Biostatistics I: Statistical tests for categorical data

Eleni-Rosalina Andrinopoulou

Department of Biostatistics, Erasmus Medical Center

e.andrinopoulou@erasmusmc.nl

@erandrinopoulou

Erasmus MC
University Medical Center Rotterdam

z-test for proportions

One-sample

Is the probability of being diagnosed with asthma now different than it was 50 years ago?

Two-sample

Is the probability of being diagnosed with asthma in the Netherlands different than in Belgium?

One sample z-test for proportions: Theory

Scenario

Is the probability of being diagnosed with asthma now different than it was 50 years ago?

Hypothesis

$H_0 : \pi = \pi_0$
$H_1 : \pi \neq \pi_0$

One sample z-test for proportions: Theory

Hypothesis

If **one-tailed**
Is the probability of being diagnosed with asthma now higher than it was 50 years ago?
$H_0 : \pi = \pi_0$
$H_1 : \pi > \pi_0$

or

Is the probability of being diagnosed with asthma now lower than it was 50 years ago?
$H_0 : \pi = \pi_0$
$H_1 : \pi < \pi_0$

One sample z-test for proportions: Theory

Test statistic

For large sample sizes, the distribution of the test statistic is approximately normal

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

- Sample proportion: p
- Population proportion: π_0
- Number of subjects: n

If continuity correction is applied: $z = \frac{p - \pi_0 + c}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$,

where

- $c = -\frac{1}{2n}$ if $p > \pi_0$
- $c = \frac{1}{2n}$ if $p < \pi_0$
- $c = 0$ if $|p - \pi_0| < \frac{1}{2n}$

One sample z-test for proportions: Theory

Sampling distribution

- z-distribution
- Critical values and p-value

Type I error

- Normally α = 0.05

Draw conclusions

- Compare test statistic (z) with the critical values$_{\alpha/2}$ or the p-value with α

If **one-tailed**: Compare test statistic with the critical value$_{\alpha}$

Bionomial test

One-sample

Is the probability of being diagnosed with asthma now different than it was 50 years ago?

- If the normal distribution cannot be used, then we need to use the binomial distribution

Chi-square test

The chi-square test tests the statistical significance of the observed relationship with respect to the expected relationship

- Two variables are related or independent
- Goodness-of-fit between observed distribution and theoretical distribution of frequencies

Fisher’s Exact Test

- Fisher’s exact test is an exact test
- Fisher’s exact test is a special case of **permutation** tests

  - Calculate the original test statistic
  - Shuffle (permute) the data and calculate the test statistic
  - Repeat the above step for every possible permutation of the sample
  - Calculate the fraction of the values of the test statistic that are as extreme or more to the original test statistic

Fisher’s Exact Test: Theory

Advantages/Disadvantages

- The advantage is that permutation tests exist for any test statistic, regardless the distribution.
- The disadvantage of this type of tests is that it can become computationally very intensive

Assumptions

- Both row and column marginal totals are fixed in advance