**Assignment 1: Data Management and Power Analysis**

Colby Price

Department of Psychiatry; Dalhousie University

PSY-6003: Fundamentals of Applied Statistics and Research Design

Dr. Igor Yakovenko

March 7, 2024

**Assignment 1: Data Management and Power Analysis**

In the following study, a total sample of 812 superheroes of *the Avengers* and their respective attributes were collected during their final battle with Thanos across two separate battlefields (North and South). A total of 12 numerical variables were measured and recorded, including each participant’s IQ, agility, speed, strength, damage resistance, flexibility, willpower rating, PTSD score, kill count, injuries, minutes spent fighting, and shots they fired. Furthermore, three categorical variables were gathered, including if the hero had superpowers, if they battled in the Northern or Southern battlefield, and if they died during the battle. Incomplete cases or cases with non-real answers were removed from our data set. From the collected data, researchers created a variable for overall effectiveness in combat (called “combat effectiveness”), which was calculated as the sum of the hero’s agility, speed, strength, and willpower scores. Management and analysis of data was done through R version 4.3.2 (2023) using “tidyverse” packages (Wickham et al., 2019).

Descriptive statistics, including means, standard deviation, and range values, for combat effectiveness, kills, and injuries were obtained for each battlefield and are presented in Table 1. Average combat effectiveness was M = 610.74 (SD = 156.14) and M = 619.30 (SD = 165.35) for the northern and southern fields respectively, showing that the southern battlefield was only slightly more effective in combat as compared to the northern battlefield. Average mean injuries were slightly higher in the northern battlefield (M = 3.80, SD = 1.20) as compared to the southern battlefield (M = 3.18, SD = 1.25). The most erroneous variable calculated was the average kills with standard deviations for the northern, southern, and combined samples being much higher than their means as shown in Table 1. The range for kills is quite high (R = (0, 176)), with many Avengers also having zero kills. As a result, the mean is a problematic way to view the average kills for each Avenger and more suitable to be calculated by mode.

A power analysis was conducted to estimate the sample size for a secondary study to denote how the presence of superpowers impact a hero’s IQ. Our research team hypothesizes that heroes with superpowers will have higher average IQs than those without. To determine if our study would have sufficient power, an independent sample t-test was run using the smallest effect size of interest to our study (d = 0.2). According to similar studies, we expect small effect sizes (d = .2-.4), making this a relevant effect size to view within our sample as well. By conducting a priori power analysis with the smallest effect size of interest, our research team should be able to capture much larger effects if they exist in our sample in addition to small effect sizes. This rationale, while being a heuristic approach, should prove effective in determining the exact power of our sample.

To determine the necessary sample size for a sufficiently powered study, our research team conducted a priori power analysis using a two-sided two-sample t-test using the “pwr” package (Champley, 2020). Using small effect sizes of d = .2, results indicated that the required sample size to achieve 80% power at a significance level of α = .05 was N = 394 per group, or a total sample size of N = 798. Thus, the sample size of N = 812 in our dataset is adequate to test our proposed study’s hypothesis. Confidence interval and significance level criteria were based on standard practices in psychology research, with confidence intervals of 80-90% being the most acceptable interval ranges in relationship studies and p values of p=.05 being standard practice in proving significance. The type of test run was based on our hypothesis in testing two separate groups (i.e., superpowered v. non-superpowered) and notifying either a positive or negative relationship. Our research team chose to use a two-sided t-test as there is a lack of a definitive relationship between IQ and the presence of a superpower, and while we suspect to see a positive relationship between superpowers and IQ, we wish to be powered to denote either a positive or negative relationship.

In addition to a priori power analysis, an equivalence test was conducted to explore if our experiment would be sufficiently powered to prove no relationship existed between superpowers and average IQ. An equivalence test is done by running a two one-sided t-tests with an upper and lower bound of our effect size (d=.2), with significance showing that our results would statistically equal zero or prove no relationship between the variables. Otherwise, insignificant values would just be inconclusive to prove the presence of no relationships. A power analysis for equivalence testing was ran using the “TOSTER” package (Caldwell, 2022; Lakens, 2017) using the same parameters as our power analysis. Results informed that a total sample size of N = 858 (per group, N = 429) is required to prove zero relationship between variables. As such, our proposed study is not sufficiently powered to prove zero relationship according to our sample size.

After conducting our analysis, our independent t-test was t = 4.25. Our research team decided to conduct an analysis of our effect size using the “effectsize” package (Ben-Shacar et al., 2020). Using a test statistic of t = 4.25 and a sample of N = 812, effect size was estimated at d = .3 at a 95% confidence interval. According to qualitative labels by Cohen’s conventions, our effect size is estimated to be within small to medium effect sizes, or between d = .2 and d = .5 respectively. As our research team conducted the test at a 95% confidence interval, our measure is precise as compared to similar studies between IQ and superpowers. A 95% confidence interval means that by retesting our sample interval, 95% of our similarly powered tests will contain the true effect size value.

**References**

Ben-Shachar M, Lüdecke D, Makowski D (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. Journal of Open Source Software, 5(56), 2815. doi: 10.21105/joss.02815

Caldwell, A.R. (2022). Exploring Equivalence Testing with the Updated TOSTER R Package. PsyArXiv. https://doi.org/10.31234/osf.io/ty8de

Champely S (2020). \_pwr: Basic Functions for Power Analysis\_. R package version 1.3-0, <https://CRAN.R-project.org/package=pwr>.

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses.Social Psychological and Personality Science, 8(4), 355-362. doi:10.1177/1948550617697177

R Core Team (2023). \_R: A Language and Environment for Statistical Computing\_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” \_Journal of Open Source Software\_, \*4\*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.

**Table 1**

*Average Combat Effectiveness, Kills, and Injuries Grouped by Battlefield*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Battlefield | Combat Effectiveness | | | Kills | | | | Injuries | | | | |
| M | SD | Range | | M | SD | Range | | M | SD | Range |
| North  South  Overall | 610.74  619.30  615.02 | 156.14  165.35  160.77 | (131, 991)  (67, 1588)  (67, 1588) | | 3.23  4.43  3.83 | 11.85  9.59  10.79 | (0, 176)  (0, 79)  (0, 176) | | 3.80  3.18  3.49 | 1.20  1.25  1.26 | (0, 5)  (0, 5)  (0, 5) |