Лекция 12. Интервальные оценки параметров генеральной совокупности

На лекции рассматриваются вопросы:

1. Понятие интервального оценивания параметров генеральной совокупности.
2. Доверительные интервалы для параметров нормального распределения (для математического ожидания, для среднего квадратического отклонения).

1. Понятие интервального оценивания параметров генеральной совокупности

Разность между генеральными характеристиками и соответствующими выборочными статистиками называется ***ошибкой выборки*** или ***ошибкой репрезентативности.***

Статистические методы позволяют оценить эту разность, которая зависит как от характеристик выборки, так и от ее объема. В процессе выборочного исследования параметры генеральной совокупности определяются в виде интервала, построенного вокруг выборочной статистики. От величины этого интервала зависит качество исследования.

Оценка неизвестного параметра θ называется ***интервальной***, если она определяется двумя числами — концами интервала $\left(\overline{\theta_1};\overline{\theta_2}\right)$, который с определенной вероятностью γ накрывает неизвестный параметр генеральной совокупности.

Интервал $\left(\overline{\theta_1};\overline{\theta_2}\right)$, накрывающий с вероятностью γ истинное значение параметра θ, называется ***доверительным интервалом***, а вероятность γ — ***надежностью оценки*** или ***доверительной вероятностью***.

Часто доверительный интервал выбирается симметричным относительно несмещенной точечной оценки $\overline{\theta}$, то есть выбирается интервал вида $\left(\overline{\theta}-\varepsilon;\overline{\theta}+\varepsilon\right)$ такой, что

$$P\left(\theta \in \left(\overline{\theta}-\varepsilon;\overline{\theta}+\varepsilon\right)\right)=P\left(\left|\theta-\overline{\theta}\right|<\varepsilon\right)=\gamma .$$

Число $\varepsilon>0$ характеризует точность оценки: чем меньше разность $\left|\theta-\overline{\theta}\right|$, тем точнее оценка.

Величина γ выбирается заранее, ее выбор зависит от конкретно решаемой задачи. Надежность принято выбирать равной 0,9; 0,95; 0,99 или

0,999. Тогда практически достоверно нахождение параметра θ в доверительном интервале $\left(\overline{\theta}-\varepsilon;\overline{\theta}+\varepsilon\right)$.

2. Доверительные интервалы для параметров нормального распределения

Пусть выборка производится из генеральной совокупности, имеющей нормальное распределение с параметрами a и σ.

1. Доверительный интервал для математического ожидания.

Доверительный интервал для математического ожидания a имеет вид:

$$\left(\overline{x_в}-\Delta;\ \overline{x_в}+\Delta\right) \text{ или } \overline{x_в}-\Delta<a<\overline{x_в}+\Delta$$

где Δ — точность оценки.

Если известно σ (при $n>30$ $S\approx\sigma$), то

$\Delta=t\cdot\dfrac{\sigma}{\sqrt{n}}$ — при повторной выборке;

$\Delta=t\cdot\dfrac{\sigma}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$ — при бесповторной выборке,

где t определяется из уравнения $\Phi(t)=\dfrac{\gamma}{2}$ по таблице функции Лапласа.

Если σ неизвестно (при $n\le 30$ вычисляют S), то

$\Delta=t_\gamma\cdot\dfrac{S}{\sqrt{n}}$ — при повторной выборке;

$\Delta=t_\gamma\cdot\dfrac{S}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$ — при бесповторной выборке,

где S — исправленное выборочное среднее квадратическое отклонение, t_γ определяется по таблице значений $t_\gamma=t(\gamma;n)$ в зависимости от надежности γ и объема выборки n.

Пример:

Из партии в 5000 электрических ламп было отобрано 300 по схеме бесповторной выборки. Средняя продолжительность горения ламп в выборке оказалась равной 1450 часов. Найти с надежностью 0,9996 доверительный интервал для средней продолжительности горения лампы всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma=20$ ч. Предполагается, что продолжительность горения ламп распределена нормально.

Решение.

По условию задачи:

$N = 5000$,
$n = 300 > 30$,
$\overline{x_в} = 1450$,
$\sigma = 20$,
$\gamma = 0{,}9996$.

Для $\gamma = 0{,}9996$ из уравнения $\Phi(t) = \dfrac{0{,}9996}{2}$ или $\Phi(t) = 0{,}4998$ по таблице значений функции Лапласа находим $t = 3{,}54$.

Так как σ известно и выборка бесповторная, то

$$\Delta = t \cdot \frac{\sigma}{\sqrt{n}} \sqrt{1 - \frac{n}{N}};$$

$$\Delta = 3{,}54 \cdot \frac{20}{\sqrt{300}} \sqrt{1 - \frac{300}{5000}} \approx 3{,}9631.$$

Тогда искомый доверительный интервал

$$1450 - 3{,}9631 < a < 1450 + 3{,}9631.$$

$$1446{,}0369 < a < 1453{,}9631.$$

Пример:

По данным девяти независимых равноточных измерений некоторой физической величины найдены выборочная средняя $\overline{x_в} = 30{,}1$ и исправленное среднее квадратическое отклонение $S = 6$. Оценить истинное значение измеряемой величины с помощью доверительного интервала с надежностью $\gamma = 0{,}99$. Предполагается, что результаты измерений распределены нормально.

Решение.

$$\Delta = t_{n-1} \cdot \frac{S}{\sqrt{n}}.$$

По таблице значений $t_\gamma = t(\gamma; n)$ по $\gamma = 0{,}99$ и $n = 9$ определяем $t_\gamma = 3{,}36$.

Тогда

$$\Delta = 3{,}36 \cdot \frac{6}{\sqrt{9}} = 6{,}72.$$

Тогда искомый доверительный интервал

$$30{,}1 - 6{,}72 < a < 30{,}1 + 6{,}72,$$

$$20{,}38 < a < 36{,}82.$$

2. Доверительный интервал для среднего квадратического отклонения.

Доверительный интервал для среднего квадратического отклонения σ имеет вид:

$$\left(S(1-q);\ S(1+q)\right) \text{ или } S(1-q) < \sigma < S(1+q) \text{ при } q < 1,$$

$$\left(0;\ S(1+q)\right) \text{ или } 0 < \sigma < S(1+q) \text{при } q > 1,$$

где q находят по таблице значений $q = q(\gamma; n)$ по заданным γ и n.

Пример:

По данным выборки объема $n = 16$ из генеральной совокупности найдено исправленное среднее квадратическое отклонение $S = 1$ нормально распределенного количественного признака. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95.

Решение.

По данным $\gamma = 0{,}95$ и $n = 16$ по таблице значений $q = q(\gamma; n)$ найдем $q = 0{,}44$. Так как $q < 1$, то получим доверительный интервал

$$1(1 - 0{,}44) < \sigma < 1(1 + 0{,}44),$$

$$0{,}56 < \sigma < 1{,}44 .$$