



РЕПРЕЗЕНТАТИВНОСТЬ ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК

ОСНОВЫ
МАТЕМАТИЧЕСКОЙ
БИОСТАТИСТИКИ

Репрезентативность

- соответствие выборочных параметров характеристикам генеральной совокупности

Ошибки оценки генеральных параметров

- не учитываемые статистическими (биометрическими методами) и учитываемые статистическими методами
- сознательные и неосознанные
- случайные и систематические

Ошибки, не учитываемые статистическими методами

- ошибки, возникающие при нарушении методики проведения эксперимента (сбора данных, анализа результатов, не соблюдение принципа пар-аналогов)
- ошибки точности и погрешности измерения (недостаточная или избыточная точность)
- ошибки внимания
- ошибки типичности (комплектования выборки не типичными объектами)

Ошибки, учитываемые статистическими методами

- ошибки обусловленные статистическим методом (статистические ошибки, ошибки репрезентативности) — закономерное следствие применения выборочного метода исследований
- чем больше объем выборки, тем меньше статистическая ошибка

Ошибка средней арифметической

$$m_M = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

- σ — стандартное отклонение
- n — количество животных в выборке
- N — количество животных в генеральной совокупности

Ошибка средней арифметической

$$\text{при } N \rightarrow \infty \quad m_M = \frac{\sigma}{\sqrt{n}}$$

- σ — стандартное отклонение
- n — количество животных в выборке

Показатель точности оценок

$$C_s = \frac{m_M}{M} \cdot 100$$

- M — средняя арифметическая
- m_M — ошибка средней арифметической

Ошибка показателя точности оценок

$$m_{C_s} = C_s \sqrt{\frac{1}{2n} + \left(\frac{C_s}{100}\right)^2}$$

- n — объем выборки
- C_s — показатель точности оценок

Ошибка медианы

$$t_{Me} = t_M \sqrt{\frac{\pi}{2n}} = 1,25331 t_M \sqrt{n}$$

- t_M — ошибка средней арифметической
- π — 3,1415926535897932384626433832795

Ошибка дисперсии

$$m_{\sigma^2} = \frac{\sigma^2}{\sqrt{2n}}$$

- σ^2 — дисперсия
- n — количество животных в выборке

Ошибка стандартного отклонения

$$m_{\sigma} = \frac{\sigma}{\sqrt{2n}}$$

- σ — среднее квадратическое отклонение
- n — объем выборки

Ошибка коэффициента вариации

$$m_{Cv} = \frac{Cv}{\sqrt{n-1}} \sqrt{\frac{1}{2} + \left(\frac{Cv}{100}\right)^2} \approx \sqrt{\frac{Cv^2}{2n}}$$

- Cv — коэффициент вариации
- n — объем выборки

Ошибка выборочной доли

$$m_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

- p и q — доли особей, имеющие и не имеющие данный альтернативный признак
- n — объем выборки

Ошибка выборочной доли с поправкой Пирсона

$$m_p = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

- p — доля особей, имеющих данный альтернативный признак
- n — объем выборки
- N — объем генеральной совокупности

Ошибка выборочной доли, выраженной в процентах

$$m_p = \sqrt{\frac{p(100 - p)}{n}}$$

- p — доля особей, имеющих данный альтернативный признак
- n — объем выборки

Ошибка абсолютной частоты

$$m_p = \sqrt{\frac{m(n - m)}{n}} = \sqrt{np(1 - p)}$$

- m — количество особей, имеющих данный альтернативный признак
- n — объем выборки
- p — доля особей, имеющих данный альтернативный признак
- n — объем выборки

Интервальная оценка репрезентативности

- Интервальная оценка — определяется двумя числами (в противоположность точечной)
- Θ (тета) — параметр
- Θ^* — его статистическая характеристика
- δ (дельта) — показатель точности оценки (максимальная разница между параметром и статистической характеристикой ($\Theta - \Theta^*$))
- γ (гамма) — надежность оценки (доверительная вероятность)

Точность оценки

- чем меньше разница $|\Theta - \Theta^*|$ между параметром и статистической характеристикой (генеральной и выборочной величиной) тем выше точность
- если $\delta > 0$ и $|\Theta - \Theta^*| < \delta$, то чем меньше δ , тем точнее оценка

Надежность (доверительная вероятность)

- надежность оценки параметра θ по статистической характеристике θ^* — вероятность при которой разница между ними меньше показателя точности оценки:

$$\gamma = P[|\theta - \theta^*| < \delta]$$

$$\gamma = P[\theta^* - \delta < \theta < \theta^* + \delta]$$

Доверительный интервал

- интервал, покрывающий неизвестный параметр с заданной надежностью

$$(\theta^* - \delta, \theta^* + \delta)$$

- доверительные границы — случайные концы $(\theta^* - \delta$ и $\theta^* + \delta)$ доверительного интервала

Принцип практической невозможности маловероятных событий

- Если случайное событие имеет очень малую вероятность, то практически можно считать, что в единичном испытании это событие не наступит (маловероятные события считают практически невозможными)
- Если случайное событие имеет вероятность, очень близкую к единице, то практически можно считать, что в единичном испытании это событие наступит (события с высокой вероятностью принимают за достоверные)

Доверительная вероятность

- достаточная вероятность для уверенного суждения о генеральных параметрах на основании выборочных данных
- значение вероятности, используемое для определения доверительного интервала

Доверительные вероятности и нормированные отклонения

P	t
0,95	1,96
0,99	2,58
0,999	3,29

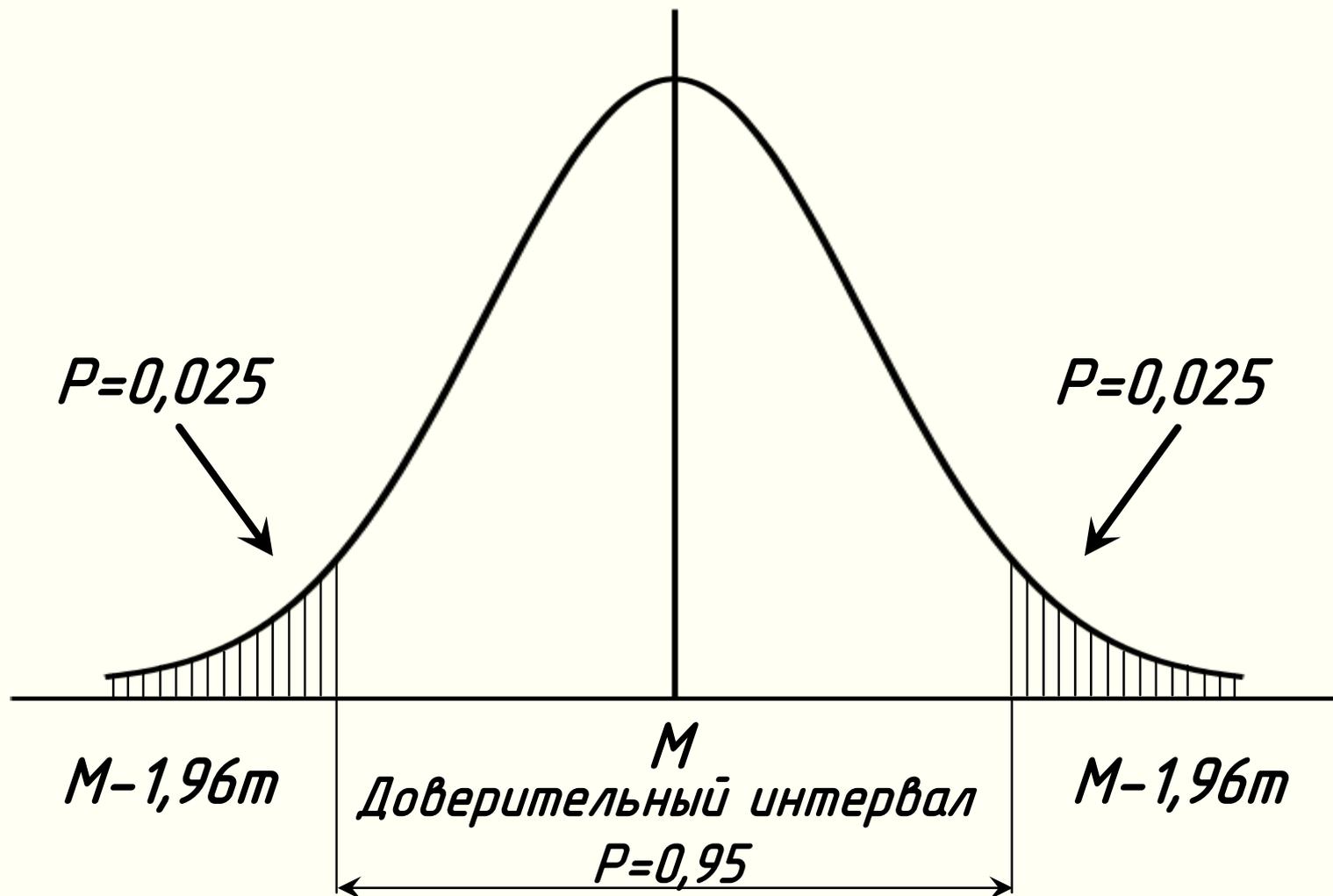
$$t = \frac{x - M}{\sigma}$$

Уровень значимости

$$\alpha = 1 - P$$

- площадь под нормальной кривой выборочного распределения, выходящая за пределы той его части, которая включает P процентов этой площади

Доверительный интервал



Доверительного интервал для генеральной средней

$$M - tm_M \leq \mu \leq M + tm_M$$

- μ — средняя арифметическая генеральной совокупности
- M — средняя арифметическая выборки
- t — нормированное отклонение
- m_M — ошибка M

Доверительный интервал для дисперсии

$$s^2 - tm_{s^2} \leq \sigma^2 \leq s^2 + tm_{s^2}$$

- σ^2 — дисперсия по генеральной совокупности
- s^2 — дисперсия по выборке
- tm_{s^2} — ошибка дисперсии

Доверительный интервал для стандартного отклонения

$$\sqrt{s^2 - tm_{s^2}} \leq \sigma \leq \sqrt{s^2 + tm_{s^2}}$$

- σ — стандартное отклонение по генеральной совокупности
- s^2 — дисперсия по выборке
- tm_{s^2} — ошибка дисперсии

Доверительны интервал коэффициента вариации

$$\frac{Cv}{1 + K\sqrt{1 + 2Cv^2}} \leq Cv \leq \frac{Cv}{1 - K\sqrt{1 + 2Cv^2}}$$

- Cv — коэффициент вариации

$$K = \frac{t}{\sqrt{2(n - 1)}}$$

- t — нормированное отклонение
- n — количество животных в выборке

Доверительны интервал для доли

$$p - tm_p \leq p \leq p + tm_p$$

- p — выборочная доля
- m_p — ошибка выборочной доли
- t — нормированное отклонение

Критерий достоверности выборочных величин

$$t_X = \frac{X}{m_X}$$

- X — выборочная величина (M, σ, Cv)
- m_X — ошибка выборочной величины (m_M, m_σ, m_{Cv})

Значения критерия достоверности по Стьюденту-Фишеру

Число степеней свободы (ν)	Доверительная вероятность (P) Уровень значимости (α)		
	P = 0.95 $\alpha = 0.05$	P = 0.99 $\alpha = 0.01$	P = 0.999 $\alpha = 0.001$
2	4.303	9.925	31.598
3	3.182	5.841	12.941
4	2.776	4.604	8.610
35	2.030	2.724	3.591
45	2.014	2.690	3.520
100	1.982	2.625	3.390
>120	1.960	2.5758	3.2905

Требования к порогу достоверности

Уровень достоверности	Характер изучаемого вопроса
$P = 0.95$	Биологически вопросы, научно-хозяйственные опыты, поисковые эксперименты
$P = 0.99$	Вопросы экономического и производственного характера, по которым будут даны рекомендации, детализация биологических явлений и закономерностей
$P = 0.999$	Изучение действия опасных для жизни препаратов и заключение о дозах безвредности

Количество степеней свободы

- $\nu(k) = (n_1 - 1)$
- если $s_1^2 = s_2^2$ $k = (n_1 - 1) + (n_2 - 1)$
- если $s_1^2 \neq s_2^2$, $n_1 = n_2$ $k = n - 1 + \frac{2n-2}{s_1^2/s_2^2 + s_2^2/s_1^2}$
- если $s_1^2 \neq s_2^2$, $n_1 \neq n_2$

$$k = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right) / \left[\frac{(s_1^2/n_1)^2}{n_1 + 1} + \frac{(s_2^2/n_2)^2}{n_2 + 1} \right] - 2$$

Достоверность разницы средних величин

$$t_D = \frac{|M_1 - M_2|}{\sqrt{m_1^2 + m_2^2}}$$

- M_1, M_2 — средние арифметические, сравниваемых выборок
- m_1, m_2 — ошибки средних арифметических

Достоверность разницы средних величин при разных дисперсиях

$$t_D = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- M_1, M_2 — средние арифметические,
- σ_1^2, σ_2^2 — дисперсии
- n_1, n_2 — количество животных

Достоверность разницы между долями

$$t_{D_p} = \frac{D_p}{m_{D_p}}$$

- D_p — разница выборочных долей
- m_{D_p} — ошибка разницы выборочных долей

Ошибка разницы между долями

$$m_{D_p} = \sqrt{m_{p_1}^2 + m_{p_2}^2}$$

• m_{p_1} и m_{p_2} — ошибки выборочных долей, сравниваемых выборок

Ошибка разницы между долями (если $\Delta p < 25\%$)

$$m_{D_p} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$m_{D_p} = \sqrt{\frac{p_1(100-p_1)}{n_1} + \frac{p_2(100-p_2)}{n_2}}$$

$$m_{D_p} = \sqrt{\frac{m_1(n_1-m_1)}{n_1} + \frac{m_2(n_2-m_2)}{n_2}}$$

Ошибка разницы между долями (если $\Delta p > 25\%$)

$$m_{D_p} = \sqrt{p(1-p) \frac{(n_1 + n_2)}{n_1 n_2}}$$

$$p = \frac{(p_1 n_1 + p_2 n_2)}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2}$$

$$m_{D_p} = \sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2} \right) \frac{(n_1 + n_2)}{n_1 n_2}}$$