# Nonlinear curve for predicting height

## Contents



## 1. Introduction

In this example we consider the fitting of a double exponential model to the simulated height measurements of 130 females between the ages of 2 and 22. Data are given in **fem_height.lsf** and the data for the first female are shown below.

fem_height.lsf

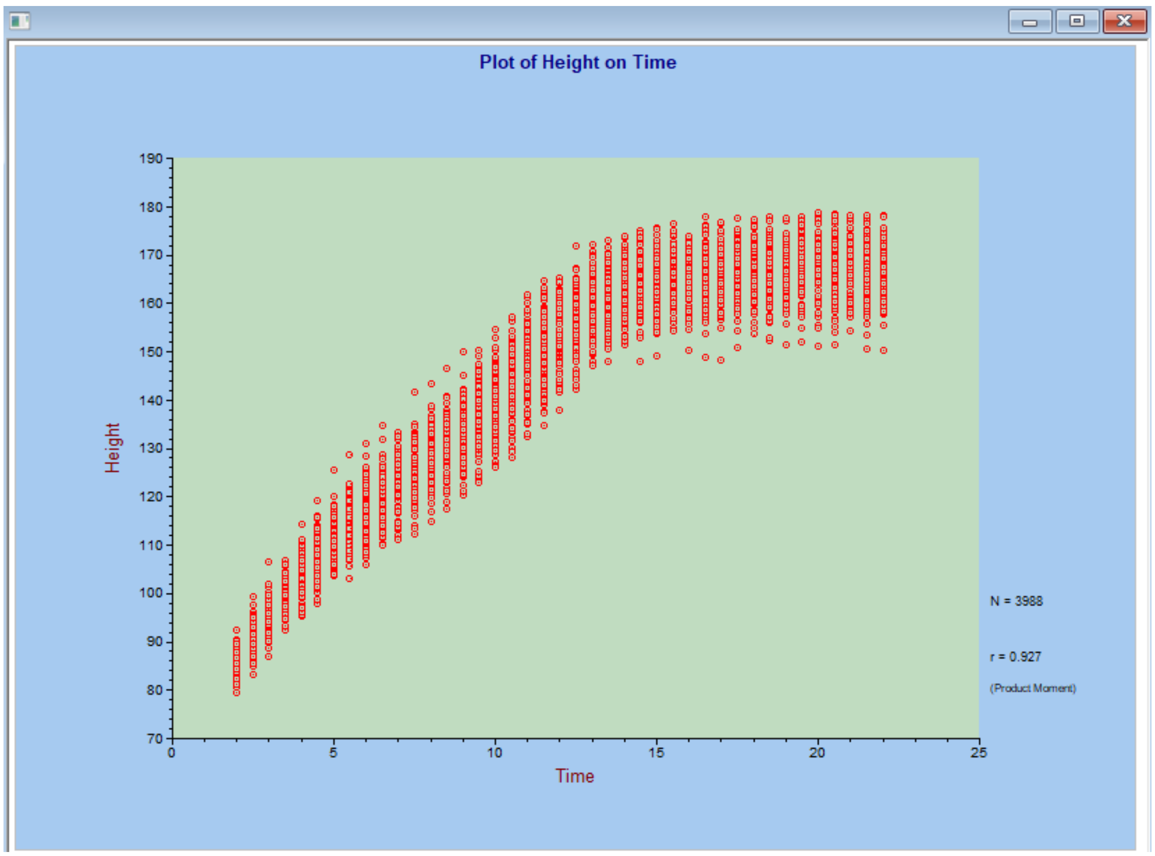
| | Case | Occasio | Height | Time |
|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 86.53 | 2.00 |
| 2 | 1.00 | 2.00 | 92.23 | 2.50 |
| 3 | 1.00 | 3.00 | 96.90 | 3.00 |
| 4 | 1.00 | 4.00 | 99.52 | 3.50 |
| 5 | 1.00 | 5.00 | 103.62 | 4.00 |
| 6 | 1.00 | 6.00 | 106.90 | 4.50 |
| 7 | 1.00 | 9.00 | 117.96 | 6.00 |
| 8 | 1.00 | 11.00 | 122.87 | 7.00 |
| 9 | 1.00 | 12.00 | 125.01 | 7.50 |
| 10 | 1.00 | 15.00 | 132.84 | 9.00 |
| 11 | 1.00 | 16.00 | 137.01 | 9.50 |
| 12 | 1.00 | 18.00 | 146.28 | 10.50 |
| 13 | 1.00 | 19.00 | 148.23 | 11.00 |
| 14 | 1.00 | 20.00 | 150.66 | 11.50 |
| 15 | 1.00 | 21.00 | 158.22 | 12.00 |
| 16 | 1.00 | 22.00 | 157.99 | 12.50 |
| 17 | 1.00 | 23.00 | 160.56 | 13.00 |
| 18 | 1.00 | 24.00 | 162.91 | 13.50 |
| 19 | 1.00 | 25.00 | 163.04 | 14.00 |
| 20 | 1.00 | 26.00 | 165.88 | 14.50 |
| 21 | 1.00 | 27.00 | 166.08 | 15.00 |
| 22 | 1.00 | 28.00 | 166.00 | 15.50 |
| 23 | 1.00 | 29.00 | 168.56 | 16.00 |
| 24 | 1.00 | 30.00 | 167.60 | 16.50 |
| 25 | 1.00 | 31.00 | 168.32 | 17.00 |
| 26 | 1.00 | 32.00 | 167.68 | 17.50 |
| 27 | 1.00 | 33.00 | 169.96 | 18.00 |
| 28 | 1.00 | 34.00 | 168.21 | 18.50 |
| 29 | 1.00 | 38.00 | 169.22 | 20.50 |

The variables of interest are:

- Case: the identifier of each individual
- Occasio: the measurement occasion
- Height: the height of the individual at that point in time
- Time: the time of measurement (in years)

Note that not all individuals had complete data over the time period: some individuals had only 26 measurements, while others had up to 35 measurements.

A scatterplot of the height measurements over time is shown below. The relationship between height and time curve is nonlinear. Moreover, the curve shows two inflection points, at approximately 6 and again at approximately 13 years of age. While a nonlinear curve like the logistic curve has a single inflection point, multiple inflection points are best handled by fitting multiple component curves.




## 2. Double logistic curve

We now fit a double logistic curve to these data. The 3988 observations, nested within 130 level-2 units, were simulated according to the following model:

$$y = b_1 / (1 + s * \exp(b_2 - b_3 * time) + c_1 / (1 + s * \exp(c_2 - c_3 * time) + e$$

The level-2 model is:

$$b_1 = \beta_1 + u_1$$
$$b_2 = \beta_2 + u_2$$
$$b_3 = \beta_3 + u_3$$
$$c_1 = \beta_3 + u_3$$
$$c_2 = \beta_4 + u_4$$
$$c_3 = \beta_5 + u_5$$

The data values used in the simulation study were:

Values used in the simulation study:

| Coefficients | Beta |
|---|---|
| b1 | 143.86781 |
| b2 | 0.13589 |
| b3 | 0.26718 |
| c1 | 24.39933 |
| c2 | 11.44060 |
| c3 | 1.01441 |

| Variance estimate | Level 1 |
|---|---|
| Sigma**2 | 0.80994 |

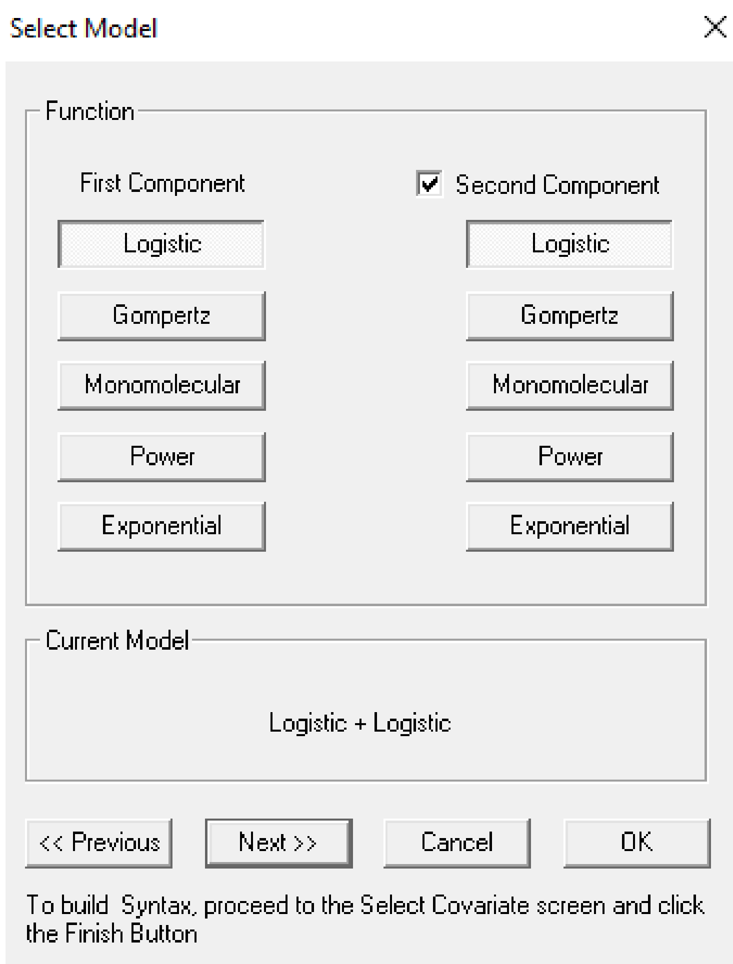
| Covariances | Level 2 |
|---|---|
| u1,u1 | 41.95183 |
| u2,u1 | 0.39510 |
| u2,u2 | 0.00809 |
| u3,u1 | -0.04130 |
| u3,u2 | 0.00024 |
| u3,u3 | 0.00066 |
| u4,u1 | -13.72547 |
| u4,u2 | -0.12930 |
| u4,u3 | 0.01351 |
| u4,u4 | 16.73571 |
| u5,u1 | 5.67229 |
| u5,u2 | 0.09164 |
| u5,u3 | -0.00396 |
| u5,u4 | -6.73733 |
| u5,u5 | 5.05311 |
| u6,u1 | 0.28877 |
| u6,u2 | 0.00612 |
| u6,u3 | 0.00110 |
| u6,u4 | -0.48735 |
| u6,u5 | 0.39385 |
| u6,u6 | 0.03602 |

The constant *s* assumes a value of 1 in this case, as we are simulating a monotonically increase in function value over time.

The syntax file for this model is shown in the syntax file **female_height_dlog.prl**. The variable Case is used as level-2 identifier (ID2).

```
OPTIONS METHOD = ML CONVERGE = 0.000010 MAXITER = 30 QUADPTS = 8;
TITLE = Double Logistic Curve fitted to male female height measurements ;
SY=fem_height.lsf;
ID1 = Occasio;
ID2 = Case;
RESPONSE = Height;
FIXED = Time;
MODEL = Logistic + Logistic;
```

The Model statement, specifying the fitting of a double logistic model, corresponds to the setting shown below on the **Select Model** dialog box accessed via the **Multilevel**, **Nonlinear Regression** option from the main menu bar.




The data summary at the top of the output file shows the number of measurements for each individual.

```
o--------------o
 | DATA SUMMARY |
 o--------------o
 NUMBER OF LEVEL 2 UNITS :      130
 NUMBER OF LEVEL 1 UNITS :     3988
```

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| N2 | : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| N1 | : | 30 | 32 | 32 | 34 | 33 | 29 | 27 | 28 |
| N2 | : | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| N1 | : | 29 | 31 | 35 | 35 | 30 | 29 | 30 | 30 |
| N2 | : | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| N1 | : | 29 | 32 | 31 | 30 | 30 | 31 | 25 | 27 |
| N2 | : | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| N1 | : | 32 | 32 | 35 | 30 | 36 | 33 | 34 | 32 |
| N2 | : | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| N1 | : | 32 | 33 | 34 | 28 | 29 | 28 | 33 | 29 |
| N2 | : | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
| N1 | : | 33 | 33 | 30 | 30 | 31 | 26 | 32 | 30 |
| N2 | : | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 |
| N1 | : | 30 | 37 | 29 | 29 | 32 | 29 | 32 | 29 |
| N2 | : | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 |
| N1 | : | 31 | 30 | 30 | 31 | 29 | 26 | 29 | 30 |
| N2 | : | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 |
| N1 | : | 30 | 35 | 30 | 26 | 33 | 28 | 33 | 30 |
| N2 | : | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| N1 | : | 27 | 27 | 31 | 29 | 32 | 31 | 28 | 31 |
| N2 | : | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 |
| N1 | : | 28 | 38 | 30 | 33 | 31 | 36 | 26 | 30 |
| N2 | : | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 |
| N1 | : | 28 | 35 | 33 | 33 | 32 | 34 | 29 | 33 |
| N2 | : | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 |
| N1 | : | 34 | 32 | 35 | 29 | 26 | 28 | 28 | 29 |
| N2 | : | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 |
| N1 | : | 31 | 27 | 27 | 34 | 33 | 30 | 33 | 33 |
| N2 | : | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
| N1 | : | 31 | 33 | 31 | 28 | 29 | 34 | 26 | 30 |
| N2 | : | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 |
| N1 | : | 29 | 28 | 29 | 32 | 32 | 29 | 29 | 29 |
| N2 | : | 129 | 130 | | | | | | |
| N1 | : | 32 | 31 | | | | | | |

The ML solution is as follows.

| Coefficients | Beta | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| b1 | 144.40476 | 0.46107 | 313.19687 | 0.00000 |
| b2 | 0.14259 | 0.00597 | 23.87648 | 0.00000 |
| b3 | 0.26813 | 0.00203 | 131.82791 | 0.00000 |
| c1 | 23.73397 | 0.31387 | 75.61814 | 0.00000 |
| c2 | 11.94765 | 0.17745 | 67.32993 | 0.00000 |
| c3 | 1.07374 | 0.01567 | 68.50994 | 0.00000 |

| Variance estimate | Level 1 | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| Sigma**2 | 0.81292 | 0.01299 | 62.57296 | 0.00000 |

| Covariances | Level 2 | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| u1,u1 | 38.97579 | 4.67259 | 8.34136 | 0.00000 |
| u2,u1 | 0.41023 | 0.05336 | 7.68757 | 0.00000 |
| u2,u2 | 0.00770 | 0.00081 | 9.55870 | 0.00000 |
| u3,u1 | -0.02083 | 0.01525 | -1.36594 | 0.17196 |
| u3,u2 | 0.00031 | 0.00019 | 1.60376 | 0.10877 |
| u3,u3 | 0.00062 | 0.00009 | 6.90328 | 0.00000 |
| u4,u1 | -11.92541 | 2.61176 | -4.56605 | 0.00000 |
| u4,u2 | -0.12552 | 0.03236 | -3.87904 | 0.00010 |
| u4,u3 | 0.00637 | 0.01032 | 0.61779 | 0.53671 |
| u4,u4 | 13.07078 | 2.02315 | 6.46060 | 0.00000 |
| u5,u1 | 3.78622 | 1.36343 | 2.77698 | 0.00549 |
| u5,u2 | 0.06869 | 0.01826 | 3.76096 | 0.00017 |
| u5,u3 | -0.00400 | 0.00575 | -0.69525 | 0.48690 |
| u5,u4 | -4.99289 | 1.01919 | -4.89889 | 0.00000 |
| u5,u5 | 4.26683 | 0.64978 | 6.56659 | 0.00000 |
| u6,u1 | 0.22012 | 0.11774 | 1.86948 | 0.06156 |
| u6,u2 | 0.00508 | 0.00159 | 3.20148 | 0.00137 |
| u6,u3 | 0.00076 | 0.00049 | 1.55961 | 0.11885 |
| u6,u4 | -0.40599 | 0.08906 | -4.55861 | 0.00001 |
| u6,u5 | 0.34800 | 0.05602 | 6.21217 | 0.00000 |
| u6,u6 | 0.03323 | 0.00506 | 6.56378 | 0.00000 |

Note: ML estimates of individual coefficients written to file THETAI.EST

The beta coefficients reported correspond closely to the values used in simulation.

The average expected height can be estimated using the formula

$$\text{Predicted}(y) = 144.40476 / (1 + \exp(0.14259 - 0.26813 * time) + 23.73397 / (1 + \exp(11.94765 - 1.07374 * time)$$

The program also writes the estimated parameters for each level-2 unit to an external text file named **thetai.est**. The contents of this file for the first 10 individuals are as follows:

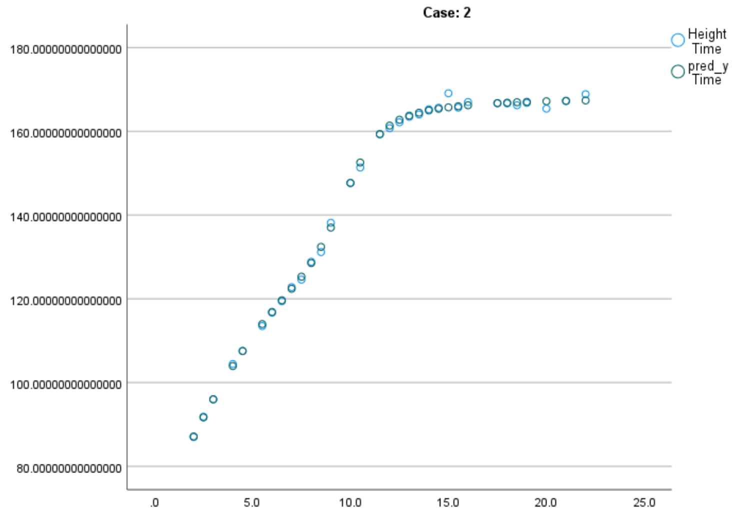
```
145.697    0.106036       0.252350    24.8169    10.3294    0.936850
137.712    0.351840E-01   0.289583    29.9058    11.2767    1.16088
```

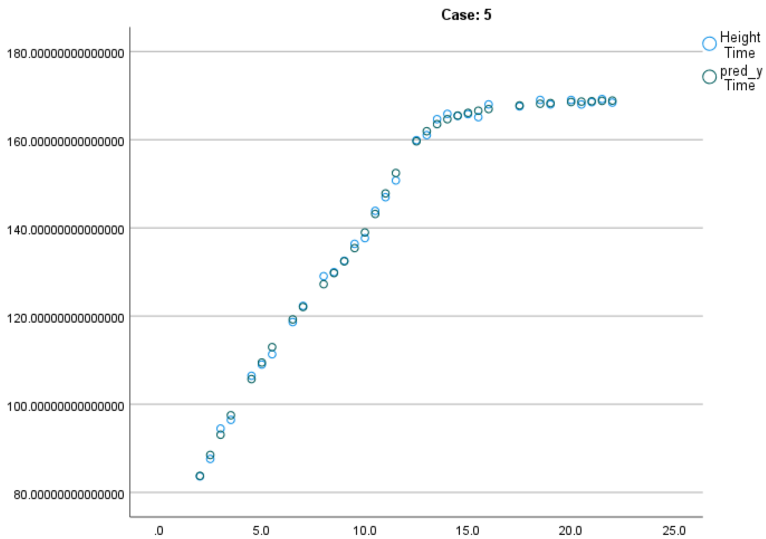
| | | | | | |
|---|---|---|---|---|---|
| 151.878 | 0.245825 | 0.270150 | 22.6760 | 13.9883 | 1.20602 |
| 147.849 | 0.185450 | 0.286616 | 24.8404 | 11.6040 | 1.15765 |
| 144.589 | 0.223534 | 0.271927 | 24.7484 | 13.0989 | 1.18217 |
| 144.996 | 0.149754 | 0.240066 | 19.3311 | 14.7623 | 1.21151 |
| 142.750 | 0.111284 | 0.270447 | 25.9821 | 12.6334 | 1.17554 |
| 145.235 | 0.150234 | 0.284730 | 25.9831 | 11.8968 | 1.12066 |
| 144.719 | 0.112014 | 0.257510 | 24.0276 | 11.7931 | 1.00305 |
| 128.401 | -0.122994E-01 | 0.259993 | 23.5280 | 10.2627 | 0.967023 |

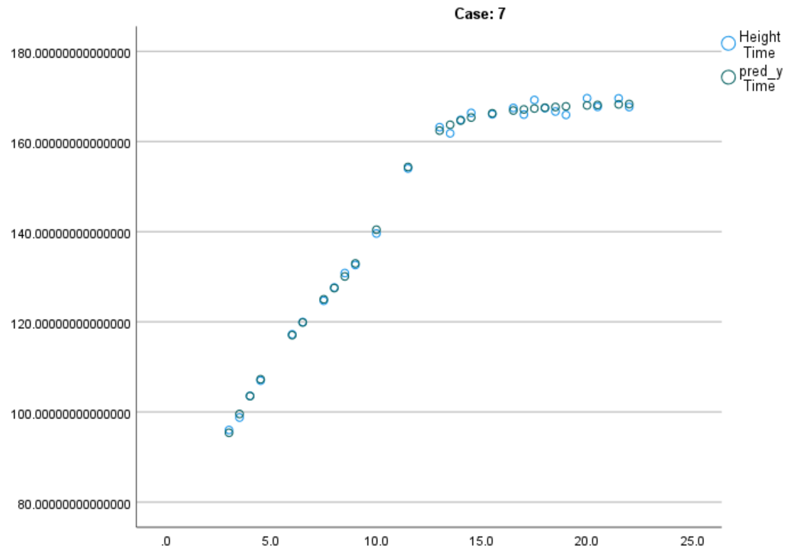
Using these results, the predicted height of, for example, the third female can be expressed as:

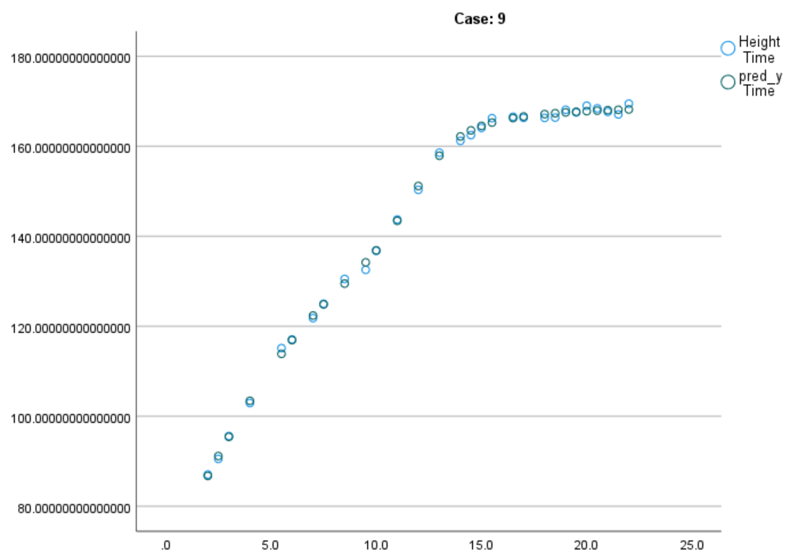
$$\text{Predicted}(y) = 151.878 / (1 + \exp(0.245825 - 0.270150 * time) + 22.6760 / (1 + \exp(13.9883 - 1.20602 * time)$$

When the observed and predicted heights for a few selected individuals are plotted, we see that the fitted curve describes the data well, even at the end of the period where there is more variation in the observed height over time.






## 3. Model misspecification

In this case, we know the model under which these data were simulated. However, if we did not, we might have opted to specify a single logistic model to describe these data. Looking at the original and estimated parameters in the previous section, we note that the parameter for `c1` is considerably smaller than for `b1`. It may thus be worth looking into whether one logistic component would do a reasonable job describing these data.

We modify the Model statement to

```
MODEL = Logistic;
```

For this model, the final results are

| Coefficients | Beta | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| b1 | 175.04940 | 0.38201 | 458.23536 | 0.00000 |
| b2 | 0.44204 | 0.00420 | 105.28139 | 0.00000 |
| b3 | 0.19440 | 0.00141 | 138.26832 | 0.00000 |

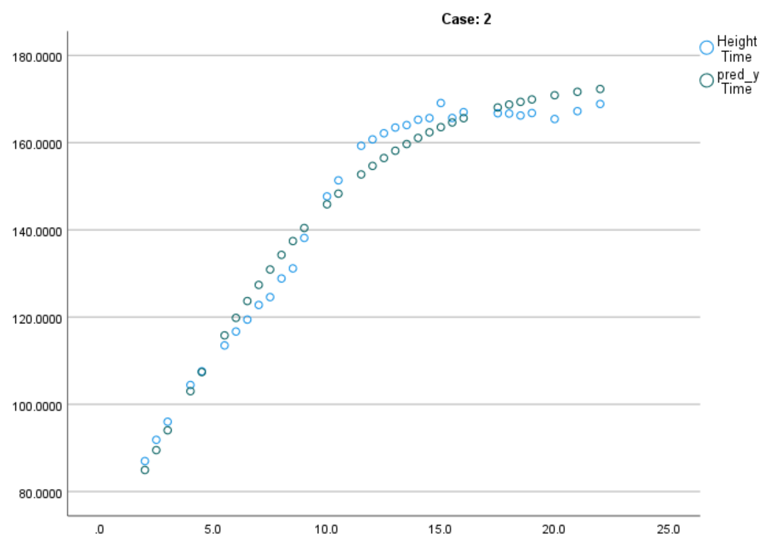
| Variance estimate | Level 1 | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| Sigma**2 | 8.74898 | 0.13928 | 62.81444 | 0.00000 |

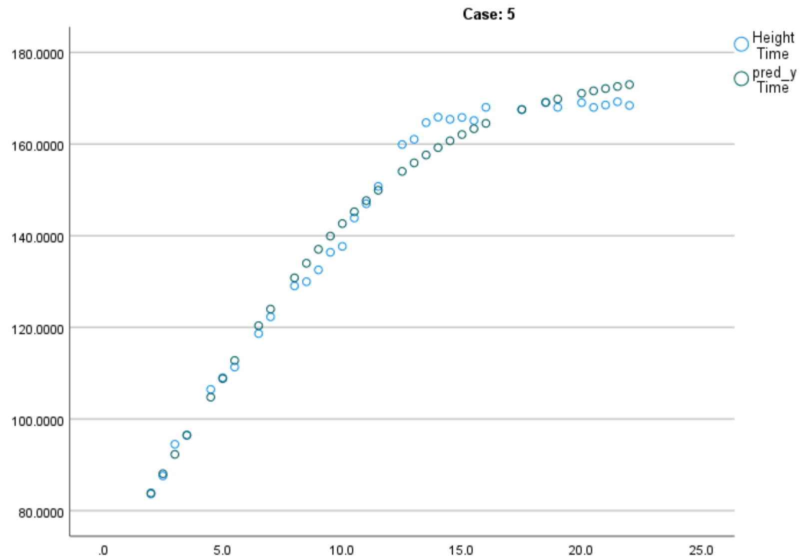
| Covariances | Level 2 | Std.Err. | Z-value | P > \|z\| |
|---|---|---|---|---|
| u1,u1 | 34.18790 | 3.31490 | 10.31341 | 0.00000 |
| u2,u1 | 0.16624 | 0.02724 | 6.10382 | 0.00000 |
| u2,u2 | 0.00260 | 0.00040 | 6.51151 | 0.00000 |
| u3,u1 | -0.03681 | 0.00920 | -4.00065 | 0.00006 |
| u3,u2 | 0.00003 | 0.00010 | 0.29646 | 0.76688 |
| u3,u3 | 0.00040 | 0.00004 | 8.94734 | 0.00000 |

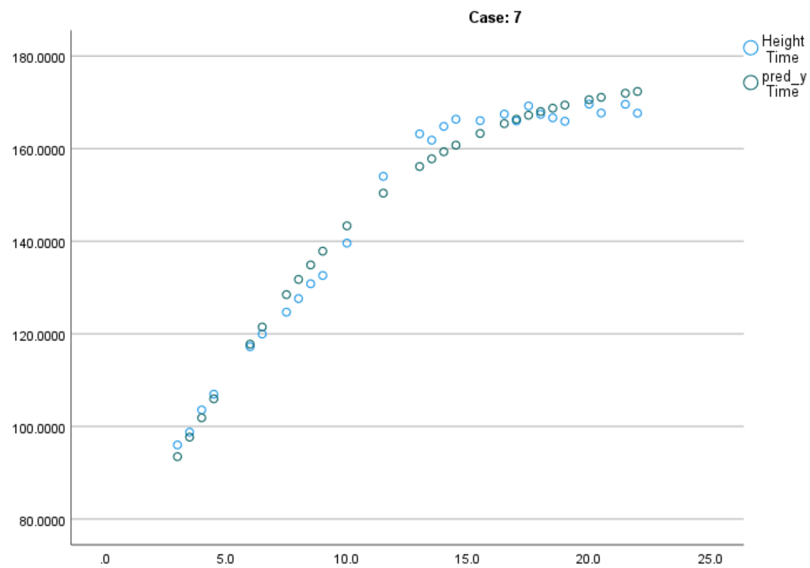
Under this model, the average predicted height can be calculated as

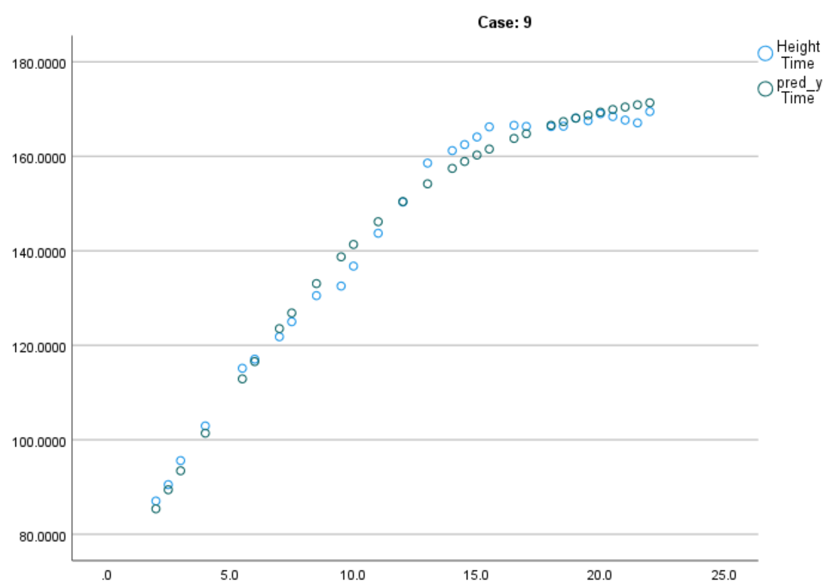
$$\text{Predicted}(y) = 175.0494 / (1 + \exp(0.44204 - 0.1944 * time)$$

If we again use the information in the additional output file **thetai.est**, we can plot the observed and predicted height for the same individuals previously considered. When we consider the scatterplots for these individuals, we see that although the basic shape of the fitted curve closely mimics that of the "observed" data, it overestimates the first inflection in the data, underestimates the second, and in general produce a higher than observed final weight. We thus conclude that while it may seem as if this model describes the data reasonably well, the second logistic component is really needed in this case where the data has two inflection points.




Case: 5
Height Time
pred_y Time
180.0000
160.0000
140.0000
120.0000
100.0000
80.0000
.0
5.0
10.0
15.0
20.0
25.0


Case: 7
Height Time
pred_y Time
180.0000
160.0000
140.0000
120.0000
100.0000
80.0000
.0
5.0
10.0
15.0
20.0
25.0


Case: 9
Height Time
pred_y Time
180.0000
160.0000
140.0000
120.0000
100.0000
80.0000
.0
5.0
10.0
15.0
20.0
25.0