ece[s]

EUROPEAN CENTRE
FOR ELECTORAL SUPPORT

PROSES

Project in Support of Enhanced Sustainability
and Electoral Integrity in Afghanistan

Brief Guidelines on Statistical Tools in Election Observation

`

`

Table of Contents

`

About the Author

The author of this guide, Łukasz Widła-Domaradzki, has 21 years of experience in analysing data and teaching statistical techniques. Łukasz is an experienced user of advanced statistical programmes such as R, SmartPLS and SPSS AMOS.

He has worked for several international research and consulting companies as well as for governmental organizations in Poland. Throughout his career, he has authored over 20 scientific articles about statistics. Łukasz was a leading expert in 2016 evaluation of Polish Aid in Kenya and worked as a Data Analyst during EU EOMs in Liberia (2017), Sierra Leone (2018), Pakistan (2018) and Nigeria (2019). In Afghanistan, he provided training in data analysis for relevant election stakeholders, under the EU-funded project PROSES, implemented by ECES.

About the Project

The European Centre for Electoral Support (ECES) is a not-for-profit foundation headquartered in Brussels which promotes sustainable democratic development through the provision of advisory services, operational support and management of large projects in the electoral and democracy assistance. ECES is implementing the Project in Support of Enhanced Sustainability and Electoral Integrity in Afghanistan (PROSES) funded by the European Union through the Instrument contributing to Peace and Stability (IcSP). PROSES outcomes are:

Outcome 1: Electoral integrity vulnerabilities are proactively identified and reduced through administrative planning and actions.

Outcome 2: Increased capacity of electoral stakeholders to apply evidence-based and effective leadership approaches to the conduct of elections and the adjudication of election disputes.

Outcome 3: Measures contributing to the accountability and inclusivity of political and electoral processes through the broad participation of relevant electoral stakeholders with specific emphasis on women.

PROSES' activities feed into a comprehensive electoral integrity strategy based on the analysis of integrity vulnerabilities from previous Afghan electoral cycles, recommendations from EU EOM/EATs and an upcoming electoral political analysis. PROSES is supporting vital electoral stakeholders' efforts in promoting enhanced integrity and credibility of the electoral process. In brief, our overall strategy is multi-stakeholder: partnering and supporting all individuals, organisations and processes (civil society, media, political parties, EMBs, Government, justice system) that can play a role as agents for reform

`

and change in the Afghan electoral process and democracy. To identify and support Afghan-led and Afghan-owned collective actions to mitigate fraud and malpractice through the work of alliances and coalitions with key Afghan players. While the tangible impact of many integrity actions will take time, the strategy intends to lay the foundations, through multiple actions and milestones, to produce changes in the mid-term and long-term by identifying agents for change and meaningful actions.[1]

Goals of these guidelines

The aim of these guidelines is to outline some basic concepts of electoral data and electoral analysis, followed by practical tips to be used as you build a database using an open-source software. These guidelines are divided in five sections. The first section introduces the concept of sample as the basis for building a database. The second section explains what election observation forms are and how they can be used to gather data for the database. The third section aims to explain more advanced statistical analysis terms for data analysis. The last section, introduces few tools for data analysis and provides some tips for statistical analysis with these tools.

SECTION I – SAMPLE

Population and sample

A population in a data set contains all members of a specified group. A sample data set contains a part, or a subset, of a population. Hence, the size of a sample is always less than the size of the population from which it is taken.

The most popular understanding of the population is the population of type or categories of people:

- Population of Afghan voters
- Population of citizens of Afghanistan
- Population of Afghan adults

But the term population is broader. In fact, we may also have other data sets that we can call populations in statistical terms:

- Population of Polling Centres
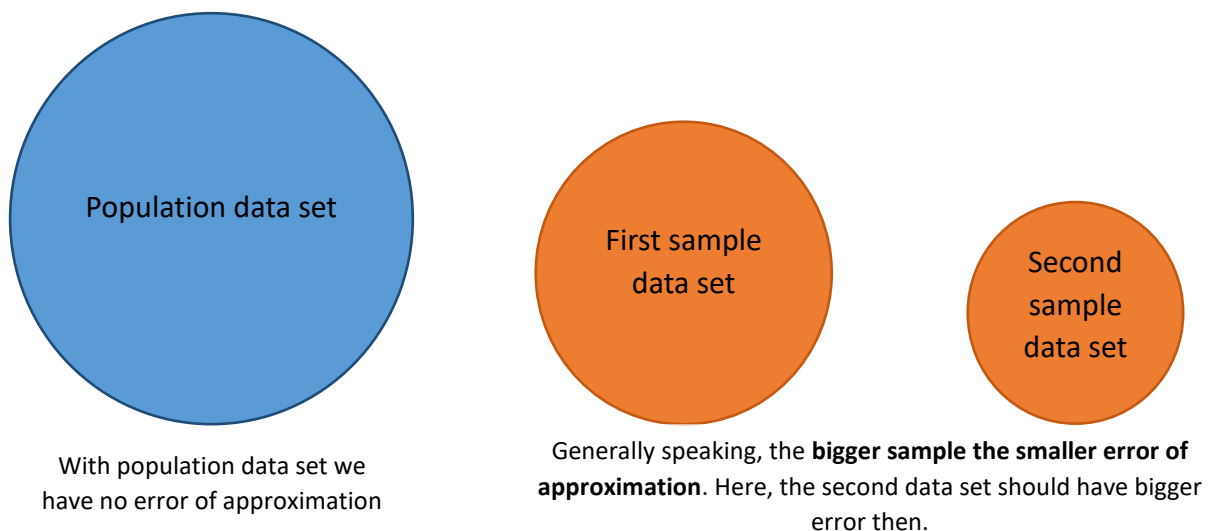- Population of opened (or closed) Polling Stations

---

[1] Further read on electoral integrity can be found at http://www.eods.eu/publications

Accordingly, we can have samples from those populations:

- Sample of Afghan voters
- Sample of citizens of Afghanistan
- Sample of Afghan adults
- Sample of Polling Centres
- Sample of opened (or closed) Polling Stations

Error of approximation

Usually we will have access only to the samples, not to the population results. After the sample is drawn and gathered, we will have to deal with an error of approximation. With samples, there will always be some level of error. We should be aware of level of the error, but the error itself is inevitable. It is not something wrong, it's just a part of statistical reality. Generally speaking, the bigger sample, the smaller the error of approximation.

Population data set

First sample data set

Second sample data set

With population data set we have no error of approximation

Generally speaking, the **bigger sample the smaller error of approximation**. Here, the second data set should have bigger error then.

Of course, the size of an error depends on many elements, not only the sample size. If the sample, even a large one, was collected without a proper sampling plan or it's not really representative, the error will be bigger. This is because the total error of approximation contains two parts:

- Sampling error
- Bias

While we can easily assess sampling error,[2] assessing the bias is more difficult. The bias derives from a poor sampling design. If, for example, we only collect a sample from the urban areas, or if we are focusing only on male polling stations or if we try to observe only those polling stations where allegedly there were some irregularities – we are exposing our sample to an unknown bias level.

[2] For example, by using online tools, such as this and this

`

The situation is quite easy to manage, if we have only a sampling error. If you are dealing with a sample, not with a population, you will encounter sampling error (SE) of approximation:

If we have population of 5'000 and draw a sample:

- Sized 50 – we will have 13,8% of SE
- Sized 100 – we will have 9,7% of SE
- Sized 500 – we will have 4,2% of SE
- Sized 1000 – we will have 2,8% of SE
- Sized 2500 – we will have 1,4% of SE
- Sized 4000 – we will have 0,7% of SE

Let's look into an example of how this works. In the chart below, you will find the approximate results for two candidates. In reality, Candidate A got 46% of all the votes and Candidate B got 42%. If we collect a very small sample, let's say only 50 cases (N=50), we can expect a sampling error at the level of 13,8%. That's why we can expect that in our sample candidate A will have a result between 32,2% and 59,8% and candidate B – between 28,2% and 55,8%. So, where is the problem? If we extract averages from those numbers, we will have the real numbers, right? Let's check: the average of 32,2 and 59,8 is: (32,2 + 59,8)/2 = 92/2 = 46. For the candidate B we will have: (28,2 + 55,8)/2 = 42. Well, the problem is that you usually draw only one sample, not an infinite number of samples, right? So, in the example with the N=50 you may get the results that candidate A got 59,8% and candidate B got 28,2% as well as the results where candidate A got 32,2% and candidate B got 55,8%. And both situations are equally possible!

With a large sample (let's say, N=4000), the accuracy of the sample will be higher. The result for candidate A may vary from 45,3% and 46,7% and for candidate B: from 41,3% to 42,7%. All those numbers are closer to the reality.
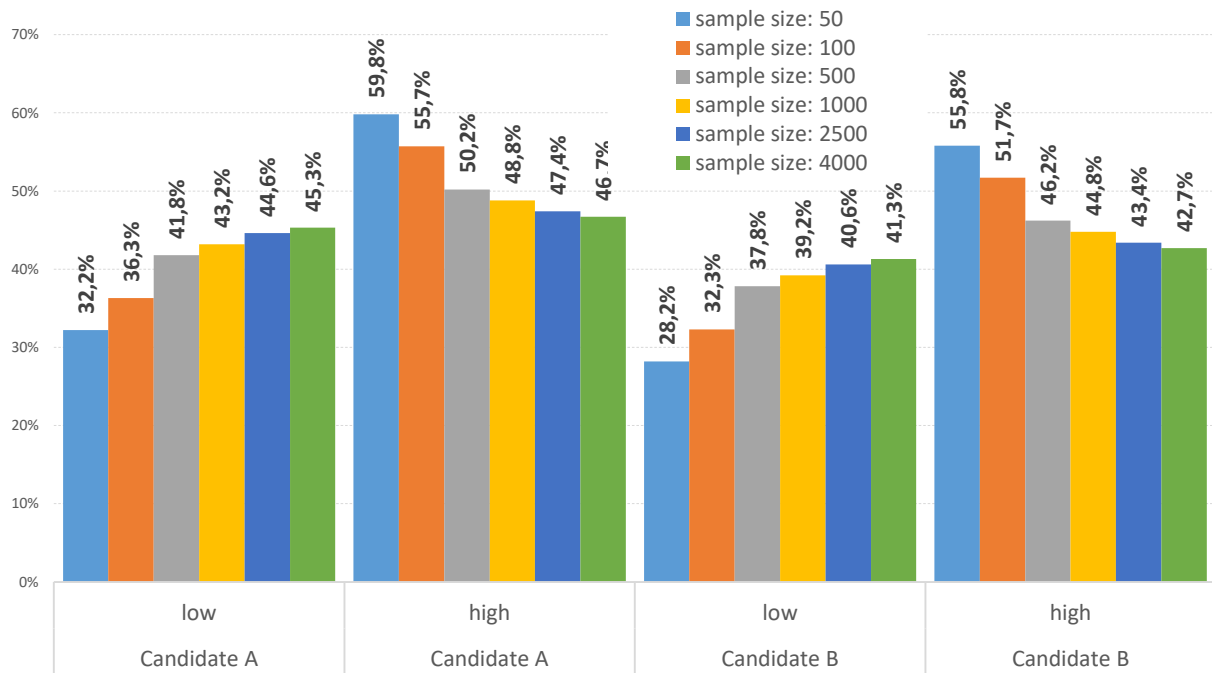
Randomness

Statisticians like randomness. Randomness of the sample is the situation where only sampling error will occur. And sampling error is known and easy to assess. Randomness is the situation when every single element of the population has exactly the same chance to be included in a drawn sample. How to make our sample random? The easiest way is:

1. Take the information of the whole population
2. Decide the size of the sample
3. Assign a random number to each element of the population
4. Select those elements with highest or lowest random number

To get the random number, you can use the simple Excel formula. It's called random number generator:

=RAND()

How to collect the sample

Once we draw our sample from the population, we have to collect the desired information from selected cases. What if the selected case is hard to reach?

This may lead to the situation when the filed observer:

- Will skip the element and will provide smaller sample (without problematic cases)
- Will replace the inaccessible element with the other population element

`

Some statisticians will just abandon the problematic case and lower the sample size. This will bring some consequences: the sampling error will be higher, but the non-random error will not appear. Sometimes we will decide to replace the case that we can't obtain information from. **But we should do this in a random manner as well.**

It is important to understand how the replacement process is implemented. If it's left to the decision of field observer, if it's done in not random manner, if it's made without any procedure – we are exposing our sample to an unknown bias.

To avoid this, we may draw two samples, using the random number generator shown above. The first sample will be our primary sample, and we should underline that this is the sample we should collect at the first place. If any element of this primary sample is inaccessible, we may replace it by the element from the second drawn sample – secondary sample. With that approach we can mitigate bias and deal only with sampling error, which is – as said before – easy to assess.

Outliers

After collecting sample, we will have our results. Let's say we collect data at the polling station level. In our sample we have 50 cases (N=50). We tried to assess the performance of the polling station. At the example below we can see most of our polling stations performed "well" (mark 4) or "very well" (mark 5). Only one case, case 31 performed "poorly" and got worse evaluation (mark 1):
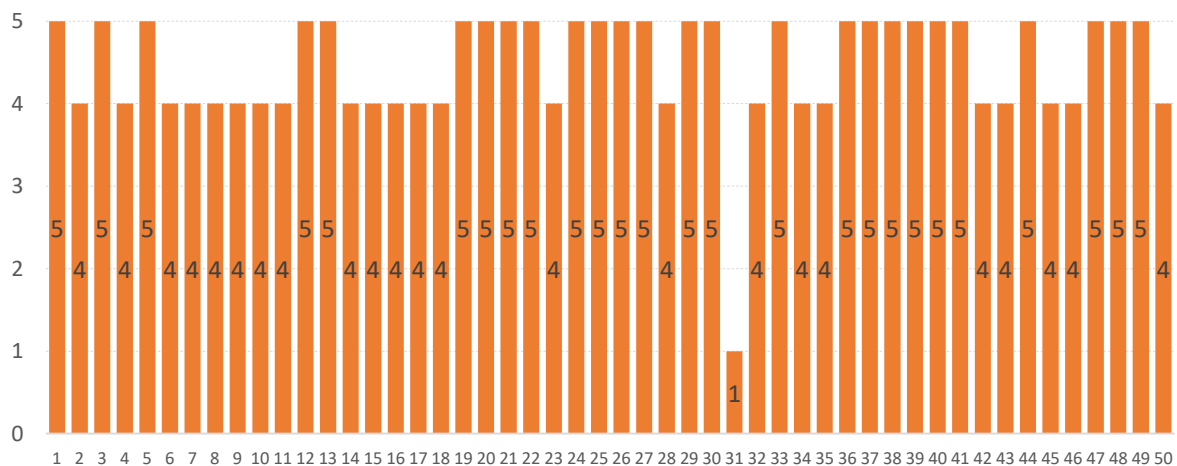
This particular polling station may be an interesting case, in which we should look into with a qualitative approach (for example we may ask observers what happened there), but from the statistical point of view, the case 31 is an outlier. An outlier may influence produced statistics. In this example the average (or mean) of the collected sample will be 4,46. If we exclude the case

`

31 it will be 4,53. It's not a big difference, but remember it is influenced by only one case!

SECTION II – FORMS

Forms

Observation forms (or questionnaires, or checklists) are your tools to collect data. Forms should be standardized, which means you should always ask questions the same way when you are filling the forms. Election observation forms are designed to gather data from the field.

We can design different forms for different stages of election i.e.: opening form, voting form, closing form, tabulation form, etc. We can also design one big form to cover all the election stages. It really depends on how you plan to observe the elections: if you will switch from polling station to polling station during the election day, you may probably design a separate forms for different stages – because it's easier for observers to fill the form and send it and go to the another polling station afterwards. If you remain in the same polling station throughout the day, you will need probably one big form for all the stages: because the situation may change and you may need to change your previous assessment

Questions

Each form will contains questions. In principle, we have three types of questions:

- Closed
- Open-ended
- Semi-closed (or precoded)

From the statistical point of view there is a hierarchy of questions, here ranged from most to least useful:

- Closed with 0/1 scale (e.g. yes/no scale)
- Closed with multiple scale (e.g. 5-point scale)
- Closed with multiple answers
- Precoded open-ended questions
- Open-ended questions

Closed questions should be the vast majority in your form. For example, when you ask:

- State of reality (e.g. is the ballot box sealed?) or
- An opinion (e.g. what is your assessment of the organization of this Polling Station?).

`

Closed questions are the main source of the statistical information. Why? Because they can easily be converted into the format number and, therefore, they can easily be transformed into charts. For example, you will have to choose the answer from the list (e.g. yes/no or very good/good/bad/very bad).

Sometimes you will be asked to answer the question with a text. This kind of question is called **open-ended**. This kind of question (e.g. 'please describe behaviour of the security forces) is sometimes very important to the people who are writing reports, but have very little use for the statistician. Usually open-ended questions can't be easily transformed into numbers (to do this you need to code them).

The solution for some open-ended questions is to precode them. If you have an idea on what the answer for your open-ended question may be, you can precode it. Let's consider the question "please describe behaviour of the security forces". It may be open-ended, but you may think of possible answers, e.g.:

- Alert
- Aggressive
- Chaotic
- Calm
- Threatening
- Other (please describe …………………………………………)

Scales

Each question should have its relevant scale. Most common are:

- Dichotomous / Dummy / 0-1 type of scale. Here the scale has only two possible answers. Usually it will be "yes/no"
- Likert scale. Likert scale is a 5-level scale. Usually:
    - Strongly disagree
    - Disagree
    - Neither agree nor disagree
    - Agree
    - Strongly agree
- 4-point scale. A scale that looks like Likert scale, but without middle level. 4-point scale is usually constructed as:
    - Strongly disagree
    - Disagree
    - Agree
    - Strongly agree
- Continuous scale. This scale is rarely used. E.g. "how many points from 0 to 100 you would give?"

`

Tips for preparing the forms

- **Do not overcomplicate**. Usually easy questions are sufficient. If you need to explain something to the observers, train them. You don't need (and they don't need) overcomplicated questions with a lot of explanation inside. If – for example – you need to ask about presence of the Presiding Officer ask: "Is the Presiding Officer present?".
- **Avoid ambiguity.** Ask one question at the time. If – for example – you have to ask if the security forces are present and ask about their behaviour, you should ask two separate questions. One about presence, and second about behaviour. Ambiguous questions are very hard – or even impossible – to analyse in a statistical manner.
- **Pay attention to the scales.** Every question should be accompanied with the relevant scale. If you are asking about presence of the Presiding Officer, do not use Likert scale!
- **Be careful with length of the form** and relevance of the questions. Too short form may be a challenge for the statistician, because there will be not much information. But avoid long forms. Ask yourself if you would really use all the questions that you are asking for? Your observers in the field can't be overwhelmed with the length of the form – they are working during E-day all around the clock.

SECTION III – ANALYSIS

Frequencies

Frequencies are the distribution of answers for ONE variable. For example: you asked if the "Presiding Officer was present at the polling station?" and you get 40 answers. Some of those answers will be "yes", some – "no". Frequencies will show you what is the proportion of "yes" and "no" for this question:

`
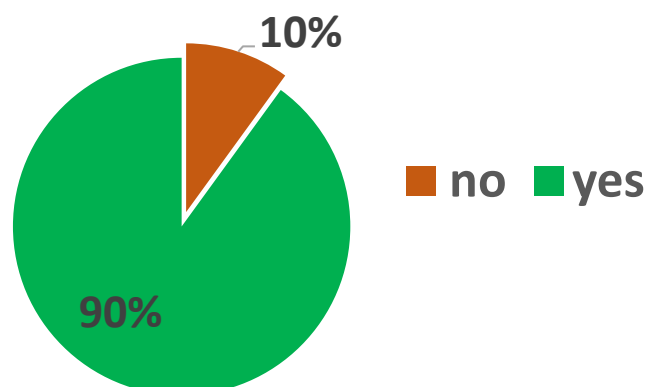
Crosstabs

Crosstabs are the distribution of answers for two (or more) variables. For example, let's assume the same question. You got 40 answers, 10 from female polling stations, 30 from male polling stations. It may look like this:
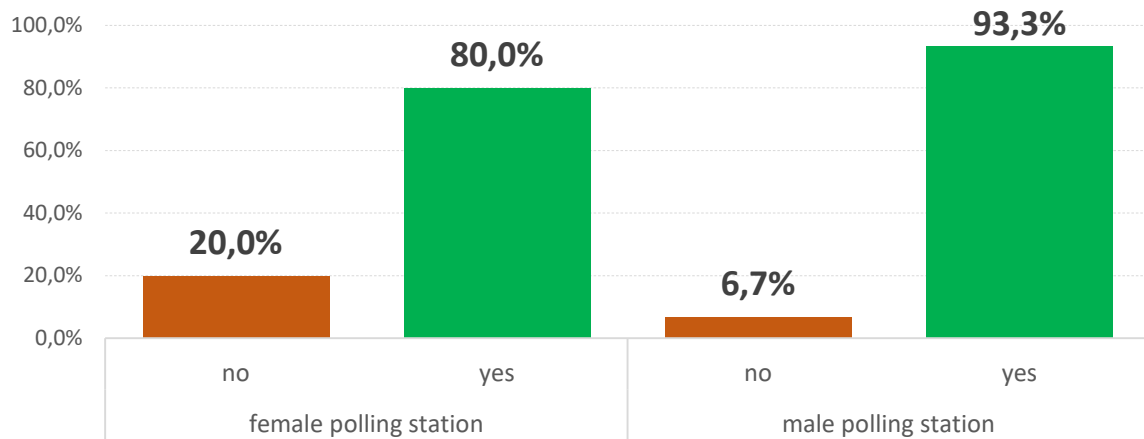

Means (averages)

For means (averages) we have to change the example. Let's say, you asked about overall assessment of the polling process at the polling station. The question was asked at a classical Likert 5-point scale (with 1 – "very bad" / 5 – "very good"). You still have 40 filled forms. Your average is just one number, instead of the 5 you had on your original scale:

Average: 3,1 on 5-point scale
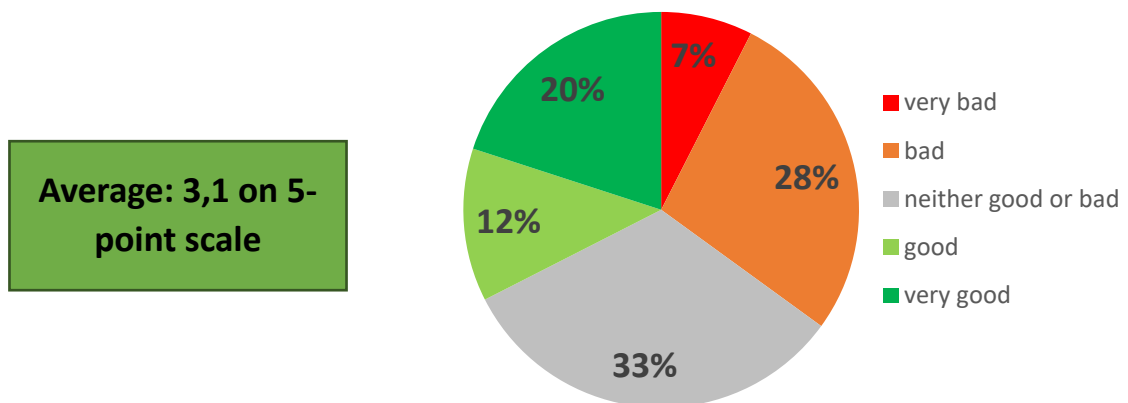
Correlations

Correlations are the first step into serious statistics. What correlation can show is the relation between two (and always two) variables. Is there a statistical tendency to have more absent polling officers at the female polling stations? Or the difference (20% to 6,7%) is just too small to have significant meaning? Let's assume the already known example:
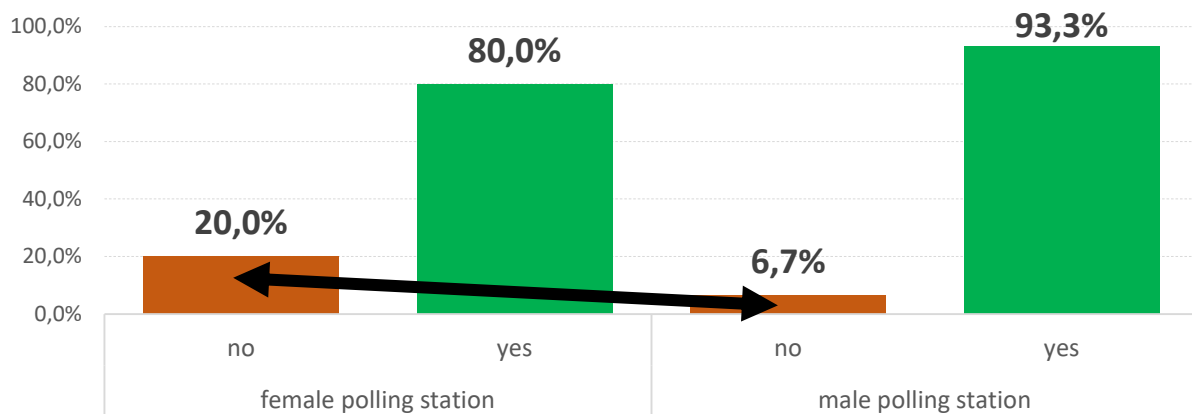
Let's assume, we still have only 40 cases in our sample. After calculations we get the answer:

- Correlation = 19,2%
- Significance of the correlation = 22,6%

That means the correlation is not statistically significant.

But correlations are sensitive to the size of the sample. If, in a given example, we increase the sample size 6 times (from 40 to 240), we can perform new calculations:

- Correlation = 19,2%
- Significance of the correlation = 0,2%

As you can see, the strength of the correlation remains at the same level, but the significance decreased. That means the correlation is statistically significant if the sample size is bigger.

SECTION IV – PRACTICAL TIPS

All of the analysis above may be performed in Excel or R software. Excel is a basic tool and it's easy to use (while is quite hard to master). R is a devoted statistical tool with a lot of possibilities, but with complicated coding skills (syntax). Below is just a quick overview of how to prepare simple statistics using Excel and R.

To use R, you need to download (and install) the program itself and download (and install) a graphical interface (called Rstudio):

- R (https://cran.r-project.org/bin/windows/base/)
- Rstudio (https://rstudio.com/products/rstudio/download/)

Task	Excel	R
Frequencies	**=COUNTIF(data!G:G,frequencies!C3)** where: - data!G:G contains data - cell C3 contain the number we would like to count (e.g. "1")	**mytable <- table(data$B0)** **mytable** where: - database is uploaded as a "data" object - "B0" is a code for variable for the frequencies - "mytable" is a newly created object and it contains frequencies
Crosstabs	No easy way to perform in Excel. The easiest way is to use **pivot tables**	First option: **table (data$B0, data$B7)** where: - database is uploaded as a "data" object - "table" is a newly created object and it contains crosstabs Second option: **table1 <- table (data$B0, data$B7)** **ftable(table1)** where: - database is uploaded as a "data" object - "table1" is a newly created object and it contains crosstabs - "ftable" is a way of showing results
Means (averages)	**=AVERAGE(B:B)** where: - column B contains data	**mean (data$B0)** where: - database is uploaded as a "data" object - "B0" is a code for variable for the mean
Correlations	To get the correlation: **=CORREL(A:A;B:B)** where: - Column A contains first variable, column B contains second variable To get the significance: **=T.DIST.2T(ABS(C*SQRT(N-2)/(1-C^2));N-2)** where: - T.DIST.2T is the main formula for testing significance - ABS is the absolute sign of the formula for correlation - SQRT is the square root - C is the correlation cell (to count separately) with =CORREL(A:A;B:B) formula - N is the sample size	First option: **rc <- rcorr(as.matrix(data), type="pearson")** **print(rc)** where: - database is uploaded as a "data" object - "rc" is a newly created object and it contains correlation with significance test. Second option: **ct <- cor.test(data[["B0"]], data[["B7"]])** **print(ct)** where: - database is uploaded as a "data" object - "ct" is a newly created object and it contains correlation for variables B0 and B7 with significance test.