Two Polynomial Regression Examples

Before beginning, run the handout installation procedure and put the file regress_display.lsp into your directory. Then, before doing any work, enter xlispstat and (you must do this each time you begin xlispstat) type

```
>(load "regress_display.lsp")
```

1. **Example 1**

   In this example, we want to investigate the effect of the number of dispensers on self-service coffee sales at a caafeteria chain. The data are:

   ```
   > (def dispensers (list 0 0 1 1 2 2 3 3 4 4 5 5 6 6))
   DISPENSERS
   > (def sales (list 508.1 498.4 568.2 577.3 651.7 657 713.4
    697.5 755.3  758.9 787.6  792.1 841.4 831.8))
   SALES
   ```

   First, we do a linear fit. Note the use of ":display" to produce nicer output.

   ```
   > (def fit1 (regression-model dispensers sales :display))

   Least Squares Estimates:


                            Coefficient    S.E.(Coef)     T-Value     P-value
   Constant                   523.800      (8.47407)       61.8       (0.0000)
   Variable 0                 54.8929      (2.35029)       23.4       (0.0000)

   R Squared:                0.978475
   Sigma hat:                 17.5879
   Number of cases:                14
   Degrees of freedom:             12
   ```

   The fit looks good, but is it a good model? A plot of the residuals versus dispensers shows a bowed shape. This is a good indication that a higher order term is needed.

   ```
   > (def resids1 (plot-points dispensers (send fit1 :residuals)))
   ```

Recall that this command plots the numbers returned by the "send" command against the variable "dispensers".

Next we fit a second order polynomial:

$$y = a + bx + cx^2,$$

where x is the number of dispensers and y is the amount of coffee sales.

```
> (def fit2 (regression-model (list dispensers (^ dispensers 2))
sales :display))

Least Squares Estimates:
                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   502.556      (4.85003)       104.       (0.0000)
Variable 0                 80.3857      (3.78605)       21.2       (0.0000)
Variable 1                -4.24881      (0.606254)      -7.01      (0.0000)

R Squared:                0.996061
Sigma hat:                 7.85795
Number of cases:                14
Degrees of freedom:             11
```

And, to push things further, a third order and a fourth order:

```
> (def fit3 (regression-model (list dispensers
(^ dispensers 2) (^ dispensers 3)) sales :display))

Least Squares Estimates:
                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   500.306      (5.35861)       93.4       (0.0000)
Variable 0                 87.8857      (8.46305)       10.4       (0.0000)
Variable 1                -7.62381      (3.45897)       -2.20      (0.0521)
Variable 2                0.375000      (0.378371)      .991       (0.3450)

R Squared:                0.996414
Sigma hat:                 7.86429
Number of cases:                14
Degrees of freedom:             10

> (def fit4 (regression-model (list dispensers (^ dispensers 2)
 (^ dispensers 3) (^ dispensers 4)) sales :display))

Least Squares Estimates:
                         Coefficient    S.E.(Coef)     T-Value     P-value
```

```
Constant                    502.751      (4.76414)       106.       (0.0000)
Variable 0                  63.8451      (13.5355)       4.72       (0.0011)
Variable 1                  13.4966      (10.4495)       1.29       (0.2287)
Variable 2                 -5.32955      (2.72457)      -1.96       (0.0822)
Variable 3                 0.475379      (0.225413)      2.11       (0.0642)

R Squared:                 0.997600
Sigma hat:                  6.78169
Number of cases:                 14
Degrees of freedom:               9

FIT4
(def x2 (^ dispensers 2))
X2
> (def x3 (^ dispensers 3))
X3
> (def x4 (^ dispensers 4))
X4
> (def x5 (^ dispensers 5))
X5
> (def fit5 (regression-model (list dispensers x2 x3 x4 x5) sales :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                    503.004      (5.01674)       100.       (0.0000)
Variable 0                  53.6431      (26.1964)       2.05       (0.0748)
Variable 1                  28.1403      (33.4657)       .841       (0.4248)
Variable 2                 -12.2816      (15.2837)      -.804       (0.4449)
Variable 3                  1.80663      (2.88492)       .626       (0.5486)
Variable 4             -8.875000E-2  (0.191684)      -.463       (0.6557)

R Squared:                 0.997662
Sigma hat:                  7.09859
Number of cases:                 14
Degrees of freedom:               8

FIT5
```

The step-wise routine "crashes" out at the fifth order, but notice that there is little improvement in R^2 or the residual plots after the second order. So we'll stop with a quadratic model:

$$y = 502.556 + 80.3857x - 4.24881x^2$$

2. **Example 2**

   Population of England and Wales since 1811. “year” is the number of years since 1800, and “pop” is the number of tens of millions of people.

```
> (def year (list 11 21 31 41 51 61 71 81 91 101 111 121 131))
YEAR
> (def pop (list 10.16 12 13.9 15.91 17.93 20.07 22.71 25.97 29  32.53 36.07 37.89 39.9
POP
> (def fit1 (regression-model year pop :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   5.63952      (0.641359)      8.79       (0.0000)
Variable 0                0.260863      (7.991426E-3)   32.6       (0.0000)

R Squared:                0.989782
Sigma hat:                 1.07810
Number of cases:                13
Degrees of freedom:             11
> (def fit2 (regression-model (list year (^ year 2)) pop :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   7.87996      (0.727238)      10.8       (0.0000)
Variable 0                0.173485      (2.351270E-2)   7.38       (0.0000)
Variable 1                 6.153347E-4  (1.612427E-4)   3.82       (0.0034)

R Squared:                0.995840
Sigma hat:                0.721460
Number of cases:                13
Degrees of freedom:             10

> (def fit2 (regression-model (list year (^ year 2)) pop :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   7.87996      (0.727238)      10.8       (0.0000)
```

```
Variable 0                0.173485      (2.351270E-2)   7.38       (0.0000)
Variable 1                 6.153347E-4  (1.612427E-4)   3.82       (0.0034)

R Squared:                0.995840
Sigma hat:                0.721460
Number of cases:                13
Degrees of freedom:             10

FIT2
> (def fit3 (regression-model (list year (^ year 2) (^ year 3)) pop :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   9.95006      (0.848257)      11.7       (0.0000)
Variable 0                 2.864428E-2  (4.938105E-2)   .580       (0.5761)
Variable 1                 3.059373E-3  (7.901596E-4)   3.87       (0.0038)
Variable 2                -1.147436E-5  (3.668320E-6)   -3.13      (0.0122)

R Squared:                0.998007
Sigma hat:                0.526401
Number of cases:                13
Degrees of freedom:              9

> (def fit4 (regression-model (list year (^ year 2) (^ year 3)
(^ year 4)) pop :display))

Least Squares Estimates:


                         Coefficient    S.E.(Coef)     T-Value     P-value
Constant                   7.04255      (0.604200)      11.7       (0.0000)
Variable 0                0.327149      (5.326007E-2)   6.14       (0.0003)
Variable 1                -5.510774E-3  (1.437317E-3)   -3.83      (0.0050)
Variable 2                 7.962430E-5  (1.490667E-5)   5.34       (0.0007)
Variable 3                -3.207699E-7  (5.217500E-8)   -6.15      (0.0003)

R Squared:                0.999652
Sigma hat:                0.233355
Number of cases:                13
Degrees of freedom:              8

FIT4
```

```
> (def fit5 (regression-model (list year (^ year 2) (^ year 3)
 (^ year 4) (^ year 5)) pop :display))

Least Squares Estimates:


                        Coefficient    S.E.(Coef)     T-Value     P-value
Constant                  7.45796      (1.06271)       7.02       (0.0002)
Variable 0               0.271376      (0.127355)      2.13       (0.0706)
Variable 1              -3.186119E-3  (5.001396E-3)   -.637       (0.5444)
Variable 2               3.889908E-5  (8.498125E-5)   .458        (0.6610)
Variable 3              -5.861357E-9  (6.481736E-7)   -9.043E-3   (0.9930)
Variable 4             -8.870664E-10  (1.819291E-9)   -.488       (0.6407)

R Squared:               0.999663
Sigma hat:               0.245336
Number of cases:               13
Degrees of freedom:             7

FIT5
```

We crashed out on the fifth order fit, so we'll settle for fourth order. The residual plot looks, if not perfect, at least better than the residuals for the lower orders.