Доверительные интервалы

n – объем выборки
e – статистическая точность

1. Нормальное распределение:

$x \in N(m, s)$

Точечные оценки:

Математическое ожидание: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Дисперсия:

$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2$, если m известно,

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$, если m неизвестно.

Если $x \in N(m, s)$, то $\bar{x} \in N\left(m, \frac{s}{\sqrt{n}}\right)$

Интервальные оценки:

Доверительный интервал – это такой интервал, который с заданной вероятностью Q накрывает истинное значение оцениваемого параметра.

Q – доверительная вероятность.

1.1. Доверительный интервал для математического ожидания, дисперсия s^2 известна.

$$\bar{x} - N_{\frac{1+Q}{2}} \frac{s}{\sqrt{n}} \le m \le \bar{x} + N_{\frac{1+Q}{2}} \frac{s}{\sqrt{n}},$$

N_a – $a \cdot 100$-процентный квантиль нормального распределения.

1.2. Объем выборки для оценки математического ожидания.

$$n \ge \frac{s^2}{e^2} N^2_{\frac{1+Q}{2}}$$

1.3. Точный доверительный интервал для математического ожидания, дисперсия s^2 неизвестна.

s^2 – оценка s^2.

$$\bar{x} - t_{\frac{1+Q}{2}}(n-1)\frac{s}{\sqrt{n}} \le m \le \bar{x} + t_{\frac{1+Q}{2}}(n-1)\frac{s}{\sqrt{n}},$$

$t_a(k)$ – $a \cdot 100$-процентный квантиль распределения Стьюдента с k степенями свободы.

1.4. Приближенный доверительный интервал для математического ожидания, дисперсия s^2 неизвестна.

При больших n можно использовать:

$$\bar{x} - N_{\frac{1+Q}{2}}\frac{s}{\sqrt{n}} \le m \le \bar{x} + N_{\frac{1+Q}{2}}\frac{s}{\sqrt{n}}$$

1.5. Доверительный интервал для дисперсии, математическое ожидание неизвестно.

$$s^2\frac{n-1}{c^2_{\frac{1+Q}{2}}(n-1)} \le s^2 \le s^2\frac{n-1}{c^2_{\frac{1-Q}{2}}(n-1)}$$

$c^2_a(k)$ – $a \cdot 100$-процентный квантиль распределения Хи-квадрат с k степенями свободы.

2. Многомерное нормальное распределение:

$$f(x)=\frac{1}{\sqrt{(2\pi)^k \det(C)}}\cdot \exp\left(-\frac{1}{2}(x-m)^T\cdot C^{-1}\cdot(x-m)\right)$$

$$x=[x_1 \quad \ldots \quad x_k]$$

$$m=[m_1 \quad \ldots \quad m_k]$$

$$C=\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_k \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma_k\sigma_1 & \rho\sigma_k\sigma_2 & \cdots & \sigma_k^2 \end{bmatrix}$$

Двумерное нормальное распределение:

$$C=\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix}$$

Точечные оценки:

$$\bar{x}_j=\frac{1}{n}\sum_{s=1}^{n} x_{sj},\ j=\overline{1,k}$$

$$c_{ij}=\frac{1}{n-1}\sum_{s=1}^{n}(x_{si}-\bar{x}_i)(x_{sj}-\bar{x}_j)$$

$$r_{ij}=\frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}$$

Интервальные оценки:

2.1. Приближенный доверительный интервал для коэффициента корреляции многомерного нормального распределения, коэффициент корреляции r известен.

$$r-N_{\frac{1+Q}{2}}\frac{1-r^2}{\sqrt{n}}\le r\le r+N_{\frac{1+Q}{2}}\frac{1-r^2}{\sqrt{n}}.$$

2.2. Объем выборки для оценки коэффициента корреляции.

$$n\ge\frac{(1-r^2)^2}{e^2}N^2_{\frac{1+Q}{2}}$$

2.3. Приближенный доверительный интервал для коэффициента корреляции двумерного нормального распределения, коэффициент корреляции r неизвестен.

r – оценка r .

$$\frac{e^{(2z-2d)}-1}{e^{(2z-2d)}+1} \le r \le \frac{e^{(2z+2d)}-1}{e^{(2z+2d)}+1},$$

$$z=\frac{1}{2}\ln\frac{1+r}{1-r},$$

$$d=N_{\frac{1+Q}{2}}\sqrt{\frac{1}{n-3}}.$$

3. Биномиальное распределение:

p – вероятность (частота)

Точечные оценки:

$v = \frac{k}{n}$ – оценка p

$k = \sum_{i=1}^{n} x_i$ – оценка числа появлений события

$x_i = 1$ с вероятностью p

$x_i = 0$ с вероятностью $1 - p$

Интервальные оценки:

3.1. Доверительный интервал для вероятности биномиального распределения, вероятность p известна.

$$v - N_{\frac{1+Q}{2}} \sqrt{\frac{p(1-p)}{n}} \le p \le v + N_{\frac{1+Q}{2}} \sqrt{\frac{p(1-p)}{n}}$$

3.2. Объем выборки для оценки частоты.

$$n \ge \frac{p(1-p)}{e^2} N^2_{\frac{1+Q}{2}}$$

3.3. Точный доверительный интервал для вероятности биномиального распределения, вероятность p неизвестна.

$$P_L(n, X) = \frac{X}{X + (n - X + 1) F_{\frac{1+Q}{2}}(2n - 2X + 2, 2X)}$$

$$P_U(n, X) = 1 - P_L(n, n - X)$$

$$P_L(n, k) \le p \le P_U(n, k)$$

$$\frac{k}{k + (n - k + 1) F_{\frac{1+Q}{2}}(2n - 2k + 2, 2k)} \le p \le \frac{(k+1) F_{\frac{1+Q}{2}}(2k + 2, 2n - 2k)}{n - k + (k+1) F_{\frac{1+Q}{2}}(2k + 2, 2n - 2k)}$$

$F_a(k_1, k_2)$ – $a \cdot 100$-процентный квантиль распределения Фишера с k_1 и k_2 степенями свободы.

3.4. Приближенный доверительный интервал для вероятности биномиального распределения, вероятность p неизвестна.

При больших n можно использовать:

$$\nu - N_{\frac{1+Q}{2}} \sqrt{\frac{\nu(1-\nu)}{n}} \le p \le \nu + N_{\frac{1+Q}{2}} \sqrt{\frac{\nu(1-\nu)}{n}}$$

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Колесников Д.Н., Сиднев А.Г., Юрганов А.А. Моделирование случайных факторов в задачах автоматики и вычислительной техники: Учебное пособие / СПбГТУ. – СПб.: Изд-во СПбГТУ, 1994. – 106 с.
2. Мятлев В.Д., Панченко Л.А., Терехин А.Т. Основы теории вероятностей. Пособие по курсу "Математические методы в биологии". – М.: Макс Пресс, 2002.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
4. Оуэн Д.Б. Сборник статистических таблиц. – М.: Вычислительный центр АН СССР, 1973. – 586 с.