Теория вероятностей и математическая статистика. CheetSheet по статистике

Автор: Сурова София, БПИ191

декабрь 2022

Точечные оценки

1) Оценка математического ожидания, выборочное среднее

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

2) Оценка дисперсии, смещённая выборочная дисперсия

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

3) Оценка дисперсии, несмещённая выборочная дисперсия

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

4) Оценка коварици, выборочная ковариация

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

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

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

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

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

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

$$\frac{\partial lnL(x_1, ..., x_n, \theta)}{\partial \theta} = 0$$

Доверительный интервал

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

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

1) оцениваем математическое ожидание θ_1
1.1) дисперсия известна $\theta_2^2 = \sigma^2$
по теореме Фишера центральная статистика

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

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

тогда доверительный интервал

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

1.2) дисперсия неизвестна
по теореме Фишера центральная статистика

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

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

тогда доверительный интервал

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

2) оцениваем дисперсию θ_2^2
2.1) математическое ожидание известно $\theta_1 = \mu$
центральная статистика

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

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

тогда доверительный интервал

$$\left(\frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,1-\alpha/2}};\ \frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,\alpha/2}}\right)$$

2.2) математическое ожидание неизвестно
по теореме Фишера центральная статистика

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

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

тогда доверительный интервал

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

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

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

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

тогда доверительный интервал

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

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

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

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

тогда доверительный интервал

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

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 => 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 $X_1, ..., X_n$

- 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. $T = \dfrac{(\overline{x} - m_0)\sqrt{n}}{\sigma}$
3. $T\Big|_{H_0} \sim N(0, 1)$
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 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. $T = \dfrac{(\overline{x} - m_0)\sqrt{n-1}}{\sqrt{s^2}}$
3. $T\Big|_{H_0} \sim t(n-1)$
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 известно μ

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

- гипотеза о дисперсии
- - математическое ожидание неизвестно

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

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. $T = \dfrac{\overline{x} - \overline{y}}{\sqrt{s_{XY}^2}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
   $s_{XY}^2 = \dfrac{n_1 s_X^2 + n_2 s_Y^2}{n_1 + n_2 - 2}$
3. $T\Big|_{H_0} \sim t(n_1 + n_2 - 2)$
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. $T = \dfrac{\widetilde{s}_X^2}{\widetilde{s}_Y^2}$
3. $T\Big|_{H_0} \sim F(n_1 - 1, n_2 - 1)$
4. для H_1 Д.О. $[F_{\alpha/2}, F_{1-\alpha/2}]$

   :)

   ты справишься

Проверка гипотезы о вероятности успеха в генеральной совокупности
Пусть $X \sim \mathbf{Bi}(1, p)$

1. $H_0 : p = p_0$
   $H_1 : p > p_0$
2. $T = \dfrac{\sum x_i - np_0}{\sqrt{np_0(1 - p_0)}}$
3. $T\Big|_{H_0} \sim N(0, 1)$
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
   $H_3 : \rho_{xy} \neq 0$ кореллированы
2. $T = \dfrac{\sqrt{n-2}\hat{\rho}_{xy}}{\sqrt{1 - \hat{\rho}_{xy}^2}}$
   $\widetilde{T} = \sqrt{n}\hat{\rho}_{xy}$ при $n \to \infty$
3. $T\Big|_{H_0} \sim t(n - 2)$
   $\widetilde{T}\Big|_{H_0} \sim N(0, 1)$
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 $\widetilde{T}$ нужно заменить квантили Стьюдента на квантили стандартного распределения, сохранив уровень

Проверка гипотезы о независимости случайных величин

1. $H_0 : \forall (i, j)\ P(A = A_i, B = B_j) = P(A = A_i)P(B = B_j)$
   $H_1 : \exists (i, j)\ P(A = A_i, B = B_j) \neq P(A = A_i)P(B = B_j)$

2. $T = n\left(\sum\limits_{i=1}^{s}\sum\limits_{j=1}^{r}\dfrac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right)$

3. $T\Big|_{H_0} \sim \chi^2((s-1)(r-1))$

4. 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}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_s	n_{s1}	$\cdots$	n_{sr}	$n_{s\cdot}$
	$n_{\cdot 1}$	$\cdots$	$n_{\cdot r}$	n