Assignment 1: Database Management and Sample Size Estimation

Christina Simms

Department of Psychiatry, Dalhousie University

PSYR 6003: Fundamentals of Applied Statistics

Dr. Selena Singh

February 26th, 2026

Method

The data were imported into RStudio after cloning the assignment repository, then the dataset was inspected prior to analysis. Data cleaning was completed using the tidyverse packages in R, where complete cases were kept using the complete.cases () function. This approach allowed all cases containing missing values to be removed in a single step, rather than filtering variables individually. Verification checks confirmed that no missing values remained across variables following the cleaning. The final usable sample consisted of 812 participants, compared to 814 participants prior to data cleaning. A new variable, CombatEffectiveness, was then created using the mutate () function as the sum of agility, speed, strength, and willpower to measure a composite score of overall combat ability.

A subset of the data was then created to include only Avengers who did not possess superpowers, who also died during battle. This subset contained 101 participants, with n = 73 located on the north battlefield, and n = 28 on the south battlefield. The subset was saved in both CSV and SPSS (.sav) formats to ensure compatibility across statistical software and reproducibility. Descriptive statistics were computed for the CombatEffectiveness, kills, and injuries variables for the overall subset and by battlefield location. Overall, the sample showed a mean Combateffectiveness of 497.54 (SD = 177.56), a mean of 2.55 kills (SD = 8.81), and a mean injuries score of 4.55 (SD = .74). The north battlefield demonstrated slightly higher combat effectiveness (M = 499.78, SD = 174.07) compared to the south battlefield (M = 491.68, SD = 189.53). The north battlefield also showed a higher average injuries score (M = 4.60, SD = .68) relative to the south battlefield (M = 4.43, SD = .88). Although the south battlefield shower a higher mean number of kills (M = 4.75) compared to the north (M = 1.71) overall combat effectiveness and injury rates were marginally higher in the north, suggesting that the north battlefield was slightly more effective in combat despite suffering from more injuries.

To evaluate which variable contributed the least to the overall mean model, a combined mean across CombatEffectiveness, kills, and injuries was calculated, after which each variable was removed one at a time to assess the change in the total mean. Unsurprisingly, removing CombatEffectiveness produced the largest change to the mean ($\Delta = 164.66$), indicating that this variable had the strongest influence on the overall average. This is likely due to this variable being measured on a substantially larger scale than the other two variables. Removing kills resulted in a moderate change ($\Delta = 82.83$), where removing injuries produced the smallest change ($\Delta = 81.83$). These results indicate that the injuries variable had the least erroneous impact on the combined mean.

A secondary analysis was then proposed to examine whether Avengers with superpowers differed in IQ compared to those without superpowers using an independent samples t-test. Two approaches were discussed to justify sample size requirements prior to analysis. First, an a priori power analysis was identified as an appropriate approach, as it estimates required sample size based on an expected effect size, alpha level, and desired statistical power, which ensured adequate sensitivity to detect meaningful group differences. Second, heuristic guielines were discusses as a supplementary approach, relying on general rules of thumb used in a particular field of research when prior information or literature is limited. While heuristics can provide an initial estimate, they are less specific as they are not tailed to our specific hypothesis and therefore should be considered secondary to a power analysis.

A medium effect size (Cohen's $d = .50$) was selected as the target effect for a preliminary power analysis. This choice was justified by the absence of prior research or pilot data examining the relationship between superpower status and IQ, making a medium effect a conservative yet reasonable assumption. The hypothesis predicted a meaningful difference between groups, which made a medium effect size more appropriate than assuming either a very small or very large effect. A power analysis using the pwr package was then conducted for a two-sample, two-tailed independent t-test. The alpha was set to .05 as this is a common cutoff used to reduce the chance of a false positive (i.e., Type I error). A target power was set to .80, which is a standard choice intended to provide a strong chance of detecting a true effect if one exists. The analysis indicated that approximately 64 participants per group are required (n = 63.77), corresponding to a total sample of roughly 128 participants. When compared to the data, the group without superpowers exceeded this requirement (n = 780), where the group with superpowers (n = 32) did not meet the minimum recommended sample size which suggests that the study would likely be underpowered for detecting the hypothesized effect.

To determine how one might establish evidence for no meaningful difference between groups, equivalence testing was discussed as an option. Specifically, the Two One-Sided Tests (TOST) approach can be used to evaluate whether observed group differences fall within predefined equivalence bounds representing non-significant effects. For this study, equivalence bounds of Cohen's $d = \pm .20$ were proposed, suggesting that any effect smaller than this would be considered insignificant given the lack of prior evidence. This approach allows us to demonstrate that any observed difference is small enough to be considered practically negligible rather than simply failing to reject the null hypothesis.

Finally, the study outcome reported an independent samples t-test statistic of 4.25. Given that the analysis compared two independent group means, a standardized mean difference was calculated as the effect size. Using the metafor package in R, the effect size was estimated as d

= .77, with a 95% confidence interval ranging from .41 to 1.12. According to Cohen's conventions, this corresponds to a medium-to-large effect size. The confidence interval did not include zero, indicating that the effect was statistically different from no effect. However, the interval was relatively wide, suggesting that the estimate was not highly precise and that there is some uncertainty regarding the exact magnitude of the true effect. Overall, the analysis suggests a meaningful difference between groups, but the precision of this estimate is limited, likely due to the large imbalance in group sample sizes.