Прикладная статистика. Лабораторная 1.

Алексей Григорьев, 674 гр. ФУПМ МФТИ

Март 2019

№2.3 *Введем оговорку, будем рассматривать не дискретную задачу, а ее непрерывный аналог: пусть время (дни) не дисретно, а непрерывно. Существенно на решение это, пожалуй, не повлияет, только на используемую модель.*

(a) Условия этой задачи схожи с постановкой классической задачи об ожидании автобуса на остановке. Здесь также могут быть введены допущения о случайности поломок электрички с некоторой фиксированной средней интенсивностью. Стандартным распределением, моделирующим подобного рода задачи, является экспоненциальное распредление $\mathrm{Exp}(\lambda)$. Это предположение согласуется с видом гистограммы предложенных данных.

(b) В предположении о том, что данные описываются экспоненциальным распределением, что эмпирически согласуется с имеющейся выборкой, задача может быть переформулирована в терминах параметра экспоненциального распредления λ (проверку гипотезы экспоненциальности распредления опустим). Учитываем, что $\mathrm{median}(\mathrm{Exp}(\lambda)) = \frac{\ln 2}{\lambda}$ и $\mathbb{D}(\mathrm{Exp}(\lambda)) = \lambda^{-2}$.

$$\begin{aligned} \text{нулевая гипотеза} \quad & H_0 : \lambda = 1/3 \\ \text{альтернатива} \quad & H_1 : \lambda < 1/3 \end{aligned}$$

(c) Данная постановка с учетом допущения экспоненциальности приводит к более удобной записи с использованием одновыборочного критерия знаков:

$$\begin{aligned} \text{нулевая гипотеза} \quad & H_0 : \mathrm{median}(X) = 3 \ln 2 \\ \text{альтернатива} \quad & H_1 : \mathrm{median}(X) > 3 \ln 2 \\ \text{статистика} \quad & T(X^n) = \sum_{i=1}^{n} [X_i > 3 \ln 2] \end{aligned}$$

(d) Для выборки $X^n = (3, 22, 13, 6, 18, 5, 6, 10, 7, 15)$ получим, что

$$T(X^n) = 10, \text{ p-val} = 0.5^{10} \approx 0.001$$

все значения больше порогового. Гипотеза отклоняется для $\alpha = 0.05$.

(e) Интервал будем строить через центральную статистику.

$$X \sim \mathrm{Exp}(\lambda) \Longrightarrow \lambda X \sim \mathrm{Exp}(1) \Longrightarrow \lambda \sum_{i=1}^{n} X_i \sim \Gamma(n, 1)$$

Таким образом, $T = \lambda n \overline{X}$ является центральной статистикой для параметра λ, поскольку распределение статистики не зависит от параметра и строго монотонно по параметру ($\lambda > 0$). Очевидно, что для оцениваемого параметра $\sigma^2 = \lambda^{-2}$ статистика $T = \frac{n}{\sigma} \overline{X}$ также является центральной. Построим центральный доверительный интервал:

$$P(t_{\alpha/2} \leq T \leq t_{1-\alpha/2}) = P\left(t_{\alpha/2} \leq \frac{n}{\sigma}\overline{X} \leq t_{1-\alpha/2}\right) = 1-\alpha$$

$$P\left(\frac{n\overline{X}}{t_{1-\alpha/2}} \leq \sigma \leq \frac{n\overline{X}}{t_{\alpha/2}}\right) = P\left(\left(\frac{n\overline{X}}{t_{1-\alpha/2}}\right)^2 \leq \sigma^2 \leq \left(\frac{n\overline{X}}{t_{\alpha/2}}\right)^2\right) = 1-\alpha$$

Окончательно получим интервал:

$$\left(\frac{n\overline{X}}{t_{1-\alpha/2}}\right)^2 \leq \sigma^2 \leq \left(\frac{n\overline{X}}{t_{\alpha/2}}\right)^2.$$

Для данной выборки $n = 10$, $\alpha = 0.05$, тогда $t_{\alpha/2} \approx 4.795$, $t_{1-\alpha/2} = 17.085$. Центральный доверительный интервал:

$$37.771 \leq \sigma^2 \leq 479.436.$$