**PSYR 6003: Assignment 1 – Data Management and Sample Size Estimation**

Julia Fraiha-Pegado

Department of Psychiatry, Dalhousie University

PSYR 6003: Fundamentals of Applied Statistics and Research Design

Dr. Igor Yakovenko

March 7th, 2024

**Assignment 1 – Data Management and Sample Size Estimation**

The present data analysis was based on the “Avengers” fictional data set created and published by Dr. Dustin Fife. This file is composed of data captured during the Avengers’ final battle against Thanos, including 15 variables describing aspects of the combat and individuals’ performances (e.g. injuries, shots taken, flexibility, etc.). This data set was cloned by Dr. Igor Yakovenko into his repository. Intrigued by the topic, we cloned it to conduct an exploratory analysis investigating the different outcomes of the battle and features of the fight.

**Cloning Procedure**

To clone this repository its URL was copied, which was then pasted into the terminal, using the code “git clone”. This populated a new folder on the desktop containing all the files from the repository.

**Data Initialization & Cleaning**

The data set was then loaded into the software package R (Version 4.2.3) running on the integrated environment RStudio (Version 2023.12.1.402). After the loading process, the data set was put into an object. Before starting we visually inspected the data. The data analysis was conducted and managed using “tidyverse” packages (Wickham et al., 2019).

The original data set contained 814 individuals. These individuals lacked “ID” rows; to facilitate identifying each Avenger, we assigned them IDs. Visual inspection also revealed that two rows had missing values (NAs), which we decided to omit, only keeping the complete case as part of the data set to ensure the accuracy and reliability of our results. This resulted in our sample decreasing by 2 (n=812).

Interested in the combat effectiveness of each Avenger, we decided to create a new variable, combining “agility”, “speed”, “strength” and “willpower”. We mutated the data using the package “dplyr” (Wickham et al., 2019). We decided to create copies of the new clean data set in both SPSS and CSV formats so that they would be available in such formats. The base R package “utils”. data was used to save it in CSV format. To do so in an SPSS format, we utilized the “haven” package (Wickham et al., 2019).

**Descriptive Statistics**

To determine how Avengers without any superpowers and who died in combat performed, we filtered the data set using the “dplyr” package (Wickham et al., 2019). We decided to look at the properties of combat effectiveness, kills and injuries to determine how individuals performed. We ran a summary analysis using the “dplyr” package (Wickham et al., 2019) to gather the results.

We obtained mean (M), standard deviation (SD) and range values for the variables combat effectiveness (M= 497.5, SD= 177.5, range= 67.3 – 946.9), number of kills (M= 3, SD= 8,8, range= 0 – 70), and injuries sustained (M= 5, SD= 0.7, range= 2 – 5).

To determine which battlefield was the most effective in combat, they were grouped by location (“North” or “South”) to see how they performed. The North battlefield had the highest mean (M= 499.8) and lowest standard deviation (SD= 174.1) when compared to the South (M= 491.7, SD= 189.5). Despite its effectiveness, the Northern battlefield was also the one accounting for the highest number of injuries sustained (M= 5, SD= 0.7, range= 2 - 5) when compared to the Southern location (M= 5, SD= 0.9, range= 2 - 5).

Additionally, we decided to examine which mean model was the most erroneous. To do that, we considered the values previously obtained for “combat effectiveness”, “kills”, and “injuries”. The results indicated that combat effectiveness was the variable to represent the most uncertainty in its mean model, represented by a high variance of values when compared to its mean (*M*= 497.5, *SD=* 177.5, range= 67.3 – 946.9).

**Sample Size Estimation**

A secondary analysis was carried out to understand the relationship between having superpowers and IQ. We hypothesized that having superpowers would be indicative of a higher average IQ. However, since we were dealing with a novel us data set, we decided that before running further analysis it would be wise to determine if the current sample size was sufficient to observe any meaningful effects. Two ways in which we could have decided to estimate the required sample size for this scenario were by a) measuring the entire population: since there supposedly is a finite number of Avengers, we can estimate the required sample (n) by measuring the entire population (N); or b) conducting an a-priori power analysis: running an effect size test based on an appropriate statistical power, to detect an effect.

**Power Analysis and Equivalence Testing**

Accounting for the exploratory nature of this analysis, we chose to run an a-priori power analysis. We powered the study for a small effect size (d= 0.2), which was determined based on the smallest effect size of interest (SESOI). The SESOI represents the smallest effect size we would consider for our findings to be deemed meaningful.

We ran an independent two-sided t-test to determine what sample was required to observe a small effect size (d=0.2). This analysis was conducted using the “pwr” package (Champely, 2020). We set the significance level to 0.05 while aiming for 80% power to capture true findings at a reasonable chance. The t-test revealed that we would require a sample of n= 394 avengers per group (or 788 total), to achieve the desired effect size of d= 0.2 powering for 80%.

To confirm if the sample had enough power to detect the equivalence of the samples, we decided to run two one-sided equivalence tests using the “TOSTER” package (Caldwell, 2022; Lakens, 2017). This analysis reported that n= 394 would only reach a power of 75.5%. To achieve the desired 80% power, we would need at least 429 Avengers, 35 more than what we had initially anticipated.

**Effect Size**

We ran an independent sample t-test and found a test statistic of t= 4.25. Interested in determining the effect size of this t-test, we ran an analysis using the “effectsize” package (Ben-Shachar, Lüdecke, & Makowski, 2020). A 95% confidence interval was considered and yielded a d= 0.30 [0.16, 0.44].

The confidence interval range [0.16, 0.44] was equivalent to a small to medium Cohen’s d effect size. Whether we can consider this estimate precise would depend on the context. Although the test did reveal a small effect size between the groups; if we consider that our SESOI was d=0.2, we would not be able to assume this estimation to be precise, since the upper bound values of the range exceed it. Based on this, determining if this estimation is precise or not must be done so with caution.

**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. https://doi.org/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* (Version 1.3-0). Retrieved from 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. https://doi.org/10.1177/1948550617697177

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., 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. https://doi.org/10.21105/joss.01686