Empirical Study

CS 345-346

Problem 2 : ClassNote Taking Application

Group 18

Name	Roll Number
Param Aryan Singh	180101055
Parth Dhananjay Bakare	180101056
Pooja Gajendra Bhagat	180101057
Tejas Prashant Khairnar	180101081

Contents

1 Identification of Research Questions **1**

2 Determination of Variables **2**

3 Choose Participants **3**

4 Experiment Design **3**

5 Data Collection **3**

6 Data Analysis **4**

1 Identification of Research Questions

The aim of our empirical research is to inspect and analyse several interfaces and quantify some of their key aspects.
To measure how good an interface is, we will measure two variables:

- **Aesthetic beauty** of the interface, which reflects satisfaction of the interface.
- The difficulty of locating **functional elements** (such as buttons) in the interface, which reflects learnability and memorability of the interface.

So, our **research question** has the following structure:
RQ 1: How usable is the interface in terms of its satisfaction, learnability and memorability?

We will further **refine this question** in the following sections, once we determine the relevant variables and factors (and their levels). This refinement would make the question **testable** enough so that it is **generalized** i.e. **balancing the tradeoff**. The research question investigates key attributes regarding usability of the system and thus, we get a comprehensive idea of how good our interface is.We will form appropriate hypotheses using this question and test them while analysing our observations.

2 Determination of Variables

We identified the following **independent variables** (or factors) whose values would affect the dependent variables:

1. **Number of sections in the interface:** The number of sections (editing space, sidebar, topbar etc) the interface is divided into.
2. **Margins and padding:** An estimate of the spacing between various elements in the interface and the text in the editing space.

The dependent variables we have choosen are:

1. **Aesthetic beauty (of the interface):**

- **Aesthetic beauty** of the interface gives a measure of the its **satisfaction**. An interface scoring high in aesthetic beauty will be pleasing to look at, easy on the eyes, and is likely to be used repeatedly by the user in future as well.
- We will measure **aesthetic beauty** on a scale of **1 to 10 (interval scale)**, with 1 being least aesthetic and 10 being most aesthetic.

2. **Difficulty (in locating functional elements):**

- Difficulty in locating functional elements gives a measure of the interface's **learnability** and **memorability**.
- It is easy to carry out the desired functionalities in an interface where **locating elements** and understanding their functions is not difficult.
- An ideal interface would have **clear and visible buttons** as functional elements, and their icons/positions in the interface would indicate their functions.
- We will measure **difficulty** on a scale of **1 to 10 (interval scale)**. User behaviour in this aspect would be better understood by recording the ease of finding functional elements on the same 1 to 10 scale, and then defining **difficulty as 11 - ease**.

Our empirical research consists of two testable questions, whose answers will answer our research question.

Testable question 1: How does the aesthetic beauty of the interface depend on the number of sections in the interface and the margin and padding of the interface?
Testable question 2: How does the difficulty in locating functional elements depend on the number of sections in the interface and the margin and padding of the interface?

RQ 2 (Research question): How do the satisfaction, learnability and memorability of the interface depend on the number of sections in the interface and the margin an padding of the interface?

H0 (Null hypothesis): The satisfaction, learnability and memorability of the interface do **not** depend on the number of sections in the interface and the margin an padding of the interface.
H1 (Alternate hypothesis): The satisfaction, learnability and memorability of the interface depend on the number of sections in the interface and the margin an padding of the interface.

3 Choose Participants

We chose **13 participants** for our research, all of them being students at some school or at some college, as they are going to be the end users of our application.
To draw some **reliable conclusion** it is advised to have **12 to 25 participants** for our research and we have 13 participants for our research which perfectly meets the criteria.

4 Experiment Design

After framing the final questions, we built a Google Form to collect data. We gathered 13 interfaces, attached them in the form and asked our participants to carefully judge the interface and rate their aesthetic beauty and ease of locating elements. To take care of the **practice effect**, we **randomised the order** of interfaces for each participant, ensuring that all interfaces were scored fairly. We used the **within-subject** method as we asked all participants to judge all the interfaces.

Interfaces – **Click Here**

5 Data Collection

Use the link below to check our recorded data. To compare the interfaces, we have used a new metric (ratio), whose significance will be explained in the next section.

Empirical Study Data Collection – **Click Here**

6 Data Analysis

An ideal interface will score **high on aesthetic beauty** and **low on difficulty**. Consider a scenario: interface X has a low aesthetic beauty score and a low difficulty score, whereas interface Y has a high aesthetic beauty score and a high difficulty score. Comparing these two interfaces would be a tricky task, since one scores higher on the first aspect and the other scores higher on the second aspect. There could always be situations where this tradeoff would need to be balanced and judging which interface is better could prove to be difficult.

To compare different interfaces using both these parameters, we designed a new metric - **the ratio of aesthetic beauty and difficulty**. Quite clearly, **a higher ratio would indicate a better interface**. This metric could be used to settle ties in a lot of situations, and help one judge a stalemate between interfaces.

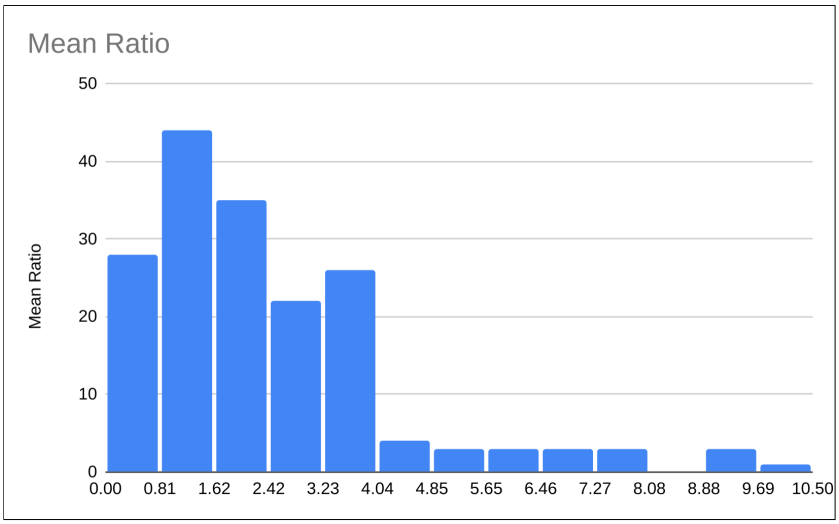
Application	Number of elements	Margin and padding	Mean Ratio
Evernote	3	Medium	4.12
Google Docs	3	Low	1.9
Google Keep	2	High	3.52
Miro	3	Medium	2.92
MS Word	4	Low	1.36
Notion	3	Low	1.58
OneNote	4	Low	1.49
Pages	4	Medium	1.48
Simple Note	3	High	2.24
Taskade	4	Medium	2.4
Xournal	4	Low	1.89
Zoho	2	High	3.32
NoteMaster	3	High	3.44

Number of elements	Mean ratio	Margin and padding	Mean ratio
2	3.42	Low	1.644
3	2.7	Medium	2.81
4	1.72	High	3.13

The table shows a summary of the data we collected. For every interface-participant pair, we calculated the **aesthetic beauty-difficulty ratio**. The mean of these ratios for a single interface is shown in the mean ratio column. In the subsequent tables, we grouped the interfaces - first by number of elements/sections (the first independent variable) and then by margin and padding (the second independent variable). After grouping the interfaces, we calculated the mean ratio for each group to study the dependence of the aesthetic beauty-difficulty ratio on the number of sections and margin and padding.

The analysis shows a **promising result** - the **ratio decreases as the number of sections increase** and the **ratio increases as the margin and padding increases**. This can be explained by the fact that a higher number of sections and less whitespaces makes the interface look more congested and thus less aesthetic. As the number of elements increases, the chance that the purpose of a specific button getting misunderstood also increases, thereby increasing the difficulty score.

The **data collected** by us has the distribution as shown in the following image.

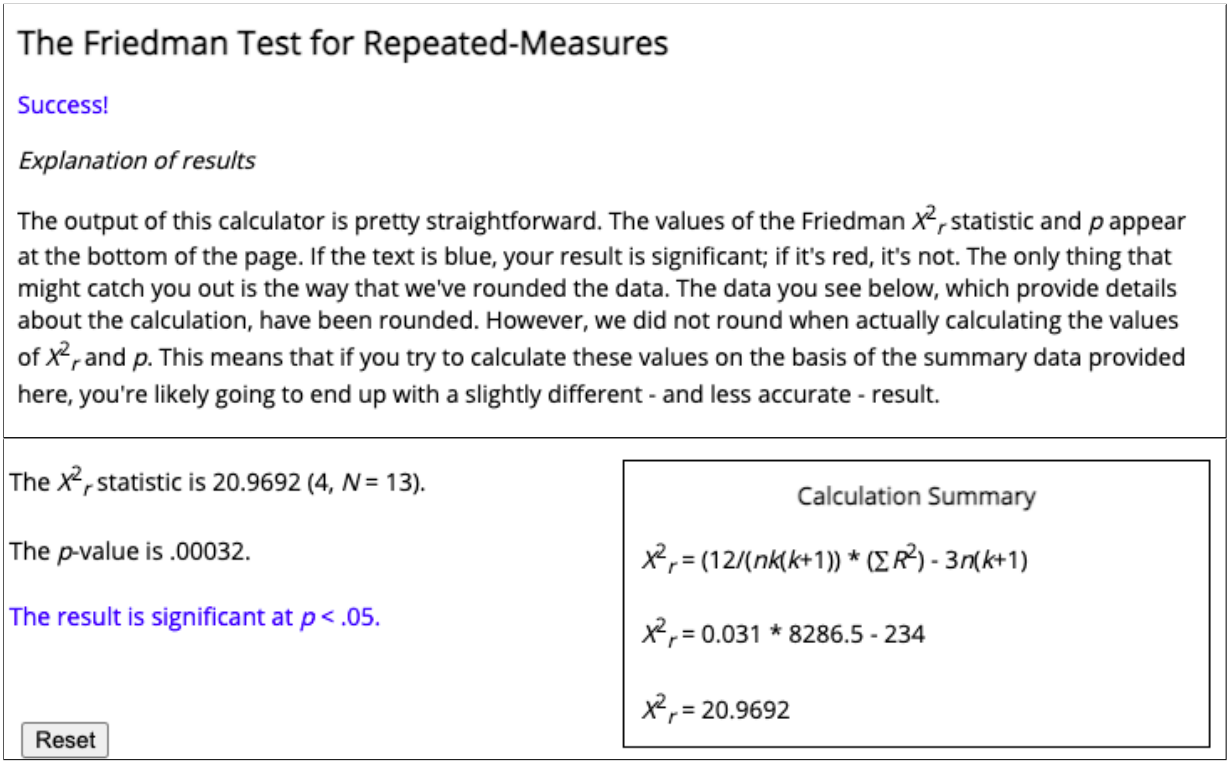


The data collected by us **does not follow the normal distribution** so we cannot apply the parametric test (t -test). So we can only apply some **non-parametric test** to our data and since we have used the within subject method for the experiment and the **factor** (number of elements) has **3 levels** this makes the data eligible for the **Friedman Test** for repeated measures.

While analysing the data using a **Friedman test**, part of the process involves checking to make sure that the data we want to analyse can actually be analysed using a Friedman test. We confirm the following **4 assumptions** before doing the Friedman's test.

Assumption 1: One factor that is measured on three or more different occasions.
Assumption 2: Group is a random sample from the population.
Assumption 3: Your dependent variable should be measured at the ordinal or continuous level. Examples of continuous variables include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.
Assumption 4: Data samples do NOT need to be normally distributed.

We used an **online calculator** to perform the Friedman test. The calculator took the mean ratios (interface-wise) as its input, and computes some relevant values to judge the significance of the collected data.



As shown in the above figure, our data passes the Friedman test, and the data for empirical research is validated. The data we collected and the inferences we concluded have a low probability of being a matter of chance.