V1A Excel2013 Model Logistic MLE 1Y1X Slides 1

Model Logistic Regression MLE 1Y1X in Excel 2013

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides and data at: www.StatLit.org/
pdf/Excel2013-Model-Logistic-MLE-1Y1X-Slides.pdf
xls/Excel2013-Model-Logistic-MLE-1Y1X-Data.xlsx

V1A Excel2013 Model Logistic MLE 1Y1X Slides 2

Background & Goals

Modelling a binary outcome (buy/look, payoff/default, go/nogo or male/female) requires logistic regression.

Doing logistic regression in Excel requires Solver. "Since its introduction in .. 1991, … Excel Solver has become the most widely distributed – and almost surely the most widely used – general-purpose optimization modeling system." www.utexas.edu/courses/lasdon/design3.htm

This presentation uses college student data: pulse.xls. This demo models gender (male) based on height.

Goals: Create graphs on slides 4 and 22.
Determine if slope is statistically significant.

V1A Excel2013 Model Logistic MLE 1Y1X Slides 3

This demo uses Height (col C) to predict Gender (col H)

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

Ave Heights:

M: 70.75" 62%

F: 65.3" 38%

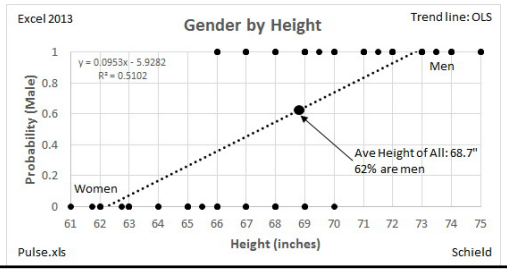
Difference: 5.35"

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	48	54	68	150	1	0	1	1
3	54	56	69	145	2	0	1	1
4	54	50	69	160	2	0	0	1
5	58	70	72	145	2	1	0	1
6	58	58	66	135	3	0	0	1
7	58	56	67	125	2	0	0	0
8	60	76	71	170	3	1	0	1
9	60	62	71	155	2	0	0	1
10	60	70	71.5	164	2	0	1	1
11	60	66	62	120	2	0	0	0
12	61	70	65.5	120	2	0	0	0
13	62	76	73.5	160	3	1	1	1
14	62	75	72	195	2	1	0	1
15	62	58	72	175	3	1	0	1
16	62	100	66	120	2	1	0	0
17	62	98	62.75	112	2	1	1	0

V1A Excel2013 Model Logistic MLE 1Y1X Slides 4

To Do: Model Gender by Height Show Trend-line and Equation

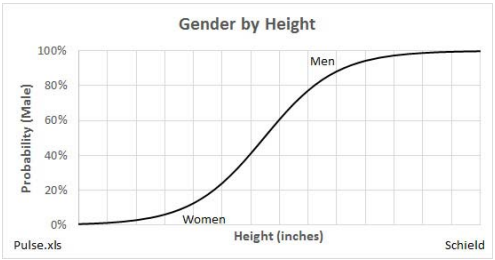
This trend-line does not satisfy the least-squares assumptions and it goes outside the valid range.



V1A Excel2013 Model Logistic MLE 1Y1X Slides 5

Intuitive idea of solution

This shape handles all heights: even if negative.



V1A Excel2013 Model Logistic MLE 1Y1X Slides 6

Outline of Approach: Five Steps

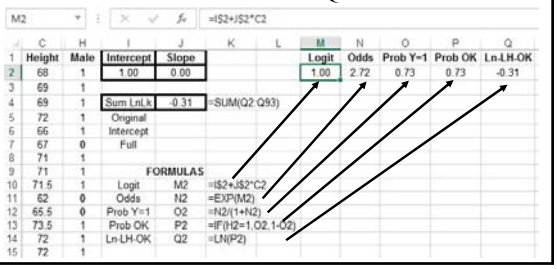
1) Prepare data for logistic MLE regression
2) Insert desired intercept
3) Use Solver to solve for intercept and slope
4) Generate various graphs
5) Test for statistical significance

To do: Get data at www.StatLit.org/xls/ Model-Logistic-MLE-1Y1X-Excel2013-Data.xlsx

V1A Excel2013 Model Logistic MLE 1Y1X Slides 7

1a) Load Data; Hide columns; Enter formula

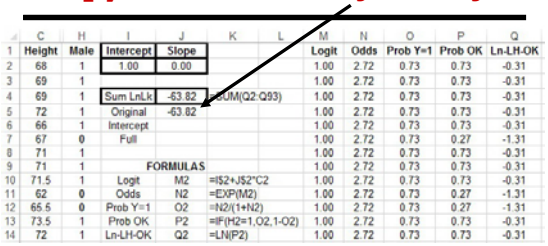
Hide columns A-B and D-G. Let I2=1; J2=0
Enter formula in M2-Q2 and J4



	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1									
4	69	1	Sum LnLk	-0.31	=SUM(Q2:Q93)						
5	72	1	Original								
6	66	1	Intercept								
7	67	0	Full								
8	71	1									
9	71	1	FORMULAS								
10	71.5	1	Logit	M2	=I$2+J$2*C2						
11	62	0	Odds	N2	=EXP(M2)						
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)						
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)						
14	72	1	Ln-LH-OK	Q2	=LN(P2)						
15	72	1									

V1A Excel2013 Model Logistic MLE 1Y1X Slides 8

1b) Select M2:Q2. Pull ↓ to 93 Copy/Paste Value of J4 onto J5



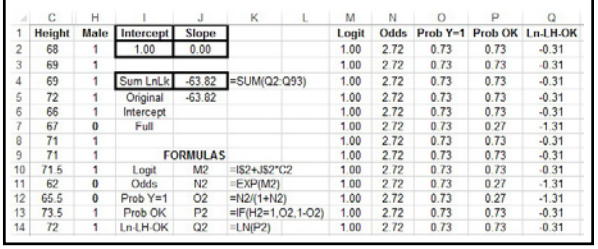
	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1					1.00	2.72	0.73	0.73	-0.31
4	69	1	Sum LnLk	-63.82	=SUM(Q2:Q93)		1.00	2.72	0.73	0.73	-0.31
5	72	1	Original	-63.82			1.00	2.72	0.73	0.73	-0.31
6	66	1	Intercept				1.00	2.72	0.73	0.73	-0.31
7	67	0	Full				1.00	2.72	0.73	0.27	-1.31
8	71	1					1.00	2.72	0.73	0.73	-0.31
9	71	1	FORMULAS				1.00	2.72	0.73	0.73	-0.31
10	71.5	1	Logit	M2	=I$2+J$2*C2		1.00	2.72	0.73	0.73	-0.31
11	62	0	Odds	N2	=EXP(M2)		1.00	2.72	0.73	0.27	-1.31
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		1.00	2.72	0.73	0.27	-1.31
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		1.00	2.72	0.73	0.73	-0.31
14	72	1	Ln-LH-OK	Q2	=LN(P2)		1.00	2.72	0.73	0.73	-0.31

Odds = Prob(Y=1)/[1-Prob(Y=1)] Range: 0 to infinity
Logit = LN(Odds). Range: -infinity to +infinity.
Logistic regress: Logit = Intercept + Slope*Height

V1A Excel2013 Model Logistic MLE 1Y1X Slides 9

1c) Review/Analyze

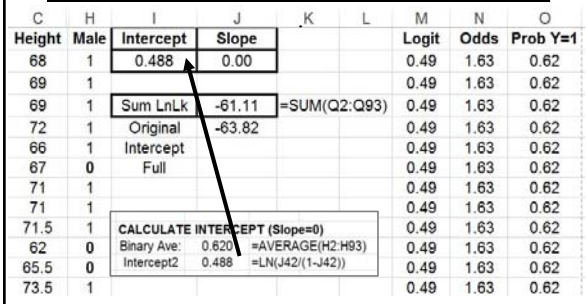
Intercept of 1 gives P(Y = 1 = male) = 73%.
But 62% of these students are male.
Step 2: Adjust intercept so P(Y=1) = 62%.



	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1					1.00	2.72	0.73	0.73	-0.31
4	69	1	Sum LnLk	-63.82	=SUM(Q2:Q93)		1.00	2.72	0.73	0.73	-0.31
5	72	1	Original	-63.82			1.00	2.72	0.73	0.73	-0.31
6	66	1	Intercept				1.00	2.72	0.73	0.73	-0.31
7	67	0	Full				1.00	2.72	0.73	0.27	-1.31
8	71	1					1.00	2.72	0.73	0.73	-0.31
9	71	1	FORMULAS				1.00	2.72	0.73	0.73	-0.31
10	71.5	1	Logit	M2	=I$2+J$2*C2		1.00	2.72	0.73	0.73	-0.31
11	62	0	Odds	N2	=EXP(M2)		1.00	2.72	0.73	0.27	-1.31
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		1.00	2.72	0.73	0.27	-1.31
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		1.00	2.72	0.73	0.73	-0.31
14	72	1	Ln-LH-OK	Q2	=LN(P2)		1.00	2.72	0.73	0.73	-0.31

V1A Excel2013 Model Logistic MLE 1Y1X Slides 10

2a) Calculate Intercept2 Paste Value into I2

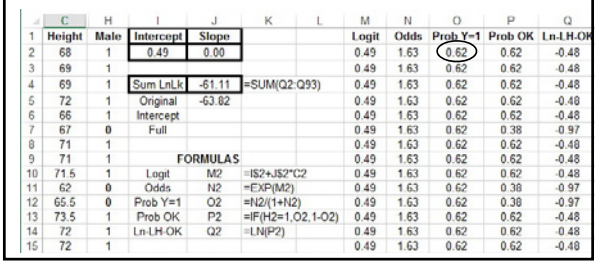


C	H	I	J	K	L	M	N	O
Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1
68	1	0.488	0.00			0.49	1.63	0.62
69	1					0.49	1.63	0.62
69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62
72	1	Original	-63.82			0.49	1.63	0.62
66	1	Intercept				0.49	1.63	0.62
67	0	Full				0.49	1.63	0.62
71	1					0.49	1.63	0.62
71	1					0.49	1.63	0.62
71.5	1	CALCULATE INTERCEPT (Slope=0)				0.49	1.63	0.62
62	0	Binary Ave:	0.620	=AVERAGE(H2:H93)		0.49	1.63	0.62
65.5	0	Intercept2	0.488	=LN(J42/(1-J42))		0.49	1.63	0.62
73.5	1					0.49	1.63	0.62

V1A Excel2013 Model Logistic MLE 1Y1X Slides 11

2b) Results are as expected: Probability of Male = 62%

If we must select a single value to predict the outcome, it would be the percentage of students who are men.

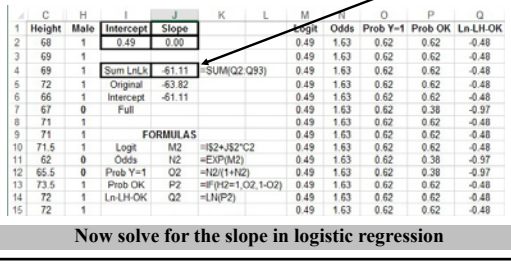


	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	0.49	0.00			0.49	1.63	0.62	0.62	-0.48
3	69	1					0.49	1.63	0.62	0.62	-0.48
4	69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62	0.62	-0.48
5	72	1	Original	-63.82			0.49	1.63	0.62	0.62	-0.48
6	66	1	Intercept				0.49	1.63	0.62	0.62	-0.48
7	67	0	Full				0.49	1.63	0.62	0.38	-0.97
8	71	1					0.49	1.63	0.62	0.62	-0.48
9	71	1	FORMULAS				0.49	1.63	0.62	0.62	-0.48
10	71.5	1	Logit	M2	=I$2+J$2*C2		0.49	1.63	0.62	0.62	-0.48
11	62	0	Odds	N2	=EXP(M2)		0.49	1.63	0.62	0.38	-0.97
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		0.49	1.63	0.62	0.38	-0.97
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		0.49	1.63	0.62	0.62	-0.48
14	72	1	Ln-LH-OK	Q2	=LN(P2)		0.49	1.63	0.62	0.62	-0.48
15	72	1					0.49	1.63	0.62	0.62	-0.48

V1A Excel2013 Model Logistic MLE 1Y1X Slides 12

2c) Copy sum of Ln-Lk-OK Copy J4; Paste value into J6

Copy "Sum LnLk" from J4 to clipboard.
Paste-Special Value in J6

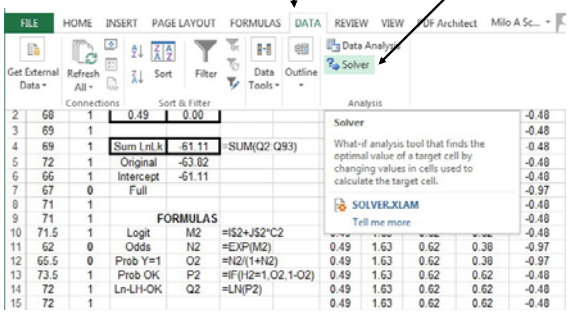


	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	0.49	0.00			0.49	1.63	0.62	0.62	-0.48
3	69	1					0.49	1.63	0.62	0.62	-0.48
4	69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62	0.62	-0.48
5	72	1	Original	-63.82			0.49	1.63	0.62	0.62	-0.48
6	66	1	Intercept	-61.11			0.49	1.63	0.62	0.62	-0.48
7	67	0	Full				0.49	1.63	0.62	0.38	-0.97
8	71	1					0.49	1.63	0.62	0.62	-0.48
9	71	1	FORMULAS				0.49	1.63	0.62	0.62	-0.48
10	71.5	1	Logit	M2	=I$2+J$2*C2		0.49	1.63	0.62	0.62	-0.48
11	62	0	Odds	N2	=EXP(M2)		0.49	1.63	0.62	0.38	-0.97
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		0.49	1.63	0.62	0.38	-0.97
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		0.49	1.63	0.62	0.62	-0.48
14	72	1	Ln-LH-OK	Q2	=LN(P2)		0.49	1.63	0.62	0.62	-0.48
15	72	1					0.49	1.63	0.62	0.62	-0.48

Now solve for the slope in logistic regression

V1A Excel2013 Model Logistic MLE 1Y1X Slides 13

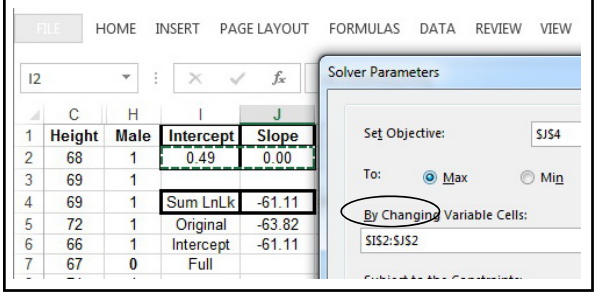
3a) Solve for Slope and Intercept From Data menu, select Solver



V1A Excel2013 Model Logistic MLE 1Y1X Slides 14

3b) Set Solver Parameters

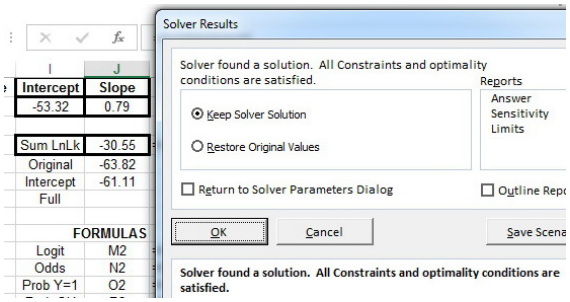
Select Intercept and slope (I2:J2) in "Changing Variable Cells".



	C	H	I	J
1	Height	Male	Intercept	Slope
2	68	1	0.49	0.00
3	69	1		
4	69	1	Sum LnLk	-61.11
5	72	1	Original	-63.82
6	66	1	Intercept	-61.11
7	67	0	Full	

V1A Excel2013 Model Logistic MLE 1Y1X Slides 15

3c) Results: All constraints and conditions are satisfied



Intercept	Slope
-53.32	0.79

Sum LnLk	-30.55
Original	-63.82
Intercept	-61.11
Full	

FORMULAS	
Logit	M2
Odds	N2
Prob Y=1	O2

V1A Excel2013 Model Logistic MLE 1Y1X Slides 16

3d) Error in row 6: short guy. Classified gals in 7, 11 & 12 OK

	C	H	L	O	P
1	Height	Male		Prob Y=1	Prob OK
2	68	1		0.61	0.61
3	69	1		0.77	0.77
4	69	1		0.77	0.77
5	72	1		0.97	0.97
6	66	1		0.24	0.24
7	67	0		0.41	0.59
8	71	1		0.94	0.94
9	71	1		0.94	0.94
10	71.5	1		0.96	0.96
11	62	0		0.01	0.99
12	65.5	0		0.18	0.82
13	73.5	1		0.99	0.99

Misclassified

V1A Excel2013 Model Logistic MLE 1Y1X Slides 17

3e) Prepare to test slope for statistical significance

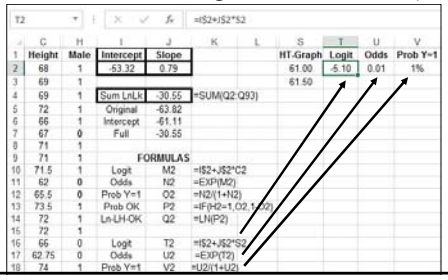
From J4, copy "Sum LnLk" to clipboard.
Paste-Special Value onto J7.

	C	H	I	J	K	L
1	Height	Male	Intercept	Slope		
2	68	1	-53.32	0.79		
3	69	1				
4	69	1	Sum LnLk	-30.55	=SUM(Q2:Q93)	
5	72	1	Original	-63.82		
6	66	1	Intercept	-61.11		
7	67	0	Full	-30.55		
8	71	1				

V1A Excel2013 Model Logistic MLE 1Y1X Slides 18

4a) Prepare data for Graphs

Start Ht-Graph at minimum of height in S2
Enter formulas for Logit, Odds and Prob(Y=1).



V1A Excel2013 Model Logistic MLE 1Y1X Slides 19

4b) Prepare X and Y data. X is Height. Y is Prob (Y=1)

Select T2:V2.
Drag down 1 row

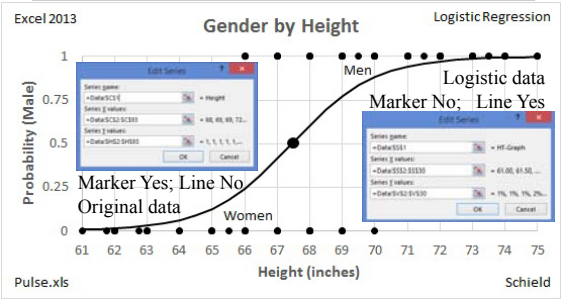
Select S2:V3.
Drag to bottom

Select columns S & V
for logistic graph.

	S	T	U	V
	HT-Graph	Logit	Odds	Prob Y=1
	61.00	-5.10	0.01	1%
	61.50	-4.71	0.01	1%
	62.00	-4.31	0.01	1%
	62.50	-3.92	0.02	2%
	63.00	-3.52	0.03	3%
	63.50	-3.12	0.04	4%
	64.00	-2.73	0.07	6%
	64.50	-2.33	0.10	9%
	65.00	-1.94	0.14	13%
	65.50	-1.54	0.21	18%
	66.00	-1.15	0.32	24%
	66.50	-0.75	0.47	32%
	67.00	-0.36	0.70	41%
	67.50	0.04	1.04	51%
	68.00	0.43	1.54	61%

V1A Excel2013 Model Logistic MLE 1Y1X Slides 20

4c) Graph Logistic Regression of Gender by Height.



Original data: Col C & H; Logistic data: Col S & V

V1A Excel2013 Model Logistic MLE 1Y1X Slides 21

5) Hypothesis test: Is non-zero slope statistically significant?

Calculate difference with 'full'; multiply by -2.
Conduct a right-tail Chi^2 test with 1 degree freedom.

	C	H	I	J	K	L
1	Height	Male	Intercept	Slope		
2	68	1	-53.32	.0.79		
3	69	1				
4	69	1	Sum LnLk	-30.55	=sum(Q2:Q9)	
5	72	1	Original	-63.82	Manual paste	
6	66	1	Intercept	-61.11	Manual paste	
7	67	0	Full	-30.55	Manual paste	
8	71	1	Difference	-30.56	=J6-J7	
9	71	1	DegFree	1	Manual entry	
10	71.5	1	P-value	5.35E-15	=CHISQ.DIST.RT(-2*J8,J9)	

Slope is statistically significant: P-value < 0.05

V1A Excel2013 Model Logistic MLE 1Y1X Slides 22

Acknowledgment and Reference

ACKNOWLEDGMENT:

This presentation closely follows the Carlberg (2012) presentation in Chapter 2: pages 21-52.

These slides present the how – step by step – of logistic regression for a single case.

Carlberg (2012) discusses the how and the why.

Schield introduced the shortcut on slide 10.

REFERENCE:

Carlberg, Conrad (2012). *Decision Analytics: Microsoft Excel*. Que Publishing.

Model Logistic Regression MLE 1Y1X in Excel 2013

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides and data at: www.StatLit.org/
pdf/Excel2013-Model-Logistic-MLE-1Y1X-Slides.pdf
xls/Excel2013-Model-Logistic-MLE-1Y1X-Data.xlsx

Background & Goals

Modelling a binary outcome (buy/look, payoff/default, go/nogo or male/female) requires logistic regression.

Doing logistic regression in Excel requires Solver. "Since its introduction in .. 1991, … Excel Solver has become the most widely distributed – and almost surely the most widely used – general-purpose optimization modeling system." www.utexas.edu/courses/lasdon/design3.htm

This presentation uses college student data: pulse.xls. This demo models gender (male) based on height.

Goals: Create graphs on slides 4 and 22.
Determine if slope is statistically significant.

This demo uses Height (col C) to predict Gender (col H)

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

Ave Heights:

M: 70.75” 62%

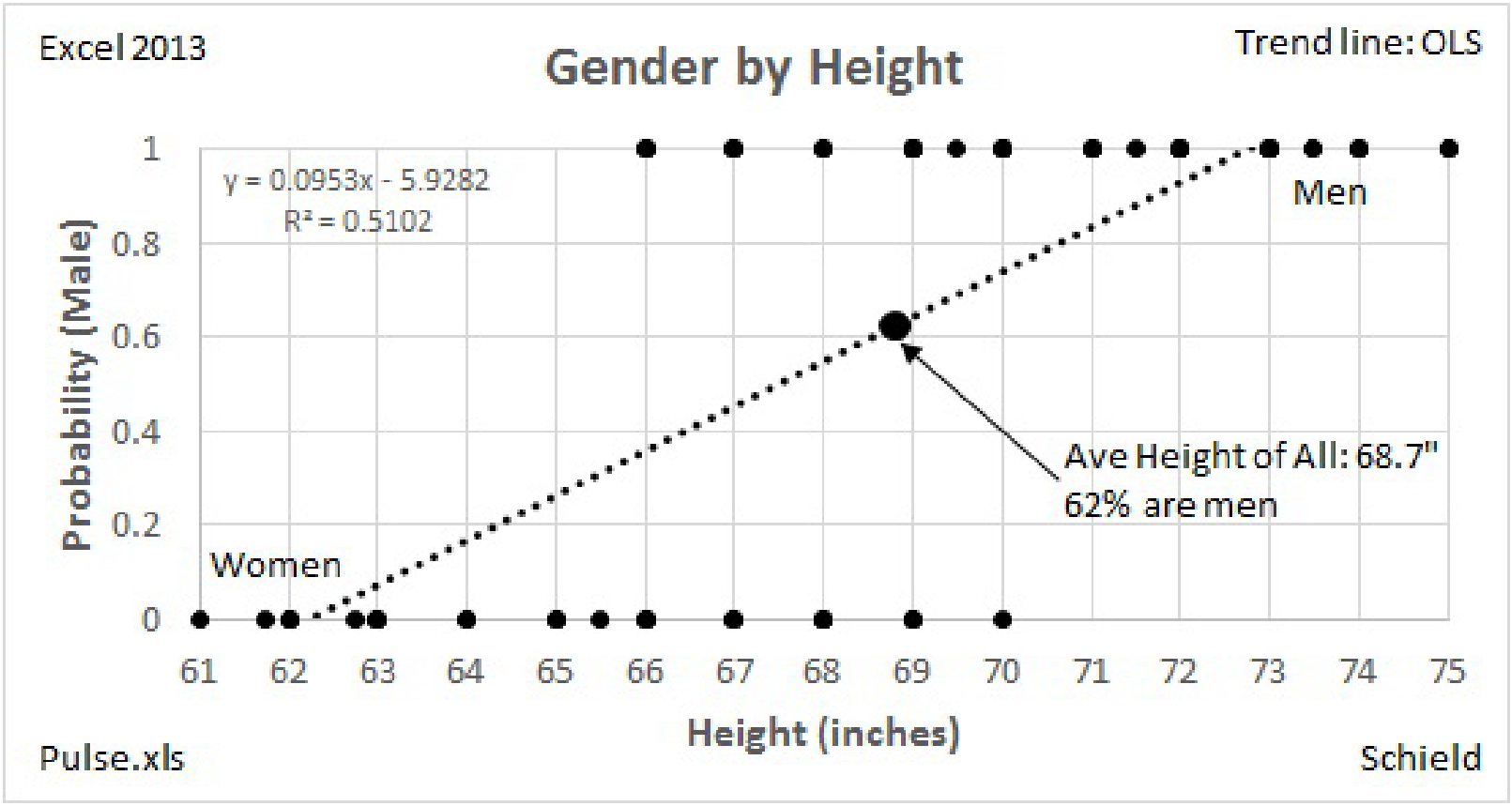
F: 65.3” 38%

Difference: 5.35”

	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	48	54	68	150	1	0	1	1
3	54	56	69	145	2	0	1	1
4	54	50	69	160	2	0	0	1
5	58	70	72	145	2	1	0	1
6	58	58	66	135	3	0	0	1
7	58	56	67	125	2	0	0	0
8	60	76	71	170	3	1	0	1
9	60	62	71	155	2	0	0	1
10	60	70	71.5	164	2	0	1	1
11	60	66	62	120	2	0	0	0
12	61	70	65.5	120	2	0	0	0
13	62	76	73.5	160	3	1	1	1
14	62	75	72	195	2	1	0	1
15	62	58	72	175	3	1	0	1
16	62	100	66	120	2	1	0	0
17	62	98	62.75	112	2	1	1	0

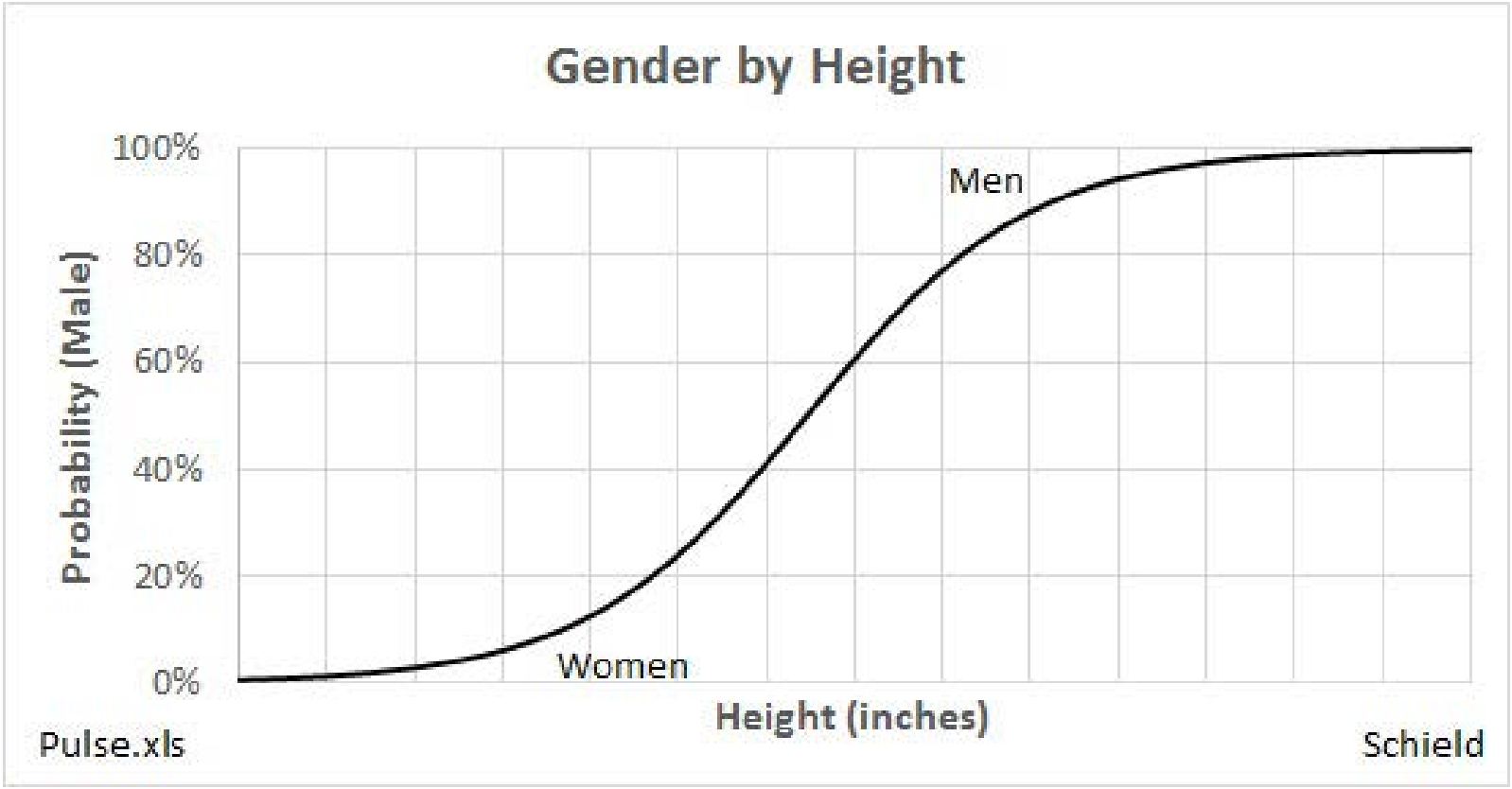
To Do: Model Gender by Height Show Trend-line and Equation

This trend-line does not satisfy the least-squares assumptions and it goes outside the valid range.



Intuitive idea of solution

This shape handles all heights: even if negative.



Outline of Approach: Five Steps

1) Prepare data for logistic MLE regression
2) Insert desired intercept
3) Use Solver to solve for intercept and slope
4) Generate various graphs
5) Test for statistical significance

To do: Get data at www.StatLit.org/xls/ Model-Logistic-MLE-1Y1X-Excel2013-Data.xlsx

1a) Load Data; Hide columns; Enter formula

Hide columns A-B and D-G. Let I2=1; J2=0
Enter formula in M2-Q2 and J4

M2 | =I$2+J$2*C2

	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1									
4	69	1	Sum LnLk	-0.31	=SUM(Q2:Q93)						
5	72	1	Original								
6	66	1	Intercept								
7	67	0	Full								
8	71	1									
9	71	1		FORMULAS							
10	71.5	1	Logit	M2	=I$2+J$2*C2						
11	62	0	Odds	N2	=EXP(M2)						
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)						
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)						
14	72	1	Ln-LH-OK	Q2	=LN(P2)						
15	72	1									

1b) Select M2:Q2. Pull ↓ to 93 Copy/Paste Value of J4 onto J5

	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1					1.00	2.72	0.73	0.73	-0.31
4	69	1	Sum LnLk	-63.82	=SUM(Q2:Q93)		1.00	2.72	0.73	0.73	-0.31
5	72	1	Original	-63.82			1.00	2.72	0.73	0.73	-0.31
6	66	1	Intercept				1.00	2.72	0.73	0.73	-0.31
7	67	0	Full				1.00	2.72	0.73	0.27	-1.31
8	71	1					1.00	2.72	0.73	0.73	-0.31
9	71	1		FORMULAS			1.00	2.72	0.73	0.73	-0.31
10	71.5	1	Logit	M2	=I$2+J$2*C2		1.00	2.72	0.73	0.73	-0.31
11	62	0	Odds	N2	=EXP(M2)		1.00	2.72	0.73	0.27	-1.31
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		1.00	2.72	0.73	0.27	-1.31
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		1.00	2.72	0.73	0.73	-0.31
14	72	1	Ln-LH-OK	Q2	=LN(P2)		1.00	2.72	0.73	0.73	-0.31

Odds = Prob(Y=1)/[1-Prob(Y=1)] Range: 0 to infinity

Logit = LN(Odds). Range: -infinity to +infinity.

Logistic regress: Logit = Intercept + Slope*Height

1c) Review/Analyze

Intercept of 1 gives P(Y = 1 = male) = 73%.
But 62% of these students are male.
Step 2: Adjust intercept so P(Y=1) = 62%.

	C	H	I	J	K	L	M	N	O	P	Q
1	**Height**	**Male**	**Intercept**	**Slope**			**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
2	68	1	1.00	0.00			1.00	2.72	0.73	0.73	-0.31
3	69	1					1.00	2.72	0.73	0.73	-0.31
4	69	1	Sum LnLk	-63.82	=SUM(Q2:Q93)		1.00	2.72	0.73	0.73	-0.31
5	72	1	Original	-63.82			1.00	2.72	0.73	0.73	-0.31
6	66	1	Intercept				1.00	2.72	0.73	0.73	-0.31
7	67	**0**	Full				1.00	2.72	0.73	0.27	-1.31
8	71	1					1.00	2.72	0.73	0.73	-0.31
9	71	1		**FORMULAS**			1.00	2.72	0.73	0.73	-0.31
10	71.5	1	Logit	M2	=I$2+J$2*C2		1.00	2.72	0.73	0.73	-0.31
11	62	**0**	Odds	N2	=EXP(M2)		1.00	2.72	0.73	0.27	-1.31
12	65.5	**0**	Prob Y=1	O2	=N2/(1+N2)		1.00	2.72	0.73	0.27	-1.31
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		1.00	2.72	0.73	0.73	-0.31
14	72	1	Ln-LH-OK	Q2	=LN(P2)		1.00	2.72	0.73	0.73	-0.31

2a) Calculate Intercept2 Paste Value into I2

C	H	I	J	K	L	M	N	O
Height	**Male**	**Intercept**	**Slope**			**Logit**	**Odds**	**Prob Y=1**
68	1	0.488	0.00			0.49	1.63	0.62
69	1					0.49	1.63	0.62
69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62
72	1	Original	-63.82			0.49	1.63	0.62
66	1	Intercept				0.49	1.63	0.62
67	**0**	Full				0.49	1.63	0.62
71	1					0.49	1.63	0.62
71	1					0.49	1.63	0.62
71.5	1					0.49	1.63	0.62
62	**0**					0.49	1.63	0.62
65.5	**0**					0.49	1.63	0.62
73.5	1					0.49	1.63	0.62

CALCULATE INTERCEPT (Slope=0)

Binary Ave:	0.620	=AVERAGE(H2:H93)
Intercept2	0.488	=LN(J42/(1-J42))

2b) Results are as expected: Probability of Male = 62%

If we must select a single value to predict the outcome, it would be the percentage of students who are men.

	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	0.49	0.00			0.49	1.63	0.62	0.62	-0.48
3	69	1					0.49	1.63	0.62	0.62	-0.48
4	69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62	0.62	-0.48
5	72	1	Original	-63.82			0.49	1.63	0.62	0.62	-0.48
6	66	1	Intercept				0.49	1.63	0.62	0.62	-0.48
7	67	0	Full				0.49	1.63	0.62	0.38	-0.97
8	71	1					0.49	1.63	0.62	0.62	-0.48
9	71	1	FORMULAS				0.49	1.63	0.62	0.62	-0.48
10	71.5	1	Logit	M2	=I$2+J$2*C2		0.49	1.63	0.62	0.62	-0.48
11	62	0	Odds	N2	=EXP(M2)		0.49	1.63	0.62	0.38	-0.97
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		0.49	1.63	0.62	0.38	-0.97
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		0.49	1.63	0.62	0.62	-0.48
14	72	1	Ln-LH-OK	Q2	=LN(P2)		0.49	1.63	0.62	0.62	-0.48
15	72	1					0.49	1.63	0.62	0.62	-0.48

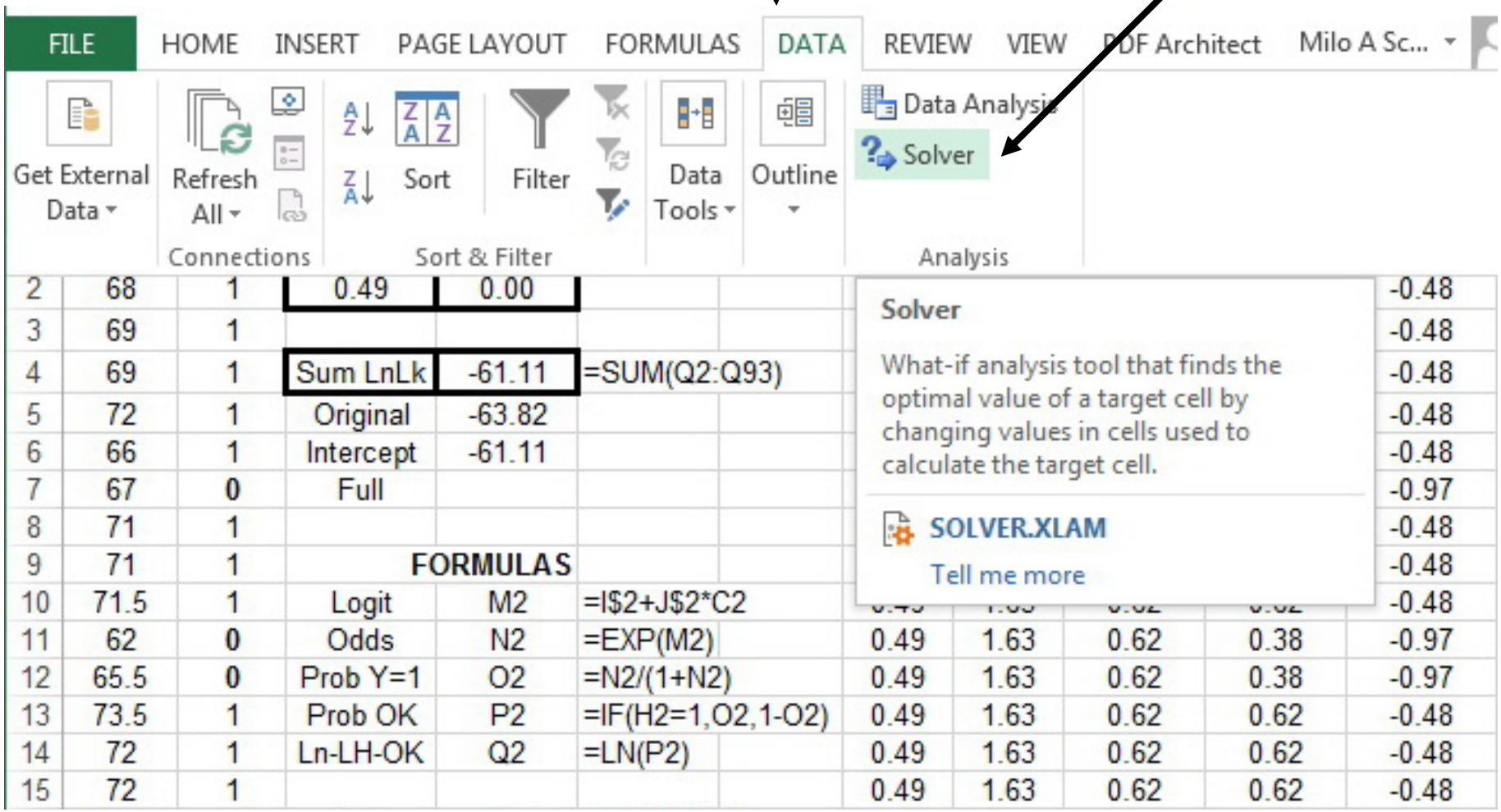
2c) Copy sum of Ln-Lk-OK Copy J4; Paste value into J6

Copy “Sum LnLk” from J4 to clipboard.
Paste-Special Value in J6

	C	H	I	J	K	L	M	N	O	P	Q
1	Height	Male	Intercept	Slope			Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
2	68	1	0.49	0.00			0.49	1.63	0.62	0.62	-0.48
3	69	1					0.49	1.63	0.62	0.62	-0.48
4	69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)		0.49	1.63	0.62	0.62	-0.48
5	72	1	Original	-63.82			0.49	1.63	0.62	0.62	-0.48
6	66	1	Intercept	-61.11			0.49	1.63	0.62	0.62	-0.48
7	67	0	Full				0.49	1.63	0.62	0.38	-0.97
8	71	1					0.49	1.63	0.62	0.62	-0.48
9	71	1		FORMULAS			0.49	1.63	0.62	0.62	-0.48
10	71.5	1	Logit	M2	=I$2+J$2*C2		0.49	1.63	0.62	0.62	-0.48
11	62	0	Odds	N2	=EXP(M2)		0.49	1.63	0.62	0.38	-0.97
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)		0.49	1.63	0.62	0.38	-0.97
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)		0.49	1.63	0.62	0.62	-0.48
14	72	1	Ln-LH-OK	Q2	=LN(P2)		0.49	1.63	0.62	0.62	-0.48
15	72	1					0.49	1.63	0.62	0.62	-0.48

Now solve for the slope in logistic regression

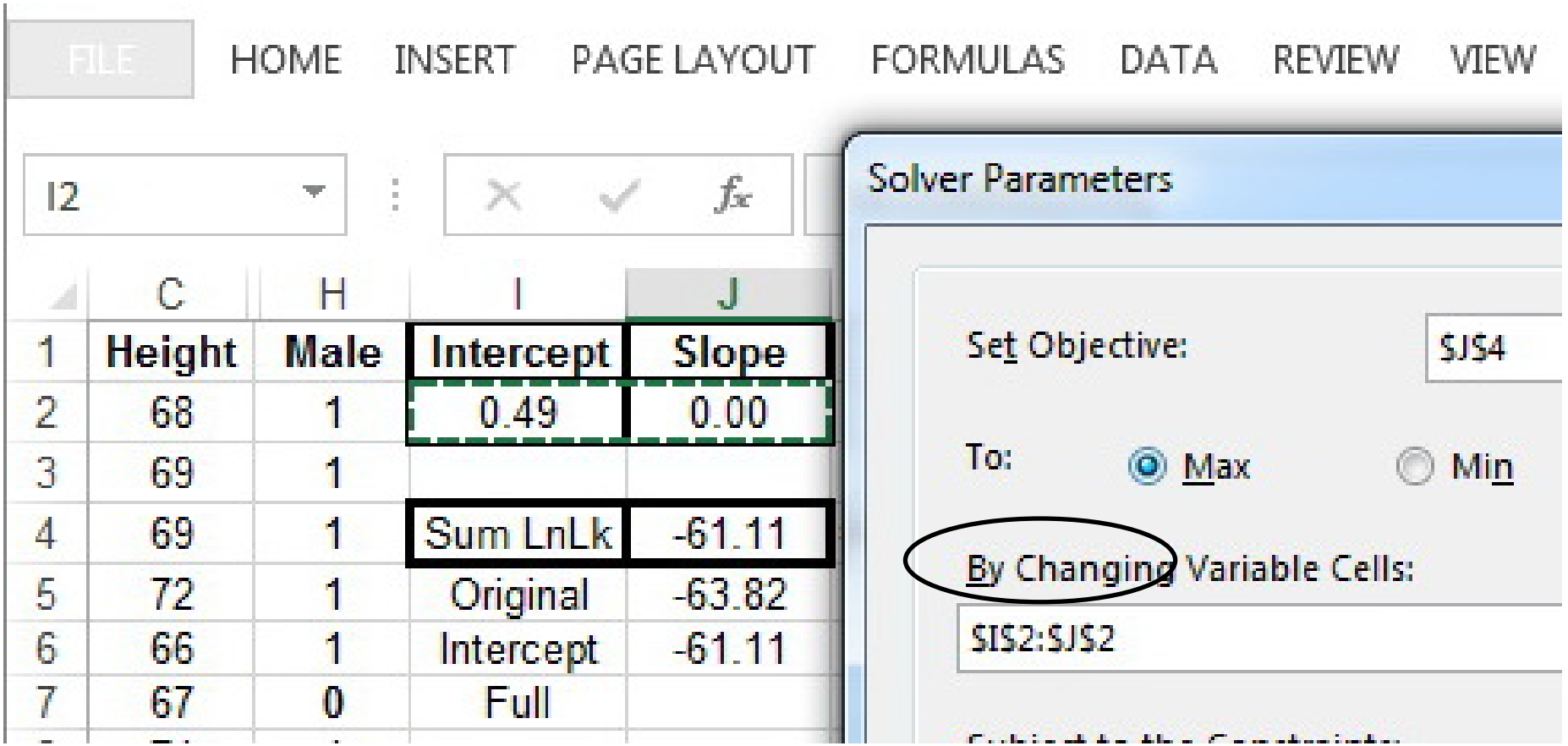
3a) Solve for Slope and Intercept From Data menu, select Solver



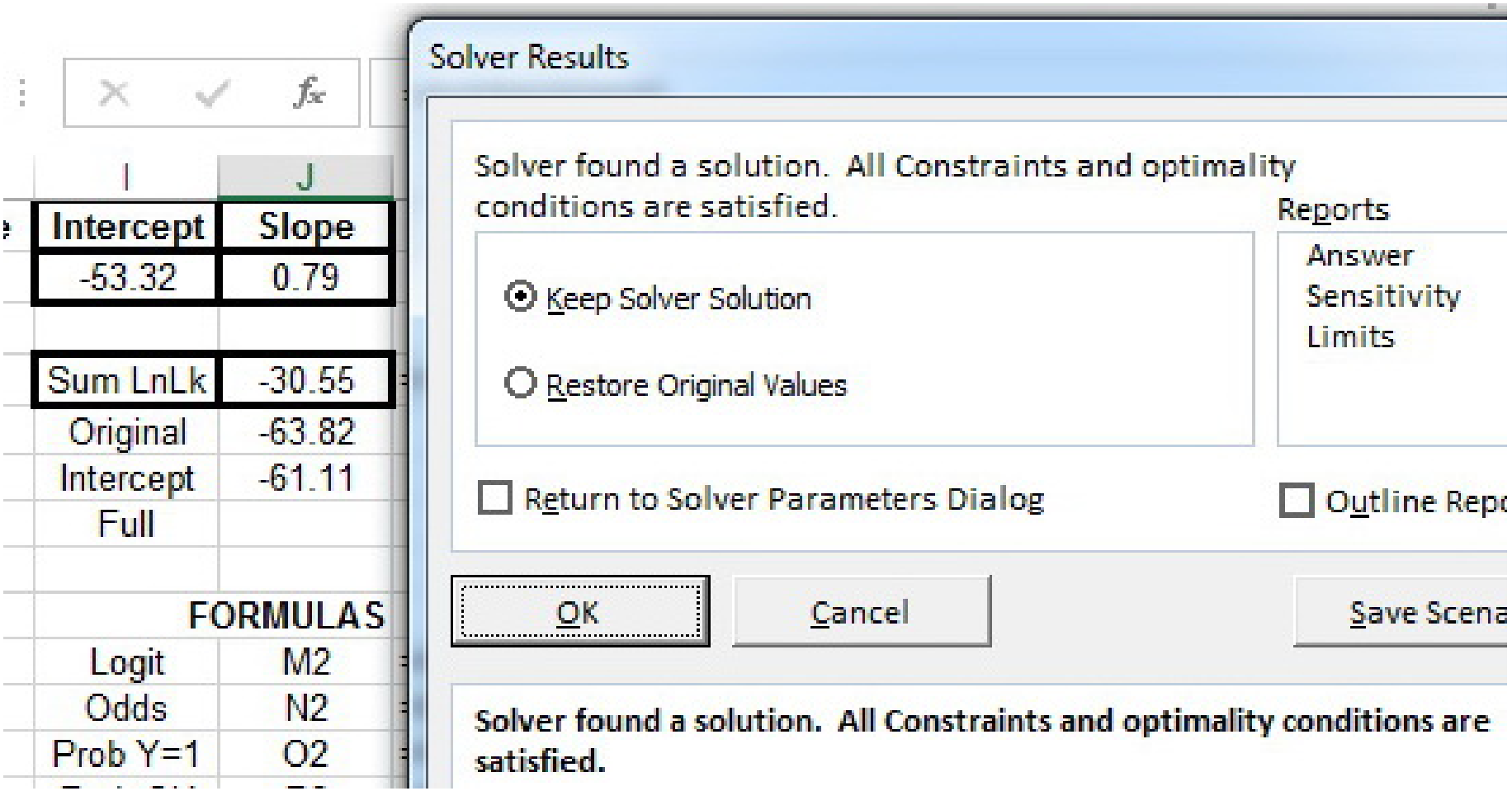
2	68	1	0.49	0.00						-0.48
3	69	1								-0.48
4	69	1	Sum LnLk	-61.11	=SUM(Q2:Q93)					-0.48
5	72	1	Original	-63.82						-0.48
6	66	1	Intercept	-61.11						-0.48
7	67	0	Full							-0.97
8	71	1								-0.48
9	71	1		FORMULAS						-0.48
10	71.5	1	Logit	M2	=I$2+J$2*C2					-0.48
11	62	0	Odds	N2	=EXP(M2)	0.49	1.63	0.62	0.38	-0.97
12	65.5	0	Prob Y=1	O2	=N2/(1+N2)	0.49	1.63	0.62	0.38	-0.97
13	73.5	1	Prob OK	P2	=IF(H2=1,O2,1-O2)	0.49	1.63	0.62	0.62	-0.48
14	72	1	Ln-LH-OK	Q2	=LN(P2)	0.49	1.63	0.62	0.62	-0.48
15	72	1				0.49	1.63	0.62	0.62	-0.48

3b) Set Solver Parameters

Select Intercept and slope (I2:J2) in “Changing Variable Cells”.



3c) Results: All constraints and conditions are satisfied



3d) Error in row 6: short guy. Classified gals in 7, 11 & 12 OK

	C	H	L	O	P
1	Height	Male		Prob Y=1	Prob OK
2	68	1		0.61	0.61
3	69	1		0.77	0.77
4	69	1		0.77	0.77
5	72	1		0.97	0.97
6	66	1		0.24	0.24
7	67	0		0.41	0.59
8	71	1		0.94	0.94
9	71	1		0.94	0.94
10	71.5	1		0.96	0.96
11	62	0		0.01	0.99
12	65.5	0		0.18	0.82
13	73.5	1		0.99	0.99

Misclassified

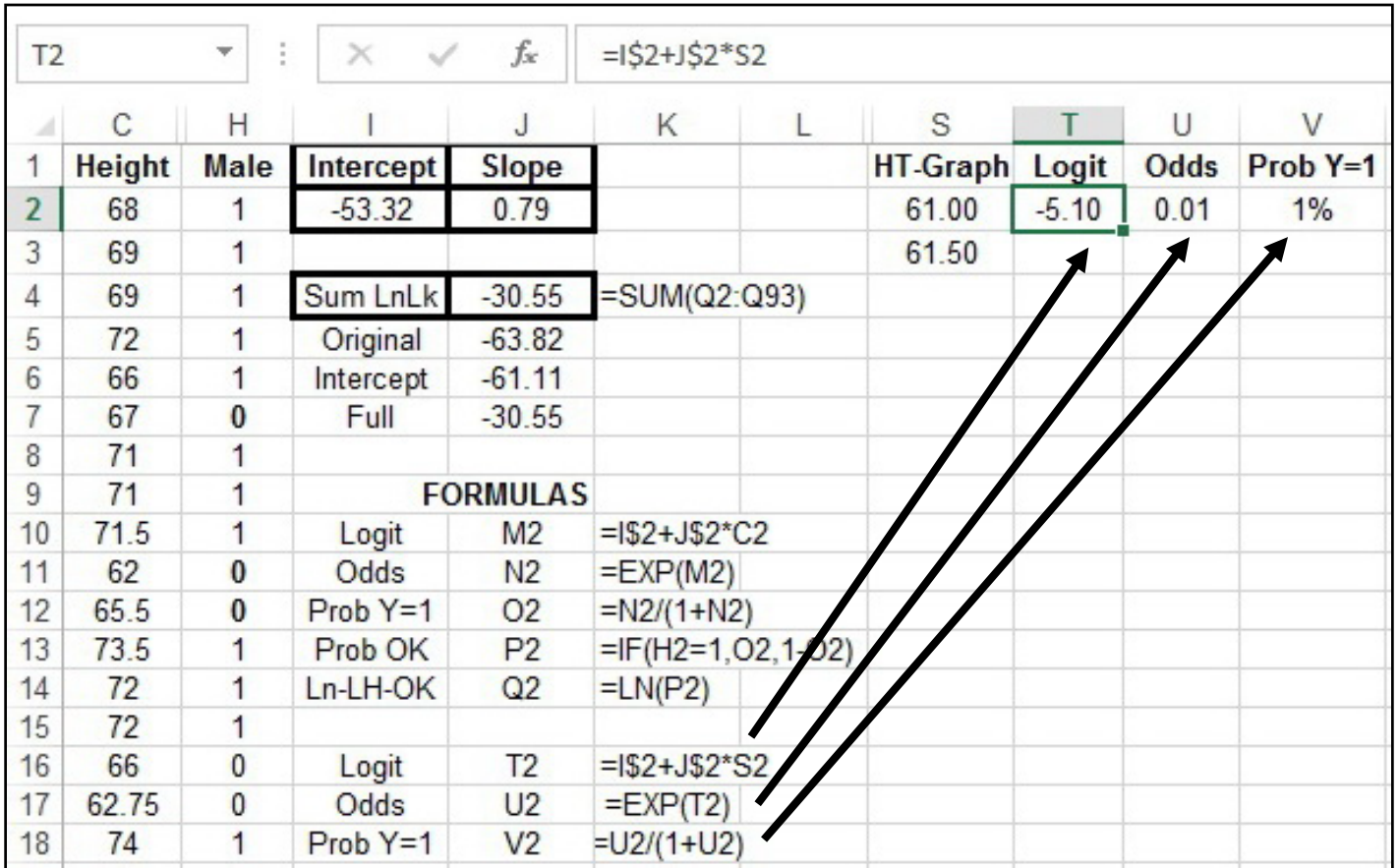
3e) Prepare to test slope for statistical significance

From J4, copy "Sum LnLk" to clipboard.
Paste-Special Value onto J7.

	C	H	I	J	K	L
1	**Height**	**Male**	**Intercept**	**Slope**		
2	68	1	-53.32	0.79		
3	69	1				
4	69	1	Sum LnLk	-30.55	=SUM(Q2:Q93)	
5	72	1	Original	-63.82		
6	66	1	Intercept	-61.11		
7	67	0	Full	-30.55	←	
8	71	1				

4a) Prepare data for Graphs

Start Ht-Graph at minimum of height in S2
Enter formulas for Logit, Odds and Prob(Y=1).



4b) Prepare X and Y data. X is Height. Y is Prob (Y=1)

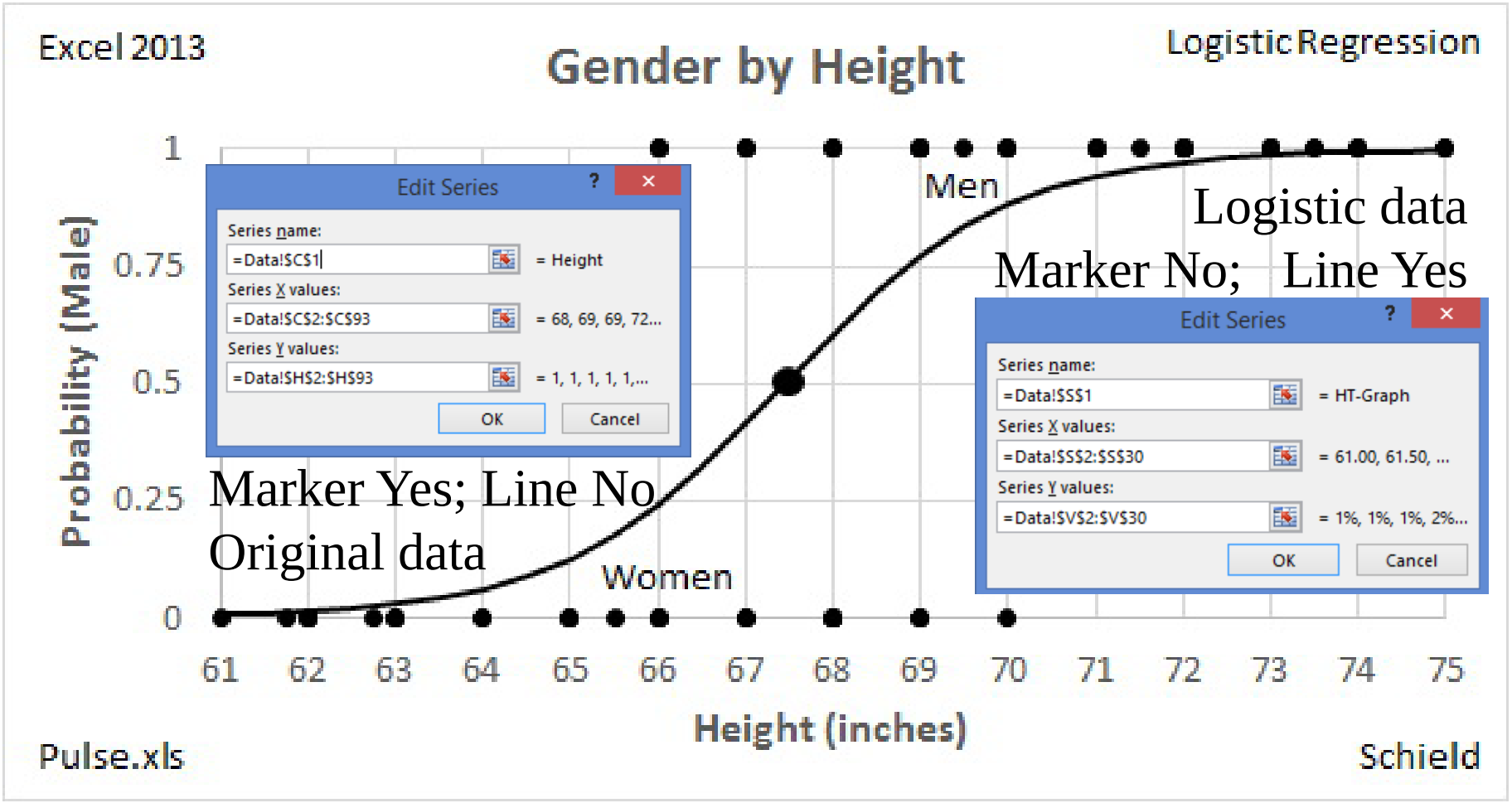
Select T2:V2.
Drag down 1 row

Select S2:V3.
Drag to bottom

Select columns S & V for logistic graph.

S	T	U	V
HT-Graph	Logit	Odds	Prob Y=1
61.00	-5.10	0.01	1%
61.50	-4.71	0.01	1%
62.00	-4.31	0.01	1%
62.50	-3.92	0.02	2%
63.00	-3.52	0.03	3%
63.50	-3.12	0.04	4%
64.00	-2.73	0.07	6%
64.50	-2.33	0.10	9%
65.00	-1.94	0.14	13%
65.50	-1.54	0.21	18%
66.00	-1.15	0.32	24%
66.50	-0.75	0.47	32%
67.00	-0.36	0.70	41%
67.50	0.04	1.04	51%
68.00	0.43	1.54	61%

4c) Graph Logistic Regression of Gender by Height.



Original data: Col C & H; Logistic data: Col S & V

5) Hypothesis test: Is non-zero slope statistically significant?

Calculate difference with ‘full’; multiply by -2.

Conduct a right-tail Chi^2 test with 1 degree freedom.

	C	H	I	J	K	L
1	**Height**	**Male**	**Intercept**	**Slope**		
2	68	1	-53.32	.0.79		
3	69	1				
4	69	1	Sum LnLk	-30.55	=sum(Q2:Q9)	
5	72	1	Original	-63.82	Manual paste	
6	66	1	Intercept	-61.11	Manual paste	
7	67	**0**	Full	-30.55	Manual paste	
8	71	1	Difference	-30.56	=J6-J7	
9	71	1	DegFree	1	Manual entry	
10	71.5	1	P-value	5.35E-15	=CHISQ.DIST.RT(-2*J8,J9)	

Slope is statistically significant: P-value < 0.05

Acknowledgment and Reference

ACKNOWLEDGMENT:

This presentation closely follows the Carlberg (2012) presentation in Chapter 2: pages 21-52.

These slides present the how – step by step – of logistic regression for a single case.

Carlberg (2012) discusses the how and the why.

Schield introduced the shortcut on slide 10.

REFERENCE:

Carlberg, Conrad (2012). *Decision Analytics: Microsoft Excel*. Que Publishing.