Model using Trendline Linear 3Factor in Excel 2013

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides at: www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-3Factor-slides.pdf

Goal: Summarize association before/after control for Gender

1. Generate Pivot table (slide 3)
2. Generate two XY charts (slides 4 and 10). Show trend-line (linear model) and R^2 as shown.

Subjects are college students.
Data is at www.statlit.org/Excel/Pulse.xlsx

Slide 4: See www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-Slides.pdf

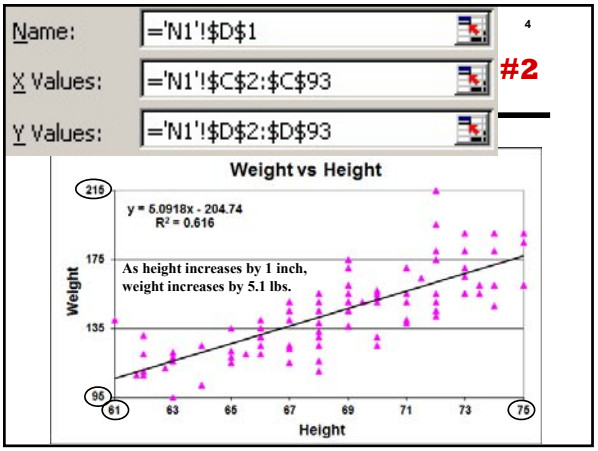
Slide 10: To put 2 series on same chart, see www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf

Generate Summary Statistics Overall and by Gender: #1

Select all data. Insert Pivot Table.
Use Male for column heading. 0=Female; 1 = Male.
Put Height & Weight in body values. Change Sum to Average.
If values spread horizontally, move ΣValues from Col to Row.

	Male?		
Data	0	1	Grand Total
Average of Height	65.4	70.8	68.7
Average of Weight	123.8	158.3	145.2

1. Average male-female weight difference: 34.5 pounds.
2. Average male-female height difference: 5.4 inches.



Analysis

1. Weight is associated with height: R^2 = 0.616
2. Average male-female weight-difference: 34.5#
3. Weight difference due to ave height difference: 5.4 inches times 5.1 pounds per inch = 27.5#.
4. The resulting Sex difference (after controlling for height) is 7 pounds. (34.5 minus 27.5)
5. But weight-height slope is confounded by sex. Solution: Analyze each gender separately.

Need separate weight-height Data for Men and for Women

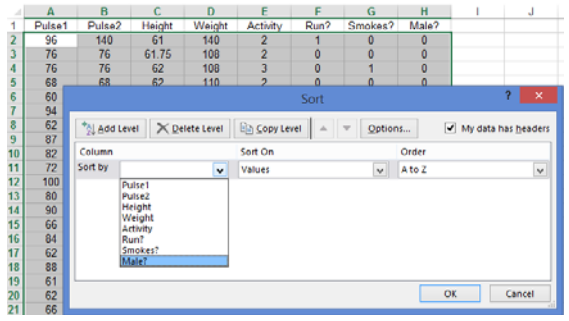
1. **Copy data to new sheet. Rename as N2. Delete pivot table and graph on N2. Copy headings A1:H1 to J1:Q1.**
2. **Change Gal headings**: Cols A, C and D; Add '-F' at end of Pulse1, Height & Weight.

	A	B	C	D	E	F	G	H
1	Pulse1-F	Pulse2	Height-F	Weight-F	Activity	Run?	Smokes?	Male?

3. **Change Guy headings**: Cols J, L and M; Add '–M' at end of Pulse1, Height & Weight.

J	K	L	M	N	O	P	Q
Pulse1-M	Pulse2	Height-M	Weight-M	Activity	Run?	Smokes?	Male?

Select data A1:H93.
Custom Sort by Male: low to high



Need separate weight-height data for Men and for Women

Move guy data (male=1) from A:H to J:Q.

Select A37:H93. Move to J2

	A	B	C	D	E	F	G	H
1	Pulse1-F	Pulse2	Height-F	Weight-F	Activity	Run?	Smokes?	Male?
2	58	56	67	125	2	0	0	0
3	60	66	62	120	2	0	0	0
4	61	70	65.5	120	2	0	0	0
5	62	100	66	120	2	1	0	0

	J	K	L	M	N	O	P	Q
7	Pulse1-M	Pulse2	Height-M	Weight-M	Activity	Run?	Smokes?	Male?
8	48	54	68	150	1	0	1	1
9	54	56	69	145	2	0	1	1
10	54	50	69	160	2	0	0	1
	58	70	72	145	2	1	0	1
	58	58	66	135	3	0	0	1

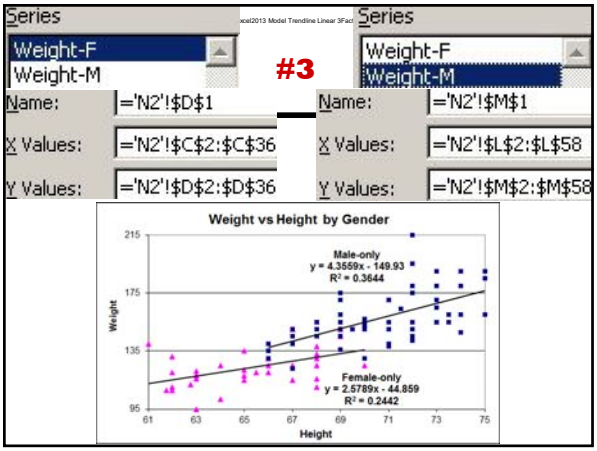
Create weight vs. height graph; Show two-series: guys & gals

Note: Guys are Male=1; Gals are Male=0.

Assignment: Put two series on same graph.
Show the trendline, equation and R^2 for each series.

For detailed instructions, see slides for XL2C at:
> www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-Slides.pdf

Weight: Min = 95#; Max = 215#
Height: Min = 65”; Max = 75”



Analysis

1. If the lines were parallel, the weight difference at any height would be due to the gender difference.
2. If the lines were parallel, the slope would be between 2.6 and 4.4 pounds per inch of height
3. If the weight-height slope was 3.5# per inch, then the weight difference due to the average height difference would be 17.9# [5.1” *3.5#/inch]
4. Given this, the sex difference in weight (after controlling for height) would be 17# [34.5-17.9]

Conclusions

1) We need a better way of modeling – one that gives the same weight-height slope for men and for women. Multivariate linear regression does this automatically.

2) Difference in gender explain part of the association between height and related variables (e.g., weight).

Failure to take into account a relevant confounder can result in associations that are spurious or associations that increase, decrease or reverse.

Moral: What you take into account matters!

Model using Trendline Linear 3Factor in Excel 2013

by
Milo Schield
Member: International Statistical Institute
US Rep: International Statistical Literacy Project
Director, W. M. Keck Statistical Literacy Project

Slides at: www.StatLit.org/pdf/
Excel2013-Model-Trendline-Linear-3Factor-slides.pdf

Goal: Summarize association before/after control for Gender

1. Generate Pivot table (slide 3)
2. Generate two XY charts (slides 4 and 10).
   Show trend-line (linear model) and R^2 as shown.

Subjects are college students.
Data is at www.statlit.org/Excel/Pulse.xlsx

Slide 4: See www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-Slides.pdf

Slide 10: To put 2 series on same chart, see www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-2Y1X-Slides.pdf

Generate Summary Statistics Overall and by Gender: #1

Select all data. Insert Pivot Table.
Use Male for column heading. 0=Female; 1 = Male.
Put Height & Weight in body values. Change Sum to Average.
If values spread horizontally, move ΣValues from Col to Row.

	Male?		
Data	0	1	Grand Total
Average of Height	65.4	70.8	68.7
Average of Weight	123.8	158.3	145.2

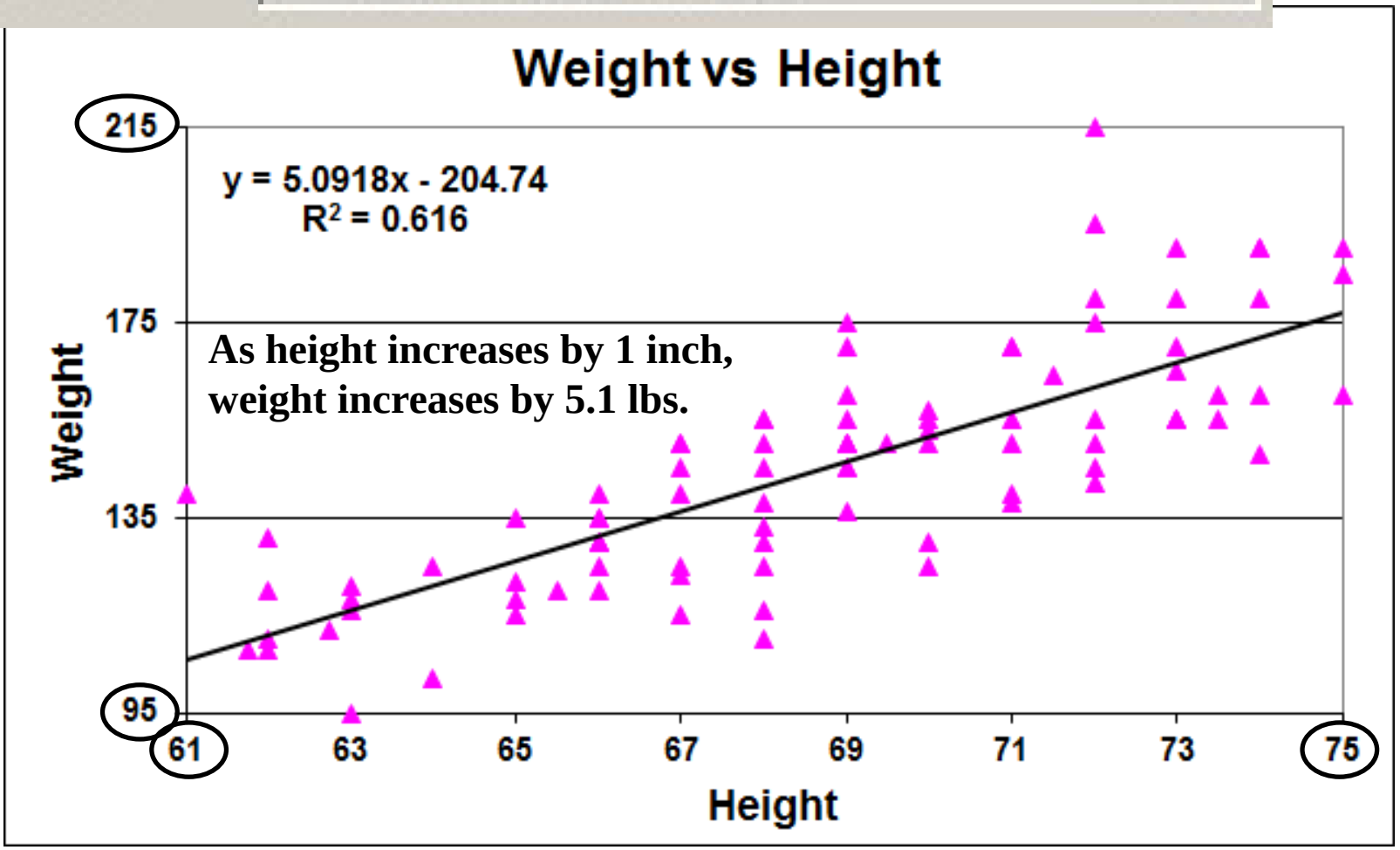
1. Average male-female weight difference: 34.5 pounds.
2. Average male-female height difference: 5.4 inches.

Name: ='N1'!$D$1

X Values: ='N1'!$C$2:$C$93

Y Values: ='N1'!$D$2:$D$93

#2



Analysis

1. Weight is associated with height: R^2 = 0.616
2. Average male-female weight-difference: 34.5#
3. Weight difference due to ave height difference: 5.4 inches times 5.1 pounds per inch = 27.5#.
4. The resulting Sex difference (after controlling for height) is 7 pounds. (34.5 minus 27.5)
5. But weight-height slope is confounded by sex. Solution: Analyze each gender separately.

Need separate weight-height Data for Men and for Women

1. **Copy data to new sheet. Rename as N2. Delete pivot table and graph on N2. Copy headings A1:H1 to J1:Q1.**

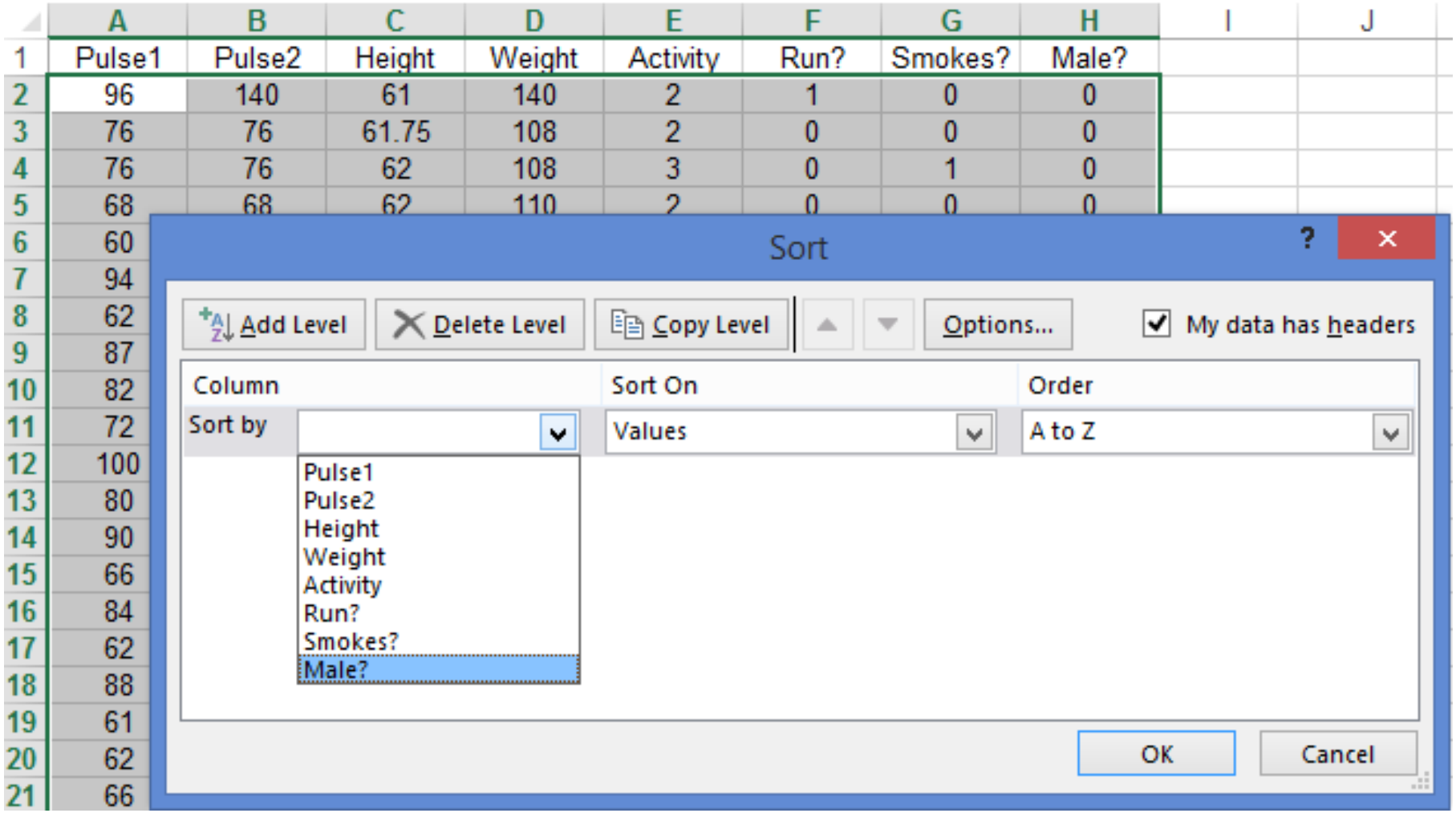
2. **Change Gal headings**: Cols A, C and D; Add '**-F**' at end of Pulse1, Height & Weight.

	A	B	C	D	E	F	G	H
1	Pulse1-F	Pulse2	Height-F	Weight-F	Activity	Run?	Smokes?	Male?

3. **Change Guy headings**: Cols J, L and M; Add '**–M**' at end of Pulse1, Height & Weight.

J	K	L	M	N	O	P	Q
Pulse1-M	Pulse2	Height-M	Weight-M	Activity	Run?	Smokes?	Male?

Select data A1:H93.
Custom Sort by Male: low to high



	A	B	C	D	E	F	G	H
1	Pulse1	Pulse2	Height	Weight	Activity	Run?	Smokes?	Male?
2	96	140	61	140	2	1	0	0
3	76	76	61.75	108	2	0	0	0
4	76	76	62	108	3	0	1	0
5	68	68	62	110	2	0	0	0
6	60							
7	94							
8	62							
9	87							
10	82							
11	72							
12	100							
13	80							
14	90							
15	66							
16	84							
17	62							
18	88							
19	61							
20	62							
21	66							

Need separate weight-height data for Men and for Women

Move guy data (male=1) from A:H to J:Q.

Select A37:H93. Move to J2

	A	B	C	D	E	F	G	H
1	Pulse1-F	Pulse2	Height-F	Weight-F	Activity	Run?	Smokes?	Male?
2	58	56	67	125	2	0	0	0
3	60	66	62	120	2	0	0	0
4	61	70	65.5	120	2	0	0	0
5	62	100	66	120	2	1	0	0
6								

J	K	L	M	N	O	P	Q
7 Pulse1-M	Pulse2	Height-M	Weight-M	Activity	Run?	Smokes?	Male?
8 48	54	68	150	1	0	1	1
9 54	56	69	145	2	0	1	1
10 54	50	69	160	2	0	0	1
58	70	72	145	2	1	0	1
58	58	66	135	3	0	0	1

Create weight vs. height graph; Show two-series: guys & gals

Note: Guys are Male=1; Gals are Male=0.

Assignment: Put two series on same graph.
Show the trendline, equation and R^2 for each series.

For detailed instructions, see slides for XL2C at:
> www.StatLit.org/pdf/Excel2013-Model-Trendline-Linear-Slides.pdf

Weight: Min = 95#; Max = 215#
Height: Min = 65”; Max = 75”

Series

Weight-F

Weight-M

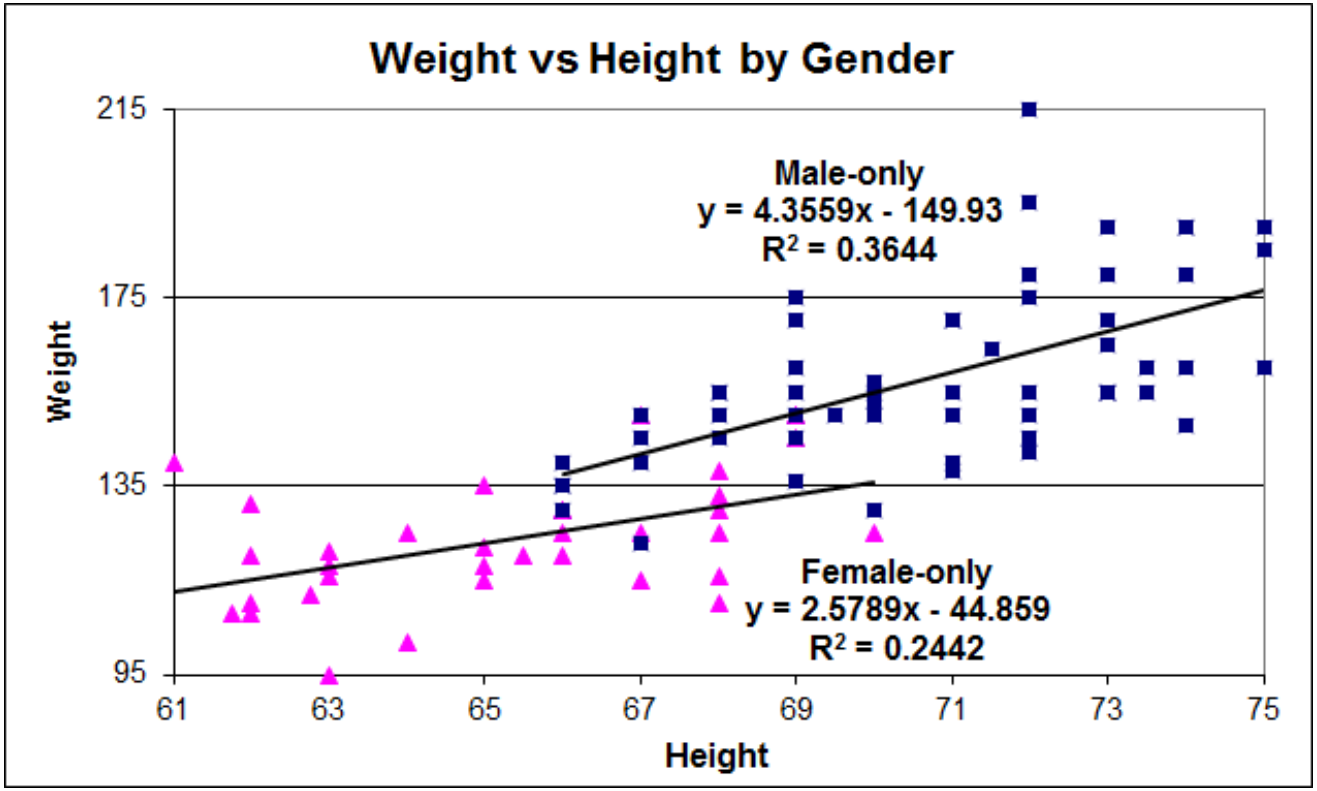
#3

Series

Weight-F

Weight-M

Name:	='N2'!$D$1	Name:	='N2'!$M$1
X Values:	='N2'!$C$2:$C$36	X Values:	='N2'!$L$2:$L$58
Y Values:	='N2'!$D$2:$D$36	Y Values:	='N2'!$M$2:$M$58



Analysis

1. If the lines were parallel, the weight difference at any height would be due to the gender difference.
2. If the lines were parallel, the slope would be between 2.6 and 4.4 pounds per inch of height
3. If the weight-height slope was 3.5# per inch, then the weight difference due to the average height difference would be 17.9# [5.1” *3.5#/inch]
4. Given this, the sex difference in weight (after controlling for height) would be 17# [34.5-17.9]

Conclusions

1) We need a better way of modeling – one that gives the same weight-height slope for men and for women. Multivariate linear regression does this automatically.

2) Difference in gender explain part of the association between height and related variables (e.g., weight).

Failure to take into account a relevant confounder can result in associations that are spurious or associations that increase, decrease or reverse.

Moral: What you take into account matters!