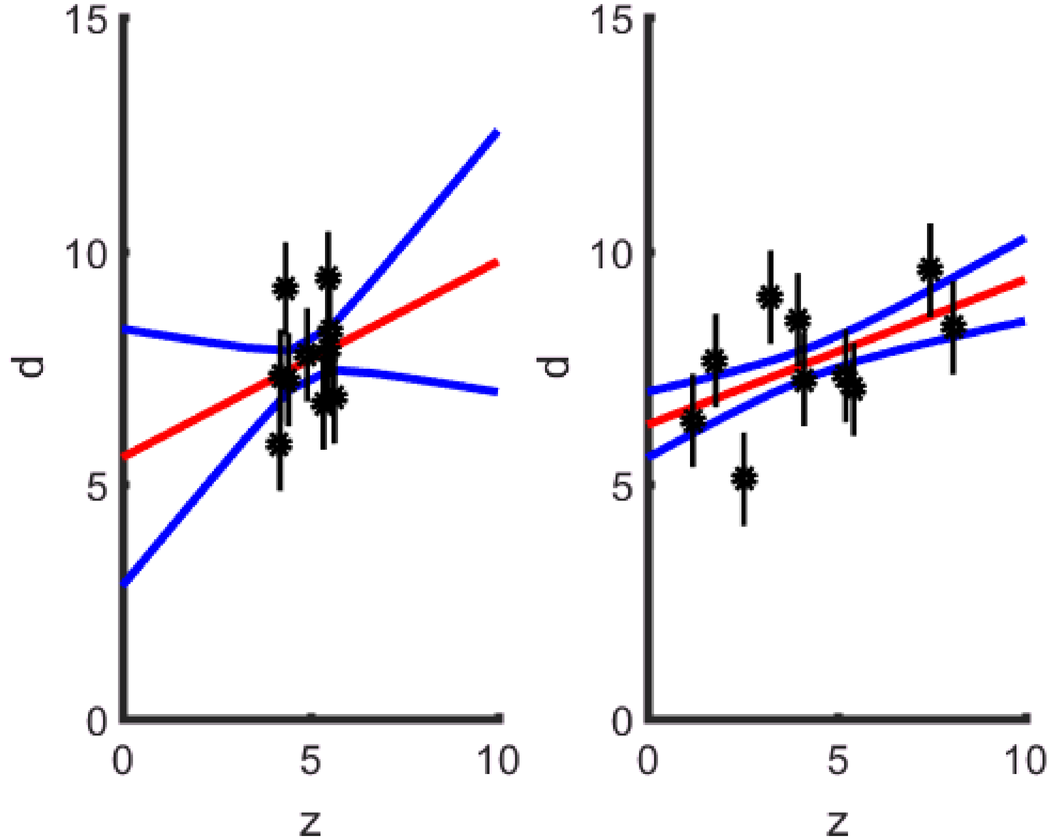
```
% gda04_03
%
% least squares fit to two cases of synthetic data
% distinguished by the spacing of the z's
% supports Figure 4.3

clear all;

% CASE 1: date clumped in middle of interval

% auxially variable z
N=10;
zmin=0;
zmax=10;
z = sort(random('Uniform',zmin+4*(zmax-zmin)/10,zmin+6*(zmax-zmin)/10,N,1));

% create synthetic observed data
% d = a + b*z + noise
a=5.0;
b=0.5;
sd=1;
dobs = a+b*z+random('Normal',0,sd,N,1);

% least squares fit
M=2;
G=[ones(N,1), z];
mest = (G'*G)\(G'*dobs);
Cm = (sd^2)*inv(G'*G); % covariance of model parameters
sm = [sqrt(Cm(1,1)), sqrt(Cm(2,2)) ]; % sqrt(variance)

% predicted data at a new set of z's
No=10;
zeval = zmin + (zmax-zmin)*[0:No-1]'/(No-1);
Go = [ones(No,1), zeval];
deval = Go*mest; % predicted data
Cdeval = Go*Cm*Go'; % its covariance
sdeval = sqrt(diag(Cdeval)); % its sqrt(variance)

% plot
figure(1);
clf;

% plot scale
pdmin=0;
pdmax=15;

% plot observed data, predicted data and +/- one-sigma error bounds
subplot(1,2,1);
set(gca,'LineWidth',3);
set(gca,'FontSize',14);
hold on;
axis( [zmin, zmax, pdmin, pdmax ]' );
plot( zeval, deval+sdeval, 'b-', 'LineWidth', 3);
plot( zeval, deval-sdeval, 'b-', 'LineWidth', 3);
plot( zeval, deval, 'r-', 'LineWidth', 3);
plot( z, dobs, 'ko', 'LineWidth', 3);
for i = [1:N]
    plot( [z(i), z(i)], [dobs(i)-sd, dobs(i)+sd], 'k-', 'LineWidth', 2);
end
```

```
xlabel('z');
ylabel('d');


% CASE 2: data spread out over whole interval

% auxially variable z
N=10;
zmin=0;
zmax=10;
z = sort(random('Uniform',zmin,zmax,N,1));

% create synthetic observed data
% d = a + b*z + noise
a=5.0;
b=0.5;
sd=1;
dobs = a+b*z+random('Normal',0,sd,N,1);

% least squares fit
M=2;
G=[ones(N,1), z]; % data kernel
mest = (G'*G)\(G'*dobs); % estimated solution
Cm = (sd^2)*inv(G'*G); % model coverance
sm = [sqrt(Cm(1,1)), sqrt(Cm(2,2)) ]; % sqrt(variance)

% predicted data at a new set of z's
No=10;
zeval = zmin + (zmax-zmin)*[0:No-1]'/(No-1);
Go = [ones(No,1), zeval]; % data kernel
deval = Go*mest; % predicted data
Cdeval = Go*Cm*Go'; % its covariance
sdeval = sqrt(diag(Cdeval)); % sqrt(variance)

% plot observed data, predicted data and +/- one-sigma error bounds
subplot(1,2,2);
set(gca,'LineWidth',3);
set(gca,'FontSize',14);
hold on;
axis( [zmin, zmax, pdmin, pdmax ]' );
plot( zeval, deval+sdeval, 'b-', 'LineWidth', 3);
plot( zeval, deval-sdeval, 'b-', 'LineWidth', 3);
plot( zeval, deval, 'r-', 'LineWidth', 3);
plot( z, dobs, 'ko', 'LineWidth', 3);
for i = [1:N]
    plot( [z(i), z(i)], [dobs(i)-sd, dobs(i)+sd], 'k-', 'LineWidth', 2);
end
xlabel('z');
ylabel('d');
```




```
% Figure 4.3 (A) The method of least squares is used to fit a straight line (red) to uncorrela
% data (black circles) with uniform variance (vertical bars, 1σ confidence limits). Since the
% are not well-separated in z, the variance of the slope and intercept is large, and consequen
% the variance of the predicted data is large as well (blue curves, 1σ confidence limits). (B)
% as (A) but with the data well-separated in z. Although the variance of the data is the same
% in (A), the variance of the intercept and slope, and consequently the predicted data, is muc
% smaller. MatLab script gda04_03.
```