

**King Hussein Faculty of Computing Sciences**

**Department of Data Science**

**Statistical Programming using R.**

**Spring 2023/2024**

**Lab #10 – Hypothesis Testing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Exercise** | **Ex1** | **Ex2** | **Ex3** |
| **Mark** | **3** | **4** | **3** |
| **Total Mark** | **/3** | **/4** | **/3** |

**Lab Objectives**

* Understand categorical and numerical statistical hypothesis testing
* Understand the assumptions of each test
* Use normality testing, equality of variance testing, and visualisations to confirm the validity of the assumptions.
* Perform Chi-squared tests of association and goodness-of-fit, Fisher Exact test, One-sample proportion test, Two-independent samples proportions, one-sample mean, two-independent samples means, and paired data t-test.

**Ungraded Steps**

* Load the tidyverse library
* Import Markteing.csv as data1, and Drug\_data.csv as data2.

***Description of the Marketing dataset:***

It contains 6 variables

1. User ID: A unique identifier assigned to each user in the dataset.
2. Test Group: Indicates whether the user is part of the test group. This column contains categorical values indicating the user's membership in either the "ad" or "psa" group.
3. Converted: A binary flag indicating whether the user was converted (TRUE) or not (FALSE).
4. Total Ads: The total number of ads viewed or interacted with by the user.
5. Most Ads Day: The day of the week when the user engaged with the most ads.
6. Most Ads Hour: The hour of the day when the user engaged with the most ads

Description of the Drug\_data dataset:

It contains 2 variables

1. First Name: The first name of an individual.
2. Drug: The specific drug or treatment associated with the individual.
3. Mem\_Score\_Before: Memory score before undergoing the drug or treatment. This score represents the individual's memory performance prior to intervention.
4. Mem\_Score\_After: Memory score after undergoing the drug or treatment. This score represents the individual's memory performance after the intervention.

**Exercise 1:**

1. Use a proper statistical test to test the association between the marketing test group and the converted in the group.
   1. Write the null and alternative hypotheses.
   2. Test at 95% level of confidence.
2. Use a proper statistical test to test the equality between two ad and psa test group in the proportion of converted .
   1. Write the null and alternative hypotheses
   2. Test at 95% level of confidence.
   3. Compare your conclusion with part 1.

**Exercise 2:**

1. Use a proper statistical test to confirm if the average most ads hour = 6.
   1. Write the null and alternative hypotheses.
   2. Test your data at 99% level of confidence.
   3. Confirm the test by constructing your own 99% confidence interval.
2. Use a proper statistical test to compare the average of each test group.
   1. Write the null and alternative hypotheses.
   2. Test your data at 99% level of confidence.

**Exercise 3:**

1. Use a proper statistical test to confirm if the campaign improved Memory scores by 2 after undergoing the drug.
   1. Write the null and alternative hypotheses.
   2. Test your data at 95% level of confidence.