

Pavel Hanes - William Gibson - Dana Hanesova\*

---

## USING MATHEMATICAL STATISTICS ETHICALLY IN RESEARCH IN THE SOCIAL SCIENCES AND HUMANITIES

*The aim of the authors is to consider some of the ethical challenges caused by the use of statistical methods for research in the humanities and social sciences (specifically education sciences). Statistics has the potential to be a useful tool in the research methodology of these sciences but, on the other hand, it can be abused and misused when the scientists are not aware of its limitations and pre-conditions. Most of the problems arise with deciding what to do to try to answer their questions and how to collect the data. How the data are collected affects how the data are analysed, and how the data will be analysed should affect how the data are collected. The authors will make an introduction to the topic by showing the tension in choosing either quantitative (statistical) or qualitative methodology for research in the human and social sciences over recent decades. Then they will present some ethical consequences of the misuse of mathematical statistics and emphasise the need for greater awareness and education about its correct use in the social sciences and humanities.*

**Keywords:** social sciences, humanities, mathematics, statistics, ethics, research

### 1. Introduction

The current globalized and neo-liberal valued worldview dictates and shapes the organization of education and research methodology across the spectrum of the contemporary sciences. The expenses for research and the university education of future scientists are controlled by the competitive political and economic interests of the parties in power. The *raison d'être* of universities and research academies seems to have been shifted from producing educated people to producing economic goods, from centres of genuine scientific research to institutions competing to meet the challenges of the free-market economy, by producing individuals who are as economically effective as possible. Kascak & Pupala [1] share Liesner's worries [2] that, "with the exception of a few elite universities, most universities would employ people primarily because of their enthusiasm for 'fashionable methods' rather than for their orientation for research".

How has this global situation influenced research in the social sciences and humanities? When we look back at its history in the last 50 years [3], [4], [5], [6], [7], it has been marked by the tense co-existence of the positivist and interpretivist methodological paradigms. The former prefers quantitative research methods, the latter qualitative ones. The former focuses on reliably measuring the impact of various processes on the final products, including

the people concerned. This methodology has typically been used by economics, pharmacy, medicine and other natural sciences, but also by psychology, sociology, education (e.g. school league tables, PISA education rankings), and even by practical theology (e.g. measuring the mutual impact of religion and culture [8]). The second paradigm is a personality-focused approach focused on comprehension and evaluation of the quality of inherent human cognitive, emotional or social processes, such as the needs, ways and reasoning of thinking, including the impact of value feedback and reflection on people in various societal contexts - e.g. social care, psychology, sociology and the education sciences. One of the objections to this kind of research is that it is difficult to generalize results to a larger population; another is its fragmentation. Because of multiple legitimate criticisms during the last decades, the interpretivist researchers have developed a whole range of methods, and combinations of them, in order to improve the validity of their qualitative (e.g. phenomenological, ethnographic, comparative) research studies.

Under the current pressure of neoliberalism, and thanks to the massive development of modern technologies and statistical software, it has become possible to process astounding amounts of meta-data; "data-mining" has become a major theme. The developed economies support extensive research samples with an unprecedented large scope of data, including meta-analyses and

---

\* <sup>1</sup>Pavel Hanes, <sup>2</sup>William Gibson, <sup>1</sup>Dana Hanesova

<sup>1</sup>Department of Theology and Catechetics, Faculty of Education, Matej Bel University, Banska Bystrica, Slovakia

<sup>2</sup>Royal Grammar School, Newcastle, England

Email: pavel.hanes@umb.sk

systematized reviews, enabling them to make optimal economic decisions in a specific social area, similarly as is the case with evidence-based medicine. It seems that the use of statistics for psychological, sociological and educational research is increasing. Another observable present style in the social sciences and humanities is a mixed research approach, either as a combination of quantitative and qualitative methodology, or the application of some qualitative methods within the quantitative approach, and vice versa [9].

When using the quantitative methodology, researchers in the social sciences and humanities have sometimes been accused of using it incorrectly. Besides other problems, there are some specific ethical issues which arise in connection with its use in these sciences. "Every question in human life can be framed ethically even though it may focus primarily on another dimension" [10, p. 46]. The existence of choices between various methodologies raises the ethical question of the motives influencing the decision-making process of their selection. It is similar to the ethical issue of "which highest principle gives value to technologies" [10, p. 46], we can ask which principle gives the value to the research methodology for a certain research problem.

The aim of this study is not to focus on the details of any particular research, but to consider some general ways in which statistics can be and have been misused, and to see if some lessons can be learnt to change or improve the situation. "The casual reader may not have time to investigate every methodological aspect of data collection, survey methods, but with the right approach, they will be able to better understand how data have been used or interpreted. The same approach may help the reader to spot statistics that are misleading in themselves" [11].

In our study we will look at this issue from two complementary viewpoints. We begin with a philosophical reflection on the question if, or to what extent, it is possible and ethically permissible to research "ontologically different subjects" with the same epistemological methods, i.e. statistical. The second approach is a summary of ideas how to prevent unethical use of statistics in the social sciences and humanities.

## 2. Reflections on the research subject matter

The challenge of the discussion on the issue of quantitative research between the natural sciences on the one hand, and the social sciences and humanities, on the other hand, lies in the very essence of their respective philosophies and their methodologies. The answer to the question of what is adequate method of scientific research is directly linked to the actual subject matter of the research in question.

The research subject matter of the social sciences are individuals as well as society and, in the case of the humanities, various areas of human experience and activities, such as language, literature, culture, philosophy, and religion. "In case

of humanities and social sciences, the issue researched is rather complex ..." [12]. So the complex phenomena under research in the social sciences and humanities should usually be viewed from interdisciplinary angles.

The philosophical problem with quantitative research of human and social phenomena lies in the necessity to reduce the complexity of the human and societal issues in question to just a few variables. Often the true and in some ways substantial dimensions of the phenomena have to be omitted. For example, the use of statistics in textual analysis might produce lots of data about the occurrence of certain words in some texts or social discourse, but to understand the real meaning requires deeper qualitative, etymological or hermeneutical content analysis. These differences between the interpretivist, qualitative approach and the positivist, or physicalist, quantitative approach to phenomena were characterized by Hogben: "Algebra is a language in which we describe the *sizes* of things in contrast to the ordinary languages which we use to describe the *sorts* of the things in the world" [13]. Statistical research can measure the number and even the intensity of people's attitudes, but it cannot explain the intrinsic motivation or worldview of a person or whether people really understand and mean what they have just said. Such reductionist approach to complex philosophical and emotional issues is obviously unethical.

Also Valco warns against the "relentless philosophical reductionism" of scientism - "the popular widely-spread conviction that modern science, modelled on natural sciences, is the only source of knowledge" [Hutchinson in 14, p. 10]. This view goes hand in hand with the fact that "the higher is explained in terms of the lower, mind in terms of brain, human social behaviour in terms of physics and chemistry. Humans are appreciated mainly for their instrumental value: earning capacity, socio-political usefulness and their excellence of giftedness." [14, p. 19]. Far from everything inevitable for the existence of individuals and societies can be explained by the same scientific method as is used by the natural sciences. As Kant said: "So the Biblical theologian ... draws his teachings not from reason, but from the Bible; the professor of law gets his, not from natural law, but from the law of the land; and the professor of medicine does not draw his method of therapy as practices on the public from the physiology of the human body but from medical regulations." [15, p. 35].

Grant agencies financing scientific research are often interested primarily in performance or profit [10, p. 46] which means that they tend to prefer quantitative research results. Given the complex research subject matter of the social sciences and humanities, the value of qualitative methods needs to be recognized, especially in the initial stages of combined research approach.

If there is a discussion between statisticians and researchers, they can clarify what is and is not possible. Then they can cooperate with a clear understanding of each other's area of

expertise. This is in line with the “cooperative principle” of Grice [16], [17] or an older idea of the “middle axiom” of Oldham [18].

In practice, this means that there is an opportunity for discussion between people from the whole spectrum of sciences. We need to be reminded of one of the basic principles of argumentation - the “principle of charity” - without which communication is virtually impossible [19, p. 197]. According to one definition this principle “...urges charitable interpretation, meaning interpretation that maximizes the truth or rationality of what others think and say” [20, p. 122]. The charity principle means that all parties try to debate and understand the strengths and weaknesses of different approaches, and try to appreciate the other point of view. Let us take, for example, Article 1 of the Charter of Fundamental Rights of the European Union: “Human dignity is inviolable. It must be respected and protected.” [21]. It is not difficult to point out differences in opinions about what this means in, say, politics or bioethics. Nevertheless, if the different methodological cultures make thorough use of what is common, they will make real progress in developing ethical research methodology in the face of dehumanizing tendencies of technological, neoliberal society.

### 3. Using statistics ethically

In the next part of our study, we present some examples and specific ideas about how to actually use statistics correctly and ethically in the social sciences and humanities.

#### 3.1 The power and the danger of statistics

There are two main dangers when using statistics: the first is to believe that statistics tell you everything truthfully, and the second is to believe that statistics tell you nothing, or that what statistics tell you are lies. In the first case, people say, “Statistics proves ...” or “Statistical techniques show ...” and bow to the statistical deity. In the second case, people use the well-known quotation: “There are lies, damned lies, and statistics.” [22], and thereby dismiss the data or the results.

The reality is that statistics is a powerful tool to deal with variability, uncertainty and limited information. Statistical techniques have their own methodology and assumptions. Like any powerful tool, statistics can be misused, accidentally or deliberately.

Some basic knowledge of statistics is important for most people nowadays, partly to avoid being fooled by the abuse of statistics. Most people doing research will use statistics in one form or another.

In research, there can sometimes be a choice between a qualitative approach and a quantitative approach. At one extreme the qualitative approach produces large amounts of non-numerical

data from a small number of respondents, and at the other, the quantitative approach produces small amounts of numerical data from a large number of respondents. It may well not be possible to analyse the data obtained through the qualitative approach by any statistical methods, and it is, therefore, a problem if the researcher is pressurized to use statistical methods. The strength of the statistical methods, if they are applied correctly, is that the results obtained from a sample may be extended to the population from which the sample was obtained, and the degree of certainty with which this may be done is known.

There are rules and regulations on the ethical use of statistics designed to prevent harm to the participants [23]. In this article we deal, instead, with the ethics of using the statistical methods correctly.

#### 3.2 The two main aspects of statistics

There are two main aspects to statistics. The first is the collection, representation and analysis of data. The second is fitting a model to the data and/or testing some hypotheses about the population from which the data came, and/or estimating some parameters of the population.

Often we have an idea, so we collect some data to check it out. When thinking about carrying out a hypothesis test on the data, we should check the requirements of the test before collecting the data, so that we collect sufficient data in the right way. Almost always this will involve randomisation in some form, either choosing a random sample from a defined population or randomising the allocation or order of treatments in an experiment.

D. Moore wrote: “The most important requirement for any procedure is that the data come from a process to which the laws of probability apply. Inference is most reliable when the data come from a probability sample or a randomised comparative experiment. Probability samples use chance to choose respondents. Randomised comparative experiments use chance to assign subjects to treatments. The deliberate use of chance ensures that the laws of probability apply to the outcomes, and this in turn ensures that statistical inference makes sense.” [24, p. 366].

Thus, the way in which we collect the data determines, or affects, our choice of test, and similarly the test we intend using should influence the way in which we collect the data. There are many disaster stories of researchers - in business and industry, as well as in academia - collecting their data and then going to a statistician to find out how to analyse their data, only to be told that the data are worthless.

### 3.3 The ethics of data collection

In practice, this first stage is usually thought to be easy but is often the most difficult and the most important. We need to ask the questions: What data should we collect? How should we collect it? How much data should we collect?

Let us give you a few specific examples.

1. Some of our UK friends were once asked to produce a questionnaire and do a survey about people's attitudes to windfarms. However, they were told to distribute the questionnaire at an open day at a windfarm. Almost certainly this would produce a biased sample, one which would not be representative of the general population. When they suggested a different, and better, methodology, they lost the contract. There did seem to be some ethical issues involved.
2. In the days of the Cold War, some citizens of the USA were asked the following two questions:
  - A. Do you think that press correspondents from the USA should be allowed the freedom to ask whatever questions they want of people in the Soviet Union?
  - B. Do you think that press correspondents from the Soviet Union should be allowed the freedom to ask whatever questions they want of people in the USA?

It turned out that the order in which the questions were asked made a difference to the responses which were obtained. If you are aware of this phenomenon, then there is an ethical question about which order you use for your questions and/or how you analyse the responses.

3. Elections are often preceded by opinion polls trying to predict the result. The failure of opinion polls to correctly predict election results is not a new phenomenon. In 1936 the *Literary Digest* magazine used a poll of about 2 million people, and predicted the wrong result. In fact, George Gallup made his own prediction before the magazine issued its poll, using a random poll sample of 50,000 people, and correctly predicted the result.

The problem with the *Literary Guild's* poll was that the sample consisted mostly of people who were magazine readers, car owners or telephone customers—and had money during the Depression. So it was not a representative sample.

In almost every situation it is essential to do a pilot experiment or a pilot survey. This enables you to identify problems before you collect the data you will actually use. It is wise to try to get a wide spread of people, etc. to test your plan. Failure to do this may invalidate your research.

What is the population you are aiming to investigate? Is it feasible to obtain data from the whole of this population (a census)?

Selecting a random sample is usually the best way to obtain a representative sample. However, this requires that you have a list of every member of the population in which you are interested (which could be people, cars, sections of road, planets, horses,

etc.) and are able to choose your sample so that every sample of your required size has the same chance of being chosen.

There are other methods which are used to try to obtain a representative sample. For example, you may use stratified sampling if your population divides into clear strata; in this case you take a random sample from each stratum. Other methods exist to save time and effort, and hence money, or when there is no list of the population in which you are interested.

Collecting good data is often difficult and can be time-consuming and expensive. If people are under pressure to save time and/or money they may be tempted to cut corners. It becomes an ethical issue.

Self-selecting samples are often biased. The people who volunteer to complete a questionnaire on a particular topic probably have a special interest in the topic. If there is a payment in one form or another for completing the questionnaire or being involved in the experiment, this may attract a particular sort of person. If a sample of people is chosen at a particular place and/or at a particular time, that may have a large influence on the type of person involved in the survey, and this could bias the outcome.

Similarly, if the researcher deliberately chooses a particular group of people, they should try to demonstrate that the group chosen is likely to be representative of the population being investigated. Here the ethical questions are: Am I really trying to answer a question, or am I actually trying to prove my point? Are questions asked in an unbiased way, or are they designed to obtain the result I want?

At least one of us has been involved in a survey where the way in which the questions were asked resulted in less than half of the students involved being able to complete the questionnaire. Only completed questionnaires were accepted. Those able to complete the questionnaire were not representative of the population being investigated.

If researchers are aware of how to collect the data, but instead choose to collect data which they know are likely to be biased in some way, then that is an ethical issue. Of course, the researcher may be aware of the issues but be unable to obtain a better sample, then that should be part of the report and it is probably not reasonable to use statistical tests, etc. on the data.

### 3.4 A recent example of good practice

In 2016 the Royal Statistical Society's *Series A Journal* published an article "Does preschool boost the development of minority children? The case of Roma children." [25]. The authors start by outlining the problem: "Social, economic and political exclusion remains an everyday challenge that ethnic minorities face in modern societies. A key reason for the 'vicious cycle of exclusion and poverty' is the gap in educational achievements caused by disadvantaged family backgrounds and residential segregation ... As a consequence, numerous policy experts suggest

providing minorities with unlimited access to the education system of the host country, in particular to the early education system.” They then go on to focus on Roma children.

They explain that data were collected by the United Nations Development Programme (UNDP), the World Bank (WB) and the European Commission (EC), the so-called UNDP-WB-EC regional Roma survey. They explain how the data were collected: “The survey was conducted in a three-stage random representative sampling process: first, in each country 110 random clusters of approximately 30 households from areas of compact Roma populations were selected; second, in each cluster, seven households were randomly chosen and the respective head of the household answered questions about the household; third, one random household member older than 15 years was selected to answer a battery of questions on status and attitudes.” The study represents an example of the use of the random sampling methods.

The population being investigated is carefully defined: “We restrict our sample to the Roma population only: to children at preschool age (3-6 years old) and to households which have not moved during the previous 5 years (which is the case for more than 95%). This restriction guarantees that the children under study have at least some time and chance to attend preschool in the location of residence.”

A control group is also defined: “For descriptive comparisons we also draw on 569 non-Roma children who live close to the Roma households under study. These non-Roma children are exposed to similar regional conditions. As such, any differences between the Roma children and the non-Roma children who were included in this sample should not result from the fact that Roma often live in regions that are most affected by poverty and unemployment.”

The authors use data collected on a large scale. Nevertheless, the principles also apply to smaller scale projects. The question being investigated is clarified, as is the population being investigated. Then a method of collecting a random sample is used to attempt to obtain a representative sample. With decent data it is then worth doing some analysis. In fact, the authors use a sophisticated method to analyse the data, but it is the initial stages from which we can learn some helpful lessons.

### 3.5 The ethics of data analysis

Once the data have been collected, it is usually necessary to clarify what the data have to say, both for ourselves and to communicate the results to others. It is often difficult to make sense of the “raw” data, so we summarise them and/or use some form of visual representation.

Even choosing which “average” to use should be a deliberate choice, as there are several options available, and may be an ethical decision. The most commonly used average is the

(arithmetic) mean. If you have one very large value, this will have quite a big effect on the mean, especially if the sample is fairly small.

For example, suppose that there is a small business with one CEO and ten employees. If the CEO is on a very high salary and the employees are on low salaries, the mean salary could still be reasonably high ... and certainly quite a bit higher than any of the salaries of the employees. In this situation a better average might be the median: the middle value when the salaries are arranged in order of size. So this average would be one of the workers’ salaries, and lower than the mean.

The median is said to be a “resistant” statistic; it is usually not affected by a few extreme values in the data set. The inter-quartile range is also a resistant statistic, whereas the standard deviation is not. So, even the choice of which average or measure of spread to use should be a deliberate decision.

Of course, it is possible to choose the average (or measure of spread) which communicates what we want to communicate, which may not be the most honest representation of our data.

### 3.6 The ethics of data representation

“A picture is worth a thousand words.” However, with the statistical software available nowadays, it is easy to produce lots of graphs, so it is still necessary to choose which graph(s) to use and not just use everything available. Pie charts are popular in the media but are they are not very useful for research.

On the other hand, some people only present their data in tables, and this makes it harder to take in what the data have to say.

When using a graph to represent data, there are many well-known tricks. For example, if we use the fictitious data below (Table 1) about the number of hours contributed by the authors of this paper, we can use two simple graphs to represent the data (Figure 1). They give very different impressions, even though both are “correct”.

Advertisers are well aware of the effects of different ways of representing data. Researchers may or may not be so aware of the effects of their choices. If they are, the choice becomes an ethical decision.

### 3.7 Probability

Probability theory underlies statistical tests. If we want to know the probability of something happening, such as a drawing pin landing point up, we do the experiment lots of times and record the proportion of times the event occurs. The “true” probability is the proportion if we were to repeat the experiment in an infinite number of times.

Table 1 Example of data representation

Contributor	Hours
Hanes	123
Gibson	120
Hanesova	121

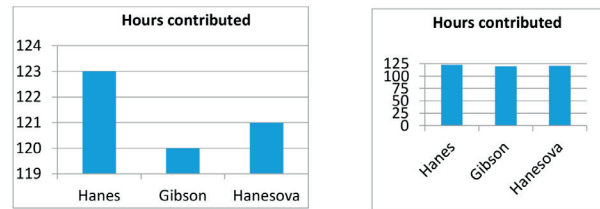


Figure 1 Examples of graph representation

In explaining the approach to statistical hypothesis tests, we shall often refer to what would happen if we did something lots of times; this is because of our definition of probability.

### 3.8 Statistical hypothesis tests, or tests of significance

Like every other discipline, statistics has its own jargon and methodology. When you understand it, communication is easier. Researchers and others are often encouraged or required to use statistics in their research, and therefore need to understand the key words and phrases. This applies even for just reading the research reports. Further, it is better and safer if the researcher understands how the statistical method works.

D. Moore says, “There is a saying among statisticians that ‘mathematical theorems are true; statistical methods are effective when used with judgement ... Effective use of statistical methods requires more than knowing ... facts. It requires even more than understanding the underlying reasoning.” [24, p. 366].

Once you have passed the investigation stage and have a hypothesis which you wish to test, it is really worth finding a statistician and discussing your ideas with him or her. Failure to do that could mean that you waste a lot of time, and perhaps waste a lot of money too. Nevertheless, understanding the framework and terminology will help your communication.

Once you have collected your data you will use it in some way to test your hypothesis. You will usually calculate a single number, the “test statistic”, from your data such as the t-value, Spearman’s Rank Correlation Coefficient or Cronbach’s Alpha. The important thing is that, if the null hypothesis is correct, the behaviour of this number is known, at least approximately.

### 3.9 Significance level and p-values

How do we decide where to put the cut-off point for accepting or rejecting the null hypothesis? We decide on a definition of “unlikely”, which may be different in different situations, but the default value is 0.05 (5%); this is our significance level. We choose the critical value (the cut-off point) so that if the null hypothesis is true the probability of getting a value in the rejection range is at most 0.05.

The p-value is usually used by researchers because they don’t normally need to make a decision; instead they are presenting to readers the strength of their evidence. The p-value is usually produced by the software, such as SPSS. “Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g. the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.” [26, p. 131]. So if the p-value is less than 0.05 you would reject the null hypothesis with a significance level of 0.05.

Taking our definition of “significant” as  $p < 0.05$  means that if the null hypothesis is correct, so there is no change or no difference in the population(s), we should expect once in 20 surveys or experiments to get a “significant” result even though the null hypothesis is correct. Of course, when others try to reproduce the effect usually none is found.

In the same way, if a researcher effectively did the same survey or experiment 20 times when, in reality, the null hypothesis was correct, we would expect one of the outcomes to be “significant”. As above, this should cause problems with repeatability and/or reproducibility.

For many years, statisticians have been concerned about researchers effectively trying to summarise the results of their research in a single p-value. This discussion has become more focused in recent years.

In 2014, G. Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

- Q: Why do so many colleges and grad schools teach  $p = 0.05$ ?  
A: Because that’s still what the scientific community and journal editors use.  
Q: Why do so many people still use  $p = 0.05$ ?  
A: Because that’s what they were taught in college or grad school.

He summarised the position as: “We teach it because it’s what we do; we do it because it’s what we teach.” [26, p. 129].

The statistician and “Simply Statistics” blogger J. Leek wrote: “The problem is not that people use p-values poorly, it is that the vast majority of data analysis is not performed by people properly trained to perform data analysis.” [Leek, 2014 in 27, p. 129].

There are many problems with p-values, despite their widespread use. In 2015 the editors of Basic and Applied Social Psychology decided to ban the use of p-values by submitting authors [28].

Table 2 Right and wrong conclusions after testing hypotheses

	Null hypothesis true	Null hypothesis false
Accept the null hypothesis	Correct conclusion	Type II error
Reject the null hypothesis	Type I error	Correct conclusion

The ASA's p-value statement of 2016 includes the following principles [26]:

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

According to Wasserstein, executive director of ASA, "there is no single, perfect way to turn data into insight. The only surprise is that anyone believes there is! Science is complex. Inference is hard work. It has been extraordinarily costly to science that the shared understanding of generations of researchers has been that a p-value, or any other single index, could provide a simple, clear, objective answer to the question: What does this data tell us?" [27].

### 3.10 Ways to be wrong!

There are important consequences which follow from using this framework for statistical tests of hypotheses; one is that we could reach the wrong conclusion. The possibilities are shown in the Table 2.

Statisticians, and those using statistical inference, should not be arrogant people, because they always know that they can be wrong.

### 3.11 Conditions for validity

We are used to having conditions of validity in many areas of life. If you buy a second-class train ticket, it is not usually valid in a first-class carriage. Similarly, there are conditions for the validity of statistical tests.

One very common condition is independence. This is saying that the result of one experiment or survey interview should not affect others. Thus it would not be appropriate to give questionnaires to a group of friends and allow them to discuss

their responses; in fact, giving the questionnaires to a group might break the condition because they could be similar.

In many situations the requirement for a hypothesis test is that the normal/Gaussian distribution models the variable of interest in the population under investigation. In practice, nothing is normally distributed. Therefore our tests need to be "robust" to non-normality, and some are more robust than others. We should test our sample to see whether it is plausibly from a normal distribution. For example, if the sample is highly skewed and/or has extreme outliers, that would suggest that the population is not reasonably normally distributed.

If our data are values on a 5-point scale, such as a Likert-type scale, that is a long way from a normal distribution. However, the Central Limit Theorem says that if we are working with the (sum or) mean of values from almost any distribution, this will be approximately normally distributed provided that our sample size is reasonably large, and 50 is usually satisfactory. So, with samples of size 50 or more we can use "z-tests" instead of "t-tests" and not worry about not having an underlying normal distribution.

Other tests have different conditions and you need to check these out before collecting your data. There are lots of stories of people collecting their data and then going to the statistician to ask what they should do with it, and being told that they are unusable.

### 3.12 The need to publish

Apparently in the notebook of the scientist R. Millikan he refers to publishing good results. There is a tendency for researchers to want to find evidence to support their hypothesis. If they are using statistical methods, researchers want to get "significant results", low p-values, in order to get their research published.

If a researcher believes strongly that they are correct, but their results don't show that, they could feel pressure to modify or adapt their data to get a significant result. This could mean changing some values, omitting values, etc.

### 3.13 The ethics of reporting

While reporting the high-powered statistical methods which have been used in research, it may be tempting to omit reference to some of the more basic deficiencies. For example, are references to multivariate regression, Cronbach's alpha and

MANCOVA designed to hide the fact that the sample from which the data were obtained was selected by the researcher and was not representative of the population being investigated? When reporting the number of people who took part in a study, do we also report the number who refused to take part, or whose responses were rejected for some reason?

When it comes to reporting our conclusion, statisticians are usually quite careful and say that the data are consistent with the null hypothesis, or that the data suggest that the null hypothesis is false! This reflects the fact that we know we could be reaching the wrong conclusion.

Although it is very impressive to refer to multivariate regression, Cronbach's alpha, Mancova, etc. do we know what we are writing about? Do we have some understanding of what is going on, or are we just copying some words from a book?

It is much easier to refute someone else's ideas than to have your own. Similarly, it is often easy to criticise the research of someone else than to get it right yourself. The aim of these comments on statistics is to try to highlight some of the things which can go wrong when using statistics in the hope that the readers will be able to avoid some of the errors mentioned.

If you use statistics as a part of your research, please make statistics an integral part. Take the time to think about your project from a statistician's viewpoint; better still, find a friendly statistician - preferably one who already understands something of your research area - and discuss the topic with her/him. Share your expertise with others. Like any other tool, statistics will work

better for you if you follow the instructions, understand how it works, and keep your knowledge up-to-date!

The magazine "Significance," produced by the Royal Statistical Society and the American Statistical Association [29] presents recent research on a very broad range of topics in a style which is accessible to most people who have some understanding of statistics. This could be a useful and interesting way of developing your statistical knowledge and expertise.

#### 4. Conclusion

This article has been written in response to the frustrations felt by some researchers in the social sciences and humanities at having to use statistical methods when they felt it was unethical, inappropriate and reductionist. On the other hand, it reflects the frustrations felt by some statisticians at the abuses and oftentimes crude simplifications of the finely tuned statistical methods by some improperly and insufficiently prepared researchers. We have reflected on some of the differences in the subject, nature, aims, and methods of research in the social sciences and humanities. We have also considered some of the basic requirements of the statistical method and ways in which researchers sometimes ignore these, or are unaware of them. We have concluded that to ignore methodological conditions for the correct statistical procedure and its applicability in the area of research is unethical.

#### References

- [1] LIESNER, A.: Education or Service? Remarks on Teaching and Learning in the Entrepreneurial University. *Educational Philosophy and Theory*, 38(4), 483-495, 2006.
- [2] KASCAK, O., PUPALA, B.: Neoliberalism in Education: Five Pictures of Critical Analyses/Neoliberalizmus vo vzdelavani: pat obrazov kritických analyz (in Slovak). *Pedagogická Orientace*, 21(1), 5-34, 2011.
- [3] ALASUUTARI, P., BICKMAN, L., BRANNEN J. (Eds.): *The SAGE Handbook of Social Research Methods*. Sage, London, 2009.
- [4] GREEN, B.: Knowledge, the Future and Education(al) Research: A New-Millennial Challenge. *Australian Educational Researcher*, 37(4), 43-62, 2010.
- [5] MACBEATH, J.: *Future of Teaching Profession*. University of Cambridge, Cambridge, p. 111, 2012.
- [6] JEONG, D. W., LEE, H. J., et al.: Shaping Education Policy Research in an Asia-Pacific Context. *Asia Pacific Educational Review*, 15(3), 367-380, 2015.
- [7] TERHART, E., BENNEWITZ, H., ROTHLAND, M. (Eds.): *Handbook of Research on the Teaching Profession/Handbuch der Forschung zum Lehrerberuf* (in German), 2. edition. Waxmann, Munster, 2014.
- [8] E.g.: ZIEBERTZ, H.-G., STERKENS, C. (Eds.): *Religion and Civil Human Rights in Empirical Perspective*. Springer, Cham, 2017; ZIEBERTZ, H.-G. (Ed.): *Gender in Islam and Christianity. Theoretical and empirical studies/Gender in Islam und Christentum. Theoretische und empirische Studien* (in German). LIT, Munster, p. 304, 2010; ZIEBERTZ, H.-G., RIEGEL, U. (Eds.): *How Teachers in Europe Teach Religion. An International Empirical Study in 16 countries*. LIT, Munster, p. 405, 2009; ZIEBERTZ, H.-G., KAY, W., RIEGEL, U. (Eds.): *Youth in Europe. Vol. 3: An International Empirical Study of the Significance of Religion for Life Orientation*. LIT, Münster, p. 380, 2009.
- [9] HANZEL, I.: Qualitative or Quantitative Methods in Social Sciences?/Kvalitatívne alebo kvantitatívne metódy v sociálnych vedách? (in Slovak). *Filozofia*, 64(7), 646-657, 2009.

- [10] HANESOVA, D., NELSON, A., BADLEY, K.: Educators in Search for the Fine Line between Use and Misuse of New Technologies. *Communications - Scientific Letters of the University of Zilina*, 19(1), 44-48, 2017.
- [11] BOLTON, P.: Statistical Literacy Guide: How to Spot Spin and Inappropriate Use of Statistics [online]. Library of the House of Commons, 2010. Available: <http://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN04446> [accessed 2008-09-25].
- [12] WALI, FAYAZ, A. S., AMRAM, R.: Use of Structural Equation Modelling in Social Science Research. *Asian Social Science*, 11(4), 371-377, 2015. DOI: 10.5539/ass.v11n4p371.
- [13] HOGBEN, L.: *Mathematics for the Million*. W. Norton & co., New York, p. 649, 1968.
- [14] VALCO, M.: Combating Scientism: Lessons from David Bentley Hart and Charles Taylor. VALCO, M., KONYA, P. *Ethical Aspects of Contemporary Scientific Research*. Kud Apokalipsa & Ceri-Sk, Ljubljana, 14-35, 2017.
- [15] KANT, I.: *The Conflict of the Faculties*. Abaris Books, New York, p. 221, 1979.
- [16] HANES, P., DE MUYNCK, B.: Deification of Technology and the Dignity of the Human Person. *Communications - Scientific Letters of the University of Zilina*, 19(1), 121-125, 2017.
- [17] GRICE, H. P.: *Studies in the Way of Words*. Harvard University Press, Cambridge, p. 394, 1995.
- [18] HANES, P.: Revivalism and Social Ethics/Revivalizmus a socialna etika (in Slovak). PF UMB, Banska Bystrica, p. 125, 2002.
- [19] "Charity is forced on us; whether we like it or not, if we want to understand others, we must count them right in most matters." DAVIDSON, D.: *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford, 1991.
- [20] FELDMAN, R.: Charity, Principle of. CRAIG, E. (Ed.): *The Shorter Routledge Encyclopedia of Philosophy*. Routledge, London, p. 1077, 2005.
- [21] EU Charter of Fundamental Rights. Article 1 - Human Dignity [online]. European Union Agency for Fundamental Rights. Available: <http://fra.europa.eu/en/charterpedia/article/1-human-dignity> [accessed 2017-09-15].
- [22] This saying was used by Mark Twain in his article Chapters from My Autobiography [online]. *The North American Review*, 185(618), p. 471, 1906. It is not clear who its author is. Available: [https://www.jstor.org/stable/25105919?seq=6#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/25105919?seq=6#page_scan_tab_contents) [accessed 2017-09-10].
- [23] DAOUD, J., AZRAM, M.: Ethical Issues When Using Statistical Analysis in Research [online]. *Sci.Int(Lahore)*, 26(4), 1841-1846, 2014. Available: [https://www.researchgate.net/publication/277771534\\_ETHICAL\\_ISSUES\\_WHEN\\_USING\\_STATISTICAL\\_ANALYSIS\\_IN\\_RESEARCH](https://www.researchgate.net/publication/277771534_ETHICAL_ISSUES_WHEN_USING_STATISTICAL_ANALYSIS_IN_RESEARCH) [accessed 2017-09-02].
- [24] MOORE, D. S.: *The Basic Practice of Statistics*. W. H. Freeman and Comp., New York, 2003.
- [25] HUBER, M., FELFE, CH.: Does Preschool Boost the Development of Minority Children? The Case of Roma Children. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, 180(2), p. 24, 2015. DOI: 10.1111/rssa.12207.
- [26] WASSERSTEIN, R. L., LAZAR, N. L.: The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129-133, 2016. DOI: 10.1080/00031305.2016.1154108.
- [27] MATTHEWS, R., WASSERSTEIN, R. L., SPIEGELHALTER, D.: The ASA's p-Value Statement, One Year On [online]. *Significance*, 14(2), 38-41, 2017. Available: <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01021.x/full> [accessed 2017-09-15].
- [28] TRAFIMOW, D., MARKS, M.: Editorial [online]. *Basic and Applied Psychology*, 37(1), 2015. Available: <http://dx.doi.org/10.1080/01973533.2015.1012991> [accessed 2017-09-13].
- [29] Online. Available: [http://www.rss.org.uk/RSS/Publications/Significance/RSS/Publications/Significance\\_sub/Significance.aspx?hkey=b75bac40-bd86-4939-b3df-3c04d21481e4](http://www.rss.org.uk/RSS/Publications/Significance/RSS/Publications/Significance_sub/Significance.aspx?hkey=b75bac40-bd86-4939-b3df-3c04d21481e4).