■ RICK HESSE, Feature Editor, Pepperdine University

Simulating Heights and Weights using Regression

by Rick Hesse, Feature Editor, Pepperdine University

For years I've given a simple simulation homework exercise to simulate the height and weight of male students to determine how many out of 200 would be eligible to fit inside Mercer Engineering School's biannual Sunrayce vehicle, a solar-powered car. It wasn't until a year or so ago that it dawned on me that I was treating the weight as independent of the student's height and so the results turned out to be much lower than they should be.

Independent Scenario

Mercer School of Engineering needed a driver for the Sunrayce solar car competition, and the requirements were that the male student needed to be between 5' 0" and 5' 8" tall, and weigh no more than 138 pounds. The assumptions for the average height, weight and the respective standard deviations are also given in the shaded cells in rows 4 and 8, as shown in Figure 1. The template simulates 200 students.

Row 5 simply computes the average and standard deviation of the 200 heights from **B17:B116**, and the minimum and maximum height simulated. Row 9 does the same with the weights. **B11:B13** takes the average of columns D, E and F to determine the probabililty of this simulation having a student with the correct height, correct weight, and passing both hurdles. **B14** multiplies 200 by **B13** (or could just count the number of 1's in **F17:F116**. In this run, there are six students who qualify (including #1 shown in row 17). **C11:C12** uses the **NORMDIST** function in Excel as follows:

C11: =NORMDIST(E4,B4,C4,1)-NORMDIST(D4,B4,C4,1)

D11: =NORMDIST(E8,B8,C8,1)

	A	B	C	D	E	F	G	H
1	SUNRAYCE DRIVER POOL							
2	Independent							
3		Mean	Std Dev	Lower	Upper			
4	Exp Height	69.00	4.30	60.00	68.00			
5	Sim Height	68.91	4.42	57.84	78.75			
6								
7		Mean	Std Dev	Lower	Upper			
8	Exp Weight	155.10	13.77		138			
9	Sim Weight	153.12	13.60	114.81	191.52			
10								
11	Sim P(Height)	34.00%	38.99%	Exp P(Height)				
12	Sim P(Weight)	12.00%	10.71%	Exp P(Weight)				
13	Prob(OK)	3.00%	4.18%	Exp Prob				
14	Sim #	6.00	8.36	Exp #				
15							RN	RN
16	Student	Height	Weight	Height OK?	Weight OK?	Driver OK?	Height	Weight
17	1	64.14	134.48	1	1	1	12.9%	6.7%
18	2	73.87	161.87	0	0	0	87.1%	68.9%
19	3	67.65	148.68	1	0	0	37.7%	32.0%
20	4	68.66	133.27	0	1	0	46.9%	5.6%
21	5	66.99	142.93	1	0	0	32.0%	18.8%
22	6	68.75	157.89	0	0	0	47.6%	58.0%

Figure 1. Independent model to simulate height and weight.

Rick Hesse
is professor of quantitative methods at Pepperdine University in the Graziadio Graduate School of Business and Management. He received his BS, MS, and DSc at Washington University School of Engineering in applied math and computer science. Dr. Hesse is the author of Managerial Spreadsheet Modeling & Analysis *and* Applied Management Science: A Quick & Dirty Approach *(with Gene Woolsey), articles in numerous journals, and software for personal computers. Rick was the first professor to be awarded the Outstanding Civilian Service Medal by the Department of the Army at West Point in 1982, and was the winner of the Decision Sciences Institute's Innovative Instructional Award in 1981.*

Rick Hesse
Pepperdine University
Graziadio Graduate School of Business and Management
Malibu, CA 90265
rickhesse@aol.com

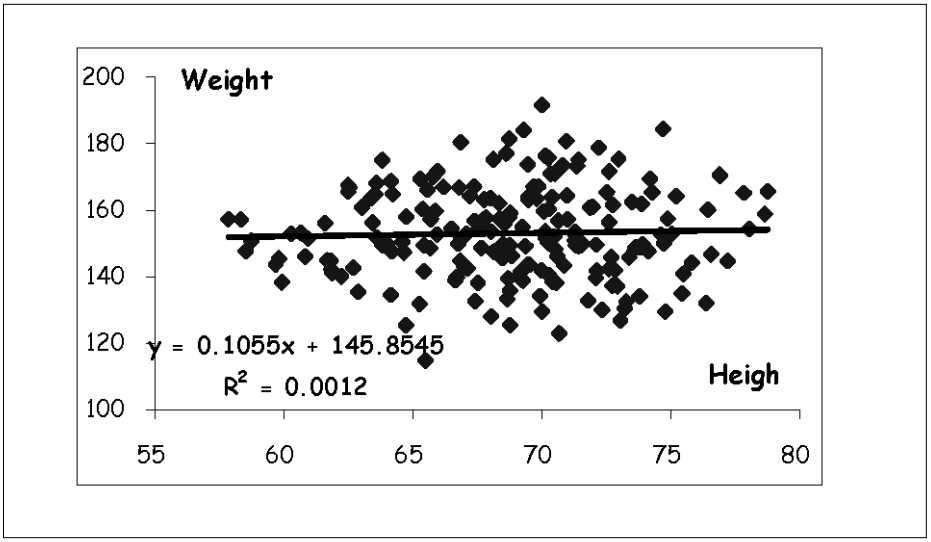

Figure 2. Scatter plot of independent heights and weights.

E11: =C11*D11

F11: =200*E11

Now let us examine the actual simulation of heights, weights and check if the student meets each requirement.

B17: =NORMINV(G17,$B$4,$C$4)

C17: =NORMINV(H17,$B$8,$C$8)

D17: =(B17>=$D$4)*(B17<=$E$4)

E17: =(C17<=$E$8)*1

F17: =D17*E17

G17: =RAND()

H17: =RAND()

The **NORMINV** function uses the appropriate random number as a percentage of the area under the normal curve and returns the simulated height or weight. **D17** and **E17** use Logic expressions (in parentheses) which return either TRUE (1) or FALSE (0). But Excel has a quirk. Unless you use the result of logic expressions in a mathematical formula, the screen will show 0/1 but the actual underlying value in Excel will be FALSE/TRUE, which is a label, and thus is really zero. If **E17** doesn't multiply by 1 (which doesn't change the result), the template will never get any student qualifying.

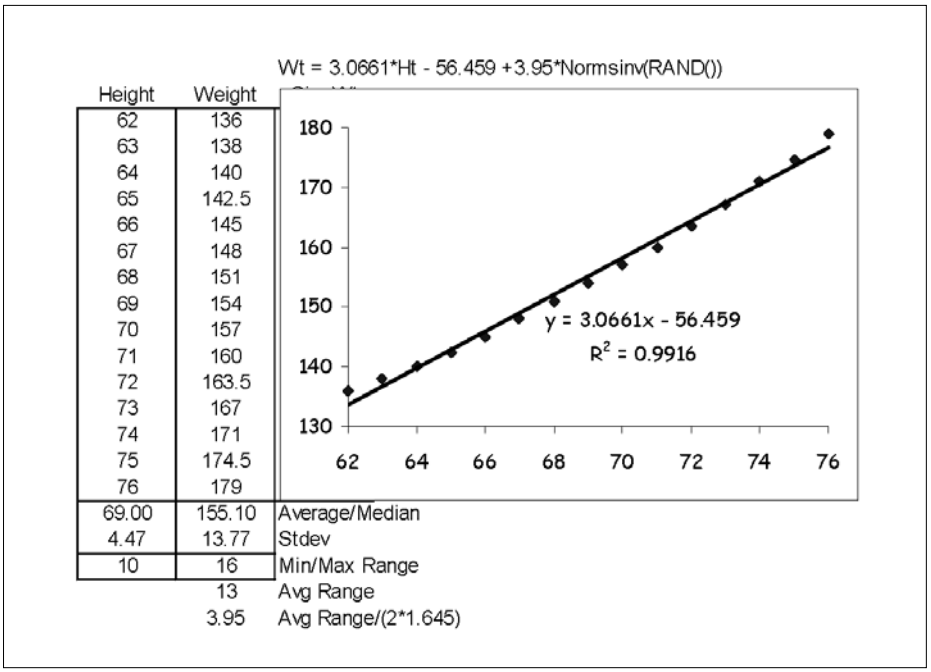

Figure 3. Regression on standard height and mid-range weights.

Graphing the Results

I have used this simulation from time to time over the years, with never a second thought about possible dangers. Then one day, in a flash of the obvious, it occurred to me that these heights and weights would be independent of each other. So I made up a quick XY Scatter plot of **B17:C116** and added the simple trend line with R^2, and my suspicions were confirmed. Figure 2 shows the results.

Even those who are not mathematically inclined know that there is a better chance of a taller person being heavier than a shorter person. But the graph shows basically no correlation (R^2 = 0.12%, r = 3.46%).

Modeling a Dependent Relationship

I "Googled" for "Height and Weight" and found a website for height and associated weight charts for medium frame young males:

www.healthchecksystems.com/heightweightchart.htm.

Figure 3 shows the results for the heights and mid-range for weights. I then did a simple XY scatter plot of the points and right clicked on the data points and selected the formula and r2 value options for the Trend.

The average (and assumed median) height is 5′ 9" (69.0") and weight is 155.10 pounds—the data used in Figure 1. To get an approximate normal distribution for heights and weights, I used the standard deviation derived from the charts, again for the Independent Simulation. For the Dependent

	A	B	C	D	E	F	G	H
1	SUNRAYCE DRIVER POOL							
2	Dependent							
3		Mean	Std Dev	Lower	Upper			
4	Exp Height	69.00	4.30	60.00	68.00			
5	Sim Height	68.91	4.42	57.84	78.75			
6	Wt = 3.0661*Ht - 56.459 +4.17*Normsinv(RAND())							
7		Mean	Std Dev	Lower	Upper			
8	Exp Weight	155.10	13.77		138			
9	Sim Weight	154.21	14.30	120.70	188.20			
10								
11	Sim P(Height)	34.00%	38.99%	Exp P(Height)				
12	Sim P(Weight)	13.00%						
13	Prob(OK)	9.50%						
14	Sim #	19.00						
15							RN	RN
16	Student	Height	Weight	Height OK?	Weight OK?	Driver OK?	Height	Weight
17	1	64.14	133.90	1	1	1	12.9%	6.7%
18	2	73.87	172.10	0	0	0	87.1%	68.9%
19	3	67.65	149.00	1	0	0	37.7%	32.0%
20	4	68.66	147.40	0	0	0	46.9%	5.6%
21	5	66.99	145.30	1	0	0	32.0%	18.8%
22	6	68.75	155.20	0	0	0	47.6%	58.0%

Figure 4. Dependent simulation.

Simulation, I found the average weight range (13.0) and divided by 2*90% z-scores (1.645) to set up the equation shown above the graph. Although the range does increase with height, I have neglected that for now.

Dependent Simulation

Figure 4 shows the Dependent Simulation, with the only change being for the simulated height and the fact that **C12:C14** really can't be computed analytically (or at least easily):

C17:=ROUND(3.0661*B17-56.459+4.17*NORMSINV(H17),1)

G17: =Independent!G17 copied G17:H116

The same random numbers are used as the Independent Simulation (just pointing to that sheet) and now the random numbers in column H are used in the normal standard inverse function for the variation in weight, but the bulk of the weight is due to the height (about 3 pounds per inch).

The simulated heights are identical in both simulations, however the weights are dramatically different, as shown in **C17:C22**. Although the probability of the weight being OK has changed only slightly from 12% to 13%, the ensuing probability of a student being qualified has risen dramatically from 3.0% to 9.5% This is because the weights are now properly associated with the height.

Final Scatter Plot

Figure 5 shows the scatter plot and regression for the Dependent Simulation, with an obvious correlation of weight to height.

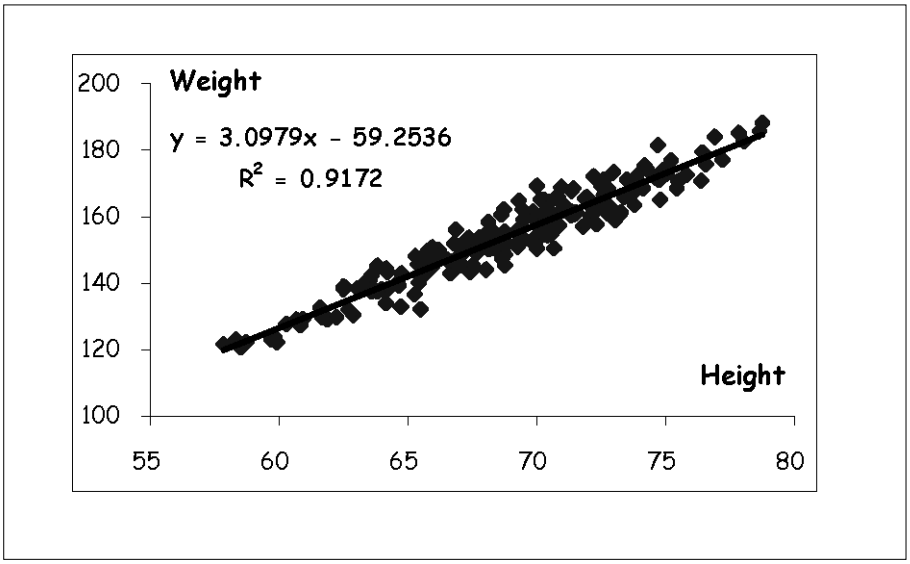

Figure 5. Scatter plot and regression for the dependent simulation.

Further Investigation

This exercise can be developed further if there is a desire to make another worksheet with a data table to collect the results of several hundred runs of both sheets and the simulated number of students who qualify as drivers. Then a statistical test could be used to determine if the Dependent Simulation produces a significantly higher number of qualified candidates. My early observations suggest twice as many. Or Crystal Ball or @Risk could run the simulations and gather the statistics.

Of course the multiplier for the random variation of weights could use some work so that the range is narrow for shorter students and then increases with an increase in height. This workbook is available on the website with the paper with the random numbers live on the Independent worksheet. Therefore the results will be similar but different than what is shown in the column.

This exercise serves two purposes: a simple Excel-based simulation and reinforcement of simple linear regression. ■