Лекція 3. Порядкові статистики та варіаційні ряди

Нехай $x_1, x_2, ..., x_n$ – вибірка, що отримана шляхом простого випадкового вибору S із генеральної сукупності G. Упорядкуємо її у зростаючому порядку:

$$x_{k_1} \le x_{k_2} \le ... \le x_{k_n} ,$$

де $k_i = \overline{1,n}, i = \overline{1,n}$. Позначимо $x_{(1)} = x_{k_1}, x_{(2)} = x_{k_2}, ..., x_{(n)} = x_{k_n}$. Тоді $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$. Випадкові величини $x_{(i)} = x_{k_i}$ називаються порядковими статистиками, а послідовність $x_{(1)}, x_{(2)}, ..., x_{(n)}$ – варіаційним рядом. Отже, порядкові статистики $x_{(i)}, i = \overline{1,n}$ є членами варіаційного ряду. Очевидно, що $x_{(1)}$ є мінімальною порядковою статистикою, а $x_{(n)}$ максимальною.

Порядкові статистики $x_{(1)}, x_{(2)}, ..., x_{(n)}$, на відміну від вибіркових значень, є залежними і мають різні функції розподілу. Легко бачити, що функція розподілу $F_k(x)$ k-ї порядкової статистики $x_{(k)}$ має такий вигляд:

$$F_k(u) = \sum_{i=k}^{n} C_n^i \left[F(u)\right]^i \left[1 - F(u)\right]^{n-i} ,$$

де $F(u) = p(x \le u)$. Випадкова величина x називається неперервною, якщо її функція розподілу є абсолютно неперервною для будь-якого $u \in R^1$. Використовуючи наведену формулу, можна довести, що для будь-якої неперервної випадкової величини x із генеральної сукупності G, ймовірність того, що вона потрапить в інтервал $J = \left(x_{(i)}, x_{(j)}\right)$ утворений i-ю і j-ю порядковими статистиками $(i < j)$, не залежить від функції розподілу $F(u)$ і дорівнює

$$p\left(x \in \left(x_{(i)}, x_{(j)}\right)\right) = \frac{j-i}{n+1} .$$

Дійсно, якщо випадкові величини ξ і η є незалежними, то

$$p(\xi < \eta) = \int_{-\infty}^{\infty} F_\xi(u)\, dF_\eta(u),$$

де $F_\xi(u)$ і $F_\eta(u)$ є функціями розподілу випадкових величин ξ і η відповідно.

Щільність розподілу *i*-ї порядкової статистики обчислюється так:

$$F_k(u) = \sum_{i=k}^{n} C_n^i \left[F(u)\right]^i \left[1 - F(u)\right]^{n-i} = \sum_{i=k}^{n} G_i(u),$$

$$F_k'(u) = \sum_{i=k}^{n} G'(u)_i\ ,$$

$$G_k'(u) = \Big[C_n^k \Big(\big(F(u)\big)^k \big(1 - F(u)\big)^{n-k} -$$

$$- \big(F(u)\big)^k (n-k) \big(1 - F(u)\big)^{n-k-1} \Big) f(u) \Big]$$

$$G_{k+1}'(u) = \Big[C_n^{k+1} \big(F(u)\big)^{k+1} \big(1 - F(u)\big)^{n-k-1} \Big]' =$$

$$= C_n^{k+1} \Big[\Big((k+1) \big(F(u)\big)^k \big(1 - F(u)\big)^{n-k-1} -$$

$$- \big(F(u)\big)^{k+1} \big(1 - F(u)\big)^{n-k-2} (n-k-1) \Big) f(u) \Big]$$

Другий доданок в попередньому виразі скорочується із першим доданком у наступному виразі.

$$-C_n^k (n-k) + C_n^{k+1} (k+1) =$$

$$= -\frac{n!(n-k)}{(n-k)!k!} + \frac{n!(k+1)}{(n-k-1)!(k+1)!} =$$

$$= -\frac{n!}{(n-k-1)!k!} + \frac{n!}{(n-k-1)!k!} = 0.$$

Зважимо на те, що

$$kC_n^k = \frac{n!k}{(n-k)k!} = \frac{(n-1)!nk}{(n-k-1)!(n-k)(k-1)!k} =$$

$$= \frac{(n-1)!}{(n-k)!(k-1)!} = nC_{n-1}^{k-1} = \frac{1}{B(k,n-k+1)}.$$

Останній член останнього доданка дорівнює нулю.

$$\left(F(u)\right)^n \left(1-F(u)\right)^0 (n-n) f(u) = 0 .$$

Отже,

$$f_k(u) = nC_{n-1}^{k-1}\left[F(u)\right]^{k-1}\left[1-F(u)\right]^{n-k} f(u) .$$

Обчислимо ймовірності $p\left(x < x^{(i)}\right)$ і $p\left(x < x^{(j)}\right)$. Використовуючи наведені вище формули, отримаємо

$$p\left(x < x_{(i)}\right) = \int_{-\infty}^{\infty} F(u)\,dF_i(u) = nC_{n-1}^{i-1}\int_{-\infty}^{\infty}\left[F(u)\right]^i\left[1-F(u)\right]^{n-i} dF(u) =$$

$$= nC_{n-1}^{i-1}\int_0^1 v^i (1-v)^{n-i}\,dv.$$

Як відомо,

$$\int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q-1)} = \frac{(p-1)!(q-1)!}{(p+q-1)!} .$$

Отже,

$$\int_0^1 x^{i+1-1}(1-x)^{n-i+1-1}\,dx = B(i+1,n-i+1) =$$

$$= \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(i+1+n+1-i)} = \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(n+2)} = \frac{i!(n-i)!}{(n+1)!}$$

$$p\left(x < x_{(i)}\right) = nC_{n-1}^{i-1}\frac{i!(n-i)!}{(n+1)!} =$$

$$= n\frac{(n-1)!\,i!\,(n-i)!}{(n-1-i+1)!\,(i-1)!\,(n+1)!} =$$

$$= \frac{n(n-1)!\,(i-1)!\,i!\,(n-i)!}{(n-i)!\,(i-1)!\,n(n+1)(n-1)!} = \frac{i}{n+1}.$$

.

Звідси випливає, що

$$p\left(x \in \left(x_{(i)}, x_{(j)}\right)\right) = p\left(x_{n+1} < x_j\right) - p\left(x_{n+1} < x_i\right) =$$

$$= \frac{j}{n+1} - \frac{i}{n+1} = \frac{j-i}{n+1}.$$

.

Таким чином, якщо випадкова величина x не залежить від вибірки $x_1, x_2, ..., x_n$, отриманої простим випадковим вибором S із генеральної сукупності G із абсолютно неперервною функцією розподілу $F(u)$, то

$$p\left(x \in \left(x_{(1)}, x_{(n)}\right)\right) = \frac{n-1}{n+1}.$$

Цей факт дозволяє побудувати довірчий інтервал для основної розподіленої маси генеральної сукупності із точно визначеним довірчим рівнем. Дійсно, оскільки довірчий рівень довірчого інтервалу $J = \left(x_{(1)}, x_{(n)}\right)$ дорівнює $\frac{n-1}{n+1}$, то для $n \geq 39$ рівень значущості цього інтервалу не перевищує 0,05.