KARPOV.COURSES >>>
КОНСПЕКТ

> Конспект > 4 урок > Проверка гипотез, параметрические статистические критерии

> Оглавление

>Проверка гипотез

Давайте для начала разберемся в идее, которая лежит за проверкой гипотез.

Основная проблема в том, что выборки между собой различаются. Но возьмем за идею, что если мы встретили малые изменения в наших метриках, то это нормально, и мы можем считать, что выборки одинаковые.

Постановка

Представим, что мы обучили модель классификации текста на спам и нормальные сообщения. И мы хотим убедиться, что новая модель лучше старой, но разница в качестве не очень большая, поэтому есть вероятность, что это шум.

Построим выборки: массивы из ошибок моделей по каждому объекту. И будем проверять, что средние значения выборок совпадают.

Сформулируем нулевую гипотезу: средние значения выборок совпадают

Альтернативная гипотеза: средние значения выборок не совпадают

После того, как выбрали гипотезу, выбираем статистику (функция от выборки). Статистику будет выбирать такую, распределение которой мы знаем, при условии, что нулевая гипотеза верна. Например, разность выборочных средних, нормированных на среднюю дисперсию выборок.

Затем мы вычисляем значение статистики на выборках. Нулевое распределение показывает, какое должно быть распределение статистики при верности нулевой гипотезы. А значит, мы можем проверить, насколько экстремальное значение статистики мы получили - посчитаем вероятность встретить такие и более экстремальные значения. Если вероятность ниже уровня значимости \alpha, считаем, что такие отклонения маловероятны и нулевую гипотезу можно отвергнуть.

Формально

Выборка $X_N, X_N \sim P$

Нулевая гипотеза — H_0

Альтернатива — H_1

Статистика $T(X_N)$ — распределена конкретным образом для H_0, каким-то другим образом для H_1

Мы заранее определили уровень α, с которым будем потом сравнивать нашу вероятность.

Значение статистики $t = T(X_N)$

Достигаемый уровень значимости $p(X_N)$— вероятность при H_0 получить $T(X_N) = t$ или более экстремальное значение $p(X_N) = P(T \geq t | H_0)$

H_0 отвергается при $p(X_N) \leq \ \alpha$

>Параметрические тесты

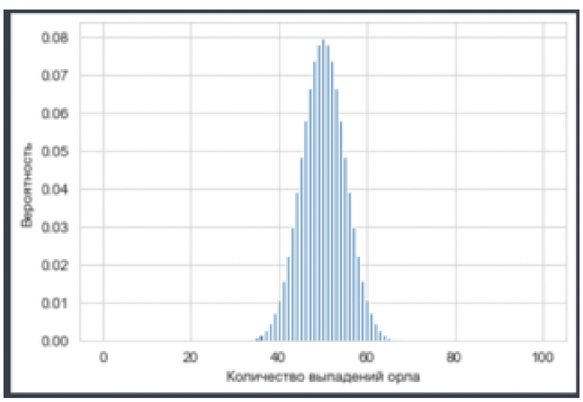
Предполагаем конкретный вид распределения, а зная распределение, можно выбрать соответствующую статистику и оценить ее распределение при условии нулевой гипотезы.

Сравнение долей

Например, мы хотим проверить, что монетка настоящая: одинаково часто падает орлом и решкой. Из 100 бросков получили 45 "решка"

Все ли хорошо с монетой?

Монетка имеет распределение Бернулли. Когда мы подкидываем монетку несколько раз, то получаем биномиальное распределение. Нулевая гипотеза - монетка нормальная (p = 1/2). Альтернативная гипотеза - монетка не нормальная. Критический уровень - 0.05. Можем построить распределение "Сколько раз нормальная монетка падает орлом, если бросать 100 раз"



Select an Image

Хотим посчитать вероятность встретить такое и более экстремальное значение, то есть:

$$p(X_N) = P(T \geq t|H_0) = P(T \leq 45|H_0) + P(T \geq 55|H_0) = 2P(T \leq 45|H_0) = 2F_{Bin(1/2)}(45)$$

Сравнение долей: z-критерий

Как мы уже говорили: биномальное распределение стремится к нормальному распределению по центральной предельной теореме.

Соответственно, для выборки, распределенной по Бернулии, можем посчитать некоторую статистику, которая будет распределена нормально. Такая статистика называется z-критерий.

Выборка: $X_1, \ldots, X_N \sim Ber(p)$

Нулевая гипотеза: $p = p_0$

Альтернативная гипотеза: $p \neq p_0$

Статистика Z_N

$$Z_N = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{N}}}$$

Нулевое распределение: $Z_N \sim N(0, 1)$

>Сравнение двух выборок: z-критерий

В предыдущих примерах мы сравнивали монетку с некоторым идеалом, для которого знали вероятность. Такие критерии называют одновыборочные.

Двухвыборочные критерии - сравнение двух выборок между собой.

Расммотрим пример:

Мы придумали новый алгоритм рекоммендаций товаров и хотим проверить, лучше ли новый алгоритм. Для этого мы для 50% будем использовать одну версию, а для остальных - вторую.

Здесь нам все так же поможет z-критерий

Выборки: $X_{11}, \ldots, X_{1N1} \sim Ber(p_1), X_{21}, \ldots, X_{2N2} \sim Ber(p_2)$ — независимые

Нулевая гипотеза: $p1 = p2$

Альтернативная гипотеза: $p1 \neq p2$

Статистика Z_N

$$Z_N = \frac{\overline{p_1} - \overline{p_2}}{\sqrt{P(1-P)(\frac{1}{n_1} + \frac{1}{n_2})}} \qquad P = \frac{\overline{p_1}n_1 + \overline{p_2}n_2}{n_1 + n_2}$$

Нулевое распределение: $ZN \sim N(0, 1)$

>Сравнение средних

Мы разобрались с тем, как сравнивать между собой выборки, представленные в виде бинарных величин, но что, если эти величины вещественные и нормально распределены?

Оказывается. что z-критерий все так же применим, но нам необходимо знать дисперсию.

Выборка: $X_1, \ldots, X - N \sim N(\mu, \sigma), \sigma$—известна

Нулевая гипотеза: $\mu = \mu_0$

Альтернативная гипотеза: $\mu \neq \mu_0$

Статистика Z_N

$$Z_N = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

Нулевое распределение: $Z_N \sim N(0, 1)$

Если дисперсия вам не известна, то среднее уже распределено не нормально, а по Стьюденту.

В таком случае возникает Т-критерий

Выборка: $X_1, \ldots, X_N \sim N(\mu, \sigma), \sigma$—неизвестна

Нулевая гипотеза: $\mu = \mu_0$

Альтернативная гипотеза: $\mu \neq \mu_0$

Статистика T_N

$$T_N = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{N}}}$$

Нулевое распределение: $T_N \sim St(N-1)$

Для сравнения двух выборок между собой также применим Т-критерий:

Выборки: $X_{11}, \ldots, X_{1N1} \sim N(\mu_1, \sigma_1), X_{21}, \ldots, X_{2N2} \sim N(\mu_2, \sigma_2), \sigma_1, \sigma_2$— не известны

Нулевая гипотеза: $\mu_1 = \mu_2$

Альтернативная гипотеза: $\mu_1 \neq \mu_2$

Статистика T_N

$$T_N = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

Нулевое распределение: $T_N \sim St(\nu), v =$ ...

>Проверка нормальности

Критерий согласия Пирсона:

Нулевая гипотеза - выборка из нормального распределения

Идея:

- Разбиваем выборку на K бинов $[a_i, a_{i+1}]$ и считаем n_i количество попадающих объектов выборки
- Сравниваем эти числа с теоретическими npi

$$p_i = F(a_{i+1}) - F(a_i)$$

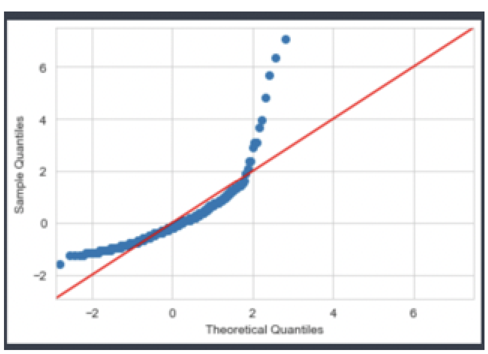
Статистика$\chi^2, K-1$ степень свободы при известных параметрах, K-3 при оценённых по выборке

Недостатки:

- Число бинов можно выбирать неоднозначно
- Только для больших выборок

QQ-Plot

На данном графике визуализируют реальные квантили и теоретические квантили, как если бы распределение было нормальным.



Критерий Шапиро-Уилка

Нулевая гипотеза та же - выборка из нормального распределения

Аналогично с QQ-Plot - сравниваем вариационный ряд с теоретическим.