Topic 9: How to do better?

University Heidelberg
Seminar: How do I lie with statistics?
Supervisor: Prof. Dr. Ullrich Köthe
Duc Anh Phi
30th January 2020

structure

1. The need for doing better
2. What we’ve learned so far
3. Your options as a researcher
4. The seven deadly sins of statistical analysis
5. Practice statistics responsibly
6. Not all the truth lies in statistics
7. Conclusion
8. Sources

The need for doing better

The prevalence of statistical reporting errors in psychology (1985–2013)

Michèle B. Nuijten[1] · Chris H. J. Hartgerink[1] · Marcel A. L. M. van Assen[1] · Sacha Epskamp[2] · Jelte M. Wicherts[1]

Why Most Published Research Findings Are False

John P. A. Ioannidis

Incongruence between test statistics and *P* values in medical papers

Emili García-Berthou & Carles Alcaraz

What we've learned so far about statistics…

17. October	Ullrich Köthe: Introduction
	Peter Hügel: How to lie with statistics (Huff)
24. October	Florian Fallenbüchel: How to lie with charts (Jones)
	Marvin Ruder: The Visual Display of Quantitative Information (Tufte)
31. October	Julius Drück: The Demon-Haunted World (Sagan)
	Claire Zhao Sun: Intersubjective reality (Harari, Sapiens)
7. November	Valentin Wüst: Reasoning Fallacies
	Mustafa Ibrahim: Statistical Fallacies
14. November	Sophie Jost: Biases in Questionaires
	Lennart Stipulkowski: "Flexible" Data Collection
21. November	Meike Steinhilber: Reproducibility of Psychological Science
	Frederik Stegmüller: Shortcomings of p-values and multiple testing bias
28. November	**Climate Week Special**
	Oliver Mehling: Climate skepticism
	Peter Lippmann: Climate attribution science
5. December	Hannes Kepler: Pitfalls of counterfactual inference
12. December	Lasse Becker-Czarnetzki: Is the winner really the best?
19. December	Karl Thyssen: Debunking myths effectively
	Mihai Verzan: Learning to avoid cognitive biases
9. January	Pingchuan Ma: Methods of epidemiology
	Raphael Hirsch: How many deaths did the Chernobyl desaster cause?
16. January	Patrick Damman: The Book of Why 1
	Jasper Henze: The Book of Why 2
23. January	Marina Walther: Health Effects of Smoking
	Aysegül Peközsoy: Populism

What we’ve learned so far about statistics

- How to lie with charts
- Flexible Data Collection:
  - HARKing
  - P-Hacking
  - Researchers Degrees of Freedom
  - Shortcomings of p-values
  - Distortion in Graphics

What we've learned so far about statistics

Fallacies:

- Base-Rate/ Prosecutor's Fallacy
- Gambler's/ Hot-Hand Fallacy
- Spurious Precision
- Sampling
- Different definitions of measured quantities
- Counterfactual Inference
- Conjunction fallacy
- Correlation does not equal causation
- Absence of evidence is not evidence of absence

Biases:

- Publication Bias
- Selection Bias
- Hindsight Bias
- Question Design
- Questionnaire Design
- Questionnaire Administration
- Multiple testing bias
- Cognitive biases

How confident are you about your statistical abilities?

Your options as a researcher

1. Become a statistician

2. Work together with a statistician

3. Do what you can do

Option 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 at 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]

- most difficult and therefore most unrealistic option

Option 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]

- develop a long-term relationship with a statistician
- choosing the right statistician is analogous to choosing a lawyer, doctor or hair stylist
- Consult them prior to commencing experiments

“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. Fischer

Option 2: Work together with a statistician

- It is critical that they achieve an understanding of your experimental goals and of the technical methods employed
- The statistician need to acquire a working knowledge of the field of research the data addresses
- 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

Option 2: Work together with a statistician

Sin 7: Failure to Rely on a Statistician or Relying too much on a Statistician

- "My statistician says…" - This statement is a double-edged sword
- Asking for assistance is a desirable maneuver
- However, 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
- 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

Option 3: Do what you can do

“Beware of false confidence. You may soon develop a smug sense of satisfaction that your work doesn’t screw up like everyone else’s” [1]

- Read up on statistics. take courses. practice.
- Plan your data analysis carefully in advance
- Follow guidelines in your scientific field
- Follow requirements by Simmons et. al. 2011

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	

source: https://www.equator-network.org/
access on Jan 27 2020 22:44

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.

[5] Simmons et. al. (2011)

The Seven Deadly Sins of Statistical Analysis

Kuzon, William M.- Urbanchek, Melanie G.- McCabe, Steven (1996) *Annals of Plastic Surgery*, 37:265-272.[1]

- authors discuss some common errors in the use of statistical analysis that are regularly observed in professional surgical literature

Sin 1: Using Parametric Analysis for Ordinal Data

- 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

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).

In 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 complete lack of the characteristic.

Sin 1: Using Parametric Analysis for Ordinal Data

- Multiplication and division 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
- Simply expressing ordinal data using integers does not justify the use of parametric statistics
- Use nonparametric statistical methods for nominal or ordinal scaled data

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 use of parametric analysis can be provided, non-parametric analysis should be employed
- for most of common parametric tests an equivalent nonparametric approach is available

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.

nonparametric methods can be used with ordinal data, do not require normally distributed data, and can be used with small sample sizes

Sin 3: Failure to Consider Type II Statistical Error

	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

Source: https://www.statisticssolutions.com/wp-content/uploads/2017/12/rachnovblog.jpg
Access on January 29 2020

Sin 3: Failure to Consider Type II Statistical Error

- acceptable β = 0.2
- the sin of failing to report β is serious
- the sin of failing to compute sample sizes based on reasonable β is fatal

$$Z_\beta = \sqrt{\frac{ND^2}{\sigma^2}} - Z_{\alpha/2}$$

where:

Z_β = standart normal value associated with β

$Z_{\alpha/2}$ = standart normal value associated with α

N = sample size

D = difference between the two sample means $(\mu_1 - \mu_2)$

and σ^2 = the variance of the differences between the sample means.

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
- Use 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

Sin 4: Using Unmodified t-Tests for Multiple Comparisons

- 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 an other mean
- However, ANOVA does not reveal which means are different from which
- If the F ratio is associated with a probability less than 0.05 then the null hypothesis is rejected

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?
- When there is more than one important covariate that could affect a particular outcome, the use of more complex regression analysis should be considered

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

- Using multivariate regression, the “significance” of each independent variables in accounting for the variation of the outcome or dependent variable could be tested
- A limitation of 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

Sin 6: Reporting Standard Error Instead of Standard Deviation

The standard error of the mean (SEM) can be expressed as:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where

σ is the standard deviation of the population.

n is the size (number of observations) of the sample.

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Sin 6: Reporting Standard Error Instead of Standard Deviation

- SEM is the standard deviation 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
- Standard deviation preferable when reporting descriptive statistics, indicating the spread of the sample data
- The practice of reporting standard error because it “looks better” is a statistical sin

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 didn’t 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 paper’s 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]

Not all the truth lies in statistics

- Remember: 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]

Conclusion

- We explored the various options you have as a researcher
- Usually it is in your interest to work together with a statistician
- You should still have a solid foundation in statistics to an extend that you can participate in the interpretation of the data in a meaningful way
- Your statistician should have working knowledge about your research
- Follow guidelines!
- Follow rules by Simmons et. al. 2011!
- Not all truth lies in statistics

literature

[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, 1996 in Annals of Plastic Surgery, 37:265-272

[3] Michèle 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 (1985–2013) 2016 in Behaviour Research Methods Volume 48, Issue 4, pp 1205–1226

[4] John P. A. Ioannidis: Why Most Published Research Findings Are False in PLoS Medicine 2. e124 September 2005

literature

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

[6] García-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

Any Questions?