Course: Applied Business Statistics

Week 2

Chapter 1. Numerical Descriptive Measures (2)

Lecturer: Udam Prang, PhD, MEd

មុខវិជ្ជា៖ ស្ថិតិធុរកិច្ចអនុវត្តន៍

សប្តាហ៍ទី២

មេរៀនទី១. ទន្វាសព័ណនាចំនួន (២)

គ្រូបង្រៀន៖ បណ្ឌិត ប្រាំង ឧត្តម

មាតិកា

៤. រង្វាស់ទីតាំងធៀប (Measures of Relative Locations)

៥. រង្វាស់ទំនាក់ទំនង (Measures of Association)

៤. រង្វាស់ទីតាំងធៀប

ពិន្ទុស្តង់ដា (Standard Scores) | ទ្រឹស្តីបទ Chebyshev | វិធានតាមការសម្គាល់ (Empirical Rule)

រង្វាស់ទីតាំងធៀប

រង្វាស់ទីតាំងធៀប អនុញ្ញាតឱ្យយើងដឹងថា តម្លៃជាក់លាក់ណាមួយ មានទីតាំងឆ្ងាយកម្រិតណាពីមធ្យម។ រង្វាស់នោះ គឺ "ពិន្ទុស្តង់ដា" ដែលជាទូទៅត្រូវបានគេហៅកាត់ថា ពិន្ទុ Z (ជាភាសាអង់គ្លេស៖ z-score)។ ពិន្ទុស្តង់ដា គ្មានឯកតានោះទេ។

ទន្ទឹមនេះ មានមធ្យោបាយពីរយ៉ាងផ្សេងទៀត៖

- តាមទ្រឹស្តីបទរបស់គណិតវិទូ ជនជាតិរុស្ស៊ី ឈ្មោះ Pafnuty Chebyshev (ហៅកាត់ថា ទ្រឹស្តីបទ Chebyshev)
- តាមវិធានតាមការសម្គាល់

ពិន្ទុស្តង់ដា

ពិន្ទុស្តង់ដា (z-score) សម្រាប់តម្លៃ x ណាមួយ មានរូបមន្តដូចខាងក្រោម៖

សម្រាប់សំណុំទិន្នន័យគំរូ៖

$$z = \frac{x - \bar{x}}{s}$$

សម្រាប់សំណុំទិន្នន័យសាកល៖

$$z = \frac{x - \mu}{\sigma}$$

ពិន្ទុស្តង់ដា

ឧទាហរណ៍. និស្សិតម្នាក់ទទួលបានពិន្ទុ ៩០ សម្រាប់មុខវិជ្ជាស្ថិតិ។ ដោយដឹងថា ពិន្ទុមធ្យម ស្មើនឹង ៧៥ ពិន្ទុ ហើយ គម្លាតស្តង់ដាស្មើនឹង ៥ ពិន្ទុ តើពិន្ទុរបស់និស្សិតម្នាក់នោះស្ថិតនៅឆ្ងាយកម្រិតណាពីពិន្ទុមធ្យម?

ពិន្ទុស្តង់ដា

ដំណោះស្រាយ

ដើម្បីដឹងថាតម្លៃនេះ មានចម្ងាយកម្រិតណាពីមធ្យម យើងត្រូវបំប្លែងពិន្ទុ ៩០ នេះ ទៅជាពិន្ទុស្តង់ដា។ យើងបាន៖

$$z = \frac{x - \mu}{\sigma} = \frac{៩០ - ៧៥}{៥} = ៣$$

ដូច្នេះ ពិន្ទុស្តង់ដា ស្មើនឹង ៣។ នេះមានន័យថា៖

ពិន្ទុ ៩០ របស់និស្សិតម្នាក់នោះ ធំជាងមធ្យម ចំនួន ៣ គម្លាតស្តង់ដា។

ទ្រឹស្តីបទ Chebyshev

ទ្រឹស្តីបទ Chebyshev កំណត់ថា៖

សម្រាប់សំណុំទិន្នន័យណាក៏ដោយ,
យ៉ាងហោចណាស់ $(1 - 1/z^2)$ នៃសំណុំទិន្នន័យ
គឺស្ថិតនៅក្នុងរវាង z គម្លាតស្តង់ដា ពីមធ្យម
ដែលក្នុងនោះ z ជាពិន្ទុស្តង់ដាដែលមានតម្លៃធំជាងមួយ។

ទ្រឹស្ដីបទ Chebyshev

ឧទាហរណ៍. និស្សិត ១០០ នាក់ បានប្រឡងឆមាសមុខវិជ្ជាស្ថិតិ។ លទ្ធផលបង្ហាញថា និស្សិតទាំងអស់ មានពិន្ទុ ជាមធ្យម ៧០ ពិន្ទុ ដោយមានគម្លាតស្តង់ដាស្មើនឹង ៥ ពិន្ទុ។ តើនិស្សិតប៉ុន្មានភាគរយ មានពិន្ទុចន្លោះពី ៦០ ពិន្ទុ ទៅ ៨០ ពិន្ទុ?

ទ្រឹស្តីបទ Chebyshev

ដំណោះស្រាយ

ដំបូង យើងត្រូវស្វែងរកពិន្ទុស្តង់ដា (ពិន្ទុ z) សម្រាប់តម្លៃ ៦០ ពិន្ទុ និង ៨០ ពិន្ទុ។

តាមការគណនា, ពិន្ទុស្តង់ដា សម្រាប់តម្លៃ ៦០ ពិន្ទុ ស្មើនឹង –២ ហើយ ពិន្ទុស្តង់ដា សម្រាប់តម្លៃ ៨០ ពិន្ទុ ស្មើនឹង ២។

យើងបាន តម្លៃ ៦០ ពិន្ទុ ទាបជាងពិន្ទុមធ្យម ចំនួន ២ គម្លាតស្តង់ដា ហើយ តម្លៃ ៨០ ពិន្ទុ ខ្ពស់ជាងពិន្ទុមធ្យម ចំនួន ២ គម្លាតស្តង់ដា។

ទ្រឹស្តីបទ Chebyshev

បន្ទាប់មកទៀត យើងយកទ្រឹស្តីបទ Chebyshev មកប្រើប្រាស់ជាយោង។

តាមទ្រឹស្តីបទ Chebyshev, z ជាតម្លៃដែលធំជាងមួយ។ យើងបាន z មានតម្លៃស្មើនឹង ២ (មិនមែន –២ នោះទេ)។

ដោយប្រើប្រាស់ទ្រឹស្តីបទ Chebyshev, យើងបាន៖

- *យ៉ាងហោចណាស់ (១ – ១/២² ឬ 0,៧៥ ឬ ៧៥%) នៃនិស្សិត មានពិន្ទុស្ថិតនៅក្នុងរវាង ២ គម្លាតស្តង់ដាពីមធ្យម ឬ ស្ថិតនៅចន្លោះ ៦0 ពិន្ទុ និង ៨0 ពិន្ទុ។*

វិធានតាមការសម្គាល់

វិធានតាមការសម្គាល់ គឺអាចអនុវត្តបានចំពោះតែសំណុំទិន្នន័យដែលមានបំណែងចែកធម្មតា (Normal Distribution) ឬបំណែងចែករាងជាជួង (Bell-shaped Distribution) ប៉ុណ្ណោះ។

- យើងនឹងសិក្សាលម្អិតបន្ថែមពីបំណែងចែកធម្មតានៅមេរៀនបន្ទាប់។

ជាគោលការណ៍ ដើម្បីដឹងថាសំណុំទិន្នន័យណាមួយមានរាងជាជួង យើងចាំបាច់ត្រូវសង់អ៊ីស្តូក្រាម (Histogram) សម្រាប់សំណុំទិន្នន័យនោះ។

វិធានតាមការសម្គាល់

វិធានតាមការសម្គាល់កំណត់ថា៖

- ប្រហែល ៦៨% នៃសំណុំទិន្នន័យ នឹងស្ថិតនៅក្នុងរវាង ១ គម្លាតស្តង់ដាពីមធ្យម
- ប្រហែល ៩៥% នៃសំណុំទិន្នន័យ នឹងស្ថិតនៅក្នុងរវាង ២ គម្លាតស្តង់ដាពីមធ្យម
- ស្ទើរតែ ១00% នៃសំណុំទិន្នន័យ នឹងស្ថិតនៅក្នុងរវាង ៣ គម្លាតស្តង់ដាពីមធ្យម

វិធានតាមការសង្កេត

ឧទាហរណ៍. ក្រុមហ៊ុនមួយមានបុគ្គលិក ២៨០ នាក់។ របាយការណ៍វាយតម្លៃសមិទ្ធកម្មការងាររបស់បុគ្គលិកប្រចាំឆ្នាំ បានបង្ហាញថា ពិន្ទុមធ្យមរបស់បុគ្គលិក ស្មើនឹង ៧៤ ពិន្ទុ និងគម្លាតស្តង់ដាស្មើនឹង ៨ ពិន្ទុ។ សន្មតថា សំណុំទិន្នន័យមានបំណែងចែករាងជាជួង តើមានបុគ្គលិកប៉ុន្មាននាក់ដែលមានពិន្ទុស្ថិតនៅចន្លោះ ៥៨ និង ៩០ ពិន្ទុ?

វិធានតាមការសង្កេត

ដំណោះស្រាយ

ដំបូង យើងត្រូវបំប្លែង តម្លៃ ៥៨ ពិន្ទុ និង តម្លៃ ៩០ ពិន្ទុ ទៅពិន្ទុស្តង់ដា (z-score)។

តាមការគណនា យើងបាន ពិន្ទុស្តង់ដា សម្រាប់ តម្លៃ ៥៨ ពិន្ទុ ស្មើនឹង –២ ហើយពិន្ទុស្តង់ដាសម្រាប់តម្លៃ ៩០ ពិន្ទុ ស្មើនឹង ២។

តាមលទ្ធផលនេះ យើងអាចកំណត់បានថា៖

- ពិន្ទុ ៥៨ ទាបជាងមធ្យមចំនួន ២ គម្លាតស្តង់ដា
- ពិន្ទុ ៩០ ខ្ពស់ជាងមធ្យមចំនួន ២ គម្លាតស្តង់ដា

វិធានតាមការសម្គាល់

ដោយសារសំណុំទិន្នន័យមានរាងជាជួង និង តាមវិធានតាមការសម្គាល់ យើងបាន៖

- តម្លៃសមមាត្រនៃបុគ្គលិកដែលទទួលបានពិន្ទុដែលស្ថិតនៅចន្លោះ ៥៨ ពិន្ទុ និង ៩០ ពិន្ទុ (ឬដែលស្ថិតនៅរវាង ២ គម្លាតស្តង់ដាពីមធ្យម) គឺ ៩៥%។

ដោយដឹងថា ចំនួនបុគ្គលិកទាំងអស់ ស្មើនឹង ២៨០ នាក់ យើងបាន៖

- ចំនួនបុគ្គលិកដែលទទួលបានពិន្ទុដែលស្ថិតនៅចន្លោះ ៥៨ ពិន្ទុ និង ៩០ ពិន្ទុ គឺ ២៦៦ នាក់ (២៨០ × ៩៥%)។

៥. រង្វាស់ទំនាក់ទំនង

កូវ៉ារ្យ៉ង់ (Covariance) | មេគុណទំនាក់ទំនង (Correlation Coefficient)

រង្វាស់ទំនាក់ទំនង

រង្វាស់ទំនាក់ទំនងរវាងអថេរពីរ មានច្រើន (Khamis, 2008)។

ប៉ុន្តែយើងនឹងផ្តោតទៅលើរង្វាស់ចំនួនពីរប៉ុណ្ណោះ តាមទ្រឹស្តីបទរបស់លោក Karl Pearson។

- ទីមួយ គឺ កូវ៉ារ្យ៉ង់ ដែលបង្ហាញពីទិសដៅ នៃទំនាក់ទំនងរវាងអថេរពីរ។
- ទីពីរ គឺ មេគុណទំនាក់ទំនង ដែលបង្ហាញពីទិសដៅ និងកម្លាំង នៃទំនាក់ទំនងរវាងអថេរពីរ។ មេគុណទំនាក់ទំនង មិនមានឯកតានោះទេ។

ខ្វាស់ទំនាក់ទំនង

យើងអាចប្រើប្រាស់រង្វាស់ទាំងពីរបាន លុះត្រាតែ៖

- ទំនាក់ទំនងរវាងអថេរទាំងពីរ មានលក្ខណៈស្របតាមបន្ទាត់ត្រង់ (ដែលអាចកត់សម្គាល់បានតាមរយៈក្រាហ្វិក)
- សំណុំទិន្នន័យនៃអថេរទាំងពីរ ជាសំណុំទិន្នន័យអនុបាត (Ratio Data)
- បំណែងចែកទិន្នន័យរបស់អថេរទាំងពីរ គឺជា បំណែងចែកធម្មតា ឬបំណែងចែកមានរាងជាជួង

កូវ៉ារ្យ៉ង់

សម្រាប់សំណុំទិន្នន័យគំរូ, កូវ៉ារ្យ៉ង់រវាងអថេរ x និង អថេរ y តាងដោយ s_{xy} និងមានរូបមន្តដូចខាងក្រោម៖

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

សម្រាប់សំណុំទិន្នន័យសាកល, កូវ៉ារ្យ៉ង់រវាងអថេរ x និង អថេរ y តាងដោយ σ_{xy} និងមានរូបមន្តដូចខាងក្រោម៖

$$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$

កូវ៉ារ្យ៉ង់

លទ្ធផលនៃការគណនាមានបីយ៉ាង៖

- ប្រសិនបើកូវ៉ារ្យ៉ង់មានតម្លៃអវិជ្ជមាន អថេរទាំងពីរមានទំនាក់ទំនងបញ្ច្រាស់ទិសគ្នា។
- ប្រសិនបើកូវ៉ារ្យ៉ង់មានតម្លៃវិជ្ជមាន អថេរទាំងពីរមានទំនាក់ទំនងស្របទិសគ្នា។
- ប្រសិនបើកូវ៉ារ្យ៉ង់មានតម្លៃស្មើនឹងសូន្យ អថេរទាំងពីរមិនមានទំនាក់ទំនងនោះទេ។

មេគុណទំនាក់ទំនង

សម្រាប់សំណុំទិន្នន័យគំរូ, មេគុណទំនាក់ទំនងរវាងអថេរ x និង អថេរ y តាងដោយ r_{xy} និងមានរូបមន្តដូចខាងក្រោម៖

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

សម្រាប់សំណុំទិន្នន័យសាកល, មេគុណទំនាក់ទំនងរវាងអថេរ x និង អថេរ y តាងដោយ ρ_{xy} និងមានរូបមន្តដូចខាងក្រោម៖

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

មេគុណទំនាក់ទំនង

តម្លៃនៃមេគុណទំនាក់ទំនង គឺស្ថិតនៅចន្លោះ –១ និង ១។

- ប្រសិនបើតម្លៃមេគុណស្មើនឹង –១ នោះមានន័យថា អថេរទាំងពីរ មានទំនាក់ទំនងបញ្ច្រាស់ទិស ហើយកម្លាំងនៃទំនាក់ទំនង មានកម្រិតខ្លាំងបំផុត។
- ប្រសិនបើតម្លៃមេគុណស្មើនឹង ១ នោះមានន័យថា អថេរទាំងពីរ មានទំនាក់ទំនងស្របទិស ហើយកម្លាំងនៃទំនាក់ទំនង មានកម្រិតខ្លាំងបំផុត។
- ប្រសិនបើតម្លៃមេគុណស្មើនឹង 0 នោះមានន័យថា អថេរទាំងពីរ គ្មានទំនាក់ទំនងនោះទេ។

ឧទាហរណ៍

តារាងនេះបង្ហាញពីចំនួនឆ្នាំដែលបុគ្គល ៨ នាក់បានចំណាយនៅក្នុងសាលា និងចំណូលប្រចាំខែដែលពួកគាត់ទទួលបានពីការបំពេញការងារ។ តើទំនាក់ទំនងរវាងចំនួនឆ្នាំនៅក្នុងសាលា និងចំណូលប្រចាំខែ មានលក្ខណៈដូចម្ដេចដែរ?

	ចំនួនឆ្នាំនៅក្នុងសាលា	ចំណូលប្រចាំខែ (ជាដុល្លារ)
បុគ្គលទី១	៩	៣០០
បុគ្គលទី២	១២	៣០០
បុគ្គលទី៣	១៦	៥០០
បុគ្គលទី៤	១៦	៧០០
បុគ្គលទី៥	១២	២០០
បុគ្គលទី៦	៩	៣០០
បុគ្គលទី៧	១៨	១០០០
បុគ្គលទី៨	២០	៨០០

ឧទាហរណ៍

ដើម្បីដឹងថាអថេរទាំងពីរមានទំនាក់ទំនងយ៉ាងដូចម្តេច យើងអាចប្រើប្រាស់រូបមន្តកូវ៉ារ្យ៉ង់ និងរូបមន្តមេគុណទំនាក់ទំនង (សន្មតថា ទំនាក់ទំនងរវាងអថេរទាំងពីរ មានលក្ខណៈស្របតាមបន្ទាត់ត្រង់ ហើយអថេរទាំងពីរមានបំណែងចែករាងជាជួង)។

យើងតាង៖

- x ជាអថេរ "ចំនួនឆ្នាំដែលបុគ្គលទាំង ៨ បានចំណាយនៅក្នុងសាលា"
- y ជាអថេរ "ចំណូលដែលបុគ្គលទាំង ៨ ទទួលបានក្នុងមួយខែ"

ឧទាហរណ៍

ដើម្បីសម្រួលដល់ការគណនាកូវ៉ារ្យ៉ង់ យើងអាចសង់តារាងដូចខាងក្រោមជាជំនួយ

No.	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
១	៩	៣០០	-៥	-២១៣	១០៦៣
២	១២	៣០០	-២	-២១៣	៤២៥
៣	១៦	៥០០	២	-១៣	-២៥
៤	១៦	៧០០	២	១៨៨	៣៧៥
៥	១២	២០០	-២	-៣១៣	៦២៥
៦	៩	៣០០	-៥	-២១៣	១០៦៣
៧	១៨	១០០០	៤	៤៨៨	១៩៥០
៨	២០	៨០០	៦	២៨៨	១៧២៥
សរុប					៧២០០

ឧទាហរណ៍

តាមរូបមន្តកូវ៉ារ្យ៉ង់ យើងបាន៖

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{៧២០០}{៨-១} = ១០២៨,៥៧១ \text{ ឆ្នាំដុល្លារ}$$

កូវ៉ារ្យ៉ង់ មានតម្លៃវិជ្ជមាន។

នេះមានន័យថា ទំនាក់ទំនងរវាងអថេរទាំងពីរ ជាទំនាក់ទំនងស្របទិស។ ដើម្បីដឹងថា កម្លាំងនៃទំនាក់ទំនងរវាងអថេរទាំងពីរខ្លាំងឬខ្សោយកម្រិតណា យើងត្រូវគណនាមេគុណទំនាក់ទំនង។

ឧទាហរណ៍

ដើម្បីគណនាមេគុណទំនាក់ទំនង យើងត្រូវស្គាល់តម្លៃកូវ៉ារ្យ៉ង់ និងតម្លៃគម្លាតស្តង់ដារបស់អថេរទាំងពីរ។

យើងបានគណនាតម្លៃកូវ៉ារ្យ៉ង់រួចរាល់ហើយ ដូច្នេះកិច្ចការដែលនៅសល់ គឺ ការគណនាតម្លៃគម្លាតស្តង់ដារបស់អថេរទាំងពីរ។

ឧទាហរណ៍

ដើម្បីគណនាគម្លាតស្តង់ដារបស់អថេរទាំងពីរ យើងអាចសង់តារាងខាងក្រោមជាជំនួយ

No.	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
១	៩	៣០០	-៥	-២១៣	២៥	៤៥១៥៦
២	១២	៣០០	-២	-២១៣	៤	៤៥១៥៦
៣	១៦	៥០០	២	-១៣	៤	១៥៦
៤	១៦	៧០០	២	១៨៨	៤	៣៥១៥៦
៥	១២	២០០	-២	-៣១៣	៤	៩៧៦៥៦
៦	៩	៣០០	-៥	-២១៣	២៥	៤៥១៥៦
៧	១៨	១០០០	៤	៤៨៨	១៦	២៣៧៦៥៦
៨	២០	៨០០	៦	២៨៨	៣៦	៨២៦៥៦
សរុប					១១៨	៥៨៨៧៥០

ឧទាហរណ៍

តាមតារាងខាងលើ យើងបានគម្លាតស្តង់ដាសម្រាប់អថេរទាំងពីរដូចខាងក្រោម៖

$$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{១១៨}{៨-១}} = ៤,១១ \text{ ឆ្នាំ}$$

$$s_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{៥៨៨៧៥០}{៨-១}} = ២៩០ \text{ ដុល្លារ}$$

ឧទាហរណ៍

តាមរូបមន្តមេគុណទំនាក់ទំនង យើងបាន៖

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\text{១០២៨,៥៧១ ឆ្នាំដុល្លារ}}{\text{៤,១១ ឆ្នាំ} \times \text{២៩០ ដុល្លារ}} = \text{០,៨៦}$$

មេគុណទំនាក់ទំនង ស្មើនឹង ០,៨៦ ជាតម្លៃវិជ្ជមាន ហើយខិតទៅរក ១។

នេះមានន័យថា ទំនាក់ទំនងរវាងអថេរទាំងពីរ ជាទំនាក់ទំនងស្របទិស ហើយកម្លាំងនៃទំនាក់ទំនងមានកម្រិតខ្លាំងគួរសម។

ឯកសារយោង

Khamis, H. (2008). Measures of association: How to choose? *Journal of Diagnostic Medical Sonography*, *24*(3), 155-162. https://doi.org/10.1177/8756479308317006

បញ្ចប់មេរៀនត្រឹមនេះ!

នៅសប្តាហ៍បន្ទាប់ យើងនឹងចាប់ផ្តើមសិក្សាមេរៀនទី២ ស្តីពី មូលដ្ឋានគ្រឹះប្រូបាប។