

In an environmental engineering study of a certain chemical reaction, the concentration of 18 separately prepared solutions was recorded at different times (three measurements at each of six times). The data are given in the following table.

| Solution Number ID | Time (Hour) X | Concentration (mg/ml) Y |
|-----------------------|------------------|----------------------------|
| 1 | 6 | 0.029 |
| 2 | 6 | 0.032 |
| 3 | 6 | 0.027 |
| 4 | 8 | 0.079 |
| 5 | 8 | 0.072 |
| 6 | 8 | 0.088 |
| 7 | 10 | 0.181 |
| 8 | 10 | 0.165 |
| 9 | 10 | 0.201 |
| 10 | 12 | 0.425 |
| 11 | 12 | 0.384 |
| 12 | 12 | 0.472 |
| 13 | 14 | 1.130 |
| 14 | 14 | 1.020 |
| 15 | 14 | 1.249 |
| 16 | 16 | 2.812 |
| 17 | 16 | 2.465 |
| 18 | 16 | 3.099 |

You will try to estimate three regression models to assess the fit of each model to the data.

a. $Y_i = \beta_0 + \beta_1 X_i + e_i$

b. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i$

c. $\ln Y_i = \beta_0 + \beta_1 X_i + e_i$

You could enter compute statements to create the matrix X and vector Y in the syntax window and employ all of the following computations – I recommend that you enter the data in an SPSS data window. Be sure to enter a column of 1s (name it “const”) to hold a place in the X matrix for computing the intercept β_0 (the constant).

For each of the models:

1. plot each of the fitted equations in a scatter plot
2. estimate the parameters: β s and $SE(\beta)$
3. determine the proportion of variation in Y accounted for by X
4. report the ANOVA table and determine the adequacy of fit for each model
5. Which one of the models is better for describing this data set? Explain.

Some SPSS Procedures for the assignment.

Note the lines in syntax to compute Time^2 (X2), the combined X which includes the constant (estimated from the column of 1s) and Time, the combined Xn (for the nonlinear model including constant, Time, Time^2), and natural log of Y (lny). To compute each model, with Time, Time^2 , and lnY, replace X in the commands below with Xn or replace Y with lny.

You will have to complete the syntax below.

```
Matrix.

Get TIME /variables time.
Get Y /variables concent.
Compute ONES = make(nrow(Y),1,1).

Compute TIME2 = TIME**2.
Compute X = {ONES,TIME}.
Compute Xn = {ONES,TIME,TIME2}.
Compute lny = ln(y).

compute meany = csum(y)/nrow(y).
compute dy = y-meany.

compute betas =          .
print betas.

compute I = Ident(nrow(x),nrow(x)).
compute sse =          .
print sse.

compute mse = sse/(nrow(X)-ncol(x)).
print mse.

compute varbeta =          .
compute sebeta = sqrt(diag(varbeta)).
print sebeta.

compute sst =          .
print sst.

compute SSR =          .
print SSR.

compute rsquare =          .
print rsquare.

End Matrix.
```

The first four rows of the SPSS file should look something like this:

| id | const | time | concent |
|----|-------|------|---------|
| 1 | 1 | 6 | .029 |
| 2 | 1 | 6 | .032 |
| 3 | 1 | 6 | .027 |
| 4 | 1 | 8 | .079 |

With the data entered in the SPSS data window, complete the scatter plots with the following procedures:

1. In the menu bar: Graphs → Scatter → Define → X-Axis (time), Y-Axis (concent) → OK
2. Double Click on the graph in the output window—this allows you to edit the graph.
3. In the menu bar: Chart → Options → Select “Fit Line”
4. Fit Line Options → Select “Linear” once then redo to select “Quadratic” on a separate graph

What you need to turn in:

1. Scatter Plots of
 - a. Linear relationship and
 - b. Quadratic relationship.
2. Output for each model (linear, quadratic, and log-y) based on the print statements in the syntax above.
3. Summarize the output from each model in an ANOVA table.
4. Report the proportion of variance in Y accounted for by X from each model.
5. Identify the model which does the best job of explaining variation – explain why you selected that model.

Additional Syntax that you might find helpful:

The syntax above can be used if you input the data into an SPSS data window. That you can call in the variables you need with the GET commands.

If you want to do the regression assignment completely within SPSS syntax and not rely on using the data window, you can use the following syntax to compute a vector of TIME values, the TIME² term, a ONES vector, and combine them into a new augmented X matrix.

```
Compute TIME = {6;6;6;8;8;8;10;10;10;12;12;12;14;14;14;16;16;16}.
```

```
Compute TIME2 = TIME**2.
```

```
Compute ONES = make(nrow(TIME),1,1).
```

```
Compute X1 = {ONES, TIME}.
```

```
Compute X2 = {ONES, TIME, TIME2}.
```

```
Compute Y = {.029;.032;.027;.079;.072;.088;.181;.165;.201;.425;.384;  
.472;1.130;1.020;1.249;2.812;2.465;3.099}.
```

```
Compute lny = ln(Y).
```

Notice that you create each vector (variable) separately, then combine them in one statement. Since each column is a vector, the final COMPUTE X statement includes each column vector separated by commas – for columns.

So, X1 will be useful for models a and c, X2 will be useful for model b, and lny will be useful for model c..