2. Статистическое оценивание параметров. Методы нахождения точечных оценок параметров: метод моментов К. Пирсона, Метод максимального правдоподобия Р. Фишера, метод наименьших квадратов.

Опр

Точечная статистистическая оцена $\hat{\theta}$ - приближенное значение этого параметра, полученное по выборке, т.е $\theta \approx \hat{\theta}$

- точечная - значит что оценка это либо число, либо точка на числовой прямой
- Статистическая - значит, что оценка получается по собранной исследователем статистике

Точечная оценка вычисляется следующим способом

$$\hat{\theta} = f(x_1, x_2, ...x_n)$$

где:

- $\hat{\theta}$ - сама оценка
- $x_1, x_2...x_n$ - наблюдения
- f(x) - представляет собой алгоритм вычисления оценки

Найти ошибку приближения не получится, т.к в большинстве случаев мы не знаем истинное значение

Но если считать, что $x_1, x_2, ...x_n$ это просто какие-то значения случайной велечины X, то $\hat{\theta}$ это функция от случайной велечины, т.е тоже случайная велечина. А для них можно считать Мат.Ожидание и Дисперсию, а также закон распеределения

Если рассматривать $\hat{\theta}$ как С.В, то можно предъявить свойства, обладая которыми, можно говорить, что приблежение хорошего качества. Эти свойства

- Состоятельность
- Несмещенность
- Эффективность

Опр

Состоятельная оценка $\hat{\theta}$ - означает, что с ростом n, её предел по вероятность равен оцениваемому параметру

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

Если короче: Чем больше наблюдений, тем меньше отклонений $\hat{\theta}$ от θ

Оценка обязательно должна быть состоятельной

Опр

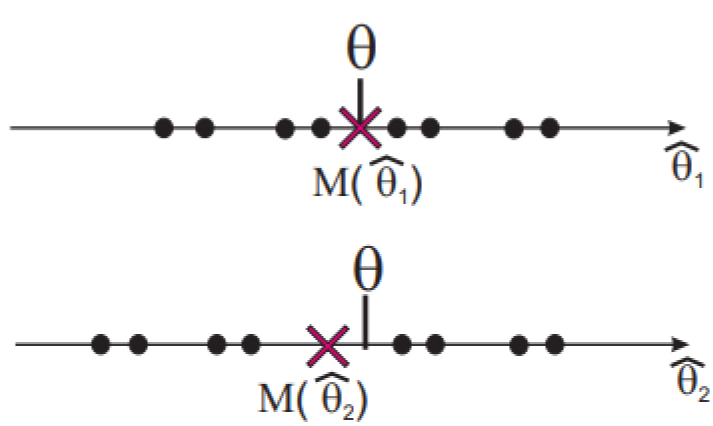
Несмещенная оценка $\hat{\theta}$ - означает, что её Мат.Ожидание равно оцениваемому параметру θ при любом фиксированном n

$$M(\widehat{\theta}) = \theta.$$

У смещенной оценки это будет не так, т.к она обладает смещением равным $M(\hat{\theta}) - \theta$

Если наглядно

- первая оценка - несмещенная
- вторая смещена



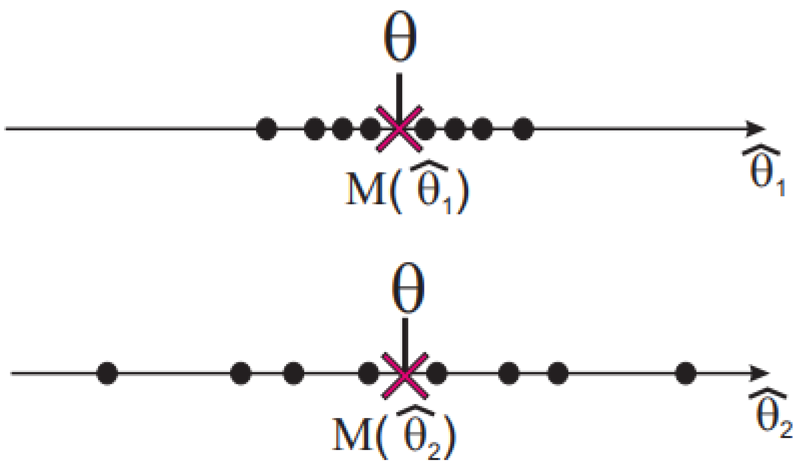
Несмещенность гарантирует отсутствие систематических ошибок при оценивании

ОПР

Эффективная оценка $\hat{\theta}$ - имеет минимальную дисперсию среди других выборочных оценок

$$D(\widehat{\theta}) = \min_{\widehat{\theta}_i \in \Theta} D(\widehat{\theta}_i),$$

Если наглядно



Неравенство Рао Крамера Фреше

Вставлю тут фотку, т.к не помню, чтобы мы его использовали при решении задач. Блестните знаниями, если захотите 😃

Вообще эта штуковина нужна для того, чтобы находить минимум из дисперсий оценок. Зная минимум можно сравнить его с дисперсией получившейся оценки и понять на сколько она хорошая

3.2. Неравенство Рао-Крамера-Фреше

Пусть $\varphi(x, \theta)$ — плостность вероятностей случайной величины X, если X — непрерывная случайная величина, и $\varphi(x_i, \theta) = P(X = x_i, \theta)$ — вероятность того, что с.в. X приняла значение x_i, если X - дискретная случайная величина.

Если $\varphi(x, \theta)$ удовлетворяет условиям регулярности:

1. область возможных значений исследуемой случайной величины, в которой
   $\varphi(x, \theta) \neq 0$, не зависит от θ;

2. в выражении $M(\theta) = \int \theta \cdot L(x_1, \ldots, x_n, \theta) dx_1 dx_2 \ldots dx_n$ и тождестве $\int L(x_1, \ldots, x_n, \theta) dx_1 dx_2 \ldots dx_n = 1$ допустимо дифференцирование по θ под знаком интеграла;

3. величина $I(\theta, x) \neq 0$,

тогда для любой оценки $\widehat{\theta}$ неизвестного параметра θ имеет место неравенство Рао-Крамера-Фреше:

$$D(\widehat{\theta}) \geq \frac{1}{n \cdot I(\theta, x)}, \tag{3.5}$$

где $D(\widehat{\theta})$ - дисперсия оценки $\widehat{\theta}$ параметра θ; $I(\theta, x)$ - количество информации Фишера о параметре θ, содержащееся в единичном наблюдении:

$$I(\theta, x) = M \left[(\ln \varphi(x, \theta))'_\theta \right]^2. \tag{3.6}$$

По правилу дифференцирования сложной функции

$$(\ln y(x))'_x = \frac{y'(x)}{y(x)}.$$

Поэтому в случае дискретной случайной величины

$$I(\theta, x) = \sum_{i=1}^{n} \left[\frac{\varphi'(x_i, \theta)}{\varphi(x_i, \theta)} \right]^2 \cdot \varphi(x_i, \theta).$$

В случае непрерывной случайной величины

$$\int^{\infty} \left[\varphi'(x, \theta) \right]^2$$

$$\overline{i=1} \lfloor \varphi(x_i, \theta) \rfloor$$

В случае непрерывной случайной величины

$$I(\theta, x) = \int_{-\infty}^{\infty} \left[\frac{\varphi'(x,\theta)}{\varphi(x,\theta)}\right]^2 \cdot \varphi(x,\theta).$$

Неравенство (5.11) позволяет найти тот минимум дисперсии, который должна иметь дисперсия оценки $D(\widehat{\theta})$, чтобы быть эффективной оценкой, т.е.

$$D_{\text{эф}}(\widehat{\theta}) = \min D(\widehat{\theta}) = \frac{1}{n \cdot I(\theta, x)}.$$

Свойства основных выборочных оценок

Частость

- Частость - точечная оценка вероятности события
  - Выборочная доля - точечная оценка генеральной доли

N - объем выборки генеральной совокупности.
M - кол-во эл-тов выборки генеральной совокупности с нужным свойством A
p = $\frac{M}{N}$ - генеральная доля или вероятность события A
w = $\frac{m}{n}$ - выборочная доля,где

- m - кол-во эл-тов взятых из выборки генеральной совокупности с нужным св-ом A, т.е m < M
- n - кол-во каких эл-тов взятых из исходной генеральной совокупности. n < N

Рассмотрим повторную выборку

Т.к выборка повторная, то вероятность извлечь элемент не зависит от ранее извлеченных эл-тов, т.е **события являются независимыми**

число m - можно представить как сумму n св X_i, где:

- X_i = 1,если X_i - обладает свойством A
- X_i = 0,если X_i - не обладает свойством A

i - это номер наблюдения

Таким образом получаем, что X_i - распределена по Закону Бернулли, а значит:

- $M(X_i) = n \cdot p$

- $D(X_i) = n \cdot p \cdot (1-p)$

Дальше приводится д-во, что выборочная доля это эффективная, состоятельная и несмещенная оценка

ятности) $p = \frac{}{N}$.

а) Состоятельность. Для испытаний Бернулли справедлива теорема Бернулли, согласно которой для любого $\varepsilon > 0$ выполняется равенство:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1$$

Сравнивая полученное выражение с определением свойства состоятельности (m/n - это $\widehat{\theta}$, p - это θ), заключаем, что $\widehat{p} = m/n$ — состоятельная оценка генеральной доли (вероятности) p.

б) Несмещенность. Зафиксируем число n испытаний Бернулли. Вычислим математическое ожидание оценки $\widehat{p}$:

$$M(\widehat{p}) = M\left(\frac{m}{n}\right) = \frac{1}{n}M(m) = \frac{1}{n} \cdot np = p$$

Таким образом, $M(\widehat{p}) = p$. По определению свойства несмещенности получаем, что $\widehat{p} = m/n$ — несмещенная оценка параметра p.

в) Эффективность. Найдем дисперсию частости:

$$D(\widehat{p}) = D\left(\frac{m}{n}\right)_{\text{n=const}} = \frac{1}{n^2}D(m) = \frac{1}{n^2}n\ p\ q = \frac{p\ q}{n} = \frac{p(1-p)}{n}$$

Так как $D(\widehat{p})$ совпадает с минимальной границей, вычисленной с помощью неравенства Рао-Крамере-Фреше (см. Пример 3.1 предыдущего раздела), то выборочная доля $\widehat{p}$ (частость) является эффективной оценкой генеральной доли (вероятности) p.

Теперь рассмотрим **Бесповторную выборку**

Теперь св X_i - будут зависимыми

X_i оказываются подчинены гипергеометрическому закону распределения, следовательно:

- $$M(X_i) = n \cdot \frac{M}{N}$$

- $$D(X_i) = n \cdot \frac{\frac{M}{N} \cdot (1 - \frac{M}{N}) \cdot (N-n)}{N-1}$$

это состоятельная, несмещенная оценка генеральной доли

Выборочное среднее $\bar{X} = \frac{1}{n} \cdot \sum X_i$

- Среднее по выборке - это точечная оценка Мат.Ожидания

Пусть из генеральной совокупности объема N отобрана случайная выборка $X_1, X_2, \ldots, X_k, \ldots, X_n$, где X_k — случайная величина, выражающая значение признака у k-го элемента выборки ($k = 1, \ldots, n$).

Рассмотрим в качестве оценки генеральной средней выборочную среднюю

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

.

При повторной выборке

X_i - снова независимые

1) Повторная выборка. Закон распределения для каждой случайной величины X_k имеет вид:

x_i	x_1	x_2	$\ldots$	x_i	$\ldots$	x_m
p_i	$\frac{N_1}{N}$	$\frac{N_2}{N}$	$\ldots$	$\frac{N_i}{N}$	$\ldots$	$\frac{N_m}{N}$

Так как из общего числа N элементов генеральной совокупности N_i элементов имеют значение признака x_i, и выборка повторная (одинаковость условий наблюдения), то вероятность $p(X_k = x_i) = \dfrac{N_i}{N}$. То есть закон распределения каждой случайной величины X_k - один и тот же, такой же как и у случайной величины X. Это означает, что

$$M(X_1) = M(X_2) = \ldots = M(X_n) = M(X) \qquad (*)$$
$$D(X_1) = D(X_2) = \ldots = D(X_n) = D(X) \qquad (**)$$

далее д-во, что среднее по выборке это состоятельная, несмещенная оценка генеральной средней с дисперсией $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

а) Состоятельность следует из теоремы Чебышева: для любого $\varepsilon > 0$, если выполняется условие независимости случайных величин (а для повторной выборки это так) и дисперсии $D(X_k)$ - конечны, то выполняется равенство:

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

Так как $M(X_k) = M(X)(k = 1, \ldots, n)$, то

$$\sum_{k=1}^{n} M(X_k) \underset{(*)}{=} \sum_{k=1}^{n} M(X) = nM(X).$$

Тогда $\dfrac{1}{n}\sum\limits_{k=1}^{n} M(X_k) = M(X)$. В результате получим:

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

Сравнивая полученное выражение с определением свойства состоятельности (положив $\widehat{\theta} = \overline{x}$ и $\theta = M(X)$), заключаем, что $\overline{x}$ — состоятельная оценка математического ожидания $M(X)$ или генеральной средней.

б) Несмещенность. Если выполняется условие $(*)$, то при любом фиксированном n имеет место следующая цепочка равенств:

$$M(\overline{x}) = M\left(\frac{1}{n}\sum_{k=1}^{n}(X_k)\right) = \frac{1}{n}\sum_{k=1}^{n} M(X_k) \underset{(*)}{=} \frac{1}{n}\sum_{k=1}^{n} M(X) = \frac{1}{n}nM(X) = M(X).$$

Таким образом, при выполнении условия $(*)$ (одинаковость условий наблюдения) справедливо равенство $M(\overline{x}) = M(X)$. Из определения несмещенности (2)(положив $\widehat{\theta} = \overline{x}$ и $\theta = M(X)$) получаем, что $\overline{x}$ — несмещенная оценка математического ожидания.

в) Эффективность. Доказывается для несмещенных оценок. Мы должны показать, что дисперсия $D(\overline{x}) = \sigma^2_{\overline{x}}$ является минимальной из всех возможных оценок (например, в качестве оценки генеральной средней может быть выбрана мода или медиана). Мы проведем доказательство для случая, когда величина X имеет нормальное распределения, т.е. $X = N(a, \sigma)$, где $a = M(X)$ и $\sigma = \sqrt{DX}$.

Для доказательства эффективности среднего $\overline{x}$ надо показать, что его дисперсия совпадает с минимальной границей, равной в случае нормального распределения σ^2/n (см. пример 2 предыдущей лекции). Вычислим дисперсию $D(\overline{x})$:

$$D(\overline{x}) = D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2} D\left(\sum_{k=1}^{n} X_k\right) \underset{(***)}{=} \frac{1}{n^2}\sum_{k=1}^{n} D(X_k) \underset{(**)}{=} \frac{1}{n^2}\sum_{k=1}^{n} D(X) =$$
$$= \frac{nD(X)}{n^2} = \frac{D(X)}{n} = \sigma^2/n$$

Таким образом, средняя $\overline{x}$ является эффективной оценкой математического ожидания $M(X)$.

Ниже будет рассмотрен пример, подтверждающий свойства выборочной средней.

Повторная выборка

X_i - зависимые СВ

Является состоятельной и несмещенной оценкой генеральной средней

Выборочная дисперсия $D(X_i) = \frac{1}{n} \cdot \sum(X_i - \bar{X})^2$

- является оценкой для дисперсии
- состоятельная и смещенная оценка генеральной дисперсии

Теорема 3.5. Выборочная дисперсия $\widehat{D}(x) = \widehat{\sigma}^2$ повторной и бесповторной выборки есть состоятельная и СМЕЩЕННАЯ оценка генеральной дисперсии σ^2.

а) Состоятельность. Примем без доказательства.

б) Смещенность. В соответствии с правилом вычисления дисперсии $\widehat{\sigma}^2 = \overline{x^2} - \overline{x}^2$. Можно показать, что средняя арифметическая и выборочная дисперсия обладают теми же свойствами, что и математическое ожидание и генеральная дисперсия [???]: если все значения признака уменьшить на одно и то же число C, то средняя уменьшится на это же число, т.е. $\overline{x - C} = \overline{x} - C$, а дисперсия на изменится:

$$\widehat{\sigma}^2 = \widehat{\sigma}^2_x = \widehat{\sigma}^2_{x-C} = \overline{(x-C)^2} - (\overline{x-C})^2 = \overline{(x-C)^2} - (\overline{x} - C)^2.$$

Полагая $C = M(X)$, получим

$$\widehat{\sigma}^2 = \overline{(X - M(X))^2} - (\overline{X} - M(X))^2.$$

1) Повторная выборка.

Для повторной выборки случайные величины $X_1, \dots, X_n$ удовлетворяют свойствам $(*)$,$(**)$ и $(***)$. Найдем математическое ожидание оценки $\widehat{\sigma}^2$:

$$M(\widehat{\sigma}^2) = M\left[\frac{1}{n}\sum_{k=1}^{n}(X_k - M(X))^2\right] - M[(\overline{x} - M(X))^2]$$

Вычислим первое из слагаемых:

$$M\left[\frac{1}{n}\sum_{k=1}^{n}(X_k - M(X))^2\right] = \frac{1}{n}\sum_{k=1}^{n} M(X_k - M(X))^2 = \frac{1}{n}\sum_{k=1}^{n} D(X_k) =$$
$$= \frac{1}{n}\sum_{k=1}^{n} D(X) = \frac{nD(X)}{n} = D(X).$$

Второе слагаемое, с учетом того, что $\overline{x}$ есть несмещенная оценка математического ожидания, т.е. $M(X) = M(\overline{x})$, и того, что дисперсия среднего арифметического $D_{\overline{x}} = \sigma^2/n$, будет

$$M[(\overline{x} - M(X))^2] = M[(\overline{x} - M(\overline{x}))^2] = D(\overline{x}) = \frac{\sigma^2}{n} = \frac{D(X)}{n}$$

Тогда

$$M(\widehat{D}(X)) = M(\widehat{\sigma}^2) = D(x) - \frac{D(X)}{n} = \frac{n-1}{n}\, D(X) = \frac{n-1}{n}\, \sigma^2.$$

Так как $M(\widehat{D}(X)) \neq D(X)$, то выборочная дисперсия в случае повторной выборки является смещенной оценкой генеральной дисперсии.

2) Бесповторная выборка.

Для бесповторной выборки случайные величины $X_1, \dots, X_n$ являются зависимыми. Приведем без доказательства формулу для математического ожидания выборочной дисперсии:

$$M(\widehat{\sigma'}^2) = \frac{n-1}{n}\frac{N}{N-1}\sigma^2 \approx \frac{n-1}{n}\sigma^2,$$

(при больших объемах N генеральной совокупности $N \approx N-1$).

Исправленная выборочная дисперсия $s^2 = \frac{n}{n-1} \cdot \sum(X_i - \bar{X})^2$

- имеет связь с обычной дисперсией $s^2 = \sigma^2 * \frac{n}{n-1}$

Повторная выборка

является состоятельной, несмещенной оценкой генеральной дисперсии

а) Состоятельность. Примем без доказательства.
б) Несмещенность. Вычислим математическое ожидание оценки s^2:

$$M(s^2) = M\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \frac{n}{n-1}M(\hat{\sigma}^2) = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2.$$

Бесповторная выборка

Замечание. Если выборка бесповторная, то случайные величины X_k являются зависимыми и s^2 является смещенной оценкой.

Методы нахождения точечных оценок

Метод моментов

Метод моментов состоит в том, что начальные и центральные **выборочные** моменты приравниваются к **теоретическим** моментам. Из решения уравнения находятся неизвестные параметры

Обычно в методе моментов используют первый начальный момент - математическое ожидание для теоретического момента и среднее арифметическое для выборочного, и второй центральный момент - дисперсия (теоретическая и выборочная, соответственно). Эти моменты используют потому,

Распределение	Вероятность	Осн. числовые характеристики
Биномиальное	$P_n(X=m)=C_n^m p^m\ q^{n-m}$	$M(X)=np; \quad D(X)=np\ q$
Пуассона	$P(X=m)=\dfrac{e^{-\lambda}\cdot\lambda^m}{m!}$	$M(X)=\lambda; \quad D(X)=\lambda$
Геометрическое	$P(X=m)=p\ q^{m-1}$	$M(X)=\dfrac{1}{p}; \quad D(X)=\dfrac{q}{p^2}$
Гипер-геометрическое	$P(X=m)=\dfrac{C_M^m\ C_{N-M}^{n-m}}{C_N^n}$	$M(X)=n\dfrac{M}{N}$ $D(X)=n\dfrac{M}{N-1}\left(1-\dfrac{M}{N}\right)\left(1-\dfrac{n}{N}\right)$

Непрерывные случайные величины Таблица 3.2.

Распределение	Плотность вероятности	Основные числовые характеристики
Нормальное	$f_N(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\ e^{-\dfrac{(x-a)^2}{2\sigma^2}}$	$M(X)=a; \quad D(X)=\sigma^2$
Показательное	$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{при} x\geq 0\\ 0 & \text{при} x<0\end{cases}$	$M(X)=\dfrac{1}{\lambda}; \quad D(X)=\dfrac{1}{\lambda^2}$
Равномерное	$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{при} x\in[a;b]\\ 0 & \text{при} x\notin[a;b]\end{cases}$	$M(X)=\dfrac{a+b}{2};$ $D(X)=\dfrac{(b-a)^2}{12}$
Гамма-распр.	$f(x)=\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}$ $\alpha>0,\ \beta>0,\ x\geq 0$	$M(X)=\alpha\beta$ $D(X)=\alpha\beta^2$

Метод максимального правдоподобия(ММП) Фишера

Пусть есть выборка $X=(x_1,x_2,...x_n)$, полученная из генеральной совокупности по закону $f(x,\theta)$

если x_i - Д.С.В, то $f(x_i,\theta)$ - вероятность p_i, если Н.С.В, то плотность

Понятно, что взависимости от выборки X, распределение $f(x,\theta)$ будет отличаться

Требуется найти такой параметр θ, чтобы вероянтость $f(x,\theta)$ лучше описывала выборку X, т.е была бы максимальной

для этого вводят специальную функцию:

$$L(x,\theta) = f(x_1,\theta) \cdot f(x_2,\theta) \cdot ...f(x_n,\theta)$$

- L - это функция правдопободия
- она работает только если X - независимые случайные велечины

Дальше ищут такое θ, чтобы достигался максимум этой функции

цию $L(x,\theta)$:

$$\widehat{\theta}: \qquad L(x,\widehat{\theta}) \to \max.$$

ищут его с помощью производной, но т.к считать производную произведения(особенно, когда там много множителей) это не для белых людей, то белые люди решили схитрить и считать производную такой функции

$$l = ln(L(x,\theta)) = ln(f(x_1,\theta)) + ln(f(x_2,\theta)) + ...ln(f(x_n,\theta))$$

По свойству получаем, что мы уже считаем производную суммы логорифмов, что значительно проще

ну и всё - теперь находим критические точки и проверяем их на экстремум(выбираем максимальную)

гайдик как находить точку максимума при 2 параметрах

Если оценивается несколько параметров. например два - θ_1 и θ_2, то есть находится точка максимума функции двух переменных — $L(\theta_1, \theta_2)$, то необ-

ходимое условие экстремума — равенство нулю частных производных:

$$\begin{cases} \dfrac{\partial L}{\partial \theta_1} = 0 \\ \dfrac{\partial L}{\partial \theta_2} = 0. \end{cases}$$

Из этой системы уравнений определяют критическую точку (θ_1, θ_2).

После нахождения критической точки необходимо доказать, что в ней действительно наблюдается максимум функции $L(\theta_1, \theta_2)$. Для этого используют достаточные условия экстремума функции. Напомним их.

Для функции одной переменной надо показать, что производная функции меняет знак с «плюса» на «минус» при переходе через найденную точку или, что вторая производная в этой точке меньше нуля.

Для функции одной переменной надо показать, что производная функции меняет знак с «плюса» на «минус» при переходе через найденную точку или, что вторая производная в этой точке меньше нуля.

Для функции двух переменных необходимо вычислить в найденной точке (θ_1, θ_2) определитель

$$\triangle = \begin{vmatrix} A & B \\ B & C \end{vmatrix}, \quad \text{где} \quad A = \frac{\partial^2 L}{\partial \theta_1^2},\ B = \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2},\ C = \frac{\partial^2 L}{\partial \theta_2^2},$$

и если $\triangle > 0$ и $A < 0$, то в точке (θ_1, θ_2) будет наблюдаться максимум функции $L(\theta_1, \theta_2)$ или ее логарифма.

Метод наименьших квадратов (МНК)

Здесь оценка определяется из поиска минимума следующей функции

$$U(\theta) = \sum_{i=1}^{n} (x_i - \theta)^2 \to \min.$$