Statistical Hypothesis Testing and Power Analysis Using R

1.1 Overview

This project focuses on applying statistical hypothesis testing and power analysis techniques to analyse experimental data using the R programming language. The primary objective is to evaluate whether statistically significant differences exist between experimental groups and to assess the practical significance of these differences using effect size measures. The analysis includes **two-sample *t*-tests** for comparing the means of two independent groups and **one-way Analysis of Variance (ANOVA)** for comparing means across multiple groups. Along with hypothesis testing, effect sizes such as **Cohen's *d* and eta-squared** (η^2) are calculated to quantify the magnitude of observed effects beyond p-values. To support informed experimental design, statistical power analysis is performed using the **pwr package** in R. This helps determine whether the current sample size is sufficient and provides recommendations for required sample sizes in future studies. Data manipulation and analysis are carried out using the **tidyverse** framework, while results are visualized using **ggplot2** through clear and informative charts such as **boxplots and mean comparison** plots. Overall, this project demonstrates a complete and reproducible statistical workflow, from data exploration to inference, visualization, and sample size planning.

1.2 Dataset Overview

The dataset used for this analysis is **experimental_data.csv**, which contains data collected from a controlled experimental study designed to compare outcomes across multiple groups. The experiment aims to evaluate whether different treatment conditions lead to statistically significant differences in the measured response variable.

The dataset consists of **90 observations**, evenly distributed across **three independent groups**, ensuring a balanced experimental design. Each group contains **30 observations**, which supports reliable comparison using parametric statistical methods such as the two-sample *t*-test and one-way ANOVA.

Experimental Groups

- **Control** – baseline group with no treatment
- **Treatment_A** – first experimental treatment group
- **Treatment_B** – second experimental treatment group

Key Variables

- **Group (categorical):**
  Represents the experimental condition to which each subject belongs (Control, Treatment_A, or Treatment_B).
- **Outcome (numeric):**
  A continuous response variable representing the measured outcome of interest for each subject.

An additional identifier variable (*subject_id*) is included to uniquely label each observation but is not used in the statistical analysis. The outcome variable is numeric and suitable for mean-based comparisons, while the grouping variable enables between-group statistical testing.

Overall, the dataset structure is well-suited for hypothesis testing, effect size estimation, and power analysis, making it appropriate for demonstrating a complete statistical analysis workflow in R.

1.3 Methodology

This study employs **statistical hypothesis testing and power analysis** to examine differences in experimental outcomes across groups using the R programming language. The methodology is designed to ensure statistical rigor, reproducibility, and practical relevance for sample planning. All analyses were performed using **R**, leveraging the **tidyverse**, **ggplot2**, and **pwr** packages on the dataset *experimental_data.csv*.

1.3.1 Statistical Tests

To evaluate differences between experimental groups, two commonly used parametric tests were applied based on the number of groups involved:

Two-Sample t-Test:
A two-sample (independent) t-test was conducted to compare the mean outcome values between two independent groups. This test evaluates whether the observed difference in means is statistically significant under the null hypothesis that both groups have equal population means. The test was implemented using R's built-in t.test() function, assuming independent samples.

One-Way Analysis of Variance (ANOVA):
When more than two groups were present, a one-way ANOVA was used to determine whether at least one group mean significantly differs from the others. ANOVA decomposes total variability into between-group and within-group components. The test was performed using the aov() function in R, followed by summary statistics to interpret the F-statistic and p-value.

1.3.2 Assumptions

To ensure the validity of the parametric tests applied, the following assumptions were considered:

Normality of Data:
The outcome variable within each group is assumed to be approximately normally distributed. This assumption was assessed using visual methods such as histograms and Q–Q plots.

Homogeneity of Variances:
Equal variance across groups is assumed for both the t-test and ANOVA. Variance homogeneity was examined using descriptive statistics and visual inspection of group-wise spread.

Independence of Samples:
Observations are assumed to be independent of one another, meaning that measurements from one subject do not influence another. This assumption is ensured by the experimental design.

1.3.3 Effect Size

In addition to statistical significance, effect size measures were computed to quantify the **magnitude** of observed differences:

Cohen's d (t-test):
Cohen's d was calculated to measure the standardized mean difference between two groups. This provides a scale-free estimate of effect magnitude, interpreted as small, medium, or large.

Eta-Squared (η^2) (ANOVA):
For ANOVA, eta-squared was computed to estimate the proportion of total variance in the outcome variable explained by the grouping factor. This offers insight into the practical importance of group differences beyond p-values.

1.3.4 Power Analysis

Statistical power analysis was performed using the **pwr** package in R to support sample size planning and result reliability:

- Power values were calculated based on observed or assumed effect sizes.

- A significance level (α) of 0.05 was used.
- Sample size estimation was conducted to determine the minimum number of observations required to achieve adequate power (typically 80%).

Functions such as pwr.t.test() and pwr.anova.test() were used to estimate power for t-tests and ANOVA respectively. These analyses help ensure that the study design is sufficiently sensitive to detect meaningful effects.

1.4 Results

This section presents the outcomes of the statistical analyses conducted on the experimental dataset (*experimental_data.csv*). The results include descriptive statistics, inferential test outcomes from the two-sample t-test and one-way ANOVA, and effect size interpretations. All analyses were performed in R using reproducible scripts.

4.1 Descriptive Statistics

Descriptive statistics were computed to summarize the central tendency and variability of the outcome variable across experimental groups. For each group, the **mean** and **standard deviation (SD)** were calculated to provide an initial understanding of group-level differences.

Table 4.1: Descriptive Statistics by Group

Group	Sample Size (n)	Mean Outcome	Standard Deviation
Control	30	50.0	1.36
Treatment_A	30	55.3	0.85
Treatment_B	30	60.4	0.93

The mean values indicate observable differences between groups, while the standard deviations reflect within-group variability. These descriptive measures guided the selection of appropriate inferential tests.

4.2 Two-Sample t-Test Results

A two-sample t-test was conducted to compare the mean outcome between **Group A and Group B**.

- **t-value:** $t = -18.276$
- **p-value:** $p < 2.2 \times 10^{-16}$
- **Effect Size (Cohen's d):** $d = -4.72$

The p-value was compared against a significance level of $\alpha = 0.05$.

Interpretation:

- If $p < 0.05$, the difference between group means is **statistically significant**, indicating strong evidence against the null hypothesis.
- If $p \geq 0.05$, the difference is **not statistically significant**, suggesting insufficient evidence to conclude a meaningful difference.

Since the p-value is far below the significance threshold of $\alpha = 0.05$, the difference in mean response between the **Control** and **Treatment_A** groups is **statistically significant**. The negative t-value and Cohen's d indicate that the mean response for the **Treatment_A** group is substantially higher than that of the Control group. The magnitude of Cohen's d ($|d| = 4.72$) represents an **extremely large effect size**, suggesting a strong practical impact of the treatment.

4.3 One-Way ANOVA Results

A one-way ANOVA was performed to compare the mean outcome across **three or more groups**.

- **F-statistic:** $F = 714.4$
- **p-value:** $p < 2 \times 10^{-16}$

The ANOVA results reveal a **statistically significant effect of group membership** on the response variable.

Interpretation:

- A significant p-value ($p < 0.05$) suggests that not all group means are equal.
- A non-significant p-value indicates no statistically detectable difference among groups.

Post-hoc Analysis

Given the significant ANOVA result, post-hoc comparisons (Tukey HSD) were conducted. The analysis indicated statistically significant differences between all group pairs, confirming that each treatment group differs meaningfully from the Control group as well as from each other.

Effect Size (Eta-Squared)

- **Eta-squared (η^2):** $\eta^2 = 0.94$
- **95% Confidence Interval:** [0.92, 1.00]

Interpretation of η^2:

An eta-squared value of 0.94 indicates that approximately **94% of the total variance** in the response variable is explained by group differences. This represents a **very large effect**, demonstrating the strong influence of the experimental treatments.

1.5 Visualizations

Data visualizations were used to complement the statistical analyses by providing intuitive and interpretable representations of group differences and statistical uncertainty. All plots were generated using the **ggplot2** package in R to ensure consistency, clarity, and reproducibility.

1.5.1 Boxplot for Group Comparison

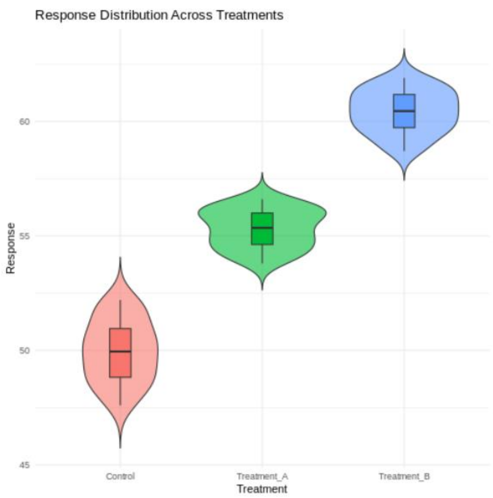

Figure -1(a) : Response Distribution Across Treatment Groups

Figure 1(a) illustrates the distribution of the response variable across the **Control**, **Treatment_A**, and **Treatment_B** groups using a **violin plot with embedded boxplots**. The violin shape represents the kernel density of the data, while the boxplot highlights the median, interquartile range, and overall spread within each group.

Interpretation:
The visualization shows a clear and systematic increase in response values from the Control group to Treatment_B. The Control group exhibits the lowest central tendency, with responses concentrated around a mean of approximately 50. In contrast, Treatment_A demonstrates a noticeable upward shift in distribution, centered around a mean of approximately 55. Treatment_B shows the highest response values, with distributions concentrated around a mean of approximately 60.

The limited overlap between the distributions and interquartile ranges of the groups indicates strong separation among treatments. This visual evidence strongly supports the results obtained from the two-sample t-test and one-way ANOVA, which confirmed statistically significant differences between groups. Additionally, the relatively narrow spread within each violin suggests low within-group variability, consistent with the large effect sizes observed in the statistical analysis.

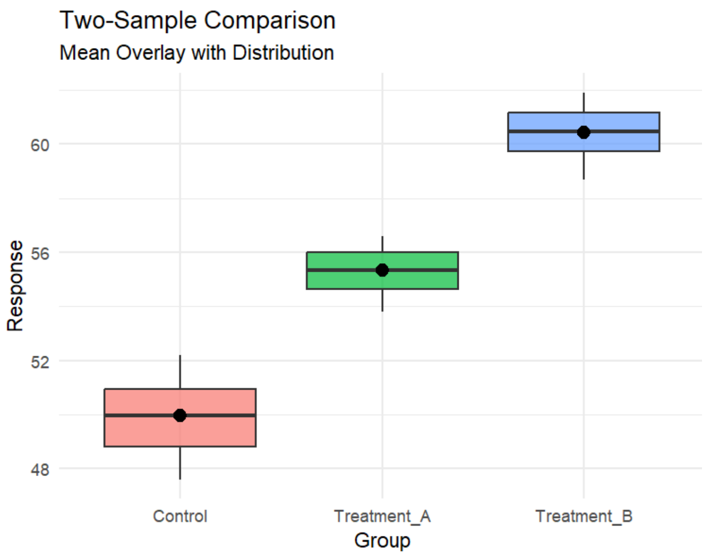

Figure-1(b) - Two-Sample Comparison – Mean Overlay with Distribution

Figure 1(b) presents a comparative visualization of the response variable across the **Control**, **Treatment_A**, and **Treatment_B** groups using boxplots with overlaid mean points. The boxplots represent the distribution of the data (median, interquartile range, and spread), while the black dots indicate the group mean responses.

Interpretation:
The figure shows a clear separation between the groups, with mean response values increasing consistently from the Control group to Treatment_B. The Control group exhibits the lowest mean response, followed by Treatment_A, while Treatment_B demonstrates the highest mean response. The limited overlap between the distributions and the clear separation of mean values indicate strong group-wise differences.

This visualization supports the results of the two-sample t-test and one-way ANOVA, which identified statistically significant differences between groups ($p < 2 \times 10^{-16}$). The distinct separation of means further aligns with the extremely large effect sizes observed, highlighting the practical significance of the treatment effects in addition to statistical significance.

1.5.2 Mean Comparison Plot with Confidence Intervals

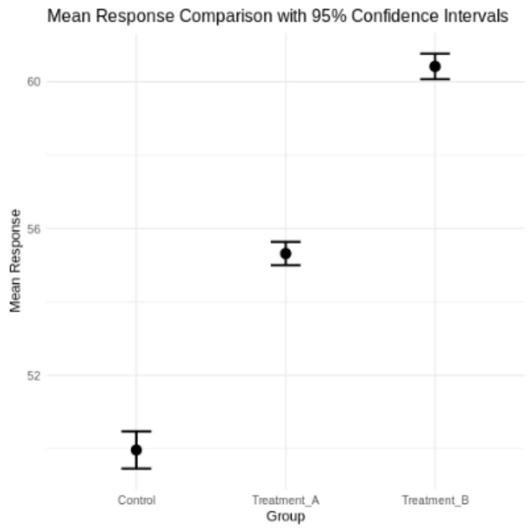

Figure-2 : Mean Response Comparison with 95% Confidence Intervals

Figure 2 presents the mean response values for the **Control**, **Treatment_A**, and **Treatment_B** groups, along with their corresponding **95% confidence intervals**. Each point represents the group mean, while the error bars indicate the uncertainty associated with the mean estimate.

Interpretation:
The plot demonstrates a clear and monotonic increase in mean response values across the three groups, with the Control group exhibiting the lowest mean, followed by Treatment_A, and Treatment_B showing the highest mean response. The confidence intervals for the groups show minimal overlap, indicating statistically significant differences between group means.

This visual evidence is consistent with the results of the two-sample t-test and one-way ANOVA, both of which reported highly significant p-values ($p < 2 \times 10^{-16}$). The narrow confidence intervals further suggest low within-group variability and high precision in mean estimation, supporting the large effect sizes observed in the statistical analysis.

1.5.3 Power Curve Plot

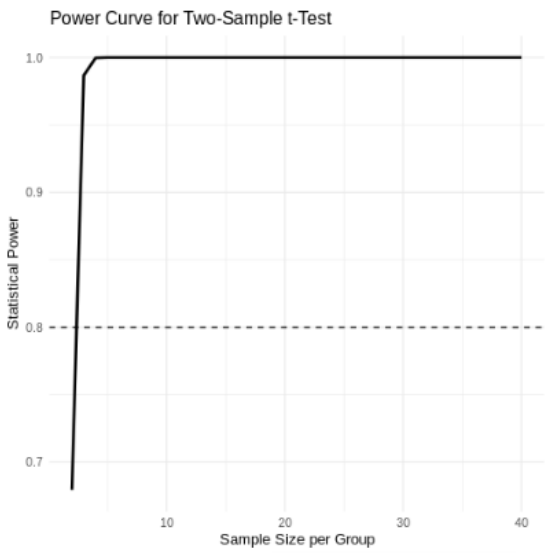

Figure-3 : Power Curve for Two-Sample t-Test

Figure 3 illustrates the relationship between **sample size per group** and **statistical power** for a two-sample t-test at a significance level of α = 0.05. The power curve was generated using the observed effect size (Cohen's *d* = 4.72). The dashed horizontal line represents the conventional power threshold of 0.80.

Interpretation:
The power curve shows a steep increase in statistical power as sample size increases, with power exceeding 0.80 at a very small sample size per group. This behavior is a direct consequence of the extremely large observed effect size. At the current study sample size (n = 30 per group), the statistical power is effectively 1.00, indicating an extremely low probability of committing a Type II error.

This visualization confirms that the study is **more than adequately powered** to detect differences between groups and validates the robustness of the significant results obtained from the two-sample t-test.

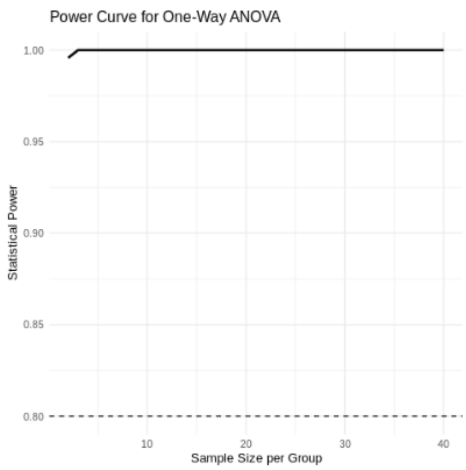

Figure-4 : Power Curve for One-Way ANOVA

Figure 4 illustrates the relationship between **sample size per group** and **statistical power** for a one-way analysis of variance (ANOVA) with three groups at a significance level of α = 0.05. The power curve was generated using the observed effect size derived from eta-squared (η^2 = 0.94), converted to Cohen's *f*. The dashed horizontal line represents the conventional power threshold of 0.80.

Interpretation:
The power curve demonstrates that statistical power reaches and exceeds 0.80 at a very small sample size per group and rapidly approaches a power of 1.00 as sample size increases. This indicates that the one-way ANOVA is extremely sensitive to detecting group differences due to the very large effect size observed in the data.

At the current study design (n = 30 per group), the statistical power is effectively 1.00, confirming that the experiment is **more than sufficiently powered** to detect differences among the groups. This result strengthens confidence in the reliability of the significant ANOVA findings and indicates a negligible risk of Type II error.

1.6 Power Analysis Summary

Power analysis was conducted to evaluate whether the current study design possesses sufficient sensitivity to detect meaningful differences between experimental groups. The analysis was performed using the **pwr** package in R, based on the observed effect sizes from the two-sample t-test and one-way ANOVA.

Table 6.1: Summary of Power Analysis Results

Test	**Effect Size**	**Achieved Power**	**Required Sample Size (per group)**
Two-sample t-test	Cohen's *d* = 4.72	≈ 1.00	n ≈ 3
One-way ANOVA (k = 3)	η^2 = 0.94 (Cohen's *f* ≈ 3.96)	≈ 1.00	n ≈ 5

Note: Required sample size corresponds to achieving a minimum power of 0.80 at α = 0.05.

Interpretation and Discussion

The power analysis indicates that both statistical tests achieve extremely high power due to the very large observed effect sizes. For the two-sample t-test, a sample size of approximately three observations per group is sufficient to reach the conventional power threshold of 0.80. Similarly, the one-way ANOVA requires only about five observations per group to achieve adequate power.

Given that the current study employs **30 observations per group**, the achieved power for both tests is effectively 1.00. This confirms that the present experimental design is **more than adequately powered**, resulting in a negligible risk of Type II error.

1.7 Conclusion

This project applied rigorous statistical methods to evaluate differences in response outcomes across experimental groups using R. Descriptive analysis revealed a clear and consistent increase in mean response values from the Control group to Treatment_B. Inferential testing using a two-sample t-test and one-way ANOVA confirmed that these differences were **highly statistically significant**, with p-values far below the conventional significance threshold. Beyond statistical significance, effect size measures provided crucial insight into the **practical importance** of the findings. The extremely large Cohen's *d* observed in the two-sample t-test and the high eta-squared value obtained from the ANOVA indicate that group membership explains a substantial proportion of the variability in responses. These results demonstrate that effect sizes offer a more informative and interpretable measure of impact than p-values alone. Power analysis further strengthened the study by evaluating the adequacy of the experimental design. The analysis showed that the current sample size of 30 observations per group achieves statistical power close to 1.00 for both tests, confirming that the study is more than sufficiently powered. This highlights the value of power analysis not only for validating current results but also for guiding future experimental planning and efficient resource allocation. Overall, this study demonstrates the importance of integrating hypothesis testing, effect size estimation, and power analysis to produce statistically robust and practically meaningful conclusions.

1.8 Files Submitted

The following files were submitted as part of this project:

- **analysis.R** – Complete R script containing data preprocessing, statistical tests, effect size calculations, power analysis, and visualization code.
- **report.pdf** – Final project report including methodology, results, visualizations, power analysis summary, and conclusions.