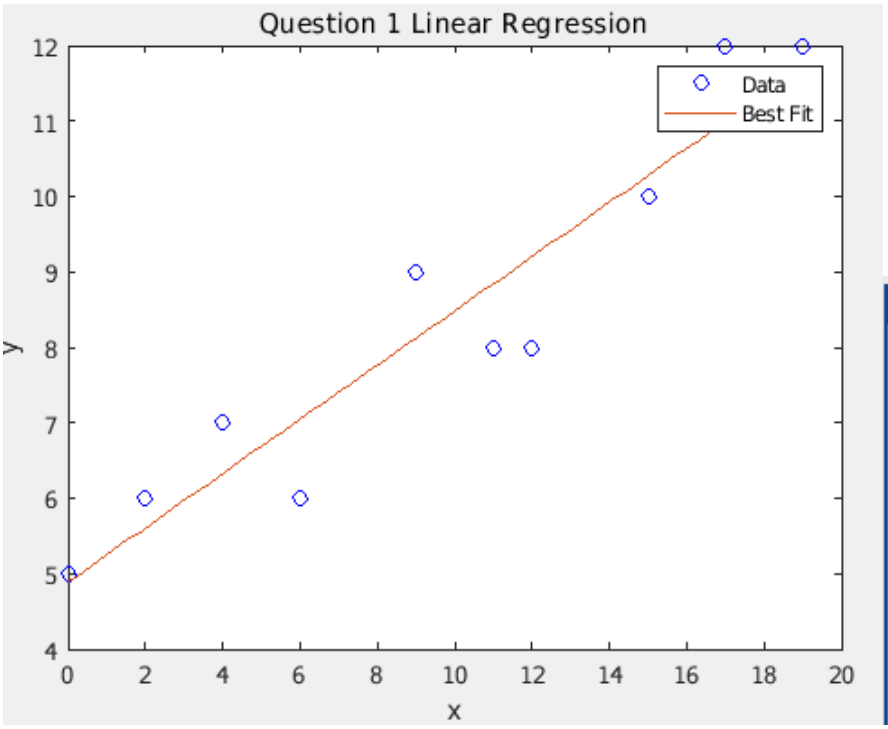
Practical Answers 7 – Linear Regression

1. Use least squares regression to fit a straight line to the following data (implement the formula from the lectures in Matlab).

x	0	2	4	6	9	11	12	15	17	19
y	5	6	7	6	9	8	8	10	12	12




```
Command Window
a0 =
        4.8881
>> a1
a1 =
        0.35915
```

2. Calculate the correlation coefficient for the best fit line from question 1.

```
Command Window
>> Q1_2_Sol
The regression coefficient, r, is: 0.94493
```

3. Repeat question 1 but use the Matlab built-in functions polyfit to find the equation and polyval to plot.

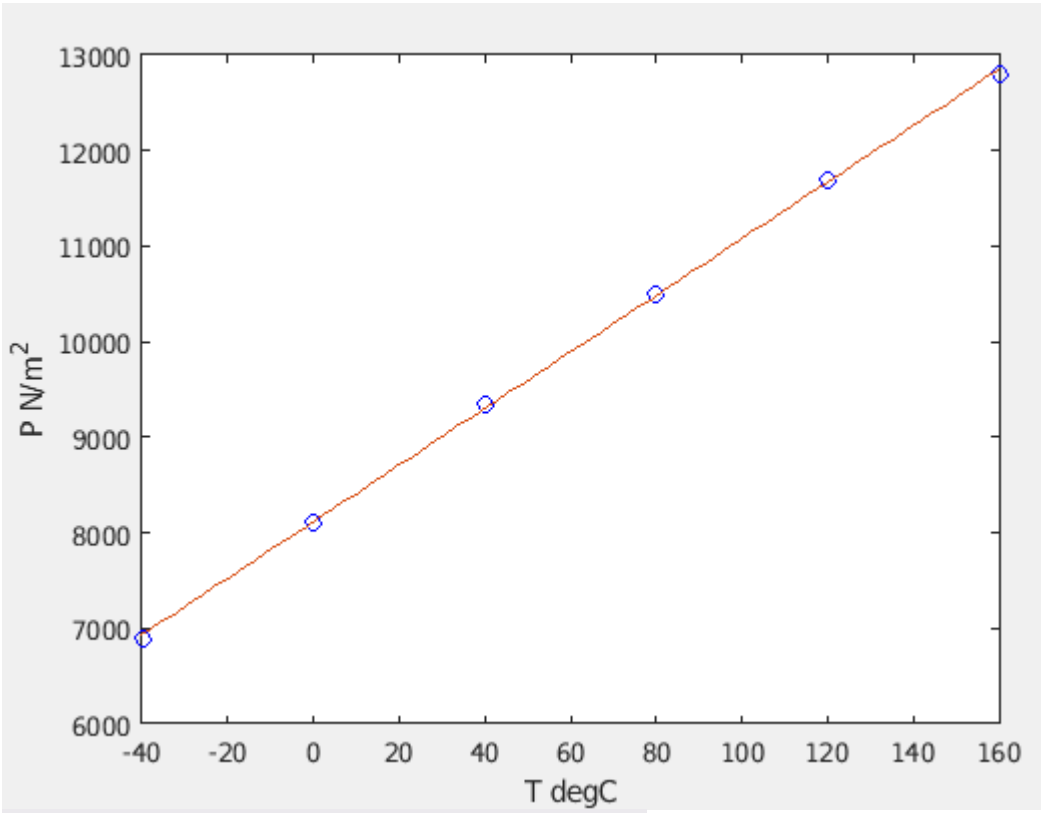
Same as above.

4. The following data were gathered to determine the relationship between pressure and temperature of a fixed volume of 1 kg of nitrogen (molar mass of 14.006747 g/mol). The volume is 10m^3. Use the ideal gas law equation to determine the

universal gas constant, R.

$$pV = nRT$$

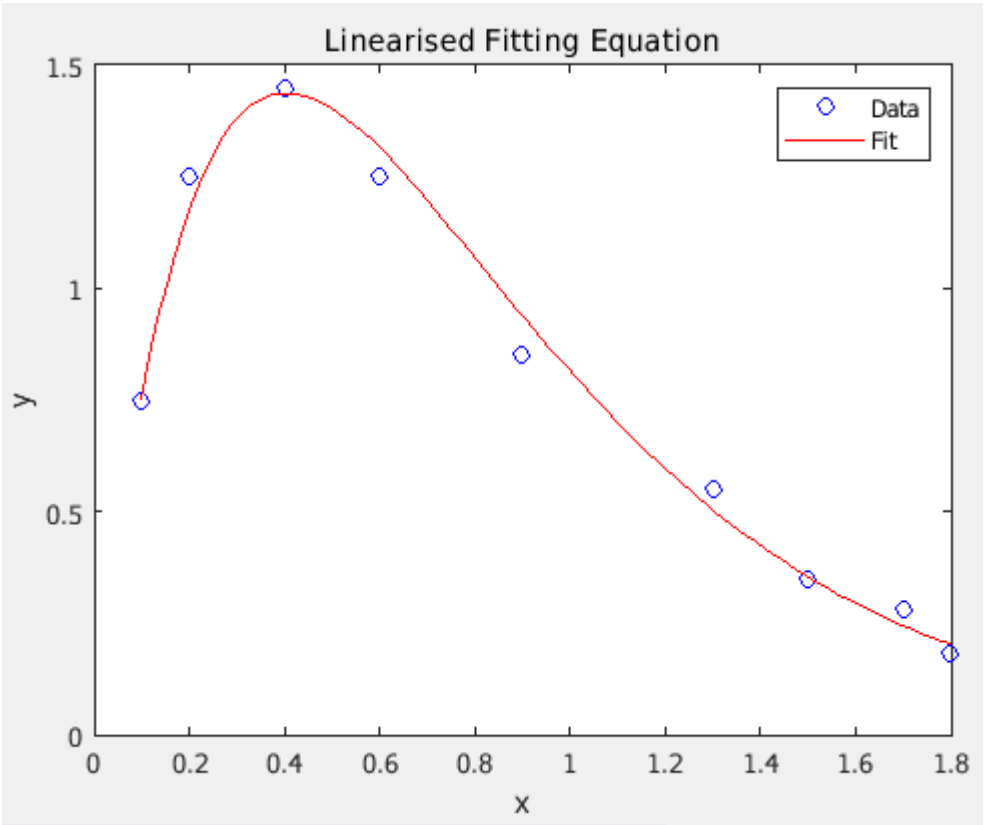
T, °C	−40	0	40	80	120	160
p, N/m²	6900	8100	9350	10,500	11,700	12,800



5. Linearise the following model and use it to estimate the parameters α_4 and β_4 for the following data. Plot the data with the line of best fit and calculate the correlation coefficient.

$$y = \alpha_4 x e^{\beta_4 x}$$

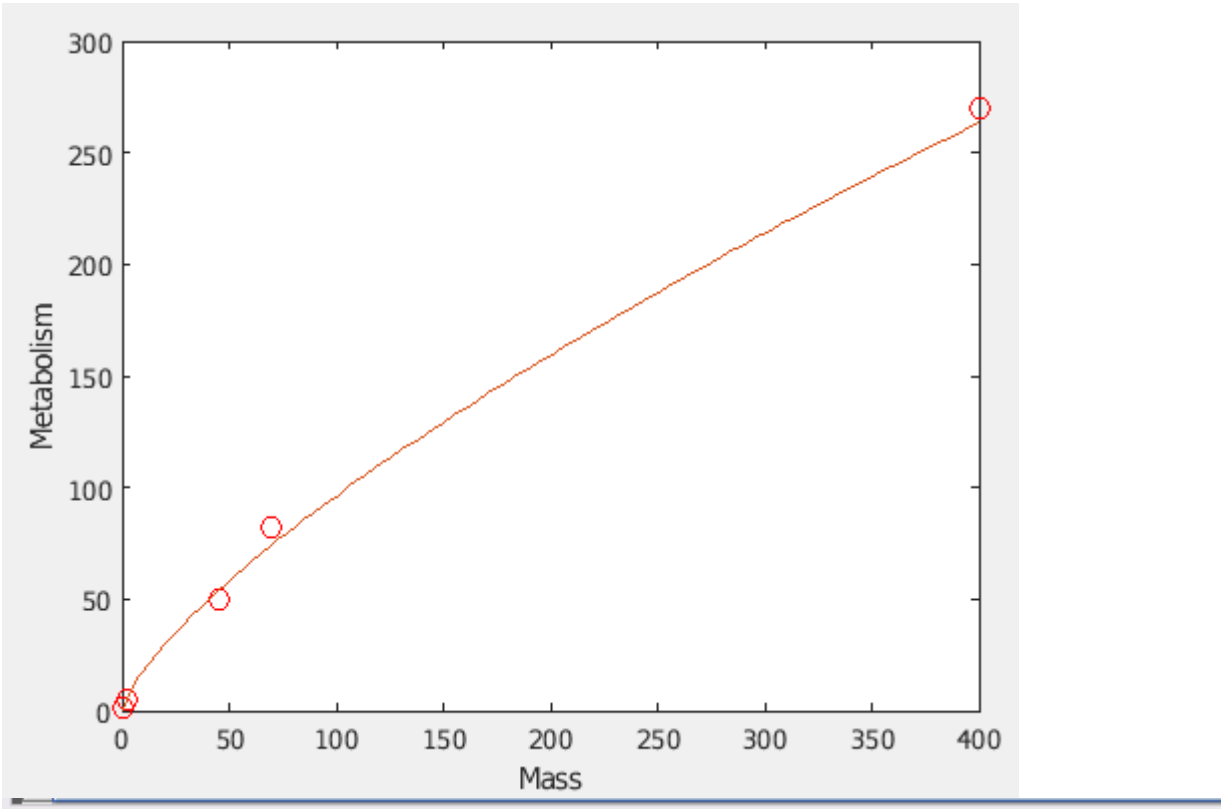
x	0.1	0.2	0.4	0.6	0.9	1.3	1.5	1.7	1.8
y	0.75	1.25	1.45	1.25	0.85	0.55	0.35	0.28	0.18




```
Command Window
>> Q5_Sol
The fitting equation is:
y = 9.6618 x*exp(-2.4733 x)
```

6. Try fitting a power equation to the following data on metabolisms of animals then using it to predict the metabolism rate of a 200kg tiger.

Animal	Mass (kg)	Metabolism (watts)
Cow	400	270
Human	70	82
Sheep	45	50
Hen	2	4.8
Rat	0.3	1.45
Dove	0.16	0.97

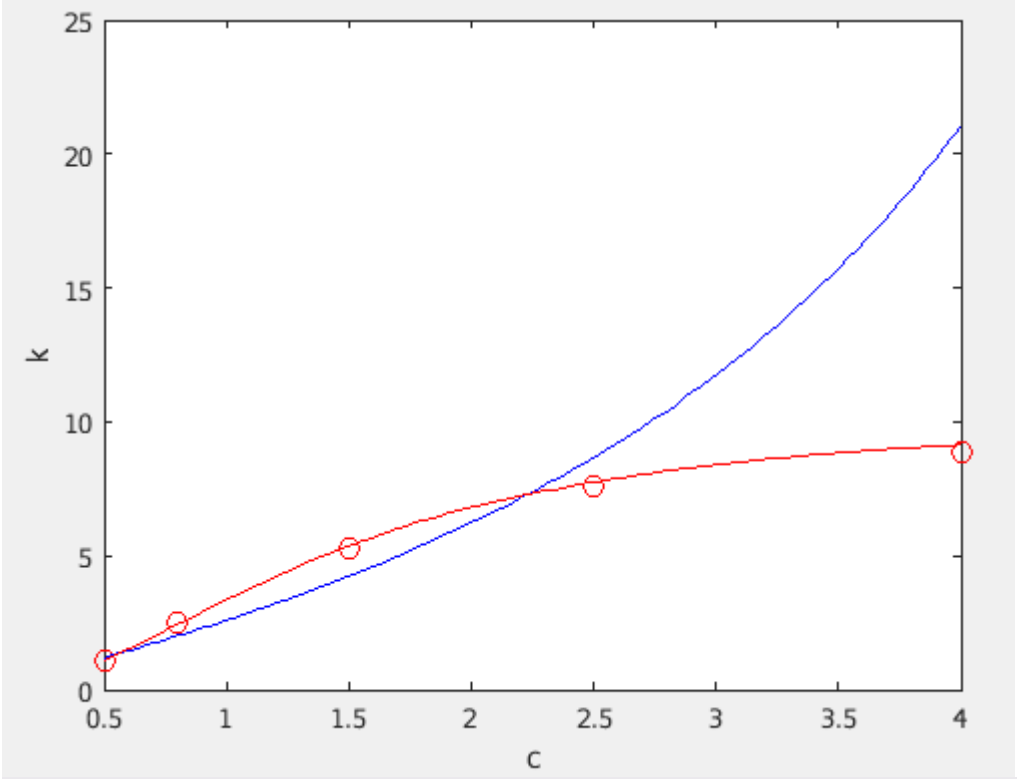


7. The following data represents bacteria growth. Two models have been proposed for this growth:

$$k = \frac{k_m c}{c_s + c} \qquad k = \frac{k_m c^2}{c_s + c^2}$$

Where k is the growth rate and c is the concentration. Fit the data below to both of these models by transforming them into linear relationships and finding the parameters k_m and c_s. Which model fits the data better? Predict the growth of bacteria when c = 2 mg/L.

c	0.5	0.8	1.5	2.5	4
k	1.1	2.5	5.3	7.6	8.9

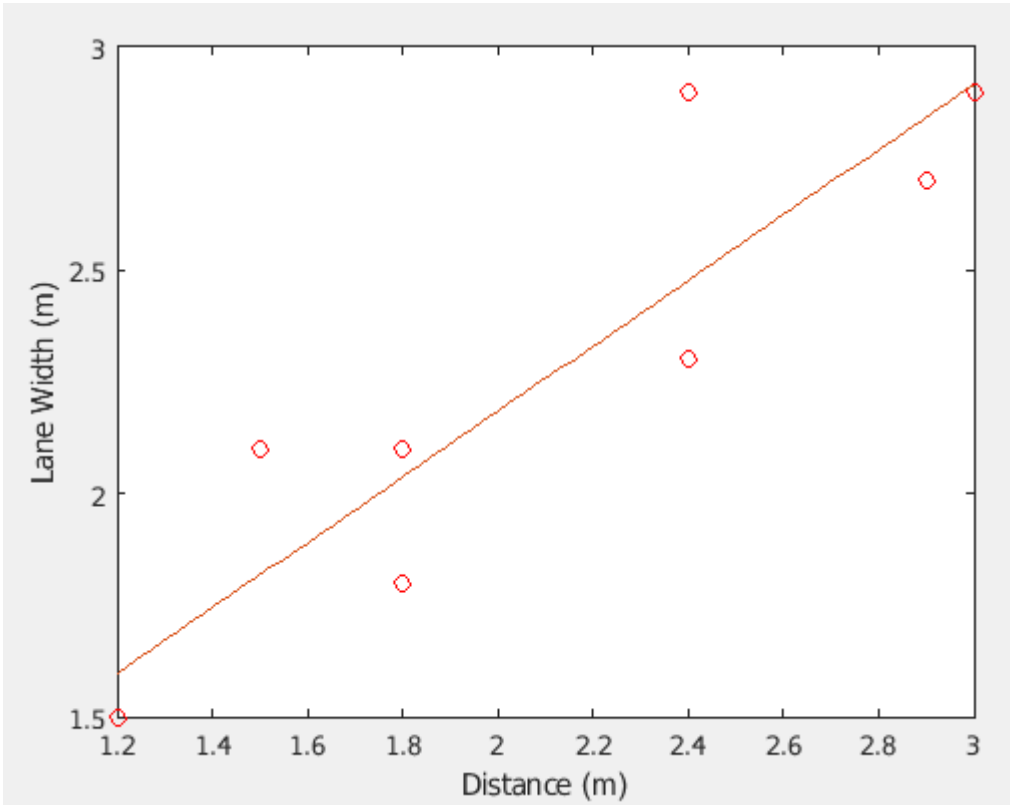



```
Command Window
>> Q7_Sol
The 2nd model is a better fit and gives the following growth for c = 2 mg/L
    6.7950
```

8. A transportation engineering study was conducted to determine the proper design of bike lanes. Data were gathered on bike lane widths and average distance between bikes and passing cars. The data from 9 streets are:

Distance, m	2.4	1.5	2.4	1.8	1.8	2.9	1.2	3	1.2
Lane Width, m	2.9	2.1	2.3	2.1	1.8	2.7	1.5	2.9	1.5

(a) Plot the data.
(b) Fit a straight line to the data with linear regression. Add this line to the plot.

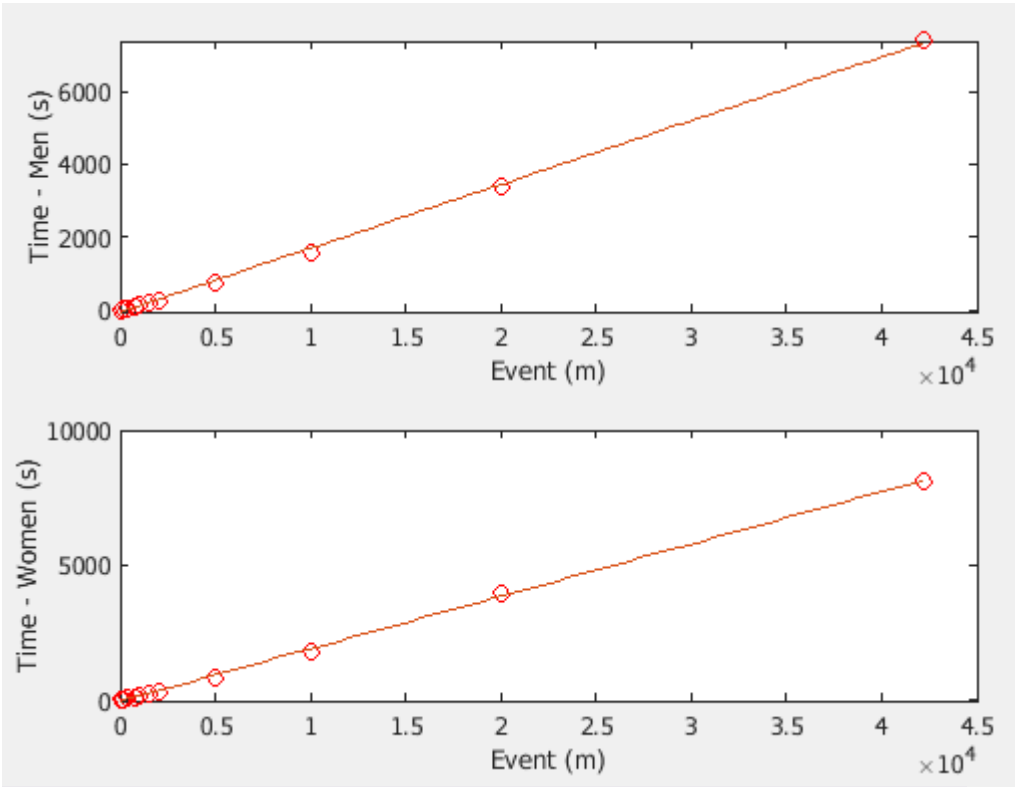


(c) If the minimum safe average distance between bikes and passing cars is considered to be 1.8m, determine the corresponding minimum lane width.

```
Command Window
>> Q8_Sol
The minimum lane width is 2.037(m)
```

9. The table below shows the 2015 world record times and holders for outdoor running. **Fit a power model** for each gender and use it to predict the record time for a half marathon (21,097.5m). Note: the actual records for the half marathon are 3503s (Tadese) and 3909s (Kiplagat) for men and women, respectively.

Event (m)	Time (s)	Men Holder	Time (s)	Women Holder
100	9.58	Bolt	10.49	Griffith-Joyner
200	19.19	Bolt	21.34	Griffith-Joyner
400	43.18	Johnson	47.60	Koch
800	100.90	Rudisha	113.28	Kratochvilova
1000	131.96	Ngeny	148.98	Masterkova
1500	206.00	El Guerrouj	230.07	Dibaba
2000	284.79	El Guerrouj	325.35	O'Sullivan
5000	757.40	Bekele	851.15	Dibaba
10,000	1577.53	Bekele	1771.78	Wang
20,000	3386.00	Gebrselassie	3926.60	Loroupe
42,195	7377.00	Kimetto	8125.00	Radcliffe




```
Command Window
>> Q9_Sol
Predicted time for men is 3630.2401 (s)
Predicted time for women is 4041.7457 (s)
```

Challenge Problems

10. Create a function file to fit a power model to some input data. Have the function return the best-fit coefficient α_2 and power β_2 along with the correlation coefficient for the untransformed model. In addition, use the subplot function to display graphs of both the transformed and untransformed equations along with the data. Test it

with the data from question 6.

11. Extend your function file from question 10 to include an option for the user to select whether they would like to fit their data to a linear model, an exponential model, a power model or saturation-growth equation. Test it with the data from the previous questions.