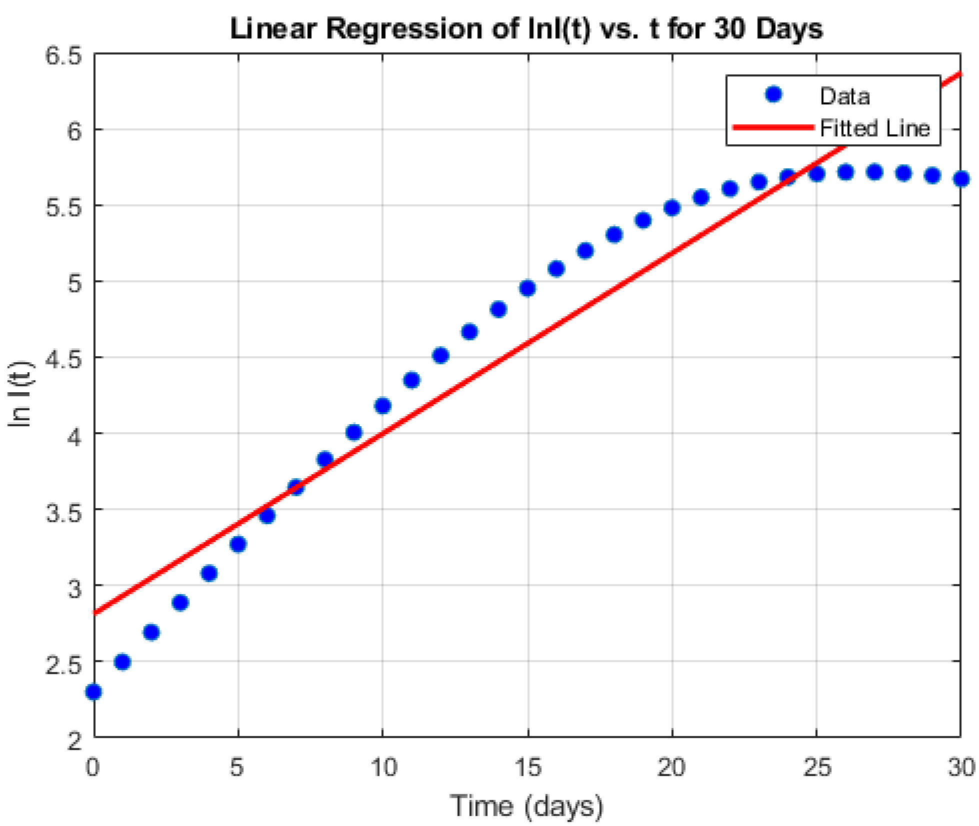
Part III: Linear Least Squares

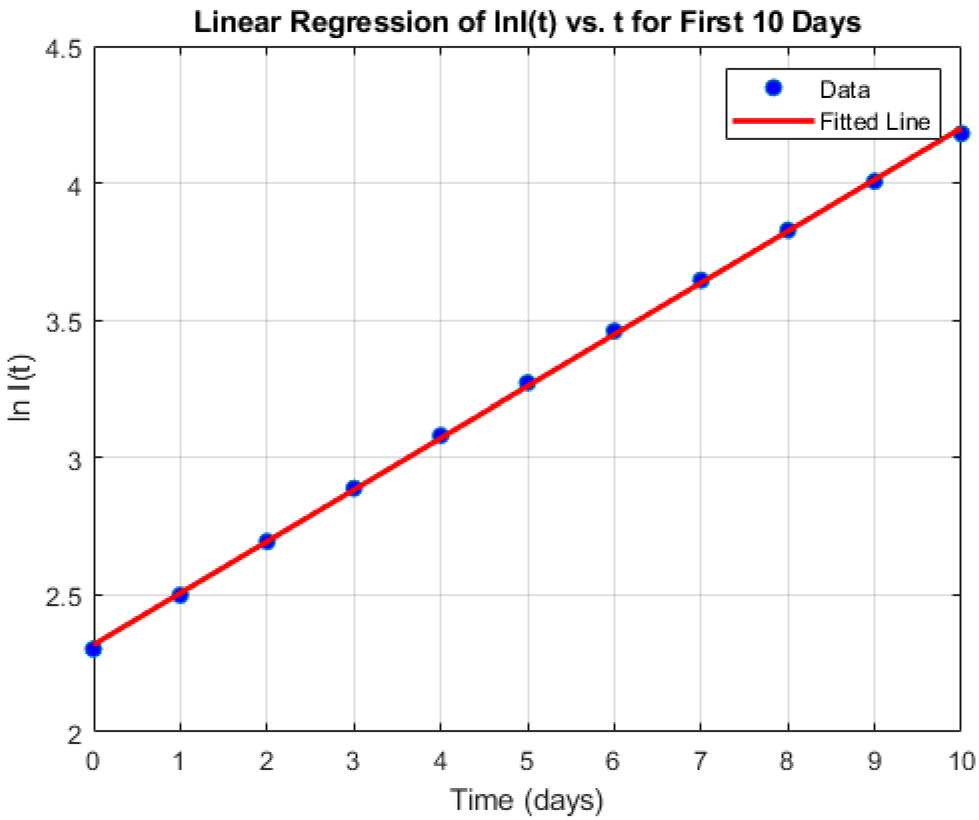
```
S(1) = 990;
I(1) = 10;
R(1) = 0;
h = 1;
Time = 30;
N_total = 1000;
Beta = 0.3;
gamma = 0.1;

N = @(S, I, R) S + I + R;
dSdt = @(N_pop, S_curr, I_curr) (-Beta / N_pop) * S_curr * I_curr;
dIdt = @(N_pop, S_curr, I_curr) (Beta / N_pop) * S_curr * I_curr - gamma *
I_curr;
dRdt = @(I_curr) gamma * I_curr;

Population(1) = S(1) + I(1) + R(1);
Population2(1) = 0;
Population3(1) = 0;
Population4(1) = 0;

for i = 1:Time
    Population(i+1) = N(S(i), I(i), R(i));
    K1Susceptible = dSdt(Population(i), S(i), I(i));
    K1Infected = dIdt(Population(i), S(i), I(i));
    K1Recovered = dRdt(I(i));
    K2S = S(i) + 0.5 * K1Susceptible * h;
    K2I = I(i) + 0.5 * K1Infected * h;
    K2R = R(i) + 0.5 * K1Recovered * h;
    Population2(i+1) = N(K2S, K2I, K2R);
    K2Susceptible = dSdt(Population2(i+1), K2S, K2I);
    K2Infected = dIdt(Population2(i+1), K2S, K2I);
    K2Recovered = dRdt(K2I);
    K3S = S(i) + 0.5 * K2Susceptible * h;
    K3I = I(i) + 0.5 * K2Infected * h;
    K3R = R(i) + 0.5 * K2Recovered * h;
    Population3(i+1) = N(K3S, K3I, K3R);
    K3Susceptible = dSdt(Population3(i+1), K3S, K3I);
    K3Infected = dIdt(Population3(i+1), K3S, K3I);
    K3Recovered = dRdt(K3I);
    K4S = S(i) + K3Susceptible * h;
    K4I = I(i) + K3Infected * h;
    K4R = R(i) + K3Recovered * h;
    Population4(i+1) = N(K4S, K4I, K4R);
    K4Susceptible = dSdt(Population4(i+1), K4S, K4I);
    K4Infected = dIdt(Population4(i+1), K4S, K4I);
    K4Recovered = dRdt(K4I);
    S(i+1) = S(i) + (h / 6) * (K1Susceptible + 2 * K2Susceptible + 2 *
K3Susceptible + K4Susceptible);
    I(i+1) = I(i) + (h / 6) * (K1Infected + 2 * K2Infected + 2 * K3Infected
+ K4Infected);
```

```
    R(i+1) = R(i) + (h / 6) * (K1Recovered + 2 * K2Recovered + 2 *
K3Recovered + K4Recovered);
end

t = 0:h:Time;
I_t = I;
ln_I_t = log(I_t);
n = length(t);

sum_t = sum(t);
sum_lnI = sum(ln_I_t);
sum_t_lnI = sum(t .* ln_I_t);
sum_t_squared = sum(t.^2);

numerator_k = n * sum_t_lnI - sum_t * sum_lnI;
denominator_k = n * sum_t_squared - sum_t^2;
k_est = numerator_k / denominator_k;

ln_I0_est = (sum_lnI - k_est * sum_t) / n;
I0_est = exp(ln_I0_est);

beta_est = (k_est + gamma) * N_total / S(1);

fprintf('Using 30 days of data:\n');
fprintf('Estimated I(0) = %.5f, True I(0) = %.5f\n', I0_est, I(1));
fprintf('Estimated beta = %.5f, True beta = %.5f\n', beta_est, Beta);

n_10 = 11;  % Number of data points from t = 0 to t = 10
t_10 = t(1:n_10);
ln_I_t_10 = ln_I_t(1:n_10);

sum_t_10 = sum(t_10);
sum_lnI_10 = sum(ln_I_t_10);
sum_t_lnI_10 = sum(t_10 .* ln_I_t_10);
sum_t_squared_10 = sum(t_10.^2);
numerator_k_10 = n_10 * sum_t_lnI_10 - sum_t_10 * sum_lnI_10;
denominator_k_10 = n_10 * sum_t_squared_10 - sum_t_10^2;
k_est_10 = numerator_k_10 / denominator_k_10;

ln_I0_est_10 = (sum_lnI_10 - k_est_10 * sum_t_10) / n_10;
I0_est_10 = exp(ln_I0_est_10);

beta_est_10 = (k_est_10 + gamma) * N_total / S(1);

fprintf('\nUsing first 10 days of data:\n');
fprintf('Estimated I(0) = %.5f, True I(0) = %.5f\n', I0_est_10, I(1));
fprintf('Estimated beta = %.5f, True beta = %.5f\n', beta_est_10, Beta);

figure;
plot(t, ln_I_t, 'o', 'MarkerFaceColor', 'b');
hold on;
ln_I_fit = ln_I0_est + k_est * t;
plot(t, ln_I_fit, 'r-', 'LineWidth', 2);
title('Linear Regression of lnI(t) vs. t for 30 Days');
```

```
xlabel('Time (days)');
ylabel('ln I(t)');
legend('Data', 'Fitted Line');
grid on;
hold off;

figure;
plot(t_10, ln_I_t_10, 'o', 'MarkerFaceColor', 'b');
hold on;
ln_I_fit_10 = ln_I0_est_10 + k_est_10 * t_10;
plot(t_10, ln_I_fit_10, 'r-', 'LineWidth', 2);
title('Linear Regression of lnI(t) vs. t for First 10 Days');
xlabel('Time (days)');
ylabel('ln I(t)');
legend('Data', 'Fitted Line');
grid on;
hold off;
```

```
Using 30 days of data:
Estimated I(0) = 16.67335, True I(0) = 10.00000
Estimated beta = 0.22065, True beta = 0.30000

Using first 10 days of data:
Estimated I(0) = 10.15965, True I(0) = 10.00000
Estimated beta = 0.29121, True beta = 0.30000
```





Published with MATLAB® R2023b

Yes, using only the first 10 days of data provides estimates of I(0) and β closer to the true value. This is because the simplified model ($S(t) \approx S_0$) holds better early on, making ln I(t) vs. t a more linear fit initially at least. From Part 1, we saw that the infected curve for seasonal influenza initially looks more linear, but over longer times, it appears more quadratic. Thus, shorter datasets give more accurate estimations for I(0) and β.