

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 1

Logistic Regression using OLS1D 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-OLS1D-Excel2013-Slides.pdf
[pdf/2015-Schield-Logistic-OLS1D-Excel2013-Demo.pdf](http://www.StatLit.org/pdf/2015-Schield-Logistic-OLS1D-Excel2013-Demo.pdf)
[Excel/2015-Schield-Logistic-OLS1D-Excel2013-Data.xls](http://www.StatLit.org/Excel/2015-Schield-Logistic-OLS1D-Excel2013-Data.xls)

XL4D: V0H 2015 Schield Logistic Regression using OLS1D 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 This presentation uses an approximation. By “nudging” the binary outcomes, one can use ordinary least-squares regression to get a decent logistic model. **Assignment: Create the logistic model (slide 9) and the logistic graphs (slides 12 and 16).**

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 3

Use Height (A) & Weight (B) to predict Gender (C)

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

6	Height	Weight	Male	Male1	LN(Odds)	yPred
7	61	140	0			
8	61.75	108	0			
9	62	108	0			
10	62	110	0			
11	62	120	0			
12	62	131	0			
13	62.75	112	0			
14	63	95	0			
15	63	116	0			
16	63	118	0			
17	63	121	0			
18	64	102	0			
19	64	125	0			
20	65	115	0			
21	65	118	0			
22	65	122	0			
23	65	135	0			
27	66	130	0			
28	66	130	1			
29	66	130	0			
30	66	135	1			
31	66	135	1			
32	66	140	1			
33	67	115	0			
34	67	123	1			
35	67	125	0			
36	67	140	1			
37	67	145	1			
38	67	150	1			
39	67	150	0			

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 4

1a) Nudge Binary Male in D7 to Eliminate Zero and One

1	A	B	C	D	E
2	Predict chance of being male given height				
3	D7 =IF(C7=0, 0.001, 0.999)				
4					
5					
6	Height	Weight	Male	Male1	LN(Odds)
7	61	140	0	0.001	

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 5

1b) Generate Ln[Odds(Male1)] in E7

B	C	D	E	F	G
Predict chance of being male given height after controlling for weight					
D7 =IF(C7=0, 0.001, 0.999)			E7 =LN(D7/(1-D7))		
Height	Weight	Male	Male1	LN(Odds)	yPred
	140	0	0.001	-6.91	

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 6

1c) Select D7:E7 Pull down to bottom: Row 98

A	B	C	D	E	F	G
Predict chance of being male given height after controlling for weight						
D7 =IF(C7=0, 0.001, 0.999)			E7 =LN(D7/(1-D7))			
Height	Weight	Male	Male1	LN(Odds)	yPred	
61	140	0	0.001	-6.91		6
61.75	108	0				7
62	108	0				8
62	110	0				9
						10

2a) From Data Bar, Select Data Analysis; Regression

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

2b) Select Input & Output. Check Labels. Press OK

The screenshot shows the 'Regression' dialog box. The 'Input Y Range' is set to 'E6:E98', the 'Input X Range' is 'A6:B98', and the 'Labels' checkbox is checked. The 'Output Range' is set to 'H16'. The 'OK' button is highlighted.

2c) OLS1 Regression

The screenshot shows the 'SUMMARY OUTPUT' table for the OLS1 regression. The 'R Square' is 0.57 and the 'Adjusted R Square' is 0.56. A text box notes: 'Main source of error: No mention of Weight in H34. Double-check H34! To fix, redo X-range in slide 8.'

	H	I	J	K	L
16	SUMMARY OUTPUT				
17	Regression Statistics				
18	Multiple R	0.75			
19	R Square	0.57			
20	Adjusted R Square	0.56			
21	Standard Error	4.49			
22	Observation	92			
31	Coefficients				
32	Intercept	-66.37	11.00	-6.03	0.00
33	Height	0.7586	0.21	3.66	0.00
34	Weight	0.1095	0.03	3.43	0.00

3a) Generate F7. Check value. Select; pull down to row 98.

The screenshot shows the spreadsheet with the formula for F7: $F7 = 1 / (1 + \exp(-\$32 - \$33 * A7 - \$34 * B7))$. The formula bar shows the formula being entered.

Height	Weight	Male	Male1	LN(Odds)	yPred
61	140	0	0.001	-6.91	0.008
61.75	108	0	0.001	-6.91	
62	108	0	0.001	-6.91	
62	110	0	0.001	-6.91	

3b) Insert Chart (XY Plot): yPred vs. Height

The screenshot shows the 'Edit Series' dialog box for the chart. The series name is 'Male|Ht+Wt!\$F56', the X values are 'Male|Ht+Wt!\$A57:\$A598', and the Y values are 'Male|Ht+Wt!\$F57:\$F598'.

3c) Chart #1 Results Add Title and textboxes

The screenshot shows a scatter plot titled 'Chance of Male Given Height and Weight'. The plot shows the relationship between Height and yPred. The formula for the probability of being male is $P(\text{male}) = 1 / (1 + \exp(-Z))$, where $Z = -66.37 + 0.759 * \text{Height} + 0.110 * \text{Weight}$. A text box notes: 'P(male) = 0.5 if Z = 0. If Z = 0, 66.37 = 0.759 * Height + 0.11 * Weight'.

**4a) Enter formula in R3 & S3
Pull R3:S3 down to Row 31**

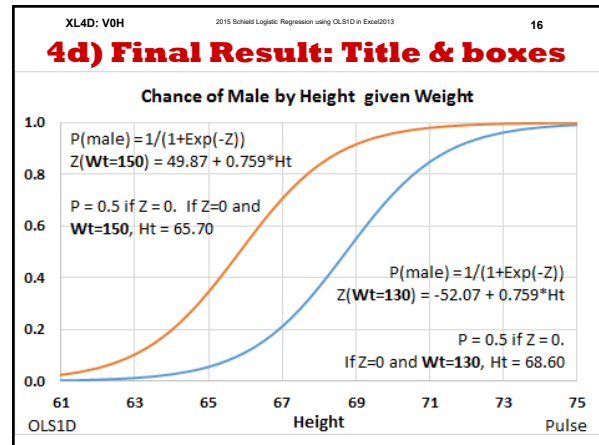
P	Q	R	S	T	U	V	W
Row	X	Y(X Wt=130)	Y(X Wt=150)	R3 = 1/(1+EXP(-I\$32-I\$33*Q3-I\$34*130))	S3 = 1/(1+EXP(-I\$32-I\$33*Q3-I\$34*150))		
3	61	0.003	0.025				
4	61.5						
5	62						
6	62.5						
7	63						
8	63.5						
9	64						
10	64.5						
11	65						
12	65.5						
13	66						
14	66.5						
15	67						

**4b) Insert XY Plot: Two Series
Y(X | Wt=130) Y(X | Wt=150)**

Name: R2 Name: S2
 X values: Q3:Q31 X values: Q3:Q31
 Y values: R3:R31 Y values: S3:S31

Series name: Series name:
 ='Male|Ht+Wt'!\$R\$2 ='Male|Ht+Wt'!\$S\$2
 Series X values: Series X values:
 ='Male|Ht+Wt'!\$Q\$3:\$Q\$31 ='Male|Ht+Wt'!\$Q\$3:\$Q\$31
 Series Y values: Series Y values:
 ='Male|Ht+Wt'!\$R\$3:\$R\$31 ='Male|Ht+Wt'!\$S\$3:\$S\$31

**4c) Format Data Series
Paint: No marker; Solid line**



- Conclusion for OLS1 Approach to Logistic Regression**
1. Plus: This OLS1 ‘nudge’ approach allows students to generate a decent solution quickly using Excel and answer relevant questions with quantitative answers.
 2. Plus: Students do not need to use different software so they can focus on interpreting the results, and it is more accurate than a linear OLS on binary data. .
 3. Minus: This Ordinary Least Squares (OLS) model using “nudged” binary outcomes gives less accurate estimates than the Maximum-Likelihood Estimation (MLE). If more accuracy is needed, find a statistician

**Appendix: Simplify Z;
Solve for Height at P=50%**

$$Z = -66.37 + 0.759 * \text{Ht} + 0.11 * \text{Wt}$$

If Wt=130, Z = -52.07 + 0.759*Ht		
Ht P=50%	68.60	=52.07/0.76
If Wt=150, Z = -49.87 + 0.759*Ht		
Ht P=50%	65.70	=49.87/0.76