Title: Assignment 1 - Data Management and Reproducible Workflow

Bala Harri

Department of Psychiatry Dalhousie University

PSYR6003 - Fundamentals of Applied Statistics

Dr Sandra Meier

2025-02-15

**Method**

The dataset will be accessed by cloning the repository from the provided GitHub link <https://github.com/iyakoven/PSYR6003-Assignment-1> .It will be inspected for missing values and unusual entries to ensure data integrity upon retrieval. Cases with missing values will be removed to maintain data quality. A new variable, combat effectiveness, will be developed as the sum of agility, speed, strength, and willpower. The dataset will be subset to include only Avengers who did not have a superpower and had died in battle. This refined dataset will be saved in both SPSS and CSV formats for further analysis.

**Justification for required sample size.**

              Power analysis will be used to determine the sample size. It ensures the study has a high probability of detecting a true effect if one exists while minimizing the risk of Type I or II errors. Using a power analysis approach will help determine the minimum sample size needed to achieve the desired statistical power. This approach will enhance the study's reliability, validity, prevent underpowered studies that may yield inconclusive results, and optimize resource allocation by avoiding unnecessarily large samples. An alternative way of determining the sample size is by using a hypothetical effect size chosen based on prior research to determine the sample size. The hypothetical effect size selected for this study is 0.8. This is because higher difference in QI is easily identified in real life compared to a small or medium IQ difference. This study targets 80% power and significance level of 0.05.

**Statistical Analysis**

An R programming language will be used to analyze data. A descriptive statistic (mean, standard deviation (SD), Maximum (Max), minimum (Min), range and Coefficient of variance will be calculated for combat effectiveness, Kills, and Injuries for the overall sample and stratified by battlefield location (North and South). Comparative analysis will be conducted to determine which battlefield exhibited the highest combat effectiveness and which had the most injuries.

**Results**

**Descriptive Statistics**

The total sample size for the study was 814. However, 2 participants had incomplete data and were excluded from the analysis, making the sample size used for the analysis 812. The mean, standard deviation (SD), and range for combat effectiveness, Kills, and Injuries were calculated separately for the entire sample and each battlefield location. The entire sample had a combat effectiveness mean of (497.53), Standard Deviation (SD) (177.56), Minimum (Min) (67.25), Maximum (Max) (946.89) and the Range is (879.64). Kills has a mean of (2.55), SD (8.81), Min (0.00), Max (79.00) range of (79.00) while Injuries has a mean of (4.55), SD (0.74), Min (2.00), Max (5.00) range 3.00. See Table 1 for the summary statistics of the entire sample. Of the three variables (combat effectiveness, kills, and injuries), kills are the most erroneous, with a Coefficient of variance of 3.44. See Table 1 for the summary statistics of the entire sample.

Table 1

*Summary Statistics showing Mean, standard deviation (SD), and range for the entire sample.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable name | Mean | SD | Min | Max | Rang |
| Combat effectiveness | 497.53 | 177.56 | 67.25 | 946.89 | 879.64 |
| Kill | 2.55 | 8.81 | 0.00 | 79.00 | 79.00 |
| Injuries | 4.55 | 0.74 | 2.00 | 5.00 | 3.00 |

**Summary Statistics according to battlefield location**

The North battlefield demonstrated combat effectiveness with a mean of (499.78), SD (174.06), Min (130.68), Max (897.06), and range (766.38). The kills at the northern battlefield had a mean (1.71), SD (4.56) Min (0), Max (34), and Range (34), while the injuries at the north had a Mean (4.60), SD (0.68) Min (2) Max (5) range (3). The southern battlefield had combat effectiveness with a mean (491.67), SD (189.52), Min (67.25), Max (946.89), and a range of (879.64). The kills at the southern location had a mean of (4.75), SD (14.99), Min (0), Max (79) and a range of (79). The southern location injuries had a mean (4.42), SD (0.8789), Min (2), Max (5), and a range (3). See Table 2 for summary statistics on battlefields according to location North and South.

At a hypothetical effect size of (d = 0.8), sample size (n = 406 per group, total N = 812) and significance level (α = 0.05), the power value obtained from the analysis is approximately 1.

**Table 2**

***Summary statistics according to the battlefields (North and South)***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Battle**  **fields** | **Mean of Combat Effectiveness** | **SD of Combat Effectiveness** | **Min of Combat Effectiveness** | **Max of Combat Effectiveness** | **Range of Combat Effectiveness** | **Mean of kills** | **SD of kills** | **Min of kills** | **Max of kills** | **Range of kills** | **Mean of injuries** | **SD of injuries** | **Min of injuries** | **Max of injuries** | **Range of injuries** |
| North | 499.78 | 174.07 | 130.68 | 897.06 | 766.38 | 1.71 | 4.57 | 0.00 | 34.00 | 34.00 | 4.60 | 0.68 | 2.00 | 5.00 | 3.00 |
| South | 491.68 | 189.53 | 67.25 | 946.89 | 879.64 | 4.75 | 14.99 | 0.00 | 79.00 | 79.00 | 4.43 | 0.88 | 2.00 | 5.00 | 3.00 |

**Table 3**

***Summary statistics of the entire sample with Coefficient of Variance***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mean of Combat Effectiveness mean** | **SD of Combat effectiveness** | **Coefficient of Variance for Combat Effectiveness** | **Mean of Kills Mean** | **SD of Kills** | **Coefficient of variance**  **Kills** | **Mean of Injuries** | **SD of injuries** | **Coefficient of variance**  **Of injuries** |
| 497.53 | 177.55 | 0.36 | 2.55 | 8.81 | 3.45 | 4.55 | 0.74 | 0.163 |

Power analysis was used for a two-sample t-test to determine the required minimum sample size at an effect size (d) = 0.8, with a significance level of 0.05 and a statistical power of 0.80. The results indicate that a minimum sample size of about 26 participants per group is needed to achieve sufficient power for detecting the specified effect. However, to ensure truly no difference between the groups (zero effect), an equivalence test was conducted. The analysis was performed with a significance level of 0.05, a statistical power of 0.80, and equivalence bounds set at -0.8 and 0.8 for Cohen’s *d*. The required sample size obtained is 27 per group or 54 in total. Consequently, indicating a power of 0.82 at a sample size of 54. Furthermore, at an independent t-test analysis of 4.25, the study found a small effect size (Cohen’s d = 0.03) with a 95% confidence interval [0.16, 0.44].

**Discussion**

The study assessed combat effectiveness, injuries, and kills among Avengers who lacked superpowers and died in battle.  The study explored the differences between the battlefields in the North and South. The results indicated that combat effectiveness was slightly higher in the North (Mean = 499.78, SD = 174.06) compared to the South (Mean = 491.67, SD = 189.52. Similarly, the average number of injuries was marginally more significant in the North (Mean = 4.60, SD = 0.68) than in the South (Mean = 4.42, SD = 0.88). However, a notable difference was observed in kills, with the South battlefield exhibiting a significantly higher kill count (Mean = 4.75, SD = 14.99) than the North (Mean = 1.71, SD = 4.57), suggesting greater combat intensity.

An independent samples t-test (t = 4.25) was provided to calculate the effect size for the compare IQ between Avengers with and without superpowers. The effect size was Cohen’s d = 0.30, with a 95% confidence interval [0.16, 0.44]. According to Cohen’s conventions, this effect size is small, indicating that although a difference was detected, it is unlikely to have practical significance. The confidence interval was moderately narrow, suggesting a reasonably precise estimate of the effect.

**Conclusion**

In conclusion, the study revealed slight regional differences in combat effectiveness and injuries among non-superpowered Avengers, with higher effectiveness and injury rates in the North but a significantly greater kill count in the South. The IQ differences between superpowered and non-superpowered Avengers has a small effect size of (0.30), which implies minimal practical relevance. The power analysis confirmed a 100% detection capability, though the power is high, it is rare and might indicate an overpowered study.