ស្ថិតិវិទ្យាមិនប៉ារ៉ាម៉ែត្រ

Non-parametric Statistics

$$
\begin{aligned}
\mathrm{H}_0 &: \tilde{\mu}_1 = \tilde{\mu}_2 \\
\mathrm{H}_1 &: \tilde{\mu}_1 \neq \tilde{\mu}_2
\end{aligned}
$$

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

ម៉ុខ ម៉ារ៉ា

នាយកដ្ឋានគណិតវិទ្យានិងស្ថិតិ
វិទ្យាស្ថានវិទ្យាសាស្ត្រនិងបច្ចេកវិទ្យា

២០២១

មាតិកា

ទំព័រ

អារម្ភកថា

សៀវភៅនេះនឹងបង្ហាញអំពីវិធីមួយចំនួននៃស្ថិតិមិនប៉ារ៉ាម៉ែត្រ ជាពិសេសតេស្តមិនប៉ារ៉ាម៉ែត្រដែលជាជម្រើសមួយជំនួសឱ្យវិធីប៉ារ៉ាម៉ែត្រនៅក្នុងករណីដែលលក្ខខណ្ឌសម្រាប់ការអនុវត្តមិនត្រូវបានផ្ទៀងផ្ទាត់។ តាមរយៈសៀវភៅនេះ អ្នកអាននឹងបានសិក្សាអំពីវិធីមួយចំនួនដូចជា ការបង្កើតចន្លោះជឿជាក់សម្រាប់មេដ្យាន, តេស្តសម្មតិកម្មមួយចំនួន៖ តេស្តសញ្ញា តេស្តលំដាប់សញ្ញាវីលកូសុន តេស្តផលបូកលំដាប់វីលកូសុន តេស្តគ្រុស្កាល់-វ៉ាលីស តេស្តស្ទ៊ីតវ៉ល-វ៉ុលហ្វូវីត តេស្តភាពស៊ីគ្នា តេស្តភាពមិនអាស្រ័យ តេស្តភាពអូម៉ូហ្សែននៃសមាមាត្រ ការគណនាមេគុណកូរ៉េឡាស្យុងស្ពៀម៉ិន និងមេគុណកូរ៉េឡាស្យុងខេនដល។ ការគណនាខ្លះត្រូវពឹងផ្អែកលើ Excel និងសុសវែរRផងដែរ។ វាមានភាពអំណោយផលល្អច្រើនក្នុងការអានសៀវភៅនេះឱ្យបានងាយយល់ បើសិនជាអ្នកអានមានចំណេះដឹងស្ថិតិវិទ្យាជាមូលដ្ឋានគ្រឹះខ្លះៗដូចជា បំណែងចែកណរម៉ាល់ តេស្តz តេស្តt និងការវិភាគវ៉ារ្យ៉ង់មួយកត្តា។

អ្នករៀបរៀងសង្ឃឹមថា សៀវភៅនេះនឹងក្លាយជាវិភាគទានចំណេះដឹងមូលដ្ឋានផ្នែកស្ថិតិវិទ្យាដល់និស្សិត អ្នកសិក្សាស្រាវជ្រាវព្រមទាំងអ្នកស្រឡាញ់ចូលចិត្តស្ថិតិវិទ្យានិងការវិភាគទិន្នន័យ។

សូមទទួលបានលទ្ធផលជាផ្លែផ្កាអំពីការអានសៀវភៅនេះ។

ភ្នំពេញ, ខែធ្នូ ឆ្នាំ២០២១

ម៉ុង ម៉ារ៉ា

សេចក្តីផ្តើម

នៅក្នុងស្ថិតិវិទ្យាពិពណ៌នា មេដ្យានត្រូវបានប្រើប្រាស់ជារង្វាស់និន្នាការកណ្តាលជាជាងមធ្យមនព្វន្ត បើសិនជាសំណុំទិន្នន័យប្រាស់ចាកឆ្ងាយពីភាពណរម៉ាល់។ ជាងនេះទៅទៀត វិធីប៉ារ៉ាម៉ែត្រនៅក្នុងស្ថិតិ ទាមទារនូវការគោរពលក្ខខណ្ឌមួយចំនួនដូចជាទំហំគំរូតាង និង ភាពណរម៉ាល់នៃទិន្នន័យដែលត្រូវធ្វើការវិភាគ។ នៅពេលដែលលក្ខខណ្ឌទាំងនោះមិនត្រូវបានគោរព ជម្រើសនៃវិធីស្ថិតិផ្សេងត្រូវប្រកាន់យក។ វិធីទាំងនោះគឺជាវិធីនៃ *ស្ថិតិមិនប៉ារ៉ាម៉ែត្រ* ដែលគេប្រើពាក្យ *វិធីមិនប៉ារ៉ាម៉ែត្រ* សម្រាប់ហៅវាផងដែរ។

ស្ថិតិមិនប៉ារ៉ាម៉ែត្រ(non-parametric statistics) គឺជាមែកធាងមួយនៃស្ថិតិវិទ្យា ដែលមិនពឹងផ្អែកលើគ្រួសារបំណែងចែកប្រូបាប៊ីលីតេដែលកំណត់ដោយប៉ារ៉ាម៉ែត្រ។ វិធីស្ថិតិមិនប៉ារ៉ាម៉ែត្រច្រើនប្រើប្រាស់ទិន្នន័យក្នុងកម្រិតលំដាប់ (ordinal-level data)។ ប៉ុន្តែទោះជាយ៉ាងណាក៏ដោយ វិធីស្ថិតិមិនប៉ារ៉ាម៉ែត្រជាវិធីដែលត្រូវបានប្រើប្រាស់ទៅលើទិន្នន័យក្នុងកម្រិតនាមករណ៍, កម្រិតលំដាប់, កម្រិតចន្លោះ(interval) ឬ កម្រិតផលធៀប(ratio)ផងដែរនៅក្នុងករណីដែលលក្ខខណ្ឌមិនគោរពតាមការប្រើប្រាស់វិធីប៉ារ៉ាម៉ែត្រ។

ជំពូកទី១

ចន្លោះជឿជាក់សម្រាប់មេដ្យាន

Confidence Interval for the Median

នៅក្នុងការពិពណ៌នាទិន្នន័យ មេដ្យានត្រូវបានចាត់ទុកជារង្វាស់តម្លៃកណ្តាល មួយនៅក្នុងចំណោមតម្លៃផ្សេងៗទៀតដូចជា មធ្យមនព្វន្ត និង ម៉ូដ។ វាក៏ត្រូវគេហៅថាជារង្វាស់ទីតាំងមួយផងដែរ។ និយមន័យមេដ្យានលើកឡើងថា ពាក់កណ្តាលនៃតម្លៃនៅក្នុងទិន្នន័យគឺស្ថិតនៅក្រោមមេដ្យាននិងពាក់កណ្តាលទៀតស្ថិតនៅលើមេដ្យាន។ ចំពោះទិន្នន័យបរិមាណដែលមានបំណែងចែកដែលមានភាពទេរខ្លាំងចាកឆ្ងាយពីសណ្ឋានជាជួង នោះមេដ្យានគឺជារង្វាស់និន្នាការកណ្តាលដែលសមស្របជាងមធ្យម។

នៅក្នុងការបង្កើតចន្លោះជឿជាក់ សម្រាប់ប៉ាន់ស្មានមធ្យម មានលក្ខខណ្ឌមួយកំណត់ថាបំណែងចែកសាកលស្ថិតិ ត្រូវតែជាបំណែងចែកណរម៉ាល់ ឬ បើមិនដូច្នោះទេ ទំហំគំរូតាងត្រូវធំជាង 30។ នៅក្នុងករណីជាក់ស្តែងជាច្រើនលក្ខខណ្ឌទាំងនេះមិនត្រូវបានផ្ទៀងផ្ទាត់។ ចំពោះបញ្ហានេះ គេត្រូវប្រើប្រាស់វិធីផ្សេងវិញ។ ការប្រើប្រាស់មេដ្យានគឺជាជម្រើសដ៏ញឹកញាប់មួយ។

សន្មតថាបំណែងចែកសាកលស្ថិតិជាបំណែងចែកជាប់។
តាង M ជាមេដ្យាន និង X ជាតម្លៃនៅក្នុងបំណែងចែក
នោះគេបាន

$$P(X \le M) = P(X \ge M) = \frac{1}{2} \quad (១.១)$$

សមីការនេះបានបង្ហាញឱ្យឃើញថា ចំពោះគំរូតាងចៃដន្យមួយ $X_1, X_2, ..., X_n$ ជ្រើសរើសចេញអំពីសាកលស្ថិតិមានដែលមានមេដ្យាន M បំណែងចែកនៃតម្លៃដែលតូចជាង M គោរពតាមបំណែងចែកទ្វេរធាដែលមានប៉ារ៉ាម៉ែត្រ n និង $p = 0.5$ ដោយមិនទាក់ទងនឹងបំណែងចែកសាកលស្ថិតិទេ។ ចំណុចនេះ មានន័យថា បើយើងតាង N^- ជាចំនួនតម្លៃដែលតូចជាង M នោះបំណែងចែកនៃ N^- គឺជាបំណែងចែកទ្វេធាដែលមានប៉ារ៉ាម៉ែត្រ n និង $p = 0.5$ ចំពោះគំរូតាងមួយដែលមានទំហំ n។ ដូច្នេះ គេអាចបង្កើតចន្លោះជឿជាក់មួយសម្រាប់មេដ្យាន ដោយប្រើប្រាស់បំណែងចែកទ្វេធា។

ចំពោះប្រូបាប៊ីលីតេ α ណាមួយ គេអាចកំណត់ a និង b ដែលផ្ទៀងផ្ទាត់

$$\begin{aligned} P(N^- \le a) &= \sum_{i=0}^{a} \binom{n}{i} (0.5)^i (0.5)^{n-i} \\ &= \sum_{i=0}^{a} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2} \end{aligned} \quad (១.២)$$

និង

$$\begin{aligned} P(N^- \ge b) &= \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i} \\ &= \sum_{i=b}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \frac{\alpha}{2} \end{aligned} \quad (១.៣)$$

តាង $X_{(1)}, X_{(2)}, ..., X_{(a)}, ..., X_{(b)}, ..., X_{(n)}$ គឺជាតម្លៃដែលបានរៀបតាមលំដាប់ ត្រូវគ្នានឹងតម្លៃ $X_1, X_2, ..., X_n$ ក្នុងទិន្នន័យ។ នោះមេដ្យាននៃសាកលស្ថិតិដែលយើងប៉ាន់ស្មាន គឺធំជាងតម្លៃ $X_{(b)}$ ដោយឱកាស $(\alpha / 2)100\%$ និង តូចជាងតម្លៃ $X_{(a)}$ ដោយឱកាស $(\alpha / 2)100\%$។ ដូច្នេះ ចន្លោះជឿជាក់ $(1-\alpha)100\%$ សម្រាប់ មេដ្យានរបស់ស្ថិតិសាកលគឺ

$$X_{(a)} < M < X_{(b)} \qquad (១.៤)$$

ជាការសង្ខេបមកវិញ ដើម្បីរកចន្លោះជឿជាក់ $(1-\alpha)100\%$ សម្រាប់មេដ្យានរបស់សាកលស្ថិតិ គេអាចធ្វើតាមជំហានខាងក្រោម (Ramachandran & Tsokos, 2009)

១. រៀបតម្លៃនៅក្នុងទិន្នន័យគំរូតាងតាមលំដាប់កើន។

២. ប្រើប្រាស់បំណែងចែកប្រូបាប៊ីលីតេទ្វេធាដែលមានប៉ារ៉ាម៉ែត្រ n ដែលជាទំហំគំរូតាង និង $p = 0.5$ ដើម្បីរកតម្លៃ a ដែលផ្ទៀងផ្ទាត់[1] $P(X \le a)$ ស្មើ $\frac{\alpha}{2}$ ឬនៅក្បែរ $\frac{\alpha}{2}$ បំផុត។

៣. គណនា $b = n + 1 - a$ ។

៤. លីមីតក្រោមរបស់ចន្លោះជឿជាក់គឺតម្លៃទី a ហើយលីមីតលើរបស់ចន្លោះជឿជាក់គឺតម្លៃទី b នៅក្នុងទិន្នន័យដែលតម្លៃត្រូវបានរៀបតាមលំដាប់នៅក្នុងជំហាន១។

---

[1] ការរកតម្លៃ a ដែលផ្ទៀងផ្ទាត់ $P(X \le a) = \frac{\alpha}{2}$ គឺតាមរយៈការប្រើប្រាស់បំណែងចែកប្រូបាប៊ីលីតេទ្វេធា។

លក្ខខណ្ឌសម្រាប់ការប្រើប្រាស់វិធីនេះ គឺបំណែងចែកសាកលស្ថិតិ គឺជាបំណែងចែកជាប់ រីឯគំរូតាងគឺជាប្រភេទគំរូតាងចៃដន្យងាយ។

ឧទាហរណ៍១.១

គំរូតាងចៃដន្យនៃសិស្សវិទ្យាល័យចំនួន50នាក់ ត្រូវបានជ្រើសរើស។ ចំនួនម៉ោងក្នុង១សប្តាហ៍ដែលពួកគេរៀនគណិតវិទ្យាបន្ថែម ក្រៅពីម៉ោងផ្លូវការក្នុងថ្នាក់ត្រូវបានកត់ត្រាដូចបង្ហាញនៅខាងក្រោម។ ចូររកចន្លោះជឿជាក់95%ប៉ាន់ស្មានមេដ្យាននៃចំនួនម៉ោងដែលសិស្សរៀនគណិតវិទ្យាបន្ថែមក្រៅពីម៉ោងផ្លូវការក្នុងថ្នាក់។

3.00	7.00	6.00	6.00	5.00
6.00	4.00	5.00	3.00	6.00
5.00	3.00	6.00	6.00	6.00
5.00	4.00	6.00	3.00	5.00
5.00	6.00	3.00	6.00	6.00
5.00	2.00	12.00	6.00	6.00
6.00	3.00	6.00	6.00	8.00
6.00	3.00	3.00	10.00	8.00
7.00	3.00	6.00	3.00	8.00
6.00	6.00	8.00	6.00	6.00

ដំណោះស្រាយ

ចំពោះចន្លោះជឿជាក់95%, $\alpha = 0.05$។ ដូច្នេះ $\alpha/2 = 0.025$។ នៅក្នុងបំណែងចែកទ្វេធាដែល $n = 20$ និង $p = 1/2$ យើងឃើញថា $P(X \le 18) = 0.032$។ ដូច្នេះ $a = 18$ នាំឱ្យបាន *ប្រូបាប៊ីលីតេកើន* (cumulative probability) មានតម្លៃ នៅក្បែរ $\alpha/2 = 0.025$ ជាងគេ។ ទិន្នន័យដែលមានតម្លៃរៀប

តាមលំដាប់កើនគឺ

2.00	3.00	6.00	6.00	6.00
3.00	4.00	6.00	6.00	6.00
3.00	4.00	6.00	6.00	7.00
3.00	5.00	6.00	6.00	7.00
3.00	5.00	6.00	6.00	8.00
3.00	5.00	6.00	6.00	8.00
3.00	5.00	6.00	6.00	8.00
3.00	5.00	6.00	6.00	8.00
3.00	5.00	6.00	6.00	10.00
3.00	5.00	6.00	6.00	12.00

នៅក្នុងទិន្នន័យនេះ តម្លៃទី18ដែលស្មើនឹង5 គឺជាតម្លៃលីមីតក្រោមរបស់ចន្លោះ។ តម្លៃទី50+1-18=33 ដែលស្មើនឹង6គឺជាលីមីលើរបស់ចន្លោះ។ ដូច្នេះគេជឿជាក់ក្នុងកម្រិត95%ថាមេដ្យាននៃចំនួនម៉ោងដែលសិស្សរៀនគណិត វិទ្យាបន្ថែមក្រៅពីម៉ោងផ្លូវការក្នុងថ្នាក់ ស្ថិតនៅក្នុងចន្លោះពី5ម៉ោងទៅ6ម៉ោងនៅក្នុងមួយសប្តាហ៍។

ការបង្កើតចន្លោះជឿជាក់ប៉ាន់ស្មានមេដ្យាន តាមវិធីដែលបង្ហាញនៅក្នុង(Bland, 2015)មានរបៀបដូចតទៅនេះ។

នៅក្នុងបំណែងចែកទ្វេជា ដែលមានប៉ារ៉ាម៉ែត្រ n និង p មធ្យមរបស់បំណែងចែកគឺ $\mu = np$ និងគម្លាតស្តង់ដាគឺ $\sigma = \sqrt{np(1-p)}$ ។

បន្ទាប់មក គណនា

$$j = np - z_{\alpha/2}\sqrt{np(1-p)} \qquad (១.៥.ក)$$

$$k = np + z_{\alpha/2}\sqrt{np(1-p)} \qquad (១.៥.ខ)$$

តម្លៃ j និង k ត្រូវបង្គត់ឡើងទៅជាចំនួនគត់ធំបន្ទាប់។ លីមីតលើរបស់ចន្លោះជឿជាក់ គឺជាតម្លៃទី j និងលីមីតក្រោមរបស់ចន្លោះជឿជាក់ គឺជាតម្លៃទី k នៅក្នុងទិន្នន័យដែលមានតម្លៃត្រូវបានរៀបតាមលំដាប់រួចហើយ។ គួររម្លឹកឡើងវិញថាមេដ្យានគឺជាកង់ទីល[2] ដែលស្ថិតនៅចំណុចកណ្តាលទិន្នន័យ។ ដូច្នេះ ចំពោះចន្លោះជឿជាក់95%នៃមេដ្យាន

$$j = 0.5n - 1.96\sqrt{0.25n} \qquad (១.៦.ក)$$

$$k = 0.5n + 1.96\sqrt{0.25n} \qquad (១.៦.ខ)$$

ចំពោះឧទាហរណ៍១.១ យើងមាន $n = 50$ នោះ

$$i = 0.5\times 50 - 1.96\sqrt{0.25\times 50} = 18.07$$

$$k = 0.5\times 50 + 1.96\sqrt{0.25\times 50} = 31.93$$

ដោយបង្គត់ឡើងទៅចំនួនគត់ធំបន្ទាប់ យើងបាន $i = 19$ និង $k = 32$ ដូច្នេះ លីមីតក្រោមនៃចន្លោះជឿជាក់គឺតម្លៃទី១៩ ដែលស្មើនឹង5 ។ រីឯលីមីតលើនៃចន្លោះជឿជាក់គឺតម្លៃទី៣២ ដែលស្មើនឹង6 មានន័យថាមេដ្យាននៃចំនួនម៉ោងដែលសិស្សរៀនគណិតវិទ្យាបន្ថែមក្រៅពីម៉ោងផ្លូវការក្នុងថ្នាក់ស្ថិតនៅក្នុងចន្លោះពី5ម៉ោង

---

[2]កង់ទីល(quantile)សំដៅដល់ចំណុចដែលចែកសំណុំទិន្នន័យ មួយទៅជាក្រុមៗទំហំស្មើៗគ្នា។ មេដ្យានជាកង់ទីលមួយ ដែលបែងចែកសំណុំទិន្នន័យជាពីរចំណែកស្មើគ្នា។ជួនកាលគេហៅហ្វ្រាក់ទីល(fractile)ជំនួសឱ្យកង់ទីល។

ទៅ៦ម៉ោងក្នុងមួយសប្តាហ៍។

លំហាត់អនុវត្តន៍

គំរូតាងចៃដន្យដែលជ្រើសរើសចេញពីសាកលស្ថិតិមួយរួមមានតម្លៃដូចខាងក្រោម។ ចូររកចន្លោះជឿជាក់៩៨%នៃមេដ្យានរបស់សាកលស្ថិតិ ដោយប្រើរូបមន្ត១.៤ និងម្តងទៀតតាមវិធីប៉ាន់ស្មានដោយប្រើរូបមន្ត១.៦.ក និង ១.៦.ខ។

ជំពូកទី២

តេស្តសញ្ញា

Sign Test

ការធ្វើតេស្តលើមធ្យម ដែលមានសម្មតិកម្មសូន្យកំណត់ដោយ $\mu = \mu_0$ (μ_0 ជាតម្លៃមធ្យមសន្មតមួយ) ទាមទារលក្ខខណ្ឌលើភាពណរម៉ាល់របស់សាកលស្ថិតិ ឬ ទំហំគំរូតាងលើសពី30។ ការធ្វើតេស្តលើមធ្យមនេះអាចប្រើតេស្តtបើមិនស្គាល់គម្លាតស្តង់ដា (σ)របស់សាកលស្ថិតិ និង ប្រើតេស្តzបើស្គាល់គម្លាតស្តង់ដារបស់សាកលស្ថិតិ។

នៅក្នុងករណីដែលលក្ខខណ្ឌទាំងពីរនេះ មិនត្រូវបានផ្ទៀងផ្ទាត់ តេស្តមិនប៉ារ៉ាម៉ែត្រត្រូវបានពិចារណាជាជម្រើសជំនួសឱ្យតេស្តt ឬ តេស្តzដែលជាវិធីប៉ារ៉ាម៉ែត្រ។

តេស្តសញ្ញាជាតេស្តមិនប៉ារ៉ាម៉ែត្រ ដែលអាចប្រើប្រាស់សម្រាប់ធ្វើតេស្តមេដ្យានរបស់សាកលស្ថិតិ។

គេតាង $\tilde{\mu}$ ជាមេដ្យានរបស់សាកលស្ថិតិ និង $\tilde{\mu}_0$ ជាតម្លៃ សន្មតនៃមេដ្យាន(តម្លៃដែលប្រើប្រាស់នៅក្នុងសម្មតិកម្មសម្រាប់ការធ្វើតេស្ត)។

សម្មតិកម្មសម្រាប់តេស្តសញ្ញា មួយគំរូតាងកំណត់ដោយ

តារាង២.១ *សម្មតិកម្មសម្រាប់តេស្ត មួយគំរូតាង*

សម្មតិកម្មសូន្យ	**សម្មតិកម្មឆ្លាស់**
	H$_1$: $\tilde{\mu} \neq \tilde{\mu}_0$ (តេស្តសងខាង)
H$_0$: $\tilde{\mu} = \tilde{\mu}_0$	H$_1$: $\tilde{\mu} > \tilde{\mu}_0$ (តេស្តខាងស្តាំ)
	H$_1$: $\tilde{\mu} < \tilde{\mu}_0$ (តេស្តខាងឆ្វេង)

នៅក្នុងន័យទូទៅ មេដ្យានសំដៅលើតម្លៃដែលនៅចំកណ្តាលទិន្នន័យ ដែលមានតម្លៃរៀបតាមលំដាប់រួចហើយ។ ចំពោះអថេរចៃដន្យ X មួយមេដ្យាន $\tilde{\mu}$ ត្រូវបានកំណត់ឡើងដោយផ្ទៀងផ្ទាត់ $P\left(X \geq \tilde{\mu}\right) = 0.5$ និង $P\left(X \leq \tilde{\mu}\right) = 0.5$ (Walpole et al.2012, p.656)។ នៅក្នុងករណីអថេរ ចៃដន្យជាប់គេបាន

$$P\left(X > \tilde{\mu}\right) = P\left(X < \tilde{\mu}\right) = 0.5 \quad (២.១)$$

នៅក្នុងដំណើរការនៃការធ្វើតេស្តតម្លៃទាំងឡាយណាដែលធំជាង $\tilde{\mu}_0$ ត្រូវជំនួសដោយសញ្ញាបូក(+) និង តម្លៃទាំងឡាយណាដែលតូចជាង $\tilde{\mu}_0$ ត្រូវបានជំនួសដោយសញ្ញាដក(−)។ បើសម្មតិកម្មសូន្យពិតនិងសាកលស្ថិតិមានភាពឆ្លុះ នោះចំនួននៃសញ្ញាបូកស្មើនឹងចំនួននៃសញ្ញាដក។ នៅពេលដែលសញ្ញាណាមួយមានចំនួនច្រើនជាងសញ្ញាមួយទៀតដោយប្រាកដ(មិនមែនដោយចៃដន្យ) នោះសម្មតិកម្មសូន្យនឹងត្រូវបានបដិសេធ។ នៅក្នុងការអនុវត្ត តម្លៃនៅក្នុងគំរូតាងដែលស្មើនឹងមេដ្យាន($\tilde{\mu}$) មិនត្រូវបានគិត(មិនរាប់បញ្ចូល) ព្រោះផលដករវាងមេដ្យាននិងតម្លៃទាំងនោះស្មើសូន្យ

(គ្មានសញ្ញា)។ នៅក្នុងករណីនេះ ទំហំគំរូតាងត្រូវបានកាត់បន្ថយ។

នៅក្នុងសៀវភៅនេះ អ្នករៀបរៀងកំណត់តម្លៃស្ថិតិតេស្តនៃតេស្តសញ្ញា គឺជាអថេរចៃដន្យទ្វេធា X ដែលតាងឱ្យចំនួនសញ្ញាបូកនៅក្នុងគំរូតាង។ បើសិនជាសម្មតិកម្មសូន្យ($\tilde{\mu} = \tilde{\mu}_0$)ពិតនោះប្រូបាប៊ីលីតេដែលនាំឱ្យបានសញ្ញាបូក ស្មើនឹងប្រូបាប៊ីលីតេដែលនាំឱ្យបានសញ្ញាដក ហើយវាស្មើនឹង0.5។ ដូច្នេះ តម្លៃស្ថិតិតេស្តគឺជាចំនួនសញ្ញាបូក ដែលជាតម្លៃរបស់អថេរចៃដន្យទ្វេធាមានប៉ារ៉ាម៉ែត្រ $p = 0.5$ និង n ។ n គឺជាចំនួនសរុបនៃសញ្ញាទាំងបូកទាំងដកនៅក្នុងគំរូតាង។ តម្លៃP(P-value)នៃតេស្តអាចគណនា ដោយប្រើប្រាស់បំណែងចែកប្រូបាប៊ីលីតេទ្វេធា។

ចំពោះតេស្តខាងឆ្វេងដែលមានសម្មតិកម្មH_0: $\tilde{\mu} = \tilde{\mu}_0$ និង H_1: $\tilde{\mu} < \tilde{\mu}_0$ សម្មតិកម្មសូន្យ(H_0)នឹងត្រូវបដិសេធ និងទទួលយកសម្មកម្មឆ្លាស់ បើសិនជាសមាមាត្រនៃចំនួនសញ្ញាបូក(+)ពិតជាតូចជាង0.5។ តម្លៃPកំណត់ដោយ

$$P - \text{value} = P\left(X \le x\right) \quad (២.២)$$

ចំពោះតេស្តខាងស្តាំដែលមានសម្មតិកម្មH_0: $\tilde{\mu} = \tilde{\mu}_0$ និង H_1: $\tilde{\mu} > \tilde{\mu}_0$ សម្មតិកម្មសូន្យនឹងត្រូវបដិសេធនិងទទួលយកសម្មតិកម្មឆ្លាស់បើសិនជាសមាមាត្រនៃចំនួនសញ្ញាបូក(+)ពិតជាធំជាង0.5។ តម្លៃPកំណត់ដោយ

$$P - \text{value} = P\left(X \ge x\right) \quad (២.៣)$$

ចំពោះតេស្តសងខាងដែលមានសម្មតិកម្មH_0: $\tilde{\mu} = \tilde{\mu}_0$ និងH_1: $\tilde{\mu} \neq \tilde{\mu}_0$ តម្លៃPកំណត់ដោយ

$$P - \text{value} = 2P\left(X \leq x\right) \text{ បើ } x < \frac{n}{2} \qquad (២.៤.ក)$$

$$P - \text{value} = 2P\left(X \geq x\right) \text{ បើ } x > \frac{n}{2} \qquad (២.៤.ខ)$$

នៅក្នុងករណីដែល n (ចំនួនសញ្ញាសរុប) ធំជាង10នោះ លក្ខខណ្ឌសម្រាប់ការប៉ាន់ស្មានណរម៉ាល់(normal approximation) ចំពោះបំណែងចែកទ្វេជាត្រូវបានផ្ទៀងផ្ទាត់(Ramachandran & Tsokos, 2009, p.610)។ គួររម្លឹកថា ការប៉ាន់ស្មានណរម៉ាល់ ចំពោះបំណែងចែកទ្វេជាអាចប្រើប្រាស់បានគឺនៅពេលដែល $np \geq 5$ និង $n\left(1 - p\right) \geq 5$ ។ ដោយសារនៅក្នុងករណីនេះ $p = 0.5$ នោះត្រឹមតែ n មានទំហំចាប់ពី10ឡើងទៅ គឺគ្រប់គ្រាន់សម្រាប់ ការប៉ាន់ស្មានណរម៉ាល់។ ដូច្នេះ តម្លៃស្ថិតិតេស្តអាចរកតាមបំណែង ចែកណរម៉ាល់ស្តង់ដា។ ឯកសារខ្លះដូចជា(Bluman, 2014, p.693)និង(Triola, 2018, p.647)ជាដើមប្រើប្រាស់ការប៉ាន់ស្មាន ណរម៉ាល់នៅក្នុងតេស្តសញ្ញា នៅក្នុងករណីដែល $n > 25$ ។

គួររម្លឹកឡើងវិញផងដែរ នៅពេលដែលការប៉ាន់ស្មាន ណរម៉ាល់ត្រូវបានប្រើប្រាស់ គេត្រូវអនុវត្តកំណែតម្រូវភាពជាប់ ដោយបូកឬដក0.5ទៅលើ ឬ ចេញពីតម្លៃ x ។ មធ្យមរបស់បំណែង ចែកកំណត់ដោយ $np = 0.5n$ និងគម្លាតស្តង់ដាស្មើនឹង $\sqrt{np\left(1 - p\right)} = \sqrt{0.25n}$ ដែល n គឺជាចំនួនសរុបនៃសញ្ញា។

ឧទាហរណ៍២.១

ក្រុមហ៊ុនលក់បរិក្ខាអគ្គិសនីមួយអះអាងថាថ្មម៉ូទ័រកាត់មួយប្រភេទ អាចប្រើប្រាស់បានរយៈពេល1.8ម៉ោង(ជាតម្លៃមេដ្យាន)។ គំរូតាងចៃដន្យមួយត្រូវបានជ្រើសរើស ហើយរយៈពេលដែលថ្មម៉ូទ័រកាត់អាចប្រើប្រាស់បានមុនពេលដែលត្រូវសាកថ្មឡើងវិញ មានដូចខាងក្រោម។

1.5 2.2 0.9 1.3 2.0 1.6 1.8 1.5 2.0 1.2 1.7

ចូរធ្វើតេស្តអំណះអំណាងខាងលើដោយប្រើកម្រិតសារៈសំខាន់0.05។

ដំណោះស្រាយ

សម្មតិកម្ម

H$_0$: $\tilde{\mu}=1.8$ (មេដ្យានស្មើ1.8ម៉ោង។)

H$_1$: $\tilde{\mu}\neq 1.8$ (មេដ្យានខុសពី1.8ម៉ោង។)

ការគណនាតម្លៃស្ថិតិតេស្ត

គណនាផលដក(ដក1.8ពីតម្លៃនីមួយៗ)យើងបាន

1.5	2.2	0.9	1.3	2.0	1.6	1.8	1.5	2.0	1.2	1.7
-0.3	0.4	-0.9	-0.5	0.2	-0.2	0	-0.3	0.2	-0.6	-0.1

តាមរយៈផលដកខាងលើ ផលដកអវិជ្ជមាន(សញ្ញាដក)មានចំនួន7 និង ផលដកវិជ្ជមាន(សញ្ញាបូក)មានចំនួន3។ ដូច្នេះយើងបាន $n=10, x=3$ និង $n/2=5$។ តម្លៃP អាចរកតាមរូបមន្ត ២.៣.ក។

$$\begin{aligned} P\text{-value} &= 2P\left(X \le 3\right) \\ &= 2\left[P(0)+P(1)+P(2)+P(3)\right] \end{aligned}$$

$$\text{P-value} = 2\left(0.001 + 0.010 + 0.044 + 0.117\right) = 0.344$$

ការរកតម្លៃP អាចធ្វើនៅក្នុងExcelតាមរយៈ

=2 *BINOM.DIST(3,10,0.5,TRUE)

ការសម្រេចចិត្ត

ដោយសារតម្លៃPស្មើ0.344 ធំជាងកម្រិតសារៈសំខាន់ α=0.05 នោះការសម្រេចចិត្តគឺមិនបដិសេធសម្មតិកម្មសូន្យទេ។

ការសង្ខេបលទ្ធផលតេស្ត

គេមិនមានភស្តុតាងគ្រប់គ្រាន់ដើម្បីបដិសេធអំណះអំណាងដែលថាមេដ្យានស្មើ1.8ម៉ោង នោះទេ។

ឧទាហរណ៍២.២

នាយកគ្រប់គ្រងមន្ទីរពេទ្យមួយអះអាងថាមេដ្យាននៃចំនួនអ្នកដែលមកពិគ្រោះជំងឺគឺ80នាក់ក្នុងមួយថ្ងៃ។ គំរូតាងចៃដន្យមួយដែលមានចំនួន20ថ្ងៃត្រូវបានជ្រើសរើស។ ចំនួនអ្នកមកពិគ្រោះជំងឺក្នុងមួយថ្ងៃៗក្នុងកំលុងពេល20ថ្ងៃនោះត្រូវបានកត់ត្រា។ តើគេអាចបដិសេធអំណះអំណាងខាងលើដែរឬទេ ចំពោះកម្រិតសារៈសំខាន់5%?

82	85	93	81	80
72	84	88	81	83
86	95	89	74	62
105	80	86	81	87

ដំណោះស្រាយ

សម្មតិកម្ម

H_0: $\tilde{\mu} = 80$ (មេដ្យាននៃចំនួនអ្នកជំងឺគឺ80នាក់។)

H_1: $\tilde{\mu} \neq 80$ (មេដ្យាននៃចំនួនអ្នកជំងឺខុសពី80នាក់។)

ការគណនា

តម្លៃមេដ្យានសន្មតគឺ80(នៅក្នុងសម្មតិកម្ម)។ ដក80 ចេញពីតម្លៃនីមួយៗក្នុងគំរូតាង និង ពិនិត្យសញ្ញា យើងបាន

+	+	+	+	0
-	+	+	+	+
+	+	+	-	-
+	0	+	+	+

ទំហំគំរូតាងដើមគឺ 20។ ចំនួនសរុបនៃសញ្ញាមាន18 ដែលចំនួនសញ្ញាដកមាន3និងសញ្ញាបូកមាន15 (សូន្យ មិនត្រូវបានរាប់)។ ដូច្នេះ $x = 15$, $n = 18$ និង $p = 0.5$ ។ តម្លៃPគណនាតាមរូបមន្ត ២.៤.៥ ដូច្នេះ

$$\text{P-value} = 2P\left(X \geq 15\right)$$

ដោយប្រើប្រាស់ការប៉ាន់ស្មានណរម៉ាល់យើងបាន

$$\text{P-value} \approx 2P\left(Z > \frac{14.5 - 18 \times 0.5}{\sqrt{18 \times 0.5\left(1 - 0.5\right)}}\right)$$

$$\text{P-value} \approx 2P\left(Z > 2.59\right) = 2\left(1 - 0.9952\right) = 0.0048$$

ការសម្រេចចិត្ត

ការសម្រេចចិត្តគឺត្រូវបដិសេធសម្មតិកម្មសូន្យព្រោះតម្លៃP តូចជាងកម្រិតសារៈសំខាន់5%។

ការសង្ខេបលទ្ធផល

គេមានភស្តុតាងគ្រប់គ្រាន់ក្នុងការបដិសេធអំណះអំណាងដែលនិយាយថាមេដ្យាននៃចំនួនអ្នកដែលមកពិគ្រោះជំងឺស្មើនឹង80នាក់ក្នុងមួយថ្ងៃ។ ដូច្នេះគេសន្និដ្ឋានក្នុងកម្រិតសារៈសំខាន់5%ថាមេដ្យាននៃចំនួនអ្នកជំងឺខុសពី80នាក់ក្នុង១ថ្ងៃ។

តេស្តសញ្ញាអាចប្រើប្រាស់ជំនួសតេស្តសម្រាប់ប្រៀបធៀបមធ្យមចំពោះគំរូតាងអាស្រ័យ(dependent sample) នៅក្នុងករណីដែលលក្ខខណ្ឌសម្រាប់តេស្តមិនត្រូវបានផ្ទៀងផ្ទាត់។

ឧទាហរណ៍២.៣

ការសិក្សាស្រាវជ្រាវមួយត្រូវបានធ្វើឡើងដើម្បីពិនិត្យមើលប្រសិទ្ធភាពនៃឱសថបញ្ចុះទម្ងន់។
គំរូតាងចៃដន្យដែលមានស្ត្រី៨នាក់ត្រូវបានជ្រើសរើសសម្រាប់ការសិក្សានេះ។
ទម្ងន់គិតជាផោនមុនការប្រើប្រាស់និងក្រោយការប្រើប្រាស់ត្រូវបានកត់ត្រា។ ចំពោះកម្រិតសារៈសំខាន់0.05 តើគេអាចសន្និដ្ឋានថាការប្រើឱសថនេះធ្វើឱ្យប្រែប្រួលទម្ងន់របស់ស្ត្រីដែរឬទេ?

អ្នកចូលរួម	A	B	C	D	E	F	G	H
មុន	187	163	201	158	139	143	198	154
ក្រោយ	178	162	188	156	133	150	175	150

ដំណោះស្រាយ

តាង $\tilde{\mu}_1$ មេដ្យានទម្ងន់(គិតជាផោន)មុនប្រើឱសថ, $\tilde{\mu}_2$ មេដ្យានទម្ងន់ក្រោយពេលប្រើឱសថ។

សម្មតិកម្ម

H$_0$: $\tilde{\mu}_1 - \tilde{\mu}_2 = 0$ (ទម្ងន់គ្មានការប្រែប្រួល។)

H$_1$: $\tilde{\mu}_1 - \tilde{\mu}_2 \neq 0$ (ទម្ងន់មានការប្រែប្រួល។)

ការគណនា

គណនាផលដក(ទម្ងន់មុនប្រើ - ទម្ងន់ក្រោយប្រើ) និងពិនិត្យសញ្ញា យើងបាន

\+ + + + + + - + +

សញ្ញាទាំងអស់មានចំនួន៩ ដែលក្នុងនោះមានសញ្ញាបូក មានចំនួន ៨ និង សញ្ញាដកមានចំនួន1។ ដូច្នេះ អថេរចៃដន្យទ្វេធា X ដែលតាងឱ្យចំនួនសញ្ញាបូកមានបំណែងចែកទ្វេធា។ បំណែងចែកនេះមាន $n = 9$ និង $p = 0.5$។ តាមរូបមន្ត២.៤.២ តម្លៃPកំណត់ដោយ

$$\begin{aligned}\text{P-value} &= 2P(X \geq 8) = 2\left[P(8) + P(9)\right] \\ &= 2(0.018 + 0.002) = 0.040\end{aligned}$$

ការសម្រេចចិត្ត

ការសម្រេចចិត្តគឺត្រូវបដិសេធសម្មតិកម្មសូន្យព្រោះតម្លៃP តូចជាងកម្រិតសារៈសំខាន់។

ការសង្ខេបលទ្ធផល

មានភស្តុតាងគ្រប់គ្រាន់ក្នុងការសន្និដ្ឋានថា ឱសថនេះមានប្រសិទ្ធភាពធ្វើឱ្យប្រែប្រួលទម្ងន់របស់ស្ត្រី(ទម្ងន់ស្រកចុះ)។

ឧទាហរណ៍២.៤

ក្រុមហ៊ុនរថយន្តឈ្មួលមួយចង់ធ្វើការសិក្សាស្រាវជ្រាវមើលថាតើការប្រើសម្បកកង់ប្រភេទផ្សេងៗគ្នាមានឥទ្ធិលទៅលើការសន្សំសំចៃប្រេងដែរឬទេ។ រថយន្ត១៦គ្រឿងបំពាក់ដោយសម្បកកង់ប្រភេទទី១ត្រូវបានបើកបរសាកល្បង។ បន្ទាប់មកអ្នកបើកបរដដែលនិងរថយន្តដដែលប៉ុន្តែបំពាក់ដោយសម្បកកង់ប្រភេទទី២ត្រូវបានបើកបរសាកល្បងម្តងទៀត ដើម្បីធ្វើការប្រៀបធៀប។ ចម្ងាយរត់បានគិតជាគីឡូម៉ែត្រ ក្នុងការប្រើប្រេង១លីត្រ ត្រូវបានកត់ត្រា។

ដោយប្រើកម្រិតសារៈសំខាន់0.05តើគេអាចសន្និដ្ឋានថាសម្បកកង់ប្រភេទទី១ ធ្វើឱ្យមានការសន្សំប្រេងជាងសម្បកកង់ប្រភេទទី២ដែរឬទេ?

រថយន្ត	ប្រភេទទី១	ប្រភេទទី២
1	4.2	4.1
2	4.7	4.9
3	6.6	6.2
4	7.0	6.9
5	6.7	6.8
6	4.5	4.4
7	5.7	5.7
8	6.0	5.8
9	7.4	6.9
10	4.9	4.9
11	6.1	6.0
12	5.2	4.9
13	5.7	5.3

14	6.9	6.5
15	6.8	7.1
16	4.9	4.8

ដំណោះស្រាយ

តាង $\tilde{\mu}_1$ មេដ្យានសម្រាប់ប្រភេទទី១ $\tilde{\mu}_2$ មេដ្យានសម្រាប់ប្រភេទទី២។ បើសិនជាការប្រើសម្បកកង់ប្រភេទទី១ជួយសន្សំសំចៃប្រេងជាងការប្រើសម្បកកង់ប្រភេទទី២នោះ $\tilde{\mu}_1 - \tilde{\mu}_2 > 0$ ។

សម្មតិកម្ម

H$_0$: $\tilde{\mu}_1 - \tilde{\mu}_2 = 0$

H$_1$: $\tilde{\mu}_1 - \tilde{\mu}_2 > 0$

ការគណនា

គណនាផលដក(តម្លៃក្នុងគំរូតាងទី១ - តម្លៃក្នុងគំរូតាងទី២)នោះយើងបានសញ្ញាបូកចំនួន11និងសញ្ញាដកចំនួន3។ ដូច្នេះ $x = 11$ ជាតម្លៃមួយនៃអថេរចៃដន្យ X ដែលមានបំណែងចែកទ្វេធាមានប៉ារ៉ាម៉ែត្រ $n = 14$ និង $p = 0.5$ ។ តាមរូបមន្ត២.៣, តម្លៃPគណនាតាម

$$\text{P-value} = P\left(X \geq 11\right)$$

ដោយប្រើការប៉ាន់ស្មានណរម៉ាល់ យើងបាន

$$\text{P-value} = P\left(X \geq 11\right) \approx P\left(Z > \frac{10.5 - 0.5\times 14}{\sqrt{14\times 0.5\left(1-0.5\right)}}\right)$$

$$= P\left(Z > 1.87\right) = 1 - 0.9693 = 0.0307$$

ការសម្រេចចិត្ត

ការសម្រេចចិត្តគឺបដិសេធសម្មតិកម្មសូន្យព្រោះតម្លៃPតូចជាងកម្រិតសារៈសំខាន់0.05។

ការសង្ខេបលទ្ធផល

គេមានភស្តុតាងគ្រប់គ្រាន់ក្នុងការសន្និដ្ឋានថា ការប្រើសម្បកកង់ប្រភេទទី១នាំឱ្យមានការសន្សំសំចៃប្រេងបានច្រើនជាងការប្រើសម្បកកង់ប្រភេទទី២។

លំហាត់អនុវត្តន៍

១. ទិន្នន័យខាងក្រោមនេះគឺជារយៈពេលគិតជានាទីដែលអ្នកជំងឺម្នាក់ត្រូវរង់ចាំមុននឹងបានជួបពិគ្រោះជាមួយគ្រូពេទ្យមួយលើកៗក្នុងចំនួន១២លើក។ ដោយប្រើតេស្តសញ្ញាជាមួយកម្រិតសារៈសំខាន់$\alpha=0.05$ ចូរធ្វើតេស្តអំណះអំណាងដែលលើកឡើងថាមេដ្យាននៃរយៈរង់ចាំមានមិនលើសពី20នាទី។

17	15	20	20	32	28
12	26	25	25	35	24

២. អ្នកត្រួតពិនិត្យគុណភាពចំណីអាហារម្នាក់ ពិនិត្យមើលដំណាប់ផ្លែឈើចំនួន១៦ក្រឡ ដើម្បីកំណត់នូវភាគរយនៃសារជាតុផ្សេងៗដែលនៅលាយឡំក្នុងដំណាប់។ ទិន្នន័យត្រូវបានកត់ត្រាដូចបង្ហាញខាងក្រោម។ ដោយប្រើប្រាស់ការប៉ាន់ស្មានណរម៉ាល់ចូរធ្វើតេស្តសញ្ញាជាមួយកម្រិតសារៈសំខាន់$\alpha=0.05$ លើសម្មតិកម្មសូន្យដែលថ្លែងថាមេដ្យាននៃភាគរយនៃសារជាតុផ្សេងៗដែលនៅលាយឡំក្នុង

ដំណាប់គឺ2.5%។

2.4	2.3	3.1	2.2	2.3	1.2	1.0	2.4
1.7	1.1	4.2	1.9	1.7	3.6	1.6	2.3

តេស្តសញ្ញាដែលជាវិធីមិនប៉ារ៉ាម៉ែត្រមិនត្រឹមតែអាចអនុវត្តបានលើទិន្នន័យបរិមាណវិស័យប៉ុណ្ណោះទេ វាក៏អាចត្រូវបានអនុវត្តលើ *ទិន្នន័យចម្លាម*(dichotomous data) ដែលនៅក្នុងនោះតម្លៃមួយប្រភេទតាងដោយសញ្ញាបូក និងមួយប្រភេទទៀតតាងដោយសញ្ញាដក ឧទាហរណ៍ «ត្រូវ» តាងដោយសញ្ញាបូក និង «ខុស»តាងដោយសញ្ញាដក។