# STAT3191/6191: Group Project Submission

**Group Number: Group 17**

(Group Leader) Vathana Khun (48302031)
Sadiqun Nur Ayan (48920959) Raja Pedapudi (46562753)

2025-09-26

## Part A: Natural Variation in Completion Times

```r
# Load required library
library(matrixStats)
library(kableExtra)
```

```r
# Simulation-Based Inference: Estimating Completion Time for Security Training
# Set seed for reproducibility
set.seed(42)

# Parameters
n <- 250              # Sample size per simulation
lambda <- 1           # Rate parameter for Exponential(1)
n_sim <- 1000         # Number of simulations

## Section 1, Part A : Natural Variation in Completion Times ##

# Step 1: Generate 1000 samples of size 250 from Exponential(1)
samples <- replicate(n_sim, rexp(n, rate = lambda))

# Step 2: Compute estimators for each sample
sample_means <- colMeans(samples)
sample_medians <- apply(samples, 2, median)
sample_trimmed_means <- apply(samples, 2, function(x) mean(x, trim = 0.1))

# Step 3: Calculate Bias, Monte Carlo Standard Deviation (MCSD), and Mean Squared Error (MSE
true_mean <- 1  # Theoretical mean of Exponential(1)
```

```r
# Bias = mean of estimates - true mean
bias_mean <- mean(sample_means) - true_mean
bias_median <- mean(sample_medians) - true_mean
bias_trimmed <- mean(sample_trimmed_means) - true_mean

# Monte Carlo Standard Deviation = standard deviation of estimates
mcsd_mean <- sd(sample_means)
mcsd_median <- sd(sample_medians)
mcsd_trimmed <- sd(sample_trimmed_means)

# Mean Squared Error = mean squared difference from true mean
mse_mean <- mean((sample_means - true_mean)^2)
mse_median <- mean((sample_medians - true_mean)^2)
mse_trimmed <- mean((sample_trimmed_means - true_mean)^2)

# Step 4: Summarise results in a table
results_partA <- data.frame(
  Estimator = c("Mean", "Median", "Trimmed Mean (10%)"),
  Bias = c(bias_mean, bias_median, bias_trimmed),
  MCSD = c(mcsd_mean, mcsd_median, mcsd_trimmed),
  MSE = c(mse_mean, mse_median, mse_trimmed)
)
results_partA%>%
  kable()%>%
  kable_classic(full_width = F, html_font = "Cambria")
```

| Estimator | Bias | MCSD | MSE |
|---|---|---|---|
| Mean | -0.0017348 | 0.0641395 | 0.0041128 |
| Median | -0.3068622 | 0.0637944 | 0.0982301 |
| Trimmed Mean (10%) | -0.1701139 | 0.0587640 | 0.0323885 |

## Section 1, Part B: Logging Error Introduces Outliers #

```r
# Step 1: Generate 1000 contaminated samples:
# 90% from Exp(1), 10% from Exp(0.02) to simulate contamination
lambda1 <- 1
lambda2 <- 0.02
contamination_rate <- 0.1
```

```r
samples_contaminated <- matrix(NA, nrow = n, ncol = n_sim)

for (i in 1:n_sim) {
  n1 <- round(n * (1 - contamination_rate))  # number from main distribution
  n2 <- n - n1                               # number from contamination

  # Generate and combine samples
  samples_contaminated[, i] <- c(rexp(n1, rate = lambda1), rexp(n2, rate = lambda2))

  # Shuffle to mix contaminated values randomly
  samples_contaminated[, i] <- sample(samples_contaminated[, i])
}

# Step 2: Compute estimators for contaminated samples
sample_means_cont <- colMeans(samples_contaminated)
sample_medians_cont <- apply(samples_contaminated, 2, median)
sample_trimmed_means_cont <- apply(samples_contaminated, 2, function(x) mean(x, trim = 0.1))

# Step 3: Calculate Bias, MCSD, MSE for contaminated data
# True mean for mixture: 0.9 * 1/lambda1 + 0.1 * 1/lambda2
true_mean_cont <- 0.9 * (1 / lambda1) + 0.1 * (1 / lambda2)  # = 5.9

bias_mean_cont <- mean(sample_means_cont) - true_mean_cont
bias_median_cont <- mean(sample_medians_cont) - true_mean_cont
bias_trimmed_cont <- mean(sample_trimmed_means_cont) - true_mean_cont

mcsd_mean_cont <- sd(sample_means_cont)
mcsd_median_cont <- sd(sample_medians_cont)
mcsd_trimmed_cont <- sd(sample_trimmed_means_cont)

mse_mean_cont <- mean((sample_means_cont - true_mean_cont)^2)
mse_median_cont <- mean((sample_medians_cont - true_mean_cont)^2)
mse_trimmed_cont <- mean((sample_trimmed_means_cont - true_mean_cont)^2)

# Step 4: Summarise contaminated results in a table
results_partB <- data.frame(
  Estimator = c("Mean", "Median", "Trimmed Mean (10%)"),
  Bias = c(bias_mean_cont, bias_median_cont, bias_trimmed_cont),
  MCSD = c(mcsd_mean_cont, mcsd_median_cont, mcsd_trimmed_cont),
  MSE = c(mse_mean_cont, mse_median_cont, mse_trimmed_cont)
)
```

```
results_partB%>%
  kable()%>%
  kable_classic(full_width = F, html_font = "Cambria")
```

| Estimator | Bias | MCSD | MSE |
|---|---|---|---|
| Mean | -0.0017397 | 1.0242196 | 1.04798 |
| Median | -5.0862454 | 0.0751779 | 25.87554 |
| Trimmed Mean (10%) | -4.8103352 | 0.0771154 | 23.14527 |