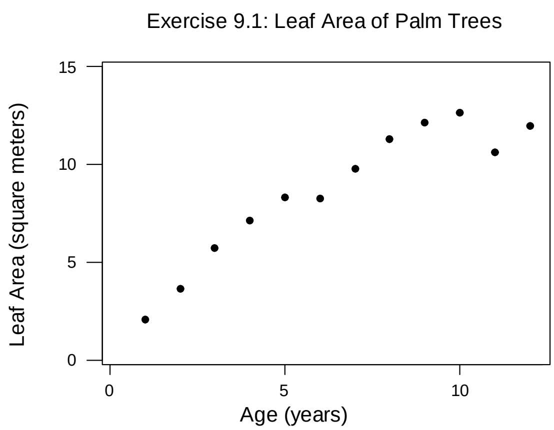
CHAPTER 9

A note on computing with SPSS (Version 11.5):

We use the SPSS software to fit the nonlinear regression models of Chapter 9. SPSS works like a spreadsheet program. We enter the data into the various columns of the spreadsheet and use the tabs: Analyze > Regression > Nonlinear. We write out the model equation and specify initial parameter values. We can save the fitted values and the residuals (also the derivatives of the objective function) into columns of the worksheet.

Several options for the iterative nonlinear estimation procedure are available. In the following examples we have used the Levenberg-Marquardt algorithm. Options for specifying the number of iterations and various convergence cutoffs are available. See the SPSS on-line help for further discussion and examples.

9.1 A graph of the leaf area against the age of the palm tree is given below.



Note that there is not an abundance of data points to determine the model. The graph indicates that the relationship between leaf area and age is not linear; a quadratic component needs to be added to the model. The estimation results for the quadratic model $y = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \varepsilon$ (Minitab output) is shown below. The quadratic coefficient is clearly needed; the estimate of the coefficient for Age**2 is -0.09616, with a significant t-ratio of -4.95.

Regression Analysis: Area (square meters) versus Age, Age2**

```
The regression equation is
Area (square meters) = - 0.123 + 2.15 Age - 0.0962 Age**2

Predictor        Coef     SE Coef          T        P
Constant      -0.1234      0.7334      -0.17    0.870
Age            2.1496      0.2594       8.29    0.000
Age**2       -0.09616     0.01942      -4.95    0.001

S = 0.7096      R-Sq = 96.6%     R-Sq(adj) = 95.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2     128.071      64.036    127.19    0.000
Residual Error     9       4.531       0.503
Total             11     132.603
```

Rasch/Sedlacek use the Gompertz model $\mathrm{y} = \mu + \varepsilon = \alpha \exp[-\beta \exp(-\gamma \mathrm{Age})] + \varepsilon$ with parameters $\alpha > 0, \beta > 0, \gamma > 0$. Before fitting this model, we need to determine suitable starting values for the iterative nonlinear parameter estimation. The graph indicates that the saturation level for large values of Age is about 15. Hence a suitable starting value for α is given by 15. For Age = 1, the response is about 2; for Age = 5, the response is roughly 7. The model equation implies $-\beta \exp(-\gamma) = \ln(2/15)$ and $-\beta \exp(-5\gamma) = \ln(7/15)$. This implies $\exp(4\gamma) = [\ln(2/15)]/[\ln(7/15)]$ and $\gamma = \{\ln[\ln(2/15))/\ln(7/15)]\}/4 \approx 0.25$. Finally, $-\beta \exp(-\gamma) = \ln(2/15)$ and $\beta = -\ln(2/15)\exp(\gamma) \approx 2.6$. The starting values $\alpha = 15$, $\beta = 2.6$ and $\gamma = 0.25$ are used in the SPSS nonlinear regression routine. The (SPSS) outcome is given below:

```
 Iteration  Residual SS           A           B           C

     1      12.59000092  15.0000000  2.60000000  .250000000
     1.1    15.64377972  11.3812687  2.34691685  .336739045
     1.2    6.515778841  13.4641276  2.14122037  .271436482
     2      6.515778841  13.4641276  2.14122037  .271436482
     2.1    6.243186484  12.0109653  2.42204992  .359733910
     3      6.243186484  12.0109653  2.42204992  .359733910
     3.1    5.136619171  12.4921144  2.50012161  .359000316
     4      5.136619171  12.4921144  2.50012161  .359000316
     4.1    5.136518308  12.4937910  2.49764737  .358922047
     5      5.136518308  12.4937910  2.49764737  .358922047
     5.1    5.136518286  12.4936881  2.49773050  .358935226

Run stopped after 11 model evaluations and 5 derivative evaluations.
Iterations have been stopped because the relative reduction between
successive residual sums of squares is at most SSCON = 1.000E-08
```

Nonlinear Regression Summary Statistics   Dependent Variable AREA

Source	DF	Sum of Squares	Mean Square
Regression	3	1023.43418	341.14473
Residual	9	5.13652	.57072
Uncorrected Total	12	1028.57070	
(Corrected Total)	11	132.60269	

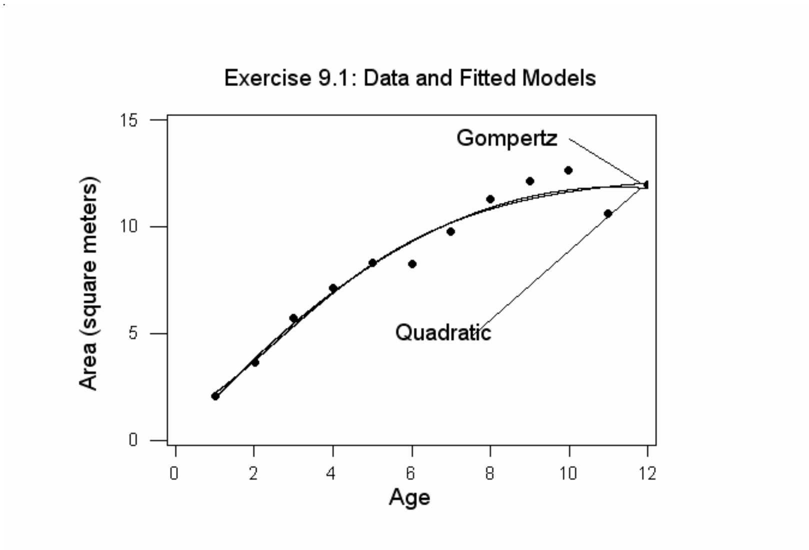
R squared = 1 - Residual SS / Corrected SS = .96126

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
A (α)	12.493688057	.683789772	10.946848127	14.040527986
B (β)	2.497730497	.440644079	1.500924338	3.494536656
C (γ)	.358935226	.066769083	.207893067	.509977385

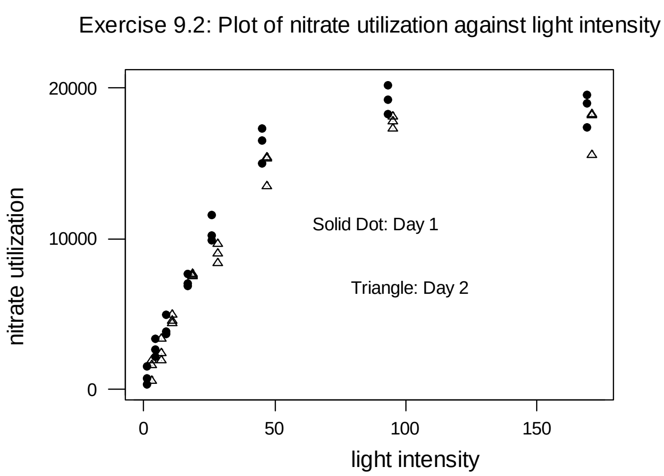
Asymptotic Correlation Matrix of the Parameter Estimates

	A	B	C
A	1.0000	-.4983	-.8306
B	-.4983	1.0000	.8339
C	-.8306	.8339	1.0000

The estimate of α is 12.5; the estimate of β is 2.5, and the estimate of γ is 0.36. All estimates are statistically significant. There is a fair amount of correlation, especially between the estimates of γ and α (-0.83) and the estimates of γ and β (0.83). The coefficient of determination (0.961) is similar to the R^2 from the quadratic regression. There is little difference between the fits of the quadratic regression (which is linear in the parameters) and the Gompertz model (which is nonlinear in the parameters). Both models lead to similar fitted curves. One difference is that the fitted values for the Gompertz model increase with age to an asymptotic value, whereas the quadratic curve starts to decrease with age after having reached a maximum. However, over the observed age range the two fitted models are virtually indistinguishable.



9.2 A scatter plot of nitrate utilization versus light intensity is shown below. We use solid circles for day 1 observations, and triangles for day 2 observations. Furthermore, we have added some jitter to the light intensity in order to emphasize the differences between the measurements of day 1 and day 2. The day 2 measurements are slightly lower, especially at increasing light intensity.



Michaelis-Menton model: Nitrate utilization reaches an asymptote of about 20,000 for large light intensity. Letting x go to infinity in the model equation

$$\frac{\beta_1 \mathrm{x}}{\beta_2 + \mathrm{x}} = \frac{\beta_1}{1 + (\beta_2 / \mathrm{x})} \approx 20{,}000$$

leads to the starting value $\beta_1 \approx 20{,}000$. Furthermore, the average nitrate utilization at light intensity 2.2 is 1075. Solving the model equation with $\beta_1 = 20{,}000$ leads to the starting value $\beta_2 = 38.7$.

Using these starting values in the SPSS nonlinear regression routine results in the following estimation results:

```
Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                 DF  Sum of Squares  Mean Square

  Regression              2  6467226758.31  3233613379.15
  Residual               46  96536195.6932  2098612.94985
  Uncorrected Total      48  6563762954.00
  (Corrected Total)      47  2076766799.92

  R squared = 1 - Residual SS / Corrected SS =     .95352

                                           Asymptotic 95 %
                          Asymptotic     Confidence Interval
  Parameter   Estimate    Std. Error     Lower         Upper

  B1        23582.527043 889.35646658 21792.345325 25372.708760
  B2        34.243774004  3.427314571 27.344947587 41.142600421

  Asymptotic Correlation Matrix of the Parameter Estimates

                  B1        B2

  B1          1.0000     .8785
  B2            .8785    1.0000
```

<u>Exponential rise model:</u> Nitrate utilization reaches an asymptote of about 20,000 for large light intensity. Letting x go to infinity in the equation for the exponential rise model leads to the starting value $\beta_1 \approx 20{,}000$. The average nitrate utilization at light intensity 2.2 is 1075. Solving the model equation with $\beta_1 = 20{,}000$ leads to the starting value $\beta_2 = -\frac{1}{2.2}\ln\left[1 - \frac{1075}{20{,}000}\right] = 0.025$. Using these starting values in the SPSS nonlinear regression program results in the estimation results:

```
Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                 DF  Sum of Squares  Mean Square

  Regression              2  6504309173.87  3252154586.93
  Residual               46  59453780.1310  1292473.48111
  Uncorrected Total      48  6563762954.00
  (Corrected Total)      47  2076766799.92
```

```
R squared = 1 - Residual SS / Corrected SS =     .97137

                                         Asymptotic 95 %
                        Asymptotic     Confidence Interval
Parameter   Estimate    Std. Error     Lower         Upper

B1        19014.305975 398.04663684 18213.079652 19815.532299
B2          .030021624   .001629334   .026741945   .033301303

Asymptotic Correlation Matrix of the Parameter Estimates

                B1        B2

B1          1.0000    -.7393
B2          -.7393    1.0000
```

Quadratic Michaelis-Menton model: Starting with $\beta_1 = 20{,}000$ and $\beta_2 = 38.7$ (from the earlier Michaelis-Menton model) and a small value for the parameter in the quadratic component ($\beta_3 = 0.1$) leads to the following results:

```
Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                 DF  Sum of Squares  Mean Square

  Regression              3  6520540397.33  2173513465.78
  Residual               45  43222556.6654   960501.25923
  Uncorrected Total      48  6563762954.00
  (Corrected Total)      47  2076766799.92

  R squared = 1 - Residual SS / Corrected SS =     .97919

                                           Asymptotic 95 %
                          Asymptotic     Confidence Interval
  Parameter   Estimate    Std. Error     Lower         Upper

  B1        66769.034924 17585.504714 31350.010284 102188.05956
  B2        137.82679758 43.735712594 49.738550634 225.91504453
  B3          .011281055   .004496402   .002224837   .020337274


 Asymptotic Correlation Matrix of the Parameter Estimates

                  B1        B2        B3

  B1          1.0000     .9964     .9941
  B2           .9964    1.0000     .9856
  B3           .9941     .9856    1.0000
```

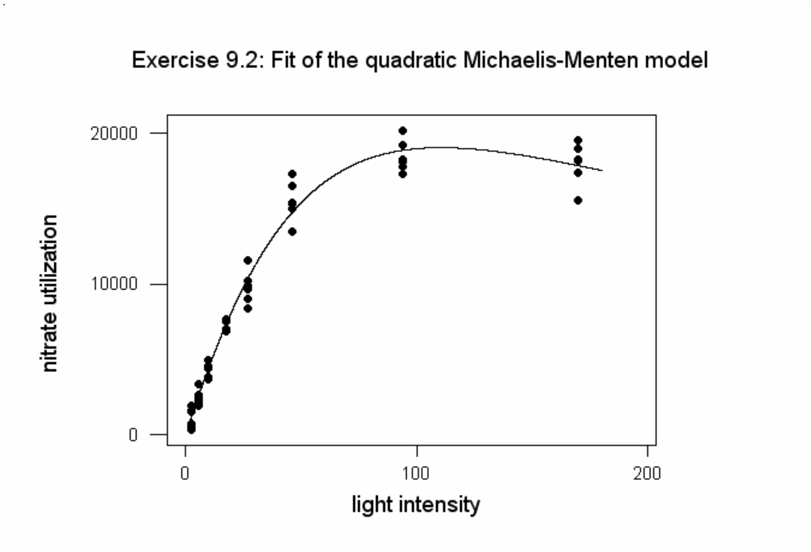
Modified exponential rise model: Using $\beta_1 = 20{,}000$ and $\beta_2 = 0.025$ from the earlier exponential rise model and a small value for $\beta_3 = 0.01$ leads to the following results:

```
Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                 DF  Sum of Squares  Mean Square

  Regression              3  6519117089.28  2173039029.76
  Residual               45  44645864.7154   992130.32701
  Uncorrected Total      48  6563762954.00
  (Corrected Total)      47  2076766799.92

  R squared = 1 - Residual SS / Corrected SS =     .97850

                                             Asymptotic 95 %
                         Asymptotic     Confidence Interval
  Parameter   Estimate    Std. Error     Lower         Upper

  B1        33551.454219 9502.1687711 14413.103896 52689.804543
  B2          .018534079   .003572151   .011339397   .025728761
  B3          .003221159   .001338559   .000525162   .005917155

  Asymptotic Correlation Matrix of the Parameter Estimates

                  B1        B2        B3

  B1          1.0000    -.9898     .9948
  B2          -.9898    1.0000    -.9741
  B3           .9948    -.9741    1.0000
```

All four models lead to large R^2. The Michaelis-Menton and its quadratic extension lead to R^2 of 0.954 and 0.979, respectively. Carrying out an F-test for the significance of the quadratic component in the Michaelis-Menton model leads to the F-statistic $F = [96{,}536{,}195 - 43{,}222{,}556]/[43{,}222{,}556/45] = 55.5$, which is highly significant. This shows that the quadratic extension represents a significant improvement.

Similarly, the exponential rise model and its extension lead to R^2 of 0.971 and 0.979, respectively. The F-test for the significance of the extra component in the exponential rise model leads to the F-statistic
$F = [59{,}453{,}780 - 44{,}645{,}864]/[44{,}645{,}864/45] = 14.9$, which is also highly significant.

The extensions are beneficial. The modified Michaelis-Menton and the modified exponential rise models perform similarly. In the following graph we show the fit of the quadratic Michaelis-Menton model; the fitted values of the modified exponential rise model are virtually indistinguishable.



Standard Michaelis-Menton model with an indicator for the change of day: The final parameter estimates in the previous Michaelis-Menton model, $\hat{\beta}_1 = 23{,}500$ and $\hat{\beta}_2 = 34.2$, are taken as the starting values in the iterative nonlinear estimation. Small values for the day indicator $\alpha_1 = -1000, \alpha_2 = -1$ are used as the starting values for the two additional parameters. The estimation results are given below:

```
Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                  DF  Sum of Squares  Mean Square

  Regression               4  6477253424.57  1619313356.14
  Residual                44  86509529.4274  1966125.66881
  Uncorrected Total       48  6563762954.00
  (Corrected Total)       47  2076766799.92

  R squared = 1 - Residual SS / Corrected SS =     .95834

                                           Asymptotic 95 %
                          Asymptotic     Confidence Interval
  Parameter   Estimate    Std. Error     Lower         Upper

  B1        24743.334444 1241.1211323 22242.019158 27244.649730
  B2        35.275400267  4.656586052 25.890667730 44.660132803
  A1        -2328.743446 1720.3472191 -5795.875448 1138.3885567
  A2        -2.172827290  6.626226364 -15.52710905 11.181454466
```

```
Asymptotic Correlation Matrix of the Parameter Estimates

                  B1        B2        A1        A2

  B1          1.0000     .8810    -.7214    -.6191
  B2           .8810    1.0000    -.6356    -.7028
  A1          -.7214    -.6356    1.0000     .8781
  A2          -.6191    -.7028     .8781    1.0000
```

The F-statistic for testing the null hypothesis $\alpha_1 = \alpha_2 = 0$ is $F = [(96{,}536{,}195 - 86{,}509{,}529)/2]/[86{,}509{,}529/44] = 2.55$. The probability value from the F(2,44) distribution is $P[F(2,44) \geq 2.55] = 1 - 0.91 = 0.09$. Hence there is only weak evidence for including a day effect. The individual confidence intervals for α_1 and α_2 cover zero, which makes the individual interpretation of the two day-effect parameters difficult. These estimates are also quite correlated.

Quadratic Michaelis-Menton model with an indicator for the change of day: The final values from the earlier quadratic model $\hat{\beta}_1 = 66{,}700, \hat{\beta}_2 = 138, \hat{\beta}_3 = 0.01$ and small values for the three parameters associated with the day indicators, $\alpha_1 = -2000, \alpha_2 = -2, \alpha_3 = 0.001$, are used as the starting values in the iterative nonlinear SPSS estimation. The estimation results are given below:

```
Run stopped after 10 model evaluations and 5 derivative evaluations.
Iterations have been stopped because the relative reduction between
successive residual sums of squares is at most SSCON = 1.000E-08


Nonlinear Regression Summary Statistics     Dependent Variable NITRATE

  Source                 DF  Sum of Squares  Mean Square

  Regression              6  6531740362.05  1088623393.67
  Residual               42  32022591.9535   762442.66556
  Uncorrected Total      48  6563762954.00
  (Corrected Total)      47  2076766799.92

 R squared = 1 - Residual SS / Corrected SS =     .98458

                                             Asymptotic 95 %
                         Asymptotic     Confidence Interval
  Parameter   Estimate    Std. Error     Lower         Upper

  B1        89797.916970 37540.345749 14038.432096 165557.40184
  B2        186.61862445 89.984553967  5.022442558 368.21480635
  A1        -38897.78690 39982.748874 -119586.2408 41790.667033
  A2        -83.09078151 96.727346453 -278.2944695 112.11290653
  B3          .016252421   .009207288  -.002328638   .034833481
  A3         -.008449660   .009916211  -.028461384   .011562065
```

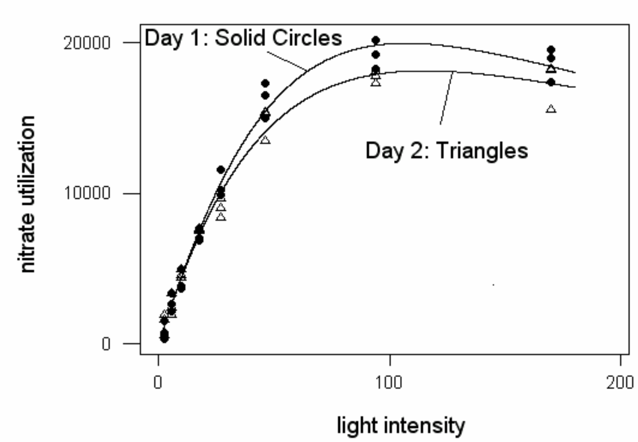
```
Asymptotic Correlation Matrix of the Parameter Estimates
```

	B1	B2	A1	A2	B3	A3
B1	1.0000	.9978	-.9389	-.9283	.9965	-.9252
B2	.9978	1.0000	-.9369	-.9303	.9913	-.9204
A1	-.9389	-.9369	1.0000	.9971	-.9356	.9953
A2	-.9283	-.9303	.9971	1.0000	-.9222	.9894
B3	.9965	.9913	-.9356	-.9222	1.0000	-.9285
A3	-.9252	-.9204	.9953	.9894	-.9285	1.0000

The F-statistic for testing the null hypothesis $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is given by $F = [(43{,}222{,}556 - 32{,}022{,}591)/3]/[32{,}022{,}591/42] = 4.90$. The probability value from the F(3,42) distribution is $P[F(3,42) \geq 4.90] = 1 - 0.995 = 0.005$, showing that the indicators for the day effect help explain the variation. Individually the three parameters are statistically insignificant and also highly correlated. This makes an individual interpretation of the estimates difficult.

The graph shown below compares the quadratic Michaelis-Menton model with and without the day indicator. The graph shows that the quadratic Michaelis-Menton model with a day indicator is capable of expressing the day differences.

Exercise 9.2: Quadratic Michaelis-Menton model with day indicator



9.3

Model 1: The logarithmic transformation of the first model leads to

$$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \ln(\varepsilon)$$

A standard multiple linear regression of $\ln(y)$ on $\ln(x_1)$ and $\ln(x_2)$ leads to the estimates of $\alpha = \ln(\beta_0), \beta_1$, and β_2 The estimate of β_0 can be obtained from $\beta_0 = \exp(\alpha)$. When carrying out the regression with the transformed variables we need to assume that the error $\ln(\varepsilon)$ satisfies the standard regression assumptions.

Model 2: Taking the reciprocal of the response in the second model leads to

$$1/y = \beta_0 + \beta_1 x + \varepsilon$$

A simple linear regression of (1/y) on x_1 leads to the estimates of β_0, β_1.

Model 3: The reciprocal of the response and a subsequent logarithmic transformation leads to the model

$$\ln[(1/y) - 1] = \beta_0 + \beta_1 x_1 + \ln(\varepsilon)$$

A simple linear regression of $\ln[(1/y) - 1]$ on x_1 leads to the estimates of β_0, β_1. We need to assume that the error $\ln(\varepsilon)$ satisfies the standard regression assumptions.

9.4 Search the literature.