

## FORM 2A- RESEARCH MASTER'S PSYCHOLOGY: RESEARCH INTERNSHIP RESEARCH PROPOSAL

### 1. GENERAL INFORMATION

#### 1.1 Student information

Student name: Arne John

Student ID card number: 13580329

Address: Cornelis van Rijplantsoen 75

Postal code and residence: 1063MD Amsterdam

Telephone number: +49 176 30793220

Email address: arne.john@student.uva.nl

#### 1.2 Supervisor information

Supervisor (*eligible for the ResMas*): Prof. Dr. Eric-Jan Wagenmakers

Note. Eligible supervisors are staff members of the Department of Psychology with a PhD-degree and appointed as examiner by the Examinations Board of the Department of Psychology.

Specialization: Psychological Methods and Statistics

#### 1.3 Other information

Date: 01.02.2022–31.08.2023

Number of ECs for the research internship: 24

Note. 1EC = 28 hours. Minimally 18EC and maximally 24EC. The total number of EC for the Internship and Thesis should be 50EC.

Ethics Review Board (ERB) code: \_\_\_\_\_

Note. See <https://www.lab.uva.nl/lab/ethics/>

*Have your research proposal signed by your supervisor (see Section 9) and submit the signed research proposal and the two peer reviews via CANVAS. Please inform by email your supervisor and the secretariat of the research master psychology ([thesis-researchmaster-psy-fmg@uva.nl](mailto:thesis-researchmaster-psy-fmg@uva.nl)) that you submitted your research proposal.*

## 2. TITLE AND SUMMARY OF THE RESEARCH PROJECT

### 2.1 Title

Auto-Stats – Continuous Text Output Explaining Statistics

### 2.2 Summary of proposal – max 150 words

Interpreting statistical output can be difficult and easily wrought with inaccuracies due to a lack of understanding for the nuances in the tools used. While statistical programs like JASP reduce the avenues for failure by allowing users to easily visualize their data and test for statistical assumptions, the issue of falsely drawn conclusions from the numeric results is not resolved. This internship proposes an automatically generated report based on the choices the user makes, with differing levels of verbosity to help users both easily report their findings and avoid common misconceptions of the statistical method used.

Word count = 96

---

## 3. PROJECT DESCRIPTION – max 1200 words (This is a strict limit).

Note: Since this project is different from the conventional format, the structure of the proposal has been adjusted accordingly.

Statistical analysis can be a daunting task. With math being the least favorite school subject (EducationQuizzes, 2020), and the reason many US college students do not graduate (<https://newsroom.ucla.edu/stories/why-so-many-u-s-students-arent-learning-math>, 2018), it stands to reason that mistakes are made in the correct interpretation of results.

However, poor understanding of statistical concepts are not limited to people with little interest or experience in mathematics. Many undergraduate and graduate students as well as academic researchers struggle with correct interpretation of Null-Hypothesis Rejection Testing and p-values (Falk & Greenbaum, 1995; Gigerenzer, 2004; Haller & Krauss, 2002; Oakes, 1986). This is troubling, given how many papers still rely on the p-value to report their findings (e.g., see Stephens et al., 2005 for an ongoing debate within ecology). As pointed out by Goodman (2008), the most common misconception about p-values is “If  $P < .05$ , the null hypothesis has only a 5% chance of being true.” It stems from a misunderstanding of where the p-value originates from; Since p-values derive from the assumption of the null-hypothesis being true, they cannot also give information about the null-hypothesis being false.

The addition of confidence intervals by the APA in the fifth and sixth edition to improve accessibility of statistics seems to have had only mild results. As studies show, confidence intervals, too, are often misunderstood and wrong conclusions are drawn from them (Hoekstra et al., 2014; Morey et al., 2016). This is because p-values, confidence intervals, and differences in Akaike's information criterion (AIC) are all linked (Murtaugh, 2014). Changing the value reported will not automatically change people's understanding of the statistical tool as a whole.

Assuming researchers do not want to switch to a Bayesian model of statistics, the easiest way to ensure correct interpretation of statistical results is to give a written report stating what one can and cannot conclude. This way, the researcher can build their reporting around correct sentences rather than relying on their memory and inferred understanding.

JASP is an open-source statistical program that combines the benefits of up-to-date developments

of statistical methods with the user-friendliness of a program with a GUI. It allows the user to conduct both frequentist and Bayesian statistics. It reduces avenues of failure by easily allowing the user to plot their data and visually discover problems, thereby avoiding misinterpretations by only looking at the descriptive statistics as can be seen in Anscombe's quartet (Anscombe, 1973)

Sometimes when people report statistics, they do it incompletely. People don't plot their data and are not using safety nets to check whether the analysis fits the data.

This project intends to expand the safety net JASP provides users when interpreting their data. It does this by taking the statistical output from JASP and creating a report about the analysis results in text form. It is intended to have three levels of verbosity. On the simplest level, it just states the results with its assumptions, and a link of the literature. On a medium level, the report also includes a simple summary of the chosen test and its statistical assumptions that the data needs to meet. On the highest level, the text also includes common misconceptions and the actual facts about hypothesis testing and parameter interpretation.

Word count = 549

---

#### 4. PROCEDURE – With a maximum of about 1000 words

##### 4.1 Operationalization

*Describe how the research questions are operationalized.*

- a) Operationalize the research questions in a clear manner into a research design/strategy.
- b) Describe the procedures for conducting the research and collecting the data.
- c) For methodological and/or simulation projects describe the design of the simulation study.

The resulting report will be done in R (R Core Team, 2023), with all the packages that the t-test module in JASP requires. The report will consist of a series of decision trees depending on the user's choices and statistical results. Specifically, the data set provided by the user will be tested for outliers, skewness, normality, and homoscedasticity, after which the parameter estimation and significance affect the returned output text.

Additionally, the user can opt into a more verbose report, which includes common misconceptions and false interpretations of t-tests, as well as a simple summary to refresh the user's memory of the nuances of a t-test.

The texts will go through two rounds of review. The first round will be feedback by the supervisor to refine the statements for each potential case; The second round will be a peer-review by other methodology department associates as well as outsiders to check the validity and potential generality of the statements. Once the review stages are complete, this auto-stat report will be implemented into JASP.

##### 4.2 Sample characteristics

- a) Indicate, given a power analysis, how many participants will be recruited. Also motivate whether the resulting number is feasible.
- b) If a subset of participants will be excluded from the analysis given their scores on dependent variables, indicate the objective criterion to do so. For example include a phrase like: "Scores on dependent variables exceeding  $\pm 3$  SD of the mean will be excluded from the analysis".
- c) If a subset of participants will be excluded from the analysis given their scores on a manipulation check item, indicate the objective criterion to do so. For example include a phrase like: "Participants scoring 15 or lower on a manipulation checks

- item, will be excluded from the analysis”.
- d) In case of a simulation study, indicate how data will be generated.

Does not apply.

### 4.3 Materials

Indicate which tests, stimuli, equipment, etc. will be used; provide sufficiently elaborate descriptions and motivate your choice. (Always report the psychometric characteristics, such as reliability and validity, if existing tests are used. If new or adapted instruments or test materials (e.g., questionnaires) will be developed, then the new instrument must be independently validated first; only then it can be used as a testing instrument. Exception to this rule is allowed in case of questionnaires that do not contain more than one question (e.g., indicate on a 5-point scale how you feel today). In case of a simulation study 4.3 might be omitted.

This internship will be focus on the Student's t-test as a trial case to see whether this report works as intended. Following this, the report will be generalized for paired samples and one-sample t-tests, as well as non-parametric versions.

### 4.4 Data analysis

Describe the data preprocessing. Indicate for each research question *separately*, how it is translated into a statistical prediction. For example: “In a repeated measures ANOVA we expect an interaction effect of the between factor x and the within factor y on the dependent variable z. Also indicate how you will correct for multiple comparisons. Only the analyses proposed here can be described as confirmatory analyses in your research report. All other have to be mentioned as exploratory.

As this internship focuses on developing a report from user data, no actual data analysis will be done. Nonetheless, the report will test for assumptions and their violations before conducting a student t-test.

Word count = 242

---

## 5. INTENDED RESULTS - max 250 words

Clarify what the implication of possible outcomes would be (per hypothesis) for the specific and general research questions as well as for the theory. Address the following in approximately 250 words:

- What are the interpretations if the results do match the expectations?
- What are the interpretations if the results do not match the expectations?
- Are there any alternative interpretations?
- Is there any practical or societal relevance? Please explain.

The goal is to implement a report that takes the results of the statistical analysis and inserts these into a text that can easily be read. This feature aims to improve the usability and simplicity of JASP by reducing the amount of errors users may accidentally make when interpreting their results. Additionally, the report tests statistical assumptions and redirects users to alternative tests if violated. The final project encompasses the student t-test, but the framework should allow easily adaptation for other statistical tests as well.

Word count = 85

## 6. WORK PLAN – max 500 words

*Describe how the research project will be executed. Who is doing what and when? Is the planning of the current project realistic, efficient and feasible?*

### 6.1 Time schedule

State the total amount of EC as noted in the internship contract (18-24 EC), 1EC stands for 28 hours work. Present and justify a time schedule in weeks, including your time investment in hours per week. Plan some spare time, and indicate what elements can be cut / reduced if necessary. Provide the intended presentation date.

February '22: 19h/week	Learning R, creating Rmd file to mimic JASP
March '22: 19h/week	Working on Rmd file, writing proposal
April '22: 20h/week	Writing proposal, adapting non-parametric tests
May' 22: 39h/week	Writing conditions, dividing texts across verbosity, BUFFER FOR PROBLEMS
June' 23: 19h/week	Re-starting project, adapting existing code, improve assumption tests
July' 23: 19h/week	Rewriting proposal, changing format to new texts
August '23: 19h/week	Introducing it into JASP, creating template for other modules
October '23: 19h/week	Sending off proposal, meet with programmer for full integration, writing of internship report

### 6.2 Infrastructure

Where will the research take place? How is access to the facilities and materials ensured?

As this internship is completely digital, there is no need for facilities. If I want to study at the building, I can reach out to the staff to let me study in EJ's room.

### 6.3 Data storage

Each researcher needs to comply with the storage protocol of the Research Institute Psychology: <http://psyres.uva.nl/content/scientific-integrity-docs/data-protocol.html>

Since the only data "collected" is the description of methodological results from experts, there is no need for data storage.

### 6.4 Budget

A € 25 budget may be used for printing costs (e.g. for the conference poster). Please go to the secretariat of the specialization of your supervisor with your receipts.

Due to unforeseeable events, the time schedule and work plan was delayed and thereby stretched longer than originally intended.

Word count = 139



---

## 7. REFERENCES

- Anscombe, F. J. (1973). Graphs in Statistical Analysis, *The American Statistician*, 27(1), 17-21, doi: 10.1080/00031305.1973.10478966
- Falk, R., & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, 5 (1), 75-98. <https://doi.org/10.1177/0959354395051004>
- Fidler, F., & Loftus, G. R. (2009). Why Figures with Error Bars Should Replace p Values: Some Conceptual Arguments and Empirical Demonstrations. *Zeitschrift für Psychologie / Journal of Psychology*, 217 (1), 27-37. <https://doi.org/10.1027/0044-3409.217.1.27>
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33 (5), 587-606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, 45 (3), 135-140. <https://doi.org/10.1053/j.seminhematol.2008.04.003>
- Haller, H., & Krauss, S. (2002). Misinterpretations of Significance: A Problem Students Share with Their Teachers? *Methods of Psychological Research Online*, 7 (1).
- Held, L. (2013). Reverse-Bayes analysis of two common misinterpretations of significance tests. *Clinical Trials (London, England)*, 10 (2), 236-242. <https://doi.org/10.1177/1740774512468807>
- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21 (5), 1157-1164. <https://doi.org/10.3758/s13423-013-0572-3>
- Knottnerus, J. A., & Bouter, L. M. (2001). The ethics of sample size: Two-sided testing and one-sided thinking. *Journal of clinical epidemiology*, 54 (2), 109-110. [https://doi.org/10.1016/s0895-4356\(00\)00276-6](https://doi.org/10.1016/s0895-4356(00)00276-6)
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23 (1), 103-123. <https://doi.org/10.3758/s13423-015-0947-8>
- Muff, S., Nilsen, E. B., O'Hara, R. B., & Nater, C. R. (2022). Rewriting results sections in the language of evidence. *Trends in Ecology & Evolution*, 37 (3), 203-210. <https://doi.org/10.1016/j.tree.2021.10.009>
- Murtaugh, P. A. (2014). In defense of P values. *Ecology*, 95 (3), 611-617. <https://doi.org/10.1890/13-0590.1>
- Oakes, M. W. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Wiley.
- Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm* (1st ed). Chapman & Hall.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1 (2), 115-129. <https://doi.org/10.1037/1082-989X.1.2.115>
- Stephens, P. A., Buskirk, S. W., Hayward, G. D., & Martinez Del Rio, C. (2005). Information theory and hypothesis testing: A call for pluralism. *Journal of Applied Ecology*, 42 (1), 4-12. <https://doi.org/10.1111/j.1365-2664.2005.01002.x>
- Tendeiro, J. N., & Kiers, H. A. L. (2019). A review of issues about null hypothesis Bayesian testing. *Psychological Methods*, 24 (6), 774-795. <https://doi.org/10.1037/met0000221>
- van Ravenzwaaij, D., & Wagenmakers, E.-J. (2021). Advantages masquerading

as "issues" in Bayesian hypothesis testing: A commentary on Tendeiro and Kiers (2019). *Psychological Methods*. <https://doi.org/10.1037/met0000415>

---

## 8. FURTHER STEPS

Make sure your supervisor submits an Ethics Checklist for your intended research to the Ethics Review Board of the Department of Psychology at <https://www.lab.uva.nl/lab/ethics/>

---

## 9. SIGNATURES

- ☐ I hereby declare that both this proposal, and its resulting internship, will only contain original material and is free of plagiarism (cf. Teaching and Examination Regulation in the research master's course catalogue).
- ☐ I hereby declare that the results section of the internship report will consist of two subsections, one entitled "confirmatory analyses" and one entitled "exploratory analyses" (one of the two subsections may be empty):
  - a) The confirmatory analysis section reports *exactly* the analyses proposed in Section 4 of this proposal
  - b) The exploratory analysis section contains not previously specified, and thus exploratory, analyses.

Location:

Student's signature:

Supervisor's signature:

.....