Topic 7

Statistical power and the computation of sample size

Sergey Mastitsky ©
Klaipeda, 28-30 September 2011

7. Power and the computation of sample size

7.1. The principles of power calculations

Five statistical things every biologist should know

Ewan Birney, Head of Nucleotide Data at European Bioinformatics Institute, EMBL

From a [post]() in his blog “Five statistical things I wished I had been taught 20 years ago”

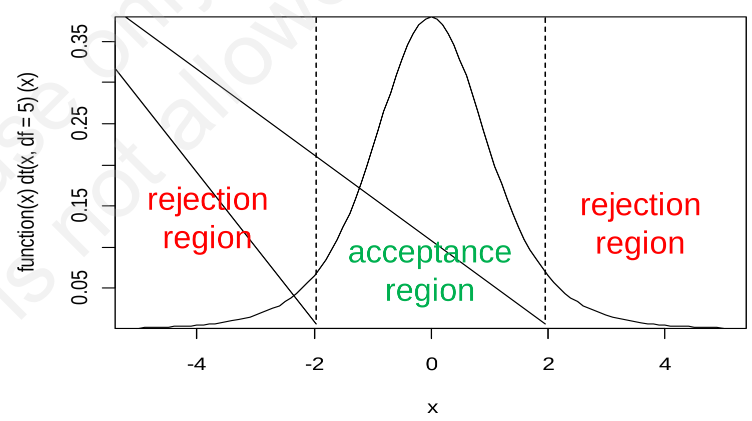
- Nonparametric statistics
- R
- The problem of multiple testing
- The relationship between P-value, effect size, and sample size
- Linear models and PCA

The problem of statistical power

- A statistical test is not able to detect a true difference if the sample size is too small as compared with the magnitude of the difference
- Thus, a sufficient amount of data should be planned and collected
- R has a number of tools for doing these kind of calculations (in particular, for one- and two-sample *t*-tests and for comparisons of proportions)

How statistical hypotheses are tested?

- Define the hypothesis to test
- Define a test statistic
- Define the acceptance and rejection regions using the selected significance level (α)
- Calculate the probability P of getting a test statistic that falls into the rejection region, and compare it against the significance level



Types of statistical errors

- As the data are sampled at random, there is always a risk of reaching a wrong conclusion:
  - Type I error: the hypothesis if correct, but the test rejects it (α-risk)
  - Type II error: the hypothesis is wrong, but the test accepts it (β-risk)

The principles of power calculations

- The risk of a type I error is the significance level α (e.g., 0.05)
- The risk of a type II error depends on the sample size and the size of effect we are trying to detect
- Power of test is the probability of rejecting a false hypothesis

A trade-off between the significance level and power of test

- In most (*but not all!*) of the practical applications a power of 80% (= β-risk of 20%) is considered acceptable
- This convention implies a one-to-four trade-off between the α-risk and β-risk, i.e.

$$\text{if } \alpha = 0.05, \text{ then}$$
$$\beta = 0.05 \times 4 = 0.20$$
$$\text{and power} = 1 - \beta = 0.80$$

7. Power and the computation of sample size

7.2. Computation of the sample size and power of one- and two-sample *t*-tests

The principles of power and sample size calculations for one- and two-sample *t*-tests

- We will not go into the theory of power and sample size calculations
- To do calculations in R, one has to deal with the following parameters:
  - desired size of effect (“delta”)
  - SD (supposed to be the same in both groups)
  - significance level (e.g., 0.05)
  - power (expressed in %)
  - sample size

An example…

- Suppose we want to perform an experiment to determine the effect of temperature on growth (dry weight) in an aquatic beetle
- There will be two temperature treatments, hence the use of the two-sample t-test
- From literature, we know that SD of dry weight values in this species is 1.8 mg
- We wish to compute a sample size that with a power of 80%, using a two-sided t-test at the 5% significance level, can detect a difference of 3.0 mg

Calculating the sample size with the function `power.t.test()`

```
> power.t.test(delta = 3.0,
    sd = 1.8,
    sig.level = 0.05,
    power = 0.8)
     Two-sample t test power calculation

              n = 6.76095
          delta = 3
             sd = 1.8
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Calculating the power

```
> power.t.test(n = 15,
    delta = 3.0,
    sd = 1.8,
    sig.level = 0.05)
```

```
     Two-sample t test power calculation

              n = 15
          delta = 3
             sd = 1.8
      sig.level = 0.05
          power = 0.9927162
    alternative = two.sided

NOTE: n is number in *each* group
```

Suppose, we wish to perform a paired t-test

```
> power.t.test(delta = 3.0,
    sd = 1.8,
    sig.level = 0.05,
    power = 0.8,
    type  = "paired")
```

```
 Paired t test power calculation

              n = 8.943881
          delta = 3
             sd = 2.8
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

Suppose, we wish to perform a one-sample t-test

```
> power.t.test(delta = 3.0,
    sd = 1.8,
    sig.level = 0.05,
    power = 0.8,
    type  = "one.sample")
```

```
 One-sample t test power calculation

              n = 8.943881
          delta = 3
             sd = 2.8
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

The same n as for a paired test

Exercises

- Examine what happens
  - with the sample size of a two-sample test if we
    - decrease the desired delta
    - increase the desired power
    - decrease the significance level
  - with the power of a two-sample test if we
    - increase the standard deviation
    - decrease the sample size
    - decrease the significance level

7. Power and the computation of sample size

7.3. Computation of the sample size and power for comparison of proportions

Power and sample size for comparison of proportions: implementation in R

- The function `power.prop.test()` is used
- Don't trust the results if any of the expected counts drop below 5
- Similar to `power.t.test()`, but `delta` and `sd` are replaced with `p1` and `p2` – the hypothesized probabilities (=proportions)
- Currently, it's not possible to consider a one-sample problem

An example…

- Suppose we are going to conduct parasitological examinations of two snail populations
- From a preliminary study, we expect that the proportions of infected individuals in these populations will be 0.20 and 0.45
- We want a power of 90%, at the significance level of 0.01
- What is the required sample size?

Calculation of the sample size for comparison of two proportions

```
> power.prop.test(power = 0.9,
    sig.level = 0.01,
    p1 = 0.20,
    p2 = 0.45)

     Two-sample comparison of proportions power calculation

              n = 101.9511
             p1 = 0.2
             p2 = 0.45
      sig.level = 0.01
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```


Author: Sergey Mastitsky

Calculation of power for comparison of two proportions

```
> power.prop.test(n = 50,
    sig.level = 0.01,
    p1 = 0.20,
    p2 = 0.45)

     Two-sample comparison of proportions power calculation

              n = 50
             p1 = 0.2
             p2 = 0.45
      sig.level = 0.01
          power = 0.5384273
    alternative = two.sided

NOTE: n is number in *each* group
```