

Тема 10. Перестановочные тесты. Бутстреп

Большинство из рассмотренных на предыдущих занятиях статистических тестов можно использовать, только если исследуемая величина удовлетворяет определенным условиям (например, нормальное распределение или равенство дисперсий между группами). Для многих параметрических тестов существуют непараметрические аналоги, которые можно использовать, если эти условия не выполняются (например, можно использовать критерий Краскела-Уолиса вместо однофакторного дисперсионного анализа, если зависимая величина имеет распределение отличное от нормального или её дисперсия не равна во всех группах). Однако не для всех статистических тестов существуют непараметрические аналоги, например не существует соответствующего аналога для двухфакторного дисперсионного анализа. Кроме того, и параметрические, и непараметрические тесты плохо работают на маленьких выборках или выборках, содержащих выбросы. Во всех этих случаях можно использовать подходы, основанные на генерации рандомизации (перестановочные тесты) и повторных выборках (бутстреп).

Перестановочные тесты

Перестановочные тесты позволяют создавать случайные выборки из реальных данных путем их «перемешивания». Допустим, вы хотите посчитать тест Стьюдента для данных, имеющих нормальное распределение. Вы рассчитали значение t -статистики. Сравнивая это значение с теоретическим распределением, имеющим при нулевой гипотезе и достаточно большом объеме выборки нормальный характер, можно рассчитать значение p , интерпретируемое как вероятность получить значение большее, чем рассчитанное. Чем меньше значение p , тем более вероятна альтернативная гипотеза о неравенстве средних значений. Однако если исходные данные имеют распределение, не являющееся нормальным, или если объем выборки мал, то t -статистика может иметь распределение, отличное от нормального, и значение p может быть рассчитано некорректно. В таком случае рассчитанное значение t -статистики можно сравнить с эмпирическим распределением, полученным на случайных выборках. Для того чтобы получить эмпирическое распределение, исходные данные «перемешивают» между двумя выборками и рассчитывают значение критерия Стьюдента t_1 , потом опять перемешивают, рассчитывают значение t_2 , и т.д. На основе этих значений строят эмпирическое распределение t и сравнивают с ним значение t , рассчитанное на реальных данных.

Выполнить перестановочные тесты можно при помощи функций, реализованных в пакете **coin**. Функции для наиболее популярных статистических тестов представлены в таблице 6:

Таблица 6. Функции из пакета **coin**, позволяющие выполнять перестановочные тесты.

Функция	Перестановочный тест
<code>oneway_test(y~A)</code>	Тест для двух и более выборок (аналог теста Стьюдента и однофакторного дисперсионного анализа)
<code>wilcox_test(y~A)</code>	Тест Уилкоксона для независимых выборок (тест Манна-Уитни)
<code>wilcoxsign_test(y1~y2)</code>	Тест Уилкоксона для зависимых выборок
<code>chisq_test(A~B)</code>	Хи-квадрат тест (критерий согласия Пирсона)
<code>cmh_test(A~B C)</code>	Тест Кокрана-Мантеля-Хензеля
<code>spearman_test(y~x)</code>	Коэффициент корреляции Спирмена

Строчными буквами в формулах представлены числовые переменные, а заглавными – категориальные.

В пакете **ImPerm** реализованы перестановочные подходы к дисперсионному и регрессионному анализу (регрессионный анализ будет разобран позднее). Выполнить соответствующие виды анализа можно при помощи функций **aovp()** и **lmp()**. Параметры функций такие же, как и для функций **aov()** и **lm()**.

Бутстреп

Бутстреп-анализ создает эмпирическое распределение статистики путем генерации многих случайных выборок с возвратом из исходной выборки. Бутстреп можно использовать для вычисления доверительного интервала измеренной величины. Допустим, у нас есть выборка из 10 измерений, и мы хотим рассчитать доверительный интервал для среднего значения. Выполним следующие вычисления:

- 1) Случайно выберем 10 наблюдений из выборки с возвратом значений после каждого выбора. Часть наблюдений исходной выборки могут войти в новую выборку несколько раз, а другая часть наблюдений – не войти совсем;
- 2) Вычислим среднее значение для полученной выборки;
- 3) Повторим шаги 1-2 (тысячу) раз;
- 4) Отсортируем тысячу выборочных средних по возрастанию;

5) Найдем выборочные средние, которые представляют собой 2,5 и 97,5 процентиля. Это и будут границы 95%-го доверительного интервала.

Выполнить бутстреп-анализ можно при помощи пакета **boot**. Для этого нужно, во-первых, написать функцию, которая вычисляет нужную статистику, во-вторых, применить функцию **boot()** к этой функции, чтобы создать бутстреп-повторности данной статистики, в-третьих, использовать функцию **boot.ci()**, чтобы вычислить доверительный интервал для искомой статистики. Рассмотрим применение функций **boot()** и **boot.ci()** на примере расчета доверительного интервала для среднего роста женщин. Используем для этого выборку роста 15 женщин (в дюймах):

```
> data(women)
> head(women)
  height weight
1     58    115
2     59    117
3     60    120
4     61    123
5     62    126
6     63    129
```

Напишем сначала функцию для расчета среднего роста на *i*-й сгенерированной выборке:

```
samplemean<-function(d, i) {mean(d[i])}
```

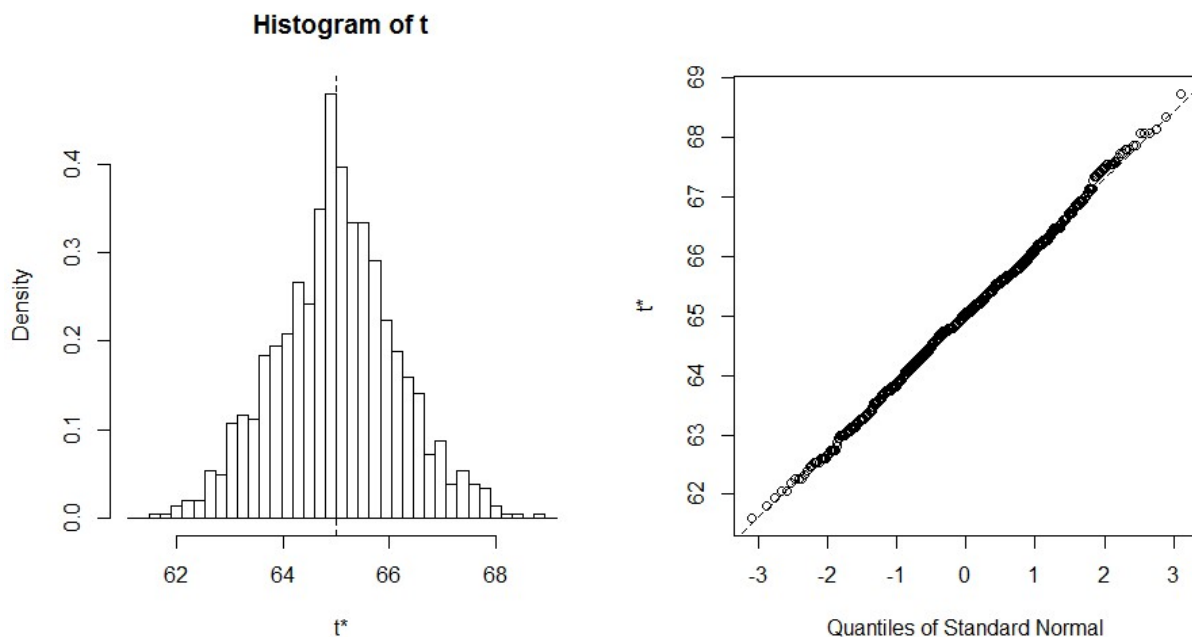
где *d* – это набор созданных при помощи бутстрепа выборок, а *i* – индекс конкретной выборки.

Далее сгенерируем тысячу выборок и рассчитаем для них средние значения:

```
mn<-boot(women$height, samplemean, R=1000)
```

Применение функции **plot()** к результатам функции **boot()** позволяет построить гистограмму из полученных средних значений и диаграмму квантилей, сравнивающую полученное эмпирическое распределение со стандартным нормальным распределением:

```
> plot(mn)
```



Как видно из диаграмм, среднее значение роста имеет нормальное распределение.

Теперь, используя функцию **boot.ci()**, вычислим 95%-й доверительный интервал для роста методом процентилей (параметр `type="perc"`):

```
> boot.ci(mn,type="perc")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = mn, type = "perc")

Intervals :
Level      Percentile
95%      (62.73, 67.46 )
Calculations and Intervals on Original scale
```

Само среднее значение можно получить следующим образом:

```
> mn$t0
[1] 65
```

Таким образом, среднее значение роста – 65 дюймов, а верхняя и нижняя граница 95%-ного доверительного интервала 62.7 и 67.5 соответственно. Это означает, что истинное значение среднего роста попадает в эти границы с 95%-й вероятностью.

Практическое задание

1) Сохраните в переменную таблицу `Arthritis.txt`. Выполните тест Кохрана-Мантеля-Хензеля для оценки наличия статистически значимой взаимосвязи между приемом

лекарства и эффектом с учетом пола пациентов. Выполните аналогичный перестановочный тест. Являются ли выявленные связи статистически значимыми?

2) Импортируйте таблицу из файла “Anorexia.txt”. В таблице представлен вес пациентов, страдающих анорексией, до и после лечения (семейная терапия). Каждая строка – это пациент. Prewt – вес до лечения. Postwt – вес после лечения. Какой статистический тест нужно использовать для оценки эффективности лечения? Выполните аналогичный перестановочный тест. Привело ли лечение к улучшению состояния пациентов?

3) Импортируйте данные из файла thalidomide.txt. В таблице представлена информация об увеличении веса (wt) пациентов с ВИЧ инфекцией при терапии талидомидом (trt) или плацебо. У части пациентов также имеется туберкулез (tb).

Выполните двухфакторный дисперсионный анализ с помощью перестановочного теста. Эффективно ли лекарство? Влияет ли наличие туберкулеза на эффективность лекарства? Если влияет, то как?

4) Импортируйте данные из файла birthwt.txt. В таблице представлена информация о влиянии двух факторов на низкий вес новорожденных: smoke – курение табака, ht – наличие артериальной гипертензии. С помощью бутстреп-анализа вычислите средний вес новорожденных и 95% доверительный интервал для курящих/некурящих и страдающих/не страдающих артериальной гипертензией матерей. Перекрываются ли доверительные интервалы, рассчитанные для курящих/некурящих матерей? Страдающих/не страдающих артериальной гипертензией? Выполните дисперсионный анализ с помощью перестановочного теста. Исходя из всех полученных данных, можно ли утверждать, что курение и наличие артериальной гипертензии влияют на вес новорожденных?

Вопросы

1. Что такое перестановочные тесты и для чего они используются?
2. Что такое бутстреп и для чего он используется?