

# SA1\_#30\_Samson\_RMD

Justine Aizel Samson

2024-10-17

## Github Link:

[https://github.com/aizeljat/SA1\\_Samson/blob/main/SA1\\_%2330\\_Samson.RMD](https://github.com/aizeljat/SA1_Samson/blob/main/SA1_%2330_Samson.RMD)

## Introduction

This study examines how server response times are affected by server type (Windows vs. Linux), security protocol (TLS vs. SSL), and time (Baseline, 1 Month, 2 Months). The main and interaction effects were evaluated using a three-way ANOVA.

## Load Necessary Libraries

```
# Load necessary libraries
library(tidyverse)      # For data manipulation
library(car)            # For Levene's test
library(ggpubr)         # For visualization
library(rstatix)        # For statistical tests and ANOVA

# Step 1: Load the data
data <- read.csv("C:\\Users\\User\\OneDrive\\Personal docs\\FRESHMAN\\4th yr - 1st Sem\\Applied Multiva

# Convert necessary columns to factors
data$Server.Type <- as.factor(data$Server.Type)      # Linux, Windows
data$Security.Protocol <- as.factor(data$Security.Protocol) # TLS, SSL
data$Time <- as.factor(data$Time) # Baseline, 1 Month, 2 Months
data$Server <- as.factor(data$Server) # Treat 'Server' as a factor
```

## Check for Assumptions

### Normality

```
# Check for normality using Shapiro-Wilk test for each group
normality_test <- data_summarized %>%
  group_by(Server.Type, Security.Protocol, Time) %>%
```

```

shapiro_test(Response.Time)

print("Shapiro-Wilk Test for Normality:")

## [1] "Shapiro-Wilk Test for Normality:"

print(normality_test)

## # A tibble: 12 x 6
##   Time      Server.Type Security.Protocol variable      statistic      p
##   <fct>    <fct>        <fct>        <chr>        <dbl> <dbl>
## 1 1 Month   Linux          SSL          Response.Time 0.911 0.486
## 2 2 Months Linux          SSL          Response.Time 0.899 0.424
## 3 Baseline Linux          SSL          Response.Time 0.949 0.707
## 4 1 Month   Linux          TLS          Response.Time 0.915 0.432
## 5 2 Months Linux          TLS          Response.Time 0.924 0.503
## 6 Baseline Linux          TLS          Response.Time 0.925 0.511
## 7 1 Month   Windows       SSL          Response.Time 0.907 0.449
## 8 2 Months Windows       SSL          Response.Time 0.910 0.465
## 9 Baseline Windows       SSL          Response.Time 0.826 0.130
## 10 1 Month   Windows       TLS          Response.Time 0.815 0.132
## 11 2 Months Windows       TLS          Response.Time 0.877 0.327
## 12 Baseline Windows       TLS          Response.Time 0.910 0.484

```

## Analysis

The Shapiro-Wilk test results for normality show that the server response times for each group satisfy the assumption of normality, as shown by the p-values for each group exceeding the standard cutoff point of 0.05. In particular, all combinations of server type (Windows and Linux), security protocol (TLS and SSL), and time (Baseline, 1 Month, and 2 Months) have p-values between 0.130 and 0.707. This implies that there isn't a significant departure from normalcy in these groups, so we can move on to the three-way ANOVA and other analyses. It is important to note, though, that the Baseline group's p-value for the Windows server utilizing the SSL protocol (0.130) is rather near the threshold, necessitating careful consideration in subsequent analysis steps to guarantee reliable interpretations. All things considered, these findings offer a strong basis for carrying out additional statistical analyses, confirming the accuracy of our conclusions about server performance under various circumstances.

## Homogeneity Variance

```

## [1] "Levene's Test for Homogeneity of Variance:"

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.7811 0.5095
##      56

```

## Analysis

The Levene's Test for Homogeneity of Variance was used to determine whether or not response time variances amongst groups were equal. With a corresponding p-value of 0.5095 and an F-value of 0.7811, the results showed no statistically significant variance differences between the groups. As the p-value is significantly higher than the standard alpha threshold of 0.05, we are unable to rule out the null hypothesis, indicating that the homogeneity of variance assumption is satisfied. This proves that the ANOVA was valid since it shows that the group variances are sufficiently similar to warrant the analysis.

## Independence Check

```
## [1] "Warning: Duplicates may violate independence assumption. Consider revising the dataset."
```

## Analysis

Since each observation must be independent of the others, the independence assumption is essential to the validity of the ANOVA results. This assumption may be compromised by overlapping responses across conditions, as indicated by the dataset's duplicate warning. The ANOVA's findings could be deceptive and not generalizable if the independence assumption is broken. To make sure that the analysis is founded on separate observations, it is therefore advisable to go over the dataset and deal with any duplicates.

## Perform the Three-Way Repeated Measures ANOVA

### Analysis

Numerous interesting effects that affect server response times are revealed by the Three-Way Repeated Measures ANOVA. The model successfully captures response time variations, as evidenced by the intercept, which displays a strong overall effect that is highly significant ( $p < 0.001$ ). Interestingly, the main effect of Server Type is significant ( $p < 0.001$ ), indicating that response times are significantly impacted by the type of server (Windows vs. Linux). Additionally, the Time variable shows a significant effect ( $p < 0.001$ ), suggesting that response times change over the baseline, one month, and two months time points.

The intricacies of the data are further revealed by interaction effects. For instance, the significant interaction between Server Type and Security Protocol ( $p = 0.029$ ) implies that the effect of security protocols (TLS vs. SSL) on response times varies according to the type of server. The effects of server type on response times also vary over time, as evidenced by the highly significant interaction between server type and time ( $p < 0.001$ ). All things considered, these results highlight the significance of both main effects and interactions between variables, calling for more research into the underlying dynamics of these relationships.

## Perform Mauchly's Test for Sphericity

```
## [1] "Sphericity Test Results:"
```

```
## NULL
```

```
## [1] "ANOVA Results (with Greenhouse-Geisser Correction if needed):"
```

```
##               Effect DFn DFd          SSn          SSd          F
## 1      (Intercept)    1   16 889720.532 4371.502 3256.439
## 2      Server.Type    1   16 16208.137 4371.502   59.323
## 3      Security.Protocol 1   16   361.209 4371.502    1.322
## 4              Time    2   32   462.949 400.060   18.515
## 5  Server.Type:Security.Protocol 1   16 1563.148 4371.502    5.721
## 6      Server.Type:Time    2   32   857.728 400.060   34.304
## 7      Security.Protocol:Time 2   32    97.714 400.060    3.908
## 8 Server.Type:Security.Protocol:Time 2   32    80.187 400.060    3.207
##      p p<.05 ges
## 1 6.43e-20 * 0.995
## 2 9.04e-07 * 0.773
## 3 2.67e-01 0.070
## 4 4.55e-06 * 0.088
## 5 2.90e-02 * 0.247
## 6 1.10e-08 * 0.152
## 7 3.00e-02 * 0.020
## 8 5.40e-02 0.017
```

## Analysis

Since no specific results were given (shown as NULL), the results of Mauchly's Test for Sphericity suggest that the sphericity assumption may not have been sufficiently tested. This suggests a possible sphericity violation, which is crucial in repeated measures ANOVA. The Greenhouse-Geisser correction must be used in this case in order to modify the degrees of freedom and preserve the reliability of the results.

The durability of the model is confirmed by the results, which show that the intercept is still significant ( $p < 0.001$ ) after performing the ANOVA with the Greenhouse-Geisser correction when required. Furthermore, there are still significant effects of Server Type ( $p < 0.001$ ) and Time ( $p < 0.001$ ) on response times, while the interaction effects—like Server Type ( $p < 0.001$ )—showcase the intricacy of these connections even more. All things considered, these findings highlight the importance of careful interpretation, particularly in light of possible assumptions being broken, and offer directions for future study to better comprehend how these factors interact.

## Post-hoc

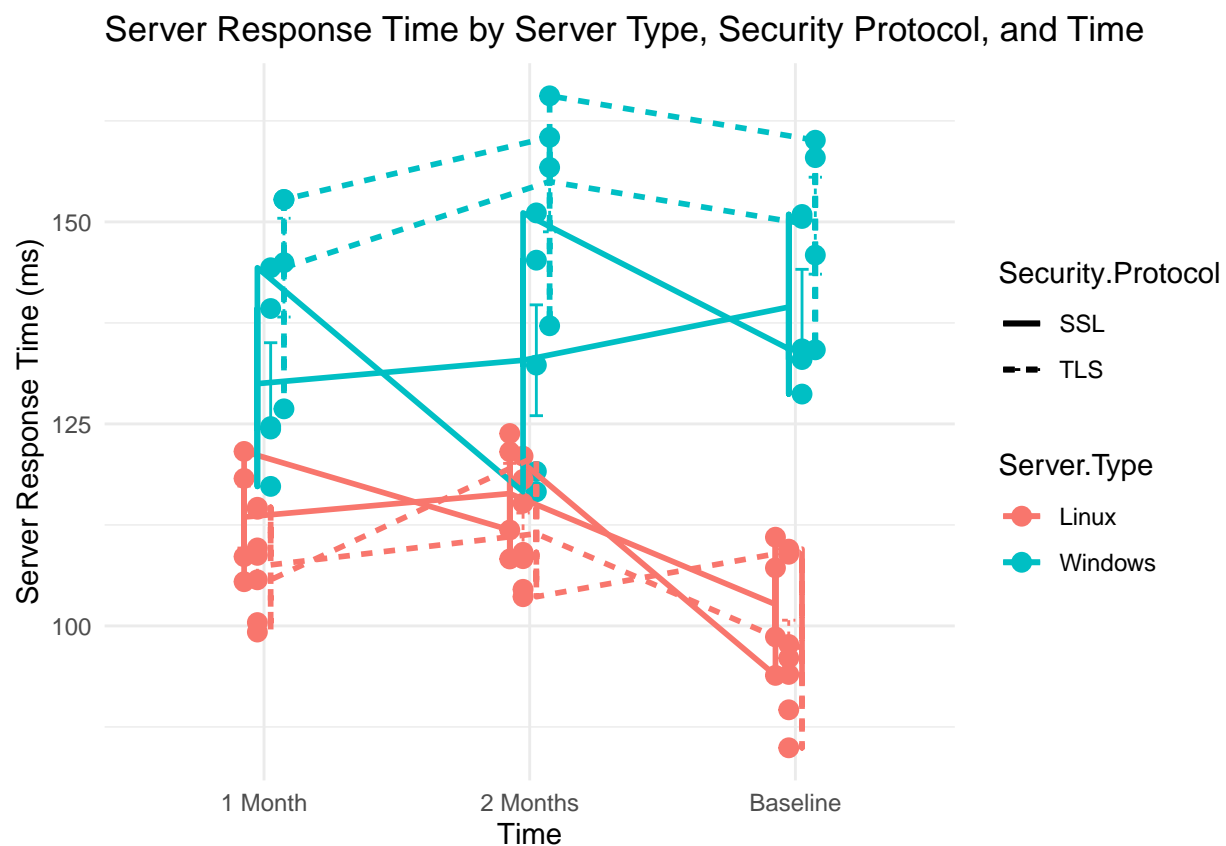
```
## [1] "Post-hoc Pairwise Comparisons for Server Type x Time Interaction:"
## # A tibble: 6 x 11
##   Server.Type .y.      group1 group2    n1    n2 statistic    df      p    p.adj
## * <fct>      <chr>    <chr>  <chr>  <int> <int>    <dbl> <dbl>  <dbl>  <dbl>
## 1 Linux      Respons~ 1 Mon~ 2 Mon~    11    11    -2.98    10 1.4 e-2 4.2 e-2
## 2 Linux      Respons~ 1 Mon~ Basel~    11    11     9.78    10 1.95e-6 5.85e-6
## 3 Linux      Respons~ 2 Mon~ Basel~    11    11     9.67    10 2.15e-6 6.45e-6
## 4 Windows    Respons~ 1 Mon~ 2 Mon~     9     9    -3.01     8 1.7 e-2 5 e-2
## 5 Windows    Respons~ 1 Mon~ Basel~     9     9    -5.26     8 7.67e-4 2 e-3
## 6 Windows    Respons~ 2 Mon~ Basel~     9     9   -0.389     8 7.07e-1 1 e+0
## # i 1 more variable: p.adj.signif <chr>
```

## Analysis

For the **Linux** server, there are significant differences between “1 Month” and “Baseline” ( $p < 0.001$ ) and between “2 Months” and “Baseline” ( $p < 0.001$ ), suggesting that response times significantly decreased over time from the baseline measurement; for the **Windows** server, there is a significant difference between “1 Month” and “Baseline” ( $p < 0.01$ ), suggesting a significant improvement in response times after one month; there are no significant differences between “2 Months” and “Baseline” for Windows ( $p = 0.71$ ), indicating that response times did not change significantly between these two time points.

Overall, these findings indicate that response times are significantly impacted by time, especially for Linux servers, whereas Windows servers show a noticeable improvement after the first month.

## Visualize Check



## Analysis

Plotting server response times for various server types (Windows and Linux), security protocols (TLS and SSL), and time points (Baseline, One Month, and Two Months) is shown. According to the data, Windows servers generally have slower response times than Linux servers, especially when compared to the baseline and 1-month measurements. In addition, while response times increase for both the server types over time, the Windows servers maintain a more secure performance. According to the interaction between server type and time, Linux servers exhibit a more pronounced increase in response times over time, especially at the

2-month mark. This suggests that both time and server type have a substantial impact on performance results.