How to do better? - Avoiding Statistical Errors in Published Research

Duc Anh Phi

WS 2019/20

Abstract

This report is about presenting the various options a researcher has to avoid statistical errors in his/her published research. Furthermore I am going to discuss some common errors in the use of statistical analysis based on "The Seven Deadly Sins of Statistical Analysis" [2].

Contents

Chapter 1

The need for doing better

There is an evident poor practice of statistical analysis in current published research. *"Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true"*, concluded Ioannidis in his controversial paper [4] in 2005, which marked the beginning of the current crisis. A year before, authors studied 181 papers from the Nature journal and 63 papers from the BMJ journal and found out that respectively 69 and 16 of these papers contain at least one statistical error. They concluded: *"This incongruence of test statistics and p-values is another example that statistical practice is generally poor, even in the most renowned scientific journals, and that the quality of papers should be more controlled and evaluated"* [6]. A recent study in 2016 [3] documents reporting errors in a sample of over 250 000 P values reported in eight major psychology journals from 1985-2013. The authors found that half of all published psychology papers that used Null hypothesis significance testing (NHST) contained at least one p-value that was inconsistent with its test statistics and degrees of freedom. One in eight papers contained a grossly inconsistent p-value that may have affected the statistical conclusion. This is just a small excerpt of papers, which show that there is a need for doing better.

Chapter 2

Your options as a researcher

You have basically three options as a researcher, which are explored in the following.

2.1 Become a statistician

"Demands placed on the modern scientist are extreme. Besides mastering their own rapidly advancing fields, most scientists are expected to be good a programming [...], designing statistical graphics, writing scientific papers, managing research groups, mentoring students, managing and archiving data, teaching, applying for grants, and peer-reviewing other scientists' work, along with the statistical skills [...]" [1]

Becoming an expert in your field is already challenging, additionally becoming an expert in another field, such as statistics is therefore unfeasible. Option 1 of becoming a statistician is the most difficult, thus the most unrealistic one.

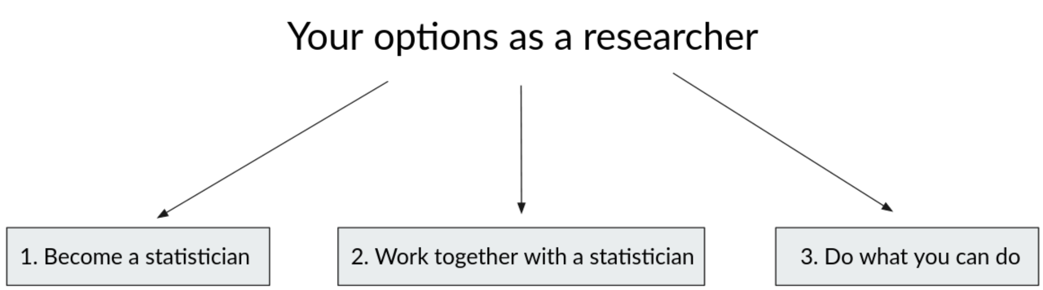

Figure 2.1: Your options as a researcher

2.2 Work together with a statistician

A competent statistician can recommend an experimental design that mitigates issues such as pseudo replication and helps you collect the right data - and the right quantity of data - to answer your research question. [1]

You should develop a long-term relationship with a statistician. Choosing the right statistician is analogous to choosing a lawyer, doctor or hair stylist [1]. Consult them prior to commencing experiments, because *"to consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of"*

- R.A. Fisher.

Additionally, it is critical that your statistician achieves an understanding of your experimental goals and of the technicals methods employed. The statistician need to acquire a working knowledge of the field of research the data addresses, because data cannot be interpreted in a vacuum. If statistical analysis is performed without an understanding of the underlying context, whatever "my statistician says" becomes completely irrelevant. [1]

Another thing to keep in mind is described in the 7th sin of the "Seven deadly sins of statistical analysis" [2]. The 7th sin is the failure to rely on a statistician of relying too much on a statistician.

"My statistician says..." - This statement is a double-edged sword. On the one hand it is a desirable maneuver to ask for assistance. However, it indicates that the researcher has little or no concept of the statistical methods being employed for the analysis of the data, preferring to abdicate all responsibility to a third party. Statistical analysis is used as means of evaluating research results and thereby used to validate important decisions. That is why it is a sin to simply "give data to the statistician" and then to get back the "results". A researcher who is seeking help, should become informed to the extent that he/she can actively participate in the interpretation of the data in a meaningful way. [2]

2.3 Do what you can do

"Beware of false confidence. You may soon develop a smug sense of satisfaction that your work does not screw up like everyone else's" [1]

If the previous options are out of reach, you are on your own. In that case you should follow the following steps:

- Read up on statistics. Take courses. Practice.
- Plan your data analysis carefully in advance.
- Identify established guidelines in your scientific field and adhere to them.
- To avoid a false-positive research result follow the proposed requirements by Simmons et. al. [5]

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	
Economic evaluations	CHEERS	

Figure 2.2: Guidelines in Health Research [7]

Table 2. Simple Solution to the Problem of False-Positive Publications

Requirements for authors

1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
3. Authors must list all variables collected in a study.
4. Authors must report all experimental conditions, including failed manipulations.
5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

Guidelines for reviewers

1. Reviewers should ensure that authors follow the requirements.
2. Reviewers should be more tolerant of imperfections in results.
3. Reviewers should require authors to demonstrate that their results do not hinge on arbitrary analytic decisions.
4. If justifications of data collection or analysis are not compelling, reviewers should require the authors to conduct an exact replication.

Figure 2.3: Requirements for Authors by Simmons et. al. [5]

Chapter 3

The seven deadly sins of statistical analysis

In this paper [2] the authors discuss some common errors in the use of statistical analysis that are regularly observed in professional surgical literature.

3.1 Sin 1: Using Parametric Analysis for Ordinal Data

To support your understanding, I need to define certain terms:

- In sampling theory a **parameter** is a variable that expresses some property of the entire population.
- Population mean, variance and standard deviation are the parameters most commonly used to describe a population.
- Sample mean, standard deviation, and variance are the corresponding descriptive statistics for a sample of data drawn from that population.

There are 4 **measurement scales**:

- **Nominal scales** simply categorize data without assigning any hierarchical order.
- **Ordinal scales** are used to rank data points hierarchically. The order, e.g. excellent ¿ good ¿ fair ¿ poor, is well defined, but the interval between each level is not certain.
- **Interval scales** have discrete defined levels and, in addition, the interval between each of the levels on the scale is well defined (and usually equal).
- **I**n a ratio scale, there is no restriction of a data point to a discrete level. Any value is permitted, including fractions. Ratio data have the additional requirement, that there must be a meaningful zero point, representing a complete lack of the characteristic.

Examples of Nonparametric Analog of Common Parametric Statistical Methods[a,b]

Type of Problem	*Type of Data*	*Parametric Methods*	*Nonparametric Methods*
Comparison of groups	One group (compared to a reference value)	z-test, t-test	Chi-squared test, Kolmogorov-Smirnov test
	Two independent groups	t-test, z-test, analysis of variance	Wilcoxon's signed rank test, median test, chi-squared test, Kolmogorov-Smirnov test, Mann-Whitney test
	Two paired or related groups	Paired t-test, z-test	Wilcoxon rank sum test, sign test
	Three or more groups	Analysis of variance, z-test	Kruskall-Wallis test, Friedman two-way analysis of variance by ranks
Association	One sample	Least-squares correlation analysis	Spearman rank correlation coefficient, Kendall's rank correlation coefficient (tau)
	More than one sample[b]	Regression analysis or logistical regression	Chi-squared test of independence

[a]Note that for each row, all the tests listed in the nonparametric column are similar in approach to all of those in
[b]Note that the chi-squared test can be applied to frequency data only. There is no direct nonparametric analog of least-squared regression analysis.

Figure 3.1: Examples of Nonparametric Methods

Multiplication and devision are used to compute the mean and variance. In order for these mathematical operations to be valid, the data must be expressed using an interval or a ratio scale. And here lies the sin: Simply expressing ordinal data using integers does not justify the use of parametric statistics. Instead you should use nonparametric statistical methods for nominal or ordinal scaled data. Note, that nonparametric methods can be used with ordinal data, do not require normally distributed data, and can be used with small sample sizes.

3.2 Sin 2: Inappropriate Use of Parametric Analysis

Before parametric analysis is appropriate certain sampling criteria must be met:

- **(1)** The study sample must be randomly drawn from a normally distributed population.
- **(2)** The sample size must be large enough to be "representative" of the study population.

Unless sufficient justification for the use of parametric analysis can be provided, nonparametric analysis should be employed. For most common parametric tests there is an equivalent nonparametric approach available (cf. Figure 3.1)

3.3 Sin 3: Failure to Consider Type II Statistical Error

Based on the sample size and differences between sample means, as well as the variances of these differences you can compute the probability beta of committing a type II statistical error (false negative), e.g. with a z-test (when comparing two groups). An acceptable beta is considered

	reject H_0	accept H_0
H_0 true	Type I error (false positive) Probability: α	Correct (true negative) Probability: 1- α
H_0 false	Correct (true positive) Probability: 1-β (Power)	Type II error (false negative) Probability: β

Figure 3.2: Types of Statistical Errors

at 20% or less. The sin of failing to report beta is serious, but the sin of failing to compute appropriate sample sizes based on a reasonable beta is fatal.

3.4 Sin 4: Using Unmodified t-Tests for Multiple Comparisons

When comparing more than 2 groups, whether they are statistically significantly different, t-tests are not appropriate. Instead you should use an analysis of variance **(ANOVA)** instead. ANOVA asks the question: "Is the variation within the dataset due to differences between groups greater than the variation due to differences within groups?". This determination is made by computing an F ratio, which is an expression of between-group variation divided by within-group variation. The probability associated with the F ratio can then be determined from standard F distributions.

ANOVA tests the null hypothesis, that all population means are equal. When the null hypothesis is rejected, at least one population mean is significantly different from at least another mean. However, ANOVA does not reveal which means are different from which. If the F ratio is associated with a probability of less than 0.05, then the null hypothesis is rejected.

3.5 Sin 5: Underutilization of Analysis of Covariance (ANCOVA), Multivariate Regression, Nonlinear Regression, and Logistic Regression

If confounding variables (covariates) could affect conclusions, **ANCOVA** is a useful technique. ANCOVA asks the question: "For our target dependent variables, is there a difference between groups if we adjust our data, taking into consideration differences between groups with regard to possible covariates?". However, when there is more than one important covariate that could affect a particular outcome, the use of a more complex regression analysis should be considered. By using a multivariate regression, the "significance" of each independent variable, in accounting for the variation of the outcome or dependent variable, could be tested. A limitation of the multivariate regression is, that the variable must be continuous. In order to consider the effect of independent categorical or non-continuous variables on a given dependent variable, logistical regression should be employed.

3.6 Sin 6: Reporting Standard Error instead of Standard Deviation

The standard error of the mean (SEM) is associated with the distribution of sample means that would be derived by repeatedly sampling n data elements from the study population. In other words, the SEM is a measure of the dispersion of sample means around the population mean. For reporting descriptive statistics and to indicate the spread of the sample data from the mean it is preferable to use the standard deviation over SEM. The practice of reporting SEM just because it ”looks better” is a statistical sin.

Chapter 4

Practice statistics responsibly

*"**Whenever we understand something that few others do, it is tempting to find every opportunity to prove it.** [...] Rather than taking the time to understand the interesting parts of scientific research, armchair statisticians snipe at news articles, using the vague description of the study regurgitated from some over enthusiastic university press release to criticize the statistical design of the research.[...] The first comments on a news article are always complaints about how they didnt control for this variable and the sample size is too small, and 9 times out of 10, the commenter never read the scientific paper to notice that their complaint was addressed in the third paragraph. This is stupid. A little knowledge of statistics is not an excuse to reject all of modern science. A research papers statistical methods can be judged only in detail and in context with the rest of its methods: study design, measurement techniques, cost constraints, and goals. Use your statistical knowledge to better understand the strengths, limitations, and potential biases of research, not to shoot down any paper that seems to misuse a p value or contradict your personal beliefs."* [1]

Chapter 5

Not all the truth lies in statistics

Conclusions supported by poor statistics can still be correct - statistical and logical errors do not make a conclusion wrong, but merely unsupported. [1] Even if statistical methods are employed and interpreted correctly, statistical analysis is still merely the computation of probabilities that will not overcome problems in methodology and may give a false sense of security. [2] Rather than interpreting statistical analysis as a ”final answer”, we should think of the result of statistical analysis as another piece of data that helps us decide whether our conceptualization is correct or incorrect. [2]

Chapter 6

Conclusion

We explored the various options you have as a researcher. Usually it is in your interest to work together with a statistician. However, you should still have a solid foundation in statistics to an extend, that you can participate in the interpretation of the research data in a meaningful way. It is crucial that your statistician has a working knowledge about your research. Additionally it is recommended that you follow established guidelines in your research area. To lower the risk of publishing a false-positive paper you should adhere to the requirements proposed by Simmons et. al. [5]. Finally, remember that not all truth lies in statistics.

Bibliography

[1] Alex Reinhart: Statistics Done Wrong - The Woefully Complete Guide

[2] Kuzon, William M.-Urbanchek, Melanie G.-McCabe, Steven: The Seven Deadly Sins of Statistical Analysis

In: Annals of Plastic Surgery 1996, 37:265-272

[3] Michle B. Nuijten, Chris H. J. Hartgerink, Marcel A. L. M. van Assen, Sacha Epskamp, Jelte M. Wicherts: The prevalence of statistical reporting errors in psychology (19852013)

In: Behaviour Research Methods 2016 Volume 48, Issue 4, pp 12051226

[4] John P. A. Ioannidis: Why Most Published Research Findings Are False

In: PLoS Medicine 2. e124 September 2005

[5] Simmons, Joseph & Nelson, Leif & Simonsohn, Uri. (2011). False-Positive Psychology. Psychological science. 22. 1359-66. 10.1177/0956797611417632.

[6] Garca-Berthou, E., Alcaraz, C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 4, 13 (2004). https://doi.org/10.1186/1471-2288-4-13

[7] Source: https://www.equator-network.org/; Access on Jan 27 2020 22:44

List of Figures