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

1. Основные определения

Генеральная средняя — это математическое ожидание $M(X)$ случайной величины X (генеральной совокупности).

Выборочная средняя ($\overline{x}$) — это среднее арифметическое значений выборки. Пусть из генеральной совокупности объема N отобрана случайная выборка $X_1, X_2, ..., X_n$, где X_k — случайная величина, выражающая значение признака у k-го элемента выборки.
Тогда выборочная средняя определяется формулой:

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

2. Оценка генеральной средней по выборочной средней

В качестве точечной оценки генеральной средней (параметра $M(X)$) рассматривается её статистический аналог — выборочная средняя $\overline{x}$.

Для обоснования качества этой оценки проверяются свойства: состоятельность, несмещенность и эффективность.

Доказательства проводятся для случая **повторной выборки**.
Условие: При повторной выборке каждый отобранный элемент возвращается, поэтому вероятность извлечения не меняется. Случайные величины X_k являются **независимыми** и

имеют тот же закон распределения, что и генеральная совокупность.
Следовательно, их характеристики совпадают с генеральными:

$$M(X_1) = M(X_2) = ... = M(X_n) = M(X)$$

$$D(X_1) = D(X_2) = ... = D(X_n) = D(X)$$

3. Свойства выборочной средней (с доказательством)

Теорема 3.3. Выборочная средняя $\overline{x}$ повторной выборки есть состоятельная и несмещенная оценка генеральной средней $M(X)$, причем ее дисперсия равна $\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}$.

А) Доказательство состоятельности

Определение: Оценка $\hat{\theta}$ называется состоятельной, если она стремится по вероятности к оцениваемому параметру θ при $n \to \infty$:

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

Доказательство:
Доказательство опирается на теорему Чебышева (Закон Больших Чисел). Для независимых случайных величин с конечными дисперсиями выполняется равенство:

$$\lim_{n \to \infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} M(X_k)\right| < \varepsilon\right) = 1$$

Преобразуем выражение. Так как $M(X_k) = M(X)$ для всех $k = 1, ..., n$, то сумма математических ожиданий:

$$\sum_{k=1}^{n} M(X_k) = \sum_{k=1}^{n} M(X) = n \cdot M(X)$$

Тогда среднее арифметическое математических ожиданий:

$$\frac{1}{n}\sum_{k=1}^{n} M(X_k) = \frac{1}{n} \cdot n \cdot M(X) = M(X)$$

Подставим полученный результат и формулу выборочной средней $\overline{x}$ в предел из теоремы Чебышева. В результате получим:

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

Сравнивая это выражение с определением состоятельности (где $\hat{\theta} = \overline{x}$, а $\theta = M(X)$), заключаем, что $\overline{x}$ — **состоятельная оценка** генеральной средней.

Б) Доказательство несмещенности

Определение: Оценка называется несмещенной, если ее математическое ожидание равно оцениваемому параметру ($M(\hat{\theta}) = \theta$) при любом объеме выборки.

Доказательство:
Вычислим математическое ожидание выборочной средней при любом фиксированном n:

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

Используем свойства математического ожидания (константу $1/n$ можно вынести за знак матожидания, матожидание суммы равно сумме матожиданий):

$$M(\overline{x}) = \frac{1}{n}\sum_{k=1}^{n} M(X_k)$$

Так как $M(X_k) = M(X)$ (условие повторной выборки), то:

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

Таким образом, справедливо равенство $M(\overline{x}) = M(X)$. Следовательно, оценка является **несмещенной**.

В) Доказательство эффективности

Определение: Несмещенная оценка называется эффективной, если она имеет минимальную дисперсию среди всех возможных оценок.

Доказательство (для нормального распределения):
Минимальная граница дисперсии для несмещенной оценки математического ожидания (вычисляется с помощью неравенства Рао-Крамера-Фреше) в случае нормального

распределения равна $\frac{\sigma^2}{n}$.
Для доказательства эффективности $\overline{x}$ надо показать, что её дисперсия совпадает с этой границей.

Вычислим дисперсию выборочной средней:

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

Используем свойства дисперсии (константа выносится в квадрате, дисперсия суммы независимых величин равна сумме дисперсий):

$$D(\overline{x}) = \frac{1}{n^2} D\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k)$$

Так как $D(X_k) = D(X) = \sigma^2$ для всех k:

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

Дисперсия $\overline{x}$ совпадает с минимально возможной границей, значит, $\overline{x}$ — **эффективная оценка**.

4. Устойчивость выборочных средних

Устойчивость оценки обеспечивается доказанными выше свойствами:

1. **Состоятельность:** Равенство $\lim_{n\to\infty} P(|\overline{x} - M(X)| < \varepsilon) = 1$ означает, что чем больше число наблюдений n, тем больше уверенность (вероятность) в незначительном отклонении оценки от неизвестного параметра. Чем больше объем информации, тем «ближе мы к истине».
2. **Малая дисперсия (Эффективность):** Формула $\sigma^2_{\overline{x}} = \frac{\sigma^2}{n}$ показывает, что с увеличением объема выборки n дисперсия (разброс) выборочной средней уменьшается. Это обеспечивает устойчивость оценки относительно генеральной средней.

Примечание (Случай бесповторной выборки):
В случае бесповторной выборки величины X_k становятся зависимыми.
Теорема 3.4: Выборочная средняя бесповторной выборки есть состоятельная и несмещенная оценка генеральной средней $M(X)$, причем ее дисперсия равна:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(1 - \frac{N-n}{N-1}\right) \approx \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)$$

(Теорема принимается без доказательства).