|  |  |
| --- | --- |
| SOP No. |  |
| Title |  |
|  |  |
| Author: |  |
|  |  |
| Reviewed: |  |
|  |  |
| Approved: |  |
|  |  |

|  |  |
| --- | --- |
| Document information | |
| Revision period | 2 year |

1. Check library loading correctness:

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(DBI)  
library(RSQLite)  
library(rmarkdown)  
library(writexl)  
  
  
if (!all(c("dplyr", "ggplot2", "DBI", "RSQLite", "rmarkdown", "writexl") %in% .packages())) {  
 stop("Some libraries failed to load.")  
} else {  
 cat("All libraries loaded successfully.\n")  
}

## All libraries loaded successfully.

1. Simulate the data and display header:

set.seed(123)  
simulated\_data <- data.frame(  
 ID = 1:100,  
 Group = sample(c("Control", "Treatment"), 100, replace = TRUE),  
 Age = round(runif(100, 20, 60), 0),  
 Score\_Pre = round(runif(100, 50, 80), 1),  
 Score\_Post = round(runif(100, 60, 90), 1)  
)  
  
head(simulated\_data)

## ID Group Age Score\_Pre Score\_Post  
## 1 1 Control 44 57.2 83.5  
## 2 2 Control 33 78.9 60.3  
## 3 3 Control 40 68.0 83.4  
## 4 4 Treatment 58 65.5 81.9  
## 5 5 Control 39 62.1 78.9  
## 6 6 Treatment 56 76.4 74.4

1. Data manipulation:

* Create a new column for score difference
* Merge the data with a new table
* Filter for Treatment group only
* Display header.

manipulated\_data <- simulated\_data %>%  
 mutate(Score\_Diff = Score\_Post - Score\_Pre)  
  
extra\_data <- data.frame(  
 ID = sample(1:100, 50),  
 Extra\_Info = sample(c("High", "Low"), 50, replace = TRUE)  
)  
merged\_data <- manipulated\_data %>%  
 left\_join(extra\_data, by = "ID")  
  
filtered\_data <- merged\_data %>%  
 filter(Group == "Treatment")  
  
head(filtered\_data)

## ID Group Age Score\_Pre Score\_Post Score\_Diff Extra\_Info  
## 1 4 Treatment 58 65.5 81.9 16.4 <NA>  
## 2 6 Treatment 56 76.4 74.4 -2.0 <NA>  
## 3 7 Treatment 57 60.9 64.7 3.8 <NA>  
## 4 8 Treatment 44 58.6 60.2 1.6 High  
## 5 11 Treatment 57 64.5 71.7 7.2 Low  
## 6 12 Treatment 32 57.6 73.9 16.3 Low

1. Statistical analysis:

* Perform a t-test on Score\_Diff between Control and Treatment groups.

t\_test\_result <- t.test(  
 Score\_Diff ~ Group,  
 data = manipulated\_data  
)  
print(t\_test\_result)

##   
## Welch Two Sample t-test  
##   
## data: Score\_Diff by Group  
## t = -1.19, df = 91.246, p-value = 0.2371  
## alternative hypothesis: true difference in means between group Control and group Treatment is not equal to 0  
## 95 percent confidence interval:  
## -7.727674 1.937384  
## sample estimates:  
## mean in group Control mean in group Treatment   
## 8.935088 11.830233

* Perform a linear regression and ANOVA on Score\_Pre and Score\_Post.

regression\_model <- lm(Score\_Post ~ Score\_Pre + Group, data = simulated\_data)  
summary(regression\_model)

##   
## Call:  
## lm(formula = Score\_Post ~ Score\_Pre + Group, data = simulated\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.5992 -7.2040 -0.1892 7.3229 15.9709   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 69.54978 6.59136 10.552 <2e-16 \*\*\*  
## Score\_Pre 0.06418 0.10016 0.641 0.523   
## GroupTreatment 2.48833 1.77812 1.399 0.165   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.8 on 97 degrees of freedom  
## Multiple R-squared: 0.02342, Adjusted R-squared: 0.003287   
## F-statistic: 1.163 on 2 and 97 DF, p-value: 0.3168

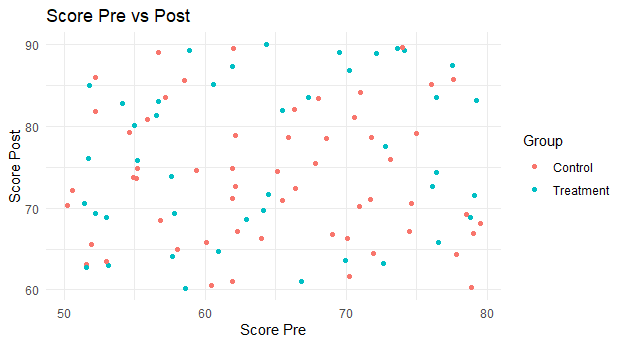
* Perform ANOVA on Score\_Post between Control and Treatment groups.

anova\_result <- aov(Score\_Post ~ Group, data = simulated\_data)  
summary(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Group 1 148 148.38 1.927 0.168  
## Residuals 98 7544 76.98

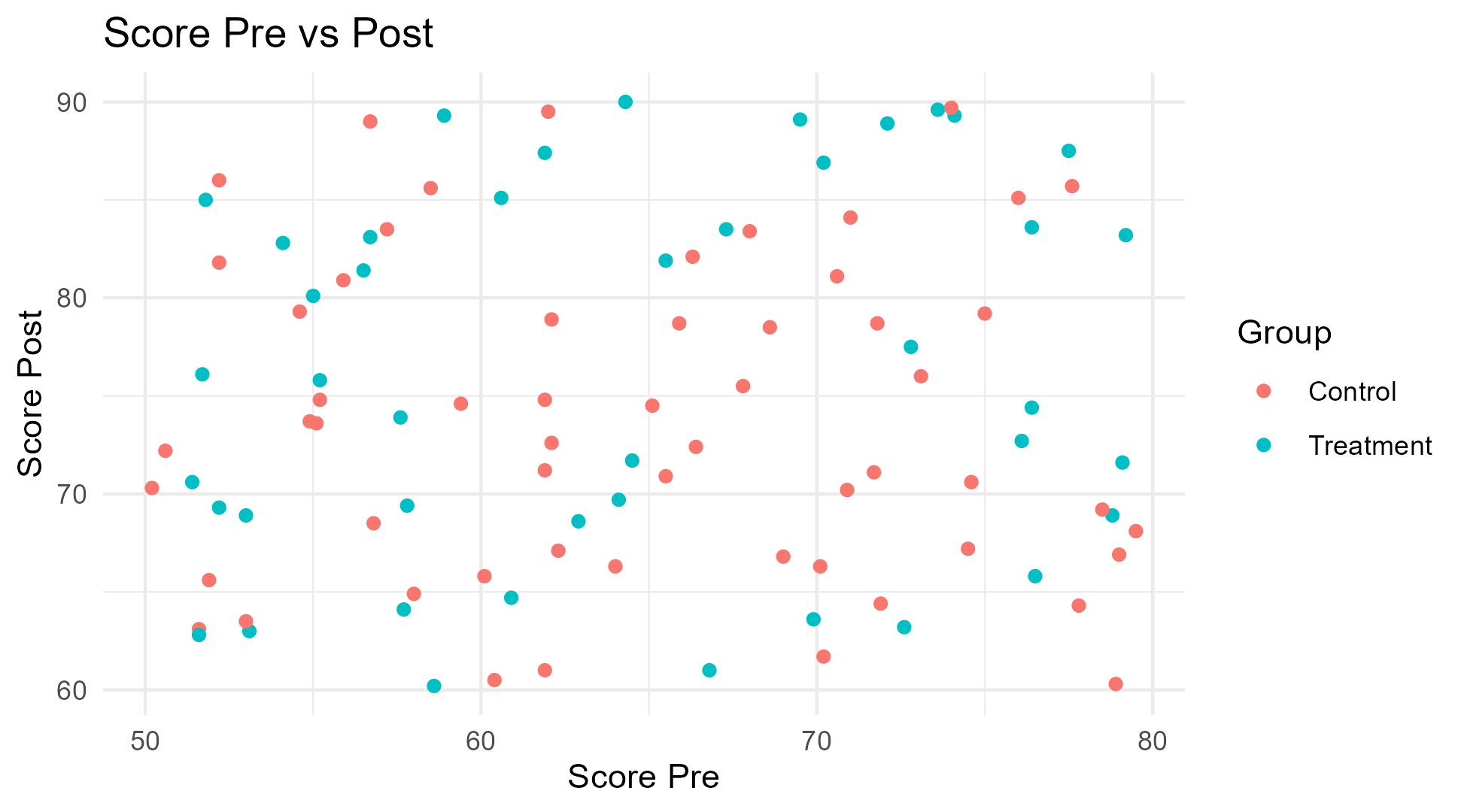
1. Graphical representation:

scatter\_plot <- ggplot(simulated\_data, aes(x = Score\_Pre, y = Score\_Post, color = Group)) +  
 geom\_point() +  
 labs(title = "Score Pre vs Post", x = "Score Pre", y = "Score Post") +  
 theme\_minimal()  
print(scatter\_plot)



1. Check correctness of data export as .png and subsequent reporting:

# Save the plot  
ggsave("scatter\_plot.png", scatter\_plot, width = 6.5, height = 3.6)



Scatter plot of Score Pre vs Post

1. Check correctness of data export as .xlsx:

# 5. Data Export and Reporting  
# Export data to Excel  
write\_xlsx(manipulated\_data, "manipulated\_data.xlsx")  
  
# Check if the file exists in the current working directory  
if (file.exists("manipulated\_data.xlsx")) {  
 cat("The file 'manipulated\_data.xlsx' exists in the working directory.\n")  
} else {  
 cat("The file 'manipulated\_data.xlsx' does NOT exist in the working directory.\n")  
}

## The file 'manipulated\_data.xlsx' exists in the working directory.