Logistic Regression using OLS1A 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-OLS1A-Excel2013-Slides.pdf
pdf/2015-Schield-Logistic-OLS1A-Excel2013-Demo.pdf
Excel/2015-Schield-Logistic-OLS1A-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: OLS1. By slightly adjusting the binary outcomes, one can use OLS regression to solve for a good logistic model.

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

This demo uses Height (col A) to predict Gender (col B)

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

	A	B	C
6	Height	Male	Male1
7	61	0	
8	61.75	0	
9	62	0	
10	62	0	
11	62	0	
12	62	0	
13	62.75	0	
14	63	0	
15	63	0	
16	63	0	
17	63	0	

Data in rows 6 to 98

	A	B	C
24	65.5	0	
25	66	1	
26	66	0	
27	66	0	
28	66	0	
29	66	1	
30	66	1	
31	66	1	
32	66	0	
33	67	1	
34	67	1	

1) Nudge Binary Male to Eliminate Zero and One

A	B	C	D	E
Predict chance of being male given height.				
	C7	=IF(B7=0, 0.001, 0.999)		
Height	Male	Male1	Odds	LN(Odds)
61	0	0.001		

2) Enter formula for Odds in D7; LN[Odds(p)] in E7

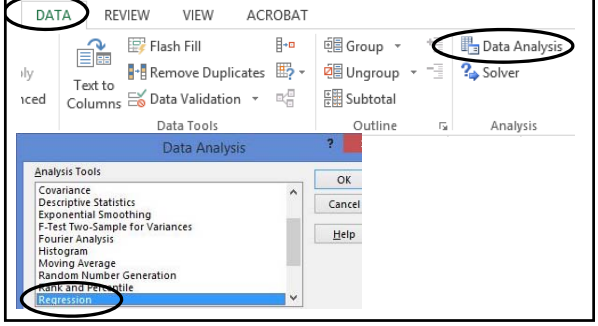
A	B	C	D	E	F	G
Predict chance of being male given height. Regress using						
	C7	=IF(B7=0, 0.001, 0.999)		E7	=LN(D7)	
	D7	=C7/(1-C7)				
Height	Male	Male1	Odds	LN(Odds)	yPred	6
61	0	0.001	0.001	-6.91		7
61.75	0					8
62	0					9

3) Select C7:E7 Drag to bottom of data: Row 98

A	B	C	D	E	F	G
Predict chance of being male given height. Regress using						
	C7	=IF(B7=0, 0.001, 0.999)		E7	=LN(D7)	
	D7	=C7/(1-C7)				
Height	Male	Male1	Odds	LN(Odds)	yPred	6
61	0	0.001	0.001	-6.91		7
61.75	0					8
62	0					9

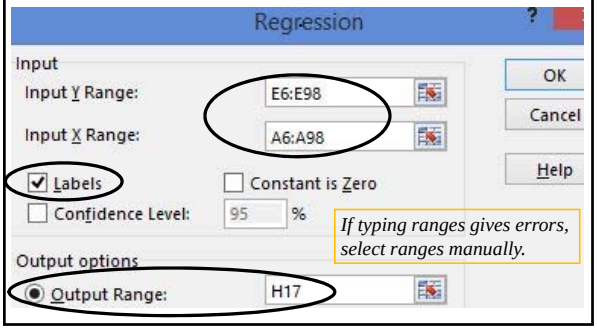
XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 7

A) From Data Bar, Select Data Analysis; Regression



XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 8

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



XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 9

C) Logistic Regression: Results Using OLS1A

16	H	I	J	K	L	M	N
17	SUMMARY OUTPUT						
18							
19	*Regression Statistics*						
20	Multiple R	0.71					
21	R Square	0.51					
22	Adjusted R Square	0.50					
23	Standard Error	4.75					
24	Observations	92					
25							
26	ANOVA						
27		*df*	*SS*	*MS*	*F*	*Significance F*	
28	Regression	1	2,111.07	2,111.07	93.75	0.00	
29	Residual	90	2,026.67	22.52			
30	Total	91	4,137.74				
31							
32		*Coefficients*	*tandard Errc*	*t Stat*	*P-value*	*Lower 95%*	*Jpper 95%*
33	Intercept	-88.80	9.35	-9.49	0.00	-107.38	-70.21
34	Height	1.32	0.14	9.68	0.00	1.05	1.59

Check to see that you get the same results in the boxes. Formatting is optional

XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 10

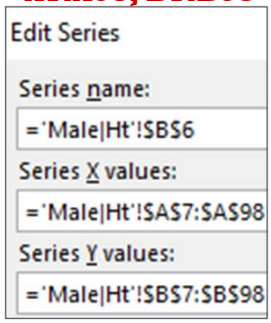
D) Generate F7: Pull F7 down to F98

D	E	F	G
; male given height. Regress using a logistic			
1, 0.999)	E7 =LN(D7)		
	F7 =1/(1+EXP(-I$33-I$34*A7))		
Odds	LN(Odds)	yPred	6
0.001	-6.91	0.000	7

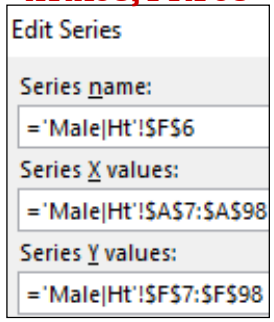
XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 11

E) Insert XY Plot. Add Two Series.

Male vs Height | A7:A98, B7:B98

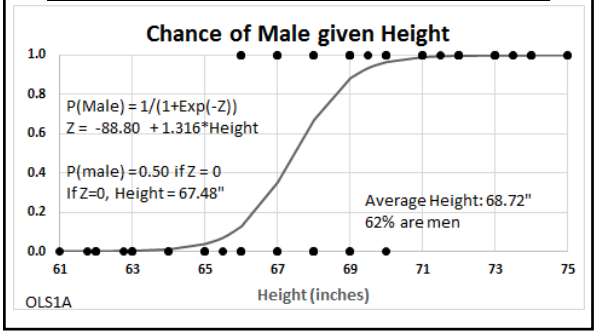


yPred vs Height | A7:A98, F7:F98



XL4A: V0E2x 2015 Schield Logistic Regression using OLS1A in Excel2013 12

E) Insert Titles & Textboxes. Format yPred with Solid Line



Logistic Regression using OLS1A 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-OLS1A-Excel2013-Slides.pdf
pdf/2015-Schield-Logistic-OLS1A-Excel2013-Demo.pdf
Excel/2015-Schield-Logistic-OLS1A-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: OLS1. By slightly adjusting the binary outcomes, one can use OLS regression to solve for a good logistic model.

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

This demo uses Height (col A) to predict Gender (col B)

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

	A	B	C
6	Height	Male	Male1
7	61	0	
8	61.75	0	
9	62	0	
10	62	0	
11	62	0	
12	62	0	
13	62.75	0	
14	63	0	
15	63	0	
16	63	0	
17	63	0	

Data in rows 6 to 98

	A	B	C
24	65.5	0	
25	66	1	
26	66	0	
27	66	0	
28	66	0	
29	66	1	
30	66	1	
31	66	1	
32	66	0	
33	67	1	
34	67	1	

1) Nudge Binary Male to Eliminate Zero and One

A	B	C	D	E
Predict chance of being male given height. F				
	C7	=IF(B7=0, 0.001, 0.999)		
Height	Male	Male1	Odds	LN(Odds)
61	0	0.001		

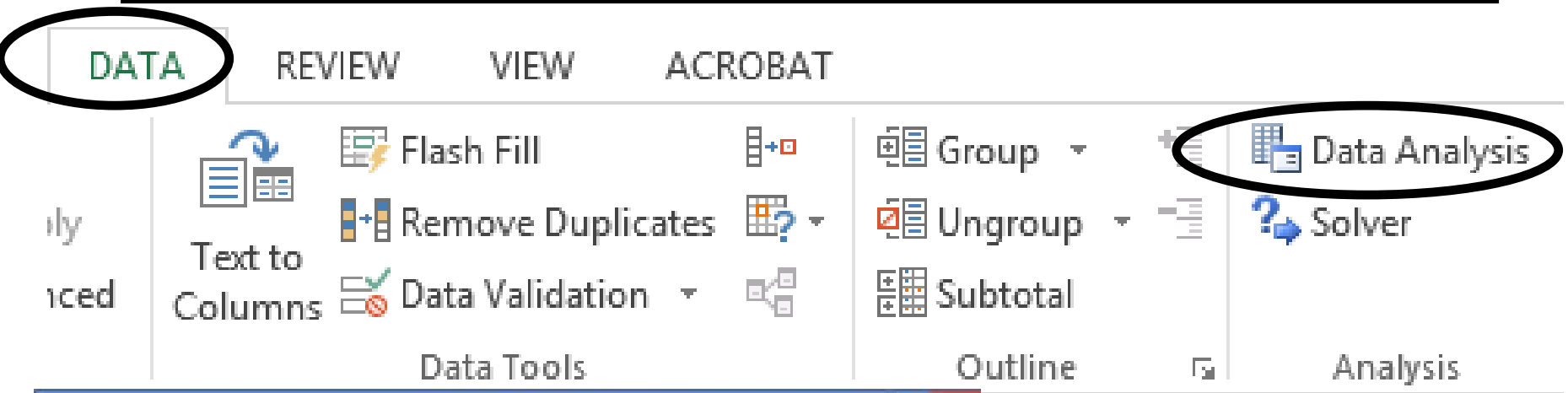
2) Enter formula for Odds in D7; LN[Odds(p)] in E7

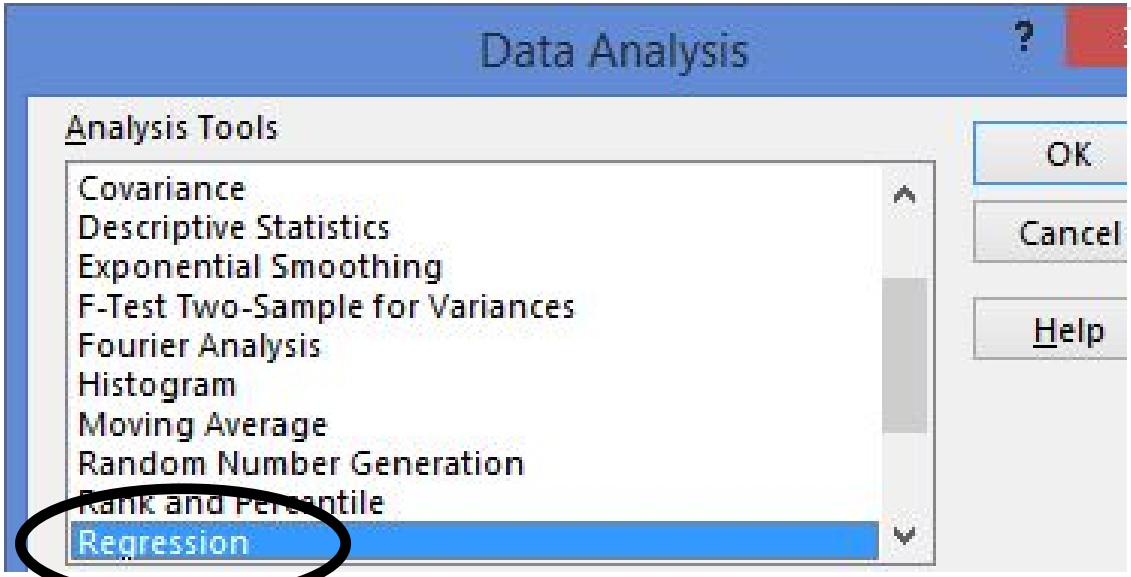
A	B	C	D	E	F	G
Predict chance of being male given height. Regress using						
C7	=IF(B7=0, 0.001, 0.999)			E7	=LN(D7)	
D7	=C7/(1-C7)					
Height	Male	Male1	Odds	LN(Odds)	yPred	6
61	0	0.001	0.001	-6.91		7
61.75	0					8
62	0					9

3) Select C7:E7
Drag to bottom of data: Row 98

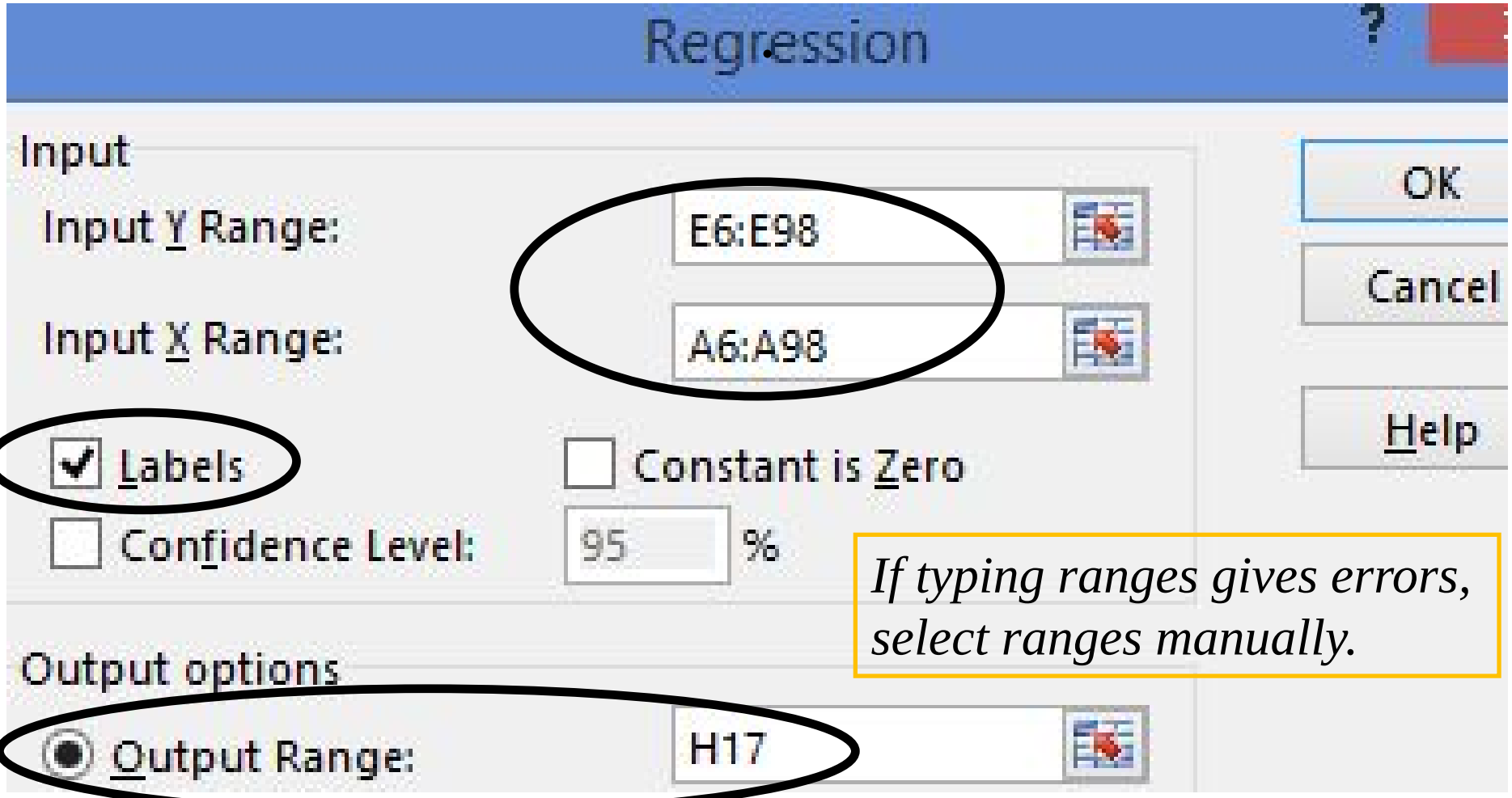
A	B	C	D	E	F	G
Predict chance of being male given height. Regress using						
	C7	=IF(B7=0, 0.001, 0.999)		E7	=LN(D7)	
	D7	=C7/(1-C7)				
Height	Male	Male1	Odds	LN(Odds)	yPred	6
61	0	0.001	0.001	-6.91		7
61.75	0					8
62	0					9

A) From Data Bar, Select Data Analysis; Regression





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



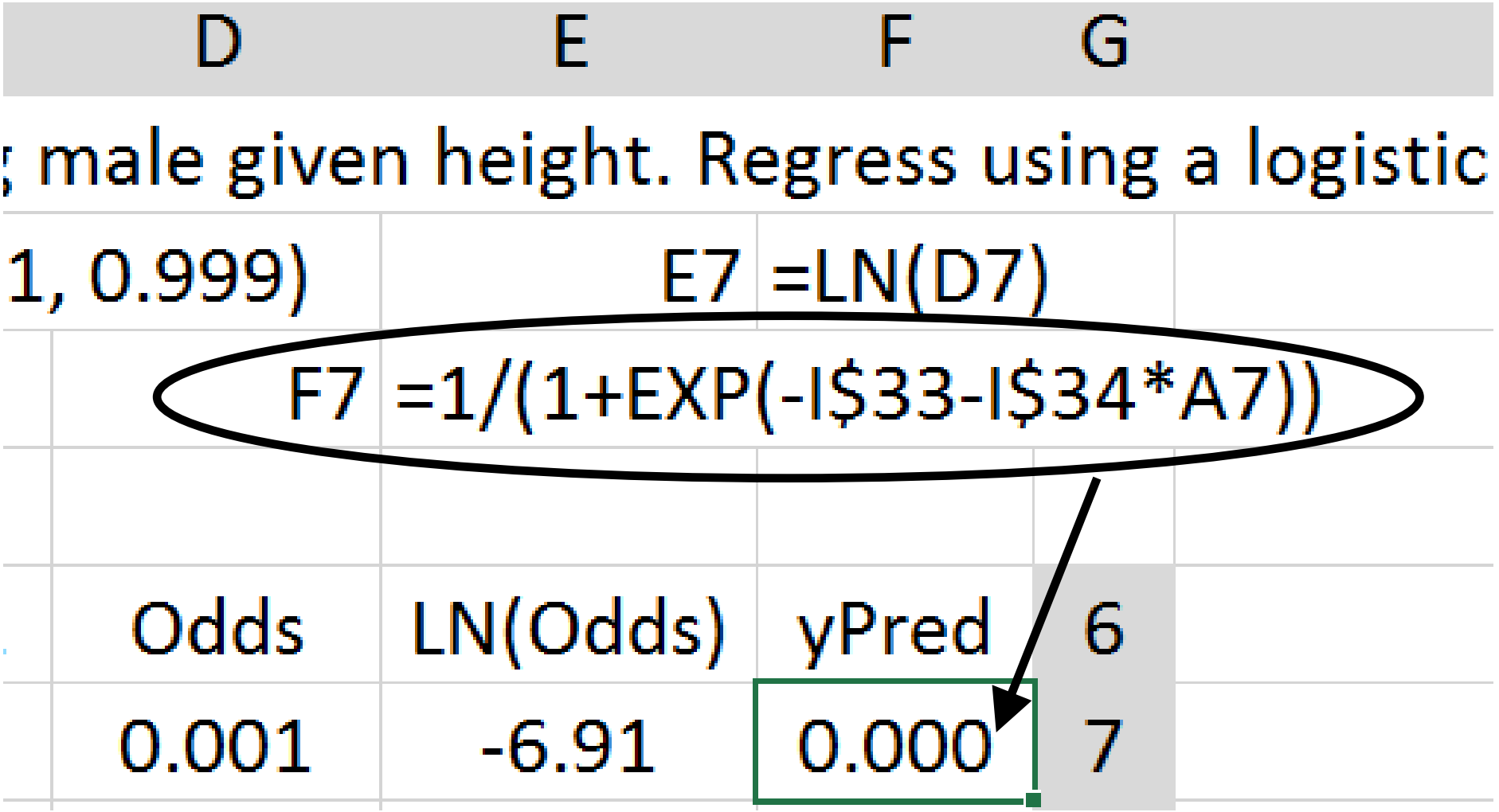
If typing ranges gives errors, select ranges manually.

C) Logistic Regression: Results Using OLS1A

16	H	I	J	K	L	M	N
17	SUMMARY OUTPUT						
18							
19	*Regression Statistics*						
20	Multiple R	0.71					
21	R Square	0.51					
22	Adjusted R Square	0.50					
23	Standard Error	4.75					
24	Observations	92					
25							
26	ANOVA						
27		*df*	*SS*	*MS*	*F*	*Significance F*	
28	Regression	1	2,111.07	2,111.07	93.75	0.00	
29	Residual	90	2,026.67	22.52			
30	Total	91	4,137.74				
31							
32		*Coefficients*	*ndard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
33	Intercept	-88.80	9.35	-9.49	0.00	-107.38	-70.21
34	Height	1.32	0.14	9.68	0.00	1.05	1.59

Check to see that you get the same results in the boxes. Formatting is optional

D) Generate F7: Pull F7 down to F98



E) Insert XY Plot. Add Two Series.

Male vs Height | A7:A98, B7:B98

Edit Series

Series name:

='Male|Ht'!$B$6

Series X values:

='Male|Ht'!$A$7:$A$98

Series Y values:

='Male|Ht'!$B$7:$B$98

yPred vs Height | A7:A98, F7:F98

Edit Series

Series name:

='Male|Ht'!$F$6

Series X values:

='Male|Ht'!$A$7:$A$98

Series Y values:

='Male|Ht'!$F$7:$F$98

E) Insert Titles & Textboxes. Format yPred with Solid Line