XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 1

Logistic Model using OLS1: Gender vs. Height & Smoker

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/2017-Schield-Logistic-OLS1B-Excel2013-Slides.pdf
pdf/2017-Schield-Logistic-OLS1B-Excel2013-Demo.pdf
Excel/2017-Schield-Logistic-OLS1B-Excel2013-Data.xlsx

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 2

Background & Goals

Modelling a binary outcome (loan vs. no-loan) uses a logistic curve/model 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 approach uses a nudge approximation: OLS1. By adjusting the binary outcomes and using a logistic model, OLS regression generates a fairly good fit.

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

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 3

Goal: Predict Gender using Height & Smoker

Column B: Smoker = 1; Column C: Male = 1;
Non-smoker = 0; Female = 0

	A	B	C
6	Height	Smokes	Male
7	61	0	0
8	61.75	0	0
9	62	0	0
10	62	0	0
11	63	0	0
12	63	0	0
13	63	0	0
14	64	0	0
15	65	0	0
16	65	0	0
17	65	0	0

Non-smoker
smoker
Female
Male

	A	B	C
	Height	Smokes	Male
68	62	1	0
69	62.75	1	0
70	63	1	0
71	64	1	0
72	65	1	0
73	66	1	1
74	67	1	1
75	67	1	1
76	68	1	1
77	68	1	1
78	68	1	0
79	69	1	1

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 4

1) Nudge Binary Male to eliminate Zero and One

	A	B	C	D	E
1					
2	Predict Gender using Height and Smoker				
3		D7 =IF(C7=1,0.999,0.001)			
4					
5					
6	Height	Smokes	Male	Male1	Odds(M)
7	61	0	0	0.001	

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 5

2) Enter formula for Odds in E7; LN[Odds(p)] in F7

	A	B	C	D	E	F
1						
2	Predict Gender using Height and Smoker					
3		D7 =IF(C7=1,0.999,0.001)				F7 =Ln(E7)
4		E7 =D7/(1-D7)				
5						
6	Height	Smokes	Male	Male1	Odds(M)	Ln(Odds)
7	61	0	0	0.001	0.001	-6.91

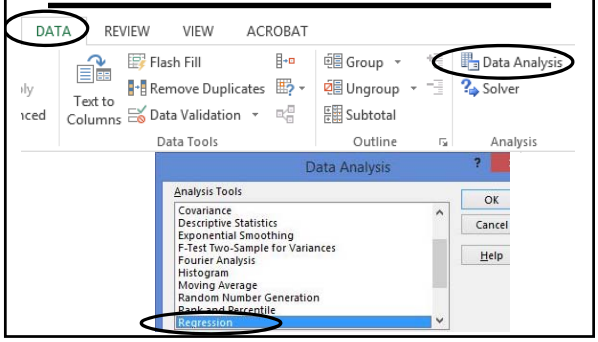
XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 6

3) Select D7:F7 Drag to bottom of data: Row 98

	A	B	C	D	E	F	G
1							
2	Predict Gender using Height and Smoker						
3		D7 =IF(C7=1,0.999,0.001)				F7 =Ln(E7)	
4		E7 =D7/(1-D7)					
5							
6	Height	Smokes	Male	Male1	Odds(M)	Ln(Odds)	yPredict
7	61	0	0	0.001	0.001	-6.91	
8	61.75	0	0				
9	62	0	0				

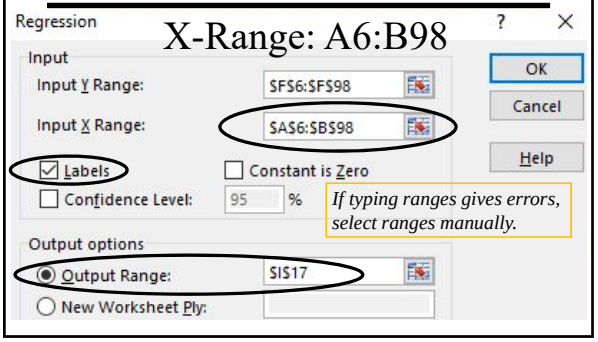
XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 7

A) From Data Bar, Select Data Analysis; Regression



XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 8

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



XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 9

C) Logistic Model: Results Using OLS1

	I	J	K	L	M	N	O P Q
16							
17	SUMMARY OUTPUT						
18							
19	*Regression Statistics*						
20	Multiple R	0.720					
21	R Square	0.518					
22	Adjusted R S	0.507					
23	Standard Err	4.733					
24	Observations	92					
25							
26	ANOVA						
27		*df*	*SS*	*MS*	*F*	*Significance F*	
28	Regression	2	2143.861	1071.930	47.847	0.000	
29	Residual	89	1993.879	22.403			
30	Total	91	4137.739				
31							
32		*Coefficients*	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*wer %r 9.r 93*
33	Intercept	-88.56	9.333	-9.489	0.000	-107.101	## ## ##
34	Height	1.31	0.136	9.624	0.000	1.037	## ## ##
35	Smokes	1.30	1.074	1.210	0.230	-0.835	## ## ##

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

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 10

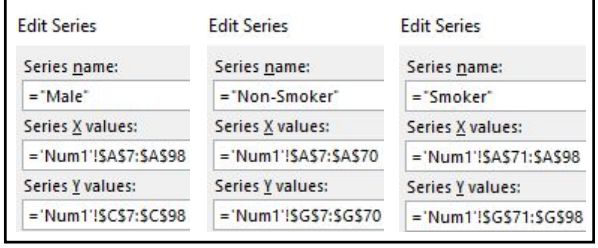
D) Generate G7. Pull G7 down to G98

E	F	G	Row	I
d Smoker				
	F7 =LN(E7)			
	G7 =1/(1+EXP(-J$33-J$34*A7-J$35*B7))			
Odds(M)	Ln(Odds)	yPredict	6	
0.001	-6.91	0.000	7	
0.001	-6.91		8	
0.001	-6.91		9	

XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 11

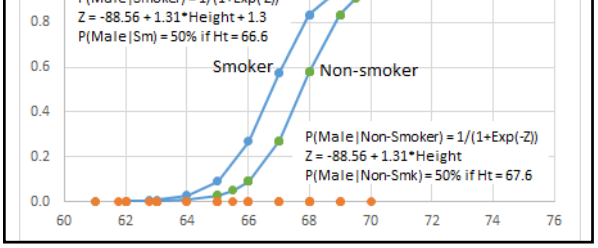
E) Insert Chart (XY Plot): Add Male, Non-smoker, Smoker

A7:A98	A7:A70	A71:A98
C7:C98	G7:G70	G71:G98



XL4B: V0C-2x 2015 Schield Logistic Regression using OLS1B in Excel2013 12

E) Add Title & Text boxes Format Smk/NS with solid lines



Logistic Model using OLS1: Gender vs. Height & Smoker

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/2017-Schield-Logistic-OLS1B-Excel2013-Slides.pdf
pdf/2017-Schield-Logistic-OLS1B-Excel2013-Demo.pdf
Excel/2017-Schield-Logistic-OLS1B-Excel2013-Data.xlsx

Background & Goals

Modelling a binary outcome (loan vs. no-loan) uses a logistic curve/model 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 approach uses a nudge approximation: OLS1. By adjusting the binary outcomes and using a logistic model, OLS regression generates a fairly good fit.

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

Goal: Predict Gender using Height & Smoker

Column B: Smoker = 1; Column C: Male = 1;
Non-smoker = 0; Female = 0

5	A	B	C
6	Height	Smokes	Male
7	61	0	0
8	61.75	0	0
9	62	0	0
10	62	0	0
11	63	0	0
12	63	0	0
13	63	0	0
14	64	0	0
15	65	0	0
16	65	0	0
17	65	0	0

Non-smoker

smoker

Female

Male

	A	B	C
	Height	Smokes	Male
68	62	1	0
69	62.75	1	0
70	63	1	0
71	64	1	0
72	65	1	0
73	66	1	1
74	67	1	1
75	67	1	1
76	68	1	1
77	68	1	1
78	68	1	0
79	69	1	1

1) Nudge Binary Male to eliminate Zero and One

1	A	B	C	D	E
2	Predict Gender using Height and Smoker				
3	D7 =IF(C7=1,0.999,0.001)				
4					
5					
6	Height	Smokes	Male	Male1	Odds(M)
7	61	0	0	0.001	

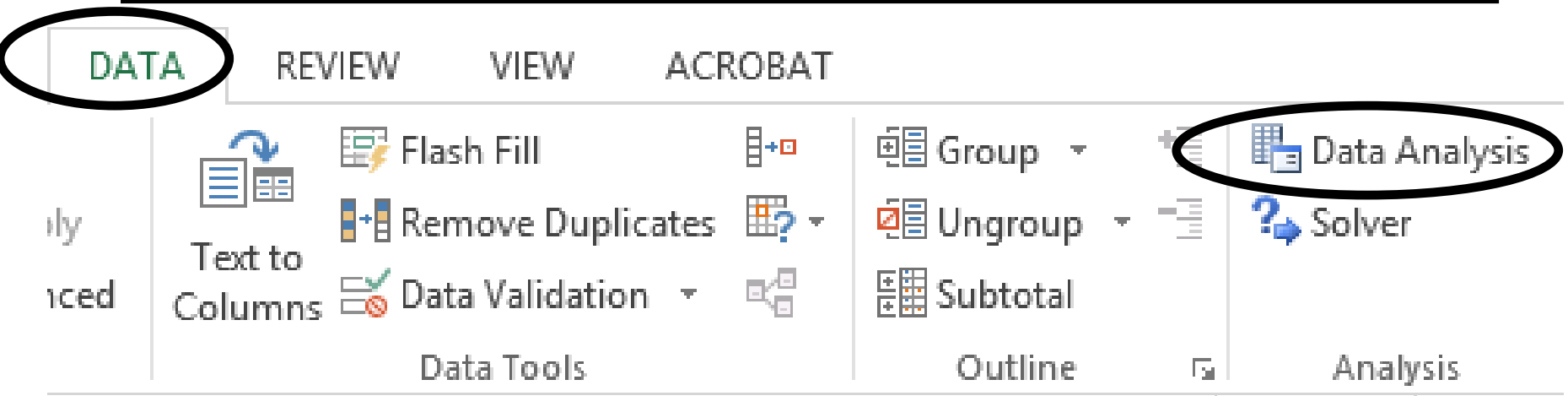
2) Enter formula for Odds in E7; LN[Odds(p)] in F7

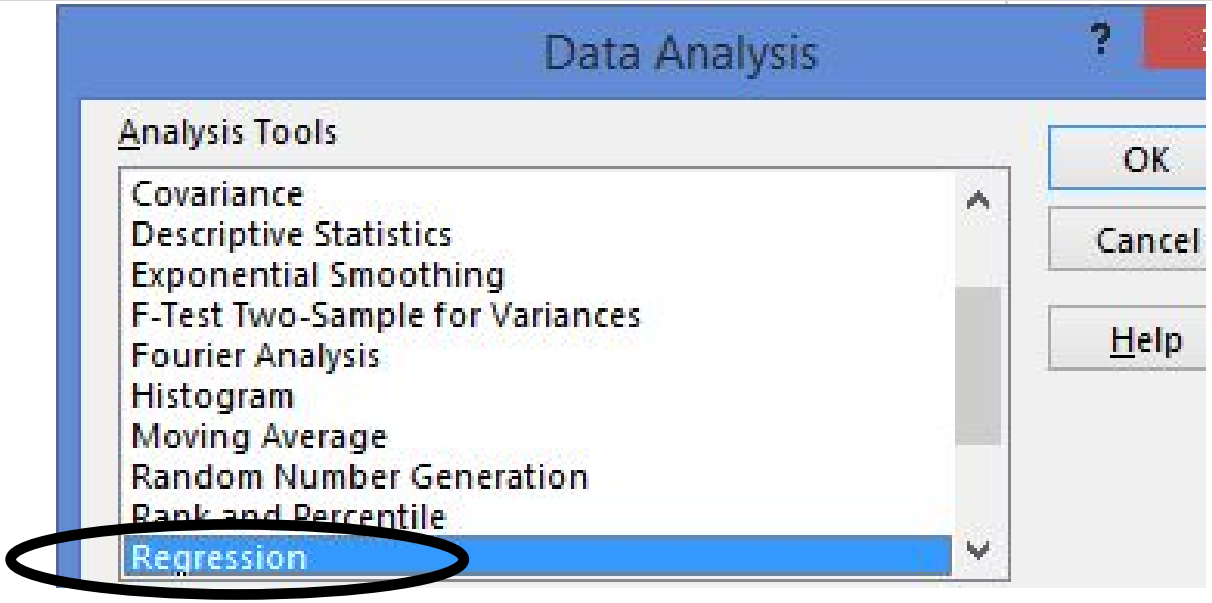
	A	B	C	D	E	F
1						
2	Predict Gender using Height and Smoker					
3	D7 =IF(C7=1,0.999,0.001)				F7 =Ln(E7)	
4	E7 =D7/(1-D7)					')
5						
6	Height	Smokes	Male	Male1	Odds(M)	Ln(Odds)
7	61	0	0	0.001	0.001	-6.91

3) Select D7:F7
Drag to bottom of data: Row 98

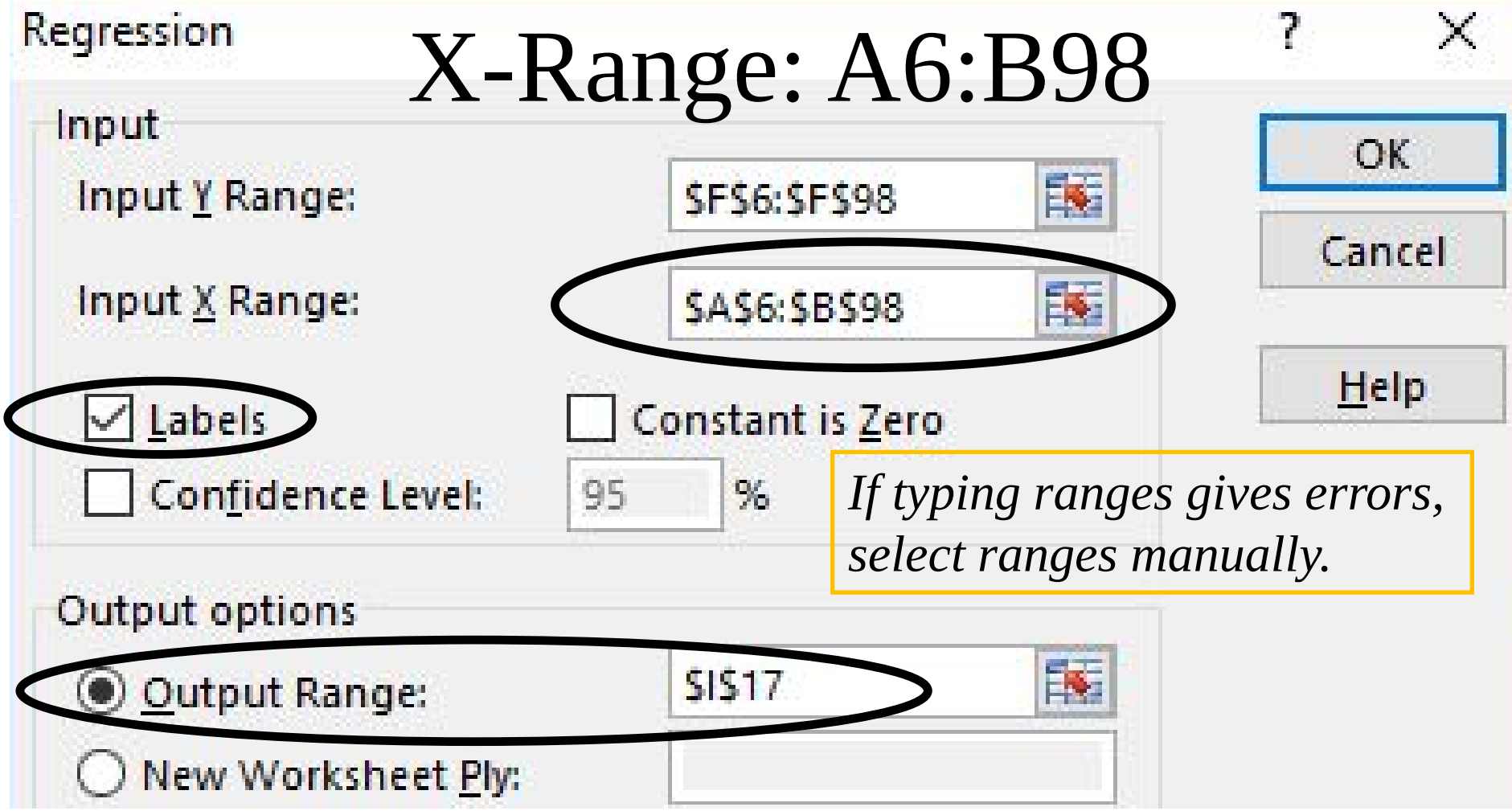
	A	B	C	D	E	F	G
1							
2	Predict Gender using Height and Smoker						
3		D7 =IF(C7=1,0.999,0.001)				F7 =Ln(E7)	
4		E7 =D7/(1-D7)					
5							
6	Height	Smokes	Male	Male1	Odds(M)	Ln(Odds)	yPredict
7	61	0	0	0.001	0.001	-6.91	
8	61.75	0	0				
9	62	0	0				

A) From Data Bar, Select Data Analysis; Regression





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

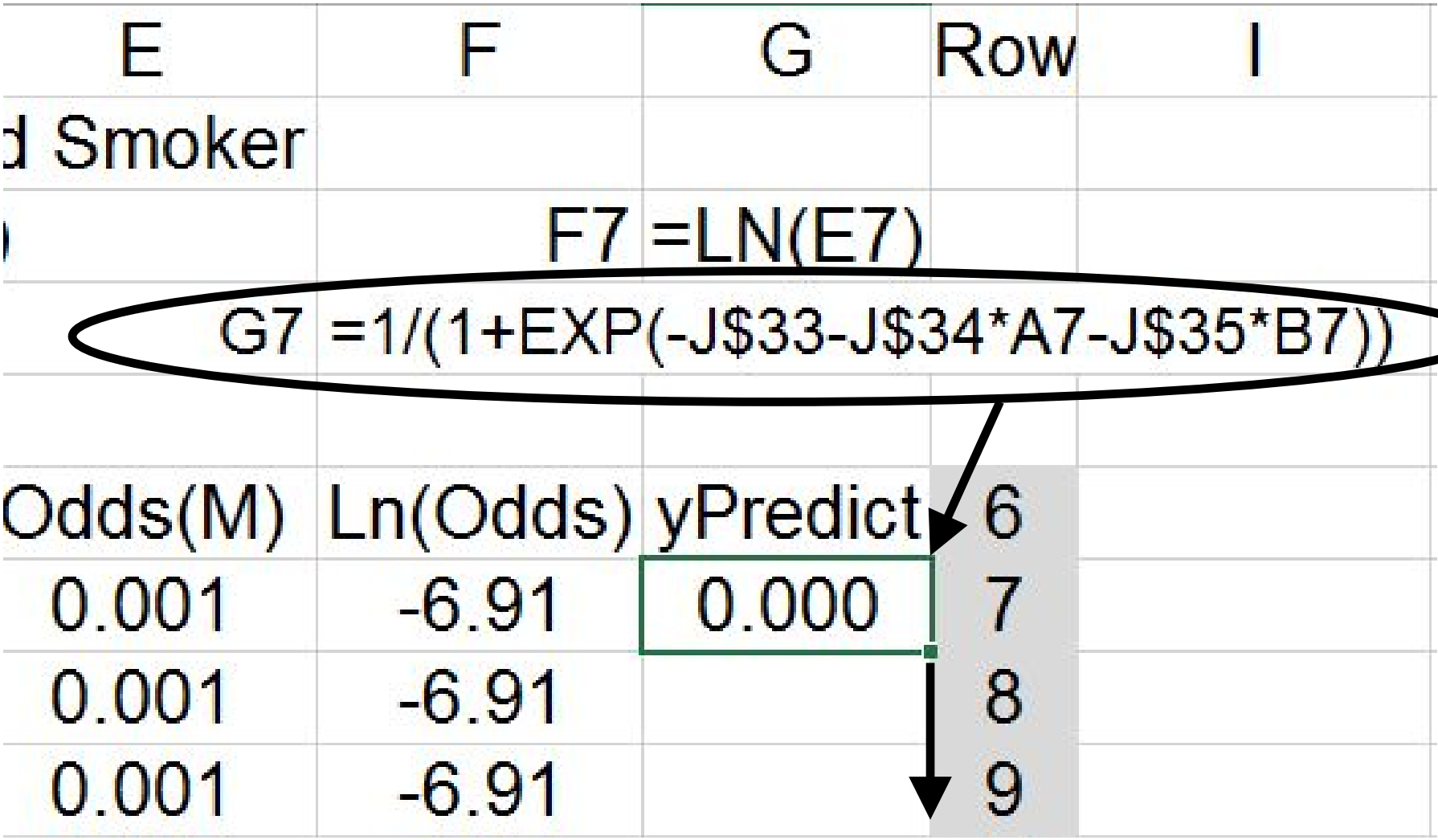


C) Logistic Model: Results Using OLS1

	I	J	K	L	M	N	O	P	Q
16									
17	SUMMARY OUTPUT								
18									
19	*Regression Statistics*								
20	Multiple R	0.720							
21	R Square	0.518							
22	Adjusted R S	0.507							
23	Standard Erro	4.733							
24	Observations	92							
25									
26	ANOVA								
27		*df*	*SS*	*MS*	*F*	*Significance F*			
28	Regression	2	2143.861	1071.930	47.847	0.000			
29	Residual	89	1993.879	22.403					
30	Total	91	4137.739						
31									
32		*Coefficients*	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*ber 9r*	*9ɔr*	*95*
33	Intercept	-88.56	9.333	-9.489	0.000	-107.101	##	##	##
34	Height	1.31	0.136	9.624	0.000	1.037	##	##	##
35	Smokes	1.30	1.074	1.210	0.230	-0.835	##	##	##

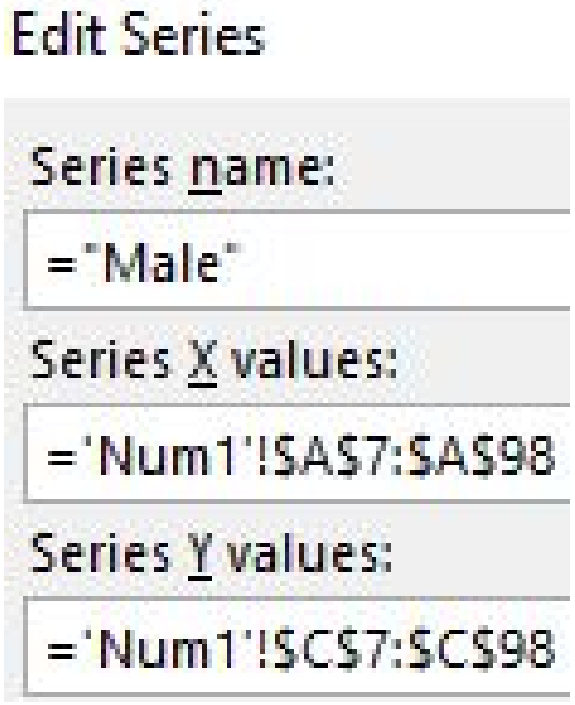
Check to see that you get the same results in the boxes. Formatting is optional

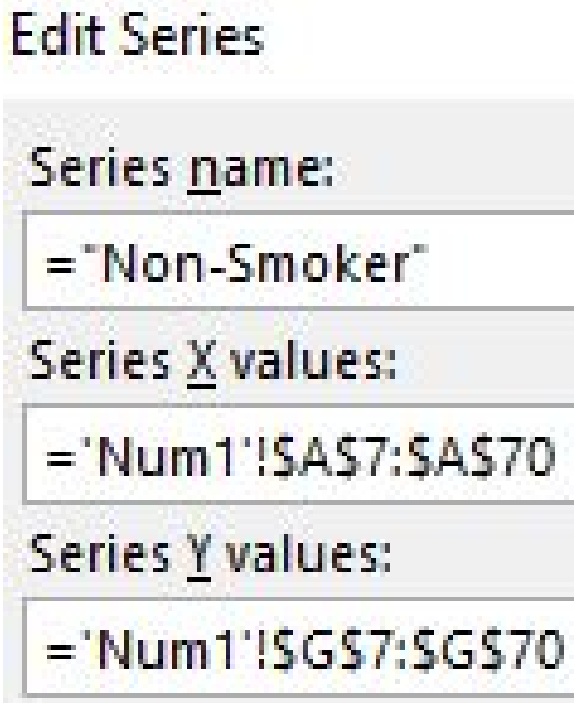
D) Generate G7. Pull G7 down to G98

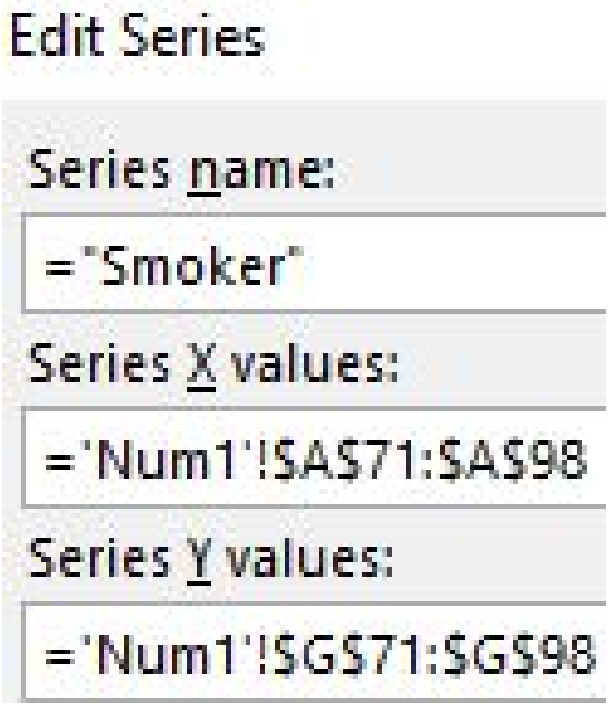


E) Insert Chart (XY Plot): Add Male, Non-smoker, Smoker

A7:A98	A7:A70	A71:A98
C7:C98	G7:G70	G71:G98







E) Add Title & Text boxes Format Smk/NS with solid lines