Статистическое моделирование и прогнозирование
Лекция 4. Проверка статистических гипотез

Семёнов Михаил Евгеньевич
к. ф.-м. н., доцент ОЭФ ИЯТШ

Томский политехнический университет
16 марта 2020 г.

Содержание

Статистическая гипотеза

Статистической гипотезой принято считать любое предположение о законе распределения случайной величины генеральной совокупности или о значениях параметров закона распределения.

Высказанное предположение, которое подлежит проверке, обозначается H_0 и называется *основной* или *нулевой* гипотезой. Наряду с основной гипотезой в рассмотрение вводится и противоречащая ей гипотеза H_1, которая называется *конкурирующей* или *альтернативной*.

Цель проверки статистической гипотезы заключается в том, чтобы установить, не противоречит ли высказанная гипотеза H_0 имеющимся выборочным данным $\{X_1, X_2, \ldots, X_n\}$.

Проверка нулевой гипотезы

Для проверки нулевой гипотезы формируется *статистический критерий* — специальная статистика $K(X_1, X_2, \ldots, X_n)$ определение которой в условиях нулевой гипотезы H_0 звестно. По известному распределению статистического критерия определяется множество значений, которые величина K принимает с вероятностью γ, близкой к единице, то есть практически достоверно. Это множество называется *областью принятия нулевой гипотезы* H_0. Дополнение этого множества образует критическую область (или область отвержения гипотезы H_0).

Рисунок 1 – Области принятия и отклонения нулевой гипотезы

Метод проверки нулевой гипотезы

По выборочным данным вычисляется наблюдаемое значение критерия $K_n = K(X_1, X_2, \ldots, X_n)$. Если значение K_n принадлежит критической области, то проверяемая гипотеза H_0 отвергается, как противоречащая выборочным данным, и принимается альтернативная гипотеза H_1. Если же K_n принадлежит области принятия нулевой гипотезы, то она принимается, как согласующаяся с выборочными данными. В этом случае говорят, что нулевая гипотеза принимается на уровне значимости $\alpha = 1 - \gamma$.

Замечание. Статистическими методами можно лишь опровергнуть выдвинутую гипотезу H_0, но нельзя её доказать.

Ошибки 1-го и 2-го рода. Мощность критерия

Уровень значимости гипотезы α характеризует вероятность совершить *ошибку первого рода*, заключающуюся в напрасном отвержении верной нулевой гипотезы: $P\{H_1|H_0\} = \alpha$ помимо этого, существует вероятность совершить *ошибку второго рода*, состоящую в напрасном принятии неверной нулевой гипотезы $P\{H_0|H_1\} = \beta$.
Дополнительную к β величину, соответствующую вероятности недопущения ошибки второго рода $P\{H_1|H_1\} = 1 - \beta$ называют *мощностью критерия.*
Заметим, что одновременное уменьшение вероятностей ошибок первого и второго рода возможно только при увеличении объёма выборки n.

Пирсона χ^2-критерий согласия

Критерием согласия называется статистический критерий проверки гипотезы о соответствии эмпирического распределения вероятностей — теоретическому.
Гипотезы: Проверяется нулевая гипотеза $H_0 : F(x) = F_0(x, \theta)$ против альтернативной $H_1 : F(x) \neq F_0(x, \theta)$, где $F_0(x, \theta)$ — теоретическая функция распределения случайной величины x; $\theta \in \mathbb{R}^m$ — m-мерный вектор в общем случае неизвестных параметров распределения θ.

Пирсона χ^2-критерий согласия

Статистика: Критерий согласия χ^2, предложенный К. Пирсоном в 1900 году, основывается на анализе группированных данных.
При этом область возможных значений реализации выборки $\{x_1, x_2, \ldots, x_n\}$ разбивают на k непересекающихся интервалов:
$x_j \in (a_0, a_k] = (a_0, a_1] \cup (a_1, a_2] \cup \ldots \cup (a_{k-1}, a_k]$ и вычисляют статистику, имеющую распределение χ^2 с числом степеней свободы $k - m - 1$:

$$X_d^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2_{k-m-1}, \quad n_i = \sum_{i=j}^{n} \mathbb{I}_{(a_{i-1}, a_i]} x_j, \tag{1}$$

где n_i — эмпирическая частота попаданий выборочных значений x_j интервал $(a_{i-1}, a_i]$; p_i — теоретическая вероятность попадания значений случайной величины X в интервал $(a_{i-1}, a_i]$: $p_i = F_0(a_i, \theta_n) - F_0(a_{i-1}, \theta_n)$, где $\theta_n \in \mathbb{R}^m$ — выборочная оценка m-мерного вектора неизвестных параметров распределения θ.

Пирсона χ^2-критерий согласия

Критерий: Если наблюдаемое значение статистики превосходит на заданном уровне значимости α квантиль распределения χ^2 с тем же числом степеней свободы:

$$X_d^2 > \chi^2_{\alpha,k-m-1},$$

то нулевая гипотеза на уровне значимости α отвергается в пользу альтернативной $H_1 : F(x) \neq F_0(x, \theta)$.
В противном случае при

$$X_d^2 \leq \chi^2_{\alpha,k-m-1}$$

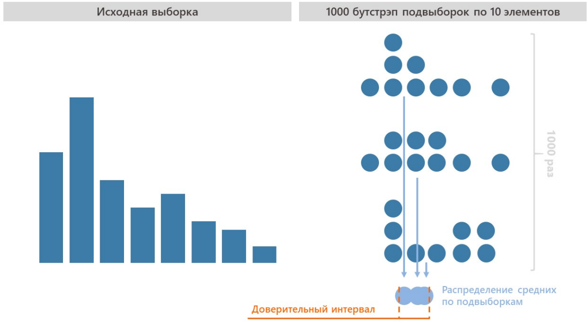
говорят, что нулевая гипотеза $H_0 : F(x) = F_0(x, \theta)$ на уровне значимости α согласуется с выборочными данными.

Бутстрэп анализ

Статистические методы проверки гипотез базируются на теоретических распределениях и работают только при соблюдении ограничений и допущений. Существует альтернативный подход, который вместо теории задействует грубую компьютерную силу — *ресемплинг*. Этот подход объединяет целый класс методов генерации дополнительных выборок из уже имеющихся. При ресемплинге расчет параметров проводится на фактических данных, а теоретическое распределения не используется.
К ресемплингу относится:

- бутстрэп анализ,
- перестановочные тесты и
- «складной нож».

Бутстрэп-процедура состоит в многократном извлечении подвыборок из эмпирического распределения. Для оценки любых параметров можно сформировать тысячи повторных бутстрэп-выборок (обычно 500-10000), каждая из которых содержит 2/3 значений исходной выборки.



Алгоритм расчета доверительного интервала при помощи бутстрэп

- Выберите 10 наблюдений из выборки. Отбор осуществляется с возвратом, то есть некоторые наблюдения могут быть выбраны несколько раз, а некоторые могут остаться невыбранными.
- Вычислите среднее для полученного набора из 10 значений.
- Повторите шаги 500 – 10000 раз.
- Отсортируйте тысячу выборочных средних по возрастанию.
- Найдите средние, которые представляют собой 2,5 и 97,5 процентили.

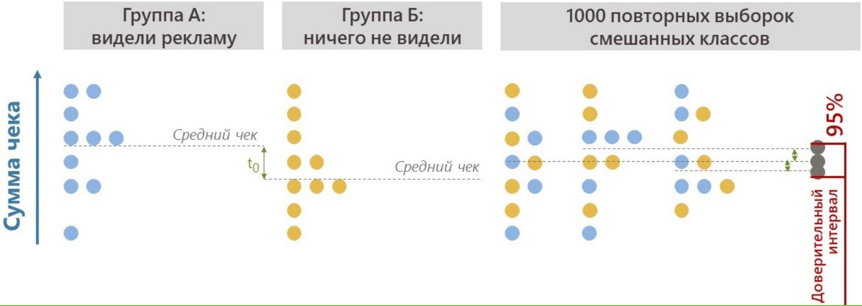
Перестановочные тесты

Перестановочные тесты – позволяют сравнивать несколько выборок между собой при помощи любых критериев.

- У нас в распоряжении 20 человек: 10 человек, которые видели рекламу (группа A), не видели рекламу (группа B). Обе группы идут в магазин.
- Измеряем сумму чека каждого участника и рассчитываем средний чек по группе A и B.
- Рассчитываем t_0 – различия между средними по группам. Наша задача определить, повлиял ли рекламный ролик на средний чек.
- Если реклама не связана с средним чеком, то неважно к какой группе относится каждый участник. Случайно перемешиваем участников и рассчитываем разницу между средними (t_n) для 1000 смешанных выборок из 10 человек.

Перестановочные тесты

- Полученное распределение и будет эмпирическим для данного эксперимента. Отсортируем 1000 t по возрастанию.
- Если t_0 не входит в центральные 95% значений эмпирического распределения, то у нас есть все основания отвергнуть нулевую гипотезу о равенстве средних значений в двух группах.



Постановка заданий

- Для реализации выборки (см. ЛР2) выполните по критерию согласия Пирсона проверку нулевой гипотезы: $H_0 : F(x) = F_0(x, (\bar{x}, s_x))$, где $F_0(x, (\bar{x}, s_x)) = \Phi(\frac{x-\bar{x}}{s_x}) + \frac{1}{2}$ при альтернативной $H_1 = \bar{H}_0$.
- Для двух частей реализации выборки: $\{x_i\}_n = \{u_j\}_m \cup \{v_k\}_l$, где $n = m + l$ – объемы полной и частичных выборок, с помощью критерия а) однородности Смирнова проверить нулевую гипотезу $H_0 : F_1(u) = F_2(v)$ при альтернативной $H_1 : F_1(u) \neq F_2(v)$, б) Фишера проверить нулевую гипотезу $H_0 : \sigma_u^2 = \sigma_v^2$ при альтернативных H_1 : а) $\sigma_u^2 \neq \sigma_v^2$, б) $\sigma_u^2 < \sigma_v^2$, в) $\sigma_u^2 > \sigma_v^2$. Уровень значимости $\alpha = 0.05$.
- Разбить выборку на две части. В первую часть включить элементы исходной выборки с четными номерами, во вторую часть – оставшиеся элементы. Для этих частей реализации выборки повторить предыдущее задание.

Постановка заданий (продолжение)

- Сгенерировать случайную величину и найти доверительный интервал для среднеарифметического с помощью бутстреп анализа. Построить график эмпирического распределения.
- Получить p-level для перестановочного теста при помощи эмпирического распределения.

Скрипт на R [Буховец et al., 2010]

Алгоритм 1 Проверка гипотезы о нормальности реализации выборочной совокупности

Вход: Выборка с.в..

```
source("D:/samples.r")
par(mfrow=c(1,2))
x <- samples(100); a <- mean(x); s <- sd(x)
b1 <- c(4:15); m1 <- table(cut(x, breaks=b1))
b2 <- c(4, 6:11, 15); m2 <- table(cut(x, breaks=b2))
b3 <- c(-Inf, 6:11, Inf); p3 <- diff(pnorm(b3, a, s))
print(m2); print(chisq.test(x=m2, p=p3))
x1 <- seq(4,15,length=300); f1 <- dnorm(x1, a, s);
hist(x, breaks=b2); rug(x); lines(x1, f1);
qqnorm(x, pch=3); qqline(x, lty=2)
```

Выход: Статистика, p-value критерия согласия Пирсона хи-квадрат. Графики

Скрипт на R [Буховец et al., 2010]

Алгоритм 2 Проверка гипотезы об однородности распределений для двух частей реализации случайной выборки по критерию Смирнова

Вход: x, u, v – исходная реализация случайной выборки и две её части

```
1: source("D:/samples.r")
2: x <- samples(); u <- x[1:49]; v <- x[50:100]; print(ks.test(u,v))
3: plot(ecdf(u), pch=25, cex=0.5, xlim=range(x)); rug(u, side=1)
4: plot(ecdf(v), pch=24, cex=0.5, add=TRUE); rug(v, side=3)
```

Выход: Статистика, p-value критерия Смирнова. Графики

Скрипт на R [Буховец et al., 2010]

Алгоритм 3 Проверка гипотезы о равенстве дисперсий двух частей реализации случайной выборки по двухвыборочному F-критерию

Вход: x, u, v – исходная реализация случайной выборки и две её части

```
source("D:/samples.r")
x <- samples(); u <- x[1:49]; v <- x[50:100]
print(var.test(u,v, alter="two"))
print(var.test(u,v, alter="le"))
print(var.test(u,v, alter="gr"))
```

Выход: Статистика, p-value критерия Фишера

Скрипт на R [Блог про статистику, 2020]

Алгоритм 4 Доверительный интервал для среднеарифметического

Вход: *data* – таблица с исходными данными

```
library("boot") # подключение библиотеки
myfn <- function(x) return(mean(x)) # функция для расчета статистики
result <- boot(data$varname, myfn, R = 999) # бутстрэп-анализ
result # текстовый вывод результатов
plot(result) # визуализация эмпирического распределения
quantile(result$t, c(0.025, 0.975))
```

Выход: Доверительный интервал

Скрипт на R [Блог про статистику, 2020]

Алгоритм 5 Вычисление p-level для перестановочного теста

Вход: *data* – таблица с наблюдениями x и y, где y – исследуемая переменная (сумма чека), x – метка группы (видели / не видели рекламу)

```
library("coin") # библиотека для перестановочных тестов
oneway_test(y ~ x, data)
```

Выход: p_level для перестановочного теста

ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Блог про статистику, (2020).
Бутстрэп анализ.
https://www.tidydata.ru/hypothesis.
[Онлайн; дата доступа 16 марта 2020].

Буховец, , Москалев, , Богатова, , and Бирючинская, (2010).
Статистический анализ данных в системе R. Учебное пособие.
ВГАУ, Воронеж.