Параметричні методи в бізнес-аналізі

Порівняння вибіркового середнього арифметичного з відомим середнім значенням генеральної сукупності

Умови застосування t-критерію: вибірка отримана із генеральної сукупності, яка має наближено нормальний розподіл з параметрами μ та σ.

Гіпотеза H0: $\mu = \mu_0$ - середнє значення μ генеральної сукупності, з якої отримано вибірку, дорівнює даному значенню μ_0 (відомому, наприклад, з попередніх досліджень)

Альтернатива: $\mu \neq \mu_0$ ***Рівень значущості:*** α

***Порядок застосування t-критерію*:**

1. Приймається передумова про нормальність, формуються гіпотези H0 та H1, задається рівень значущості α;
2. Отримують вибірку обсягу *n*;
3. Обчислюється вибіркове середнє арифметичне $\overline{x}_{виб}$ та виправлена вибіркова дисперсія s^2;
4. Визначається значення t-критерію за формулою

$$t_{набл} = \frac{\left|\overline{x}_{виб} - \mu_0\right|}{S / \sqrt{n}}$$

$v=n-1$ – ступені волі

Порядок застосування t-критерію:

5. За таблицями знаходиться $t_{табл}$ - критичне значення t-критерію при рівні значущості α та числі ступенів волі *ν=n-1*;
6. Робиться висновок: якщо $t_{набл} \geq t_{табл}$, то вибіркове середнє $\bar{x}_{виб}$ значущо відрізняється від μ_0 на рівні значущості α, й у цій ситуації s^2 відхиляється гіпотеза Н0, тобто вважається, що вибірку взято з іншої генеральної сукупності, для якої $\mu \neq \mu_0$;

Якщо $t_{набл} < t_{табл}$, то на заданому рівні рівниця не значуща й зберігається гіпотеза Н0.

Параметричний тест z-тест (якщо відомі середнє значення та дисперсія генеральної сукупності) :

Відповідно до однієї з основних теорем статистики – *центральної граничної теореми*, розподіл середніх значень вибірок, що вилучаються із однієї тієї ж сукупності при достатньо великому *n* відповідає нормальному розподілу. Середнє значення усіх вибіркових середніх буде дорівнювати середньому значенню сукупності (μ), а стандартне відхилення вибіркових середніх складе величину

$\sigma_X = \dfrac{s}{\sqrt{n}}$, S – стандартне відхилення вибіркової сукупності

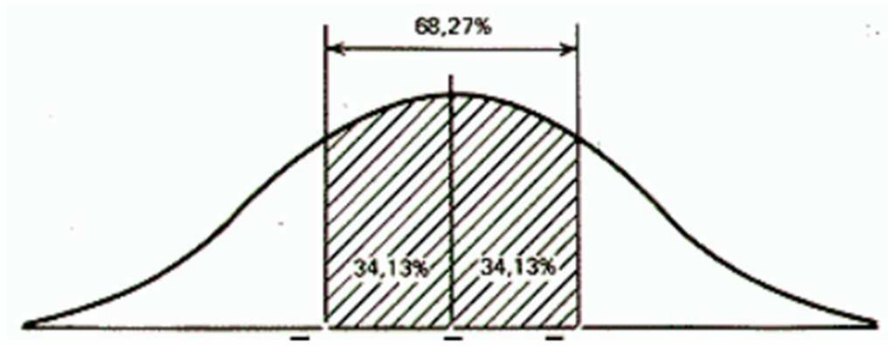
Параметричний тест z-тест:

Емпіричне значення z-критерію, показує, наскільки вибіркове середнє відрізняється від середнього генеральної сукупності у одиницях стандартного відхилення на визначається за формулою

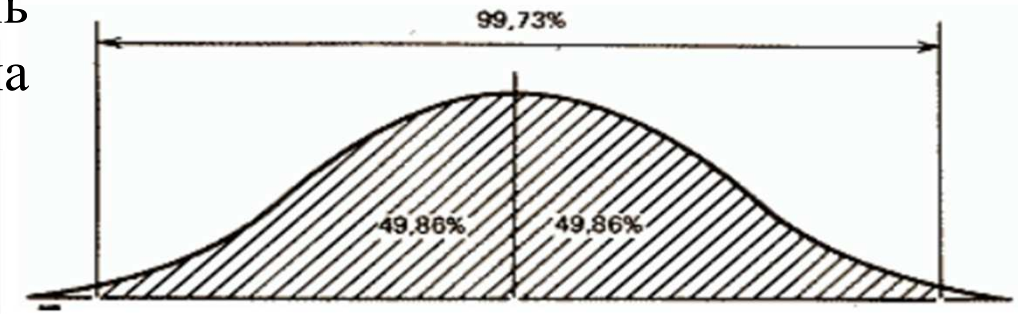
$$z = \frac{\bar{X} - \mu}{\sigma_X}$$

Параметричний тест z-тест:

При нормальному розподілі 68,27% результатів, розміщуються у межах одного стандартного відхилення по обидва боки від середнього значення, поза залежністю від величини стандартного відхилення.



У межах трьох стандартних відхилень уміщується майже уся генеральна сукупність – 99,73%.



Порівняння двох вибіркових середніх значень для зв'язанних вибірок

Умови застосування критерію: $d_i = x_i - y_i$ - різниця зв'язанних пар результатів вимірювання. Передумова про нормальний розподіл μ_d, σ_d

Гіпотеза H0: $\mu_d = 0$

Альтернатива: $\mu_d \neq 0$ ***(для двобічного критерію)***

$\mu_d > 0$ ***(однобічна альтернатива)***

Рівень значущості: α

Порядок застосування критерію:

1. Приймається передумова про нормальний розподіл різниць d_i, формуються гіпотези H0 та H1, обирається рівень значущості α;
2. Отримують дві вибірки (ряди зв'язанних пар спостережень) обсягу *n*;
3. Обчислюється середнє арифметичне $\bar{d}_{виб}$ та виправлена вибіркова дисперсія s_d^2;
4. Визначається значення t-критерію за формулою

$$t_{набл} = \frac{\bar{d}_{виб}}{S_d / \sqrt{n}}$$

Порядок застосування критерію:

5. За таблицями знаходиться $t_{табл}$ - критичне значення t-критерію при рівні значущості α та числі ступенів волі *ν=n-1*;
6. Робиться висновок: якщо $t_{набл} \geq t_{табл}$, то спостерігається відмінність на рівні значущості α.

Якщо $t_{набл} < t_{табл}$, то на заданому рівні значущості рівниця не значуща й зберігається гіпотеза Н0.

Порівняння двох вибіркових дисперсій із нормальних сукупностей

Умови застосування F-критерію: обидві вибірки незалежні й отримані з нормально розподілених генеральних сукупностей з параметрами $\mu_x, \sigma_x, \mu_y, \sigma_y$

Гіпотеза Н0: $\sigma_x^2 = \sigma_y^2$

Альтернатива: $\sigma_x^2 \neq \sigma_y^2$ ***(двобічний критерій)***

$\sigma_x^2 > \sigma_y^2$ ***(однобічна альтернатива)***

Рівень значущості: α

Порядок застосування F-критерію:

1. Приймається передумова про нормальний розподіл генеральних сукупностей, формуються гіпотези H0 та H1, обирається рівень значущості α;
2. Отримують дві незалежні вибірки з сукупностей X та Y обсягом n_x, n_y відповідно;
3. Обчислюється значення виправлених вибіркових дисперсій s_x^2, s_y^2. Більшу з цих дисперсій позначають s_1^2, а меншу - s_2^2;

***Порядок застосування критерію*:**

4. Визначається значення F-критерію за формулою

$$F_{набл} = \frac{S_1^2}{S_2^2}$$

5. Обчислене значення F-критерію порівнюється з критичним значенням $F_{крит}$ при заданому рівні значущості α та числі ступенів волі $v_1 = n_1 - 1, v_2 = n_2 - 1$;
6. Робиться висновок: якщо $F_{набл} \geq F_{крит}$, то дисперсії значно відрізняються при заданому рівні значущості α. У протилежному випадку гіпотеза про рівність двох дисперсій не відхиляється.