

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 1

Logistic Model using OLS1: Gender vs. Height & Smoker

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/2017-Schield-Logistic-OLS1B-Excel2013-Slides.pdf
www.StatLit.org/pdf/2017-Schield-Logistic-OLS1B-Excel2013-Demo.pdf
www.StatLit.org/excel/2017-Schield-Logistic-OLS1B-Excel2013-Data.xlsx*

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 2

Background & Goals

Modelling a binary outcome (loan vs. no-loan) uses a logistic curve/model 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

This approach uses a nudge approximation: OLS1. By adjusting the binary outcomes and using a logistic model, OLS regression generates a fairly good fit.

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

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 3

Goal: Predict Gender using Height & Smoker

Column B: Smoker = 1; Column C: Male = 1; Non-smoker = 0; Female = 0

| | A | B | C | |
|----|--------|--------|------|------------|
| 5 | Height | Smokes | Male | |
| 7 | 61 | 0 | 0 | Non-smoker |
| 8 | 61.75 | 0 | 0 | smoker |
| 9 | 62 | 0 | 0 | |
| 10 | 62 | 0 | 0 | Female |
| 11 | 63 | 0 | 0 | |
| 12 | 63 | 0 | 0 | Male |
| 13 | 63 | 0 | 0 | |
| 14 | 64 | 0 | 0 | |
| 15 | 65 | 0 | 0 | |
| 16 | 65 | 0 | 0 | |
| 17 | 65 | 0 | 0 | |

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 4

1) Nudge Binary Male to eliminate Zero and One

| | A | B | C | D | E |
|---|--|--------|------|-------|---------|
| 1 | | | | | |
| 2 | Predict Gender using Height and Smoker | | | | |
| 3 | D7 =IF(C7=1,0.999,0.001) | | | | |
| 4 | | | | | |
| 5 | | | | | |
| 6 | Height | Smokes | Male | Male1 | Odds(M) |
| 7 | 61 | 0 | 0 | 0.001 | |

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 5

2) Enter formula for Odds in E7; LN[Odds(p)] in F7

| | A | B | C | D | E | F |
|---|--|--------|------|-------|---------|----------|
| 1 | | | | | | |
| 2 | Predict Gender using Height and Smoker | | | | | |
| 3 | D7 =IF(C7=1,0.999,0.001) F7 =Ln(E7) | | | | | |
| 4 | E7 =D7/(1-D7) | | | | | |
| 5 | | | | | | |
| 6 | Height | Smokes | Male | Male1 | Odds(M) | Ln(Odds) |
| 7 | 61 | 0 | 0 | 0.001 | 0.001 | -6.91 |

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 6

3) Select D7:F7 Drag to bottom of data: Row 98

| | A | B | C | D | E | F | G |
|---|--|--------|------|-------|---------|----------|----------|
| 1 | | | | | | | |
| 2 | Predict Gender using Height and Smoker | | | | | | |
| 3 | D7 =IF(C7=1,0.999,0.001) F7 =Ln(E7) | | | | | | |
| 4 | E7 =D7/(1-D7) | | | | | | |
| 5 | | | | | | | |
| 6 | Height | Smokes | Male | Male1 | Odds(M) | Ln(Odds) | yPredict |
| 7 | 61 | 0 | 0 | 0.001 | 0.001 | -6.91 | |
| 8 | 61.75 | 0 | 0 | | | | |
| 9 | 62 | 0 | 0 | | | | |

A) From Data Bar, Select Data Analysis; Regression

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

C) Logistic Model: Results Using OLS1

| ANOVA | | | | | |
|------------|----|----------|----------|--------|----------------|
| | df | SS | MS | F | Significance F |
| Regression | 2 | 2143.861 | 1071.930 | 47.847 | 0.000 |
| Residual | 89 | 1993.879 | 22.403 | | |
| Total | 91 | 4137.739 | | | |

D) Generate G7. Pull G7 down to G98

| Odds(M) | Ln(Odds) | yPredict |
|---------|----------|----------|
| 0.001 | -6.91 | 0.000 |
| 0.001 | -6.91 | |
| 0.001 | -6.91 | |

E) Insert Chart (XY Plot): Add Male, Non-smoker, Smoker

A7:A98 A7:A70 A71:A98
C7:C98 G7:G70 G71:G98

E) Add Title & Text boxes Format Smk/NS with solid lines

$P(\text{Male}|\text{Smoker}) = 1/(1+\text{Exp}(-Z))$
 $Z = -88.56 + 1.31 * \text{Height} + 1.3$
 $P(\text{Male}|\text{Sm}) = 50\% \text{ if Ht} = 66.6$

$P(\text{Male}|\text{Non-Smoker}) = 1/(1+\text{Exp}(-Z))$
 $Z = -88.56 + 1.31 * \text{Height}$
 $P(\text{Male}|\text{Non-Smk}) = 50\% \text{ if Ht} = 67.6$