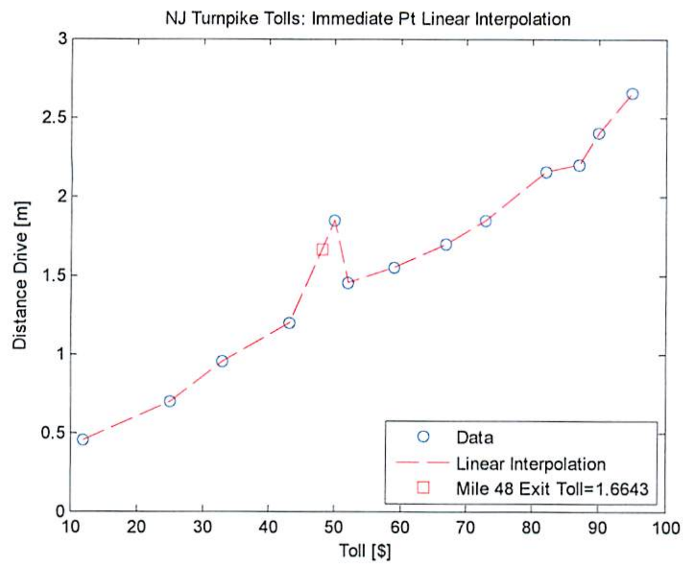
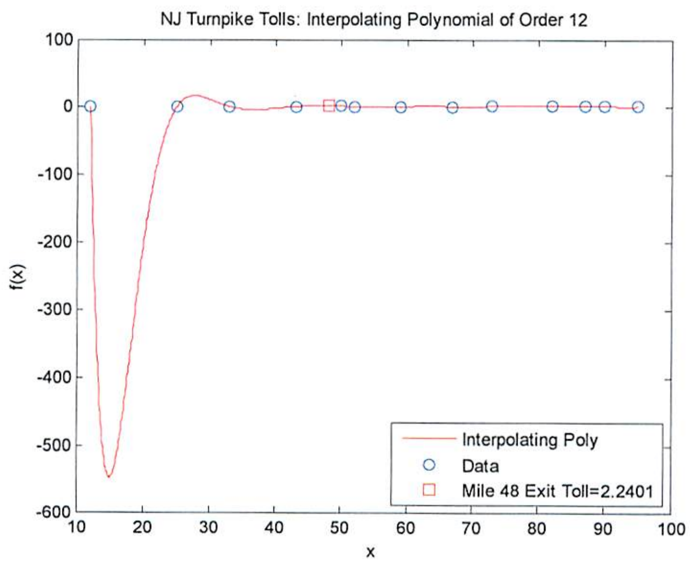
I. Introductory Example: Use NJ Turnpike data to determine toll for new exit at mile 48. **[Go1]**. See Example 5.22 in Section 5.8, p 420, Bradie, "A Friendly Introduction to Numerical Anlysis".

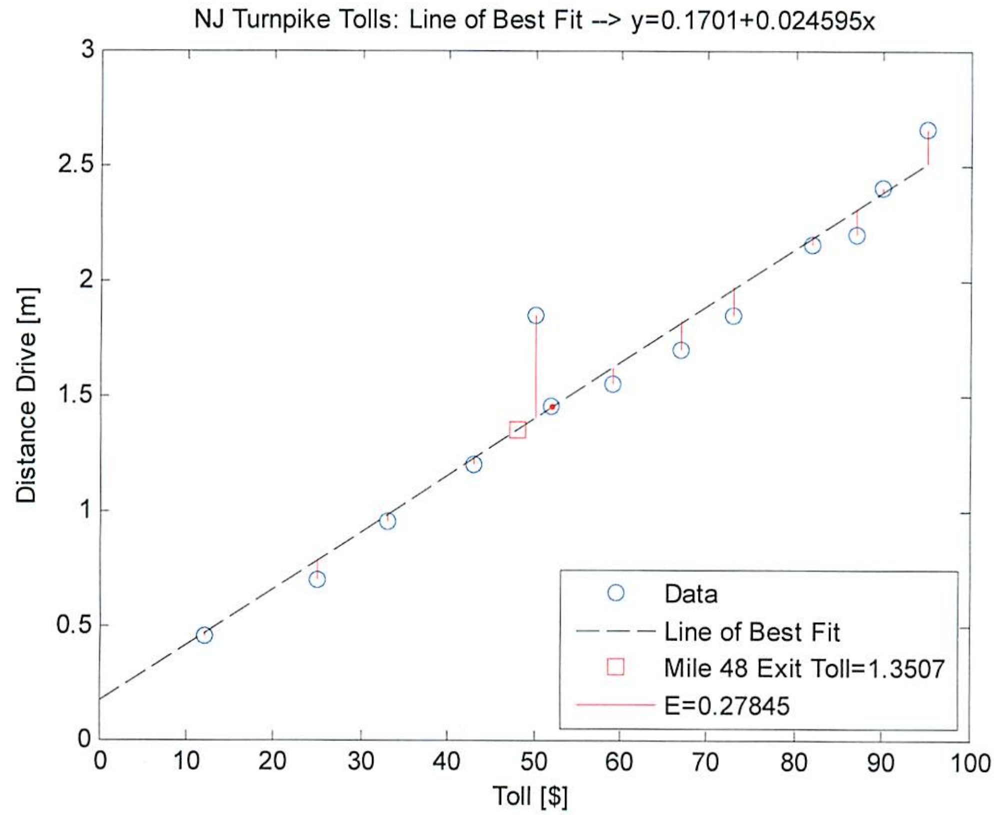
A. Immediate Point Linear Interpolation. : $T_{48} = \frac{T_{50}-T_{43}}{50-43}(48-43)$



B. Interpolation Polynomials:



C. Line of Best Fit: Lenear Regression



II Deriving Linear Regression Formulas

A. Error Between a Point & Line

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$$

y_i ↑ data; $\hat{y}_i$ ↑ linear estimat

B Define Total Error E

$$E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - (a + bx_i)\right]^2$$

C. $\frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = 0$

$$\Rightarrow \begin{cases} \frac{\partial E}{\partial a} = -2\left[\sum_{i=1}^{n} [y_i - (a+bx_i)]\right] = 0 \\ \frac{\partial E}{\partial b} = -2\sum_{i=1}^{n} [y_i - (a+bx_i)]\, x_i = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n} y_i - an - b\sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 \end{cases}$$

$$\begin{cases} an + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

Let:
$X = \sum x_i$
$X2 = \sum x_i^2$
$Y = \sum y_i$
$XY = \sum x_i y_i$

Therefore: (Finding b)

$$\begin{cases} an + bX = Y & \text{multiply by } X \\ aX + bX^2 = XY & \text{multiply by } n \end{cases}$$

$$\begin{cases} anX + b(X)(X) = (X)(Y) \\ anX + bnX^2 = nXY \end{cases}$$

careful
$(X)(X) \neq X^2$
$(X)(Y) \neq XY$

$$\Rightarrow bnX^2 - b(X)(X) = nXY - (X)(Y)$$

$$\Rightarrow b = \frac{nXY - (X)(Y)}{nX^2 - (X)(X)} \Rightarrow$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$\Rightarrow$ finding a $\Rightarrow$

$$an + bX = Y \Rightarrow an = Y - bX$$

$$\Rightarrow a = \frac{Y}{n} - b\frac{X}{n} \Rightarrow a = \frac{1}{n}\sum_{i=1}^{n} Y_i - b\frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\Rightarrow \boxed{a = \bar{y} - b\bar{x}}$$

note $\bar{y} = \text{mean}(y) = \frac{1}{n}\sum_{i=1}^{n} y_i$

D) Let Matlab Do The Work

$\sum_{i=1}^{n} x_i$ = sum(x)

$\bar{x}$ = mean(x)

$\sum y_i$ = sum(y)

$\bar{y}$ = mean(y)

$\sum x_i^2$ = sum(x.^2)

$\sum x_i y_i$ = sum(x.*y)

↓

See code on following page.

III. Linear Regression Code

```
function [a,b,f]=linreg(x,y);
%Find's Line of Best Fit for XY Data;

%plottin raw data;
plot(x,y,'o');

%getting parameters for linear regression
n=length(x);
X=sum(x);
X2=sum(x.^2);
Y=sum(y);
XY=sum(x.*y);
b=(n.*XY-X.*Y)./(n.*X2-X.^2);
a=mean(y)-b*mean(x);

%plot line of best fit,
hold on;
y0=a+b*0; ymax=a+b*max(x);
plot([0,max(x)],[y0,ymax],'k--');

%Returning function for further use;
f=[num2str(a),'+',num2str(b),'.*x'];
f=inline(f);
```

IV Exponential Behavior:

A) Suppose: $y = ab^x$

B) Then $\ln(y) = \ln(ab^x)$

$$\Rightarrow \ln(y) = \ln(a) + \ln(b^x)$$

$$\Rightarrow \boxed{\ln(y) = \ln(a) + x\ln(b)}$$

C) now if we plot $\ln(y)$ vs x (i.e semi-log plot) we will see linear behavior

D) linear regression will solve for $\ln(a)$ & $\ln(b)$

E) to put these back in original formula must recover a, & b i.e.

$$a = e^{\ln(a)}, \quad b = e^{\ln(b)}$$

V Power Behavior - Similarly if $y = ax^b$

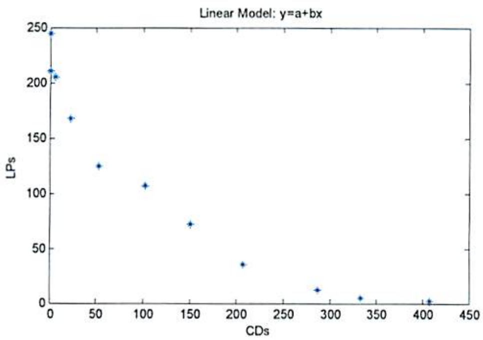
A) $\ln(y) = \ln(ax^b) = \ln(a) + \ln(x^b)$

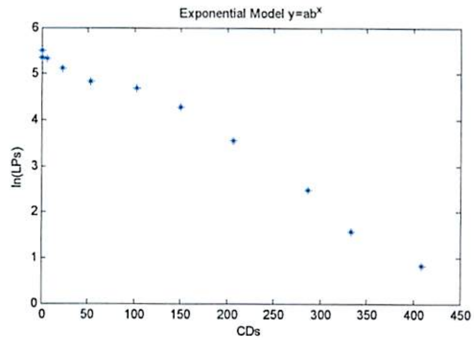
$$\Rightarrow \boxed{\ln(y) = \ln(a) + b\ln(x)}$$

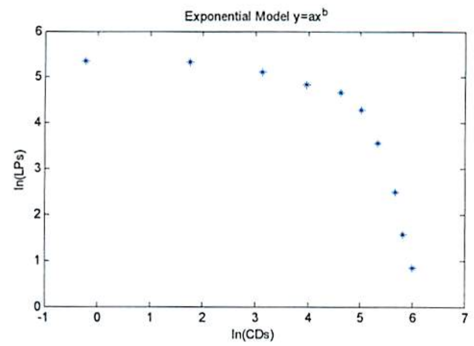
B) Plot $\ln(y)$ vs $\ln(x)$ (log-log plot) & linear behavior emerges

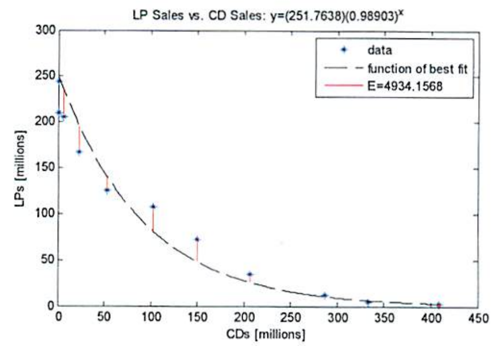
C) Regression solves for $\ln(a)$, b.

VI. CD Sales vs LP Sales (Example 5.26, p. 426)









VII. Code for Example

```
function go2
% Example 5.25 LP vs CD Sales, Section 5.8, p. 425,
% Bradie - " A Friendly Introduction to Numerical Analysis"
%
% Explores adapting linear regression for power and exponential functions
% using semi-log an log-log plots;
close all;

%data
x=[0 .8 5.8 23 53 102 150 207 287 333 408]; %CD Sales [millions]
y=[244 210 205 167 125 107 72 35 12 4.8 2.3]; %Album Sales [millions]

subplot(2,2,2); %the exponential model --> y=ab^x
    plot(x,log(y),'*');
    title('Exponential Model y=ab^x');xlabel('CDs');ylabel('ln(LPs)');
    disp('Hit Return for Power Model');

subplot(2,2,3); %the power model worked best --> y=ax^b
    plot(log(x),log(y),'*');
    title('Exponential Model y=ax^b');xlabel('ln(CDs)'); ylabel('ln(LPs)');
    disp('Hit Return to Continue');
    pause;

subplot(2,2,4); %fitting the exponential model
    %getting parameters for linear regression on exponential model
    xx=x; yy=log(y); %finding parameters for a linear log-log plot;
    n=length(xx);
    X=sum(xx);
    X2=sum(xx.^2);
    Y=sum(yy);
    XY=sum(xx.*yy);
    b=(n.*XY-X.*Y)./(n.*X2-X.^2),%ln(y)=ln(a)+bx;
    a=mean(yy)-b*mean(xx),
    a=exp(a); b=exp(b) %converting a to use in y=ax^b;

    %plot curve of best fit,
    xx=0:1:408; yy=a.*b.^xx;
    plot(x,y,'*');
    hold on;
    plot(xx,yy,'k--');
    title(['LP Sales vs. CD Sales: y=(',num2str(a),')(',num2str(b),')^x']);
    xlabel('CDs [millions]'); ylabel('LPs [millions]');

    E=0;
    f=[num2str(a),'.*',num2str(b),'.^x'];
    f=inline(f);
    for i=1:length(x);
        E=E+(y(i)-f(x(i)))^2;
        plot([x(i),x(i)],[y(i),f(x(i))],'r');
    end
    legend('data','function of best
fit',['E=',num2str(E)],'Location','NorthEast');
```