3.3. P-value, или надёжность гипотезы

Величина, которая называется p-value (p-значение, p-уровень значимости, p-критерий), в последнее время стала очень широко применяться в статистических исследованиях. Однако определения этой величины весьма туманны и зачастую просто искажают её природу. В некоторых монографиях и учебниках это понятие также вводится и называется по-разному: «критический уровень» (Э. Леман), «фактически достигаемый уровень семейства критериев» (А. А. Боровков), «реально достигнутый уровень значимости критерия» (Н. И. Чернова). В этом разделе мы попытаемся объяснить, что такое p-value, и исследовать её свойства.

Начнём с очень простого примера. Хорошо известно, что для проверки простой гипотезы $H\colon \xi \sim \mathbf{N}(0,1)$ против сложной альтернативы $K\colon \xi \sim \mathbf{N}(\mu,1)$, $\mu > 0$ (мы рассматриваем одномерную выборку ξ), существует РНМК, критическое множество которого имеет вид $Z_\alpha = \{x > C_\alpha\}$, где C_α связана с ошибкой α первого рода (вероятностью ошибочно отклонить гипотезу) равенством

$$\alpha = P_0(\xi \in Z_\alpha) = P_0(\xi > C_\alpha) = \int_{C_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz. \tag{3.90}$$

Задавая приемлемый уровень ошибки $\alpha = \alpha_0$ и находя из уравнения (3.90) значение $C = C_{\alpha_0}$, мы получаем РНМК, который отклоняет гипотезу всякий раз, когда реализация x с.в. ξ больше C_{α_0}.

Если $x \leqslant C_{\alpha_0}$, то мы гипотезу принимаем. Однако, очевидно, имеется некоторая разница между реализацией $x \ll C_{\alpha_0}$ и реализацией $x \approx C_{\alpha_0}$, но которая тем не менее принадлежит множеству принятия гипотезы, $x \leqslant C_{\alpha_0}$. По всей видимости, в первом случае мы будем принимать гипотезу с большей уверенностью, чем во втором, потому что значение $x \ll C_{\alpha_0}$ находится «глубоко внутри» множества принятия гипотезы, а значение $x \approx C_{\alpha_0}$ близко к границе множества принятия гипотезы и критического множества. Достаточно немного увеличить величину α_0, и согласно нашему критерию реализация $\xi = x$ уже будет противоречить гипотезе.

Возникает идея количественно охарактеризовать эту степень уверенности в принятии гипотезы величиной

$$\alpha(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz.$$

В самом деле, для $x \ll C_{\alpha_0}$ эта величина достаточно большая (и тем больше, чем дальше x от C_{α_0}), а для $x \approx C_{\alpha_0}$ мы имеем $\alpha(x) \approx \alpha_0$, т. е. величина $\alpha(x)$ мала (больше, но почти равна ошибке критерия, которую, конечно, выбирают малой).

Рассмотрим всё семейство критических множеств, параметризованное величиной ошибки:

$$Z_\alpha = \{x > C_\alpha\}, \qquad \alpha \in [0, 1].$$

Нетрудно заметить, что

$$\alpha(x) = \inf_{Z_\alpha\colon x \in Z_\alpha} \alpha = \inf_{Z_\alpha\colon x \in Z_\alpha} \int_{Z_\alpha} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz. \tag{3.91}$$

Действительно, интеграл

$$\int_{Z_\alpha} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = \int_{C_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz$$

тем меньше, чем больше C_α. При этом мы требуем, чтобы реализация x лежала в множестве Z_α, т. е. $x > C_\alpha$. Таким образом, чтобы найти точную нижнюю грань в (3.91), нужно найти точную верхнюю грань значений C_α, которые удовлетворяют неравенству $C_\alpha < x$; очевидно, что эта точная верхняя грань равна x:

$$\inf_{Z_\alpha\,:\, x \in Z_\alpha} \int_{Z_\alpha} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = \\ = \sup_{C_\alpha\,:\, C_\alpha < x} \int_{C_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz.$$

Величина $\alpha(x)$, заданная уравнением (3.91), и равна p-value для данного критерия и данной реализации x. Часто говорят, что она равна минимальной вероятности ошибочно отклонить гипотезу по наблюдению $\xi = x$.

Можно также провести рассуждения «в обратном направлении», начав с величины p-value. Предположим, что по реализации x мы нашли $\alpha(x)$ по формуле (3.91). Теперь поставим вопрос: какова ошибка критерия, который обязательно отклоняет гипотезу по реализации x? Имеем $x \in Z_\alpha$, следовательно, $x > C_\alpha$, и тогда

$$\alpha = \int_{C_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz \geqslant \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = \alpha(x)$$

Таким образом, если $\alpha(x)$ велика, то ошибаться, отклоняя гипотезу по реализации x, мы будем с вероятностью $\alpha \geqslant \alpha(x)$, т. е. с очень большой вероятностью. Если, наоборот, $\alpha(x)$ мала, то, отклоняя гипотезу по реализации x, мы ошибаемся не с такой большой вероятностью.

Однако все эти рассуждения верны для фиксированной реализации x и тем самым для фиксированной величины $\alpha(x)$. Поэтому рассуждения о какой-либо вероятности ошибочного решения теряют смысл. В самом деле, для понимания, чему равна вероятность, нам нужно много раз применить свой критерий (и тогда доля случаев ошибочных решений будет примерно равна вероятности), но всякий раз мы будем иметь разные реализации x и, следовательно, разные значения $\alpha(x)$. Другими словами, p-value – это случайная величина, заданная как функция $\alpha(\xi)$ от случайной выборки, и любая аккуратная интерпретация p-value требует исследования её распределения.

P-value как случайная величина систематически и достаточно полно была исследована в работах Ю. П. Пытьева и его учеников. Он назвал эту случайную величину *надёжностью статистической гипотезы* (точнее, но длиннее было бы назвать эту величину надёжностью решения о принятии статистической гипотезы).

Вернёмся к рассмотренному примеру и рассмотрим с.в $\alpha(\xi)$, которая рассчитывается по формуле (3.92) для $\alpha(x)$, когда $\xi = x$. Найдём распределение надёжности в случаях, когда верна гипотеза и когда верна альтернатива. Для нашего примера

$$\alpha(\xi) = \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz.$$

Сначала заметим, что для любого распределения с.в. ξ и для любого $b \in (0,1)$ в силу строго монотонной зависимости интеграла от нижнего предела

$$P(\alpha(\xi) < b) = P(\xi > x_b), \quad \text{где} \quad \alpha(x_b) = b. \tag{3.92}$$

Перепишем последнее уравнение как

$$b = \int_{x_b}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = P_0(\xi > b). \tag{3.93}$$

Пусть верна гипотеза. Сравнивая уравнение (3.93) с первым равенством в (3.92), видим, что $P_0(\alpha(\xi) < b) = b$ для любого $b \in (0, 1)$. Таким образом, если верна гипотеза, то функция распределения и плотность вероятности надёжности задаются как

$$F_\alpha(b) = b, \quad p_\alpha(b) = F'_\alpha(b) = 1, \qquad 0 < b < 1.$$

Мы получили следующий результат: *при верной простой гипотезе надёжность распределена равномерно*, $\alpha(\xi) \sim \mathbf{U}[0, 1]$.

Пусть теперь верна альтернатива, $\xi \sim \mathbf{N}(\mu, 1)$, $\mu > 0$. Из (3.92) получаем

$$F_\alpha(b) = P(\alpha(\xi) < b) = P(\xi > x_b) = \int_{x_b}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(z-\mu)^2/2}\, dz,$$

$$p_\alpha(b) = \frac{dF_\alpha(b)}{db} = -\frac{1}{\sqrt{2\pi}}\, e^{-(x_b-\mu)^2/2} \cdot \frac{dx_b}{db}.$$

При этом уравнение (3.93) для x_b остаётся прежним, дифференцируя его по b, имеем

$$1 = \frac{d}{db} \int_{x_b}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz = -\frac{1}{\sqrt{2\pi}}\, e^{-x_b^2/2} \cdot \frac{dx_b}{db}.$$

Отсюда

$$\frac{dx_b}{db} = \left(-\frac{1}{\sqrt{2\pi}}\, e^{-x_b^2/2} \right)^{-1},$$

что после подстановки в формулу для $p_\alpha(b)$ даёт

$$p_\alpha(b) = \frac{e^{-(x_b-\mu)^2/2}}{e^{-x_b^2/2}} = e^{2x_b\mu} \cdot e^{-\mu^2/2}.$$

Из (3.93) следует, что если $b \to 1$, то $x_b \to -\infty$, если $b \to 0$, то $x_b \to +\infty$. Отсюда получаем предельное поведение плотности вероятности надёжности при верной альтернативе $\mu > 0$:

$$\lim_{b \to 1} p_\alpha(b) = \lim_{x_b \to -\infty} e^{2x_b\mu} \cdot e^{-\mu^2/2} = 0,$$

$$\lim_{b \to 0} p_\alpha(b) = \lim_{x_b \to +\infty} e^{2x_b\mu} \cdot e^{-\mu^2/2} = +\infty.$$

Таким образом, при *верной альтернативе надёжность с большой вероятностью близка к* 0 *и с малой вероятностью близка к* 1.

Полученные результаты показывают, что для рассмотренного примера интуитивные представления о p-value в целом верны. Если $\alpha(\xi)$ велика, то это свидетельствует в пользу верной гипотезы, но не потому, что при верной гипотезе распределение надёжности сосредоточено в области значений, близких к 1, а потому что при верной альтернативе такие значения маловероятны. Напротив, низкие значения надёжности свидетельствуют в пользу альтернативы, но опять же потому, что при верной альтернативе вероятность наблюдать низкие значения $\alpha(\xi)$ достаточно высока по сравнению с той же вероятностью при верной гипотезе.

Однако всё это верно лишь в рассмотренном случае. Для других задач и других критериев требуется отдельное исследование распределения надёжности.

Тем не менее можно доказать некоторое общее утверждение. Рассмотрим простую гипотезу $H\colon \xi \sim P_0(\,\cdot\,)$ и предположим, что задано семейство нерандомизированных критериев для её проверки, параметризованное величиной ошибки $\alpha \in [0,1]$. Обозначим критические множества для таких критериев как Z_α, $\alpha \in [0,1]$. Наложим на семейство критических множеств $\{Z_\alpha\}$ следующие естественные условия:

1) множество Z_α из этого семейства существует и единственно для любого $\alpha \in [0,1]$;

2) семейство монотонно по α: если $\alpha_1 \leqslant \alpha_2$, то $Z_{\alpha_1} \subset Z_{\alpha_2}$.

Зададим надёжность гипотезы как

$$\alpha(\xi) = \inf_{\alpha\colon \xi \in Z_\alpha} \alpha = \inf_{Z_\alpha\colon \xi \in Z_\alpha} P_0(\xi \in Z_\alpha). \tag{3.94}$$

Теорема 10. *Если при верной гипотезе функция распределения надёжности* $F_\alpha(b) = P_0(\alpha(\xi) < b)$, $0 \leqslant b \leqslant 1$, *непрерывна всюду на отрезке* $[0,1]$, *то* $\alpha(\xi) \sim \mathbf{U}[0,1]$.

Доказательство. Рассмотрим множество $\{\xi\colon \alpha(\xi) < b\}$ при фиксированном $b \in (0,1)$. Если $\alpha(\xi) < b$, то по определению точной нижней грани среди α, по которым вычисляется точная нижняя грань (3.94) (тех, для которых $\xi \in Z_\alpha$), найдётся значение α, такое что $\alpha(\xi) \leqslant \alpha < b$. Тогда $\xi \in Z_\alpha \subset Z_b$. Таким образом,

$$\{\xi\colon \alpha(\xi) < b\} \subset \{\xi \in Z_b\}.$$

Если $\alpha(\xi) > b$, то $\xi \notin Z_b$, потому что если $\xi \in Z_b$, то Z_b – одно из множеств, по которым вычисляется точная нижняя грань в (3.94), следовательно,

$$\alpha(\xi) \leqslant P_0(\xi \in Z_b) = b.$$

Получили очевидное противоречие с $\alpha(\xi) > b$. Таким образом,

$$\{\xi\colon \alpha(\xi) > b\} \subset \{\xi \notin Z_b\},$$

и, переходя к дополнениям, мы имеем

$$\{\xi \in Z_b\} \subset \overline{\{\xi\colon \alpha(\xi) > b\}} = \{\xi\colon \alpha(\xi) \leqslant b\},$$

Объединяя полученные включения, получаем

$$\{\xi\colon \alpha(\xi) < b\} \subset \{\xi \in Z_b\} \subset \{\xi\colon \alpha(\xi) \leqslant b\}.$$

Отсюда для любого распределения с.в. ξ

$$P(\alpha(\xi) < b) \leqslant P(Z_b) \leqslant P(\alpha(\xi) \leqslant b\}.$$

Если верна гипотеза, то $P_0(\alpha(\xi) \leqslant b\} = b$, и по условию теоремы функция распределения надёжности непрерывна, следовательно,

$$F_\alpha(b) = P(\alpha(\xi) < b) \leqslant P_0(Z_b) = b \leqslant \\ \leqslant P_0(\alpha(\xi) \leqslant b\} = F_\alpha(b+0) = F_\alpha(b).$$

Как результат, $F_\alpha(b) = b$ для всех $b \in (0,1)$. Теорема доказана.

Если задано семейство статистических критериев проверки простой гипотезы, параметризованное величиной $\alpha \in (0,1)$ ошибки первого рода, то мы можем ввести надёжность гипотезы следующим образом. Для каждой реализации выборки $x \in \mathsf{R}$ положим, как и выше,

$$\alpha(x) = \inf_{\alpha\colon x \in Z_\alpha} \alpha,$$

где Z_α – критическое множество статистического критерия с ошибкой α. Надёжность – это случайная величина $\alpha(\xi)$, равная $\alpha(x)$, когда $\xi = x$. В частности, в случае простой гипотезы $\xi \sim L_0(\,\cdot\,)$ и простой альтернативы $\xi \sim L_1(\,\cdot\,)$ существует НМК, и тогда надежность задаётся формулами

$$\alpha(\xi) = \inf_{\alpha\colon \xi \in Z_\alpha} \alpha,$$

$$Z_\alpha = \big\{x \in \mathsf{R}\colon L_1(x) > C_\alpha L_0(x)\big\}, \qquad C_\alpha\colon \int_{Z_\alpha} L_0(x)\,dx = \alpha.$$