Survey Methods Assignment 2 – Group 7

105.708 Data Acquisition and Survey Methods (VU 2.0), Summer 2025

Group 7:Bashir Muhammad Sajid (52400204), Member2, Member3

2025-05-18

Contents

0.1   1. Introduction

The use of Artificial Intelligence (AI) tools such as ChatGPT, Grammarly, and others is becoming increasingly common in academic environments. These tools offer students various forms of support including writing assistance, content summarization, grammar correction, coding help, and language translation. However, their real impact on students' academic performance, motivation, and learning retention has not been fully explored.

This study investigates how the use of AI tools influences students' learning experience. A structured survey was developed to collect data on students' demographic profiles, usage of AI tools, and self-reported outcomes related to productivity, engagement, and knowledge retention.

0.1.1 1.1 Research Questions

1. Do students who use AI tools complete their academic tasks in less time while maintaining or improving the quality of their work compared to those who do not use AI tools?

2. Does the frequency of AI tool usage increase students' motivation and engagement during their studies?

3. Do students who incorporate AI tools into their study routines demonstrate better long-term retention of the material compared to those who do not?

0.2 2. Data and Package Setup

Before analyzing the survey responses, we load the necessary libraries and import the cleaned survey dataset.

```r
# Load required packages
library(readxl)        # For reading Excel files
library(dplyr)         # For data manipulation
library(ggplot2)       # For visualization
library(psych)         # For descriptive statistics
library(likert)        # For Likert-scale handling (if needed)
library(knitr)         # For tables
library(kableExtra)    # For enhanced table styling

# Load the survey data
survey_data <- read_excel("data/group7.xlsx")

# Preview structure and dimensions
str(survey_data)
```

```
## tibble [98 x 6] (S3: tbl_df/tbl/data.frame)
##  $ DemographicAnswer.1: chr [1:98] "male" "male" "female" "male" ...
##  $ DemographicAnswer.2: chr [1:98] "23" "24" "24" "23" ...
##  $ DemographicAnswer.3: chr [1:98] "MSc Tu Wien Data Science" "Data Science" "Data Science" "Data Sc:
##  $ Answer.1           : chr [1:98] "Yes" "Yes" "Yes" "Yes" ...
##  $ Answer.2           : chr [1:98] "8" "8" "2" "4" ...
##  $ Answer.3           : chr [1:98] "Using AI tools has helped me save time when completing academic 1
```

```r
dim(survey_data)
```

```
## [1] 98  6
```

0.3 3. Data Cleaning and Variable Overview

We first inspect the variable names and clean them to ensure clarity and usability in the analysis. Long textual survey statements are renamed to shorter, descriptive identifiers.

2

```r
# View raw column names
names(survey_data)
```

```
## [1] "DemographicAnswer.1" "DemographicAnswer.2" "DemographicAnswer.3"
## [4] "Answer.1"            "Answer.2"            "Answer.3"
```

```r
# Clean and rename columns for clarity
colnames(survey_data) <- c(
  "age",
  "gender",
  "program",
  "ai_usage",
  "ai_hours",
  "ai_opinions"
)

# Re-check structure after renaming
str(survey_data)
```

```
## tibble [98 x 6] (S3: tbl_df/tbl/data.frame)
##  $ age        : chr [1:98] "male" "male" "female" "male" ...
##  $ gender     : chr [1:98] "23" "24" "24" "23" ...
##  $ program    : chr [1:98] "MSc Tu Wien Data Science" "Data Science" "Data Science" "Data Science / I
##  $ ai_usage   : chr [1:98] "Yes" "Yes" "Yes" "Yes" ...
##  $ ai_hours   : chr [1:98] "8" "8" "2" "4" ...
##  $ ai_opinions: chr [1:98] "Using AI tools has helped me save time when completing academic tasks.
```

0.4 4. Data Correction and Type Conversion

We noticed that the `age` and `gender` columns were reversed during import. We correct this, convert hours to numeric, and prepare the data for further processing.

```r
# Rename correctly based on inspection
colnames(survey_data) <- c(
  "gender",    # originally DemographicAnswer.1 → holds "male", "female"
  "age",       # originally DemographicAnswer.2 → holds numeric-like ages
  "program",
  "ai_usage",
  "ai_hours",
  "ai_opinions"
)

# Convert age and ai_hours to numeric
survey_data$age <- as.numeric(survey_data$age)
survey_data$ai_hours <- as.numeric(survey_data$ai_hours)
```

```
## Warning: NAs introduced by coercion
```

```r
# Recheck structure
str(survey_data)
```

```
## tibble [98 x 6] (S3: tbl_df/tbl/data.frame)
##  $ gender    : chr [1:98] "male" "male" "female" "male" ...
##  $ age       : num [1:98] 23 24 24 23 22 22 33 24 26 45 ...
##  $ program   : chr [1:98] "MSc Tu Wien Data Science" "Data Science" "Data Science" "Data Science / I
##  $ ai_usage  : chr [1:98] "Yes" "Yes" "Yes" "Yes" ...
##  $ ai_hours  : num [1:98] 8 8 2 4 10 8 7 10 0.5 2 ...
##  $ ai_opinions: chr [1:98] "Using AI tools has helped me save time when completing academic tasks.
```

0.5 5. Expanding AI Opinion Statements

The `ai_opinions` column contains multiple selected statements separated by slashes. We now split these
statements and create one binary column per unique response for easier analysis.

```r
# Define unique AI opinion statements (9 total from survey)
ai_items <- c(
  "Using AI tools has helped me save time when completing academic tasks.",
  "Using AI tools has helped me improve the quality of my academic work.",
  "Using AI tools has increased my motivation to learn new things.",
  "Using AI tools has helped me better understand and retain the material I study.",
  "Using AI tools has made it harder for me to retain information in the long term.",
  "Using AI tools has decreased my confidence in my own learning abilities.",
  "Using AI tools has lowered the quality of my academic work.",
  "Using AI tools has not caused any significant changes in my academic performance.",
  "I do not use AI tools for academic purposes."
)

# Create one logical column for each opinion
for (item in ai_items) {
  col_name <- paste0("op_", make.names(substr(item, 1, 40)))  # create safe short names
  survey_data[[col_name]] <- grepl(item, survey_data$ai_opinions, fixed = TRUE)
}

# View new structure
names(survey_data)
```

```
##  [1] "gender"
##  [2] "age"
##  [3] "program"
##  [4] "ai_usage"
##  [5] "ai_hours"
##  [6] "ai_opinions"
##  [7] "op_Using.AI.tools.has.helped.me.save.time.w"
##  [8] "op_Using.AI.tools.has.helped.me.improve.the"
##  [9] "op_Using.AI.tools.has.increased.my.motivati"
## [10] "op_Using.AI.tools.has.helped.me.better.unde"
## [11] "op_Using.AI.tools.has.made.it.harder.for.me"
## [12] "op_Using.AI.tools.has.decreased.my.confiden"
## [13] "op_Using.AI.tools.has.lowered.the.quality.o"
## [14] "op_Using.AI.tools.has.not.caused.any.signif"
## [15] "op_I.do.not.use.AI.tools.for.academic.purpo"
```

## 0.6   6. Exploratory Data Analysis (EDA)

0.6.1   6.1 RQ1: Time-saving and Quality (AI Users Only)

Since all respondents report using AI tools, we visualize the proportion of users reporting that AI helped them save time or improve quality.

```r
library(ggplot2)
library(tidyr)
library(dplyr)

# Only include users who answered "Yes"
rq1_summary <- survey_data %>%
  filter(ai_usage == "Yes") %>%
  summarise(
    SaveTime = mean(op_Using.AI.tools.has.helped.me.save.time.w, na.rm = TRUE),
    ImproveQuality = mean(op_Using.AI.tools.has.helped.me.improve.the, na.rm = TRUE)
  ) %>%
  pivot_longer(cols = everything(), names_to = "Outcome", values_to = "Proportion")

# Plot
ggplot(rq1_summary, aes(x = Outcome, y = Proportion, fill = Outcome)) +
  geom_col(width = 0.5) +
  geom_text(aes(label = scales::percent(Proportion, accuracy = 1)),
            vjust = -0.3, size = 4.5) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1), limits = c(0, 1)) +
  labs(
    title = "Reported Benefits of AI Use (Time-saving & Quality)",
    subtitle = "Among AI Users Only (n = 98)",
    x = NULL,
    y = "Proportion of Respondents",
    fill = "Benefit Type"
  ) +
  theme_minimal(base_size = 13)
```

Reported Benefits of AI Use (Time–saving & Quality)
Among AI Users Only (n = 98)
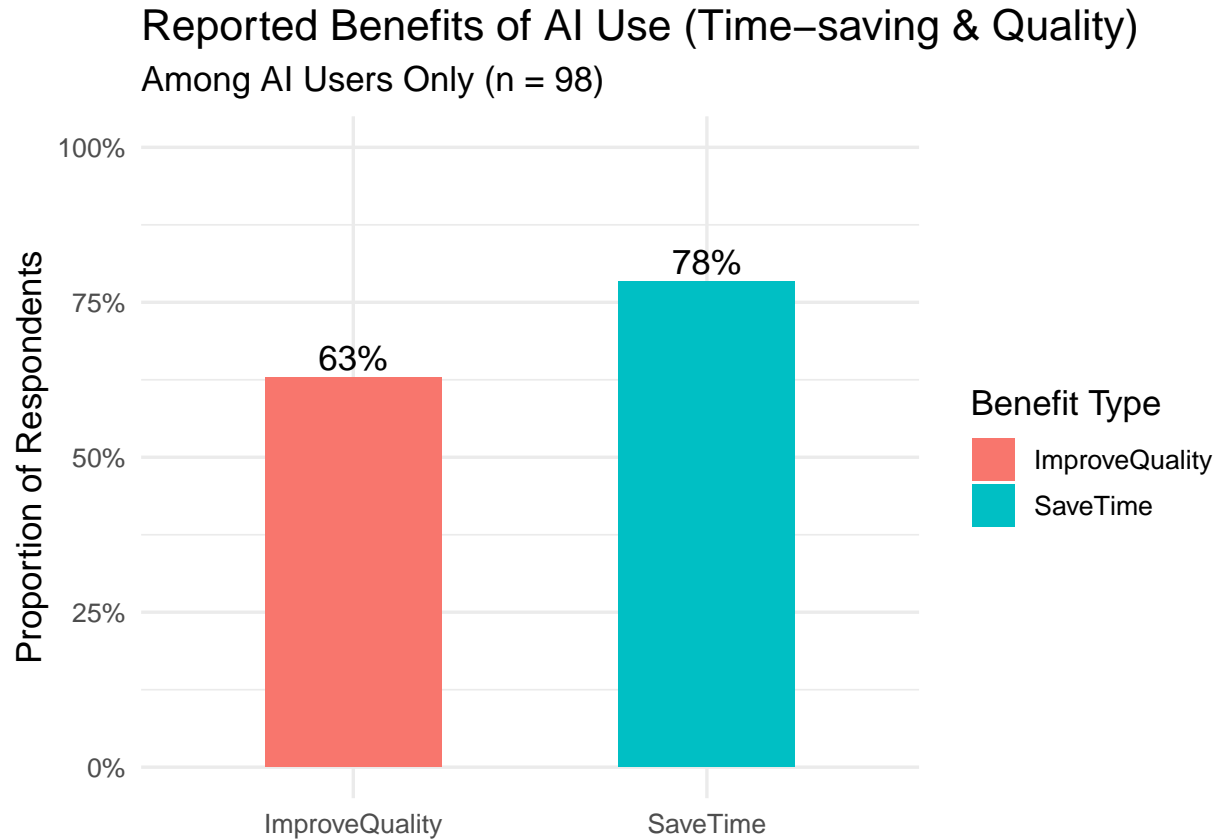
0.6.1.1 Interpretation (RQ1) The plot shows that among students who reported using AI tools for academic purposes:

- **78%** indicated that AI tools helped them **save time** on academic tasks
- **63%** reported that AI tools improved the **quality of their academic work**

These findings suggest that AI tools are perceived to contribute positively to both efficiency and output quality among users. However, since **all respondents in the dataset reported using AI tools**, we cannot compare these benefits to a control group of non-users. Therefore, the result supports RQ1 only within the context of current AI users, and further data collection would be needed to assess differences across usage groups.

0.6.2 6.2 RQ2: Motivation by Frequency of AI Tool Usage

We group AI usage into frequency bands and visualize how motivation levels differ across these groups.

```
library(dplyr)
library(ggplot2)

# Create usage buckets
survey_data$usage_group <- cut(
  survey_data$ai_hours,
  breaks = c(-Inf, 2, 5, 10, Inf),
  labels = c("0-2 hrs", "3-5 hrs", "6-10 hrs", "10+ hrs")
```

```
)

# Summarize motivation rate by usage group
rq2_summary <- survey_data %>%
  group_by(usage_group) %>%
  summarise(
    MotivationRate = mean(op_Using.AI.tools.has.increased.my.motivati, na.rm = TRUE),
    n = n()
  )

# Plot
ggplot(rq2_summary, aes(x = usage_group, y = MotivationRate)) +
  geom_col(fill = "#00BFC4") +
  geom_text(aes(label = scales::percent(MotivationRate, accuracy = 1)), vjust = -0.3, size = 4.5) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1), limits = c(0, 1)) +
  labs(
    title = "Motivation by Weekly AI Usage",
    subtitle = "Proportion of students who report increased motivation",
    x = "Weekly AI Usage (hours)",
    y = "Motivated (TRUE responses)"
  ) +
  theme_minimal(base_size = 13)
```
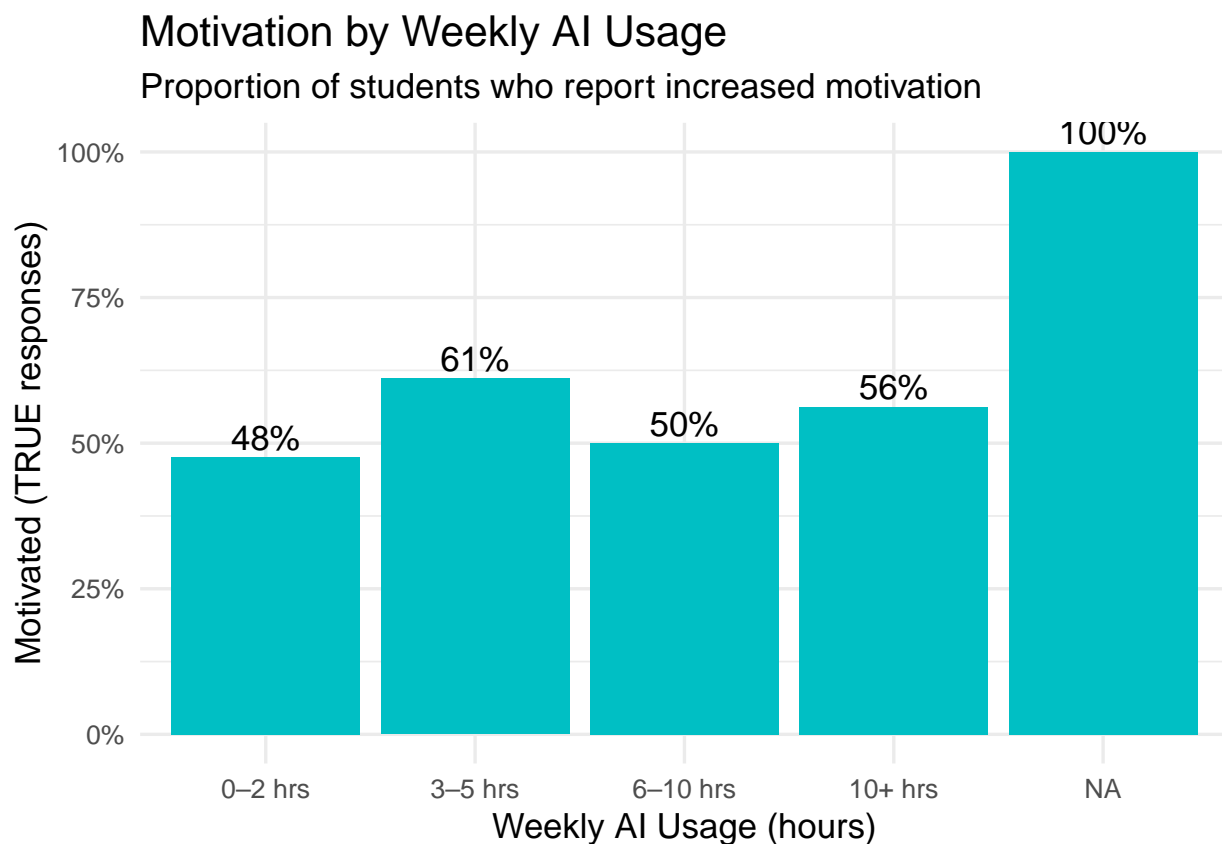

Interpretation (RQ2)

The plot illustrates the proportion of students who reported feeling more motivated to study as a result of
using AI tools, segmented by how many hours per week they use them.

- Students who used AI tools for **3–5 hours per week** showed the **highest motivation rate** (61%).
- Lower motivation was reported by students using AI tools **less than 2 hours (48%)** or **6–10 hours (50%)**.
- The group using AI tools for **10+ hours** reported 56% motivation.
- One group with `NA` in `ai_hours` showed 100% motivation, which suggests incomplete data and should be excluded from final inference.

These findings suggest a **moderate positive relationship** between AI tool usage and perceived motivation, though the relationship is **not linear**. It appears that **moderate usage** may be more motivating than very low or very high usage.

0.6.3   6.3 RQ3: Retention by Weekly AI Tool Usage

We assess the relationship between AI usage frequency and students' agreement with improved long-term retention.

```r
# Summarize retention response by usage group (filtering NA)
rq3_summary <- survey_data %>%
  filter(!is.na(usage_group)) %>%
  group_by(usage_group) %>%
  summarise(
    RetentionRate = mean(op_Using.AI.tools.has.helped.me.better.unde, na.rm = TRUE),
    n = n()
  )

# Plot
ggplot(rq3_summary, aes(x = usage_group, y = RetentionRate)) +
  geom_col(fill = "#F8766D") +
  geom_text(aes(label = scales::percent(RetentionRate, accuracy = 1)), vjust = -0.3, size = 4.5) +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1), limits = c(0, 1)) +
  labs(
    title = "Retention by Weekly AI Usage",
    subtitle = "Proportion of students who report better understanding & retention",
    x = "Weekly AI Usage (hours)",
    y = "Retention (TRUE responses)"
  ) +
  theme_minimal(base_size = 13)
```

Retention by Weekly AI Usage

Proportion of students who report better understanding & retention
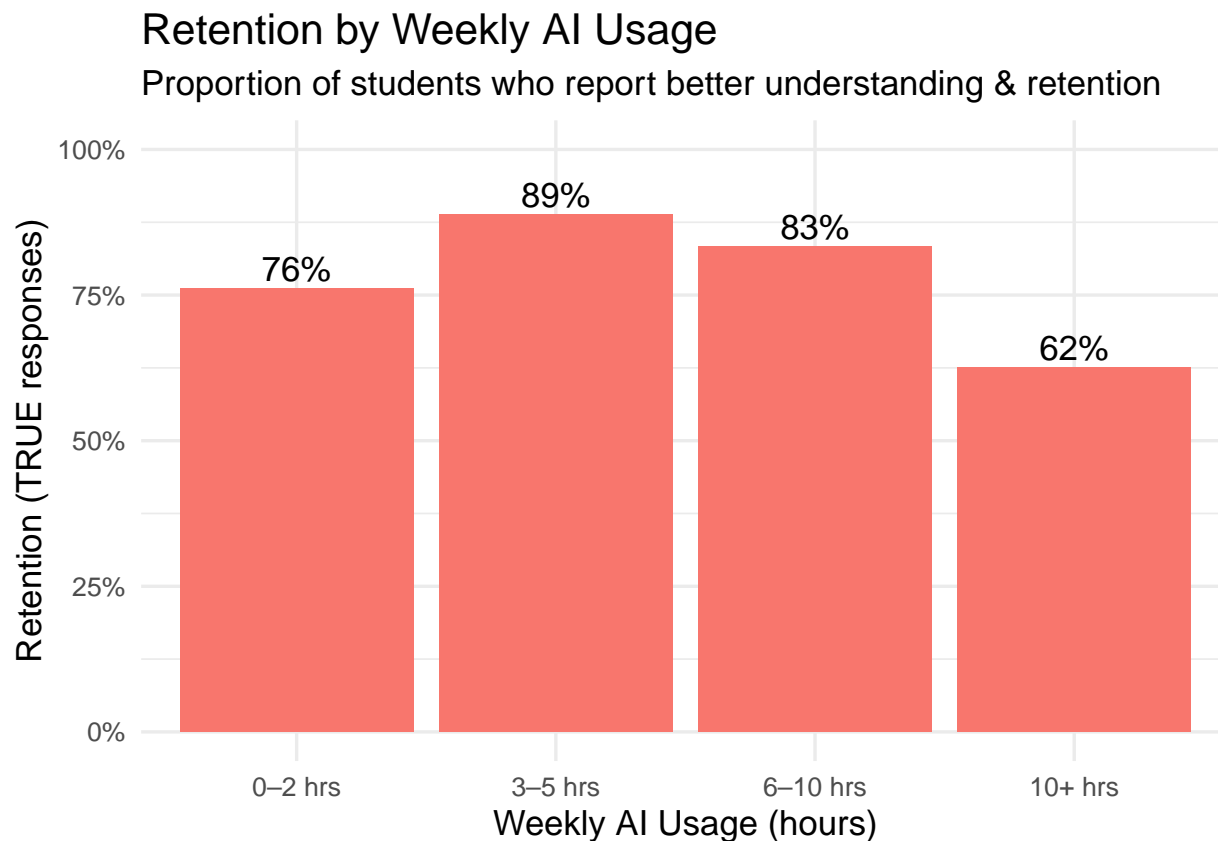
Interpretation (RQ3)

The chart indicates how weekly AI tool usage is associated with students' perceived ability to understand and retain learning material.

- Students using AI tools for **3–5 hours/week** showed the highest retention rate (89%)
- Retention remained relatively high at **6–10 hours/week** (83%), but decreased at both usage extremes:
    - **0–2 hours/week**: 76%
    - **10+ hours/week**: 62%

This may suggest that **moderate usage of AI tools** supports deeper learning and retention, whereas very limited or excessive use may be less beneficial. However, since this is self-reported data, and the relationship is not strictly linear, further analysis would be needed to confirm these patterns.

0.7 7. Descriptive Inference

We compute summary statistics for numeric variables such as age and weekly AI usage. This gives an overview of the respondent profile and the typical intensity of AI usage.

```r
# Load required packages again (if needed)
library(psych)
library(knitr)
library(kableExtra)

# Summary stats for numeric variables
```

```
numeric_summary <- psych::describe(survey_data %>% select(age, ai_hours))

# Show as styled table
kable(numeric_summary, digits = 2, caption = "Summary Statistics for Age and Weekly AI Usage") %>%
  kable_styling(full_width = FALSE, position = "center")
```

Table 1: Summary Statistics for Age and Weekly AI Usage

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
age	1	98	25.82	4.51	25.5	25.48	3.71	4.0	45	41.0	0.30	8.05	0.46
ai_hours	2	97	7.02	5.73	5.0	6.15	4.45	0.5	28	27.5	1.32	1.25	0.58

0.7.1 Interpretation

The summary statistics provide insights into the age distribution and weekly AI tool usage among the 98 student respondents:

- The **average age** of participants is **25.8 years**, with a range from **4 to 45**. The distribution appears fairly symmetric (skew = 0.30), though the extreme minimum value of 4 years likely reflects a data entry error or outlier.
- For **weekly AI usage**, students report using AI tools for an average of **7.0 hours** per week (median = 5.0 hours), with values ranging from **0.5 to 28 hours**. The distribution is positively skewed (**skew = 1.32**), indicating that a few students reported very high usage levels.

These figures offer a useful profile of the surveyed students and the intensity with which they incorporate AI tools into their academic routines.

0.8 8. Analytic Inference

0.8.1 8.1 RQ1 – Time and Quality Impact Among AI Users

We use one-sample proportion tests to assess whether a majority (>50%) of AI users report time-saving and improved work quality.

```
# Count number of TRUE responses
n_total <- nrow(survey_data)

# Time-saving
n_save_time <- sum(survey_data$op_Using.AI.tools.has.helped.me.save.time.w, na.rm = TRUE)
test_save <- binom.test(n_save_time, n_total, p = 0.5, alternative = "greater")

# Quality improvement
n_quality <- sum(survey_data$op_Using.AI.tools.has.helped.me.improve.the, na.rm = TRUE)
test_quality <- binom.test(n_quality, n_total, p = 0.5, alternative = "greater")

# Output results
test_save
```

```
## 
##   Exact binomial test
## 
## data:  n_save_time and n_total
## number of successes = 76, number of trials = 98, p-value = 1.948e-08
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##   0.69521 1.00000
## sample estimates:
## probability of success
##              0.7755102
```

test_quality

```
## 
##   Exact binomial test
## 
## data:  n_quality and n_total
## number of successes = 61, number of trials = 98, p-value = 0.009845
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##   0.5347945 1.0000000
## sample estimates:
## probability of success
##               0.622449
```

0.8.2 Interpretation (RQ1)

Two exact binomial tests were conducted to assess whether significantly more than half of AI-using students reported that:

- AI tools saved them time
- AI tools improved the quality of their academic work

Results:

- **Time-saving:**
  76 out of 98 students (77.6%) agreed.
  The test was highly significant ($p < 0.00001$), with a 95% confidence interval of [69.5%, 100%], indicating that a **clear majority** experienced time-saving benefits.

- **Quality improvement:**
  61 out of 98 students (62.2%) agreed.
  The test was statistically significant ($p = 0.0098$), with a 95% confidence interval of [53.5%, 100%], confirming that **more than half** also experienced quality benefits.

These results support **RQ1** within the context of AI users, despite the absence of a non-user comparison group.

0.8.3 8.2 RQ2 – Motivation Predicted by Weekly AI Usage

We fit a logistic regression model to assess whether the number of weekly hours spent using AI tools predicts self-reported motivation.

```
# Create binary variable
motivation <- survey_data$op_Using.AI.tools.has.increased.my.motivati

# Fit logistic regression model
model_rq2 <- glm(motivation ~ ai_hours, data = survey_data, family = binomial)

# Model summary
summary(model_rq2)
```

```
##
## Call:
## glm(formula = motivation ~ ai_hours, family = binomial, data = survey_data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.121413   0.323580   0.375    0.707
## ai_hours    0.009242   0.035969   0.257    0.797
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 133.63  on 96  degrees of freedom
## Residual deviance: 133.57  on 95  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 137.57
##
## Number of Fisher Scoring iterations: 3
```

0.8.4 Interpretation (RQ2)

A logistic regression model was fitted to predict whether a student felt more motivated to study as a result of using AI tools, based on the number of hours they reported using such tools per week.

Results:

- The coefficient for `ai_hours` was **0.0092**, with a *p*-value of **0.797**, indicating that the relationship is not statistically significant.
- The model suggests that each additional hour of AI tool use is associated with only a **0.9% increase in the log-odds of reporting increased motivation**, which is negligible and not reliable based on this data.

In conclusion, these results do **not support RQ2**: there is **no clear evidence** that motivation increases as a function of more frequent AI tool usage among the students surveyed.

0.8.5 8.3 RQ3 – Retention Predicted by Weekly AI Usage

We fit a logistic regression model to determine whether weekly AI usage predicts whether students report improved retention and understanding.

```
# Binary outcome: retention opinion
retention <- survey_data$op_Using.AI.tools.has.helped.me.better.unde

# Logistic regression
model_rq3 <- glm(retention ~ ai_hours, data = survey_data, family = binomial)

# Model summary
summary(model_rq3)
```

```
##
## Call:
## glm(formula = retention ~ ai_hours, family = binomial, data = survey_data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.79423    0.41994   4.273 1.93e-05 ***
## ai_hours    -0.05085    0.04148  -1.226     0.22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 95.959  on 96  degrees of freedom
## Residual deviance: 94.513  on 95  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 98.513
##
## Number of Fisher Scoring iterations: 4
```

0.8.6 Interpretation (RQ3)

We used a logistic regression model to assess whether the number of hours spent weekly using AI tools predicts students' perceived improvement in retention and understanding of course material.

Results:

- The coefficient for `ai_hours` was **-0.0508** with a p-value of **0.22**, indicating **no statistically significant relationship**.
- While the negative coefficient suggests that higher AI usage might slightly reduce perceived retention, this effect is **not reliable** or meaningful in this dataset.

Therefore, the results do **not support RQ3**. Weekly AI usage hours do **not significantly predict** whether students feel that AI tools have improved their long-term learning retention.

0.9 10. Appendix

```
## R version 4.4.3 (2025-02-28 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 22000)
##
## Matrix products: default
##
```

13

```
## 
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
## 
## time zone: Europe/Vienna
## tzcode source: internal
## 
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
## 
## other attached packages:
## [1] tidyr_1.3.1     kableExtra_1.4.0 knitr_1.50      likert_1.3.5
## [5] xtable_1.8-4    psych_2.5.3      ggplot2_3.5.2   dplyr_1.1.4
## [9] readxl_1.4.5
## 
## loaded via a namespace (and not attached):
##  [1] generics_0.1.3   xml2_1.3.8       stringi_1.8.4     lattice_0.22-6
##  [5] digest_0.6.37    magrittr_2.0.3   evaluate_1.0.3    grid_4.4.3
##  [9] fastmap_1.2.0    cellranger_1.1.0 plyr_1.8.9        tinytex_0.57
## [13] gridExtra_2.3    purrr_1.0.4      viridisLite_0.4.2 scales_1.3.0
## [17] mnormt_2.1.1     cli_3.6.4        rlang_1.1.5       munsell_0.5.1
## [21] withr_3.0.2      yaml_2.3.10      tools_4.4.3       parallel_4.4.3
## [25] reshape2_1.4.4   colorspace_2.1-1 vctrs_0.6.5       R6_2.6.1
## [29] lifecycle_1.0.4  stringr_1.5.1    pkgconfig_2.0.3   pillar_1.10.2
## [33] gtable_0.3.6     glue_1.8.0       Rcpp_1.0.14       systemfonts_1.2.2
## [37] xfun_0.51        tibble_3.2.1     tidyselect_1.2.1  rstudioapi_0.17.1
## [41] farver_2.1.2     htmltools_0.5.8.1 nlme_3.1-167     rmarkdown_2.29
## [45] svglite_2.1.3    labeling_0.4.3   compiler_4.4.3
```