

4.6: How good is the proportion?

Proportions are frequent in the data analysis, especially for categorical variables. How to check how well the sample proportion corresponds with population proportion?

Here is an example. In hospital, there was a group of 476 patients undergoing specific treatment and 356 among them are smokers (this is the old data). In average, proportion of smokers is slightly less than in our group (70% *versus* 75%, respectively). To check if this difference is real, we can run the *proportions test*:

(We used [two.sided](#) option to check both variants of inequality: larger and smaller. To check one of them (“one tail”), we need [greater](#) or [less](#)^[1].)

Confidence interval is narrow. Since the null hypothesis was that “true probability of is equal to 0.7” and p-value was less than 0.05, we *reject* it in favor to alternative hypothesis, “true probability of is not equal to 0.7”. Consequently, proportion of smokers in our group is *different* from their proportion in the whole hospital.

Now to the example from foreword. Which candidate won, A or B? Here the proportion test will help again^[2]:

According to the confidence interval, the real proportion of people voted for candidate A varies from 100% to 47%. This might change completely the result of elections!

Large p-value suggests also that we cannot reject the null hypothesis. We must conclude that “true p is not greater then 0.5”. Therefore, using only that data it is *impossible* to tell if candidate A won the elections.

This exercise is related with phyllotaxis (Figure 4.7.1), botanical phenomenon when leaves on the branch are distributed in accordance with the particular rule. Most amazingly, this rule (*formulas of phyllotaxis*) is quite often the *Fibonacci rule*, kind of fraction where numerators and denominators are members of the famous Fibonacci sequence. We made R function [Phyllotaxis\(\)](#) which produces these fractions:

In the open repository, there is a data file [phyllotaxis.txt](#) which contains measurements of phyllotaxis in nature. Variables [N.CIRCLES](#) and [N.LEAVES](#) are numerator and denominator, respectively. Variable [FAMILY](#) is the name of plant family. Many formulas in this data file belong to “classic” Fibonacci group (see above), but some do not. Please count proportions of non-classic formulas per family, determine which family is the most deviated and check if the proportion of non-classic formulas in this family is statistically different from the average proportion (calculated from the whole data).

This page titled 4.6: How good is the proportion? is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by Alexey Shipunov via source content that was edited to the style and standards of the LibreTexts platform.