ការយល់ដឹងអំពីប៊ូតស្ត្រេបស្ថិតិ

Understanding Statistical Bootstrapping

ម៉ុង ម៉ារ៉ា

មន្ត្រីនាយកដ្ឋានគណិតវិទ្យានិងស្ថិតិ
វិទ្យាស្ថានវិទ្យាសាស្ត្រនិងបច្ចេកវិទ្យា
រាជបណ្ឌិត្យសភាកម្ពុជា

២០២៤

មូលន័យសង្ខេប

ប៊ូតស្ត្រែបស្ថិតិគឺជាបច្ចេកទេសជ្រើសរើសគំរូតាងសាឡើងវិញដ៏មានភាពបត់បែននិងមានអានុភាពដែលត្រូវបានប្រើប្រាស់យ៉ាងទូលំទូលាយនៅក្នុងការវិភាគទិន្នន័យទំនើប។ អត្ថបទនេះផ្តល់នូវការយល់ដឹងយ៉ាងងាយនៃរបៀបដែលប៊ូតស្ត្រែបមានដំណើរការដោយផ្តោតលើការអនុវត្តរបស់វាសម្រាប់ការប៉ាន់ស្មានមធ្យមតាមរយៈចន្លោះជឿជាក់ ។ តាមរយៈការបង្កើតគំរូតាងសាឡើងវិញជាច្រើនពីសំណុំទិន្នន័យដើម ប៊ូតស្ត្រែបអនុញ្ញាតឱ្យអ្នកប្រើប្រាស់ទាញសន្និដ្ឋានដ៏រឹងមាំដោយពុំចាំបាច់មានលក្ខខណ្ឌសន្មតលើប៉ារ៉ាម៉ែត្រ។ ការពិភាក្សានៅក្នុងអត្ថបទរួមមានវិធីសាស្រ្តបឋមចំនួនបី៖ ចន្លោះជឿជាក់ប៊ូតស្ត្រែបមូលដ្ឋាន ប៊ូតស្ត្រែបពែសង់ទីល និង ប៊ូតស្ត្រែបស្ទូដិន។ វិធីនីមួយៗត្រូវបានបង្ហាញជាមួយនិងឧទាហរណ៍មួយជំហានម្តងៗ ដោយគូសបញ្ជាក់ពីទ្រឹស្តី ជំហានគណនា និងប្រសិទ្ធភាព។ អត្ថបទនេះមានគោលបំណងបំភ្លឺឱ្យយល់ច្បាស់អំពីវិធីប៊ូតស្ត្រែបនិងបង្ហាញពីការប្រើប្រាស់ជាក់ស្តែងរបស់វា ដោយផ្តល់ដល់អ្នកអាននូវសមត្ថភាពអនុវត្តវិធីទាំងនេះដោយទំនុកចិត្តនៅក្នុងសេណារីយ៉ូការវិភាគទិន្នន័យក្នុងពិភពពិត។

ការយល់ដឹងអំពី ប៊ូតស្ត្រេបស្ថិតិ

១. សេចក្តីផ្តើម

នៅក្នុងការវិភាគទិន្នន័យ ស្ថិតិវិទូ និង អ្នកសិក្សាស្រាវជ្រាវប្រឈមជាញឹកញាប់ជាមួយបញ្ហានៃការទាញសន្និដ្ឋានអំពីសាកលស្ថិតិដោយផ្អែកលើទិន្នន័យគំរូតាងដែលមានភាពមិនគ្រប់លក្ខណៈ។ វិធីស្ថិតិសន្និដ្ឋានធម្មតាដូចជាចន្លោះជឿជាក់ និង តេស្តសម្មតិកម្ម តែងតែទាមទារលក្ខខណ្ឌចំពោះទិន្នន័យដូចជាលក្ខខណ្ឌភាពណរម៉ាល់ ឬ គំរូតាងទំហំធំ។ ទោះជាយ៉ាងណាក៏ដោយ លក្ខខណ្ឌទាំងនេះមិនទៀងតែផ្ទៀងផ្ទាត់ជានិច្ចកាលនោះទេ ជាពិសេសនៅក្នុងសេណារីយ៉ូនៃការអនុវត្តជាក់ស្តែងដែលទិន្នន័យច្រើនតែមានភាពរញ៉េរញ៉ៃ មិនពេញលេញ ឬមានភាពទេរ។ នៅក្នុងករណីទាំងនេះ ការប្រើវិធីស្ថិតិធម្មតាអាចនាំទៅដល់លទ្ធផលដែលមិនអាចជឿជាក់និងមានលម្អៀងច្រើន។ ដោយហេតុផលទាំងនេះហើយដែល ប៊ូតស្ត្រេបស្ថិតិក្លាយជាវិធីដែលមានសារៈប្រយោជន៍យ៉ាងពិសេសជាមួយនិងភាពបត់បែននិងភាពខ្លាំងរបស់វា។

គោលបំណងរបស់អត្ថបទនេះគឺផ្តល់ការយល់ដឹងដ៏ទូលំទូលាយមួយអំពីវិធីប៊ូតស្ត្រេប។ នៅចុងបញ្ចប់នៃការពិភាក្សានៅក្នុងអត្ថបទ អ្នកអាននឹងទទួលបានការយល់ដឹងកាន់តែស៊ីជម្រៅនូវរបៀបដែល ប៊ូតស្ត្រេបលើកកម្ពស់ការជឿជាក់លើការប៉ាន់ស្មាននិងការសន្និដ្ឋានស្ថិតិដែលធ្វើឱ្យវាកាន់តែមានសារៈប្រយោជន៍នៅក្នុងការវិភាគទិន្នន័យ។

២. តើប៊ូតស្ត្រេបមានដំណើរការយ៉ាងដូចម្តេច?

ប៊ូតស្ត្រេបគឺជាវិធីដ៏មានប្រសិទ្ធភាពទាក់ទងនឹងការជ្រើសរើសគំរូតាងសាឡើងវិញ។ ការជ្រើសរើសគំរូតាងសាឡើងវិញនេះត្រូវបានធ្វើម្តងហើយម្តងទៀត ដដែលៗចេញពីទិន្នន័យដើមដែលមានស្រាប់។ គំនិតជាគន្លឹះនៅក្នុងប៊ូតស្ត្រេបគឺការជ្រើសរើសគំរូតាងសាឡើងវិញដែលគំរូតាងនេះជាប្រភេទធ្វើឡើងដោយដាក់ចូលទៅវិញ មានន័យថាសង្កេតនីមួយៗនៅក្នុងទិន្នន័យដើមមានប្រូបាប៊ីលីតេស្មើៗគ្នានៅក្នុងការបញ្ចូល(មួយដង ឬ លើសពីមួយដង)ឬមិនបញ្ចូលសោះទៅក្នុងគំរូតាងប៊ូតស្ត្រេប។ ដំណើរការនេះមានលក្ខណៈជាការធ្វើត្រាប់តាមរបៀបដែលគេជ្រើសរើស(ដោយការដាក់ចូលទៅវិញ)នូវគំរូតាងជាច្រើនដែលមានទំហំដូចៗគ្នាចេញពីសាកលស្ថិតិតែមួយ ដោយមិនទាមទារការប្រមូលទិន្នន័យអ្វីផ្សេងទៀតបន្ថែម។ វិធីនេះជួយប៉ាន់ស្មានរបាយតម្លៃស្ថិតិ(sampling distribution of a statistic) ដោយមិនទាមទារលក្ខខណ្ឌអ្វីច្រើនចំពោះសាកលស្ថិតិនៅក្នុងការសិក្សា(Efron, 1979)។

វិធីប៊ូតស្ត្រេបមានដំណើរការជាជំហានៗដូចខាងក្រោម៖

១) ចាប់ផ្តើមដំណើរការជាមួយគំរូតាងដើមមានទំហំnដែលជាបណ្តុំនៃសង្កេតដែលបានពីការប្រមូលក្នុងពិសោធន៍ឬការស្ទង់មតិ។ ឧទាហរណ៍ បើសិនជាយើងមានទិន្នន័យ៥០តម្លៃនោះ វាគឺជាទិន្នន័យមូលដ្ឋានសម្រាប់គំរូតាងសាឡើងវិញឬគំរូតាងប៊ូតស្ត្រេប(Davison & Hinkley, 1997a; Efron, 1979)។

២) បង្កើតគំរូតាងថ្មីទំហំnតាមរយៈការជ្រើសរើសដោយចៃដន្យនិងដាក់ចូលទៅវិញនូវសង្កេតចេញពីទិន្នន័យគំរូតាងដើម។ វាមានន័យថាសង្កេតនីមួយៗអាចលេចឡើងច្រើនដងនៅក្នុងគំរូតាងថ្មីឬអាចមិនលេចឡើងសោះ។(Davison & Hinkley, 1997a)

៣) គណនាតម្លៃស្ថិតិដែលត្រូវការ(អាចជា មធ្យម មេដ្យាន ឬគម្លាតស្តង់ដា)របស់គំរូតាងសាឡើងវិញ(គំរូតាងប៊ូតស្រ្ទេប)។

៤) បន្តធ្វើដដែលៗនូវការជ្រើសរើសគំរូតាងសាឡើងវិញព្រមទាំងការគណនាតម្លៃស្ថិតិឱ្យបានចំនួនច្រើនដង (ឧទាហរណ៍ ចំនួន១០០០ដងឬច្រើនជាងនេះ) (DiCiccio & Efron, 1996)។

៥) បន្ទាប់ពីការធ្វើដដែលៗរួចមក យើងសម្រេចបាននូវបាយតម្លៃស្ថិតិ។ របាយនេះហៅថារបាយប៊ូតស្រ្ទេប (bootstrap distribution) ហើយវាដើរតួជារបាយតម្លៃស្ថិតិដែលបានមកពីការជ្រើសរើសគំរូតាងម្តងហើយម្តងទៀតចំនួនច្រើនដងចេញពីសាកលស្ថិតិមួយ ។ គេប្រើប្រាស់វាដើម្បីប៉ាន់ស្មានរបាយគំរូតាងនៃតម្លៃស្ថិតិ(sampling distribution of the statistic)ដោយមិនពឹងផ្អែកលើលក្ខខណ្ឌភាពណរម៉ាល់ទេ(DiCiccio & Efron, 1996; Efron, 1979)។

៦) របាយប៊ូតស្រ្ទេបអាចត្រូវបានប្រើដើម្បីប៉ាន់ស្មានរង្វាស់មួយចំនួនដូចជាល្មៀងស្តង់ដាដើម្បីបង្កើតចន្លោះជឿជាក់ ឬធ្វើតេស្តសម្មតិកម្មជាដើម អាស្រ័យទៅតាមគោលដៅនៃការវិភាគទិន្នន័យ(DiCiccio & Efron, 1996)។

ឧទាហរណ៍ថាយើងមានសំណុំទិន្នន័យតូចមួយអំពីកម្ពស់គិតជាសង់ទីម៉ែត្ររបស់បុគ្គលចំនួន៥នាក់។ គំរូតាងដើមកំណត់ដោយ{155, 163, 170, 158, 162}។ យើងជ្រើសរើសគំរូតាងដែលមានទំហំ n =5 ចេញពីគំរូតាងនេះនិងបានលទ្ធផលដូចខាងក្រោម៖

- គំរូតាងសាឡើងវិញទី១ {55, 55, 70, 63, 62} មានមធ្យមស្មើ 61
- គំរូតាងសាឡើងវិញទី២ {58, 63, 63, 70, 55} មានមធ្យមស្មើ 61.8
- គំរូតាងសាឡើងវិញទី៣ {63, 62, 70, 70, 58} មានមធ្យមស្មើ 64.6

ដំណើរការនេះត្រូវបន្តធ្វើដដែលៗច្រើនដង អាចដល់១០០០ដងឬលើសពីនេះដើម្បីបង្កើតរបាយនៃមធ្យមគំរូតាង។ ដូច្នេះនៅចុងបញ្ចប់នៃដំណើរការនេះយើងទទួលបានរបាយប៊ូតស្រ្ទេបនៃមធ្យមគំរូតាងដែលយើងអាចប្រើប្រាស់វាដើម្បីប៉ាន់ស្មានបម្រែបម្រួល(ល្មៀងស្តង់ដា)នៃមធ្យមគំរូតាង គណនាចន្លោះជឿជាក់ឬអនុវត្តតេស្តសម្មតិកម្ម។

៣. ឧទាហរណ៍៖ ការប៉ាន់ស្មានមធ្យម

ឧទាហរណ៍នេះប្រើប្រាស់ទិន្នន័យគំរូតាងខាងលើដែលនិយាយអំពីកម្ពស់គិតជាសង់ទីម៉ែត្ររបស់បុគ្គលចំនួន៥នាក់: {155, 163, 170, 158, 162}។ យើងនឹងជ្រើសរើសគំរូតាងសាឡើងវិញដោយដាក់ចូលទៅវិញ ចំនួន

១០០០គំរូតាង រួចគណនាមធ្យមនៃគំរូតាងនីមួយៗ។ ការប៉ាន់ស្មានដោយតម្លៃប្រមាណចន្លោះអាចត្រូវបានកំណត់តាមរបៀបមួយចំនួនដូចជា៖

៣.១ ចន្លោះជឿជាក់ប៊ូតស្ត្រែបមូលដ្ឋាន (Basic Bootstrap Confidence Interval)

ប៊ូតស្ត្រែបមូលដ្ឋានគឺជាទម្រង់ការងារដែលងាយមួយនៅក្នុងការបង្កើតចន្លោះជឿជាក់ដោយប្រើកង់ទីល (quantile) ចេញពីរបាយប៊ូតស្ត្រែបនៃប៉ារ៉ាម៉ែត្រ។

ជំហាននៅក្នុងវិធីប៊ូតស្ត្រែបមូលដ្ឋានរួមមាន៖

1) ចំពោះគំរូតាងដើមដែលមានទំហំn គេគណនាស្ថិតិដែលត្រូវការ នៅក្នុងករណីឧទាហរណ៍នេះ វាគឺជាមធ្យម។
2) ជ្រើសរើសគំរូតាងសាឡើងវិញដោយដាក់ចូលទៅវិញ (resampling with replacement) ទៅតាមចំនួនដែលត្រូវការ ឧទាហរណ៍ ១០០០គំរូតាង ក្នុងនោះគំរូតាងនីមួយៗមានទំហំn។
3) គណនាស្ថិតិដែលត្រូវការ (មធ្យម) ចេញពីគំរូតាងសាឡើងវិញ។
4) កំណត់កង់ទីល $\alpha/2$ និង $1-\alpha/2$ ពីរបាយប៊ូតស្ត្រែបនៃតម្លៃស្ថិតិ ឧទាហរណ៍ កង់ទីល2.5%និង 97.5%សម្រាប់ចន្លោះជឿជាក់៩៥%។
5) ចន្លោះជឿជាក់កំណត់ដោយ (Davison & Hinkley, 1997b)

$$\left[2\hat{\theta}-\theta^{*}_{1-\alpha/2}, 2\hat{\theta}-\theta^{*}_{\alpha/2}\right] \qquad (1)$$

ដែល $\hat{\theta}$ តាងឱ្យស្ថិតិ (មធ្យម) ដែលគណនាពីគំរូតាងដើម និង θ^{*} មេគុណប៊ូតស្ត្រែបសម្រាប់កង់ទីលដែលគណនានៅក្នុងជំហានទី4។

កូដRសម្រាប់អនុវត្តវិធីនេះ បង្ហាញដូចខាងក្រោម៖

```
# សម្រាប់ការធ្វើឡើងវិញ
set.seed(123)
# ទិន្នន័យគំរូតាងដើម
data <- c(155, 163, 170, 158, 162)
# ចំនួនគំរូតាងប៊ូតស្ត្រែប
B <- 1000
# កម្រិតសារៈសំខាន់សម្រាប់ចន្លោះជឿជាក់៩៥%
alpha <- 0.05
# គណនាមធ្យមគំរូតាងដើម
theta_hat <- mean(data)
# បង្កើតគំរូតាងប៊ូតស្ត្រែបនិងគណនាមធ្យមរបស់វា
bootstrap_means <- replicate(B, mean(sample(data, size = length(data),
replace = TRUE)))
# គណនាកង់ទីលនៃមធ្យមក្នុងរបាយប៊ូតស្ត្រែប
quantile_lower <- quantile(bootstrap_means, alpha / 2)
quantile_upper <- quantile(bootstrap_means, 1 - alpha / 2)
```

```
# គណនាលីមីតក្រោមនិងលីមីតលើរបស់ចន្លោះជឿជាក់
LCL <- 2 * theta_hat - quantile_upper
UCL <- 2 * theta_hat - quantile_lower
# ការបង្ហាញលទ្ធផល
cat("Original Mean (theta_hat):", theta_hat, "\n")
cat("Reverse Bootstrap 95% Confidence Interval: [", LCL, ",", UCL,
"]\n")
```

៣.២ ប៊ូតស្ត្រែបពែសង់ទីល (Percentile Bootstrap)

ប៊ូតស្ត្រែបពែសង់ទីលមានដំណើរការស្រដៀងនឹងប៊ូតស្ត្រែបមូលដ្ឋានដោយប្រើពែសង់ទីលនៃរបាយប៊ូតស្ត្រែប។ វិធីនេះមានដំណើរការល្អនៅក្នុងករណីដែលរបាយប៊ូតស្ត្រែបមានភាពឆ្លុះនិងប្រមូលផ្តុំគ្នាជុំវិញស្ថិតិដែលយើងសង្កេត (Efron, 1982)។ វិធីប៊ូតស្ត្រែបពែសង់ទីលប្រើប្រាស់រូបមន្ត (Davison & Hinkley, 1997b, p. 203)

$$\left[\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2}\right] \qquad (2)$$

ដែល θ^* ជាមេគុណប៊ូតស្ត្រែប។

```
# សម្រាប់ការធ្វើឡើងវិញ
set.seed(123)
# ទិន្នន័យគំរូតាងដើម
data <- c(155, 163, 170, 158, 162)
# ចំនួនគំរូតាងប៊ូតស្ត្រែប
B <- 1000
# កម្រិតសារៈសំខាន់សម្រាប់ចន្លោះជឿជាក់៩៥%
alpha <- 0.05
# បង្កើតគំរូតាងប៊ូតស្ត្រែបនិងគណនាមធ្យមរបស់វា
bootstrap_means <- replicate(B, mean(sample(data, size = length(data),
replace = TRUE)))
# លីមីតក្រោមនិងលីមីតលើនៃចន្លោះជឿជាក់
LCL <- quantile(bootstrap_means, alpha / 2)
UCL <- quantile(bootstrap_means, 1 - alpha / 2)
# ការបង្ហាញលទ្ធផល
cat("Bootstrap Percentile 95% Confidence Interval: [", LCL, ",", UCL,
"]\n")
```

៣.៣ ប៊ូតស្ត្រែបស្ទូដិន (Studentized Bootstrap)

ប៊ូតស្ត្រែបស្ទូដិនអាចហៅឈ្មោះម្យ៉ាងទៀតប៊ូតស្ត្រែបt គឺជាវិធីកម្រិតខ្ពស់មួយសម្រាប់បង្កើតចន្លោះជឿជាក់ដោយការធ្វើកែសម្រួលបម្រែបម្រួលនៃតម្លៃស្ថិតិដែលគណនា។ វិធីនេះរាប់បញ្ចូលបម្រែបម្រួលល្អៀងស្តង់ដានៅក្នុងគំរូតាងប៊ូតស្ត្រែបទាំងអស់។ លក្ខណៈនេះហើយដែលធ្វើឱ្យវិធីនេះមានភាពខ្លាំងនិងជឿជាក់បានកាន់តែខ្ពស់បើប្រៀបធៀបនឹងវិធីប៊ូតស្ត្រែបផ្សេងៗទៀត។ វិធីនេះមានដំណើរការដូចតទៅនេះ៖

1) គណនាមធ្យមនៃគំរូតាងដើម $\hat{\theta}$ និងល្មៀងស្តង់ដារបស់វា $SE(\hat{\theta})$។

2) ជ្រើសរើសគំរូតាងឡើងវិញ(ដោយដាក់ចូលទៅវិញ)។ ចំពោះគំរូតាងឡើងវិញនីមួយៗគណនាមធ្យម θ^* ល្មៀងស្តង់ដារបស់វា $SE(\theta^*)$ និងស្ថិតិស្ទូដិន $t^* = \dfrac{\theta^* - \hat{\theta}}{SE(\theta^*)}$ ។

3) គណនាកង់ទីលនៃស្ថិតិស្ទូដិន(ដូចជា $t^*_{\alpha/2}$ និង $t^*_{1-\alpha/2}$)ដើម្បីប្រើប្រាស់វាជាតម្លៃវិនិច្ឆ័យក្នុងការគណនាចន្លោះជឿជាក់។

4) ចន្លោះជឿជាក់ $(1-a)100\%$ នៃប៉ារ៉ាម៉ែត្រ កំណត់ដោយ(Tibshirani & Efron, 1993)

$$\left[\hat{\theta} - t^*_{1-\alpha/2} SE(\hat{\theta}), \hat{\theta} - t^*_{\alpha/2} SE(\hat{\theta})\right] \quad (3)$$

```
# សម្រាប់ការធ្វើឡើងវិញ
set.seed(123)
#ទិន្នន័យគំរូតាងដើម
data <- c(155, 163, 170, 158, 162)
n <- length(data)
#កំណត់ចំនួនគំរូតាង
B <- 1000
#កំណត់កម្រិតសារៈសំខាន់សម្រាប់ចន្លោះជឿជាក់៩៥%
alpha <- 0.05
# គណនាមធ្យមគំរូតាងដើម
theta_hat <- mean(data)
# គណនាគម្លាតស្តង់ដា
SE_hat <- sd(data) / sqrt(n)
# រក្សាទុកស្ថិតិស្ទូដិនt
bootstrap_t <- numeric(B)
for (i in 1:B) {
 #បង្កើតគំរូតាងប៊ូតស្រ្ទែប
  bootstrap_sample <- sample(data, size = n, replace = TRUE)
   #គណនាមធ្យមប៊ូតស្រ្ទែបនិងគម្លាតស្តង់ដាប៊ូតស្រ្ទែប
  theta_b <- mean(bootstrap_sample)
  SE_b <- sd(bootstrap_sample) / sqrt(n)

 #គណនាស្ថិតិស្ទូដិន
  bootstrap_t[i] <- (theta_b - theta_hat) / SE_b
}
# គណនាកង់ទីលនៃស្ថិតិស្ទូដិន
t_lower <- quantile(bootstrap_t, alpha / 2)
t_upper <- quantile(bootstrap_t, 1 - alpha / 2)
```

```
#គណនាលីមីតក្រោមនិងលីមីតលើនៃចន្លោះជឿជាក់
LCL <- theta_hat - t_upper * se_hat
UCL <- theta_hat - t_lower * se_hat
# បង្ហាញលទ្ធផល
cat("Bootstrap Percentile 95% Confidence Interval:[", LCL, ",", UCL,
"]\n")
```

៤ សេចក្តីសន្និដ្ឋាន

នៅក្នុងអត្ថបទនេះ យើងបានស្វែងយល់អំពីគោលគំនិតគន្លឹះនៃប៊ូតស្ត្រែបស្ថិតិនិងការអនុវត្តរបស់វានៅក្នុងការប៉ាន់ស្មានប៉ារ៉ាម៉ែត្រសាកលស្ថិតិជាមួយនិងការបង្កើតចន្លោះជឿជាក់។ អត្ថបទចាប់ផ្តើមដោយការបង្ហាញអំពីវិធីប៊ូតស្ត្រែបដោយការសង្កត់ធ្ងន់លើលក្ខណៈមិនប៉ារ៉ាម៉ែត្ររបស់វិធីនិងសមត្ថភាពរបស់វាក្នុងការផ្តល់នូវការប៉ាន់ស្មានដែលមានរឹងមាំដោយមិនទាមទារនូវលក្ខខណ្ឌលើរបាយទិន្នន័យ។

បន្ទាប់មកយើងបានពិនិត្យមើលនូវរបៀបដែលវិធីប៊ូតស្ត្រែបមានដំណើរការប្រព្រឹត្តទៅតាមរយៈការជ្រើសរើសគំរូតាងសាឡើងវិញដោយមិនដាក់ចូលទៅវិញចំពោះទិន្នន័យគំរូតាងដើម ដើម្បីបង្កើតគំរូតាងក្លែងជាច្រើន។ ដំណើរការនេះនាំឱ្យអ្នកប្រើប្រាស់អាចប៉ាន់ស្មានបាននូវរបាយគំរូតាងតម្លៃស្ថិតិ ក៏ដូចជាគណនាចន្លោះជឿជាក់ នៅក្នុងករណីដែលវិធីស្ថិតិធម្មតាមិនអាចអនុវត្តបាន។ ជាងនោះទៅទៀត យើងបានបង្ហាញវិធីចំនួនបីសម្រាប់បង្កើតចន្លោះជឿជាក់សម្រាប់មធ្យមដោយប្រើប៊ូតស្ត្រែប៖ ចន្លោះជឿជាក់ប៊ូតស្ត្រែបមូលដ្ឋាន ប៊ូតស្ត្រែបពៃសង់ទីល និងប៊ូតស្ត្រែបស្ទូដិន។

ថ្វីបើវិធីប៊ូតស្ត្រែបគឺជាឧបករណ៍ដែលមានភាពអាចបត់បែនបាននិងមានដ៏អានុភាពក៏ដោយ ក៏វាមានដែនកំណត់របស់វាដែរ។ គុណវិបត្តិចម្បងមួយគឺប្រពលភាពនៃការគណនា ជាពិសេសសម្រាប់សំណុំទិន្នន័យធំ ឬស្ថិតិស្មុគស្មាញដែលទាមទារឱ្យមានគំរូតាងសាឡើងវិញយ៉ាងទូលំទូលាយ។ លើសពីនេះទៀតវិធីប៊ូតស្ត្រែបសន្មតថាគំរូតាងគឺជាតំណាងឱ្យសាកលស្ថិតិ ដែលមានន័យថាលទ្ធផលអាចនឹងនាំឱ្យមានការយល់ច្រឡំប្រសិនបើទិន្នន័យដើមមានភាពលំអៀង ឬមានភាពជាតំណាងមិនបានត្រឹមត្រូវល្អ។ ជាងនេះទៅទៀត សម្រាប់សំណុំទិន្នន័យដែលមានភាពទេរខ្លាំងឬមានទំហំតូច វិធីកម្រិតខ្ពស់ដូចជា "bias-corrected and accelerated (BCa) bootstrap" គឺភាពចាំបាច់ដើម្បីកែលម្អភាពត្រឹមត្រូវដល់ការវិភាគ។

សរុបសេចក្តីមក ប៊ូតស្ត្រែបស្ថិតិគឺជាវិធីដែលមានសារៈសំខាន់សម្រាប់ការវិភាគទិន្នន័យ តាមរយៈការផ្តល់ឧបករណ៍ដ៏រឹងមាំសម្រាប់ការសន្និដ្ឋានស្ថិតិនៅពេលដែលវិធីប៉ារ៉ាម៉ែត្រមិនអាចអនុវត្តបាន។ ទោះជាយ៉ាងណាក៏ដោយ អ្នកប្រើប្រាស់ត្រូវយល់ដឹងឱ្យបានច្បាស់អំពីលក្ខណៈនិងតម្រូវការនៃការគណនាដើម្បីប្រើវិធីនេះប្រកប

ដោយប្រសិទ្ធិភាព។ តាមរយៈការយល់ដឹងអំពីចំណុចខ្លាំងនិងអំពីភាពមានដែនកំណត់របស់វិធី អ្នកប្រើប្រាស់ទាញយកអត្ថប្រយោជន៍ច្រើនបំផុតពីវិធីសម្រាប់ការវិភាគទិន្នន័យឱ្យបានកាន់តែត្រឹមត្រូវ។

ឯកសារយោង

Davison, A. C., & Hinkley, D. V. (1997a). *Bootstrap methods and their application* (Issue 1). Cambridge university press.

Davison, A. C., & Hinkley, D. V. (1997b). *Bootstrap methods and their application* (Issue 1). Cambridge university press.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–228.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. In *The Jackknife, the Bootstrap and Other Resampling Plans*. https://doi.org/10.1137/1.9781611970319

Tibshirani, R. J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, *57*(1), 1–436.