Logistic Regression: MLE with 2 inputs, 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/2015-Schield-Logistic-MLE1C-Excel2013-Demo.pdf
pdf/2015-Schield-Logistic-MLE1C-Excel2013-Slides.pdf
xls/2015-Schield-Logistic-MLE1C-Excel2013-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 graph on slide 17.
Determine if slopes are statistically significant.

Use Height (A) and Weight (B) to predict Gender (col C)

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

Ave Heights:

M: 70.75" 62%

F: 65.3" 38%

Difference: 5.35"

A	B
Height	Male
68	1
69	1
69	1
72	1
66	1
67	0
71	1
71	1
71.5	1
62	0
65.5	0

Outline of Approach: Four Steps

1) Prepare data for logistic MLE regression
   Insert desired intercept
2) Use Solver to solve for intercept and slopes
3) Test for statistical significance
4) Generate graph

To do: Get data at www.StatLit.org/Excel/ 2015-Schield-Logistic-MLE1C-Excel2013-Data.xlsx

1a) Load Data; Find Mean(Y). Set Intercept

Enter formula for F21 and F22.

	E	F	G
19			
20	GENDER & INTERCEPT #1		
21	Male-Pctg	0.62	=AVERAGE(C3:C94)
22	Intercept#1	0.4877	=LN(F21/(1-F21))

Copy value from F22; Paste in E3

Row	E	F	G
2	Intercept	SlopeHt	SlopeWt
3	0.4877	0.0000	0.0000

1b) Enter formula for I3-M3 Select I3:M3. Pull to row 94.

Row	E	F	G	H	I	J	K	L	M
2	Intercept	SlopeHt	SlopeWt		Logit	Odds	Prob Y=1	Prob OK	Ln-LH-OK
3	0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
4									
5	Sum LnLk	-0.97	=SUM(M3:M94)						
6	Sum LnLk1		Sum #1: Manual						
7	Sum Ln Lk2		Sum #2: Solver MLE						
8	Chi-Sq	0.00	=-2*(F6-F7)						
9	P-Value	1	=CHISQ.DIST.RT(F8,1)						
10									
11	FORMULAS & TEXT: Enter, Copy Down								
12	Logit	I3	=E$3+F$3*A3+G$3*B3						
13	Odds	J3	=EXP(I3)						
14	Prob Y=1	K3	=J3/(1+J3)						
15	Prob OK	L3	=IF(C3=1,K3,1-K3)						
16	Ln-LH-OK	M3	=LN(L3)						

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 7

1c) Results are as expected. Probability of male = 0.62

Row	E	F	G	H	I	J	K	L	M
2	**Intercept**	**SlopeHt**	**SlopeWt**		**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
3	0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
4					0.49	1.63	0.62	0.38	-0.97
5	Sum LnLk	-61.11	=SUM(M3:M94)		0.49	1.63	0.62	0.38	-0.97
6	Sum LnLk1		Sum #1: Manual		0.49	1.63	0.62	0.38	-0.97
7	Sum Ln Lk2		Sum #2: Solver MLE		0.49	1.63	0.62	0.38	-0.97
8	Chi-Sq	0.00	=-2*(F6-F7)		0.49	1.63	0.62	0.38	-0.97
9	P-Value	1	=CHISQ.DIST.RT(F8,1)		0.49	1.63	0.62	0.38	-0.97
10					0.49	1.63	0.62	0.38	-0.97
11	**FORMULAS & TEXT: Enter, Copy Down**				0.49	1.63	0.62	0.38	-0.97
12	Logit	I3	=E$3+F$3*A3+G$3*B3		0.49	1.63	0.62	0.38	-0.97
13	Odds	J3	=EXP(I3)		0.49	1.63	0.62	0.38	-0.97
14	Prob Y=1	K3	=J3/(1+J3)		0.49	1.63	0.62	0.38	-0.97
15	Prob OK	L3	=IF(C3=1,K3,1-K3)		0.49	1.63	0.62	0.38	-0.97
16	Ln-LH-OK	M3	=LN(L3)		0.49	1.63	0.62	0.38	-0.97
17					0.49	1.63	0.62	0.38	-0.97
18					0.49	1.63	0.62	0.38	-0.97

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 8

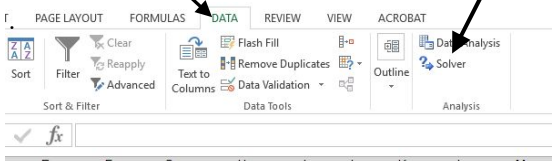
1d) Copy "Value" of F5 onto F6

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	0.4877	0.0000	0.0000	
4				
5	Sum LnLk	-61.11	=SUM(M3:M94)	
6	Sum LnLk1	-61.11	Sum #1: Manual	
7	Sum Ln Lk2		Sum #2: Solver MLE	
8	Chi-Sq	122.23	=-2*(F6-F7)	
9	P-Value	2E-28	=CHISQ.DIST.RT(F8,1)	

To add Solver to the Excel Data menu in the Analysis section:
1) Select File, Options and Add-Ins. Select "Solver Add-in".
2) Under "Manage" menu, select "Excel Add-ins" Press GO.
3) Insert Check in "Solver Add-In" check box. Press OK.

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 9

2a) Solve for Slope and Intercept From Data menu, select Solver



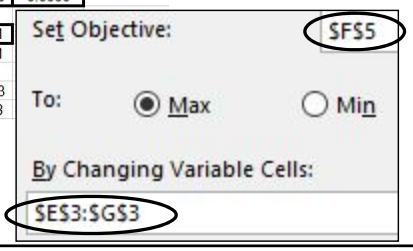
E	F	G	H	I	J	K	L	M
Intercept	**SlopeHt**	**SlopeWt**		**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
				0.49	1.63	0.62	0.38	-0.97
Sum LnLk	-61.11	=SUM(M3:M94)		0.49	1.63	0.62	0.38	-0.97
Sum LnLk1	-61.11	Sum #1: Manual		0.49	1.63	0.62	0.38	-0.97
Sum Ln Lk2		Sum #2: Solver MLE		0.49	1.63	0.62	0.38	-0.97
Chi-Sq	122.23	=-2*(F6-F7)		0.49	1.63	0.62	0.38	-0.97
P-Value	2E-28	=CHISQ.DIST.RT(F8,1)		0.49	1.63	0.62	0.38	-0.97

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 10

2b) Set Solver Parameters. GRC Nonlinear. Press Solve

Select Objective Cell (F5) and Variable Cells (E3:G3)

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	0.4877	0.0000	0.0000	
4				
5	Sum LnLk	-61.11		
6	Sum LnLk1	-61.11		
7	Sum Ln Lk2			
8	Chi-Sq	122.23		
9	P-Value	2E-28		

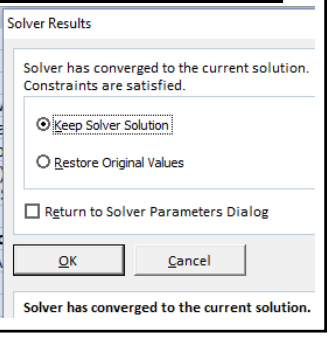


"GRC Non-Linear" is default method

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 11

2c) Results: All constraints & conditions satisfied. Press OK

E	F	G
Intercept	**SlopeHt**	**SlopeWt**
-41.3971	0.3817	0.1146
Sum LnLk	-23.45	=SUM(M3:M
Sum LnLk1	-61.11	Sum #1: Ma
Sum Ln Lk2		Sum #2: So
Chi-Sq	122.23	=-2*(F6-F7)
P-Value	2E-28	=CHISQ.DI
FORMULAS & TEXT: Enter, Co		
Logit	I3	=E$3+F$3*A
Odds	J3	=EXP(I3)
Prob Y=1	K3	=J3/(1+J3)



V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 12

3a) Copy Value of F5 onto F7

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	-41.3971	0.3817	0.1146	
4				
5	Sum LnLk	-23.45	=SUM(M3:M94)	
6	Sum LnLk1	-61.11	Sum #1: Manual	
7	Sum Ln Lk2	-23.45	Sum #2: Solver MLE	
8	Chi-Sq	75.33	=-2*(F6-F7)	
9	P-Value	4E-18	=CHISQ.DIST.RT(F8,1)	

Ready for test of null hypothesis: Slopes are zero

V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 13

3b) Hypothesis test: Slopes statistically significant?

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

Sum Ln Lk2	-23.45	Sum #2: Solver MLE
Chi-Sq	75.33	=-2*(F6-F7)
P-Value	4E-18	=CHISQ.DIST.RT(F8,1)

Slopes are statistically significant:
P-value < 0.05

Note: 4E-18 means move the decimal point 18 places to the left: 0.000000000000000004

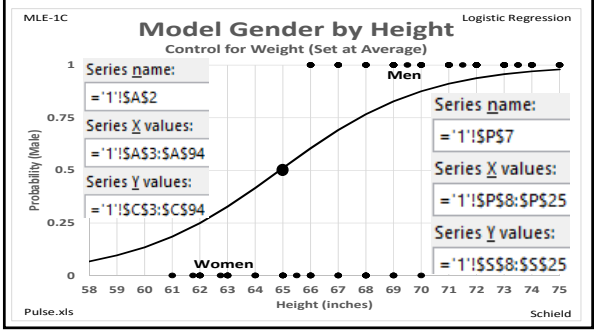
V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 14

4a) Setup Q8:Q25 and R5. Enter formula R8:T8. Pull down

	P	Q	R	S	T	U
2	Q8	Enter X manually		R8	=E$3+F$3*P6+G$3*Z$6	
3	S8	=EXP(Q6)		T8	=R6/(1+R6)	
4						
5	Wt-Average		145.15	=AVERAGE(B3:B94)		
6						
7		X-Ht	Logit	Odds	Prob Y=1	
8		58.00	-2.63	0.07	7%	
9		59.00	-2.25	0.11	10%	
10		60.00	-1.87	0.15	13%	
11		61.00	-1.48	0.23	18%	

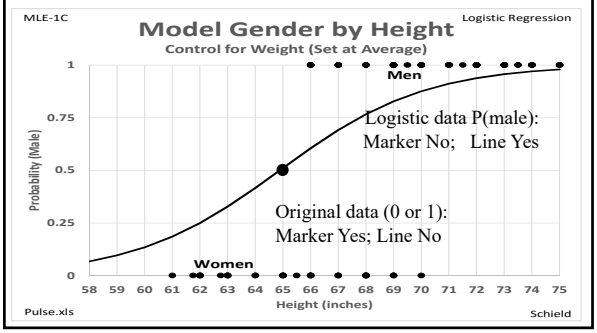
V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 15

4b) Graph Logistic Regression of Gender by Height.



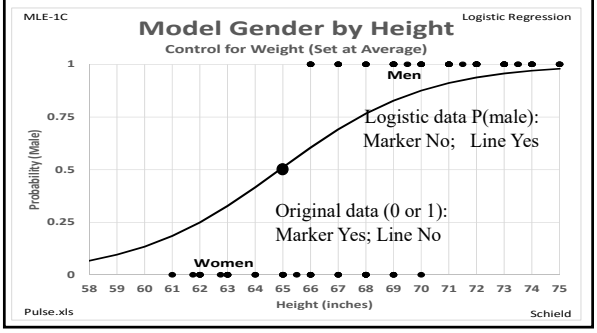
V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 16

4c) Graph Logistic Regression Format Data Series



V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 17

4c) Graph Logistic Regression Format Data Series



V0D 2015 Schield Logistic MLE 1C Excel2013 Slides 18

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 5.

REFERENCE:

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

Logistic Regression: MLE with 2 inputs, 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/2015-Schield-Logistic-MLE1C-Excel2013-Demo.pdf
pdf/2015-Schield-Logistic-MLE1C-Excel2013-Slides.pdf
xls/2015-Schield-Logistic-MLE1C-Excel2013-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 graph on slide 17.
Determine if slopes are statistically significant.

Use Height (A) and Weight (B) to predict Gender (col C)

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

Ave Heights:

M: 70.75” 62%

F: 65.3” 38%

Difference:

5.35”

A	B
Height	Male
68	1
69	1
69	1
72	1
66	1
67	0
71	1
71	1
71.5	1
62	0
65.5	0

Outline of Approach: Four Steps

1) Prepare data for logistic MLE regression
   Insert desired intercept
2) Use Solver to solve for intercept and slopes
3) Test for statistical significance
4) Generate graph

To do: Get data at www.StatLit.org/Excel/ 2015-Schield-Logistic-MLE1C-Excel2013-Data.xlsx

1a) Load Data; Find Mean(Y). Set Intercept

Enter formula for F21 and F22.

19	E	F	G
20	**GENDER & INTERCEPT #1**		
21	Male-Pctg	0.62	=AVERAGE(C3:C94)
22	Intercept#1	0.4877	=LN(F21/(1-F21))

Copy value from F22; Paste in E3

Row	E	F	G
2	**Intercept**	**SlopeHt**	**SlopeWt**
3	0.4877	0.0000	0.0000

1b) Enter formula for I3-M3 Select I3:M3. Pull to row 94.

Row	E	F	G	H	I	J	K	L	M
2	**Intercept**	**SlopeHt**	**SlopeWt**		**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
3	0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
4									
5	Sum LnLk	-0.97	=SUM(M3:M94)						
6	Sum LnLk1		Sum #1: Manual						
7	Sum Ln Lk2		Sum #2: Solver MLE						
8	Chi-Sq	0.00	=-2*(F6-F7)						
9	P-Value	1	=CHISQ.DIST.RT(F8,1)						
10									
11	**FORMULAS & TEXT: Enter, Copy Down**								
12	Logit	I3	=E$3+F$3*A3+G$3*B3						
13	Odds	J3	=EXP(I3)						
14	Prob Y=1	K3	=J3/(1+J3)						
15	Prob OK	L3	=IF(C3=1,K3,1-K3)						
16	Ln-LH-OK	M3	=LN(L3)						

1c) Results are as expected. Probability of male = 0.62

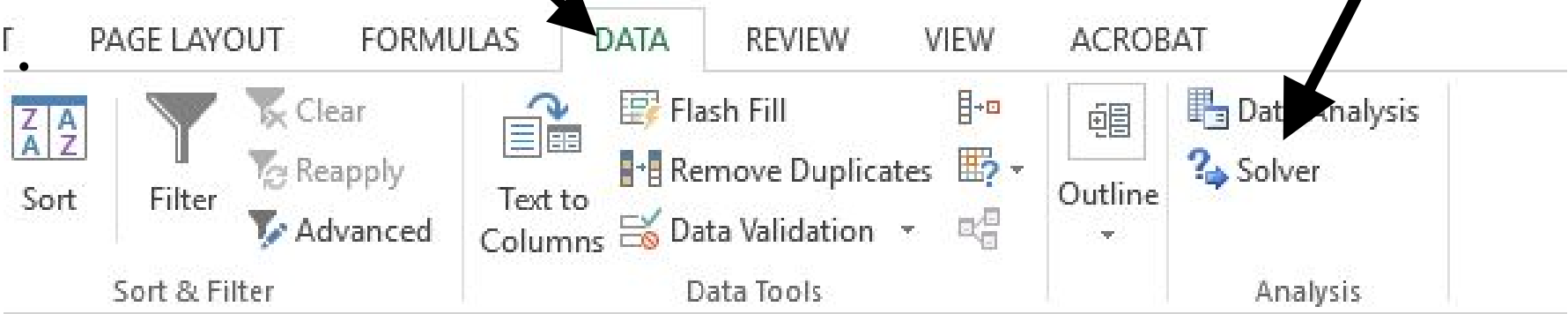
Row	E	F	G	H	I	J	K	L	M
2	**Intercept**	**SlopeHt**	**SlopeWt**		**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
3	0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
4					0.49	1.63	0.62	0.38	-0.97
5	Sum LnLk	-61.11	=SUM(M3:M94)		0.49	1.63	0.62	0.38	-0.97
6	Sum LnLk1		Sum #1: Manual		0.49	1.63	0.62	0.38	-0.97
7	Sum Ln Lk2		Sum #2: Solver MLE		0.49	1.63	0.62	0.38	-0.97
8	Chi-Sq	0.00	=-2*(F6-F7)		0.49	1.63	0.62	0.38	-0.97
9	P-Value	1	=CHISQ.DIST.RT(F8,1)		0.49	1.63	0.62	0.38	-0.97
10					0.49	1.63	0.62	0.38	-0.97
11	**FORMULAS & TEXT: Enter, Copy Down**				0.49	1.63	0.62	0.38	-0.97
12	Logit	I3	=E$3+F$3*A3+G$3*B3		0.49	1.63	0.62	0.38	-0.97
13	Odds	J3	=EXP(I3)		0.49	1.63	0.62	0.38	-0.97
14	Prob Y=1	K3	=J3/(1+J3)		0.49	1.63	0.62	0.38	-0.97
15	Prob OK	L3	=IF(C3=1,K3,1-K3)		0.49	1.63	0.62	0.38	-0.97
16	Ln-LH-OK	M3	=LN(L3)		0.49	1.63	0.62	0.38	-0.97
17					0.49	1.63	0.62	0.38	-0.97
18					0.49	1.63	0.62	0.38	-0.97

1d) Copy "Value" of F5 onto F6

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	0.4877	0.0000	0.0000	
4				
5	Sum LnLk	-61.11	=SUM(M3:M94)	
6	Sum LnLk1	-61.11	Sum #1: Manual	
7	Sum Ln Lk2		Sum #2: Solver MLE	
8	Chi-Sq	122.23	=-2*(F6-F7)	
9	P-Value	2E-28	=CHISQ.DIST.RT(F8,1)	

To add Solver to the Excel Data menu in the Analysis section:
1) Select File, Options and Add-Ins. Select "Solver Add-in".
2) Under "Manage" menu, select "Excel Add-ins" Press GO.
3) Insert Check in "Solver Add-In" check box. Press OK.

2a) Solve for Slope and Intercept From Data menu, select Solver

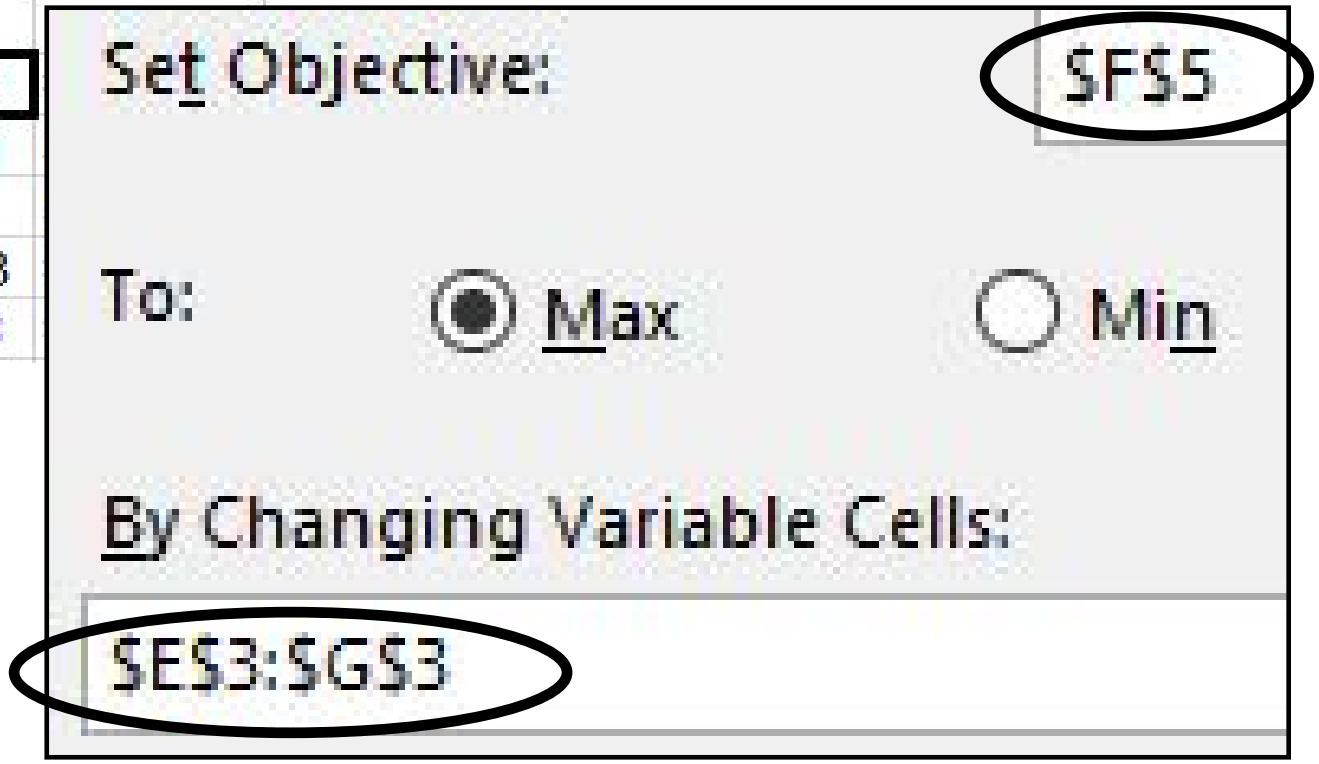


E	F	G	H	I	J	K	L	M
Intercept	**SlopeHt**	**SlopeWt**		**Logit**	**Odds**	**Prob Y=1**	**Prob OK**	**Ln-LH-OK**
0.4877	0.0000	0.0000		0.49	1.63	0.62	0.38	-0.97
				0.49	1.63	0.62	0.38	-0.97
Sum LnLk	-61.11	=SUM(M3:M94)		0.49	1.63	0.62	0.38	-0.97
Sum LnLk1	-61.11	Sum #1: Manual		0.49	1.63	0.62	0.38	-0.97
Sum Ln Lk2		Sum #2: Solver MLE		0.49	1.63	0.62	0.38	-0.97
Chi-Sq	122.23	=-2*(F6-F7)		0.49	1.63	0.62	0.38	-0.97
P-Value	2E-28	=CHISQ.DIST.RT(F8,1)		0.49	1.63	0.62	0.38	-0.97

2b) Set Solver Parameters. GRC Nonlinear. Press Solve

Select Objective Cell (F5) and Variable Cells (E3:G3)

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	0.4877	0.0000	0.0000	
4				
5	Sum LnLk	-61.11		
6	Sum LnLk1	-61.11		
7	Sum Ln Lk2			
8	Chi-Sq	122.23		
9	P-Value	2E-28		



"GRC Non-Linear" is default method

2c) Results: All constraints & conditions satisfied. Press OK

E	F	G
Intercept	**SlopeHt**	**SlopeWt**
-41.3971	0.3817	0.1146
Sum LnLk	-23.45	=SUM(M3:M
Sum LnLk1	-61.11	Sum #1: Ma
Sum Ln Lk2		Sum #2: So
Chi-Sq	122.23	=-2*(F6-F7)
P-Value	2E-28	=CHISQ.DI
FORMULAS & TEXT: Enter, Co		
Logit	I3	=E$3+F$3*A
Odds	J3	=EXP(I3)
Prob Y=1	K3	=J3/(1+J3)

Solver Results

Solver has converged to the current solution.
Constraints are satisfied.

(•) Keep Solver Solution

( ) Restore Original Values

[ ] Return to Solver Parameters Dialog

OK Cancel

Solver has converged to the current solution.

3a) Copy Value of F5 onto F7

Row	E	F	G	H
2	**Intercept**	**SlopeHt**	**SlopeWt**	
3	-41.3971	0.3817	0.1146	
4				
5	Sum LnLk	-23.45	=SUM(M3:M94)	
6	Sum LnLk1	-61.11	Sum #1: Manual	
7	Sum Ln Lk2	-23.45	Sum #2: Solver MLE	
8	Chi-Sq	75.33	=-2*(F6-F7)	
9	P-Value	4E-18	=CHISQ.DIST.RT(F8,1)	

Ready for test of null hypothesis: Slopes are zero

3b) Hypothesis test: Slopes statistically significant?

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

Sum Ln Lk2	-23.45	Sum #2: Solver MLE
Chi-Sq	75.33	=-2*(F6-F7)
P-Value	4E-18	=CHISQ.DIST.RT(F8,1)

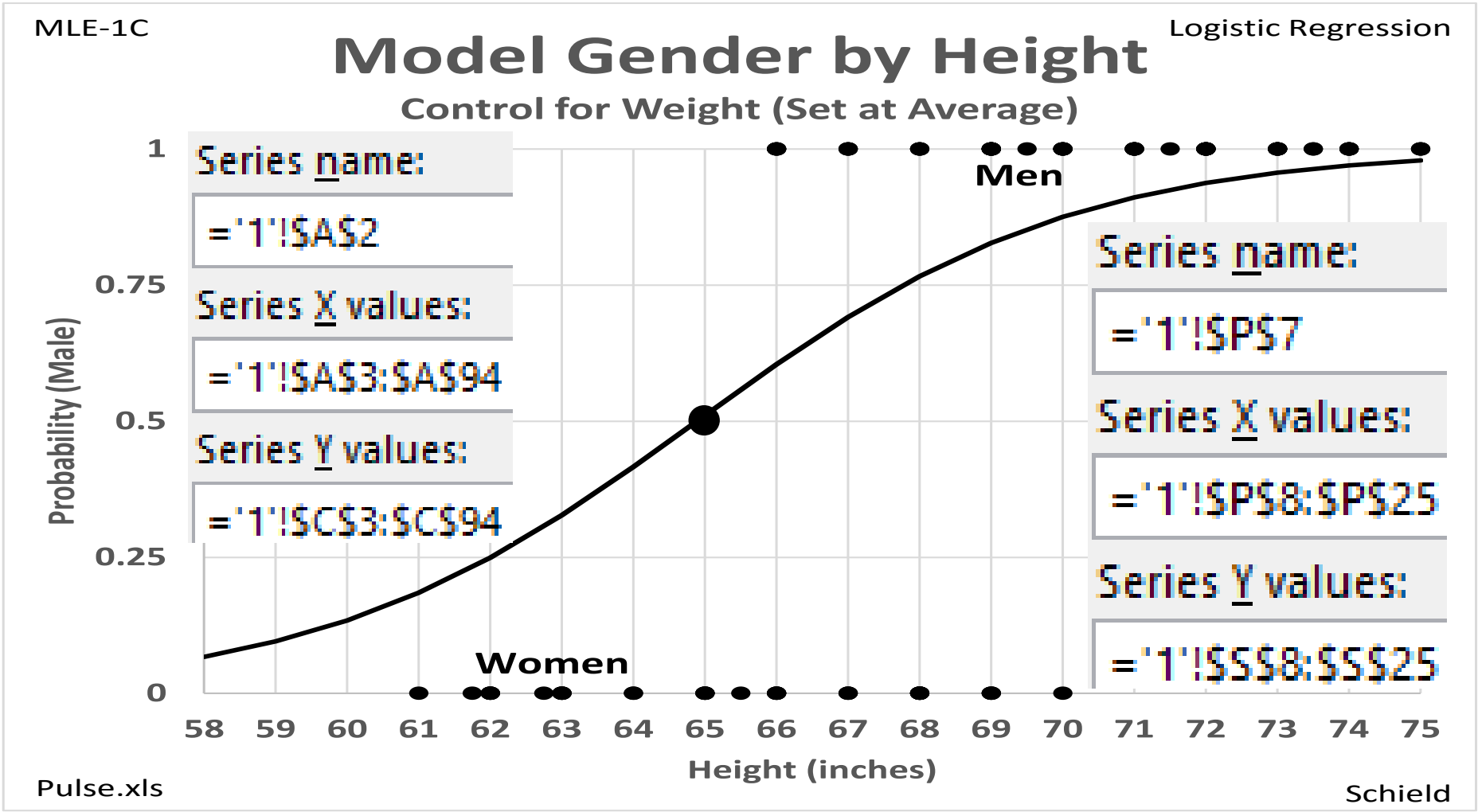
Slopes are statistically significant:
P-value < 0.05

Note: 4E-18 means move the decimal point 18 places to the left: 0.000000000000000004

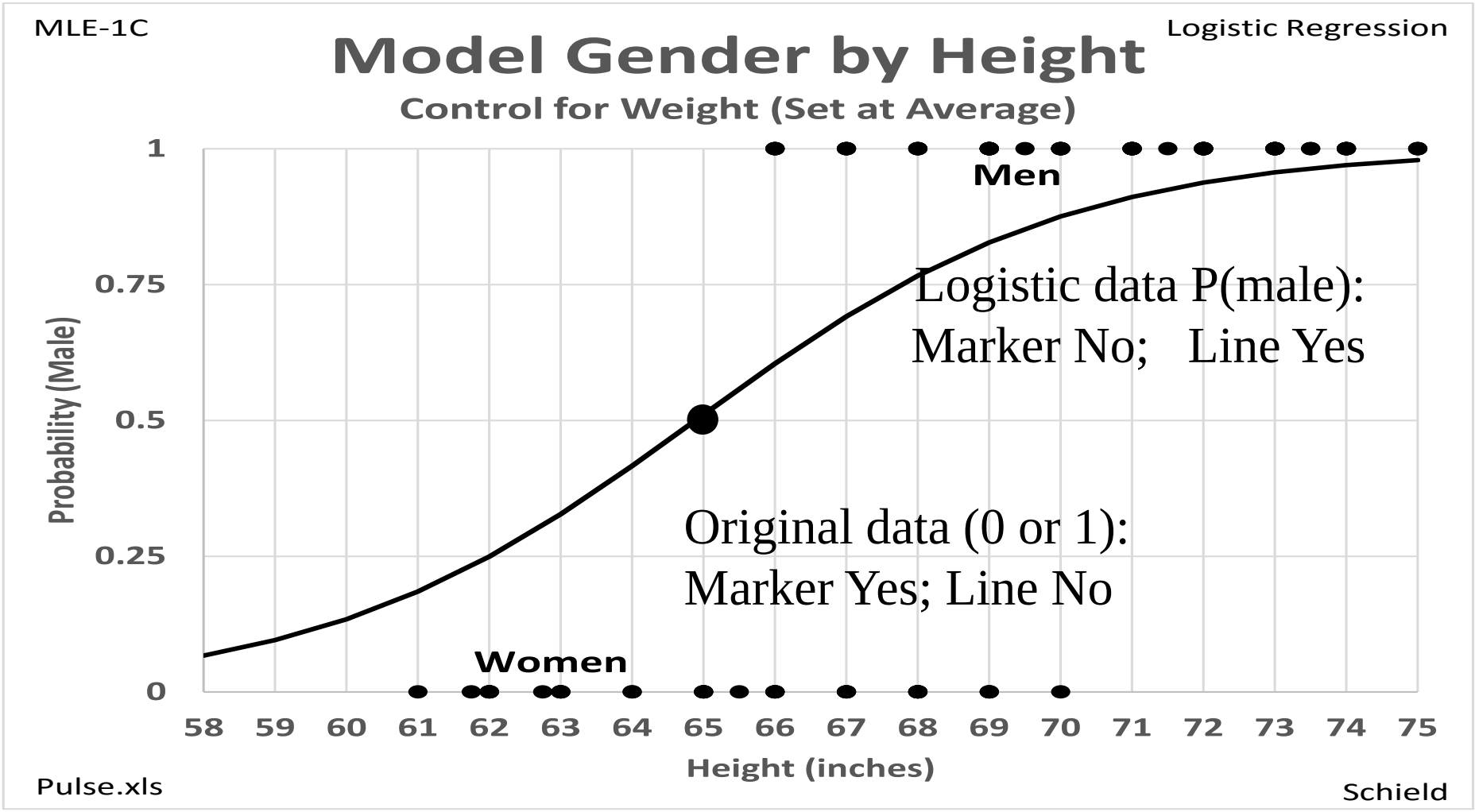
4a) Setup Q8:Q25 and R5. Enter formula R8:T8. Pull down

O	P	Q	R	S	T	U
2	Q8	Enter X manually		R8	=E$3+F$3*P6+G$3*Z$6	
3	S8	=EXP(Q6)		T8	=R6/(1+R6)	
4						
5	Wt-Average		145.15	=AVERAGE(B3:B94)		
6						
7		**X-Ht**	**Logit**	**Odds**	**Prob Y=1**	
8		58.00	-2.63	0.07	7%	
9		59.00	-2.25	0.11	10%	
10		60.00	-1.87	0.15	13%	
11		61.00	-1.48	0.23	18%	

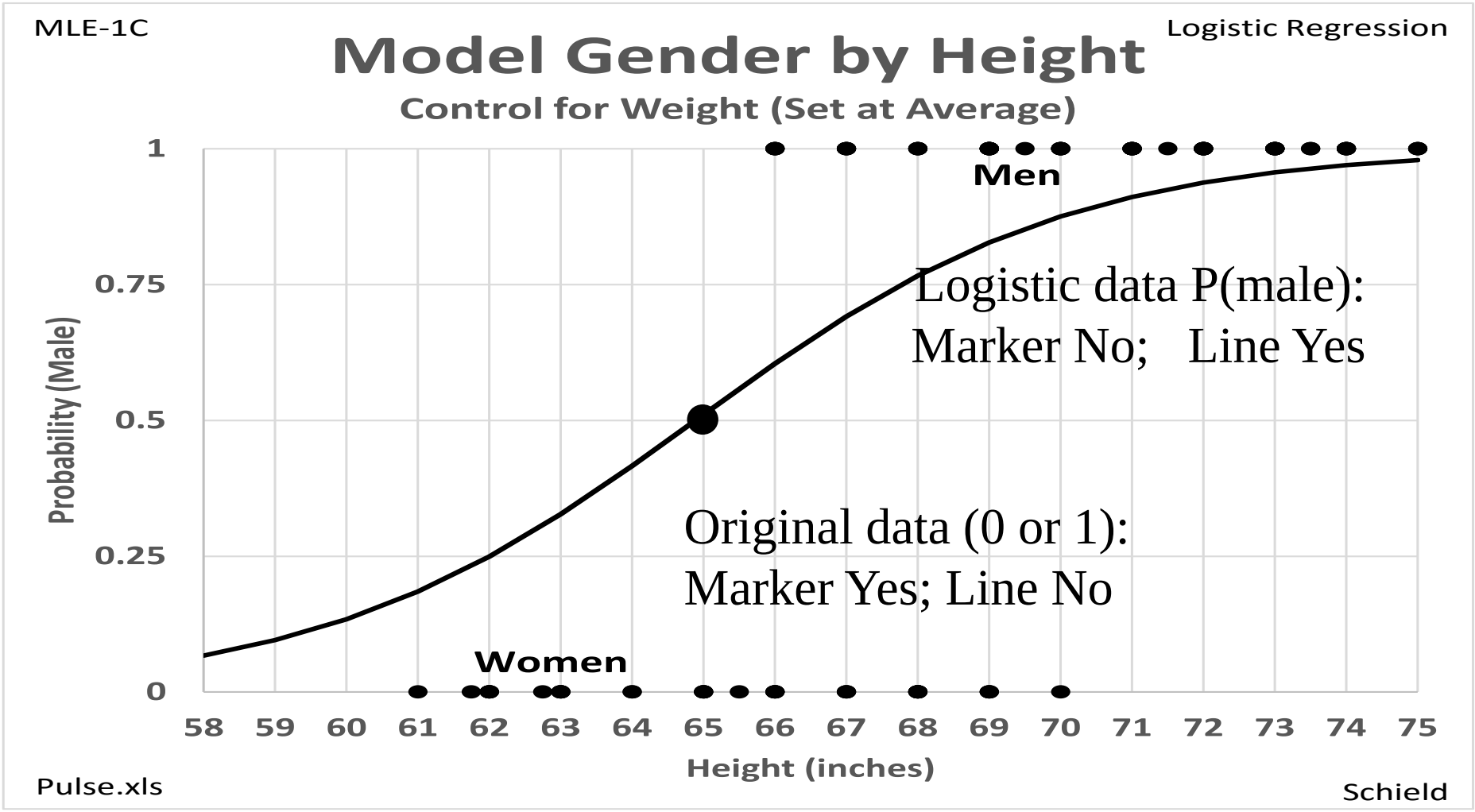
4b) Graph Logistic Regression of Gender by Height.



4c) Graph Logistic Regression Format Data Series



4c) Graph Logistic Regression Format Data Series



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 5.

REFERENCE:

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