# МЕТОД ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ДЛЯ ПАРНОГО СРАВНЕНИЯ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ И ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

*Цейтлин Н.А.*

Фирма «CuBe Matrix», ФРГ: tseitlin@gmx.net

Метод доверительных интервалов (МДИ) был предложен [1, 2] для проверки гипотез о равенстве пар однотипных параметров нормально распределённых случайных величин (СВ) - математических ожиданий (МО) и среднеквадратичных отклонений (СО).

**Цель работы** - воспользоваться МДИ также и для проверки гипотез о согласии пар непрерывных генеральных функции распределений (ГФР) СВ.

**Даны** m независимых выборок $\{y_{ji}\}_{i=1}^{N_j}$ значений $y_{ji}$ СВ $Y_j$ (j = 1, 2, …, m) равных объёмов - по $N_j = N$ элементов – каждая, распределенных нормально, $Y_j \sim N(\nu_{Yj}, \sigma_{Yj})$, имеющих непрерывные ГФР $F_i(y) = P(Y_i \le y)$, где: $\nu_{Yj}$ – МО; $\sigma_{Yj}$ – СО.

**Необходимо** с помощью МДИ [1, 2] проверить $U = C_m^2 = m(m-1)/2$ нулевых гипотез $H_{0nq}$ о равенстве пар **параметров** $\nu_{Yn}$ и $\nu_{Yq}$ или $\sigma_{Yn}$ и $\sigma_{Yq}$, или - о согласии пар ГФР $F_n(y)$ и $F_q(y)$:

$$H_{0nq}: u_n(y) = u_q(y), \ (1 \le n < q \le U) \qquad (1)$$

против $U$ альтернативных гипотез $H_{1nq}$: $u_n(y) \ne u_q(y)$, где $u_n(y)$, $u_q(y)$ – **параметры** или ГФР.

МДИ заключается в сравнении на графике двусторонних ДИ, построенных в виде пар плеч $\delta_{jP/2}$ (отрезков прямых с двух сторон от точечных оценок параметров) с такой доверительной вероятностью (1-P), чтобы на априори заданном [2, 5] уровне значимости $\alpha_к$ читатель мог визуально определить значимость различия сравниваемых пар параметров. Если ДИ перекрываются, то соответствующая гипотеза $H_{0nq}$ не отклоняется, если же - не перекрываются, то - отклоняется в пользу $H_{1nq}$.

Критическим событием является равенство суммы плеч $\delta_{nP/2} + \delta_{qP/2}$ и критического расстояния $\Delta_\lambda$ между сравниваемыми оценками параметров, т. е.

$$\delta_{nP/2} + \delta_{qP/2} = \Delta_\lambda; \ (1 \le n < q \le U), \qquad (2)$$

где $\lambda = \lambda_к$ – критическая вероятность, и $\lambda_к = \lambda(\alpha_к)$. Для проверки гипотез (1) относительно МО в виде $H_{0nq}$: $\nu_{Yn} = \nu_{Yq}$ при равных СО $(\sigma_{Yn} = \sigma_{Yq} = \sigma_Y)$ Р. Фишер, основываясь на теореме Бонферрони, предложил [3] использовать верхний $\lambda_к$-предел статистики Стьюдента $t_{\lambda;mf}$, где $\lambda = \lambda_к = \alpha_к/(2U)$; mf - число степеней свободы. Это предложение и формула (2) позволили [1] для построения 100(1-P)-процентных ДИ использовать соотношение

$$t_{P/2;mf} = 2^{-0,5} t_{\lambda;mf}; \ \lambda = \lambda_к = \alpha_к/(2U), \qquad (3)$$

где $t_{P/2,mf}$ - верхний P/2-предел распределения Стьюдента. При $N \to \infty$ (а практически - при $N \ge 25$) эта зависимость принимает вид [2, с. 225]:

$$Z_{P/2} = 2^{-0,5} Z_\lambda; \ \lambda = \lambda_к = \alpha_к/(2U), \qquad (4)$$

где $Z_{P/2}$ и $Z_\lambda$ - верхние P/2- и **λ**-пределы распределения Гаусса. Аналогичное соотношение для построения ДИ СО в МДИ было получено при проверке гипотез (1) относительно СО в виде $H_{0nq}$: $\sigma_{Yn} = \sigma_{Yq}$ [2]. Для проверки гипотез (1) в виде $H_{0nq}$: $F_n(y) = F_q(y)$ о согласии пар ГФР **предлагается** воспользоваться соотношением, аналогичным (4):

$$D_{P/2;N} = 2^{-0,5} D_{\lambda;N}; \ \lambda = \lambda_к = \alpha_к/(2U); \ \lambda < 0{,}2, \qquad (5)$$

где $D_{P/2;N}$ и $D_{\lambda;N}$ - верхние P/2- и λ-пределы распределения Колмогорова-Смирнова. Подставив зависимость $D_{\lambda;N} = ((-0{,}5\ln(\lambda/2))/N)^{0,5}$, справедливую [4, с. 372] для непрерывной ГФР $F(y)$, в (5), получим

$$D_{P/2;N} = ((-\ln(\alpha_к/2m(m-1))/4N)^{0,5}; \ N > 70. \qquad (6)$$

Данные для **примера** (рис.) заимствованы из маркетингового исследования [5]; $N_1 = 78$; $N_2 = 120$; по ф-ле (6): $D_{P/2;78} = ((-\ln(0{,}05/2\times2(2-1))/4\times78)^{0,5} = 0{,}119$; $D_{P/2;120} = 0{,}0955$. Нижние $F_i^-$ и верхние $F_i^+$ границы ДИ построены путём сдвига ЭФР $F_i$ вниз и вверх на величины $D_{P/2;78} = 0{,}119$ и $D_{P/2;120} = 0{,}0955$ соответственно; медианы распределений (при $F = 0{,}5$) $y_{м1} = 175$ с и $y_{м2} = 308$ с различаются значимо. Однако различия времени пребывания в павильоне напитков менее 110 с и более 316 с между 15% и 30% мужчин и женщин соответственно, не значимы.

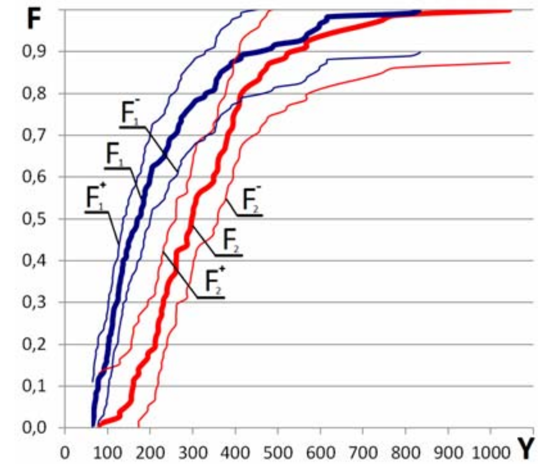


Рис. Сравнение распределений времени пребывания мужчин $F_1$ и женщин $F_2$ в павильоне напитков: Y – время, с; $F_i$ – эмпирические ФР (ЭФР); $F_i^-$ и $F_i^+$ – нижние и верхние границы соответствующих ДИ; m = 2; $\alpha_к = 0.05$.

## Литература

1. Цейтлин Н. А. Проверка гипотез методом доверительных интервалов. В кн. Методы математической статистики в основной химии. Труды, Т. 55, НИОХИМ, Харьков, 1981, с. 82-89.
2. Цейтлин Н. А. Из опыта аналитического статистика. - М.: Солар, 2007. - 906 с. www.cubematrix.com/oldsite/anlagen/as.pdf.
3. Fisher R. M. The design of experiments. - London: Oliver and Boyd, 1935.- 360 p.
4. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Справочное издание. Основы моделирования и первичная обработка данных. - М.: Финансы и статистика, 1983. - 471 с.
5. Горбач А. Н., Цейтлин Н. А. Покупательское поведение: анализ спонтанных последовательностей и регрессионных моделей в маркетинговых исследованиях. - Киев: Освіта України, 2011. - 298 с. http://www.cubematrix.com/oldsite/anlagen/asp.pdf.