# 5. Генеральная и выборочная средняя. Оценка генеральной средней по выборочной средней. Устойчивость выборочных средних. Свойства выборочной средней (с доказательством)

## 1. Основные определения

**Генеральная средняя ($M(X)$)** — это среднее арифметическое значений признака всей генеральной совокупности. В вероятностном смысле это математическое ожидание случайной величины $X$:

$$M(X) = a$$

**Выборочная средняя ($\overline{x}$)** — это среднее арифметическое значений выборки. Если получена выборка объема $n$ со значениями $x_1, x_2, ..., x_n$, то она вычисляется по формуле:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

## 2. Оценка генеральной средней по выборочной средней

Параметры генеральной совокупности (в том числе $M(X)$) обычно неизвестны. Задача статистики — найти их точечные оценки по выборке.
В качестве наилучшей точечной оценки генеральной средней принимают выборочную среднюю $\overline{x}$.

Это подтверждается основными методами нахождения оценок:

- **Метод моментов:** Приравнивая теоретическое матожидание выборочному среднему, получаем $\overline{x}$.

- **Метод максимального правдоподобия:** Для нормального распределения оценкой параметра $a$ является $\overline{x}$.
- **Метод наименьших квадратов:** Величина, минимизирующая сумму квадратов отклонений от данных выборки, также равна $\overline{x}$.

# 3. Свойства выборочной средней (с доказательством)

Чтобы оценка считалась качественной, она должна быть **состоятельной**, **несмещенной** и **эффективной**.
Рассмотрим свойства $\overline{x}$ для случая **повторной выборки**.

*Условие:* При повторной выборке элементы отбираются независимо, поэтому случайные величины $X_i$ имеют те же характеристики, что и генеральная совокупность:

$$M(X_i) = M(X), \quad D(X_i) = D(X) = \sigma^2$$

## А) Состоятельность

**Определение:** Оценка состоятельна, если при $n \to \infty$ она стремится по вероятности к истинному параметру (то есть с ростом выборки мы всё ближе к истине).

**Доказательство:**
Согласно закону больших чисел (теорема Чебышева), для независимых случайных величин с одинаковым матожиданием и ограниченной дисперсией среднее арифметическое сходится по вероятности к математическому ожиданию:

$$\lim_{n \to \infty} P\left(|\overline{x} - M(X)| < \epsilon\right) = 1$$

Следовательно, $\overline{x}$ — состоятельная оценка.

## Б) Несмещенность

**Определение:** Оценка несмещенная, если ее математическое ожидание равно оцениваемому параметру ($M(\overline{x}) = M(X)$). Это гарантирует отсутствие систематической ошибки (завышения или занижения).

**Доказательство:**
Найдем математическое ожидание выборочной средней:

$$M(\overline{x}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

Используем свойства матожидания (выносим константу $1/n$ и заменяем матожидание суммы суммой матожиданий):

$$M(\overline{x}) = \frac{1}{n}\sum_{i=1}^{n} M(X_i)$$

Так как $M(X_i) = M(X)$:

$$M(\overline{x}) = \frac{1}{n} \cdot n \cdot M(X) = M(X)$$

**Вывод:** Оценка несмещенная.

## В) Эффективность (для нормального распределения)

**Определение:** Несмещенная оценка эффективна, если она имеет наименьшую возможную дисперсию среди всех оценок.

**Доказательство:**

1. Нижняя граница дисперсии для оценки матожидания (неравенство Рао-Крамера-Фреше) равна $\frac{\sigma^2}{n}$.
2. Найдем дисперсию выборочной средней $\overline{x}$:

$$D(\overline{x}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

   Так как $X_i$ независимы, дисперсия суммы равна сумме дисперсий:

$$D(\overline{x}) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

3. Так как дисперсия $\overline{x}$ совпадает с минимальной границей, оценка является эффективной.

# 4. Устойчивость выборочных средних

Устойчивость оценки обеспечивается её свойствами:

1. **Состоятельность:** Гарантирует, что при большом объеме данных погрешность стремится к нулю.
2. **Малая дисперсия:** Дисперсия $D(\overline{x}) = \frac{\sigma^2}{n}$ убывает с ростом $n$. Это означает, что выборочные средние, полученные по разным выборкам одного объема, будут мало отличаться друг от друга и группироваться около истинного среднего.

**Примечание (случай бесповторной выборки):**
Если выборка бесповторная (элементы не возвращаются), результаты наблюдений становятся зависимыми.
Для этого случая:

- $\overline{x}$ остается **состоятельной** и **несмещенной** оценкой.
- Дисперсия вычисляется иначе и она меньше, чем при повторной выборке (оценка точнее):

$$D(\overline{x}) = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)$$

*(Теорема принимается без доказательства).*