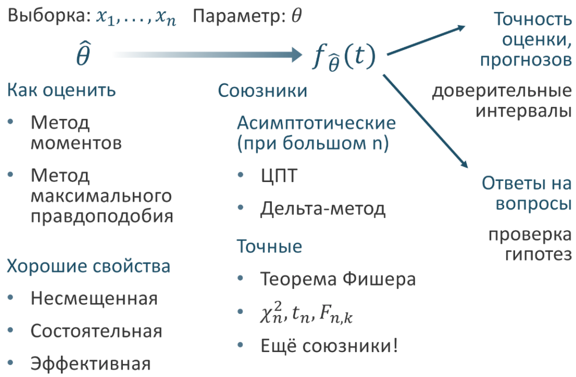
Шпаргалка по математической статистике

Схема математической статистики



Метод Моментов

Чтобы использовать метод моментов, необходимо предположить две вещи:

1. $X_1, ..., X_n \sim iid$, *Inependet and Identically Distributed random values*
2. Предположить расспределение из которого пришла выборка

Момент $E(X_i^k)$ зависит от неизвестного параметра θ: $E(X_i^k) = f(\theta)$
Оценкой метода моментов называется случайная величина: $\hat{\theta}_{MM} = f^{-1}(\overline{X^k})$
Замечание: *Нам позволяет так делать ЗБЧ, который говорит, что* $\overline{X} = E(X)$
Замечание: *Обычно хватает среднего первого порядка, но иногда нужно использовать более высокие порядки (например если среднее равно нулю)*

Свойства оценок

Несмещенность

Оценка называется **несмещённой**, если её математическое ожидание равно оцениваемому параметру: $E(\hat{\theta}) = \theta$

Смещение оценки это разница между её математическим ожиданием и её реальным значением: $bias(\hat{\theta}) = E(\hat{\theta}) - \theta$

Состоятельность

Оценка называется **состоятельной**, если она сходится по вероятности к истинному значению параметра при $n \to \infty$: $\hat{\theta} \xrightarrow{p} \theta$

Асимптотическая несмещенность

Оценка называется асимтотически несмещенной, если её математическое ожидание сходится к оцениваемому параметру при $n \to \infty$, т.е: $E(\hat{\theta}) \xrightarrow[n\to\infty]{} \theta$

Эффективность

Обычно оценки сравнивают между собой с помощью квадратичной ошибки: $MSE = E(\hat{\theta} - \theta)^2$
Чем меньше, тем лучше
Статистик хочет получить:

- несмещенную оценку - хочет в среднем не ошибаться при фиксированном размере выборки
- состоятельную оценку - хочет при большом обьеме выборки быть близко к реальности
- оценку с маленькой средней квадратичной ошибкой

Разложение на смещение и разброс

$MSE = E(\hat{\theta} - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \theta)^2 = Var(\hat{\theta}) + bias^2(\hat{\theta})$

В классе всех возможных оценок наилучшей в смысле среднеквадратического подхода не существует, но можно зафиксировать смещение и найти оценку с наименьшей дисперсией(а.к.а. разбросом). Такая оценка называется эффективной в классе со смещением $bias(\hat{\theta})$. Для функции потерь MSE существует теоретическая нижняя граница, ее называют неравенством Рао-Фреше-Крамера.

Неравенство Рао-Фреше-Крамера

Если оценка параметра несмещена и выполнены условия регулярности:

- Область определения случайной велечины не зависит от параметра θ
- Можно брать производные от функции плотности
- Существует конечная $J(\theta) = E(\frac{\partial \ln f(x,\theta)}{\partial \theta})^2$

То верно: $Var(\hat{\theta}) \geq \frac{1}{n \cdot J(\theta)}$.

Если оказалось, что $Var(\hat{\theta}) = \frac{1}{n \cdot J(\theta)}$., то $\hat{\theta}$ - эффективная оценка для θ
Замечание: *Если* $\hat{\theta}$ *- смещенная, то в числителе вместо 1 стоит* $(1 + bias'\theta)^2$*, а в остальном все аналогично*

Доверительные интервалы

Интервал $[\theta_L; \theta_U]$ называется **доверительным интервалом** для параметра θ, с уровнем доверия $1 - \alpha$, если при бесконечном повторении эксперимента в $100 \cdot (1 - \alpha)$ процентах случаев этот интервал будет накрывать значение параметра θ
Величину α называют уровнем значимости
Зачем нужны доверительные интервалы?

- Точечная оценка делается по случайной выборке, а значит возникает неопределенность (т.к. выборка отличается от случая к случаю)
- Нужно делать выводы в каком-то диапазоне
- Доверительный интервал показывает, насколько мы уверены в точечной оценке

Замечание: *Обычно ищут наиболее короткий интервал, т.к. уже значит точнее*

Асимптотический доверительный интервал для среднего

- ЦПТ позволяет построить интервал для любого среднего т.к. для ЦПТ не важно распределение величины
- Наблюдаем $X_1, ..., X_n$
- **Предполагаем:** X_i независимо и одинаково распределены, число n велико, **нет выбросов**

$\overline{x} \overset{asy}{\sim} N(\mu, \frac{\sigma^2}{n}) \underset{\text{цент.}}{\Leftrightarrow} \overline{x} - \mu \overset{asy}{\sim} N(0, \frac{\sigma^2}{n}) \underset{\text{норм.}}{\Leftrightarrow} \frac{\overline{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \overset{asy}{\sim} N(0, 1)$

$p(\overline{x} - z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}) = 1 - \alpha$
или более просто: $\overline{x} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}$

Дельта-метод

Если:
$X_1, ..., X_n \sim iid$, $E(X_1) = \mu, Var(X_1) = \sigma^2$ и $g(t)$ - дифференцируемая функция, тогда:
$g(\overline{x}) \sim N(g(\mu), \frac{\sigma^2}{n} \cdot g'(\mu)^2)$,
тогда доверительный интервал: $g(\overline{x}) \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} g'(\mu)$

Асимптотический доверительный интервал для дисперсии

$s^2 = \frac{n}{n-1}\hat{\sigma}^2 \Rightarrow s^2 \sim N(\sigma^2, \frac{\mu_4 - \sigma^4}{n})$, $\mu_4 = E[(X_i - \mu)^4]$ Затем можно построить доверительный интервал, более подробно об интервалах для дисперсий смотри в разделе точных интервалов

Хи-квадрат распределение

$X_1, ..., X_k \sim iid\ N(0,1)$
$Y = X_1^2 + ... + X_k^2 \sim \chi_k^2$
Характеристики:
$E(Y) = k,\ Var(Y) = 2k$
Плотность: $f(x) = \frac{1}{2^{\frac{k}{2}} \cdot \Gamma(\frac{k}{2})} \cdot x^{\frac{k}{2}-1} \cdot e^{-\frac{x}{2}},\ x \geq 0$

Распределение Стьюдента

$X_0 \sim iid\ N(0,1),\ Y \sim X_k^2$
$Z = \frac{X_0}{\sqrt{Y/k}} \sim t(k)$
Характеристики:
$E(Z) = 0,\ Var(Z) = \frac{k}{k-2},\ k > 2$
Плотность: $f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi k} \cdot \frac{k}{2}} \cdot (1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$

Распределение Фишера

$X \sim X_k^2,\ Y \sim X_k^2$
$Z = \frac{\sqrt{X/k}}{\sqrt{Y/m}} \sim F(k,m)$
Характеристики:
$E(Z) = \frac{m}{m-2},\ m > 2,\ Var(Z) = \frac{2m^2(k+m-2)}{k(m-2)^2(m-4)},\ m > 4$
Плотность: *тут очень большая формула*

Теорема Фишера

Пусть $X_1, ..., X_k \sim iid\ N(0,1)$, тогда

1. Выборочное среднее $\overline{x}$ и дисперсия s^2 независимы
2. $\frac{(n-1) \cdot s^2}{\sigma^2}$ имеет χ^2 - распределение с $n-1$ степенью свободы т.е. $\frac{(n-1) \cdot s^2}{\sigma^2} \sim \chi_{n-1}^2$

Точные доверительные интервалы для нормальных выборок

Обязательное условие: $X_1, ..., X_n \sim iid\ N(\mu, \sigma^2)$

Точный доверительный интервал для среднего

- **Дисперсия известна** σ^2 - известна и
  $\hat{\mu} = \overline{x} = \frac{X_1 + ... + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$
  Тогда доверительный интервал:
  $p(\overline{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$ или
  проще: $\overline{x} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- **Дисперсия не известна** σ^2 - не известна и
  $\hat{\mu} = \overline{x} \sim N(\mu, \frac{\sigma^2}{n})$
  $\frac{\overline{x} - \mu}{\sqrt{\frac{s^2}{n}}} \sim t(n-1)$
  Тогда доверительный интервал:
  $p(\overline{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}) = 1 - \alpha$ или
  проще: $\overline{x} \pm t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

Точный vs Асимптотический

Асимптотический	Точный
• Союзник: ЦПТ	• Союзники: теорема Фишера, t-распределение
• Работает при большом n	• Работает при любом n
• Выборка независимая, без аномалий	• Выборка независимая из нормального распределения

Точные доверительные интервалы для разности средних

Дисперсии неизвестны (асимптотика)

$\overline{x} - \overline{y} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$

Дисперсии известны (точный)

$\overline{x} - \overline{y} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$

Дисперсии неизвестны, но равны (точный)

$\overline{x} - \overline{y} \pm t(n_x + n_- 2)_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n_x} + \frac{s^2}{n_y}}$

Дисперсии неизвестны и не равны (примерный)

$\overline{x} - \overline{y} \pm t(v)_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$

где v - приближенное распределение Уэлча

Точные доверительные интервалы для дисперсии

- Мат. ожидание известно
  $p(\frac{n \cdot s^2}{\chi_n^2(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{n \cdot s^2}{\chi_n^2(\frac{\alpha}{2})}) = 1 - \alpha$
- Мат. ожидание НЕизвестно
  $p(\frac{(n-1) \cdot s^2}{\chi_{(n-1)}^2(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{(n-1) \cdot s^2}{\chi_{(n-1)}^2(\frac{\alpha}{2})}) = 1 - \alpha$

Точный доверительный интервал для отношения дисперсий независимых выборок

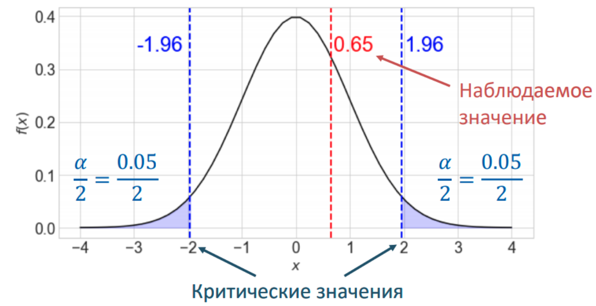
$\frac{s_m^2}{s_n^2} F_{n-1,m-1}(\frac{\alpha}{2}) \leq \frac{\sigma_m^2}{\sigma_n^2} \leq \frac{s_m^2}{s_n^2} F_{n-1,m-1}(1 - \frac{\alpha}{2})$

Проверка гипотез

Гипотеза - утверждение, которое возникло в нашей голове и которое мы собираемся проверить по данным.

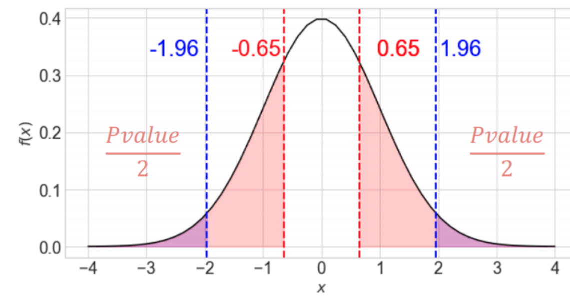
Процедура проверки гипотез

1. Фиксируем уровень значимости α (вероятность зря отвергнуть нулевую гипотезу)
2. Формируем нулевую H_0 и альтернативную H_a гипотезы
3. Выбираем союзника для проверки гипотезы
4. Находим наблюдаемое значение
5. Находим критическое значение с помощью союзника
6. Сравниваем наблюдаемое значение с критическим и делаем выводы



p-значение

p-значение - это достигаемый уровень значимости



- если $p\_value > \alpha$, то наблюдаемое значение попало в область между кретическими, а значит гипотеза **не отвергается**
- если $p\_value < \alpha$, то наблюдаемое значение не попало в область между кретическими, а значит гипотеза **отвергается**

Вопрос: *Какой уровень значимости надо брать, чтобы гипотеза впервые отверглась?*
Ответ: *равный p-значению*

Ошибки 1-ого и 2-ого рода

	H_0 верна	H_0 неверна	
H_0 не отвергается	ok	β	ошибка 2 рода
H_0 отвергается	α	ok	
	ошибка 1 рода		

$\alpha = P(H_0$ отвергнута $|$ H_0 верна$)$
$\beta = P(H_0$ не отвергнута $|$ H_0 не верна$)$
Величину $1 - \beta$ называют **мощностью** критерия
Замечание: *При уменьшении ошибки первого рода всегда возрастает ошибка второго рода*

Параметрические критерии

Включают в себя расчёт параметров конкретного распределения (т.е. предполагаем, что данные пришли из определенного распределения)

Критерии о долях

Z-критерий для доли

$X_1, ..., X_N \sim iidBern(p)$
$H_0 : p = p_0$
$H_a : p \neq p_0$
ЦПТ: $\hat{p} \underset{H_0}{\overset{asy}{\sim}} N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

Критерий для проверки: $z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$

Z-критерий для разности независимых долей

$X_1, ..., X_N \sim iidBern(p_x)$
$Y_1, ..., Y_N \sim iidBern(p_y)$
Выборки независимые
$H_0 : p_x = p_y$, $H_a : p_x \neq p_y$
$z = \frac{\hat{p_x}-\hat{p_y}}{\sqrt{P(1-P)\cdot(\frac{1}{n_x}+\frac{1}{n_y})}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$, $P = \frac{m_x+m_y}{n_x+n_y}$, где m_i - число 1 в выборке

Z-критерий для разности зависимых долей

$X_1, ..., X_N \sim iidBern(p_x)$
$Y_1, ..., Y_N \sim iidBern(p_y)$
Выборки независимые
$H_0 : p_x = p_y$, $H_a : p_x \neq p_y$

		X	
		0	1
Y	0	a	b
	1	c	d

$z = \frac{c-b}{\sqrt{c+b-\frac{(c-b)^2}{n}}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$

Критерии о средних

Z-критерий для средних

$X_1, ..., X_N \sim iid(\mu, \sigma^2)$
$H_0 : \mu = \mu_0$
$H_a : \mu \neq \mu_0$
ЦПТ: $\overline{x} \underset{H_0}{\overset{asy}{\sim}} N\left(\mu_0, \frac{\hat{\sigma^2}}{n}\right)$

Критерий для проверки: $z = \frac{\overline{\mu}-\mu_0}{\sqrt{\frac{\hat{\sigma^2}}{n}}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$

Замечание: *Данный критерий может быть использован для любых распределений, а вот следующий только для нормального*

t-критерий для средних

$X_1, ..., X_N \sim iidN(\mu, \sigma^2)$
$H_0 : \mu = \mu_0$
$H_a : \mu \neq \mu_0$
σ^2 - известна
ЦПТ: $\overline{x} \underset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right)$
Критерий для проверки: $z = \frac{\overline{\mu}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} \underset{H_0}{\sim} N(0,1)$

σ^2 - НЕ известна
ЦПТ: $\overline{x} \underset{H_0}{\sim} N\left(\mu_0, \frac{\sigma^2}{n}\right)$
Критерий для проверки: $z = \frac{\overline{\mu}-\mu_0}{\sqrt{\frac{s^2}{n}}} \underset{H_0}{\sim} t(n-1)$

Z-критерий для разности средних

$X_1, ..., X_N \sim iid(\mu_1, \sigma_1^2)$
$Y_1, ..., Y_N \sim iid(\mu_2, \sigma_2^2)$
Выборки независимые
$H_0 : \mu_1 = \mu_2$
$H_a : \mu_1 \neq \mu_2$
ЦПТ: $\overline{X_1} - \overline{X_2} \underset{H_0}{\overset{asy}{\sim}} N\left(0, \frac{s_1^2}{n} + \frac{s_2^2}{m}\right)$
Критерий для проверки: $z = \frac{\overline{X_1}-\overline{X_2}}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}} \underset{H_0}{\overset{asy}{\sim}} N(0,1)$

Точные критерии для разности средних

Критерий для проверки:

дисперсии известны	$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \underset{H_0}{\sim} N(0,1)$	Нормальное распределение
дисперсии неизвестны, но равны	$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s^2}{n}+\frac{s^2}{m}}} \underset{H_0}{\sim} t(n+m-2)$	Распределение Стьюдента
дисперсии неизвестны	$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \underset{H_0}{\sim} t(\nu)$ $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)}+\frac{s_2^4}{n_2^2(n_2-1)}}$	Распределение Уэлча

Критерии о дисперсиях

χ^2 - критерий для дисперсии

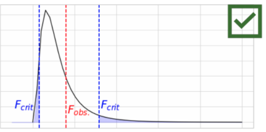
$X_1, ..., X_N \sim iidN(\mu, \sigma^2)$
μ - известно
$H_0 : \sigma^2 = \sigma_0^2$
$H_a : \sigma^2 > \sigma_0^2$
Критерий для проверки:
$\frac{n*s^2}{\sigma_0^2} = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma_0^2} \underset{H_0}{\sim} \chi_n^2$

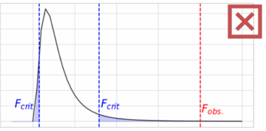
- Высокая дисперсия связана с риском и нестабильностью
- Мы хотим знать, принимает ли дисперсия своё значение ниже σ_0^2
- Из-за этого в качестве альтернативы обычно используют правостороннюю

μ - НЕ известно
$H_0 : \sigma^2 = \sigma_0^2$
$H_a : \sigma^2 > \sigma_0^2$
Критерий для проверки:
$\frac{(n-1)*s^2}{\sigma_0^2} = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma_0^2} \underset{H_0}{\sim} \chi_{(n-1)}^2$

Тест Фишера для отношения дисперсий

$X_1, ..., X_n \sim iid\ N(\mu_x, \sigma_x^2)$
$Y_1, ..., Y_m \sim iid\ N(\mu_y, \sigma_y^2)$
Выборки независимые
H_0: $\sigma_x = \sigma_y$
H_a: $\sigma_x \neq \sigma_y$
Критерий для проверки:
$\frac{s_x^2 \cdot \sigma_y^2}{s_y^2 \cdot \sigma_x^2} \underset{H_0}{=} \frac{s_x^2}{s_y^2} \underset{H_0}{\sim} F_{n-1,m-1}$





Критерий точный, используется предположение о нормальности выборки

Непараметрические критерии

Такие критерии помогают не зацикливаться на том из какого распределения пришла выборка

Критерий знаков

Идея: Превратить выборку в нули и елиницы => можем использовать биномиальное распределение
Минусы: Теряем часть информации зашитой в выборку
Пример: Критерий знаков(одновыборочный)
$X_1, ..., X_n \sim iid$
$H_0 : Median(X) = m_0$
$H_a : Median(X) \neq m_0$
Критерий для проверки:
$T = \sum_{i=1}^{n} [X_i > m_0] \sim Bin(0.5, n)$
Пример: Критерий знаков(двухвыборочный)
$X_1, ..., X_n \sim iid$ $Y_1, ..., Y_n \sim iid$
Выборки связанные
$H_0 : P(X > Y) = 0.5$
$H_a : P(X > Y) \neq 0.5$
Критерий для проверки:
$T = \sum_{i=1}^{n} [X_i > Y_i] \sim Bin(0.5, n)$

Ранговые критерии

$x_1, x_2, ..., x_n$ - выборка
$x_{(1)} \leq x_{(3)} \leq ... \leq x_{(n)}$ - упорядочим по возрастанию
Правила выставления ранга:

1. Порядковый номер наблюдения - ранг
2. Если встречаются несколько одинаковых значений, им присваивается одинаковое значение рагна, равное среднему арифметическому их порядквых номеров

Критерий Уилкоксона (одновыборочный)
$X_1, ..., X_n \sim iid$
$F_X(x)$ - симметрична относительно медианы
$H_0 : Median(X) = m_0$
$H_a : Median(X) \neq m_0$
Критерий для проверки:
$W = \sum\limits_{i=1}^{n} rank(|X_i - m_0|) * sign(X_i - m_0)$ - табличное распределение
Критерий Уилкоксона (двухвыборочный)
Предполагаем, что распределения одинаковые, хотим проверить сдвинутые ли они $X_1, ..., X_n \sim iid$
$Y_1, ..., Y_n \sim iid$
Выборки связанные
$H_0 : Median(X - Y) = 0$
$H_a : Median(X - Y) \neq 0$
Критерий для проверки:
$W = \sum\limits_{i=1}^{n} rank(|X_i - Y_i|) * sign(X_i - Y_i)$ - табличное распределение
Критерий Манна-Уитни(двухвыборочный)
$X_1, ..., X_{n_x} \sim iid$
$Y_1, ..., Y_{n_y} \sim iid$
Выборки независимые
$n_x \leq n_y$
$H_0 : f_X(x) = f_Y(y)$
$H_a : f_X(x) = f_Y(y + m), m \neq 0$
Объединим обе выборки в одну общую и посчитаем для всех чисел ранги
Критерий для проверки:
$W = \sum\limits_{i=1}^{n_x} rank(X_i)$ - табличное распределенрие

Бутстрап

Идея метода: имеющаяся выборка - это единственная информация об истинном распределении данных

Доверительный интервал Эфрона

Бутстрапируем: оценку неизвестного параметра
Сэмплируем: $x_1^*, \dots, x_n^*$
Считаем: $\hat{\theta}^*$
Повторяем: B раз
Строим распределение: $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$
Интервал: $[\, \hat{\theta}^*_{\frac{\alpha}{2}};\ \hat{\theta}^*_{1-\frac{\alpha}{2}} \,]$

Если распределение выборки несимметрично, такой доверительный интервал усиливает смещение, присущее изначальной выборке

Доверительный интервал Холла

Бутстрапируем: Отклонение оценки от истинного значения
Сэмплируем: $x_1^*, \dots, x_n^*$
Считаем: $\hat{q}_i^* = \hat{\theta}^* - \hat{\theta}$
Повторяем: B раз
Строим распределение: $\hat{q}_1^*, \dots, \hat{q}_B^*$
Интервал: $[\hat{\theta} - \hat{q}^*_{1-\frac{\alpha}{2}};\ \hat{\theta} - \hat{q}^*_{\frac{\alpha}{2}}]$

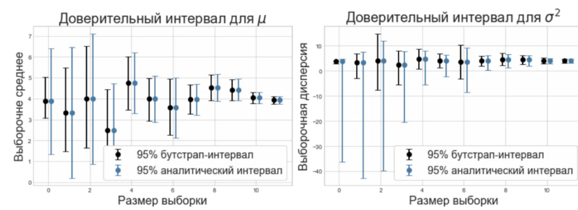
t-процентильный доверительный интервал

Бутстрапируем: t - статистику
Сэмплируем: $x_1^*, \dots, x_n^*$
Считаем: $t_i^* = \dfrac{(\hat{\theta}^* - \hat{\theta})}{se(\hat{\theta}^*)}$
Повторяем: B раз
Строим распределение: $t_1^*, \dots, t_B^*$
Интервал: $[\, \hat{\theta} - t^*_{1-\frac{\alpha}{2}} \cdot se(\hat{\theta});\ \hat{\theta} - t^*_{\frac{\alpha}{2}} \cdot se(\hat{\theta})]$

Если бутстрапировать $|\ \hat{\theta}^* - \hat{\theta}\ |$, можно получить симметричный интервал

Проблемы бутстрапа

- Чтобы бутстрап сработал, выборка должна быть репрезентативной



- Если исходная выборка маленькая, бутстраповский доверительный интервал будет уже аналитического, так как в выборке недостаточно "неопределенности"
- Если в данных есть структура (регрессия, временные ряды), бутстрап нужно устроить так, чтобы учитывать её $\Rightarrow$ разные виды бутстрапа
- Бутстрап ненадёжно работает в хвостах распределения из-за маленького числа наблюдений: мы можем хорошо оценить медиану, но не 99% квантиль
- Если у распределения тяжёлые хвосты, бутстрап может работать некорректно и в средиземье

Проверка гипотез при прмощи бутсрапа

H_0: $\theta = \theta_0$
H_a: $\theta > \theta_0$

t-статистика: $t = \dfrac{(\hat{\theta} - \theta_0)}{se(\hat{\theta})}$

Бутстрап-аналог: $t^* = \dfrac{(\hat{\theta}^* - \hat{\theta})}{se(\hat{\theta}^*)}$

- Гипотеза отвергается, если $t_{obs} > t^*_{1-\alpha}$
- По аналогии можно проверять гипотезы против других альтернатив
- Для более сложных гипотез есть специальные бутстраповские алгоритмы проверки

Критерии согласия

Эмпирическая функция распределение

Функция распределения – функция, которая определяет вероятность события $X \leq x$, то есть $F(x) = P(X \leq x)$
Эмпирическая функция распределения – функция, которая определяет для каждого x частоту события $X \leq x$, то есть $\hat{F}_n(x) = \hat{P}(X \leq x) = \frac{1}{n}\sum\limits_{i=1}^{n}[X_i \leq x]$, [] - индикатор

Свойства эмпирической функции распределения

1. Несмещённость: $\mathbb{E}(\hat{F}_n(x)) = F_X(x)$
2. Состоятельность: $\operatorname*{plim}\limits_{n\to\infty} \hat{F}_n(x) = F_X(x)$
3. Асимптотическая нормальность:

$$\hat{F}_n(x) \overset{asy}{\sim} N\left(F_X(x), \frac{\hat{F}_n(x)\cdot\left(1-\hat{F}_n(x)\right)}{n}\right)$$

Критерий Колмогорова

Используется для проверки гипотез о непрерывных распределениях

- Критерии согласия - критерий о виде неизвестного закона распределения: $H_0 : X \sim F_x(x)$, H_a : гипотеза H_0 неверна
- Распределение случайной величины описывается её функцией распределения: $H_0 : F_x(x) = F_0(x)$, $H_a : F_x(x) \neq F_0(x)$

Статистика Колмогорова:

$$D_n(X_1, \dots, X_n) = \sup_x |\, F_0(x) - \hat{F}_n(x)\, |$$

При справедливости нулевой гипотезы, распределение статистики D_n одинаково для любых **непрерывных** распределений, при этом его функция распределения:

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}\cdot D_n \leq z) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2z^2}$$

Распределение Колмогорова – наш новый союзник, на её основе мы можем построить критерий

$X_1, \dots, X_n \sim iid\ F_X(x)$
H_0: $F_X(x) = F_0(x)$
H_a: $F_X(x) \neq F_0(x)$

Критерий для проверки:

$$K_n = \sqrt{n}\cdot\sup_x |\, F_0(x) - \hat{F}_n(x)\, |$$

$K_n \overset{asy}{\underset{H_0}{\sim}}$ распределение Колмогорова

$K_{набл.} > K_{1-\alpha}$

$K_{набл.} \leq K_{1-\alpha}$

Критерий применяется только для непрерывных распределений

- Критические значения рассчитываются по таблицам, составленным для распределения Колмогорова

Замечание: *Для проверки гипотезы об однородности двух выборок, можно также пользоваться этой идеей, просто считать расстояние между двумя эмпирическими функциями*

Замечание:

- *Чтобы проверять гипотезы о распределениях, нужно научиться считать между ними расстояния*
- *Критерий Колмогорова ищет расстояние как супремум и позволяет проверять гипотезы о распределении и однородности выборок*
- *Критерии Крамера-Мизеса и Андерсона-Дарлинга помогают специфицировать критерий либо для "средиземья" либо для "крайнеземья" (хвостов)*
- *Эти критерии работают только для непрерывных распределений*
- *Неизвестные параметры распределения фиксируются в нулевой гипотезе*
- *Если мы оцениваем параметры по выборке, распределения Колмогорова не будет одинаковым для всех распределений*
- *Для различных распределений можно строить уточнения теста Колмогорова*

Критерий Пирсона

Используется для проверки гипотез о дискретных распределениях

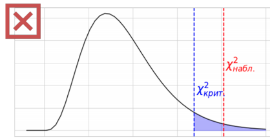
- Выборка $X_1, \dots, X_n$ из дискретного распределения
- Предполагаем, что случайная величина принимает s значений с какими-то вероятностями

X	z_1	z_2	...	z_s	- возможные значения
$\mathbb{P}(X = z)$	$p_1(\theta)$	$p_2(\theta)$	...	$p_s(\theta)$	- теоретические вероятности
$\#(X_i = z)$	ν_1	ν_2	...	ν_s	- эмпирические частоты

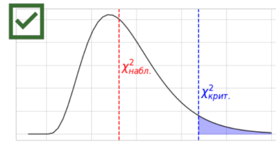
- Нужно сравнить все эмпирические частоты с теоретическими $\hat{\theta}$ - состоятельная оценка параметров распределения, k- их количество

$X_1, \dots, X_n \sim iid F_X(x)$
$H_0 : F_x(x) = F_0(x)$
$H_a : F_x(x) \neq F_0(x)$

Критерий для проверки: $\sum\limits_{i=1}^{s}\frac{(v_i - n*p_i(\hat{\theta}))^2}{n*p_i(\hat{\theta})} \overset{asy}{\underset{H_0}{\sim}} \chi^2_{s-k-1}$



$\chi^2_{набл.} > \chi^2_{s-k-1}(1-\alpha)$



$\chi^2_{набл.} \leq \chi^2_{s-k-1}(1-\alpha)$

Критерий обычно применяют для дискретных распределений

Замечание:

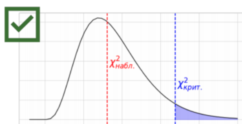
- *Критерий Пирсона не состоятелен против всех альтернатив, т.е. бывают распределения, которые он не может отличить друг от друга*
- *Можно попробовать использовать критерий Пирсона для непрерывных распределений*
- *Для этого нужно будет разбить все возможные значения непрерывной случайной величины на бины (как на гистограмме)*
- *езультат работы теста будет зависеть от числа выбранных бинов $\Rightarrow$ лучше пользоваться для непрерывных распределений другими критериями*

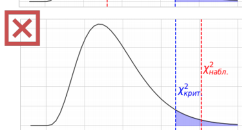
Гипотезы об однородности выборок
Критерий Пирсона по аналогии с критерием Колмогорова можно использовать для проверки гипотез об однородности выборок
$X_1, \dots, X_{n_x} \sim iidF_X(x)$ $Y_1, \dots, Y_{n_y} \sim iidF_Y(x)$
$H_0 : F_X(x) = F_Y(x)$
$H_a : F_X(x) \neq F_Y(x)$

- Нужно сравнивать эмпирические частоты между собой

X, Y	z_1	z_2	...	z_s
$\#(y_i = z)$	w_1	w_2	...	w_s
$\#(x_i = z)$	ν_1	ν_2	...	ν_s

$$\sum_{j=1}^{s}\frac{\left(\frac{v_j}{n_x} - \frac{w_j}{n_y}\right)^2}{n_x + n_y} \overset{asy}{\underset{H_0}{\sim}} \chi^2_{s-1}$$



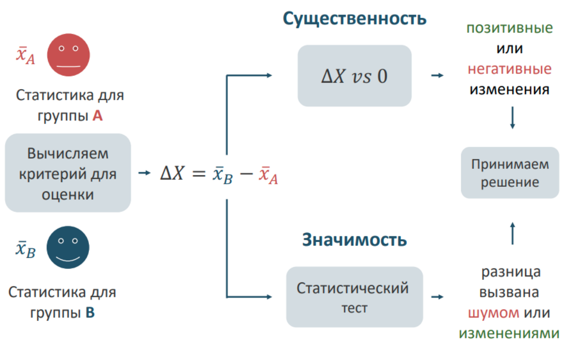


- Можно использовать критерий для непрерывных распределений, но тогда придется дробить данные на бины

Замечание:

- *Его можно использовать и для непрерывных величин, но результат будет зависеть от числа выбранных групп*

A/B тесты



Типичные метрики:

- Уникальные пользователи за сессию
- Клики на пользователя, клики на один запрос
- Среднее время пользователя на сайте
- Возвращаемость пользователя
- Средний чек
- Средний трафик
- Средняя разница между ценой товара и его себестоимостью (маржа)

Значимость – статистический тест говорит нам, что изменения в метрике неслучайны.
Существенность – насколько изменения большие по своей величине, насколько большой размер эффекта (изменение метрики), который мы ловим.
Замечание:
АБ-тест используется для проверки идей на группе пользователей. При проведении АБ-теста мы должны ответить на ряд вопросов:

1. *Что является целевой метрикой?*
2. *На какое увеличение мы рассчитываем?*
3. *Какой критерий мы используем для проверки результата на статистическую значимость?*
4. *Как должен выглядеть дизайн эксперимента, как разбить пользователей на группы?*
5. *Как долго должен идти эксперимент?*

Множественная проверка гипотез

Проблема: Проверяем две гипотезы:
$H_0 : \mu_1 = \mu_2 = \mu_3$
В таком случае можем ошибиться сразу в двух местах(т.е. в одной или в другой):
P(ошибочно отвергнуть хотябы одну из H_0) =
$1 - P$(не ошибиться ни в одной) $= 1-(1-\alpha)^2 = 2\alpha - \alpha^2 > \alpha$
Т.е. получается наша ошибка первого рода растет, а в случае n гипотез будет рост $1-(1-\alpha)^n$
Как можем корректировать?
Неравенстов Бонферонни
$P(A+B) \le P(A) + P(B)$ - т.е. можем проосто тестировать каждую гипотезу на уровне $\frac{\alpha}{n}$ чтобы в худшем случае получить общий уровень значимости α
Минусы?

- Из-за коррекции уровня значимости возникают проблемы с мощностью тестов
- Чем больше гипотез проверяется, тем ниже шансы отклонить неверные гипотезы
- Более того, из-за презумпции нулевой гипотезы для более низкого уровня значимости нам нужно собрать большее число наблюдений, чтобы зафиксировать значимое отклонение от нулевой гипотезы

Вывод: Нужно улучшать метод коррекции
Рассмотрим случай, когда мы проверяем n гипотез

	верных H_{0i}	неверных H_{0i}
не отвергнутых H_{0i}	U	T
отвергнутых H_{0i}	V	S

- Неверно отклонили V гипотез, неверно не отклонили T гипотез
- На практике пытаются контролировать обобщения ошибки первого рода, например: FWER и FDR

Групповая вероятность ошибки, FWER (Family-Wise Error Rate) – это вероятность совершить хотя бы одну ошибку первого рода $FWER = P(V > 0)$
Ожидаемая доля ложны отклонения, FDR (False Discovery Rate) – это математическое ожидание числа ошибок первого рода к общему числу отклонений нулевой гипотезы $FDR = E(\frac{V}{V+S})$
Т.е. FWER пытается контролировать ошибку первого рода, когда отрицаем верную гипотезу, а FDR пытается контролировать обобщение ошибки второго рода

Метод Холма

- Поправка Бонферрони пытается контролировать FWER (вероятность хотя бы одной ошибки 1 рода)
- Метод Холма – улучшение поправки Бонферрони, обладает более высокой мощностью
- Отсортируем гипотезы по получившимся P-значениям по возрастанию: $p_{(1)} \le p_{(2)} \le ... \le p_{(k)}$
- Возьмём для них: $\alpha_{(1)} = \frac{\alpha}{k}, \alpha_{(2)} = \frac{\alpha}{k-1}, ..., \alpha_{(i)} = \frac{\alpha}{k-i+1}, ..., \alpha_{(k)} = \alpha$
- Если $p_{(1)} \ge \alpha_{(1)}$, все нулевые гипотезы не отвергаются, иначе отвергаем первую и продолжаем
- Если $p_{(2)} \ge \alpha_{(2)}$, все оставшиеся нулевые гипотезы не отвергаются, иначе отвергаем вторую и продолжаем
- Идём, пока не кончатся гипотезы
- Метод Холма обеспечивает контроль FWER на уровне α
- Метод Холма оказывается мощнее корректировки Бонферрони, так как его уровни значимости меньше

Метод Бенджамини-Хохберга

- Отсортируем гипотезы по получившимся P-значениям по возрастанию: $p_{(1)} \le p_{(2)} \le ... \le p_{(k)}$
- Возьмём для них: $\alpha_{(1)} = \frac{\alpha}{k}, \alpha_{(2)} = \frac{2\alpha}{k}, ..., \alpha_{(i)} = \frac{i\alpha}{k}, ..., \alpha_{(k)} = \alpha$
- Если $p_{(k)} \le \alpha_{(k)}$, отвергнуть все гипотезы, иначе не отвергнуть k - ую и продолжить
- Если $p_{(k-1)} \le \alpha_{(k-1)}$, отвергнуть все гипотезы, иначе не отвергнуть (k-1) - ую и продолжить
- Идём, пока не кончатся гипотезы
- Для любой процедуры множественного тестирования гипотез FDR $\le$ FWER
- Метод Бенджамини-Хохберга обычно оказывается более мощным, чем методы контролирующие FWER
- Он отвергает не меньше гипотез с теми же α_i
- Это происходит за счёт того, что метод позволяет допустить большее число ошибок первого рода

А сколько наблюдений нужно?

- Необходимое количество наблюдений зависит от размеров ошибок первого и второго рода, а также от размера эффекта
- Фиксируем уровень значимости (ошибку 1 рода), на которую мы согласны
- Подбираем соотношение между минимальным размером эффекта, желаемой мощностью и объёмом выборки
- В выборе соотношении помогает заказчик эксперимента, у него обычно есть ограничения, с которыми нам придётся работать (количество магазинов, длительность АБ-теста и т.п.)

Таблица эффекта-ошибки

Ошибка 1/2 рода $\alpha = \beta$

		0.1%	1%	5%	10%
размер эффекта	1%	много данных			
	1.5%				
	3%				
	5%				
	10%				мало данных

Совокупность этих трёх параметров (ошибка 1/2 рода, размер эффекта) позволяют рассчитать необходимый для эксперимента объём выборки.

Анализ мощности
До эксперимента:

- Какой нужен объём выборки, чтобы найти различия с разумной степенью уверенности
- Различия какой величины мы можем найти, если известен объём выборки

После эксперимента:

- смогли бы мы найти различия с помощью нашего эксперимента, если бы величина эффекта была равна Δ

Метрики для A/B тестирования

- Показатель для улучшения – метрика
- Метрики бывают разными, они конструируются в зависимости от бизнес-задачи
- Иногда метрики привязаны к деньгам
- Чаще всего денежные метрики грубые (слабо реагируют на изменения либо, надо очень много времени, чтобы их измерить)
- Из-за этого чистым денежным метрикам предпочтительнее промежуточные метрики

Роли метрик

- **Ключевые метрики** – отражают ключевые цели сервиса, должны обладать безусловной направленностью (если метрика растёт, сервису всегда становится лучше/хуже)
- **Основные критерии оценки изменений** – должны быть согласованы с ключевыми метриками, должны быть чувствительными
- **Ограничительные метрики** – должны быть чувствительны и согласованы с тем, что нельзя ломать
- **Целевые метрики эксперимента** – должны выбираться в зависимости от смысла эксперимента

Желательные свойства метрик

- **Согласованность** – метрика должна быть согласована с целями сервиса и его ключевыми метриками
- **Направленность** – если значение метрики изменилась, должна быть чёткая интерпретация этого изменения (хорошо это или плохо)
- **Чувствительность (sensitivity)** – способность метрики отражать статистически значимую разницу между контрольной и тестовой группами, когда она есть. Чем выше чувствительность, тем меньше данных нужно, чтобы обнаружить статистически-значимые изменения.
  Пример: метрики, основанные на деньгах слабо реагируют на изменения
- **Стабильность** – метрика должна быть чувствительной и согласованной с тем, что нельзя ломать. Если у метрики высокая дисперсия, то для того, чтобы уловить значимый эффект, надо собирать много данных
  Пример: розничный торговый оборот магазина может колебаться в очень широких диапазонах. Чтобы уменьшить его дисперсию, обычно смотрят торговый оборот отдельных отделов.
-

Примеры:
Лояльность пользователя:

- Число пользовательских сессий
- Время, которое юзер проводит в сервисе

Имеют чёткую направленность
Хорошие предикторы для долгосрочного успеха продукта
Обладают слабой чувствительностью
Активность пользователя:

- Число кликов за сессию
- Длина пользовательской сессии

Обладают сильной чувствительностью
Обладают неоднозначной направленностью
Пример: клики пользователей в рекомендательной системе отражают как позитивные, так и негативные сигналы

- С одной стороны, они говорят, что пользователю нравится пользоваться продуктом
- С другой, они говорят, что у нас много кликбейтного контента
- Метрики с чёткой интерпретацией часто обладают низкой чувствительностью