Scuola estiva AIP di Metodologia 2023

Intelligenza Artificiale per la psicometria: Metodi e applicazioni

Seminario sull'Etica della Ricerca

Centro Residenziale Universitario di Bertinoro

01/09/2023

Replicabilità e riproducibilità della ricerca

Gian Mauro Manzoni

Membro della Commissione Etica dell'AIP

Sezione Psicologia Clinica e Dinamica

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Med 2(8): e124.

"There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and ..".

"…, a research finding is less likely to be true when the **studies are smaller**; when **effect sizes are smaller**; when there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is a **greater number and lesser preselection of tested relationships**; when there is **greater financial and other interest**; …

Estimating the reproducibility of psychological science

Open Science Collaboration*

Science, vol 349, issue 6251

"Only 36% of 100 associations that resulted statistically significant in previous studies were statistically significant in the new studies".

"The replication success was positively related to the size of the original effect and .. the effect sizes were about half the magnitude of those reported in the original articles".

Questionable research practices (QRPs)

Table 1. Questionable Research Practices (QRPs) and self-admission rates in percentages for US [25] and Italian psychologists.

	US		Italian Association of Psychology	
QRP	Self-admission rate (*N*)	95% CI	Self-admission rate (*N*)	95% CI
1. In a paper, failing to report all of a study's dependent measures	63.4 (486)	59.1–67.7	47.9 (219)	41.3–54.6
2. Deciding whether to collect more data after looking to see whether the results were significant	55.9 (490)	51.5–60.3	53.2 (222)	46.6–59.7
3. In a paper, failing to report all of a study's conditions	27.7 (484)	23.7–31.7	16.4 (219)	11.5–21.4
4. Stopping collecting data earlier than planned because one found the result that one had been looking for	15.6 (499)	12.4–18.8	10.4 (221)	6.4–14.4
5. In a paper, "rounding off" a *p* value (e.g., reporting that a *p* value of .054 is less than .05)	22.0 (499)	18.4–25.7	22.2 (221)	16.7–27.7
6. In a paper, selectively reporting studies that "worked"	45.8 (485)	41.3–50.2	40.1 (217)	33.6–46.6
7. Deciding whether to exclude data after looking at the impact of doing so on the results	38.2 (484)	33.9–42.6	39.7 (219)	33.3–46.2
8. In a paper, reporting an unexpected finding as having been predicted from the start	27.0 (489)	23.1–30.9	37.4 (219)	31.0–43.9
9. In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)	3.0 (499)	1.5–4.5	3.1 (223)	0.9–5.4
10. Falsifying data	0.6 (495)	0.0–1.3	2.3 (220)	0.3–4.2

Agnoli F, Wicherts JM, Veldkamp CLS, Albiero P, Cubelli R (2017)

Questionable research practices among italian research psychologists. PLoS ONE 12(3)

La replicabilità

Dr. Ulrich Schimmack Blogs about Replicability

https://replicationindex.com/

DEFINITION OF REPLICABILITY:

In empirical studies with sampling error, replicability refers to the probability of a study with a significant result to produce a significant result again in an exact replication study of the first study using the same sample size and significance criterion (Schimmack, 2017).

La replicabilità

Brian A. Nosek et al. (2022). *Annual Review of Psychology*

Replicability, Robustness and Reproducibility in Psychological Science

DEFINITION OF REPLICABILITY:

Replicability refers to testing the reliability of a prior finding using different data and the same analysis strategy (Natl. Acad. Sci. Eng. Med. 2019*).

** Reproducibility and Replicability in Science. Washington, DC: Natl. Acad. Press*

La replicabilità

FURTHER DEFINITIONS OF REPLICABILITY:

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

An attempt by a second researcher to replicate a previous study is an effort to determine whether applying the same methods to the same scientific question produces similar results.

La riproducibilità

Brian A. Nosek et al. (2022). *Annual Review of Psychology*

Replicability, Robustness and Reproducibility in Psychological Science

DEFINITION OF REPRODUCIBILITY:

Reproducibility refers to testing the reliability of a prior finding using the same data and the same analysis strategy (Natl. Acad. Sci. Eng. Med. 2019*).

** Reproducibility and Replicability in Science. Washington, DC: Natl. Acad. Press*

La riproducibilità

OPERATIONAL DEFINITION OF REPRODUCIBILITY:

If someone applies the same analysis to the same data, the same result should occur. Reproducibility tests can fail for two reasons:

1) Process reproducibility: when the original analysis cannot be repeated because of the unavailability of data, code, information needed to recreate the code, or necessary software or tools.
2) Outcome reproducibility: when the reanalysis obtains a different result than the one reported originally. This can occur because of an error in either the original or the reproduction study.

La riproducibilità

Whereas an outcome reproducibility failure suggests that the original result may be wrong, a process reproducibility failure merely indicates that the original result cannot be verified.

Either reason challenges credibility and increases uncertainty about the value of investing additional resources to replicate or extend the findings.

Sharing data and code reduces process reproducibility failures, which can reveal more outcome reproducibility failures.

La riproducibilità

RESEARCH EVIDENCE

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. Psychological Methods, 26(5), 527–546

Authors successfully reproduced just 70% of the 232 findings analyzed, and 18 of those were reproduced only after deviating from the analysis reported in the original papers.