

Содержание

1	Пределевые теоремы и законы больших чисел	2
2	Вариационные ряды и их характеристики	2
2.1	Генеральная и выборочная совокупности, объём выборки.	2
2.2	Вариационный ряд, варианта, частота. Виды вариационных рядов. Гистограмма, полигон.	2
2.3	Формулы числовых характеристик. Эмперическая функция распределения (ЭФР). Свойства ЭФР.	4
2.4	Эмперическая функция распределения ЭФР. Свойства ЭФР. . .	5
3	Оценки параметров распределения	6
3.1	Понятия статистики, оценки, выборочной характеристики.	6
3.2	Несмешённые, состоятельные и эффективные оценки	6
3.3	Теорема о несмешённой состоятельной оценке мат. ожидания .	7
3.4	Дополнительная информация	8
3.5	Теорема о несмешённой и состоятельной оценке дисперсии . .	8
3.6	Теорема о несмешённой и состоятельной оценке функции распределения	9
3.7	Теорема об эффективной оценке математического ожидания .	11
3.8	Теорема о единственности эффективной оценки	12
3.8.1	Определение регулярной параметрической модели. Информация Фишера.	13
3.9	Неравенство Рао-Крамера. Построение эффективной по Рао-Крамеру оценки для пуассоновского распределения	14
4	Методы построения оценок	14
4.1	Метод моментов. Пример (биномиальное распределение)	14
4.2	Функция правдоподобия. Метод максимального правдоподобия. Пример (биномиальное распределение)	16
5	Регрессионный анализ	17
5.1	Постановка задачи	17
5.2	Построение выборочного уравнения регрессии	19
5.3	Выборочный коэффициент корреляции, его свойства	20
5.4	Проверка гипотезы о значимости выборочного коэффициента корреляции	21
5.5	Теорема Гаусса-Маркова	21
5.6	Статические свойства МНК оценок	22
5.7	Анализ вариации	22

1 Предельные теоремы и законы больших чисел

2 Вариационные ряды и их характеристики

2.1 Генеральная и выборочная совокупности, объём выборки.

Рассмотрим постановку задачи математической статистики: по результатам наблюдения за некоторой случайной величиной ξ требуется сделать выводы о неизвестном законе распределения этой величины $\mathcal{L}(x, \theta)$ либо о неизвестных параметрах $\theta_1, \dots, \theta_n$ известного распределения.

Пусть ξ — случайная величина с некоторой (теоретической) функцией распределения $F_\xi(x) = P\{\xi < x\}$, $x \in R$.

Определение:

Совокупность n независимых одинаково распределённых случайных величин X_1, X_2, \dots, X_n называется выборкой (выборочной совокупностью), извлечённой из распределения случайной величины ξ .

Определение:

Под генеральной совокупностью понимается множество всех возможных значений случайной величины ξ .

Определение:

Объёмом совокупности называется количество всех её элементов, объём выборки или выборочной совокупности обозначается n , генеральной совокупности — N .

2.2 Вариационный ряд, варианта, частота. Виды вариационных рядов. Гистограмма, полигон.

Определение:

Пусть x_1, x_2, \dots, x_n — выборка из генеральной совокупности значений.

Вариационным рядом называется последовательность $x_1^*, x_2^*, \dots, x_n^*$ элементов выборки расположенных в порядке неубывания, т.е. $x_1^* \leq x_2^* \leq \dots \leq x_n^*$.

x_i^* — варианта.

n_i — частота появления варианты x_i^* в выборке.

Определение:

Точечным вариационным рядом называется:

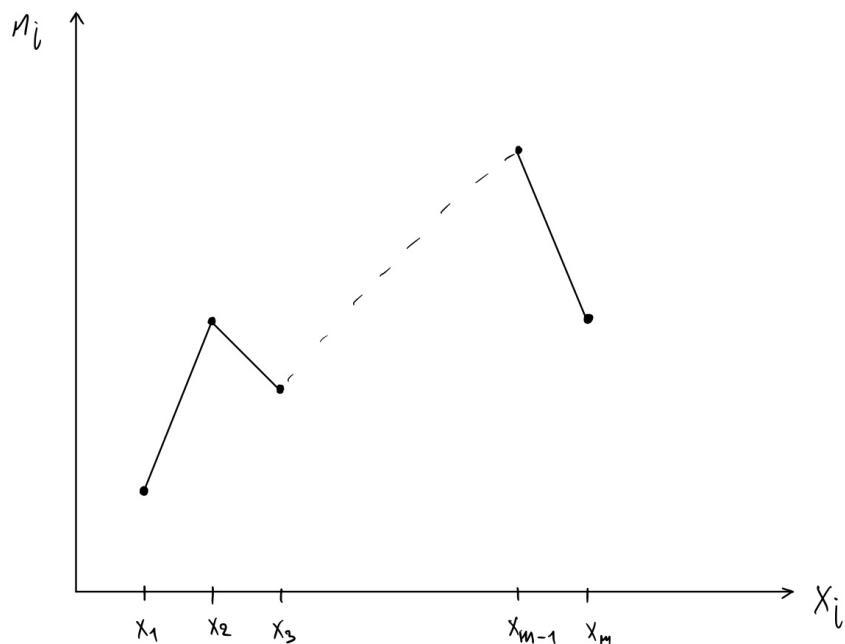
x_i	x_1	x_2	\dots	x_m
n_i	n_1	n_2	\dots	n_m

x_i — варианта, n_i — частота соответствующей варианты.

m — количество групп (различных вариантов (вариант в таблице)).

$n = \sum_{i=1}^m n_i$, где n_i — объём выборки.

Для графического представления точечных вариационных рядов используется полигон частот — ломанная с вершинами в точках (x_i, n_i) .



Определение:

Интервальным вариационным рядом называется:

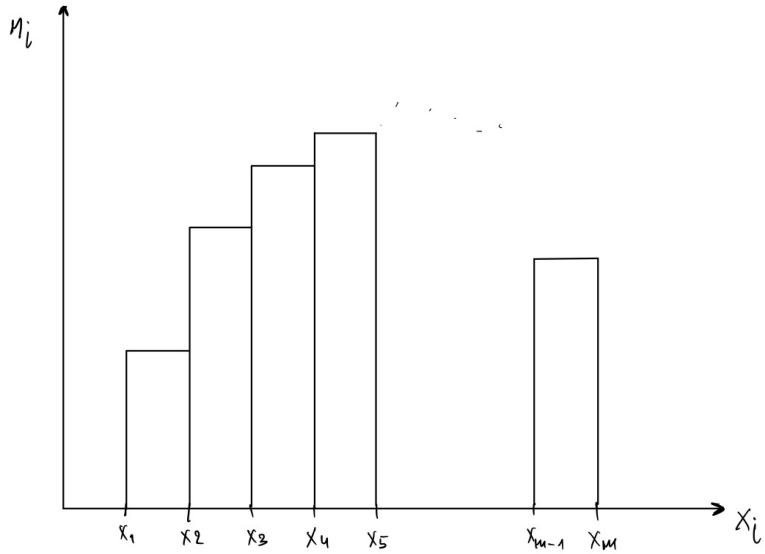
x_i	$[x_1, x_2]$	$(x_2, x_3]$	\dots	$(x_m, x_{m+1}]$
n_i	n_1	n_2	\dots	n_m

x_i — варианты, n_i — частота.

m — количество групп (интервалов).

$n = \sum_{i=1}^m n_i$, где n_i — объём выборки.

Для графического представления интервальных вариационных рядов используется гистограмма частот — фигура, составленная из прямоугольников, одной стороной которых служат интервалы $(x_i, x_{i+1}]$, а длина второй равна n_i .



2.3 Формулы числовых характеристик. Эмперическая функция распределения (ЭФР). Свойства ЭФР.

Определение:

Выборочным средним называется величина:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i * n_i,$$

где m — количество групп в точечном или интервалов в интервальном вариационном ряду, n_i — частота, т.е. количество элементов выборки, принадлежащих i -той группе или i -тому интервалу, x_i — варианта для точечного ряда и середина i -того интервала для интервального ряда.

Определение:

Выборочной дисперсией (смешённой) называется величина:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2.$$

Она характеризует среднее из квадратов отклонений наблюдаемой величины от выборочного среднего. Величина $S = \sqrt{S^2}$ называется выборочным средним

квадратическим отклонением (смешённым) величин выборки от выборочного среднего.

Если данные представлены в виде точечного или интервального вариационного ряда, то для вычисления используют формулу:

$$S^2 = \frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^2 n_i.$$

или

$$S^2 = \left(\frac{1}{n} \sum_{i=1}^m x_i^2 n_i \right) - \bar{x}^2.$$

Здесь m — количество групп в точечном или интервалов в интервальном вариационных рядах, n_i — частота. т.е. количество элементов выборки, принадлежащих i -той группе или i -тому интервалу, x_i — варианта для точечного ряда и середина i -того интервала для интервального ряда.

Определение:

Выборочной дисперсией (несмешённой) называется величина:

$$\bar{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})^2.$$

Аналогично, величина $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$ называется выборочным несмешённым средним квадратическим отклонением.

Очевидно, что смешённая и несмешённая выборочные дисперсии связаны формулой:

$$\bar{\sigma}^2 = \frac{n}{n-1} S^2.$$

2.4 Эмперическая функция распределения ЭФР. Свойства ЭФР.

Эмперическая функция распределения ЭФР

Эмперической функцией распределения (ЭФР) называется функция

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n e(x - X_i),$$

где $e(x) = 1$, при $x > 0$ $e(x) = 0$, при $x \leq 0$.

Таким образом, если X_i , то $e(x) = 1$, если $X_i \geq x$, то $e(x) = 0$, а сумма $e(x - X_i)$ будет равна количеству элементов выборки, которые приняли значение, строго меньше некоторого $x \in R$.

Пусть x_1, x_2, \dots, x_n — реализация выборки X_1, X_2, \dots, X_n , т.е. наблюдавшиеся значения случайной величины ξ .

Обозначим $\mu(x)$ — число элементов выборки, строго меньших $x \in R$. Тогда эмпирическая функция распределения $\tilde{F}_n(x)$ может быть определена как

$$\tilde{F}_n(x) = \frac{\mu(x)}{n}.$$

Свойства ЭФР:

1. $0 \leq \tilde{F}_n(x) \leq 1$, т.к. $0 \leq \mu(x) \leq n$;
2. неубывающая непрерывная слева функция;
3. $\tilde{F}_n(x)$ — ступенчатая функция для всех типов распределений;
4. $\tilde{F}_n(x)$ сходится по распределению к $F_\xi(x)$.

3 Оценки параметров распределения

3.1 Понятия статистики, оценки, выборочной характеристики.

Определение:

Пусть $g(t_1, \dots, t_n)$ — непрерывная функция. Оценкой θ назовём $\tilde{\theta} = g(X_1, \dots, X_n)$. Если $g(X_1, \dots, X_n) = T$ некоторая функция, то T — статистика.

Определение:

Выборочными характеристиками называются функции от наблюдений (точечные оценки), приближённо оценивающие соответствующие числовые характеристики случайной величины.

3.2 Несмешённые, состоятельные и эффективные оценки

Определение:

Оценка $\tilde{\theta}_n$ называется несмешённой оценкой параметра θ , если $M(\tilde{\theta}_n) = \theta$.

Определение:

Оценка $\tilde{\theta}_n$ называется состоятельной оценкой параметра θ , если $\tilde{\theta}_n$ сходится по вероятности к θ .

Определение:

Оценка $\tilde{\theta}_n$ называется эффективной, или оптимальной, или наилучшей несмешённой оценкой с минимальной дисперсией (НОМД), если $M(\tilde{\theta}_n) = \theta$ и $D(\tilde{\theta}_n) = \inf_{\tilde{\theta}_n^*} D\tilde{\theta}_n^*$.

3.3 Теорема о несмешённой состоятельной оценке мат. ожидания

Пусть $\xi \sim L(x, \theta)$, $L(x, \theta)$ — закон распределения известен с точностью до параметра.

$$\bar{\theta} = (\Theta_1, \dots, \Theta_n)$$

По результатам X_1, \dots, X_n наблюдений за ξ требуется построить оценку θ .

Теорема:

Пусть $X_1, \dots, X_n \sim L_\xi(x, \theta)$, где ξ — случайная величина с $M\xi = a < +\infty$, $D\xi = \sigma^2$. Тогда выборочное среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ является несмешённой и состоятельной оценкой $M\xi$.

Доказательство:

$$\begin{aligned} M\bar{x} &= M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n Mx_i = \\ &= |x_i \text{ — н. о. р. случайной величины} \Rightarrow Mx_i = M\xi \quad \forall i| = \frac{1}{n} \sum_{i=1}^n M\xi = \\ &= \frac{1}{n} * n * a = a \Rightarrow \bar{x} \text{ — несмешённое} \end{aligned}$$

Состоятельность $\forall \epsilon > 0 \quad P\{|\bar{x} - a| < \epsilon\} \rightarrow_{n \rightarrow \infty} 1$

По неравенству Чебышёва:

$$\begin{aligned} \forall \epsilon > 0 \quad P\{|\bar{x} - a| < \epsilon\} &\geq 1 - \frac{D\bar{x}}{\epsilon^2} = 1 - \frac{1}{\epsilon^2} D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \\ &= |\text{н. о. р. } Dx_i = \sigma^2| = \end{aligned}$$

$$= 1 - \frac{1}{n^2} \epsilon^2 D\xi = 1 - \frac{n\sigma^2}{n^2 \epsilon^2} = 1 - \frac{\sigma^2}{n\epsilon^2} \rightarrow_{n \rightarrow \infty} 1 - 0$$

Т.к. $P(A) \leq 1 \quad \forall A$, то $\lim_{n \rightarrow \infty} P\{|\bar{x} - a| < \epsilon\} = 1 \Rightarrow \bar{x} \text{ — состоятельная.}$

3.4 Дополнительная информация

Определение:

Выборочной дисперсией (смешённой) называется величина $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Для группированных данных: $S^2 = \frac{1}{n} \sum_{i=1}^m (x_i - \bar{x})^2 * n_i$.

Определение:

Выборочной дисперсией (несмешённой) называется оценка $\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Для группированных данных: $\tilde{\sigma}^2 = \frac{n}{n-1} * S^2$.

3.5 Теорема о несмешённой и состоятельной оценке дисперсии

Теорема:

Пусть $X_1, \dots, X_n \sim L(x, \theta)$, где $M\xi = a < +\infty$, $D\xi = \sigma^2$. Тогда несмешённой и состоятельной оценкой $D\xi$ является величина $\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Доказательство:

1. Доказем несмешённость ($M\bar{g}^2 = g^2$):

$$M\bar{g}^2 = M \left(\frac{1}{n-1} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \right) = \frac{1}{n-1} \left(\sum_{i=1}^n (Mx_i^2 - 2M(x_i\bar{x}) + M\bar{x}^2) \right) \quad (\textcircled{2})$$

Найдём значения членов:

$$Mx_i^2 = M\xi^2 = D\xi + (M\xi)^2 = g^2 + a^2 \quad \boxed{\begin{aligned} D\xi &= M\xi^2 - (M\xi)^2 \\ g^2 &= M\xi^2 - a^2 \Rightarrow M\xi^2 = g^2 + a^2 \end{aligned}}$$

$$\begin{aligned} Mx_i\bar{x} &= M \left(x_i \cdot \frac{1}{n} \sum_{j=1}^n x_j \right) = \frac{1}{n} \left(M(x_1 x_1) + M(x_1 x_2) + \dots + Mx_i^2 + \dots + \right. \\ &\quad \left. + M(x_1 x_n) \right) \stackrel{\text{H.O.P.}}{=} \frac{1}{n} \left(Mx_i Mx_1 + Mx_i Mx_2 + \dots + Mx_i^2 + \dots + Mx_i Mx_n \right) = \\ &= \frac{1}{n} \left(a^2 + a^2 + \dots + (g^2 + a^2) + \dots + a^2 \right) = \frac{n a^2 + g^2}{n} = a^2 + \frac{g^2}{n} \end{aligned}$$

$$M\bar{x}^2 = M\bar{x} \cdot \bar{x} = M \left(\frac{1}{n} \sum_{i=1}^n x_i \bar{x} \right) = \frac{1}{n} \sum_{i=1}^n Mx_i \bar{x} = \frac{1}{n} \cdot n \cdot \left(a^2 + \frac{g^2}{n} \right) = a^2 + \frac{g^2}{n}$$

$$\begin{aligned} \textcircled{2} \quad \frac{1}{n-1} \cdot n \left(g^2 + a^2 - 2 \left(a^2 + \frac{g^2}{n} \right) + \left(a^2 + \frac{g^2}{n} \right) \right) &= \frac{n}{n-1} \left(g^2 + a^2 - a^2 - \frac{g^2}{n} \right) = \\ &= \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot g^2 = g^2 \end{aligned}$$

T.R. $M\bar{g}^2 = g^2$, то оценка несмешённая.

Составность не доказывается.

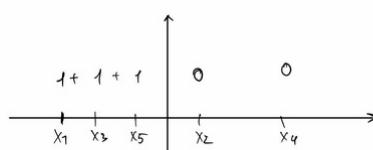
3.6 Теорема о несмешённой и состоятельной оценке функции распределения

Определение:

Энпримеская функция распределения: $\tilde{F}_n(x) = \frac{M_H(x)}{n}$; $M_H(x)$ — кол-во элементов выборки строго меньших x .

$$\text{Пусть } e(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad \sum_{i=1}^n e(x - x_i)$$

$$\text{Таким образом, } \tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n e(x - x_i).$$



Теорема:

Пусть $X_1, \dots, X_n \sim L_\xi(x, \theta)$, где $F_\xi(x)$ — функция распределения случайной величины ξ . Тогда эмпирическая функция распределения $\tilde{F}_n(x)$ является несмещённой и состоятельной оценкой функции распределения $F_\xi(x)$.

Доказательство:

Покажем несмещённость ($M(\tilde{F}_n(x)) = F_\xi(x) = P\{\xi < x\}$):

$$M(\tilde{F}_n(x)) = M\left(\frac{1}{n} \sum_{i=1}^n e(x - x_i)\right) = \frac{1}{n} \sum_{i=1}^n M(e(x - x_i)) \quad \text{□}$$

Построим ряд р-я вида $e(x - x_i)$:

$e(x - x_i)$	0	1
P	$P\{x - x_i \leq 0\}$	$P\{x - x_i > 0\}$

$$P\{x - x_i \leq 0\} = P\{x \leq x_i\} = P\{x_i \geq x\} = 1 - F_x(x)$$

$$P\{x - x_i > 0\} = P\{x > x_i\} = P\{x_i < x\} = F_{X_i}(x)$$

$$M(e(x - x_i)) = 0 \cdot (1 - F_{X_i}(x)) + 1 \cdot F_{X_i}(x) = F_{X_i}(x) \quad \forall i$$

$$M(e^2(x - x_i)) = 0^2(1 - F_{X_i}(x)) + 1^2 \cdot F_{X_i}(x) = F_{X_i}(x)$$

$$De(x - x_i) = F_{X_i}(x) - F_{X_i}^2(x) = \underbrace{F_{X_i}(x)}_{0 \leq \cdot \leq 1} \underbrace{(1 - F_{X_i}(x))}_{0 \leq \cdot \leq 1}$$

$$\text{□} \quad \frac{1}{n} \sum_{i=1}^n F_{X_i}(x) = \left[\frac{x_i -}{\text{M.O.P.}} \right] = \frac{1}{n} \cdot n F_\xi(x) = F_\xi(x)$$

Покажем состоятельность $\tilde{F}_n(x) \xrightarrow{P} F_\xi(x)$:

По неравенству Чебышева:

$$\begin{aligned} \forall \varepsilon > 0 \quad P\{|\tilde{F}_n(x) - F_\xi(x)| < \varepsilon\} &\geq 1 - \frac{D\tilde{F}_n(x)}{\varepsilon^2} = 1 - \frac{1}{\varepsilon^2} D\left(\frac{1}{n} \sum_{i=1}^n e(x - x_i)\right) = \\ &= 1 - \frac{1}{n^2 \varepsilon^2} \cdot n F_\xi(x) (1 - F_\xi(x)) = 1 - \frac{F_\xi(x)(1 - F_\xi(x))}{n \varepsilon^2} \xrightarrow{n \rightarrow \infty} 1 - 0 \end{aligned}$$

T.R. $P(A) \leq 1$, то $\tilde{F}_n(x) \xrightarrow{P} F_\xi(x)$

3.7 Теорема об эффективной оценке математического ожидания

Определение:

Пусть $\tilde{\theta}_n$ — несмешённая оценка. $\tilde{\theta}_n$ — называется эффективной оценкой, если

$$D\tilde{\theta}_n = \inf_{\tilde{\theta}_n^*} D\tilde{\theta}_n^*$$

Теорема:

Пусть $X_1, \dots, X_n \sim L_\xi(x, \theta)$, где ξ — случайная величина с $M\xi = a < +\infty$. Тогда выборочное среднее $\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$ является эффективной оценкой $M\xi$ в классе линейных несмешённых оценок.

Доказательство:

Расс-и производимо ми оценку $\tilde{\theta}_n = \sum_{i=1}^n \alpha_i X_i$:

$$\sum_{i=1}^n \alpha_i = 1; \alpha_i > 0 \forall i$$

Найдем критерий оценки $\tilde{\theta}_n$:

$$D\tilde{\theta}_n = D\sum_{i=1}^n \alpha_i X_i = \sum_{i=1}^n \alpha_i DX_i = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \sum_{i=1}^n \alpha_i^2 G^2 = G^2 \sum_{i=1}^n \alpha_i^2 = G^2 g(\alpha_1, \dots, \alpha_n)$$

$D\tilde{\theta}_n$ достигает мин на тех α_i , при которых $g(\alpha_1, \dots, \alpha_n) \rightarrow \min$

Ряд як-то воспользовался методом лагранжа поиском условного минимума

$$g(\alpha_i) = \sum_{i=1}^n \alpha_i^2$$

$$L(\alpha_i, \lambda) = \sum_{i=1}^n \alpha_i^2 + \lambda \left(\sum_{i=1}^n \alpha_i - 1 \right)$$

константа лагранжа

Найдем производные $L(\alpha_i, \lambda)$, λ -const лагранжа.

$$\begin{cases} \frac{\partial L(\alpha_i, \lambda)}{\partial \lambda} = \sum_{i=1}^n \alpha_i - 1 = 0 \\ \frac{\partial L(\alpha_i, \lambda)}{\partial \alpha_i} = 2\alpha_i + \lambda = 0 \Rightarrow \alpha_i = -\frac{\lambda}{2} \end{cases} \quad \Rightarrow \quad \sum_{i=1}^n \frac{\lambda}{2} - 1 = 0 \Rightarrow$$

$$\Rightarrow \frac{n\lambda}{2} = -1; \lambda = -\frac{2}{n} \Rightarrow \alpha_i = -\left(\frac{-2}{n}\right) \cdot \frac{1}{2} \Rightarrow \alpha_i = \frac{1}{n} \Rightarrow$$

$$\Rightarrow g(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \alpha_i^2 \text{ имеет мин значение, если } g(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^n \left(\frac{1}{n}\right)^2$$

$$\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \frac{1}{n} = 1$$

следовательно, залогом оценкой является $\tilde{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$.

3.8 Теорема о единственности эффективной оценки теорема:

Пусть $X_1, \dots, X_n \sim L_\xi(x, \theta)$, где θ — параметр распределения, и пусть $\tilde{\theta}(\bar{X}_n)$ и $\hat{\theta}_n(\bar{X}_n)$ — две эффективные оценки θ . Тогда $\tilde{\theta}_n(\bar{X}_n) = \hat{\theta}_n(\bar{X}_n)$. То есть $P\{(x_1, \dots, x_n) : \tilde{\theta}_n(\bar{X}_n) \neq \hat{\theta}_n(\bar{X}_n)\} = 0$.

Доказательство:

Док-во: т.к. $\hat{\theta}_1$ и $\hat{\theta}_2$ — эфр., то по опр. $M\hat{\theta}_1 = M\hat{\theta}_2 = \theta$. $D\hat{\theta}_1$ и $D\hat{\theta}_2$ есть инт. $D\hat{\theta}_1$, т.к. инт. единст., то $D\hat{\theta}_1 = D\hat{\theta}_2$. Рассм. $\hat{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$:

$$M\hat{\theta} = \frac{1}{2}(M\hat{\theta}_1 + M\hat{\theta}_2) = \frac{1}{2}(\theta + \theta) = \theta \Rightarrow \hat{\theta} — \text{нестанд. оценка}$$

$$D\hat{\theta} = \frac{1}{4}(D\hat{\theta}_1 + D\hat{\theta}_2 + 2\text{cov}(\hat{\theta}_1, \hat{\theta}_2)) = \frac{1}{2}(D\hat{\theta}_1 + \text{cov}(\hat{\theta}_1, \hat{\theta}_2)) \quad (*)$$

$$|\text{cov}(\hat{\theta}_1, \hat{\theta}_2)| = |M(\hat{\theta}_1 - M\hat{\theta}_1)(\hat{\theta}_2 - M\hat{\theta}_2)| \leq \sqrt{M(\hat{\theta}_1 - M\hat{\theta}_1)^2} \cdot \sqrt{M(\hat{\theta}_2 - M\hat{\theta}_2)^2} =$$

$$= \sqrt{D\hat{\theta}_1} \cdot \sqrt{D\hat{\theta}_2} = D\hat{\theta}_1, \text{ т.е. } |\text{cov}(\hat{\theta}_1, \hat{\theta}_2)| \leq D\hat{\theta}_1,$$

$$D\hat{\theta} = D\hat{\theta}_1 = \frac{1}{2}|D\hat{\theta}_1 + \text{cov}(\hat{\theta}_1, \hat{\theta}_2)| \leq \frac{1}{2}(|D\hat{\theta}_1| + |\text{cov}(\hat{\theta}_1, \hat{\theta}_2)|) \leq \frac{1}{2}(|D\hat{\theta}_1| + |D\hat{\theta}_1|) =$$

$$= D\hat{\theta}_1 = \text{инт. } D\hat{\theta}_1 \Rightarrow D\hat{\theta} = D\hat{\theta}_1$$

Решим (*): $D\hat{\theta}_1 = \frac{1}{2}(D\hat{\theta}_1 - \text{cov}(\hat{\theta}_1, \hat{\theta}_2)) \Rightarrow D\hat{\theta}_1 = \text{cov}(\hat{\theta}_1, \hat{\theta}_2)$

Коэф. корреляции $\text{cov}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{cov}(\hat{\theta}_1, \hat{\theta}_1)}{\sqrt{D\hat{\theta}_1} \sqrt{D\hat{\theta}_2}} = \frac{D\hat{\theta}_1}{D\hat{\theta}_1} = 1 \Rightarrow \hat{\theta}_2 = a\hat{\theta}_1 + b$

$$M\hat{\theta}_2 = aM\hat{\theta}_1 + b \quad (\text{нестанд.})$$

$$\begin{aligned} \theta = a\hat{\theta}_1 + b \\ 1 \cdot \theta + 0 = a\hat{\theta}_1 + b \end{aligned} \left. \begin{aligned} \Rightarrow a = 1, b = 0 \\ \Rightarrow \hat{\theta}_1 = \hat{\theta}_2 \end{aligned} \right\} \Rightarrow \text{эфр. оценка единст.} \Rightarrow ! \blacksquare$$

3.8.1 Определение регулярной параметрической модели. Информация Фишера.

Определение:

Пусть ξ — наблюдаемая случайная величина.

Выборочным пространством X_n называется $\{\bar{X}_n = (X_1, \dots, X_n)\}$.

Параметрическим множеством называется Θ — множество значений параметра θ .

Определение:

Параметрической моделью называется

$$\{\mathbf{X}_n; F_\xi(x, \theta) : \theta \in \Theta\}$$

Определение:

Параметрическая модель называется регулярной, если:

1. параметрическое множество Θ — открытое множество;
2. носитель распределён, то есть множество $A = \{x : f(x) > 0\}$ — не зависит от параметров;
3. $\forall \theta$ и $x \in A$ существует производная $\frac{\partial f(x, \theta)}{\partial \theta} < +\infty$;
4. $\forall \theta \in \Theta \quad M\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right) = 0 \quad M\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2 < +\infty$;

5. Дважды допустимо дифференцирование под знаком интеграла по параметру.

Определение:

Пусть ξ — случайная величина с функцией плотности $f(x, \theta)$. Тогда информацией Фишера для случайного вектора \bar{X}_n называется $I_n(\theta) = M\left(\frac{\partial}{\partial \theta} \ln f(x, \theta)\right)^2$

3.9 Неравенство Рао-Крамера. Построение эффективной по Рао-Крамеру оценки для пуассоновского распределения

Неравенство Рао-Крамера:

Пусть дана регулярная параметрическая модель. Тогда для любой несмешённой оценки $\tilde{\theta}$ параметра θ $D\tilde{\theta} \geq \frac{1}{nI_n(\theta)}$.

Определение:

Величину $e(\theta) = \frac{1}{nI(\theta)*D\theta}$ называют показателем эффективности по Рао-Крамеру.

Для любой несмешённой оценки $0 < e(\theta) \leq 1$. Если $e(\theta) = 1$, то $\tilde{\theta}_n$ называется эффективной по Рао-Крамеру, следовательно, просто эффективна.

Построение эффективной по Рао-Крамеру оценки для пуассоновского распределения:

$(X_1, \dots, X_n) \sim P_{\theta}, s(1)$. Построить эф. (оптимальную) оценку λ .

$$P\{\xi = k\} = \frac{e^{-\lambda} \lambda^k}{k!} \Rightarrow P\{X_i = k\} = P\{\xi = k\}$$

$$M_{X_i} = \lambda, D_{X_i} = \lambda^2 \text{ (из ТБ)} \quad M_{X_i^2} = \lambda^2 + \lambda$$

Следовательно $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ — оценка λ .

$$M\bar{X} = \lambda, D\bar{X} = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^n D_{X_i} = \frac{1}{n^2} \cdot n \cdot \lambda = \frac{\lambda}{n}.$$

Найдём $I(1)$ — в одном наблюдении:

$$I(1) = M\left(\frac{\partial \ln\left(\frac{e^{-\lambda} \lambda^x}{x!}\right)}{\partial \lambda}\right)^2 = M\left(\frac{\partial}{\partial \lambda}(\lambda \ln \lambda - \lambda - \ln \lambda!)^2\right) = M\left(\frac{\lambda}{\lambda} - 1\right)^2 = M\left(\frac{\lambda^2}{\lambda^2} - \frac{2\lambda}{\lambda} + 1\right) = \frac{1}{\lambda^2} M\lambda^2 - \frac{2}{\lambda} M\lambda + 1 = \frac{1}{\lambda^2} (\lambda^2 + 1) - \frac{2}{\lambda} \cdot \lambda + 1 = 1 + \frac{1}{\lambda} - 2 + 1 = \frac{1}{\lambda}.$$

Вычислим $e(\bar{X}) = \frac{1}{n} \cdot \lambda \cdot \frac{1}{\lambda} = 1 \Rightarrow \bar{X}$ — оптимальная оценка параметра λ .

4 Методы построения оценок

4.1 Метод моментов. Пример (биномиальное распределение)

Определение:

Начальным теоретическим моментом k -го порядка называется $m_k = M\xi^k$, если $M|\xi|^k < +\infty$.

Определение:

Центральным теоретическим моментом k -го порядка называется $\mu_k = M(\xi - M\xi)^k$, если $M|\xi| < +\infty$.

Определение:

Выборочным начальным моментом k -го порядка называется $\tilde{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$.

Определение:

Выборочным центральным моментом k -го порядка называется $\tilde{\mu}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^k$.

Метод моментов:

Метод моментов состоит в том, что за оценку параметров принимается решение системы уравнений:

$$\begin{cases} m_k = \tilde{m}_k \\ \mu_k = \tilde{\mu}_k \end{cases} \quad (1)$$

Пример (биномиальное распределение):

Пусть $X_1, \dots, X_n \sim B(m, p)$

m - кол-во недавне. испытаний Бернуlli

p - веро-ть успеха

Получить $\tilde{P}_{m,n}$

Решение: $\mu_1 = \tilde{m}_1 \Rightarrow M\xi = \bar{x} \Rightarrow mp = \bar{x} \Rightarrow p = \frac{\bar{x}}{m}$.

Таким образом, $\tilde{P}_{m,n} = \frac{\bar{x}}{m} = \frac{1}{m \cdot n} \sum_{i=1}^m X_i$

4.2 Функция правдоподобия. Метод максимального правдоподобия. Пример (биномиальное распределение)

Определение:

Пусть $X_1, \dots, X_n \sim L_\xi(x, \theta)$, причём $f_\xi(x, \theta)$ — известна, а θ неизвестен.

Функцией правдоподобия называют $L(\bar{X}_n, \theta) = \prod_{i=1}^n f(X_i, \theta) = f(X_1, \theta) * \dots * f(X_n, \theta)$.

Метод максимального правдоподобия:

Рассмотрим $\ln L(\bar{X}_n, \theta) = \sum_{i=1}^n \ln f(X_i, \theta)$.

Метод максимального правдоподобия состоит в том, что за оценку параметра θ принимается точка максимума функции правдоподобия.

Алгоритм:

1. Построить функцию $L(\bar{X}_n, \theta)$;
2. Взять её логарифм $\ln L(\bar{X}_n, \theta)$;
3. $\frac{\partial \ln L(\bar{X}_n, \theta)}{\partial \theta} = 0 \Rightarrow \theta$;
4. Если значение второй производной строго меньше ноля, то оценка параметра $\tilde{\theta}$ есть θ_0 .

$$\frac{\partial^2 \ln L(\bar{X}_n, \theta)}{\partial \theta^2}|_{\theta_0} < 0, \text{ то } \tilde{\theta} = \theta_0$$

Пример (биномиальное распределение):

Пример (Построим оценку $\hat{P}_{\text{ннр}}$ параметра p ВИМ (n, p)):

$$1. L(\bar{x}_n; \theta) = C_n^{x_1} p^{x_1} q^{n-x_1} \cdot C_n^{x_2} p^{x_2} q^{n-x_2} \cdots \cdot C_n^{x_n} p^{x_n} q^{n-x_n} =$$

$$= \left(\prod_{i=1}^n C_n^{x_i} \right) \cdot p^{\bar{n}\bar{x}} q^{n-n\bar{x}}$$

$$\text{т.к. } x_1 + \dots + x_n = \sum_{i=1}^n x_i = n \cdot \left(\frac{1}{n} \sum_{i=1}^n x_i \right) = n\bar{x}$$

$$2. \ln L(\bar{x}_n; p) = \sum_{i=1}^n \ln C_n^{x_i} + n\bar{x} \ln p + (n-n\bar{x}) \ln (1-p)$$

$$3. \frac{\partial}{\partial p} \ln L(\bar{x}_n; p) = \frac{n\bar{x}}{p} - \frac{n\mu - n\bar{x}}{1-p} = 0 \quad [0 < p < 1]$$

$$\bar{x} - p\bar{x} - (\mu - \bar{x})p = 0$$

$$p_0 = \frac{\bar{x}}{\mu}.$$

$$4. \frac{\partial^2 L(\bar{x}_n; p)}{\partial p^2} = -\frac{n\bar{x}}{p^2} - \frac{n\mu - n\bar{x}}{(1-p)^2} \Big|_{p_0} = -n \left(\frac{\bar{x}}{\bar{x}^2/\mu^2} + \frac{(\mu - \bar{x})}{(\mu - \bar{x})^2/\mu^2} \right) = -n^2 \mu \left(\frac{1}{\bar{x}} + \frac{1}{\mu - \bar{x}} \right) =$$

$$= -n^2 \mu \left(\frac{\mu - \bar{x} + \bar{x}}{\bar{x}(\mu - \bar{x})} \right) = -\frac{\mu^2 \bar{x}^2}{\bar{x}(\mu - \bar{x})} > 0 \Rightarrow$$

$$\Rightarrow \hat{p}_{\text{ннр}} = p_0 = \frac{\bar{x}}{\mu}.$$

Оценки по методу ММП не являются линейными по параметру, но они состоят из линейных и асимптотически согласованных.

5 Регрессионный анализ

5.1 Постановка задачи

подготовительная информация:

Выборочные средние имеют вид:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

Выборочные дисперсии (смешённые) имеют вид:

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2, \quad S_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{y})^2.$$

Величины $S_X = \sqrt{S_X^2}$ и $S_Y = \sqrt{S_Y^2}$ представляют собой выборочные средние квадратические отклонения (смешённые) величин ξ и η .

Выборочное среднее (совместное) есть величина:

$$\bar{xy} = \frac{1}{n} \sum_{i=1}^n X_i Y_i$$

Коэффициент корреляции (выборочный):

$$r_{\text{выб}}(\text{далше } r_v) = \frac{\bar{xy} - \bar{x} * \bar{y}}{\sqrt{S_x^2} \sqrt{S_y^2}}$$

Постановка задачи:

Пусть производятся наблюдения некоторого процесса, который описывается двумя случайными величинами (ξ, η) .

Тогда результаты наблюдений представляют собой двумерную выборку $(X_1, Y_1), \dots, (X_n, Y_n)$.

Множество точек с координатами (X_i, Y_i) называют полем корреляции или диаграммой рассеяния. По виду поля корреляции делают предположение о связи между величинами ξ и η .

Как известно, если коэффициент корреляции $r = \text{Cov}(\xi, \eta) / \sqrt{(D\xi D\eta)}$ равен нулю, то величины ξ и η некоррелированы, если $|r| = 1$, то величины ξ и η линейно связаны $\eta = a\xi + b$.

Следовательно, при изучении связи величин необходимо оценить отличие выборочного коэффициента корреляции от нуля.

Проверим гипотезу о значимости коэффициента корреляции, т.е. $H_0 : "r = 0"$ против гипотезы $H_1 : "r \neq 0"$.

Статистика $t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$ имеет распределение Стьюдента с $n - 2$ степенями свободы.

Для заданного уровня значимости критическое значение равно $t_{\text{кр}} = t(\frac{\alpha}{2}; n - 2)$.

Если $t_{\text{набл}} < t_{\text{кр}}$, принимают гипотезу $H_0 : "r = 0"$, в противном случае H_0 отвергают, и делают вывод о значимом отличии коэффициента корреляции от нуля.

Если установлено, что коэффициент значимо отличается от нуля, то линейное уравнение должно достаточно близко описывать имеющуюся между величинами, но неизвестную связь.

5.2 Построение выборочного уравнения регрессии

Пусть наблюдалась выборка $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, и выборочный коэффициент значимо отличается от нуля.

Найдем коэффициенты уравнения $Y = aX + b$, которое наилучшим образом аппроксимирует $Y = f(X)$ ($\eta = f(\xi)$).

Величина Y называется зависимой переменной, признаком, величина X называется независимой переменной, фактором, регрессором.

В парной регрессионной модели зависимая переменная зависит только от одного регрессора.

Оценим коэффициенты уравнения методом наименьших квадратов (МНК).

Будем обозначать \hat{Y} – вычисленные (прогнозные) значения.

Согласно МНК требуется найти такие значения оценок параметров \hat{a} и \hat{b} , чтобы была минимальной сумма квадратов отклонений прогнозных значений от наблюдаемых:

$$L(\hat{a}, \hat{b}) = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - (\hat{a} + \hat{b}X_t))^2 \rightarrow \min.$$

Значит, для нахождения оценки параметров парной регрессионной модели МНК необходимо найти экстремум (минимум) функции двух аргументов.

Запишем необходимые условия экстремума:

$$\begin{cases} \frac{\partial L}{\partial \hat{a}} = -2 \sum_{t=1}^n (Y_t - \hat{a} - \hat{b}X_t) = 0, \\ \frac{\partial L}{\partial \hat{b}} = -2 \sum_{t=1}^n X_t(Y_t - \hat{a} - \hat{b}X_t) = 0, \end{cases}$$

Раскрывая скобки получим:

$$\begin{cases} \sum_{t=1}^n Y_t - \hat{a}n - \hat{b} \sum_{t=1}^n X_t = 0, \\ \sum_{t=1}^n X_t Y_t - \hat{a} \sum_{t=1}^n X_t - \hat{b} \sum_{t=1}^n X_t^2 = 0. \end{cases}$$

Из первого уравнения системы имеем оценку параметра a :

$$\hat{a} = \frac{1}{n} \sum_{t=1}^n Y_t - \hat{b} \frac{1}{n} \sum_{t=1}^n X_t = \bar{y} - \hat{b}\bar{x}.$$

Преобразуем второе уравнение системы и подставим полученную оценку \hat{a}

$$\sum_{t=1}^n X_t Y_t - \hat{a} \sum_{t=1}^n X_t - \hat{b} \sum_{t=1}^n X_t^2 = n\bar{xy} - \hat{a}\bar{x} - n\hat{b}(S_x^2 + \bar{x}^2) = 0$$

$$\bar{xy} - \bar{x}\bar{y} + \hat{b}\bar{x}^2 - \hat{b}S_x^2 - \hat{b}\bar{x}^2 = 0$$

Отсюда получаем оценку параметра b :

$$\hat{b} = \frac{\bar{xy} - \bar{x}\bar{y}}{S_x^2} = \frac{\bar{xy} - \bar{x}\bar{y}}{S_x S_y} \frac{S_y}{S_x} = r \frac{S_y}{S_x},$$

Таким образом решение системы уравнений имеет вид

$$\begin{cases} \hat{b} = r \frac{S_y}{S_x} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases}$$

Уравнение регрессии Y на X имеет вид:

$$\bar{y}_x = \bar{y} + r \frac{S_y}{S_x} (x - \bar{x})$$

Уравнение регрессии X на Y имеет вид

$$\bar{x}_y = \bar{x} + 1/r \frac{S_x}{S_y} (y - \bar{y})$$

Заметим, что каждое из уравнений имеет \bar{x} своим решением, т.е. графики проходят через точку (\bar{x}, \bar{y})

5.3 Выборочный коэффициент корреляции, его свойства

Определения:

$$r_{\text{выб}}(\text{далее } r_v) = \frac{\bar{xy} - \bar{x} * \bar{y}}{\sqrt{S_x^2} \sqrt{S_y^2}}$$

Свойства выборочного коэффициента корреляции:

1. $|\tilde{r}| \leq 1$
2. Если $|\tilde{r}| \approx 0$, то x и y слабо коррелируемы. Если известно, что x и y независимы, то $r = 0$.
3. ξ, η линейно связаны $\iff |\tilde{r}| = 1$.

5.4 Проверка гипотезы о значимости выборочного коэффициента корреляции

Рассм. выборочный коэффициент корреляции $r = \frac{\bar{xy} - \bar{x} \cdot \bar{y}}{S_x S_y}$, где S_x, S_y — смещенные оценки дисперсии.

Значимость коэффициента корр. опр. по шкаламе $H_0: "r=0"$ против гип. $H_1: "r \neq 0"$. Ур. знач-ти $\alpha = (0,05; 0,01)$.

$$t_{\text{набл.}} = r \sqrt{\frac{n-2}{1-r^2}} \quad t_{\text{кр.}} \sim T\left(\frac{n-2}{2}; n-2\right) - p\text{-е Стьюдента}$$

Если $|t_{\text{набл.}}| < t_{\text{кр.}}$, то гип. H_0 следует принять. Отп. о., коэффициент корр. недоказано отн. от 0 \Rightarrow ур-е регрессии недоказано. Иначе ($|t_{\text{набл.}}| > t_{\text{кр.}}$) принять H_1 (H_0 отвергнуто): r значимо отн. от 0.

5.5 Теорема Гаусса-Маркова

Теорема Гаусса-Маркова: пусть $(X_t, Y_t), t=1, n$, — рез-т набл.. Предн., что $Y = f(X)$, но в рез-те набл. знач. Y_t изм-т с ошибкой, т.е. $Y_t = f(X_t) + \varepsilon_t$. Пусть $f(X)$ — линейная ф-ция. Тогда $Y_t = a + bX_t + \varepsilon_t, t=1, n$, Y_t — зависимая пер. (объясняемая), X_t — независ. пер. (регрессор), ε_t — си. вен. (сл. ошибка).

1) Спецификация регрессионной модели

$$Y_t = a + bX_t + \varepsilon_t, t=1, n$$

2) X_t — детерминированная вен.

3) $M\varepsilon_t = 0 \quad M\varepsilon_t^2 = D\varepsilon_t = \sigma^2$ не зав. от t

$M\varepsilon_t \varepsilon_s = 0, s \neq t$ (некорр-ть ошибок модели)

недав-ть \Rightarrow корр-ть \Leftarrow обратное неверно

3*) $\varepsilon_t \sim N(0; \sigma^2)$ — норм. модель

Теорема Гаусса-Маркова: в пред-ях 1, 2, 3 оценки \hat{a} и \hat{b} , полученные МНК, имеют норм. дисперсию в классе норм. оценок.

5.6 Статистические свойства МНК оценок

Статистические свойства МНК - оценок.

Будет рассмотрена линейная регрессионная модель, т.е. $\varepsilon_t \sim N(0; \sigma^2)$.
Пусть $\bar{\varepsilon} \sim N(0; \sigma^2 I_n)$, где $\bar{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ - вектор ошибок модели, ε_t - н.о. р. с. в..

1) $y_t = a + b x_t + \varepsilon_t \Rightarrow M y_t = a + b x_t$
 $\vec{y} \sim N(a + b \bar{x}; \sigma^2 I_n), \quad \vec{y} = (y_1, \dots, y_n)$

$$\hat{a} \sim N(a; \sigma^2 \frac{\sum x_t^2}{n \sum x_t^2}) \quad \hat{b} \sim N(b; \sigma^2 \frac{1}{\sum x_t^2})$$

2) Дисперсия ошибок ε_t
 $\bar{\varepsilon} \sim N(0; \sigma^2 I_n); \quad \frac{(n-2)S^2}{\sigma^2} \sim \chi^2(n-2)$
 $S^2 = \sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^2; \quad \bar{\varepsilon} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t$

3) Дисперсия ошибки S^2 , оценки \hat{a} и \hat{b} - независимые с.в..

5.7 Анализ вариации

Анализ вариации зависимой переменной в регрессии.

$y_t = a + b x_t + \varepsilon_t$ наблюд.

$\hat{y} = \hat{a} + \hat{b} x_t$ предсказ. (прогноз.)

Рассмотрим дисперсию

$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (y_t - \hat{y}_t + \hat{y}_t - \bar{y})^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 + \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 + 2 \sum (y_t - \hat{y}_t)(\hat{y}_t - \bar{y})$$

$$\sum (y_t - \bar{y})^2 = \sum (y_t - \hat{y}_t)^2 + \sum (\hat{y}_t - \bar{y})^2$$

Total Sum of Squares Error Sum of Squares Regression Sum of Squares

Частота - м. детерминации $R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}, \quad 0 \leq R^2 \leq 1$.

$R^2 = 1 - \text{Частота наименее объясняемой вариацией}$

$R^2 \approx 0 - \text{Частота неизменности}$

R^2 имеет распред. Фишера $F(n, m) = \frac{\frac{F}{n}(\sum_{t=1}^n (\hat{y}_t - \bar{y})^2)}{\frac{m}{n}(\sum_{t=1}^n (y_t - \hat{y}_t)^2)}$