**Hypothesis testing in Experiment Outcomes**

1. Understand the concept of hypothesis testing and its role in assessing experiment outcomes.
   1. Explore the purpose of hypothesis testing in analyzing experiment results.
   2. Familiarize with key terminologies related to hypothesis testing.
   3. Learn the process of hypothesis testing and power calculation.
   4. Conduct hypothesis testing using R to evaluate experiment outcomes.

**Code:**

**#Null Hypothesis (H0):**

**#**H0: There is no significant relationship between a customer's age and the number of web purchases made.

**#Alternative Hypothesis (Ha):**

**#**Ha: There is a significant relationship between a customer's age and the number of web purchases made.

**#Step 1: Loading and Preparing the Data**

# Load necessary libraries library (ggplot2)

# Load the dataset

superstore\_data <- read.csv("D:/MSc DS/Semester 1/Retail Market Analysis/Practical/superstore\_data.csv") selected\_data <- superstore\_data[, c("Id","Year\_Birth", "Marital\_Status", "Education", "Dt\_Customer", "Recency", "NumStorePurchases", "NumWebPurchases", "NumWebVisitsMonth")]

selected\_data <- unique(selected\_data)

# Assuming 'selected\_data' is your dataset

**# Step2: Perform the t-test to assess the relationship between age and web purchases**

t\_test\_result <- t.test(selected\_data$Year\_Birth, selected\_data$NumWebPurchases) # View the t-test results

print(t\_test\_result)

# Calculate the mean difference between the groups (effect size for t-test)

mean\_difference <- mean(selected\_data$Year\_Birth) - mean(selected\_data$NumWebPurchases) standard\_deviation <- sqrt((var(selected\_data$Year\_Birth) + var(selected\_data$NumWebPurchases)) / 2) effect\_size <- mean\_difference / standard\_deviation

# View the effect size print(effect\_size)

**#Step3: Power Calculation**

# Calculate the statistical power for the t-test assuming a sample size library(pwr)

sample\_size <- 100 # You should input an appropriate sample size power <- pwr.t.test(n = sample\_size, d = effect\_size, sig.level = 0.05,

power = NULL, type = "two.sample", alternative = "two.sided")

# View the power calculation results print(power)

**Interpretation and Implications:**

**T-Test Result:** The extremely small p-value (< 0.0001) indicates strong evidence against the null hypothesis, suggesting a highly significant difference between the mean age and the mean number of web purchases. The confidence interval also confirms this significant difference, indicating that the means of the two groups (age and web purchases) are not equal.

**Two Sample t-test:**

* t-value: 7558.7

* Degrees of Freedom (df): 2479.1

* p-value: < 2.2e-16 (extremely small)

* Confidence Interval: 95% CI for the difference in means: (1964.211, 1965.231)

* Sample Estimates: Mean of 'Year\_Birth': 1968.805804, Mean of 'NumWebPurchases': 4.084821

Power Calculation: The calculated statistical power of 1 indicates a high probability of correctly detecting the observed effect (difference in means) if it truly exists in the population. A power of 1 suggests that the study is well-powered to detect the effect, leaving almost no chance of a Type II error (false negative).

* Sample Size (n): 100 in each group * Effect Size (d): 225.8605

* Significance Level (sig.level): 0.05 * Power (power): 1 (High)