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). Обычно обозначается как параметр a или θ.

Выборочная средняя ($\overline{x}$) — это среднее арифметическое значений выборки. Если 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, то выборочная средняя вычисляется по формуле:

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

2. Оценка генеральной средней по выборочной средней

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

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

Доказательства приведем для случая **повторной выборки**.
Условие: При повторной выборке результаты наблюдений $X_1, X_2, ..., X_n$ можно рассматривать

как независимые случайные величины, имеющие тот же закон распределения, что и генеральная совокупность. Следовательно:

$$M(X_k) = M(X) = a$$

$$D(X_k) = D(X) = \sigma^2$$

3. Свойства выборочной средней (с доказательством)

Теорема: Выборочная средняя $\overline{x}$ повторной выборки есть состоятельная, несмещенная и эффективная оценка генеральной средней $M(X)$.

А) Доказательство состоятельности

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

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

Доказательство:
Опирается на теорему Чебышева (Закон Больших Чисел). Для суммы независимых случайных величин с одинаковым математическим ожиданием и ограниченной дисперсией справедливо утверждение:

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

Так как $M(X_k) = M(X)$, то среднее арифметическое математических ожиданий равно:

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

Подставляя это в предел, получаем:

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

Это означает, что $\overline{x}$ является состоятельной оценкой.

Б) Доказательство несмещенности

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

Доказательство:
Найдем математическое ожидание выборочной средней:

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

Используя свойства математического ожидания (константу $1/n$ можно вынести, матожидание суммы равно сумме матожиданий):

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

Так как $M(X_k) = M(X)$ для всех k:

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

Вывод: $M(\overline{x}) = M(X)$, следовательно, оценка несмещенная.

В) Доказательство эффективности (для нормального распределения)

Определение: Несмещенная оценка $\hat{\theta}$ называется эффективной, если она имеет минимальную дисперсию среди всех возможных несмещенных оценок данного параметра.

Неравенство Рао-Крамера:
Нижняя граница дисперсии для любой несмещенной оценки определяется неравенством:

$$D(\hat{\theta}) \geq \frac{1}{n \cdot I(\theta)}$$

где $I(\theta)$ — информация Фишера. Для нормального распределения $N(a, \sigma^2)$ информация о

параметре a равна $I(a) = \frac{1}{\sigma^2}$.
Следовательно, минимально возможная дисперсия равна:

$$D_{min} = \frac{1}{n \cdot \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

Доказательство:
Вычислим дисперсию выборочной средней $\overline{x}$:

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

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

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

Так как $D(X_k) = \sigma^2$:

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

Так как дисперсия оценки совпадает с нижней границей Рао-Крамера ($D(\overline{x}) = D_{min}$), то $\overline{x}$ — эффективная оценка генеральной средней нормального распределения.

4. Устойчивость выборочных средних

В прикладной статистике под устойчивостью (или **робастностью**) понимают способность оценки сохранять свои свойства и точность при небольших отклонениях от предполагаемой модели распределения, в том числе при наличии **выбросов** (аномальных значений).

Является ли выборочная средняя устойчивой?
Нет, выборочная средняя $\overline{x}$ **не является устойчивой (робастной)** оценкой.
Это связано с тем, что она линейно зависит от всех значений выборки. Появление даже одного экстремально большого значения (выброса) может существенно сместить $\overline{x}$, исказив представление о центре распределения.

Условия для использования выборочной средней:
Выборочная средняя является оптимальной оценкой только при выполнении следующих условий:

1. **Нормальность распределения:** Генеральная совокупность должна иметь нормальный закон распределения.
2. **Симметричность:** Распределение должно быть симметричным.
3. **Отсутствие выбросов:** В выборке не должно быть «загрязняющих» данных.

Если эти условия не выполняются (например, есть тяжелые хвосты или выбросы), целесообразнее использовать более устойчивые характеристики положения, например, выборочную медиану.

Примечание (Случай бесповторной выборки):
Если выборка бесповторная, результаты наблюдений становятся зависимыми.
В этом случае $\overline{x}$ остается состоятельной и несмещенной, но ее дисперсия меньше (оценка точнее):

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

(Теорема принимается без доказательства).