ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»»

Факультет компьютерных наук

Департамент программной̆ инженерии

Домашнее задание

по курсу Эргономика человеко-машинного взаимодействия

по направлению подготовки бакалавров

09.03.04 «Программная инженерия»

Выполнили:

студенты группы БПИ152

образовательной̆ программы

09.03.04 «Программная инженерия»

Суровцев М.О.

Калянов К.М.

Шадрин М.Д.

Москва, 2018

[Введение 3](#_Toc516199790)

[Приложение 3](#_Toc516199791)

[Предполагаемые изменения 3](#_Toc516199792)

[Гипотеза 3](#_Toc516199793)

[!Переменные 3](#_Toc516199794)

[Планирование эксперимента 3](#_Toc516199795)

[**Участники** 4](#_Toc516199796)

[**Обоснование выбора стиля тестирования** 4](#_Toc516199797)

[**Статистические тесты** 4](#_Toc516199798)

[**Прототип** 5](#_Toc516199799)

[Исследование 6](#_Toc516199800)

[!Проведение исследования 6](#_Toc516199801)

[!Результаты исследования 6](#_Toc516199802)

[!Заключение 12](#_Toc516199803)

[Ссылки 12](#_Toc516199804)

# 

# **Введение**

## **Приложение**

В рамках данной работы было выбрано приложение «Hamster» (Hamster)[1] — агрегатор досуга для смартфонов на ОС Android, основанный на применении психологических тестов и критериев настроения.

Данное приложение является результатом труда одного из авторов выполненной работы. Исследование проводилось с целью определить, какой дизайн приложения позволит пользователям более доступно воспринимать инструкции приложения, не вызывая желания покинуть приложение раньше, чем оно выдаст результирующий список форм досуга.

## **Предполагаемые изменения**

В данном исследовании мы планируем изучить изменение продолжительности пользовательской сессии и восприятия приложения с помощью редизайна. В приложении мы изменим стартовое окно с хомячком, а также изменим окно, определяющее способ подбора досуга.

# **Гипотеза**

Мы предполагаем, что редизайн увеличит продолжительность времени нахождения в приложении и получения удовлетворения от его использования.

## **Переменные**

1. Независимые переменные(Independent (IV)):
   1. Дизайн стартового окна (оригинальный или обновленный);
   2. Дизайн окна выбора способа подбора досуга (оригинальный или обновленный);
2. Зависимые переменные (Dependent variables (DV)):
   1. Продолжительность пользовательской сессии в приложении;
   2. Рейтинг приложения пользователем.

## **Планирование эксперимента**

Самый оптимальный вариант для сравнения различных дизайнов приложение - A/B-тестирование. Однако так как в нашем приложении мы рассматриваем влияние изменения сразу двух экранов приложения, то нами было выполнено A/B/C/D-тестирование.

В данном случае мы каждому участнику тестирования предложили воспользоваться приложением со случайно предоставленной конфигурацией из экранов приложения: часть пользователей взаимодействовала со старым дизайном приложения, часть работала только с одним измененным экраном, а часть пользовалась приложением с полностью обновленным дизайном.

## **Участники**

Для данного исследования мы попросили 135 человек воспользоваться нашим приложением со своих смартфонов и зафиксировали их время нахождения в приложении и оценку, которую они поставили приложению.

Если время и рейтинг увеличится, значит тест выполнен. Для чистоты эксперимента мы попытались выбрать людей, которым максимально необходимо это приложение. Поэтому для нашей группы были выбраны люди от 14 до 26 лет, для большинства из которых вопросов подбора досуга является крайне актуальным.

## **Обоснование выбора стиля тестирования**

Мы считаем, что редизайн в нашем случаем будет оптимален, так как мы сделали привлекательнее главную страницу приложения, а также более ясно и конкретно поставили сформулировали вопрос о выборе способа подбора досуга, чтобы пользователь мог на более интуитивном уровне пользоваться приложением.

## **Статистические тесты**

В данной работе мы будем сравнивать среднее время пользовательской сессии и рейтинг, полученный от пользователя, на старой и новых версиях приложения. Для проверки наших гипотез мы выполним проверку гипотезы о равенстве математического ожидания времени сессии старой версии и новых версий, а также построим линейную регрессионную модель, чтобы оценить влияние каждого из обновленных экранов по отдельности и их влияние в сумме.

# **Прототип**

В качестве прототипа мы взяли текущую версию приложения. Изменения в дизайне стартовой страницы приложения можно просмотреть на Рисунке 1 и Рисунке 2.

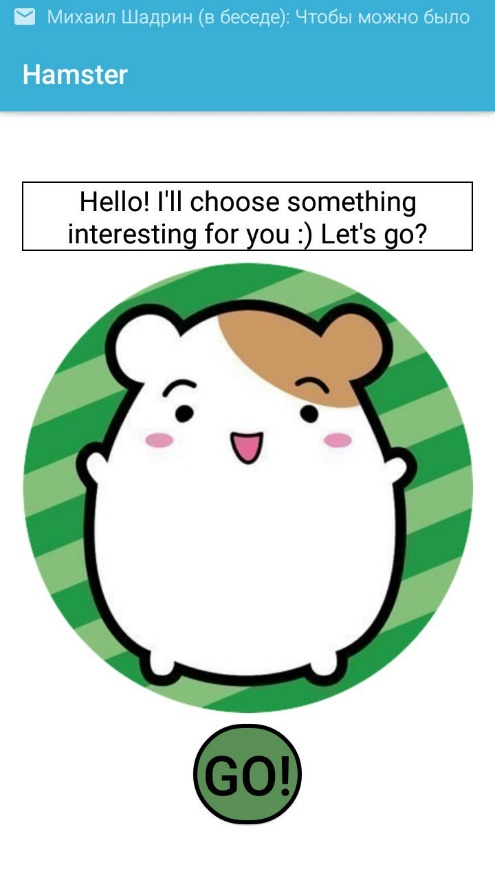


Рис.1. Стартовая страница. Исходный вариант.



Рис.2. Стартовая страница. Новый вариант.

Изменения в дизайне страницы выбора способа подбора досуга можно наблюдать на Рисунке 3 и Рисунке 4.

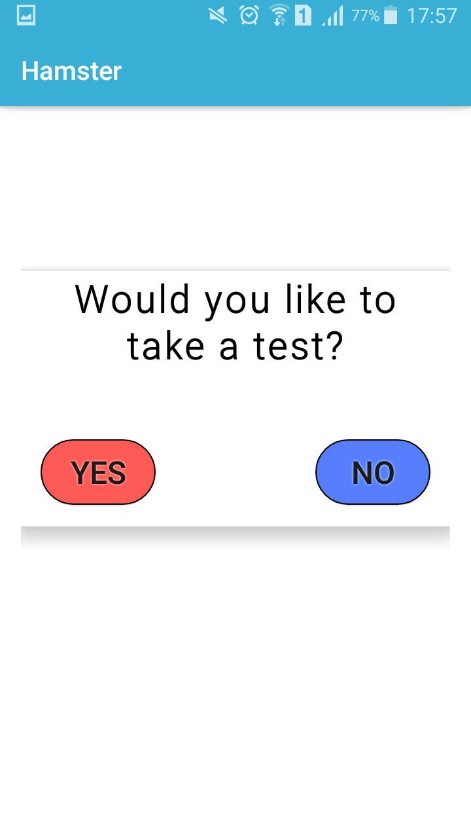


Рис.3. Страница выбора способа подбора досуга. Новый вариант.



Рис.4. Страница выбора способа подбора досуга. Новый вариант.

# **Исследование**

## **Проведение исследования**

После проведения эксперимента мы получили данные, представленные в файле result\_data.csv, где:

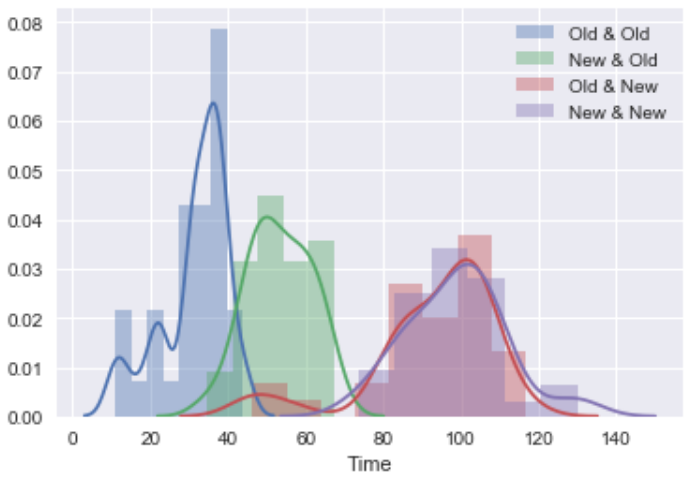
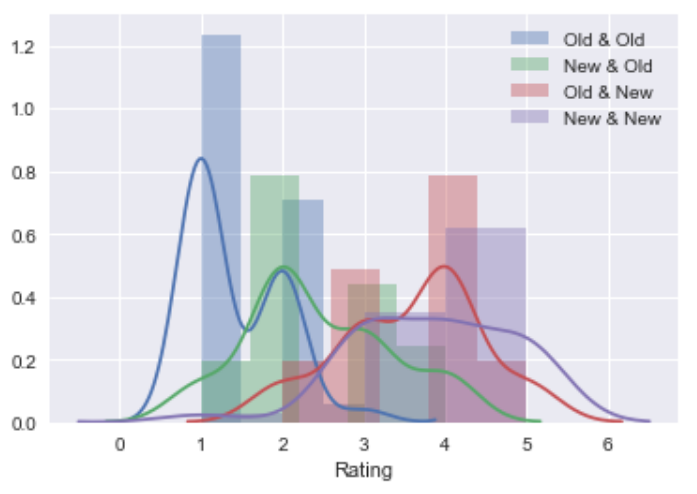
* Time — продолжительность пользовательской сессии (в секундах);
* Rating — оценка пользователем приложения о 5-балльной шкале;
* Start — дамми-переменная дизайна стартового экрана: 0 соответствует исходному дизайну, а 1 — новому;
* Choose — дамми-переменная дизайна экрана выбора способа подбора досуга: 0 соответствует исходному дизайну, а 1 — новому;

Из всех полученных данных мы попытались убрать выбросы, но в данном исследовании их не было обнаружено. На полученных данных были проведены следующие тесты:

1. Рассчитаны описательные статистики для объясняемых переменных;
2. Построены гистограммы и box-plot графики для переменных Time и Rating для каждой из конфигурации экранов;
3. Рассчитана корреляция между временем пользовательской сессии, пользовательским рейтингом приложения и конфигурацией экранов приложения;
4. Проведены тесты Стьюдента между выборкой, соответствующей исходному дизайну, и выборками с обновленным дизайном;
5. Построена линейная регрессия между временем пользовательской сессии и конфигурацией экранов приложения.

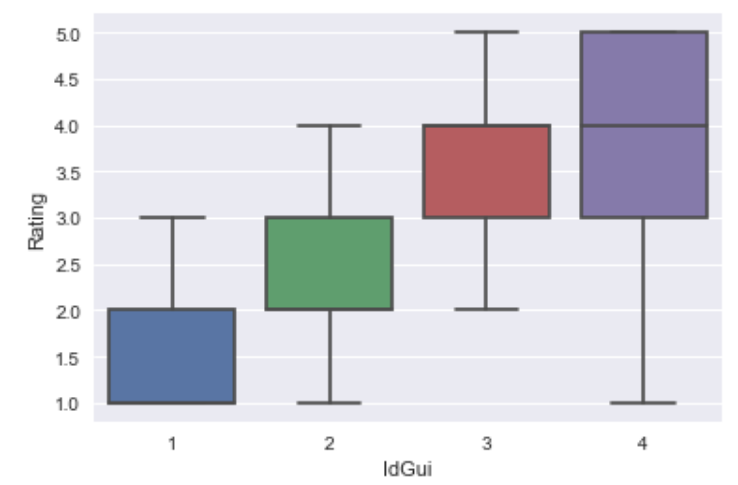
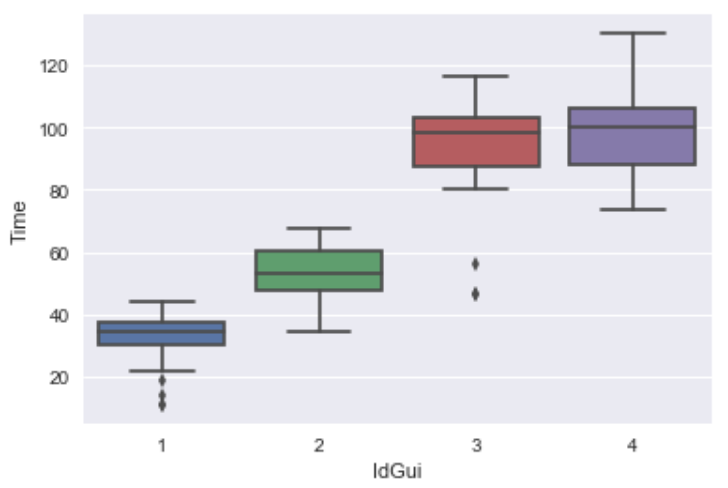
# **Результаты исследования**

Как видно из рисунков 5 и 6, уже хотя бы при одном из изменений наблюдается заметное увеличение продолжительности пользовательской сессии, равно как и рост пользовательских оценок.

*Рис.5, 6. Гистограммы времени сессии и пользовательского рейтинга*

Однако среди данных о времени сессии присутствуют выбросы. Рассмотрим полученные значения на box-plot’ах, чтобы убедиться в этом.



*Рис.7,8. Box-plot графики времени сессии и пользовательского рейтинга*

Действительно, в полученных выборках были обнаружены выбросы от которых нам пришлось избавиться для дальнейшего корректного анализа.

Затем мы исследовали описательные статистика переменных, рассчитав их среднее значение и стандартное отклонение. Полученные значения отображены в Таблице 1.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Среднее значение | | | | Дисперсия | | | |
| Old & Old | New & Old | Old & New | New & New | Old & Old | New & Old | Old & New | New & New |
| Time | 31.472059 | 53.285735 | 92.840706 | 98.827618 | 8.637791 | 8.076869 | 16.384042 | 12.673557 |
| Rating | 1.411765 | 2.441176 | 3.588235 | 3.852941 | 0.556920 | 0.894128 | 0.856973 | 0.957660 |

*Таблица 1. Среднее значение и стандартное отклоение времени сессии и рейтинга пользователя.*

Невооруженным взгядом уже заметна значительная разница в средних показателях в зависимости от того, какие экраны демонстрируются пользователю. Тем не менее, для того чтобы убедиться в правоте нашей гипотезы, мы провели три t-теста. В качестве нулевой гипотезы мы предположили, что математическое ожидание времени пользовательской сессии в приложении с исходным дизайном не отличается от времени пользовательской сессии в приложении с изменным дизайном. Как альтернативную гипотезу мы выдвинули предположение, что матожидание времени сессии все-таки меньше в случае с исходным дизайном. Все гипотезы проверялись на уровне значимости 1%.

Для проверки гипотез использовался метод ttest\_ind для двустороннего t-теста из библиотеки SciPy. То, что данный метод предназначен для двустороннего тестирования нисколько не останавливает нас в его использовании, так как рассчитанное p-значение может быть использовано для получения p-значения результата одностороннего теста.

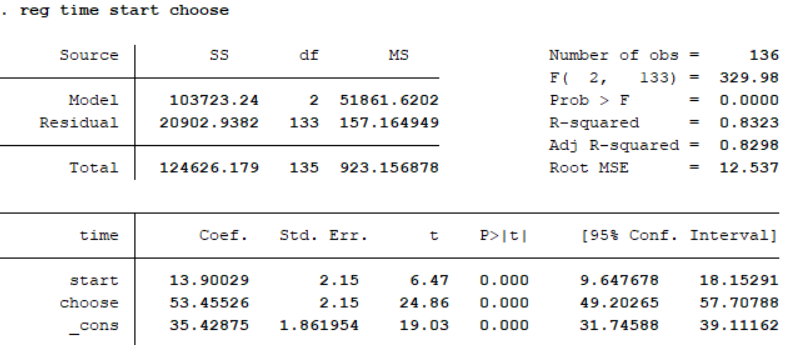
В ходе проведения теста мы получили следующие значения, с которыми можно ознакомиться в Таблице 2.

|  |  |  |  |
| --- | --- | --- | --- |
| H0 | M(Old & Old) = M(New & Old) | M(Old & Old) = M(New & Old) | M(Old & Old) = M(New & New) |
| H1 | M(Old & Old) < M(New & Old) | M(Old & Old) < M(New & Old) | M(Old & Old) < M(New & New) |
| α | 0,01 | | |
| p-value | 7.540846906526524e-25 | 3.954106151999242e-16 | 2.2240787076754402e-33 |
| t-статистика | -19.320056944483294 | -10.755766517271494 | -25.607436609652122 |
| p-value / 2 && t < 0 | Да | Да | Да |
| Вывод | Отвергаем H0 в пользу H1 | Отвергаем H0 в пользу H1 | Отвергаем H0 в пользу H1 |

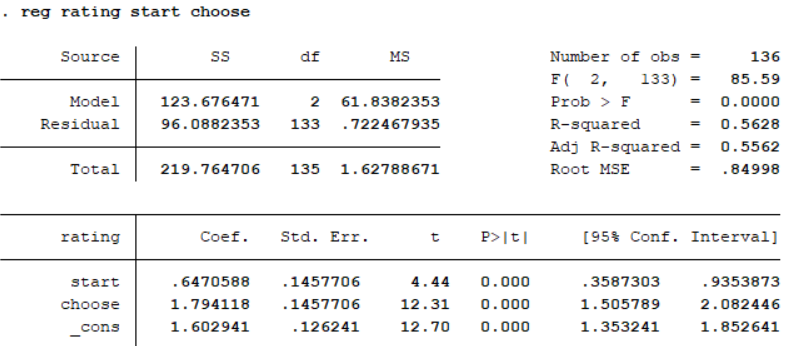
*Таблица 2. Результаты t-теста*

Таким образом, мы убедились, что внесенные нами изменения, действительно, значительно повлияли на продолжительность пользовательской сессии.

Однако будет крайне интересно узнать, как каждый из критериев повлиял на продолжительность сессии и пользовательский рейтинг. Для этого мы воспользовались инстурментом для аналитики Stata и посторили регрессионные модели (Рисунок 9, 10)

**

*Рисунок 9. Регрессионная модель time ~ start + choose*

**

*Рисунок 10. Регрессионная модель rating ~ start + choose*

Обе регрессионные модели проходят F-тест на адекватность регрессии в целом на уровне значимости 1%, а также t-тест на значимость каждого коэффициента регресси по отдельности на том же уровне значимости.

По результатам выполненного регрессионного анализа можно заметить, что значительно большее влияние на восприятие приложения вносит конкретизация инструкции для пользователя, нежели изменения в дизайне стартового окна.

Рассчитанные коэффициенты корреляции Спирмена так же демонстрируют значительно более ярко выраженную зависимость между дизайном экрана выбора способа подбора досуга и продолжительностью сессии и рейтингом, чем дизайн стартового экрана (Рисунок 11).

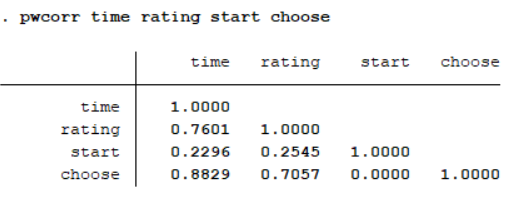


Рисунок 11. Коэффициенты корреляции Спирмена

1. **Заключение**

В данном исследовании удалось выявить взаимосвязь между изменениями в дизайне приложения и продолжительностью пользовательских сессий и рейтингом приложения от пользователей.

Можно предположить, что приведенные нами изменения На дальнейшем восприятии приложения и помочь выйти на тот уровень дизайна интерфейса, который позволит с удобством пользоваться данным продуктом.