំហាត់ ៤០

មេរៀនទី៤
រឺគ្រ្រេសស្យុងពហុៈ ការធ្វើតេស្ត

៤.១ បំណែងចែកប៉ាន់គំរូនៃOLS Estimators ៤៣
៤.២ ការធ្វើតេស្តសម្មតិកម្ម (t-test) ៤៣
៤.២.១ តេស្តខាងស្តាំ ៤៥
៤.២.២ តេស្តឆ្វេង ៤៦
៤.២.៣ តេស្តទាំងសងខាង ៤៨
៤.២.៤ ការធ្វើតេស្តសម្មតិកម្មផ្សេងទៀត ៤៩
៤.៣ ការគណនា$p - Value$សំរាប់$t - Test$ ៥១
៤.៤ ចន្លោះជឿជាក់ ៥២
៤.៥ ធ្វើតេស្តសម្មតិកម្មអំពីការបញ្ចូលគ្នានៃបណ្តាប៉ារ៉ាមែត្រលីនេអ៊ែរតែមួយ ៥៣
៤.៦ ការធ្វើតេស្តលីនេអ៊ែរពហុ: $F - test$ ៥៥
៤.៧ $F - test$ សារៈសំខាន់នៃរឺគ្រ្រេសស្យុងទាំងមូល ៥៧
សំណួរ និង លំហាត់ ៥៨

មេរៀនទី៥
រឺគ្រ្រេសស្យុងពហុៈ លក្ខណៈបន្ថែម

៥.១ ការជះឥទ្ធិពលនៃប្រភេទទិន្នន័យទៅលើលទ្ធផលនៃម៉ូដែលរឺគ្រ្រេសស្យុង ៦២
៥.១.១ ទម្រង់ឡូការីត ៦៤
៥.១.២ ម៉ូដែលជាមួយអញ្ញត្តិឯករាជ្យលើកជាការេ (Models with Quadratics) .. ៦៥
៥.១.៣ ម៉ូដែលជាមួយលក្ខខណ្ឌអន្តរកម្ម(Interaction Terms) ៦៦
៥.២ គណនាតម្លៃR^2 (Goodness-of-Fit) ៦៩
៥.២.១ តម្លៃR^2ដែលត្រូវកែសម្រួល (*Adjusted* R^2) ៧០
៥.២.២ ការប្រើប្រាស់$Adjusted\ R^2$ដើម្បីជ្រើសរើសរវាង$Nonnested\ Models$... ៧១
៥.៣ ព្យាករណ៍ និង វិភាគតម្លៃលំអៀង ៧៣
៥.៣.១ ចន្លោះជឿជាក់សម្រាប់ការព្យាករណ៍ ៧៣
៥.៣.២ ការព្យាករណ៍y ករណីអញ្ញត្តិមិនឯករាជ្យគឺ log(y) ៧៥
លំហាត់ ៧៧

មេរៀនទី៦
រឺគ្រេស្យុងពហុៈ ពត៌មានបែបគុណភាព

៦.១ បរិយាយពត៌មានបែបគុណភាព ៧៨
៦.២ អញ្ញត្តិឯករាជ្យជាមួយដាំមីតែមួយ(Single Dummy) ៧៩
៦.៣ អន្តរកម្មក្នុងចំណោមបណ្តាអញ្ញត្តិដាំមី ៨០
៦.៤ ការបញ្ចូលពត៌មានបែបអរឌីណាល(Ordinal Information)
ដោយប្រើ Dummy Variables ៨២
៦.៥ អន្តរកម្មរវាងអញ្ញត្តិដាំមី (dummy variable) និង មិនមែនដាំមី ៨៤
៦.៦ ការធ្វើតេស្តអេគ្វីវ៉ាឡង់(equivalence)នៃរឺគ្រេស្យុងពីរ ៨៥
៦.៧ ម៉ូដែលប្រូបាប៊ីលីតេលីនេអ៊ែរ(The Linear Probability Model) ៨៩
លំហាត់ ៩២

មេរៀនទី៧
រឺគ្រេស្យុងពហុៈ Heteroskedasticity

៧.១ ការកើតឡើងនៃheteroskedasticityចំពោះOLS estimator ៩៤
៧.២ ការស្វែងរកheteroskedasticity ៩៦
៧.២.១ ការអនុវត្តន៍ Breusch-Pagan test សម្រាប់heteroskedasticity ៩៧
៧.២.២ ការអនុវត្តន៍ White Test សម្រាប់heteroskedasticity ៩៩
៧.៣ Heteroskedasticity-Robust inference ក្រោយពេលប៉ាន់ស្មានOLS ១០១
៧.៣.១ គណនាHeteroskedasticity-Robust standard error ១០១
៧.៣.២ គណនាHeteroskedasticity-Robust *F* statistic ១០៣
៧.៣.៣ គណនាHeteroskedasticity-Robust *LM* Tests ១០៣
៧.៤ ការប៉ាន់ស្មានWeighted Least Squares (WLS) ១០៤
៧.៤.១ Heteroskedasticityត្រូវបានគេស្គាល់ ១០៤
៧.៤.២ Heteroskedasticityត្រូវបានប៉ាន់ស្មាន: Feasible GLS ១០៧
លំហាត់ ១០៩

មេរៀនទី៨
រឺគ្រេស្យុងពហុៈ Functional Form

៨.១ កំហុសឆ្គងទម្រង់មុខងារ (Functional Form Misspecification) ១១១
៨.២ RESETជាតេស្តទូទៅមួយសម្រាប់Functional Form Misspecification ១១៤
៨.៣ កំហុសឆ្គងក្នុងការវាស់វែង(Errors of Measurement) ១១៥

៨.៣.១ កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិមិនឯករាជ្យ ១១៥
៨.៣.២ កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិឯករាជ្យ ១១៦
៨.៤ ការស្វែងរកកំហុសឆ្គង (Specification Errors) ១១៦
លំហាត់ ១១៨

ឯកសារយោង ១២២
តារាងស្ថិតិ ១២៤

មេរៀនទី១
ចរិតនៃរឺគ្រេស្យុង

ក្នុងការស្រាវជ្រាវផ្នែកសេដ្ឋកិច្ច ហិរញ្ញវត្ថុ គ្រប់គ្រង និង ផ្នែកដទៃទៀត មួយចំនួនជាស្រាវជ្រាវបែបបរិមាណ(quantitative research) ដូច្នេះវាតម្រូវឲ្យមានការវិភាគទិន្នន័យ។ រឺគ្រេស្យុង វាមានសារៈសំខាន់ទាក់ទងនឹងការវិភាគទិន្នន័យនៃការស្រាវជ្រាវបែបបរិមាណ។ ម្យ៉ាងទៀត វាជាមូលដ្ឋានគ្រឹះសម្រាប់និស្សិតដែលសិក្សាមុខវិជ្ជាមួយចំនួនផងដែរ ដូចជា អេកូណូមេទ្រិច (econometrics) ហិរញ្ញវត្ថុសហគ្រាស (quantitative corporate finance)។ល។ មេរៀននេះ ពិភាក្សាលក្ខណៈទូទៅនៃរឺគ្រេស្យុង (regression)។

១.១ អ្វីទៅជារឺគ្រេស្យុង?

រឺគ្រេស្យុង គឺជាវិធីសាស្ត្រស្ថិតិដ៏សំខាន់ សម្រាប់អង្កេតទៅលើទំនាក់ទំនងក្នុងចំណោមអញ្ញត្តិ (variables) ឬ អថេរទាំងឡាយ។

- អំឡុងពេលដែលមានបញ្ហាកូវីត១៩កើតឡើង នៅក្នុងប្រទេសកម្ពុជា គ្រឹះស្ថានអប់រំភាគច្រើនដំណើរការសិក្សាតាមអនឡាញ។ ក្រោយពីដំណើរការសិក្សាបានមួយរយៈពេល ក្នុងនាមជាថ្នាក់ដឹកនាំសាលាមួយចង់ដឹងថា តើកត្តាអ្វីខ្លះដែលជំរុញឲ្យការសិក្សាតាមអនឡាញមានប្រសិទ្ធភាព?
- ក្រុមហ៊ុនអាជីវកម្មមួយ បរិមាណលក់បានធ្លាក់ចុះ អ្នកគ្រប់គ្រងចង់ដឹងថា តើកត្តាអ្វីខ្លះដែលជះឥទ្ធិពលដល់ការលក់?

ផ្អែកលើសំណួរទាំងនេះ គេអាចបង្កើតជាសមីការ ឬ ជាម៉ូដែលរឺគ្រេស្យុង ដើម្បីបញ្ជាក់អំពីទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ (dependent variable) និង អញ្ញត្តិឯករាជ្យ (independent variables)មួយ ឬ ច្រើនជាងមួយ ។

យើងតាងy ជាអញ្ញត្តិ ឬ អថេរមិនឯករាជ្យ និង x ជាអញ្ញត្តិ ឬ អថេរឯករាជ្យ។ ម៉ូដែលរឺគ្រេស្យុងដែលមានអញ្ញត្តិឯករាជ្យតែមួយ គេហៅថារឺគ្រេស្យុងសាមញ្ញ(simple regression)។

$$y = \beta_0 + \beta_1 x_1 + u \qquad (១.១)$$

រីឯម៉ូដែលរឺគ្រេស្យុងដែលមានអញ្ញត្តិឯករាជ្យច្រើនជាងមួយ គេហៅថារឺគ្រេស្យុងពហុ (multiple regression)។

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \qquad (១.២)$$

- β_0:ប៉ារ៉ាម៉ែត្រអ៊ីនធឺរសិប(intercept parameter)គឺជាប៉ារ៉ាម៉ែត្រក្នុងម៉ូដែលរឺគ្រេស្យុងដែលផ្តល់តម្លៃពឹងទុកនៃអញ្ញត្តិមិនឯករាជ្យ (dependent variable) នៅពេលរាល់ អញ្ញត្តិឯករាជ្យ

(independent variables) ស្មើសូន្យ។ ជួនកាលតម្លៃរបស់វា(β_0) ត្រូវបានចាត់ទុកថាថេរ (constant term)។

- $\beta_1, \beta_2, \ldots\ldots, \beta_k$: ប៉ារ៉ាម៉ែត្រជម្រាល (slope parameter) គឺជាមេគុណ (coefficient) នៃអញ្ញត្តិឯករាជ្យ។
- u: អញ្ញត្តិលំអៀង (error term) គឺតំណាងឲ្យអញ្ញត្តិដទៃទៀត ក្រៅពីអញ្ញត្តិឯករាជ្យ ដែលជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ ប៉ុន្តែមិនត្រូវបានអង្កេត (unobserved other factors)។

១.២ ការវិភាគដោយអំណះអំណាង (Empirical analysis)

នៅក្នុងទ្រឹស្តីសេដ្ឋកិច្ច ក៏ដូចជាទ្រឹស្តីអាជីវកម្មភាគច្រើនបង្កើតបណ្តាប្រយោគ(statements) ឬ សម្មតិកម្ម (hypotheses)បែបបរិមាណ។ ឧទាហរណ៍ ទ្រឹស្តីមីក្រូសេដ្ឋកិច្ចថ្លែងថា ថ្លៃផលិតផលប្រើប្រាស់កើនឡើង នោះរំពឹងថា តម្រូវការផលិតផលថយចុះ។ ដូច្នេះ ទ្រឹស្តីសេដ្ឋកិច្ចបង្ហាញនូវទំនាក់ទំនងអវិជ្ជមានរវាងថ្លៃ និង តម្រូវការបរិមាណនៃផលិតផល ដែលគេហៅថា ទ្រឹស្តីតម្រូវការ។ ប៉ុន្តែ ទ្រឹស្តីមិនបានផ្តល់នូវការវាស់វែងជាតួលេខនៃភាពខ្លាំង ឬ ខ្សោយនៃទំនាក់ទំនងរវាងថ្លៃ និង តម្រូវការឡើយ។ វាមិនបានប្រាប់ថា នៅពេលថ្លៃកើនឡើង១ដុល្លារក្នុង១ឯកតានៃផលិតផល តើរំពឹងថា តម្រូវការប្រើប្រាស់ផលិតផលថយចុះប៉ុន្មានឯកតា។ ហេតុនេះ Regression model analysisផ្តល់នូវការសិក្សាទៅលើទំនាក់ទំនងរវាងថ្លៃ និង តម្រូវការដែលបង្ហាញលទ្ធផលជាតួលេខ គេហៅថា Empirical analysis ឬ Empirical research។

Empirical analysis គឺជាការសិក្សាដែលប្រើប្រាស់ការវិភាគទិន្នន័យ ដើម្បីធ្វើតេស្តទៅលើទ្រឹស្តីប៉ាន់ស្មានទំនាក់ទំនង ឬ កំណត់ប្រសិទ្ធភាពនៃគោលនយោបាយ។ ដូច្នេះ វាបង្ហាញអំពីទំនាក់ទំនងក្នុងទម្រង់ជាសមីការដោយផ្អែកលើregression model។

១.៣ ម៉ូដែលរ៉ឺហ្គ្រេស្យុង (Regression Model Analysis)

ការវិភាគម៉ូដែលរ៉ឺហ្គ្រេស្យុងរួមមានជំហានដូចខាងក្រោម:

ជំហានទី១: កំណត់បញ្ហា (Problem statement)

ការវិភាគរ៉ឺហ្គ្រេស្យុង (regression analysis)ចាប់ផ្តើមជាមួយការស្វែងរកបញ្ហា (problem) តាមរយៈការអង្កេត សំភាសន៍ ឬ ការអានឯកសារផ្សេងៗ និង ចាំបាច់ត្រូវដោះស្រាយវា។ ឧទាហរណ៍ ការអានរបាយការណ៍ហិរញ្ញវត្ថុរបស់បណ្តាក្រុមហ៊ុន នោះនឹងទទួលបានពត៌មានអំពីការធ្លាក់ចុះនៃប្រាក់ចំណេញ។ សំណួរស្រាវជ្រាវ (research questions)គឺ តើមូលហេតុ ឬ កត្តាអ្វីខ្លះដែលបណ្តាលឲ្យប្រាក់ចំណេញធ្លាក់ចុះ? តើវាគួរតែត្រូវបានដោះស្រាយយ៉ាងដូចម្តេច? ហេតុនេះ យើងត្រូវស្រាវជ្រាវស្វែងរកដំណោះស្រាយបញ្ហា ឬ រកចម្លើយនៃសំណួរដោយការកំណត់គោលបំណងនៃការស្រាវជ្រាវ (research objectives)ឲ្យបានច្បាស់លាស់។ ទាំងនេះគឺជាជំហានដំបូង និង សំខាន់បំផុតនៃការ

វិភាគរ៉ឺគ្រេស្យុង ពីព្រោះការកំណត់បញ្ហា សំណួរស្រាវជ្រាវ និង គោលបំណងមិនច្បាស់លាស់ វាធ្វើឲ្យការសំយោគទ្រឹស្តី (theoretical framework) សម្រាប់បង្កើតសម្មតិកម្ម (hypotheses) ឬ បណ្តាអញ្ញត្តិផ្សេងៗ និង ឈានទៅដល់ការជ្រើសរើសវិធីសាស្ត្រវិភាគទិន្នន័យមិនបានត្រឹមត្រូវ នោះនាំឲ្យលទ្ធផលនៃការស្រាវជ្រាវបរាជ័យ។

ជំហានទី២: ការជ្រើសរើសអញ្ញត្តិ (Selection of potentially relevant variables)

ក្រោយពេលកំណត់បញ្ហា ជំហានបន្ទាប់គឺ ការធ្វើសំយោគទ្រឹស្តី (theoretical framework)តាមរយៈការអានអត្ថបទដែលបានបោះពុម្ពផ្សាយដោយអ្នកស្រាវជ្រាវ និង សៀវភៅដែលពាក់ព័ន្ធ ឬ ពិភាក្សាជាមួយអ្នកជំនាញផងដែរ សម្រាប់ជាមូលដ្ឋានក្នុងការបង្កើតសម្មតិកម្ម (hypotheses) និង ឈានទៅដល់ការជ្រើសរើសបណ្តាអញ្ញត្តិសំខាន់ៗ។ ជាទូទៅ y តំណាងឲ្យអញ្ញត្តិមិនឯករាជ្យ (dependent variable) និង $x_1, x_2, \dots, x_k$តំណាងឲ្យអញ្ញត្តិឯករាជ្យ (independent variables) ជាអញ្ញត្តិដែលអាច ឬ ជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ (y)។ ឧទាហរណ៍ យោងលើអត្ថបទស្រាវជ្រាវផ្សេងៗបង្ហាញថា កត្តាជះឥទ្ធិពលទៅលើលទ្ធភាពនៃប្រាក់ចំណេញ (profitability)មានដូចជា ទំហំក្រុមហ៊ុន (Malik, 2011) ទុនបង្វិល (Malik, 2011) ប្រសិទ្ធភាពនៃការគ្រប់គ្រងទ្រព្យសម្បត្តិសរុប (Warrad and Al Omari, 2015) លទ្ធភាពទូទាត់បំណុលរយៈពេលខ្លី (Mohd Zaid et al.,2014) និង កម្រិតនៃការប្រើប្រាស់បំណុល (Wald, 1999)។

ជំហានទី៣: ប្រមូលទិន្នន័យ

បន្ទាប់ពីជ្រើសរើសបណ្តាអញ្ញត្តិសំខាន់ៗ យើងត្រូវតែប្រមូលទិន្នន័យសម្រាប់វិភាគ។ ចំពោះempirical analysisនៃឧទាហរណ៍ខាងលើ យើងត្រូវការទិន្នន័យដូចជា កម្រិតនៃប្រាក់ចំណេញ (profitability) តាងដោយreturn on assets (ROA) វាស់វែងជាភាគរយ ទំហំក្រុមហ៊ុន (firm size) តាងដោយទ្រព្យសម្បត្តិសរុប (total assets) វាស់វែងជាដុល្លារ ទុនបង្វិល (working capital - WC) វាស់វែងជាដុល្លារ ប្រសិទ្ធភាពនៃការគ្រប់គ្រងទ្រព្យសម្បត្តិសរុប(company efficiency) តាងដោយ total assets turnover ratio (ATR) វាស់វែងជាចំនួនដង(times) លទ្ធភាពទូទាត់បំណុលរយៈពេលខ្លី(liquidity) តាងដោយ quick ratio (QR) វាស់វែងជាចំនួនដង(times) កម្រិតនៃការប្រើប្រាស់បំណុល (leverage) តាងដោយ debt ratio (DR) វាស់វែងជាចំនួនដង(times)។

ទិន្នន័យមានបីប្រភេទ:

ក. ទិន្នន័យសេរីនៃកាល (time series data)

ទិន្នន័យសេរីនៃកាលរួមមានការអង្កេត(observations)ទៅលើអញ្ញត្តិមួយ ឬ អញ្ញត្តិជាច្រើន ក្នុងរយៈពេលណាមួយ (អាចជាច្រើនថ្ងៃ សប្តាហ៍ ខែ...) (over time)។ ឧទាហរហ៍ យើងកំណត់ទិន្នន័យរបស់ក្រុមហ៊ុនAដូចខាងក្រោម:

obs.no	year	ROA	TA	WC	ATR	QR	DR
1	2011	34	1,924,657	231,200	2.19	1.24	0.34
2	2012	42	2,028,765	358,400	3.28	1.28	0.39
3	2013	39	2,232,100	362,340	4.12	1.86	0.42
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
9	2019	27	2,219,880	218,920	2.16	0.96	0.49
10	2020	25	1,998,560	189,650	1.42	0.68	0.52

ខ. ទិន្នន័យឆ្លងកាត់ផ្នែក (Cross-sectional data)

Cross-sectional dataរួមមានសំណាកនៃមនុស្ស គ្រួសារ ក្រុមហ៊ុន ទីក្រុង ប្រទេស។ល។ ដែលទទួលបានក្នុងពេលជាក់លាក់មួយ (a given point in time)។ លក្ខណៈសំខាន់មួយនៃcross-sectional dataគឺ យើងអាចសន្មតថា ទិន្នន័យទទួលបានពីសំណាកចៃដន្យ (random sampling) ដែលតំណាងឲ្យpopulation។ ឧទាហរណ៍ ក្នុងការវិភាគកត្តាជះឥទ្ធិពលទៅលើកម្រិតនៃប្រាក់ចំណេញ សន្មតថា យើងលើកយក៤០ក្រុមហ៊ុន(random sampling)ដែលតំណាងឲ្យក្រុមហ៊ុនជាច្រើនដោយកំណត់ឆ្នាំជាក់លាក់គឺ ២០១៧។

obs.no	ឈ្មោះក្រុមហ៊ុន	ROA	TA	WC	ATR	QR	DR
1	A	0.28	2,268,000	268,900	2.19	1.78	0.39
2	B	0.30	1,908,674	190,600	1.98	1.65	0.43
3	C	0.27	1,899,650	188,000	2.02	1.44	0.47
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
39	SA	0.32	2,439,000	269,000	2.45	1.90	0.36
40	SM	0.26	2,190,340	210,328	1.80	1.75	0.42

គ. ទិន្នន័យបែបផ្ទាំណែល (Panel data)

Panel dataគឺជាសំណុំនៃទិន្នន័យដែលរួមមាន time serries សម្រាប់cross-sectional dataនីមួយៗ។ ឧទាហរហ៍ យើងកំណត់សំណុំនៃទិន្នន័យរបស់ក្រុមហ៊ុនទាំង៤០សម្រាប់អំឡុងពេល១០ឆ្នាំ។

obs.no	ឈ្មោះក្រុមហ៊ុន	year	ROA	TA	WC	ATR	QR	DR
1	A	2011	34	1,924,657	231,200	2.19	1.24	0.34
2	A	2012	42	2,028,765	358,400	3.28	1.28	0.39
3	A	2013	39	2,232,100	362,340	4.12	1.86	0.42
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
7	A	2017	0.28	2,268,000	268,900	2.19	1.78	0.39
.	.	.	.	.	.	.	.	.
9	A	2019	27	2,219,880	218,920	2.16	0.96	0.49
10	A	2020	25	1,998,560	189,650	1.42	0.68	0.52
11	B	2011						
12	B	2012						
13	B	2013						
.	B	.	.	.	.	.	.	.
.	B	.	.	.	.	.	.	.
.	B	.	.	.	.	.	.	.
17	B	2017	0.30	1,908,674	190,600	1.98	1.65	0.43
.	B	.	.	.	.	.	.	.
19	B	2019						
20	B	2020						
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
331	SM	2011						
332	SM	2012						
333	SM	2013						
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

.	.	.	.	.	.	.	.	.
337	SM	2017	0.26	2,190,340	210,328	1.80	1.75	0.42
.	.	.	.	.	.	.	.	.
339	SM	2019						
400	SM	2020						

<u>សៀវភៅនេះ</u> យើងសិក្សាអំពី**ការវិភាគម៉ូដែលរ៉ឺហ្គ្រេស្យុង**តែជាមួយcross-sectional data។

ជំហានទី៤: កំណត់ម៉ូដែល (Model Specification)

ទម្រង់នៃម៉ូដែលរ៉ឺហ្គ្រេស្យុងដែលបង្ហាញអំពីទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ និង បណ្តាអញ្ញត្តិឯករាជ្យអាចត្រូវបានកំណត់ដោយបណ្តាអ្នកជំនាញ ដោយផ្អែកលើចំណេះដឹង ឬ គោលបំណងរបស់ពួកគាត់ក្នុងវិស័យផ្សេងៗ។ បន្ទាប់មក ម៉ូដែលជាមួយការកំណត់សម្មតិកម្មអាចត្រូវបានបញ្ជាក់ ឬ បដិសេធតាមរយៈការវិភាគទិន្នន័យដែលប្រមូលបាន។ កត់សំគាល់ថា ម៉ូដែលចាំបាច់ត្រូវកំណត់ក្នុងទម្រង់មួយដែលអាចយកទៅប្រើប្រាស់បាន។ យើងចាំបាច់ជ្រើសរើសទម្រង់នៃមុខងារ (function): $y = f(x_1, x_2, \dots, x_k) + u$។ មុខងារនេះអាចបែងចែកចេញជាពីរប្រភេទ: លីនេអ៊ែរ (linear) និង មិនមែនលីនេអ៊ែរ (nonlinear)។ ឧទាហរណ៍ មុខងារលីនេអ៊ែរគឺ:

$$y = \beta_0 + \beta_1 x_1 + u, \qquad (១.៣)$$

មុខងារមិនមែនលីនេអ៊ែរគឺ:

$$y = \beta_0 + e^{\beta_1 x_1} + u, \qquad (១.៤)$$

ម៉ូដែលនីមួយៗខាងក្រោមគឺជាលីនេអ៊ែរ

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u,$$

$$y = \beta_0 + \beta_1 \log(x) + u,$$

ពីព្រោះ ក្នុងករណីនីមួយៗ ប៉ារ៉ាម៉ែត្រ(parameters)ដែលបញ្ចូលមានលក្ខណៈជាលីនេអ៊ែរ ទោះបីទំនាក់ទំនងរវាង y និង x គឺមិនមែនលីនេអ៊ែរក៏ដោយ។ ម៉ូដែលទាំងពីរ យើងអាចសរសេរឡើងវិញដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

$$y = \beta_0 + \beta_1 x_1 + u,$$

ដែលសមីការទី១ យើងមាន $x_1 = x$ និង $x_2 = x^2$ រីឯ សមីការទី២ យើងមាន $x_1 = \log(x)$។ នេះគឺអញ្ញត្តិទាំងឡាយត្រូវបានបំលាស់ប្តូរ (transform)។ រាល់មុខងារមិនមែនលីនេអ៊ែរ (nonlinear functions)អាចត្រូវបានបំលាស់ប្តូរទៅជាមុខងារលីនេអ៊ែរ (linear functions)ដែលត្រូវបានហៅ

ថា*linearizable* functions។ ម៉្យាងវិញទៀត មិនមែនរាល់nonlinear functionsអាចបម្លាស់ប្តូរទៅជាមុខងារលើនេអ៊ែរឡើយ ដូចជា ករណី(១.៤)។ ការបែងចែកប្រភេទផ្សេងៗនៃការវិភាគរ៉ឺហ្គ្រេស្យុងគឺ:

តារាង១.១

ប្រភេទនៃរ៉ឺហ្គ្រេស្យុង	លក្ខណៈ
Univariate	មានអញ្ញត្តិមិនឯករាជ្យតែមួយ និង វាមានលក្ខណៈបែបបរិមាណ
Multivariate	មានអញ្ញត្តិមិនឯករាជ្យច្រើនជាងមួយ និង វាមានលក្ខណៈបែបបរិមាណ
Simple	មានអញ្ញត្តិឯករាជ្យតែមួយ
Multiple	មានអញ្ញត្តិឯករាជ្យច្រើនជាងមួយ
Linear	រាល់ប៉ារ៉ាម៉ែត្រដែលបញ្ចូលទៅក្នុងសមីការមានលក្ខណៈជាបន្ទាត់ត្រង់(linearly)
Nonlinear	ទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ និង អញ្ញត្តិឯករាជ្យមួយចំនួនគឺមិនមែនជាលីនេអ៊ែរ (nonlinear) ឬ ប៉ារ៉ាម៉ែត្រមួយចំនួនមិនអាចបម្លាស់ប្តូរទៅជាលីនេអ៊ែរឡើយ។
Analysis of Variance	រាល់អញ្ញត្តិឯករាជ្យទាំងឡាយមានលក្ខណៈបែបគុណភាព
Analysis of Covariance	អញ្ញត្តិឯករាជ្យមួយចំនួនមានលក្ខណៈបែបបរិមាណ និង មួយចំនួនទៀតមានលក្ខណៈបែបគុណភាព
Logistic	វាមានអញ្ញត្តិមិនឯករាជ្យតែមួយ និង មានលក្ខណៈបែបគុណភាព

ជំហានទី៥: ការប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រ (Parameter Estimation)

ក្រោយពេលម៉ូដែលត្រូវបានកំណត់ និង ទិន្នន័យត្រូវបានប្រមូល កិច្ចការបន្ទាប់គឺត្រូវតែប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រនៃម៉ូដែលផ្អែកលើទិន្នន័យដែលបានប្រមូល។ វិធីសាស្ត្រប៉ាន់ស្មានដែលគេនិយមប្រើបំផុតគឺវិធីសាស្ត្រordinary least squares (OLS)។ ចំពោះOLS ដើម្បីឲ្យការអនុវត្តន៍បានត្រឹមត្រូវ វាត្រូវឆ្លើយតបនឹងលក្ខខណ្ឌ(assumptions)មួយចំនួន (លក្ខខណ្ឌទាំងនោះត្រូវបានលើកយកមកសិក្សាក្នុងមេរៀនក្រោយៗ)។នៅក្នុងសៀវភៅនេះ យើងនឹងពិភាក្សា ordinary least squaresតែជាមួយទិន្នន័យបែបcross-sectional។

ជំហានទី៦: ភាពសមរម្យនៃម៉ូដែល (Model Fitting)

ជំហានបន្ទាប់ក្នុងការវិភាគគឺ ត្រូវតែប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រ៉ឺហ្គ្រេស្យុង ឬ បន្សឹម៉ូដែលទៅនឹងទិន្នន័យ ដែលបានប្រមូលដោយប្រើប្រាស់វិធីសាស្ត្រដែលបានជ្រើសរើស (ឧទាហរណ៍ ordinary least squares)។ ការប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រ$\beta_1, \beta_2, \dots, \beta_k$នៃ(១.២)ត្រូវបានតំណាងដោយ$\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$។ បន្ទាប់មកម៉ូដែលរ៉ឺហ្គ្រេស្យុងក្លាយទៅជា:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k, \qquad (១.៥)$$

$\hat{\beta}_j$ (j = 1, 2,...,k) តំណាងឲ្យប៉ារ៉ាម៉ែត្រដែលត្រូវបានប៉ាន់ស្មាន និង តម្លៃ$\hat{y}$ គេហៅថា*fitted value*។ នៅក្នុងទិន្នន័យ យើងអាចគណនា$\hat{y}_i$ដោយប្រើប្រាស់(១.៥)។ ហេតុនេះ យើងអាចសរសេរបាន សមីការដូចខាងក្រោម:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}, \; i = 1,2, \dots, n \qquad (១.៦)$$

ដែល$x_{i1}, x_{i2}, \dots, x_{ik}$គឺជាតម្លៃនៃអញ្ញត្តិឯករាជ្យនៃការអង្កេតនីមួយៗ។

កត់សំគាល់ថា (១.៥)អាចត្រូវបានប្រើប្រាស់ដើម្បីព្យាករណ៍តម្លៃy ចំពោះតម្លៃណាមួយនៃx_i ដែលទិន្នន័យរបស់វា(x_i)មិនត្រូវបានគេអង្កេត។ ដូច្នេះ $\hat{y}$គេហៅថា តម្លៃដែលត្រូវបានព្យាករណ៍ (predicted value)។

ជំហានទី៧: វាយតម្លៃ និង ជ្រើសរើសម៉ូដែល

សុពលភាពនៃវិធីសាស្ត្រស្ថិតិ (ដូចជា វិភាគរ៉ឺហ្គ្រេស្យុង) ផ្អែកទៅលើលក្ខខណ្ឌ (assumptions)។ ជាទូទៅ លក្ខខណ្ឌត្រូវបានបង្កើតឡើងដែលបរិយាយទិន្នន័យ និង ម៉ូដែល។ ភាពត្រឹមត្រូវនៃលទ្ធផលវិភាគ និង សន្និដ្ឋានទទួលបានពីការវិភាគជាមួយការឆ្លើយតបនឹងលក្ខខណ្ឌ (assumptions)។ មុនពេលប្រើប្រាស់(១.៥)សម្រាប់គោលបំណងណាមួយ ដំបូងយើងចាំបាច់កំណត់ ថាតើ(១.៥)ឆ្លើយតបនឹងលក្ខខណ្ឌដែលបានកំណត់ដែរឬទេ (សិក្សានៅក្នុងមេរៀនក្រោយ)។ យើងចាំបាច់ត្រូវឆ្លើយសំណួរខាងក្រោម:

ក. អ្វីខ្លះជាលក្ខខណ្ឌដែលចាំបាច់សំរាប់OLS?
ខ.តើយើងត្រូវធ្វើដូចម្តេច ដើម្បីដឹងថាតើOLSបានឆ្លើយតបជាមួយនឹងលក្ខខណ្ឌទាំងនោះដែរ ឬ ទេ?
គ. ក្រោយពេលដំណើរការOLS យើងរកឃើញថា លទ្ធផលនៃOLSមិនបានឆ្លើយតប ជាមួយ លក្ខខណ្ឌណាមួយ តើយើងត្រូវធ្វើដូចម្តេច?

កត់សំគាល់ថា ការធ្វើតេស្តទៅលើសុពលភាពនៃលក្ខខណ្ឌ (assumptions)ត្រូវបានអនុវត្ត *មុនពេល* ធ្វើការសន្និដ្ឋានណាមួយដែលទទួលបានពីការវិភាគម៉ូដែលរ៉ឺហ្គ្រេស្យុង(OLS)។ ការវិភាគ

រឺហ្គ្រេស្សុងត្រូវបានគេចាត់ទុកជាដំណើរការអន្តរកម្ម (interactive process) មានន័យថា លទ្ធផល ឬ មុខចេញ(outputs)ត្រូវបានប្រើដើម្បីព្យាបាល(diagnose) កំណត់សុពលភាព(validation) វិនិច្ឆ័យ(criticize) និង កែសម្រួលមុខចូល(inputs)។ ដំណើការត្រូវតែធ្វើម្តងហើយម្តងទៀត រហូតទាល់តែទទួលបាននូវលទ្ធផល ឬ មុខចេញ(outputs)ប្រកបដោយភាពពេញចិត្ត វាមានន័យថា យើងនឹងទទួលបានម៉ូដែល (estimate model) មួយដែលឆ្លើយតបរាល់លក្ខខណ្ឌ (assumptions) ទាំងឡាយ និង បន្សុីគ្នាជាមួយទិន្នន័យបានល្អ។

១.៤ គោលបំណងនៃម៉ូដែលរឺហ្គ្រេស្សុង

ការកំណត់ម៉ូដែលរឺហ្គ្រេស្សុងគឺជាផលិតផលសំខាន់បំផុតនៃការវិភាគ។ វាសង្ខេបនូវទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យy និង បណ្តាអញ្ញត្តិឯករាជ្យ$x_1, x_2, \dots, x_k$។ វាត្រូវបានគេប្រើប្រាស់សម្រាប់គោលបំណងមួយចំនួន ដូចជា សំរាប់វិភាគប្រសិទ្ធភាព ឬ ការជះឥទ្ធិពលនៃគោលនយោបាយ ឬ ព្យាករណ៍តម្លៃផ្សេងៗ។ល។

សំណួរ

១. ចូរបែងចែកអញ្ញត្តិនីមួយៗដូចខាងក្រោមជា អញ្ញត្តិដែលមានលក្ខណៈបែបបរិមាណ ឬ គុណភាព:

ក. ទីតាំងភូមិសាស្ត្រ	ច. ចំនួនកូនក្នុងគ្រួសារ
ខ. ថ្លៃទំនិញ	ឆ. ពណ៌សម្បុរនៃមនុស្ស
គ. សីតុណ្ហភាព	ជ. ការប្រើប្រាស់ប្រេង
ឃ. អត្រាការងារ	ឈ. គណបក្សនយោបាយដែលអ្នកចូលចិត្ត
ង. រស់ជាតិអាហារ	ញ. ចំណូលរបស់ក្រុមហ៊ុន

២. ចូរលើកជាឧទាហរណ៍ក្នុងវិស័យណាមួយដែលអ្នកចូលចិត្តជាមួយការវិភាគរ៉ឺហ្គ្រេស្យុង ដោយឆ្លើយនូវសំណួរមួយចំនួនដូចខាងក្រោម:

ក. កំណត់សំណួរស្រាវជ្រាវ

ខ. កំណត់អញ្ញត្តិមិនឯករាជ្យ និង បណ្តាអញ្ញត្តិឯករាជ្យ

គ. បែងចែកអញ្ញត្តិនីមួយៗជា អញ្ញត្តិដែលមានលក្ខណៈបែបបរិមាណ ឬ គុណភាព

ឃ. កំណត់ តើប្រភេទនៃរ៉ឺហ្គ្រេស្យុង (តារាង១.១ ខាងលើ) អាចត្រូវបានប្រើប្រាស់ដើម្បីវិភាគទិន្នន័យ?

ង. បង្កើតជាទម្រង់នៃម៉ូដែល និង កំណត់ប៉ារ៉ាម៉ែត្ររបស់វា។

៣. នៅក្នុងសំណុំនីមួយៗនៃបណ្តាអញ្ញត្តិខាងក្រោម តើមួយណាជាអញ្ញត្តិមិនឯករាជ្យ និង ណាខ្លះជាអញ្ញត្តិឯករាជ្យ? ចូរពន្យល់

ក. ថ្លៃ តម្រូវការ និង គុណភាពនៃទំនិញ

ខ. ចំនួនបំពង់ផ្សែង និង ការប្រើប្រាស់ប្រេងនៃរថយន្ត

គ. តម្រូវការ និង ការផ្គត់ផ្គង់ទំនិញ

ឃ. ទ្រព្យសម្បត្តិសរុបរបស់ក្រុមហ៊ុន អត្រាប្រាក់ចំណេញនៃទ្រព្យរបស់ម្ចាស់ (return on equity) និង ចំណូលពីការលក់ទំនិញ។

ង. ទម្ងន់របស់មនុស្សម្នាក់ ថាតើគាត់ជាអ្នកជក់បារី ឬ មិនជក់បារី និង ថាតើគាត់កើតមហារីកបំពង់ក ឬ មិន។

ច. ប្រាក់ឈ្នួល បទពិសោធន៍ការងារ រយៈពេលនៃការសិក្សា កម្រិតនៃការសិក្សា និង ភេទ

ឆ. ការចំណាយពេលធ្វើការងារ ការចំណាយពេលក្នុងការងារ និង អាយុ

ជ. ប្រាក់បៀវត្សរបស់អ្នកគ្រប់គ្រង និង លទ្ធផលនៃដំណើរការរបស់ក្រុមហ៊ុន

៤. ចំពោះអញ្ញត្តិនីមួយៗក្នុងសំណួរទី៣:

ក. ចូរបែងចែកអញ្ញត្តិនីមួយៗជាអញ្ញត្តិដែលមានលក្ខណៈបែបបរិមាណ ឬ គុណភាព

ខ. តើប្រភេទនៃរ៉ឺហ្គ្រេស្យុង (តារាង១.១ ខាងលើ) អាចត្រូវបានប្រើប្រាស់ដើម្បីវិភាគទិន្នន័យ?

៥. ឧបមាថា ក្រុមហ៊ុនមួយចំនួនបានសម្រេចចិត្តបង្កើនថ្លៃផលិតផលដោយសារថ្លៃវត្ថុធាតុដើមសម្រាប់ផលិតនូវផលិតផលទាំងនោះបានកើនឡើង។

ក. នៅពេលថ្លៃផលិតផលកើនឡើង តើវាជះឥទ្ធិពលយ៉ាងដូចម្តេចទៅលើតម្រូវការ?

ខ. ចូរបកស្រាយករណីខាងលើនេះដោយភ្ជាប់ជាមួយដំហានទាំងឡាយនៃការវិភាគម៉ូដែលរ៉ឺប្រ្តេស្សុងដែលបានអធិប្បាយក្នុងមេរៀន។

ករណីសិក្សា

ក្នុងអំឡុងពេលដែលគ្រោះមហន្តរាយកើតឡើងដោយវីរុសកូវីដ១៩ វិស័យអប់រំនៅកម្រិតឧត្តមសិក្សាដែលពីមុនដំណើរការក្នុងថ្នាក់ផ្ទាល់ ត្រូវផ្លាស់ទៅជាការសិក្សាពីចម្ងាយសម្រាប់និស្សិត។ ហេតុនេះ វាអាចធ្វើឲ្យគុណភាព និង លទ្ធផលនៃការសិក្សារបស់និស្សិតមានការប្រែប្រួល។ ក្នុងនាមជាអ្នកជំនាញក្នុងវិស័យអប់រំ អ្នកចង់ដឹងថា តើគួរតែធ្វើដូចម្តេចដើម្បីរក្សាបាននូវគុណភាព និង លទ្ធផលនៃការសិក្សាពីចម្ងាយរបស់និស្សិតឲ្យបានល្អ។ យោងបញ្ហាដែលបានកើតឡើង និង គោលបំណងខាងលើ ចូរអ្នកបង្កើតនូវគម្រោងស្រាវជ្រាវ (research proposal)មួយជាempirical researchដោយភ្ជាប់ជាមួយដំហានទាំងឡាយនៃការវិភាគម៉ូដែលរ៉ឺប្រ្តេស្សុង។

មេរៀនទី២
រ៉ឺគ្រ្តេស្យុងសាមញ្ញ

ម៉ូដែលរ៉ឺគ្រ្តេស្យុងសាមញ្ញ (simple regression model) អាចត្រូវបានប្រើប្រាស់ដើម្បីសិក្សាទំនាក់ទំនងរវាងអញ្ញត្តិពីរ។ ជួនកាលវាមានប្រយោជន៍ជាឧបករណ៍វិភាគបែបអំណះអំណាង (Empirical analysis)។ ការយល់ដឹងអំពីម៉ូដែលរ៉ឺគ្រ្តេស្យុងសាមញ្ញ វាជាមូលដ្ឋានដំបូងដែលឈានទៅសិក្សាម៉ូដែលរ៉ឺគ្រ្តេស្យុងពហុ (multiple regression model)។

២.១ សេចក្តីផ្តើម

យើងចាប់ផ្តើមជាមួយករណីសាមញ្ញដោយការសិក្សាទៅលើទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ (dependent variable)ដែលតាងដោយ y និង អញ្ញត្តិឯករាជ្យ (independent variable)មានតែមួយតាងដោយ x។ ម៉ូដែលរ៉ឺគ្រ្តេស្យុងសាមញ្ញ (simple regression model) គេអាចហៅថាម៉ូដែលរ៉ឺគ្រ្តេស្យុងដែលមានអញ្ញត្តិពីរ (two-variable linear regression model ឬ bivariate linear regression model) ពីព្រោះវាមានតែអញ្ញត្តិx និង y ។ក្នុងមេរៀននេះឧទាហរណ៍មួយចំនួននឹងត្រូវបានបង្ហាញថា នៅពេលអញ្ញត្តិxប្រែប្រួល តើវាធ្វើឲ្យអញ្ញត្តិ y ប្រែប្រួលយ៉ាងដូចម្តេច?

ក្នុងការសរសេរម៉ូដែលមួយ ដែលនឹងពន្យល់អំពីអញ្ញត្តិy ដែលអាស្រ័យនឹងអញ្ញត្តិx យើងត្រូវតែប្រឈមមុខជាមួយបញ្ហាបី។ ទី១ ខណៈពេលដែលមានភាពមិនច្បាស់លាស់កើតឡើងទៅលើទំនាក់ទំនងរវាងx និងy តើយើងអនុញ្ញាតឲ្យអញ្ញត្តិ ឬ កត្តាដទៃ ជះឥទ្ធិពលទៅលើ y យ៉ាងដូចម្តេចដែរ? ទី២ តើអ្វីទៅជាសារៈប្រយោជន៍នៃទំនាក់ទំនងរវាងx និងy? ទី៣ តើយើងជឿជាក់ថា x និងy មានទំនាក់ទំនងគ្នាដែរឬទេ?

ម៉ូដែលរ៉ឺគ្រ្តេស្យុងសាមញ្ញមានទម្រង់ដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + u \qquad (២.១)$$

កត់សំគាល់: អត្ថន័យនៃβ_0 β_1 និង u ត្រូវបានបង្ហាញក្នុងមេរៀនទី១។

បើកត្តាដទៃទៀត (u) ត្រូវបានកំណត់ថាថេរ នាំឲ្យបំរែបំរួលនៃ u គឺស្មើសូន្យ ($\Delta u = 0$)។ បន្ទាប់មកគេកំណត់ថា ការជះឥទ្ធិពលនៃx ទៅលើy គួសបានជាបន្ទាត់ត្រង់ (linear effect)។

$$\Delta y = \beta_1 \Delta x_1 \text{ប្រសិន } \Delta u = 0 \quad (២.២)$$

ដូច្នេះបំរែបំរួលy(Δy) គឺស្មើបំរែបំរួលx (Δx)គុណនឹង β_1។ មានន័យថា β_1គឺជាប៉ារ៉ាម៉ែត្រជម្រាល (slop parameter) ឬ ទំហំនៃទំនាក់ទំនងរវាង y និង x ដោយកំណត់កត្តាដទៃទៀត(u)ថេរ។

ឧទាហរណ៍ ២.១: ទម្ងន់ទារក និង ការជក់បារី

ឧបមាថា ទម្ងន់ទារកដែលទើបនឹងកើតត្រូវបានកំណត់ដោយម៉ូដែល:

$$bwght = \beta_0 + \beta_1 cigs + u \qquad (២.៣)$$

ដូច្នេះ $y = bwght$ (ទម្ងន់ទារកដែលទើបនឹងកើត គិតជាអោន) និង $x = cigs$ (ចំនួនបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះ គិតជាចំនួនដើមក្នុងមួយថ្ងៃ) ដោយកំណត់កត្តាជទៃទៀតថេរ។ លំអៀង (u) រួមមានកត្តាជទៃទៀតថេរដូចជា ប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារ (annual family income) ទារកដែលទើបនឹងកើតជាកូនទី (the bird order of this child)។ល។ មេគុណ β_1 វាស់វែងការជះឥទ្ធិពលនៃ *cigs* ទៅលើ *bwght*។ ដោយកំណត់កត្តាជទៃទៀតថេរ (u): $\Delta bwght = \beta_1 \Delta cigs$។

ឧទាហរណ៍ ២.២: ថ្លៃផ្ទះ និង ទំហំផ្ទះ

សន្មតថា ថ្លៃផ្ទះត្រូវបានកំណត់ដោយម៉ូដែល:

$$price = \beta_0 + \beta_1 sqrmt + u \qquad (២.៤)$$

ដូច្នេះ $y = price$ (ថ្លៃផ្ទះនីមួយៗ គិតជាដុល្លារ) និង $x = sqrmt$ (ទំហំផ្ទះ គិតជាម៉ែត្រការេ) ដោយកំណត់កត្តាជទៃទៀតថេរ។ លំអៀង (u) រួមមានកត្តាជទៃទៀតថេរដូចជា ទំហំដីសង់ផ្ទះ (size of the lot) ចំនួនបន្ទប់គេង (number of bedrooms)។ល។ មេគុណ β_1 វាស់វែងការជះឥទ្ធិពលនៃ *sqrmt* ទៅលើ *price*។ ដោយកំណត់កត្តាជទៃទៀតថេរ (u): $\Delta price = \beta_1 \Delta sqrmt$។

ការសន្មតជ៏សំខាន់ (key assumption): ចំពោះផបផ្សៃលសិន (population) តម្លៃមធ្យមនៃ u គឺស្មើសូន្យ ឬ មិនមានទំនាក់ទំនងកើតឡើងរវាង u និង x។

$$E(u|x) = E(u) = 0 \qquad (២.៥)$$

ពន្យល់ (២.៥) ភ្ជាប់ជាមួយឧទាហរណ៍ ២.១ (ទម្ងន់ទារក និង ការជក់បារី)។ សន្មតថា u ដូចជាប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារ (annual family income = faminc)។ ផ្អែកលើ(២.៥) កំណត់ថាកម្រិតមធ្យមនៃ *faminc* មិនមានទំនាក់ទំនងជាមួយនឹង *cigs*។ ឧទាហរណ៍ ប្រសិន $E(faminc|9)$ តំណាងឲ្យប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារទាំងឡាយដែលម្តាយបានជក់បារីក្នុងអំឡុងពេលមានផ្ទៃពោះ ចំនួន៩ដើមក្នុងមួយថ្ងៃ។ $E(faminc|១៤)$ តំណាងឲ្យប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារទាំងឡាយដែលម្តាយបានជក់បារីក្នុងអំឡុងពេលមានផ្ទៃពោះ ចំនួន១៤ដើមក្នុងមួយថ្ងៃ។ ផ្អែកលើ (២.៥): $E(faminc|9)$ និង $E(faminc|១៤)$ គឺត្រូវតែដូចគ្នា។ ជាការពិត កម្រិតមធ្យមប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារនីមួយៗត្រូវតែស្មើគ្នា ទោះបីម្តាយបានជក់បារីក្នុងអំឡុងពេលមានផ្ទៃពោះក្នុងមួយ

ថ្ងៃចំនួនប៉ុន្មានដើមក៏ដោយ។ បើយើងគិតថា ចំនួនមធ្យមប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារនីមួយៗ (*faminc*)កើនឡើងជាមួយចំនួនបារីដែលម្តាយបានជក់បារីក្នុងអំឡុងពេលមានផ្ទៃពោះ (*cigs*) នោះ *(២.៥)គឺខុស*។ នៅពេលយើងមិនអាចអង្កេតប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារ (annual family income = faminc) យើងគ្មានមធ្យោបាយដឹង ថាតើប្រាក់ចំណូលប្រចាំឆ្នាំក្នុងគ្រួសារនីមួយៗស្មើគ្នា ឬអត់ សម្រាប់រាល់គ្រួសារដែលម្តាយបានជក់បារីក្នុងអំឡុងពេលមានផ្ទៃពោះ។ ប៉ុន្តែ(២.៥) យើងត្រូវតែអនុវត្តវា មុនពេលវិភាគម៉ូដែលរីគ្រេសិន (regression analysis)។

ផ្អែកលើ (២.៥): $E(u|x) = 0$ នាំឲ្យទាញបាន:

$$E(y|x) = \beta_0 + \beta_1 x \qquad (២.៦)$$

២.២ ការប៉ាន់ស្មានមេគុណ

ក្នុងម៉ូដែលរ៉ីគ្រេស្យុងសាមញ្ញ (២.១) តើប៉ារ៉ាម៉ែត្រ β_0 និង β_1 ត្រូវបានប៉ាន់ស្មានយ៉ាងដូចម្តេច។ ដើម្បីកំណត់ β_0 និង β_1 យើងត្រូវការចំនួនសំណាក (sample) បានពីផបផ្សេលេសិន (population)។ $\{(x_i, y_i): i = 1, \dots, n\}$ តំណាងឲ្យសំណាកចៃដន្យ (random sample) ចំនួន *n* បានពី population។

$$y_i = \beta_0 + \beta_1 x_i + u \qquad (២.៧)$$

ការសន្មតដ៏សំខាន់ (key assumption): តម្លៃមធ្យមនៃ u គឺ ស្មើសូន្យ។

$$E(u) = 0 \qquad (២.៨)$$

ដូច្នេះ គេទទួលបាន:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \qquad (២.៩)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (២.១០)$$

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right)/n \text{ និង } \bar{y} = \left(\sum_{i=1}^{n} y_i\right)/n \qquad (២.១១)$$

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\right) \Big/ \left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right) \qquad (២.១២)$$

លក្ខខណ្ឌ:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$$

ឧទាហរណ៍ ២.៣: គណនា $\widehat{\beta}_0$ និង $\widehat{\beta}_1$

Leland (1972) បានបង្កើតនូវទំនាក់ទំនងរវាងបរិមាណតម្រូវការទំនិញ (D) គិតជាខ្នង់ពាន់ឯកតា និង ថ្លៃទំនិញក្នុង១ឯកតា (R)គិតជាដុល្លារ ក្នុងស្ថានភាពមិនទៀងទាត់ដូចខាងក្រោម:

$$f(R, D, u) = 0 \quad (២.១៣)$$

ផ្អែកលើអនុគមន៍ (២.១៣) យើងអាចទាញបានជាម៉ូដែលរ៉ឺហ្គ្រេស្យុងសាមញ្ញដូចខាងក្រោម និង ដោយរំពឹងថា $\beta_1 < 0$:

$$D_i = \beta_0 + \beta_1 R_i + u$$

ខាងក្រោមគឺជាទិន្នន័យដែលសន្មតទាក់ទងនឹងបរិមាណតម្រូវការទំនិញ (D) និង ថ្លៃទំនិញក្នុង១ឯកតា (R):

តារាង១.១

Obs.	$D_i = y_i$	$R_i = x_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	124	12	0	-10	0	0
2	89	17	5	-45	-225	25
3	146	9	-3	12	-36	9
4	118	15	3	-16	-48	9
5	80	20	8	-54	-432	64
6	186	7	-5	52	-260	25
7	120	14	2	-14	-28	4
8	140	10	-2	6	-12	4
9	195	5	-7	61	-427	49
10	142	11	-1	8	-8	1
Sum	$\bar{y} = 134$	$\bar{x} = 12$			-1,476	190

$$(2.12) \Rightarrow \hat{\beta}_1 = -\frac{1,476}{190} = -7.76842$$
$$(2.10) \Rightarrow \hat{\beta}_0 = 134 - (-7.76842)(12) = 227.2211$$

ដូច្នេះ យើងទទួលបានបន្ទាត់ត្រង់រ៉ឺហ្គ្រេស្យុង (OLS regression line) ដូចខាងក្រោម:

$$\Rightarrow \widehat{D} = 227.2211 - 7.76842\, R$$

ការប៉ាន់ស្មានទទួលបាននៅក្នុង (២.១០) និង (២.១២) ត្រូវបានហៅថាការប៉ាន់ស្មាន ordinary least squares (OLS) នៃ β_0 និង β_1។ នាំឲ្យគេបានតម្លៃប៉ាន់ស្មាននៃ y:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (២.១៤)$$

តម្លៃសំណល់ (residual "$\hat{u}_i$") សម្រាប់ការអង្កេតនីមួយៗ (observation i) គឺជាភាពខុសគ្នារវាងតម្លៃជាក់ស្តែង y_iនិង តម្លៃប៉ាន់ស្មាន (fitted value "$\hat{y}_i$"):

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad (២.១៥)$$

យើងជ្រើសរើស β_0 និង β_1 ដើម្បីកំណត់ផលបូកនៃតម្លៃសំណល់លើកជាការេ (sum of squared residuals = SSR) និង តម្លៃរបស់វា (SSR)ត្រូវតែអប្បបរិមា:

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \quad (២.១៦)$$

ក្នុងករណីភាគច្រើន $\hat{\beta}_1$អាចគណនាដូចខាងក្រោម:

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x} \quad (២.១៧)$$

ផ្អែកលើ(២.១៧) យើងទាញបាន:

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x \quad (២.១៨)$$

(២.១៦)អាចបញ្ជាក់ថា បម្រែបម្រួល xយើងអាចព្យាករណ៍ទៅលើបម្រែបម្រួល y។ ក្នុងករណី $\hat{\beta}_1 > 0$ បើ xកើនឡើង នោះធ្វើឲ្យ yកើនឡើងផងដែរ។ ផ្ទុយទៅវិញ ក្នុងករណី $\hat{\beta}_1 < 0$ បើ xកើនឡើង នោះធ្វើឲ្យ yចុះ។

ឧទាហរណ៍ ២.៤: ទម្ងន់ទារក និង ការជក់បារី

នៅសហរដ្ឋអាមេរិច គេបានសិក្សាស្រាវជ្រាវទៅលើស្ត្រីចំនួន១,៣៨៨នាក់ដែលមានផ្ទៃពោះ និងបានផ្តល់កំណើតកូនរបស់ពួកគាត់។ ក្នុងនោះមានអញ្ញត្តិសំខាន់ពីរ គឺ អញ្ញត្តិមិនឯករាជ្យ (ទម្ងន់ទារកដែលចាប់កំណើត គិតជាអោន "ounces") និង អញ្ញត្តិឯករាជ្យ (ចំនួនបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះ គិតជាចំនួនដើមក្នុងមួយថ្ងៃ)។ អ្នកស្រាវជ្រាវចង់ដឹងថា តើសកម្មភាពជក់បារីរបស់ស្ត្រីក្នុងអំឡុងពេលមានផ្ទៃពោះ ជះឥទ្ធិពលទៅដល់ទម្ងន់ទារកដែលចាប់កំណើតយ៉ាងដូចម្តេច? ខាងក្រោមគឺជាទម្រង់នៃម៉ូដែលរ៉ឺហ្គ្រេស្យុងសាមញ្ញ (simple regression model) សម្រាប់ជាមូលដ្ឋានក្នុងការសិក្សា:

$$bwght = \beta_0 + \beta_1 cigs + u$$

✓ *bwght*: ទម្ងន់ទារកដែលទើបនឹងកើត គិតជាអោន

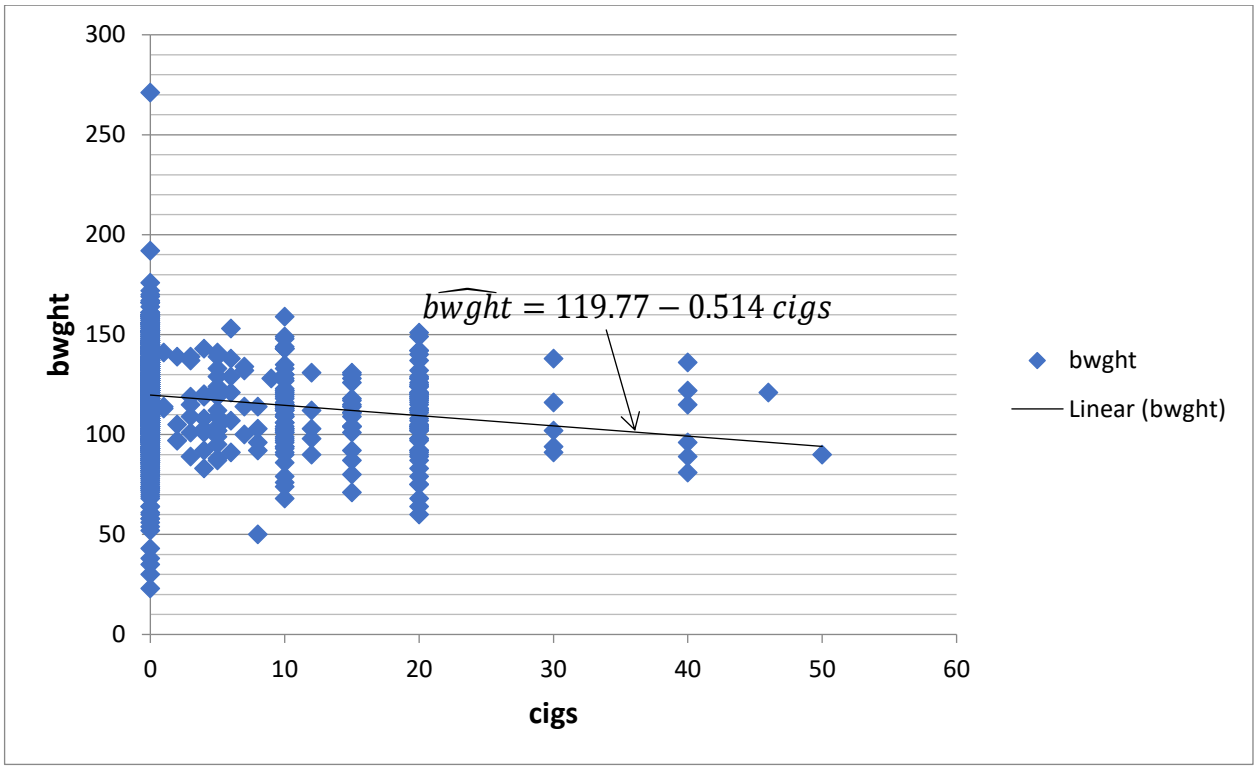
✓ *cigs*: ចំនួនបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះ គិតជាចំនួនដើមក្នុងមួយថ្ងៃ

ប៉ារ៉ាម៉ែត្រជម្រាល (β_1)វាស់វែងបម្រែបម្រួលទម្ងន់ទារកដែលទើបនឹងកើត ($bwght$) នៅពេលមានបម្រែបម្រួលបរិមាណបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះ ដោយរំពឹងថា $\beta_1 < 0$។

ផ្អែកលើទិន្នន័យ (Wooldridge, ២០១៦) ទទួលបាន បន្ទាត់ត្រង់រ៉ឺហ្គ្រេស្យុង (OLS regression line) ដូចខាងក្រោម:

$$\widehat{bwght} = 119.77 - 0.514\, cigs$$

រូបភាព ២.១



យើងប្រើប្រាស់ "bwght hat" ដើម្បីបញ្ជាក់ថា នេះជាសមីការប៉ាន់ស្មាន(an estimated equation)។ តើសមីការនេះត្រូវបានបកស្រាយយ៉ាងដូចម្តេច? *ដំបូង* ប្រសិន *cigs* = 0 (ម្តាយមិនបានជក់បារីទាល់តែសោះក្នុងអំឡុងពេលមានផ្ទៃពោះ) *បន្ទាប់មក* $\widehat{bwght}$ (ទម្ងន់ទារកដែលត្រូវបានព្យាករណ៍) គឺ ១១៩.៧៧អោន (Intercept ឬ $\hat{\beta}_0$ =១១៩.៧៧)។ បន្ទាប់មកទៀត យើងអាចសរសេរបំរែបំរួលទម្ងន់ទារកដែលត្រូវបានព្យាករណ៍ មានទំនាក់ទំនងជាមួយបម្រែបម្រួលរិមាណបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះ: $\Delta\widehat{bwght} = -0.514\ (\Delta cigs)$។ នេះមានន័យថា បើបរិមាណបារីដែលម្តាយបានជក់ក្នុងអំឡុងពេលមានផ្ទៃពោះកើនឡើង១០០ដើម ($\Delta cigs = ១00$) នោះអាចព្យាករណ៍ថាទម្ងន់ទារកដែលទើបនឹងថយចុះអស់៥១.៤អោន ($\Delta\widehat{bwght} = -៥១.៤$អោន)។

ឧទាហរណ៍ ២.៥: ថ្លៃផ្ទះ និង ទំហំផ្ទះ

ឧបមាថា អ្នកជំនាញអចលនទ្រព្យម្នាក់ គាត់ត្រូវការប៉ាន់ស្មានថាតើតម្លៃផ្ទះប្រែប្រួលដោយសារកត្តាអ្វីខ្លះ? ក្នុងនោះកត្តាដំបូងដែលគាត់ចង់ដឹង តើទំហំផ្ទះដែរឬទេ? ដូច្នេះ ម៉ូដែលរ៉ឺហ្គ្រេស្យុងសមញ្ញមួយត្រូវបានបង្កើតឡើង ដែលបង្ហាញដូចខាងក្រោម:

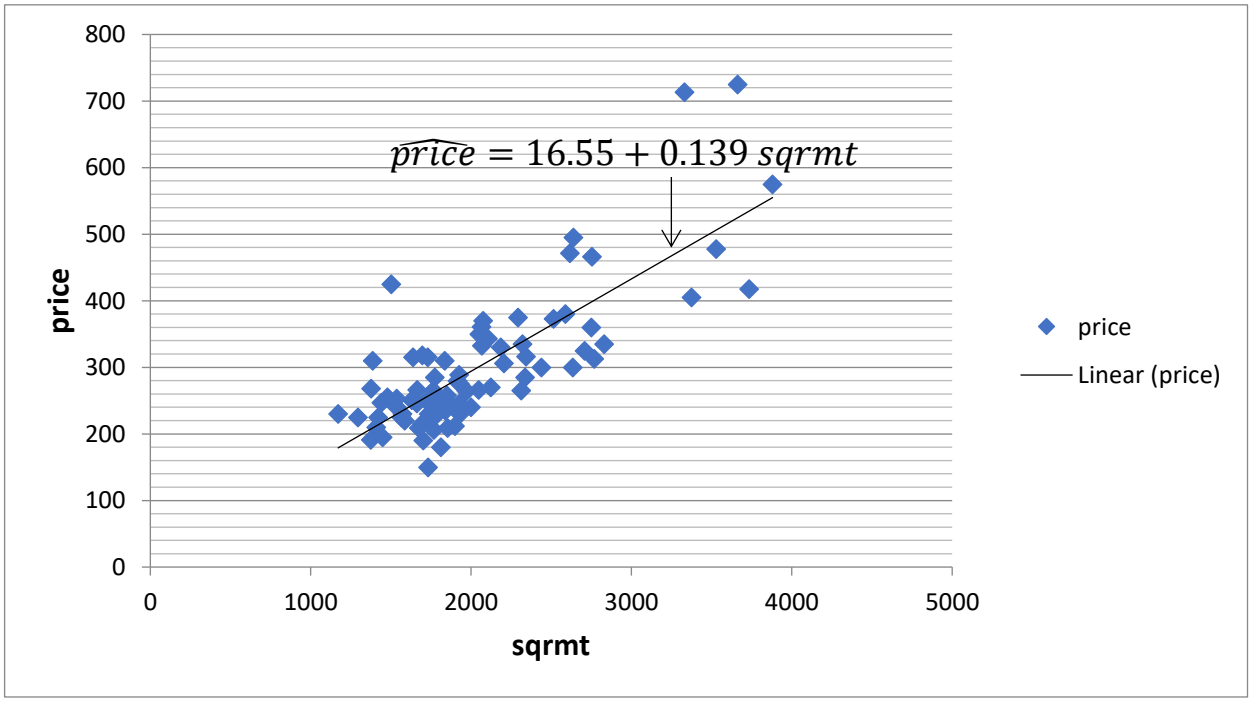
$$price = \beta_0 + \beta_1 sqrmt + u$$

ក្នុងនោះ $price$ គឺជាថ្លៃផ្ទះនីមួយៗ គិតជាពាន់ដុល្លារ និង $sqrmt$ គឺជាទំហំផ្ទះ គិតជាម៉ែត្រការេ។ ប៉ារ៉ាម៉ែត្រជម្រាល (β_1)វាស់វែងបម្រែបម្រួលថ្លៃផ្ទះនីមួយៗ ($price$) អាស្រ័យដោយបម្រែបម្រួលទំហំផ្ទះ ដោយរំពឹងថា $\beta_1 > 0$។

ផ្អែកលើទិន្នន័យដែលទទួលបានពីការអង្កេតផ្ទះចំនួន៨០ គេអាចបង្កើតជាបន្ទាត់ត្រង់រ៉ឺហ្គ្រេស្យុង (OLS regression line) ដូចខាងក្រោម:

$$\widehat{price} = 16.55 + 0.139\ sqrmt$$

រូបភាព ២.២



"price hat" ត្រូវបានប្រើប្រាស់ដើម្បីបញ្ជាក់ថា នេះជាសមីការប៉ាន់ស្មាន (an estimated equation) ។ តើសមីការនេះត្រូវបានបកស្រាយយ៉ាងដូចម្តេច? យើងអាចបកស្រាយថា បម្រែបម្រួលតម្លៃផ្ទះដែលត្រូវបានព្យាករណ៍ មានទំនាក់ទំនងជាមួយបម្រែបម្រួលបរិមាណទំហំផ្ទះ: $\Delta\widehat{price} = 0.139\ (\Delta sqrmt)$។ នេះមានន័យថា បើទំហំផ្ទះកើនឡើង១០០ម៉ែត្រការេ (Δsqrmt $= 900$) នោះអាចព្យាករណ៍ថា ថ្លៃផ្ទះនឹងកើនឡើង១៣.៩ ពាន់ដុល្លារ។

២.៣ ទាយតម្លៃកំរិតតាពឿជាអំនៃការព្យាករណ៍

សំរាប់រ៊ីហ្គ្រេស្យុងដែលមានលក្ខណៈជាបន្ទាត់ត្រង់ (OLS regression) តម្លៃជាក់ស្តែង(y_i)គឺស្មើតម្លៃព្យាករណ៍($\hat{y}_i$) បូកតម្លៃសំណល់($\hat{u}_i$)។

$$y_i = \hat{y}_i + \hat{u}_i \qquad (២.១៩)$$
$$\Rightarrow \hat{u}_i = y_i - \hat{y}_i$$

- ✓ ប្រសិន $\hat{u}_i > 0 \Rightarrow$ ការព្យាករណ៍ខ្វះ
- ✓ ប្រសិន $\hat{u}_i < 0 \Rightarrow$ ការព្យាករណ៍លើស

ជាគោលការណ៍ $\hat{u}_i = 0$ មានន័យ រាល់តម្លៃជាក់ស្តែង(y_i)គឺ ស្មើរាល់តម្លៃព្យាករណ៍($\hat{y}_i$)។ ប៉ុន្តែ ជាទូទៅ $\hat{u}_i$មិនស្មើសូន្យឡើយ។

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad (២.២០)$$

(២.២០): មធ្យមនៃតម្លៃសំណល់មានន័យថា មធ្យមនៃតម្លៃព្យាករណ៍ (sample average of the fitted value) ស្មើគ្នាមធ្យមនៃតម្លៃជាក់ស្តែង (sample average of y_i)។

ផ្អែកលើ *ឧទាហរណ៍ ២.៣* រ៊ីហ្គ្រេស្យុងលីនេអ៊ែរ (regression linear) ដែលយើងទទួលបាន:

$$\widehat{D} = 227.2211 - 7.76842\ R \qquad (២.២១)$$

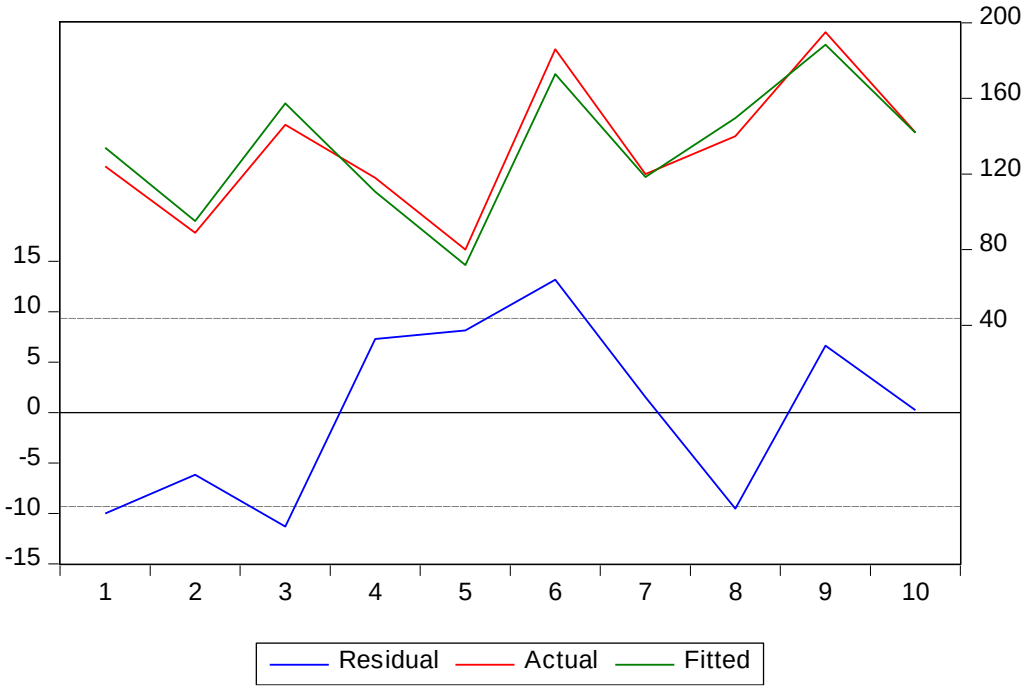
សំណួរសួរថា ប្រសិនរ៊ីហ្គ្រេស្យុងលីនេអ៊ែរ(២.២១)នេះត្រូវបានយកទៅប្រើប្រាស់ សម្រាប់ព្យាករណ៍បរិមាណតម្រូវការទំនិញ ($\widehat{D}$) តើលទ្ធផលនៃការព្យាករណ៍ជឿជាក់ដែរឬទេ?

យើងអាចពិនិត្យមើលវិធីដូចខាងក្រោម:

ទី១: បើលក្ខខណ្ឌ (២.២០)ត្រូវបានឆ្លើយតប នោះលទ្ធផលនៃការព្យាករណ៍គួរឲ្យជឿជាក់។

ទី២: មើល**រូបភាព ២.៣** ករណីបន្ទាត់សំណល់ (residual line) ប្រែប្រួលនៅចន្លោះ (-១០,+ ១០) នោះវាជាការប្រសើរសម្រាប់ការព្យាករណ៍បម្រែបម្រួលបរិមាណតម្រូវការទំនិញ ($\widehat{D}$) ទាក់ទងនឹងបម្រែបម្រួលថ្លៃទំនិញក្នុង១ឯកតា (R)។

រូបភាព ២.៣



២.៤ ការគណនាតម្លៃ R^2

ចំពោះម៉ូដែលរ៉ីហ្គ្រេស្យុង (regression model) យើងមិនត្រឹមតែត្រូវដឹងអំពីទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ (y) និង អញ្ញត្តិឯករាជ្យ (x) ប៉ុណ្ណោះទេ តែត្រូវដឹងអំពីគុណភាពបន្សុំនៃម៉ូដែលជាមួយនឹងទិន្នន័យ (goodness of fit to the data) ផងដែរ។

ក្រោយពីគណនាតម្លៃប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រ ($\hat{\beta}_0$ និង $\hat{\beta}_1$) យើងអាចគណនាចំណុចផ្សេងៗដូចខាងក្រោម:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (២.២២)$$

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad (២.២៣)$$

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 \qquad (២.២៤)$$

- ✓ ក្នុងនោះ *SST* គឺវាស់វែងបម្រែបម្រួលសំណាក (sample) សរុបនៃតម្លៃជាក់ស្តែង (y_i)។
- ✓ *SSE* គឺវាស់វែងបម្រែបម្រួលសំណាក (sample) សរុបនៃតម្លៃព្យាករណ៍ ($\hat{y}_i$)។
- ✓ *SSR* គឺវាស់វែងបម្រែបម្រួលសំណាក (sample) សរុបនៃតម្លៃសំណល់ ($\hat{u}_i$)។

ដូច្នេះ បម្រែបម្រួលសរុបនៃតម្លៃជាក់ស្តែង (y_i) ជាផលបូកនៃបម្រែបម្រួលដែលអាចពន្យល់បាន និង បម្រែបម្រួលដែលមិនអាចពន្យល់បាន។

$$SST = SSE + SSR \qquad (២.២៥)$$

សមីការ (២.២៥)កើតឡើងពីការអង្កេតដូចខាងក្រោម:

$$y_i = \hat{y}_i + (y_i - \hat{y}_i)$$

Observed = Fit + Deviation from fit

ដក $\bar{y}$ ពីអង្គទាំងពីរ យើងបាន:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

Deviation from mean = Deviation due to fit + Residual

ផ្អែកលើសមីការ (២.២៥) $R^2 = SSE/SST$ អាចត្រូវបានបកស្រាយជាសមាមាត្រនៃបំរែបំរួលសរុបនៃអញ្ញត្តិមិនឯករាជ្យ (y) ដែលត្រូវបានពន្យល់ដោយអញ្ញត្តិឯករាជ្យ (x)។

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \qquad (២.២៦)$$

ឧទាហរណ៍ ២.៦: គណនាតម្លៃR^2

ដោយប្រើប្រាស់ទិន្នន័យ និង លទ្ធផលរ៉ឺហ្គ្រេស្យុងម៉ូដែល ($\widehat{D} = 227.2211 - 7.76842\,R$) ក្នុងឧទាហរណ៍ ២.៣ ដើម្បីគណនាតម្លៃR^2។

តារាង១.២

Obs.	$D_i = Y_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$\hat{u}_i$	$\hat{u}_i^2$
1	124	-10	100	134.000	0	0	-10	100
2	89	-45	2,025	95.158	-38.842	1,508.701	-6.15789	37.91967
3	146	12	144	157.305	23.305	543.123	-11.3053	127.809
4	118	-16	256	110.695	-23.305	543.123	7.30526	53.36687
5	80	-54	2,916	71.8526	-62.147	3,862.299	8.14737	66.37961
6	186	52	2,704	172.842	38.842	1,508.701	13.1579	173.1302
7	120	-14	196	118.463	-15.537	241.3984	1.53684	2.361884
8	140	6	36	149.537	15.537	241.3984	-9.53684	90.95136
9	195	61	3,721	188.379	54.379	2,957.076	6.62105	43.83834
10	142	8	64	141.768	7.768	60.34182	0.23158	0.053629
Sum	$\bar{y} = \mathbf{134}$		Σ = 12,162			Σ = 11,466.16		Σ = 695.8105

សំគាល់: $\hat{u}_i = y_i - \hat{y}_i$

$$\Rightarrow R^2 = \frac{\text{១១,៤៦៦.១៦}}{\text{១២,១៦២}} = 0.\text{៩៤២៨ ឬ } \mathbf{94.28\%}$$

R^2 = ៩៤.២៨% អាចបកស្រាយថា បម្រែបម្រួលបរិមាណតម្រូវការទំនិញ (D)ប្រហែល ៩៤.២៨%ត្រូវបានពន្យល់ដោយ ថ្លៃទំនិញក្នុង១ឯកតា (R)។

២.៥ ឯកតាខាស់វែង និង ទម្រង់(functional form)នៃម៉ូដែលរ៉ឺគ្រេស្យុង

មានបញ្ហាសំខាន់ពីរក្នុងការអនុវត្តន៍គឺ (១) តើបំរែបំរួលបណ្តាឯកតាវាស់វែងនៃអញ្ញត្តិមិនឯករាជ្យ (y) ឬ អញ្ញត្តិឯករាជ្យ (x) ឬ អញ្ញត្តិទាំងពីរ ជះឥទ្ធិពលដល់ការប៉ាន់ស្មានOrdinary Least Square (OLS) យ៉ាងដូចម្តេច? (២) តើទម្រង់(functional form)នៃម៉ូដែលរ៉ឺគ្រេស្យុងត្រូវបង្កើតយ៉ាងដូចម្តេច?

២.៥.១ ការជះឥទ្ធិពលនៃបម្រែបម្រួលឯកតាខាស់វែងទៅលើOLS

ក្នុង *ឧទាហរណ៍ ២.៣* យើងវាស់វែងបរិមាណតម្រូវការទំនិញ (D) គិតជាខ្នង់ពាន់ឯកតា រីឯថ្លៃទំនិញក្នុង១ឯកតា (R)គិតជាដុល្លារ និង ទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{D} = 227.2211 - 7.76842\,R \qquad (២.២៧)$$

វាសំខាន់ដែលត្រូវដឹងថា តើអញ្ញាត្តិ D និង R ត្រូវបានវាស់វែងយ៉ាងដូចម្តេច ដើម្បីយល់អំពីលទ្ធផលនៃការប៉ាន់ស្មានក្នុងសមីការ(២.២៧)?

យើងត្រូវតែដឹងថា នៅពេលឯកតាវាស់វែងនៃអញ្ញត្តិមិនឯករាជ្យ និង ឯករាជ្យប្រែប្រួល នោះវានឹងធ្វើឲ្យការប៉ាន់ស្មាននៃOrdinary Least Square (OLS)ទាំងមូលប្រែប្រួលតែម្តង។ ក្នុង *ឧទាហរណ៍ ២.៣* សន្មតថា យើងវាស់វែងតម្រូវការទំនិញ (D)ជាចំនួនបរិមាណ (មិនមែនគិតជាខ្នង់ពាន់ឯកតា)។ ដូច្នេះ យើងបានសមីការដែលប៉ាន់ស្មានដូចខាងក្រោម:

$$\widehat{D} = 227{,}221.1 - 7{,}768.42\,R \qquad (២.២៨)$$

យើងទទួលបានintercept ($\hat{\beta}_0$) និង ប៉ារ៉ាម៉ែត្រជម្រាល ($\hat{\beta}_1$) ក្នុង(២.២៨) ដោយយក intercept ($\hat{\beta}_0$) និង ប៉ារ៉ាម៉ែត្រជម្រាល ($\hat{\beta}_1$) ក្នុង(២.២៧)គុណនឹង១,០០០។ នេះមានន័យថាការបកស្រាយលទ្ធផលក្នុងសមីការ(២.២៧) និង (២.២៨) គឺដូចគ្នា។ ពិនិត្យមើលសមីការ (២.២៨) ប្រសិន

R = 0 បន្ទាប់មក $\widehat{D}$ = ២២៧,២២១.១។ ដូច្នេះ បរិមាណតម្រូវការទំនិញ (D)ដែលព្យាករណ៍គឺ ២២៧,២២១.១ដុល្លារ [តម្លៃដូចគ្នាជាមួយនឹងសមីការ (២.២៧)]។ ជាងនេះទៅទៀត បើថ្លៃទំនិញក្នុង១ឯកតាកើនឡើង១ដុល្លារ នោះបរិមាណតម្រូវការទំនិញ (D)ដែលព្យាករណ៍នឹងថយចុះ ៧,៧៦៨.៤២ដុល្លារ។

ជាទូទៅ វាមានភាពងាយស្រួលយល់ដឹងថាតើអ្វីកើតឡើងចំពោះការប៉ាន់ស្មាន intercept ($\hat{\beta}_0$) និង ប៉ារ៉ាម៉ែត្រជម្រាល ($\hat{\beta}_1$) នៅពេលអញ្ញត្តិមិនឯករាជ្យប្រែប្រួលឯកតាវាស់វែងរបស់វា។ ប្រសិនអញ្ញត្តិមិនឯករាជ្យ (dependent variable) ត្រូវបានគុណជាមួយចំនួនថេរ c (មានន័យថា តម្លៃនីមួយៗនៃសំណាក (sample) ត្រូវបានគុណជាមួយចំនួនថេរ c) បន្ទាប់មក intercept ($\hat{\beta}_0$) និង ប៉ារ៉ាម៉ែត្រជម្រាល ($\hat{\beta}_1$)នៃOLS ក៏ត្រូវបានគុណជាមួយចំនួនថេរ cផងដែរ។ ករណី (២.២៨) ចំនួនថេរ c គឺស្មើ ១,០០០។

យើងអាចប្រើប្រាស់*សមីការ (២.២៨)* ផងដែរ ដើម្បីស្វែងយល់ថា តើអ្វីនឹងកើតឡើង នៅពេលឯកតាវាស់វែងនៃអញ្ញត្តិឯករាជ្យ(independent variable) ត្រូវបានប្រែប្រួល។ ក្នុងសមីការ (២.២៨) ថ្លៃទំនិញក្នុង១ឯកតាគិតជាដុល្លារអាមេរិច (សន្មតថា ១ដុល្លារអាមេរិចស្មើនឹង៤,០០០រៀល)។ នៅពេលយើងបម្រែបម្រួលថ្លៃទំនិញក្នុង១ឯកតា(independent variable) ពីដុល្លារអាមេរិចទៅជាប្រាក់រៀលខ្មែរ ចំពោះតម្រូវការទំនិញ (D)ជាចំនួនបរិមាណ(dependent variable)ដូច*សមីការ (២.២៨)* ដដែល។ ករណីនេះ ប៉ារ៉ាម៉ែត្រជម្រាល($\hat{\beta}_1$):

$$\hat{\beta}_1 = \frac{-7{,}768.42}{4{,}000} = -1.942105$$

ដូច្នេះ យើងបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$\widehat{D} = 227{,}221.1 - 1.942\ R \qquad (២.២៩)$$

សមីការ(២.២៩) បើថ្លៃទំនិញក្នុង១ឯកតាកើនឡើងមួយរៀល នោះបរិមាណតម្រូវការទំនិញថយចុះ១.៩៤២ឯកតា។

កត់សំគាល់ថា ការប្រែប្រួលពី(២.២៨)ទៅ(២.២៩) អញ្ញត្តិឯករាជ្យ(independent variable) ត្រូវបានចែកជាមួយ៤,០០០។ ជាទូទៅ ប្រសិនអញ្ញត្តិឯករាជ្យត្រូវបានចែក ឬ គុណជាមួយចំនួនថេរ c (some nonzero constant) បន្ទាប់មក ប៉ារ៉ាម៉ែត្រជម្រាល($\hat{\beta}_1$)នៃOLS ត្រូវបានគុណ ឬ ចែកជាមួយចំនួនថេរ c ជាបន្តបន្ទាប់។

នៅក្នុងផ្នែកមុន យើងកំណត់R^2 ជាgoodness of fit សំរាប់OLS។ យើងអាចសួរថា នៅពេលឯកតាវាស់វែងនៃអញ្ញត្តិមិនឯករាជ្យ ឬ អញ្ញត្តិឯករាជ្យប្រែប្រួល តើR^2ប្រែប្រួលដែរឬទេ ?
ចំលើយគឺថា Goodness of fit នៃម៉ូដែលរ៉ឺហ្គ្រេស្យុងមិនផ្អែកទៅលើឯកតាវាស់វែងនៃអញ្ញត្តិ។ មានន័យថា បម្រែបម្រួលនៃR^2 មិនទាក់ទងនឹងបម្រែបម្រួលនៃឯកតាវាស់វែងនៃអញ្ញត្តិមិនឯករាជ្យ ឬ អញ្ញត្តិឯករាជ្យឡើយ។

២.៥.២ ការបញ្ចូលNonlinearitiesក្នុងម៉ូដែលរ៉ឺហ្គ្រេស្យុងសាមញ្ញ

ក្នុងករណីខ្លះ ទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ និង អញ្ញត្តិឯករាជ្យ មិនមែនសុទ្ធតែមានលក្ខណៈជាលីនេអ៊ែរឡើយ។ ដូច្នេះ យើងនឹងបញ្ចូលnonlinearitiesក្នុងម៉ូដែលរ៉ឺហ្គ្រេស្យុង។

ជាញឹកញាប់ អ្នកស្រាវជ្រាវបានបញ្ចូលទម្រង់ឡូការីត (logarithmic form) ទៅក្នុងម៉ូដែលរ៉ឺហ្គ្រេស្យុង។ *ហេតុអ្វីត្រូវធ្វើដូច្នេះ?*

ឧទាហរណ៍ ២.៧: នៅឆ្នាំ១៩៩៨ អាលតុង និង មៀលល័របានសិក្សាទៅលើការជះឥទ្ធិពលនៃបទពិសោធន៍ការងារ(work experience) ទៅលើប្រាក់ឈ្នួលនារី (female wages) និង ការផ្គត់ផ្គង់កំលាំងពលកម្ម (labor supply)។ រកឃើញថា បទពិសោធន៍ការងារជះឥទ្ធិពលទៅលើប្រាក់ឈ្នួលនារី និង ការផ្គត់ផ្គង់កំលាំងពលកម្ម។ ហេតុដូច្នេះ នៅក្នុងឧទាហរណ៍នេះ យើងលើកយកសំណាកចំនួន ៥២៦នាក់ ដើម្បីសិក្សាទៅលើការជះឥទ្ធិពលនៃបទពិសោធន៍ការងារទៅលើប្រាក់ឈ្នួល និង សរសេរបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$wage = \beta_0 + \beta_1 exp. + u$$

ក្នុងនោះអញ្ញត្តិមិនឯករាជ្យប្រាក់ឈ្នួល (wage) គិតជា២០,០០០រៀលក្នុង១ម៉ោង និង អញ្ញត្តិឯករាជ្យបទពិសោធន៍ការងារ (exp)គិតជាឆ្នាំ។ ប៉ារ៉ាម៉ែត្រជម្រាល (β_1)វាស់វែងបម្រែបម្រួលប្រាក់ឈ្នួលអាស្រ័យដោយបម្រែបម្រួលបទពិសោធន៍ការងារ ដោយរំពឹងថា $\beta_1 > 0$។

ផ្អែកលើទិន្នន័យដែលទទួលបានពីការអង្កេតចំនួន៥២៦នាក់ គេអាចបង្កើតជាបន្ទាត់ត្រង់រ៉ឺហ្គ្រេស្យុង (OLS regression line) ដូចខាងក្រោម:

$$\widehat{wage} = 5.373 + 0.031\ exp \qquad (២.៣០)$$

យោងលើ(២.៣០) យើងអាចបកស្រាយថា បម្រែបម្រួលប្រាក់ឈ្នួលដែលត្រូវបានព្យាករណ៍ មានទំនាក់ទំនងជាមួយបម្រែបម្រួលបទពិសោធន៍ការងារ: $\Delta\widehat{wage} = 0.031\ (\Delta exp)$។ នេះមានន័យថាបើបទពិសោធន៍ការងារកើនឡើង១ឆ្នាំ នោះអាចព្យាករណ៍ថាប្រាក់ឈ្នួលនឹងកើនឡើង៦២០រៀល។ ក្នុងករណី អ្នកធ្វើការមានបទពិសោធន៍ការងារខុសគ្នាគិតជាចំនួនឆ្នាំ ការប្រើប្រាស់មេគុណ0.0៣១ សម្រាប់វាស់វែងទំហំនៃបំរែបំរួលប្រាក់ឈ្នួល គឺប្រហែលជាពុំសមហេតុផលឡើយ ពីព្រោះបទពិសោធន៍ការងារគិតជាចំនួនឆ្នាំខុសគ្នា ប្រាក់ឈ្នួលគិតជារៀលក៏អាចខុសគ្នាដែរ។

ហេតុដូច្នេះ យើងអាចបំពាក់ទម្រង់ឡូការីត(logarithmic form)នៃ(២.៣០) ទៅលើអញ្ញត្តិមិនឯករាជ្យប្រាក់ឈ្នួល (wage) និង ទទួលបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$\log(wage) = \beta_0 + \beta_1 exp. + u, \qquad (២.៣១)$$

ប្រសិន$\Delta u = 0$, នោះ $\%\Delta wage \approx (100 \times \beta_1)\Delta exp.$ (២.៣២)

សមីការ(២.៣២)មានន័យថា បើបទពិសោធន៍ការងារកើន១ឆ្នាំ នោះអាចព្យាករណ៍ថា ប្រាក់ឈ្នួលក៏កើនឡើងប្រហែល១០០$\times\beta_1$ភាគរយ។ ផ្អែកលើ (២.៣១) យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\log\widehat{(wage)} = 1.5490 + 0.0044\ exp.$$

មានន័យថាបើបទពិសោធន៍ការងារកើនមួយឆ្នាំ គេអាចព្យាករណ៍ថា ប្រាក់ឈ្នួលប្រហែលកើនឡើង០.៤៤%។

តារាង១.៣

សង្ខេបទម្រង់ទាក់ទងនឹងឡូការីត			
ម៉ូដែល	**អញ្ញត្តិមិនឯករាជ្យ**	**អញ្ញត្តិមិនឯករាជ្យ**	**ការបកស្រាយβ_1**
level-level	y	x	$\Delta y = \beta_1 \Delta x$
level-log	y	log(x)	$\Delta y = (\beta_1/100)\%\Delta x$
log-level	log(y)	x	$\%\Delta y = (100\beta_1)\Delta x$
log-log	log(y)	log(x)	$\%\Delta y = \beta_1 \%\Delta x$

២.៦ លក្ខខណ្ឌនៃម៉ូដែលរ៉ឺហ្គ្រេស្យុង (Assumptions for OLS regression model)

ដើម្បីចៀសវាងកំហុសឆ្គងកើតឡើង ម៉ូដែលរ៉ឺហ្គ្រេស្យុងត្រូវគោរពតាមលក្ខខណ្ឌដូចខាងក្រោម:

លក្ខខណ្ឌទី១: ម៉ូដែលរ៉ឺហ្គ្រេស្យុងត្រូវតែជាអនុគមន៍លីនេអ៊ែរទាក់ទងនឹងប៉ារ៉ាម៉ែត្រ (Linear in parameters)។

លក្ខខណ្ឌទី២: ប៉ាន់គំរូដែលត្រូវបានជ្រើសរើសមានលក្ខណៈដោយចៃដន្យ (Random Sampling)។

លក្ខខណ្ឌទី៣: រាល់បម្រែបម្រួលតម្លៃនៃអញ្ញត្តិឯករាជ្យ (Sample Variation in the Explanatory Variable) គឺមិនដូចគ្នាឡើយ។ នេះមានន័យថា ប្រសិនគំលាតបទដ្ឋាន (Standard Deviation)នៃអញ្ញត្តិឯករាជ្យស្មើសូន្យ នោះលក្ខខណ្ឌទី៣ មិនត្រូវបានឆ្លើយតបឡើយ។

លក្ខខណ្ឌទី៤: ចំពោះម៉ូដែលរ៉ឺហ្គ្រេស្យុង $y = \beta_0 + \beta_1 x_1 + u$ ។ លក្ខខណ្ឌគឺ Zero Conditional Mean មានន័យថា តម្លៃមធ្យមនៃ u គឺ ស្មើសូន្យ ឬ មិនមានទំនាក់ទំនងកើតឡើងរវាង u និង x។

$$E(u|x) = E(u) = 0$$

ក្នុងករណីលក្ខខណ្ឌនេះមិនត្រូវបានអនុវត្តន៍តាម នោះតម្លៃនៃប៉ារ៉ាម៉ែត្រ (β_1)នឹងមានកំហុសឆ្គង។

លក្ខខណ្ឌទី៥: វ៉ារីយង់ (variance)នៃ error *u* គឺ ថេរ (*Homoskedasticity*)។

$$Var(u|x) = \sigma^2$$

ករណីវ៉ារីយង់ (variance)នៃ error *u* គឺ ប្រែប្រួលថេរ (*heteroskedasticity*) វាធ្វើឲ្យការអនុវត្តន៍តេស្តមួយចំនួនពុំមានសុពលភាព (*t*-test *F*-test និង *LM*-test)។

លំហាត់

២.១. តារាងខាងក្រោមរួមមាន បរិមាណទំនិញដែលបានលក់ចេញ (Q) គិតជាឯកតា និង ថ្លៃលក់ទំនិញក្នុង១ឯកតា (P) គិតជាដុល្លារ។

ក្រុមហ៊ុន	Q	P
1	34	9
2	26	10
3	68	4
4	36	9
5	19	12
6	56	5
7	28	10
8	64	4
9	24	11
10	48	8

ក. ចូរប៉ាន់ស្មានទំនាក់ទនងរវាង Q និង P ជាមួយការប្រើប្រាស់OLSដោយទទួលបានintercept ($\hat{\beta}_0$) និង មេគុណប៉ាន់ស្មាន ($\hat{\beta}_1$)ក្នុងសមីការ:

$$\hat{Q} = \hat{\beta}_0 + \hat{\beta}_1 P.$$

បញ្ជាក់ទិសដៅនៃទំនាក់ទនង។ ករណីនេះ តើinterceptមានសារៈសំខាន់ដែរឬទេ? ចូរពន្យល់។ ប្រសិនថ្លៃទំនិញកើន២ដុល្លារ តើបរិមាណលក់ប្រែប្រួលប៉ុន្មាន?

ខ. ចូរគណនាតម្លៃ$\hat{Q}$ និង តម្លៃសំណល់ (residual) សំរាប់ក្រុមហ៊ុននីមួយៗដែលត្រូវបានអង្កេត។ បន្ទាប់មក ចូរផ្ទៀងផ្ទាត់ថា ផលបូកនៃតម្លៃសំណល់ប្រហែល ឬ ស្មើសូន្យ។

គ. នៅពេលថ្លៃទំនិញស្មើ១៤ដុល្លារ តើព្យាករណ៍ថា បរិមាណលក់ស្មើប៉ុន្មាន?

ឃ. ចំពោះក្រុមហ៊ុនទាំង១០ខាងលើ តើបំរែបំរួលនៅក្នុងបរិមាណលក់ប៉ុន្មាន ដែលត្រូវបានពន្យល់ដោយថ្លៃទំនិញ? ចូរបកស្រាយ។

២.២. ក្រុមហ៊ុនតូចមួយបានជួលអ្នកជំនាញម្នាក់ដើម្បីព្យាករណ៍បរិមាណការលក់នៃទំនិញ ក្នុងរយៈពេលមួយខែ ប្រសិនថ្លៃលក់កើនឡើងដល់១២ដុល្លារក្នុង១កតា។ អ្នកជំនាញបានកត់ត្រារាល់ទិន្នន័យថ្លៃទំនិញ និង បរិមាណលក់ក្នុងរយៈពេល១ឆ្នាំ។ អ្នកជំនាញបានកត់សំគាល់ថា រយៈពេល១ឆ្នាំមុន ថ្លៃទំនិញគិតជាមធ្យមគឺ ១០ដុល្លារ និង បរិមាណលក់ទំនិញដែលបានលក់ចេញគឺ ១,៤០០ឯកតាគិតជាមធ្យមក្នុងរយៈពេលមួយខែ។ ផ្អែកលើលទ្ធផលនៃម៉ូដែលរ៉ឺហ្គ្រេស្យុងសាមញ្ញមួយ ខ្ញុំព្យាករណ៍ថា ប្រសិនថ្លៃលក់កើនឡើងដល់១២ដុល្លារក្នុង១កតា នោះបរិមាណលក់ទំនិញដែលបានលក់ចេញគឺ ១,០០០ឯកតា"។

ក. ចូរបង្ហាញម៉ូដែលរ៉ឺគ្រ្ស្សុងសាមញ្ញដែលត្រូវបានប្រើប្រាស់ដោយអ្នកជំនាញសំរាប់ព្យាករណ៍បរិមាណការលក់នៃទំនិញ។

ខ. ចូរគូសក្រាហ្វិកនៃរ៉ឺគ្រ្ស្សុងលីនេអ៊ែរដែលទទួលបានពីសំណួរ(ក)។

២.៣. ឧបមាថា យើងសិក្សាទៅលើទំនាក់ទំនងរវាងថ្លៃសិក្សាសំរាប់និស្សិតម្នាក់ៗ (ACS) គិតជាខ្ទង់រយដុល្លារ និង ចំនួននិស្សិត (enroll) គិតជាខ្ទង់រយ ដែលបានចុះឈ្មោះចូលរៀននៅសាកលវិទ្យាល័យ ដោយទទួលបានម៉ូដែលរ៉ឺគ្រ្ស្សុងដូចខាងក្រោម:

$$\widehat{enroll} = 24.69 - 1.24\ ACS$$

ក. នៅពេលACSស្មើសូន្យ តើ*enroll*ដែលបានព្យាករណ៍ស្មើប៉ុន្មាន? ចុះបើ ACSស្មើ៤ តើ*enroll*ដែលបានព្យាករណ៍ស្មើប៉ុន្មាន? ចូរបកស្រាយទៅលើភាពខុសគ្នា។

ខ. តើរ៉ឺគ្រ្ស្សុងសាមញ្ញនេះ វាបញ្ជាក់ភាពចាំបាច់ដែលបង្ហាញនូវទំនាក់ទំនងរវាងថ្លៃសិក្សាសំរាប់និស្សិតម្នាក់ៗ និង ចំនួននិស្សិតដែលបានចុះឈ្មោះចូលរៀននៅសាកលវិទ្យាល័យដែរឬទេ? ចូរពន្យល់។

គ. បើចំនួននិស្សិតស្មើ៣,៤០០នាក់ តើថ្លៃសិក្សាស្មើប៉ុន្មាន? ចូរពិភាក្សា។

២.៤. សន្មតថា យើងទទួលបានទិន្នន័យថ្លៃជួលផ្ទះ(RP) និង ទំហំផ្ទះ(HS) នៅតំបន់មួយដូចខាងក្រោម:

ផ្ទះ	**RP (ដុល្លារ/ខែ)**	**HS (ម៉ែត្រការ៉េ)**
1	1,200	340
2	900	256
3	1,500	400
4	2,200	550
5	3,000	600
6	2,700	480
7	3,600	640
8	890	230
9	2,100	400
10	680	220
11	1,980	450
12	2,950	490

ក. ផ្អែកលើទិន្នន័យផ្ទះនីមួយៗខាងលើ ចូរកំណត់ចំណុចប្រសព្វរវាងថ្លៃជួលផ្ទះ និង ទំហំផ្ទះនៅក្នុងអ័ក្សឈរ និង ដេក។

ខ. ផ្អែកលើទិន្នន័យ ចូរកំណត់ម៉ូដែលរ៉ឺហ្គ្រេស្យុង (linear model): $RP = \beta_0 + \beta_1 HS + u$។ បន្ទាប់មក ចូរបកស្រាយលទ្ធផលដែលប៉ាន់ស្មាន និង គូសរ៉ឺហ្គ្រេស្យុងលីនេអ៊ែរ។

គ. ផ្អែកលើទិន្នន័យ ចូរកំណត់ម៉ូដែល (quadratic model): $RP = \beta_0 + \beta_1 HS^2 + u$។ ប្រសិនទំហំផ្ទះកើនឡើង១០ម៉ែត្រការ៉េ តើថ្លៃជួលផ្ទះកើនប៉ុន្មាន?

ឃ. ចូរគូសក្រាហ្វិចនៃlinear model និង quadratic model។ តើម៉ូដែលមួយណាដែលបន្សុីជាមួយទិន្នន័យបានល្អជាងគេ? ប្រៀបធៀបផលបូកការេនៃសំណល់ (sum of squared residuals: SSR)។ តើមួយណាតូចជាងគេ?

មេរៀនទី៣
វិភាគរ៉ឺគ្រ្ស្សុងពហុៈ ការប៉ាន់ស្មាន

ក្នុងមេរៀនទី២ គឺ វិភាគរ៉ឺគ្រ្ស្សុងសាមញ្ញ ដែលសិក្សាអំពីការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យតែមួយ (a single independent variable) ទៅលើអញ្ញត្តិមិនឯករាជ្យ (dependent variable)។ ការស្រាវជ្រាវ (an empirical work) ជាមួយការប្រើម៉ូដែលរ៉ឺគ្រ្ស្សុងសាមញ្ញ វាមានការលំបាកក្នុងការសន្និដ្ឋានអំពីការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យដែលមានតែមួយ (a single independent variable) ទៅលើអញ្ញត្តិមិនឯករាជ្យ ពីព្រោះជាញឹកញាប់ ការប្រើម៉ូដែលរ៉ឺគ្រ្ស្សុងសាមញ្ញ ច្រើនតែបរាជ័យក្នុងការអនុវត្តន៍លក្ខខណ្ឌទី៤នៃ OLS regression model។

ម៉ូដែលរ៉ឺគ្រ្ស្សុងពហុ គឺជារ៉ឺគ្រ្ស្សុងដែលមានអញ្ញត្តិឯករាជ្យច្រើនជាងមួយ។ គេអាចសិក្សាអំពីការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យទាំងឡាយនោះទៅលើអញ្ញត្តិមិនឯករាជ្យក្នុងពេលដំណាលគ្នាតែម្តង។ វាមានសារៈសំខាន់ក្នុងការធ្វើតេស្តទ្រឹស្តីហិរញ្ញវត្ថុ សេដ្ឋកិច្ច ក៏ដូចជាវាយតំលៃការជះឥទ្ធិពលនៃគោលនយោបាយ តាមរយៈការវិភាគទិន្នន័យ (nonexperimental data)។ យើងអាចធ្វើតេស្ត ថាតើមានទំនាក់ទំនងទៅវិញទៅមកដែលកើតឡើងរវាងអញ្ញត្តិឯករាជ្យទាំងនោះដែរឬទេ។ ដូច្នេះ រ៉ឺគ្រ្ស្សុងពហុ (multiple regression) អាចត្រូវបានគេប្រើប្រាស់ដើម្បីបង្កើតម៉ូដែលសំរាប់ព្យាករណ៍អញ្ញត្តិមិនឯករាជ្យបានកាន់តែប្រសើរ។

៣.១ រ៉ឺគ្រ្ស្សុងពហុ (multiple regression)

ឧទាហរណ៍ទី១

ឧបមាថា យើងចង់ធ្វើការសិក្សាការជះឥទ្ធិពលនៃកត្តាទាំងឡាយទៅលើការចំណាយពេលធ្វើការនៅផ្ទះរបស់ស្ត្រីដែលរៀបការហើយ ដោយអង្កេតស្ត្រីចំនួន២០នាក់ និងទទួលបានទិន្នន័យដូចខាងក្រោម:

HW	Room	Log(HW)	Children	education	Heduc
1435	2	3.156852	3	12	21
678	3	2.83123	2	16	22
986	5	2.993877	3	19	19
1680	6	3.225309	4	9	20
239	2	2.378398	2	26	26
1986	7	3.297979	5	8	17
1460	7	3.164353	3	11	13
1498	4	3.175512	4	10	17
257	2	2.409933	0	16	18
870	4	2.939519	2	17	21

370	3	2.568202	1	24	28
274	2	2.437751	0	25	27
598	4	2.776701	2	23	17
309	2	2.489958	2	19	18
1650	3	3.217484	3	9	24
1640	4	3.214844	4	10	25
1690	6	3.227887	4	10	20
1397	6	3.145196	3	12	28
1360	5	3.133539	3	14	26
986	3	2.993877	2	20	18

តារាង៣.១

ក្នុងនោះ HW (housewife)គឺជាចំនួនម៉ោងដែលធ្វើការនៅផ្ទះរបស់ស្ត្រីក្នុងរយៈពេល១ឆ្នាំ Roomsគឺជាចំនួនបន្ទប់នៅក្នុងផ្ទះ Childrenគឺជាចំនួនកូនក្រោមអាយុ១៦ឆ្នាំ Educationគឺជារយៈពេលនៃការសិក្សារបស់ស្ត្រីគិតជាឆ្នាំ និង Heduc (Husband's education)គឺជារយៈពេលនៃការសិក្សារបស់ប្តីនៃស្ត្រីនោះគិតជាឆ្នាំ។ ដូច្នេះ យើងអាចសរសេរបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងពហុដូចខាងក្រោម:

$$\log(HW) = \beta_0 + \beta_1 rooms + \beta_2 children + \beta_3 education + \beta_4 heduc + u$$

ផ្អែកលើទិន្នន័យខាងលើ យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\log(HW) = 2.811 + 0.032\ room + 0.102\ children - 0.024\ education + 0.005\ heduc$$

តើយើងបកស្រាយលទ្ធផលនៃសមីការខាងលើយ៉ាងដូចម្តេច? ការប៉ាន់ស្មានសំខាន់គឺមេគុណជម្រាល (slope coefficients: β_j)នៃអញ្ញត្តិចំនួនបន្ទប់ ចំនួនកូនក្រោមអាយុ១៦ឆ្នាំ រយៈពេលនៃការសិក្សារបស់ស្ត្រី និង រយៈពេលនៃការសិក្សារបស់ប្តីនៃស្ត្រីនោះ ដែលវាបង្ហាញអំពីទំហំនៃការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យទាំងនេះទៅលើការចំណាយពេលធ្វើការនៅផ្ទះរបស់ស្ត្រី។

៣.១.១ ម៉ូដែលមានអញ្ញត្តិឯករាជ្យពីរ

បើធៀបជាមួយរីហ្គ្រេស្យុងសាមញ្ញ តើរីហ្គ្រេស្យុងពហុអាចត្រូវបានប្រើប្រាស់ដើម្បីដោះស្រាយបញ្ហាអ្វីខ្លះ? ដើម្បីឆ្លើយសំណួរនេះ យើងនឹងលើកយកឧទាហរណ៍ដូចខាងក្រោម ដើម្បីបកស្រាយ។

ឧទាហរណ៍ទី២

ក្នុងមេរៀនទី២ យើងបានវិភាគកត្តាដែលជះឥទ្ធិពលទៅលើថ្លៃផ្ទះ ដោយលើកយកកត្តាតែមួយគត់ គឺ ទំហំផ្ទះ។ រីឯមេរៀននេះ យើងបន្ថែមកត្តាមួយទៀតគឺ ទំហំដីសង់ផ្ទះ (lotsize) គិតជាម៉ែត្រការេ។

យើងអាចសរសេរបានរ៉ឺហ្គ្រេស្យុងពហុដូចខាងក្រោម:

$$price = \beta_0 + \beta_1 sqrmt + \beta_2 lotsize + u \qquad (៣.១)$$

ដូច្នេះ ថ្លៃផ្ទះ (*price*) ត្រូវបានកំណត់ដោយអញ្ញត្តិឯករាជ្យពីរ គឺ ទំហំផ្ទះ (*sqrmt*) ទំហំដីសង់ផ្ទះ (*lotsize*) និង កត្តាដ៏ទៃទៀតដែលមិនត្រូវបានអង្កេត (*u*)។ មេគុណ (β_1)វាស់វាងទំហំការជះឥទ្ធិពលនៃទំហំផ្ទះទៅលើថ្លៃផ្ទះ។ រីឯ មេគុណ (β_2)វាស់វាងទំហំការជះឥទ្ធិពលនៃទំហំដីសង់ផ្ទះទៅលើថ្លៃផ្ទះ។

តាមគោលការណ៍អេកូណូមេទ្រីច (econometrics) នៅពេលបកស្រាយអំពីការជះឥទ្ធិពលនៃទំហំផ្ទះទៅលើថ្លៃផ្ទះ យើងត្រូវកំណត់ថាអញ្ញត្តិទំហំដីសង់ផ្ទះថេរ ឬ នៅពេលបកស្រាយអំពីការជះឥទ្ធិពលនៃទំហំដីសង់ផ្ទះទៅលើថ្លៃផ្ទះ យើងក៏ត្រូវកំណត់ថាអញ្ញត្តិទំហំផ្ទះថេរដែរ។

នៅក្នុងរ៉ឺក្រ្ទេស្យុង(៣.១) ការបញ្ចូលអញ្ញត្តិទំហំដីសង់ផ្ទះ យើងអាចដឹងអំពីការជះឥទ្ធិពលរបស់វាទៅលើថ្លៃផ្ទះ។ ម៉្យាងទៀត វាប្រហែលសំខាន់ការបញ្ចូលអញ្ញត្តិទំហំដីសង់ផ្ទះទៅក្នុងរ៉ឺក្រ្ទេស្យុង(៣.១) ពីព្រោះ ទំហំផ្ទះអាចមានទំនាក់ទំនងជាមួយទំហំដីសង់ផ្ទះ គឺថាទំហំដីសង់ផ្ទះកាន់តែធំ គេអាចមានលទ្ធភាពសង់ផ្ទះកាន់តែធំ។ ករណីរ៉ឺក្រ្ទេស្យុងសាមញ្ញ អញ្ញត្តិទំហំដីសង់ផ្ទះត្រូវបានចញ្ចូលទៅក្នុង *error term* (*u*) ដូច្នេះវាធ្វើឲ្យអញ្ញត្តិទំហំផ្ទះមានទំនាក់ទំនងជាមួយ*error term*។ ជាលទ្ធផល នាំឲ្យមេគុណ(β_1) មានកំហុសឆ្គងកើតឡើង ពីព្រោះ ផ្អែកលើលក្ខខណ្ឌទី៤នៃមេរៀនទី២ គឺ $E(u|sqrmt) = 0$។

ក្នុងឧទាហរណ៍(៣.១) ក្រៅពីអញ្ញត្តិទំហំផ្ទះ តើអញ្ញត្តិដែលអាចអង្កេតបាន (observable variables)ដទៃទៀតត្រូវបានបញ្ចូលទៅក្នុងម៉ូដែលរ៉ឺក្រ្ទេស្យុងយ៉ាងដូចម្តេច។ យើងអាចសរសេរម៉ូដែលរ៉ឺក្រ្ទេស្យុង ដែលមានអញ្ញត្តិឯករាជ្យចំនួនពីរដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

ដែល:

β_0: អ៊ីនធ័រសេប (intercept)

β_1: វាស់វែងបម្រែបម្រួល y ទាក់ទងនឹងx_1 ដោយកំណត់កត្តាដទៃទៀតថេរ

β_2: វាស់វែងបម្រែបម្រួល y ទាក់ទងនឹងx_2 ដោយកំណត់កត្តាដទៃទៀតថេរ

ម៉ូដែលរ៉ឺក្រ្ទេស្យុងពហុក៏មានសារៈសំខាន់ផងដែរចំពោះការបង្កើតfunctional relationshipsរវាងអញ្ញត្តិទាំងឡាយ។ សន្មតថា ថ្លៃផ្ទះ (*price*)ជាអនុគមន៍លើកជាការេ (quadratic function) នៃទំហំផ្ទះ (*sqrmt*):

$$price = \beta_0 + \beta_1 sqrmt + \beta_2 sqrmt^2 + u, \quad (៣.២)$$

ក្នុងនេះ u ជាកត្តាដទៃទៀតដែលជះឥទ្ធិពលទៅលើថ្លៃផ្ទះ។ ក្នុងម៉ូដែលនេះ ថ្លៃផ្ទះផ្តោតទៅលើកត្តាដែលត្រូវបានអង្កេត(observed factor)តែមួយគត់គឺទំហំផ្ទះ។ ដូច្នេះ វាហាក់ដូចជាម៉ូដែលរ៉ឺក្រ្ទេស្យុង

សាមញ្ញ (simple regression model)។ ប៉ុន្តែ វាពុំមែនជាម៉ូដែលរ៉ឺគ្រ្ស្សុងសាមញ្ញទេ ពីព្រោះ វាមានអញ្ញត្តិឯករាជ្យចំនួនពីរគឺ $sqrmt$ និង $sqrmt^2$។

ក្នុងសមីការ(៣.១) β_1វាស់វែងទំហំការជះឥទ្ធិពលនៃទំហំផ្ទះទៅលើថ្លៃផ្ទះ។ ចំណែកសមីការ (៣.២) ប៉ារ៉ាម៉ែត្រβ_1មិនវាស់វែងទំហំការជះឥទ្ធិពលនៃទំហំផ្ទះ(*sqrmt*)ទៅលើថ្លៃផ្ទះ(*price*)ឡើយ ក្នុងខណៈពេលកំណត់$sqrmt^2$ថេរ ពីព្រោះប្រសិនទំហំផ្ទះប្រែប្រួល បន្ទាប់មកទំហំផ្ទះលើកជាការេក៏ប្រែប្រួល។ ហេតុនេះ ពេលទំហំផ្ទះប្រែប្រួល វាធ្វើឲ្យថ្លៃផ្ទះ(*price*)ក៏ប្រែប្រួលប្រហែល:

$$\frac{\Delta price}{\Delta sqrmt} \approx \beta_1 + 2\beta_2 sqrmt$$

ម៉ូដែលជាមួយអញ្ញត្តិឯករាជ្យចំនួនពីរ លក្ខខណ្ឌសំខាន់(key assumption)អំពីទំនាក់ទំនងរវាង u ជាមួយx_1 និង x_2 គឺ

$$E(u|x_1, x_2) = 0 \quad (៣.៣)$$

ផ្អែកលើ(៣.៣) តើយើងអាចបកស្រាយឧទាហរណ៍ទី១យ៉ាងដូចម្តេច? ក្នុងសមីការ(៣.១) លក្ខខណ្ឌគឺ $E(u|sqrmt,\ lotsize) = 0$។ នេះបញ្ជាក់ថា កត្តាដទៃទៀត(u)ដែលជះឥទ្ធិពលទៅលើថ្លៃផ្ទះ វាមិនមានទំនាក់ទំនាក់ជាមួយទំហំផ្ទះ និង ទំហំដីសង់ផ្ទះឡើយ។

ចំពោះសមីការ(៣.២) នៅពេលអញ្ញត្តិទំហំផ្ទះ(*sqrmt*)ត្រូវបានគេដឹង ទំហំផ្ទះលើកជាការេ ($sqrmt^2$) ក៏ត្រូវបានគេដឹងផងដែរ។ ដូច្នេះ លក្ខខណ្ឌ $E(u|sqrmt,\ sqrmt^2) = 0$ និង $E(u|sqrmt) = 0$ គឺដូចតែគ្នា។ ប៉ុន្តែ លក្ខខណ្ឌ $E(u|sqrmt) = 0$ មានភាពច្បាស់លាស់ជាង។

៣.១.២ ម៉ូដែលជាមួយ"k"អញ្ញត្តិឯករាជ្យ (k independent variables)

ក្នុងការវិភាគរ៉ឺគ្រ្ស្សុងពហុមានអញ្ញត្តិឯករាជ្យមួយចំនួន(x_j)ជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ (y)។ ម៉ូដែលរ៉ឺគ្រ្ស្សុងពហុអាចសរសេរដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៣.៤)$$

ដែល: β_0 គឺជាអ៊ីនជ័រសិប (intercept)
β_1 គឺជាប៉ារ៉ាម៉ែត្រដែលទាក់ទងនឹងx_1
β_2 គឺជាប៉ារ៉ាម៉ែត្រដែលទាក់ទងនឹងx_2 និង បន្តបន្ទាប់

នៅពេលមាន"k"អញ្ញត្តិឯករាជ្យ (k independent variables) និង អ៊ីនជ័រសិប (intercept) នោះសមីការ(៣.៤)មាន"$k + 1$"ប៉ារ៉ាម៉ែត្រផុបផ្យូលេសិន (population parameters)។ យើងអាចហៅប៉ារ៉ាម៉ែត្រទាំងឡាយ (លើកលែង intercept) គឺ ប៉ារ៉ាម៉ែត្រជំរាល(slop parameters)។

អថេរ "u" គឺជា $error\ term$ ឬ $disturbance$។ វាជាកត្តាដទៃទៀត (ក្រៅពី $x_1, x_2, \dots, x_k$) ដែលជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ (y)។

នៅពេលម៉ូដែលរ៉ឺហ្គ្រេស្យុងពហុទូទៅត្រូវបានអនុវត្តន៍ យើងត្រូវតែដឹង ថាតើប៉ារ៉ាម៉ែត្រទាំងឡាយត្រូវបានបកស្រាយយ៉ាងដូចម្តេច។ ក្នុងឧទាហរណ៍ទី២ យើងបានវិភាគកត្តាដែលជះឥទ្ធិពលទៅលើថ្លៃផ្ទះ ដោយលើកយកកត្តាដូចជា ទំហំផ្ទះ (*sqrmt*) និង ទំហំដីសង់ផ្ទះ (lotsize) ដោយអញ្ញត្តិទាំងពីរគិតជាម៉ែត្រការ៉េ។ ។ ត្រង់ចំណុចនេះ យើងបន្ថែមអញ្ញត្តិឯករាជ្យចំនួនបន្ទប់គេង($bdrms$)។ ម៉ូដែលត្រូវបានសរសេរដូចខាងក្រោម:

$$price = \beta_0 + \beta_1 sqrmt + \beta_2 lotsize + \beta_3 bdrms + u \qquad (៣.៥)$$

នេះគឺជាម៉ូដែលរ៉ីហ្គ្រេស្យុងពហុដែលមានបីអញ្ញត្តិឯករាជ្យ $(k = 3)$ ក្នុងនោះ $x_1 = sqrmt$, $x_2 = lotsize,\ x_3 = bdrms$។

ម៉ូដែលរ៉ឺហ្គ្រេស្យុងពហុទូទៅលក្ខខណ្ឌសំខាន់គឺ:

$$E(u|x_1, x_2, \dots, x_k) = 0 \qquad ៣.៦)$$

លក្ខខណ្ឌ(៣.៦)មានន័យថា រាល់កត្តាទាំងឡាយដែលមិនត្រូវបានអង្កេត (unobserved error term, u) មិនមានទំនាក់ទំនងជាមួយអញ្ញត្តិឯករាជ្យ (x_j)ឡើយ។ ករណីលក្ខខណ្ឌនេះមិនត្រូវអនុត្តន៍តាម វានឹងនាំឲ្យម៉ូដែលរ៉ឺហ្គ្រេស្យុង (regression model ឬ ordinary least squares: OLS)មានកំហុសឆ្គងកើតឡើង។

៣.២ យន្តការ និង ការបកស្រាយ Ordinary Least Squares

តើសមីការOLSត្រូវបានបកស្រាយយ៉ាងដូចម្តេច?

ដំបូងយើងពិចារណាម៉ូដែលជាមួយអញ្ញត្តិឯករាជ្យពីរ ($k = 2$)។ សមីការOLSដែលត្រូវប៉ាន់ស្មានត្រូវបានសរសេរដូចខាងក្រោម:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2, \qquad (៣.៧)$$

- ✓ $\hat{\beta}_0$គឺជាការប៉ាន់ស្មាននៃβ_0
- ✓ $\hat{\beta}_1$គឺជាការប៉ាន់ស្មាននៃβ_1
- ✓ $\hat{\beta}_2$គឺជាការប៉ាន់ស្មាននៃβ_2

តើ$\hat{\beta}_0$ $\hat{\beta}_1$ និង $\hat{\beta}_2$ ត្រូវបានគណនាយ៉ាងដូចម្តេច?

$$\hat{\beta}_1 = \left(\sum_{i=1}^{n} \hat{r}_{i1} y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{r}_{i1}^2\right) \qquad (៣.៨)$$

ដែល $\hat{r}_{i1}$គឺជារីស៊ីឌ្យូល(OLS residuals)បានពីរ៉ឺហ្គ្រេស្យុងនៃx_1ជាអនុគមន៍x_2 ដូចខាងក្រោម:

$$\hat{x}_1 = \hat{\gamma}_0 + \hat{\gamma}_2 x_2$$

ដូច្នេះ យើងទទួលបាន $\hat{r}_{i1} = x_{i1} - \hat{x}_{i1}$

$$\hat{\beta}_2 = \left(\sum_{i=1}^{n} \hat{r}_{i2} y_i\right) \Big/ \left(\sum_{i=1}^{n} \hat{r}_{i2}^2\right) \qquad (\text{៣.៩})$$

ដែល $\hat{r}_{i2}$គឺជារីស៊ីឌ្យូល(OLS residuals)បានពីរ៉ឺហ្គ្រេស្យុងនៃx_2ជាអនុគមន៍x_1 ដូចខាងក្រោម:

$$\hat{x}_2 = \hat{\gamma}_0 + \hat{\gamma}_1 x_1$$

ដូច្នេះ យើងទទួលបាន $\hat{r}_{i2} = x_{i2} - \hat{x}_{i2}$

$$\text{រីឯ } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \qquad \text{៣.១០)}$$

$$\bar{y} = \left(\sum_{i=1}^{n} y_i\right)/n, \quad \bar{x}_1 = \left(\sum_{i=1}^{n} x_i\right)/n \quad \text{និង } \bar{x}_2 = \left(\sum_{i=1}^{n} x_2\right)/n$$

ក្នុងករណី ម៉ូដែលជាមួយអញ្ញត្តិឯករាជ្យពីរឡើងទៅ សមីការOLSដែលត្រូវប៉ាន់ស្មាន ត្រូវបានសរសេរដូចខាងក្រោម:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k, \qquad (\text{៣.១១})$$

ការគណនា$\hat{\beta}_1$តាមរូបមន្ត(៣.៨)ដែរ។ ប៉ុន្តែ $\hat{r}_{i1}$គឺជារីស៊ីឌ្យូល(OLS residuals)បានពីរ៉ឺហ្គ្រេស្យុងនៃx_1ជាអនុគមន៍x_2 x_3.... x_kដូចខាងក្រោម:

$$\hat{x}_1 = \hat{\gamma}_0 + \hat{\gamma}_2 x_2 + \hat{\gamma}_3 x_3 + \cdots + \hat{\gamma}_k x_k$$

គណនា$\hat{\beta}_2$តាមរូបមន្ត(៣.៩)ដែរ។ ប៉ុន្តែ $\hat{r}_{i2}$គឺជារីស៊ីឌ្យូល(OLS residuals)បានពីរ៉ឺហ្គ្រេស្យុងនៃx_2ជាអនុគមន៍x_1 x_3.... x_kដូចខាងក្រោម:

$$\hat{x}_2 = \hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_3 x_3 + \cdots + \hat{\gamma}_k x_k$$

វិធីសាស្ត្រគណនា$\hat{\beta}_j$ដទៃទៀត អនុវត្តន៍ដូចខាងលើដែរ។

ឧទាហរណ៍ទី៣

វិភាគកត្តាដែលជះឥទ្ធិពលទៅលើថ្លៃផ្ទះ(*price*) ដោយលើកយកកត្តាដូចជា ទំហំផ្ទះ (*sqrmt*) និង ទំហំដីសង់ផ្ទះ (lotsize) ដោយអញ្ញត្តិទាំងពីរគិតជាម៉ែត្រការេ និង ចំនួនបន្ទប់គេង(*bdrms*) ដោយសិក្សាទៅលើផ្ទះចំនួន៨០ យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{price} = -27.373 + 0.119 sqrmt + 0.002 lotsize + 18.390 bdrms \qquad (\text{៣.១២})$$

តើយើងបកស្រាយសមីការ(៣.១២)យ៉ាងដូចម្តេច? ការប៉ាន់ស្មានសំខាន់គឺមេគុណជម្រាល (slope coefficients)នៃអញ្ញត្តិទំហំផ្ទះ ទំហំដីសង់ផ្ទះ និង ចំនួនបន្ទប់គេង ដែលវាបង្ហាញអំពីទំហំនៃការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យទាំងនេះទៅលើថ្លៃផ្ទះ។

ករណីរក្សាទំហំដីសង់ផ្ទះ និង ចំនួនបន្ទប់គេងថេរ គេទទួលបាន $\Delta\widehat{price} = 0.119\Delta sqrmt$។ ឧបមាថា ថ្លៃផ្ទះគិតជាខ្នង់ពាន់ដុល្លារ រីឯទំហំផ្ទះគិតជាម៉ែត្រការេ។ គេអាចប៉ាន់ស្មានថា បើទំហំផ្ទះកើនឡើង១ម៉ែត្រការេ នោះថ្លៃផ្ទះរំពឹងកើនឡើង១១៩ដុល្លារ។ យើងក៏អាចធ្វើការបកស្រាយបែបប្រៀបធៀបផងដែរ។ ផ្ទះទី១មានទំហំធំជាងផ្ទះទី២ចំនួន១០ម៉ែត្រការេ នោះយើងអាចប៉ាន់ស្មានថា ផ្ទះទី១មានតម្លៃថ្លៃជាងផ្ទះទី២ប្រហែល១,១៩០ដុល្លារ (១០គុណនឹង១១៩)។

ករណីរក្សាទំហំផ្ទះ និង ចំនួនបន្ទប់គេងថេរ គេទទួលបាន $\Delta\widehat{price} = 0.002\Delta lotsize$។ ឧបមាថា ថ្លៃផ្ទះគិតជាខ្នង់ពាន់ដុល្លារ រីឯទំហំដីសង់ផ្ទះគិតជាម៉ែត្រការេ។ គេអាចប៉ាន់ស្មានថា បើទំហំដីសង់ផ្ទះកើនឡើង១ម៉ែត្រការេ នោះថ្លៃផ្ទះរំពឹងកើនឡើងតែ២ដុល្លារ។

ផ្អែកលើលទ្ធផលខាងលើ យើងឃើញថា មានភាពខុសគ្នាខ្លាំងទាក់ទងទៅនឹងទំហំនៃការជះឥទ្ធិពលរបស់អញ្ញត្តិទំហំផ្ទះ (មេគុណជម្រាលស្មើ0.១១៩) និង ទំហំដីសង់ផ្ទះ (មេគុណជម្រាលស្មើ ០.០០២)ទៅលើថ្លៃផ្ទះ។ សំណួរគឺថ តើលទ្ធផលខាងលើគួរឲ្យជឿជាក់ដែរឬទេ? បើយល់ថា លទ្ធផលខាងលើមិនគួរឲ្យជឿជាក់ ដូច្នេះវាគួរតែត្រូវបានគេបកស្រាយមូលហេតុឲ្យបានច្បាស់លាស់ជាការល្អ។

ជួនកាលនៅពេលដំណាលគ្នា វាមានអញ្ញត្តិឯករាជ្យ(independent variables)លើសពីមួយដែលប្រែប្រួល និង ជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ(dependent variable)។ ផ្អែកលើ(៣.១២) កំណត់ចំនួនបន្ទប់គេងថេរ ដូច្នេះយើងអាចសរសេរបានដូចខាងក្រោម:

$$\Delta\widehat{price} = 0.119\Delta sqrmt + 0.002\Delta lotsize$$

ការវាស់វែងអញ្ញត្តិដូចខាងលើ ប្រសិនទំហំផ្ទះកើនឡើង១ម៉ែត្រការេ និង ចំនួនបន្ទប់គេងកើនឡើង ១ នោះគេអាចប៉ាន់ស្មានថា ថ្លៃផ្ទះនឹងកើនឡើងចំនួន១,១២១ដុល្លារ (0.១១៩ + 0.00២)× ១,000។

៣.៣ តំលៃព្យាករណ៍ និង សំណល់ (OLS Fixed Values and Residuals)

ក្រោយពេលដំណើរការរ៉ឺគ្រ្ស្យុង(៣.៤) យើងទទួលបានរ៉ឺគ្រ្ស្យុងបន្ទាត់ត្រង់(OLS regression line) ដែលវាត្រូវបានគេប្រើប្រាស់ដើម្បីកំណត់តម្លៃប៉ាន់ស្មាន(fitted or predicted value)សំរាប់ការអង្កេត(observation)នីមួយៗ ដូចខាងក្រោម:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}, \qquad (៣.១៣)$$

ដែល "i" គឺជាចំនួនអង្កេត និង "k" គឺជាចំនួនអញ្ញត្តិឯករាជ្យ។

ជាទូទៅ តម្លៃជាក់ស្តែង(y_i) និង តម្លៃប៉ាន់ស្មាន$(\hat{y}_i)$ គឺមិនស្មើគ្នាឡើយសំរាប់ការអង្កេត (observation)ណាមួយ។ ភាពខុសគ្នាត្រូវបានគេហៅថាសំណល់(residuals) ឬ កំហុសឆ្គងនៃការព្យាករណ៍(prediction error) ដែលតម្លៃមធ្យមរបស់វាលើកជាការត្រូវតែអប្បបរិមា។

$$\hat{u}_i = y_i - \hat{y}_i \qquad (៣.១៤)$$

- ✓ ប្រសិន$\hat{u}_i > 0$ មានន័យថា ការព្យាករណ៍ខ្វះ
- ✓ ប្រសិន$\hat{u}_i < 0$ មានន័យថា ការព្យាករណ៍លើស

៣.៤ ការគណនាR^2 (Goodness-of-Fit)

ដូចក្នុងមេរៀនទី២ដែរ:

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$SSR \equiv \sum_{i=1}^{n} \hat{u}_i^2 \text{ និង } SST = SSE + SSR$$

ទម្រង់ពេញរេញនៃ SST: total sum of squares

SSE: explained sum of squares

SSR: residual sum of squares

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \qquad (៣.១៥)$$

គេអាចនិយាយថា R^2 គឺជាទំនាក់ទំនងមេគុណលើកជាការរវាងតម្លៃជាក់ស្តែង(y_i) និង តម្លៃប៉ាន់ស្មាន$(\hat{y}_i)$។ នោះគឺ

$$R^2 = \left(\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2 \Big/ \left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2\right) \qquad (៣.១៦)$$

យោងទៅលើ(៣.១២)នៃ**ឧទាហរណ៍ទី២** យើងអាចសរសេរជាថ្មីដូចខាងក្រោម:

$$\widehat{price} = -27.373 + 0.119sqrmt + 0.002lotsize + 18.390bdrms$$

$$n = 80,\ R^2 = 67.85\%$$

$R^2 = 67.85\%$ មានន័យថា បម្រែបម្រួលតម្លៃផ្ទះប្រហែល៦៧.៨៥%ត្រូវបានពន្យល់ដោយទំហំផ្ទះ (*sqrmt*) ទំហំដីសង់ផ្ទះ (lotsize) និង ចំនួនបន្ទប់គេង$(bdrms)$ ជាមួយគ្នា។ ចំណែក៣២.១៥% ត្រូវបានពន្យល់ដោយកត្តាដទៃទៀត(u)ដែលមិនត្រូវបានលើកយកមកសិក្សា។

៣.៥ លក្ខខណ្ឌនៃរ៉ឺហ្គ្រេស្យុងពហុ (Multiple regression model)

ដើម្បីចៀសវាងកំហុសឆ្គងកើតឡើង រ៉ឺហ្គ្រេស្យុងពហុត្រូវគោរពតាមលក្ខខណ្ឌដូចខាងក្រោម:

លក្ខខណ្ឌទី១: លីនេអ៊ែរ (Linear in Parameters)

ម៉ូដែលត្រូវបានសរសេរ:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៣.១៧)$$

$\beta_0,\ \beta_1,\ldots,\beta_k$ ជាប៉ារ៉ាម៉ែត្រដែលមិនទាន់ត្រូវបានគេដឹង (unknown parameters) និង u ជាលំអៀងដែលមិនត្រូវបានអង្កេត (unobservable random error)។ តាមលក្ខខណ្ឌទី១ (៣.១៧)ត្រូវតែជាអនុគមន៍លីនេអ៊ែរ។

លក្ខខណ្ឌទី២: ប៉ាន់គំរូចៃដន្យ (Random Sampling)

យើងមានប៉ាន់គំរូចៃដន្យដែលត្រូវបានអង្កេតចំនួន $n\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, 2, \ldots, n\}$, ដែល i គឺជាចំនួនអង្កេតពី១ដល់nរីឯ k គឺជាចំនួនអញ្ញត្តិឯករាជ្យ។ រ៉ឺហ្គ្រេស្យុងពហុអាចសរសេរដូចខាងក្រោម:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u, \qquad (៣.១៨)$$

លក្ខខណ្ឌទី៣: No Perfect Collinearity

គ្មានអញ្ញត្តិឯករាជ្យថេរ (constant)ឡើយជាមួយប៉ាន់គំរូ (sample) និង មិនមានទំនាក់ទំនងជាលីនេអ៊ែរពិតប្រាកដ(no exact linear relationships)ក្នុងចំណោមអញ្ញត្តិឯករាជ្យទាំងឡាយឡើយ។ ប្រសិនអញ្ញត្តិឯករាជ្យក្នុង(៣.១៧) មានទំនាក់ទំនងជាលីនេអ៊ែរពិតប្រាកដ(exact linear relationships) ជាមួយគ្នា នោះម៉ូដែលរងផលប៉ះពាល់ពី**perfect collinearity** និង ម៉ូដែល (ordinary least squares) មិនអាចយកទៅប្រើប្រាស់សំរាប់ការព្យាករណ៍បានឡើយ។

លក្ខខណ្ឌទី៤: Zero Conditional Mean

រំពឹងថាតម្លៃលំអៀង (error u)ស្មើសូន្យ ឬ មិនមានទំនាក់ទំនងកើតឡើងរវាងអញ្ញត្តិឯករាជ្យ និង u។

$$E(u|x_1, x_2, \ldots, x_k) = 0$$

ប្រសិនលក្ខខណ្ឌទី៤មិនត្រូវអនុវត្តន៍តាម នោះគេនិយាយថា ទម្រង់នៃម៉ូដែល(៣.១៧)កំណត់មិនបានត្រឹមត្រូវ(functional form misspecification) និង ធ្វើឲ្យការព្យាករណ៍មានកំហុសឆ្គងកើតឡើង។

លក្ខខណ្ឌទី៥: អូម៉ូស្កេតដាសស្ទីស៊ីធី (Homoskedasticity)

វ៉ារីយ៉ង់នៃលំអៀង (error u)ថេរ និង និយាយម៉្យាងទៀត:

$$Var(u|x_1, x_2, \ldots, x_k) = \sigma^2$$

វាមានន័យថា វ៉ារីយ៉ង់នៃលំអៀងដែលមិនត្រូវបានគេអង្កេត (unobserved error u)មិនពឹងផ្អែកទៅលើអញ្ញត្តិឯករាជ្យ $x_1, x_2, \dots, x_k$ឡើយ។

ករណីផ្ទុយពីលក្ខខណ្ឌទី៥នេះ ម៉ូដែលកើតឡើងនូវ heteroskedasticity នោះធ្វើឲ្យ *t*-test *F*-test និង *LM*-test មិនមានសុពលភាព។

លក្ខខណ្ឌទី១ដល់លក្ខខណ្ឌទី៥ គេហៅថា Gauss-Markov assumptions សំរាប់cross-sectional regression និង គេទទួលបាន:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \qquad (៣.១៩)$$

- ✓ $j = 1, 2, \dots., k$ (ចំនួនអញ្ញត្តិឯករាជ្យ)
- ✓ $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ គឺជាបម្រែបម្រួលប៉ាន់គំរូសរុប(total sample variation) ដែលទាក់ទងx_j និង R_j^2 គឺជាR^2 ដែលទទួលបានពីx_j ជាអនុគមន៍នៃអញ្ញត្តិឯករាជ្យដទៃទៀត។

ទំហំនៃ $Var(\hat{\beta}_j)$គឺមានសារៈសំខាន់ក្នុងការអនុវត្តន៍ នៅពេលវាកាន់តែធំ:

- ✓ នោះលទ្ធផលនៃការព្យាករណ៍កាន់តែលែងមានភាពជឿជាក់។
- ✓ ចន្លោះជឿជាក់ (confidence intervals) ក៏កាន់តែធំ។
- ✓ ភាពត្រឹមត្រូវនៃការធ្វើតេស្ត (hypotheses testing) កាន់តែថយចុះ។

៣.៦ ការប៉ាន់ស្មានវ៉ារីយ៉ង់ (σ^2)នៃម៉ូដែលOLS

ផ្អែកលើ(៣.១៨) យើងអាចទាញបាន:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}$$

ជាលទ្ធផល:

$$\hat{\sigma}^2 = \left(\sum_{i=1}^{n} \hat{u}_j^2\right) \Big/ (n - k - 1) = \frac{SSR}{n - k - 1} \qquad (៣.២០)$$

ផ្អែកលើ(៣.១៩) យើងអាចទាញបានគំលាតបទដ្ឋាន (standard deviation)នៃ$\hat{\beta}_j$ ជាឬសការេនៃវ៉ារីយ៉ង់ (σ^2) ដូចខាងក្រោម:

$$sd(\hat{\beta}_j) = \sigma / \left[SST_j(1 - R_j^2)\right]^{1/2}$$

ផ្អែកលើគំលាតបទដ្ឋាន យើងអាចទាញបានលំអៀងបទដ្ឋាន(standard error)ដូចខាងក្រោម:

$$se(\hat{\beta}_j) = \hat{\sigma} / \left[SST_j(1 - R_j^2)\right]^{1/2} \qquad (៣.២១)$$

សំណួរ

៣.១ ចំពោះម៉ូដែលរ៉ឺហ្គ្រេស្យុង ចូរឆ្លើយសំណួរដូចខាងក្រោម:

ក. តើវាត្រូវបានគេប្រើដើម្បីអ្វី?

ខ. តើintercept (β_0)មានអត្ថន័យយ៉ាងដូចម្តេច?

គ. តើមេគុណ ($\beta_1, \beta_1, \beta_2, .., \beta_k$) បញ្ជាក់អំពីអ្វី?

ឃ. តើR^2មានអត្ថន័យយ៉ាងដូចម្តេច?

៣.២ តើចំណុចណាក្នុងចំណោមខាងក្រោម ដែលបណ្តាលឲ្យOLS estimatorsមានកំហុសកើតឡើង?

ក. ក្នុងម៉ូដែលមានអញ្ញត្តិឯករាជ្យពីរដោមួយមេគុណទំនាក់ទនងរបស់វាស្មើ៩៨%

ខ. អញ្ញត្តិសំខាន់មួយត្រូវបានលុបចោល

គ. Heteroskedasticity

៣.៣ ក្នុងការសិក្សាទៅលើលទ្ធភាពប្រាក់ចំណេញ(profitability)របស់ក្រុមហ៊ុនមួយចំនួនដែលមានទំហំប្រហាក់ប្រហែលគ្នា យើងយល់ដឹងថា បម្រែបម្រួលប្រាក់ចំណេញទាក់ទងនឹងថ្លៃដើមទំនិញលក់ចេញ(CGS) ចំណាយការលក់(SE) ចំណាយរដ្ឋបាល(AE) និង ចំណាយទូទៅ(GE)។ សំរាប់ក្រុមហ៊ុននីមួយៗមានផលបូករាល់ចំណាយទាំងបួនប្រភេទគិតជាមធ្យមស្មើ៤៦៨,៩០០ ដុល្លារ។ យើងអាចសរសេរបានជាម៉ូដែលដូចខាងក្រោម:

$$Profitability = \beta_0 + \beta_1 CGS + \beta_2 SE + \beta_3 AE + \beta_4 GE + u,$$

ក. នៅក្នុងម៉ូដែល តើសមរម្យដែរឬទេ ខណៈពេលដែលចំណាយការលក់(SE)ប្រែប្រួល បែជាថ្លៃដើមទំនិញលក់ចេញ(CGS) ចំណាយរដ្ឋបាល(AE) និង ចំណាយទូទៅ(GE)ត្រូវបានកំណត់ថេរ?

ខ. ចូរពន្យល់ ហេតុអ្វីម៉ូដែលនេះមិនឆ្លើយតបលក្ខខណ្ឌទី៣គឺ perfect collinearity?

គ. ហេតុនេះ ដើម្បីឲ្យប៉ារ៉ាម៉ែត្រទាំងឡាយមានប្រយោជន៍ក្នុងការបកស្រាយ និង ឆ្លើយតបលក្ខខណ្ឌទី៣ តើត្រូវកែសម្រួលម៉ូដែលឡើងវិញយ៉ាងដូចម្តេច?

លំហាត់

៣.១ ក្រុមហ៊ុនលក់ឧបករណ៍អេឡិចត្រូនិចមួយទទួលការបញ្ជាទិញពីអតិថិជនតាមរយៈអ៊ីម៉ែល។ បច្ចុប្បន្ននេះ អ្នកគ្រប់គ្រងកំពុងពិនិត្យមើលដំណើការដឹកជញ្ជូនពីឃ្លាំងទៅកាន់អតិថិជន និងចង់សិក្សាកត្តាទាំងឡាយដែលជះឥទ្ធិពលទៅលើថ្លៃដើមជញ្ជូន (DC)។ ទិន្នន័យដែលប្រមូលបានរយៈពេល១៤ខែបញ្ជាក់អំពីថ្លៃដើមជញ្ជូន (DC)គិតជាខ្នង់ពាន់ដុល្លារ ចំណូលពីការលក់(sales)ខ្នង់ពាន់ដុល្លារ និង ចំនួនការបញ្ជាទិញ (order)គិតជាចំនួនដង។ ហេតុនេះ យើងសរសេរបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$DC = \beta_0 + \beta_1 sales + \beta_2 order + u$$

DC	Sales	Order
42.38	256	2890
62.24	412	3689
78.45	512	5190
64.68	434	4012
74.23	478	4369
34.56	189	1790
52.89	301	3189
67.45	486	4924
84.34	690	5310
72.56	502	4865
70.84	524	4623
53.89	359	2980
59.54	387	4012
82.45	678	5126

ក. ផ្អែកលើទិន្នន័យ ចូរបង្កើតម៉ូដែលរ៉ឺហ្គ្រេស្យុង

ខ. បកស្រាយអត្ថន័យនៃមេគុណβ_1 និង β_2។

គ. ចូរព្យាករណ៍ថ្លៃដើមជញ្ជូនប្រចាំខែ ប្រសិនចំណូលពីការលក់ស្មើ៣៥,០០០ដុល្លារ និង ចំនួនការបញ្ជាទិញស្មើ៣,២០០។

៣.២ យោងលើអត្ថបទស្រាវជ្រាវផ្សេងៗបង្ហាញថា កត្តាជះឥទ្ធិពលទៅលើលទ្ធភាពនៃប្រាក់ចំណេញ (ROA)មានដូចជា ទំហំក្រុមហ៊ុន (logsales) ទុនបង្វិល (working capital: WC) ប្រសិទ្ធភាពនៃការគ្រប់គ្រងទ្រព្យសម្បត្តិសរុប (total asset turnover) លទ្ធភាពទូទាត់បំណុលរយៈពេលខ្លី (quick ratio) និង កំរិតនៃការប្រើប្រាស់បំណុល (debt ratio)។ ដូច្នេះ យើងអាចបង្កើតបានម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 log(sales) + \beta_2 WC + \beta_3 TAT + \beta_4 QR + \beta_5 DR + u,$$

ដែល ROA log(sales) និង DRគិតជាភាគរយ WCគិតជាដុល្លារ រីឯ TAT និង QR គិតជាចំនួនដង។

ក. ហេតុអ្វីយើងរំពឹងថា$\beta_1 > 0$?

ខ. ចំពោះប៉ារ៉ាម៉ែត្រដទៃទៀត តើយើងរំពឹងថា វាមានសញ្ញាបែបណា? ចូរពន្យល់។

៣.៣ ផ្អែកលើទិន្នន័យនៅតារាង៣.១ និង លទ្ធផលរបស់វា ចូរធ្វើការបកស្រាយសំណួរខាងក្រោម:

ក. បកស្រាយអត្ថន័យនៃមេគុណβ_j។

ខ. ប្រសិន រយៈពេលនៃការសិក្សារបស់ស្ត្រីកើន២ឆ្នាំ តើការចំណាយពេលធ្វើការនៅផ្ទះរបស់ស្ត្រីនោះថយចុះប៉ុន្មាន?

គ. តើមានអញ្ញត្តិឯករាជ្យដទៃទៀតដែរឬទែ ដែលរំពឹងថាជះឥទ្ធិពលទៅលើការចំណាយពេលធ្វើការនៅផ្ទះរបស់ស្ត្រី ព្រមទាំងមានទំនាកទំនងជាមួយអញ្ញត្តិឯករាជ្យដែលមានស្រាប់នៅក្នុងសមីការ? ចូរពន្យល់។

មេរៀនទី៤
ទិសាគរីព្រ្ហេស្យុងពហុៈ ការធ្វើតេស្ត

មេរៀននេះបរិយាយអំពីការធ្វើតេស្តសម្មតិកម្ម(hypotheses testing)ប៉ារ៉ាម៉ែត្រនីមួយៗ ឬ ច្រើនជាងមួយនៃម៉ូដែលរ៉ីព្រ្ហេស្យុង។ តើក្រុមនៃអញ្ញត្តិឯករាជ្យអាចលុបចេញពីម៉ូដែលបានដែរឬទេ។

៤.១ បំណែងចែកប៉ាន់គំរូនៃOLS Estimators

តាំងពីចាប់ផ្តើមមេរៀនទី១រហូតដល់ចំណុចនេះ យើងបានដឹងអំពីលក្ខខណ្ឌទាំងឡាយនៃម៉ូដែលរ៉ីព្រ្ហេស្យុង។ បន្ថែមពីលើនេះ ដើម្បីអនុវត្តន៍តេស្តស្ថិតិ (statistical inference) យើងចាំបាច់ត្រូវយល់អំពីបំណែងចែកប៉ាន់គំរូ (sampling distribution)នៃ$\hat{\beta}_j$។

នៅពេលយើងចង់កែសំរួលតម្លៃនៃអញ្ញត្តិឯករាជ្យទាក់ទងនឹងប៉ាន់គំរូឲ្យកាន់តែប្រសើរឡើង វាចាំបាច់ត្រូវយល់ដឹងឲ្យច្បាស់ថា បំណែងចែកប៉ាន់គំរូនៃOLS Estimatorsផ្អែកទៅលើបំណែងចែកនៃលំអៀង (error distribution)។ ដើម្បីដំណើរការបំណែងចែកប៉ាន់គំរូ (sampling distribution)នៃ$\hat{\beta}_j$ឲ្យកាន់តែមានភាពងាយស្រួល យើងសន្មតថា នៅក្នុងផុបផ្សំលែសិន(population) លំអៀងដែលមិនត្រូវបានអង្កេត(unobserved error) គឺមានបំណែងចែកធម្មតា(normal distribution)។ នេះជាលក្ខខ័ណ្ឌណកម៉ាល់លីធី(**normality assumption**)នៃលំអៀង(u)។

លក្ខខណ្ឌទី៦: ណរម៉ាល់លីធី (Normality)

លំអៀងផុបផ្សំលែសិន (population error u) គឺមិនមានទំនាក់ទំនងជាមួយបណ្តាអញ្ញត្តិឯករាជ្យ $x_1, x_2, \dots, x_k$ និង មានលក្ខខណ្ឌណរម៉ាល់លីធីជាមួយតម្លៃមធ្យមស្មើសូន្យ (zero mean) និង វ៉ារីយ៉ង់ថេរ $(variance\ \sigma^2)$: $\boldsymbol{u \sim Normal(0, \sigma^2)}$។

ចំពោះការអនុវត្តន៍ម៉ូដែលរ៉ីព្រ្ហេស្យុងជាមួយទិន្នន័យបែបcross-sectional លក្ខខណ្ឌទី១ដល់លក្ខខណ្ឌទី៦ត្រូវបានគេហៅថា **classical linear model (CLM) assumptions**។ គេអាចសរសេរសង្ខេបដូចខាងក្រោម:

$$y|x \sim Normal(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \sigma^2),$$

៤.២ ការធ្វើតេស្តសម្មតិកម្មនៃប៉ារ៉ាម៉ែត្រនីមួយៗ: *t*-test

ម៉ូដែលរ៉ីព្រ្ហេស្យុងអាចសរសេរដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៤.១)$$

យើងសន្មតថា ម៉ូដែលនេះបានឆ្លើយតបនឹងCLM assumptions។ Ordinary Least Squares (OLS) កំណត់បាននូវតម្លៃប៉ាន់ស្មាន (β_j)ប្រកបដោយភាពសមហេតុផល(unbiased estimators) ។ ក្នុងផ្នែកនេះ យើងសិក្សា តើការធ្វើតេស្តសម្មតិកម្មនៃប៉ារ៉ាម៉ែត្រ(β_j)ត្រូវបានអនុវត្តន៍យ៉ាងដូចម្តេច។ ដំបូង យើងបង្កើតសម្មតិកម្មអំពីតម្លៃប៉ាន់ស្មាន(β_j) និងបន្ទាប់មកប្រើប្រាស់វិធីសាស្ត្រស្ថិតិ (statistical inference) ដើម្បីធ្វើតេស្តសម្មតិកម្ម។

នៅក្រោមលក្ខខណ្ឌទី១ដល់លក្ខខណ្ឌទី៦ (CLM assumptions) យើងបង្កើតបានវិធីសាស្ត្រសំរាប់តេស្តសម្មតិកម្មដូចខាងក្រោម:

$$(\hat{\beta}_j - \beta_j)/se(\hat{\beta}_j) \sim t_{n-k-1}, \qquad (៤.២)$$

ដែល $n-k-1$ គឺជាdegree of freedom (df) និង $standard\ error\ (se)$គឺជាលំអៀងស្តង់ដា។

(៤.២)មានសារៈសំខាន់សំរាប់ធ្វើតេស្តសម្មតិកម្មទាក់ទងនឹងβ_j។ ក្នុងការអនុវត្តន៍ យើងចាប់អារម្មណ៍ធ្វើតេស្តសម្មតិកម្មដើម(null hypothesis):

$$H_0: \beta_j = 0, \qquad (៤.៣)$$

ដែល j គឺជាអញ្ញត្តិឯករាជ្យ k ណាមួយ។ រីឯβ_jគឺវាស់វែងការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យ(x_j) ណាមួយ ទៅលើអញ្ញត្តិមិនឯករាជ្យ(y) ដោយកំណត់អញ្ញត្តិឯករាជ្យដទៃទៀតថេរ។

ឧទាហរណ៍ទី១ យើងពិចារណាសមីការទាក់ទងនឹងការចំណាយពេលក្នុងការគេងដែលត្រូវបានសិក្សាដោយលោកបិតដល និង ហាម៉ើម៉ែស នៅឆ្នាំ១៩៩០ ទៅលើមនុស្សចំនួន៧០៦នាក់។ អ្នកស្រាវជ្រាវទាំងពីរ គាត់ចង់ដឹងអំពីទំនាក់ទំនងរវាងការចំណាយពេលគេង $(sleep)$ និង ចំណាយពេលក្នុងការងារ$(total\ work)$ តាងដោយ$totwrk$។ គាត់ក៏បានបញ្ចូលកត្តាដទៃទៀតដែលជះឥទ្ធិពលទៅលើការចំណាយពេលគេងចូលទៅក្នុងសមីការផងដែរ។

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u, \qquad (៤.៤)$$

ក្នុងនោះអញ្ញត្តិ $sleep$ និង $totwrk$ គិតជាចំនួននាទីក្នុង១សប្តាហ៍ រីឯ ការចំណាយពេលសិក្សា $(educ)$ និង អាយុ (age)គិតជាឆ្នាំ។

សម្មតិកម្មដើមគឺ $H_0: \beta_1 = 0$ មានន័យថា ចំណាយពេលក្នុងការងារ$(totwrk)$ មិនជះឥទ្ធិពលទៅលើការចំណាយពេលគេងឡើយ ខណៈពេលអញ្ញត្តិដទៃទៀតដូចជា ការចំណាយពេលសិក្សា $(educ)$ និង អាយុ (age)ថេរ។ ក្រោយពេលធ្វើតេស្ត បើលទ្ធផលរកឃើញថា សម្មតិកម្មដើមពិត នោះបញ្ជាក់ថា ចំណាយពេលក្នុងការងារ$(totwrk)$ គឺពិតជាមិនជះឥទ្ធិពលទៅលើការចំណាយពេលគេងឡើយ។

វិធីសាស្ត្រស្ថិតិសំរាប់ធ្វើតេស្ត(៤.៣)ត្រូវបានគេហៅថា "$\boldsymbol{t\ statistic\ of\ \hat{\beta}_j}$" និងត្រូវបានកំណត់ដូចខាងក្រោម:

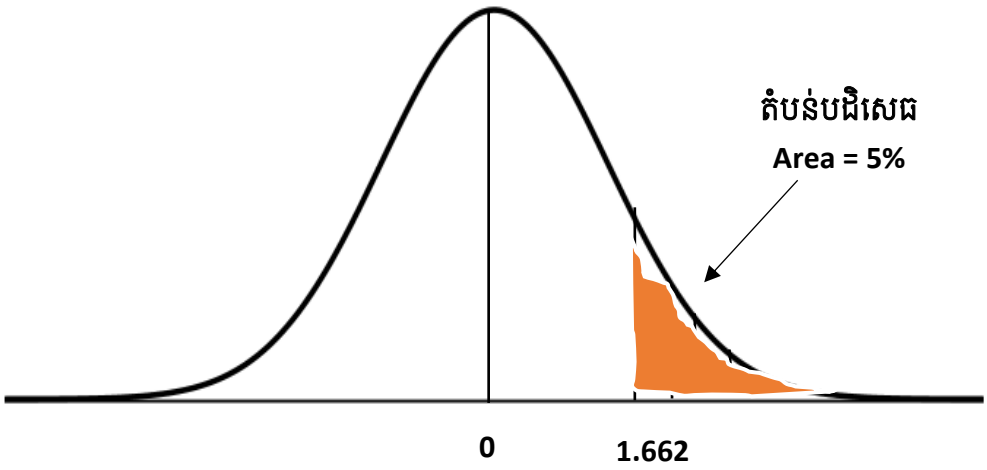
$$t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}, \qquad (៤.៥)$$

៤.២.១ ការធ្វើតេស្តសម្មតិកម្មខាងស្តាំ

ដើម្បីកំណត់គោលការណ៍បដិសេធសម្មតិកម្មដើម(H_0) យើងចាំបាច់ធ្វើការសំរេចចិត្តទៅលើសម្មតិកម្មផ្ទុយ(alternative hypothesis: H_1)។ ជាដំបូង ពិចារណាទំរង់ធ្វើតេស្តសម្មតិកម្មខាងស្តាំដូចខាងក្រោម:

$$H_1: \beta_j > 0, \qquad (៤.៥)$$

ក្នុងការយល់ស្រប ឬ បដិសេធសម្មតិកម្ម (H_0ឬ H_1) តើយើងគួរតែជ្រើសរើសគោលការណ៍យ៉ាងដូចម្តេច? ដំបូងយើងត្រូវតែកំណត់កំរិតសារៈសំខាន់ជាស្ថិតិ ($statistical\ significance\ level$) នៅឯ១% ៥% ឬ ១០% និង និយាយម្យ៉ាងទៀតគឺ កំណត់ប្រូបាបបីលីតេនៃការយល់ស្រប ឬ បដិសេធសម្មតិកម្ម។ ជាទូទៅ គេកំណត់យកកំរិត៥%។



រូបភាព៤.១: គោលការណ៍បដិសេធនៅឯ៥%សំរាប់សម្មតិកម្មផ្ទុយ ($H_1: \beta_j > 0$)ជាមួយ៨៨df

ចំពោះការធ្វើតេស្តសម្មតិកម្មខាងស្តាំ តំលៃប៉ាន់ស្មាន (expected value)នៃ$t_{\hat{\beta}_j}$គឺ វិជ្ជមាន ($\beta_j > 0$)។ ដូច្នេះ ដើម្បីបដិសេធសម្មតិកម្មដើម ($H_0: \beta_j = 0$) និងយល់ស្របសម្មតិកម្មផ្ទុយ ($H_1: \beta_j > 0$) យើងត្រូវស្វែងរកតំលៃវិជ្ជមាន និងជាចំនួនធំគ្រប់គ្រាន់ (sufficiently large)នៃ$t_{\hat{\beta}_j}$ មានន័យថា:

$$t_{\hat{\beta}_j} > t_{n-k-1} = c \qquad (៤.៦)$$

ឧបមាថា $degree\ of\ freedom(df) = n - k - 1 = 88$ និងកំរិតសារៈសំខាន់ជាស្ថិតិនៅឯ៥%។ ដូច្នេះ យើងទទួលបានតំលៃ $C = 1.662$។ ប្រសិន $t_{\hat{\beta}_j} < 1.662$ យើងយល់ស្របសម្មតិកម្មដើម ($H_0: \beta_j = 0$)។ បើ $t_{\hat{\beta}_j} > 1.662$ យើងបដិសេធសម្មតិកម្មដើម និង យល់ស្របសម្មតិកម្មផ្ទុយ ($H_1: \beta_j > 0$)។

ឧទាហរណ៍ទី២ នៅពេលប៉ាន់ស្មានសមីការប្រាក់ឈ្នួល (wage equations) យើងរំពឹងថា កម្មកររក្មេងៗដែលមិនមានបទពិសោធន៍ការងារនឹងទទួលបានប្រាក់ឈ្នួលទាប រីឯកម្មកដែលមានបទពិសោធន៍ការងារ(experience)កាន់តែច្រើន នឹងទទួលបានប្រាក់ឈ្នួលកាន់តែខ្ពស់។ យើងបញ្ចូលការសិក្សា (year of education) ទៅក្នុងសមីការផងដែរៈ

$$Wage = \beta_0 + \beta_1 exper + \beta_2 educ + u, \qquad (៤.៧)$$

ដោយអង្កេតទៅលើកម្មករចំនួន១,០០០នាក់ យើងទទួលបានលទ្ធផលដូចខាងក្រោមៈ

$$\widehat{wage} = -8.140 + 0.133\ exper + 2.322\ educ$$
$$(1.802)\quad (0.027)\qquad (0.141)$$
$$n = 1{,}000,\ R^2 = 0.217,$$

តួលេខដែលនៅខាងក្រោមមេគុណដែលប៉ាន់ស្មាន (estimated coefficients) គឺជាលំអៀងស្តង់ដា (standard error)។ សមីការនេះអាចត្រូវបានប្រើប្រាស់ដើម្បីធ្វើតេស្ត ថាតើយល់ស្រប ឬ បដិសេធសម្មតិកម្ម (H_0ឬ H_1) សំរាប់អញ្ញត្តិបទពិសោធន៍ការងារ(exper) ដោយអញ្ញត្តិការសិក្សា (educ) ថេរ។ យើងអាចសរសេរបាន $H_0: \beta_{exper} = 0$ រីឯ $H_1: \beta_{exper} > 0$។

ចំនួនអង្កេត(n) និង អញ្ញត្តិឯករាជ្យ(k_j) ស្មើ1,000 និង 2 រៀងគ្នា។ ដូច្នេះ degree of freedom (df = n – k – 1) ស្មើនឹង 997(df = 1,000 - 2 - 1)។ ហេតុនេះ គេទាញបានតំលៃ១.៦៤៥ និង ២.៣២៦ (critical value ឬ C) នៅឯកំរិត៥% និង ១% រៀងគ្នា។ រីឯ តំលៃ $\boldsymbol{t}\ statistic$ សំរាប់ $\hat{\beta}_{exper}$គឺ

$$t_{exper} = 0.133/0.027 \approx 4.93 > c = 2.326$$

ដូច្នេះ $\hat{\beta}_{exper}$ ឬ exper មានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត១%។ បន្ទាប់មកទាញបានថា បទពិសោធន៍ការងារ(exper)ជះឥទ្ធិពលជាវិជ្ជមានទៅលើប្រាក់ឈ្នួល(wage)។ បើកម្មកកាន់តែមានបទពិសោធន៍ការងារ នោះគេរំពឹងថា ប្រាក់ឈ្នួលដែលកម្មកទទួលបានក៏កាន់តែកើនឡើង។

៤.២.២ ការធ្វើតេស្តសម្មតិកម្មខាងឆ្វេង

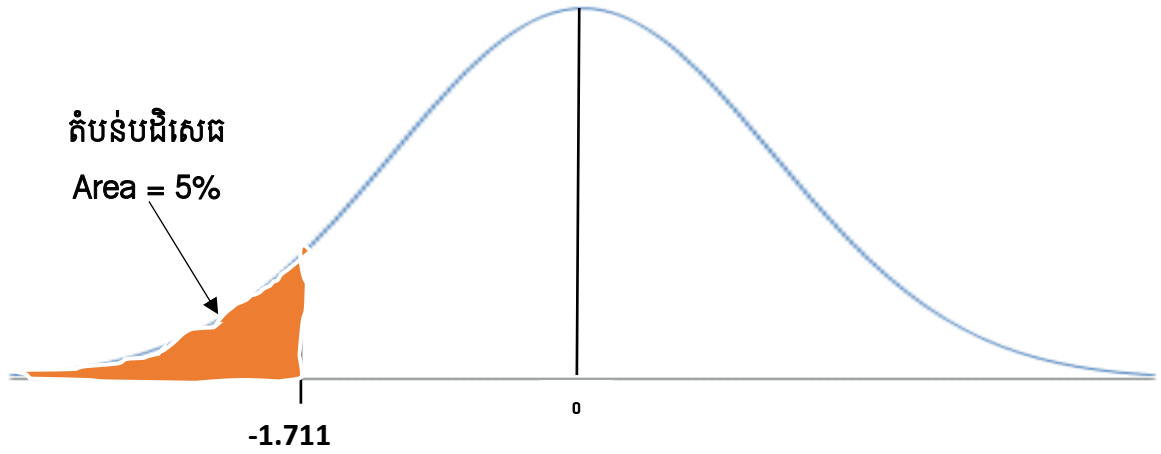
ករណីធ្វើតេស្តសម្មតិកម្មខាងឆ្វេង សម្មតិកម្មផ្ទុយ(alternative hypothesis: H_1)គឺ ប៉ារ៉ាម៉ែត្រតូចជាងសូន្យ។

$$H_1: \beta_j < 0, \qquad (៤.៨)$$

សម្មតិកម្មផ្ទុយ(H_1)ត្រូវបានយល់ស្របនៅពេល:

$$t_{\hat{\beta}_j} < -c, \qquad (៤.៩)$$

ប្រសិនកំរិតសារៈសំខាន់ស្ថិតិគឺ ៥% និង $df = 24$ នោះ $c = 1.711$។ ដូច្នេះ $H_0: \beta_j = 0$ ត្រូវបានបដិសេធ ដោយទទួលយក$H_1: \beta_j < 0$ នៅឯកំរិត៥% ប្រសិន$t_{\hat{\beta}_j} < -1.711$។



រូបភាព៤.២: គោលការណ៍បដិសេធនៅឯ៥%សំរាប់សម្មតិកម្មផ្ទុយ ($H_1: \beta_i < 0$)ជាមួយ២៤df

ឧទាហរណ៍ទី៣ យោងពី ***ឧទាហរណ៍ទី១*** យើងអាចប្រើប្រាស់ទិន្នន័យដើម្បីធ្វើតេស្តសម្មតិកម្មដើមដែលថា *ចំណាយពេលក្នុងការងារ(totwrk) មិនជះឥទ្ធិពលទៅលើការចំណាយពេលគេង(sleep)ឡើយ* និង សម្មតិកម្មផ្ទុយ *ចំណាយពេលក្នុងការងារ(totwrk) ជះឥទ្ធិពលអវិជ្ជមានទៅលើការចំណាយពេលគេង(sleep)*។ យើងអាចសរសេរបាន $H_0: \beta_{totwrk} = 0$ រីឯ $H_1: \beta_{totwrk} < 0$។

សមីការដែលត្រូវបានប៉ាន់ស្មាន និងលំអៀងស្តង់ដា(*standard error*)ត្រូវបានបង្ហាញដូចខាងក្រោម:

$$\widehat{sleep} = 3638.245 - 0.148\ totwrk - 11.134\ educ + 2.20\ age$$
$$(112.275)\ (0.0167) \qquad (5.884) \qquad (1.446)$$
$$n = 706,\ R^2 = 0.1134$$

ចំនួនអង្កេត(n) និង អញ្ញត្តិឯករាជ្យ(k_j) ស្មើ 706 និង 3 រៀងគ្នា។ ដូច្នេះ កំរិតសេរីភាព (*degree of freedom*: $df = n - k - 1$) ស្មើនឹង 702(df = 706 - 3 - 1)។ ហេតុនេះ គេទាញបានតំលៃ1.645 (critical value ឬ C) នៅឯកំរិត5%។ រីឯ តំលៃ ***t*** *statistic* សំរាប់ $\hat{\beta}_{totwrk}$គឺ

$$t_{exper} = -0.148/0.0167 \approx -8.86 < -c = -1.645$$

ដូច្នេះ $\hat{\beta}_{totwrk}$ ឬ totwrk មានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត៥%។ បន្ទាប់មកទាញបានថា *ចំណាយពេលក្នុងការងារ*$(totwrk)$ *ជះឥទ្ធិពលអវិជ្ជមានទៅលើការចំណាយពេលគេង*$(sleep)$។ ដូច្នេះ គេអាចព្យាករណ៍ថា បើម្នាក់ៗចំណាយពេលធ្វើការកាន់តែច្រើន នោះគាត់មានលទ្ធភាពគេងកាន់តែតិច។

បើអញ្ញត្តិការចំណាយពេលសិក្សា $(educ)$ និង អាយុ (age)ថេរ នោះនាំឲ្យ:

$$\Delta\widehat{sleep} = -0.148\,\Delta totwrk$$

ករណីចំណាយពេលក្នុងការងារមួយសប្តាហ៍កើនឡើង៣ម៉ោង (ឬ 180នាទី) នោះការគេងថយចុះប្រហែល២៧នាទីក្នុងមួយសប្តាហ៍។

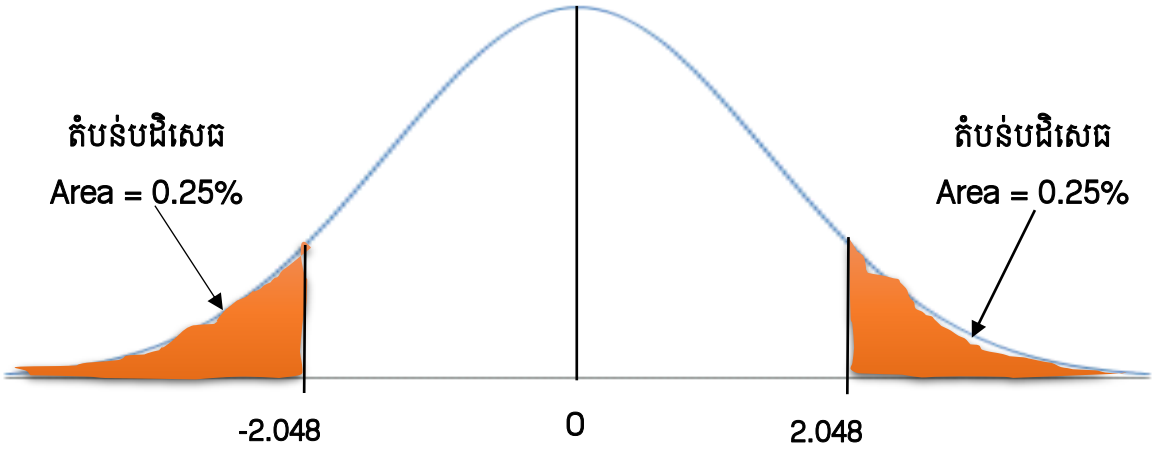
៤.២.៣ ការធ្វើតេស្តសម្មតិកម្មទាំងសងខាង

ក្នុងការអនុគ្គន៍ ជារួមសម្មតិកម្មដើម(null hypothesis)គឺ $H_0: \beta_j = 0$។ ចំពោះការធ្វើតេស្តសម្មតិកម្មទាំងសងខាង សម្មតិកម្មផ្ទុយ(alternative hypothesis: H_1)គឺ:

$$H_1: \beta_j \neq 0, \qquad (៤.១០)$$

សម្មតិកម្មផ្ទុយ(H_1)ត្រូវបានយល់ស្របនៅពេល:

$$\left|t_{\hat{\beta}_j}\right| > c, \qquad (៤.១១)$$



រូបភាព៤.៣: គោលការណ៍បដិសេធនៅឯ5%សំរាប់សម្មតិកម្មផ្ទុយ ($H_1: \beta_i \neq 0$)ជាមួយ28df

កត់សំគាល់: ករណីមានអញ្ញត្តិx_jណាមួយជះឥទ្ធិពលទៅលើអញ្ញត្តិy ដោយមិនប្រាកដថា តើការជះឥទ្ធិពលនោះវិជ្ជមាន ឬ អវិជ្ជមានឡើយ ពីព្រោះដោយសារមិនមានទ្រឹស្តីដែលបញ្ជាក់អំពីសញ្ញានៃβ_j។ ហេតុនេះត្រូវជ្រើសរើស*ការធ្វើតេស្តសម្មតិកម្មទាំងសងខាង*។

ប្រសិនកំរិតសារៈសំខាន់ស្ថិតិគឺ 5% និង $df = 28$ នោះ $c = 2.048$។ ដូច្នេះ $H_0: \beta_j = 0$ត្រូវបានបដិសេធ ដោយទទួលយក$H_1: \beta_j \neq 0$ នៅឯកំរិត៥% ប្រសិន$\left|t_{\hat{\beta}_j}\right| > 2.048$។

ឧទាហរណ៍ទី៤ យោងពី *ឧទាហរណ៍ទី៣*

ករណីមានអញ្ញត្តិអាយុ(age) មិនប្រាកដថា តើការជះឥទ្ធិពលនោះវិជ្ជមាន ឬ អវិជ្ជមាន*ទៅលើការចំណាយពេលគេង*$(sleep)$ឡើយ ពីព្រោះដោយសារមិនមានទ្រឹស្តី ឬ របកគំហើញនៃការស្រាវជ្រាវពីមុនដែលបញ្ជាក់អំពីសញ្ញានៃβ_jរបស់អាយុឲ្យបានច្បាស់លាស់។ ហេតុនេះត្រូវជ្រើសរើស*ការធ្វើតេស្តសម្មតិកម្មទាំងសងខាង*។ យើងអាចសរសេរបាន $H_0: \beta_{age} = 0$ រីឯ $H_1: \beta_{age} \neq 0$។

ចំនួនអង្កេត(n) និង អញ្ញត្តិឯករាជ្យ(k_j) ស្មើ706 និង 3 រៀងគ្នា។ ដូច្នេះ កំរិតសេរីភាព $(degree\ of\ freedom: df = n - k - 1)$ ស្មើនឹង 702(df = 706 - 3 - 1)។ ហេតុនេះ គេទាញបានតំលៃ១.៩៦០ (critical value ឬ C) នៅឯកំរិត៥%។ តំលៃដាច់ខាត $\boldsymbol{t}\ statistic$ សំរាប់ $\hat{\beta}_{age}$គឺ

$$\left|t_{\hat{\beta}_j}\right| = 2.20/1.446 = 1.521 < c = 1.960$$

ដូច្នេះ $\hat{\beta}_{age}$ ឬ ageមិនមានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត៥%ឡើយ មានន័យថា *អាយុ*(age) *មិនជះឥទ្ធិពលទៅលើការចំណាយពេលគេង*$(sleep)$។

៤.២.៤ ការធ្វើតេស្តសម្មតិកម្មផ្សេងទៀត

ជារួមសម្មតិកម្មដើម(null hypothesis)គឺ $H_0: \beta_j = 0$។ ជួនកាលយើងចង់ធ្វើតេស្ត ថាតើβ_jស្មើនឹងតំលៃថេរ(a_j)ណាមួយដែរឬទេ។ ករណីនេះសម្មតិកម្មដើមគឺ:

$$H_0: \beta_j = a_j \qquad (៤.១២)$$

រីឯ $\boldsymbol{t}\ statistic$ សំរាប់ $\hat{\beta}_j$គឺ

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}\ (\text{ឬ } c)$$

ប្រសិនសម្មតិកម្មដើមគឺ $H_0: \beta_j = 2$ និង សម្មតិកម្មផ្ទុយគឺ $H_1: \beta_j > 2$ បន្ទាប់មកយើងត្រូវស្វែងរកតំលៃ t_{n-k-1} (ឬ c) ជាមួយការធ្វើតេស្តសម្មតិកម្មខាងស្តាំ។ យើងបដិសេធសម្មតិកម្មដើម(H_0) និងទទួលយកសម្មតិកម្មផ្ទុយ(H_1) ប្រសិន$t_{\hat{\beta}_j} > c$។ ក្នុងករណីនេះ យើងនិយាយថា $\hat{\beta}_j$មានសារៈសំខាន់ជាស្ថិតិធំជាងពីរនៅឯកំរិតសារៈសំខាន់ណាមួយដែលត្រូវបានជ្រើសរើស។

ឧទាហរណ៍ទី៥

ក្រុមហ៊ុនធានារ៉ាប់រងអាយុជីវិតមួយមានបំណងក្នុងការពិនិត្យទំនាក់ទំនងរវាងចំនួនធានារ៉ាប់រងអាយុជីវិតគ្រួសារមួយ$(insurance)$ និង ចំណូលរបស់គ្រួសារ$(income)$ ដោយអង្កេតទៅលើចំនួន ២០គ្រួសារជាមួយទិន្នន័យ(គិតជាខ្នង់ពាន់ដុល្លា)ដូចខាងក្រោម:

No.	Insurance	Income
1	90	25
2	165	40
3	220	60
4	145	30
5	114	29
6	175	41
7	145	37
8	192	46
9	395	105
10	339	81
11	230	57
12	262	72
13	570	140
14	100	23
15	210	55
16	243	58
17	335	87
18	299	72
19	305	80
20	205	48

ផ្អែកលើទិន្នន័យ យើងទទួលបានសមីការដូចខាងក្រោម:

$$\widehat{insurance} = 6.855 + 3.880\ income$$
$$(7.383)\ \ (0.112)$$
$$n = 20, R^2 = 0.9852$$

អ្នកគ្រប់គ្រងក្រុមហ៊ុនថ្លែងថា រាល់ពេលចំណូលរបស់គ្រួសារ($income$)កើនឡើង1,000ដុល្លារ នោះចំនួនជានារ៉ាប់រងអាយុជីវិតគ្រួសារមួយ($insurance$)នឹងកើនឡើង4,000ដុល្លារ។ ដើម្បីបញ្ជាក់អ្វីដែលអ្នកគ្រប់គ្រងក្រុមហ៊ុនថ្លែងថាតើពិត ឬ មិនពិត ជាដំបូងយើងកំណត់សម្មតិកម្មដើម និង ផ្នែយក៏ដូចជា $\boldsymbol{t}\ statistic$ សំរាប់ $\hat{\beta}_{income}$ដូចខាងក្រោម:

$$H_0: \beta_{income} = 4$$
$$H_1: \beta_{income} \neq 4$$

$$t_{\widehat{\beta}_{income}} = \frac{3.88 - 4}{0.112} = -1.071$$

ចំនួនអង្កេត(n) និង អញ្ញត្តិឯករាជ្យ(k_j) ស្មើ20 និង 1 រៀងគ្នា។ ដូច្នេះ កំរិតសេរីភាព ($degree\ of\ freedom$: $df = n - k - 1$) ស្មើនឹង 18(df = 20 - 1 - 1)។ ហេតុនេះ គេទាញបានតំលៃ2.101 (critical value ឬ C) នៅឯកំរិត5%។

នៅពេល $t = -1.071 > -c = -2.101$។ ដូច្នេះសម្មតិកម្មដើមត្រូវបានទទួលយក មានន័យថា អ្វីដែលអ្នកគ្រប់គ្រងក្រុមហ៊ុនថ្លែងគឺជាការពិត។

៤.៣ ការគណនា$p - Value$សំរាប់$t\ Test$

ក្នុងការធ្វើតេស្ត ប្រសិនយើងទទួលបានតំលៃp ($p -$ value) យើងអាចកំណត់បាននូវលទ្ធផលនៃការធ្វើតេស្ត ដោយប្រៀបធៀប $p -$ valueជាមួយកំរិតសារៈសំខាន់ជាស្ថិតិដែលបានជ្រើសរើស(α)នៅឯ1% 5% ឬ 10%។ តំលៃpប្រែប្រួលនៅចន្លោះ0 និង 1។

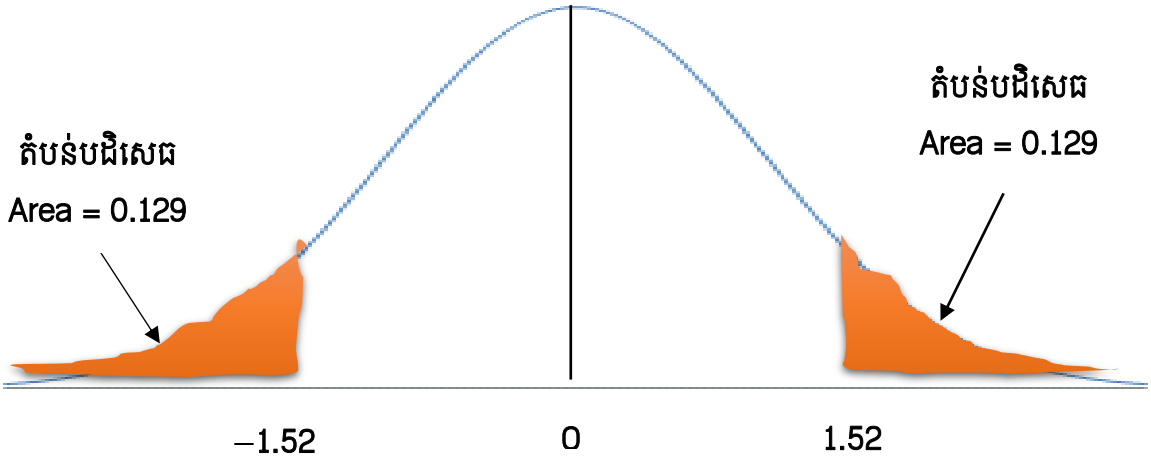
- ✓ ប្រសិន $p \leq \alpha \Rightarrow$ បដិសេធH_0 ដោយទទួលយកH_1
- ✓ ប្រសិន $p > \alpha \Rightarrow$ បដិសេធH_1 ដោយទទួលយកH_0

ដើម្បីឲ្យមានភាពងាយស្រួល តំលៃpត្រូវបានគណនាតាមរយៈការដោយប្រើប្រាស់កំព្យូទ័រ(statistical software)ជាជំនួយដូចជា Eviews, Stata, Excel។ល។ ផ្អែកលើបណ្តាឧទាហរណ៍(ទី២ ទី៣ និង ទី៤)ខាងលើ វាត្រូវបានកំណត់ដូចខាងក្រោម:

តេស្តសម្មតិកម្មខាងស្តាំ	$p = P[t_{(997)} \geq 4.93]$	$p = 0.000$
តេស្តសម្មតិកម្មខាងឆ្វេង	$p = P[t_{(702)} \leq -8.86]$	$p = 0.000$
តេស្តសម្មតិកម្មទាំងសងខាង	$p = P[t_{(702)} \geq 1.52] + P[t_{(702)} \leq -1.52]$	$p = 0.129$

ឧបមាថាយើងជ្រើសរើសយកកំរិតសារៈសំខាន់ជាស្ថិតិនៅឯ៥% ($\propto = 5\%$)។ សំរាប់ករណីទាំងបីខាងលើ យើងអាចសន្និដ្ឋានដូចខាងក្រោម:

តេស្តសម្មតិកម្មខាងស្តាំ	$p = 0.000 < 5\%$	បដិសេធH_0 ដោយទទួលយកH_1
តេស្តសម្មតិកម្មខាងឆ្វេង	$p = 0.000 < 5\%$	បដិសេធH_0 ដោយទទួលយកH_1
តេស្តសម្មតិកម្មទាំងសងខាង	$p = 0.129 > 5\%$	បដិសេធH_1 ដោយទទួលយកH_0



រូបភាព៤.៤: តម្លៃpនៅឯ5%សំរាប់តេស្តសម្មតិកម្មទាំងសងខាង ពេល$t = 1.52$ និង $df = 702$

៤.៤ ចន្លោះជឿជាក់ (Confidence Intervals)

ផ្អែកលើclassical linear model (CLM) assumptions គេអាចបង្កើតចន្លោះជឿជាក់(CI)សំរាប់β_j (population parameterβ_j)។ ចន្លោះជឿជាក់ត្រូវបានគេហៅថា ការប៉ាន់ស្មានចន្លោះ (interval estimation)ផងដែរ ពីព្រោះវាផ្តល់នូវចន្លោះនៃតម្លៃសំរាប់β_j។

$$\hat{\beta}_j \pm c * se(\hat{\beta}_j), \qquad (៤.១៣)$$

ដើម្បីទទួលបានតម្លៃ c យើងត្រូវតែដឹង*degree of freedom* (df) និង កំរិតជឿជាក់(90% 95% ឬ 99%)។ បន្ទាប់មកតម្លៃ cត្រូវបានទទួលពីបំណែងចែកt_{n-k-1}។

សន្មតថា $df = n - k - 1 = 30$ និង សំរាប់β_jណាមួយ កំរិតជឿជាក់គឺ95%។ ដូច្នេះ យើងទទួលបានចន្លោះជឿជាក់គឺ: $[\hat{\beta}_j - 2.042 * se(\hat{\beta}_j), \hat{\beta}_j + 2.042 * se(\hat{\beta}_j)]$។
ផ្អែកលើបណ្តាឧទាហរណ៍(ទី២ ទី៣ និង ទី៤) យើងកំណត់បានចន្លោះជឿជាក់ដូចខាងក្រោម:

- ✓ **ឧទាហរណ៍ទី២** (សម្មតិកម្មខាងស្តាំ)
  $[0.133 - 2.326 * (0.027), 0.133 + 2.326 * (0.027)] = (+0.07, +0.196)$
- ✓ **ឧទាហរណ៍ទី៣** (សម្មតិកម្មខាងឆ្វេង)
  $[-0.148 - 1.645 * (0.0167), -0.148 + 1.645 * (0.0167)] = (-0.175, -00.121)$
- ✓ **ឧទាហរណ៍ទី៤** (សម្មតិកម្មទាំងសងខាង)
  $[2.20 - 1.96 * (1.446), 2.20 + 1.96 * (1.446)] = (-0.634, +5.034)$

កត់សំគាល់

សម្មតិកម្ម	**ចន្លោះជឿជាក់**	**សញ្ញា**	**លទ្ធផលតេស្តសម្មតិកម្ម**
ខាងស្តាំ	$(+0.07, +0.196)$	$(+, +)$	បដិសេធH_0 ដោយ*ទទួលយក*H_1

ខាងឆ្វេង	$(-0.175, -0.121)$	$(-,-)$	បដិសេធH_0 ដោយ*ទទួលយក*H_1
ទាំងសងខាង	$(-0.634, +5.034)$	$(-,+)$	បដិសេធH_1 ដោយទទួលយកH_0

៤.៥ ធ្វើតេស្តសម្មតិកម្មអំពីការបញ្ចូលគ្នានៃបណ្តាប៉ារ៉ាម៉ែត្រលីនេអ៊ែរតែមួយ

(a Single Linear Combination of the Parameters)

នៅផ្នែក៤.២ និង៤.៤ខាងលើ យើងបានបរិយាយអំពីការធ្វើតេស្តសម្មតិកម្ម និង កំណត់ចន្លោះជឿជាក់អំពីβ_jនីមួយៗដាច់ចេញពីគ្នា។ ប៉ុន្តែក្នុងការអនុវត្តន៍ ជួនកាលβ_jត្រូវបានគេបញ្ចូលគ្នាតែមួយក្នុងពេលតែមួយដោយការប្រើ*t*-test។ តើតេស្តរបៀបនេះត្រូវបានអនុវត្តយ៉ាងដូចម្តេច?

ឧទាហរណ៍ទី៦

ដើម្បីបង្ហាញវិធីសាស្ត្រនេះ យើងពិចារណាសមីការមួយដែលពន្យល់អំពីប្រាក់ខែរបស់នាយកក្រុមហ៊ុនជាអនុគមន៍នៃចំណូលប្រចាំឆ្នាំរបស់ក្រុមហ៊ុន ប្រាក់ចំណេញដែលទទួលពីការប្រើប្រាស់ទ្រព្យម្ចាស់ (return on equity, *roe*, គិតជាភាគរយ) និង ប្រាក់ចំណេញដែលទទួលពីការវិនិយោគទៅលើទ្រព្យសម្បត្តិសរុប(return on assets, *roa*, គិតជាភាគរយ)។ គេបានសមីការដូចខាងក្រោម:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roa + \beta_3 roe + u, \qquad (៤.១៤)$$

យើងចង់ដឹងថា តើប្រាក់ចំណេញដែលទទួលបានពីការប្រើប្រាស់ទ្រព្យម្ចាស់ (return on equity, *roe*) និង ប្រាក់ចំណេញដែលទទួលពីការវិនិយោគទៅលើទ្រព្យសម្បត្តិសរុប(return on assets, *roa*) ដែរឥទ្ធិពលទៅលើប្រាក់ខែរបស់នាយកក្រុមហ៊ុនកំរិតដូចគ្នាដែរឬទេ? ហេតុនេះ គេសរសេរបានសម្មតិកម្មដើម:

$$H_0: \beta_2 = \beta_3 \qquad (៤.១៥)$$

ឧបមាថា កំរិតនៃការដេះឥទ្ធិពលនៃប្រាក់ចំណេញដែលទទួលបានពីការវិនិយោគទៅលើទ្រព្យសម្បត្តិសរុបរបស់ក្រុមហ៊ុន(return on assets, *roa*)ទៅលើប្រាក់ខែរបស់នាយកក្រុមហ៊ុនតូចជាងប្រាក់ចំណេញដែលទទួលបានពីការប្រើប្រាស់ទ្រព្យម្ចាស់ (return on equity, *roe*) យើងសរសេរបានសម្មតិកម្មផ្ទុយដូចខាងក្រោម:

$$H_1: \beta_2 < \beta_3 \qquad (៤.១៦)$$

សម្មតិកម្ម(៤.១៥) និង (៤.១៦) ផ្តោតទៅលើការធ្វើតេស្តប៉ារ៉ាម៉ែត្រពីរបញ្ចូលក្នុងពេលតែមួយតែម្តងគឺ β_2និង β_3 ដូច្នេះគេមិនអាចធ្វើតេស្តប៉ារ៉ាម៉ែត្រβ_2និង β_3ដាច់ចេញពីគ្នាឡើយ (*individual t*-test)។ ករណីដើម្បីធ្វើតេស្តប៉ារ៉ាម៉ែត្រពីរបញ្ចូលក្នុងពេលតែមួយ យើងសរសេរបានសម្មតិកម្មដើមនិង សម្មតិកម្មផ្ទុយឡើងវិញ: $H_0: \beta_2 - \beta_3 = 0$ និង $H_1: \beta_2 - \beta_3 < 0$។ ក្នុងការប៉ាន់ស្មាន លំអៀង

សំណាក (sampling error) គេកំណត់បានភាពខុសគ្នានេះ ដោយចែកជាមួយលំអៀងស្តង់ដា (standard error):

$$t = (\hat{\beta}_2 - \hat{\beta}_3)/se(\hat{\beta}_2 - \hat{\beta}_3) \qquad (៤.១៧)$$

ក្នុងការធ្វើតេស្តសម្មតិកម្ម យើងត្រូវជ្រើសរើសកំរិតសារៈសំខាន់ (១0% ៥% ឬ ១%) ជាមួយ*df* (*degree of freedom*) ដើម្បីទទួលបានតំលៃ*C* (*critical value*)។ លទ្ធផលទទួលស្គាល់សម្មតិកម្ម ផ្ទុយ (៤.១៦)នៅពេល $t < -c$។

យោងលើ ***ឧទាហរណ៍ទី៦*** ឧបមាថាក្រុមហ៊ុនចំនួន២0៩ត្រូវបានលើកយកមកសិក្សា និងទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\log(\widehat{salary}) = 4.072 + 0.2944\log(sales) + 0.0053\,roa + 0.0119\,roe \qquad (៤.១៨)$$
$$(0.286)\ (0.0319) \qquad (0.0011) \qquad (0.0039)$$
$$n = 209, R^2 = 0.3545$$

យោងលើ(៤.១៨) ភាពខុសគ្នាដែលត្រូវបានប៉ាន់ស្មានរវាងប្រាក់ចំណេញដែលទទួលបានពីការវិនិយោគទៅលើទ្រព្យសម្បត្តិសរុបរបស់ក្រុមហ៊ុន(return on assets, *roa*) និង ប្រាក់ចំណេញដែលទទួលបានពីការប្រើប្រាស់ទ្រព្យម្ចាស់ (return on equity, *roe*)គឺ $\hat{\beta}_2 - \hat{\beta}_3 = -0.0066$ ដែលវាគ្រាន់តែជាភាគយកនៃ*t*-test ក្នុង(៤.១៧)។

កំណត់ប៉ារ៉ាម៉ែត្រថ្មីមួយជាភាពខុសគ្នារវាងβ_2 និង β_3: $\theta_1 = \beta_2 - \beta_3$។ បន្ទាប់មក យើងត្រូវធ្វើតេស្ត:

$$H_0: \theta_1 = 0 \text{ និង } H_1: \theta_1 < 0, \qquad (៤.១៩)$$

យោងលើ(៤.១៨) យើងទាញបាន $t = \hat{\theta}_1/se(\hat{\theta}_1)$។

ពីព្រោះ $\theta_1 = \beta_2 - \beta_3 \Rightarrow \beta_2 = \theta_1 + \beta_3$។ ហេតុនេះសមីការ(៤.១៤) គេអាចសរសេរឡើងវិញដូចខាងក្រោម:

$$\log(salary) = \beta_0 + \beta_1\log(sales) + (\theta_1 + \beta_3)roa + \beta_3 roe + u$$

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \theta_1 roa + \beta_3(roa + roe) + u, \qquad (៤.២0)$$

ផ្អែកលើ(៤.២0) ប៉ារ៉ាម៉ែត្រសំខាន់ដែលត្រូវធ្វើតេស្តគឺ θ_1 និង យើងតាងសរុបភាគរយនៃប្រាក់ចំណេញដែលទទួលបានពីការវិនិយោគទៅលើទ្រព្យសម្បត្តិសរុបរបស់ក្រុមហ៊ុន(return on assets, *roa*) និង ប្រាក់ចំណេញដែលទទួលបានពីការប្រើប្រាស់ទ្រព្យម្ចាស់ (return on equity, *roe*)ដោយ $return$: $return = roa + roe$ នោះគេបាន:

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \theta_1 roa + \beta_3 return + u, \qquad (៤.២១)$$

ផ្អែកលើទិន្នន័យ ក្រុមហ៊ុនចំនួន២០៩ត្រូវបានលើកយកមកសិក្សា និងទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{\log(salary)} = 4.072 + 0.2944\log(sales) - 0.0066\,roa + 0.0119\,return \quad (4.22)$$
$$(0.286)\ (0.0319)\qquad (0.0044)\qquad (0.0039)$$
$$n = 209, R^2 = 0.3545$$

t-test សំរាប់ការធ្វើតេស្ត(៤.១៥)គឺ $-0.0066/0.0044 = -1.50$ និង តំលៃC (*critical value*) គឺ1.960 នៅឯកំរិតសារៈសំខាន់ 5% ជាមួយ*df* (*degree of freedom*), 20។ គេទទួលបាន $t = -1.50 > -c = -1.960$ ឬ តំលៃ$p = 13.39\% > 5\%$ ដូច្នេះ មិនមានភស្តុតាងក្នុងការបដិសេធសម្មតិកម្មដើម(៤.១៥)ឡើយ។

កត់សំគាល់: លទ្ធផលបញ្ជាក់ថាត្រឹមត្រូវ ពីព្រោះអ៊ីនជ័រស៊ិប(*intercept*) និង តម្លៃមេគុណដែលបានប៉ាន់ស្មាន(*slop estimate*)នៃអញ្ញត្តិ$log(sales)$ជាមួយលំអៀងស្តង់ដា(*standard error*)របស់វានៅក្នុង(៤.១៨) និង (៤.២២)គឺដូចគ្នា។

៤.៦ ការធ្វើតេស្តលីនេអ៊ែរពហុ: *F test*

យើងនឹងធ្វើតេស្ត ថាតើក្រុមនៃអញ្ញត្តិឯករាជ្យទាំងឡាយ(*independent variables*)ពិតជាមិនជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ(*dependent variables*)មែនដែរឬទេ? សម្មតិកម្មដើមគឺថាសំណុំនៃអញ្ញត្តិឯករាជ្យមួយមិនជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ។

ដើម្បីបកស្រាយ ហេតុអ្វីការធ្វើតេស្តសារៈសំខាន់នៃសំណុំនៃអញ្ញត្តិឯករាជ្យមានប្រយោជន៍ ***ឧទាហរណ៍ទី១*** ត្រូវបានពិភាក្សាបន្តដោយការបន្ថែមអញ្ញត្តិរៀបការ(*married*) ក្នុងនោះអ្នកដែលរៀបការតាងដោយលេខ១ (*married = 1*) និង អ្នកដែលមិនទាន់រៀបការតាងដោយលេខ០ដូចខាងក្រោម:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 married + u, \quad (4.23)$$

ផ្អែកលើ(៤.២៣) យើងបានសម្មតិកម្មដូចតទៅ:

សម្មតិកម្មដើម $H_0: \beta_3 = 0, \beta_4 = 0,$ (៤.២៤)

សម្មតិកម្មផ្ទុយ $H_1: H_0$ គឺមិនពិត (៤.២៥)

យោងលើការអង្កេតមនុស្សចំនួន៧០៦នាក់ គេទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{sleep} = 3{,}594.60 - 0.148\,totwrk - 10.931\,educ + 2.186\,age + 49.420\,married$$
$$(118.00)\ (0.017)\qquad (5.886)\qquad (1.445)\qquad (41.241) \quad \textbf{(4.26)}$$
$$n = 706, R^2 = 0.1152$$

ដូច្នេះ $t_{age} = 2.186/1.445 = 1.513$ និង $t_{married} = 49.42/41.241 = 1.198$។ ដោយអនុវត្តន៍តេស្តសម្មតិកម្មទាំងសងខាងជាមួយកំរិតសារៈសំខាន់ស្ថិតិគឺ ៥% និង $df = ៧០១$ នោះ $c =$ ១.៩៦០។ នេះបញ្ជាក់ថា មិនមានសារៈសំខាន់ជាស្ថិតិកើតឡើងទាំងអញ្ញត្តិអាយុ(*age*) និង រៀបការ (*married*) ទៅលើការគេង(*sleep*) នៅឯកំរិត៥%ឡើយ ករណីលទ្ធផលតេស្តនីមួយៗដាច់ចេញពីគ្នា(*individual test*)។

បើអញ្ញត្តិអាយុ(*age*) និង រៀបការ(*married*)ត្រូវបានដកចេញពីសមីការ គេទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{sleep} = 3{,}756.22 - 0.149\ totwrk - 13.504\ educ \qquad (៤.២៧)$$
$$(81.29)\ (0.017) \qquad (5.680)$$
$$n = 706, R^2 = 0.1104$$

សមីការ(៤.២៦)ត្រូវបានគេហៅថា *linear unrestrictions* (*ur*) *model* និង (៤.២៧)គេហៅថា *linear restrictions* (*r*) *model*។ សំណួរសួរថា ***តើវាមានសារៈសំខាន់ជាស្ថិតិកើតឡើងដែរឬទេ ប្រសិនអញ្ញត្តិអាយុ(age) និង រៀបការ(married)ត្រូវបានបញ្ចូលគ្នា?*** ដើម្បីឆ្លើយសំណួរនេះ ជាដំបូងយើងត្រូវ គណនាតំលៃ*F* និង បន្ទាប់មកត្រូវប្រៀបធៀបតំលៃ*F* ជាមួយ*critical value* (*C*) តាមរូបមន្តដូចខាងក្រោម:

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n - k - 1)} = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/df_{ur}} \sim F_{q,\ n-k-1}(or\ C) \qquad (៤.២៨)$$

$$\text{បើ} \quad F < C \quad \Rightarrow \text{យល់ស្របសម្មតិកម្មដើម}$$
$$F > C \quad \Rightarrow \text{បដិសេធសម្មតិកម្មដើម និងយល់ស្របសម្មតិកម្មផ្ទុយ} \qquad (៤.២៩)$$

$$q = df_r - df_{ur}, \quad df = n - k - 1$$

តាមរូបមន្ត(២.២៨) គេទទួលបានតម្លៃ*F*:

$$F = \frac{(0.1152 - 0.1104)/2}{(1 - 0.1152)/(706 - 4 - 1)} \sim F_{2,\ 701}\ (or\ C)$$

$$= 1.905 < C = 3\ \ at\ 5\%$$

លទ្ធផលបង្ហាញថា *យល់ស្របតាមសម្មតិកម្មដើម*(H_0)។ ដូច្នេះ វាមិនមានសារៈសំខាន់ជាស្ថិតិកើតឡើង ករណីអញ្ញត្តិអាយុ(age) និង រៀបការ(married)ត្រូវបានបញ្ចូលគ្នា ទៅលើការគេង(*sleep*) នៅឯកំរិត៥%ឡើយ។

គណនាតម្លៃ*F* គេអាចអនុវត្តតាមរូបមន្តផ្សេងទៀតដូចខាងក្រោម:

$$F = \frac{(SSR_{ur} - SSR_r)/q}{SSR_{ur}/(n-k-1)} \qquad (២.៣០)$$

ដែល SSR: ផលបូកការេនៃសំណល់ $(Residual\ Sum\ of\ Squares)$

៤.៧ *F test* សារៈសំខាន់នៃរ៉ឺគ្រេស្យុងទាំងមូល

ម៉ូដែលរ៉ឺគ្រេស្យុង សម្មតិកម្មដើម និង សម្មតិកម្មផ្ទុយត្រូវបានសរសេរដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$$

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0, \qquad (២.៣១)$$

$$H_1: \beta_1 \neq 0, \beta_2 \neq 0, \ldots, \beta_k \neq 0 \qquad (២.៣២)$$

ក្នុងករណីទទួលស្គាល់សម្មតិកម្មដើម(២.៣១) យើងបានម៉ូដែល(*restricted model*): $y = \beta_0 + u$, ពាល់អញ្ញត្តិឯករាជ្យ(x_j)ត្រូវបានលុបចោលពីសមីការ។ ដូច្នេះ $R^2 = 0$ និង គ្មានបំរែបំរួលនៃyត្រូវបានពន្យល់ឡើយ ពីព្រោះគ្មានអញ្ញត្តិឯករាជ្យ(x_j)។ ហេតុនេះ ***F test*** សំរាប់ធ្វើតេស្ត (២.៣១) គឺដូចខាងក្រោម:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,\ n-k-1}(or\ C) \qquad (២.៣៣)$$

យោងពី ***ឧទាហរណ៍ទី៣*** យើងអាចគណនាតម្លៃFដូចខាងក្រោម:

$$F = \frac{0.1134/3}{(1-0.1134)/(706-3-1)} = 30$$

ដោយកំណត់យកកំរិតសារៈសំខាន់ស្ថិតិគឺ 5% និង $df = 706-3-1 = 702$ នោះ $c = 2.60$ $\left(F_{2,706-3-1}\ or\ C\right)$។

បើ $F < C$ $\Rightarrow$ យល់ស្របសម្មតិកម្មដើម

$F > C$ $\Rightarrow$ បដិសេធសម្មតិកម្មដើម និង យល់ស្របសម្មតិកម្មផ្ទុយ (៤.៣៤)

លទ្ធផល $F = 30 > C = 2.60$ $\Rightarrow$ យល់ស្របសម្មតិកម្មផ្ទុយ។ ដូច្នេះ ម៉ូដែលទាំងមូលមានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត5% (*overall significance of the regression*) ឬ យ៉ាងហោចណាស់មានអញ្ញត្តិឯករាជ្យ (x_j)មួយដែរ ដែលជះឥទ្ធិពលទៅលើអញ្ញត្តិមិនឯករាជ្យ(y)។

សំណួរ

៤.១ ក. តើលទ្ធផលនៃFtestប្រាប់អំពីអ្វី?

ខ. អ្វីទៅជាសម្មតិកម្មទាំងពីរនៃFtest?

គ. ដើម្បីឲ្យលទ្ធផលនៃFtestមានសារៈសំខាន់ជាស្ថិតិ តើតម្លៃR^2ត្រូវតែទាប ឬ ខ្ពស់? ហេតុអ្វី?

៤.២ ក. តើអ្វីជា$t$test?

ខ. ប្រសិនលទ្ធផលនៃFtestមិនមានសារៈសំខាន់ជាស្ថិតិ តើយើងអាចដំណើការបន្តទៅមុខ និង អនុវត្ត$t$testសំរាប់β_jនីមួយៗដែរឬទេ? ហេតុអ្វី?

គ. ប្រសិនអ្នកត្រូវការទទួលបានលទ្ធផលនៃការព្យាករណ៍ដ៏ល្អបំផុត ហេតុអ្វីបានជាអ្នកមិនដាក់បន្ថែមបណ្តាអញ្ញត្តិសំខាន់ៗដែលអ្នកបានគិត?

៤.៣ ផលិតភាព(Productivity)គឺជាបញ្ហាសំខាន់។ ចំពោះនិយោជិកម្នាក់ៗ អ្នកមានទិន្នន័យអំពីផលិតភាព ក៏ដូចជាបណ្តាកត្តាដទៃទៀត។ អ្នកចង់ដឹងថា តើកត្តាទាំងនេះពន្យល់អំពីបម្រែបម្រួលនៃផលិតភាពពីនិយោជិកម្នាក់ៗទៅម្នាក់ទៀតបានប៉ុន្មាន។

ក. ចូរសរសេរម៉ូដែលរ៉ឺគ្រ្ស្សុងដែលបង្ហាញអំពីទំនាក់ទំនងរវាងផលិតភាព និង កត្តាទាំងនេះ។

ខ. តើយើងត្រូវប្រើប្រភេទនៃតេស្តអ្វី ដើម្បីពន្យល់អំពីបំរែបំរួលនៃផលិតភាពពីនិយោជិកម្នាក់ៗទៅម្នាក់ទៀត?

៤.៤ ដើម្បីជួយទៅដល់យុទ្ធនាការណ៍ម៉ាកេធីញសំរាប់ហាងលក់ទំនិញរបស់អ្នកដែលទាក់ទងនឹងការកំណត់អតិថិជនគោលដៅ (age level)។ អ្នកចង់ស្វែងរក ថាតើមានសារៈសំខាន់ជាស្ថិតិនូវភាពខុសគ្នានៃអាយុរបស់អតិថិជនដោយគិតជាមធ្យមសំរាប់អាយុ45ឆ្នាំដែរឬទេ។ ក្នុងចំណោមអតិថិជនទាំងអស់ដែលស្ថិតនៅតំបន់នោះ អ្នកធ្វើការអង្កេតទៅលើអតិថិជនចំនួន100នាក់។

ក. ចូរកំណត់សម្មតិកម្មដើម និង សម្មតិកម្មផ្ទុយសំរាប់តេស្តទាំងសងខាង។

ខ. ក្នុងការអនុវត្តតេស្តទាំងសងខាងនៅឯកំរិតសារៈសំខាន់5% ចូរអធិប្បាយអំពីលទ្ធផលនៃតេស្ត។

លំហាត់

៤.១ ផ្អែកលើទិន្នន័យក្នុងលំហាត់ ៣.១នៃមេរៀនទី៣ ចូរឆ្លើយសំណួរដូចខាងក្រោម:

ក. តើមានសារៈសំខាន់ជាស្ថិតិកើតឡើងនូវទំនាក់ទំនងរវាងថ្លៃដើមដឹកជញ្ជូន និង ចំណូលពីការលក់ ក៏ដូចជាការបញ្ជាទិញនៅឯកំរិត5%ដែរឬទេ ?

ខ. ចូរបកស្រាយអត្ថន័យនៃតម្លៃ p (p-value)

គ. ចូរកំណត់ជឿជាក់នៅឯ95%សំរាប់ β_{sales} និង β_{order}

ឃ. តើបម្រែបម្រួលនៃថ្លៃដើមដឹកជញ្ជូនប្រហែលប៉ុន្មាន ដែលត្រូវបានពន្យល់ដោយចំណូលពីការលក់ និង ការបញ្ជាទិញ

៤.២ អ្នកបានសំរេចចិត្តបង្កើតផលិតផលថ្មីមួយប្រភេទ និង អ្នកចង់ធ្វើតេស្ត ថាតើចំនួនភាគរយនៃអតិថិជនតិចជាង ឬ ច្រើនជាង20%នៅក្នុងសហគមន៍មួយបានចាប់អារម្មណ៍ទៅលើការទិញផលិតផលនោះដែរឬទេ។ ប្រសិនមានភស្តុតាងគ្រប់គ្រាន់ឆ្លើយតបនឹងតំរូវការ អ្នកនឹងបង្កើតផលិតផលនោះ។ អ្នកបានអង្កេតទៅលើមនុស្សចំនួន500នាក់ និង រកឃើញថា 32%នៃអ្នកដែលរស់នៅក្នុងសហគមន៍ត្រូវការផលិតផលថ្មីនោះ។

ក. ករណីនេះ ហេតុអ្វីវាសមហេតុផលសំរាប់តេស្តតែម្ខាង(one-sided test) ?

ខ. ថ្លែងអំពីសម្មតិកម្មដើម និង សម្មតិកម្មដើមសំរាប់តេស្តតែម្ខាង

គ. អនុវត្តតេស្តនៅឯកំរិតសារៈសំខាន់5% និង អធិប្បាយអំពីលទ្ធផលនៃតេស្ត

ឃ. អនុវត្តតេស្តនៅឯកំរិតសារៈសំខាន់1% និង អធិប្បាយអំពីលទ្ធផលនៃតេស្ត

៤.៣ យោងលើលំហាត់៣.២នៃមេរៀនទី៣ ឧបមាថា យើងអង្កេតទៅលើ45ក្រុមហ៊ុន និង ទទួលបានទិន្នន័យ (ក្រុមហ៊ុនដែលមានទីតាំងនៅតំបន់ពាណិជ្ជកម្មតំលៃស្មើ1 និង ផ្ទុយពីនេះស្មើសូន្យ) ដូចខាងក្រោម:

ROA	sales	WC	TAT	QR	DR	location	log(sales)	log(wc)
12.45	54987	14987	2.01	1.01	0.44	0	4.74026	4.175715
23.1	175098	75098	4.35	2.35	0.35	1	5.2432812	4.875628
17.89	60898	20898	3.89	2.89	0.48	1	4.784603	4.320105
24.21	1123870	123870	4.39	3.37	0.37	1	6.0507161	5.092966
14.56	342895	142895	2.13	1.13	0.49	0	5.5351612	5.155017
25.22	1345890	445890	4.12	2.12	0.38	1	6.1290096	5.649228
34.31	3737890	737890	7.98	3.12	0.56	1	6.5726265	5.867992
32.62	2832900	832900	5.98	2.98	0.48	1	6.4522312	5.920593
12.97	234560	94560	1.78	0.78	0.78	0	5.370254	4.975707
10.35	342579	42579	1.24	1.21	0.21	0	5.5347607	4.629195
9.78	132798	32798	1.98	0.98	0.38	0	5.1231915	4.515847
11.89	453908	103908	1.28	0.97	0.47	0	5.6569678	5.016649
10.45	232000	92000	1.09	0.97	0.37	0	5.365488	4.963788
32.19	3121908	921908	5.12	2.32	0.32	1	6.4944201	5.964688

11.13	297890	97890	1.09	1.9	0.45	0	5.4740559	4.990738
12.45	321900	81900	1.1	0.79	0.39	1	5.507721	4.913284
13.32	289000	89000	1.58	1.124	0.42	1	5.4608978	4.94939
22.65	1321890	921890	3.89	2.14	0.49	1	6.1211953	5.964679
30.12	2935690	1235690	5.12	2.57	0.57	1	6.4677102	6.09191
26.32	937890	937890	3.89	2.26	0.26	1	5.9721519	5.972152
17.69	321900	91900	2.79	1.87	0.87	1	5.507721	4.963316
18.69	543200	123200	2.9	1.29	0.48	1	5.7349598	5.090611
23.68	1432900	932900	3.27	2.58	0.58	1	6.1562159	5.969835
18.45	568900	98900	2.98	1.49	0.49	1	5.7550359	4.995196
14.34	1278900	978900	1.19	1.14	46	0	6.1068366	5.990738
24.78	1345678	945678	3.89	1.98	0.43	1	6.1289412	5.975743
28.35	90800	70800	4.24	2.59	0.38	1	4.9580858	4.850033
21.16	65790	25790	4.12	2.48	0.48	1	4.8181599	4.411451
26.65	78900	18900	4.22	2.89	0.54	1	4.897077	4.276462
28.45	908859	308859	4.46	2.29	0.39	1	5.9584965	5.48976
14.36	23567	13567	1.89	1.19	0.38	0	4.3723043	4.132484
18.38	32456	22456	2.23	2.57	0.46	1	4.511295	4.351332
17.98	213890	93890	2.19	1.69	0.67	0	5.3301905	4.972619
30.21	1934500	234500	4.98	3.12	0.29	1	6.2865687	5.370143
29.44	187600	97600	4.68	2.86	0.29	1	5.2732328	4.98945
25.19	82460	22460	3.78	2.24	0.58	1	4.9162433	4.35141
20.89	23580	13580	3.12	1.89	0.42	1	4.3725438	4.1329
13.98	12890	6890	2.58	1.59	0.38	0	4.1102529	3.838219
31.24	187600	87600	5.46	2.87	0.49	1	5.2732328	4.942504
23.17	76900	36900	3.89	2.24	0.49	1	4.8859263	4.567026
28.79	125690	65690	3.99	1.98	0.41	1	5.0993007	4.817499
25.79	679000	279000	3.24	2.14	0.29	1	5.8318698	5.445604
23.14	95890	45890	2.99	1.58	0.58	1	4.9817733	4.661718
18.97	55890	25890	2.19	2.68	0.68	1	4.7473341	4.413132
34.19	2288960	128960	6.26	3.24	0.29	1	6.3596382	5.110455

យើងពិចារណាសមីការដើម្បីពន្យល់អពីកត្តាទាំងឡាយដែលជះឥទ្ធិពលទៅលើលទ្ធភាពប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 \log(SALES) + \beta_2 \log(WC) + \beta_3 TAT + \beta_4 QR + \beta_5 DR + u$$

ក. តើអ្នករំពឹងថា សញ្ញាសំរាប់$\beta_1, \beta_2,$ និង β_3វិជ្ជមាន ឬ អវិជ្ជមាន ?

ខ. រាល់ប៉ារ៉ាម៉ែត្រនៃម៉ូដែល ចូរកំណត់សម្មតិកម្មដើម និង ផ្ទុយនៃlog(sales) log(WC) TAT QR និង DRជាមួយROA។

គ. ប្រើប្រាស់ទិន្នន័យខាងលើ សមីការខាក្រោមត្រូវបានទទួលដោយOLS:

$$\widehat{ROA} = \underset{(3.601)}{-0.067} - \underset{(1.558)}{2.840}\log(SALES) + \underset{(1.638)}{4.307}\log(WC) + \underset{(0.465)}{3.460}TAT + \underset{(0.887)}{1.873}QR - \underset{(0.060)}{0.007}DR$$

$$n = 45, R^2 = 0.8978$$

ចូរធ្វើតេស្តសម្មតិកម្មទាំងឡាយដែលថ្លែងក្នុងសំណួរ(ខ)នៅឯកំរិតសារៈសំខាន់៥%។

ឃ. តើយើងអាចបដិសេធH$_0$: $\beta_{\log(WC)} = 2, \beta_{TAT} = 1.2, \beta_{QR} = 1$ នៅឯកំរិតសារៈសំខាន់៥% ដែរឬទេ ?

ង. ផ្អែកលើOLSក្នុងសំណួរ(គ) ចូរកំណត់ចន្លោះជឿជាក់(confidence interval)សំរាប់រាល់β_j ។

ច. ផ្អែកលើលទ្ធផលនៃសំណួរ(គ) យើងទទួលបានOLSថ្មី:

$$\widehat{ROA} = \underset{(3.388)}{-1.989} + \underset{(0.680)}{1.648}\log(WC) + \underset{(0.452)}{3.312}\,\text{TAT} + \underset{(0.898)}{1.978}\,QR$$

$$n = 45, R^2 = 0.8891$$

តើយើងអាចបដិសេធ H$_0$: $\beta_{\log(SALES)} = \beta_{DR} = 0$ នៅឯកំរិតសារៈសំខាន់៥%ដែរឬទេ ?

មេរៀនទី៥
រ៉ឺគ្រេស្យុងពហុៈ លក្ខណៈបន្ថែម

ក្នុងមេរៀនមុន យើងបានសិក្សារ៉ឺគ្រេស្យុងលីនេអ៊ែរពហុ (multiple linear regression) ជាមួយវិធីសាស្ត្រសំរាប់វិភាគទិន្នន័យ ដែលជួយឲ្យយើងយល់ដឹងទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ (y) និងអញ្ញត្តិឯករាជ្យ (x_j) ឬ បើអញ្ញត្តិx_jណាមួយប្រែប្រួល តើវាធ្វើឲ្យអញ្ញត្តិ yប្រែប្រួលយ៉ាងដូចម្តេច? ក៏ដូចជាការព្យាករណ៍អញ្ញត្តិ y។ ម៉ូដែលត្រូវបានប៉ាន់ស្មានដូចខាងក្រោម:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

យើងចាប់ផ្តើមគិតអំពីការបង្កើតម៉ូដែលជាមួយការធ្វើតេស្តសមមូតិកម្មសំរាប់ប៉ារ៉ាម៉ែត្ររ៉ឺគ្រេស្យុង, $(\beta_1, \beta_2, \ldots, \beta_k)$។ តេស្តទាំងនេះអាចកំណត់ តើអញ្ញត្តិx_jណាខ្លះដែលសំខាន់សំរាប់ម៉ូដែលy និង គួរតែត្រូវបានបញ្ចូលទៅក្នុងម៉ូដែល ហើយផ្ទុយទៅវិញ តើអញ្ញត្តិx_jណាខ្លះដែលមិនសូវសំខាន់សំរាប់ម៉ូដែល y និង គួរតែត្រូវបានដកចេញពីម៉ូដែលដែរឬទេ។

ខណៈពេលក្នុងការសម្រេចចិត្ត ថាតើអញ្ញត្តិx_jណាខ្លះដែលគួរតែត្រូវបានបញ្ចូលទៅក្នុងម៉ូដែល និងតើអញ្ញត្តិx_jណាខ្លះដែលគួរតែត្រូវបានដកចេញពីម៉ូដែល រាល់វិធីសាស្ត្រមួយចំនួនត្រូវបានពិភាក្សានៅក្នុងមេរៀនមុនរួចមកហើយ ដូចជា*F*-testជាដើមសំរាប់ម៉ូដែលលីនេអ៊ែរ(linear regression model) ។ ជួនកាល ម៉ូដែលមិនមានលក្ខណៈជាលីនេអ៊ែរ(nonlinear regression model) ដែលវាត្រូវបានសិក្សានៅក្នុងមេរៀននេះ។

៥.១ ការជះឥទ្ធិពលនៃរង្វាស់រង្វាល់ទិន្នន័យទៅលើស្ថិតិOrdinary Least Squares

ឥឡូវនេះ យើងបង្ហាញរង្វាស់រង្វាល់ទិន្នន័យនៃអញ្ញត្តិមិនឯករាជ្យ (y) និង អញ្ញត្តិឯករាជ្យ (x_j) និងការជះឥទ្ធិពលរបស់វាទៅលើលំអៀងស្តង់ដារ (*standard errors*) *t*-test *F*-test និង ចន្លោះជឿជាក់ (*confident intervals*)។

យោងលើ***ឧទាហរណ៍ទី១*** ដែលបានពិភាក្សានៅក្នុង***មេរៀនទី៤*** យើងអាចសរសេរបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$\widehat{sleep} = \hat{\beta}_0 + \hat{\beta}_1 totwrk + \hat{\beta}_2 educ + \hat{\beta}_3 age \qquad (\text{៥.១})$$

ក្នុងនោះអញ្ញត្តិ $sleep$ និង $totwrk$ គិតជាចំនួននាទីក្នុង១សប្តាហ៍ រីឯ ការចំណាយពេលសិក្សា $(educ)$ និង អាយុ (age)គិតជាឆ្នាំ។ ក្នុងករណី អញ្ញត្តិ $sleep$ និង $totwrk$ ដែលគិតជាចំនួននាទី នឹងត្រូវបានបំលែងទៅជាម៉ោង នោះយើងទទួលបានសមីការប៉ាន់ស្មានថ្មីដូចខាងក្រោម:

$$\widehat{sleep}/60 = \hat{\beta}_0/60 + \hat{\beta}_1(totwrk/60) + (\hat{\beta}_2/60)educ + (\hat{\beta}_3/60)age \qquad (๕.๒)$$

ដោយអង្កេតទៅលើមនុស្សចំនួន៧០៦នាក់ សមីការ(๕.๑) និង (๕.๒) បង្ហាញលទ្ធផលដូចខាង

តារាង ๕.๑ *ការជះឥទ្ធិពលនៃរង្វាស់រង្វាល់ទិន្នន័យ*

អញ្ញត្តិមិនឯករាជ្យ	(๑) ***sleep***	(๒) ***sleep***
អញ្ញត្តិឯករាជ្យ		
totwrk.....	-๐.๑๔๘๔	-๐.๑๔๘๔
	(๐.๐๑๗๐)	(๐.๐๑๗๐)
educ.....	-๑๑.๑๓๓๘	-๐.๑๘๕๖
	(๕.๘๘๔๖)	(๐.๐๙๘๐)
age.....	๒.๑๙๙๙	๐.๐๓๖๗
	(๑.๔๔๕๗)	(๐.๐๒๔๑)
Intercept.....	๓,๖๓๘.๒๕	๖๐.๖๓๗๔
	(๑๑๒.๒๘)	(๑.๘๗๑๓)
Observation.....	๗๐๖	๗๐๖
R-Squared.....	๐.๑๑๓๔	๐.๑๑๓๔
F-statistic.....	๒๙.๙๑๘๘๙	๒๙.๙๑๘๘๙
Pro. (*F*-stat).....	๐.๐๐๐...	๐.๐๐๐...

ផ្អែកលើ**តារាង ๕.๑** លំអៀងស្តង់ដារ (*standard errors*)ត្រូវបានបង្ហាញក្នុងសញ្ញាវង់ក្រចក។ សំរាប់**ម៉ូដែលទី๑** ប្រសិនក្នុងមួយសប្តាហ៍ចំណាយពេលធ្វើការថែមចំនួន๑๒๐នាទី នោះគេអាចព្យាករណ៍ថា ការគេងត្រូវបានកាត់បន្ថយអស់ប្រហែល๐.๑๔๘๔(๑๒๐) = ๑๗.๘๑នាទី។ *t*-test

លើចំណាយពេលធ្វើការ(*totwrk*)គឺ−0.1484/0.0170 = −8.729។ ដូច្នេះ អញ្ញត្តិ *totwrk* មានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត៥%។

ឥឡូវ យើងបំលែងទិន្នន័យនៃអញ្ញត្តិចំណាយពេលធ្វើការ(*totwrk*) និង អញ្ញត្តិការចំណាយពេលគេង (*sleep*)ដែលគិតជានាទី ទៅជាម៉ោងវិញ។ ហេតុនេះ ទិន្នន័យនៃអញ្ញត្តិទាំងពីរត្រូវចែកជាមួយ ៦០ ពីព្រោះ ១ម៉ោងស្មើ៦០នាទី រីឯទិន្នន័យការសិក្សា(*educ*) និង អាយុ(*age*) រង្វាស់រង្វាល់របស់វាត្រូវបានរក្សាដដែល គឺគិតជាឆ្នាំ។ លទ្ធផលទទួលបានបង្ហាញនៅក្នុង**ម៉ូដែលទី២** និង អធិប្បាយសង្ខេបដូចខាងក្រោម:

- ✓ ប៉ារ៉ាម៉ែត្រ (β_{totwrk}) និង លំអៀងស្តង់ដារ (*standard errors*)នៃអញ្ញត្តិចំណាយពេលធ្វើការ (*totwrk*) រវាង**ម៉ូដែលទី១** និង **ម៉ូដែលទី២** នៅរក្សាដដែល។
- ✓ ចំណែក β_0 β_{educ} β_{age} និង លំអៀងស្តង់ដារនៃអញ្ញត្តិការសិក្សា(*educ*) ក៏ដូចជាអញ្ញត្តិអាយុ(*age*)នៃ**ម៉ូដែលទី១**ធំជាង**ម៉ូដែលទី២** ចំនួន៦០ដង។
- ✓ រីឯ $R^2, F-stat,$ និង $Pro.(F-stat.)$ រវាង**ម៉ូដែលទី១** និង **ម៉ូដែលទី២** នៅតែស្មើគ្នា។
- ✓ ចន្លោះជឿជាក់ (*confident intervals*)នៃអញ្ញត្តិចំណាយពេលធ្វើការ(*totwrk*) រវាង**ម៉ូដែលទី១** និង **ម៉ូដែលទី២** នៅរក្សាដដែល។ ប៉ុន្តែ ចំពោះអញ្ញត្តិការសិក្សា(*educ*) និង អញ្ញត្តិអាយុ(*age*) ចន្លោះជឿជាក់មានការប្រែប្រួល អាស្រ័យដោយបម្រែបម្រួលនៃប៉ារ៉ាម៉ែត្រ និងលំអៀងស្តង់ដាររបស់វា នៅក្នុង**ម៉ូដែលទី១** និង **ម៉ូដែលទី២**។

៥.១.១ ឧទ្រង់ឡូការីត

យោងលើ ***ឧទាហរណ៍ទី៥*** ដែលបានពិភាក្សានៅក្នុង***មេរៀនទី៤***

ករណីទី១

យើងបំពាក់ឡូការីតទៅលើអញ្ញត្តិមិនឯករាជ្យជានារ៉ាប់រងអាយុជីវិត($insurance: ins$) និង អញ្ញត្តិឯករាជ្យចំណូលរបស់គ្រួសារ($income$) ជាមួយសមីការដូចខាងក្រោម:

$$\log(ins) = \beta_0 + \beta_1 \log(income) + u, \qquad (\text{៥.៣})$$

ដោយអង្កេតទៅលើចំនួន២០គ្រួសារជាមួយទិន្នន័យ(គិតជាខ្នង់ពាន់ដុល្លា) គេទទួលបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$\widehat{log(ins)} = 1.558 + 0.958\, log(income) \qquad (\text{៥.៤})$$
$$(0.133)\ \ (0.033)$$
$$n = 20, R^2 = 0.9787$$

ផ្អែកលើ(៥.៤) ប្រសិនចំណូលកើនឡើង១% នោះជានាវ៉ាប់រងនឹងកើនឡើង0.៩៥៨%។

ករណីទី២

យើងបំពាក់ឡូការីតទៅលើអញ្ញត្តិមិនឯករាជ្យជានាវ៉ាប់រងអាយុជីវិត$(insurance: ins)$ ប៉ុន្តែមិនបំពាក់ឡូការីតទៅលើអញ្ញត្តិឯករាជ្យចំណូលរបស់គ្រួសារ$(income)$ ជាមួយសមីការដូចខាងក្រោម:

$$\log(ins) = \beta_0 + \beta_1 income + u, \qquad (៥.៥)$$

ករណីនេះ $\%\Delta(ins) = 100.\beta_1$។ ដោយអង្កេតទៅលើចំនួន២០គ្រួសារជាមួយទិន្នន័យ(គិតជាខ្នង់ពាន់ដុល្លា) គេទទួលបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$\widehat{log(ins)} = 4.458 + 0.015\ income \qquad (៥.៦)$$
$$(0.074)\ \ (0.0011)$$
$$n = 20, R^2 = 0.9115$$

ផ្អែកលើ(៥.៦) ប្រសិនចំណូលកើនឡើង១,០០០ដុល្លារ នោះជានាវ៉ាប់រងនឹងកើនឡើង១.៥0%។ ក្នុងការអនុវត្តន៍ លទ្ធផលនៃការប៉ាន់ស្មាននេះអាចមានភាពមិនសមហេតផល។ ហេតុនេះ គេអាចធ្វើការប៉ាន់ស្មានដោយអនុវត្តវិធីសាស្ត្រដូចខាងក្រោម:

$$\%\Delta ins = 100.\left[exp\left(\hat{\beta}_1 \Delta income\right) - 1\right] \qquad (៥.៧)$$

ដូច្នេះ ប្រសិនចំណូលកើនឡើង១,០០០ដុល្លារ (ដោយគិតជាខ្នង់ពាន់ដុល្លា: $\Delta income = 1$) នោះជានាវ៉ាប់រងនឹងកើនឡើង: $\%\Delta ins = 100.[exp(0.015) - 1] = 1.51\%$។

កត់សំគាល់: ករណីទិន្នន័យនៃអញ្ញត្តិមានតំលៃស្មើសូន្យ ឬ អវិជ្ជមាន គេមិនបំពាក់ឡូការីតទៅលើអញ្ញត្តិនោះឡើយ។

៥.១.២ ម៉ូដែលជាមួយអញ្ញត្តិឯករាជ្យលើកជាការេ (Models with Quadratics)

ទម្រង់ដែលអញ្ញត្តិឯករាជ្យលើកជាការេ (quadratic functions) ត្រូវបានគេប្រើប្រាស់ជាញឹកញាប់នៅវិស័យជាច្រើន ជាពិសេសជំនាញសេដ្ឋកិច្ច ដើម្បីបង្ហាញអំពីការដេះឥទ្ធិពលកើនឡើង ឬ ថយចុះ (marginal effects) ដោយលើកយកករណីសាមញ្ញមកបង្ហាញដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u, \qquad (៥.៨)$$

ពីព្រោះការសរសេរម៉ូដែលក្នុងទំរង់បែបនេះ គោលបំណងសំខាន់គឺ*marginal effects* ហេតុនេះ ប៉ារ៉ាម៉ែត្រβ_1 វាមិនវាស់វែងការជះឥទ្ធិពលនៃ xទៅលើ yឡើយ។ សមីការអាចត្រូវបានសរសេរដូចខាងក្រោម:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2, \qquad (\text{៥.៩})$$

បន្ទាប់មក យើងអាចធ្វើការប៉ាន់ស្មានដូចខាងក្រោម:

$$\Delta\hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x, \text{ ដូច្នេះ } \Delta\hat{y}/\Delta x \approx \hat{\beta}_1 + 2\hat{\beta}_2 x, \qquad (\text{៥.១០})$$

ដូច្នេះ មេគុណ(slope)នៃទំនាក់ទំនងរវាងx និង y គឺផ្អែកទៅលើតំលៃនៃx និង មេគុណដែលបានប៉ាន់ស្មានគឺ $\hat{\beta}_1 + 2\hat{\beta}_2 x$។

យោងលើ***ឧទាហរណ៍ទី៥*** ដែលបានពិភាក្សានៅក្នុង***មេរៀនទី៤*** ដោយអង្កេតទៅលើចំនួន២០គ្រួសារជាមួយទិន្នន័យ(គិតជាខ្នង់ពាន់ដុល្លា) គេទទួលបានសមីការដូចខាងក្រោម:

$$\widehat{ins} = 23.278 + 3.336\, income + 0.0037 income^2 \qquad (\text{៥.១១})$$
$$(14.679)\ (0.437) \qquad (0.0028)$$
$$n = 20, R^2 = 0.9865$$

ឧបមាថា t_{income^2}មានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត៥% ជាមួយt_{income}ក៏មានសារៈសំខាន់ជាស្ថិតិ ប្រសិនចំណូលកើនឡើងពី1,000ដុល្លារ ទៅ 2,000ដុល្លារ គេអាចប៉ាន់ស្មានថា ធានារ៉ាប់រងអាយុជីវិតក៏នឹងកើនឡើងប្រហែល $3.336 + 2(0.0037)(1) \approx 3.343$ ឬ 3,343 ដុល្លារ។

៥.១.៣ ម៉ូដែលជាមួយលក្ខខណ្ឌអន្តរកម្ម(Interaction Terms)

នៅពេលទំនាក់ទំនងរវាងអញ្ញត្តិឯករាជ្យx និង អញ្ញត្តិមិនឯករាជ្យ y អាស្រ័យដោយបម្រែបម្រួលតម្លៃនៃអញ្ញត្តិ M យើងអាចបង្កើតរ៉ឺគ្រេស្យុងលីនេអ៊ែរ(multiple linear regression) ជាមួយលក្ខខណ្ឌអន្តរកម្ម ដោយដាក់បញ្ចូលxM ចូលទៅក្នុងម៉ូដែល។ ដូច្នេះ គេសរសេរបានម៉ូដែលដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x + \beta_2 M + \beta_3 xM + u, \qquad (\text{៥.១២})$$

ការជះឥទ្ធិពលដោយផ្នែក*(*partial effect)នៃx ទៅលើ y គឺ:

$$\Delta y/\Delta x = \beta_1 + \beta_3 M \qquad (\text{៥.១៣})$$

ប្រសិន $\beta_3 > 0$ បន្ទាប់មក(៥.១៣)បញ្ជាកថា xកើនឡើង នោះវាធ្វើឲ្យ yកាន់តែកើនឡើង ក្នុងពេលដែលតំលៃMកើនឡើង។ នេះគឺជាការជះឥទ្ធិពលអន្តរកម្មរវាងx និង M។ សរុបមក ការជះឥទ្ធិពលនៃx ទៅលើ y យើងត្រូវតែវាយតំលៃ(៥.១៣)ជាមួយតំលៃM ដោយគិតជាតំលៃមធ្យម(mean value)។

ដូចបានពិភាក្សា β_1 គឺជាការជះឥទ្ធិពលដោយផ្នែក(partial effect)នៃx ទៅលើ y ខណៈពេល $M = 0$។ យើងអាចបង្កើតប៉ារ៉ាម៉ែត្រថ្មីសំរាប់ម៉ូដែលដូចខាងក្រោម:

$$y = \alpha_0 + \delta_1 x + \delta_2 M + \beta_3 (x - \mu_1)(M - \mu_2) + u,$$

ដែលμ_1 គឺជាមធ្យមប៉ូពុយឡាស្យុង(population mean)នៃ x និង μ_2 គឺជាមធ្យមប៉ូពុយឡាស្យុង(population mean)នៃ M។ δ_1គឺជា partial effectនៃx ទៅលើ y ជាមួយតំលៃមធ្យម(mean value)នៃM។ ដូច្នេះ គេទទួលបាន $\delta_1 = \beta_1 + \beta_3 \mu_2$។

ឧទាហរណ៍ ឧបមាថាមានក្រុមហ៊ុនមួយចំនួនដំណើរការលក់គ្រឿងបរិភោគសំរាប់អតិថិជន។ ក្នុងនាមជាអ្នកស្រាវជ្រាវចង់ដឹងអំពីកត្តាជះឥទ្ធិលទៅលើចំណូល(sales)របស់ក្រុមហ៊ុនក្នុងតំបន់មួយ ដោយបានធ្វើការអង្កេតក្រុមហ៊ុនចំនួន៣២។ យោងលើរបកគំហើញពីមុន និង ទ្រឹស្តីផ្សេងៗ កត់សំគាល់ឃើញថា វាមានកត្តាសំខាន់ៗដែលទាក់ទងនឹងបំរែបំរួលចំណូល ដូចជា ថ្លៃទំនិញ(price)គិតជាមធ្យម ចំណាយផ្សាយពាណិជ្ជកម្ម(advertising គិតខ្ទង់ពាន់) ចំនួនហាង(stores) ចំនួននិយោជិកផ្នែកលក់ (Staffs) គុណភាពទំនិញ(quality) ចំណង់ចំណូលចិត្ត(satisfaction) កំរិតជីវភាព(living standard)។ល។ ដោយផ្តោតតែទៅលើថ្លៃទំនិញ និង ចំណាយផ្សាយពាណិជ្ជកម្ម យើងសរសេរបានម៉ូដែលដូចខាងក្រោម:

$$Sales = \beta_0 + \beta_1 Price + \beta_2 Adv + \beta_3 Price * Adv + u,$$

ដោយប្រើប្រាស់ទិន្នន័យខាងលើ គេទទួលបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

Obs.	Sales	Price	Adv	Stores	Staffs	log(sales)	log(price)	Log(adv)
1	1,215	36	24	12	129	7.102499356	3.583518938	3.1780538
2	2,189	28	45	11	239	7.691200098	3.33220451	3.8066625
3	6,765	10	54	21	138	8.819517541	2.302585093	3.988984
4	4,231	3	21	78	179	8.350193651	1.098612289	3.0445224
5	3,453	6	32	49	122	8.146998697	1.791759469	3.4657359
6	2,452	8	45	45	210	7.804659297	2.079441542	3.8066625

7	3,136	21	12	67	134	8.050703381	3.044522438	2.4849066
8	5,238	12	32	98	127	8.563695025	2.48490665	3.4657359
9	6,458	9	29	12	134	8.773074951	2.197224577	3.3672958
10	1,234	34	17	98	132	7.118016204	3.526360525	2.8332133
11	2,231	24	23	67	123	7.710205194	3.17805383	3.1354942
12	3,854	11	65	43	245	8.256866849	2.397895273	4.1743873
13	2,436	22	32	21	189	7.798112629	3.091042453	3.4657359
14	7,236	10	68	34	165	8.886823847	2.302585093	4.2195077
15	4,231	32	32	67	245	8.350193651	3.465735903	3.4657359
16	7,329	9	43	89	178	8.89959436	2.197224577	3.7612001
17	4,231	14	42	54	134	8.350193651	2.63905733	3.7376696
18	2,544	7	31	39	156	7.841492924	1.945910149	3.4339872
19	9,453	4	87	78	252	9.15408743	1.386294361	4.4659081
20	6,454	8	45	65	178	8.772455373	2.079441542	3.8066625
21	5,324	15	46	78	198	8.57998018	2.708050201	3.8286414
22	7,564	8	27	49	295	8.93115543	2.079441542	3.2958369
23	2,439	43	33	48	190	7.799343398	3.761200116	3.4965076
24	2,457	13	23	68	139	7.806696373	2.564949357	3.1354942
25	1,435	42	15	92	154	7.268920128	3.737669618	2.7080502
26	7,345	12	43	34	165	8.901775089	2.48490665	3.7612001
27	4,234	8	65	56	189	8.350902452	2.079441542	4.1743873
28	5,232	22	34	78	145	8.562548893	3.091042453	3.5263605
29	6,657	21	47	12	187	8.803424212	3.044522438	3.8501476
30	3,453	12	24	19	156	8.146998697	2.48490665	3.1780538
31	2,440	23	23	56	156	7.799753318	3.135494216	3.1354942
32	6,458	14	68	78	194	8.773074951	2.63905733	4.2195077

$$Sales = 3{,}626.53 - 81.815Price + 55.142Adv + 0.109Price * Adv \qquad (๕.๑๔)$$
$$(1{,}263.05)\ (64.092) \qquad (29.350) \qquad (2.051)$$
$$n = 32,\ R^2 = 0.5122$$

យើងត្រូវតែបកស្រាយសមីការនេះដោយប្រយ័ត្នប្រយែង។ បើគ្រាន់តែមើលទៅលើមេគុណនៃថ្លៃទំនិញ(price) នោះយើងយើញថា ថ្លៃទំនិញជះឥទ្ធិពលជាអវិជ្ជមានទៅលើចំណូលរបស់ក្រុមហ៊ុន។ ការសន្និដ្ឋានបែបនេះមិនត្រឹមត្រូវឡើយ។ យើងត្រូវតែប្រយ័ត្នប្រយែង ដោយមិនត្រូវមើលទៅលើតំលៃ

ប៉ាន់ស្មាននៃβ_1 និង β_3ដាច់ដោយឡែកពីគ្នាឡើយ ដោយសន្និដ្ឋានថា យើងមិនអាចបដិសេធ $H_0: \beta_1 = 0, \beta_3 = 0$ (ពីព្រោះ *t-stat* មិនមានសារៈសំខាន់ជាស្ថិតិ)។ ការពិត ***p-value*** សំរាប់***F-test*** នៃសម្មតិកម្មរួម(joint hypothesis) គឺ $0.000138 < 5\%$ ដូច្នេះប្រាកដណាស់ យើងបដិសេធH_0 នៅឯកំរិត5%។

តើយើងគួរតែប៉ាន់ស្មានការជះឥទ្ធិពលដោយផ្នែក(partial effect)នៃថ្លៃទំនិញទៅលើចំណូលរបស់ក្រុមហ៊ុនយ៉ាងដូចម្តេច? យើងត្រូវតែភ្ជាប់ជាមួយចំណាយផ្សាយពាណិជ្ជកម្ម ដើម្បីទទួលបានការជះឥទ្ធិពលដោយផ្នែក(partial effect)នៃថ្លៃទំនិញទៅលើចំណូល។ តំលៃមធ្យម(mean value)នៃការផ្សាយពាណិជ្ជកម្មគឺ 38.334 ហេតុនេះ ការជះឥទ្ធិពលដោយផ្នែក(partial effect)នៃថ្លៃទំនិញទៅលើចំណូលគឺ -81.815 + 0.109(38.334) ≈ -77.63។ អាស្រ័យដោយថ្លៃទំនិញវាស់វែងជាដុល្លារ និងចំណូលក៏វាស់វែងជាដុល្លារជាមួយតំលៃគិតជាខ្នង់ពាន់ ប្រសិនថ្លៃទំនិញកើនមួយដុល្លារ នោះការថយចុះនៃចំណូលគឺ ប្រហែល77,630ដុល្លារ។

តើតម្លៃប៉ាន់ស្មាន -77.63 មានសារៈសំខាន់ជាស្ថិតិខុសពីសូន្យដែរឬទេ? យើងត្រូវដំណើរការរ៉ឺគ្រ្តេស្យុងម្តងទៀត ដោយជំនួស $Price * Adv$ ជាមួយ $Price * (Adv - 38.334)$។ បន្ទាប់មកគេសរសេរបានសមីការដូចខាងក្រោម:

$$Sales = 3{,}626.53 - 77.63\ Price + 55.14\ Adv + 0.11\ Price * (Adv - 38.334) \quad (\text{៥.១៥})$$
$$(1{,}263.05)\ \ (36.62)\qquad (29.34)\qquad (2.05)$$
$$n = 32, R^2 = 0.5122$$

យោងលើ(៥.១៥) គេទទួលបាន$t_{price} = 77.63/36.62 = 2.12 >$ តំលៃ$C = 2.045$ នៅឯកំរិត 5%។ ដូច្នេះ ថ្លៃទំនិញជះឥទ្ធិពលជាអវិជ្ជមានទៅលើចំណូលរបស់ក្រុមហ៊ុនជាមួយតំលៃមធ្យមនៃការផ្សាយពាណិជ្ជកម្ម នៅឯកំរិតសារៈសំខាន់ជាស្ថិតិ5%។

៥.២ គណនាតម្លៃR^2 (Goodness-of-Fit)

សំរាប់ម៉ូដែលលីនេរអ៊ែរ(linear model)មិនមានលក្ខខ័ណ្ឌដែលកំណត់ថា **តម្លៃR^2** ត្រូវតែធំជាងចំនួនណាមួយឡើយ។ វាគ្រាន់តែប៉ាន់ស្មានថា តើបម្រែបម្រួលនៃអញ្ញត្តិy ត្រូវបានពន្យល់ដោយអញ្ញត្តិ $x_1, x_2, \ldots x_k$ចំនួនប៉ុន្មាន? យើងបានឃើញមកហើយមានម៉ូដែលរ៉ឺគ្រ្តេស្យុងខ្លះតម្លៃR^2របស់វាតូច។

តម្លៃR^2តូចបញ្ជាក់ថា លំអៀងវ៉ារីយ៉ង់(error variance)គឺមានតម្លៃធំទាក់ទងនឹងវ៉ារីយ៉ង់នៃy ដែលមានន័យថា យើងប្រហែលជាលំបាកក្នុងការប៉ាន់ស្មានβ_jឲ្យបានច្បាស់លាស់។ ប្រសិនចំនួនសំណាក(sample size)គ្រប់គ្រាន់ ប្រហែលជាយើងអាចប៉ាន់ស្មានការជះឥទ្ធិពលបានច្បាស់លាស់ទោះបីជាលំអៀងវ៉ារីយ៉ង់(error variance)ធំក៏ដោយ។

៥.២.១ គណនាតម្លៃ$Adjusted\ R^2$

ក្នុងការអនុវត្តន៍ $Adjusted\ R^2$ត្រូវបានគេលើកយកមកបង្ហាញជាញឹកញាប់ និង មានរូបមន្តដូចខាងក្រោម:

$$\bar{R}^2 = 1 - [SSR/(n-k-1)]/[SST/(n-1)]$$

$$= 1 - \hat{\sigma}^2/[SST/(n-1)] \qquad (\text{៥.១៦})$$

ដែល $\hat{\sigma}^2 = \left(\sum_{i=1}^{n} \hat{u}_i^2\right) \Big/ (n-k-1) = SSR/(n-k-1)$

$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik},$

$Explained\ Sum\ of\ Squares(SSE) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

$Residual\ Sum\ of\ Squares(SSR) = \sum_{i=1}^{n} \hat{u}_i^2$

$Total\ Sum\ of\ Squares(SST) = \sum_{i=1}^{n} (y_i - \bar{y})^2 = SSE + SSR$

ផ្អែកលើទិន្នន័យខាងលើដែលទាក់ទងនឹងកត្តាជះឥទ្ធិពលទៅលើបំរែបំរួលចំណូល ដូចជា ថ្លៃទំនិញ (price)គិតជាមធ្យម និង ចំណាយផ្សាយពាណិជ្ជកម្ម(advertising គិតខ្នង់ពាន់) យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

Dependent Variable: SALES

Observations: 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3590.308	1045.557	3.433873	0.0018
PRICE	-78.78646	28.98685	-2.718007	0.0110
ADV	56.36198	18.01878	3.127957	0.0040

R^2	0.512181	S.E. of regression	1595.718
Adjusted R^2	***0.478539***	Sum squared residual	73843164
F-statistic	15.22416	Prob(F-statistic)	0.000030

ទំនាក់ទនងរវាង $Adjusted\ R^2$ និង R^2 គឺ

$$\bar{R}^2 = 1 - (1 - R^2)(n-1)/(n-k-1) \qquad (\text{៥.១៧})$$

យោងលើ(៥.១៧) $\bar{R}^2 = 1 - (1 - 0.512181)(32-1)/(32-2-1) = \mathbf{0.478538}$

៥.២.២ ការប្រើប្រាស់$Adjusted\ R^2$ដើម្បីជ្រើសរើសរវាង$Nonnested\ Models$

ក្នុងករណីខ្លះ យើងចង់ជ្រើសរើសម៉ូដែលមួយក្នុងនោះមានអញ្ញត្តិឯករាជ្យដែលមិនមានសារៈសំខាន់ជាស្ថិតិ តាមរយៈ $adjusted\ R^2$។ ផ្អែកលើទិន្នន័យខាងលើ យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

តារាង ៥.២ *ការជះឥទ្ធិពលនៃចំណូលក្រុមហ៊ុន*

អញ្ញត្តិមិនឯករាជ្យ អញ្ញត្តិឯករាជ្យ	(១) *Sales*	(២) *Sales*
price.....	-79.062	-81.136
	(29.239)	(29.339)
adv.....	57.274	49.988
	(18.242)	(20.264)
stores.....	6.912	
	(10.896)	
staffs.....		5.160
		(7.439)
Intercept.....	3,188.873	2,974.667
	(1,237.294)	(1,385.920)
Observation.....	32	32
Adjusted R^2.....	0.4676	0.4690

ម៉ូដែលទាំងពីរនេះ(តារាង ៥.២)គឺជា***nonnested models*** ពីព្រោះ ម៉ូដែលមួយគឺវាមិនមែនជាករណីពិសេសនៃម៉ូដែលដទៃទៀត ក្នុងនោះទាំងអញ្ញត្តិចំនួនហាង(stores) ក៏ដូចជាចំនួននិយោជិកផ្នែកលក់ (Staffs) មិនមានសារៈសំខាន់ជាស្ថិតិ($individual\ t-test$)។ ចំណែក*F*-stat. ដែលបានសិក្សានៅក្នុងមេរៀនទី៤ អនុញ្ញាតឲ្យយើងធ្វើតេស្ត***nested models***: ម៉ូដែលមួយ(the restricted model)គឺជាករណីពិសេសនៃម៉ូដែលដទៃទៀត(the unrestricted model)។

នៅក្នុងរ៉ឺស៊ុីព្រ្ចេស្សុងចំណូលក្រុមហ៊ុន $\bar{R}^2$សំរាប់រ៉ឺស៊ុីព្រ្ចេស្សុងដែលមានអញ្ញត្តិចំនួនហាង(stores)គឺ 0.4676 និង $\bar{R}^2$សំរាប់រ៉ឺស៊ុីព្រ្ចេស្សុងដែលមានអញ្ញត្តិចំនួននិយោជិកផ្នែកលក់ (Staffs)គឺ 0.4690។ ដូច្នេះ ផ្អែកលើadjusted R^2 ម៉ូដែលជាមួយអញ្ញត្តិចំនួននិយោជិកផ្នែកលក់ (Staffs) ប្រសើរជាងម៉ូដែលជាមួយអញ្ញត្តិចំនួនហាង(stores)បន្តិចបន្តួចបំផុត។

ក្នុងការជ្រើសរើសម៉ូដែលដោយផ្អែកលើការប្រៀបធៀប$\bar{R}^2$ វាមានសារៈសំខាន់ទាក់ទងនឹងទម្រង់ផ្សេងៗគ្នា(different functional forms)នៃម៉ូដែល។ ផ្អែកលើទិន្នន័យខាងលើ យើងទទួលបាន និងសិក្សាទៅលើលទ្ធផលនៃម៉ូដែលទាំងពីរដូចខាងក្រោម:

តារាង ៥.៣ *ការជះឥទ្ធិពលនៃចំណូលក្រុមហ៊ុន*

អញ្ញត្តិមិនឯករាជ្យ អញ្ញត្តិឯករាជ្យ	(១) log(sales)	(២) log(sales)
log(price).....	-0.496	
	(0.124)	
price.....		-0.021
		(0.029)
price^2.....		-0.0003
		(0.0007)
Intercept.....	9.556	8.724
	(0.336)	(0.267)
Observation.....	32	32
Adjusted R^2.....	0.3254	0.4027

នៅក្នុង**តារាង៥.៣** ម៉ូដែលទី១ យើងបំពាក់ឡូការីតទៅលើទាំងអញ្ញត្តិចំណូលក្រុមហ៊ុន និង ថ្លៃទំនិញ។ រីឯ ម៉ូដែលទី២ យើងបំពាក់ឡូការីតទៅលើអញ្ញត្តិចំណូលក្រុមហ៊ុន និង ថ្លៃទំនិញត្រូវបានលើកជាការ។ ជាលទ្ធផល ម៉ូដែលទី១មាន$\bar{R}^2$ស្មើនឹង0.3254 តូចជាង$\bar{R}^2$របស់ម៉ូដែលទី២ដែលស្មើនឹង0.4027។ ដូច្នេះ ទម្រង់នៃម៉ូដែលទី២ដែលលើកអញ្ញត្តិឯករាជ្យជាការមានភាពប្រសើរជាងទម្រង់ម៉ូដែលទី១។

វាមានលក្ខខណ្ឌជាមួយការប្រើប្រាស់$\bar{R}^2$សំរាប់ជ្រើសរើស***nonnested models*** គឺថា យើងមិនអាចប្រើប្រាស់វាក្នុងករណីដែលអញ្ញត្តិមិនឯករាជ្យមានទម្រង់ផ្សេងគ្នា(different functional forms for the dependent variable)។

៥.៣ ការព្យាករណ៍ និង វិភាគតម្លៃសំណល់ (Predictions and Residual Analysis)

លទ្ធផលនៃការព្យាករណ៍ទទួលបានពីOLS estimatorsមានសារៈសំខាន់សំរាប់អនុវត្តន៍ និងវាប្រែប្រួលអាស្រ័យនឹងបំរែបំរួលសំណាក (sampling variation)។ យើងបង្ហាញបន្ថែមថា តើចន្លោះជឿជាក់ (confidence intervals) សំរាប់ការព្យាករណ៍មួយត្រូវបានកំណត់ពីបន្ទាត់ត្រង់រ៉ឺគ្រេស្យុងយ៉ាងដូចម្តេច?

៥.៣.១ ចន្លោះជឿជាក់សំរាប់ការព្យាករណ៍

ឧបមាថា យើងទទួលបានសមីការសំរាប់ប៉ាន់ស្មានដូចខាងក្រោម:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \qquad (៥.១៧)$$

នៅពេលយើងបញ្ចូលតម្លៃជាក់ស្តែងនៃបណ្តាអញ្ញត្តិឯករាជ្យ(x_j) យើងទទួលបានការព្យាករណ៍សំរាប់អញ្ញត្តិមិនឯករាជ្យ(y) ដែលជាការប៉ាន់ស្មាននៃតម្លៃរំពឹងទុក (expected value)នៃy ផ្អែកលើបណ្តាតម្លៃនៃអញ្ញត្តិឯករាជ្យ(x_j)។ តាង $a_1, a_2, \dots, a_k$ ជាតម្លៃជាក់ស្តែងនៃបណ្តាអញ្ញត្តិឯករាជ្យ(x_j) ជាមួយប៉ារ៉ាម៉ែត្រដែលយើងចង់ប៉ាន់ស្មានគឺ:

$$\theta_0 = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k \qquad (៥.១៨)$$

$$= E(y|x_1 = a_1, x_2 = a_2, \dots, x_k = a_k)$$

ការប៉ាន់ស្មាននៃθ_0 គឺ: $\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 a_1 + \hat{\beta}_2 a_2 + \cdots + \hat{\beta}_k x_k \qquad (៥.១៩)$

ដើម្បីទទួលបានចន្លោះជឿជាក់(confidence intervals)នៃθ_0 យើងត្រូវការលំអៀងស្តង់ដារ(standard error)សំរាប់$\hat{\theta}_0$។ បន្ទាប់មក យើងអាចបង្កើតបានចន្លោះជឿជាក់នៅឯ95% ដោយប្រើប្រាស់គោលការណ៍(the rule of thumb): $\widehat{\boldsymbol{\theta}}_{\mathbf{0}} \pm \mathbf{2} * \boldsymbol{se}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}})$ ជាមួយdegree of freedom (df)។

តើយើងទទួលបានលំអៀងស្តង់ដារ(standard error)សំរាប់$\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$យ៉ាងដូចម្តេច? ដើម្បីទទួលបានវាយើងត្រូវការលំអៀងស្តង់ដារ(standard error)ដែលកើតឡើងពីបញ្ចូលគ្នា(linear combination)មួយនៃOLS estimator។ ផ្អែកលើ(៥.១៨) គេអាចសរសេរបាន $\beta_0 = \theta_0 - \beta_1 a_1 - \cdots \beta_k a_k$ បញ្ចូលជាមួយសមីការ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ នោះគេទទួលបាន:

$$y = \theta_0 + \beta_1(x_1 - a_1) + \beta_2(x_2 - a_2) + \cdots + \beta_k(x_k - a_k) + u, \qquad (៥.២០)$$

យើងដកតម្លៃa_j ពី x_j និង បន្ទាប់មកដំណើការរ៉ឺគ្រេស្យុងដូចខាងក្រោម:

$$y_i \; on \; (x_{i1} - a_1), \dots, (x_{ik} - a_k), i = 1, 2, \dots, n. \qquad (៥.២១)$$

តម្លៃដែលព្យាករណ៍ក្នុង(៥.១៩) និង លំអៀងស្តង់ដា(standard error)របស់វា ត្រូវបានទទួលពីអ៊ីនធ័រសិប (intercept or constant)ក្នុងរ៉ឺហ្គ្រេស្យុង(៥.២១)។

ផ្អែកលើទិន្នន័យខាងលើ យើងកំណត់ចន្លោះជឿជាក់នៃចំណូលរបស់ក្រុមហ៊ុន ដោយផ្អែកលើកត្តាផ្សេងៗ ដូចជា ថ្លៃទំនិញ(price) និង ចំណាយផ្សាយពាណិជ្ជកម្ម(advertising គិតខ្នង់ពាន់) ដោយដំណើរការជាបន្តបន្ទាប់ដូចខាងក្រោម:

$$sales = 3{,}600.349 - 79.145\ price + 56.246\ adv \qquad (៥.២២)$$
$$(1{,}042.765)\ \ (28.936)\qquad (17.982)$$
$$n = 32, R^2 = 0.5122, \bar{R}^2 = 0.4785$$

នៅពេលថ្លៃទំនិញ(price) = $22 និង ចំណាយផ្សាយពាណិជ្ជកម្ម(advertising គិតខ្នង់ពាន់) = $24 តើរំពឹងថា ចំណូលក្រុមហ៊ុនស្មើប៉ុន្មាន? ករណីនេះងាយស្រួលក្នុងការគណនា។ ប៉ុន្តែ យើងមិនអាចប្រើប្រាស់(៥.២២)ដោយផ្ទាល់ ដើម្បីគណនាចន្លោះជឿជាក់ (confidence intervals)សំរាប់ចំណូលក្រុមហ៊ុនដែលបានរំពឹងទុកឡើយ។

វិធីសាស្ត្រដើម្បីគណនាចន្លោះជឿជាក់គឺថា ត្រូវកំណត់បណ្តាអញ្ញត្តិឯករាជ្យថ្មី: $pricen = price - 22$ និង $advn = adv - 24$។ នៅពេលអនុវត្តជាមួយបណ្តាអញ្ញត្តិឯករាជ្យថ្មី យើងក៏ទទួលបានរ៉ឺហ្គ្រេស្យុងថ្មីដូចខាងក្រោម:

$$sales = 3{,}209.077 - 79.145\ pricen + 56.246\ advn \qquad (៥.២៣)$$
$$(367.128)\ \ (28.936)\qquad (17.982)$$
$$n = 32, R^2 = 0.5122, \bar{R}^2 = 0.4785$$

ភាពខុសគ្នារវាងរ៉ឺហ្គ្រេស្យុង(៥.២៣) និង (៥.២២) គឺ អ៊ីនធ័រសិប (intercept) ជាមួយមេគុណ (β_0) និង លំអៀងស្តង់ដា(standard error)របស់វាដែលយើងចង់បាន។ រីឯ មេគុណ(β_j) និងលំអៀងស្តង់ដារបស់អញ្ញត្តិឯករាជ្យទាំងពីរ $R-$ squared ក្នុងរ៉ឺហ្គ្រេស្យុងទាំងពីរស្មើគ្នារៀងខ្លួន។ នេះបញ្ជាក់ថា បម្លាស់ប្តូរពី(៥.២២)ទៅជា (៥.២៣) គឺ ត្រឹមត្រូវ។

ផ្អែកលើលទ្ធផលខាងលើ ដោយជ្រើសរើសយកចន្លោះជឿជាក់នៅឯ95% យើងទទួលបានចន្លោះជឿជាក់សំរាប់ចំណូលក្រុមហ៊ុនដែលបានរំពឹងទុកគឺ

$$3{,}209.077 \pm 2.045(367.128) \approx 2{,}458.30\ to\ 3{,}959.85$$

វិធីសាស្ត្រខាងលើនេះ កំណត់ចន្លោះជឿជាក់ទាក់ទងការប៉ាន់ស្មាននៃ$E(y|x_1, x_2, \dots, x_k)$ចំពោះតម្លៃណាមួយនៃបណ្តាអញ្ញត្តិឯករាជ្យ។ ម៉្យាងទៀត ចន្លោះជឿជាក់ត្រូវបានកំណត់សំរាប់តម្លៃមធ្យមនៃy ទាក់ទងនឹងសំណាកទាំងមូលជាមួយសំណុំនៃអញ្ញត្តិឯករាជ្យណាមួយ។

ក្នុងការបង្កើតចន្លោះជឿជាក់សំរាប់លទ្ធផលដែលមិនបានដឹង(unknown outcomes)នៃy យើងត្រូវតែពិនិត្យមើលវ៉ារីយ៉ង់នៃunobserved errorនៃកត្តាដទៃទៀត(unobserved factors)ដែលជះឥទ្ធិពលទៅលើy។

៥.៣.២ ការព្យាករណ៍y ករណីអញ្ញត្តិមិនឯករាជ្យគឺ log(y)

ក្នុងការសិក្សា ការបំលែងអញ្ញត្តិជាឡូការីត(log transformation) ត្រូវបានគេប្រើប្រាស់ជាញឹកញាប់។ យើងនឹងបង្ហាញអំពីការព្យាករណ៍ y នៅពេល log(y)គឺអញ្ញត្តិមិនឯករាជ្យ (dependent variable)។

ដើម្បីទទួលបាននូវលទ្ធផលនៃការព្យាករណ៍ យើងកំណត់ *logy* = log(y) គឺថា log នៃ y ត្រូវបានព្យាករណ៍ទាក់ទងម៉ូដែល:

$$logy = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៥.២៤)$$

យោងលើOLS estimators, គេអាចសរសេរបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$\widehat{logy} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k, \qquad (៥.២៥)$$

ប្រសិន(៥.២៤)ឆ្លើយតបតាមលក្ខខណ្ឌទី១ ដល់ ទី៦ របស់ម៉ូដែលលីនេអ៊ែរ (classical linear model) នោះគេទទួលបាន:

$$E(y|x) = exp(\sigma^2/2) * exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k),$$

x គឺជាអញ្ញត្តិឯករាជ្យ (independent variable) និង σ^2ជាវ៉ារីយ៉ង់(variance)នៃ u។ ក្នុងរកណី $\boldsymbol{u \sim Normal(0, \sigma^2)}$ នោះតំលៃរំពឹងទុកនៃ$exp(u)$ គឺជា $exp(\sigma^2/2)$ បន្ទាប់មកទទួលបាន:

$$\hat{y} = exp(\hat{\sigma}^2/2)exp\left(\widehat{logy}\right), \qquad (៥.២៦)$$

លទ្ធផល(៥.២៦)កើតឡើង ក្នុងករណី $\boldsymbol{u \sim Normal(0, \sigma^2)}$ មានន័យថា error term (u)ត្រូវតែជាបំណែងចែកធម្មតា(normality)។ យើងអាចសរសេរបានដូចខាងក្រោម:

$$E(y|x) = \alpha_0 exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k), \qquad (៥.២៧)$$

α_0 ជាតំលៃរំពឹងទុកនៃ $exp(u)$។ ដូច្នេះ តំលៃព្យាករណ៍នៃy គឺ:

$$\hat{y} = \hat{\alpha}_0 exp\left(\widehat{logy}\right), \qquad (៥.២៨)$$

ជាសរុប ដើម្បីព្យាករណ៍y នៅពេលអញ្ញត្តិមិនឯករាជ្យគឺ log(y) គេត្រូវអនុវត្តតាមវិធីសាស្ត្រជាបន្តបន្ទាប់ដូចខាងក្រោម:

1. ស្វែងរកតម្លៃ $\widehat{logy}$ តាមរយៈ $logy = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$
2. សំរាប់ការអង្កេតនីមួយៗ (each observation i) ចូរកំណត់ $\widehat{m}_i = exp\left(\widehat{log}_i\right),$

3. អនុវត្តរ៉ឺគ្រ្សែសុងy ទៅលើអញ្ញត្តិ $\hat{m}$ ដោយមិនមានអ៊ីនធ័រសិប (without an intercept)។ មេគុណនៃអញ្ញត្តិ $\hat{m}$ គឺជាតម្លៃប៉ាន់ស្មាននៃ$\hat{\alpha}_0$។
4. ដោយអនុវត្តន៍តាម(៥.២៨) នោះយើងនឹងទទួលបានតម្លៃប៉ាន់ស្មាន $\hat{y}$។

លើកយកឧទាហរណ៍ខាងលើ ដែលសិក្សាទៅលើកត្តាសំខាន់ៗដែលទាក់ទងនឹងបំរែបំរួលចំណូល (sales) ដូចជា ថ្លៃទំនិញ(price) និងចំណាយផ្សាយពាណិជ្ជកម្ម(advertising)ជាមួយការបំពាក់ឡូការីតទៅលើអញ្ញត្តិទាំងឡាយនោះ ដោយសរសេរបានជាម៉ូដែលដូចខាងក្រោម:

$$\log(sales) = \beta_0 + \beta_1 \log(price) + \beta_2 \log(adv) + u,$$

ផ្អែកលើទិន្នន័យខាងលើ យើងទទួលបានសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$\log \widehat{(sales)} = 7.207 - 0.343 \log(price) + 0.550 \log (adv)$$

$$(0.773)\ (0.118) \qquad (0.168) \qquad (៥.២៩)$$

$$n = 32, R^2 = 0.5237,$$

បន្ទាប់មក យើងទទួលបាន $\hat{m}_i = exp(\widehat{lsales_i})$ សំរាប់ការអង្កេតនីមួយៗនៃសំណាក (sample)។ ដោយអនុវត្តន៍រ៉ឺគ្រ្សែសុង sales ទៅលើ $\hat{m}$ ដោយមិនមានអ៊ីនធ័រសិប (without an intercept) យើងទទួលបាន $\hat{\alpha}_0 \approx 1.041$។

យើងអាចប្រើប្រាស់ $\hat{\alpha}_0 \approx 1.041$ ជាមួយ(៥.២៩) ដើម្បីព្យាករណ៍ចំណូលទាក់ទងនឹងតម្លៃនីមួយៗនៃទំនិញ និង ការផ្សាយពាណិជ្ជកម្ម។ ឧបមាថា ថ្លៃទំនិញជាមធ្យមគឺ 24ដុល្លារក្នុង1ឯកតា និង ចំណាយផ្សាយពាណិជ្ជកម្ម (គិតខ្ចង់ពាន់)គឺ 34ដុល្លារ។ ដូច្នះ ការព្យាករណ៍សំរាប់ $\log \widehat{(sales)}$គឺ $7.207 - 0.343 \log(24) + 0.550 \log (34) \approx 7.576$។ ហេតុនេះ ចំណូលដែលព្យាករណ៍គឺ $1.041 * \exp (7.576) \approx 2{,}030.592$ ឬ 2,030,592ដុល្លារ។

លំហាត់

៥.១ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ យើងទទួលបានសមីការ(OLS)
ដូចខាងក្រោម:

$$ROA = 17.7278 + 0.00290\,WC - 0.00000000201WC^2$$
$$(1.5670)\ (0.00133)\qquad (0.00000000126)$$
$$n = 45, R^2 = 0.2138$$

ក. នៅចំណុចណាដែលការដេះឥទ្ធិពលលំអៀង (marginal effect)នៃWCទៅលើROA ក្លាយទៅជាអវិជ្ជមាន?
តើយើងគួររក្សាទុកWC^{2}ក្នុងម៉ូដែលដែរឬទេ? ចូរពន្យល់

គ. WCmilគឺជាWCត្រូវបានដែលវាស់វែងជាខ្នង់លានដុល្លារ:
WCmil = WC/1,000,000។ ចូរសរសេរសមីការប៉ាន់ស្មានជាមួយWCmil និង WCmil^{2}ជាអញ្ញត្តិឯករាជ្យម្តងទៀត។

ឃ. សំរាប់របាយការណ៍លទ្ធផល តើសមីការមួយណាដែលអ្នកចូលចិត្ត?

៥.២ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ ដើម្បីប៉ាន់ស្មានសមីការលទ្ធភាពនៃប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 \log(SALES) + \beta_2 \log(WC) + \beta_3 TAT + u$$

ក. ចូររាយការណ៍លទ្ធផលនៃម៉ូដែល។ បកស្រាយβ_1 β_2 និង β_3។ តើβ_1 β_2 និង β_3មានសារៈសំខាន់ជាស្ថិតិដែរឬទេ?

ខ. ចូរធ្វើតេស្តសម្មតិកម្មដើម$H_0: \beta_{log(WC)} \geq 0.1$ និង
សម្មតិកម្មផ្ទុយ$H_1: \beta_{log(WC)} < 0.1$។

គ. ចូរកំណត់ជឿជាក់នៅឯ៩៥%សំរាប់$\beta_{\log(SALES)}$ និង β_{TAT}។

ឃ. ចូររាយការណ៍លទ្ធផលនៃម៉ូដែលឡើងវិញដោយបន្ថែម$log(SALES) \times log(WC)$ $[log(SALES)]^2$ និង $[log(WC)]^2$។

ង. សំរាប់ម៉ូដែលថ្មី ចូរបង្ហាញអំពីការដេះឥទ្ធិពលលំអៀង(marginal effect)នៃ $log(SALES)$ទៅលើ ROA និង $log(WC)$ទៅលើ ROA។

ច. ចូរប៉ាន់ស្មានឥទ្ធិពលលំអៀងនៃ$log(WC)$ទៅលើ ROAសំរាប់ក្រុមហ៊ុនពីរ។ ក្រុមហ៊ុនទី១ មាន $log(SALES) = 10\%, log(WC) = 20\%$ និង ក្រុមហ៊ុនទី២មាន $log(SALES) = 15\%, log(WC) = 15\%$។

មេរៀនទី ៦
រ៉ឺគ្រ្តេស្យុងពហុៈ ជាមួយពត៌មានបែបគុណភាព

កាលពីមេរៀនមុនៗ រាល់អញ្ញត្តិឯករាជ្យ(independent variables) និង មិនឯករាជ្យ (dependent variable)នៃម៉ូដែលរ៉ឺគ្រ្តេស្យុង(multiple regression model)ជាប្រភេទបរិមាណ (quantitative information) ដូចជា ចំណូល ថ្លៃទំនិញ។ល។ ក្នុងករណីនីមួយៗ រង្វាស់រង្វាល់នៃ អញ្ញត្តិគឺជាពត៌មានដែលមានសារៈសំខាន់។ ក្នុងការស្រាវជ្រាវ(empirical work) កត្តាបែបគុណភាព (qualitative factors) ក៏ត្រូវបានបញ្ចូលក្នុងម៉ូដែលរ៉ឺគ្រ្តេស្យុងផងដែរ ដូចជា ភេទ ឬ ពណ៌សម្បុរនៃ បុគ្គលម្នាក់ៗ ប្រភេទអង្គភាព ទីតាំង ។ល។ ដូច្នេះ នៅក្នុងម៉ូដែលរ៉ឺគ្រ្តេស្យុងនីមួយៗអាចមានទាំង អញ្ញត្តិបែបបរិមាណ និង គុណភាព។ មេរៀននេះ យើងនឹងបញ្ចូលពត៌មានបែបគុណភាពទៅក្នុង ម៉ូដែលរ៉ឺគ្រ្តេស្យុងជាមួយការបង្ហាញអំពីរបៀបបកស្រាយ និង សារៈសំខាន់របស់វាទាក់ទងនឹងការ អនុវត្តន៍។

៦.១ បរិយាយពត៌មានបែបគុណភាព

បណ្តាកត្តាគុណភាពទាក់ទងនឹងទម្រង់នៃពត៌មានជាគូ(binary information) ដូចជា មនុស្ស ប្រុស ឬ ស្រី អ្នកប្រើទូរសព្ទ ឬ មិនប្រើ នៅលីវ ឬ រៀបការ។ រាល់ពត៌មានខាងលើនេះអាចត្រូវបានហៅ ថា អញ្ញត្តិជាគូ(binary variable) ឬ សូន្យ-មួយ (zero-one variables)។ ជាទូទៅ វាត្រូវបានគេហៅ ថា dummy variables។

ក្នុងការកំណត់dummy variables យើងត្រូវតែសំរេចចិត្តថា មួយណាកំណត់តម្លៃស្មើមួយ និង មួយ ណាកំណត់តម្លៃស្មើសូន្យ។ ឧទាហរណ៏ ក្នុងការសិក្សាអំពីកត្តាជះឥទ្ធិលទៅលើចំណូលរបស់ក្រុមហ៊ុន យើងលើកយកក្រុមហ៊ុនពីរប្រភេទ(binary variable) គឺ លក់ដុំ(wholesaler) និង លក់រាយ (retailer) មកបរិយាយ ដោយកំណត់តម្លៃស្មើមួយសំរាប់ក្រុមហ៊ុនលក់ដុំ ចំណែកក្រុមហ៊ុនលក់រាយ កំណត់តម្លៃស្មើសូន្យ។

តារាង ៦.១

Obs.	Sales	Price	Adv	Stores	Firm size	Business	Gender	ROA
1	1,215	36	12	12	1	0	1	12
2	2,189	28	28	11	0	0	1	22
3	6,765	10	42	21	1	1	0	21
4	4,231	3	21	78	1	1	1	32

5	3,453	6	68	49	0	0	1	30
6	2,452	8	45	45	1	0	1	29
7	3,136	21	12	67	0	0	1	12
8	5,238	12	32	98	0	1	0	19
9	6,458	9	29	12	1	1	0	26
10	1,234	34	17	98	1	0	1	10
11	2,231	24	23	67	1	0	1	15
12	3,854	11	29	43	0	0	1	17
13	2,436	22	32	21	1	0	1	13
14	7,236	10	34	34	0	1	0	18
15	4,231	32	10	67	1	1	1	11
16	7,329	9	42	89	1	1	0	21
17	4,231	14	44	54	0	1	1	22
18	2,544	7	31	39	1	0	1	14
19	9,453	4	65	78	0	1	0	30
20	6,454	8	28	65	1	1	0	18
21	5,324	15	36	78	0	1	0	22
22	7,564	8	27	49	1	1	0	23
23	2,439	43	9	48	1	0	1	8
24	2,457	13	23	68	0	0	1	20
25	1,435	42	15	92	1	0	0	17
26	7,345	12	43	34	1	1	0	23
27	4,234	8	32	56	1	1	1	26
28	5,232	22	29	78	0	1	0	19
29	6,657	21	30	12	1	1	0	16
30	3,453	12	35	19	1	0	1	20
31	2,440	23	23	56	0	0	1	14
32	6,458	14	32	78	0	1	0	18

៦.២ អញ្ញត្តិឯករាជ្យជាមួយដាំមីតែមួយ (Single Dummy)

តើយើងបញ្ចូលពត៌មានជាគូ (binary information) ទៅក្នុងម៉ូដែលរ៉ឺហ្គ្រេស្យុងយ៉ាងដូចម្តេច ? ឧទាហរណ៍ ពិចារណាម៉ូដែលនៃអំពីកត្តាជះឥទ្ធិលទៅលើចំណូលរបស់ក្រុមហ៊ុនដូចខាងក្រោម:

$$\log(sales) = \beta_0 + \beta_1 \log(price) + \delta_0 \text{wholesaler} + u, \qquad (៦.១)$$

យើងប្រើប្រាស់ δ_0 ជាប៉ារ៉ាម៉ែត្រទៅលើប្រភេទនៃក្រុមហ៊ុន សំរាប់បកស្រាយប៉ារ៉ាម៉ែត្រនៃdummy variables។ ក្នុងម៉ូដែល(៦.១) ចំពោះអញ្ញត្តិប្រភេទនៃក្រុមហ៊ុន យើងកំណត់តម្លៃស្មើមួយសំរាប់ក្រុមហ៊ុនលក់ដុំ ចំណែកក្រុមហ៊ុនលក់រាយកំណត់តម្លៃស្មើសូន្យ (ក្រុមហ៊ុនលក់រាយជា base group)។ ហេតុនេះ ប៉ារ៉ាម៉ែត្រδ_0ត្រូវបានបកស្រាយដូចខាងក្រោម:

- ✓ δ_0 គឺជាភាពខុសគ្នារវាងក្រុមហ៊ុនលក់ដុំ និង ក្រុមហ៊ុនលក់រាយទាក់ទងនឹងចំណូល
- ✓ $\delta_0 > 0 \Rightarrow$ ចំណូលរបស់ក្រុមហ៊ុនលក់ដុំ ធំជាង ក្រុមហ៊ុនលក់រាយ
- ✓ $\delta_0 < 0 \Rightarrow$ ចំណូលរបស់ក្រុមហ៊ុនលក់ដុំ តិចជាង ក្រុមហ៊ុនលក់រាយ

យើងអាចកំណត់សម្មតិកម្មដើម(H_0) និង ផ្ទុយ(H_1)ដូចខាងក្រោម:

$$H_0: \delta_0 = 0$$
$$H_1: \delta_0 \neq 0$$

ផ្អែកលើ***តារាង៦.១***គេទទួលបានលទ្ធផលនៃម៉ូដែលទាក់ទងនឹងចំណូលរបស់ក្រុមហ៊ុនដូចខាងក្រោម:

$$\widehat{\log(sales)} = 8.460 - 0.242 \log(price) + 0.808 \text{ wholesaler} \qquad (៦.២)$$
$$(0.238) \quad (0.078) \qquad\qquad (0.103)$$
$$n = 32, R^2 = 0.7916$$

មេគុណនៃ wholesalerគឺសំខាន់ ពីព្រោះ វាបង្ហាញអំពីភាពខុសគ្នានៃចំណូលរវាងក្រុមហ៊ុនលក់ដុំ និង ក្រុមហ៊ុនលក់រាយដោយគិតជាមធ្យម។ មេគុណ(coefficient) ស្មើនឹង0.808 មានន័យថា ចំណូលនៃក្រុមហ៊ុនលក់ដុំរំពឹងថា ច្រើនជាងចំណូលនៃក្រុមហ៊ុនលក់រាយប្រហែល80.8%។

ពីព្រោះ $t = 0.808/0.103 = 7.86 > c = 2.763$ នៅឯ1% ហេតុនេះសម្មតិកម្មដើម(H_0)ត្រូវបានបដិសេធ និង ទទួលយកសម្មតិកម្មផ្ទុយ(H_1)។ ដូច្នេះ វាមានសារៈសំខាន់ជាស្ថិតិកើតឡើងនូវភាពខុសគ្នានៃចំណូលរបស់ក្រុមហ៊ុនលក់ដុំ និង ក្រុមហ៊ុនលក់រាយនៅឯ1%។

៦.៣ អន្តរកម្មក្នុងចំណោមបណ្តាអញ្ញត្តិដាំមី(Interactions Involving Dummy Variables)

នៅក្នុងម៉ូដែលរ៉ឺហ្គ្រេស្យុងដែលមានអញ្ញត្តិដាំមីមួយចំនួន យើងអាចធ្វើអន្តរកម្មបញ្ចូលគ្នារវាងអញ្ញត្តិដាំមីទាំងនោះ។ ឧទាហរណ៍ ប្រភេទនៃក្រុមហ៊ុន យើងកំណត់តំលៃស្មើមួយសំរាប់ក្រុមហ៊ុនតម្លៃ(wholesaler = 1) និង ក្រុមហ៊ុនលក់រាយកំណត់តម្លៃស្មើសូន្យ (retailer = 0)។ រីឯ នាយកក្រុមហ៊ុន

កេទស្រីកំណត់ស្មើមួយ (female = 1) និង កេទប្រុសស្មើសូន្យ (male = 0)។ ដូច្នេះ យើងអាចធ្វើអន្តរកម្មបញ្ចូលគ្នារវាងប្រភេទនៃក្រុមហ៊ុន និង កេទ (gender) គឺ នាយកក្រុមហ៊ុនលក់ដុំកេទស្រី (wholesaler * female)។ ហេតុនេះ យើងសរសេរបានម៉ូដែលដូចខាងក្រោម:

$$log(sales) = \beta_0 + \beta_1\, log(price) + \delta_0 whsaler + \delta_1 female + \gamma(whsaler * female) + u, \qquad (៦.៣)$$

វាសំខាន់ត្រូវតែបកស្រាយឲ្យបានត្រឹមត្រូវ នៅពេលមានអន្តរកម្មរវាងdummy variables ដោយការអនុត្តន៍ទម្រង់ដូចខាងក្រោម:

$$E(lsales|lprice) = \begin{cases} \beta_0 + \beta_1 lprice & retailer - male \\ (\beta_0 + \delta_0) + \beta_1 lprice & wholesaler - male \\ (\beta_0 + \delta_1) + \beta_1 lprice & retailer - female \\ (\beta_0 + \delta_0 + \delta_1 + \gamma) + \beta_1 lprice & wholesaler - female \end{cases}$$

ផ្អែកលើ ***តារាង ៦.១*** យើងទទួលបានលទ្ធផលនៃម៉ូដែលទាក់ទងនឹងចំណូលរបស់ក្រុមហ៊ុនដូចខាងក្រោម:

$$\log(sales) = \underset{(0.330)}{8.117} - \underset{(0.066)}{0.227}\log(price) + \underset{(0.242)}{1.226}\, whsaler + \underset{(0.232)}{0.321}\, female - \underset{(0.262)}{0.787}\,(whsaler * female) \qquad (៦.៤)$$

$$n = 32, R^2 = 0.8694$$

យោងលើ(៦.៤) យើងអាចបកស្រាយដូចខាងក្រោម:

- ✓ ដោយប្រើប្រាស់ $t - test$ បង្ហាញថា វាមានសារៈសំខាន់ជាស្ថិតិកើតឡើងសំរាប់ប្រភេទនៃក្រុមហ៊ុន។ ដូច្នេះ នៅក្នុងក្រុមហ៊ុនលក់ដុំដែលមានអ្នកគ្រប់គ្រងជាបុរស (wholesaler male)អាចរកចំណូលឲ្យក្រុមហ៊ុនគិតជាមធ្យមខ្ពស់ជាងក្រុមហ៊ុនលក់រាយដែលមានអ្នកគ្រប់គ្រងជាបុរស (retailer male)ប្រហែល122.60%។
- ✓ ដោយប្រើប្រាស់ $t - test$ បង្ហាញថា វាមិនមានសារៈសំខាន់ជាស្ថិតិកើតឡើងចំពោះកេទ (gender)ឡើយ។ ហេតុនេះ វាមិនមានភស្តុតាងដែលបញ្ជាក់ថា នៅក្នុងក្រុមហ៊ុនលក់រាយដែលមានអ្នកគ្រប់គ្រងជានារី (retailer female)អាចរកចំណូលឲ្យក្រុមហ៊ុនគិតជាមធ្យមខ្ពស់ជាងក្រុមហ៊ុនលក់រាយដែលមានអ្នកគ្រប់គ្រងជាបុរស (retailer male)ឡើយ។

- ✓ នៅក្នុងក្រុមហ៊ុនលក់ដុំដែលមានអ្នកគ្រប់គ្រងជានារី (wholesaler female)អាចរកចំណូលឲ្យក្រុមហ៊ុនគិតជាមធ្យមខ្ពស់ជាងក្រុមហ៊ុនលក់រាយដែលមានអ្នកគ្រប់គ្រងជាបុរស (retailer male) ប្រហែល76%។

៦.៤ ការបញ្ចូលពត៌មានបែបអរឌីណាល (Ordinal Information) ដោយប្រើ Dummy Variables

តាមទ្រឹស្តី បរិមាណទំនិញដែលលក់ចេញ វាប្រែប្រួលទៅតាមតំរូវការ មានន័យថា បើតំរូវការនៃអតិថិជនទាក់ទងនឹងទំនិញណាមួយកើនឡើង នោះបរិមាណលក់ទំនិញនោះក៏កើនឡើង ឬ វាធ្វើឲ្យចំណូលរបស់ក្រុមហ៊ុនកើនឡើងផងដែរ។ ឧបមាថា តំរូវការរបស់អតិថិជនចែកជាកំរិត គឺ

មិនត្រូវការទាល់តែសោះ (n_0)	= 0
ត្រូវការបន្តិចបន្តួច (n_1)	= 1
ត្រូវការមធ្យម (n_2)	= 2
ត្រូវការបង្គួរ (n_3)	= 3
ត្រូវការខ្លាំង (n_4)	= 4

Obs.	Sales	Price	n_1	n_2	n_3	n_4
1	1,215	36	1	0	0	0
2	2,189	28	1	0	0	0
3	6,765	10	0	0	1	0
4	4,231	3	0	1	0	0
5	3,453	6	0	1	0	0
6	2,452	8	1	0	0	0
7	3,136	21	0	1	0	0
8	5,238	12	0	0	1	0
9	6,458	9	0	0	1	0
10	1,234	34	1	0	0	0
11	2,231	24	1	0	0	0
12	3,854	11	0	1	0	0
13	2,436	22	1	0	0	0
14	7,236	10	0	0	0	1

15	4,231	32	0	1	0	0
16	7,329	9	0	0	0	1
17	4,231	14	0	1	0	0
18	2,544	7	1	0	0	0
19	9,453	4	0	0	0	1
20	6,454	8	0	0	1	0
21	5,324	15	0	0	1	0
22	7,564	8	0	0	0	1
23	2,439	43	1	0	0	0
24	2,457	13	1	0	0	0
25	1,435	42	1	0	0	0
26	7,345	12	0	0	0	1
27	4,234	8	0	1	0	0
28	5,232	22	0	0	1	0
29	6,657	21	0	0	1	0
30	3,453	12	0	1	0	0
31	2,440	23	1	0	0	0
32	6,458	14	0	0	1	0

នេះគឺជាឧទាហរណ៍នៃordinal variable។ សំណួរដែលយើងត្រូវឆ្លើយគឺ តើអញ្ញត្តិកំរិតនៃតំរូវការទំនិញពីអតិថិជនត្រូវបានដាក់បញ្ចូលទៅក្នុងម៉ូដែលរ៉ឺគ្រ្សែងយ៉ាងដូចម្តេច ដើម្បីពន្យល់អំពីចំណូលរបស់ក្រុមហ៊ុន?

លទ្ធភាពដែលអាចធ្វើបាន យើងអាសរសេរសមីការដូចខាងក្រោម:

$$log(sales) = \beta_0 + \beta_1 d + \text{កត្តាដទៃទៀត},$$

ដោយរក្សាកត្តាដទៃទៀតថេរ នៅពេលកំរិតនៃតំរូវការកើនឡើង១ឯកតា បន្ទាប់មកចំណូលរបស់ក្រុមហ៊ុនប្រែប្រួលβ_1។ ប៉ុន្តែ ការបកស្រាយបែបនេះ វាពុំសូវសមហេតុផល ពីព្រោះ ទោះបីជាគម្លាតនៃបំរែបំរួលនៃតំរូវការពីមួយកំរិតទៅមួយកំរិតស្មើគ្នាក៏ដោយ (បំរែបំរួលពី n_1ទៅn_2 ស្មើ បំរែបំរួលពី n_3 ទៅn_4 ស្មើ១) ប៉ុន្តែ បម្រែបម្រួលចំណូលខុសគ្នា។

ដំណោះស្រាយដែលសមហេតុផលគឺថា កំរិតនៃតំរូវការទំនិញពីអតិថិជនត្រូវកំណត់ជា dummy variables តាមវិធីសាស្ត្រដូចខាងក្រោម:

$n_1 = 1$	ប្រសិនបើ	$n = 1$	និង	ក្រៅពីនេះ	$n_1 = 0$
$n_2 = 1$	ប្រសិនបើ	$n = 2$	និង	ក្រៅពីនេះ	$n_2 = 0$
$n_3 = 1$	ប្រសិនបើ	$n = 3$	និង	ក្រៅពីនេះ	$n_3 = 0$
$n_4 = 1$	ប្រសិនបើ	$n = 4$	និង	ក្រៅពីនេះ	$n_4 = 0$

ដូច្នេះ យើងអាចប៉ាន់ស្មានម៉ូដែលខាងក្រោម:

$$\log(sales) = \beta_0 + \delta_1 n_1 + \delta_2 n_2 + \delta_3 n_3 + \delta_4 n_4 + \text{កត្តាដទៃទៀត}, \qquad (៦.៥)$$

ក្នុងម៉ូដែល(៦.៥) យើងបញ្ចូល dummy variablesចំនួន៤ ពីព្រោះ កំរិតនៃតម្រូវការទំនិញពីអតិថិជនមាន៥ប្រភេទ។ កំរិតនៃតម្រូវការ "មិនត្រូវការទាល់តែសោះ (n_0)" ជាមួយតម្លៃស្មើសូន្យ មិនត្រូវបានបញ្ចូលទៅក្នុង(៦.៥)ឡើយ និង កំណត់ជាក្រុមគោល (base group)។ យើងអាចបកស្រាយមេគុណដូចខាងក្រោម:

- ✓ δ_1 គឺជាភាពខុសគ្នារវាង n_1 និង n_0 ទាក់ទងនឹងចំណូលក្រុមហ៊ុន
- ✓ δ_2 គឺជាភាពខុសគ្នារវាង n_2 និង n_0 ទាក់ទងនឹងចំណូលក្រុមហ៊ុន
- ✓ δ_3 គឺជាភាពខុសគ្នារវាង n_3 និង n_0 ទាក់ទងនឹងចំណូលក្រុមហ៊ុន
- ✓ δ_4 គឺជាភាពខុសគ្នារវាង n_4 និង n_0 ទាក់ទងនឹងចំណូលក្រុមហ៊ុន

៦.៥ អន្តរកម្មរវាងអញ្ញត្តិដាំមី (dummy variable) និង មិនមែនដាំមី

ក្នុងការសិក្សាអំពីកត្តាជះឥទ្ធិលទៅលើចំណូលរបស់ក្រុមហ៊ុន យើងលើកយកក្រុមហ៊ុនពីរប្រភេទ (binary variable) គឺ លក់ដុំ(wholesaler) និង លក់រាយ(retailer) មកបរិយាយ ដោយកំណត់តម្លៃស្មើមួយសំរាប់ក្រុមហ៊ុនលក់ដុំ ចំណែកក្រុមហ៊ុនលក់រាយកំណត់តម្លៃស្មើសូន្យ (dummy variable)។

ជាមួយគ្នានេះ អញ្ញត្តិអន្តរកម្មរវាងថ្លៃទំនិញជាមួយប្រភេទក្រុមហ៊ុន (price × whsaler) ក៏ត្រូវបានលើកយកមកបញ្ចូលក្នុងម៉ូដែលខាងក្រោមផងដែរ:

$$sales = \beta_0 + \delta whsaler + \beta_1 price + \gamma(price \times whsaler) + u, \qquad (៦.៦)$$

ហេតុនេះ ចំណូលអាចត្រូវបានបែងចែកទៅតាមប្រភេទក្រុមហ៊ុនទាំងពីរដូចខាងក្រោម:

$$E(sales|price) = \begin{cases} (\beta_0 + \delta) + (\beta_1 + \gamma)price & wholesaler = 1 \\ \beta_0 + \beta_1 price & retailer = 0 \end{cases}$$

- ✓ ករណីថ្លៃទំនិញកើនឡើង១ដុល្លារ ចំណូលរបស់ក្រុមហ៊ុនលក់ដុំគឺ $(\beta_1 + \gamma)$ រីឯ ចំណូលរបស់ក្រុមហ៊ុនលក់រាយគឺ β_1។
- ✓ ភាពខុសគ្នារវាងចំណូលរបស់ក្រុមហ៊ុនលក់ដុំ និង រាយ គឺ δ ។

ផ្អែកលើទិន្នន័យ***តារាង ៦.១*** យើងទទួលបានលទ្ធផលនៃសមីការដូចខាងក្រោម:

$$sales = 3{,}445.00 + 3{,}687.60\ whsaler - 44.57\ price - 35.12\ price * whsaler \quad (៦.៧)$$
$$(589.49) \quad (805.23) \quad (23.47) \quad (45.18)$$
$$n = 32, R^2 = 0.7756$$

ដូច្នេះ យើងអាចបែងចែកលទ្ធផលប៉ាន់ស្មាននៃសមីការដូចខាងក្រោម:

ក្រុមហ៊ុនលក់ដុំ $sales = (3{,}445.00 + 3{,}687.60) - (44.57 + 35.12)price$
$= 7{,}132.60 - 79.69\ price$

ក្រុមហ៊ុនលក់រាយ $sales = 3{,}445 - 44.57\ price$

ចំពោះ(៦.៧) ប្រសិនរាល់អញ្ញត្តិទាំងឡាយមានសារៈសំខាន់ជាស្ថិតិ យើងអាចបកស្រាយដូចខាងក្រោម:

- ✓ ក្រុមហ៊ុនលក់ដុំរំពឹងថា ទទួលបានចំណូលខ្ពស់ជាក្រុមហ៊ុនលក់រាយប្រហែល3,687,600ដុល្លារ
- ✓ ខណៈពេលថ្លៃទំនិញកើនមួយដុល្លារ នោះចំណូលរបស់ក្រុមហ៊ុនលក់ដុំថយចុះ 79,690 ដុល្លារ និង ចំណូលរបស់ក្រុមហ៊ុនលក់រាយថយចុះ 44,570 ដុល្លារ។

៦.៦ ការធ្វើតេស្តអេគ្វីវ៉ាឡង់(equivalence)នៃរ៉ឺហ្គ្រេស្យុងពីរ

បណ្តាឧទាហរហណ៍មុនបានអធិប្បាយអន្តរកម្មរវាងdummy variablesជាមួយអញ្ញត្តិឯករាជ្យដទៃទៀត។ ជួនកាលយើងត្រូវការធ្វើតេស្តសម្មតិកម្មដើម(null hypothesis)ដែលមានពីរក្រុម (two populations or groups)ផ្សេងគ្នាជាមួយរ៉ឺហ្គ្រេស្យុងដូចគ្នា(same regression function) និងសម្មតិកម្មផ្ទុយ(alternative hypothesis)ដែលមានស្លូប(slopes)ទាំងឡាយក៏ខុសគ្នា។

ផ្អែកលើទិន្នន័យក្នុង***តារាង ៦.១*** យើងចង់ធ្វើតេស្តថាតើម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចគ្នា(same regression model)ពន្យល់អំពីប្រសិទ្ធភាពនៃការប្រើប្រាស់ទ្រព្យសម្បត្តិសរុបក្នុងការបង្កើតប្រាក់ចំណេញគិតជាភាគរយ(ROA)ទាក់ទងក្រុមហ៊ុនទាំងពីរប្រភេទនោះដែរឬទេ។ សមីការគឺ:

$$ROA = \beta_0 + \beta_1 price + \beta_2 adv + u,$$

យើងអាចបញ្ចូលdummy variable(ទំហំក្រុមហ៊ុន គឺ ក្រុមហ៊ុនតូចស្មើសូន្យ ក្រុមហ៊ុនមធ្យមស្មើមួយ)ជាមួយអ៊ីនធ័រសិប(intercept)មួយផ្សេង។ បើស្លូប(slope)ផ្តោតទៅលើប្រភេទក្រុមហ៊ុន យើងអាចធ្វើអន្តរកម្មរវាងទំហំក្រុមហ៊ុន(ក្រុមហ៊ុនមធ្យម) ជាមួយអញ្ញត្តិឯករាជ្យទៅក្នុងសមីការ។

ប្រសិនយើងចាប់អារម្មណ៍ក្នុងការធ្វើតេស្ត ថាតើវាមានភាពខុសគ្នាណាមួយរវាងក្រុមហ៊ុនតូច និងមធ្យម នោះយើងត្រូវតែកំណត់ម៉ូដែលមួយមាន intercept និង slopes ខុសគ្នាសំរាប់ទាំងពីរក្រុម(two populations or groups)ដូចខាងក្រោម:

$$ROA = \beta_0 + \delta_0 firm + \beta_1 price + \delta_1 firm.price + \beta_2 adv + \delta_2 firm.adv + u \quad (៦.៨)$$

ប៉ារ៉ាម៉ែត្រδ_0 គឺជាភាពខុសគ្នានៃ intercept រវាងក្រុមហ៊ុនតូច និង មធ្យម រីឯδ_1គឺជាភាពខុសគ្នានៃ slope ទាក់ទងនឹងថ្លៃទំនិញរវាងក្រុមហ៊ុនតូច និង មធ្យម។ ចំណែកδ_2គឺជាភាពខុសគ្នានៃ slope ទាក់ទងនឹងការចំណាយផ្សាយពាណិជ្ជកម្ម។ ម៉ូដែលរ៉ឺហ្គ្រេស្យុងដូចគ្នា(same regression model) ពន្យល់ចំណូលក្រុមហ៊ុនតូច និង មធ្យម យើងអាចកំណត់សម្មតិកម្មដើម(null hypothesis)ដូចខាងក្រោម:

$$H_0: \delta_0 = 0, \delta_1 = 0, \delta_2 = 0. \quad (៦.៩)$$

ប្រសិន δ_jណាមួយខុសពីសូន្យ នោះបញ្ជាក់ថា ម៉ូដែលគឺខុសគ្នារវាងក្រុមហ៊ុនតូច និង មធ្យម។ ផ្អែកលើទិន្នន័យក្នុង***តារាង ៦.១*** យើងទទួលបានលទ្ធផលប៉ាន់ស្មាន(unrestricted model)ដូចខាងក្រោម:

ទិន្នន័យទាំងមូល

$$\widehat{ROA} = \underset{(6.267)}{6.306} + \underset{(7.989)}{19.982} firm + \underset{(0.222)}{0.125} price - \underset{(0.245)}{0.528} firm.price + \underset{(0.097)}{0.345} adv - \underset{(0.154)}{0.353} firm.adv \quad (៦.១០)$$
$$n = 32, R^2 = 0.6854, \bar{R}^2 = 0.6249, SSR = 365.23$$

ទិន្នន័យក្រុមហ៊ុនលក់ដុំ

$$\widehat{ROA} = \underset{(8.902)}{3.644} + \underset{(9.773)}{29.524} firm + \underset{(0.314)}{0.183} price - \underset{(0.337)}{0.844} firm.price + \underset{(0.137)}{0.389} adv - \underset{(0.170)}{0.506} firm.adv \quad (៦.១១)$$
$$n = 17, R^2 = 0.8055, \bar{R}^2 = 0.7171, SSR = 81.352$$

ទិន្នន័យក្រុមហ៊ុនលក់រាយ

$$\widehat{ROA} = 7.396 - 17.718firm + 0.099price + 0.205firm.price$$
$$(7.384)\ (14.869)\qquad (0.264)\qquad (0.359)$$
$$+0.331adv + 0.408firm.adv \qquad (៦.១២)$$
$$(0.114)\qquad (0.307)$$
$$n = 15, R^2 = 0.78, \bar{R}^2 = 0.6579, SSR = 126.172$$

វាមានការលំបាកក្នុងការបកស្រាយនូវភាពខុសគ្នារវាងក្រុមហ៊ុនតូច និង មធ្យម ករណីលំអៀងស្តង់ដា(standard errors)នៃអញ្ញត្តិទំហំក្រុមហ៊ុន(firm) និង អញ្ញត្តិអន្តរកម្ម(interaction terms)មានចំនួនធំ។ ចំពោះ(៦.១១) ប្រសិនយើងមើលតែទៅលើអញ្ញត្តិទំហំក្រុមហ៊ុន(firm)ជាមួយអញ្ញត្តិដទៃទៀតថេរ(holding other factors fixed) រួចប៉ាន់ស្មានថា ក្រុមហ៊ុនមធ្យមមាន ROAធំជាងក្រុមហ៊ុនតូចចំនួន29.524% នោះការប៉ាន់ស្មានបែបនេះ វាជាកំហុសឆ្គង។ វាមានន័យថា ការប៉ាន់ស្មានបែបនេះកើតឡើងបានតែក្នុងករណីដែលអញ្ញត្តិprice និង adv ស្មើសូន្យ។ ឧបមាថា ជាមធ្យមថ្លៃទំនិញក្នុង១ឯកតា(price)ស្មើ10ដុល្លារ និង ចំណាយផ្សាយពាណិជ្ជកម្ម(adv)ស្មើ22ពាន់ដុល្លារ។ ដូច្នេះ ភាពខុសគ្នាដែលបានប៉ាន់ស្មានរវាងក្រុមហ៊ុនតូច និង មធ្យម ទាក់ទងនឹងROA គឺ 29.524 – 0.844(10) – 0.506(22) ≈ 9.952 គឺ យើងអាចប៉ាន់ស្មានថា ក្រុមហ៊ុនមធ្យមទទួលបានROA ធំជាងក្រុមហ៊ុនតូចប្រហែល 9.952%។

សំរាប់ម៉ូដែលរ៉ឺគ្រេស្យុងមួយដែលមាន kអញ្ញត្តិឯករាជ្យ និង interceptមួយ ដោយយើងសន្មតថាមានពីរក្រុមគឺ d = 1 និង d = 2។ បន្ទាប់មក យើងត្រូវការធ្វើតេស្ត ថាតើintercept និង រាល់ស្លូប(slope)នៃក្រុមទាំងពីរដូចគ្នាដែរឬទេ។ ម៉ូដែលដូចខាងក្រោម:

$$y = \beta_{d,0} + \beta_{d,1}x_1 + \beta_{d,2}x_2 + \cdots + \beta_{d,k}x_k + u, \qquad (៦.១៣)$$

ដែល d = 1 (ក្រុមហ៊ុនលក់ដុំ) និង d = 2 (ក្រុមហ៊ុនលក់ដុរាយ)។ ក្នុង(៦.១៣) សម្មតិកម្មត្រូវបានកំណត់ថា បៃតតា (β)នីមួយៗនៃក្រុមទាំងពីរគឺ ស្មើគ្នា ចំពោះ k + 1 restrictions (ក្នុងឧទាហរណ៍ខាងលើ k + 1 = 3)។ រីឯ unrestricted modelមានdegree of freedomស្មើ n – 2(k + 1) [ក្នុងឧទាហរណ៍ខាងលើ 17 – 2(3) = 11] ជាមួយdummy variableមួយក្រុម(firm) និង interaction termsចំនួនk (firm.price និង firm.adv) ដោយបន្ថែមទៅលើ intercept និង បណ្តាអញ្ញត្តិខ្លួនវាផ្ទាល់(price និង adv)។ សំខាន់គឺ sum of squared residuals (SSR)នៃunrestricted model ជាផលបូកSSRនៃក្រុមទាំងពីរ។ តាងSSR$_1$គឺជា sum of squared residualsដែលទទួលបានពី(៦.១៣)សំរាប់ក្រុមទី១មានសំណាកn$_1$។ តាងSSR$_2$គឺជា sum of squared residualsដែលទទួលបានពី(៦.១៣)សំរាប់ក្រុមទី២មានសំណាកn$_2$។ នៅក្នុងឧទាហរណ៍មុន ក្រុម១តំណាងក្រុមហ៊ុនមធ្យម និង

ក្រុម២តំណាងក្រុមហ៊ុនតូច ដូច្នេះ sum of squared residualsសំរាប់unrestricted modelគឺ $SSR_{ur} = SSR_1 + SSR_2$។ រីឯ sum of squared residuals (SSR_p) សំរាប់restricted modelគឺជាSSRដែលទទួលបានពីការបញ្ចូលគ្នានៃសំណាករបស់ក្រុមទាំងពី និង បន្ទាប់មក ដំណើការសមីការធម្មតាមួយ។ ក្រោយពេលទទួលបានSSR យើងអាចគណនាតម្លៃFតាមរូបមន្តដូចខាងក្រោម:

$$F = \frac{[SSR_p - (SSR_1 + SSR_2)]}{SSR_1 + SSR_2} \cdot \frac{[n - 2(k+1)]}{k+1} \qquad (៦.១៤)$$

ដែលn គឺជាសរុបចំនួនសំណាក។ នៅក្នុងអេកូណូមេទ្រីច(econometrics) តម្លៃស្ថិតិនៃF ត្រូវបានហៅថា **Chow statistic**។ វាមានសុពលភាពនៅក្រោមលក្ខខណ្ឌheteroskedasticity មានន័យថាផ្អែកលើសម្មតិកម្មដើម (null hypothesis) លំអៀងវ៉ារីយ៉ង់(error variance) សំរាប់ក្រុមទាំងពីរត្រូវស្មើគ្នា។

ផ្អែកលើទិន្នន័យក្នុង***តារាង ៦.១*** យើងអាចអនុវត្ត**Chow statistic** តាមរយៈសមីការ(restricted model)ដូចខាងក្រោម:

ក្រុមហ៊ុនលក់ដុំ $ROA = 27.249 - 0.547\ price + 0.030\ adv$ (៦.១៥)

$(3.857)\ (0.131)\qquad (0.080)$

$n = 17,\ R^2 = 0.6395, \bar{R}^2 = 0.5880, SSR = 150.779$

ក្រុមហ៊ុនមធ្យម $ROA = 33.168 - 0.660\ price - 0.117\ adv$ (៦.១៦)

$(4.867)\ (0.148)\qquad (0.120)$

$n = 10, R^2 = 0.7550, \bar{R}^2 = 0.6850, SSR = 75.480$

ក្រុមហ៊ុនតូច $ROA = 3.644 + 0.183\ price + 0.389\ adv$ (៦.១៧)

$(3.967)\ (0.140)\qquad (0.061)$

$n = 7,\ R^2 = 0.9460, \bar{R}^2 = 0.9190, SSR = 5.872$

$$F = \frac{[150.779 - (75.480 + 5.872)]}{75.480 + 5.872} \cdot \frac{11}{3} \approx 3.13$$

ដូច្នេះ យើងមិនអាចបដិសេដសម្មតិកម្មដើមឡើយ គឺថា ម៉ូដែលរ៉ឺហ្គ្រេស្យុងទាក់ទងនឹង*ROA* ប្រហែលមិនខុសគ្នាឡើយរវាងក្រុមហ៊ុនតូច និង មធ្យម។

៦.៧ ម៉ូដែលប្រូបាប៊ីលីតេលីនេអ៊ែរ (The Linear Probability Model)

ជួនកាលលទ្ធផលដែលកើតឡើងក្នុងការអនុវត្តន៍គឺ លទ្ធផលទ្វេភាគ (binary outcome) មានន័យថា អញ្ញត្តិមិនឯករាជ្យ (dependent variable) y ត្រូវបានកំណត់តម្លៃសូន្យ និង មួយ។ ឧទាហរណ៏ y អាចត្រូវបានកំណត់ដើម្បីបញ្ជាក់ ថាតើក្រុមហ៊ុនមួយត្រូវបានកាន់កាប់ដោយក្រុមហ៊ុនដទៃទៀតដែរឬទេ។ ក្នុងឧទាហរណ៏នេះ យើងអាចកំណត់ y = 1 គឺជាលទ្ធផលមួយក្នុងចំណោមលទ្ធផលទាំងឡាយ និង y = 0 តំណាងឲ្យលទ្ធផលផ្សេងទៀត។ ជាទូទៅ ម៉ូដែលរ៉ឺហ្គ្រេស្សុងមានទំរង់ដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៦.១៨)$$

នៅពេលម៉ូដែលរ៉ឺហ្គ្រេស្សុងមានអញ្ញត្តិមិនឯករាជ្យ y ជាអញ្ញត្តិទ្វេភាគ (binary variable) តើយើងអាចបកស្រាយវាយ៉ាងដូចម្តេច? តំលៃ y ប្រែប្រួលពី សូន្យទៅមួយ ឬ ពីមួយទៅសូន្យ។

ចំណុចសំខាន់គឺថា នៅពេល y ជាអញ្ញត្តិទ្វេភាគ (binary variable) ដោយកំណត់តម្លៃសូន្យ និង មួយ នោះគឺ $P(y = 1|x) = E(y|x)$: ប្រូបាប៊ីលីតេនៃ"ភាពជោគជ័យ" (the probability of "success") គឺថា ជាប្រូបាប៊ីលីតេដែល y = 1 វាជាតំលៃរំពឹងទុកនៃ y។ ដូច្នេះ យើងទទួលបានសមីការសំខាន់:

$$P(y = 1|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \qquad (៦.១៩)$$

នោះគឺ ប្រូបាប៊ីលីតេនៃភាពជោគជ័យ ឬ $p(x) = P(y = 1|x)$ គឺជាមុខងារលីនេអ៊ែរនៃx_j។ អាស្រ័យដោយ ផលបូកនៃប្រូបាប៊ីលីតេត្រូវតែស្មើមួយ ហេតុនេះ$P(y = 0|x) = 1 - P(y = 1|x)$ក៏ជាមុខងារលីនេអ៊ែរនៃ$x_j$ផងដែរ។

ម៉ូដែលរ៉ឺហ្គ្រេស្សុងមានអញ្ញត្តិមិនឯករាជ្យ y ជាអញ្ញត្តិទ្វេភាគ (binary dependent variable) ត្រូវបានហៅថា ម៉ូដែលប្រូបាប៊ីលីតេលីនេអ៊ែរ (**linear probability model – LPM**) ពីព្រោះ ម៉ូដែលទ្វេភាគ (a binary response model) ជាមួយប្រូបាប៊ីលីតេនៃអញ្ញត្តិមិនឯករាជ្យកំណត់តម្លៃស្មើមួយ គឺជាលីនេអ៊ែរទាក់ទងនឹងប៉ារ៉ាម៉ែត្រβ_j។ នៅក្នុងLPM β_jវាស់វែងបម្រែបម្រួលប្រូបាប៊ីលីតេនៃភាពជោគជ័យនៅពេលx_jប្រែប្រួលជាមួយកត្តាដទៃទៀតថេរ:

$$\Delta P(y = 1|x) = \beta_j \Delta x_j. \qquad (៦.២០)$$

ដូច្នេះ ម៉ូដែលរ៉ឺហ្គ្រេស្សុងអនុញ្ញាតឲ្យអ្នកប្រើប្រាស់ប៉ាន់ស្មានការជះឥទ្ធិពលនៃអញ្ញត្តិឯករាជ្យ(x_j) ទៅលើព្រឹត្តិការណ៍បែបគុណភាព។ យន្តការនៃOLSគឺដូចគ្នាពីមុន។ យើងសរសេរជាសមីការប៉ាន់ស្មាន ដូចជា:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k,$$

យើងត្រូវតែចងចាំថា $\hat{y}$គឺជាប្រូបាប៊ីលីតេនៃភាពជោគជ័យដែលត្រូវបានព្យាករណ៍។ ហេតុនេះ $\hat{\beta}_0$គឺជាប្រូបាប៊ីលីតេនៃភាពជោគជ័យដែលត្រូវបានព្យាករណ៍ នៅពេលx_jនីមួយៗត្រូវបានកំណត់ស្មើសូន្យ។ ចំណែក β_jវាស់វែងបម្រែបម្រួលដែលត្រូវបានព្យាករណ៍ដែលទាក់ទងនឹងប្រូបាប៊ីលីតេនៃភាពជោគជ័យ នៅពេលx_1កើនឡើងមួយឯកតា។

ដើម្បីបកស្រាយម៉ូដែលប្រូបាប៊ីលីតេលីនេអ៊ែរ(linear probability model)ឲ្យបានត្រឹមត្រូវ យើងត្រូវតែយល់ដឹង តើពាក្យ "ភាពជោគជ័យ (success)"មានន័យយ៉ាងដូចម្តេច។ វាជាការល្អដែលកំណត់តម្លៃនៃអញ្ញត្តិមិនឯករាជ្យស្មើមួយ (event y = 1)។ ឧទាហរណ៍ អតិថិជនម្នាក់កំពុងធ្វើការសម្រេចចិត្តថា តើគាត់គួរតែជ្រើសក្រុមហ៊ុន A ឬ B។ កំណត់អញ្ញត្តិក្រុមហ៊ុន A:

$$A = \begin{cases} 1 & \text{ប្រសិនក្រុមហ៊ុនA ត្រូវបានជ្រើសរើស} \\ 0 & \text{ប្រសិនក្រុមហ៊ុនB ត្រូវបានជ្រើសរើស} \end{cases}$$

តំលៃរំពឹងទុកនៃអញ្ញត្តិនេះគឺ $\Delta P(A = 1|x) = \beta_j \Delta x_j = P_A =$ ប្រូបាប៊ីលីតេដែលក្រុមហ៊ុន A ត្រូវបានជ្រើសរើសផ្អែកលើកត្តាមួយចំនួន។ តើការសម្រេចចិត្តរបស់អតិថិជនផ្អែកលើកត្តាអ្វីខ្លះ? សមាមាត្រផលធៀបរវាងថ្លៃទំនិញរបស់ក្រុមហ៊ុនA និង B (PR) គឺជាកត្តាសំខាន់មួយ។ នៅពេលថ្លៃទំនិញរបស់ក្រុមហ៊ុនAកើនឡើង នោះប្រូបាប៊ីលីតេនៃការជ្រើសរើសទំនិញរបស់ក្រុមហ៊ុនAគួរតែថយចុះ។ នៅមានកត្តាដទៃទៀតក៏ប្រហែលជះឥទ្ធិពលដល់ការសម្រេចចិត្តរបស់អតិថិជនផងដែរ ដូចជា ការតាំងបង្ហាញទំនិញទាំងនេះ។ តាង អញ្ញត្តិ DA គឺការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនA និង DB គឺការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនB។ ម៉្យាងទៀត វាមានតម្លៃស្មើមួយករណីមានការតាំងបង្ហាញ រីឯតម្លៃស្មើសូន្យករណីមិនមានការតាំងបង្ហាញ។ យើងរំពឹងថា វត្តមាននៃការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនA នឹងធ្វើឲ្យប្រូបាប៊ីលីតេនៃការទិញទំនិញរបស់ក្រុមហ៊ុនAកើនឡើង និង វត្តមាននៃការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនBនឹងធ្វើឲ្យប្រូបាប៊ីលីតេនៃការទិញទំនិញរបស់ក្រុមហ៊ុនAថយចុះ។ ឧបមាថា យើងមានទិន្នន័យដូចខាងក្រោម:

P_A	1	1	0	1	1	1	1	0	1	0	1	1	0	1
P	0.8	1.2	1.3	0.7	0.9	1.3	1.1	1.4	1.3	1.2	0.6	0.9	1.2	1.1
R	9	4	2	9	6	9	2	5	2	6	9	5	4	2
DA	1	0	0	1	1	1	1	0	1	0	1	1	0	1
DB	0	0	1	1	0	0	1	1	0	1	0	0	1	1

ផ្អែកលើទិន្នន័យទាំងនេះ យើងទទួលបានលទ្ធផលនៃសមីការប៉ាន់ស្មានដូចខាងក្រោម:

$$P_A = 0.589 - 0.124PR + 0.633DA - 0.285DB$$
$$(0.478)\ (0.348) \qquad (0.176) \qquad (0.142)$$
$$n = 14, R^2 = 0.8025$$

ជាលទ្ធផល មេគុណនៃសមាមាត្រផលធៀបរវាងថ្លៃទំនិញរបស់ក្រុមហ៊ុនA និង B (PR)មិនមានសារៈសំខាន់ជាស្ថិតិឡើយ ទោះបីវាបង្ហាញទំនាក់ទំនងផ្ទុយជាមួយប្រូបាប៊ីលីតេនៃការជ្រើសរើសទំនិញរបស់ក្រុមហ៊ុនA។ មេគុណនៃអញ្ញត្តិពីទៀតគឺ DA និង DB បង្ហាញនូវសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត1% និង 5% រៀងគ្នា។ ដូច្នេះ វត្តមាននៃការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនAនឹងធ្វើឲ្យប្រូបាប៊ីលីតេនៃការទិញទំនិញរបស់ក្រុមហ៊ុនAកើនឡើង0.633 និង វត្តមាននៃការតាំងបង្ហាញទំនិញរបស់ក្រុមហ៊ុនBនឹងធ្វើឲ្យប្រូបាប៊ីលីតេនៃការទិញទំនិញរបស់ក្រុមហ៊ុនAថយចុះ0.285។

លំហាត់

៦.១ តារាងខាងក្រោមបង្ហាញលទ្ធផលនៃការវិភាគម៉ូដែលរ៉ឺហ្គ្រេស្សុងដែលត្រូវបានបង្កើតឡើងដើម្បីពន្យល់អំពីប្រាក់បៀវត្សន៍ (គិតជាខ្នង់ពាន់ដុល្លារក្នុង១ឆ្នាំ) របស់នាយកប្រតិបត្តិចំនួន៣២ក្រុមហ៊ុនដោយផ្អែកលើចំណូល (គិតជាខ្នង់លានដុល្លារក្នុង១ឆ្នាំ) និង ប្រភេទក្រុមហ៊ុន(ធនាគារ សណ្ឋាគារ ផ្សារទំនើប និង ធានារ៉ាប់រង)។ ប្រភេទក្រុមហ៊ុនគឺជាdummy variables និង ធានារ៉ាប់រងជាbase group។

លទ្ធផលនៃmultiple regressionសំរាប់ប្រាក់បៀវត្សន៍របស់នាយកប្រតិបត្តិ

ម៉ូដែលរ៉ឺហ្គ្រេស្សុងគឺ:

$$\begin{aligned} \text{Salary} &= 854.7634 \\ &+ 0.0228 * \text{sales} \\ &- 189.1290 * \text{Bank} \\ &- 210.2450 * \text{Hotel} \\ &- 124.6489 * \text{Supermarket} \end{aligned}$$

$R^2 = 0.3488$

Variable	Coefficient	Standard error
Constant	854.7634	126.3209
Sales	0.0228	0.078
Bank	− 189.1290	45.0865
Hotel	− 210.2450	168.2360
Supermarket	− 124.6489	178.7650

ក. តើចំណូល និង ប្រភេទក្រុមហ៊ុនមានការជះឥទ្ធិពលជាសំខាន់ទៅលើប្រាក់បៀវត្សន៍ដែរឬទេ?

ខ. ចូរធ្វើតេស្តទៅលើសម្មតិកម្មដើមH_0: ចំណូលមិនជះឥទ្ធិពលទៅលើប្រាក់បៀវត្សន៍ និង សម្មតិកម្មផ្ទុយH_1: ចំណូលជះឥទ្ធិពលទៅលើប្រាក់បៀវត្សន៍នៅឯកំរិតសារៈសំខាន់៥%។ បន្ទាប់មក ចូរធ្វើតេស្តទៅលើសម្មតិកម្មនៃប្រភេទក្រុមហ៊ុននីមួយៗ (ធនាគារ សណ្ឋាគារ និង ផ្សារទំនើប) ទាក់ទងនឹងប្រាក់បៀវត្សន៍ធៀបជាមួយក្រុមហ៊ុនធានារ៉ាប់រង។

គ. ចូរបកស្រាយបំរែបំរួលនៃចំណូលទៅលើប្រាក់បៀវត្សន៍។

ឃ. ផ្អែកទៅលើមេគុណ (regression coefficient) តើប្រាក់បៀវត្សន៍របស់នាយកប្រតិបត្តិធនាគាទាបជាង ឬ ខ្ពស់ជាងធានារ៉ាប់រងចំនួនប៉ុន្មាន?

៦.២ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ ដើម្បីប៉ាន់ស្មានសមីការលទ្ធភាពនៃប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 \log(WC) + \beta_2 TAT + \beta_3 TAT^2 + \beta_4 DR + \beta_5 LOCATION + u$$

ក. ប៉ាន់ស្មានសមីការនៃ*ROA*

ខ. ចូរធ្វើតេស្តសម្មតិកម្មនៃរាល់β_jនៅឯកំរិតសារៈសំខាន់៥%

គ. បកស្រាយមេគុណ(β_j)នីមួយៗ

ឃ. ផ្អែកលើសំណួរ(ខ) បើរកឃើញថា វាមានβ_jមួយចំនួនមិនមានសារៈសំខាន់ជាស្ថិតិ តើយើងអាចដកវាចេញពីម៉ូដែលក្នុងសំណួរ(ក)ដែរឬទេ ?

មេរៀនទី ៧
ហេធីរ៉ូស្កេដាស្ទីស៊ីតធី (Heteroskedasticity)

ចំពោះOLS estimator វ៉ារីយ៉ង់(variance)នៃunobserved errorត្រូវតែថេរ (homoskedasticity assumption)។ បើវាមិនឆ្លើយតមតាមលក្ខខណ្ឌនេះ នោះគេហៅថា heteroskedasticity។ ក្នុងករណីកើតឡើងនូវheteroskedasticity វានាំឲ្យលទ្ធផលនៃការធ្វើតេស្ត មួយចំនួនដូចជា *t, F*,....មិនមានសុកលភាពឡើយ។

៧.១ ការកើតឡើងនៃheteroskedasticityចំពោះOLS estimator

ម៉ូដែលរ៉ឺហ្គ្រេស្យុងលីនេអ៊ែរគឺ:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៧.១)$$

នៅក្នុងមេរៀនទី៣ ភាពមិនលំអៀងនៃOLS estimator $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ ត្រូវបានបង្ហាញនៅក្រោមលក្ខខណ្ឌគោសម៉ាកកូវ (Gauss-Markov assumptions)ទី១ដល់ទី៤ (MLR.1 ដល់ MLR.4)។ រីឯលក្ខខណ្ឌទី៥(MLR.5)គឺ homoskedasticity assumptionដែលថ្លែងថា error varianceថេរ:

$$Var(u|x_1, x_2, \dots, x_k) = \sigma^2 \qquad (៧.២)$$

ទ្រឹស្តីបទGauss-Markovបានបញ្ជាក់ថា OLS estimatorជាលីនេអ៊ែរដែលមិនមានភាពលំអៀង (best linear unbiased)សំខាន់ក៏អាស្រ័យដោយhomoskedasticity assumption។ ប្រសិន $Var(u|x)$មិនថេរ នោះOLSលែងជាលីនេអ៊ែរដែលមិនមានភាពលំអៀង(best linear unbiased estimator: BLUE)។

ឧទាហរណ៍ ៧.១ ឧបមាថាមានក្រុមហ៊ុនមួយចំនួនដំណើរការលក់គ្រឿងបរិភោគសំរាប់អតិថិជន។ ក្នុងនាមជាអ្នកស្រាវជ្រាវចង់ដឹងអំពីកត្តាជះឥទ្ធិលទៅលើចំណូល(sales)របស់ក្រុមហ៊ុនក្នុងតំបន់មួយ ដោយបានធ្វើការអង្កេតក្រុមហ៊ុនចំនួន៣២។ យោងលើរបកគំហើញពីមុន និង ទ្រឹស្តីផ្សេងៗ កត់សំគាល់ឃើញថា វាមានកត្តាសំខាន់ៗដែលទាក់ទងនឹងបម្រែបម្រួលចំណូល ដូចជា ថ្លៃទំនិញ (price)គិតជាមធ្យម ចំណាយផ្សាយពាណិជ្ជកម្ម(advertising គិតខ្នង់ពាន់) ចំនួនហាង(stores) ចំនួននិយោជិកផ្នែកលក់ (Staffs) គុណភាពទំនិញ(quality) ចំណង់ចំណូលចិត្ត(satisfaction) កំរិតជីវភាព(living standard)។ល។ ដោយផ្តោតតែទៅលើថ្លៃទំនិញ និង ចំណាយផ្សាយពាណិជ្ជកម្ម យើងសរសេរបានម៉ូដែលដូចខាងក្រោម:

$$Sales = \beta_0 + \beta_1 Price + \beta_2 Adv + \beta_3 stores + \beta_4 staffs + u, \qquad (៧.៣)$$

ដូច្នេះ យោងលើលក្ខខណ្ឌទី៥(MLR.5) ម៉ូដែលរ៉ឺហ្គ្រេស្យុង(៧.៣)ត្រូវឆ្លើយតបដូចខាងក្រោម:

$$Var(u|price, adv) = \sigma^2$$

ឧបមាថា យើងមានទិន្នន័យដូចខាងក្រោម:

Obs.	Sales	Price	Adv	Stores	Staffs	log(sales)	log(price)	Log(adv)
1	1,215	36	24	12	129	7.102499356	3.583518938	3.1780538
2	2,189	28	45	11	239	7.691200098	3.33220451	3.8066625
3	6,765	10	54	21	138	8.819517541	2.302585093	3.988984
4	4,231	3	21	78	179	8.350193651	1.098612289	3.0445224
5	3,453	6	32	49	122	8.146998697	1.791759469	3.4657359
6	2,452	8	45	45	210	7.804659297	2.079441542	3.8066625
7	3,136	21	12	67	134	8.050703381	3.044522438	2.4849066
8	5,238	12	32	98	127	8.563695025	2.48490665	3.4657359
9	6,458	9	29	12	134	8.773074951	2.197224577	3.3672958
10	1,234	34	17	98	132	7.118016204	3.526360525	2.8332133
11	2,231	24	23	67	123	7.710205194	3.17805383	3.1354942
12	3,854	11	65	43	245	8.256866849	2.397895273	4.1743873
13	2,436	22	32	21	189	7.798112629	3.091042453	3.4657359
14	7,236	10	68	34	165	8.886823847	2.302585093	4.2195077
15	4,231	32	32	67	245	8.350193651	3.465735903	3.4657359
16	7,329	9	43	89	178	8.89959436	2.197224577	3.7612001
17	4,231	14	42	54	134	8.350193651	2.63905733	3.7376696
18	2,544	7	31	39	156	7.841492924	1.945910149	3.4339872
19	9,453	4	87	78	252	9.15408743	1.386294361	4.4659081
20	6,454	8	45	65	178	8.772455373	2.079441542	3.8066625
21	5,324	15	46	78	198	8.57998018	2.708050201	3.8286414
22	7,564	8	27	49	295	8.93115543	2.079441542	3.2958369
23	2,439	43	33	48	190	7.799343398	3.761200116	3.4965076
24	2,457	13	23	68	139	7.806696373	2.564949357	3.1354942
25	1,435	42	15	92	154	7.268920128	3.737669618	2.7080502
26	7,345	12	43	34	165	8.901775089	2.48490665	3.7612001
27	4,234	8	65	56	189	8.350902452	2.079441542	4.1743873

28	5,232	22	34	78	145	8.562548893	3.091042453	3.5263605
29	6,657	21	47	12	187	8.803424212	3.044522438	3.8501476
30	3,453	12	24	19	156	8.146998697	2.48490665	3.1780538
31	2,440	23	23	56	156	7.799753318	3.135494216	3.1354942
32	6,458	14	68	78	194	8.773074951	2.63905733	4.2195077

៧.២ ការស្វែងរកheteroskedasticity

បច្ចុប្បន្ននេះ មានវិធីសាស្ត្រមួយចំនួនដែលត្រូវបានប្រើប្រាស់ដើម្បីស្វែងរកheteroskedasticity។ ក្នុងមេរៀននេះ យើងនឹងលើកយកវិធីសាស្ត្រមួយចំនួនដែលគេនិយមអនុវត្តមកបង្ហាញ ដោយចាប់ផ្តើមជាមួយម៉ូដែលលីនេអ៊ែរ:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u,$$

លក្ខខណ្ឌពីទី១ដល់ទី៤ (MLR.1 ដល់ MLR.4)នៅតែត្រូវបានអនុវត្តន៍។ ជាក់ស្តែង យើងសន្មតថា $E(u|x_1, x_2, \dots, x_k) = 0$ ដូច្នេះ OLSមិនមានកំហុសឆ្គង និង ប្រកបដោយនិរន្តរ៍ភាព។

សម្មតិកម្មដើម (null hypothesis)នៃលក្ខខណ្ឌទី៥(MLR.5)ពិត ក្នុងករណីដែល:

$$H_0: Var(u|x_1, x_2, \dots, x_k) = \sigma^2, \qquad (៧.៤)$$

ប្រសិន យើងមិនអាចបដិសេធ(៧.៤) នៅឯកំរិតសារៈសំខាន់ដូចជា 5% ឬ 1% នោះយើងអាចសន្និដ្ឋានបានថា វាមិនមានបញ្ហាheteroskedasticityកើតឡើង។

អាស្រ័យដោយយើងសន្មតថា $E(u) = 0$ និង $Var(u|x) = E(u^2|x)$ ដូច្នេះ សម្មតិកម្មដើមនៃhomoskedasticityស្មើនឹង:

$$H_0: E(u^2|x_1, x_1, \dots, x_k) = E(u^2) = \sigma^2$$

ហេតុនេះដើម្បីធ្វើតេស្តសក្ខខណ្ឌhomoskedasticity យើងចាំបាច់ធ្វើតេស្ត ថាតើu^2មានទំនាក់ទំនងជាមួយអញ្ញត្តិx_jមួយ ឬ ពីរ...ដែរឬទេ។ ប្រសិន សម្មតិកម្មដើម(H_0)មិនត្រូវបានឆ្លើយតប នោះតម្លៃរំពឹងទុកនៃu^2មានទំនាក់ទំនងជាមួយអញ្ញត្តិx_jណាមួយ។ ដើម្បីដំណោះស្រាយនូវបញ្ហានេះ យើងត្រូវតែសន្មតអនុគមន៍លីនេអ៊ែរមួយដូចខាងក្រោម:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v, \qquad (៧.៥)$$

vគឺជាerror termជាមួយ $E(v|x_j) = 0$។ ចូរចាប់អារម្មណ៍ជាមួយអញ្ញត្តិឯករាជ្យ(u^2)នៃសមីការ(៧.៥) ដែលវាជាការរៃនៃerror termនៃសមីការ(៧.១)។ សម្មតិកម្មដើម(H_0)នៃhomoskedasticityគឺ

$$H_0: \delta_1 = \delta_2 = \cdots = \delta_k = 0 \qquad (\text{៧.៥})$$

ផ្អែកលើសម្មតិកម្មដើម(H_0)នេះ យើងសន្មតថា vមិនមានទំនាក់ទំនងជាមួយ$x_1, x_2, \ldots, x_k$ឡើយ។ បន្ទាប់មក ការធ្វើតេស្តF ឬ *LM* statisticsសំរាប់សារៈសំខាន់ទាំងមូល(overall significance)នៃអញ្ញត្តិឯករាជ្យ(x_j)ក្នុងការពន្យល់ទៅលើu^2 វាអាចត្រូវបានប្រើប្រាស់ដើម្បីធ្វើតេស្ត(៧.៥)។ ទាំងF និង *LM* statisticsអាចអនុវត្តបាន ទោះបីជាu^2មិនមានលក្ខណៈជាបំណែងចែកធម្មតា(normal distribution) ក៏ដោយ។ ក្នុងសំណាក(sample) បើយើងអាចអង្កេតu^2 បន្ទាប់មកងាយស្រួលគណនាF និង *LM* statistics ដោយអនុវត្តOLS estimatorsនៃu^2ជាអនុគមន៏ជាមួយ$x_1, x_2, \ldots, x_k$ ដោយប្រើប្រាស់រាល់ការអង្កេត(n observations)។

ដូចបានបង្ហាញខាងលើ យើងមិនដែលដឹងលំអៀងជាស្តែង(actual errors)ទាំងឡាយនៅក្នុងប៉ូពុយឡាស្យុង(population)ឡើយ។ ប៉ុន្តែ យើងបានប៉ាន់ស្មានលំអៀង(errors)ទាំងនោះ។ ដូច្នេះយើងអាចប៉ាន់ស្មានសមីការដូចខាងក្រោម:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + error \qquad (\text{៧.៦})$$

និង គណនាF ឬ *LM* statisticsសំរាប់សារៈសំខាន់បញ្ចូលគណនា(joint significance)នៃ$x_1, x_2, \ldots, x_k$។

ទាំងF និង *LM* statisticsផ្អែកទៅលើR^2ដែលទទួលបានពីរ៉ីហ្គ្រេស្យុង(៧.៦) នោះគឺ $R^2_{\hat{u}^2}$ ដែលខុសពីR^2ដែលទទួលបានពីរ៉ីហ្គ្រេស្យុង(៧.១)។ បន្ទាប់មក Fstatisticគឺ

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)} \sim F_{k,n-k-1} \qquad (\text{៧.៧})$$

ដែល kគឺជាចំនួននៃអញ្ញត្តិ$x_1, x_2, \ldots, x_k$ ក្នុង(៧.៦)ដូចគ្នានឹង(៧.១)ដែរ។

ចំណែក *LM* statisticសំរាប់heteroskedasticityគឺ

$$LM = n * R^2_{\hat{u}^2} \sim x^2_k \qquad (\text{៧.៨})$$

ការធ្វើតេស្ត*LM* statisticត្រូវបានគេហៅថា **Breusch-Pagan test for heteroskedasticity (BP test)** និង វិធីសាស្ត្រនេះត្រូវបានអនុវត្តដូចខាងក្រោម:

៧.២.១ ការអនុវត្តន៏ Breusch-Pagan test សំរាប់ heteroskedasticity:

១. ប៉ាន់ស្មានម៉ូដែល(៧.១)ដោយOLS។ បន្ទាប់មកទទួលបានការរៃនៃOLS residuals ($\hat{u}^2$) មួយសំរាប់ការអង្កេតនីមួយៗ (one for each observation)។

២. ដំណើរការរ៉ឺគ្រ្ហេស្សុង(៧.៦) និង កត់ត្រាR² ពីរ៉ឺគ្រ្ហេស្សុងនេះ ($R^2_{\hat{u}^2}$)។

៣. កំណត់F statistic ឬ LM statistic និង គណនាតំលៃP ($P-$ value)ដោយការប្រើប្រាស់ បំណែងចែក$F_{k,n-k-1}$ និង បំណែងចែកx_k^2។ ប្រសិន តំលៃP តូចជាងកំរិតសារៈសំខាន់ដែលបានជ្រើសរើស បន្ទាប់មកយើងបដិសេធសម្មតិកម្មដើម(H_0)នៃhomoskedasticity(៧.៥)។

យោងលើ ***ឧទាហរណ៍ ៧.១*** និង ទិន្នន័យខាងលើ យើងបកស្រាយheteroskedasticityនៃសមីការចំណូល(sales)របស់ក្រុមហ៊ុនដូចខាងក្រោម:

$$\widehat{sales} = \underset{(1,554.390)}{2,541.099} - \underset{(29.651)}{81.091}price + \underset{(20.530)}{50.915}adv + \underset{(10.999)}{7.075}stores + \underset{(7.520)}{5.262}staffs \quad (៧.៩)$$
$$n = 32, R^2 = 0.5277$$

សមីការនេះមិនបានបង្ហាញឲ្យឃើញថា តើវាមានបញ្ហាheteroskedasticityកើតឡើង ឬមិនកើតឡើងនោះទេ។ ដូច្នេះ យើងត្រូវដំណើរការរ៉ឺគ្រ្ហេស្សុង$\hat{u}^2$នៃ(៧.៩)ជាអនុគមន៍ទៅលើprice adv stores និង staffs នោះយើងទទួលបាន$R^2_{\hat{u}^2}$= 0.3638។ ដោយ n = 32 និង k = 4 នោះF statisticគឺ

$$F = \frac{0.3638/4}{(1-0.3638)/(32-4-1)} \approx 3.86 > F_{4,\ 32-4-1} = 2.73 \text{ នៅឯកំរិត } 5\%$$

ចំណែកឯ LM statisticគឺ LM = 32 * 0.3638 = 11.64 > x_k^2 = 9.49 នៅឯកំរិត 5% ជាមួយតម្លៃP ($P-$ value) = 0.0132 តូចជាង 5%។ ដាលទ្ធផល វាមានបញ្ហាheteroskedasticityកើតឡើង។

ឥឡូវយើងសាកល្បងបំពាក់ទម្រង់ឡូការីត (logarithmic functional form)ទៅលើអញ្ញត្តិនៃសមីការ និង ទទួលបានលទ្ធផលដូចខាងក្រោម:

$$\widehat{\log(sales)} = \underset{(1.785)}{5.516} - \underset{(0.123)}{0.334}\log(price) + \underset{(0.197)}{0.540}\log(adv)$$
$$+\underset{(0.261)}{0.117}\log(stores) + \underset{(0.820)}{0.226}\log(staffs) \quad (៧.១០)$$
$$n = 32, R^2 = 0.5288$$

សមីការ(៧.១០)មិនបានបង្ហាញឲ្យឃើញថា តើវាមានបញ្ហាheteroskedasticityកើតឡើង ឬ មិនកើតឡើងនោះទេ។ ដូច្នេះ យើងត្រូវដំណើរការរ៉ឺគ្រ្ហេស្សុង$\hat{u}^2$នៃសមីការនេះជាអនុគមន៍ទៅលើlog(price) log(adv) log(stores) និង log(staffs) នោះយើងទទួលបាន$R^2_{\hat{u}^2}$= 0.1732។ ដោយ n = 32 និង k = 4 នោះF statisticគឺ

$$F = \frac{0.1732/4}{(1-0.1732)/(32-4-1)} \approx 1.41 < F_{4,\ 32-4-1} = 2.73 \text{ នៅឯកំរិត } 5\%$$

ចំណែកឯ *LM* statisticគឺ $LM = 32 * 0.1732 = 5.54 < x_k^2 = 9.49$ នៅឯកំរិត 5% ជាមួយតម្លៃ *P* (*P* – value) = 0.2562 ធំជាង 5%។ ជាលទ្ធផល វាមិនមានបញ្ហា**heteroskedasticity**កើតឡើង។ ដូច្នេះ យើងបរាជ័យក្នុងការបដិសេធសម្មតិកម្មដើម(H_0)នៃhomoskedasticityនៃម៉ូដែលជាមួយបំពាក់ទម្រង់ឡូការីត (logarithmic functional form) ទៅលើអញ្ញត្តិរបស់វា។

កត់សំគាល់: ប្រសិន$\hat{u}^2$នៃសមីការដែលមានអញ្ញត្តិឯករាជ្យតែមួយ(simple regression model) នោះវិធីសាស្ត្រក្នុងការធ្វើតេស្តចំពោះ**heteroskedasticity** គឺគ្រាន់តែអនុវត្តt statisticធម្មតាទៅលើអញ្ញត្តិឯករាជ្យ(independent variable)។ ករណីរកឃើញមានសារៈសំខាន់ជាស្ថិតិកើតឡើង នោះបញ្ជាក់ថា វាមានបញ្ហា**heteroskedasticity**កើតឡើង។

៧.២.២ ការអនុវត្តន៍ White Test សំរាប់ heteroskedasticity:

រាល់គំលាតស្តង់ដា(standard errors) និងការធ្វើតេស្តស្ថិតិ(test statistics)នៃOLSមានសុពលភាពលុះត្រាតែវាឆ្លើយតបទៅនឹងGauss-Markov assumptions ដែលក្នុងនោះលក្ខខណ្ឌទី៥(MLR.5)គឺ homoskedasticity assumption: $Var(u|x_1, x_2, \dots, x_k) = \sigma^2$។ ដើម្បីធ្វើតេស្ត **heteroskedasticity** អ្នកអនុវត្តអាចប្រើវិធីសាស្ត្រផ្សេងទៀតដោយបន្ថែមការេនៃបណ្តាអញ្ញត្តិឯករាជ្យ (x_j^2) និង $(x_j x_h \; for \; j \neq h)$ទៅក្នុងសមីការ(៧.៦) (White, 1980)។ *ឧទាហរណ៍* នៅពេលម៉ូដែលមួយដែលមានអញ្ញត្តិឯករាជ្យពីរ($k = 2$) White testត្រូវបានផ្តោតលើការប៉ាន់ស្មាន:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 + \delta_5 x_1 x_2 + error. \qquad (៧.១១)$$

White Test សំរាប់ **heteroskedasticity** គឺជា *F*ឬ *LM* statistic ជាមួយសម្មតិកម្មដើម(H_0): រាល់ δ_jក្នុងសមីការ(៧.១១)ស្មើសូន្យ លើកលែងδ_0។ នៅពេលសមីការដែលមានអញ្ញត្តិឯករាជ្យកាន់តែច្រើន វាមានការលំបាកក្នុងការអនុវត្តន៍។ ដើម្បីជាជំនួយឲ្យការអនុវត្តន៍មានភាពងាយស្រួល យើងប្រើប្រាស់ OLS fitted value $(\hat{y})$ក្នុងការធ្វើតេស្ត **heteroskedasticity**។

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots \hat{\beta}_k x_{ik},$$

ប្រសិន លើករាល់តំលៃ$\hat{y}_i$ជាការេ នោះយើងទទួលបានរាល់បណ្តាអញ្ញត្តិឯករាជ្យ(x_j^2) និង $(x_j x_h \; for \; j \neq h)$។ ដូច្នេះ ការការធ្វើតេស្ត**heteroskedasticity** គេអាចអនុវត្តដោយប៉ាន់ស្មានសមីការ:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + error, \qquad (៧.១២)$$

យើងអាចប្រើប្រាស់*F* ឬ *LM* statistic សំរាប់សម្មតិកម្មដើម(H_0)គឺ $H_0: \delta_1 = 0, \delta_2 = 0$ ក្នុងសមីការ(៧.១២)។

ករណីពិសេសនៃ White Test សំរាប់ heteroskedasticity:

១. ប៉ាន់ស្មានម៉ូដែល(៧.១)ដោយOLS។ បន្ទាប់មកទទួលបានរាល់OLS residuals ($\hat{u}$) និង fitted value ($\hat{y}$)។ គណនាOLS residuals និង fitted valueជាការេ ($\hat{u}^2$ និង $\hat{y}^2$)។

២. ដំណើរការរ៉ឺគ្រ្ចេស្យុង(៧.១២) និង កត់ត្រាR² ពីរ៉ឺគ្រ្ចេស្យុងនេះ ($R^2_{\hat{u}^2}$)។

៣. កំណត់*F* statistic ឬ *LM* statistic និង គណនាតម្លៃ*P* (*P*– value)ដោយការប្រើប្រាស់បំណែងចែក$F_{2,n-3}$ និង បំណែងចែកx_2^2។ ប្រសិន តំលៃ*P* តូចជាងកំរិតសារៈសំខាន់ដែលបានជ្រើសរើស បន្ទាប់មកយើងបដិសេធសម្មតិកម្មដើម(H_0)នៃhomoskedasticity: $H_0: \delta_1 = 0, \delta_2 = 0$។

ដូច្នេះ *F* statisticគឺ

$$F = \frac{R^2_{\hat{u}^2}/2}{(1 - R^2_{\hat{u}^2})/(n-3)} \sim F_{2,n-3} \qquad (៧.១៣)$$

White Test សំរាប់heteroskedasticityនៃសមីការចំណូល(sales)របស់ក្រុមហ៊ុន:

១. ផ្អែកលើសមីការ(៧.៩) យើងទាញបានតម្លៃresiduals ($\hat{u}$) និង ($\widehat{sales}$)។ បន្ទាប់មកលើកតម្លៃទាំងពីរជាការេ ($\hat{u}^2$ និង $\widehat{sales}^2$)។

២. ដំណើរការរ៉ឺគ្រ្ចេស្យុង: $\hat{u}^2 = \delta_0 + \delta_1\widehat{sales} + \delta_2\widehat{sales}^2 + error,$ និង ទាញបាន$R^2_{\hat{u}^2} = 0.1949$

៣. ផ្អែកលើ(៧.១៣) កំណត់បានតម្លៃ*F* statisticដូចខាងក្រោម:

$$F = \frac{0.1949/2}{(1-0.1949)/(32-2-1)} \approx 3.51 > F_{2,29} = 3.33 \text{ នៅឯកំរិត } 5\%$$

ចំណែកឯ *LM* statisticគឺ *LM* = 32 *0.1949 = 6.24 > x_2^2 = 5.99 នៅឯកំរិត 5% ជាមួយតម្លៃ*P* (*P*– value) = 0.043 តូចជាង 5%។ ជាលទ្ធផល វាមានបញ្ហា**heteroskedasticity**កើតឡើង។

ឥឡូវយើងសាកល្បងបំពាក់ទម្រង់ឡូការីត (logarithmic functional form)ទៅលើអញ្ញត្តិនៃសមីការ(៧.៣) ជាមួយការអនុវត្តន៍ដូចខាងក្រោម:

១. ផ្អែកលើសមីការ(៧.១០) យើងទាញបានតំលៃresiduals ($\hat{u}$) និង log ($\widehat{sales}$)។ បន្ទាប់មកលើកតំលៃទាំងពីរជាការេ ($\hat{u}^2$ និង log $(\widehat{sales})^2$)។

២. ដំណើរការរ៉ឺគ្រ្ចេស្យុង: $\hat{u}^2 = \delta_0 + \delta_1 \log(\widehat{sales}) + \delta_2 \log(\widehat{sales})^2 + error,$ និង ទាញបាន $R^2_{\hat{u}^2} = 0.0134$

៣. ផ្អែកលើ(៧.១៣) កំណត់បានតម្លៃF statisticដូចខាងក្រោម:

$$F = \frac{0.0134/2}{(1-0.0134)/(32-2-1)} \approx 0.197 < F_{2,29} = 3.33$$ នៅឯកំរិត 5%

ចំណែកឯ *LM* statisticគឺ *LM* = 32 *0.0134 = 0.4288 < x_2^2 = 5.99 នៅឯកំរិត 5% ជាមួយតម្លៃ*P* (*P*– value) = 0.822 ធំជាង 5%។ ដូច្នេះ យើងបរាជ័យក្នុងការបដិសេធសម្មតិកម្មដើម(H_0)នៃhomoskedasticityនៃម៉ូដែលជាមួយបំពាក់ទម្រង់ឡូការីត (logarithmic functional form) ទៅលើអញ្ញត្តិរបស់វា។

៧.៣ Heteroskedasticity-Robust inference ក្រោយពេលប៉ាន់ស្មានOLS

ប្រសិនHeteroskedasticityត្រូវបានរកឃើញដោយការប្រើវិធីសាស្ត្រ Breusch-Pagan test ឬ White Test ដំណោះស្រាយមួយក្នុងចំណោមផ្សេងៗដែលត្រូវបានអនុវត្តគឺ heteroskedasticity-robust statisticsក្រោយពេលប៉ាន់ស្មានOLS។

យើងត្រូវតែសម្រេចចិត្ត ថាតើគួរតែបោះបង់ចោលOLS estimatorsដែរឬទេ ពីព្រោះ ការធ្វើតេស្តសម្មតិកម្ម វាមានសារៈសំខាន់ និង OLS inferenceមិនមានសុពលភាព ករណីមានបញ្ហា heteroskedasticityកើតឡើង។ ហេតុនេះ អ្នកជំនាញ (econometricians)បានសិក្សា តើstandard error, *t*, *F* និង *LM* statisticsត្រូវបានកែតំរូវយ៉ាងដូចម្តេច ដើម្បីឲ្យវាមានសុពលភាព ក្នុងពេលមានបញ្ហា heteroskedasticityកើតឡើង។ វិធីសាស្ត្រដែលអ្នកជំនាញបានបង្កើតឡើងនេះ គេហៅថានិតិវិធី*heteroskedasticity-robust* និង វាមានសុពលភាពសំរាប់អនុវត្តន៍ ក្នុងករណីការស្រាវជ្រាវផ្អែកលើចំនួនសំណាកច្រើន (large samples) សំរាប់រ៉ឺគ្រេស្យុង:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

៧.៣.១ គណនាHeteroskedasticity-Robust standard error

យោងលើលំក្ខខណ្ឌគោសម៉ាកកូវ (Gauss-Markov assumptions)ទី១ដល់ទី៤ (MLR.1 ដល់ MLR.4) estimatorនៃ$Var(\hat{\beta}_j)$ដែលប្រកបដោយសុពលភាពគឺ

$$\widehat{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}, \qquad (៧.១៤)$$

- ដែល $\hat{r}_{ij}$តំណាងឲ្យលំអៀង(residual)ទីi^{th}នៃរ៉ឺគ្រេស្យុងx_jជាអនុគមន៍ជាមួយអញ្ញត្តិឯករាជ្យ(independent variables)ដទៃទៀត។
- SSR_j^2គឺជាផលបូកការេនៃលំអៀង(sum of squared residuals)នៃរ៉ឺគ្រេស្យុងyជាអនុគមន៍ជាមួយរាល់អញ្ញត្តិឯករាជ្យ(x_j)ទាំងឡាយ។

- $\hat{u}_i$គឺជាOLS residualsនៃរ៉ឺហ្គ្រេស្យុងyជាអនុគមន៍ជាមួយរាល់អញ្ញត្តិឯករាជ្យ(x_j)ទាំងឡាយ។

ដូច្នេះ ឫសការេ (square root)នៃ$\widehat{Var}(\hat{\beta}_j)$ត្រូវបានគេហៅថា**heteroskedasticity-robust standard error** សំរាប់$\hat{\beta}_j$។

នៅពេលទទួលបាន**heteroskedasticity-robust standard error** គេនឹងទាញបាន **heteroskedasticity-robust *t* statistic** ដូចខាងក្រោម:

$$t = \frac{estimate - hypothesized\ value}{standard\ error} \qquad (៧.១៥)$$

ភាពខុសគ្នារវាងOLS t statisticធម្មតា និង heteroskedasticity-robust t statistic គឺជាstandard error ពីព្រោះ យើងនៅតែប្រើប្រាស់មេគុណប៉ាន់ស្មាន(OLS estimates) និង ជ្រើសរើសhypothesized value។

យោងលើសមីការ(៧.១០) (Log sales equation with heteroskedasticity-robust standard errors) យើងអាចប្រៀបធៀបOLS standard errorsធម្មតាជាមួយheteroskedasticity-robust standard errors។

$$\widehat{\log(sales)} = \underset{\substack{(1.785)\\[1.901]}}{6.516} - \underset{\substack{(0.123)\\[0.112]}}{0.334}\log(price) + \underset{\substack{(0.197)\\[0.188]}}{0.540}\log(adv)$$

$$+\underset{\substack{(0.261)\\[0.292]}}{0.117}\log(stores) + \underset{\substack{(0.820)\\[0.919]}}{0.226}\log(staffs) \qquad (៧.១៦)$$

$$n = 32, R^2 = 0.5288$$

វង្វង់ក្រចក( )ជាstandard errorsធម្មតា និង heteroskedasticity-robust standard errors ស្ថិតក្នុងវង់ដង្កៀប[ ]។ នៅក្នុងសមីការ(៧.១៦)មានចំនុចមួយចំនួនបានបង្ហាញគឺ

១. អញ្ញត្តិដែលមានសារៈសំខាន់ជាស្ថិតិ(statistical significance)ដោយប្រើ t statisticធម្មតា គឺនៅតែមានសារៈសំខាន់ជាស្ថិតិនៅពេលប្រើheteroskedasticity-robust t statistic ពីព្រោះ standard errorsទាំងពីរគូរមិនខុសគ្នាច្រើនទេ។

២. Robust standard errorsអាចធំជាង ឬ តូចជាងstandard errorsធម្មតា។ ភាគច្រើន Robust standard errorsធំជាងstandard errorsធម្មតា។

កត់សំគាល់

- ✓ ប្រសិន លក្ខខណ្ឌhomoskedasticityត្រូវបានឆ្លើយតប និង errors (u)មានលក្ខណៈជាបំណែងចែកធម្មតា(normal distribution) នោះ *t* statisticធម្មតាក៏មានបំណែងបែក *t* ច្បាស់លាស់(*exact t* distributions)។
- ✓ Robust standard errors និង robust *t* statisticsត្រូវបានអនុវត្ត ក្នុងករណីដែលសំណាក (sample size) មានចំនួនច្រើន ពីព្រោះ បើចំនួនសំណាកតិច នោះrobust *t* statisticsមិនអាចខិតទៅរក *exact t* distributions។
- ✓ ជាទូទៅ heteroskedasticity-robust standard errorsអាចអនុវត្តបានតែជាមួយcross-sectional data។

Heteroskedasticity-robustក៏អាចអនុវត្តទៅលើ *F* និង *LM* statistics។ ក្នុងនោះ heteroskedasticity-robust *F* statisticត្រូវបានគេហៅថា *heteroskedasticity-robust Wald statistic*។

៧.៣.២ គណនាHeteroskedasticity-Robust *F* statistic

សន្មតថា យើងមានបំណងធ្វើតេស្តសម្មតិកម្មដើម: $H_0: \beta_{\log(stores)} = 0, \beta_{\log(staffs)} = 0$។ នៅពេលយើងទទួលបាន$R^2 = 0.5237$ពីrestricted modelជាមួយ heteroskedasticity-robust និង $R^2 = 0.5288$ពីunrestricted modelក៏ជាមួយheteroskedasticity-robustផងដែរ បន្ទាប់មក heteroskedasticity-robust *F* statisticគឺ

$$F = \frac{(0.5288 - 0.5237)/2}{(1 - 0.5288)/(32 - 4 - 1)} \approx 0.146$$

ហេតុនេះ យើងបរាជ័យក្នុងការបដិសេធសម្មតិកម្មដើម(H_0)។

៧.៣.៣ គណនាHeteroskedasticity-Robust *LM* Tests

ពិចារណាម៉ូដែលខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u,$$

យើងចង់ធ្វើតេស្ត$H_0: \beta_3 = 0, \beta_4 = 0$។ ក្នុងការគណនា*LM* statisticធម្មតា ដំបូងយើងប៉ាន់ស្មានrestricted model (ម៉ូដែលមិនមានx_3 និង x_4) ដើម្បីទទួលបានresiduals ($\tilde{u}$)។ បន្ទាប់មក យើងដំណើរការសមីការ $\tilde{u}$ជាអនុគមន៍នៃរាល់អញ្ញត្តិឯករាជ្យ ($x_1, x_2, x_3,$ និង x_4) និង $LM = n * R_u^2$ ដែល R_u^2គឺជាR²ធម្មតានៃសមីការខាងលើ។

ដំណើរការជាមួយheteroskedasticity-robust *LM* statistic គឺវាមានជំហានជាច្រើន។ មធ្យោបាយមួយដើម្បីគណនាheteroskedasticity-robust *LM* statistic គឺត្រូវការបណ្តារ៉ឺហ្គ្រេស្យុង OLS។ យើងកំណត់residuals($\tilde{r}_1$)នៃរ៉ឺហ្គ្រេស្យុងx_3ជាអនុគមន៍នៃx_1និង x_2 ហើយ residuals($\tilde{r}_2$)នៃរ៉ឺហ្គ្រេស្យុងx_4ជាអនុគមន៍នៃx_1និង x_2។ ជំហានចុងក្រោយ ដំណើការរ៉ឺហ្គ្រេស្យុងដូចខាងក្រោម:

$$1 \; on \; \tilde{r}_1\tilde{u}, \tilde{r}_2\tilde{u}, \qquad (\text{៧.១៧})$$

មិនមានinterceptឡើយ។ រាល់ការអង្កេត(observations)ទាំងឡាយ អញ្ញត្តិមិនឯករាជ្យមានតម្លៃស្មើមួយ។ Robust *LM* statisticគឺ $n - SSR_1$ ដែលSSR_1ជាSSRនៃ(៧.១៧)។

A Heteroskedasticity-Robust *LM* Statistic:

១. កំណត់ residuals($\tilde{u}$)ពីrestricted model។

២.ផ្អែកលើសម្មតិកម្មដើម ដំណើការរ៉ឺហ្គ្រេស្យុងនៃអញ្ញត្តិឯករាជ្យនីមួយៗដែលមិនត្រូវបានដាក់បញ្ចូលជាអនុគមន៍នៃបណ្តាអញ្ញត្តិឯករាជ្យទាំងឡាយដែលត្រូវបានដាក់បញ្ចូល។ ប្រសិន មានចំនួនqនៃអញ្ញត្តិឯករាជ្យដែលមិនត្រូវបានដាក់បញ្ចូល នេះទទួលបានចំនួនqនៃ residuals $(\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_q)$។

៣. គណនា $\tilde{r}_j\tilde{u}$ នីមួយៗ សំរាប់រាល់ការអង្កេតទាំងឡាយ(all observations)។

៤. ដំណើរការរ៉ឺហ្គ្រេស្យុងនៃ 1ជាអនុគមន៍នៃ$\tilde{r}_1\tilde{u}, \tilde{r}_2\tilde{u}, \dots, \tilde{r}_q\tilde{u},$ ដោយមិនមានintercept។ Heteroskedasticity-robust *LM* statisticគឺ $n - SSR_1$, ដែលSSR_1ជាSSRធម្មតាដែលទទួលបានពីរ៉ឺហ្គ្រេស្យុងចុងក្រោយនេះ។ ផ្អែកលើH_0, *LM* ត្រូវប្រៀបធៀបជាមួយx_q^2។ បើ *LM* តូចជាងx_q^2 នោះបរាជ័យក្នុងការបដិសេធសម្មតិកម្មដើម។

៧.៤ ការប៉ាន់ស្មានWeighted Least Squares (WLS)

ក្រោយពេលប៉ាន់ស្មានដោយប្រើOLS estimators ប្រសិន Heteroskedasticityត្រូវបានរកឃើញតាមរយៈវិធីសាស្ត្រ **Breusch-Pagan test** ឬ **White test** នោះគេអាចឆ្លើយតបបញ្ហានេះជាមួយនឹងការអនុវត្តន៍heteroskedasticity-robust statistics។ ម៉្យាងទៀត គេអាចប្រើម៉ូដែល Weighted Least Squares (WLS) ពីព្រោះ វាមានប្រសិទ្ធភាពជាងOLS។ ការអនុវត្តន៍WLS គេទទួលបានស្ថិតិt និង F($t$and Fstatistics)គឺជាបំណែងចែកt និង F($t$and Fdistributions)។

៧.៤.១ Heteroskedasticityត្រូវបានគេស្គាល់

$\boldsymbol{x}$ គឺជាអញ្ញត្តិឯករាជ្យទាំងឡាយ(independent variables)ក្នុងសមីការ(៧.១) និង សន្មតថា

$$Var(u|x) = \sigma^2 h(x), \qquad (\text{៧.១៨})$$

$h(x)$គឺជាមុខងារនៃអញ្ញត្តិ$\boldsymbol{x}$ទាំងឡាយដែលកំណត់នូវheteroskedasticity។ នៅពេលតម្លៃ (variances)ត្រូវតែវិជ្ជមាន នោះ$h(x) > 0$សំរាប់រាល់តម្លៃនៃអញ្ញត្តិឯករាជ្យទាំងឡាយ។ សន្មតថាមុខងារ $h(x)$ត្រូវបានគេស្គាល់។

នៅពេលយើងតាង $\boldsymbol{x}$ គឺជាអញ្ញត្តិឯករាជ្យទាំងឡាយសំរាប់ការអង្កេតi (observation i) និង h_iប្រែប្រួលជាមួយការអង្កេតនីមួយៗ ពីព្រោះអញ្ញត្តិឯករាជ្យទាំងឡាយប្រែប្រួលតាមរយៈការអង្កេតទាំងឡាយ។ ហេតុនេះ យើងអាចសរសេរបាន $\sigma_i^2 = Var(u_i|x_i) = \sigma^2 h(x_i) = \sigma^2 h_i$។

តើគេអាចប្រើប្រាស់ពត៌មានក្នុងសមីការ(៧.១៨) ដើម្បីប៉ាន់ស្មានβ_jយ៉ាងដូចម្តេច? គេអាចកំណត់សមីការដើម(៧.១៩)ដែលរួមមានheteroskedastic errors និង កែសំរួលទៅជាសមីការមួយដែលមានhomoscedastic errors ដូចខាងក្រោម:

$$y_i = \beta_0 + \beta_i x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \qquad (៧.១៩)$$

នៅពេលh_iគឺជាមុខងារនៃx_i នោះ$u_i/\sqrt{h_i}$មិនមានទំនាក់ទំនងជាមួយx_iឡើយ។ បន្ថែមពីនេះ នៅពេល$Var(u_i|x_i) = E(u_i^2|x_i) = \sigma^2 h_i$ នោះវ៉ារីយ៉ង់នៃ$u_i/\sqrt{h_i}$ គឺ σ^2:

$$E\left((u_i/\sqrt{h_i})^2\right) = E(u_i^2)/h_i = (\sigma^2 h_i)/h_i = \sigma^2,$$

គេអាចចែកសមីការ(៧.១៩)ជាមួយ$\sqrt{h_i}$ នោះទទួលបាន:

$$y_i/\sqrt{h_i} = \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \beta_2(x_{i2}/\sqrt{h_i}) + \cdots + \beta_k(x_{ik}/\sqrt{h_i}) + (u_i/\sqrt{h_i}) \qquad (៧.២០)$$

ឬ

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \cdots + \beta_k x_{ik}^* + u_i^*, \qquad (៧.២១)$$

ដែល$x_{i0}^* = 1/\sqrt{h_i}$ និង បណ្តាអញ្ញត្តិដទៃទៀតតំណាងឲ្យអញ្ញត្តិដើមនៃ(៧.១៩)ចែកនឹង$\sqrt{h_i}$។

សមីការ(៧.២១)គឺ $E(u_i^*|x_i^*) = 0$ និង $Var(u_i^*|x_i^*) = \sigma^2$។ ប្រសិនសមីការដើម(៧.១)ឆ្លើយតបGauss-Markov assumptionsពីទី១ដល់ទី៤ បន្ទាប់មកការបំលាស់ប្តូរទៅជាសមីការ(៧.២១) ឆ្លើយតបGauss-Markov assumptionsពីទី១ដល់ទី៥។ បើ u_iមានបំណែងចែកធម្មតា (normal distribution) បន្ទាប់មកu_i^*ក៏មានបំណែងចែកធម្មតាជាមួយវ៉ារីយ៉ង់σ^2។ ដូច្នេះ សមីការ (៧.២១) ឆ្លើយតបGauss-Markov assumptionsពីទី១ដល់ទី៦។

ផ្អែកលើ*Gauss-Markov assumptions* OLSត្រូវតែជាសមីការប៉ាន់ស្មានដែលមិនមានកំហុសឆ្គង (best linear unbiased estimator: BLUE)។ ប៉ារ៉ាម៉ែត្រដែលបានប៉ាន់ស្មាននៅក្នុងសមីការដើម (៧.១)ត្រូវតែខុសគ្នាពីសមីការ(៧.២១)។ ប៉ារ៉ាម៉ែត្រβ_j^* គឺជាឧទាហរណ៍នៃgeneralized least

squares (GLS)estimators។ GLS estimatorsត្រូវបានគេយកទៅអនុវត្តសំរាប់កែតម្រូវ heteroskedasticity និង វាត្រូវបានគេហៅថា **weighted least squares (WLS) estimators**។

ឧទាហរណ៍ ពិចារណារ៉ឺគ្រ្វេស្យុងនៃចំណូលរបស់ក្រុមហ៊ុន:

$$sales = \beta_0 + \beta_1 price + \beta_2 adv + \beta_3 stores + \beta_4 staffs + u, \qquad (៧.២១)$$

$$Var(u_i|price_i, adv_i, stores_i, staffs_i) = \sigma^2 adv_i, \qquad (៧.២២)$$

ហេតុនេះ $h(x) = h(adv) = adv$: វ៉ារីយ៉ង់នៃerror គឺសមាមាត្រទៅនឹងកំរិតនៃadv។ នេះមានន័យថា បំរែបំរួលនៃចំណូលកើនឡើង នៅពេលការផ្សាយពាណិជ្ជកម្មកើនឡើង (បើ $\beta_2 > 0$)។ ពីព្រោះ $adv > 0$ វ៉ារីយ៉ង់ក្នុងសមីការ (៧.២២) ក៏វិជ្ជមានផងដែរ។ គំលាតបទដ្ឋាន (standard deviation) នៃu_iទៅលើadv_iគឺ $\sigma\sqrt{adv_i}$។

សមីការ(៧.២១)ត្រូវបានបំលាស់ប្តូរដូចខាងក្រោម:

$$sales_i/\sqrt{adv_i} = \beta_0(1/\sqrt{adv_i}) + \beta_1(price_i/\sqrt{adv_i}) + \beta_2\sqrt{adv_i} + \beta_3(stores_i/\sqrt{adv_i}) + \beta_4(staffs/\sqrt{adv_i}) + u_i^*,$$

ផ្អែកលើទិន្នន័យខាងលើ យើងទទួលបានលទ្ធផលដូចខាងក្រោម:

Dependent Variable: sales

	(1) OLS	(2) WLS
price.....	-81.09	-82.48
	(29.65)	(24.04)
adv.....	50.92	47.11
	(20.53)	(20.70)
stores.....	7.08	5.83
	(10.99)	(9.86)
staffs.....	5.26	9.65
	(7.52)	(6.77)

intercept.....	2,541.10	2,013.29
	(1,554.39)	(1,419.18)
Observation.....	32	32
R-Squared.....	0.5277	0.5776
Heteroskedasticity (BP test):		
F-statistic.....	3.86	2.45
P-value.....	**0.0132**	**0.0705**

យើងឃើញថា ម៉ូដែលOLSធម្មតាមានបញ្ហាheteroskedasticity $(p - value < 0.05)$។ ប៉ុន្តែនៅពេលបំលាស់ប្តូរទៅជាម៉ូដែលWLS លទ្ធផលគឺ homoskedasticity $(p - value > 0.05)$។

៧.៤.២ Heteroskedasticityត្រូវបានប៉ាន់ស្មាន: Feasible GLS

វាពិបាកស្វែងរកមុខងារនៃ $h(x_i)$ ពីព្រោះ ទម្រង់ត្រឹមត្រូវនៃ heteroskedasticityមិនបានបញ្ជាក់ឲ្យច្បាស់លាស់។ យ៉ាងណាក៏ដោយ យើងអាចបង្កើតម៉ូដែលមុខងារ h និង ប្រើប្រាស់ទិន្នន័យដើម្បីប៉ាន់ស្មានបណ្តាប៉ារ៉ាម៉ែត្រដែលគេមិនស្គាល់(unknown parameters)ក្នុងម៉ូដែលនោះគឺ $\hat{h}_i$។ ការប្រើប្រាស់ $\hat{h}_i$ជំនួសឲ្យh_iនៅក្នុងការបម្លាស់ប្តូរGLS នេះទទួលបាននូវការប៉ាន់ស្មានមួយ ដែលគេហៅថា **feasible GLS (FGLS) estimator**។

សន្មតថា

$$Var(u|x) = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k), \qquad (៧.២៣)$$

ដែល$x_1, x_2, \ldots, x_k$គឺជាអញ្ញត្តិឯករាជ្យក្នុងម៉ូដែល(៧.១) និង δ_jគឺជាបណ្តាប៉ារ៉ាម៉ែត្រដែលគេមិនស្គាល់(unknown parameters)។ ហេតុនេះ យោងពីផ្នែកមុន $h(\boldsymbol{X}) = exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)$. បណ្តាម៉ូដែលលីនេអ៊ែរមិនអាចធានាថា តំលៃដែលព្យាករណ៍គឺវិជ្ជមានឡើយ និង ដើម្បីអនុវត្តWLS បណ្តាវ៉ារីយ៉ង់ដែលបានប៉ាន់ស្មានត្រូវតែវិជ្ជមាន។

ប្រសិន បណ្តាប៉ារ៉ាម៉ែត្រδ_jត្រូវបានគេស្គាល់ បន្ទាប់មកយើងអនុវត្តWLS។ ប៉ុន្តែ វាមិនប្រាកដឡើយ។ ដូច្នេះ តើយើងអាចប៉ាន់ស្មានδ_jយ៉ាងដូចម្តេច? សំខាន់ យើងនឹងបំលាស់ប្តូរសមីការនេះទៅជាទម្រង់លីនេអ៊ែរដែលអាចត្រូវបានប៉ាន់ស្មានដោយOLS។

ក្រោមលក្ខខណ្ឌ(៧.២៣) យើងអាចសរសេរ:

$$u^2 = \sigma^2 exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k)v,$$

ដែល vមានមធ្យមស្មើមួយ និង $\boldsymbol{X} = (x_1, x_2, \dots, x_k)$។ យើងសន្មតថា vមិនមានទំនាក់ទំនងជាមួយ$\boldsymbol{X}$ និង សរសេរដូចខាងក្រោម:

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + e \qquad (\text{៧.២៤})$$

ដែល eមានតម្លៃមធ្យមស្មើសូន្យ និង វាមិនមានទំនាក់ទំនងជាមួយ$\boldsymbol{X}$ឡើយ។ Intercept (α_0)ខុសពី δ_0។ នៅពេល(៧.២៤)ឆ្លើយតបទៅនឹងGauss-Markov assumptions យើងទទួលបានការប៉ាន់ស្មានដែលមិនមានកំហុស(unbiased estimators)នៃδ_jដោយOLS។

ជាធម្មតា យើងត្រូវតែជំនួសunobserved u ជាមួយ OLS residuals នោះយើងអាចដំណើរការរ៉ឺរ៉ូស្យុង:

$$\log(\hat{u}^2) \text{ ជាអនុគមន៍នៃ} x_1, x_2, \dots, x_k \qquad (\text{៧.២៥})$$

ជាស្តែង អ្វីដែលយើងត្រូវការពីរ៉ឺរ៉ូស្យុងនេះគឺ តម្លៃដែលសមស្រប(fitted values) និង តាងដោយ $\hat{g}_i$។ បន្ទាប់មក យើងប៉ាន់ស្មានh_iគឺ

$$\hat{h}_i = exp(\hat{g}_i). \qquad (\text{៧.២៦})$$

និតិវិធីFeasible GLS ដើម្បីកែតម្រូវHeteroskedasticity:

១. ដំណើរការរ៉ឺរ៉ូស្យុងyជាអនុគមន៍នៃ$x_1, x_2, \dots, x_k$ និង ទទួលបានresiduals, $\hat{u}$។

២. បង្កើត$\log(\hat{u}^2)$ ដោយដំបូងលើកជាការេនៃ residuals, $\hat{u}$ និង បន្ទាប់មក បំពាក់ជាឡូការីត។

៣. ដំណើរការរ៉ឺរ៉ូស្យុង(៧.២៥) និង ទទួលបានfitted values, $\hat{g}$។

៤. បំលែង fitted values, $\hat{g}$នៃ(៧.២៦)ទៅជាអ៊ិចស្ប៉ូណង់ស្យែល: $\hat{h} = exp(\hat{g})$។

៥. ប៉ាន់ស្មានសមីការ: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ ដោយWLS ជាមួយការប្រើប្រាស់ $1/\hat{h}$

លំហាត់

៧.១ ពិចារណាម៉ូដែលលីនេអ៊ែរដើម្បីពន្យល់លទ្ធភាពនៃប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 \log(SALES) + \beta_2 \log(WC) + \beta_3 TAT + \beta_4 DR + u,$$

$$E(u|log(SALES), \log(WC), TAT, DR) = 0$$

$$Var(u|log(SALES), \log(WC), TAT, DR) = \sigma^2 \mathrm{DR}^2$$

ចូរសរសេរសមីការបំលាស់ប្តូរដែលមានhomoscedastic error term។

៧.២ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ យើងទទួលបានសមីការប៉ាន់ស្មានលទ្ធភាពនៃប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \underset{(3.119)\ [2.919]}{1.006} + \underset{(0.682)\ [0.621]}{1.175 \log(WC)} + \underset{(0.322)\ [0.434]}{3.547\ TAT} + \underset{(0.059)\ [0.022]}{0.044\ DR} + \underset{(1.072)\ [1.117]}{3.473\ location}$$

$$n = 45, R^2 = 0.9019$$

ទាំងstandard errorsធម្មតាជាមួយវង់ក្រចក និង heteroskedasticity-robust standard errorsជាមួយវង់ឃ្នាប។

ក. តើមានភាពខុសគ្នារវាងstandard errorsទាំងពីរប្រភេទនេះដែរឬទេ?

ខ. តើβ_jណាខ្លះដែលមានសារៈសំខាន់ជាស្ថិតិ និង β_jណាខ្លះដែលមិនមានសារៈសំខាន់ជាស្ថិតិនៅឯកំរិត៥%?

គ. ករណីកត្តាផ្សេងទៀតថេរ ប្រសិនTATកើន២ តើROAកើនឡើងប៉ុន្មាន?

ឃ. ចូរបកស្រាយ$\beta_{location}$។

ង. ចូរគណនាតម្លៃព្យាករណ៍នៃក្រុមហ៊ុនដែលមានលេខរៀងទី១២ក្នុងចំណោមទាំង៤៥ក្រុមហ៊ុន និង វែកញែកលទ្ធផល។

៧.៣ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ ដើម្បីប៉ាន់ស្មានសមីការលទ្ធភាពនៃប្រាក់ចំណេញដូចខាងក្រោម:

$$ROA = \beta_0 + \beta_1 \log(WC) + \beta_2 \log(SALES) * \log(WC) + \beta_3 TAT + \beta_4 DR + u,$$

ក. ចូរបង្កើតសមីការប៉ាន់ស្មាន (OLS estimators)

ខ. ចូរបង្កើតផ្លូត(plot)នៃresiduals។ តើមានភស្តុតាងដែលបង្ហាញថា heteroskedasticityកើតឡើងដែរឬទេ?

គ. ចូរធ្វើតេស្តដែលបញ្ជាក់ថា heteroskedasticityបានកើតឡើងដែរឬទេ ដោយប្រើប្រាស់វិធីសាស្ត្រBreusch-Pagan test។ តើមាន heteroskedasticityបានកើតឡើងនៅឯកំរិតសារៈសំខាន់៥%ដែរឬទេ?

ឃ. ផ្អែកលើBreusch-Pagan testក្នុងសំណួរ(គ) បើរកឃើញថា heteroskedasticityបានកើតឡើង។ ចូរអនុវត្តវិធីសាស្ត្រheteroskedasticity-robust standard error។

ង. ផ្អែកលើBreusch-Pagan testក្នុងសំណួរ(គ) បើរកឃើញថា heteroskedasticityបានកើតឡើង។ ចូរបង្កើតម៉ូដែល Weighted Least Squares (WLS) ឬ ចូរប្រើនិតីវិធីFeasible GLS ។

ច. ផ្អែកលើលទ្ធផលនៅសំណួរ(ង) ចូរប្រើប្រាស់វិធីសាស្ត្រBreusch-Pagan testដើម្បីបញ្ជាក់នូវបញ្ហាheteroskedasticity។

មេរៀនទី ៨
កំណត់ទម្រង់នៃម៉ូដែលរ៉ឺគ្រេស្យុង

ម៉ូដែលមួយគួរដាក់បញ្ចូលតែបណ្តាអញ្ញត្តិសំខាន់(core variables)ស្របជាមួយទ្រឹស្តីដែលកំណត់ (ម៉ូដែលត្រូវផ្អែកលើtheoretical framework) និង គួរតែមិនមានទំនាក់ទំនងជាមួយerror term, u។ ម៉្យាងទៀត វាមានបញ្ហាមួយចំនួនដែលធ្វើឲ្យម៉ូដែលមានចំនុចខ្វះខាត គេហៅថាspecification errors និង វាកើតឡើងដោយសារបញ្ហាដូចជា ការលុបចោលអញ្ញត្តិសំខាន់ ការបញ្ចូលអញ្ញត្តិដែលមិនចាំបាច់ទៅក្នុងម៉ូដែល ការបង្កើតទម្រង់ម៉ូដែលមិនត្រឹមត្រូវ (wrong functional form) និង វាស់វែងអញ្ញត្តិមិនបានត្រឹមត្រូវ (errors of measurement)។ ហេតុនេះ បញ្ហា specification errorsគឺជាបញ្ហាសំខាន់បំផុតដែលអ្នកស្រាវជ្រាវប្រហែលជួបប្រទះក្នុងការអនុវត្តន៍។

នៅក្នុងមេរៀននេះ យើងនឹងពិភាក្សាបញ្ហាកាន់តែសំខាន់នៃទំនាក់ទំនងរវាង error, u និង អញ្ញត្តិឯករាជ្យមួយ ឬ ច្រើនជាងមួយ។ ប្រសិន error, u មានទំនាក់ទំនងជាមួយx_jណាមួយ នោះx_jត្រូវបានគេហៅថា endogenous explanatory variable។

ជាទូទៅ ការកើតឡើងនូវទំនាក់ទំនងរវាង error, u ជាមួយx_jណាមួយ គឺបណ្តាលមកពីការលុបចោលអញ្ញត្តិx_jសំខាន់ណាមួយចេញពីម៉ូដែលរ៉ឺគ្រេស្យុង។ ហេតុនេះ វាធ្វើឲ្យមានកំហុសឆ្គង និង មិនមាននិន្នាការកើតឡើងចំពោះOLS estimators។ ម៉្យាងទៀត ការលុបចោលអញ្ញត្តិx_jសំខាន់ណាមួយចេញពីម៉ូដែលរ៉ឺគ្រេស្យុង នោះធ្វើឲ្យម៉ូដែលរងផលប៉ះពាល់ពីកំហុសឆ្គងទម្រង់មុខងារ (functional form misspecification)។

៨.១ កំហុសឆ្គងទម្រង់មុខងារ (Functional Form Misspecification)

ម៉ូដែលរ៉ឺគ្រេស្យុងពហុ (multiple regression model)រងផលប៉ះពាល់ពីfunctional form misspecification នៅពេលវាបកស្រាយមិនបានត្រឹមត្រូវទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ និងអញ្ញត្តិឯករាជ្យ។ ឧបមាថា យើងមានទិន្នន័យក្រុមហ៊ុនទិញលក់ទំនិញចំនួន៥០ និង ចង់សិក្សាកត្តាជះឥទ្ធិពលទៅលើរចនាសម្ព័ន្ធមូលធន(capital structure)។ ផ្អែកលើអត្ថបទស្រាវជ្រាវផ្សេងៗ យើងកំណត់បានម៉ូដែលដូចខាងក្រោម:

$$CS = \beta_0 + \beta_1 AS + \beta_2 \log(SIZE) + \beta_3 PROFIT + \beta_4 \log(GROWTH) + u, \qquad (៨.១)$$

- ✓ រចនាសម្ព័ន្ធមូលធន(CS)គឺ អនុបាតនៃសរុបបំណុល (រយៈពេលខ្លី និង វែង) ធៀបជាមួយ ទ្រព្យសម្បត្តិសរុប។
- ✓ រចនាសម្ព័ន្ធទ្រព្យសម្បត្តិ(asset structure: AS) គឺ អនុបាតនៃទ្រព្យសម្បត្តិរូបីយ៍(tangible assets) ធៀបជាមួយទ្រព្យសម្បត្តិសរុប។
- ✓ ទំហំនៃក្រុមហ៊ុន(SIZE) តាងដោយចំនួនទ្រព្យសម្បត្តិសរុបក្រុមហ៊ុនគិតជាដុល្លារ។

- ✓ កំរិតប្រាក់ចំណេញ(PROFIT)គឺ ជាអនុបាតនៃប្រាក់ចំណេញមុនពេលបង់ពន្ធធៀបនឹង ទ្រព្យសម្បត្តិសរុប។
- ✓ កំណើនក្រុមហ៊ុន(GROWTH)គឺ ជាបម្រែបម្រួលនៃទ្រព្យសម្បត្តិសរុបរវាងដើម និង ចុងឆ្នាំ។
- ✓ ទំហំក្រុមហ៊ុន: S1 (មធ្យម = 1 និង តូច = 0)
- ✓ ទំហំក្រុមហ៊ុន: S2 (ធំ = 1 និង តូច = 0)

ប៉ុន្តែ នៅពេលយើងលុបអញ្ញត្តិlog(GROWTH) វានឹងកើតឡើងនូវfunctional form misspecification នោះធ្វើឲ្យ$\beta_0, \beta_1, \beta_2,$ និង β_3មានកំហុសផងដែរ។ ម៉្យាងទៀត ចំនួននៃកំហុសឆ្គងនេះ វាអាស្រ័យនឹងទំហំនៃβ_4 និង ទំនាក់ទំនងក្នុងចំណោម AS, log(SIZE), PROFIT, និង log(GROWTH)។

តារាងទី១

CS	AS	SIZE	PROFIT	GROWTH	Log(SIZE)	Log(GROWTH)
0.46	0.26	22678	0.32	184	4.355605	2.264817823
0.34	0.15	23468	0.34	128	4.370476	2.10720997
0.42	0.27	11475	0.42	180	4.059753	2.255272505
0.49	0.14	341204	0.34	97	5.533014	1.986771734
0.5	0.28	40367	0.32	60	4.606026	1.77815125
0.54	0.67	235106	0.4	48	5.371264	1.681241237
0.32	0.48	239034	0.45	8620	5.37846	3.935507266
0.4	0.49	461236	0.38	265	5.663923	2.423245874
0.48	0.68	339087	0.46	368	5.530311	2.565847819
0.34	0.42	5632	0.34	370	3.750663	2.568201724
0.38	0.48	14256	0.29	2349	4.153998	3.370883017
0.46	0.68	267833	0.52	987	5.427864	2.994317153
0.36	0.4	434744	0.34	4560	5.638234	3.658964843
0.52	0.78	24535	0.42	897	4.389786	2.952792443
0.58	0.79	3425	0.19	2356	3.534661	3.372175286
0.49	0.72	3463	0.28	1398	3.539452	3.145507171
0.47	0.68	2805	0.42	1298	3.447933	3.113274692
0.51	0.72	13467	0.29	1276	4.129271	3.105850674
0.39	0.47	2346	0.39	3210	3.370328	3.506505032
0.58	0.69	17420	0.32	2420	4.241048	3.383815366
0.45	0.54	2379	0.35	2135	3.376394	3.329397879
0.4	0.59	239872	0.28	2187	5.37998	3.339848783

0.34	0.48	457093	0.41	1200	5.660005	3.079181246
0.52	0.79	78530	0.39	4231	4.895036	3.626443025
0.47	0.8	348988	0.4	2359	5.54281	3.372727941
0.49	0.77	4320	0.39	4320	3.635484	3.635483747
0.4	0.56	238903	0.29	3290	5.378222	3.517195898
0.39	0.68	2642	0.26	2310	3.421933	3.36361198
0.47	0.59	239867	0.19	3270	5.379971	3.514547753
0.54	0.66	328746	0.41	3290	5.51686	3.517195898
0.41	0.56	436740	0.38	2380	5.640223	3.376576957
0.37	0.46	132202	0.29	3260	5.121238	3.5132176
0.48	0.64	322904	0.42	3210	5.509073	3.506505032
0.59	0.74	322906	0.27	4280	5.509076	3.631443769
0.64	0.48	326903	0.19	4320	5.514419	3.635483747
0.47	0.56	323469	0.22	3280	5.509833	3.515873844
0.49	0.58	78650	0.28	3520	4.895699	3.546542663
0.36	0.43	239804	0.32	3210	5.379856	3.506505032
0.59	0.69	564806	0.4	4300	5.751899	3.633468456
0.29	0.34	347808	0.29	52	5.54134	1.716003344
0.58	0.78	453804	0.25	2190	5.656868	3.340444115
0.46	0.72	345092	0.34	2390	5.537935	3.378397901
0.49	0.69	349808	0.26	2369	5.54383	3.374565061
0.42	0.68	129864	0.32	3212	5.113489	3.506775537
0.34	0.45	340977	0.46	2369	5.532725	3.374565061
0.42	0.49	238055	0.38	18122	5.376677	4.258206126
0.45	0.58	347903	0.26	90	5.541458	1.954242509
0.62	0.68	280386	0.27	980	5.447756	2.991226076
0.47	0.58	459803	0.34	894	5.662572	2.951337519
0.34	0.44	140486	0.44	258	5.147633	2.411619706

ការលុបបំបាត់មុខងារនៃបណ្តាអញ្ញត្តិឯករាជ្យ(independent variables) គឺមិនមែនជាវិធីសាស្ត្រតែមួយ ដែលម៉ូដែលមួយរងផលប៉ះពាល់ពីfunctional form misspecification។ ឧទាហរណ៍ ប្រសិន (៨.១)គឺ ជាម៉ូដែលពិតប្រាកដមួយ ដែលឆ្លើយតបGauss-Markov assumptionsពីទី១ដល់ទី៤ ប៉ុន្តែយើងបែរជាប្រើប្រាស់បំណុលសរុបគិតជាដុល្លារ (តំណាងឲ្យរចនាសម្ព័ន្ធមូលធន: CS) ជំនួសឲ្យអនុបាតនៃបំណុលសរុប (រយៈពេលខ្លី និង វែង) ធៀបជាមួយទ្រព្យសម្បត្តិសរុប ជាអញ្ញត្តិមិនឯករាជ្យ

(dependent variable) បន្ទាប់មកយើងនឹងមិនទទួលបានការប៉ាន់ស្មានដែលមិនមានកំហុសឆ្គង ឬ ប្រកបដោយនិរន្តរ៍ភាពនៃpartial effectsឡើយ។

៨.២ RESETជាតេស្តឆ្លៅមួយសំរាប់Functional Form Misspecification

នៅឆ្នាំ១៩៦៩ អ្នកជំនាញeconometrics ឈ្មោះRamseyបានបង្កើតទ្រឹស្តីមួយហៅថា **regression specification error test (RESET)** ដើម្បីកំណត់នូវ**functional form misspecification** សំរាប់អនុវត្តជាមួយម៉ូដែល:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \qquad (៨.២)$$

ដើម្បីស្វែងរក**functional form misspecification** វិធីសាស្ត្រ**RESET test**បន្ថែមfitted values($\hat{y}$)ទៅក្នុងសមីការ(៨.២)។

ដើម្បីអនុវត្តRESET យើងត្រូវតែសម្រេចចិត្តថា តើfitted values($\hat{y}$)មានចំនួនប៉ុន្មានដែលត្រូវបញ្ចូលទៅក្នុងសមីការ(៨.២)។ គ្មានចំលើយដែលត្រឹមត្រូវសំរាប់សំណួរនេះទេ ប៉ុន្តែទម្រង់លើកជាការ និង គីបនៃfitted values($\hat{y}$)មានប្រយោជន៍បំផុតក្នុងការអនុវត្តន៏។ ដូច្នេះ សមីការ(៨.២)ជាមួយទម្រង់លើកជាការ និង គីបនៃfitted values($\hat{y}$)គឺ:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + error, \qquad (៨.៣)$$

សម្មតិកម្មដើម(H_0)គឺ **សមីការ(៨.២)កំណត់បានត្រឹមត្រូវ**។ ហេតុនេះ RESETគឺ *F* statistic សំរាប់ការធ្វើតេស្តសម្មតិកម្មដើម(H_0): $\delta_1 = 0$ និង $\delta_2 = 0$ នៅក្នុងម៉ូដែល(៨.៣)។ សារៈសំខាន់ជាស្ថិតិនៃ*F* statisticបញ្ជាក់ថា វាមានបញ្ហាfunctional form misspecificationកើតឡើង។ Degree of freedom (df)ក្នុងម៉ូដែល(៨.៣)គឺ $n - k - 1 - 2 = n - k - 3$។ ផ្អែកលើសម្មតិកម្មដើម (និង Gauss-Markov assumptions) បំណែងចែកនៃ*F* statisticគឺប្រហែល$F_{2,n-k-3}$។ ចំណែក *LM* statisticក៏អាចអនុវត្ត (និង បំណែងចែកនៃchi-squareមានdfស្មើ២)ផងដែរ។

យោងលើទិន្នន័យនៅក្នុង**តារាងទី១** យើងទទួលបានលទ្ធផលនៃសមីការរចនាសម្ព័ន្ធមូលធន (capital structure)ដូចខាងក្រោម:

Dependent Variable: CS			
Independent Variables	**Model (1)**	**Model (2)**	**Model (3)**
intercept.....	0.372	0.399	0.427
	(0.083)	(0.058)	(0.091)
AS.....	0.268	0.270	0.312
	(0.059)	(0.060)	(0.066)

SIZE.....		2.79E-08	
		(5.99E-08)	
log(SIZE).....	0.006		0.007
	(0.013)		(0.013)
PROFIT.....	-0.286	-0.285	-0.303
	(0.130)	(0.131)	(0.129)
GROWTH.....		-2.19E-06	
		(3.58E-06)	
log(GROWTH).....			-0.026
			(0.018)
Observations.....	50	50	50
R-squared.....	0.3628	0.3674	0.3897
RESET test:			
F statistic.....	4.310 > 3.23	3.631 > 3.23	*1.504 < 3.23*
LM statistic.....	8.946 > 5.99	7.803 > 5.99	*3.380 < 5.99*
p-value.....	0.0195 < 0.05	0.0349 < 0.05	*0.2337 > 0.05*

ដូច្នេះ លទ្ធផលក្នុងតារាងខាងលើបញ្ជាក់ថា ម៉ូដែលទី១ និង ទី២មានបញ្ហា **functional form misspecification**កើតឡើង ក្នុងនោះម៉ូដែលទី១មិនបានបញ្ចូលអញ្ញត្តិ log(GROWTH) និង ម៉ូដែលទី២មិនបានបំពាក់ឡូការីតទៅលើអញ្ញត្តិ SIZE និង GROWTHឡើយ។ រីឯ ម៉ូដែលទី៣មិនមានបញ្ហា**functional form misspecification**កើតឡើងទេ ក្រោយពេលបំលាស់ប្តូរអញ្ញត្តិ SIZE និង GROWTHទៅជាទំរង់ឡូការីត។

៨.៣ កំហុសឆ្គងក្នុងការវាស់វែង(Errors of Measurement)

ការវាស់វែងអញ្ញត្តិមិនឯករាជ្យ(dependent variable) ឬ ឯករាជ្យ(independent variables) មិនបានត្រឹមត្រូវ វាក៏ជាចំនុចខ្វះខាតនៃម៉ូដែលរ៉ឺហ្គ្រេស្យុង។

៨.៣.១ កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិមិនឯករាជ្យ

បើកំហុសឆ្គងកើតឡើងតែទៅលើការវាស់វែងអញ្ញត្តិមិនឯករាជ្យ គេអាចបញ្ជាក់ថា:

- ✓ OLS estimatorsមិនមានកំហុសឆ្គងឡើយ។
- ✓ វ៉ារីយ៉ង់នៃOLS estimatorsក៏មិនមានកំហុសឆ្គងដែរ។

- ✓ ប៉ុន្តែ វ៉ារីយ៉ង់ដែលបានប៉ាន់ស្មាននៃestimators (ករណីកំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិមិនឯករាជ្យ) គឺ ធំជាងវ៉ារីយ៉ង់នៃករណីដែលការវាស់វែងអញ្ញត្តិមិនឯករាជ្យមិនមានកំហុសឆ្គង ពីព្រោះ កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិមិនឯករាជ្យបូកបន្ថែមerror, u។

ដូច្នេះ ការកើតឡើងកំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិមិនឯករាជ្យប្រហែលជាមិនមានបញ្ហាច្រើនក្នុងការអនុវត្តន៍ឡើយ។

៨.៣.២ កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិឯករាជ្យ

ករណីនេះ បណ្តាលឲ្យកើតឡើង:

- ✓ OLS estimatorsមានកំហុសឆ្គង គឺថា វានៅតែមានកំហុសឆ្គងកើតឡើង ទោះបីជា សំណាក (sample size) មានចំនួនច្រើនក៏ដោយ។
- ✓ OLS estimatorsក៏មិនមាននិរន្តរ៍ភាព(inconsistent)ផងដែរ។

នេះបញ្ជាក់ថា កំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិឯករាជ្យជាបញ្ហាធ្ងន់ធ្ងរ និង ត្រូវរាយការណ៍វា ដើម្បីធ្វើការកែតំរូវ។ ប៉ុន្តែ វាមិនងាយស្រួលក្នុងការស្វែងរកឡើយ។ ឧទាហរណ៍ ទិន្នន័យតម្លៃទំនិញលើទីផ្សារងាយស្រួលក្នុងការស្វែងរក ប៉ុន្តែ ទិន្នន័យចំណូលទាក់ទងនឹងសកម្មភាពល្មើសច្បាប់ពិបាកក្នុងការស្វែងរកឲ្យបានច្បាស់លាស់។

ប្រសិនបើមានការកើតឡើងនូវកំហុសឆ្គងក្នុងការវាស់វែងអញ្ញត្តិឯករាជ្យ ការកែតម្រូវ ឬ វិធីសាស្ត្រសំរាប់អនុវត្តន៍គឺ **instrumental** ឬ **proxy variables**។ អញ្ញត្តិទាំងនេះមានទំនាក់ទំនងខ្លាំងជាមួយអញ្ញត្តិឯករាជ្យ(x) ប៉ុន្តែមិនមានទំនាក់ទំនងជាមួយerror term, uឡើយ។

៨.៤ ការស្វែងរកកំហុសឆ្គង(Specification Errors)

នៅក្នុងការអនុវត្តន៍ បើមានកំហុសឆ្គងកើតឡើង យើងត្រូវតែដឹងថា តើត្រូវស្វែងរកវាយ៉ាងដូចម្តេច។ បន្ទាប់មកត្រូវតែកែតម្រូវវា។ ឧទាហរណ៍ បើរកឃើញថា អញ្ញត្តិសំខាន់ណាមួយត្រូវបានដកចេញពីម៉ូដែល ដូច្នេះត្រូវដាក់បញ្ចូលវាវិញជាមួយទិន្នន័យគ្រប់គ្រាន់។

សម្មតថា ម៉ូដែលមួយមានប្រាំអញ្ញត្តិដូចខាងក្រោម:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u, \qquad (៨.៤)$$

បើទ្រឹស្តីបញ្ជាក់ថា រាល់អញ្ញត្តិឯករាជ្យទាំងបួនកំណត់ ឬ ជះឥទ្ធិពលទៅលើអញ្ញត្តិ y យើងគួរតែរក្សាអញ្ញត្តិឯករាជ្យទាំងនេះក្នុងម៉ូដែល ទោះបីជាក្រោយពេលធ្វើតេស្ត(empirical testing) រកឃើញថាមេគុណនៃអញ្ញត្តិ x មួយ ឬ ពីរក្នុងចំណោមទាំងបួនមិនមានសារៈសំខាន់ជាស្ថិតិក៏ដោយ។ ដូច្នេះ មិនមានអញ្ញត្តិដែលមិនសំខាន់ត្រូវបានដាក់បញ្ចូលក្នុងម៉ូដែល(៨.៤)ឡើយ។

ទោះយ៉ាងណាក៏ដោយ ដើម្បីការពារកំហុសឆ្គងទាក់ទងនឹងការលុបបំបាត់អញ្ញត្តិឯករាជ្យសំខាន់ណាមួយចោល យើងអាចដាក់ចញ្ចូលអញ្ញត្តិ(control variables)ទៅក្នុងម៉ូដែល។ ប្រសិន control variablesមិនមានសារៈសំខាន់ជាស្ថិតិ និង ការដកចេញពីម៉ូដែលមិនធ្វើឲ្យមានការប្រែប្រួលធំដុំទៅដល់លទ្ធផលប៉ាន់ស្មាន នោះគេអាចដកវាចេញពីម៉ូដែលដោយមានការពន្យល់ឲ្យបានច្បាស់លាស់។ ឧបមាថា ក្នុងម៉ូដែល(៨.៤) **អញ្ញត្តិ**x_3 និង x_4 ជាcontrol variables និង ក្រោយការអនុវត្តន៍ empirical testing **រកឃើញថា** ពួកវាមិនមានសារៈសំខាន់ជាស្ថិតិ និង ចង់ដកវាចេញពីម៉ូដែល។ ដូច្នេះ យើងបង្កើតសម្មតិកម្មដើម$H_0: \beta_3 = 0, \beta_4 = 0$។ ក្រោយការអនុវត្តន៍*F* statistic បើរកឃើញថា H_0ពិត គេអាចលុប**អញ្ញត្តិ**x_3 និង x_4ចេញពីម៉ូដែល។

ហេតុនេះ ការស្វែងរកវត្តមានអញ្ញត្តិដែលមិនចាំបាច់ វាមិនមែនជាការលំបាកឡើយ។ ប៉ុន្តែការសំខាន់បំផុតគឺ យើងទទួលយកតែម៉ូដែលណាជា *ម៉ូដែលពិតប្រាកដ* (true model)។ ការប្រើប្រាស់ t និង *F* tests យើងអាចស្វែងរក ថាតើអញ្ញត្តិ xណាខ្លះដែលមានទំនាក់ទំនងពិតប្រាកដជាមួយម៉ូដែលនោះ។ ប៉ុន្តែ យើងមិនគួរប្រើប្រាស់ t និង *F* tests ដើម្បីបង្កើតម៉ូដែលម្តងហើយម្តងទៀតឡើយ នោះគឺយើងមិនអាចនិយាយថា ដំបូង yមានទំនាក់ទំនងជាមួយx_3 ពីព្រោះ β_3មានសារៈសំខាន់ជាស្ថិតិ និងបន្ទាប់មកបន្ថែមx_4ចូលទៅក្នុងម៉ូដែលដោយរក្សាx_4ក្នុងម៉ូដែល ប្រសិនវាមានសារៈសំខាន់ជាស្ថិតិ។ វិធីសាស្ត្រនេះគឺជា**stepwise regression**។

ឧទាហរណ៍ ចំណូល និង ការថែរក្សាសុខភាព គឺជាអញ្ញត្តិ ឬ កត្តាសំខាន់ដែលកំណត់គុណភាពជីវិត ទោះបីជាយើងមិនច្បាស់ក្នុងការបញ្ចូលវាទៅក្នុងម៉ូដែលក៏ដោយ។ ការមិនដាក់បញ្ចូលអញ្ញត្តិទាំងនេះទៅក្នុងម៉ូដែល វាអាចធ្វើឲ្យកើតឡើងនូវកំហុសឆ្គង។ ដូច្នេះ ដើម្បីបន្ថែមវា(data mining)ទៅក្នុងម៉ូដែល យើងចាំបាច់ត្រូវកំណត់ទម្រង់(functional form)ឲ្យបានសមស្របនៃទំនាក់ទំនងរវាងអញ្ញត្តិមិនឯករាជ្យ (គុណភាពជីវិត) និង អញ្ញត្តិឯករាជ្យ (ចំណូល និង ការថែរក្សាសុខភាព)។ *ការបង្កើតម៉ូដែលត្រូវតែផ្អែកលើទ្រឹស្តី* និង *ការវាស់វែងដោយគ្មានទ្រឹស្តីប្រៀបដូចជាការធ្វើដំណើរវង្វេងផ្លូវ*។

លំហាត់

៨.១ ឧបមាថា ក្រុមប្រឹក្សាភិបាលបានពិចារណាទៅលើមូលដ្ឋាននៃការកែសម្រួលប្រាក់បៀវត្តន៍របស់នាយកប្រតិបត្តិក្រុមហ៊ុនមួយចំនួន។ ពិចារណាទៅលើពត៌មានខាងក្រោមដែលទាក់ទងនឹងប្រាក់បៀវត្តន៍របស់នាយកប្រតិបត្តិ ចំណូលរបស់ក្រុមហ៊ុន និង ផលចំណេញដែលទទួលបានពីការវិនិយោគរបស់ម្ចាស់ភាគហ៊ុន។

Salary ($000)	Sales ($000)	ROE	log(salary)	log(sales)
48	1985	16.34	1.681241	3.297761
110	3365	29.59	2.041393	3.526985
105	3535	28.12	2.021189	3.548389
165	3445	37.56	2.217484	3.537189
139	3205	32.18	2.143015	3.505828
103	2830	20.3	2.012837	3.451786
115	3245	21.5	2.060698	3.511215
155	3645	34.12	2.190332	3.561698
142	4054	32.28	2.152288	3.607884
150	3940	34.24	2.176091	3.595496
71	3190	27.2	1.851258	3.503791
76	2144	16.89	1.880814	3.331225
65	2110	17.89	1.812913	3.324282
100	2420	23.7	2	3.383815
84	2490	27.2	1.924279	3.396199
58	1755	39.1	1.763428	3.244277
88	2605	28	1.944483	3.415808
92	2865	24	1.963788	3.457125
139	3570	30.12	2.143015	3.552668
95	3155	20.5	1.977724	3.498999
90	2678	28	1.954243	3.427811
63	2215	34.7	1.799341	3.345374
66	1800	36.1	1.819544	3.255273
80	1915	35.7	1.90309	3.282169

85	2965	20.2	1.929419	3.472025
90	3420	23.9	1.954243	3.534026
65	2380	29.9	1.812913	3.376577
67	3250	30.4	1.826075	3.511883
74	1980	36	1.869232	3.296665
110	2800	22.6	2.041393	3.447158
67	2950	16.45	1.826075	3.469822
95	2560	19.58	1.977724	3.40824
75	2210	20.12	1.875061	3.344392
67	1850	14.26	1.826075	3.267172
100	2615	32.9	2	3.417472
67	1965	38	1.826075	3.293363
120	2930	29.66	2.079181	3.466868
60	1968	18.34	1.778151	3.294025
70	2070	18.68	1.845098	3.31597
116	2900	25.4	2.064458	3.462398
75	2542	14.23	1.875061	3.405176
68	1985	12.98	1.832509	3.297761
88	2395	19	1.944483	3.379306
82	2720	18.64	1.913814	3.434569
80	2670	27.4	1.90309	3.426511
88	2890	22.3	1.944483	3.460898
79	2625	19.68	1.897627	3.419129
85	3465	17.6	1.929419	3.539703
65	3465	14.89	1.812913	3.539703
105	3380	28.68	2.021189	3.528917

ក. យើងអាចបង្កើតបានម៉ូដែលដែលទាក់ទងនឹងប្រាក់បៀវត្តន៍របស់នាយកប្រតិបត្តិដូចខាងក្រោម:

$$SALARY = \beta_0 + \beta_1 SALES + \beta_2 ROE + u,$$

១. ផ្អែកលើទិន្នន័យ ចូរបង្កើតម៉ូដែលប៉ាន់ស្មានទៅលើប្រាក់បៀវត្តន៍។

២. ចូរធ្វើតេស្តទៅលើβ_1 និង β_2 នៅឯកំរិតសារៈសំខាន់៥%។

៣. បើចំណូលរបស់ក្រុមហ៊ុនកើនឡើង400,000ដុល្លារ តើគេអាចប៉ាន់ស្មានថា ប្រាក់បៀវត្តន៍កើនឡើង ឬ ថយចុះប៉ុន្មាន?

៤. ដោយប្រើប្រាស់RESET test តើមានកស្តុតាងបញ្ជាក់ថា functional form misspecificationនៅក្នុងម៉ូដែលនេះដែរឬទេ?

ខ. យើងអាចបង្កើតបានម៉ូដែលដែលទាក់ទងនឹងប្រាក់បៀវត្តន៍របស់នាយកប្រតិបត្តិដូចខាងក្រោម:

$$SALARY = \beta_0 + \beta_1 \log(SALES) + \beta_2 ROE + u,$$

ដោយប្រើប្រាស់RESET test តើមានកស្តុតាងបញ្ជាក់ថា functional form misspecificationនៅក្នុងម៉ូដែលនេះដែរឬទេ?

គ. យើងអាចបង្កើតបានម៉ូដែលដែលទាក់ទងនឹងប្រាក់បៀវត្តន៍របស់នាយកប្រតិបត្តិដូចខាងក្រោម:

$$\log(SALARY) = \beta_0 + \beta_1 \log(SALES) + \beta_2 ROE + u,$$

ដោយប្រើប្រាស់RESET test តើមានកស្តុតាងបញ្ជាក់ថា functional form misspecificationនៅក្នុងម៉ូដែលនេះដែរឬទេ?

៨.២ នៅពេលប្រើប្រាស់ចំនួនការអង្កេត n = 129 ដើម្បីប៉ាន់ស្មានម៉ូដែល

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

នោះយើងទទួលបាន$R^2 = 0.4289$។ នៅពេលបន្ថែមអញ្ញត្តិx_3 និង x_4 នោះ$R^2 = 0.4326$។ តើមានកស្តុតាងបញ្ជាក់ថា functional form misspecificationនៅក្នុងម៉ូដែលនេះដែរឬទេ?

៨.៣ ប្រើប្រាស់ទិន្នន័យក្នុងលំហាត់៤.៣នៃមេរៀនទី៤ ចូរឆ្លើយសំណួរដូចខាងក្រោម:

ក. ប៉ាន់ស្មានម៉ូដែល

$$ROA = \beta_0 + \beta_1 \log(SALES) + \beta_2 \log(WC) + \beta_3 TAT + u,$$

ប្រសិន$\log(SALES)$ $\log(WC)$ និង TATប្រែប្រួល តើ*ROA*មានការប្រែប្រួលយ៉ាងដូចម្តេច?

ខ. តើម៉ូដែលក្នុងសំណួរ(ក)មានលក្ខណៈគ្រប់គ្រាន់ដែរឬទេ បើផ្អែកលើRESET?

គ. បន្ថែមអញ្ញត្តិQR និង DR ចូលទៅក្នុងម៉ូដែលក្នុងសំណួរ(ក) ចូរប៉ាន់ស្មានម៉ូដែលឡើងវិញ។ តើ$\log(SALES)$ $\log(WC)$ និង TAT មានការជះឥទ្ធិពលយ៉ាងដូចម្តេចROA?

ឃ. តើម៉ូដែលក្នុងសំណួរ(គ)មានលក្ខណៈគ្រប់គ្រាន់ដែរឬទេ បើផ្អែកលើRESET?

ង. បន្ថែមអញ្ញត្តិlocationចូលទៅក្នុងម៉ូដែលក្នុងសំណួរ(គ) ចូរប៉ាន់ស្មានម៉ូដែលឡើងវិញ។ តើ$\log(SALES)$ $\log(WC)$ និង TAT មានការដេះដូរទ្វិពលយ៉ាងដូចម្តេចROA?

ច. តើម៉ូដែលក្នុងសំណួរ(ង)មានលក្ខណៈគ្រប់គ្រាន់ដែរឬទេ បើផ្អែកលើRESET?

ឆ. ចំពោះម៉ូដែលរ៉ឺក្រេស្យុងដែលបានប៉ាន់ស្មានក្នុងសំណួរ(គ)និង(ង) ចូរកំណត់ទៅលើបម្រែបម្រួលនៃROA ក្នុងករណីដូចខាងក្រោម:

   ១. $\Delta \log(WC) = +10\%$

   ២. $\Delta TAT = +1.24$

   ៣. តើការលុបអញ្ញត្តិlocation វាធ្វើឲ្យមានកំហុសឆ្គងកើតឡើងដែរឬទេ? ហេតុអ្វី?

ជ. យោងលើការបកស្រាយខាងលើ ក្នុងចំណោមម៉ូដែលនៅសំណួរ(ក) (គ) និង (ង) តើម៉ូដែលណាមួយ វាសមស្របជាងគេ?

ឯកសារយោង

Biddle, J. E., and D. S. Hamermesh (1990), *Sleep and the Allocation of Time*, Journal of Political Economy 98, 922–943.

Daskalakis, N. and Psillakib, M. (2008), *Do country or firm factors explain capital structure ? Evidence from SMEs in France and Greece*, Applied Financial Economics, 2008, 18, 87–97

Hill, R. Carter, Griffiths, William E., and Lim, G. C (2018), *Principles of Econometrics*. 5th ed. John Wiley & Sons, Inc.

Malik, H. (2011), *Determinants of insurance company profitability: an analysis of insurance sector of Pakistan*, Academic Research International, Vol. 1 No. 3, pp. 315-321.

Mohd Zaid, N.A., Wan Ibrahim, W.M.F. and Zulqernain, N.S. (2014), *The determinants of profitability: evidence from Malaysian construction companies*, paper presented at the Proceedings of 5th Asia-Pacific Business Research Conference, Kuala Lumpur, February 17-18, pp. 1-13.

Panno, A. (2003), *An empirical investigation on the determinants of capital structure: the UK and Italian experience*, Applied Financial Economics, 13, 97–112.

Samprit Chatterjee and Ali S. Hadi (2012), *Regression Analysis by Example*. 5th ed. John Wiley & Sons, Inc.

Sekaran, U. and Bougie, R. (2016), *Research Methods for Business: A Skill-Building Approach*. 7th ed. John Wiley & Sons Ltd.

Wald, J.K. (1999), *How firm characteristics affect capital structure: an international comparison*, Journal of Financial Research, Vol. 22 No. 2, pp. 161-87.

Warrad, L. and Al Omari, R. (2015), *The impact of turnover ratios on Jordanian services sectors´ performance*, Journal of Modern Accounting and Auditing, Vol. 11 No. 2, pp. 77-85.

Wooldridge, J. M. (2016), Econometric Analysis of Cross Section and Panel Data. 6th ed. Cengage Learning.

តារាងស្ថិតិ

តារាង A: The t-Distribution

df \ t	0.01	0.02	0.05	0.1	0.2	0.3	0.5	0.7	0.9	α	Two tails test
	0.005	0.01	0.025	0.05	0.1	0.15	0.25	0.35	0.45	$\alpha/2$	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		

តារាង B: The Chi-Square Distribution

α / df	**0.1**	**0.05**	**0.01**
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
11	17.275	19.675	24.725
12	18.549	21.026	26.217
13	19.812	22.362	27.688
14	21.064	23.685	29.141
15	22.307	24.996	30.578
16	23.542	26.296	32.000
17	24.769	27.587	33.409
18	25.989	28.869	34.805
19	27.204	30.144	36.191
20	28.412	31.410	37.566
21	29.615	32.671	38.932
22	30.813	33.924	40.289
23	32.007	35.172	41.638
24	33.196	36.415	42.980
25	34.382	37.652	44.314
26	35.563	38.885	45.642
27	36.741	40.113	46.963
28	37.916	41.337	48.278
29	39.087	42.557	49.588
30	40.256	43.773	50.892
40	51.805	55.758	63.691
50	63.167	67.505	76.154
60	74.397	79.082	88.379
70	85.527	90.531	100.425
80	96.578	101.879	112.329
90	107.565	113.145	124.116
100	118.498	124.342	135.807

តារាង.១C 10% Critical Values of the *F* Distribution

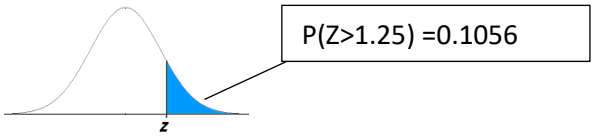
Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

តារាង.២C 5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

តារាង.៣C 1% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32



តារាង D: The Normal Distribution

Z	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

គាំទ្រថវិកាលើការរៀបរៀង និពន្ធ និង កែលម្អដោយ

ក្រសួងអប់រំ យុវជន និងកីឡា

មូលនិធិស្រាវជ្រាវ គំនិតច្នៃប្រឌិត និង នវានុវត្តន៍