Точечные оценки

- Распределение исследуемого признака X характеризуется целым рядом числовых теоретических характеристик, или параметров: $\Theta_1, \Theta_2, \ldots, \Theta_n$.
- Сюда относятся среднее значение (мат. ожидание), дисперсия, среднее кв. отклонение, моменты и т.д.
- На практике распределение изучаемого признака, как правило, неизвестно или же известно частично.

Поэтому необходимо указать некоторую функцию от исходных данных (**статистику**), для использования *вместо неизвестного теоретического параметра* в качестве его приближения.

Это приближенное значение параметра Θ называется **точечной оценкой** и обозначается Θ^*.

Точечные оценки

Выборочное среднее

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Выборочная дисперсия

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$$

Выборочное среднее квадратическое (стандартное) отклонение

$$S = +\sqrt{S^2}$$

Свойства оценок

Для одного и того же параметра можно построить разные оценки.

Например, оценкой $a = MX$ может служить среднее арифметическое, полусумма максимума и минимума и т. д.

Требования, которые предъявляются к оценкам (свойства):

- Несмещенность
- Состоятельность
- Эффективность

Свойства оценок

Всякая оценка неизвестного параметра Θ по выборке является функцией от выборочных данных $\Theta^* = \Theta^*(x_1, x_2, \dots, x_n)$.

Величины $x_i (i = 1, \dots, n)$ можно рассматривать как случайные величины.

Тогда и оценку $\Theta^* = \Theta^*(x_1, x_2, \dots, x_n)$ можно рассматривать как случайную величину, имеющую *распределение и числовые характеристики.*

Результаты наблюдений $x_i (i = 1, \dots, n)$ независимые и одинаково распределенные, поэтому

$$M(x_1) = M(x_2) = \cdots = M(x_n) = M(X) = a$$
$$D(x_1) = D(x_2) = \cdots = D(x_n) = D(X) = \sigma^2$$

Несмещенность оценок

Оценка Θ^* параметра Θ называется **несмещенной**, если математическое ожидание оценки Θ^* совпадает с оцениваемым параметром Θ:

$$M\Theta^* = \Theta$$

В противном случае оценка называется смещенной.

Условие иногда называют условием отсутствия систематической ошибки и его смысл в том, что при многократном использовании вместо параметра Θ его оценки Θ^* среднее значение ошибки приближения $\Theta \approx \Theta^*$равно нулю.

Несмещенность оценок

Несмещенность оценки:

$$M\Theta^* = \Theta$$

Выборочное среднее $\bar{x}$ является несмещенной оценкой для теоретического (генерального) среднего $a = MX$.

$$M(\bar{x}) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n} M(x_i)\right) = \frac{1}{n}\left(\sum_{i=1}^{n} a\right) = a$$

Несмещенность оценок

Рассмотрим выборочную дисперсию S^2 , предварительно преобразовав ее:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}[(x_i - a) - (\bar{x} - a)]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - \frac{2}{n}(\bar{x} - a)\sum_{i=1}^{n}(x_i - a) + \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - \frac{2(\bar{x} - a)(\bar{x} - a)n}{n} + \frac{n(\bar{x} - a)^2}{n} =$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - (\bar{x} - a)^2$$

Несмещенность оценок

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - (\bar{x} - a)^2, \qquad \text{где } a = MX$$

Найдем мат. ожидание S^2:

$$M(S^2) = \frac{1}{n}\sum_{i=1}^{n} M(x_i - a)^2 - M(\bar{x} - a)^2 = \frac{1}{n}\sum_{i=1}^{n} D(x_i) - D(\bar{x})$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sigma^2 - D(\bar{x}) = \sigma^2 - D(\bar{x})$$

$$D(\bar{x}) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(x_i) = \frac{\sigma^2}{n}$$

Несмещенность оценок

$$M(S^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Выборочная дисперсия S^2 является *смещенной* оценкой для теоретической (генеральной) дисперсии $\sigma^2 = D(X)$

Несмещенной оценкой теоретической дисперсии σ^2 является величина

$$\dot{S}^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

называемая **«исправленной» выборочной дисперсией**

Точечные оценки

Выборочное среднее

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Выборочная дисперсия (смещенная)

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$$

Исправленная выборочная дисперсия

$$\dot{S}^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Состоятельность оценок

Оценка Θ^* параметра Θ называется **состоятельной**,
если с ростом объема выборки n она сходится по вероятности
к оцениваемому параметру Θ:

$$\Theta^* \xrightarrow{p} \Theta \quad \text{при } n \to \infty$$

Состоятельность оценок

Выборочное среднее $\bar{x}$ является *состоятельной* оценкой для теоретического (генерального) среднего $a = MX$, поскольку согласно закону больших чисел среднее арифметическое n независимых одинаково распределенных сл. в. $x_1, x_2, \dots, x_n$ сходится по вероятности при $n \to \infty$ к их общему мат. ожиданию $a = MX$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} a \quad \text{при } n \to \infty$$

Состоятельность оценок

Исправленная дисперсия $\dot{S}^2$ и выборочная дисперсия S^2 являются *состоятельными* оценкой для теоретической дисперсии $\sigma^2 = DX$. Действительно, по закону больших чисел, среднее арифметическое n независимых одинаково распределенных сл. в. $(x_1 - a)^2, (x_2 - a)^2, \dots, (x_n - a)^2$ сходится по вероятности при $n \to \infty$ к их общему мат. ожиданию $M(X - a)^2 = \sigma^2$:

$$\dot{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - a)^2 \xrightarrow{p} \sigma^2 \quad \text{при } n \to \infty$$

Свойства оценок

Выберем в качестве оценки параметра $\Theta = a = MX$ величину $\Theta^* = x_1$ - первое наблюденное значение признака X.

Оценка является несмещенной:

$$M(\Theta^*) = M(x_1) = M(X) = a$$

Но не является состоятельной:

Неверно, что: $x_1 \xrightarrow{p} \sigma^2$ при $n \to \infty$

Эффективность оценок

Оценка Θ_1^* параметра Θ называется **более эффективной**, чем оценка Θ_2^*, если:

$$M(\Theta_1^* - \Theta)^2 < M(\Theta_2^* - \Theta)^2$$

Для несмещенных оценок неравенство эквивалентно:

$$D\Theta_1^* < D\Theta_2^*$$

Наилучшей оценкой параметра Θ среди всех несмещенных оценок является та из них, которая обладает минимальной дисперсией. Такая оценка называется эффективной.

Интервальные оценки

Точечные оценки Θ^* дают приближенное значение неизвестного параметра и используются, когда нужно назвать некоторое число вместо неизвестного параметра Θ. Зачастую точное определение неизвестного параметра Θ по данной выборке невозможно.

Доверительный интервал (интервальная оценка) –

интервал $(\Theta^* - \varepsilon; \Theta^* + \varepsilon)$, в который с заданной вероятностью γ входит истинное значение параметра Θ.

Надежность (доверительная вероятность) $\boldsymbol{\gamma}$ –
некоторая задаваемая вероятность. Обычно γ задается как 0.9, 0.95, 0.99.

$$P\{|\Theta^* - \Theta| < \varepsilon\} = \gamma$$

Интервальные оценки

$$P\{|\Theta^* - \Theta| < \varepsilon\} = \gamma$$

ε – **точность оценки** Θ^* параметра Θ.

В общем случае доверительный интервал может быть несимметричен:

$$P\{\Theta_1 < \Theta < \Theta_2\} = \gamma, \qquad 0 < \gamma < 1$$

Отличие интервального оценивания от точечного:

- Доверительный интервал как оценка менее «точен», т.к. указывается множество возможных значений Θ
- Утверждение "$\Theta \in (\Theta_1, \Theta_2)$ с вероятностью γ“ – истина, в то время как событие "$\Theta = \Theta^*$", как правило, ложное

Интервальные оценки

Интервальное оценивание параметров нормального распределения:

1. Интервальная оценка **среднего** нормального распределения при *известной* дисперсии
2. Интервальная оценка **среднего** нормального распределения при *неизвестной* дисперсии

3. Интервальная оценка **дисперсии** нормального распределения при *известном* среднем
4. Интервальная оценка **дисперсии** нормального распределения при *неизвестном* среднем

Интервальная оценка среднего при известной дисперсии

Пусть исследуемый признак имеет нормальное распределение с **неизвестным** средним a и **известной** дисперсией σ^2,

$X \sim N(a, \sigma)$.

Эффективной точечной оценкой параметра a является выборочное среднее $\bar{x}$.

Построение доверительного интервала основано на теореме.

Интервальная оценка среднего при известной дисперсии

Теорема:

Пусть $x_1, x_2, \dots, x_n$ – результаты *n* независимых наблюдений нормально распределенного признака X,
т.е. $x_i \sim N(a, \sigma), i = 1,2, \dots, n$, значение σ известно.

Тогда статистика

$$U = \frac{\bar{x} - a}{\sigma / \sqrt{n}}$$

имеет *<u>стандартное нормальное распределение</u>*: $U \sim N(0,1)$

Интервальная оценка среднего при известной дисперсии

$$U = \frac{\bar{x} - a}{\sigma/\sqrt{n}}$$

Пусть $u_{1-\alpha/2}$ – квантиль порядка $(1 - \alpha/2)$ стандартного нормального распределения. Поскольку $u_{\alpha/2} = -u_{1-\alpha/2}$, то:

$$P\{|U| < u_{1-\alpha/2}\} = P\{-u_{1-\alpha/2} < U < u_{1-\alpha/2}\} =$$

$$= P\{u_{\alpha/2} < U < u_{1-\alpha/2}\} = 1 - \alpha = \gamma$$

вероятность попадания в доверительный интервал

Интервальная оценка среднего при известной дисперсии

$$U = \frac{\bar{x} - a}{\sigma/\sqrt{n}} \qquad P\{-u_{1-\alpha/2} < U < u_{1-\alpha/2}\} = 1 - \alpha = \gamma$$

Решим неравенства относительно a:

$$-u_{1-\alpha/2} < \frac{\bar{x} - a}{\sigma/\sqrt{n}} < u_{1-\alpha/2}$$

$$\bar{x} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Построенный доверительный интервал симметричен относительно выборочного среднего, при этом его точность

$$\varepsilon = u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Интервальная оценка среднего при известной дисперсии

Пример:

Фирма коммунального хозяйства желает выборочным методом оценить среднюю квартплату за квартиры определенного типа с **надежностью** не менее 0,99 и **точностью**, не меньшей 10 руб.

Предполагая, что квартплата имеет нормальное распределение со средним квадратическим **отклонением**, не превышающим 30 руб., найти <u>минимальный объем</u> выборки, необходимый для решения поставленной задачи.

Известно: $\gamma = 0.99$ $\quad \varepsilon \geq 10$ $\quad \sigma \leq 30.$ $\quad$ Найти: n

$P\{|\bar{x} - a| < \varepsilon\} \geq \gamma,$ где $\bar{x}$ и a – генеральное и выборочное средние

Интервальная оценка среднего при известной дисперсии

Известно: $\gamma = 0.99 \qquad \varepsilon \geq 10 \qquad \sigma \leq 30.$ Найти: n

$$P\{|\bar{x} - a| < 10\} \geq 0.99$$

$$P\left\{\bar{x} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

$\alpha = 1 - \gamma = 0.01 \qquad u_{1-\alpha/2} = u_{0.995} = 2.576$ (по таблице)

$$\varepsilon = u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n = \left(\frac{u_{1-\alpha/2}\ \sigma}{\varepsilon}\right)^2 = \left(\frac{2.576 * 30}{10}\right)^2 \approx 59.72$$

Но так как с увеличением надежности и точности растет n,
то $n \geq 59.72$ и $n_{min} = 60$.

Интервальная оценка среднего при неизвестной дисперсии

Пусть исследуемый признак имеет нормальное распределение $X \sim N(a, \sigma)$, причем значения параметров a и σ^2 неизвестны.

По выборке объема найдем эффективную оценку параметра a – выборочное среднее $\bar{x}$ и оценку параметра σ^2 – выборочную дисперсию $\dot{S}^2$.

Построение интервальной оценки для a в данном случае основано на теореме.

Интервальная оценка среднего при неизвестной дисперсии

Теорема:

Пусть $x_1, x_2, \dots, x_n$ – результаты *n* независимых наблюдений нормально распределенного признака X,
т.е. $x_i \sim N(a, \sigma), i = 1,2, \dots, n$, значение σ **неизвестно**.

Тогда статистика

$$t = \frac{\bar{x} - a}{\dot{S}/\sqrt{n-1}},$$

где $\dot{S} -$ несмещенное выборочное среднее квадратическое отклонение,

имеет *распределение Стьюдента* с $(n-1)$ степенями свободы.

Интервальная оценка среднего при неизвестной дисперсии

$$t = \frac{\bar{x} - a}{\dot{S}/\sqrt{n-1}}$$

Пусть $t_{1-\alpha/2}(n-1)$ – квантиль порядка $(1 - \alpha\ /\ 2)$ распределения Стьюдента с $(n-1)$ степенями свободы. Поскольку распределение симметрично, то:

$$P\left\{|t| < t_{1-\alpha/2}(n-1)\right\} = P\left\{-t_{1-\alpha/2}(n-1) < t < t_{1-\alpha/2}(n-1)\right\}$$

$$= P\left\{t_{\alpha/2}(n-1) < t < t_{1-\alpha/2}(n-1)\right\} = 1 - \alpha = \gamma$$

вероятность попадания в доверительный интервал

Интервальная оценка среднего при неизвестной дисперсии

$$t = \frac{\bar{x} - a}{S/\sqrt{n-1}} \qquad P\{-t_{1-\alpha/2}(n-1) < U < t_{1-\alpha/2}(n-1)\} = 1 - \alpha = \gamma$$

Решим неравенства относительно a:

$$-t_{1-\alpha/2}(n-1) < \frac{\bar{x} - a}{S/\sqrt{n-1}} < t_{1-\alpha/2}(n-1)$$

$$\bar{x} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} < a < \bar{x} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

Построенный доверительный интервал симметричен относительно выборочного среднего, при этом его точность

$$\varepsilon = t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

Интервальная оценка среднего при неизвестной дисперсии

Пример:

Предположим, что выработка X на одного станочника механического цеха в отчетном году (в % к предыдущему) имеет нормальное распределение с неизвестными параметрами.

Построить доверительный интервал надежности 0.95
для средней выработки на одного рабочего-станочника.

n = 117

Выработка, %	**80÷90**	**90÷100**	**100÷110**	**110÷120**	**120÷130**	**130÷140**	**140÷150**
Кол-во раб.	8	15	46	29	13	3	3

Интервальная оценка среднего при неизвестной дисперсии

Точечные оценки найдены ранее:

$$\bar{x} = 108.85 \qquad S = 12.65$$
$$n = 117 \qquad \gamma = 0.95$$

$$P\left\{\bar{x} - t_{1-\alpha/2}(n-1)\frac{\dot{S}}{\sqrt{n-1}} < a < \bar{x} + t_{1-\alpha/2}(n-1)\frac{\dot{S}}{\sqrt{n-1}}\right\} = 1 - \alpha = \gamma$$

$$\alpha = 0.05 \qquad t_{0.975}(116) = 1.98$$

$$108.85 - 1.98\frac{12.65}{\sqrt{116}} < a < 108.85 + 1.98\frac{12.65}{\sqrt{116}}$$

$$106.52 < a < 111.17$$

Интервальная оценка среднего

При известной дисперсии:

$$P\left\{\bar{x} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < a < \bar{x} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha = \gamma$$

При неизвестной дисперсии:

$$P\left\{\bar{x} - t_{1-\alpha/2}(n-1)\frac{\dot{S}}{\sqrt{n-1}} < a < \bar{x} + t_{1-\alpha/2}(n-1)\frac{\dot{S}}{\sqrt{n-1}}\right\} = 1 - \alpha = \gamma$$

Интервальная оценка дисперсии при известном среднем

Пусть исследуемый признак $X \sim N(a, \sigma)$, где значение параметра a **известно**, а дисперсия σ^2 **неизвестна**.

Эффективной оценкой дисперсии σ^2 в этом случае является величина

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - a)^2$$

Доверительный интервал надежности $\gamma = 1 - \alpha$

для σ^2 строится на основе теоремы.

Интервальная оценка дисперсии при известном среднем

Теорема:

Пусть $x_1, x_2, \dots, x_n$ – результаты *n* независимых наблюдений нормально распределенного признака X,
т.е. $x_i \sim N(a, \sigma), i = 1,2, \dots, n$, значение a **известно**.

Тогда статистика

$$H = \frac{nS^2}{\sigma^2},$$

имеет *распределение хи-квадрат* с n степенями свободы

Интервальная оценка дисперсии при известном среднем

$$P\{\chi^2_{\alpha/2}(n) < H < \chi^2_{1-\alpha/2}(n)\} = 1 - \alpha$$

$$\chi^2_{\alpha/2}(n) < \frac{nS^2}{\sigma^2} < \chi^2_{1-\alpha/2}(n)$$

Решим неравенства относительно σ^2:

$$\frac{nS^2}{\chi^2_{1-\alpha/2}(n)} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha/2}(n)}$$

$$\sqrt{\frac{nS^2}{\chi^2_{1-\alpha/2}(n)}} < \sigma < \sqrt{\frac{nS^2}{\chi^2_{\alpha/2}(n)}}$$

Построенные интервалы не симметричны относительно S^2.

Интервальная оценка дисперсии при неизвестном среднем

Пусть исследуемый признак $X \sim N(a, \sigma)$, где значение параметров a и σ^2 **неизвестно**.

Эффективной оценкой дисперсии σ^2 в этом случае является несмещенная выборочная дисперсия:

$$\dot{S}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{x} - a)^2$$

Доверительный интервал надежности $\gamma = 1 - \alpha$ для σ^2 строится на основе теоремы.

Интервальная оценка дисперсии при неизвестном среднем

Теорема:

Пусть $x_1, x_2, \dots, x_n$ – результаты *n* независимых наблюдений нормально распределенного признака X,
т.е. $x_i \sim N(a, \sigma), i = 1,2, \dots, n$, значение a **неизвестно**.

Тогда статистика

$$H = \frac{(n-1)\dot{S}^2}{\sigma^2},$$

имеет *распределение хи-квадрат* с $(n-1)$ степенями свободы

Интервальная оценка дисперсии при неизвестном среднем

$$P\left\{\chi^2_{\alpha/2}(n-1) < \frac{(n-1)\dot{S}^2}{\sigma^2} < \chi^2_{1-\alpha/2}(n-1)\right\} = 1-\alpha$$

Решим неравенства относительно σ^2:

$$\frac{(n-1)\dot{S}^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)\dot{S}^2}{\chi^2_{\alpha/2}(n-1)}$$

$$\sqrt{\frac{(n-1)\dot{S}^2}{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < \sqrt{\frac{(n-1)\dot{S}^2}{\chi^2_{\alpha/2}(n-1)}}$$

Построенные интервалы не симметричны относительно $\dot{S}^2$.

Интервальная оценка дисперсии

При известном среднем:

$$P\left\{\frac{nS^2}{\chi^2_{1-\alpha/2}(n)} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha/2}(n)}\right\} = 1 - \alpha = \gamma$$

При неизвестном среднем:

$$P\left\{\frac{(n-1)\dot{S}^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)\dot{S}^2}{\chi^2_{\alpha/2}(n-1)}\right\} = 1 - \alpha = \gamma$$

Интервальные оценки

№ п/п	Параметр	Предполо-жения	Оценка параметра	Границы доверительного интервала
1	a	σ^2 известна	$\bar{x}$	$\bar{x} \mp u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
2	a	σ^2 неизвестна	$\bar{x}$	$\bar{x} \mp t_{1-\alpha/2}(n-1) \dfrac{\dot{S}}{\sqrt{n-1}}$
3	σ^2	a известно	S^2	$\dfrac{nS^2}{\chi^2_{1-\alpha/2}(n)}, \dfrac{nS^2}{\chi^2_{\alpha/2}(n)}$
4	σ^2	a неизвестно	$\dot{S}^2$	$\dfrac{(n-1)\dot{S}^2}{\chi^2_{1-\alpha/2}(n-1)}, \dfrac{(n-1)\dot{S}^2}{\chi^2_{\alpha/2}(n-1)}$

Интервальные оценки

- Строго говоря, рассмотренные интервальные оценки применимы только к выборкам, извлеченным из нормальной генеральной совокупности.
- Однако исследования показали, что эти оценки являются (особенно при больших объемах выборок) весьма устойчивыми к отклонениям исследуемых генеральных совокупностей от нормальных.
  То есть они могут применяться и к выборкам из генеральных совокупностей, не имеющих нормальное распределение.
- При этом следует учесть, что истинное значение надежности используемого доверительного интервала будет незначительно отличаться от заданного.

Интервальные оценки

Пример:

Приближенное значение неизвестного среднего тарифа a рабочих механического цеха определяется выборочным средним, т.е. $a \approx \bar{x} = 4.44$.

Для оценки параметра a с заданной надежностью $\gamma = 0.99$ следует построить для него доверительный интервал.

Тарифный разряд	**1**	**2**	**3**	**4**	**5**	**6**
Кол-во рабочих	4	6	12	16	44	18

Интервальные оценки

Пример:

$\bar{x} = 4.44 \quad \gamma = 0.99 \quad n = 100 \quad S = 1.3$ (найдено ранее)

Поскольку выборка извлечена из произвольной генеральной совокупности с неизвестной дисперсией, а ее объем n достаточно велик, то соответствующий доверительный интервал будет иметь приближенно надежность 0.99.

$$\alpha = 1 - \gamma = 0.01 \Rightarrow t_{1-\alpha/2}(n-1) = t_{0.995}(99) \approx 2.626$$

$$\bar{x} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}} < a < \bar{x} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n-1}}$$

$$4.44 - 2.626\frac{1.3}{\sqrt{99}} < a < 4.44 + 2.626\frac{1.3}{\sqrt{99}} \quad \Rightarrow \quad 4.1 < a < 4.781$$

Интервальная оценка вероятности события

Наилучшей точечной оценкой вероятности p события A является относительная частота (доля)

$$p^* = \nu_n / n\,,$$

где n – общее число независимых испытаний Бернулли, в каждом из которых событие A может произойти с вероятностью p или не произойти с вероятностью $q = 1 - p$, а ν_n – число испытаний, в которых произошло событие A.

Интервальная оценка вероятности события

Построим **интервальную оценку вероятности** p, отвечающую надежности γ. Для этого воспользуемся тем фактом, что при большом числе испытаний n в силу интегральной теоремы Муавра-Лапласа оценка $p^* = \nu_n/n$ распределена приближенно по нормальному закону со средним p и дисперсией pq/n.

Следовательно, при большом n распределение величины

$$\frac{p^* - p}{\sqrt{pq \, / \, n}}$$

близко к стандартному нормальному.

Пусть $u_{1-\alpha/2}$ - квантиль порядка $1 - \alpha/2$ стандартного распределения.

Интервальная оценка вероятности события

Тогда при достаточно большом n

$$P\left\{\left|\frac{p^* - p}{\sqrt{pq/n}}\right| < u_{1-\alpha/2}\right\} \approx 1 - \alpha$$

Разрешая неравенства относительно p, можно получить доверительный интервал, при большом n приближенно соответствующий доверительной вероятности $\gamma = 1 - \alpha$:

$$p^* - u_{1-\alpha/2}\sqrt{p^*(1-p^*)\ /\ n} < p < p^* + u_{1-\alpha/2}\sqrt{p^*(1-p^*)\ /\ n}$$

Доверительный интервал симметричен относительно p^* и его точность

$$\varepsilon = u_{1-\alpha/2}\sqrt{p^*(1-p^*)\ /\ n}$$

Интервальная оценка вероятности события

Пример: Событие А в серии из 100 испытаний Бернулли произошло 78 раз. Найти интервальную оценку для вероятности p события А с надежностью 0.9.

Точечная оценка: $\boldsymbol{p^* = 78/100 = 0.78}$

$$p^* - \varepsilon < p < p^* + \varepsilon \qquad \varepsilon = u_{1-\alpha/2}\sqrt{p^*(1-p^*)\;/\;n}$$

При $\alpha = 1 - \gamma = 0.1\; u_{1-\alpha/2} = u_{0.95} = 1.645$

$$\varepsilon = u_{1-\alpha/2}\sqrt{p^*(1-p^*)\;/\;n} = 1.645\sqrt{\frac{0.78(1-0.78)}{100}} \approx 0.068$$

С надежностью 0.9 доверительным интервалом для p будет

(0.712, 0.848)