Group Name **Students**	Group 22 2112632 5670628 5647195 5655350 5640152 5663354
Module Code	IB98D0
Module Title	Advanced Data Analysis
Submission Deadline	13-Feb-2025 12:00:00 PM
Date Submitted	12-Feb-2025 19:55:48 PM
Word Count	1928
Number of Pages	7
Question Attempted	Your team needs to conduct an analysis of the initial experimental data and write a report discussing your analysis and results, considering the goal of your company. You also need give recommendations to your Analytics Department manager and the Executive team.
Have you used Artificial Intelligence (AI) in any part of this assignment?	No
If you have ticked "Yes" above, please briefly outline below which AI tool you have used, and what you have used it for. Please note, you must also reference the use of generative AI correctly within your assessment, in line with the guidance provided in your student handbook.	

Advanced Data Analysis Assignment Report

IB98D0 – Group 22

Student IDs: 2112632, 5655350, 5670628, 5640152, 5647195. 5663354

Executive Summary

This report quantifies the impact of the implementation of an experimental new computer model on the decision-making of a small subset of loan officers within the loan review department. Our findings suggest that this new model significantly improves loan officers' error rates, both decreasing the rate at which approved loans are defaulted on and decreasing the rate at which loans that would have been paid back are rejected. We find a 31.7% reduction in the rate at which loan officers approved loans that later defaulted compared to a control. We also find a 44.8% reduction in the rate at which loan officers rejected loans that would have been paid back when compared to the same control.

While these findings are promising, we cannot currently give a robust estimate of the financial impact of the new computer model. As a result, we recommend that this experiment be scaled up to a larger segment of the loan review department and the analytics department be given access to information surrounding the face values, interest rates, durations and outstanding losses from defaults of the loans of all those participating in the experiment. We would add the further stipulations that with the larger experiment we will need to collaborate closely with the loan review department to ensure the validity of the experiment is maintained.

Background

The consumer lending company is currently facing high costs from loan officers approving loans that have defaulted later (type II errors). The standard process for approving loans involves three stages: an initial independent decision by the loan officer, a review of the loan application by a computer model and a final decision by the loan officer informed by both their initial prediction and the computer model's assessment. In response to concerns that the existing computer model is outdated, our department has developed a new model which aims to better assist loan officers, leading to an overall goal of reduced error rates.

This experiment aims to quantify how the new computer model has impacted decision-making in a small-scale pilot study. The results of this experiment should be used to inform strategic decisions concerning future implementations and testing of this model, such as if the experiment should continue running, are current results sufficient and should new experiments be developed with alternative designs.

Experiment Design

To evaluate the effectiveness of this new computer model an A/B test was conducted within the Loan Review Department. This randomly assigned loan officers to one of two groups to determine which computer model they would use. Those in the treatment group had their final decisions informed by the new computer model and those in the control group were informed by the current model.

Over the past 10 days we have been collecting real-time data on a variety of metrics concerning the number of decisions made, the error counts and types in the loan officers' decisions and the utilisation of the respective computer models in these decisions. As a result, we are likely to have captured weekly-level cycles in this data and are able to effectively target our results to an undiluted target population of loan officers.

Overall Evaluation Criteria

The Overall Evaluation Criteria (OEC) we used in this analysis was the type II error rate of each loan officer's final decisions after referring to the computer model. We specifically focused on type II error rates as our primary evaluation measure as these defaulted loans have inflicted large financial losses on the firm and this measure is likely to be more robust than metrics based on estimations of likely unavailable data (such as whether loans that were never issued would've defaulted or not). We used loan officers as our randomisation unit to attempt to maintain unit independence (at a day-by-day level, the differing individual risk preferences of loan officers would violate this assumption) while maximising our statistical power and relevancy of the experiment. We used loan officers who made a final decision after being informed by the computer model as our target population as this prevents dilution of the effect of the different computer models with users who did not use the model.

Additionally, we have provided the effect on type I error rate normalised by the same randomisation unit and target population to be used as a supplementary metric. This represents the rate at which loan officers denied loans that would have been paid back in their final decisions. This is an important metric to consider when it is available as these good loans are the main source of revenue to a consumer lending company and while limiting the amount of loans we approve that default will limit financial losses it cannot come at the cost of restricting revenues through rejecting too many good loans.

Hypothesis

Our analysis considered the null hypothesis to be that the implementation of the new computer model did not reduce our OEC (type II error rate of final decisions per loan officer). In the presence of significant evidence to the contrary, we would reject this in favour of the alternative hypothesis that the implementation of the new computer model did decrease our OEC.

Data Preparation

Data Cleaning & Preprocessing

To ensure valid comparisons, loan officers who did not interact with the computer model (complt_fin = 0) were excluded from the dataset. This step ensured that all included officers actively engaged with the model.

Aggregation at the Loan Officer Level

Given that randomization was conducted at the loan officer level, the data was aggregated accordingly. This approach ensured that performance metrics were standardized across officers. Specifically, we computed the mean **Type II error rate** (false approvals) and **Type I error rate** (false rejections) per loan officer.

Metric Calculation

The final Type II error rate per officer (**typeII_fin_rate_per_officer**) was calculated as:

typeII_fin_rate_per_officer = typeII_fin_sum/complt_fin_sum

Similarly, the final Type I error rate per officer (**typeI_fin_rate_per_officer**) was computed as:

typeI_fin_rate_per_officer = typeI_fin_sum/complt_fin_sum

These error rates were then grouped by variant (**Control/Treatment**) to facilitate statistical comparisons and assess the impact of the new computer model on decision-making accuracy.

Methodology

In order to test the impact of the new computer model we compared the mean type II error rates in the final decisions of the control and treatment groups using a one-sided Welch's two-sample t-test. We opted for a Welch's t-test as the two groups were found to have different variances and a one-sided t-test as the decisions on implementation only change if there is a decrease in error rate. If the error rate does not significantly change or increases, then either way the experiment would be stopped and no further testing would be conducted (assuming valid results and a high-powered test). To ensure that this was directly measuring the effectiveness of the computer model on decision making we restricted our target population to those loan officers that used their corresponding model in all their loan decisions.

After finding the difference in means was statistically significant, indicating that the implementation of the computer model did have an impact on the loan officers' decisions, we used Cohen's d analysis to estimate the size of this effect. Further, we would use this metric in an analysis of the power of the test we conducted. This would ensure our experimentation was sufficiently rigorous to support its conclusions being used in strategic decision making.

This process was then repeated in the same fashion considering the type I error rate, with the only other change to the design being the usage of a two-sided t-test as we may be willing to incur some lost revenue if the reduction in financial losses is large enough.

Results

Our analysis would suggest that the mean type II error rate of loan officers' final decisions in the treatment group was significantly lower than that of the control group at the 5% significance level $(t_{10.894} = 3.610, p = 0.002)$. This would suggest that the new computer model resulted in a 3.83 (confidence interval $[1.923, \infty]$) percentage point reduction in type II error rate, which is a 31.7% reduction relative to the control group.

Computing the Cohen's d for standardization and reproducibility purposes we found the difference to correspond to a value of -1.76 (confidence interval $[-2.58, -0.92]$). This would suggest that the effect size is very large, using a baseline of a value of 0.8 to indicate a large effect.

This effect size was so large in fact that it yielded a power of the test of 99.88% despite the small sample sizes of 10 loan officers in the control group and 28 in the treatment group. This would suggest that with the observed effect size we have a 99.88% chance that we have correctly rejected the null hypothesis.

The corresponding analysis on the type I error rates of loan officers' final decisions suggested the new computer model did yield significantly different error rates than the current model at the 5% significance level($t_{10.927} = 3.949, p = 0.002$). We found the difference to be equivalent to a 16.3 percentage point reduction (confidence interval $[7.205, 25.378]$) in type I error rate in the treatment group when compared to the control group, equivalent to a 44.8% reduction.

This led to a Cohen's d of -1.92 (confidence interval $[-2.75, -1.06]$) which is even larger than that of our OEC. As a result, we would not only expect the company's financial losses from default to fall significantly, but would also expect the proportion of loans that would be paid back being approved to rise too, raising their revenues.

Recommendations

Analytical Recommendations

Extrapolating from the mean loan value (assumed to be the central point of the provided range of $20,000-$35,000) implementation of this computer model could lead to roughly estimated cost savings of $1,053 per loan officer. While this is promising, we do not have a lot of contextual data here and it would be important to see more information around how the system performs in different areas or with different types of loan. To better tailor our experiment to the business use case we would want to understand the direct impact of the model on revenues and costs to the company.

Considering this we would recommend stopping the experiment in its current state and scaling it up to a wider component of the company. This larger experiment would have a larger sample size that is equally split between control and treatment groups and more available information, at the very least providing the financial loss associated with each default, the face value of the loans, the duration and the interest rate. Additionally, this new experiment should last up to at least one year to ensure longer cyclical components can be considered, for example there may be seasonal aspects of the consumer loan market that the new model could struggle with.

Business Recommendations

To ensure a smooth rollout of the new computer model, we recommend a phased implementation strategy starting with a pilot in selected regions. This approach minimises disruption and allows for adjustments based on feedback and performance. Additionally, a comprehensive training program for loan officers is essential to ensure they can effectively use and trust AI recommendations leading to better results for the business.

Continuous monitoring of the model's performance is crucial. Establishing the Key Performance Indicators (KPIs) to track the impact of new computer model and refine the system as needed. Improved loan accuracy will lead to faster approvals and higher customer satisfaction, boosting retention and strengthening relationships.

Finally, the deployment must adhere to industry regulations. Involve compliance and legal teams to ensure alignment with regulatory requirements and mitigate risks. By adopting this approach, the company can achieve substantial financial gains while maintaining operational efficiency, customer trust, and regulatory compliance.

Appendix

```
library(tidyverse)
library(effectsize)

# Load Data
loan <- read_csv("ADAproject_-5_data.csv", guess_max = 1000)

# Check data types of columns
spec(loan)
```

```
## cols(
##   Variant = col_character(),
##   loanofficer_id = col_character(),
##   day = col_double(),
##   typeI_init = col_double(),
##   typeI_fin = col_double(),
##   typeII_init = col_double(),
##   typeII_fin = col_double(),
##   agree_init = col_double(),
##   agree_fin = col_double(),
##   conflict_init = col_double(),
##   conflict_fin = col_double(),
##   revised_per_ai = col_double(),
##   revised_agst_ai = col_double(),
##   fully_complt = col_double(),
##   confidence_init_total = col_double(),
##   confidence_fin_total = col_double(),
##   complt_init = col_double(),
##   complt_fin = col_double(),
##   ai_typeI = col_double(),
##   ai_typeII = col_double(),
##   badloans_num = col_double(),
##   goodloans_num = col_double()
## )
```

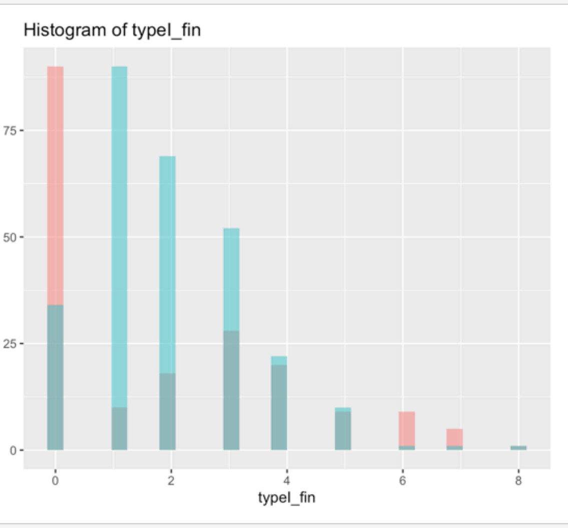
```
# Set categorical variable
loan <- loan %>% mutate(variant = factor(Variant),
                        day = factor(day))
```

Initial EDA

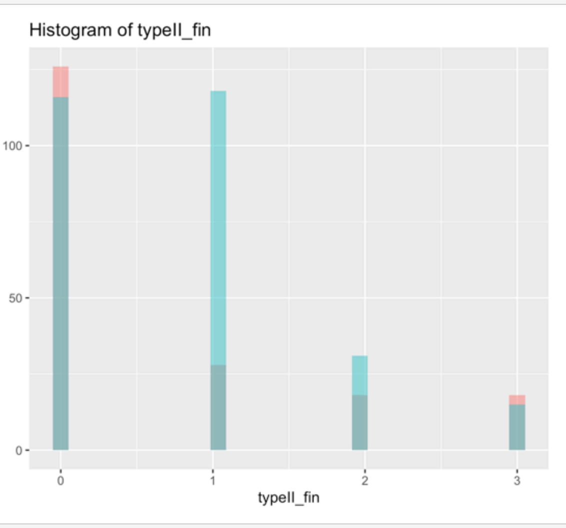
```
# Check sample sizes of both variants
loan %>%
  group_by(variant) %>%
  summarize(sample_size = n_distinct(loanofficer_id) )
```

```
## # A tibble: 2 × 2
##   variant   sample_size
##   <fct>           <int>
## 1 Control            19
## 2 Treatment          28
```

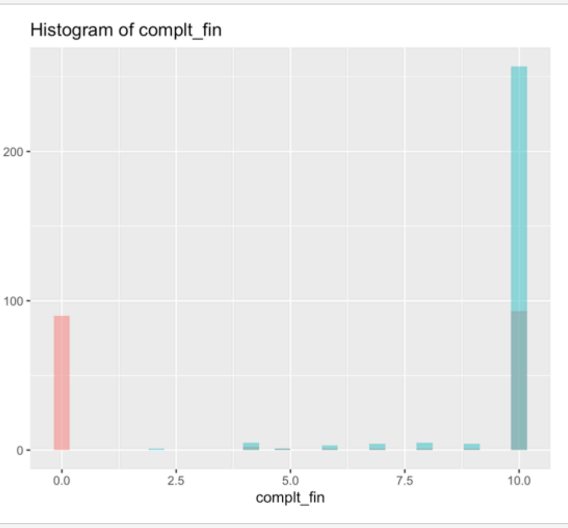
```
# Select useful variables for analysis
loan_selected <- loan %>% select(variant,loanofficer_id,day, typeI_fin,typeII_fin,complt_fin)

# Check the distribution of numeric variables
ggplot(loan_selected, aes(x = typeI_fin, fill = variant)) +
  geom_histogram(bins = 30, position = "identity", alpha = 0.5) +
  labs(title = "Histogram of typeI_fin",
       x = "typeI_fin",
       y = "Count")
```




```
ggplot(loan_selected, aes(x = typeII_fin, fill = variant)) +
  geom_histogram(bins = 30, position = "identity", alpha = 0.5) +
  labs(title = "Histogram of typeII_fin",
       x = "typeII_fin",
       y = "Count")
```




```
ggplot(loan_selected, aes(x = complt_fin, fill = variant)) +
  geom_histogram(bins = 30, position = "identity", alpha = 0.5) +
  labs(title = "Histogram of complt_fin",
       x = "complt_fin",
       y = "Count")
```




```
# description
( su_stats <- loan_selected %>%
  group_by(variant) %>%
  summarise(
    count = n(),
    mean_typeI_fin = mean(typeI_fin),
    sd_typeI_fin = sd(typeI_fin),
    min_typeI_fin = min(typeI_fin),
    max_typeI_fin = max(typeI_fin),
    mean_typeII_fin = mean(typeII_fin),
    sd_typeII_fin = sd(typeII_fin),
    min_typeII_fin = min(typeII_fin),
    max_typeII_fin = max(typeII_fin),
    mean_complt_fin = mean(complt_fin),
    sd_complt_fin = sd(complt_fin),
    min_complt_fin = min(complt_fin),
    max_complt_fin = max(complt_fin)
  )
)
```

```
## # A tibble: 2 × 14
##   variant   count mean_typeI_fin sd_typeI_fin min_typeI_fin max_typeI_fin
##   <fct>     <int>          <dbl>        <dbl>         <dbl>         <dbl>
## 1 Control     190           1.85         2.13             0             8
## 2 Treatment   280           1.94         1.38             0             8
## # ℹ 8 more variables: mean_typeII_fin <dbl>, sd_typeII_fin <dbl>,
## #   min_typeII_fin <dbl>, max_typeII_fin <dbl>, mean_complt_fin <dbl>,
## #   sd_complt_fin <dbl>, min_complt_fin <dbl>, max_complt_fin <dbl>
```

```
# check complt_fin == 0 records
loan_selected %>%
  filter(complt_fin == 0) %>%
  group_by(variant,loanofficer_id) %>%
  count()
```

```
## # A tibble: 9 × 3
## # Groups:   variant, loanofficer_id [9]
##   variant loanofficer_id     n
##   <fct>   <chr>          <int>
## 1 Control 2udootyt          10
## 2 Control itlmccd6          10
## 3 Control keltu0gq          10
## 4 Control l31kzq2d          10
## 5 Control q5pea8jk          10
## 6 Control rfinwi4z          10
## 7 Control rot0rb2t          10
## 8 Control ujxyy9v2          10
## 9 Control zqr650mp          10
```

Here we can see that there are 9 loan officers in the control group who are not interacting with the AI models throughout the entire experiment period (10 days). While this is telling that uptake is far larger for the treatment group than the control, it also confounds the currently used OEC of change in final TypeII rate, meaning these values must be removed.

Data Preparation

```
# Exclude records with complt_fin == 0, because these officers have not interacted with AI predictions.
loan_filtered <- loan_selected %>% filter(complt_fin != 0)

# Check sample size after excluding records with complt_fin == 0,
(sample_size.af <- loan_filtered %>%
  group_by(variant) %>%
  summarize(
    loanofficer_count = n_distinct(loanofficer_id)
    )
)
```

```
## # A tibble: 2 × 2
##   variant   loanofficer_count
##   <fct>                 <int>
## 1 Control                  10
## 2 Treatment                28
```

```
# Aggregate the data to the officer-level, because randomization unit of this experiment is the officer
officer_level_data <- loan_filtered %>%
  group_by(variant, loanofficer_id) %>%
  summarise(
     typeI_fin_sum = sum(typeI_fin),
     typeII_fin_sum = sum(typeII_fin),
     complt_fin_sum = sum(complt_fin),
     .groups = "drop"
  )


# calculate typeII_fin error rate and typeI_fin error rate for each officer
officer_level_data <- officer_level_data %>%
 mutate(
     typeII_fin_rate_per_officer = typeII_fin_sum/complt_fin_sum, #oec
     typeI_fin_rate_per_officer = typeI_fin_sum/complt_fin_sum  #supplementary metric
  )

officer_level_data
```

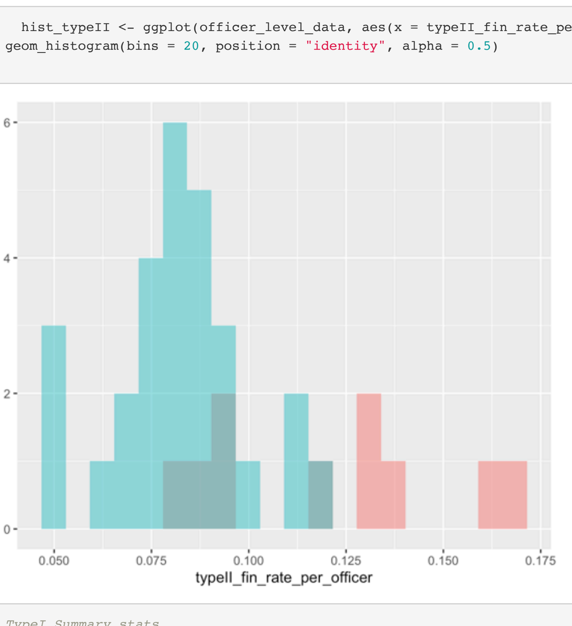
```
[illegible]
```

Data Analysis: Hypothesis Testing

[illegible]

```
[illegible]
```

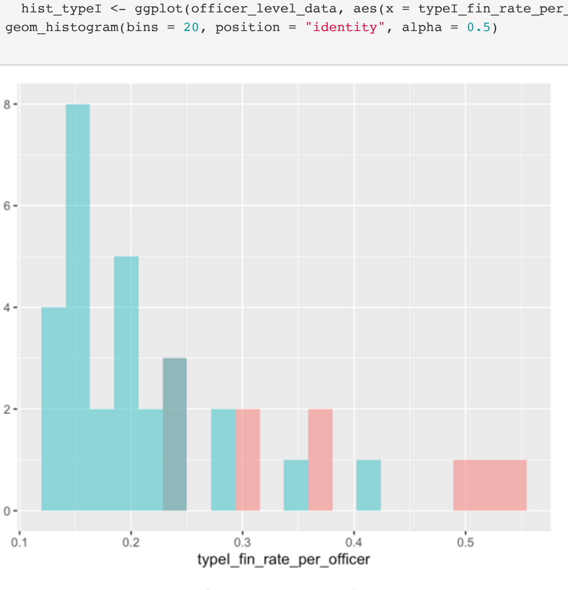
```
[illegible]
```




```
[illegible]
```

```
[illegible]
```

```
[illegible]
  geom_histogram(bins = 20, position = "identity", alpha = 0.5)
)
```



As we can see, standard deviation (and therefore variance) differs between these two groups, so we need to do Welch's t-tests.

Run Welch's two-sample t-tests to examine if there's sig. difference between 2 Variants

Final TypeII Rate

```
t.test(
  typeII_fin_rate_per_officer ~ variant,
  data = officer_level_data,
  var.equal = FALSE,
  alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  typeII_fin_rate_per_officer by variant
## t = 3.6098, df = 10.894, p-value = 0.002081
## alternative hypothesis: true difference in means between group Control and group Treatment is greater than 0
## 95 percent confidence interval:
##  0.0192293       Inf
## sample estimates:
##   mean in group Control mean in group Treatment
##              0.12088708              0.08258573
```

Here we see that there is a significant difference in means with p = 0.002081. Treatment (new AI model) significantly *decreased* typeII_fin_rate_per_officer compared to Control (existing AI model).

Final TypeI Rate

```
t.test(
  typeI_fin_rate_per_officer ~ variant,
  data = officer_level_data,
  var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  typeI_fin_rate_per_officer by variant
## t = 3.9494, df = 10.927, p-value = 0.002304
## alternative hypothesis: true difference in means between group Control and group Treatment is not equal to 0
## 95 percent confidence interval:
##  0.07204903 0.25377808
## sample estimates:
##   mean in group Control mean in group Treatment
##               0.3630571               0.2001435
```

Here we see that there is a significant difference in means with p = 0.002304. Treatment (new AI model) significantly *decreased* typeI_fin_rate_per_officer compared to Control (existing AI model).

Data Analysis

Compute Difference in Mean OEC between Variants

```
# Compute mean OEC (final typeII rate) for each Variant
mean_OEC_each_Variant <- officer_level_data %>%
  group_by(variant) %>%
  summarise(mean_typeII_fin_rate_per_officer = mean(typeII_fin_rate_per_officer))

# View mean OEC
print(mean_OEC_each_Variant)
```

```
## # A tibble: 2 × 2
##   variant   mean_typeII_fin_rate_per_officer
##   <fct>                                <dbl>
## 1 Control                              0.121
## 2 Treatment                            0.0826
```

```
# Compute pairwise % differences in OEC between variants
pairwise_diff <- mean_OEC_each_Variant %>%
  summarise(
    Diff = mean_typeII_fin_rate_per_officer[variant == "Treatment"] - mean_typeII_fin_rate_per_officer[variant ==
"Control"],
    Percentage = (Diff / mean_typeII_fin_rate_per_officer[variant == "Control"]) *100
  )

# View pairwise differences
print(pairwise_diff)
```

```
## # A tibble: 1 × 2
##      Diff Percentage
##     <dbl>      <dbl>
## 1 -0.0383      -31.7
```

Treatment (new AI model) significantly *decreased* (p = 0.002081) typeII_fin_rate_per_officer compared to Control (existing AI model) by *31.68%*.

Compute Difference in Mean Additional Metric between Variants

```
# Compute Mean Additional Metric (final typeI rate) for each Variant
mean_additional_metric_each_Variant <- officer_level_data %>%
  group_by(variant) %>%
  summarise(mean_typeI_fin_rate_per_officer = mean(typeI_fin_rate_per_officer))

# View mean Additional Metric (final typeI rate)
print(mean_additional_metric_each_Variant)
```

```
## # A tibble: 2 × 2
##   variant   mean_typeI_fin_rate_per_officer
##   <fct>                               <dbl>
## 1 Control                             0.363
## 2 Treatment                           0.200
```

```
# Compute pairwise % differences in Additional Metric (final typeI rate) between variants
pairwise_diff.ad <- mean_additional_metric_each_Variant %>%
  summarise(
    Diff = mean_typeI_fin_rate_per_officer[variant == "Treatment"] - mean_typeI_fin_rate_per_officer[variant == "
Control"],
    Percentage = (Diff / mean_typeI_fin_rate_per_officer[variant == "Control"]) *100
  )

# View pairwise differences
print(pairwise_diff.ad)
```

```
## # A tibble: 1 × 2
##     Diff Percentage
##    <dbl>      <dbl>
## 1 -0.163      -44.9
```

Treatment (new AI model) significantly *decreased* (p = 0.002304) typeI_fin_rate_per_officer compared to Control (existing AI model) by *44.87%*.

The decrease of 31.68% on typeII_fin_rate_per_officer and 44.87% on typeI_fin_rate_per_officer by treatment compared to control is *practically significant*.

Data Analysis: Compute & Interpret Effect Size (Cohen's d)

Effect size: Control vs Treatment

OEC - final typeII rate

```
Control = officer_level_data$typeII_fin_rate_per_officer[officer_level_data$variant == "Control"]
Treatment = officer_level_data$typeII_fin_rate_per_officer[officer_level_data$variant == "Treatment"]

cohens_d(Treatment, Control)
```

```
## Cohen's d |         95% CI
## --------------------------
## -1.76     | [-2.58, -0.92]
##
## - Estimated using pooled SD.
```

Additional Metric - final typeI rate

```
Control.ad = officer_level_data$typeI_fin_rate_per_officer[officer_level_data$variant == "Control"]
Treatment.ad = officer_level_data$typeI_fin_rate_per_officer[officer_level_data$variant == "Treatment"]

cohens_d(Treatment.ad, Control.ad)
```

```
## Cohen's d |         95% CI
## --------------------------
## -1.92     | [-2.75, -1.06]
##
## - Estimated using pooled SD.
```

Interpreting Effect Sizes:

```
effectsize::interpret_cohens_d(-1.76) # OEC
```

```
## [1] "large"
## (Rules: cohen1988)
```

```
effectsize::interpret_cohens_d(-1.92) # Additional Metric
```

```
## [1] "large"
## (Rules: cohen1988)
```

Both effect sizes in OEC and additional metric are large.

Treatment (new AI model) significantly *reduced* ($p < 0.05$, $d = -1.76$) typeII_fin_rate_per_officer compared to Control (existing AI model) by 31.68%.

Calculate the Power (Based only on OEC)

```
library(pwr)
d <- 1.76  # Cohen's d
p <- 0.05  # p_value set to the threshold
n1 <- 10  # control size
n2 <- 28  # treatment size

# Calculate Power
power_result <- pwr.t2n.test(d = d, n1 = n1, n2 = n2, sig.level = p, alternative = "greater")
print(power_result$power)
```

```
## [1] 0.9988189
```