

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 1

## Logistic Regression using OLS1A in Excel 2013

by  
**Milo Schield**

*Member: International Statistical Institute  
US Rep: International Statistical Literacy Project  
Director, W. M. Keck Statistical Literacy Project*

*Slides, output and data at: [www.StatLit.org/pdf/2015-Schield-Logistic-OLS1A-Excel2013-Slides.pdf](http://www.StatLit.org/pdf/2015-Schield-Logistic-OLS1A-Excel2013-Slides.pdf)  
[www.StatLit.org/pdf/2015-Schield-Logistic-OLS1A-Excel2013-Demo.pdf](http://www.StatLit.org/pdf/2015-Schield-Logistic-OLS1A-Excel2013-Demo.pdf)  
[www.StatLit.org/Excel/2015-Schield-Logistic-OLS1A-Excel2013-Data.xlsx](http://www.StatLit.org/Excel/2015-Schield-Logistic-OLS1A-Excel2013-Data.xlsx)*

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 2

## Background & Goals

Modelling a binary outcome (loan vs. no-loan) requires logistic regression to avoid meaningless predictions.

Doing an exact logistic regression in Excel requires Solver and involves many steps. For details, see [www.statlit.org/pdf/Excel2013-Schield-Logistic-MLE1A-Slides.pdf](http://www.statlit.org/pdf/Excel2013-Schield-Logistic-MLE1A-Slides.pdf)

This presentation uses an approximation: OLS1. By slightly adjusting the binary outcomes, one can use OLS regression to solve for a good logistic model.

**Assignment: Create the logistic model (slide 9) and the logistic graph (slide 12).**

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 3

## This demo uses Height (col A) to predict Gender (col B)

Column B: 0=Female, 1 = Male (circled)

Data in rows 6 to 98

|    | A      | B    | C     |
|----|--------|------|-------|
| 6  | Height | Male | Male1 |
| 7  | 61     | 0    |       |
| 8  | 61.75  | 0    |       |
| 9  | 62     | 0    |       |
| 10 | 62     | 0    |       |
| 11 | 62     | 0    |       |
| 12 | 62     | 0    |       |
| 13 | 62.75  | 0    |       |
| 14 | 63     | 0    |       |
| 15 | 63     | 0    |       |
| 16 | 63     | 0    |       |
| 17 | 63     | 0    |       |
| 24 | 65.5   | 0    |       |
| 25 | 66     | 1    |       |
| 26 | 66     | 0    |       |
| 27 | 66     | 0    |       |
| 28 | 66     | 0    |       |
| 29 | 66     | 1    |       |
| 30 | 66     | 1    |       |
| 31 | 66     | 1    |       |
| 32 | 66     | 0    |       |
| 33 | 67     | 1    |       |
| 34 | 67     | 1    |       |

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 4

## 1) Nudge Binary Male to Eliminate Zero and One

| A  | B    | C     | D    | E        |
|--|------|-------|------|----------|
| Predict chance of being male given height. f |      |       |      |          |
| C7 =IF(B7=0, 0.001, 0.999)                   |      |       |      |          |
| Height                                       | Male | Male1 | Odds | LN(Odds) |
| 61   | 0    | 0.001 |      |          |

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 5

## 2) Enter formula for Odds in D7; LN[Odds(p)] in E7

Predict chance of being male given height. Regress using

C7 =IF(B7=0, 0.001, 0.999)      E7 =LN(D7)

D7 =C7/(1-C7)

| A      | B    | C     | D     | E        | F     | G |
|--------|------|-------|-------|----------|-------|---|
| Height | Male | Male1 | Odds  | LN(Odds) | yPred |   |
| 61     | 0    | 0.001 | 0.001 | -6.91    |       |   |
| 61.75  | 0    |       |       |          |       |   |
| 62     | 0    |       |       |          |       |   |

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 6

## 3) Select C7:E7 Drag to bottom of data: Row 98

Predict chance of being male given height. Regress using

C7 =IF(B7=0, 0.001, 0.999)      E7 =LN(D7)

D7 =C7/(1-C7)

| A      | B    | C     | D     | E        | F     | G |
|--------|------|-------|-------|----------|-------|---|
| Height | Male | Male1 | Odds  | LN(Odds) | yPred |   |
| 61     | 0    | 0.001 | 0.001 | -6.91    |       |   |
| 61.75  | 0    |       |       |          |       |   |
| 62     | 0    |       |       |          |       |   |

**A) From Data Bar, Select Data Analysis; Regression**

The screenshot shows the Excel 2013 ribbon with the 'DATA' tab selected. In the 'Data Tools' group, the 'Data Analysis' icon is circled. Below the ribbon, the 'Data Analysis' task pane is open, and the 'Regression' option is selected and circled.

**B) Select Data, Labels, Output Range. Press OK**

The screenshot shows the 'Regression' dialog box. The 'Input Y Range' is set to E6:E98 and the 'Input X Range' is set to A6:A98. The 'Labels' checkbox is checked. The 'Output Range' is set to H17. A yellow box contains the text: 'If typing ranges gives errors, select ranges manually.'

**C) Logistic Regression: Results Using OLS1A**

The screenshot shows the 'SUMMARY OUTPUT' table for the regression analysis. The 'Adjusted R Square' is 0.50. A yellow box highlights the 'Adjusted R Square' value with the text: 'Check to see that you get the same results in the boxes. Formatting is optional.'

| ANOVA      |    |          |          |       |                |
|------------|----|----------|----------|-------|----------------|
|            | df | SS       | MS       | F     | Significance F |
| Regression | 1  | 2,111.07 | 2,111.07 | 93.75 | 0.00           |
| Residual   | 90 | 2,026.67 | 22.52    |       |                |
| Total      | 91 | 4,137.74 |          |       |                |

**D) Generate F7: Pull F7 down to F98**

The screenshot shows the formula bar with the formula  $F7 = 1/(1+EXP(-I$33-I$34*A7))$  entered. Below, the spreadsheet shows the formula applied to cell F7, resulting in a value of 0.000. A yellow box highlights the formula bar with the text: 'male given height. Regress using a logistic 1, 0.999) E7 =LN(D7)'. The spreadsheet also shows 'Odds' as 0.001, 'LN(Odds)' as -6.91, and 'yPred' as 0.000.

**E) Insert XY Plot. Add Two Series. Male vs Height | yPred vs Height A7:A98, B7:B98 | A7:A98, F7:F98**

The screenshot shows two 'Edit Series' dialog boxes. The first series is named 'Male|Ht' and uses the range 'Male|Ht'!\$B\$6:\$B\$98 for X values and 'Male|Ht'!\$B\$7:\$B\$98 for Y values. The second series is named 'Male|Ht' and uses the range 'Male|Ht'!\$F\$6:\$F\$98 for X values and 'Male|Ht'!\$F\$7:\$F\$98 for Y values.

**E) Insert Titles & Textboxes. Format yPred with Solid Line**

The screenshot shows an XY plot titled 'Chance of Male given Height'. The x-axis is 'Height (inches)' ranging from 61 to 75. The y-axis is 'OLS1A' ranging from 0.0 to 1.0. A solid line represents the logistic curve. Text boxes provide the formula  $P(\text{Male}) = 1/(1+Exp(-Z))$ , the equation  $Z = -88.80 + 1.316 * \text{Height}$ , and the note 'Average Height: 68.72" 62% are men'.

# **Logistic Regression using OLS1A in Excel 2013**

---

by  
**Milo Schield**

*Member: International Statistical Institute*

*US Rep: International Statistical Literacy Project*

*Director, W. M. Keck Statistical Literacy Project*

*Slides, output and data at: [www.StatLit.org/](http://www.StatLit.org/)*

*[pdf/2015-Schild-Logistic-OLS1A-Excel2013-Slides.pdf](#)*

*[pdf/2015-Schild-Logistic-OLS1A-Excel2013-Demo.pdf](#)*

*[Excel/2015-Schild-Logistic-OLS1A-Excel2013-Data.xlsx](#)*

# Background & Goals

---

Modelling a binary outcome (loan vs. no-loan) requires logistic regression to avoid meaningless predictions.

Doing an exact logistic regression in Excel requires Solver and involves many steps. For details, see [www.statlit.org/pdf/Excel2013-Schild-Logistic-MLE1A-Slides.pdf](http://www.statlit.org/pdf/Excel2013-Schild-Logistic-MLE1A-Slides.pdf)

This presentation uses an approximation: OLS1. By slightly adjusting the binary outcomes, one can use OLS regression to solve for a good logistic model.

**Assignment: Create the logistic model (slide 9) and the logistic graph (slide 12).**

# This demo uses Height (col A) to predict Gender (col B)

Column B: 0=Female, 1 = Male (circled)

|    | A      | B    | C     |
|----|--------|------|-------|
| 6  | Height | Male | Male1 |
| 7  | 61     | 0    |       |
| 8  | 61.75  | 0    |       |
| 9  | 62     | 0    |       |
| 10 | 62     | 0    |       |
| 11 | 62     | 0    |       |
| 12 | 62     | 0    |       |
| 13 | 62.75  | 0    |       |
| 14 | 63     | 0    |       |
| 15 | 63     | 0    |       |
| 16 | 63     | 0    |       |
| 17 | 63     | 0    |       |

Data  
in  
rows  
6 to 98

|    | A    | B | C |
|----|------|---|---|
| 24 | 65.5 | 0 |   |
| 25 | 66   | 1 |   |
| 26 | 66   | 0 |   |
| 27 | 66   | 0 |   |
| 28 | 66   | 0 |   |
| 29 | 66   | 1 |   |
| 30 | 66   | 1 |   |
| 31 | 66   | 1 |   |
| 32 | 66   | 0 |   |
| 33 | 67   | 1 |   |
| 34 | 67   | 1 |   |

# 1) Nudge Binary Male to Eliminate Zero and One

| A  | B                       | C     | D    | E        |
|--|-------------------------|-------|------|----------|
| Predict chance of being male given height. f |                         |       |      |          |
| C7   | =IF(B7=0, 0.001, 0.999) |       |      |          |
| Height                                       | Male                    | Male1 | Odds | LN(Odds) |
| 61   | 0                       | 0.001 |      |          |

## 2) Enter formula for Odds in D7; LN[Odds(p)] in E7

| A  | B                       | C     | D     | E        | F       | G |
|--|-------------------------|-------|-------|----------|---------|---|
| Predict chance of being male given height. Regress using |                         |       |       |          |         |   |
| C7   | =IF(B7=0, 0.001, 0.999) |       |       | E7       | =LN(D7) |   |
| D7   | =C7/(1-C7)              |       |       |          |         |   |
| Height   | Male                    | Male1 | Odds  | LN(Odds) | yPred   |   |
| 61   | 0                       | 0.001 | 0.001 | -6.91    |         | 6 |
| 61.75  | 0                       |       |       |          |         | 7 |
| 62   | 0                       |       |       |          |         | 8 |
|  |                         |       |       |          |         | 9 |

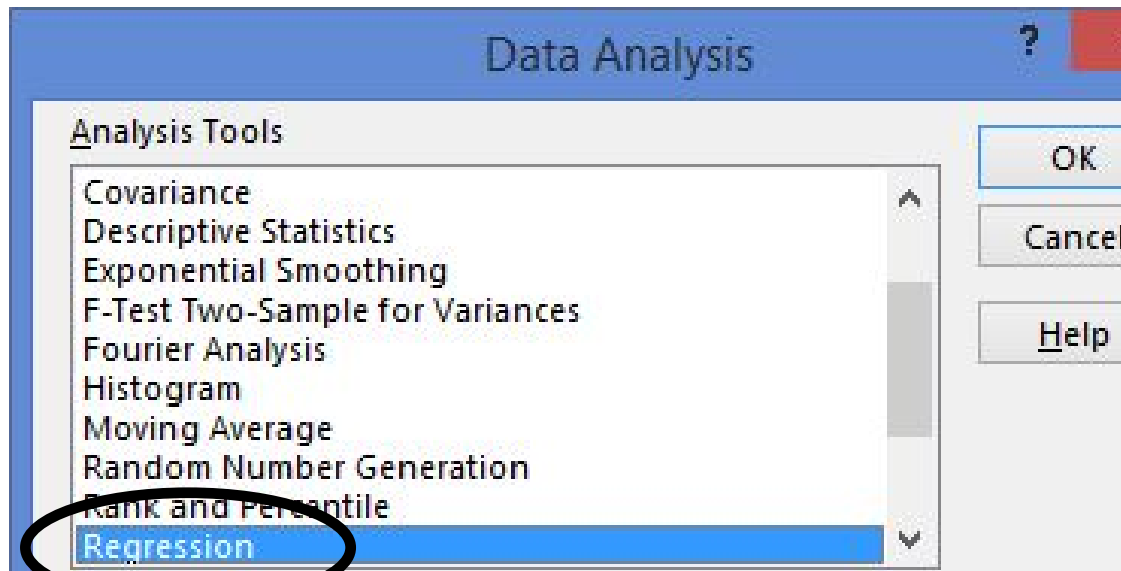
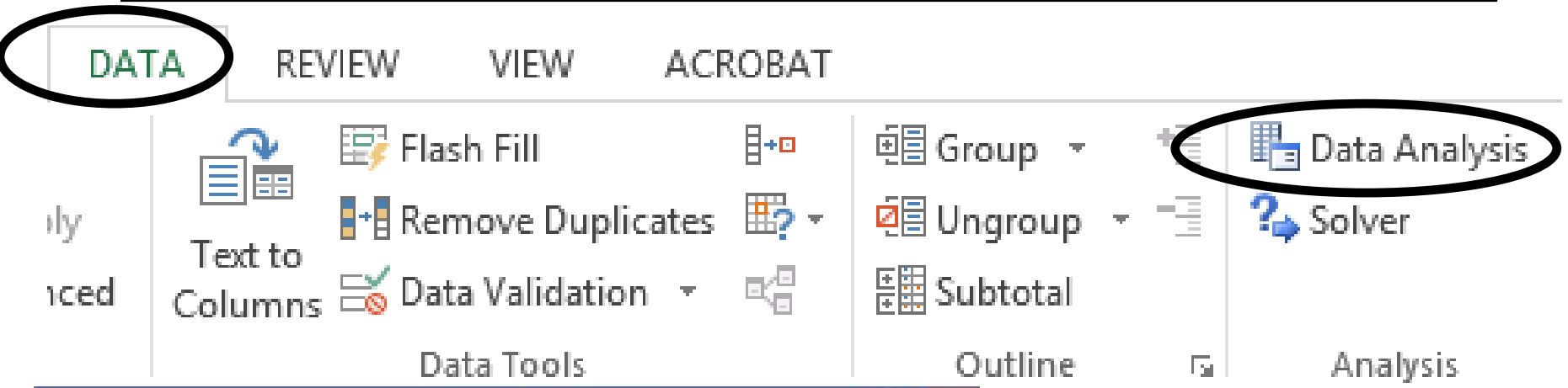
# 3) Select C7:E7

## Drag to bottom of data: Row 98

| A  | B                       | C     | D     | E        | F       | G |
|--|-------------------------|-------|-------|----------|---------|---|
| Predict chance of being male given height. Regress using |                         |       |       |          |         |   |
| C7   | =IF(B7=0, 0.001, 0.999) |       | E7    |          | =LN(D7) |   |
| D7   | =C7/(1-C7)              |       |       |          |         |   |
| Height   | Male                    | Male1 | Odds  | LN(Odds) | yPred   |   |
| 61   | 0                       | 0.001 | 0.001 | -6.91    |         | 6 |
| 61.75  | 0                       |       |       |          |         | 7 |
| 62   | 0                       |       |       |          |         | 8 |
|  |                         |       |       |          |         | 9 |



# A) From Data Bar, Select Data Analysis; Regression



# B) Select Data, Labels, Output Range. Press OK

Regression

Input

Input Y Range: E6:E98

Input X Range: A6:A98

Labels

Constant is Zero

Confidence Level: 95 %

Output options

Output Range: H17

OK

Cancel

Help

*If typing ranges gives errors, select ranges manually.*

# C) Logistic Regression: Results Using OLS1A

|    | H                            | I                   | J                     | K             | L              | M                     | N                |
|----|------------------------------|---------------------|-----------------------|---------------|----------------|-----------------------|------------------|
| 16 |                              |                     |                       |               |                |                       |                  |
| 17 | SUMMARY OUTPUT               |                     |                       |               |                |                       |                  |
| 18 |                              |                     |                       |               |                |                       |                  |
| 19 | <i>Regression Statistics</i> |                     |                       |               |                |                       |                  |
| 20 | Multiple R                   | 0.71                |                       |               |                |                       |                  |
| 21 | R Square                     | 0.51                |                       |               |                |                       |                  |
| 22 | Adjusted R Square            | 0.50                |                       |               |                |                       |                  |
| 23 | Standard Error               | 4.75                |                       |               |                |                       |                  |
| 24 | Observations                 | 92                  |                       |               |                |                       |                  |
| 25 |                              |                     |                       |               |                |                       |                  |
| 26 | ANOVA                        |                     |                       |               |                |                       |                  |
| 27 |                              | <i>df</i>           | <i>SS</i>             | <i>MS</i>     | <i>F</i>       | <i>Significance F</i> |                  |
| 28 | Regression                   | 1                   | 2,111.07              | 2,111.07      | 93.75          | 0.00                  |                  |
| 29 | Residual                     | 90                  | 2,026.67              | 22.52         |                |                       |                  |
| 30 | Total                        | 91                  | 4,137.74              |               |                |                       |                  |
| 31 |                              |                     |                       |               |                |                       |                  |
| 32 |                              | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>P-value</i> | <i>Lower 95%</i>      | <i>Upper 95%</i> |
| 33 | Intercept                    | -88.80              | 9.35                  | -9.49         | 0.00           | -107.38               | -70.21           |
| 34 | Height                       | 1.32                | 0.14                  | 9.68          | 0.00           | 1.05                  | 1.59             |

*Check to see that you get the same results in the boxes. Formatting is optional*

# D) Generate F7: Pull F7 down to F98

| D  | E          | F     | G |
|--|------------|-------|---|
| ; male given height. Regress using a logistic  |            |       |   |
| 1, 0.999)  | E7 =LN(D7) |       |   |
| <div style="border: 2px solid black; border-radius: 50%; padding: 10px; display: inline-block;">                     F7 =1/(1+EXP(-I\$33-I\$34*A7))                 </div> |            |       |   |
| Odds   | LN(Odds)   | yPred | 6 |
| 0.001  | -6.91      | 0.000 | 7 |

# **E) Insert XY Plot. Add Two Series.**

## **Male vs Height | yPred vs Height**

### **A7:A98, B7:B98 | A7:A98, F7:F98**

Edit Series

Series name:

= 'Male|Ht'!\$B\$6

Series X values:

= 'Male|Ht'!\$A\$7:\$A\$98

Series Y values:

= 'Male|Ht'!\$B\$7:\$B\$98

Edit Series

Series name:

= 'Male|Ht'!\$F\$6

Series X values:

= 'Male|Ht'!\$A\$7:\$A\$98

Series Y values:

= 'Male|Ht'!\$F\$7:\$F\$98

# E) Insert Titles & Textboxes. Format yPred with Solid Line

