

Теория вероятностей и математическая статистика. БПИ201.

CheetSheet по статистике

Автор: Сурова София, БПИ191

21 декабря 2021

Точечные оценки

1) Оценка математического ожидания, выборочное среднее

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

2) Оценка дисперсии, смещённая выборочная дисперсия

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

3) Оценка дисперсии, несмещённая выборочная дисперсия

$$\tilde{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

4) Оценка ковариации, выборочная ковариация

$$\hat{k}_{X,Y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

5) Оценка коэффициента корреляции, выборочный коэффициент корреляции

$$\hat{\rho}_{X,Y} = \frac{\hat{k}_{X,Y}}{\sqrt{s_X^2 s_Y^2}}$$

Метод максимального правдоподобия для оценивания параметров

Функция правдоподобия для непрерывной случайной величины

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

Функция правдоподобия для дискретной случайной величины, где $P(x_i, \theta) = P(X = x_i)$

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P(x_i, \theta)$$

Найти оценку $\hat{\theta}$ максимального правдоподобия можно, выразив параметр в уравнении

$$\frac{\partial \ln L(x_1, \dots, x_n, \theta)}{\partial \theta} = 0$$

Доверительный интервал

$1 - \alpha$ = уровень доверия, уровень надежности, доверительная вероятность

Пусть $X \sim N(\theta_1, \theta_2^2)$, n - количество наблюдений в выборке X_1, \dots, X_n

1) оцениваем математическое ожидание θ_1

1.1) дисперсия известна $\theta_2^2 = \sigma^2$

по теореме Фишера центральная статистика

$$G = \frac{(\bar{x} - \theta_1)\sqrt{n}}{\sigma} \sim N(0, 1)$$

$$P(-z_{1-\alpha/2} < G < z_{1-\alpha/2}) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\bar{x} - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}; \bar{x} + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}} \right)$$

1.2) дисперсия неизвестна

по теореме Фишера центральная статистика

$$G = \frac{(\bar{x} - \theta_1)\sqrt{n-1}}{\sqrt{s^2}} \sim t(n-1)$$

$$P(t_{n-1, \alpha/2} < G < t_{n-1, 1-\alpha/2}) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\bar{x} - \frac{t_{n-1, 1-\alpha/2} \sqrt{s^2}}{\sqrt{n-1}}; \bar{x} + \frac{t_{n-1, 1-\alpha/2} \sqrt{s^2}}{\sqrt{n-1}} \right)$$

2) оцениваем дисперсию θ_2^2

2.1) математическое ожидание известно $\theta_1 = \mu$

центральная статистика

$$G = \sum_{i=1}^n \left(\frac{x_i - \mu}{\theta_2} \right)^2 \sim \chi^2(n)$$

$$P(\chi_{n, \alpha/2}^2 < G < \chi_{n, 1-\alpha/2}^2) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n, 1-\alpha/2}^2}; \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n, \alpha/2}^2} \right)$$

2.2) математическое ожидание неизвестно

по теореме Фишера центральная статистика

$$G = \frac{ns^2}{\theta_2^2} \sim \chi^2(n-1)$$

$$P(\chi_{n-1, \alpha/2}^2 < G < \chi_{n-1, 1-\alpha/2}^2) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\frac{ns^2}{\chi_{n-1, 1-\alpha/2}^2}; \frac{ns^2}{\chi_{n-1, \alpha/2}^2} \right)$$

Пусть $X \sim N(\mu_1, \sigma_1^2)$ и $Y \sim N(\mu_2, \sigma_2^2)$ независимы

3) оцениваем разность математических ожиданий $\mu_1 - \mu_2$ 3.1) дисперсии известны σ_1^2, σ_2^2
центральная статистика

$$G = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

$$P(-z_{1-\alpha/2} < G < z_{1-\alpha/2}) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\bar{x} - \bar{y} - z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \bar{x} - \bar{y} + z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

3.2) дисперсии неизвестны, но известно, что они равны $\sigma_1^2 = \sigma_2^2$
центральная статистика

$$G = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{s_{XY}^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \quad s_{XY}^2 = \frac{n_1 s_X^2 + n_2 s_Y^2}{n_1 + n_2 - 2}$$

$$P(t_{n_1+n_2-2, \alpha/2} < G < t_{n_1+n_2-2, 1-\alpha/2}) = 1 - \alpha$$

тогда доверительный интервал

$$\left(\bar{x} - \bar{y} - t_{n_1+n_2-2, 1-\alpha/2} \sqrt{s_{X,Y}^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; \bar{x} - \bar{y} + t_{n_1+n_2-2, 1-\alpha/2} \sqrt{s_{X,Y}^2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Проверка гипотез

0. Фиксируем уровень значимости α

1. Определяем основную гипотезу H_0 и альтернативную гипотезу H_A
2. Выбираем статистику
3. Определяем, какое будет распределение у статистики, если верна основная гипотеза
4. Определяем границы критической и доверительной области
5. Вычисляем значение статистики на нашей выборке. Если оно попало в критическую область \Rightarrow есть основания отвергнуть нулевую гипотезу в пользу альтернативной.

Проверка гипотез о параметрах нормального распределения

Пусть $X \sim N(\mu, \sigma^2)$, n - количество наблюдений в выборке X_1, \dots, X_n

- гипотеза о математическом ожидании
- дисперсия известна σ^2

1. $H_0 : \mu = m_0$
 $H_1 : \mu < m_0$
 $H_2 : \mu > m_0$
 $H_3 : \mu \neq m_0$
2. $T = \frac{(\bar{x} - m_0)\sqrt{n}}{\sigma}$
3. $T|_{H_0} \sim N(0, 1)$
4. для H_1 Д.О. $[z_\alpha, +\infty)$
для H_2 Д.О. $(-\infty, z_{1-\alpha}]$
для H_3 Д.О. $[z_{\alpha/2}, z_{1-\alpha/2}]$
Примечание. $z_\alpha = -z_{1-\alpha}$

- гипотеза о дисперсии
- математическое ожидание известно μ

1. $H_0 : \sigma^2 = \sigma_0^2$
 $H_1 : \sigma^2 \neq \sigma_0^2$
2. $T = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0^2} \right)^2$
3. $T|_{H_0} \sim \chi^2(n)$
4. для H_1 Д.О. $[\chi_{\alpha/2}^2, \chi_{1-\alpha/2}^2]$

- гипотеза о математическом ожидании
- дисперсия неизвестна

1. $H_0 : \mu = m_0$
 $H_1 : \mu < m_0$
 $H_2 : \mu > m_0$
 $H_3 : \mu \neq m_0$
2. $T = \frac{(\bar{x} - m_0)\sqrt{n-1}}{\sqrt{s^2}}$
3. $T|_{H_0} \sim t(n-1)$
4. для H_1 Д.О. $[t_\alpha, +\infty)$
для H_2 Д.О. $(-\infty, t_{1-\alpha}]$
для H_3 Д.О. $[t_{\alpha/2}, t_{1-\alpha/2}]$
Примечание. $t_\alpha = -t_{1-\alpha}$

- гипотеза о дисперсии
- математическое ожидание неизвестно

1. $H_0 : \sigma^2 = \sigma_0^2$
 $H_1 : \sigma^2 \neq \sigma_0^2$
2. $T = \frac{ns^2}{\sigma_0^2}$
3. $T|_{H_0} \sim \chi^2(n-1)$
4. для H_1 Д.О. $[\chi_{\alpha/2}^2, \chi_{1-\alpha/2}^2]$

Проверка гипотез о разности равенстве параметров двух нормальных распределений

Пусть $X \sim N(\mu_1, \sigma_1^2)$ и $Y \sim N(\mu_2, \sigma_2^2)$ независимы

- гипотеза о математических ожиданиях

- - известно, что $\sigma_1^2 = \sigma_2^2$

1. $H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2$
 $H_1 : \mu_1 - \mu_2 < 0 \Leftrightarrow \mu_1 < \mu_2$
 $H_2 : \mu_1 - \mu_2 > 0 \Leftrightarrow \mu_1 > \mu_2$
 $H_3 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow \mu_1 \neq \mu_2$

2. $T = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma_0^2} \right)^2$

3. $T|_{H_0} \sim \chi^2(n)$

4. для H_1 Д.О. $[t_\alpha, +\infty)$
для H_2 Д.О. $(-\infty, t_{1-\alpha}]$
для H_3 Д.О. $[t_{\alpha/2}, t_{1-\alpha/2}]$

- гипотеза о дисперсиях

- - математическое ожидание неизвестно

1. $H_0 : \sigma_1^2 = \sigma_2^2$
 $H_1 : \sigma_1^2 \neq \sigma_2^2$

2. $T = \frac{\tilde{s}_X^2}{\tilde{s}_Y^2}$

3. $T|_{H_0} \sim F(n_1 - 1, n_2 - 1)$

4. для H_1 Д.О. $[F_{\alpha/2}, F_{1-\alpha/2}]$

.)

Проверка гипотезы о вероятности успеха в генеральной совокупности

Пусть $X \sim \text{Bi}(1, p)$

1. $H_0 : p = p_0$
 $H_1 : p > p_0$

2. $T = \frac{\sum x_i - np_0}{\sqrt{np_0(1-p_0)}}$

3. $T|_{H_0} \sim N(0, 1)$

4. для H_1 Д.О. $(-\infty, z_{1-\alpha})$

Проверка гипотезы о коррелированности случайных величин

Пусть (X, Y) гауссовский случайный вектор

1. $H_0 : \rho_{xy} = 0$
 $H_1 : \rho_{xy} < 0$ отрицательно коррелированы
 $H_2 : \rho_{xy} > 0$ положительно коррелированы
 $H_3 : \rho_{xy} \neq 0$ коррелированы

2. $T = \frac{\sqrt{n-2}\hat{\rho}_{xy}}{\sqrt{1-\hat{\rho}_{xy}^2}}$

$$\tilde{T} = \sqrt{n}\hat{\rho}_{xy} \text{ при } n \rightarrow \infty$$

3. $T|_{H_0} \sim t(n-2)$
 $\tilde{T}|_{H_0} \sim N(0, 1)$

4. для H_1 Д.О. $[t_\alpha, +\infty)$
для H_2 Д.О. $(-\infty, t_{1-\alpha}]$
для H_3 Д.О. $[t_{\alpha/2}, t_{1-\alpha/2}]$

Примечание. Для \tilde{T} нужно заменить квантили Стьюдента на квантили стандартного распределения, сохранив уровень

Проверка гипотезы о независимости случайных величин

- $H_0 : \forall (i, j) \ P(A = A_i, B = B_j) = P(A = A_i)P(B = B_j)$
 $H_1 : \exists (i, j) \ P(A = A_i, B = B_j) \neq P(A = A_i)P(B = B_j)$
- $T = n \left(\sum_{i=1}^s \sum_{j=1}^r \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right)$
- $T \Big|_{H_0} \sim \chi^2((s-1)(r-1))$
- для H_1 Д.О. $(-\infty, \chi^2_{1-\alpha}]$

$A \backslash B$	B_1	\cdots	B_r	
A_1	n_{11}	\cdots	n_{1r}	$n_{1.}$
\vdots	\vdots	\ddots	\vdots	\vdots
A_s	n_{s1}	\cdots	n_{sr}	$n_{s.}$
	$n_{.1}$	\cdots	$n_{.r}$	n