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
Excel/2015-Schield-Logistic-OLS1D-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-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).

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

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

	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			

	A	B	C
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

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

	A	B	C	D	E
1					
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	

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

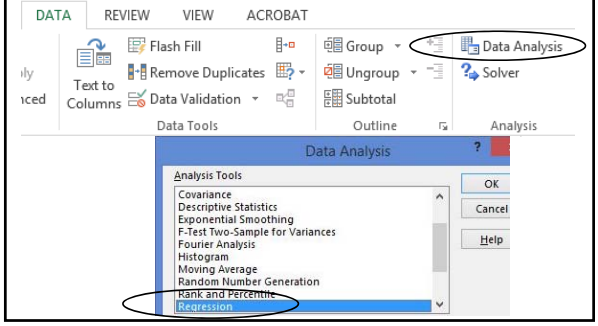
	B	C	D	E	F	G
	ct chance of being male given height after controllir					
	D7	=IF(C7=0, 0.001, 0.999)			E7	=LN(D7/(1-D7))
ht	Weight	Male	Male1	LN(Odds)	yPred	6
	140	0	0.001	-6.91		7

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						
	D7	=IF(C7=0, 0.001, 0.999)		E7	=LN(D7/(1-D7))	
Height	Weight	Male	Male1	LN(Odds)	yPred	6
61	140	0	0.001	-6.91		7
61.75	108	0				8
62	108	0				9
62	110	0				10

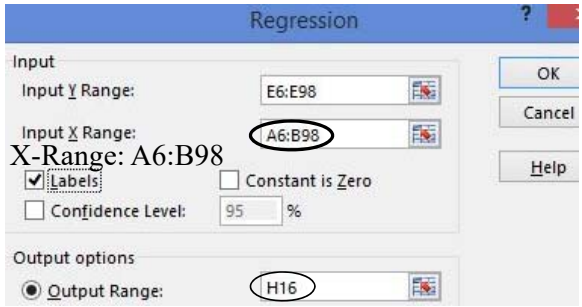
XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 7

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



XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 8

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



XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 9

2c) OLS1 Regression

	H	I	J	K	L
15					
16	SUMMARY OUTPUT				
17					
18	*Regression Statistics*				
19	Multiple R	0.75			
20	R Square	0.57			
21	Adjusted R S	0.56			
22	Standard Err	4.49			
23	Observation	92			

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
31		*Coefficients*	*Std Error*	*t Stat*	*P-value*
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

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 10

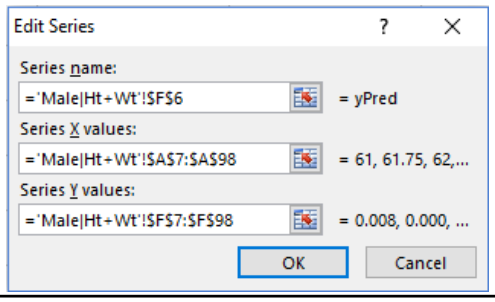
3a) Generate F7. Check value. Select; pull down to row 98.

A	B	C	D	E	F	G
Predict chance of being male given height after controllin						
	D7 =IF(C7=0, 0.001, 0.999)			E7 =LN(D7/(1-D7))		
	F7 =1/(1+EXP(-I$32-I$33*A7 -I$34*B7))					
Height	Weight	Male	Male1	LN(Odds)	yPred	6
61	140	0	0.001	-6.91	0.008	7
61.75	108	0	0.001	-6.91		8
62	108	0	0.001	-6.91		9
62	110	0	0.001	-6.91		10

XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 11

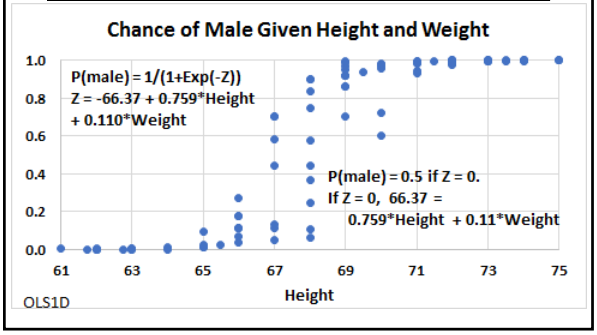
3b) Insert Chart (XY Plot): yPred vs. Height

X values: A7:A98. Y values: F7: F98



XL4D: V0H 2015 Schield Logistic Regression using OLS1D in Excel2013 12

3c) Chart #1 Results Add Title and textboxes



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)				
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						

R3 =1/(1+EXP(-I$32-I$33*Q3-I$34*130))
S3 =1/(1+EXP(-I$32-I$33*Q3-I$34*150))

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

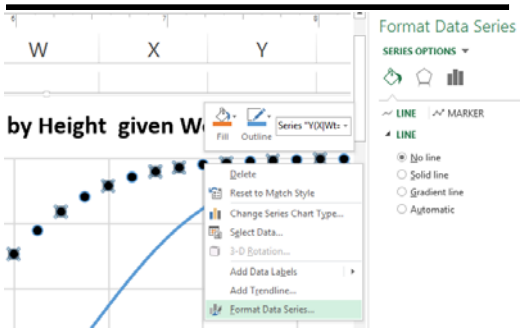
Name: R2
X values: Q3:Q31
Y values: R3:R31

Series name:
='Male|Ht+Wt'!$R$2
Series X values:
='Male|Ht+Wt'!$Q$3:$Q$31
Series Y values:
='Male|Ht+Wt'!$R$3:$R31

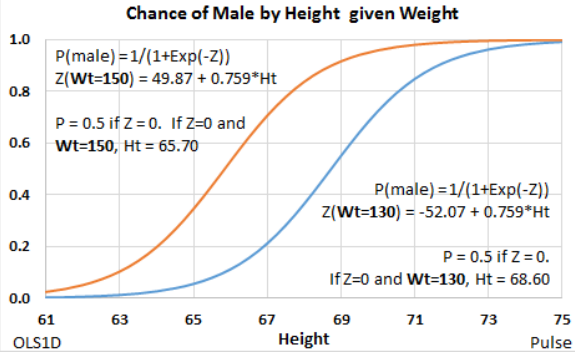
Name: S2
X values: Q3:Q31
Y values: S3:S31.

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

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



4d) Final Result: Title & boxes



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*Ht + 0.11*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

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
Excel/2015-Schield-Logistic-OLS1D-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-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).

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			

	A	B	C
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

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

	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	

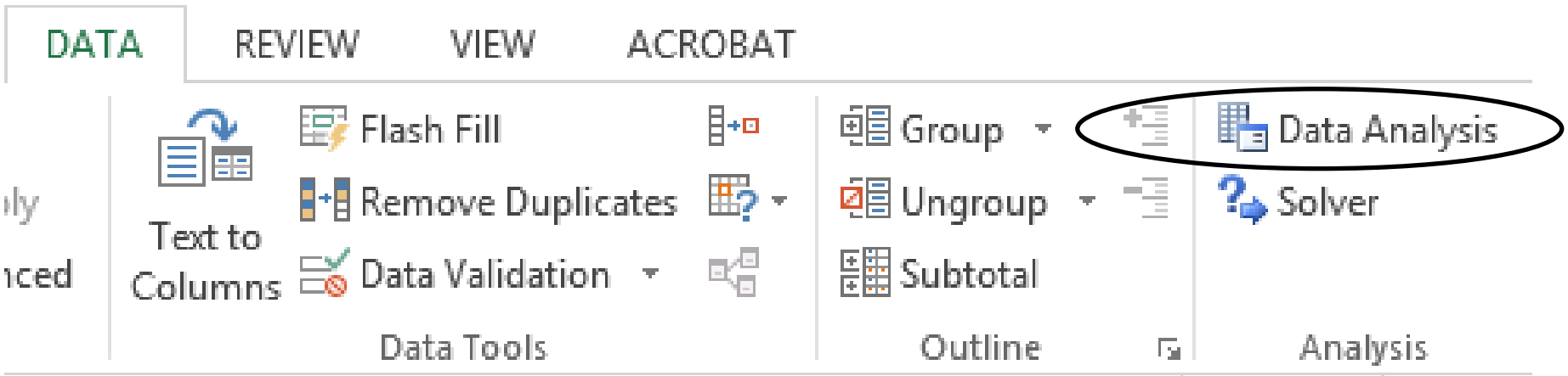
1b) Generate Ln[Odds(Male1)] in E7

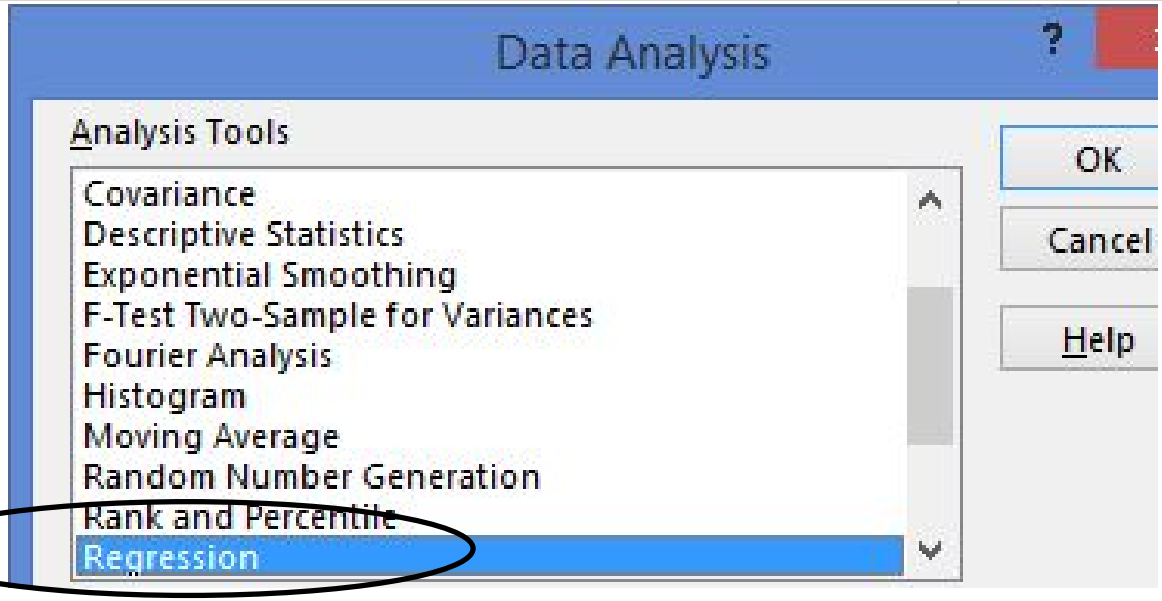
	B	C	D	E	F	G
ct chance of being male given height after controllin						
D7	=IF(C7=0, 0.001, 0.999)			E7	=LN(D7/(1-D7))	
ht	Weight	Male	Male1	LN(Odds)	yPred	6
	140	0	0.001	-6.91		7

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

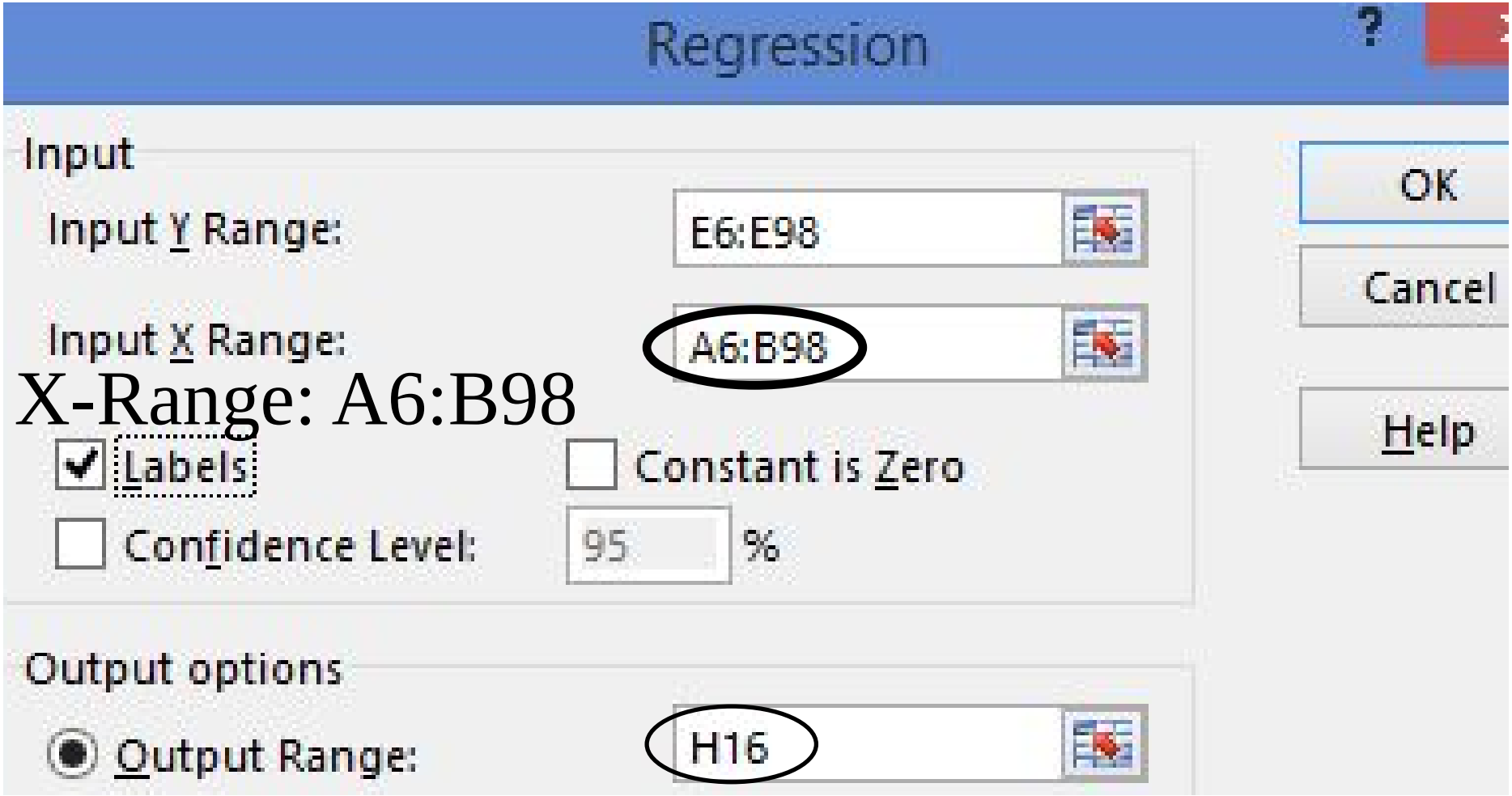
A	B	C	D	E	F	G
Predict chance of being male given height after controlling f						
	D7 =IF(C7=0, 0.001, 0.999)			E7 =LN(D7/(1-D7))		
Height	Weight	Male	Male1	LN(Odds)	yPred	6
61	140	0	0.001	-6.91		7
61.75	108	0				8
62	108	0				9
62	110	0				10

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





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



2c) OLS1 Regression

15	H	I	J	K	L
16	SUMMARY OUTPUT				
17					
18	*Regression Statistics*				
19	Multiple R	0.75			
20	R Square	0.57			
21	Adjusted R S	0.56			
22	Standard Err	4.49			
23	Observation	92			

Main source of error:

No mention of Weight in H34.

Double-check H34!

To fix, redo X-range in slide 8.

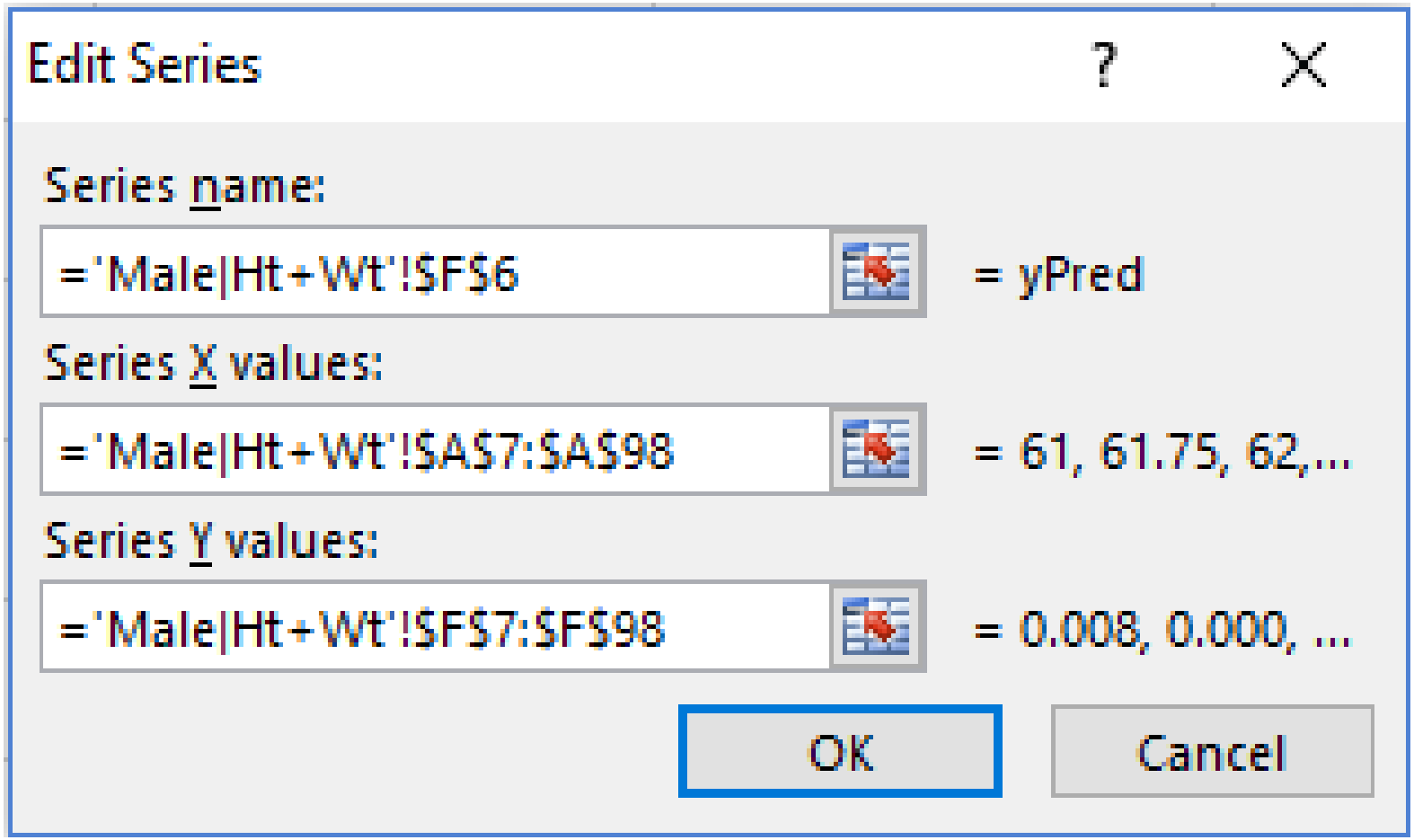
	H	I	J	K	L
31		*Coefficients*	*Std Error*	*t Stat*	*P-value*
32	Intercept	-66.37	11.00	-6.03	0.00
33	Height	0.7586	0.21	3.66	0.00
34	Weight	0.1095	0.03	3.43	0.00

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

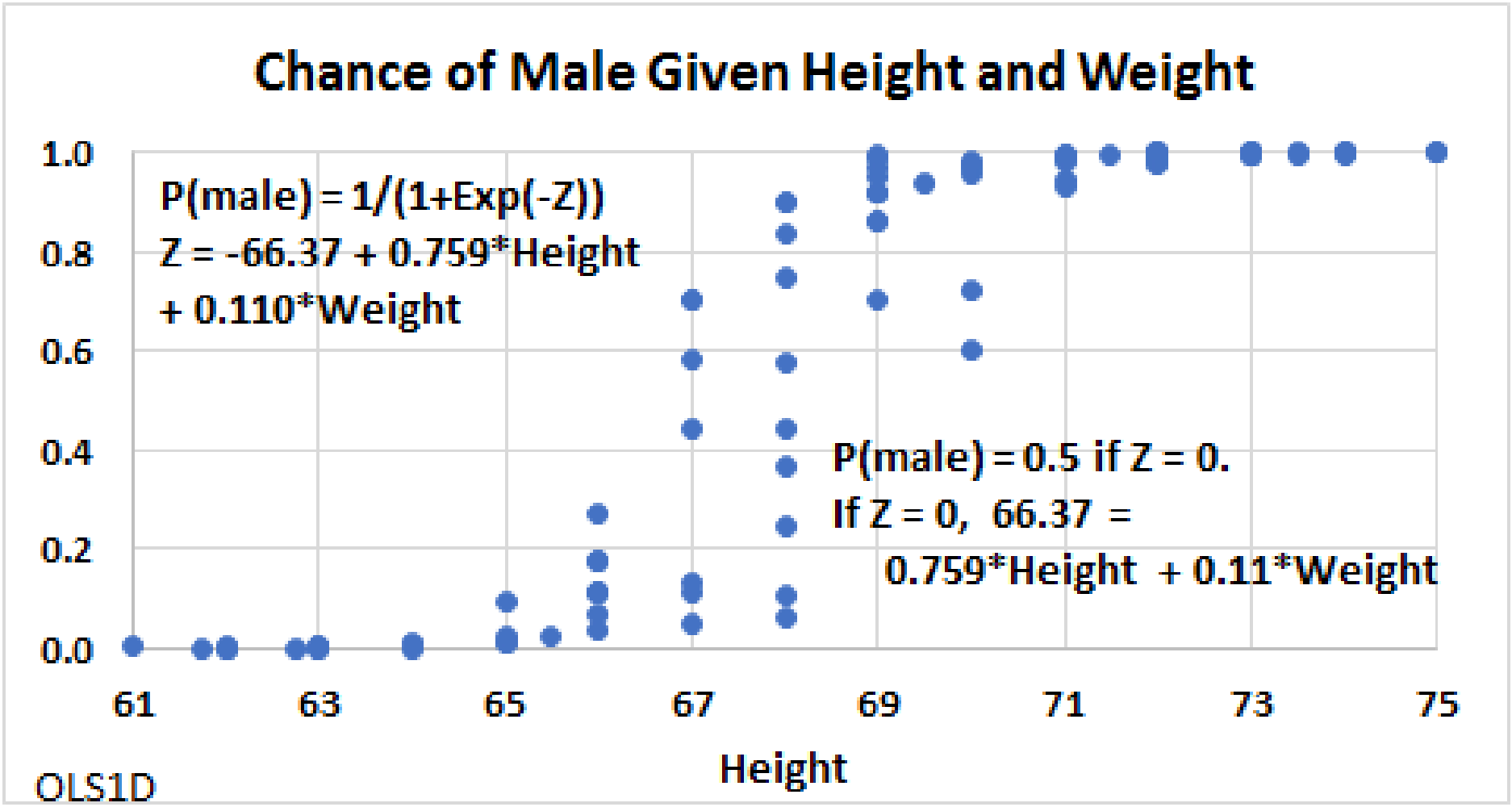
A	B	C	D	E	F	G
Predict chance of being male given height after controllin						
	D7 =IF(C7=0, 0.001, 0.999)			E7 =LN(D7/(1-D7))		
	F7 =1/(1+EXP(-I$32-I$33*A7 -I$34*B7))					
Height	Weight	Male	Male1	LN(Odds)	yPred	6
61	140	0	0.001	-6.91	0.008	7
61.75	108	0	0.001	-6.91		8
62	108	0	0.001	-6.91		9
62	110	0	0.001	-6.91		10

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

X values: A7:A98. Y values: F7: F98



3c) Chart #1 Results
Add Title and textboxes



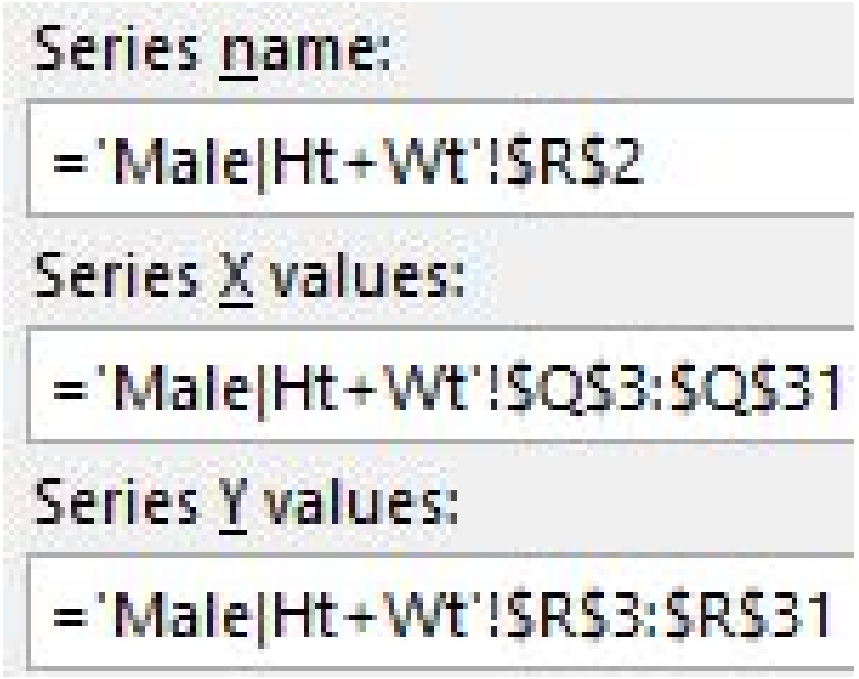
4a) Enter formula in R3 & S3 Pull R3:S3 down to Row 31

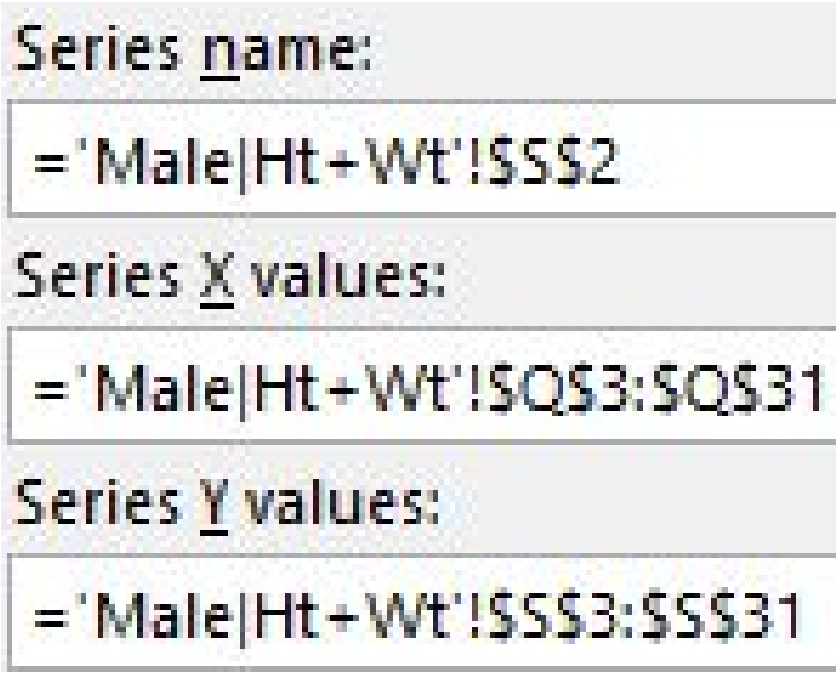
P	Q	R	S	T	U	V	W
Row	X	Y(X|Wt=130)	Y(X|Wt=150)	R3	=1/(1+EXP(-I$32-I$33*Q3-I$34*130))		
3	61	0.003	0.025	S3	=1/(1+EXP(-I$32-I$33*Q3-I$34*150))		
4	61.5						
5	62						
6	62.5						
7	63						
8	63.5						
9	64						
10	64.5						
11	65						
12	65.5						
13	66						
14	66.5						
15	67						

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

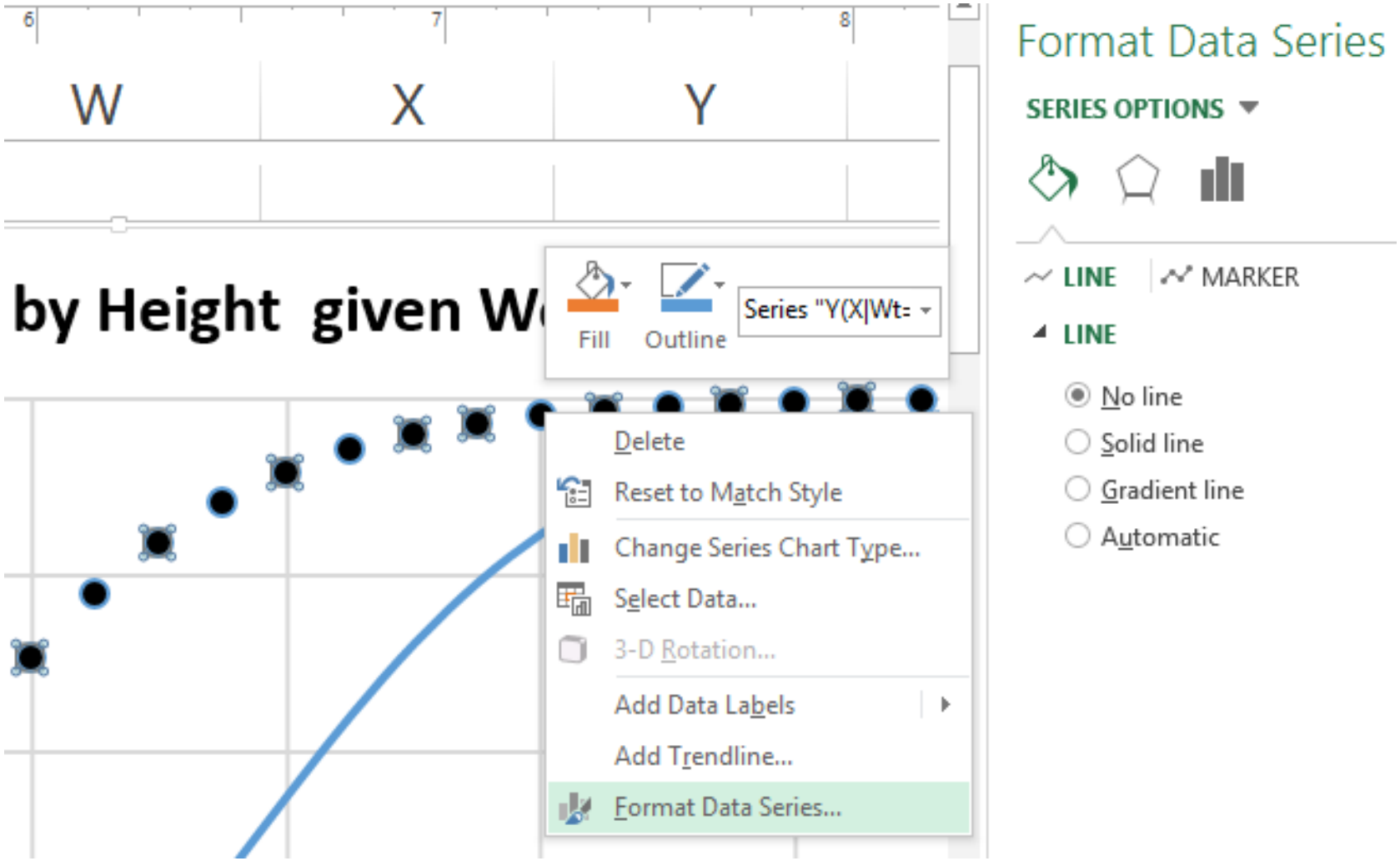
Name: R2
X values: Q3:Q31
Y values: R3:R31

Name: S2
X values: Q3:Q31
Y values: S3:S31.

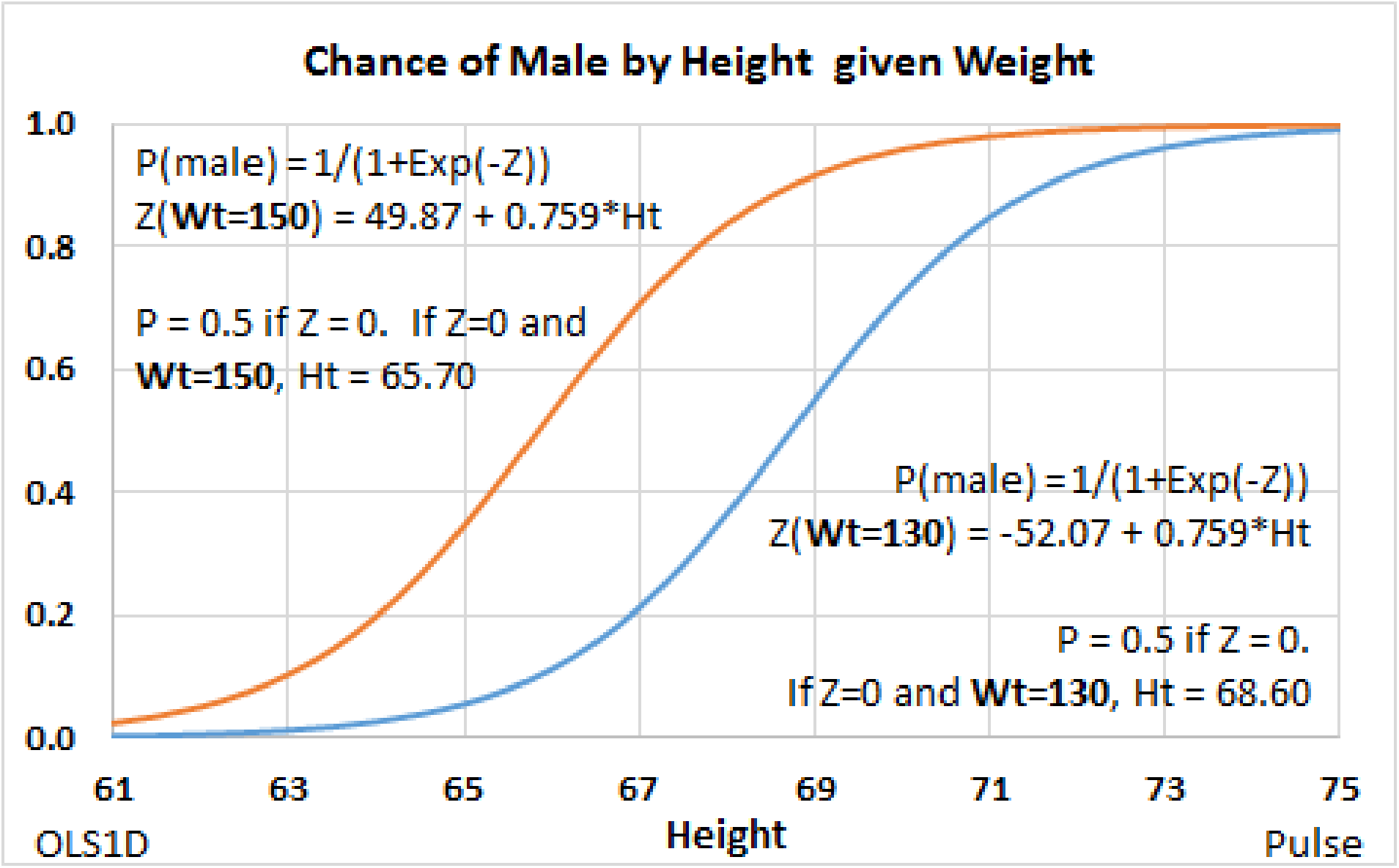




4c) Format Data Series
Paint: No marker; Solid line



4d) Final Result: Title & boxes



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*Ht + 0.11*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