Липецкий государственный технический университет

Кафедра прикладной математики

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ МАТЕМАТИЧЕСКИХ ИССЛЕДОВАНИЙ

Лекция 8

8. Анализ мощности

Составитель - Сысоев А.С., к.т.н., доц.

Липецк – 2021

Outline

8.1. Проверка статистических гипотез

Анализ мощности позволяет определить *размер выборки*, необходимый для выявления *эффекта заданной величины* с заданной *долей уверенности*.

Гипотезы: H_0 - нулевая, H_1 - альтернативная.

Допуская, что нулевая гипотеза справедлива, вычисляют вероятность получения для *генеральной совокупности* наблюдаемого в выборке или большего значения статистики.

Пример: влияние разговоров по мобильному телефону на время реакции водителя.

H_0 : $\mu_1 - \mu_2 = 0$ при H_1: $\mu 1 - \mu 2 \neq 0$.

Статистика

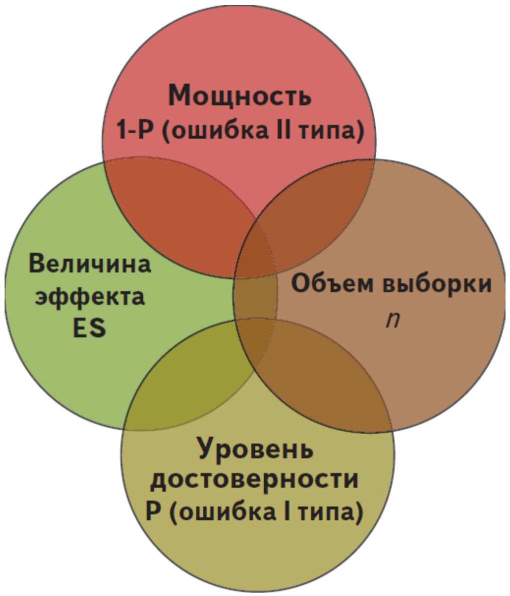
$$\left(\overline{X}_1 - \overline{X}_2\right) / \left(\frac{S}{\sqrt{n}}\right)$$

где $\overline{X}_1$ и $\overline{X}_2$ – это средние для каждой выборки значения времени реакции, **S** – это стандартное отклонение для объединенной выбор ки, **n** – число участников в каждой группе.

8.1. Проверка статистических гипотез

Если вероятность не достигает некоторого принятого критического значения (например, $p < 0.05$), то отвергают нулевую гипотезу в пользу альтернативной. Заранее определенное критическое значение (0.05) называется ***уровнем значимости теста***.

	Отвергаем H_0	**Не смогли отвергнуть H_0**
H_0 справедлива	Ошибка I типа	Верно
H_0 ложна	Верно	Ошибка II типа



размер выборки – число наблюдений при каждом типе воздействия или в каждой группе

уровень значимости (альфа) – вероятность совершения статистической ошибки первого типа. Уровень значимости можно также трактовать как вероятность нахождения несуществующей закономерности

мощность теста (вероятность совершения статистической ошибки второго типа). Мощность можно также понимать как вероятность обнаружения существующей закономерности

размер эффекта – величина эффекта, который является предметом альтернативной гипотезы

8.1. Проверка статистических гипотез

Задача при исследовании – повысить мощность статистических тестов, сохранив приемлемый уровень значимости и работая с минимально возможной выборкой.

Четыре величины (размер выборки, уровень значимости, мощность и величина эффекта) тесно взаимосвязаны. *Зная значения любых трех, можно определить четвертую*.

Функция	Вычисляет мощность для
`pwr.2p.test()`	Двух пропорций (равные объемы выборок)
`pwr.2p2n.test()`	Двух пропорций (неодинаковые объемы выборок)
`pwr.anova.test()`	Сбалансированного однофакторного дисперсионного анализа
`pwr.chisq.test()`	Теста хи-квадрат
`pwr.f2.test()`	Общей линейной модели
`pwr.p.test()`	Пропорции (одна выборка)
`pwr.r.test()`	Корреляции
`pwr.t.test()`	Тестов Стьюдента (одна выборка, две выборки, для зависимых переменных)
`pwr.t2n.test()`	Тестов Стьюдента (две выборки разного объема)

8.2. Процедура анализа мощности

8.2.1. Тесты Стьюдента

Пример сравнить среднее время реакции для водителей с мобильным телефоном и для водителей без отвлекающих факторов.

```
pwr.t.test(n=, d=, sig.level=, power=, type=, alternative=),
```

- `n` – размер выборки;
- `d` – величина эффекта, выраженная в стандартизованной разнице средних значений, которая рассчитывается по формуле

$$d = \frac{(\mu_1 - \mu_2)}{\sigma},$$

где μ_1 – среднее значение для первой группы, μ_2 – среднее значение для второй группы, σ – дисперсия общей ошибки;
- `sig.level` – уровень значимости (по умолчанию 0.05);
- `power` – уровень мощности;
- `type` – тип теста: для двух выборок ("`two.sample`", по умолчанию), для одной выборки ("`one.sample`"), для зависимых выборок ("`paired`");
- `alternative` – двухсторонний ("`two.sided`", по умолчанию) или односторонний ("`less`" или "`greater`").

8.2. Процедура анализа мощности

8.2.1. Тесты Стьюдента

```
> library(pwr)
> pwr.t.test(d=.8, sig.level=.05, power=.9, type="two.sample",
alternative="two.sided")
     Two-sample t test power calculation
              n = 34
              d = 0.8
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
 NOTE: n is number in *each* group
```

Замечание: сравнивая две группы, установим обнаружение различие между выборочными средними в 0.5 стандартного отклонения. Можем исследовать только по 20 человек.

```
> pwr.t.test(n=20, d=.5, sig.level=.01, type="two.sample",
  alternative="two.sided")
     Two-sample t test power calculation
              n = 20
              d = 0.5
      sig.level = 0.01
          power = 0.14
    alternative = two.sided
 NOTE: n is number in *each* group
```

```
pwr.t2n.test(n1=, n2=, d=, sig.level=, power=, alternative=)
```

8.2. Процедура анализа мощности

8.2.2. Дисперсионный анализ

Пример: для однофакторного дисперсионного анализа, включающего пять групп, нужно найти размер групп, при котором мощность составляет 0.8, размер эффекта – 0.25 и уровень значимости – 0.05.

```
pwr.anova.test(k=, n=, f=, sig.level=, power=)
```

где **k** – число групп, а **n** – размер выборки в каждой группе.

Для однофакторного дисперсионного анализа величина эффекта отражена в параметре `f`, который рассчитывается по формуле

$$f = \sqrt{\frac{\sum_{i-1}^{k} \boldsymbol{p}_i \times (\mu_i - \mu)^2}{\sigma^2}},$$

где $p_i = n_i/N_i$,
n_i – число наблюдений в группе i,
N – общее число наблюдений,
μ_i – среднее значение для группы i,
μ – общее среднее,
σ^2 – дисперсия ошибок внутри групп.

```
> pwr.anova.test(k=5, f=.25, sig.level=.05, power=.8)
     Balanced one-way analysis of variance power calculation
              k = 5
              n = 39
              f = 0.25
      sig.level = 0.05
          power = 0.8
 NOTE: n is number in each group
```

Пример: связь между депрессией и одиночеством.

$$H_0: \rho \le 0.25 \text{ и } H_1: \rho > 0.25$$

```
> pwr.r.test(r=.25, sig.level=.05, power=.90, alternative="greater")
     approximate correlation power calculation (arctangh transformation)

              n = 134
              r = 0.25
      sig.level = 0.05
          power = 0.9
    alternative = greater
```

Нужно установить степень депрессии и одиночества у 134 человек, чтобы быть на 90% уверенными в том, что будет отвергнута нулевую гипотезу, если она действительно окажется ложной.

Пример: влияет ли стиль руководства на удовлетворенность сотрудников своей работой наряду с уровнем зарплаты и чаевыми. Стиль руководства описан четырьмя переменными, а размер заплаты и чаевых – тремя. На основании имеющегося опыта можно сказать, что удовлетворение от работы примерно на 30% определяется доходом. С практической точки зрения интересно, добавит ли стиль руководства хотя бы 5% к этому значению. Сколько людей нужно опросить, чтобы выявить такой размер эффекта с 90% уверенности при равном 0.05 уровне значимости?

```
pwr.f2.test(u=, v=, f2=, sig.level=, power=)
```

где **u** и **v** – это числитель и знаменатель степеней свободы, а **f2** – размер эффекта, который вычисляется по формуле

$$f^2 = \frac{R^2}{(1-R^2)},$$

где R^2 – это квадрат множественного коэффициента корреляции для генеральной совокупности, или по формуле

$$f^2 = \frac{(R_{AB}^2 - R_A^2)}{(1 - R_{AB}^2)},$$

где R_A^2 – дисперсия в генеральной совокупности, учтенная набором переменных A,
R_{AB}^2 – дисперсия в генеральной совокупности, учтенная переменными из наборов A и B.

```
> pwr.f2.test(u=3, f2=0.0769, sig.level=0.05, power=0.90)
     Multiple regression power calculation
              u = 3
              v = 184.2426
             f2 = 0.0769
      sig.level = 0.05
          power = 0.9
```

В случае множественной регрессии знаменатель степеней свободы равен **N** – **k** – 1, где **N** – это число наблюдений, а **k** – число независимых переменных.

Пример: обычное лекарство облегчает симптомы болезни у 60% пациентов. Новое (и более дорогое) средство будет иметь успех на рынке, если оно улучшит состояние 65% больных. Сколько испытуемых нужно исследовать, чтобы можно было обнаружить указанное выше различие между лекарствами по эффективности?

```
pwr.2p.test(h=, n=, sig.level=, power=),
```

где **h** – величина эффекта, **n** – общий объем выборки для двух групп.

```
pwr.2p2n.test(h =, n1 =, n2 =, sig.level=, power=)
```

$$h = 2\arcsin\left(\sqrt{p_1}\right) - 2\arcsin\left(\sqrt{p_2}\right)$$

```
> pwr.2p.test(h=ES.h(.65, .6), sig.level=.05, power=.9,
              alternative="greater")
     Difference of proportion power calculation for binomial
     distribution (arcsine transformation)
              h = 0.1033347
              n = 1604.007
      sig.level = 0.05
          power = 0.9
    alternative = greater
 NOTE: same sample sizes
```

Тесты хи-квадрат часто используются для оценки связи между двумя категориальными переменными. *Нулевая гипотеза обычно подразумевает, что переменные независимы, а альтернативная гипотеза – что они зависимы*.

```
pwr.chisq.test(w= , N= , df= , sig.level= , power= )
```

где **w** – размер эффекта, **N** – общий объем выборки, а **df** – число степеней свободы.

$$w = \sqrt{\sum_{i-1}^{m} \frac{(p0_i - p2_i)^2}{p0_i}},$$

где $p0_i$ – вероятность в ячейке таблицы сопряженности при справедливой H_0, $p1_i$ – вероятность в ячейке таблицы сопряженности при справедливой H_1.

Пример: связь между этнической принадлежностью и продвижением по службе.

Этническая принадлежность	Продвинулись по службе	Не продвинулись
Белые	0.42	0.28
Афроамериканцы	0.03	0.07
Латиноамериканцы	0.10	0.10

```
> prob <- matrix(c(.42, .28, .03, .07, .10, .10), byrow=TRUE, nrow=3)
> ES.w2(prob)
[1] 0.1853198

> pwr.chisq.test(w=.1853, df=2, sig.level=.05, power=.9)
     Chi squared power calculation
              w = 0.1853
              N = 368.5317
             df = 2
      sig.level = 0.05
          power = 0.9
 NOTE: N is the number of observations
```

8.2. Процедура анализа мощности

8.2.7. Выбор размера эффекта

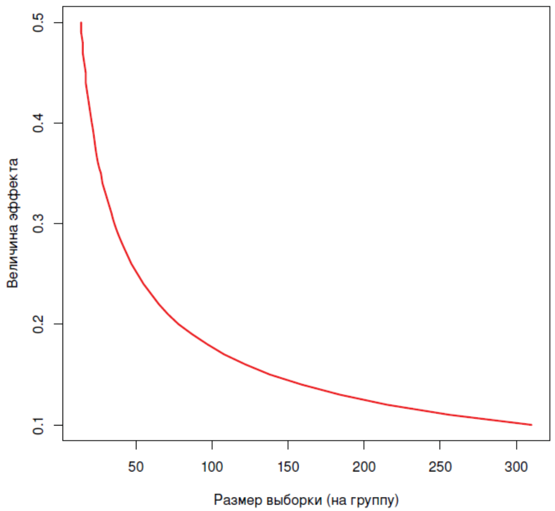
Статистический метод	Мера размера эффекта	Предлагаемые ориентировочные значения для разных размеров эффектов		
		Малый	Средний	Большой
Тест Стьюдента	d	0.20	0.50	0.80
Дисперсионный анализ	f	0.10	0.25	0.40
Линейные модели	f2	0.02	0.15	0.35
Тест пропорций	h	0.20	0.50	0.80
Тест хи-квадрат	w	0.10	0.30	0.50

8.2. Процедура анализа мощности

8.2.7. Выбор размера эффекта

Пример: сравнить пять групп при помощи однофакторного дисперсионного анализа с уровнем значимости 0.05.

```
library(pwr)
es <- seq(.1, .5, .01)
nes <- length(es)
samsize <- NULL
for (i in 1:nes){
    result <- pwr.anova.test(k=5, f=es[i], sig.level=.05, power=.9)
    samsize[i] <- ceiling(result$n)
}
```



Библиографический список

Кабаков Р. К. (2014) R в действии. Анализ и визуализация данных на языке R Издательство: ДМК Пресс, 580 с.

Шипунов А. Б., Балдин Е. М., Волкова П. А., Коробейников А. И., Назарова С. А., Петров С. В., Суфиянов В. Г. (2012) Наглядная статистика. Используем R! - М.: ДМК Пресс, 298 с.