Типовой расчет №2
по математической статистике.
Интервальная оценка

Ким В.Р.
Группа М3207
Вариант №5

Теоретическая справка

Вероятность попадания в доверительный интервал = доверительная вероятность γ:

$$P(\hat{\theta} - \delta < \theta < \hat{\theta} + \delta) = \gamma,$$

отношение $\frac{t_\gamma s}{\sqrt{n}} = \gamma$ называется точностью оценки.

- **Если σ (генеральное СКО) известно**: границы доверительного интервала определяются как
$$\bar{x} + - \frac{t_\gamma \sigma}{\sqrt{n}},$$
t_γ можно найти из соотношения
$$2\Phi(t_\gamma) = \gamma,$$

- **Если σ неизвестно:** границы интервала вычисляются по похожей формуле:
$$\bar{x} + - \frac{t_\gamma s}{\sqrt{n}},$$
но вместо σ здесь s - его несмещенная оценка.
А t_γ - значение t-распределения (распр. Стьюдента) с $n-1$ степенями свободы (табличное значение)

- **Чтобы найти доверительный интервал для СКО**, используем распределение Хи-квадрат. Оно связано с СКО следующим образом
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$
а доверительный интервал для СКО определяется так:
$$\frac{(n-1)s^2}{\chi^2_{\alpha_1,\, k}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\alpha_2,\, k}},$$
где $k = n-1$ - число степеней свободы, $\alpha_1 = \frac{1-\gamma}{2}$, $\alpha_2 = \frac{1+\gamma}{2}$.

Задача 1. Построить доверительные интервалы при значениях надежности $\gamma = 0,95$, $\gamma = 0,99$, $\gamma = 0,999$ для математического ожидания случайной величины, распределенной по нормальному закону. Выборка $n = 50$.

1. Каким образом величина доверительной вероятности влияет на ширину доверительного интервала?

2. Как увеличить точность оценки при заданном значении надежности?

Решение:
σ неизвестно. t_γ находим по таблице, для $n - 1 = 49$ степеней свободы:
$t_{0.95} \approx 2.008$, $t_{0.99} \approx 2.678$, $t_{0.999} \approx 3.501$.
Интервалы:

$$\mu_{95} \in (\bar{x} - 2.008\frac{s}{\sqrt{50}};\ \bar{x} + 2.008\frac{s}{\sqrt{50}})$$
$$\mu_{99} \in (\bar{x} - 2.678\frac{s}{\sqrt{50}};\ \bar{x} + 2.678\frac{s}{\sqrt{50}})$$
$$\mu_{999} \in (\bar{x} - 3.501\frac{s}{\sqrt{50}};\ \bar{x} + 3.501\frac{s}{\sqrt{50}})$$

1. Больше γ - шире интервал. Формально: t_γ возрастает одновременно с $\gamma \to$ возрастает и модуль $\bar{x} + -\frac{t_\gamma \sigma}{\sqrt{n}}$. Логически - чтобы с большими гарантиями покрыть истинное значение - нужно взять больший интервал.

2. Нужно уменьшить доверительный интервал, за счет увеличения выборки (больше $n \to$ меньше $\gamma \to$ - у‘же интервал)

Задача 2. Шестикратное взвешивание изделия из ценного материала дало следующие результаты (в граммах):
5.825, 5.844, 5.846, 5.850, 5.857, 5.861.
Предполагая, что результаты измерений распределены по нормальному закону, требуется:

1. Найти точечные оценки математического ожидания и среднего квадратического отклонения;

2. Найти 99%-ный доверительный интервал, покрывающий истинный вес изделия;

3. Найти 95%-ный доверительный интервал, покрывающий среднее квадратическое отклонение;

4. Найти предельную погрешность, которую мы допускаем, считая истинный вес изделия равным средней арифметической (доверительную вероятность принять равной 0,98).

Решение:

1. Точечная оценка мат. ожидания - выборочное среднее:

$$\bar{x} = \frac{(5.825 + 5.844 + 5.846 + 5.850 + 5.857 + 5.861)}{6} \approx 5.847$$

Точечная оценка СКО - корень из исправленной выборочной дисперсии:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{5}\sum_{i=1}^{n}(x_i - 5.847)^2}$$

$$= \sqrt{\frac{0.022^2 + 2 \cdot 0.003^2 + 0.001^2 + 0.01^2 + 0.014^2}{5}} \approx 0.013$$

2. $\gamma = 0.99$, то таблице: $t_{0.99}(5) \approx 6.86$.
   Границы интервала:

$$\bar{x} + -\frac{t_\gamma s}{\sqrt{n}}, = 5.847 + -\frac{6.86 \cdot 0.013}{\sqrt{6}} \to \approx (5.81; 5.88)$$

3. Найдем доверительный интервал с помощью распределения Хи-квадрат:
   $k = n - 1 = 5$,

$$\alpha_1 = \frac{1-\gamma}{2} = \frac{0.05}{2} = 0.025, \quad \chi^2_{0.025,5} = 12.8,$$

$$\alpha_2 = \frac{1+\gamma}{2} = \frac{1.95}{2} = 0.975 \quad \chi^2_{0.975,5} = 0.831$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha_1,k}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{\alpha_2,k}}; \quad \frac{5 \cdot 0.013^2}{12.8} \le \sigma^2 \le \frac{5 \cdot 0.013^2}{0.831};$$

$$0.0000660156 \le \sigma^2 \le 0.00101685;$$

$$\sqrt{0.0000660156} \le \sigma \le \sqrt{0.00101685}.$$

4. $\gamma = 0.98$

$$\Delta = \frac{t_\gamma s}{\sqrt{n}} = \frac{4.03 \cdot 0.013}{\sqrt{6}} = 0.21$$

Задача 3. Случайная величина ξ (число сорняков в пробе зерна) распределена по закону Пуассона. Ниже приведено распределение семян сорняков в n пробах зерна x_i – количество сорняков в одно пробе. n_i - число проб, содержащих x_i семян сорняков.
Построить доверительный интервал для параметра λ с надежностью $0,95$

x_i	0	1	2	3	4	5	6	7
n_1	396	361	173	48	11	7	3	1

Решение:
Закон Пуассона:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$n = \sum(n_i) = 1000$. Для распределения Пуассона мат. ожидание равно λ, тогда оценку $\hat{\lambda}$ можно найти как выборочное среднее:

$$\hat{\lambda} = \frac{361 + 173 \cdot 2 + 48 \cdot 3 + 44 + 35 + 18 + 7}{1000} = 0.955$$

Выборка большая, можем использовать асимптотическое приближение. По ЦПТ выборочное среднее имеет асимптотически нормальное распределение $N(\mu, \sigma^2)$, найдем его параметры:

1. $\mu = \lambda$ по свойству распределения Пуассона
2. Найдём дисперсию суммы $\sum_{i=1}^{n} X_i$. Так как X_i независимы:

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i) = \sum_{i=1}^{n} \lambda = n\lambda.$$

Используем полученное значение для нахождения дисперсии оценки $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$:

$$D(\hat{\lambda}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot n\lambda = \frac{\lambda}{n}.$$

Итого: $\hat{\lambda} \sim N\left(\lambda, \frac{\lambda}{n}\right)$.

Теперь можно записать доверительный интервал λ в общем виде:

$$(\hat{\lambda} - z_{\frac{3+\gamma}{2}} \cdot \sqrt{D(\hat{\lambda})};\ \hat{\lambda} + z_{\frac{1+\gamma}{2}} \cdot \sqrt{D(\hat{\lambda})})$$

Вычислим для нашего случая: Определим стандартную ошибку:

$$\sqrt{D(\hat{\lambda})} = \sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{0.955}{1000}} \approx 0.0309.$$

Для доверительной вероятности $\gamma = 0.95$, краевые значения определяются как:

$$\alpha = 1 - \gamma = 0.05, \quad \frac{\alpha}{2} = 0.025.$$

Тогда используем квантиль стандартного нормального распределения:

$$z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96.$$

Тогда доверительный интервал для λ имеет вид:

$$\hat{\lambda} \pm z_{0.975} \cdot SE \quad \to \quad 0.955 \pm 1.96 \cdot 0.0309.$$

Вычисляем погрешность:

$$1.96 \cdot 0.0309 \approx 0.0605.$$

Таким образом, доверительный интервал:

$$0.955 - 0.0605 \leq \lambda \leq 0.955 + 0.0605,$$

Ответ: $(0.8945;\ 1.0155)$

Задача 4. Для экспоненциального распределения “со сдвигом” , имеющего плотность

$$f(x;\theta) = \begin{cases} e^{-x-\theta}, & x \geq \theta \\ 0, & x < \theta \end{cases} \tag{1}$$

по выборке объема n построить интервальную оценку параметра θ (сдвиг) с доверительной вероятностью γ

Решение:
Сделав замену переменной $y = x - \theta$, получаем, что $y \geq 0$ и функция плотности

$$f_Y(y) = e^{-y}, \quad y \geq 0,$$

это плотность стандартного экспоненциального распределения Exp(1), $\lambda = 1$.
Пусть $X_1, X_2, \ldots, X_n$ — независимая выборка из данного распределения. Заметим, что условие $x \geq \theta$ для всех наблюдений позволяет нам использовать минимальное значение выборки как статистику, содержащую всю информацию о сдвиге θ. Обозначим

$$X_{min} = \min\{X_1, X_2, \ldots, X_n\}.$$

Каждое наблюдение X_i можно представить в виде

$$X_i = \theta + Y_i, \quad \text{где } Y_i \sim \text{Exp}(1),$$

где $Y_1, Y_2, \ldots, Y_n$ — независимые случайные величины.

$$Y_i \sim \text{Exp}(1), \quad \text{с плотностью } f_{Y_i}(y) = e^{-y},\ y \geq 0$$

Обозначим $Z = \min\{Y_1, Y_2, \ldots, Y_n\}$ Тогда минимум выборки - $X_{min} = \theta + Z, Z = X_{min} - \theta$

Функция распределения минимума

Рассмотрим $Z = \min\{Y_1, Y_2, \ldots, Y_n\}$.
Найдём её функцию распределения $F_Z(z)$.

$$\begin{aligned}
F_Z(z) &= P(Z \leq z) = 1 - P(Z > z) \\
&= 1 - P(\min\{Y_1, \ldots, Y_n\} > z) \\
&= 1 - P(Y_1 > z \cap \cdots \cap Y_n > z) \\
&= 1 - \prod_{i=1}^{n} P(Y_i > z) \quad \text{(по независимости)} \\
&= 1 - \prod_{i=1}^{n} e^{-z} \quad \text{(т.к. } P(Y_i > z) = e^{-z}\text{)} \\
&= 1 - e^{-nz}, \quad z \geq 0
\end{aligned}$$

Мы ищем такое множество I, что

$$P(\theta \in I) = \gamma.$$

Используя распределение Z, найдем квантиль z_γ такой, что $P(Z \leq z_\gamma) = \gamma$.
Имея функцию распределения $F_Z(z) = 1 - e^{-nz}$ получаем:

$$1 - e^{-nz_\gamma} = \gamma \quad \Rightarrow \quad e^{-nz_\gamma} = 1 - \gamma,$$

выразим z_γ

$$z_\gamma = -\frac{1}{n} \ln(1 - \gamma).$$

Подставим $Z = X_{min} - \theta$:

$$P\Big(X_{min} - \theta \leq z_\gamma\Big) = \gamma \Leftrightarrow P\Big(\theta \geq X_{min} - z_\gamma\Big) = \gamma.$$

Так как сверху θ ограничена X_{min}, получаем $\theta \in \Big(X_{min} - z_\gamma,\, X_{min}\Big)$.

Подставляя z_γ:

$$P\left(X_{min} - \frac{1}{n}\ln\frac{1}{1-\gamma} \le \theta \le X_{min}\right) = \gamma.$$

(перепишем $-\frac{1}{n}\ln(1-\gamma) = \frac{1}{n}\ln\left(\frac{1}{1-\gamma}\right)$)

Итого, интервал для параметра θ с доверительной вероятностью γ имеет вид:

$$\theta \in \left(X_{min} - \frac{1}{n}\ln\left(\frac{1}{1-\gamma}\right),\, X_{min}\right).$$