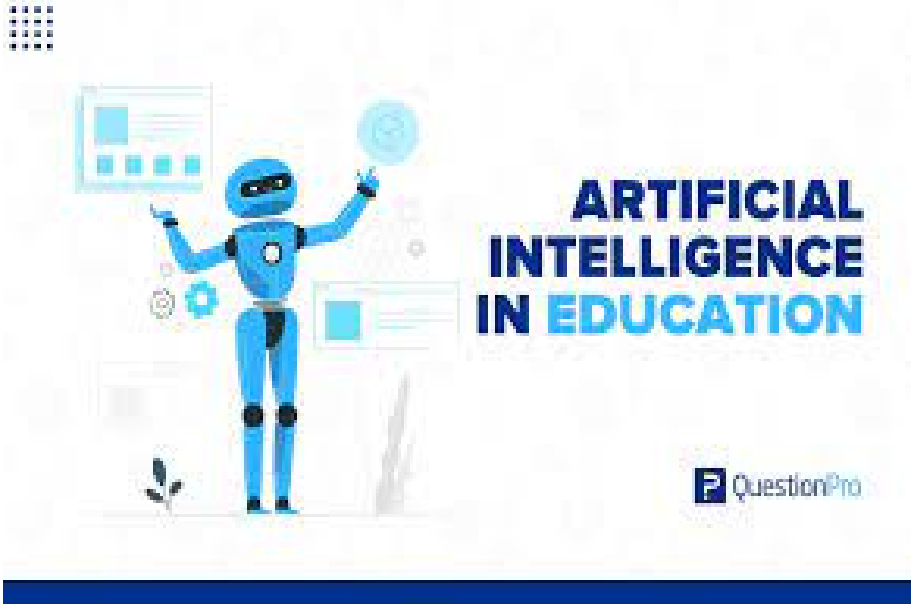
University of Toronto Mississauga

STA304: Technical Report

Shivank Goel, Freya Shah, Xavier Ibanez-Padron, Niharika Patibandla, Deepika Mandapalli, Hasan Nasir

Dated: 14th November 2023



Topic: Analysing the impact and outcomes of integration of Artificial Intelligence in student's learning

Introduction

As AI tools have revolutionized the educative process, understanding the intricate ways in which they affect learning is of utmost importance. From self-learning and proof checking all the way to cheating and plagiarism, knowing the variety of ways these new tools are utilized is necessary to maintain the integrity of all academia. Our part in helping answer this crucial question is to ask this smaller question:

> Does there exist a positive relationship between the usage of AI tools and the perception of said tools within the STA304 student populace?

In the context of this question, our hypotheses are:

(Null Hypothesis) $H_0 : p = 0$

(Alternative Hypothesis) $H_A : p > 0$

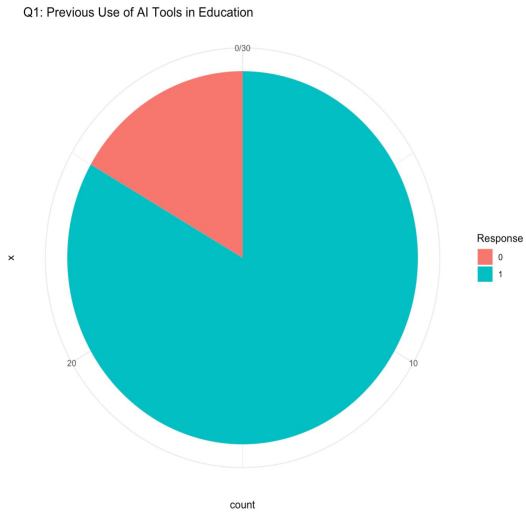
Where p represents the proportion of students with a positive perception of AI based on their AI usage.

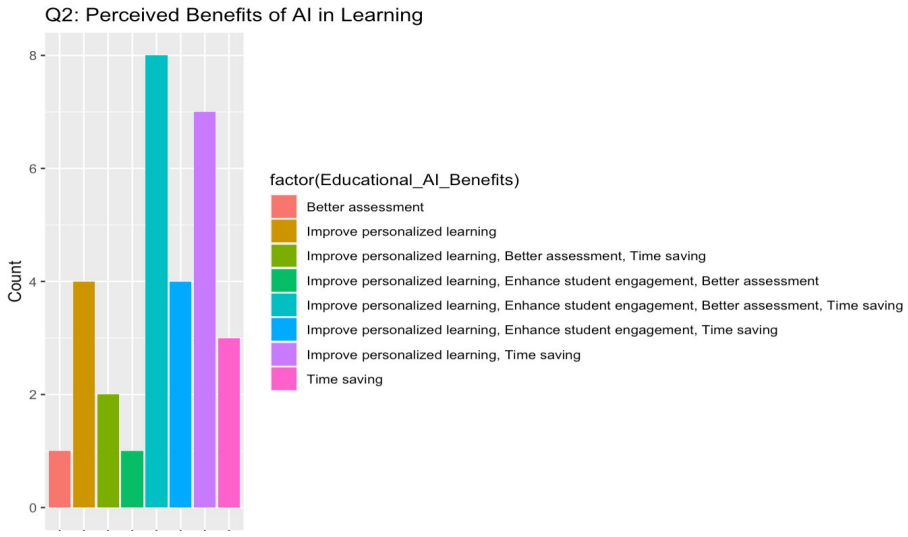
Methodology

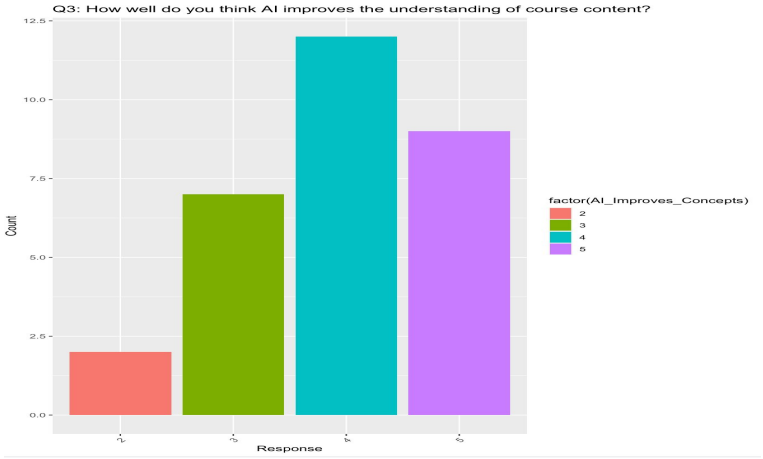
During the 2023 Fall semester, an online questionnaire was distributed to students enrolled in STA304. It utilized simple random sampling (SRS) to sample from the aforementioned population ($N = 200$), exemplified by the random selection of participants with the assurance of proper population representation. This included methods such as randomly selecting lecture participants based on seating, alongside an anonymous post on the official course Piazza. The final response count was ($n = 30$), representing our sample size. The questionnaire itself was composed of 7 short answer questions concerning basic opinions on AI tools such as average AI usage for studying, perceived accuracy of AI tools, and methods of AI utilization for learning.

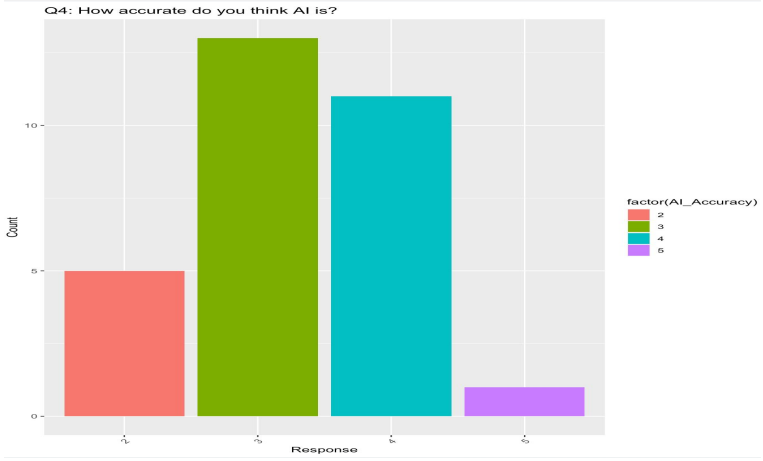
Analysis

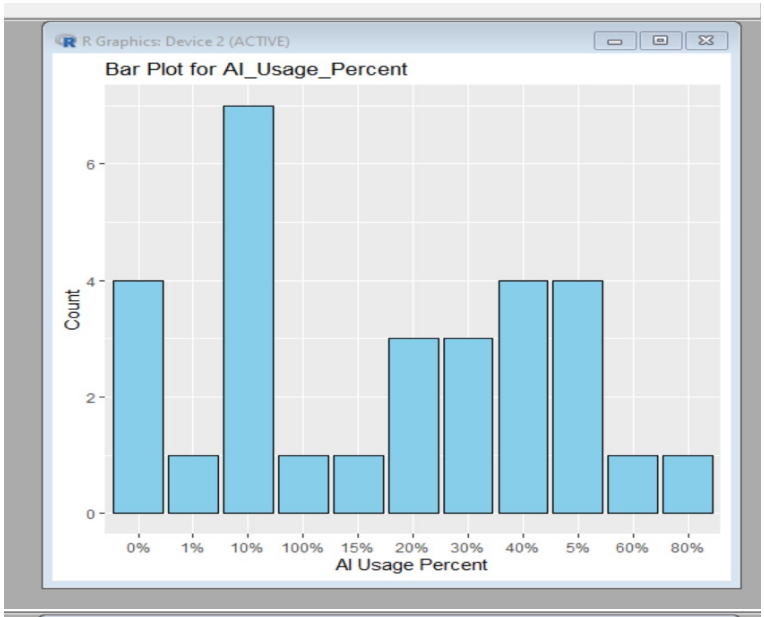
Provided below are the relevant chart/plots for the results of our questionnaire.

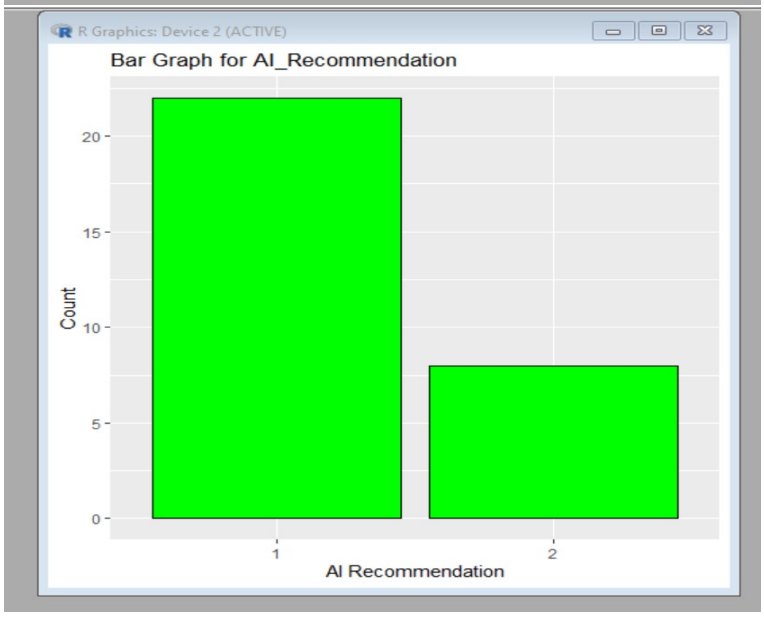





Q3: How well do you think AI improves the understanding of course content?
Count
Response
factor(AI_Improves_Concepts)
2
3
4
5


Q4: How accurate do you think AI is?
Count
Response
factor(AI_Accuracy)
2
3
4
5


R Graphics: Device 2 (ACTIVE)
Bar Plot for AI_Usage_Percent
Count
0
2
4
6
0%
1%
10%
100%
15%
20%
30%
40%
5%
60%
80%
AI Usage Percent


R Graphics: Device 2 (ACTIVE)
Bar Graph for AI_Recommendation
Count
0
5
10
15
20
1
2
AI Recommendation

For our study, our student population has a size of $N = 200$. As we already have a previously established sample size of $n = 30$, we will be using this value to calculate the ultimate bound on the error of estimation.

As the sampling method utilized was simple random sampling, to calculate the bound we will be using the following formula:

$$B = 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(1 - \frac{n}{N}\right)}$$

We will set $\hat{p} = \hat{q} = 0.5$, as those values are unknown and have therefore been chosen to maximize the bound. Substituting the known values, we have:

$$B = 2\sqrt{\frac{0.5 \cdot 0.5}{30-1}\left(1 - \frac{30}{200}\right)} \approx 0.17$$

Thus we have a bound on the error of estimation of approximately 0.17.

Before performing any tests, we must first assert the following assumptions:

- Each observation (response to the questionnaire) is independent, meaning no response influences another
- The data was randomly sampled, which is assured as we performed Simple Random Sampling (SRS)
- Our responses are considered categorical data, as they divide responses into distinct categories

Test 1: Chi-Square Test of Independence

One aspect of this study was to analyze the dependancy between certain question pairs, namely between some combination of these:

- Q.3: How well do you think AI improves the understanding of course content?
- Q.4: How accurate do you think AI is?
- Q.6: What percent of your study time is spent using AI tools?

- Q.7: Would you recommend the use of AI tools in education to your peers?

As the dataset consisted of responses indicating traits such as whether they had used AI tools for educational purposes, would recommend it for education purposes, or improved their understanding of course content, a chi-square test of independence was performed on the results of each question pair above to see whether one result influenced another. This was done to study any potential dependencies between the perceptions and experiences related to AI usage in education. The results were as follows:

Question Pairs	**X-squared value**	**degree of freedom**	**p-value**
Question 6, 7	14.537	10	0.1499
Question 3, 6	34.045	30	0.279
Question 4, 6	56.663	30	0.002289

Test 2: Multinomial Logistics Regression

Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables.

We perform logistic regression with respect to predicting the recommendation of using AI tools in education based on whether individuals have previously used AI tools for educational purposes.

Assumptions to assert:

- Both of our multinomial logistic regression variables are independent.
- There is a linear relationship between any continuous independent variables and the logit transformation of the dependent variable.

The outcome variable is AI_Recommendation and the predictor variable is AI_Familiarity. The predictor variable has two categorical values of 'Yes,' and 'No'.

(Intercept): This represents the baseline or intercept for the category that's used as the reference. Here, 0.4055049 is the intercept

Variable	Coefficient	Std. Error
Intercept	0.406	0.913
Have you previously used AI tools for educational purpose?	-1.793	1.041

Table 1: Summary of Logistic Regression Results

for the reference category.

The coefficient (-1.7918457) indicates the effect of AI_Familiarity on the log odds of AI_Recommendation. A negative coefficient suggests that as AI_Familiarity increases (from 0 to 1, i.e., from not using AI tools to using AI tools previously), the log odds of recommending AI tools decrease.

Statistical Reference

Question	Mean	Median	Standard Deviation	Confidence Interval
How well do you think AI improves the understanding of course content? (1 to 5)	3.93	4	0.91	(3.59, 4.27)
How accurate do you think AI is? (1 to 5)	3.29	3	0.76	(2.99, 3.58)
What percent of your study time is spent using AI tools?	22.28%	10%	24.52%	(12.95%, 31.60%)

Table 2: Summary Statistics of Responses with Confidence Intervals

Discussion/Results

1. Chi - square Analysis

(i) Comparing questions 6 and 7, we see that as the p-value from the test is above 0.05, there is no significant association between the usage of AI and the likelihood of recommending AI to a peer.

(ii) Comparing questions 6 and 3, we see that as the p-value from the test is above 0.05, there is no significant association between the usage of AI and the likelihood of AI improving understanding of course content.

(iii) Comparing questions 6 and 4, we see that as the p-value from the test is below 0.05, there **is** a significant association between the usage of AI and the perceived accuracy of AI.

Overall, results from these tests indicate that there is no significant relationship between the usage of AI tools and the perception of said tools within the STA304 student populace.

2. Multinomial Logistics Regression

(i) The logistic regression revealed a negative coefficient (-1.793) for the predictor variable "Have you previously used AI tools for educational purposes?"

(ii) This negative coefficient suggests that as the usage of AI tools for educational purposes increases, the likelihood of recommending AI tools decreases.

Overall, results from this test indicate that if at all, there is a negative relationship between the usage of AI tools and the perception of said tools within the STA304 student populace.

Conclusion

1. **Perception and Usage:**

   - The study indicates no evidence of a significant relationship between students' usage of AI tools for educational purposes and their perception of these tools
   - Those who haven't used AI tools previously tend to be more inclined to recommend their use in education compared to those who have prior experience using them.

2. **Limitations:**

   - The sample size (n=30) might not fully represent the entire STA304 student population (N=200), potentially affecting the generalizability of findings.
   - A potential source of bias that may have affected the data collection stage is non-response bias, as the questionnaire was partially administered on a voluntary basis, allowing students to choose whether or not to respond.

Given these test results and despite these limitations, we thus fail to reject the null hypothesis.

References

https://stats.oarc.ucla.edu/r/dae/multinomial-logistic-regression/

https://www.r-bloggers.com/2020/05/multinomial-logistic-regression-with-r/

https://youtu.be/QvnsTXfPenU?si=tSj1rVP19FGPdCQw

https://statsandr.com/blog/chi-square-test-of-independence-in-r/

STA304 - Hypothesis Testing & Statistical Test Interpretation Review.pdf

Appendix

Listing 1: R code for chi-square statistical independance test

```
> rm(list=ls())
> library(tidyverse)
> library(nnet)
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")
> my_data = read.csv("Dataset.csv")
> percent_times = my_data$AI_Usage_Percent
> recommendation = my_data$AI_Recommendation
> contingency_table = table(percent_times, recommendation)
> chi_square_result = chisq.test(contingency_table, simulate
    .p.value = TRUE)
> print(chi_square_result)

        Pearson's Chi-squared test with simulated p-value (
            based on 2000 replicates)

data:  contingency_table
X-squared = 14.537, df = NA, p-value = 0.1059

> rm(list=ls())
> library(tidyverse)
> library(nnet)
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")
> my_data = read.csv("Dataset.csv")
> percent_times = my_data$AI_Usage_Percent
> recommendation = my_data$AI_Recommendation
> contingency_table = table(percent_times, recommendation)
> chi_square_result = chisq.test(contingency_table)
Warning message:
In chisq.test(contingency_table) :
  Chi-squared approximation may be incorrect
> print(chi_square_result)

        Pearson's Chi-squared test

data:  contingency_table
X-squared = 14.537, df = 10, p-value = 0.1499

> rm(list=ls())
> library(tidyverse)
> library(nnet)
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")
> my_data = read.csv("Dataset.csv")
> percent_times = my_data$AI_Usage_Percent
> recommendation = my_data$AI_Recommendation
> improve_concepts = my_data$AI_Improves_Concepts
> accuracy = my_data$AI_Accuracy
```

```
> contingency_table = table(percent_times, recommendation)
> contingency_table2 = table(percent_times, improve_concepts
    )
> contingency_table3 = table(accuracy, percent_times)
> chi_square_result = chisq.test(contingency_table)
Warning message:
In chisq.test(contingency_table) :
  Chi-squared approximation may be incorrect
> chi_square_result2 = chisq.test(contingency_table2)
Warning message:
In chisq.test(contingency_table2) :
  Chi-squared approximation may be incorrect
> chi_square_result3 = chisq.test(contingency_table3)
Warning message:
In chisq.test(contingency_table3) :
  Chi-squared approximation may be incorrect
> print(chi_square_result)

        Pearson's Chi-squared test

data:  contingency_table
X-squared = 14.537, df = 10, p-value = 0.1499

> print(chi_square_result2)

        Pearson's Chi-squared test

data:  contingency_table2
X-squared = 34.045, df = 30, p-value = 0.279

> print(chi_square_result3)

        Pearson's Chi-squared test

data:  contingency_table3
X-squared = 56.663, df = 30, p-value = 0.002289
```

Listing 2: R code for Logistics Regression test

```
> rm(list=ls())
> library(tidyverse)
> library(nnet)
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")
> my_data = read.csv("Dataset.csv")
> str(my_data)
'data.frame':   30 obs. of  8 variables:
 $   ..Timestamp            : chr  "10-16-2023 17:58:37" "
    10-16-2023 19:44:35" "10-16-2023 20:25:52" "10-17-2023
    11:48:54" ...
 $ AI_Familiarity          : int  1 1 0 1 1 1 1 1 1 1 ...
```

```
 $ Educational_AI_Benefits : chr  "Time saving" "Improve
    personalized learning" "Improve personalized learning,
    Enhance student engagement, Time saving" "Improve
    personalized learning" ...
 $ AI_Improves_Concepts    : int  5 4 3 4 3 5 4 5 3 5 ...
 $ AI_Accuracy             : int  4 4 2 3 4 5 3 3 3 4 ...
 $ AI_Education_Utilization: chr  "As a student, to
    summarize concepts I did not understand in class" "To
    help me understand questions." "N/A" "Explain new
    concepts in simpler terms" ...
 $ AI_Usage_Percent        : chr  "5%" "30%" "0%" "5%" ...
 $ AI_Recommendation       : int  1 1 2 2 1 1 1 1 1 2 ...
> logistic_model = multinom(AI_Recommendation  ~ AI_
    Familiarity, data = my_data)
# weights:  3 (2 variable)
initial  value 20.794415
final  value 15.875119
converged
> summary(logistic_model)
Call:
multinom(formula = AI_Recommendation ~ AI_Familiarity, data
    = my_data)

Coefficients:
                   Values Std. Err.
(Intercept)     0.4055049 0.9128746
AI_Familiarity -1.7918457 1.0408395

Residual Deviance: 31.75024
AIC: 35.75024
>
```

Listing 3: R code for statistical information

```
> library(tidyverse)
> setwd("C:\Users\shiva\OneDrive\Desktop\3rd year\sta304")
Error: '\U' used without hex digits in character string
    starting ""C:\U"
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")''
Error: unexpected string constant in "setwd("C:/Users/shiva/
    OneDrive/Desktop/3rd year/sta304")''"
> setwd("C:/Users/shiva/OneDrive/Desktop/3rd year/sta304")
> data = read.csv("Dataset.csv")
> study_time_percentages <- c(5, 30, 0, 5, 5, 100, 20, 10,
    10, 30, 0, 10, 20, 1, 5, 40, 20, 10, 30, 80, 10, 0, 40,
    10, 0, 40, 40, 15, 60)
>
> # Calculate standard deviation
> std_dev <- sd(study_time_percentages)
> std_dev
```

```
[1] 24.52098
>
> mean_percentages <- mean(study_time_percentages)
> mean_percentages
[1] 22.27586
> median_percentages <- median(study_time_percentages)
> median_percentages
[1] 10
> confidence_interval <- t.test(study_time_percentages)$conf
    .int
> confidence_interval
[1] 12.94858 31.60314
attr(,"conf.level")
[1] 0.95
> accuracy = c
    (4,4,2,3,4,5,3,3,3,4,4,3,3,2,3,4,4,3,2,4,3,3,3,3,2,4,3,4)
> mean(accurancy)
Error in mean(accurancy) : object 'accurancy' not found
> mean_accuracy = mean(accuracy)
> mean_accuracy
[1] 3.285714
> median_accuracy = median(accuracy)
> median_accuracy
[1] 3
> std_dev = sd(accuracy)
> std_dev
[1] 0.7628962
> confidence_interval <- t.test(accuracy)$conf.int
> confidence_interval
[1] 2.989894 3.581535
attr(,"conf.level")
[1] 0.95
> AI_understanding = c
    (5,4,3,4,3,5,4,5,3,5,3,4,4,3,4,5,2,4,2,5,4,4,5,4,3,5,5,4
,3,4)
> mean_AI_understanding = mean(AI_UN)
Error in mean(AI_UN) : object 'AI_UN' not found
> mean_AI_understanding = mean(AI_understanding)
> mean_AI_understanding
[1] 3.933333
> median_AI_understanding = median(AI_understanding)
> median_AI_understanding
[1] 4
> std_dev_AI_understanding = sd(AI_understanding)
> std_dev_AI_understanding
[1] 0.9071871
> confidence_interval_AI_understanding = t.test(AI_
    understanding)
> confidence_interval_AI_understanding

```

```
        One Sample t-test

data:  AI_understanding
t = 23.748, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 3.594584 4.272083
sample estimates:
mean of x
 3.933333
```