 More Detailed Comments on G&S CHAPTER ONE: Reducing Confounding (Bias) and Increasing Precision (Making Comparisons FAIRER and SHARPER)

FIRST VISIT TO MARS

G&S' s dataset used in chapter 1 is a good example of the need for "synthetic" or "mathematical model" control to compensate for the lack of experimental control.

If a lab researcher were interested in the link between those two factors (height and water consumption) and variations in weight, (s)he would have chosen the 12 subjects more carefully. The researcher would have made sure that the four subjects in each water consumption group were the same height. For example, (s)he might have stipulated the following "grid"for the 12 subjects.

WATER	HEIGHT			
0 cups	30 cm	35 cm	40 cm	45 cm
10 cups	30 cm	35 cm	40 cm	45 cm
20 cups	30 cm	35 cm	40 cm	45 cm

The authors took the cheaper, unstructured, approach to selecting the 12 subjects and ended up with the following values of the "design" variables

WATER	WEIGHT				Average HEIGHT	Average WEIGHT
0 cups	29 cm	32 cm	35 cm	40 cm	34.00 cm	8.45
10 cups	34 cm	38 cm	41 cm	44 cm	39.25 cm	10.95
20 cups	35 cm	40 cm	45 cm	46 cm	41.5 cm	12.80

(The matrix consisting of 12 rows, and 2 columns containing the values for water and height for each of the 12 individuals, is called the "design matrix" for the study)

If one fitted a regression of weight on water consumption, ***without*** taking into account how tall each Martian was, one would obtain the following fitted regression

$$\hat{W} = 8.6 \text{ g} + 0.22 \text{ g/cup } C$$

The various fits yield the following relationships of Weight to ...

HEIGHT (H)	CUPS OF WATER (C)
0.39 g/cm H if **ignore** C	0.22 g/cup C if **ignore** H
0.28 g/cm H if **allow for** C	0.11 g/cup C if **allow for** H

As G&S note at the bottom of p5, "by accounting for the independent effect of drinking, we conclude that weight is less sensitive to height than before [i.e., 0.28 g/cm rather than 0.39 g/cm]".

The same holds true for how sensitive weight is to drinking: 0.11 g/cup if we account for the independent effect of height, but 0.22 g/cup if we ignore height, and simply regress Weight on C alone. [check this for yourself]

Why do the univariate and multivariate answers differ?

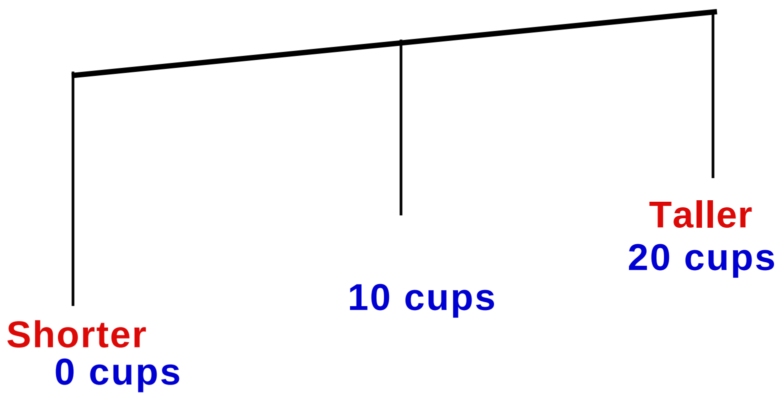
The answer lies in the pattern in the 12 (height, cups) data pairs shown in Fig 1-2B, and in the row by row summary statistics given in the table above. First consider the "crude" or univariate slope of 0.22 g/cup obtained by ignoring height.

The average height in the 20 cup a day group (41.5 cm) is larger than in the 10 cup a day group (39.25 cm) which in turn is larger than the average of 34 cm in the 0 cup a day group. Thus, in the comparison of weights in these three groups, **all other factors are not equal**. i.e., we are not "comparing like with like". Our crude comparison of weights across the water groups is really a **mix** of two comparisons:

1. A comparison of weights of groups consuming different amounts of water [our objective]

2. A comparison of weights of groups who are of different average heights [*not* our objective].

and the answer 0.22 g/cup is therefore a mix of two answers. In epidemiologic (and originally statistical) jargon, the 0.22 g/cup is confounded, i.e., a melange. The effects of the two factors are "mixed up", "distorted", "confounded", "confused", "not properly separated".



[Reminds me of the story about the two children - one Protestant, one Catholic - in Belfast].

It is helpful that the very first data example, even before multiple regression is formally introduced, illustrates this "bias of examining relationships one variable at a time" and that one can, by simply inspecting the data, see where the "upward bias" (line 5 page 6) comes from. If one fully understands this example, one should be able to explain in words -- to a scientist who has never taken a multiple regression course -- what one of the main uses of multiple regression is in biomedicine.

The same bias is operating with height, where the "crude" slope is 0.39 g/cm. The 0.39 g/cm is too high because a naive "univariate" comparison of taller and shorter Martians is also [unfortunately] a comparison of Martians who drink more and Martians who drink less.

Even though G&S do not put the two figures 1-1 and 1-2 side by side, it is instructive to do so and to see the "geometry of confounding". (see next page)

I am a bit surprised that G&S didn't drive home more fully the reason why the 0.39 is "biased *upward*". They simply say that it contains effects that it contains effects that are *"due to both height and drinking water"*.

Persons who have taken a first course in epidemiology often accuse speakers at conferences and rounds that their results are "biased", or "confounded" or "unadjusted for other "factors". But if you ask these "smart-asses" to say whether the bias is upwards or downwards, i.e., the adjustment would shift the estimate downwards or upwards, many of them are at a loss to say which way it goes. If descriptive data are presented, it should be possible to anticipate the magnitude and direction of the bias.

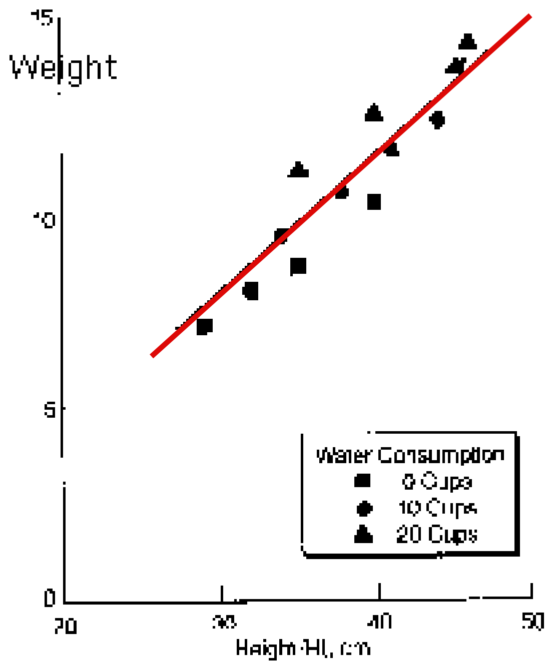
"With the multivariate analysis (i.e., using height and water consumption), the data points cluster more closely around the lines in Fig. 1-2 than the line in Fig. 1-1".

(3rd and 4th last lines in Page 5)

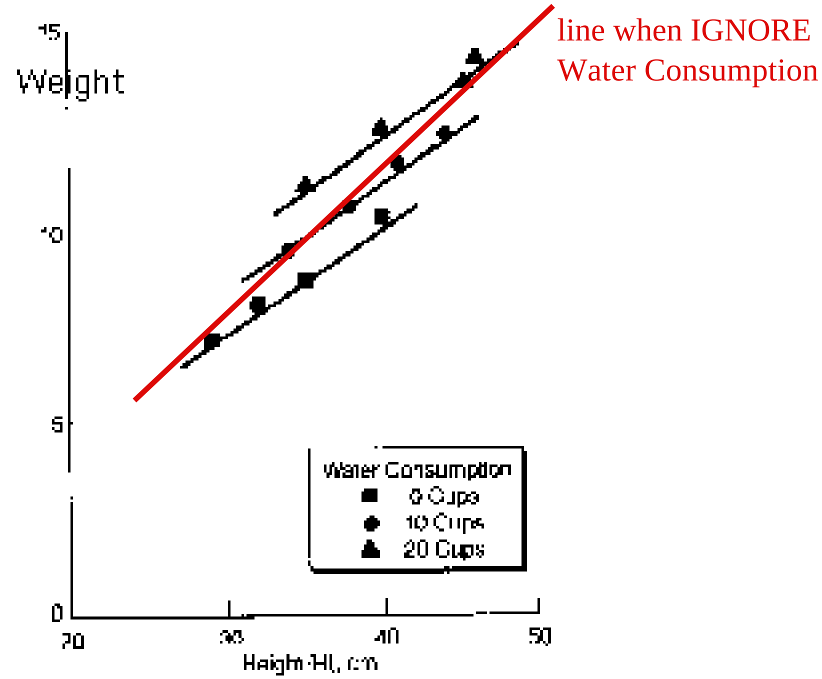
A key point here is that if there is less "residual" variation, then the estimates of the slopes will also be less variable, i.e., more reliable or more stable. We will see this explicitly in the formulae for the standard errors of estimated slopes (b's) in chapters two and three.

Thus, quite apart from "de-biasing" comparisons, another important use of multiple regression -- EVEN FOR DATA COLLECTED IN CONTROLLED EXPERIMENTS -- is to remove unwanted variation in responses caused by KNOWN factors, so that parameter estimates of interest are SHARPER (more precise).

*See the article on "Appropriate Uses of Multivariate Analysis" in the 697 web page for more on "**SHARPER** and FAIRER Comparisons".*



Slope of Weight on Height is **TOO STEEP**



Slope of Weight on Height (for a given level of Water Consumption) is **JUST RIGHT**

DUMMIES (second hand smokers) ON MARS

Coming back to the data in Fig 1-3 [Bias & Precision ..]

What would happen if...

we performed a simple, i.e., univariate t test on the 8 y's from the "clean air" and the 8 y's from the "secondhand smoke" groups?

(1) Would we get the 2.9 g difference in the average weights of the two groups that G&S get in their multivariate analysis? Would the "effect" be bigger or smaller? And

(2), would the difference be as statistically significant (i.e., have as low a p-value as in the multivariate analysis)?

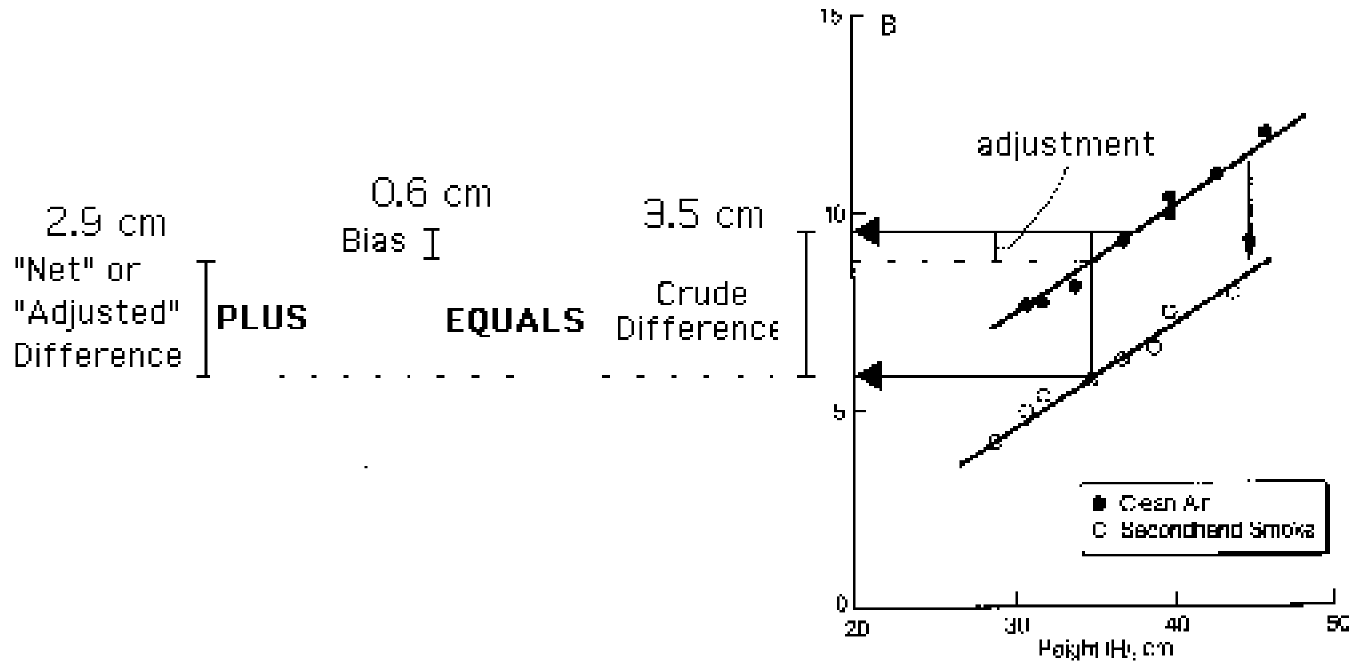
(1) Effect Estimate

The answer to (1) is No. The "crude" difference which ignored height is greater than 2.9. This is because those exposed to second hand smoke were shorter to begin with (unless it was the second hand smoke that stunted their growth!!), and so the crude difference in weight would be the sum of 2.9 g vertical difference (the "like with like" comparison) and another piece arising from the fact that the second hand smoke group are to the left of the others on the height axis. You need to become practiced on this "directional" reasoning concerning biases.

Geometrically, this can be shown as follows (regression estimates made from data extracted from Fig. 1-3).

Use of Multiple Regression for Adjustment

addition to Fig 1-3 p 7 of G&S



The crude difference is approximately 3.5 g, some 2.9 g of which is "due" to the independent effect of smoke, and 0.6 g of which is an "artifact"or "bias". The bias arises from the fact that the "smoke" group are shorter than the "clean air" group (35.9 cm vs 38.1 cm) by approximately 2.2 cm. From equation 1.3, each 1 cm difference in height translates to a 0.27 g difference in weight. Thus, the "imbalance" of 2.2 cm translates to approximately 2.2 cm × .27 g/cm = 0.69 g, which cross-checks with the 3.5 g = 2.9 g + 0.6 g above.

(2) What about the t statistics and p-values?

Here is a comparison of the estimated difference, along with the corresponding standard error (SE), t statistic and p value when the comparison is made "univariate-ly" and "multivariate-ly".

Analysis With. Height...	Estimate of Weight Difference (g)	SE	T-Statistic	P-Value
Ignored	-3.5	0.7	-4.8	0.0003
Accounted For	-2.9	0.2	-14.9	<0.0001

Thus, even with the "assist" from using a "larger than it should be" estimate (-3.5) in the numerator of the t-statistic, the t-statistic from this univariate (crude) analysis is still less extreme (t= -4.8) than that from the multivariate analysis which uses a smaller (adjusted) numerator (-2.9). The reason stems from the much smaller Standard Error (SE) in the multivariate analysis. Within each group, there is a large amount of height-explained variation in weight, which can be removed by INCLUDING height in the regression model. Since the standard errors are directly proportional to the residual variation, the removal of serious amounts of extraneous variation can "sharpen" comparisons to a considerable extent.

Seeing multivariate analyses as "t-tests carried out on adjusted data".

As I hinted at in the very end of the article on "Appropriate Uses of Multivariate Analysis", one can view the coefficients of interest (and their associated standard errors and p-values) from a multivariate analysis if they came from comparisons of means (carried out by a simple t-test) on the "*adjusted*" Y's.

To see this, consider equation 1-3 once more

$$\hat{W} = -0.8 \text{ g} + 0.27 \text{ g/cm } H - 2.9 \text{ g } D$$

In a simple t-test, one is presumably comparing "like with like", or in this case Martians who are all the same height or have the same distribution of height. But with the regression equation, we can do this "virtually", by adding to or subtracting from each Martian's weight an individualized correction factor. This individualized factor reflects (a) how much below or above average that individual is and (b) -- all other relevant factors being equal -- how much heavier are persons who are 1 cm taller (than their counterparts).

If we apply this "homogenizing" or "leveling the playing field" procedure to the data in Fig. 1-3, we get the following adjusted weight values (adjusted to a common height of 37 cm)

	CLEAN AIR GROUP				SECOND HAND SMOKE GROUP			
Subject	H	ΔH	Actual W	Corrected W	H	ΔH	Actual W	Corrected W
1	31	-6	7.6	**9.3**	29	-8	4.1	**6.3**
2	32	-5	7.7	**9.1**	31	-6	4.9	**6.5**
3	34	-3	8.1	**8.8**	32	-5	5.3	**6.7**
4	37	0	9.2	**9.5**	35	-2	5.7	**6.2**
5	41	+4	10.3	**8.8**	37	0	6.2	**6.2**
6	41	+4	9.9	**9.2**	39	+2	6.5	**6.0**
7	43	+6	10.9	**9.8**	40	+3	7.4	**6.6**
8	46	+9	11.9	**9.8**	44	+7	7.9	**6.0**
Average	38		9.5	**9.2**	36		6.0	**6.3**
SD	*5.3*		*1.6*	***0.20***	*5.1*		*1.3*	***0.26***
Crude Difference							-3.5	
SE (difference)							0.72	
Adjusted Difference								**-2.9**
SE (difference)								**0.12**

It is not essential that one adjusts everyone's weight to a common middle height of 37 cm. A simpler approach is to adjust the weights in one group to the average height in the other group. Since we are using a *common* exchange rate across all heights, the result is the same, no matter what value of height we adjust them to.

This is the same process as is described in Fig. 3 of the "Appropriate Uses"article.

One last way of viewing the -2.9 g.

One could also carry out the analysis by pairing up the subjects on height, and performing a paired t-test. Where a close pair-match is not possible, one could use a correction factor to make a pair comparable. One could think of the results of the multivariate analysis as arising from a pairing that uses a common weight-height exchange rate across all heights.