EMPIRICAL STUDY

CS346: SOFTWARE ENGINEERING LABORATORY

Group 11
Project 7: Paint Application

Siddhartha Jain 180101078
Varhade Amey Anant 180101087
Pulkit Changoiwala 180101093
Samay Varshney 180101097

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati

Contents

1 Research Question Formulation **3**
1.1 Questions . . . 3

2 Identification of Variables **4**
2.1 RQ1 . . . 4
2.2 RQ2 . . . 5

3 Experimental Setup **7**
3.1 Detemining Levels For The Factors: . . . 7
3.1.1 RQ1 . . . 7
3.1.2 RQ2 . . . 8
3.2 User Profile . . . 9
3.3 Number Of Participants . . . 10
3.3.1 Pilot Study: . . . 10
3.3.2 Empirical Data: . . . 10
3.4 Task . . . 10
3.4.1 RQ1: . . . 10
3.4.2 RQ2: . . . 10

4 Empirical Data **12**
4.1 RQ1 . . . 12
4.2 RQ2 . . . 14

5 Analysis of Empirical Data **16**

1 Research Question Formulation

1.1 Questions

RQ1. How does the ease of drawing (on a scale of 1-10) depend upon the number of objects, object types and size of the toolbox?

Why is it testable?

- **Lack of Ambiguity:** We are specifying the usability score using a numerical scale i.e. quantitative instead of leaving it completely qualitative. Moreover, we have broken down the usability into the number of image operations present on the screen, the number of types of operations, and the size of the image. This helps the user to make a better decision while choosing the usability score.
- **Measurable Quantities:** We cannot measure the usability of drawing and for that, we are measuring it between 1-10 scale. Also, it is possible to measure the number of objects, their types, number of operations, and image size.

RQ2. How do the usability score (on a scale of 1-10) depend on the number of image operations, their type and the size of the image on an interface?

Why is it testable?

- **Lack of Ambiguity:** We are specifying the usability score using a numerical scale i.e. quantitative instead of leaving it completely qualitative. Moreover, we have broken down the usability into the number of image operations present on the screen, the number of types of operations, and the size of the image. This helps the user to make a better decision while choosing the usability score.
- **Measurable Quantities:** We cannot measure the usability of drawing and for that, we are measuring it between 1-10 scale. Also, it is possible to measure the number of objects, their types, number of operations, and image size.

2 Identification of Variables

2.1 RQ1

Independent Variables:

- **Number of objects in toolbox:** It can take any integer. We use nominal scale for measurement by assigning (arbitrarily) an integer (one) to each object and then add them up.

- **Object types in the toolbox:** It can take- “Image only”, “text and image”. We use nominal scale by assigning value 1 to “Image only”, 2 to “text and image”.

- **Toolbox size:** It can take any ratio value ranging between 0 and 1. We assign a ratio scale to this variable: ratio of the area covered by the tool box and area of the screen.

Dependent Variables:

- **Ease of drawing:** It can take a value between 1 to 10 (integers only). We use a likert-type scale, which is a ordinal scale, to record ratings.

Control Variables:

- **Drawing Board:** Drawing board color might affect the ease of drawing but we don’t want to know the relationship between drawing board and ease of drawing. Thus, we are making it a control variable.

- **Age Of The User:** Age of the user might affect the ease of drawing but we don’t want to know the relationship between age and ease of drawing. Thus, we are making it a control variable.

- **Skill Of User:** User’s past experience with drawing and his/her drawing skill might affect the results of empirical research. But for our research study we are not interested in its relationship with ease of drawing thus, we are taking it as a control variable.

- **Reference Image:** Some images will be easy to draw while others are not but we don’t want to know the relationship between image and ease of drawing.Thus, we are making it a control variable keeping an image fixed.

- **Mobile Screen Brightness:** Mobile Screen Brightness: Brightness can affect the ease of drawing, as with more brightness users might be able to do better as compared to low brightness. Thus, we are making it a control variable as we are not interested in the relationship of brightness with ease of drawing.

Confounding Variables:

- **Practice/learning effect:** When participants rate multiple interfaces/objects of similar kind then participants start gaining experience and grow some expectations about the nature of the objects. Experience and expectations get amplified with each trial thus it might potentially affect the way participants rate the interface/object.

2.2 RQ2

Independent Variables:

- **The number of image operations:** It can take any integer value. We use the nominal scale for measurement by assigning (arbitrarily) an integer (one) to each object and then add them up.
- **Type of operations:** It can take - "Image only", "text and image". We use nominal scale by assigning value 1 to "Image only", 2 to "text and image".
- **Image size:** It can take any ratio value ranging between 0 and 1. We assign a ratio scale to this variable: ratio of the area covered by image and area of the drawing board.

Dependent Variables:

- **Usability score:** It can take a value between 1 to 10 (integers only). We use a likert-type scale, which is a ordinal scale, to record ratings.

Control Variables:

- **Drawing Board:** Drawing board color might affect the ease of drawing but we don't want to know the relationship between drawing board and ease of drawing. Thus, we are making it a control variable.
- **Age Of The User:** Age of the user might affect the ease of drawing but we don't want to know the relationship between age and ease of drawing. Thus, we are making it a control variable.
- **Mobile Screen Brightness:** Mobile Screen Brightness: Brightness can affect the ease of drawing, as with more brightness users might be able to do better as compared to low brightness. Thus, we are making it a control variable as we are not interested in the relationship of brightness with ease of drawing.

Confounding Variables:

- **Practice/learning effect:** When participants rate multiple interfaces/objects of similar kind then participants start gaining experience and grow some expectations about the nature of the objects. Experience and expectations get amplified with each trial thus it might potentially affect the way participants rate the interface/object.

3 Experimental Setup

3.1 Detemining Levels For The Factors:

3.1.1 RQ1

Independent Variables

- **Number of objects in toolbox:** We have performed Pilot Study for this variable.
  Pilot Study:
  We survey popular interfaces - Evernote, Windows Paint, Google Jamboard, and record the number of objects* in those and take average of them.
  - Evernote: 11
  - Windows Paint: 13
  - Google Jamboard: 9

  Average: $(11 + 13 + 9)/3 = 11$
  **The features/number of objects are selected and counted considering the age group of users.*
  Average will serve as threshold - any value of the variable below 11 indicates a toolbox with less objects and value above denotes a toolbox with more objects. Therefore, the decided levels for the factor (number of objects) are: *9, 11 and 13* considering the result of the Pilot Study.

- **Object types in the toolbox:** We have 2 object types, image only and text image, so they are the levels for the object type factor.

- **Toolbox size:** We have performed Pilot Study for this variable.
  Pilot Study:
  We survey popular interfaces - Evernote, Windows Paint, Google Jamboard, and record the toolbox size* (in terms of area ratio) in those and take the average of them.
  - Evernote: 2
  - Windows Paint: 3
  - Google Jamboard: 1

  Average: $(2 + 3 + 1)/3 = 2$
  **We scaled the drawing board area to a set reference value and calculated the area for the toolbox using, ratio*(reference area). Reference area was taken to be 10.*
  Average will serve as threshold - any value of the variable below 2 indicates a

toolbox with less size and value above denotes a toolbox with more size. Therefore, the decided levels for the factor are: *1, 2 and 3* considering the result of the Pilot Study.

Therefore, we have altogether 3 x 2 x 3 = 18 test conditions for our study.

Control Variables

- **Drawing Board:** We use white colour of the drawing board as reference.
- **Age of the User:** All the users we use for our data collections are of age 7.
- **Skill of the User:** All the users must have atleast 1 and atmost 2 months of experience in using digital apps.
- **Reference Image:** We fix the image to draw for all the test conditions.
- **Mobile Screen Brightness:** We fix the brightness of the screen to 50

3.1.2 RQ2

Independent Variables

- **The number of image operations:** We have performed Pilot Study for this variable.
  Pilot Study:
  We survey popular Mobile interfaces of Gallery - Samsung, iPhone, LG and record the number of image operations in those and take average of them.
  - Samsung: 5
  - iPhone: 4
  - LG: 4

  Average: floor((5 + 4 + 4)/3) = 4
  Average will serve as threshold - any value of the variable below 4 indicates a gallery with less operations on image and value above denotes a gallery with more operations on image. Therefore, the decided levels for the factor are: *3, 4 and 5* considering the result of the Pilot Study.
- **Types of operations:** We have 2 operation types, image only and text image, so they are the levels for the operation type factor.
- **Image size:** We have performed Pilot Study for this variable.
  Pilot Study:
  We survey popular Mobile interfaces of Gallery - Samsung, iPhone, LG and record the image size* (in terms of area ratio) operations in those and take average of them.
  - Samsung: 8

– iPhone: 8

– LG: 10

Average: $(8 + 8 + 10)/3 = 8$
**We scaled the screen area to a set reference value and calculated the area for the image using, ratio*(reference area). Reference area was taken to be 10.*
Average will serve as threshold - any value of the variable below 8 indicates an image with smaller size on image and value above denotes image with larger size. Therefore, the decided levels for the factor are: *7, 8 and 9* considering the result of the Pilot Study.

Therefore, we have altogether 3 x 2 x 3 = 18 test conditions for our study.

Control Variables

- **Drawing Board:** We use white colour of the drawing board as reference.
- **Age of the User:** All the users we use for our data collections are of age 7.
- **Mobile Screen Brightness:** We fix the brightness of the screen to 50

3.2 User Profile

We have developed our paint app for children of the age group 5 to 9. Considering that we are looking for following user profile for our participants.

- **Age:** We will take 7 years old users as our participants. As it lies in the average age group of our target age group.
- **Goals:** We are looking for participants which are interested in drawing or motivated to learn paint and drawing.
- **Needs:** We are taking those participants who are looking for a digital app for painting. Also, users which want a game based learning method to learn new skills.
- **Skill Set:** We have participants with average skill set for our study. As it ensures that our app caters to the average audience, which by far constitutes the larger segment compared to its extreme counterparts.
- **Other Factors:** We are having participants who have access to a smartphone(their own or their parent's device). Participants with healthy condition: No color blindness and no disability in using hands.

  The above described user profile represents our target audience, thus doing empirical study for the participants with above user profile will ensure that we are modeling behaviour of our whole audience.

3.3 Number Of Participants

3.3.1 Pilot Study:

For the pilot study of determining the number of objects and size of the toolbox, we selected 5 participants who have experience in all the three selected interfaces - Evernote, Windows Paint, and Google Jamboard.
For the pilot study of determining the number of image operations in the gallery, we selected 5 participants who have experience in all the three selected popular gallery interfaces of the mobile - Samsung, iPhone, and LG. Instead of selecting multiple brands of a single company, we decided to choose different company models to have a more reliable conclusion for our study.

3.3.2 Empirical Data:

To draw reliable conclusions from empirical data and considering the number of test conditions of our research question, we use 15 participants.

3.4 Task

3.4.1 RQ1:

- **Task Description:** Task for RQ1 is to rate how easy it is to draw on a scale of 1-10.
- **Assignment of Task:** ***within-subject method*** We use the "within-subject" method for the assignment of tasks to the participants because we have 18 test conditions in total and it might not be troublesome for the participant to rate all the test conditions, In this method, we ask each participant to rate all test conditions.
  For the assignment of the task, we have in total 15 participants. We give each participant all 18 conditions to rate.

3.4.2 RQ2:

- **Task Description:** Task for RQ2 is to rate how usable the interface of the gallery is on a scale of 1-10.
- **Assignment of Task:** ***within-subject method*** We use the "within-subject" method for the assignment of tasks to the participants because we have 18 test conditions in total and it might not be troublesome for the participant to rate all the test conditions, In this method, we ask each participant to rate all test conditions.
  For the assignment of the task, we have in total 15 participants. We give each participant all 18 conditions to rate.

Avoiding Practice Effect: Now we assign tasks using Latin Square Method: Each user will get a different sequence.
To avoid the influence of practice effect in each group of between-subject method, we counterbalance sequence in which tasks are given to the participants. Task sequence for each participant is different from the other participants. The most systematic method to accomplish it is to use a Latin Square Method. In this method, we organize a sequence of tasks given to the participants in the form of a square matrix, with the condition that each task occurs only once in each row and column. This makes sure that the task sequence for each participant is different from the other participants and nullifies the practice effect on our collected empirical data.

4 Empirical Data

4.1 RQ1

No	objects	type	size
1	9	image	1
2	9	image	2
3	9	image	3
4	9	text and image	1
5	9	text and image	2
6	9	text and image	3
7	11	image	1
8	11	image	2
9	11	image	3
10	11	text and image	1
11	11	text and image	2
12	11	text and image	3
13	13	image	1
14	13	image	2
15	13	image	3
16	13	text and image	1
17	13	text and image	2
18	13	text and image	3

Rating															
No	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**
1	8	7	7	9	7	6	7	7	9	9	7	6	7	6	7
2	8	5	6	8	8	6	6	5	5	8	8	5	5	5	6
3	9	6	6	5	8	8	8	5	8	6	6	6	6	8	8
4	9	10	7	8	8	7	8	8	10	7	8	8	7	9	10
5	9	9	8	8	8	10	10	9	10	8	9	8	10	8	9
6	7	8	8	8	8	5	8	5	5	8	5	6	8	8	5
7	8	5	8	8	5	5	6	8	8	5	8	6	6	6	8
8	8	9	6	9	7	6	7	6	9	6	6	9	9	7	6
9	7	6	5	5	6	8	5	6	5	6	8	8	6	8	5
10	8	9	6	9	7	9	7	6	9	9	6	7	6	6	6
11	8	7	9	6	6	9	6	9	9	7	6	7	9	9	6
12	9	5	6	5	8	6	5	5	6	5	5	6	5	6	6
13	6	4	5	7	5	4	4	7	7	4	5	7	5	5	4
14	5	3	4	3	3	6	4	3	4	3	4	6	4	3	3
15	5	6	6	6	6	3	4	4	4	3	3	3	4	4	4
16	7	8	8	8	8	5	8	5	8	8	5	5	5	6	5
17	6	5	7	4	4	4	5	7	4	7	4	5	7	7	5
18	6	4	4	4	7	4	7	5	4	5	5	4	5	7	5

4.2 RQ2

No	Image operations	Type of operations	Image size
1	3	image	7
2	3	image	8
3	3	image	9
4	3	text and image	7
5	3	text and image	8
6	3	text and image	9
7	4	image	7
8	4	image	8
9	4	image	9
10	4	text and image	7
11	4	text and image	8
12	4	text and image	9
13	5	image	7
14	5	image	8
15	5	image	9
16	5	text and image	7
17	5	text and image	8
18	5	text and image	9

Rating															
No	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**P15**
1	7	6	6	8	6	5	6	6	8	8	6	5	6	5	6
2	7	5	6	8	8	6	6	5	5	8	8	5	5	5	6
3	8	7	7	6	9	9	9	6	9	7	7	7	7	9	9
4	8	9	6	6	7	6	6	6	9	6	7	7	6	6	9
5	9	7	7	8	8	10	10	7	10	7	7	8	10	8	7
6	10	10	10	10	10	8	10	8	8	10	8	9	10	10	8
7	6	4	7	7	4	4	5	7	7	4	7	5	5	5	7
8	6	7	4	7	5	4	5	4	7	4	4	7	7	5	4
9	7	6	5	5	6	8	5	6	5	6	8	8	6	8	5
10	8	9	6	9	7	9	7	6	9	9	6	7	6	6	6
11	9	8	10	7	7	10	7	10	10	8	7	8	10	10	7
12	9	7	8	7	10	8	7	7	8	7	7	8	7	8	8
13	7	5	6	8	6	5	5	8	8	5	6	8	6	6	5
14	7	5	6	5	5	8	6	5	6	5	6	8	6	5	5
15	8	9	9	9	9	6	7	7	7	6	6	6	7	7	7
16	7	8	8	8	8	5	8	5	8	8	5	5	5	6	5
17	7	6	8	5	5	5	6	8	5	8	5	6	8	8	6
18	8	6	6	6	9	6	9	7	6	7	7	6	7	9	7

5 Analysis of Empirical Data

RQ1:

Formation of Hypothesis:

H0(null hypothesis): The ease of drawing does not depend on the number of objects, object types and size of the toolbox.
H1(alternative hypothesis): The ease of drawing depends on the number of objects, object types and size of the toolbox.
Statistical Significance Test (Hypothesis Test): This test is used to ascertain the reliability of the empirical data. Condition 2 of the parametric test fails as we use a likert-type scale, which is a ordinal scale, instead of an interval scale for measurement of dependent variable, so we use non-parametric test for testing.
We use within-subject design with 3 factors, each having two or more levels. So, the non-parametric test satisfying this condition is the **Friedman Test**.

Python Code For Friedman Test

```
from scipy.stats import friedmanchisquare

# RQ1
# enter the data
data1 = [8 , 7 , 7 , 9 , 7 , 6 , 7 , 7 , 9 , 9 , 7 , 6 , 7 , 6 , 7 ]
data2 = [8 , 5 , 6 , 8 , 8 , 6 , 6 , 5 , 5 , 8 , 8 , 5 , 5 , 5 , 6 ]
data3 = [9 , 6 , 6 , 5 , 8 , 8 , 8 , 5 , 8 , 6 , 6 , 6 , 6 , 8 , 8 ]
data4 = [9 , 10 , 7 , 8 , 8 , 7 , 8 , 8 , 10 , 7 , 8 , 8 , 7 , 9 , 10 ]
data5 = [9 , 9 , 8 , 8 , 8 , 10 , 10 , 9 , 10 , 8 , 9 , 8 , 10 , 8 , 9 ]
data6 = [7 , 8 , 8 , 8 , 8 , 5 , 8 , 5 , 5 , 8 , 5 , 6 , 8 , 8 , 5 ]
data7 = [8 , 5 , 8 , 8 , 5 , 5 , 6 , 8 , 8 , 5 , 8 , 6 , 6 , 6 , 8 ]
data8 = [8 , 9 , 6 , 9 , 7 , 6 , 7 , 6 , 9 , 6 , 6 , 9 , 9 , 7 , 6 ]
data9 = [7 , 6 , 5 , 5 , 6 , 8 , 5 , 6 , 5 , 6 , 8 , 8 , 6 , 8 , 5 ]
data10 = [8 , 9 , 6 , 9 , 7 , 9 , 7 , 6 , 9 , 9 , 6 , 7 , 6 , 6 , 6 ]
data11 = [8 , 7 , 9 , 6 , 6 , 9 , 6 , 9 , 9 , 7 , 6 , 7 , 9 , 9 , 6 ]
data12 = [9 , 5 , 6 , 5 , 8 , 6 , 5 , 5 , 6 , 5 , 5 , 6 , 5 , 6 , 6 ]
data13 = [6 , 4 , 5 , 7 , 5 , 4 , 4 , 7 , 7 , 4 , 5 , 7 , 5 , 5 , 4 ]
data14 = [5 , 3 , 4 , 3 , 3 , 6 , 4 , 3 , 4 , 3 , 4 , 6 , 4 , 3 , 3 ]
data15 = [5 , 6 , 6 , 6 , 6 , 3 , 4 , 4 , 4 , 3 , 3 , 3 , 4 , 4 , 4 ]
data16 = [7 , 8 , 8 , 8 , 8 , 5 , 8 , 5 , 8 , 8 , 5 , 5 , 5 , 6 , 5 ]
data17 = [6 , 5 , 7 , 4 , 4 , 4 , 5 , 7 , 4 , 7 , 4 , 5 , 7 , 7 , 5 ]
data18 = [6 , 4 , 4 , 4 , 7 , 4 , 7 , 5 , 4 , 5 , 5 , 4 , 5 , 7 , 5 ]

# perform the Friedman Test
stat, pvalue = friedmanchisquare(data1, data2, data3, data4, data5, data6,
                         data7, data8, data9, data10,data11, data12,
                         data13, data14, data15, data16, data17, data18)

print('Statistics = %.3f, pvalue = %.30f' % (stat, pvalue))

# interpret the result
alpha = 0.05
if pvalue > alpha:
   print('Same distributions (fail to reject H0)')
else:
   print('Different distributions (reject H0)')
```

Result:
Statistics = 136.536
pvalue = 0.000000000000000000010251012064
Different distributions (reject H0)

The value of test statistics is 136.536 and corresponding p-value is less than 0.05, so we can reject the null hypothesis. In other words, we have sufficient evidence to support the alternative hypothesis.

RQ2:

Formation of Hypothesis:

H0(null hypothesis): The usability score does not depend on the number of image operations, their type, and the size of the image on an interface.
H1(alternative hypothesis): The usability score depends on the number of image operations, their type, and the size of the image on an interface.

Statistical Significance Test(Hypothesis Test): This test is used to ascertain the reliability of the empirical data. Condition 2 of the parametric test fails as we use a likert-type scale, which is a ordinal scale, instead of an interval scale for measurement of dependent variable, so we use non-parametric test for testing.
We use within-subject design with 3 factors, each having two or more levels. So, the non-parametric test satisfying this condition is the **Friedman Test**.

Python Code For Friedman Test

```
from scipy.stats import friedmanchisquare

# RQ2
data1 = [7 , 6 , 6 , 8 , 6 , 5 , 6 , 6 , 8 , 8 , 6 , 5 , 6 , 5 , 6 ]
data2 = [7 , 5 , 6 , 8 , 8 , 6 , 6 , 5 , 5 , 8 , 8 , 5 , 5 , 5 , 6 ]
data3 = [8 , 7 , 7 , 6 , 9 , 9 , 9 , 6 , 9 , 7 , 7 , 7 , 7 , 9 , 9 ]
data4 = [8 , 9 , 6 , 6 , 7 , 6 , 6 , 6 , 9 , 6 , 7 , 7 , 6 , 6 , 9 ]
data5 = [9 , 7 , 7 , 8 , 8 , 10 , 10 , 7 , 10 , 7 , 7 , 8 , 10 , 8 , 7 ]
data6 = [10 , 10, 10 , 10 , 10 , 8 , 10 , 8 , 8 , 10 , 8 , 9 , 10 , 10 , 8]
data7 = [6 , 4 , 7 , 7 , 4 , 4 , 5 , 7 , 7 , 4 , 7 , 5 , 5 , 5 , 7 ]
data8 = [6 , 7 , 4 , 7 , 5 , 4 , 5 , 4 , 7 , 4 , 4 , 7 , 7 , 5 , 4 ]
data9 = [7 , 6 , 5 , 5 , 6 , 8 , 5 , 6 , 5 , 6 , 8 , 8 , 6 , 8 , 5 ]
data10 = [8 , 9 , 6 , 9 , 7 , 9 , 7 , 6 , 9 , 9 , 6 , 7 , 6 , 6 , 6 ]
data11 = [9 , 8 , 10 , 7 , 7 , 10 , 7 , 10 , 10 , 8 , 7 , 8 , 10 , 10 , 7 ]
data12 = [9 , 7 , 8 , 7 , 10 , 8 , 7 , 7 , 8 , 7 , 7 , 8 , 7 , 8 , 8 ]
data13 = [7 , 5 , 6 , 8 , 6 , 5 , 5 , 8 , 8 , 5 , 6 , 8 , 6 , 6 , 5 ]
data14 = [7 , 5 , 6 , 5 , 5 , 8 , 6 , 5 , 6 , 5 , 6 , 8 , 6 , 5 , 5 ]
data15 = [8 , 9 , 9 , 9 , 9 , 6 , 7 , 7 , 7 , 6 , 6 , 6 , 7 , 7 , 7 ]
data16 = [7 , 8 , 8 , 8 , 8 , 5 , 8 , 5 , 8 , 8 , 5 , 5 , 5 , 6 , 5 ]
data17 = [7 , 6 , 8 , 5 , 5 , 5 , 6 , 8 , 5 , 8 , 5 , 6 , 8 , 8 , 6 ]
data18 = [8 , 6 , 6 , 6 , 9 , 6 , 9 , 7 , 6 , 7 , 7 , 6 , 7 , 9 , 7 ]

# perform the Friedman Test
stat, pvalue = friedmanchisquare(data1, data2, data3, data4, data5, data6,
                         data7, data8, data9, data10,data11, data12,
                         data13, data14, data15, data16, data17, data18)

print('Statistics = %.3f, pvalue = %.30f' % (stat, pvalue))

# interpret the result
alpha = 0.05
if pvalue > alpha:
    print('Same distributions (fail to reject H0)')
else:
    print('Different distributions (reject H0)')
```

Result:
Statistics = 110.560
pvalue = 0.000000000000000094662258640761
Different distributions(reject H0)

The value of test statistics is 1110.560 and corresponding p-value is less than 0.05, so we can reject the null hypothesis. In other words, we have sufficient evidence to support the alternative hypothesis.