STAT 8122 Descriptive Statistics

Graphical Presentation of Data Variable using MATLAB

%% Pie chart

```
Frequency =[42 50 32 55 9 12];

Cathegories=["lung","breast", "colon", "Prostate", "Melanoma", "Bladder"];

pie (Frequency)

legend ('lung','breast', 'colon', 'Prostate', 'Melanoma', 'Bladder');
```

%% Bar chart

```
x = ["Algebra" "English" "Physics" "Biology"];

y = [26 30 19 24];

bar(x,y)

xlabel('Class');

ylabel('Frequency');
```

%% Histogram

```
x = randn(1000,1);

h = histogram(x);
```

MATLAB codes to compute the Variance and the correlation coefficient

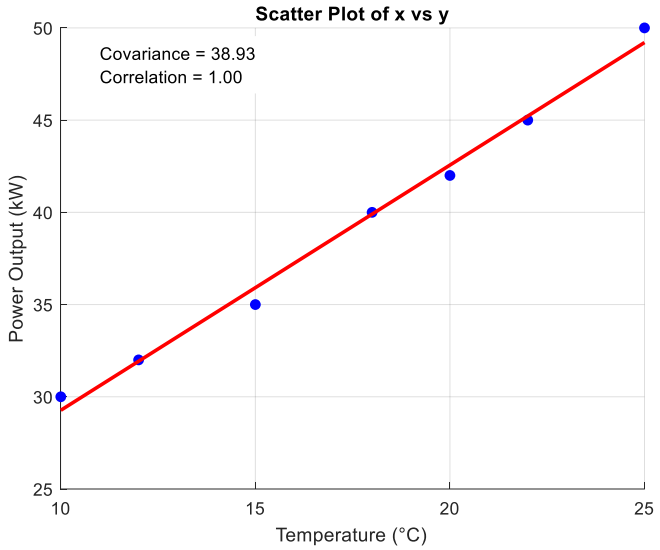
```
clear all
clc
% Define vectors
x = [11 22 32 41 51];
y = [8.2 32.8 82.0 144.4 236.2];

% Convert to column vectors
x = x(:);
y = y(:);

% Compute means
mean_x = mean(x);
mean_y = mean(y);

% Compute deviations
Xi = x - mean_x;
```

```
Yi = y - mean_y;

% Compute intermediate values
Xi2 = Xi.^2;
Yi2 = Yi.^2;
XiYi = Xi .* Yi;

% Display table
T = table(x, y, Xi, Yi, Xi2, Yi2, XiYi)

% Compute covariance (sample covariance)
cov_xy = sum(XiYi) / (length(x) - 1);

% Correlation coefficient (r)
corr_value = corr(x, y);

% Display result
fprintf('Covariance: %.4f\n', cov_xy);
fprintf('Correlation coefficient: %.4f\n', corr_value);

plot(x,y);
```

MATLAB codes to compute the Variance and the correlation coefficient with the regression line

```
clear all
clc
% Sample data
x = [10 12 15 18 20 22 25];
y = [30 32 35 40 42 45 50];

% Convert to column vectors (recommended)
x = x(:);
y = y(:);

% Covariance
cov_xy = cov(x, y);
covariance_value = cov_xy(1,2)

% Correlation coefficient (r)
corr_value = corr(x, y)

% Scatter plot
figure
scatter(x, y, 'filled','b')
xlabel('Temperature (°C)')
ylabel('Power Output (kW)')
title('Scatter Plot of x vs y')
```

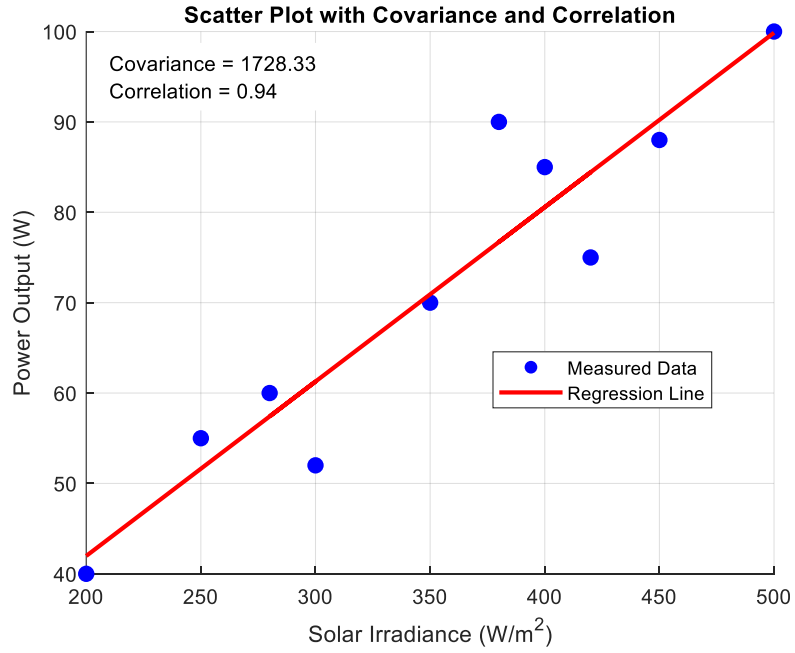
```
grid on
hold on

% Regression line
p = polyfit(x, y, 1);
y_fit = polyval(p, x);
plot(x, y_fit, 'r', 'LineWidth', 2)

% Display covariance and correlation on the plot
txt = { ...
    ['Covariance = ' num2str(covariance_value, '%.2f')], ...
    ['Correlation = ' num2str(corr_value, '%.2f')]};

text(min(x)+1, max(y)-2, txt, 'FontSize', 10, 'BackgroundColor', 'w')

% hold off
```



MATLAB codes to compute the Variance and the correlation coefficient with the regression line

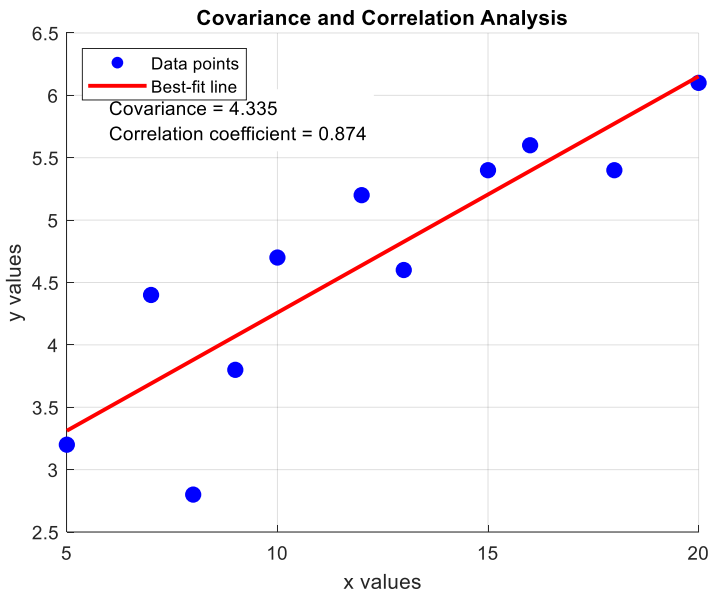
```
% Clear workspace
clc;
clear;

% More scattered sample data
% Example: Solar irradiance (x) vs power output (y)
x = [200 250 300 280 350 400 420 380 450 500];   % Irradiance (W/m^2)
y = [40 55 52 60 70 85 75 90 88 100];            % Power output (W)

% Covariance
C = cov(x, y);
cov_xy = C(1,2)
```

```
% Correlation coefficient
R = corrcoef(x, y);
corr_xy = R(1,2)

% Scatter plot
figure;
scatter(x, y, 60, 'b', 'filled');
hold on;

% Best-fit regression line
p = polyfit(x, y, 1);
y_fit = polyval(p, x);
plot(x, y_fit, 'r', 'LineWidth', 2);

% Graph formatting
xlabel('Solar Irradiance (W/m^2)');
ylabel('Power Output (W)');
title('Scatter Plot with Covariance and Correlation');
grid on;

% % Display covariance and correlation on plot
text(min(x)+10, max(y)-5, ...
    {['Covariance = ', num2str(cov_xy, '%.2f')], ...
     ['Correlation = ', num2str(corr_xy, '%.2f')]}, ...
     'FontSize', 10, 'BackgroundColor', 'w');

legend('Measured Data', 'Regression Line', 'Location', 'best');
hold off;
```



MATLAB codes to compute the Variance and the correlation coefficient with the regression line

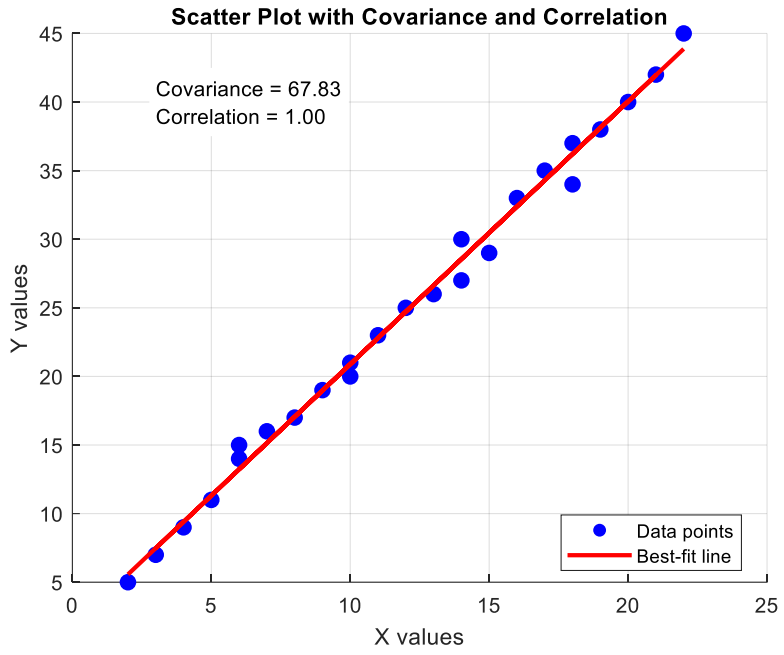
```
% Clear workspace and command window
clc;
clear;

% Given data
x = [5 7 8 9 10 12 13 15 16 18 20];
y = [3.2 4.4 2.8 3.8 4.7 5.2 4.6 5.4 5.6 5.4 6.1];

% Covariance matrix
C = cov(x, y);
cov_xy = C(1,2)

% Correlation coefficient matrix
R = corrcoef(x, y);
corr_xy = R(1,2)

% Scatter plot
figure;
scatter(x, y, 70, 'b', 'filled');
hold on;

% Linear regression (best-fit line)
p = polyfit(x, y, 1);
y_fit = polyval(p, x);
plot(x, y_fit, 'r', 'LineWidth', 2);

% Graph labels and formatting
xlabel('x values');
ylabel('y values');
title('Covariance and Correlation Analysis');
grid on;

% Display numerical results on the graph
text(6, 5.8, ...
    {['Covariance = ', num2str(cov_xy, '%.3f')], ...
     ['Correlation coefficient = ', num2str(corr_xy, '%.3f')]}, ...
     'FontSize', 10, 'BackgroundColor', 'w');

legend('Data points', 'Best-fit line', 'Location', 'northwest');
hold off;
```



MATLAB codes to compute the Variance and the correlation coefficient with the regression line

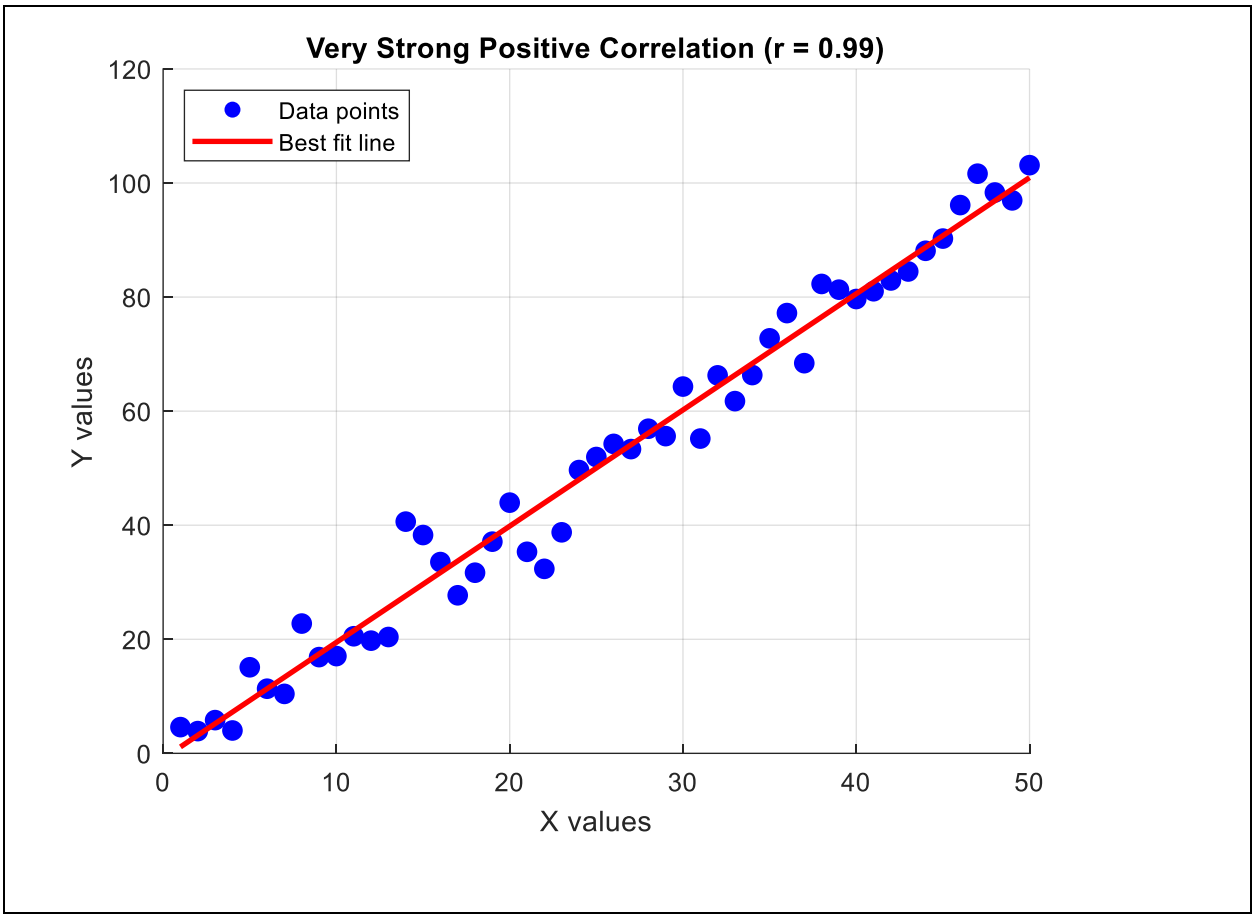
```
% Clear workspace and command window
clear all;
clc;

% Generated 25 data points (moderately scattered)
x = [2 4 6 8 10 12 14 16 18 20 3 5 7 9 11 13 15 17 19 21 6 10 14 18 22];

y = [5 9 14 17 21 25 30 33 37 40 7 11 16 19 23 26 29 35 38 42 15 20 27 34 45];


% Covariance

C = cov(x, y);          % Covariance matrix
cov_xy = C(1,2);        % Covariance value


% Correlation Coefficient

R = corrcoef(x, y);     % Correlation matrix
corr_xy = R(1,2);       % Correlation coefficient

% Display results

fprintf('Covariance between x and y = %.4f\n', cov_xy);
fprintf('Correlation coefficient between x and y = %.4f\n', corr_xy);
```

```
% Scatter plot and regression line
figure;
scatter(x, y, 60, 'b', 'filled');
hold on;

% Best-fit line
p = polyfit(x, y, 1);
y_fit = polyval(p, x);
plot(x, y_fit, 'r', 'LineWidth', 2);

grid on;
xlabel('X values');
ylabel('Y values');
title('Scatter Plot with Covariance and Correlation');
legend('Data points', 'Best-fit line', 'Location', 'best');

% Display values on the plot
text(min(x)+1, max(y)-5, ...
    sprintf('Covariance = %.2f\nCorrelation = %.2f', cov_xy, corr_xy), ...
    'FontSize', 10, 'BackgroundColor', 'w');
```



MATLAB codes to compute the Variance and the correlation coefficient with the regression line

```
clear all;
 clc;


% Generate strongly correlated data

n = 50;                       % Number of data points
x = linspace(1, 50, n)       % Independent variable
noise = randn(1, n) * 5;     % Small random noise
y = 2*x + noise             % Dependent variable (strong correlation)

% Covariance
C = cov(x, y);               % Covariance matrix
cov_xy = C(1,2)             % Covariance between x and y

% Correlation coefficient
R = corrcoef(x, y);          % Correlation coefficient matrix
r = R(1,2)                  % Correlation coefficient

% Display results
fprintf('Covariance between x and y = %.2f\n', cov_xy);
fprintf('Correlation coefficient r = %.2f\n', r);

% Scatter plot with best-fit line
figure;
scatter(x, y, 60, 'b', 'filled');
hold on;

% Linear regression line
p = polyfit(x, y, 1);
y_fit = polyval(p, x);
plot(x, y_fit, 'r', 'LineWidth', 2);

% Labels and title
grid on;
xlabel('X values');
ylabel('Y values');
title(['Very Strong Positive Correlation (r = ', num2str(r,'%.2f'), ')']);
legend('Data points', 'Best fit line', 'Location', 'northwest');
```


Very Strong Positive Correlation (r = 0.99)
Data points
Best fit line
Y values
X values
0
20
40
60
80
100
120
0
10
20
30
40
50