3E3 Probability and Statistics

Data Analysis Assignment 2 and 3

Group 29: Kerem Aktaş, Emma Bagnall, Joseph Dwan, Megi Murvanidze, Finn O'Connor

Table of Contents

Group Assignment 2 – Fitting a Probability Distribution

Introduction

This assignment carries on from the data collected in Part A of Assignment 1, and builds on the analysis conducted in Part B. In the previous assignment, each group compiled 9 datasets representing measurements taken at different times of the day across 3 different locations, denoted as A, B, and C. For this exercise we took these datasets and consolidated them into 3 individual ones, each representing their respective location data.

We used MATLAB to extract the data from our original dataset and create 3 new CSV files each representing each location. We then plotted 3 histograms for each of the 3 locations. The primary objective was to fit suitable continuous probability distributions to these datasets and evaluate which distribution best models the observed data at each location. The distributions used were the log-normal, Weibull, and exponential distributions. Each of the histograms plotted were overlayed with each of the distributions. This resulted in 3 different distributions being superimposed on each of the histograms. Through visual analysis, estimating the parameters of each distribution, and comparing their goodness-of-fit, we were able to determine which distributions gave the best representation of each respective dataset.

Through this process, we aimed to gain a deeper insight into the probabilistic structure of the data collected at each location and determine the most appropriate distribution model for each case.

Methods

The first task was to compile the data using MATLAB. From the previous assignment, we ended up with 9 sets of data for each of the 3 locations. We took this data and consolidated it into 3 individual datasets, one for each location. The larger dataset was analysed using our code and separated into 3 new CSV files.

```
data = readmatrix("soundlevel_Template.csv",'Range','2:151');

% Define relative column indices for times of day
morning_rel = [4, 19, 34, 49];
noon_rel = [9, 24, 39, 54];
afternoon_rel = [14, 29, 44, 59];

% Column offsets for each location block
offset_A = 0;
offset_B = 60;
offset_C = 120;

% Full absolute column indices
A_morning_cols   = offset_A + morning_rel;
A_noon_cols      = offset_A + noon_rel;
```

```
A_afternoon_cols = offset_A + afternoon_rel;

B_morning_cols   = offset_B + morning_rel;
B_noon_cols      = offset_B + noon_rel;
B_afternoon_cols = offset_B + afternoon_rel;

C_morning_cols   = offset_C + morning_rel;
C_noon_cols      = offset_C + noon_rel;
C_afternoon_cols = offset_C + afternoon_rel;

% Compute averages across selected columns
A_morning   = mean(data(:, A_morning_cols), 2);
A_noon      = mean(data(:, A_noon_cols), 2);
A_afternoon = mean(data(:, A_afternoon_cols), 2);

B_morning   = mean(data(:, B_morning_cols), 2);
B_noon      = mean(data(:, B_noon_cols), 2);
B_afternoon = mean(data(:, B_afternoon_cols), 2);

C_morning   = mean(data(:, C_morning_cols), 2);
C_noon      = mean(data(:, C_noon_cols), 2);
C_afternoon = mean(data(:, C_afternoon_cols), 2);

% Combine results
avg_A = [A_morning, A_noon, A_afternoon];
avg_B = [B_morning, B_noon, B_afternoon];
avg_C = [C_morning, C_noon, C_afternoon];

% Save results
writematrix(avg_A, 'Location_A_Averages.csv');
writematrix(avg_B, 'Location_B_Averages.csv');
writematrix(avg_C, 'Location_C_Averages.csv');

dataA = readmatrix('Location_A_Averages.csv', 'Range', '2:151');
dataB = readmatrix('Location_B_Averages.csv', 'Range', '2:151');
dataC = readmatrix('Location_C_Averages.csv', 'Range', '2:151');

datasets = {dataA, dataB, dataC};
location_names = {'Location A', 'Location B', 'Location C'};
```

Each of these new datasets were plotted as histograms. Through visual analysis, we could see that each set had a different distribution of data and skewness. However, in order to try and get a more in-depth insight, we superimposed a log-normal, Weibull, and exponential distribution to each of the 3 histograms.

This enabled us to visually compare distribution models for our data and determine which one gave the best representation. To better understand our data, we estimated the parameters of each distribution. We then used these parameters to calculate the probability density functions (PDFs) of each distribution.

The formulae for each PDF are shown below:

Lognormal PDF:

Parameters = μ and σ^2

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}, \qquad \text{for } y \geq 0.$$

Weibull PDF:

Scale parameter = $\delta > 0$
Shape parameter = $\beta > 0$

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-\left(\frac{x}{\delta}\right)^{\beta}}, \qquad \text{for } x > 0.$$

Exponential PDF:

Parameter = λ

$$f(x) = \lambda e^{-\lambda x}, \qquad \text{for } 0 \leq x < \infty$$

We were able to calculate the parameters and probability density functions for each distribution across each of the 3 locations by implementing the above formulae into our MATLAB script. This whole process was done by using one large for-loop.

```
% Initialize struct to hold parameters
all_params = struct();

% Plot histograms and fitted distributions
for i = 1:3
    data = datasets{i}(:);

    figure;
    histogram(data, 20, 'Normalization', 'pdf', 'FaceColor', [0.7, 0.7, 0.9]);
    hold on;

    x = linspace(min(data), max(data), 1000);

    % Log-normal Distribution
    lognormal_dist = fitdist(data, 'Lognormal');
    y_logn = pdf(lognormal_dist, x);
    plot(x, y_logn, 'LineWidth', 2, 'DisplayName', 'Log-normal');

    % Weibull Distribution
    weibull_dist = fitdist(data, 'Weibull');
    y_weibull = pdf(weibull_dist, x);
    plot(x, y_weibull, 'LineWidth', 2, 'DisplayName', 'Weibull');
```

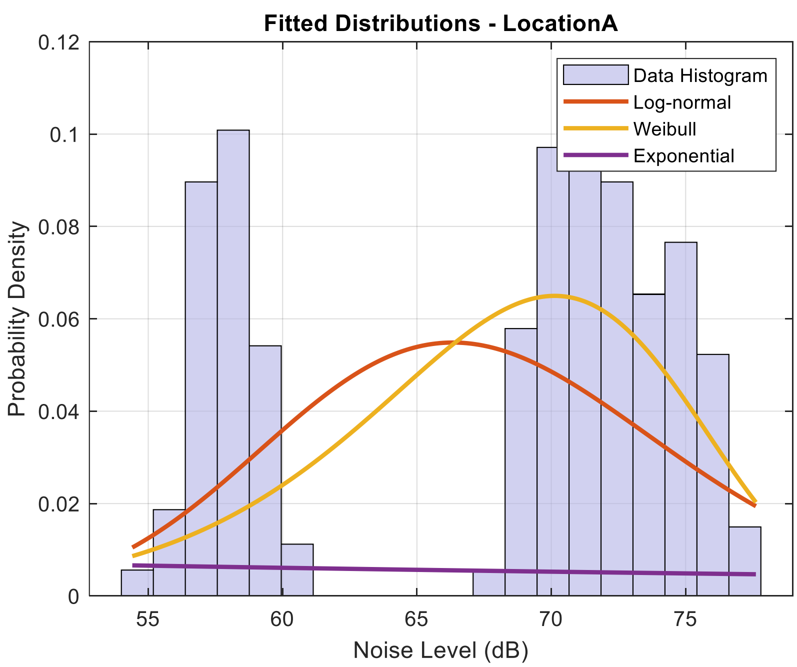
```
    % Exponential Distribution
    exponential_dist = fitdist(data, 'Exponential');
    y_exp = pdf(exponential_dist, x);
    plot(x, y_exp, 'LineWidth', 2, 'DisplayName', 'Exponential');

    title(['Fitted Distributions - ', location_names{i}]);
    xlabel('Noise Level (dB)');
    ylabel('Probability Density');
    legend('Data Histogram', 'Log-normal', 'Weibull', 'Exponential');
    grid on;
    hold off;

    % Extract parameters
    mu_logn = lognormal_dist.mu;
    sigma_logn = lognormal_dist.sigma;

    delta_weibull = weibull_dist.A; % Scale parameter
    beta_weibull  = weibull_dist.B; % Shape parameter

    mean_exp = exponential_dist.mu;
    lambda_exp = 1 / mean_exp;

    % Store in struct using dynamic array
    location = location_names{i};

    all_params.(location).Lognormal.mu = mu_logn;
    all_params.(location).Lognormal.sigma = sigma_logn;

    all_params.(location).Weibull.alpha = delta_weibull;
    all_params.(location).Weibull.beta = beta_weibull;

    all_params.(location).Exponential.mean = mean_exp;
    all_params.(location).Exponential.lambda = lambda_exp;

    % Display the parameters
    fprintf('\nParameters for %s:\n', location);
    fprintf('Log-normal:     mu = %.4f, sigma = %.4f\n', mu_logn, sigma_logn);
    fprintf('Weibull:        delta (scale) = %.4f, beta (shape) = %.4f\n',
delta_weibull, beta_weibull);
    fprintf('Exponential:    mean = %.4f, lambda = %.4f\n', mean_exp, lambda_exp);

    % Calculating the PDFs for each distribution
    pdf_logn = pdf(lognormal_dist, data);
    pdf_weibull = pdf(weibull_dist, data);
    pdf_exp = pdf(exponential_dist, data);

    % Display the PDFs
    fprintf('\nProbability Density Function statistics for %s:\n', location);
    fprintf('Log-normal PDF:  mean = %.4f, max = %.4f\n', mean(pdf_logn),
max(pdf_logn));
    fprintf('Weibull PDF:     mean = %.4f, max = %.4f\n', mean(pdf_weibull),
max(pdf_weibull));
    fprintf('Exponential PDF: mean = %.4f, max = %.4f\n', mean(pdf_exp),
max(pdf_exp));
end
```

Results

After analysing the data and fitting the distributions, we yielded numerical and graphical data. The continuous probability distributions were fitted to each of the histograms, representing each of our chosen locations. The probability density functions (PDFs) were also calculated, along with their parameters.

Location A:

Histogram with Fitted Distributions:



Parameter Calculations:

```
Parameters for LocationA:
Log-normal:     mu = 4.2066, sigma = 0.1090
Weibull:        delta (scale) = 70.6029, beta (shape) = 12.4268
Exponential:    mean = 67.5191, lambda = 0.0148
```

Probability Density Function (PDF) Calculations:

```
Probability Density Function statistics for LocationA:
Log-normal PDF:  mean = 0.0348, max = 0.0540
Weibull PDF:     mean = 0.0425, max = 0.0650
Exponential PDF: mean = 0.0055, max = 0.0066
```

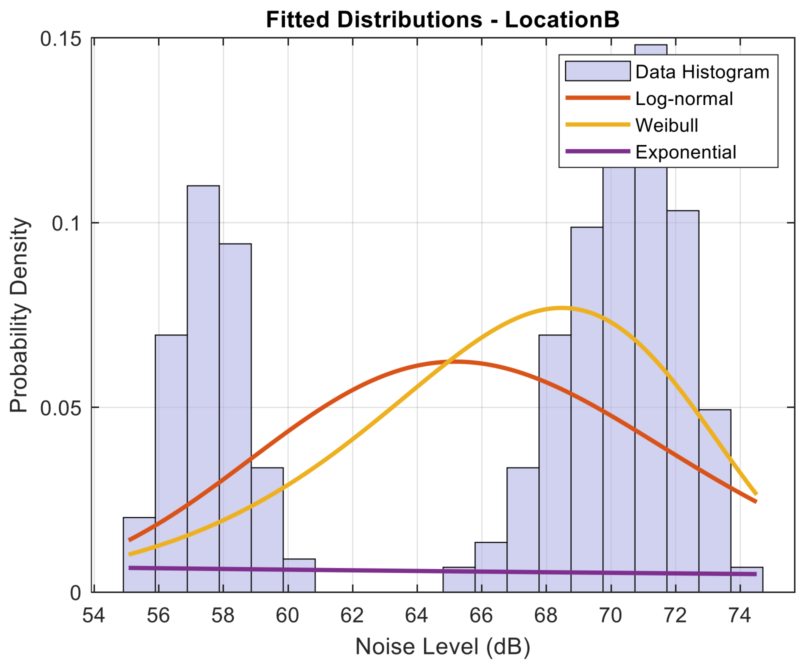
Log-normal: $f(y) = \frac{1}{0.11y\sqrt{2\pi}} e^{-\frac{(\ln(y)-4.21)^2}{2(0.11)^2}}$

Weibull: $f(x) = \frac{12.43}{70.60}\left(\frac{x}{70.60}\right)^{12.43-1} e^{-\left(\frac{x}{70.60}\right)^{12.43}}$

Exponential: $f(x) = 0.0148e^{-0.0148x}$

Location B:

Histogram with Fitted Distributions:



Parameter Calculations:

```
Parameters for LocationB:
Log-normal:     mu = 4.1867, sigma = 0.0976
Weibull:        delta (scale) = 68.8180, beta (shape) = 14.3562
Exponential:    mean = 66.1099, lambda = 0.0151
```

Probability Density Function (PDF) Calculations:

```
Probability Density Function statistics for LocationB:
Log-normal PDF:  mean = 0.0395, max = 0.0623
Weibull PDF:     mean = 0.0501, max = 0.0769
Exponential PDF: mean = 0.0056, max = 0.0066
```

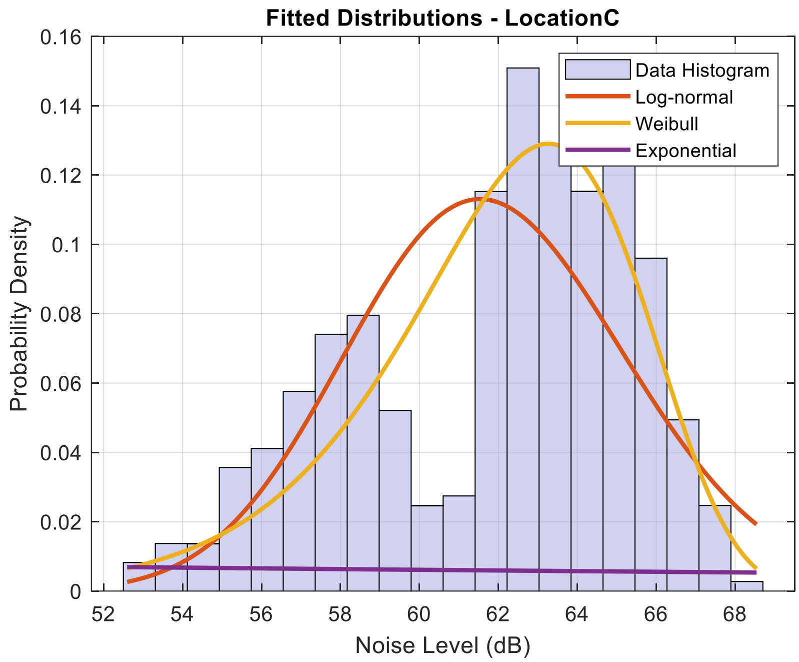
Log-normal: $f(y) = \frac{1}{0.10y\sqrt{2\pi}} e^{-\frac{(\ln(y)-4.19)^2}{2(0.10)^2}}$

Weibull: $f(x) = \frac{14.36}{68.82}\left(\frac{x}{68.82}\right)^{14.36-1} e^{-\left(\frac{x}{68.82}\right)^{14.36}}$

Exponential: $f(x) = 0.0151e^{-0.0151x}$

Location C:

Histogram with Fitted Distributions:



Parameter Calculations:

```
Parameters for LocationC:
Log-normal:     mu = 4.1229, sigma = 0.0573
Weibull:        delta (scale) = 63.4026, beta (shape) = 22.2165
Exponential:    mean = 61.8381, lambda = 0.0162
```

Probability Density Function (PDF) Calculations:

```
Probability Density Function statistics for LocationC:
Log-normal PDF:  mean = 0.0769, max = 0.1130
Weibull PDF:     mean = 0.0861, max = 0.1290
Exponential PDF: mean = 0.0060, max = 0.0069
```

Log-normal: $f(y) = \frac{1}{0.06y\sqrt{2\pi}} e^{-\frac{(\ln(y)-4.12)^2}{2(0.06)^2}}$

Weibull: $f(x) = \frac{22.22}{63.40}\left(\frac{x}{63.40}\right)^{22.22-1} e^{-\left(\frac{x}{63.40}\right)^{22.22}}$

Exponential: $f(x) = 0.0069e^{-0.0069x}$

Discussion

The fitted distribution plots for Locations A, B and C allow us to visualise, side by side, how well each theoretical model replicates the empirical noise data collected on campus. Three broad themes emerge:

1. The exponential model is inadequate: Across all sites the purple exponential curve sits almost flush with the horizontal axis, not capturing the modal regions or the heavy right tails of the histograms.
2. Either a Weibull or a log-normal distribution is needed, depending on the site: The orange log-normal and gold Weibull curves provide far closer fits, but their relative success is site-specific, reflecting differences in underlying noise-generation mechanisms.
3. Single component models struggle: Even the “best” single distribution choice sometimes struggles with secondary peaks (most evident at Locations A and B), suggesting that if finer precision were required, a mixture model could be explored. For the scope of Assignment 2, however, a single distribution that minimises bias in the upper tail is sufficient.

Location A – Best Fit: Weibull

The histogram for Location A is clearly bimodal, with clusters near 57 dB and 71 dB. The Weibull curve, flexibly shaped by a high $\beta \approx 12$, follows the dominant upper mode and produces the smallest visual error, especially in the critical 70–75 dB range. The log-normal underestimates the left-hand peak and rolls off too sharply above the main mode and the exponential remains negligible throughout the dataset.

Implication: The narrowly clustered high-volume peak suggests a steady continuous noise source (maybe a coffee machine) occasionally punctuated by lower-level events (such as background chatter). While the extra bump at ≈ 57 dB signals a second process, the Weibull’s shape parameter stretches just enough to cover it better than a log-normal.

Location B – Best Fit: Log-normal

Location B also shows a double peaked shape. The log-normal reproduces both the overall skewness and the gradual tail decay beyond 70 dB. The Weibull fits the left shoulder but decays too steeply, under-predicting higher-level events that matter for comfort criteria and again the exponential again to register meaningful probability mass.

Implication: Periodic spikes potentially from bursts of conversation drive the right tail. Because the log-normal arises naturally from multiplicative processes and retains heavier tails, it is the most defensible model for Location B.

Location C – Best Fit: Log-normal

Location C is less bimodal than the previous two and shows a more defined peak around 63dB. The log-normal hugs both the modal region and the tail, providing the closest overall fit. The Weibull models the peak well but decays too rapidly on either side. The exponential again is a pointless fit.

Implication: Noise here is less bimodal but more variable than at the other sites, characteristics well captured by the log-normal distribution.

Conclusion

The analysis confirms that campus-noise data are best described by flexible, skew-aware distributions rather than the simple exponential model. A Weibull distribution most closely captures the bimodal pattern observed at Location A, while a log-normal distribution provides the most faithful representation of the heavier-tailed noise found at Locations B and C.

Parameter estimates such as $\beta \approx 12$ for the Weibull at Location A and $\sigma \approx 0.10$ for the log-normal at Location B confirm that the underlying sound environments differ noticeably in both variance and tail weight. Although single component fits suffice for this assignment, the bimodality seen in A and B suggests that if this assignment was conducted again with different plot types there could be a benefit from mixture models to disentangle overlapping noise sources more precisely.

Overall, adopting site specific distributions lays a solid statistical foundation for the hypothesis testing tasks in Assignment 3, where accurate modelling of percentile sound levels will be essential for comparing locations and scenarios across the wider campus.

Group Assignment 3 - Comparison of Datasets

Introduction

This study aims to determine which campus locations are noisier by statistically analysing sound level data collected from both cafeterias and campus gates. As part of our investigation, we initially investigated the noise levels across three campus cafeterias including the Forum Café (location A), the Buttery (location B) and Goldsmith Hall (location C) during different times of the day. To complete the comparison, we obtained recorded noise data from group 22, who analysed the noise levels at three campus gates including the Front Gate (location A), Nassau St. Gate (location B) and Pearse St. Gate (location C). This allowed for both within scenario (cafeterias only) and between scenario (cafeterias Vs gates) comparisons.

Using statistical hypothesis testing, particularly two sample t - tests, we analysed the differences in average noise level across the locations. The aim was to determine weather any of the observed differences in noise levels were statistically significant and ultimately, to identify which specific locations are associated with higher noise levels on campus.

Methods

Data import and preprocessing

Noise level data were stored in csv files. The MATLAB `readmatrix` function was used to extract numerical values.

The datasets were structured such that noise readings from different locations and times of day were grouped into specific column blocks. For each location (A, B and C) columns were grouped in offsets of 60, and within each block, the following column indices represented the different times of day:

- Morning: columns – 4, 19, 34, 49
- Noon: columns – 9,24,39,54
- Afternoon: columns – 14,29,44,59

The `mean` function was applied across these columns to obtain a single average noise level per time slot for each observation (150 rows). This process was repeated for all three locations in both scenarios, resulting in structured datasets. Each dataset had dimensions of 150x3, corresponding to the three time periods.

```
data_cafes = readmatrix("soundlevel_Template.csv",'Range','2:151');
data_gates = readmatrix("Dataset_gates.csv",'Range','2:151');

% Define relative column indices for times of day
morning_rel = [4, 19, 34, 49];
noon_rel = [9, 24, 39, 54];
afternoon_rel = [14, 29, 44, 59];

% Column offsets for each location block
offset_A = 0;
offset_B = 60;
offset_C = 120;

% Full absolute column indices
A_morning_cols   = offset_A + morning_rel;
A_noon_cols      = offset_A + noon_rel;
A_afternoon_cols = offset_A + afternoon_rel;

B_morning_cols   = offset_B + morning_rel;
B_noon_cols      = offset_B + noon_rel;
B_afternoon_cols = offset_B + afternoon_rel;

C_morning_cols   = offset_C + morning_rel;
C_noon_cols      = offset_C + noon_rel;
C_afternoon_cols = offset_C + afternoon_rel;

% Compute averages across selected columns for cafes
A_morning_cafes   = mean(data_cafes(:, A_morning_cols), 2);
A_noon_cafes     = mean(data_cafes(:, A_noon_cols), 2);
A_afternoon_cafes = mean(data_cafes(:, A_afternoon_cols), 2);

B_morning_cafes   = mean(data_cafes(:, B_morning_cols), 2);
B_noon_cafes      = mean(data_cafes(:, B_noon_cols), 2);
B_afternoon_cafes = mean(data_cafes(:, B_afternoon_cols), 2);

C_morning_cafes   = mean(data_cafes(:, C_morning_cols), 2);
C_noon_cafes      = mean(data_cafes(:, C_noon_cols), 2);
C_afternoon_cafes = mean(data_cafes(:, C_afternoon_cols), 2);

% Combine results for cafes
dataA_cafes = [A_morning_cafes, A_noon_cafes, A_afternoon_cafes];
dataB_cafes = [B_morning_cafes, B_noon_cafes, B_afternoon_cafes];
dataC_cafes = [C_morning_cafes, C_noon_cafes, C_afternoon_cafes];

% Compute averages across selected columns for gates
A_morning_gates   = mean(data_gates(:, A_morning_cols), 2);
A_noon_gates     = mean(data_gates(:, A_noon_cols), 2);
A_afternoon_gates = mean(data_gates(:, A_afternoon_cols), 2);

B_morning_gates   = mean(data_gates(:, B_morning_cols), 2);
B_noon_gates     = mean(data_gates(:, B_noon_cols), 2);
B_afternoon_gates = mean(data_gates(:, B_afternoon_cols), 2);

C_morning_gates   = mean(data_gates(:, C_morning_cols), 2);
C_noon_gates      = mean(data_gates(:, C_noon_cols), 2);
C_afternoon_gates = mean(data_gates(:, C_afternoon_cols), 2);

% Combine results for gates
dataA_gates = [A_morning_gates, A_noon_gates, A_afternoon_gates];
```

```
dataB_gates = [B_morning_gates, B_noon_gates, B_afternoon_gates];
dataC_gates = [C_morning_gates, C_noon_gates, C_afternoon_gates];
```

Data reshaping for analysis

To prepare the data for hypothesis testing, the 150x3 matrices for each location were reshaped into single column vectors (450x1) using the `(:)` operator. This created continuous distributions of noise levels across all times of data for each location.

```
% convert to one column data sets
A_cafes = dataA_cafes(:);
B_cafes = dataB_cafes(:);
C_cafes = dataC_cafes(:);

A_gates = dataA_gates(:);
B_gates = dataB_gates(:);
C_gates = dataC_gates(:);
```

Location comparison

To evaluate whether there is a significance difference in noise levels between locations, two-sample t-test was performed using 450 data points from each location (combining morning, noon and afternoon readings).

Null Hypothesis (H_o): The mean noise levels between locations are equal

Alternative Hypothesis (H_1): The mean noise level between locations is not equal.

The MATLAB function `ttest2` was used:

```
% comparing location B and C

[H_BC, p_BC, ci_BC, stats_BC] = ttest2(B_cafes, C_cafes);


fprintf(' Location Comparison - B vs C\n');
fprintf('  p-value = %.4f\n  t-statistic = %.4f\n\n', p_BC, stats_BC.tstat);

% comparing location A and C

[H_AC, p_AC, ci_AC, stats_AC] = ttest2(A_cafes, C_cafes);

fprintf(' Location Comparison - A vs C\n');
fprintf('  p-value = %.4f\n  t-statistic = %.4f\n\n', p_AC, stats_AC.tstat);
```

Mean noise levels for each location were also calculates and compared to identify which location was objectively noisier.

```
% which location is noisier

mean_A_cafes = mean(A_cafes);
mean_B_cafes = mean(B_cafes);
mean_C_cafes = mean(C_cafes);

fprintf('  Mean Noise Levels:\n');
fprintf('  A = %.2f dB\n', mean_A_cafes);
fprintf('  B = %.2f dB\n', mean_B_cafes);
fprintf('  C = %.2f dB\n\n', mean_C_cafes);

% Find the maximum mean and identify the noisiest location
[maxMean, idx] = max([mean_A_cafes, mean_B_cafes, mean_C_cafes]);
locations = {'A', 'B', 'C'};
fprintf(' Noisiest Location: %s\n\n', locations{idx})
```

Scenario comparison

To examine whether campus gates are generally noisier than student cafeterias, we conducted a two-sample t-test comparing the average noise levels recorded in each scenario. The cafeteria dataset was formed by combining all noise measurements from locations A, B, and C across morning, noon and afternoon time periods. Similarly, the campus gate dataset included noise levels recorded at various gates during the same time intervals.

Null Hypothesis (H_o): The mean noise levels at cafeterias and campus gates are equal

Alternative Hypothesis (H_1): The mean noise level at cafeterias and campus gates are different.

```
% scenario comparison
cafeteria_all = [A_cafes; B_cafes; C_cafes];
gates_all = [A_gates; B_gates; C_gates];

% Perform two-sample t-test between cafeteria and gate noise
[H_scenario, p_scenario, ci_scenario, stats_scenario] = ttest2(cafeteria_all,
gates_all);

fprintf('  Scenario Comparison - Cafeterias vs Gates\n');
fprintf('  p-value = %.4f\n  t-statistic = %.4f\n\n', p_scenario,
stats_scenario.tstat);
```

Mean values for each scenario were calculated and compared to determine which scenario – cafeterias or gates – was noisier on average.

```
% Calculate and display average sound levels for both scenarios
mean_cafe = mean(cafeteria_all);
mean_gate = mean(gates_all);

fprintf('  Mean Noise Levels:\n');
fprintf('  Cafeterias = %.2f dB\n', mean_cafe);
fprintf('  Gates      = %.2f dB\n', mean_gate);

% Print final conclusion based on which scenario is noisier
if mean_gate > mean_cafe
    fprintf(' Conclusion: Campus gates are noisier than student cafeterias.\n');
else
    fprintf(' Conclusion: Student cafeterias are noisier than campus gates.\n');
end
```

Results

Location comparison

comparison of locations B and C:

```
Location Comparison – B vs C
 p-value = 0.0000
 t-statistic = 12.6726
```

comparison of locations A and C

```
Location Comparison – A vs C
 p-value = 0.0000
 t-statistic = 15.2131
```

```
Mean Noise Levels:
A = 67.52 dB
B = 66.11 dB
C = 61.84 dB

 Noisiest Location: A
```

Scenario comparison

```
Scenario Comparison – Cafeterias vs Gates
p–value = 0.0000
t–statistic = –10.4377

Mean Noise Levels:
Cafeterias = 65.16 dB
Gates      = 67.23 dB
Conclusion: Campus gates are noisier than student cafeterias.
```

Discussion

Location comparison

Comparison of locations B and C

Since the p- value (p<0.0001) is far below the standard significance level of 0.05, we reject the null hypothesis. This means that there is strong statistical evidence to conclude that the average noise levels at location B and C are significantly different. Mean values at location B and C were 66.1099 dB and 61.8381 dB, respectively.

For this scenario, the means are not comparable in the statistical sense. The extremely low p-value and high t-statistics suggest a clear and meaningful difference in the average sound levels recorded at two different locations.

This difference may be due to environmental factors such as crowd density, layout or proximity to busy walkways, which could be explored in future analysis.

Comparison of locations A and C

Given very small p-values, we reject the null hypothesis. This result provides strong statistical evidence that the average noise levels at locations A and C are significantly different.

The means between locations A and C are not statistically comparable. The very high t-statistic and near-zero p-value indicate a clear and meaningful difference in average noise levels. This suggests that either Location A is significantly nosier than C, or vice versa.

To better understand the direction of this difference, mean values at location A and C were calculated - 67.5191 dB and 61.8381 dB, respectively. Making location A considerably noisier than location C.

Which location is noisier?

Among the three locations, location A (Forum) has the highest average noise level, making it the noisiest location overall. It is followed by location B (Buttery), which is slightly quitter, and then location C (Goldhall), which is notably less noisy compared to the others.

The difference between A and B is relatively small (~1.4 dB), but the difference between A and C is substantial (~5.7 dB), which aligns with the results from the hypothesis tests that showed statistically significant differences in mean noise levels.

This suggests that location A might be experiencing more consistent crowding or environmental noise throughout the day. These findings are important for understanding how different cafeteria environments vary in terms of ambient noise. Such data could be useful for campus planning.

Scenario comparison

Since the p-value is well below the 0.05 significance threshold, we reject the null hypothesis. This means there is strong statistical evidence that the noise levels at campus gates and cafeterias are not equal.

The negative t-statistic indicates that the mean noise level at cafeterias is lower than that at campus gates. This is further confirmed by the mean values: campus gates had an average noise level of 67.23 dB, which is approximately 2.07 dB higher than the cafeterias' mean of 65.16 dB.

This result suggest that campus gates are noisier environments compared to student cafeterias. This could be attributes to several possible factors, such as high foot and vehicle traffic at gates, especially during peak entry and exit hours, environmental noise from roads, construction or surrounding urban activity and transient noise spikes from groups entering/leaving or from traffic horns and engines.

Conclusion

This study provides strong statistical evidence that noise levels vary significantly across campus locations and between different campus environments. Between the cafeterias, the Forum Café (location A) revealed to be the noisiest cafeteria, followed by the Buttery (location B), and with Goldsmith Hall (location C) being the quietest.

When comparing the two scenarios, it showed that the campus gates are significantly nosier than the cafeterias with a difference of over 2dB in average sound levels. These findings highlight the impact of environmental factors such as nearby traffic, pedestrian congestion and surrounding urban activity.

Overall, the results of this study highlight how location and surrounding affect ambient noise levels on campus. These findings can be valuable for campus planning, particularly when designing quiet study areas or managing traffic flow in high-noise areas.