Logistic Regression using an OLS Shortcut

Height	Male	Linear	Logistic
61	0	-0.12	0.008
61.75	0	-0.04	0.014
62	0	-0.02	0.017
62	0	-0.02	0.017
62	0	-0.02	0.017
62	0	-0.02	0.017
62.75	0	0.05	0.029
63	0	0.07	0.034
63	0	0.07	0.034
63	0	0.07	0.034
63	0	0.07	0.034
64	0	0.17	0.070
64	0	0.17	0.070
65	0	0.27	0.137
65	0	0.27	0.137
65	0	0.27	0.137
65	0	0.27	0.137
65.5	0	0.31	0.188
66	1	0.36	0.251
66	0	0.36	0.251
66	0	0.36	0.251
66	0	0.36	0.251
66	1	0.36	0.251
66	1	0.36	0.251
66	1	0.36	0.251
66	0	0.36	0.251
67	1	0.46	0.414
67	1	0.46	0.414
67	0	0.46	0.414
67	0	0.46	0.414
67	1	0.46	0.414
67	1	0.46	0.414
67	0	0.46	0.414
68	1	0.55	0.599
68	0	0.55	0.599
68	1	0.55	0.599
68	1	0.55	0.599
68	0	0.55	0.599
68	0	0.55	0.599
68	0	0.55	0.599
68	0	0.55	0.599
68	1	0.55	0.599
69	1	0.65	0.759
69	1	0.65	0.759
69	1	0.65	0.759
69	0	0.65	0.759
69	0	0.65	0.759
69	1	0.65	0.759
69	1	0.65	0.759
69	1	0.65	0.759
69	1	0.65	0.759
69	0	0.65	0.759
69.5	1	0.69	0.821

Goal: Model a binary outcome involving a single continuous predictor. V2E

Approach: Use an OLS shortcut that avoids complexity of MLE

For a comparison of OLS with MLE, see www.StatLit.org/pdf/2014-Schield-Logistic-MLE-OLS-Excel2013-Slides.pdf

Step-by-step procedure

1. Enter continuous predictor data in col A, binary outcome in col B.
2. Sort data in first 2 columns in ascending order by Col A
3. Generate appropriate summaries to create linear and logistic.
4. Enter linear in C2 (logistic in D2) and copy C2:D2 to bottom of data
5. Create XY-plot with 3 series. Format and title as needed.
6. Answer three questions

	F	G	H
	3) Summarize data		**Enter formulas from Col H (shown below) into Col G b**
15	Predictor	Height	=A1
16	Mean0	65.4	=AVERAGEIFS(A$2:A$1000,B$2:B$1000,0)
17	Mean1	70.8	=AVERAGEIFS(A$2:A$1000,B$2:B$1000,1)
18	Difference	5.354	=G17-G16
19	**Regress binary Y on continuous X with OLS.**		**Note: X=Xo if P(Y) = 0.5**
20	OLS Intercept	-5.928	=INTERCEPT(B$2:B$1000, A$2:A$1000)
21	OLS Slope	0.0953	=SLOPE(B$2:B$1000, A$2:A$1000)
22	Xo	67.46	=(0.5-G20)/G21 X=Xo if P(Y) = 0.5
23	R-squared	0.51	=CORREL(A2:A1000, B2:B1000)^2
24			
25	If R^2 > 0.4, you can use the following logistic shortcut.		
26	**4) Copy H27 into C2; copy H28 into D2. Remove quote mark from each.**		
27	Linear	C2	=G$20 + G$21*A2 Copy/paste into C2
28	Logistic	D2	=1/(1+exp(4*(G$22-A2)/G$18)) Copy/paste into D2
29			

5) Create XY Plot:

1.2	Ymax
-0.2	Ymin

Three series	Linear	Logistic	Male
Markers	No	No	Yes
Line	Dashed	Solid	No

Based on the logistic, what is the chance that a randomly selected subject is male …

Q1 given that they are 62" tall? _____

Q2 given that they are 65" tall? _____

Q3 given that they are 70" tall?