Exercises for Chapter 8

8.1 A study of the relationship between rough weight (X) and finished weight (Y) of castings was made. A sample of 12 casings was examined and the data presented below:

X Rough weight	Y Finished Weight
3.715	3.055
3.685	3.020
3.680	3.050
3.665	3.015
3.660	3.010
3.655	3.015
3.645	3.005
3.630	3.010
3.625	2.990
3.620	3.010
3.610	3.005
3.595	2.985

a) Make a scatter plot to see the relationship between X and Y

b) Use the method of least squares to calculate the coefficients in the simple linear regression model Y=a+bX

c) Calculate the standard errors of the estimated coefficients and determine if they are significant at the 95% confidence level

8.2 An equation is to be developed from which we can predict the gasoline mileage of an automobile based on its weight and the temperature at the time of operation. The model being estimated is: $Y = b_0 + b_1X_1 + b_2X_2$

The following data are available:

Car Number	X_1 Weight tons	X_2 Temperature °F	Y Miles per Gallon
1	1.35	90	17.9
2	1.90	30	16.5
3	1.70	80	16.4
4	1.80	40	16.8
5	1.30	35	18.8
6	2.05	45	15.5
7	1.60	50	17.5
8	1.80	60	16.4
9	1.85	65	15.9
10	1.40	30	18.3

a) Estimate the coefficients in the model using least squares. Use Eqn. 8.36.
Note that all spreadsheet programs have functions for matrix multiplication, matrix transposition, and matrix inversion.
Verify your answers using the regression function in a spreadsheet program.

b) Estimate the std. deviation of the "errors", s, by calculating the predicted value (using the coefficients from Part (a)) at each data point, and then calculating the standard deviation of the residuals (the differences between the data points and the predictions).
How many degrees of freedom does the estimate, s, have?
Verify your answers using the spreadsheet regression function output from Part (a).

c) Calculate the statistical significance of the coefficients, b_1 and b_2.

d) Calculate the predicted mileage (mpg) at $X_1 = 1.8$ tons and $X_2 = 70$ °F. Also calculate the error limits on that prediction. The error limits are usually just reported as $\hat{y} \pm 2\, s_{\hat{Y}}$ for simplicity.

e) Calculate R^2 for the model.
Verify your answers using the spreadsheet regression function output from Part (a).

f) Check your model assumptions by using plots of residuals. These should include a half-Normal plot and plots versus X_1, X_2, and $\hat{Y}$. Make sure you comment on what you learn/verify from each graph.

8.3 Using the data in Table 3.7 and the method of least squares fit the equation

$$\log(NO_2) = Y = b_0 + b_1X_1 + b_2 X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

to the data and verify that the coefficients obtained are identical to those on page 74.

8.4 In a project on corrosion resistance of steel plates a 10% solution of hydrochloric acid (HCl) was run over coated steel plates for different times and temperatures and the weight loss measured. The data is recorded below.

Sample	Time (X_1)	Temp (X_2)	Weight Loss Y
1	4	160	0.00068
2	4	160	0.00760
3	4	160	0.00810
4	4	180	0.00960
5	4	180	0.00920
6	4	180	0.00910
7	4	200	0.01150
8	4	200	0.01330
9	4	200	0.01240
10	6	160	0.00900
11	6	160	0.02090
12	6	160	0.03870
13	6	180	0.01000
14	6	180	0.01060
15	6	180	0.07640
16	6	200	0.01480
17	6	200	0.03940
18	6	200	0.01300
19	8	160	0.00760
20	8	160	0.00770
21	8	160	0.00830

a) Fit the equation $Y = b_o + b_1X_1 + b_2X_2$ to the data using the method of least squares and determine if any of the coefficients are significant at the 95% confidence level.

b) Calculate the predicted values and residuals from the model you fit in a) and plot residuals versus X_1, residuals versus X_2, residuals versus predicted values, and a normal or half normal plot of residuals as shown in Section 8.6.

c) Fit the equation $Y = b_o + b_1X_1 + b_2X_2 + b_{11}X_1^2$ to the data using the method of least squares and determine if any of the coefficients are significant at the 95% confidence level.

d) Calculate the residuals and predicted values from the model you fit in c) and make the same plots as in b)

e) Fit the equation $\log_e(Y) = b_o + b_1X_1 + b_2X_2 + b_{11}X_1^2$ to the data using the method of least squares and determine if any of the coefficients are significant at the 95% confidence level.

f) Calculate the residuals and predicted values from the model you fit in e) and make the same plots as in d)

g) Which model do you prefer and why?

8.5 Since humidity influences evaporation, a knowledge of the relationship will allow a painter, when applying water based paints, to adjust his/her spray gun to account for humidity. The following data were obtained (based on "Evaporation During Spray out of a Typical Water-Reducible Paint at Various Humidities", *Journal of Coating Technology*, Vol 65, 1983):

Run	X Relative Humidity	Y Solvent Evaporation	Run	X Relative Humidity	Y Solvent Evaporation
1	35.3	11.0	14	39.1	9.6
2	29.7	11.1	15	46.8	10.9
3	30.8	12.5	16	48.5	9.6
4	58.8	8.4	17	59.3	10.1
5	61.4	9.3	18	70.0	8.1
6	71.3	8.7	19	70.0	6.8
7	74.4	6.4	20	74.4	8.9
8	76.7	8.5	21	72.1	7.7
9	70.7	7.8	22	58.1	8.5
10	57.5	9.1	23	44.6	8.9
11	46.4	8.2	24	33.4	10.4
12	28.9	12.2	25	28.6	11.1
13	28.1	11.9			

a) Estimate the coefficients in the model, $Y = a + bX$, using least squares.

b) Estimate the std. deviation of the "errors", s, by calculating the predicted value (using the coefficients from Part (a)) at each data point, and then calculating the standard deviation of the residuals (the differences between the data points and the predictions). How many degrees of freedom does the estimate, s, have?

c) Calculate the 95% confidence intervals for the intercept, a, and the slope, b. Hint: the 95% confidence intervals are $\hat{a} \pm t\, s_{\hat{a}}$ and $\hat{b} \pm t\, s_{\hat{b}}$.

d) Calculate error limits on for the predictions of the equation at X = 25, 50, and 75. The error limits are usually just reported as $\hat{y} \pm 2\, s_{\hat{Y}}$ for simplicity.

8.6 Verify that if Equation 8.41 ($\mathbf{V}(\hat{\mathbf{B}}) = (\mathbf{X}^T\mathbf{X})^{-1}\,\sigma^2$) is used for the case where the model is $Y = a + bX$, the standard deviation of $\hat{b}$ (i.e. $s_{\hat{b}}$) is the same as given in Eqn 8.19.

8.7 You have the following four X-Y pairs (data points):

X	Y
5	3
6	3
8	5
9	5

a) What are the coefficients for the best fitting straight line?

b) What is your estimate of the standard deviation of the data, s?

c) Are the intercept and the slope statistically significant?

8.8 You purchase a chemical raw material that has trace amounts of impurity that is detrimental to your process (namely it poisons the catalyst). You have a specification on the maximum amount of impurity that your supplier is allowed to ship to you, but you do not want to rely on your supplier to only ship you good material. So you test each batch of raw material when it comes in. The bad news is that the test is expensive. The good news is that you think you found a cheap replacement for the test, namely measuring the absorbence at a specific frequency. You took the data below to test your proposed new analytical method.

X Absorbence	Y Concentration, ppm
0.0335	3.91
0.0489	7.81
0.0571	15.63
0.0488	31.25
0.0827	62.5
0.1662	125
0.3174	250
0.5927	500
0.8877	750
1.1705	1000

(a) What is the best straight line calibration line based on your data?

(b) Is a straight line an adequate model?

(c) How good is the new method? In other words, what is the error in a prediction (particularly at low concentrations / absorbences)?